

N° Ordre...../FT/UMBB/2026

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA  
RECHERCHE SCIENTIFIQUE

UNIVERSITE M'HAMED BOUGARA-BOUMERDES



**Faculté de Technologie**

**Thèse de Doctorat**

Présentée par :

**Hafida BOUMERIDJA**

En vue de l'obtention du diplôme de **DOCTORAT Sciences en :**

**Filière : Génie Electrique**

**Spécialité : Electronique**

**Développement d'un système d'aide au diagnostic  
intelligent pour la gynécologie**

*Devant le jury composé de :*

| Nom et Prénoms            | Grade      | Etablissement      | Qualité          |
|---------------------------|------------|--------------------|------------------|
| HARRAR Khaled             | Professeur | UMB Boumerdes      | <b>Président</b> |
| AMMAR Mohammed            | Professeur | UMB Boumerdes      | <b>Directeur</b> |
| MESSAOUDI Nouredine       | Professeur | UMB Boumerdes      | <b>Examineur</b> |
| BENLATRACHE Mohamed Salah | Professeur | UAB Mila           | <b>Examineur</b> |
| LAZOUNI Mohammed El Amine | MCA        | UABB Tlemcen       | <b>Examineur</b> |
| OMARI Tahar               | MCA        | UBB Ain Témouchent | <b>Examineur</b> |

Année académique : 2025-2026



## DÉDICACES

---

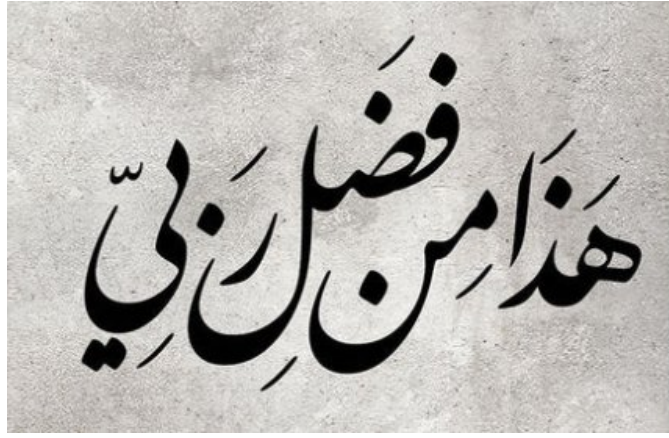
*À la mémoire de mes parents bien-aimés,  
qui rêvaient de ce jour.*

*À mon cher époux, ma force et mon refuge,  
qui a su faire renaître mes rêves  
et illuminer ce chemin de sa présence indéfectible,  
ce travail est autant le tien que le mien.*

*Et à ma fille,  
mon ancre et ma plus grande fierté.*







## REMERCIEMENTS

---

Je fais partie de ceux qui croient fermement qu'il n'est de force ni de puissance que par Dieu. C'est donc par Lui que je commence, en Le remerciant de m'avoir accordé la volonté, la patience et la grâce d'être entourée de personnes remarquables qui ont, chacune à sa façon, contribué à la réalisation de cette thèse de doctorat.

Je tiens tout d'abord à exprimer ma sincère reconnaissance à l'Université M'Hamed Bougara de Boumerdes pour le cadre académique et les ressources mis à ma disposition tout au long de ce parcours doctoral.

Mes remerciements les plus sincères vont au Professeur Ammar Mohammed, mon Directeur de Thèse, pour sa rigueur, sa disponibilité et sa confiance indéfectible — bien plus qu'un encadrement, un véritable élan qui a porté cette thèse jusqu'à son terme.

Je remercie chaleureusement le Professeur Messaoudi Noureddine, Chef du Département Ingénierie des Systèmes Électriques, pour son soutien bienveillant et les conditions favorables qu'il a offertes à la réalisation de ce travail. Ma reconnaissance va également au Professeur Akroum Hamza, Directeur du Laboratoire d'Ingénierie des Systèmes et Télécommunications, pour le cadre de recherche stimulant et la dynamique scientifique qui ont nourri chaque étape de ce parcours.

Je remercie sincèrement les membres du jury de cette thèse : le Professeur Harrar Khaled pour avoir accepté d'en assurer la présidence, ainsi que le Professeur Benlatrache Mohamed Salah de l'Université de Mila, le Docteur Lazouni Mohammed El Amine de l'Université de

---

Tlemcen et le Docteur Omari Tahar de l'Université de Aïn Témouchent, pour le temps précieux consacré à l'évaluation de ce travail et la qualité de leurs observations.

Je tiens également à remercier chaleureusement le Professeur Meraihi Yacine, pour ses encouragements à m'inscrire en doctorat et sa bienveillance constante qui m'a portée jusqu'au bout.

Ma profonde gratitude va aux Docteurs Benouada, Benamer, Bouzar, Chebaani et Lardjani pour leur précieuse contribution à l'évaluation du volet clinique de cette thèse, leur expertise et leur générosité intellectuelle ayant considérablement enrichi ce travail. Une mention toute particulière est adressée au Docteur Hassna Chebaani, ma belle-sœur, qui n'a jamais compté son temps et a répondu présente à chacune de mes sollicitations avec un dévouement sans faille, et au Docteur Karima Benouada, dont la bienveillance et l'amour sincère de la science ont été parmi les plus gratifiants de ce parcours.

À toute ma famille et à ma belle-famille, deux familles réunies par les liens du cœur autant que par ceux du sang, dont la chaleur, la patience et les encouragements sans faille ont été autant de lumières sur un chemin parfois semé d'embûches — et tout particulièrement à ma sœur Karima, dont la présence bienveillante et le soutien indéfectible ont été un ancrage silencieux qui a porté chacune des étapes de cette thèse.

À mes chères amies, piliers silencieux de ce parcours, dont les mots d'encouragement et la présence aux moments de doute comme aux moments de joie ont été d'une valeur inestimable — et tout particulièrement à Samia O., dont la générosité et le dévouement ont dépassé l'amitié pour devenir une aide concrète et précieuse jusqu'aux derniers instants de ce travail.

Je remercie enfin tous ceux qui, de près ou de loin, ont contribué à l'aboutissement de ce travail — mes collègues du département, les techniciens, ingénieurs et personnel administratif, et toute personne dont la présence, le geste ou le mot, même fugace, a semé sur ce chemin un peu de lumière et de sérénité.

---

*À tous ceux dont le nom n'apparaît pas ici mais dont le cœur a accompagné ce travail — recevez, en ces quelques lignes, l'expression de ma reconnaissance la plus sincère.*



**ملخص**

تتجلى اليوم الفجوة في التشخيص قبل الولادة بشكل صارخ نتيجة التباين التكنولوجي بين المؤسسات الصحية، لا سيما في المناطق ذات الموارد المحدودة. إذ يؤدي الاعتماد على أجهزة سونار (إيكو) متهاكة إلى إضعاف القدرة على رصد التشوهات الجنينية، مما يضع الطبيب أمام تحدي تفسير صور مشوشة تفتقر إلى الموثوقية المطلوبة. ونظراً لأن عوائق الميزانية تحول دون تحديث البنية التحتية، يصبح استكشاف حلول الذكاء الاصطناعي المسار الأنسب لترسيخ العدالة في الرعاية الصحية.

انطلاقاً من هذه التحديات الاقتصادية والاجتماعية، يبحث هذا العمل في مدى فاعلية تقنية "الدقة الفائقة" (Super-Resolution) المدعومة بالذكاء الاصطناعي في تحسين جودة صور السونار الضعيفة. تهدف هذه المقاربة إلى "إثراء" البيانات البصرية المتاحة، مما يمنح المعدات القديمة عمراً جديداً ويضمن للمريضات فرصاً متكافئة في الحصول على تشخيص دقيق.

لقد أخضعنا عدة نماذج من "التعلم العميق" (Deep Learning) للاختبار لقياس مدى مرونتها وجدواها التشخيصية. وتعتمد منهجيتنا في جوهرها على تقييم مزدوج؛ يجمع بين الاختبارات الإحصائية والخبرة السريرية المعمقة. وقد أثبتت النتائج قدرة هذه التقنيات على تعزيز اليقين التشخيصي، مما يفتح آفاقاً لرفع مستوى الأداء الطبي دون أن يظل ذلك رهناً بمدى حداثة الأجهزة المتوفرة.

**Résumé**

Aujourd'hui, l'inégalité face au diagnostic prénatal est directement exacerbée par l'écart technologique qui sépare les structures de soins dans les milieux à ressources limitées. L'utilisation d'échographes obsolètes limite la détection d'anomalies fœtales, obligeant le clinicien à interpréter des images trop dégradées pour être fiables. Le remplacement des infrastructures étant fortement entravé par des contraintes budgétaires persistantes, l'exploration de solutions d'amélioration par l'intelligence artificielle représente la voie indispensable pour rétablir l'équité des soins. Face aux enjeux socio-économiques et au coût des équipements, cette recherche explore l'efficacité de la Super-Résolution par IA pour améliorer les échographies fœtales dégradées en milieux limités. Cette démarche vise à "enrichir" l'information visuelle existante, offrant ainsi une seconde vie au matériel obsolète et une meilleure équité de soins aux patientes.

Nous avons mis à l'épreuve plusieurs architectures de "Deep Learning" afin d'en mesurer la robustesse et la pertinence diagnostique. Le cœur de la méthodologie réside dans une évaluation croisée : aux tests statistiques s'est ajoutée une expertise clinique approfondie. Les résultats valident la capacité de ces méthodes à augmenter la certitude diagnostique, offrant ainsi une perspective de performance équitable sans dépendre de la modernité du matériel.

**Abstract**

Today, inequality in prenatal diagnosis is directly exacerbated by the technological gap separating healthcare facilities in resource-limited settings. The persistent use of obsolete ultrasound machines constrains the detection of fetal anomalies, forcing the clinician to interpret images too degraded for reliable assessment. As infrastructure replacement is strongly impeded by persistent budgetary constraints, the exploration of AI-driven enhancement solutions represents the essential pathway to restore equity of care. Given these socio-economic challenges and the equipment costs, this research investigates the efficacy of AI-based Super-Resolution (SR) in improving degraded fetal ultrasound scans within constrained environments. This approach aims to "enrich" existing visual information, thus providing obsolete equipment with a second life and ensuring better equity of care for patients.

We evaluated multiple Deep Learning architectures to measure their robustness and diagnostic relevance. The core of the methodology resided in a cross-sectional evaluation: statistical tests were complemented by in-depth clinical expertise. The results validate the methods' capacity to augment diagnostic certainty, thereby offering a perspective of equitable performance regardless of equipment modernity.

# Table des matières

|                                                                                                              |           |
|--------------------------------------------------------------------------------------------------------------|-----------|
| Dédicaces                                                                                                    | i         |
| Remerciements                                                                                                | ii        |
| Résumés                                                                                                      | iv        |
| Liste des figures                                                                                            | xi        |
| Liste des tableaux                                                                                           | xiv       |
| Liste des abréviations                                                                                       | xv        |
| Liste des abréviations                                                                                       | xv        |
| Introduction Générale                                                                                        | 2         |
| <b>1 État de l'art de la Super-Résolution d'Images : des Approches Classiques aux Applications Médicales</b> | <b>6</b>  |
| 1.1 Introduction . . . . .                                                                                   | 6         |
| 1.2 Fondements théoriques et méthodologiques de la super-résolution . . . . .                                | 7         |
| 1.3 Évolution chronologique des paradigmes en super-résolution . . . . .                                     | 10        |
| 1.3.1 Bilan des paradigmes SR . . . . .                                                                      | 18        |
| 1.4 Analyse Systématique de la Littérature Médicale . . . . .                                                | 19        |
| 1.4.1 Perspectives synthétiques des revues de littérature . . . . .                                          | 23        |
| 1.4.2 Convergences méthodologiques et implications pour l'imagerie ultrasonore . . . . .                     | 23        |
| 1.5 Transfert d'apprentissage et fine-tuning en imagerie médicale . . . . .                                  | 25        |
| 1.5.1 Application du transfert d'apprentissage à la super-résolution . . . . .                               | 27        |
| 1.6 Conclusion . . . . .                                                                                     | 29        |
| <b>2 Ultrasons et imagerie diagnostique : fondements, intérêts et restrictions</b>                           | <b>34</b> |
| 2.1 Introduction . . . . .                                                                                   | 34        |
| 2.2 Fondements . . . . .                                                                                     | 35        |
| 2.2.1 Principes de l'échographie . . . . .                                                                   | 35        |
| 2.2.2 Bases physiques et principes de propagation des ultrasons . . . . .                                    | 36        |
| 2.2.2.1 Nature des ondes ultrasonores et domaine fréquentiel . . . . .                                       | 36        |
| 2.2.2.2 Propagation des ultrasons dans les tissus biologiques . . . . .                                      | 37        |
| 2.2.3 Génération et détection des ultrasons . . . . .                                                        | 37        |
| 2.2.3.1 Modes d'émission et d'imagerie . . . . .                                                             | 38        |
| 2.2.4 Définition et caractéristiques de l'image ultrasonore . . . . .                                        | 41        |
| 2.2.4.1 Définition . . . . .                                                                                 | 41        |
| 2.2.4.2 Caractéristiques principales . . . . .                                                               | 41        |

|          |                                                                                                                                     |           |
|----------|-------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.2.4.3  | Bruit de Speckle . . . . .                                                                                                          | 43        |
| 2.3      | Intérêts de l'imagerie ultrasonore . . . . .                                                                                        | 44        |
| 2.3.1    | Caractère non invasif et innocuité . . . . .                                                                                        | 44        |
| 2.3.2    | Accessibilité et faible coût . . . . .                                                                                              | 44        |
| 2.3.3    | Imagerie en temps réel et guidage interventionnel . . . . .                                                                         | 45        |
| 2.3.4    | Résolution des structures superficielles . . . . .                                                                                  | 45        |
| 2.4      | Restrictions de l'imagerie ultrasonore . . . . .                                                                                    | 46        |
| 2.4.1    | Dépendance opérateur . . . . .                                                                                                      | 46        |
| 2.4.2    | Atténuation et pénétration limitée . . . . .                                                                                        | 47        |
| 2.4.3    | Artefacts et bruit de speckle . . . . .                                                                                             | 48        |
| 2.4.3.1  | Le bruit de speckle . . . . .                                                                                                       | 48        |
| 2.4.3.2  | Les artefacts principaux . . . . .                                                                                                  | 49        |
| 2.4.4    | Champ de vision restreint . . . . .                                                                                                 | 50        |
| 2.4.4.1  | Origines techniques . . . . .                                                                                                       | 50        |
| 2.4.4.2  | Conséquences en pratique clinique . . . . .                                                                                         | 50        |
| 2.4.5    | Limites diagnostiques pour certaines pathologies . . . . .                                                                          | 51        |
| 2.4.5.1  | Limitations liées à la résolution et à la pénétration . . . . .                                                                     | 51        |
| 2.4.5.2  | Limitations liées aux artefacts et au contraste . . . . .                                                                           | 51        |
| 2.4.5.3  | Limitations liées au champ de vision et à l'opérateur . . . . .                                                                     | 52        |
| 2.5      | Conclusion . . . . .                                                                                                                | 52        |
| <b>3</b> | <b>Architectures de super-résolution : CNN, GAN et Transformers</b>                                                                 | <b>54</b> |
| 3.1      | Introduction . . . . .                                                                                                              | 54        |
| 3.2      | Enjeux et contraintes de la super-résolution en imagerie ultrasonore . . . . .                                                      | 55        |
| 3.2.1    | La nature spécifique du Speckle . . . . .                                                                                           | 55        |
| 3.2.2    | Anisotropie de la réponse impulsionnelle . . . . .                                                                                  | 55        |
| 3.2.3    | Fidélité structurelle et risque d'hallucination . . . . .                                                                           | 55        |
| 3.3      | Modèle <i>Dual Back-Projection-based Internal Learning</i> : L'approche par appren-<br>tissage interne . . . . .                    | 56        |
| 3.3.1    | Architecture . . . . .                                                                                                              | 56        |
| 3.3.2    | Apprentissage Interne . . . . .                                                                                                     | 57        |
| 3.3.3    | Originalité méthodologique . . . . .                                                                                                | 57        |
| 3.4      | Synthèse comparative des modèles SR internes (CNN) . . . . .                                                                        | 57        |
| 3.5      | Modèles de super-résolution à base de GAN . . . . .                                                                                 | 59        |
| 3.5.1    | A-ESRGAN . . . . .                                                                                                                  | 59        |
| 3.5.2    | BSRGAN . . . . .                                                                                                                    | 61        |
| 3.5.3    | Real-ESRGAN . . . . .                                                                                                               | 64        |
| 3.6      | Critères de comparaison des générateurs GAN : A-ESRGAN, BSRGAN, Real-<br>ESRGAN . . . . .                                           | 66        |
| 3.7      | Modèles de super-résolution à base de Transformers . . . . .                                                                        | 67        |
| 3.7.1    | SwinIR . . . . .                                                                                                                    | 68        |
| 3.7.2    | Swin2SR . . . . .                                                                                                                   | 70        |
| 3.7.3    | HAT . . . . .                                                                                                                       | 71        |
| 3.8      | Synthèse comparative des modèles SR . . . . .                                                                                       | 74        |
| 3.9      | Conclusion . . . . .                                                                                                                | 74        |
| <b>4</b> | <b>Évaluation quantitative, fonctionnelle et clinique des modèles de super-<br/>résolution appliqués à l'imagerie échographique</b> | <b>78</b> |
| 4.1      | Introduction . . . . .                                                                                                              | 78        |
| 4.1.1    | Objectifs et Contributions de la Partie I . . . . .                                                                                 | 79        |

|          |                                                                                                          |            |
|----------|----------------------------------------------------------------------------------------------------------|------------|
| 4.1.2    | Objectifs et Contributions de la Partie II . . . . .                                                     | 79         |
| 4.2      | Partie I – Étude quantitative et perceptuelle des approches d'apprentissage interne et externe . . . . . | 79         |
| 4.2.1    | Typologie des modèles de super-résolution étudiés . . . . .                                              | 80         |
| 4.2.2    | Jeux de données et diversité des contextes d'acquisition . . . . .                                       | 81         |
| 4.2.3    | Pipeline méthodologique et stratégie d'évaluation . . . . .                                              | 81         |
| 4.2.4    | Métriques d'évaluation quantitative et perceptuelle . . . . .                                            | 81         |
| 4.2.5    | Résultats expérimentaux – Partie I . . . . .                                                             | 84         |
| 4.3      | Partie II : Vers une analyse perceptuelle et statistique multidimensionnelle . . .                       | 98         |
| 4.3.1    | Méthodologie . . . . .                                                                                   | 99         |
| 4.3.2    | Évaluation des performances des modèles de super-résolution . . . . .                                    | 105        |
| 4.3.3    | Méthodologie d'analyse statistique . . . . .                                                             | 109        |
| 4.3.4    | Validation clinique par évaluation experte en aveugle . . . . .                                          | 117        |
| 4.4      | Conclusion . . . . .                                                                                     | 119        |
| <b>5</b> | <b>Fine-Tuning et Généralisation de Modèles de Super-Résolution pour l'Imagerie Échographique Fœtale</b> | <b>122</b> |
| 5.1      | Introduction . . . . .                                                                                   | 122        |
| 5.2      | Apprentissage par Transfert . . . . .                                                                    | 123        |
| 5.2.1    | L'Affinage ( <i>Fine-Tuning</i> ) : Principes Généraux . . . . .                                         | 124        |
| 5.3      | Méthodologie Expérimentale . . . . .                                                                     | 126        |
| 5.3.1    | Architectures de Modèles Sélectionnées . . . . .                                                         | 126        |
| 5.3.2    | Stratégie adoptée dans cette étude . . . . .                                                             | 127        |
| 5.3.3    | Jeux de données pour le Fine-Tuning . . . . .                                                            | 127        |
| 5.3.4    | Protocole de Fine-Tuning . . . . .                                                                       | 129        |
| 5.4      | Évaluation des modèles fine-tunés : Résultats et Discussion . . . . .                                    | 131        |
| 5.4.1    | Analyse du Processus de Fine-Tuning . . . . .                                                            | 131        |
| 5.4.2    | Évaluation Quantitative Globale et Perceptuelle . . . . .                                                | 133        |
| 5.4.3    | Analyse des Performances et Validation Statistique . . . . .                                             | 140        |
| 5.4.4    | Généralisation dans des contextes cliniques à ressources limitées . . . . .                              | 145        |
| 5.5      | Conclusion . . . . .                                                                                     | 146        |
|          | <b>Conclusion Générale</b>                                                                               | <b>148</b> |
|          | <b>Annexes</b>                                                                                           | <b>169</b> |
|          | Annexe A . . . . .                                                                                       | 169        |
|          | 5.5.1 Mécanisme d'extraction des caractéristiques . . . . .                                              | 169        |
|          | 5.5.2 Formalisation de la perte perceptuelle . . . . .                                                   | 169        |
|          | 5.5.3 Définitions des fonctions de perte utilisées (voir Chapitre 5) . . . . .                           | 169        |
|          | 5.5.4 Métriques Basées sur la Matrice de Confusion . . . . .                                             | 170        |
|          | Annexe B . . . . .                                                                                       | 171        |
|          | 5.5.5 Méthodologie générale . . . . .                                                                    | 171        |
|          | 5.5.6 Analyses descriptives . . . . .                                                                    | 171        |
|          | 5.5.7 Analyses inférentielles . . . . .                                                                  | 171        |
|          | 5.5.8 Taille d'effet . . . . .                                                                           | 171        |
|          | 5.5.9 Corrélations entre métriques . . . . .                                                             | 171        |
|          | 5.5.10 Résumé des indicateurs . . . . .                                                                  | 171        |
|          | 5.5.11 Cadre d'évaluation statistique ( Corrélation/Variance) . . . . .                                  | 171        |
|          | Annexe C . . . . .                                                                                       | 173        |
|          | Annexe D . . . . .                                                                                       | 174        |
|          | 5.5.12 Caractérisation de l'architecture VGG-19 . . . . .                                                | 174        |

|                                                                            |     |
|----------------------------------------------------------------------------|-----|
| 5.5.13 Mécanisme d'extraction des caractéristiques . . . . .               | 174 |
| 5.5.14 Formalisation de la perte perceptuelle . . . . .                    | 174 |
| 5.5.15 Définitions des fonctions de perte utilisées (Chapitre 5) . . . . . | 174 |
| Annexe E . . . . .                                                         | 176 |

# Table des figures

|             |                                                                                                                                                                     |    |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1.1  | Frise chronologique des paradigmes de la super-résolution, de la régularisation par priors explicites aux modèles de diffusion génératifs. . . . .                  | 10 |
| Figure 1.2  | Schéma conceptuel du modèle de dégradation et du problème inverse en super-résolution avant l'ère du deep learning. . . . .                                         | 11 |
| Figure 1.3  | Schémas comparatifs des mécanismes de super-résolution : Approche Locale (Interpolation) et Approche Non-Locale (Patches). . . . .                                  | 12 |
| Figure 1.4  | Pipeline du modèle de super-résolution : SRCNN [45] . . . . .                                                                                                       | 14 |
| Figure 1.5  | Schéma fonctionnel du modèle <b>SRGAN</b> [99] . . . . .                                                                                                            | 15 |
| Figure 1.6  | Pipeline fonctionnel du modèle <b>IPT</b> [27]. . . . .                                                                                                             | 17 |
| Figure 1.7  | Pipeline du modèle SR3 : entraînement, inférence et cascade multi-échelles. . . . .                                                                                 | 18 |
| Figure 1.8  | Diagramme de sélection des études selon PRISMA (2020), distinguant revues systématiques et études expérimentales. . . . .                                           | 20 |
| Figure 1.9  | Diagramme PRISMA (2020) – Processus de sélection systématique des 31 études sur le <i>Transfer Learning</i> et le <i>Fine-Tuning en imagerie médicale</i> . . . . . | 26 |
| Figure 2.1  | Schéma d'émission-réception d'une onde ultrasonore pulsée. . . . .                                                                                                  | 36 |
| Figure 2.2  | Principe de génération d'une onde ultrasonore par effet piézo-électrique. . . . .                                                                                   | 38 |
| Figure 2.3  | Échographie en <b>mode A</b> . . . . .                                                                                                                              | 39 |
| Figure 2.4  | Coupe transversale du crâne fœtal en échographie <b>mode B</b> . . . . .                                                                                            | 39 |
| Figure 2.5  | Valve mitrale en échographie <b>mode M</b> . . . . .                                                                                                                | 40 |
| Figure 2.6  | Illustrations des principaux types de Doppler utilisés en échographie fœtale. . . . .                                                                               | 41 |
| Figure 2.7  | Image échographique affectée par le bruit de speckle, . . . . .                                                                                                     | 44 |
| Figure 2.8  | Comparaison des modalités d'imagerie : . . . . .                                                                                                                    | 46 |
| Figure 2.9  | Variabilité inter- et intra-observateur. . . . .                                                                                                                    | 47 |
| Figure 2.10 | Illustration de l'assistance par intelligence artificielle dans l'évaluation échographique hépatique (d'après Zhou et al.[212]). . . . .                            | 47 |
| Figure 2.11 | Schéma de l'atténuation ultrasonore.[180] . . . . .                                                                                                                 | 48 |
| Figure 2.12 | Comparaison d'images échographiques selon la fréquence de la sonde.[20] . . . . .                                                                                   | 49 |
| Figure 2.13 | Obstacles anatomiques à la pénétration ultrasonore.[42] . . . . .                                                                                                   | 49 |
| Figure 2.14 | Exemples de sondes échographiques usuelles . . . . .                                                                                                                | 51 |
| Figure 2.15 | Principaux mouvements de manipulation de la sonde échographique[43] : . . . . .                                                                                     | 52 |
| Figure 3.1  | Architecture globale du DBPISR[90]. . . . .                                                                                                                         | 58 |
| Figure 3.2  | Architecture globale de l'A-ESRGAN [194] . . . . .                                                                                                                  | 59 |
| Figure 3.3  | Discriminateur U-Net avec attention simple . . . . .                                                                                                                | 61 |
| Figure 3.4  | Schéma bloc d'A-ESRGAN . . . . .                                                                                                                                    | 62 |



|             |                                                                                                                                                                                                                                                                |     |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 3.5  | Schéma bloc de BSRGAN . . . . .                                                                                                                                                                                                                                | 62  |
| Figure 3.6  | Schéma bloc de Real-ESRGAN . . . . .                                                                                                                                                                                                                           | 65  |
| Figure 3.7  | Architecture globale de SwinIR[105] . . . . .                                                                                                                                                                                                                  | 68  |
| Figure 3.8  | Architecture générale de Swin2SR[35]. . . . .                                                                                                                                                                                                                  | 70  |
| Figure 3.9  | Architecture globale du modèle HAT[29]. . . . .                                                                                                                                                                                                                | 72  |
| Figure 3.10 | Schéma global de la méthodologie expérimentale . . . . .                                                                                                                                                                                                       | 76  |
| Figure 4.1  | Schéma méthodologique du pipeline expérimental pour l'évaluation des approches de super-résolution. . . . .                                                                                                                                                    | 80  |
| Figure 4.2  | ConvNeXt-Base [115] : (a) Architecture Globale; (b) Détail d'un bloc ConvNeXt; (c) Mécanisme d'échantillonnage descendant ( <i>down-sampling</i> ). . . . .                                                                                                    | 83  |
| Figure 4.3  | Comparaison entre les images échographiques originales et reconstruites : (a) DBPISR, (b) BSRGAN, (c) Real-ESRGAN, (d) SwinIR-M et (e) SwinIR-L. . . . .                                                                                                       | 87  |
| Figure 4.4  | Zooms d'images échographiques du fémur obtenus via cinq modèles de super-résolution (jeu de données Malawi) : (a) Image originale, (b–f) Résultats des modèles SR. . . . .                                                                                     | 89  |
| Figure 4.5  | Boxplots des métriques PSNR, SSIM et (PSNR/BRISQUE) pour les cinq modèles de super-résolution : (a) BSRGAN, (b) Real-ESRGAN, (c) SwinIR-M, (d) SwinIR-L et (e) DBPISR. . . . .                                                                                 | 90  |
| Figure 4.6  | Boxplots de PSNR, SSIM et (PSNR/BRISQUE) pour les cinq modèles SR : (a) BSRGAN, (b) Real-ESRGAN, (c) SwinIR-M, (d) SwinIR-L et (e) DBPISR. . . . .                                                                                                             | 91  |
| Figure 4.7  | Nuages de points PSNR–SSIM pour les cinq pays étudiés : (a) Algérie, (b) Égypte, (c) Ghana, (d) Malawi et (e) Ouganda. . . . .                                                                                                                                 | 92  |
| Figure 4.8  | Comparaison des performances des modèles de super-résolution pour cinq pays africains : Algérie, Égypte, Ghana, Malawi et Ouganda. . . . .                                                                                                                     | 95  |
| Figure 4.9  | Comparaison des évaluations des modèles de super-résolution : (a) appréciation individuelle d'un expert médical, (b) consensus général de tous les experts confondus . . . . .                                                                                 | 97  |
| Figure 4.10 | Comparaisons agrandies des reconstructions anatomiques : (a) Région fémorale issue du dataset <i>SPAIN</i> et (b) Région cérébrale fœtale issue du dataset <i>AFRICA</i> , avec Bicubic (LR), Vérité terrain (HR), HAT, Swin2SR, BSRGAN et AESRGAN-1D. . . . . | 110 |
| Figure 4.11 | Performances des modèles de super-résolution les mieux classés sur cinq jeux de données échographiques . . . . .                                                                                                                                               | 113 |
| Figure 4.12 | Classement clinique complet des huit modèles évalués par l'experte médicale. Le score final intègre les pénalités appliquées aux critères critiques de résolution spatiale et de confiance diagnostique. . . . .                                               | 117 |
| Figure 4.13 | Nuages de points illustrant la relation entre les métriques perceptuelles objectives et les scores moyens d'opinion (MOS). . . . .                                                                                                                             | 118 |
| Figure 4.14 | Corrélation des métriques quantitatives et perceptuelles avec l'évaluation clinique (MOS). . . . .                                                                                                                                                             | 119 |
| Figure 5.1  | Illustration hiérarchique des couches d'un réseau convolutionnel avec indication des couches gelées et entraînaibles pour le <i>Fine-Tuning</i> . . . . .                                                                                                      | 125 |

|            |                                                                                                                                                                              |     |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 5.2 | Évolution des performances pendant le fine-tuning : comparaison du PSNR des modèles HAT et SwinIR-M (a) et évolution du SSIM pour le modèle HAT (b). . . . .                 | 132 |
| Figure 5.3 | Comparaison globale des modèles HAT et SwinIR-M après fine-tuning : distributions multi-métriques (PSNR, SSIM, LPIPS, GMSD, FID, NIQE) sur les cinq jeux de données. . . . . | 136 |
| Figure 5.4 | Zoom d’une région fémorale sur une image échographique fœtale (base de données AFRICA). . . . .                                                                              | 137 |
| Figure 5.5 | Comparaison de l’Indice de Préservation des Contours (EPI) sur les Images AFRICA et ISTHMIC. . . . .                                                                         | 138 |
| Figure 5.6 | Distribution globale des métriques de qualité sur l’ensemble des jeux de données évalués. . . . .                                                                            | 139 |
| Figure 5.7 | Comparaison détaillée des performances de HAT et SwinIR-M métrique par métrique. . . . .                                                                                     | 142 |

# Liste des tableaux

|             |                                                                                                                                                            |    |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tableau 1.1 | Jeux de données naturels utilisés pour l’entraînement et l’évaluation en super-résolution. . . . .                                                         | 9  |
| Tableau 1.2 | Jeux de données biomédicaux de référence pour la super-résolution d’images. . . . .                                                                        | 9  |
| Tableau 1.3 | Taxinomie des métriques d’évaluation en super-résolution . . .                                                                                             | 10 |
| Tableau 1.4 | Évolution des méthodes pré-deep learning pour la super-résolution                                                                                          | 13 |
| Tableau 1.5 | Synthèse transversale des paradigmes de super-résolution : objectifs, atouts, limites et modèles représentatifs. . . . .                                   | 19 |
| Tableau 1.6 | Évolution chronologique des approches de super-résolution et applications associées en imagerie médicale (2020–2025) . . . .                               | 22 |
| Tableau 1.7 | Synthèse de 23 revues sur la SR médicale par Deep Learning : objectifs et améliorations rapportées . . . . .                                               | 24 |
| Tableau 1.8 | Synthèse de 9 revues majeures sur le transfert d’apprentissage en imagerie médicale. . . . .                                                               | 28 |
| Tableau 1.9 | Études empiriques exhaustives sur le transfer learning et le fine-tuning en imagerie médicale (22 articles). . . . .                                       | 30 |
| Tableau 2.1 | Résumé des modes échographiques avec exemples illustratifs . .                                                                                             | 40 |
| Tableau 2.2 | Résumé des modes Doppler en échographie foetale : usage clinique et exemples . . . . .                                                                     | 42 |
| Tableau 2.3 | Synthèse des principaux artefacts en échographie et de leur origine physique. . . . .                                                                      | 50 |
| Tableau 3.1 | Synthèse comparative des méthodes de super-résolution interne/aveugle (CNN-based). <sup>1</sup> . . . . .                                                  | 58 |
| Tableau 3.2 | Forces, limitations et impact médical des modèles GAN. . . . .                                                                                             | 66 |
| Tableau 3.3 | Synthèse comparative des architectures GAN pour la super-résolution aveugle et leur impact médical. . . . .                                                | 66 |
| Tableau 3.4 | Forces, limitations et impact médical des modèles Transformers en imagerie médicale. . . . .                                                               | 73 |
| Tableau 3.5 | Comparaison bibliographique des principaux modèles de super-résolution évoqués en imagerie médicale. . . . .                                               | 75 |
| Tableau 4.1 | Performances des modèles de super-résolution : évaluation de la qualité d’image dans les pays étudiés . . . . .                                            | 84 |
| Tableau 4.2 | Scores NIQE et BRISQUE par pays et par modèle . . . . .                                                                                                    | 85 |
| Tableau 4.3 | Protocole de classification d’images à partir des images originales et des images super-résolues . . . . .                                                 | 86 |
| Tableau 4.4 | Performance des modèles de super-résolution pour la classification d’images sur le jeu de données espagnol (images originales et super-résolues) . . . . . | 86 |

|              |                                                                                                                                                                                                                  |     |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tableau 4.5  | Performance des modèles de super-résolution pour la classification d’images selon les pays de l’étude (images originales et super-résolues) . . . . .                                                            | 86  |
| Tableau 4.6  | Performances des modèles de super-résolution : Algérie, Égypte, Ghana, Malawi et Ouganda, classés du meilleur au pire . . . . .                                                                                  | 88  |
| Tableau 4.7  | Impact de la super-résolution sur les performances de classification des images échographiques (cas espagnol). . . . .                                                                                           | 93  |
| Tableau 4.8  | Impact de la super-résolution sur les métriques de classification : Algérie, Égypte, Ghana, Malawi et Ouganda. . . . .                                                                                           | 93  |
| Tableau 4.9  | Jeu de données <i>FPUS23</i> [148] . . . . .                                                                                                                                                                     | 100 |
| Tableau 4.10 | Caractéristiques des jeux de données espagnol et africain [21, 161]                                                                                                                                              | 101 |
| Tableau 4.11 | Caractéristiques du jeu de données <i>PSFHS</i> pour la super-résolution [26] . . . . .                                                                                                                          | 101 |
| Tableau 4.12 | Caractéristiques du jeu de données <i>ISTHMIC</i> Contraction [54]                                                                                                                                               | 102 |
| Tableau 4.13 | Origine des modèles de super-résolution pré-entraînés utilisés dans la partie II de l’étude. . . . .                                                                                                             | 103 |
| Tableau 4.14 | Performances moyennes de chaque modèle, agrégées sur l’ensemble des cinq jeux de données. . . . .                                                                                                                | 106 |
| Tableau 4.15 | Résumé des mesures statistiques . . . . .                                                                                                                                                                        | 111 |
| Tableau 4.16 | Scores globaux des modèles de super-résolution selon les jeux de données. . . . .                                                                                                                                | 111 |
| Tableau 4.17 | Métriques d’évaluation et score composite utilisés pour l’analyse comparative. . . . .                                                                                                                           | 115 |
| Tableau 4.18 | Classement des modèles par métrique de performance. Les valeurs reportées sont les moyennes brutes calculées sur l’ensemble des jeux de données. Les meilleurs scores pour chaque métrique sont en gras. . . . . | 115 |
| Tableau 4.19 | Métriques normalisées et Score Équilibré. . . . .                                                                                                                                                                | 115 |
| Tableau 5.1  | Travaux représentatifs sur l’apprentissage par transfert et le fine-tuning en imagerie médicale (sélection). . . . .                                                                                             | 126 |
| Tableau 5.2  | Paramètres utilisés pour le fine-tuning des modèles HAT et SwinIR-M. . . . .                                                                                                                                     | 127 |
| Tableau 5.3  | Caractéristiques des jeux de données utilisés. . . . .                                                                                                                                                           | 128 |
| Tableau 5.4  | Résultats comparatifs avant/après fine-tuning sur Spain_val. .                                                                                                                                                   | 134 |
| Tableau 5.5  | Performances comparatives des modèles fine-tunés (HAT et SwinIR-M) sur l’ensemble des jeux de données externes ( <b>AFRICA</b> , <b>ISTHMIC</b> , <b>FPUS23</b> , <b>CHINESE</b> et <b>HC</b> ). . . . .         | 135 |
| Tableau 5.6  | Métriques comparatives des modèles SR sur différents jeux de données. . . . .                                                                                                                                    | 140 |
| Tableau 5.7  | Correspondances entre défis échographiques et composants de HAT. . . . .                                                                                                                                         | 140 |
| Tableau 5.8  | Évolution des métriques objectives, perceptuelles, génératives et sans référence sur l’ensemble des jeux de données, en configuration <i>baseline</i> (BL) et après fine-tuning (FT). . . . .                    | 144 |
| Tableau 5.9  | Comparaison des performances de HAT sur deux cohortes d’échographie fœtale. La significativité statistique est évaluée à $p < 0.05$ . Cohen’s $d$ quantifie la taille d’effet. . . . .                           | 145 |
| Tableau 5.10 | Métriques d’évaluation pour la classification binaire et la segmentation. . . . .                                                                                                                                | 170 |

|              |                                                                                              |     |
|--------------|----------------------------------------------------------------------------------------------|-----|
| Tableau 5.11 | Résumé des indicateurs statistiques utilisés. . . . .                                        | 171 |
| Tableau 5.12 | Performance moyenne des modèles sur tous les jeux de données<br>étudiés. . . . .             | 173 |
| Tableau 5.13 | Exemple de p-valeurs obtenues avant application de la correc-<br>tion de Bonferroni. . . . . | 176 |

# Liste des abréviations

- AAPM** Jeu de données de scanners à faible dose utilisé pour la recherche en débruitage et reconstruction
- AdaptFormer** Transformer avec modules d'adaptation légers
- ADNI** Initiative sur la Neuroimagerie de la Maladie d'Alzheimer / Alzheimer's Disease Neuroimaging Initiative
- AIFT** Extension adaptative du SIFT
- AlexNet** Réseau convolutionnel profond pour la classification d'images
- arXiv** Archive électronique de prépublications scientifiques
- AUROC** Aire sous la courbe ROC, indicateur de performance globale d'un classifieur
- AUTOMAP** Reconstruction automatique d'images IRM directement depuis les données de l'espace  $k$
- B-mode** Mode Brightness en échographie
- BASICVSR** Basic Video Super-Resolution Network
- BiT** Modèle de transfert à grande échelle pré-entraîné
- Bit-FIT** Fine-tuning des seuls termes de biais
- BRATS** Jeu de données d'IRM cérébrales annotées pour la segmentation des tumeurs
- BreakHis** Base de données d'images histopathologiques du cancer du sein
- BRISQUE** Évaluation aveugle sans référence de la qualité d'image spatiale (Blind/Referenceless Image Spatial Quality Evaluator)
- BSD100** Jeu de données BSD comprenant 100 images naturelles
- Camelyon** Base de données de métastases ganglionnaires en histopathologie
- CAMUS** Jeu de données d'échocardiographies annotées pour l'analyse et la segmentation cardiaque
- CEUS** Technique d'ultrasons avec contraste
- cGAN** GAN Conditionnel
- CHAOS** Base de données pour la segmentation d'organes abdominaux
- CheXpert** Base de données de radiographies thoraciques
- CNN** Réseau de Neurones Convolutifs
- CNR** Rapport Contraste sur Bruit (Contrast-to-Noise Ratio)
- Cond-GAN** GAN conditionnel pour images médicales
- CONV** Convolution (opération mathématique sur les images)
- CONVNEXT** ConvNeXt, architecture de réseau neuronal convolutif moderne pour la classification d'images (ConvNeXt, modern convolutional neural network architecture for image classification)

- Covid-CT** Base de données de tomodensitométrie thoracique COVID-19
- CSRIP** Traitement d'images par super-résolution convolutive
- CT** Tomodensitométrie (imagerie par scanner)
- custom-WDSR** Super-résolution  $2\times/4\times$  d'images et vidéos échographiques 2D avec préservation du bruit
- CycleGAN** Réseau antagoniste génératif cyclique
- DCSRN** Réseau convolutif profond dilaté pour super-résolution
- DEEP LEARNING** Apprentissage Profond (Deep Learning)
- Deep-ULM** Localisation de microbulles par apprentissage profond en ultrasons
- DenseNet** Réseau convolutionnel à connexions denses pour classification d'images (Densely Connected Convolutional Network)
- DenseNet-121** DenseNet à 121 couches
- DnCNN** Réseau de neurones convolutifs pour débruitage
- DOMAIN SHIFT** Transfert de domaine (Domain Shift)
- DRN** Réseau Résiduel Profond (Deep Residual Network)
- DRRN** Réseau Résiduel Récurrent Profond (Deep Residual Recurrent Network)
- DSC** Dice Similarity Coefficient
- EDSR** Réseau profond amélioré de super-résolution d'images basé sur CNN
- EDVR** Enhanced Deformable Video Restoration
- EfficientNetB0** CNN optimisé par compound scaling
- ENL** Nombre Équivalent de Looks (Equivalent Number of Looks)
- EPI** Indice de Préservation des Contours (Edge Preservation Index)
- ESRGAN** Réseau antagoniste génératif amélioré pour la super-résolution
- EUSR** Super-résolution améliorée par mise à l'échelle
- FA-GAN** GAN sensible aux caractéristiques
- FID** Distance de Fréchet des distributions d'images (Fréchet Inception Distance)
- FINE-TUNING** Ajustement fin d'un modèle pré-entraîné
- FLICKR2K** Jeu de données composé d'images naturelles haute résolution, utilisé pour l'entraînement en super-résolution
- FPUS** Échographie ultrarapide par ondes planes
- FS-GAN** GAN conditionnel pour super-résolution
- FSRCNN** Réseau de neurones convolutifs pour super-résolution rapide
- Fundus** Base de données d'images du fond d'œil
- FUSPR** Fetal Ultrasound Standard Plane Recognition
- GAN** Réseau Antagoniste Génératif (Generative Adversarial Network)
- GAN 3D** Super-résolution d'IRM cérébrale par réseaux antagonistes génératifs 3D
- GAN-CIRCLE** Réseau antagoniste génératif avec contraintes identité, résiduelles et cycliques pour CT
- GMSD** Déviation de Similarité des Magnitudes de Gradient (Gradient Magnitude Similarity Deviation)

**GoogLeNet** Réseau convolutionnel à modules Inception

**Grad-CAM** Méthode de visualisation basée sur les gradients

**HAT** Transformeur à Attention Hybride (Hybrid Attention Transformer)

**Heatmaps** Cartes de visualisation d'intensité

**HERO-SR** Super-résolution optimisée par résidus haute efficacité, modèle CNN

**HR** Haute Résolution (High Resolution)

**IA** Intelligence Artificielle (Artificial Intelligence)

**ImageNet** Base de données d'images composée de plus de 14 millions d'images annotées manuellement, couvrant plus de 20 000 catégories organisées selon la hiérarchie WordNet ; le sous-ensemble ILSVRC comprend environ 1,2 million d'images réparties en 1 000 classes

**Inception** Architecture Inception (Inception Network)

**IPT** Image Processing Transformer

**IXI** Jeu de données d'IRM cérébrales multi-séquences acquis sur des sujets sains

**LapSRN** Réseau de super-résolution basé sur une pyramide laplacienne (Laplacian Pyramid Super-Resolution Network)

**LEARN** Réseau de reconstruction basé sur une évaluation d'experts apprise (Learned Experts' Assessment-based Reconstruction Network)

**LOLA11** Jeu de données de scanners thoraciques annotés pour la segmentation pulmonaire

**LoRA** Adaptation par décomposition de rang faible

**LP** Stratégie d'entraînement consistant à geler l'extracteur de caractéristiques

**LPGAN** Réseau antagoniste génératif à pyramide de Laplace

**LPIPS** Distance perceptuelle apprise sur des patches d'images (Learned Perceptual Image Patch Similarity)

**LR** Basse Résolution (Low Resolution)

**LSTM** Réseau de neurones récurrent pour traiter des données séquentielles et capturer des dépendances à long terme.

**LUNA16** Analyse des nodules pulmonaires 2016 (Lung Nodule Analysis 2016)

**MAP-MRF SR** Super-résolution multi-images basée sur une estimation Maximum a Posteriori avec prior de type champ aléatoire de Markov

**MedSRGAN** Réseau antagoniste génératif spécialisé pour la super-résolution multimodale médicale

**Meta-learning** Apprentissage méta, méthode permettant à un modèle de s'adapter rapidement à de nouvelles tâches

**MIASSR** Apprentissage méta + GAN pour super-résolution d'IRM et CT

**MIRN** Réseau résiduel amélioré multiple pour la super-résolution d'IRM

**MobileNetV3** CNN léger pour applications mobiles

**MOS** Score Moyen d'Opinion (Mean Opinion Score)

**MSE** Erreur quadratique moyenne (Mean Squared Error)

**mSPCN-ULM** Réseau subpixel modifié pour l'imagerie ultrasonore de microbulles (ULM)

**NCSR** Représentation parcimonieuse centralisée non locale pour la super-résolution d'images



**NIQE** Évaluation naturelle de la qualité d'image (Natural Image Quality Evaluator)

**NT** Transfert négatif

**OASIS** Base de données ouverte d'imagerie cérébrale

**OCAB** Overlapping Cross-Attention Block (bloc d'attention croisée avec recouvrement)

**OCT** Tomographie par cohérence optique

**PEFT** Méthodes de fine-tuning à faible coût paramétrique

**PET** Tomographie par émission de positons

**POCUS** Ultrasons point-of-care (échographie au point de soin)

**PRISMA 2020** Éléments de rapport préférés pour les revues systématiques et les méta-analyses, version 2020

**PSF** Fonction de flou (Point Spread Function)

**PSNR** Rapport Signal sur Bruit de Crête (Peak Signal-to-Noise Ratio)

**r** Coefficient de corrélation de Pearson (Pearson correlation coefficient)

**RCAN** Réseau de neurones convolutifs résiduel avec attention sur les canaux pour la super-résolution d'images

**RDN** Réseaux denses résiduels pour la super-résolution d'images médicales (IRM, CT)

**REAL-ESRGAN** Real-World Enhanced Super-Resolution GAN (extension d'ESRGAN adaptée aux dégradations réelles)

**RELU** Rectified Linear Unit (fonction d'activation)

**ResNet** Réseau Résiduel (Residual Network)

**ResNet50V2** Réseau résiduel profond à 50 couches (version V2)

**Restormer** Transformeur pour restauration et super-résolution d'images médicales

**RRDB** Residual-in-Residual Dense Block (bloc dense résiduel imbriqué)

**SET14** Jeu de données de référence composé de 14 images naturelles pour l'évaluation quantitative en super-résolution

**SET5** Jeu de données de référence composé de 5 images naturelles, utilisé pour l'évaluation en super-résolution

**Siamese** Architecture à deux branches partageant les poids

**SICNN** Réseau de neurones convolutifs invariant à l'échelle

**SIFT** Méthode d'extraction de descripteurs invariants à l'échelle

**SMOTE** Méthode de suréchantillonnage synthétique des classes minoritaires

**SOUP-GAN** Super-résolution 3D d'IRM par réseau antagoniste génératif avec perte perceptuelle pour produire des coupes fines et nettes

**SPADE-GAN** Réseau antagoniste génératif avec normalisation adaptative spatiale

**Sparse SR** Super-résolution d'images basée sur la représentation parcimonieuse, qui modélise les images haute résolution comme une combinaison linéaire de patches appris à partir d'exemples

**SR** Super-résolution (Super-Resolution)

**SR3** Super-résolution par raffinements répétés, modèle basé sur diffusion

- SRCNN** Réseau de neurones convolutifs pour la super-résolution d'images
- SRDenseNet** Réseau dense pour super-résolution d'images (Super-Resolution Dense Network)
- SRFormerV2** Transformeur efficace et léger pour la super-résolution d'images
- SRGAN** Réseau antagoniste génératif pour la super-résolution
- SSIM** Indice de similarité structurelle (Structural Similarity Index Measure)
- StableSR** Stable Super-Résolution, modèle basé sur la diffusion pour l'amélioration d'images réelles
- StyleGAN** Réseau antagoniste génératif pour synthèse d'images médicales
- SVM** Machine à vecteurs de support
- SwinFIR** Swin Transformer révisé avec convolution de Fourier rapide et entraînement amélioré pour la super-résolution d'images
- SwinFPN** Swin Transformer with Feature Pyramid Network
- SwinGAN** Réseau antagoniste génératif utilisant un générateur basé sur Swin Transformer pour la traduction multi-modale d'images médicales
- SWINIR** Transformeur Swin pour la restauration d'images
- T-GAN** Réseau antagoniste génératif combinant CNN résiduel et Transformeur pour super-résolution d'IRM et CT faible dose
- TRANSFORMERS** Architecture basée sur le mécanisme d'attention (Attention-based Architecture)
- TransGAN** Réseau antagoniste génératif basé sur transformeur pour IRM et CT
- TransUNet** Combinaison de transformeur et U-Net pour segmentation IRM et CT
- U-NET** réseau de neurones convolutif pour la segmentation d'images biomédicales
- ULM** Microscopie de Localisation par Ultrasons (Ultrasound Localization Microscopy)
- URBAN100** Jeu de données contenant 100 images de scènes urbaines riches en structures fines, utilisé pour l'évaluation en super-résolution
- VAEs** Auto-encodeurs variationnels
- VGG** Visual Geometry Group, réseau de neurones profond pour la vision par ordinateur
- VGG-19** Réseau convolutionnel profond à 19 couches
- ViT-SR** Vision Transformer pour la super-résolution (Vision Transformer for Super-Resolution)
- ViT-V-Net** Vision Transformer V-Net
- WGAN-GP** Wasserstein GAN avec gradient penalty
- ZERO-SHOT** Apprentissage sans données annotées spécifiques (Zero-shot Learning)



# Introduction Générale

---

# Introduction Générale

Au-delà du simple constat visuel, l'imagerie médicale s'est métamorphosée en une métrologie du vivant. En cristallisant la complexité biologique en indicateurs quantifiables, elle s'est imposée comme le socle souverain de la décision clinique. Ce saut qualitatif procède d'une symbiose étroite entre la sensibilité des capteurs et l'intelligence du traitement numérique, capable aujourd'hui de forcer les limites physiques de l'acquisition. Pourtant, si cette excellence du signal consacre la suprématie des grands centres hospitaliers, elle agit aussi, en négatif, comme le révélateur d'une fracture technologique mondiale.

Derrière les progrès technologiques se dessine une réalité incontournable : sans ressources suffisantes, l'innovation ne profite pas à tous ceux qui en ont besoin. L'accès aux soins reste une question de géographie : de l'Afrique subsaharienne aux zones rurales isolées, l'obsolescence du matériel et les pannes récurrentes figent la qualité des diagnostics[86, 60]. Ce fossé n'épargne pas les pays développés, où l'écart se creuse entre l'excellence des pôles universitaires et la précarité des structures de proximité[144]. L'enjeu technique s'efface ainsi derrière une réalité plus urgente : celle de réduire les inégalités face aux standards de soins.

Par son innocuité et sa simplicité, l'échographie a certes vocation à être le pilier universel du suivi prénatal[67], mais cette promesse se heurte aux limites du signal. Mais la réalité du terrain impose un paradoxe : avoir un échographe ne signifie pas pouvoir réaliser un diagnostic[56, 74]. Un signal dégradé, noyé dans le bruit et manquant de finesse, peut masquer une malformation fœtale et rendre l'examen inutile. Face à l'impossibilité de remplacer partout les vieux parcs matériels, les outils classiques de nettoyage d'image atteignent leurs limites : ils ne peuvent pas inventer l'information manquante.

Malgré l'essor des technologies, la course à l'innovation se heurte à une vérité incontournable : tout progrès reste limité par les moyens disponibles. Si les modèles actuels excellent à restaurer l'information, ils se heurtent souvent à la variabilité du signal médical. La question clé est simple mais cruciale : les algorithmes d'IA peuvent-ils préserver ce qui fait l'information diagnostique d'une image dégradée ?

Il s'agit de concevoir des outils de super-résolution assez robustes pour être utiles aux praticiens, malgré les contraintes techniques. Pour cela, nous ne nous contentons pas de mesures mathématiques ; nous soumettons nos algorithmes au regard des médecins. Nous étudions également comment adapter ces modèles à des contextes où les données sont rares (*fine-tuning*), une étape indispensable pour un déploiement futur comme application embarquée.

L'enjeu est de garantir que les modèles de super-résolution conservent la valeur diagnostique nécessaire en pratique clinique, au-delà de la seule qualité esthétique des images.

Notre travail s'inscrit ainsi dans une démarche globale visant à améliorer la qualité des images échographiques fœtales par super-résolution (SR), en tenant compte des contraintes spécifiques de cette modalité — *speckle*, variabilité des acquisitions et rareté des données haute résolution. L'objectif est de renforcer l'aide au diagnostic en fournissant aux cliniciens des images plus interprétables, préservant les structures anatomiques essentielles. Cette ambition prend tout son sens dans les environnements aux ressources limitées (zones isolées ou pays sous-dotés), où l'échographie, malgré sa disponibilité, souffre d'une qualité d'image souvent insuffisante pour un diagnostic fiable.

À la lumière des avancées récentes en SR appliquée à l'imagerie médicale - notamment les travaux portant sur des architectures profondes robustes et généralisables - nous adoptons une approche méthodologique fondée sur le transfert d'apprentissage et le *fine-tuning*. Cette stratégie vise à spécialiser des modèles initialement conçus pour d'autres modalités ou contextes, afin de préserver la fidélité des structures anatomiques tout en limitant l'apparition d'artefacts cliniquement trompeurs — un verrou majeur pour une adoption en routine diagnostique.

La démarche scientifique de cette thèse s'articule autour de trois phases de contributions majeures, visant à transformer la super-résolution par *deep learning* en un outil cliniquement fiable pour l'échographie fœtale :

### 1. Caractérisation des paradigmes et validation de la préservation sémantique

Une première contribution fondamentale réside dans l'établissement des fondements techniques de la super-résolution appliquée au domaine fœtal.

Nos travaux ont permis de :

- **Évaluer la généralisabilité sur données multicentriques** : Nous avons éprouvé la robustesse des modèles face à la diversité des équipements via l'exploitation de bases de données hétérogènes.
- **Comparer les paradigmes d'apprentissage** : Cette phase a permis de caractériser les compromis entre l'apprentissage interne (*zero-shot*) et externe en utilisant de grandes bases de données. À l'aide de classificateurs de pointe (ConvNeXt), nous avons pu vérifier si les plans anatomiques fœtaux restaient identifiables après super-résolution, attestant ainsi de la préservation de l'information sémantique diagnostique.
- **Établir une corrélation clinique initiale** : Une évaluation basée sur le **score d'opinion moyen (MOS) avec référence** a permis d'ancrer les performances quantitatives dans une réalité clinique en confrontant, via un panel d'experts, les reconstructions aux images originales.

### 2. Rupture technologique et modélisation prédictive du jugement expert

Dans un second temps, nous avons exploré le potentiel des architectures émergentes et formalisé la fiabilité du diagnostic. Cette étape apporte les contributions suivantes :

- **Validation de l'ascendant des architectures Transformers** : Nous avons mis en évidence la supériorité du modèle *HAT* sur les réseaux antagonistes (GAN), offrant un équilibre optimal entre acuité visuelle et fidélité structurelle.
- **Formalisation d'un protocole d'évaluation multidimensionnel** : Introduction d'un cadre d'analyse tridimensionnel (fidélité structurelle, qualité perceptuelle et validité diagnostique) destiné à contrôler que l'augmentation de résolution préserve la morphologie fœtale sans introduire d'hallucinations numériques ni d'artefacts cliniquement significatifs.
- **Analyse statistique de l'expertise clinique** : Nous avons conduit une étude en aveugle via un protocole **MOS sans référence**, révélant une forte corrélation entre le jugement médical et l'indice **GMSD** ( $r = 0,876$ ,  $p = 0,004$ ). Cette métrique explique à elle seule 76,7% de la variance des scores cliniques ( $R^2 = 0,767$ ), objectivant ainsi le lien entre performance algorithmique et perception diagnostique.

### 3. Spécialisation et résilience pour le déploiement en environnement contraint

Enfin, nos efforts ont porté sur l'optimisation des modèles pour permettre leur déploiement opérationnel dans des contextes cliniques aux ressources limitées :

- **Stratégie de spécialisation par *fine-tuning* intégral** : Mise au point d'un protocole d'adaptation spécifique reposant sur le prétraitement manuel de corpus haute résolution pour garantir la fidélité des structures apprises.
- **Résilience face à la rareté des données** : Nous avons démontré l'efficacité de cette adaptation sur des cohortes limitées et des données intentionnellement dégra-

dées, assurant la généralisation du modèle dans des contextes cliniques variés.

- **Évaluation par métriques acoustiques** : L'intégration d'indicateurs spécifiques à l'échographie (ENL, CNR, EPI) a permis de valider quantitativement le gain en contraste et la préservation des contours indispensables à la pratique radiologique.

Pour atteindre ces objectifs, la thèse s'articule en cinq chapitres complémentaires, chacun abordant une étape clé de la démarche scientifique :

- **Chapitre 1** : Il dresse un état de l'art critique de la super-résolution d'images médicales, examinant le passage des approches conventionnelles aux modèles profonds, afin d'isoler les outils algorithmiques capables de restaurer avec fidélité les structures anatomiques complexes. Il couvre également les stratégies de transfert d'apprentissage et de fine-tuning, leviers essentiels pour adapter ces modèles aux spécificités du domaine échographique fœtal en contexte de données limitées.
- **Chapitre 2** : Il présente les fondements de l'imagerie ultrasonore — principes de propagation, modes d'imagerie et caractéristiques du signal (speckle, artefacts) — ainsi que ses avantages cliniques (non invasif, temps réel, faible coût) et ses limitations inhérentes (dépendance opérateur, atténuation, champ de vision restreint), qui justifient le recours aux méthodes de super-résolution abordées dans les chapitres suivants.
- **Chapitre 3** : Il établit le socle méthodologique des Chapitres 4 et 5 en présentant les architectures de super-résolution retenues selon deux paradigmes : l'apprentissage interne, représenté par le CNN DBPISR, et l'apprentissage externe, décliné en réseaux GANs (BSRGAN, Real-ESRGAN, A-ESRGAN) et en Transformers (SwinIR, Swin2SR, HAT). Les fondements théoriques, atouts et limites de chaque architecture sont analysés au regard des contraintes spécifiques de l'imagerie ultrasonore : nature du speckle, anisotropie de la PSF et risque d'hallucination anatomique.
- **Chapitre 4** : Il soumet les modèles de super-résolution retenus à une évaluation hybride en deux phases. La première, combinant métriques objectives, classification automatique des plans anatomiques fœtaux par **ConvNeXt** et protocole **MOS avec référence** auprès d'un panel de quatre expertes médicales, valide la préservation de l'information sémantique diagnostique. La seconde, fondée sur un cadre d'analyse tridimensionnel et un protocole **MOS sans référence** mené par une experte médicale senior, hiérarchise les modèles selon leur valeur diagnostique perçue et établit, via une analyse statistique, la significativité des différences entre modèles ainsi qu'une forte corrélation entre perception clinique et indice GMSD ( $r = 0,876$ ,  $R^2 = 0,767$ ).
- **Chapitre 5** : Cette étape finale permet de juger à la fois la fiabilité des modèles et leur transférabilité vers des conditions quasi réelles, où les contraintes opérationnelles deviennent prépondérantes. La démarche s'articule autour de deux axes critiques : (i) l'adaptabilité des solutions via *fine-tuning* pour pallier la limitation des données, et (ii) l'évaluation de la généralisation sur des jeux de tests exigeants, incluant des cas d'acquisition particulièrement dégradés, afin d'identifier les modèles les plus fiables pour un déploiement en contexte contraint.

En conclusion, cette thèse répond à la problématique posée en déployant une démarche scientifique articulée autour de trois contributions complémentaires — fondements techniques, rupture technologique et déploiement opérationnel — visant à transformer la super-résolution par *deep learning* en un outil cliniquement fiable pour l'échographie fœtale. À terme, ces travaux constituent un premier pas vers l'intégration de la super-résolution comme outil d'aide au diagnostic embarqué en gynécologie-obstétrique, accessible aux praticiens opérant dans des environnements aux ressources limitées, là où l'amélioration de la qualité d'image échographique peut avoir l'impact clinique le plus décisif sur le suivi prénatal et la détection précoce des malformations fœtales.



# CHAPITRE 1

---

# Chapitre 1

## État de l’art de la Super-Résolution d’Images : des Approches Classiques aux Applications Médicales

### 1.1 Introduction

La résolution spatiale limite intrinsèquement la discrimination des structures anatomiques fines en routine clinique. Face aux barrières physiques des chaînes d’acquisition, la super-résolution (SR) logicielle permet d’inverser le modèle de dégradation  $I_{LR} = D(I_{HR}) + n$  où  $D$  incarne l’opérateur de dégradation, synthèse du flou et du sous-échantillonnage, et  $n$  un processus de bruit.

Le passage des méthodes de régularisation analytique au paradigme de l’apprentissage profond a radicalement déplacé le problème : il ne s’agit plus de résoudre une équation de parcimonie, mais d’apprendre des distributions conditionnelles  $P(I_{HR}|I_{LR})$  sur de vastes corpus. Cette approche permet de restaurer les hautes fréquences inaccessibles aux modèles statistiques classiques, mais introduit des défis critiques de fidélité et de généralisation.

L’évolution des architectures profondes (CNN, GAN, Transformers) a déplacé le curseur de la simple réduction d’erreur pixel (PSNR) vers la reconstruction de textures perceptuellement convaincantes. Toutefois, l’efficacité de ces modèles sur des images naturelles ne garantit pas leur viabilité clinique, car la Super-Résolution médicale exige une intégrité anatomique absolue, excluant toute hallucination de détails. La SR clinique impose un arbitrage entre fidélité distortionnelle, qualité visuelle et surtout intégrité anatomique. Contrairement aux domaines grand public, la priorité est ici d’éviter l’hallucination de détails par le réseau : chaque pixel reconstruit doit correspondre à une réalité physiologique.

Si la Super-Résolution a prouvé son efficacité sur des bases de données conventionnelles comme DIV2K, son application clinique reste limitée par la spécificité du signal médical. Contrairement aux images naturelles, l’imagerie médicale exige une approche ancrée dans la physique de l’acquisition, capable de distinguer le bruit délétère des signatures structurelles du vivant. Réussir le transfert technologique impose donc un changement de priorité : la fidélité anatomique et la sûreté du diagnostic doivent désormais primer sur le rendu visuel.

L’échographie fœtale cristallise ces enjeux, cumulant un speckle multiplicatif intense et une forte variabilité opérateur-dépendante. Ce *Domain shift* permanent fragilise la robustesse des modèles lors du déploiement sur des systèmes hétérogènes. Que ce soit pour la restauration du mode B, l’imagerie Doppler ou la localisation de microbulles (ULM), la SR ultrasonore nécessite une hybridation entre l’apprentissage profond et les modèles d’acquisition. Ce travail explore cette convergence, avec l’objectif de transformer des systèmes d’imagerie d’entrée de gamme en outils diagnostiques fiables via une optimisation algorithmique adaptée aux contraintes de



ressources.

## Objectifs et Organisation du chapitre

Face à cette complexité, notre objectif est de proposer une cartographie critique et opérationnelle de l'état de l'art. L'architecture suit une logique de focalisation progressive :

1. **Trajectoire historique et conceptuelle** : des approches variationnelles aux modèles profonds contemporains, avec une analyse des hypothèses, des gains et des limites propres à chaque paradigme.
2. **Transfert et adaptation de domaine** : passerelle entre images naturelles et données médicales via fine-tuning, fonctions de coût hybrides et augmentations réalistes ; discussion des compromis robustesse/généralisation.
3. **Étude ciblée de l'échographie** : jeux de données, métriques à pertinence clinique et jalons méthodologiques pour les trois voies de SR (B-mode, Doppler sans contraste, ULM).

Sans prétendre à l'exhaustivité, ce chapitre établit le socle théorique et méthodologique de notre contribution expérimentale : un pipeline de SR adapté à l'échographie fœtale, articulant modèles SOTA et contraintes cliniques.

## 1.2 Fondements théoriques et méthodologiques de la super-résolution

La super-résolution s'énonce formellement comme un problème inverse de restauration d'image, où l'on cherche à estimer une image haute résolution  $I_{HR}$  à partir de son observation dégradée  $I_{LR}$ . La relation fondamentale s'exprime par l'équation de dégradation :

$$I_{LR} = D(I_{HR}) + \eta \quad (1.1)$$

où  $D$  représente l'opérateur de dégradation composite – typiquement une convolution par une fonction de flou (PSF) suivie d'un sous-échantillonnage – et  $\eta$  modélise le bruit additif, souvent supposé gaussien dans les formulations canoniques.

Le caractère *mal posé* de la reconstruction provient de la non-injectivité de  $D$  : une même observation  $I_{LR}$  peut être reproduite par une multiplicité de  $\hat{I}_{HR}$ .

### a) Modèle de dégradation

La super-résolution (SR) consiste à estimer une image haute résolution  $I_{HR}$  à partir d'une observation  $I_{LR}$  affectée par un flou spatial, un sous-échantillonnage et un bruit. Cette dégradation peut être modélisée de manière générale par :

$$I_{LR} = (I_{HR} * k) \downarrow_s + n_{\text{eff}}, \quad (1.2)$$

où  $k$  est le noyau de flou (PSF),  $\downarrow_s$  l'opérateur de décimation, et  $n_{\text{eff}}$  le bruit ou la perturbation aléatoire. Selon la modalité,  $n_{\text{eff}}$  peut être : additif gaussien (imagerie optique), Poissonien (faible flux photonique), Rician (IRM) ou multiplicatif (speckle, échographie), ce dernier pouvant être approximé par un modèle additif après transformation logarithmique.

Deux cadres sont définis : SR non-aveugle dont les paramètres de dégradation sont connus et SR aveugle avec des paramètres inconnus ou partiellement connus. Tout écart entre le modèle de dégradation utilisé à l'entraînement et la dégradation réelle (*degradation mismatch*) entraîne une baisse significative de performance.

La reconstruction de  $I_{HR}$  peut se formaliser par un critère de type MAP :

$$\hat{I}_{HR} = \arg \min_I \mathcal{L}_{\text{data}}(D(I), I_{LR}) + \lambda \mathcal{R}(I), \quad (1.3)$$

où  $\mathcal{L}_{\text{data}}$  quantifie la cohérence avec le modèle de dégradation et  $\mathcal{R}$  encode un *a priori* sur l'image (variation totale, parcimonie, ou prior implicite appris dans les réseaux profonds).

## b) Taxinomie des approches de super-résolution

La super-résolution se décline selon les hypothèses sur l'acquisition et l'objectif de reconstruction.

**SISR.** Le SISR ou *Single Image Super-Resolution* consiste à inférer les détails haute fréquence d'une image à partir d'une unique version dégradée. Les réseaux profonds abordent ce problème mal posé en apprenant un *a priori de contenu* à partir de données massives, garantissant à la fois cohérence structurelle et fidélité visuelle. L'évolution des architectures — des premiers CNN (SRCNN) [45] aux modèles résiduels (EDSR, RCAN)[106, 209]), puis aux approches perceptuelles (SRGAN, ESRGAN)[99, 194]) et aux Transformers (SwinIR[105]) — illustre la maturation du domaine, depuis l'optimisation pixel à pixel jusqu'à la reconstruction perceptuellement cohérente. Une synthèse historique utile est donnée par [141].

**MISR.** Le problème de *Multi-Image Super-Resolution* (MISR) consiste à estimer une image haute résolution à partir d'un ensemble d'observations basse résolution obtenues sous différents points de vue ou décalages sous-pixeliques. Chaque image fournit une contrainte partielle sur la scène originale ; leur alignement et leur fusion permettent d'en reconstituer une estimation plus précise. Cette approche, décrite notamment par [55], repose sur l'hypothèse de redondance entre acquisitions successives et constitue une extension naturelle du cadre SISR [141].

**VSR.** En présence de séquences vidéo, la *Video Super-Resolution* (VSR) exploite la cohérence temporelle entre trames successives afin d'enrichir la reconstruction spatiale. Les modèles spatio-temporels précoces [22] ont été suivis par des architectures à alignement déformable et attentionnelle (EDVR [191]), puis par des réseaux à propagation bidirectionnelle (BasicVSR [24]) qui exploitent efficacement les dépendances à long terme dans le temps.

**3D/Volumique.** En imagerie biomédicale, la forte anisotropie hors-plan — typiquement présente dans les acquisitions IRM ou tomographiques — motive le recours à la super-résolution volumique et à l'interpolation d'inter-coupes. Ces approches visent à restaurer la continuité anatomique dans l'axe axial ou sagittal, initialement par des méthodes statistiques [66, 65], puis par des modèles génératifs tridimensionnels tels que les GAN 3D [156].

## c) Cadre d'évaluation et métriques de performance

L'évaluation des algorithmes de SR nécessite une approche multi-critères reflétant les différentes dimensions de la qualité d'image :

**Jeux de données et transfert inter-domaines** La super-résolution (SR) ne peut être évaluée indépendamment des données qui la fondent. Les modèles profonds apprennent les régularités statistiques des images qui leur sont présentées ; leur capacité de généralisation dépend donc étroitement de la cohérence entre distribution d'entraînement et distribution clinique.

Le développement et la validation des algorithmes s'appuient sur des corpus spécialisés :

■ **Jeux de données naturels :**

- *DIV2K*, *Flickr2K* (entraînement)
- *Set5*, *Set14*, *Urban100* (benchmark)
- Offrent une grande variété et complexité texturale.

■ **Jeux de données médicaux :**

- **IRM** : *IXI*, *BraTS* (tumeurs cérébrales)
- **Échographie** : *CAMUS* (cardiaque), *FPUS* (foetale, in-house)
- **CT** : *LoLa11*, *AAPM*

Les jeux de données naturels demeurent le standard empirique de la super-résolution d’images. Ils se déclinent en ensembles d’*entraînement*, destinés à l’apprentissage supervisé, et en *benchmarks* de référence pour l’évaluation. Ils ont constitué le socle des principales architectures de super-résolution, de *SRCNN* à *SwinIR* [45, 106, 209, 105].

La variété structurelle et la complexité texturale de ces jeux de données ont permis aux réseaux d’apprendre des représentations visuelles généralisables. Néanmoins, les modèles de dégradation qu’ils imposent — flou isotrope, bruit additif gaussien, illumination uniforme — relèvent d’une idéalisation qui ne correspond pas aux phénomènes physiques de l’imagerie médicale. Le Tableau 1.1 en présente les principaux représentants et leur rôle respectif.

TABLEAU 1.1 – Jeux de données naturels utilisés pour l’entraînement et l’évaluation en super-résolution.

| Type         | Dataset  | Rôle principal                         | Particularité                                                                                        |
|--------------|----------|----------------------------------------|------------------------------------------------------------------------------------------------------|
| Entraînement | DIV2K    | Entraînement standard (800 images 2K). | Grande variété naturelle et haute résolution.                                                        |
| Entraînement | Flickr2K | Extension de DIV2K.                    | Scènes issues de Flickr, diversité accrue.                                                           |
| Benchmark    | Set5     | Test historique.                       | 5 images emblématiques : <i>baby</i> , <i>bird</i> , <i>butterfly</i> , <i>head</i> , <i>woman</i> . |
| Benchmark    | Set14    | Évaluation moyenne.                    | 14 images aux textures variées.                                                                      |
| Benchmark    | Urban100 | Benchmark difficile.                   | Scènes urbaines, textures périodiques fines.                                                         |

Les jeux de données biomédicaux quant à eux, présentent des défis distincts, marqués par leur volume limité, leur complexité physique et leur hétérogénéité acquisitionnelle. Chaque modalité impose ses spécificités : les volumes IRM multicontrastes de *IXI* et *BraTS* [125], les données tomodensitométriques à faible dose de *LoLa11* et *AAPM* [123], ou les séquences échographiques de *CAMUS* [98] et d’un jeu *in-house* foetal (*FPUS*) illustrent la diversité des enjeux.

TABLEAU 1.2 – Jeux de données biomédicaux de référence pour la super-résolution d’images.

| Domaine     | Dataset       | Type d’imagerie               | Particularité                                                     | Utilisation en SR                                          |
|-------------|---------------|-------------------------------|-------------------------------------------------------------------|------------------------------------------------------------|
| IRM         | <i>IXI</i>    | Cerveau sain, multi-contraste | Données publiques, variabilité anatomique naturelle               | SR volumique et interpolation inter-coupes                 |
| IRM         | <i>BraTS</i>  | Cerveaux pathologiques        | Tumeurs cérébrales, annotations expertes                          | Reconstruction fidèle des structures tumorales             |
| Échographie | <i>CAMUS</i>  | Cardiaque                     | Données cliniques annotées, bruit de speckle réaliste             | SR robuste au bruit et maintien de la cohérence anatomique |
| Échographie | <i>FPUS</i>   | Foetale (in-house)            | Données propriétaires, acquisitions bruitées et non standardisées | SR en contexte obstétrical bruité                          |
| CT          | <i>LoLa11</i> | Thorax faible dose            | Concours international (Low-Dose CT Grand Challenge 2011)         | SR pour reconstruction réaliste à dose réduite             |
| CT          | <i>AAPM</i>   | Thorax faible dose            | Benchmark AAPM 2016, paires basse/haute dose                      | SR et débruitage multi-dose                                |

Les signatures bruitées propres à chaque modalité — notamment le speckle en échographie et le bruit ricien en IRM —, associées à l’anisotropie de résolution et à la variabilité induite par

l'opérateur, imposent une hétérogénéité intrinsèque qui complique la convergence et la généralisation des approches d'apprentissage profond. L'écart statistique entre images naturelles et médicales justifie le recours à des stratégies d'adaptation spécifiques. L'*adaptation de domaine* et le *fine-tuning* cherchent à aligner les distributions tout en respectant la physique du signal [210]. La modélisation réaliste des dégradations [207] et l'intégration des connaissances d'acquisition offrent des pistes prometteuses pour un transfert plus fiable vers le domaine clinique.

Le progrès ne dépend plus seulement du volume de données, mais de leur pertinence. Des jeux représentatifs et des modèles physiquement cohérents constituent la clé d'une super-résolution médicalement crédible.

Les principaux ensembles utilisés dans la littérature, regroupés dans le Tableau 1.2, illustrent la diversité des contextes d'imagerie — de l'IRM cérébrale à la tomодensitométrie thoracique — et les défis propres à chaque modalité.

La disparité de taille, de qualité d'annotation et de caractéristiques modales entre ces jeux de données souligne la nécessité de stratégies d'adaptation de domaine pour le transfert des modèles entraînés sur des images naturelles vers des applications médicales.

**Cadre d'évaluation et métriques de performance** L'évaluation des performances doit ainsi combiner plusieurs dimensions de la qualité d'image, comme résumé dans le Tableau 1.3.

TABLEAU 1.3 – Taxinomie des métriques d'évaluation en super-résolution

| Catégorie                   | Métriques                 | Portée et limitations                                                                    |
|-----------------------------|---------------------------|------------------------------------------------------------------------------------------|
| <b>Fidélité pixel</b>       | PSNR, SSIM                | Mesures référencées sensibles aux distorsions, peu corrélées à la perception             |
| <b>Qualité perceptuelle</b> | LPIPS, FID                | Métriques apprises reflétant la similarité dans des espaces de caractéristiques profonds |
| <b>Sans référence</b>       | NIQE, BRISQUE             | Estiment la naturalité sans vérité terrain, utiles pour les cas réels                    |
| <b>Validation clinique</b>  | MOS, scores diagnostiques | Évaluation par experts, critère ultime mais coûteux et subjectif                         |

## 1.3 Évolution chronologique des paradigmes en super-résolution

L'histoire de la super-résolution illustre le passage progressif de la modélisation analytique à l'apprentissage profond, puis à la génération probabiliste. Des approches fondées sur des *priors explicites* aux architectures génératives contemporaines, chaque paradigme a redéfini la manière de concilier fidélité métrique et réalisme perceptuel dans la reconstruction d'images. La Figure 1.1 en propose une synthèse chronologique.

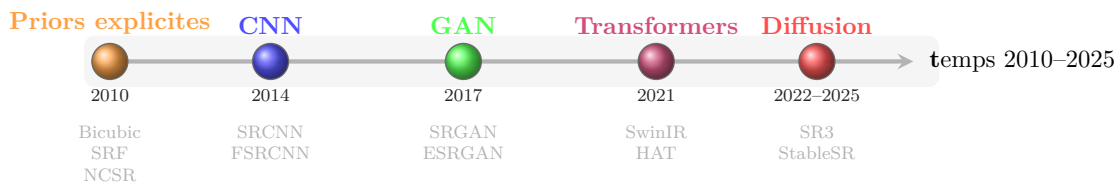


FIGURE 1.1 – Frise chronologique des paradigmes de la super-résolution, de la régularisation par priors explicites aux modèles de diffusion génératifs.

## a) L'ère pré-deep learning : régularisation par priors explicites

Avant la généralisation du deep learning, la super-résolution était abordée comme un problème inverse classique, où l'image observée  $I_{LR}$  résultait de l'action d'un opérateur de dégradation  $D$  appliqué à l'image originale  $I_{HR}$ . Cet opérateur modélisait la réponse instrumentale du système, associant flou du capteur, sous-échantillonnage spatial et bruit additif.

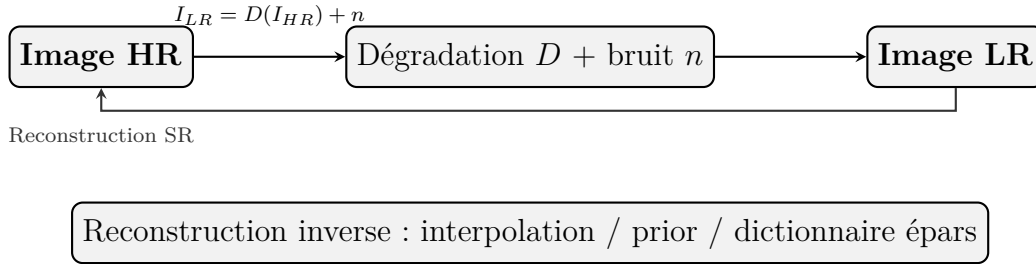


FIGURE 1.2 – Schéma conceptuel du modèle de dégradation et du problème inverse en super-résolution avant l'ère du deep learning.

Pour inverser cette transformation, les méthodes reposaient sur des *priors* analytiques visant à imposer la régularité du signal, tels que la variation totale ou la parcimonie dans une base d'ondelettes. Comme le montre la Figure 1.2, les approches initiales — de l'interpolation classique aux méthodes par patches [58] — ont incarné cette période “pré-apprenante”, marquée par une tension entre stabilité numérique et perte perceptuelle.

La figure 1.3(a) illustre les méthodes d'interpolation classiques — bilinéaire [84], bicubique [127] et Lanczos [50] — qui permettent de reconstruire les pixels manquants à partir des voisins immédiats. La bilinéaire prend la moyenne des quatre pixels voisins, la bicubique utilise seize pixels et une fonction cubique pour adoucir les contours, et Lanczos recourt à une fonction de sinc tronquée pour améliorer la préservation des détails et limiter les artefacts d'aliasing [141].

Une avancée majeure en super-résolution a été introduite par les méthodes à base de patches [58], qui exploitent la redondance locale pour restaurer les détails manquants (figure 1.3(b)). Chaque patch de l'image basse résolution est recherché dans une base d'exemples haute résolution et remplacé par le patch le plus similaire, permettant de conserver textures et structures orientées. Cette méthode reste néanmoins limitée par la qualité et la diversité de la base d'exemples.

1. Bicubic = interpolation bicubique ; SRF = *Super-Resolution Forests* [160] ; NCSR = *Nonlocally Centralized Sparse Representation* [48].

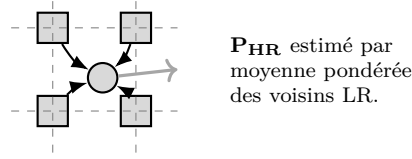
2. SRCNN = *Super-Resolution CNN* [45] ; FSRCNN = *Fast SRCNN* [47].

3. SRGAN = *Super-Resolution GAN* [99] ; ESRGAN = *Enhanced SRGAN* [194].

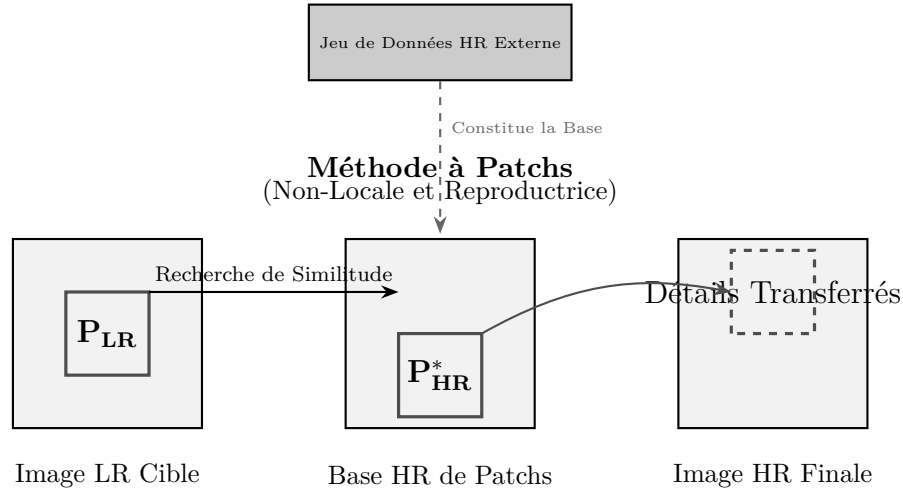
4. SwinIR = *Swin Transformer for Image Restoration* [105] ; HAT = *Hybrid Attention Transformer* [29].

5. SR3 = *Super-Resolution via Repeated Refinement* [155] ; StableSR = modèle de diffusion pour la SR (2023–2025).

**Méthode d'Interpolation**  
(Locale et Réplicative - Bilineaire, Bicubique, etc.)



(a) Pixel  $\mathbf{P}_{\text{HR}}$  estimé à partir des pixels LR voisins immédiats (Interpolation classique).



(b) Patch  $\mathbf{P}_{\text{LR}}$  reconstruit avec les détails d'un patch  $\mathbf{P}_{\text{HR}}^*$  trouvé dans une base HR externe.

FIGURE 1.3 – Schémas comparatifs des mécanismes de super-résolution : Approche Locale (Interpolation) et Approche Non-Locale (Patches).

Les interpolations classiques, rapides et faciles à mettre en œuvre, lissent les contours et suppriment les détails fins. Même avec les patches, il est impossible de générer de nouvelles structures, et la qualité perceptuelle globale reste ainsi restreinte.

À partir de 2010, les modèles de dictionnaires épars [203] établirent un nouveau standard. En supposant que chaque patch d'image pouvait être représenté par une combinaison parcimonieuse d'atomes dans un dictionnaire appris, ces approches reliaient de façon explicite les représentations basse et haute résolution. Elles offraient un cadre théorique élégant fondé sur la régularisation  $\ell_1$ , mais restaient limitées par leur capacité à capturer la complexité statistique des textures naturelles. Les reconstructions produites tendaient souvent à être sur-lissées et manquaient de réalisme perceptuel, soulignant la difficulté d'imposer un prior universellement valable pour des données riches et hétérogènes.

Dans cette première phase de développement, l'évolution des méthodes de super-résolution peut être retracée en deux grandes familles successives :

1. **Les approches purement interpolatives**, fondées sur des hypothèses de continuité locale du signal,
2. **Et les méthodes à *priors explicites***, qui cherchent à modéliser la structure statistique de l'image par des contraintes analytiques ou des modèles probabilistes.

Le tableau 1.4 récapitule les principales contributions de cette période et met en lumière la progression graduelle vers des formes de régularisation de plus en plus apprenantes.

TABLEAU 1.4 – Évolution des méthodes pré-deep learning pour la super-résolution

| Année | Travail / Méthode                                                     | Apport principal et évolution                                                                             |
|-------|-----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| 1979  | Interpolation Lanczos [50]                                            | Noyau sinc fenêtré : meilleure restitution des hautes fréquences mais artefacts possibles.                |
| 1981  | Interpolation bicubique [84]                                          | Introduit des dérivées continues pour mieux préserver les contours — devient la référence historique.     |
| 1980s | Interpolation bilinéaire [62]                                         | Première méthode numérique simple ; moyenne pondérée des 4 voisins — rapide mais lissage excessif.        |
| 2001  | Li et al., <i>Interpolation dirigée par les contours (NEDI)</i> [102] | Première interpolation adaptative orientée structure ; exploite les corrélations directionnelles locales. |
| 2002  | Freeman et al., <i>Example-Based SR</i> [58]                          | Apprentissage d’une base d’exemples LR–HR ; amorce le paradigme statistique.                              |
| 2004  | Chang et al., <i>Neighbor Embedding SR</i> [25]                       | Projection locale dans un espace de voisinage ; améliore la cohérence spatiale.                           |
| 2004  | Farsiu et al., <i>MAP-MRF SR</i> [55]                                 | Formule bayésienne multi-images intégrant un prior de Markov pour la stabilité.                           |
| 2009  | Glasner et al., <i>Internal Patch Recurrence</i> [61]                 | Exploite l’auto-similarité interne de l’image ; SR auto-supervisée sans base externe.                     |
| 2010  | Yang et al., <i>Sparse Representation SR</i> [203]                    | Dictionnaires couplés LR/HR et régularisation $\ell_1$ ; apprentissage parcimonieux.                      |
| 2012  | Katsuki et al., <i>Bayesian SR with Gaussian MRF Prior</i> [82]       | Prior MRF gaussien causal pour renforcer cohérence structurelle et réalisme des textures.                 |

## b) 2014–2016 : émergence des réseaux convolutifs profonds

Les approches à *priors explicites* offraient un cadre analytique solide, mais peinaient à capturer la complexité et les corrélations fines des images naturelles. Le deep learning a renversé cette logique : en apprenant directement des “priors implicites” à partir de données, les réseaux convolutifs ont transformé la super-résolution en un problème d’apprentissage supervisé.

### SRCNN ou *Super-Resolution Convolutional Neural Network* [45]

Le modèle SRCNN inaugure le paradigme des CNN pour la super-résolution. Son principe est simple : apprendre la fonction de mappage entre images basse résolution (LR) et haute résolution (HR) directement à partir des données, sans hypothèse analytique.

Son architecture comprend trois couches convolutives *end-to-end* :

- **Extraction de caractéristiques** ( $9 \times 9$  Conv + ReLU) : capture des motifs locaux et des textures.
- **Mappage non linéaire** ( $5 \times 5$  Conv + ReLU) : apprend la correspondance LR  $\rightarrow$  HR.
- **Reconstruction** ( $5 \times 5$  Conv) : combine les caractéristiques afin de générer l’image HR finale.

L’apprentissage repose sur la perte MSE, favorisant la fidélité pixel-à-pixel mais pouvant entraîner un lissage perceptuel.

Les limites intrinsèques du SRCNN incluent :

- une architecture peu profonde, limitant la capacité d’extraction de détails complexes ;
- une dépendance à l’interpolation bicubique initiale, susceptible d’altérer l’information fine ;
- des reconstructions lissées, parfois dépourvues de microstructures perceptuellement réalistes.

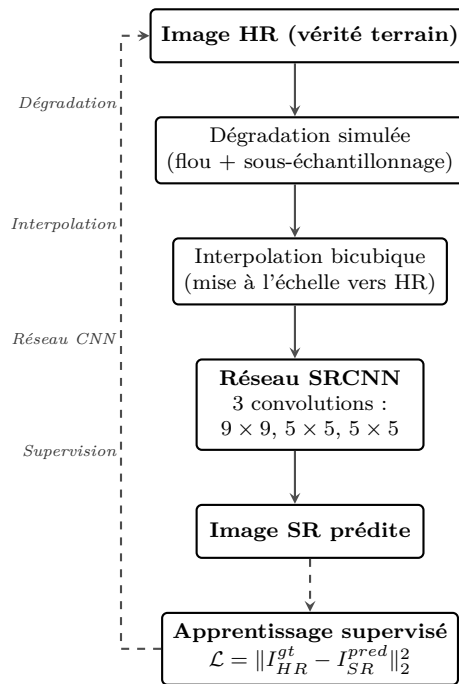


FIGURE 1.4 – Pipeline du modèle de super-résolution : SRCNN [45]. *L'image basse résolution est d'abord interpolée par bicubique, puis passée dans un réseau CNN à trois couches pour produire l'image super-résolue. L'apprentissage se fait par supervision MSE.*

— **Évolutions ultérieures** Les travaux suivants, tels que VDSR [89] ou EDSR [106], ont amélioré la vitesse, la profondeur et la stabilité des réseaux, mais ces évolutions restent principalement incrémentales par rapport au SRCNN : elles n'inversent pas les limites fondamentales liées à la supervision MSE et au lissage perceptuel.

— **Limites cliniques des CNN** En imagerie médicale, les CNN entraînés avec MSE présentent des limitations spécifiques :

- Lissage des textures fines, pouvant masquer des microstructures diagnostiquement pertinentes
- Dépendance à la vérité terrain LR-HR, limitant la généralisation à des acquisitions réelles
- Hallucination faible, mais risque de perte de détails subtils essentiels pour le diagnostic
- Limites des métriques classiques (PSNR, SSIM) pour évaluer la qualité perceptuelle, comme le souligne Kapo et al. [81] : des valeurs élevées ne garantissent pas toujours une bonne qualité visuelle des images

Cette analyse justifie le passage aux paradigmes génératifs (GANs), capables d'améliorer le réalisme perceptuel tout en conservant la cohérence diagnostique.

### c) GANs et approches perceptuelles (2017)

Dès 2017, une nouvelle orientation émerge : les modèles *génératifs* et *perceptuels*. Inspirés par les réseaux adversariaux (GANs) et les pertes perceptuelles basées sur des caractéristiques profondes, ces modèles ne se contentent plus de reconstruire les pixels : ils visent à restituer la réalité perçue, plaçant la qualité visuelle au centre de l'optimisation.

Le modèle SRGAN [99] constitue le socle de cette lignée d'architectures, marquant la première application des réseaux antagonistes génératifs à la super-résolution d'images. L'idée clé est d'apprendre à tromper un discriminateur afin de produire des images indiscernables de vraies images, plutôt que de simplement minimiser l'erreur pixel à pixel.



Le générateur et le discriminateur sont entraînés conjointement selon la fonction de coût composite :

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{contenu}} + \lambda_{\text{adv}} \mathcal{L}_{\text{adversaire}}, \quad (1.4)$$

où  $\mathcal{L}_{\text{contenu}}$  assure la cohérence perceptuelle via un réseau VGG et  $\mathcal{L}_{\text{adversaire}}$  encourage le photoréalisme des résultats.

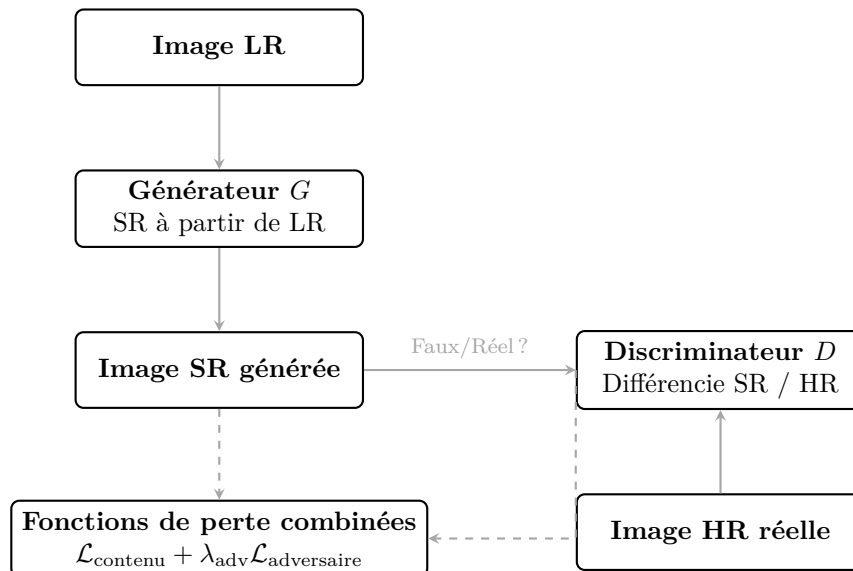


FIGURE 1.5 – Schéma fonctionnel du modèle **SRGAN** [99], combinant apprentissage perceptuel et adversarial.

— **Évolutions postérieures** SRGAN a constitué le point de départ des méthodes perceptuelles, bien que limité par une certaine instabilité et un manque de fidélité texturale. Face aux limites du modèle SRGAN, l'ESRGAN [194] propose une solution plus robuste : le bloc RRDB.

Le mécanisme « Residual-in-Residual » stabilise l'entraînement et capture des détails fins, produisant des images de haute netteté, indispensables pour la rigueur clinique. Cette architecture résiduelle à plusieurs niveaux remplace la normalisation par lots par une structure de connexions denses imbriquées, permettant de capturer des caractéristiques plus profondes sans lisser les détails.

En complément de cette avancée structurelle, ESRGAN intègre un discriminateur relativiste, qui évalue si une image est plus réaliste que la moyenne des images réelles, plutôt que de porter un jugement absolu. On obtient ainsi des textures plus riches et des détails bien mieux définis.

Enfin, le modèle Real-ESRGAN [192] pousse cette logique de généralisation encore plus loin en utilisant un processus de dégradation à plusieurs étapes (high-order degradation). En combinant un entraînement synthétique évolué et un affinement sur des données réelles, il permet d'étendre la super-résolution à une plus vaste diversité de clichés, garantissant une meilleure consistance visuelle quel que soit le support d'origine.

— **Applications médicales** Les modèles SRGAN améliorent significativement la résolution des CT scan, IRM et échographies, révélant des détails fins indispensables aux analyses cliniques. Leur utilité est maximale en radioprotection (basse dose) et en santé numérique, où ils transforment des acquisitions dégradées en supports décisionnels robustes.

— **Limites** Le risque majeur réside dans la génération d'artefacts synthétiques pouvant tromper le diagnostic. L'efficacité du modèle restant tributaire des jeux de données sources, l'insuf-

fisance des métriques classiques impose de coupler l'évaluation quantitative à une expertise humaine pour garantir l'intégrité des images produites.

## d) Transformers (2021–2025)

Succédant à la révolution perceptuelle des GANs, les Transformers ont redéfini l'état de l'art en super-résolution en remplaçant la localité des convolutions par la globalité de l'attention. Là où les CNNs excellent dans la reconstruction des textures fines mais échouent à maintenir une cohérence structurelle sur de vastes régions, les mécanismes d'auto-attention permettent d'intégrer des dépendances à longue portée.

L'application de modèles comme le SwinIR a démontré qu'une modélisation contextuelle hybride — alliant la précision locale et la vision globale — surpasse les méthodes traditionnelles.

L'évolution récente (2021–2025) montre une tendance vers la réduction de la complexité quadratique de l'attention pour rendre ces modèles plus agiles.

— **Évolution des architectures (2021–2025)** Le passage aux Transformers marque une transition de la reconstruction locale vers la modélisation contextuelle globale. Trois modèles illustrent cette progression :

- IPT [27] : Précurseur exploitant une auto-attention globale sur patchs pour assurer la cohérence structurelle. (Voir Figure 1.6)
- SwinIR [105] : Introduit une attention hiérarchique par fenêtres décalées (Shifted Windows), optimisant le compromis entre performance et coût computationnel.
- HAT [29] : Hybride l'attention spatiale et de canal via des fenêtres chevauchantes (OCAB), mobilisant plus de pixels pour stabiliser les hautes fréquences.

— **Applications et limites en santé** Malgré leur expressivité, l'intégration des Transformers en imagerie médicale se heurte à des limites critiques :

- Complexité computationnelle : La croissance quadratique de l'attention limite le traitement de volumes 3D haute résolution (IRM/CT) et l'usage en temps réel (échographie).
- Risque diagnostique : Moins contraints par la localité que les CNN, ces modèles peuvent générer des hallucinations structurelles (artéfacts anatomiques plausibles mais erronés), induisant des risques de faux diagnostics.
- Biais inductif et données : Leur faible biais structurel natif exige des jeux de données massifs, souvent rares en environnement clinique.
- Évaluation clinique : La corrélation limitée entre les métriques mathématiques (PSNR/SSIM) et la réalité pathologique souligne la nécessité de validations qualitatives expertes.

## e) Modèles de diffusion (2022–2025)

Contrairement aux méthodes directes, les modèles de diffusion reconstruisent l'image haute résolution par touches successives. Ils partent d'un état aléatoire et retirent progressivement le bruit pour faire apparaître les détails, en s'appuyant sur l'image basse résolution comme guide pour garantir la fidélité du contenu.

- SR3 [155] : Architecture pionnière utilisant un U-Net pour prédire itérativement le bruit à retirer, permettant une restitution exceptionnelle des textures fines par cascades d'agrandissement.
- Hero-SR [189] : Version optimisée visant à réduire la latence par compression du nombre d'étapes de débruitage, facilitant l'usage en conditions réelles.

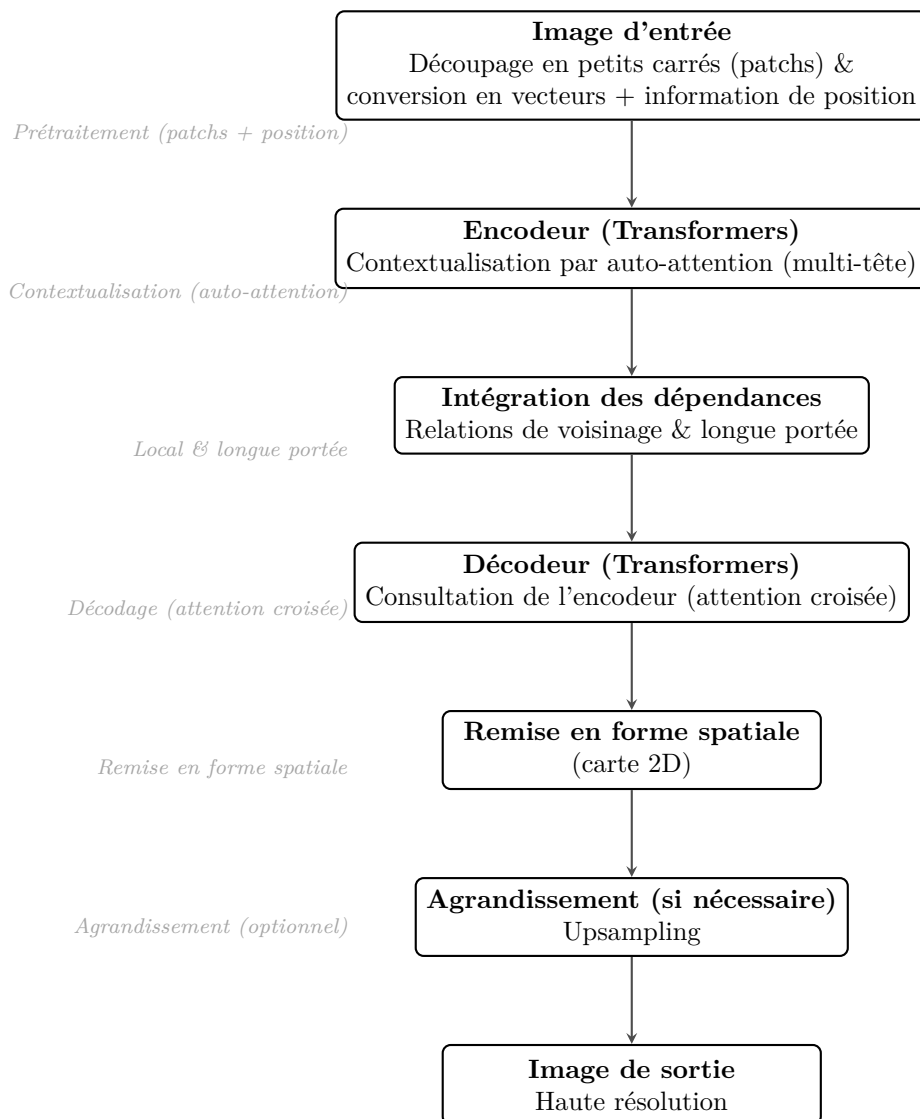


FIGURE 1.6 – Pipeline fonctionnel du modèle **IPT** [27]. *Détails des modules : (1) Encodage — découpage de l'image en patches et transformation en vecteurs de caractéristiques (embeddings) ; (2) Représentation partagée — blocs Transformer empilés avec auto-attention multi-têtes pour capturer les dépendances globales ; (3) Décodage — reconstruction de l'image haute résolution par décodeur symétrique ; (4) Supervision — apprentissage multi-tâche puis transfert fin sur la super-résolution.*

— **Potentiel et limites médicales** En imagerie diagnostique (IRM, CT), la diffusion excelle dans la restauration de la continuité tissulaire et des structures complexes. Toutefois, trois obstacles majeurs subsistent :

- **Lourdeur computationnelle** : Le caractère itératif ralentit la reconstruction, limitant les usages en temps réel.
- **Fidélité clinique** : La puissance générative induit un risque d'hallucinations (création de détails anatomiques erronés).
- **Évaluation** : L'obsolescence des métriques PSNR/SSIM pour juger de la pertinence diagnostique impose une validation par expertise humaine.

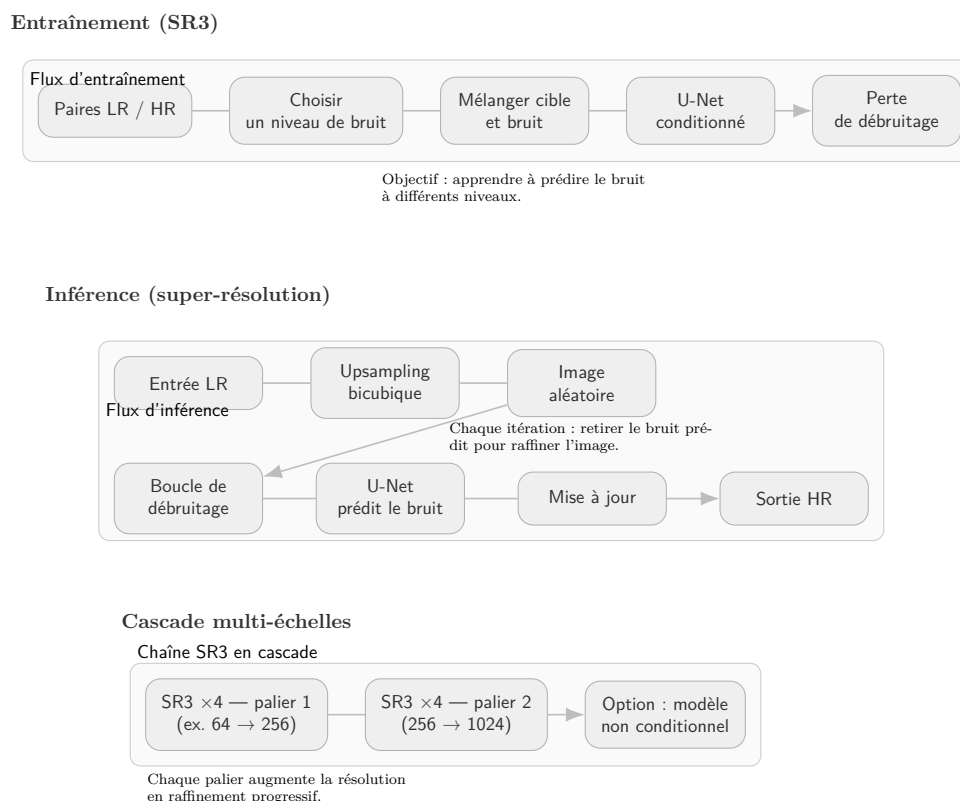


FIGURE 1.7 – Pipeline du modèle SR3 : entraînement, inférence et cascade multi-échelles.

### 1.3.1 Bilan des paradigmes SR

La synthèse du Tableau 1.5 met en évidence l'évolution des approches en imagerie médicale, illustrant une transition progressive des méthodes conventionnelles vers des paradigmes fondés sur l'intelligence artificielle profonde.

L'arrivée des réseaux convolutionnels (CNN), avec des architectures pionnières comme SRCNN ou EDSR, a d'abord permis de stabiliser les gains en fidélité métrique (PSNR/SSIM). Or, ces modèles homogénéisent les détails, au détriment du réalisme textural.

Pour pallier ce manque de détails, les approches perceptuelles (basées sur des pertes de contenu) et adversariales (SRGAN, ESRGAN) ont déplacé le curseur vers le réalisme visuel. Si ces dernières excellent à reconstruire des textures plausibles, elles introduisent parfois des artefacts ou une instabilité d'entraînement qui compliquent leur déploiement en milieu clinique.

Aujourd'hui, le domaine vit une nouvelle transition avec les Transformers (comme SwinIR) et les modèles de diffusion. Ces architectures ne se contentent plus d'un traitement local : elles capturent des dépendances contextuelles globales, offrant une cohérence structurelle inédite. Néanmoins, ce saut qualitatif impose un défi de taille : le coût computationnel et la fiabilité des détails générés.

En imagerie médicale, où toute erreur peut avoir des conséquences cliniques, l'enjeu ne se limite pas à l'apparence visuelle de l'image, mais à sa fidélité et à son authenticité diagnostique.

Si le réalisme progresse, l'authenticité diagnostique demeure conditionnée par les exigences de la modalité ultrasonore.

Dans ce contexte, la section suivante propose une analyse structurée de la littérature récente (2020–2025) consacrée à la super-résolution en imagerie médicale, avec un focus particulier sur l'échographie, afin d'identifier les tendances méthodologiques dominantes et les limites persistantes de l'état de l'art. La maturité algorithmique du traitement d'images n'assure pas encore son adoption en milieu médical. Contrairement aux images naturelles, les données médicales sont acquises dans des conditions fortement contraintes et répondent à des exigences diagnostiques

TABLEAU 1.5 – Synthèse transversale des paradigmes de super-résolution : objectifs, atouts, limites et modèles représentatifs.

| Paradigme                               | Objectif principal                                                  | Forces majeures                                                                               | Limites identifiées                                                                                                      | Modèles phares                           |
|-----------------------------------------|---------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| <b>Priors explicites</b> (pré-DL)       | Formulation analytique du problème inverse                          | Rigueur mathématique, interprétabilité, stabilité numérique                                   | Incapacité à modéliser la complexité statistique des textures naturelles ; forte dépendance aux hypothèses de régularité | MAP-MRF [55], NCSR [48], Sparse SR [203] |
| <b>CNNs</b> (2014–2020)                 | Apprentissage direct du mapping LR→HR                               | Haute fidélité métrique, rapidité d'inférence, supervision explicite                          | Lissage perceptuel dû à la perte MSE, champ réceptif limité, difficulté à restituer les microstructures                  | SRCNN, VDSR, EDSR, RCAN, RDN             |
| <b>GANs et perceptuels</b> (2017–2021)  | Optimisation orientée perception visuelle                           | Réalisme des textures, qualité perceptuelle élevée, apprentissage multi-objectifs             | Instabilité de l'entraînement, biais perceptuels, difficulté à garantir la fidélité photométrique                        | SRGAN, ESRGAN, GAN, Real-ESRGAN          |
| <b>Transformers</b> (2021–2025)         | Capture des dépendances globales et locales par attention           | Modélisation contextuelle à grande échelle, cohérence structurelle, adaptabilité multi-tâches | Coût mémoire élevé, complexité quadratique, dépendance à des données massives                                            | IPT, SwinIR, HAT, SRFormerV2             |
| <b>Modèles de diffusion</b> (2022–2025) | Génération stochastique de détails fins par raffinements successifs | Photo-réalisme, robustesse aux dégradations complexes, diversité des reconstructions          | Temps d'inférence long, coût computationnel élevé, difficulté de contrôle du contenu                                     | SR3, StableSR, Hero-SR                   |

strictes, où toute altération de la structure peut affecter l'interprétation clinique. L'imagerie ultrasonore, par sa nature bruitée et ses artefacts complexes, constitue un cas d'étude critique pour évaluer la fiabilité de ces méthodes.

## 1.4 Analyse Systématique de la Littérature Médicale

Dans le prolongement des paradigmes que nous venons d'exposer, cette section examine comment la super-résolution s'adapte concrètement aux contraintes du diagnostic médical. Nous portons une attention particulière aux travaux publiés entre 2020 et 2025, période où les modèles se sont spécialisés pour traiter des modalités complexes, notamment l'échographie.

Nous synthétisons ici l'état de l'art de la super-résolution ultrasonore en dégagant les grandes tendances méthodologiques. L'hétérogénéité des publications a conduit automatiquement à limiter l'analyse aux travaux les plus pertinents pour le diagnostic.

### 1) Méthodologie d'analyse

Face à la prolifération rapide des contributions scientifiques et à l'hétérogénéité des modalités, des architectures et des protocoles d'évaluation, une analyse narrative seule ne permet plus d'identifier clairement les tendances structurantes et les lacunes méthodologiques du domaine. Conformément aux lignes directrices PRISMA 2020, une revue systématique a été adoptée afin d'assurer une analyse rigoureuse, traçable et reproductible.

Nous avons ciblé la littérature de 2020 à 2025 pour identifier les approches de super-résolution dédiées à l'imagerie ultrasonore. Cette sélection s'articule autour de deux ensembles complémentaires : — Un corpus de **23 articles pionniers**, spécifiquement dédiés à la super-résolution en ultrason. — Un corpus de **23 revues de littérature** fournissant le contexte

scientifique plus large, avec une attention marquée pour les techniques ultrasonores.

**Procédure de recherche et critères PRISMA 2020.** La revue de la littérature a été conduite en suivant les recommandations PRISMA 2020, à partir des principales bases de données internationales. Nous avons restreint l’analyse principale à la période 2020–2025 pour capturer l’émergence des architectures de pointe en super-résolution médicale, tout en intégrant les travaux fondateurs de la période 2016–2019 afin de retracer la trajectoire méthodologique ayant conduit aux architectures modernes.

La stratégie de recherche reposait sur l’intersection des mots tels que super-résolution, *deep learning* et modalités médicales. L’inclusion d’*arXiv* répond au besoin d’intégrer l’état de l’art immédiat, par-delà les délais de publication classiques. Les critères de rétention privilégient les modèles de super-résolution neuronale ayant fait l’objet d’une évaluation quantitative rigoureuse, condition sine qua non de notre analyse. Ont été exclus les articles sans rigueur expérimentale ou déconnectés des enjeux clinique. Lorsque le texte intégral n’était pas accessible, l’évaluation s’est appuyée sur l’*abstract*, conformément aux usages pour les revues à visée technologique.

Le processus global de recherche, filtrage et catégorisation des publications est illustré par le diagramme 1.8.

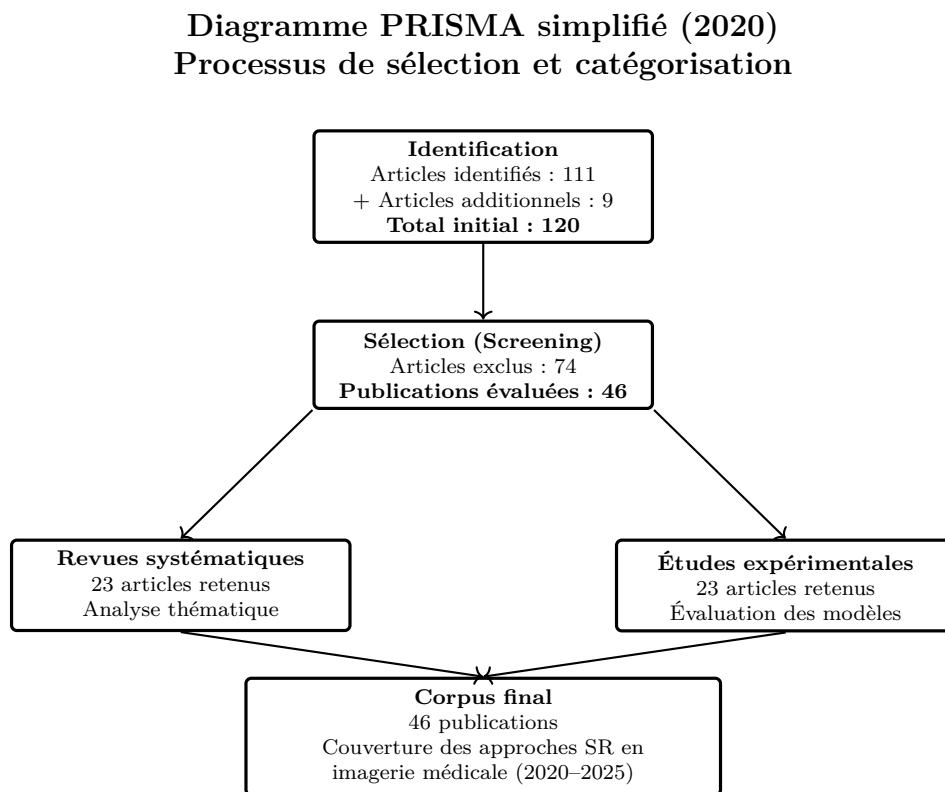


FIGURE 1.8 – Diagramme de sélection des études selon PRISMA (2020), distinguant revues systématiques et études expérimentales.

Le diagramme illustre la complémentarité entre revues systématiques et études expérimentales, respectivement dédiées à la synthèse conceptuelle et à la validation pratique des approches de super-résolution. Il retrace le cheminement critique du corpus — de l’identification initiale à la catégorisation finale — conformément aux exigences méthodologiques du protocole PRISMA (2020).

## 2) Évolution chronologique des approches en imagerie médicale

L'évolution des approches de super-résolution en imagerie médicale suit une dynamique particulière : la grande majorité des architectures fondatrices ont d'abord été conçues et validées pour les images naturelles en vision par ordinateur, avant d'être adoptées et adaptées par la communauté médicale. Ce schéma de transfert — de la vision générale vers l'imagerie clinique — constitue le fil conducteur de l'évolution présentée ci-après, rythmée par l'apparition successive des réseaux convolutionnels (CNN), des réseaux antagonistes génératifs (GAN), puis des modèles à base de transformers.

Les premières architectures à diffuser vers l'imagerie médicale sont issues du paradigme CNN et des réseaux résiduels profonds. Dès 2014, SRCNN [45] établit qu'un réseau convolutionnel peut apprendre directement la correspondance entre images basse et haute résolution, surpassant les méthodes d'interpolation classiques. Depuis, ce domaine a connu une évolution rapide et continue, avec des architectures de plus en plus profondes, des mécanismes d'apprentissage résiduel, puis l'introduction des réseaux antagonistes génératifs et des modèles à base de transformers, chacun apporté d'abord par la vision par ordinateur avant d'être progressivement adopté par la communauté médicale.

Dans le cadre de cette revue, une attention particulière est portée aux travaux publiés entre 2020 et 2025, période marquée par une forte diversification des approches de super-résolution appliquées à l'imagerie médicale, notamment en échographie. Ces travaux intègrent des problématiques telles que la gestion du bruit speckle, la super-résolution aveugle, l'apprentissage auto-supervisé ou encore les approches hybrides combinant super-résolution et segmentation. Le tableau 1.6 recense 23 travaux représentatifs de cette évolution, montrant comment les avancées algorithmiques ont progressivement amélioré la résolution spatiale, la fidélité anatomique et l'efficacité computationnelle des modèles de super-résolution médicale.

Cette chronologie démontre une accélération notable des innovations, avec une convergence récente vers les architectures hybrides et les transformers. La période 2016-2019 fut dominée par les CNN et GAN fondamentaux, tandis que les années 2020-2024 ont vu l'émergence de modèles spécialisés intégrant des contraintes physiques et des mécanismes d'attention avancés, particulièrement bénéfiques pour l'imagerie ultrasonore où la préservation des microstructures est cruciale.

**Lecture de cette trajectoire** – L'analyse de cette chronologie fait apparaître une accélération remarquable des innovations, scandée par quatre moments méthodologiques :

- 2016-2019 : Ère des fondations – CNN et GANs établissent le socle de l'apprentissage profond appliqué
- 2020-2021 : Spécialisation modalitaire – adaptation des architectures génériques aux contraintes médicales
- 2022-2023 : Révolution attentionnelle – émergence des Transformers et hybridations sophistiquées
- 2024 : Préfiguration diffusion – premiers modèles probabilistes à très haute fidélité

Cette progression n'est pas seulement technique ; elle traduit une maturation épistémologique : nous sommes passés d'une optimisation de la similarité pixelique à une recherche de la cohérence anatomique, puis à la modélisation de la vraisemblance physiologique.

Si l'analyse chronologique met en lumière la dynamique d'innovation algorithmique, elle ne permet toutefois pas, à elle seule, d'identifier les invariants conceptuels et les consensus méthodologiques qui se dégagent à l'échelle du domaine. Au-delà des performances individuelles, l'examen des revues de littérature permet d'extraire les paradigmes dominants — du CNN pur aux architectures hybrides — structurant l'état de l'art actuel.

TABLEAU 1.6 – Évolution chronologique des approches de super-résolution et applications associées en imagerie médicale (2020–2025)

| Année | Auteurs                 | Architecture                                           | Modalités             | Contribution principale                                 | Métriques                                                      |
|-------|-------------------------|--------------------------------------------------------|-----------------------|---------------------------------------------------------|----------------------------------------------------------------|
| 2020  | Yang et al. [204]       | Res2Net + SRCNN; défloutage + SR                       | Écho-endoscopie (EUS) | Défloutage + SR EUS                                     | PSNR 25.98; SSIM 0.88                                          |
|       | Temiz et al. [179]      | CNN léger (DECUSR, 39k)                                | Ultrasons B-mode      | SR ultrasonore mobile                                   | SR $\times 2$                                                  |
|       | Kim et al. [88]         | VDSR                                                   | Ultrasons             | Adaptation VDSR US                                      | SR $\times 2$                                                  |
|       | Mostafiz et al. [131]   | SRCNN + VGG19 + GWT/LBP + SVM                          | Échographie hépatique | SR pour classification                                  | Acc. 98.40%                                                    |
| 2021  | Liu et al. [113]        | CycleGAN-SR                                            | Ultrasons             | Super-résolution appariée basée sur CycleGAN            | non PSNR : 32.49–35.22 dB; SSIM : 0.876–0.919                  |
|       | Ding et al. [41]        | Zero-shot CycleGAN                                     | Thyroïde / ultrasons  | Super-résolution sans dataset d'entraînement            | PSNR 40.76; SSIM 0.99                                          |
| 2022  | Liu et al. [112]        | Apprentissage résiduel progressif + estimation du flou | Ultrasons B-mode      | SR aveugle + dégradation                                | PSNR 33.84; SSIM 0.898                                         |
|       | Ahmad et al. [4]        | GAN multi-branche                                      | Ultrasons             | Amélioration des textures                               | SR $\times 2$                                                  |
| 2023  | Wang et al. [202]       | SwinGAN                                                | IRM                   | Transformer GAN SR                                      | SSIM 0.92–0.97                                                 |
|       | Cammarasana et al. [23] | WDSR adapté au bruit speckle                           | Ultrasons             | SR + préservation du speckle                            | PSNR 23.8–24.2                                                 |
|       | Zhu et al. [215]        | MIASSR                                                 | IRM / CT              | GAN de super-résolution basé sur le méta-apprentissage  | PSNR 30.9–43.2                                                 |
| 2024  | Du et al. [49]          | T-GAN (Transformer + GAN)                              | IRM / CT              | SR 4 $\times$ avec transfert de texture via Transformer | PSNR/SSIM : 35.26 dB/0.9496 (genou); 34.69 dB/0.9353 (abdomen) |
|       | Singh et al. [166]      | Sub-RENet                                              | Ultrasons             | SR multi-échelle                                        | SR $\times 2$                                                  |
|       | Sharma et al. [162]     | PDCNCF                                                 | Ultrasons             | CNN + CWT                                               | PSNR > 42.1                                                    |
|       | Alam et al. [5]         | GAN + VGG19 + GWT + LBP                                | Ultrasons             | SR + classification                                     | Acc. 98.40%                                                    |
|       | Lyu et al. [118]        | WSRGAN                                                 | Ultrasons             | Reconstruction haute fréquence                          | Gain qualitatif                                                |
| 2025  | Jiang et al. [80]       | USBSR                                                  | Ultrasons             | SR aveugle avec apprentissage de la dégradation         | SR $\times 4$                                                  |
|       | Bentaleb et al. [15]    | CV-SS-CNN                                              | Ultrasons IQ          | SR à partir des signaux IQ                              | PSNR 23.15                                                     |
|       | Lin et al. [107]        | ZSG-Net                                                | Thyroïde / CAMUS      | SR + diagnostic                                         | Dice 0.965                                                     |
|       | Borad et al. [17]       | ESRGAN-lite                                            | Ultrasons             | ESRGAN stable                                           | Stable training                                                |
|       | Xu et al. [200]         | RDC-GAN                                                | Multi-organes         | SR + classification                                     | PSNR 32.55–34.11                                               |
|       | Tambde et al. [176]     | SRGAN + CatBoost                                       | PCOS                  | SR + classification                                     | Acc. 99%                                                       |
|       | Chi et al. [33]         | BSRTU-Net                                              | Thyroïde              | SR Aveugle thyroid                                      | PSNR / SSIM $\uparrow$                                         |



### 1.4.1 Perspectives synthétiques des revues de littérature

Face à la prolifération des contributions scientifiques, l’appréhension synthétique du domaine devient un exercice aussi nécessaire que périlleux. Notre analyse de 23 revues systématiques – résumée dans le Tableau 1.7 – vise précisément à dégager les lignes de force structurelles qui organisent ce paysage foisonnant.

L’examen comparatif témoigne d’une transition d’une phase d’innovation algorithmique fragmentée vers une phase de consolidation méthodologique. Qu’il s’agisse de CNN, de GAN ou de Transformers, la super-résolution médicale se structure désormais autour d’un cadre conceptuel invariant alliant précision multi-échelle et validation par métriques de perception avancées.

### 1.4.2 Convergences méthodologiques et implications pour l’imagerie ultrasonore

L’analyse comparative des deux corpus — études expérimentales et revues de littérature — révèle des tendances structurantes qui informent directement le développement d’algorithmes de super-résolution dédiés à l’imagerie ultrasonore. Une convergence majeure réside dans l’adaptation impérative aux spécificités physiques du signal, condition nécessaire à toute exploitation clinique fiable.

La spécialisation modalitaire constitue l’un des axes majeurs mis en évidence par les revues récentes.

**Spécialisation modalitaire :** La spécialisation modalitaire observée dans la littérature récente met en évidence une prédominance de l’IRM et du CT, deux modalités devenues incontournables pour l’évaluation des approches de super-résolution en raison de leur large disponibilité et de leur richesse diagnostique. L’échographie, incluant le Doppler et le CEUS (échographie ultrasonore avec agent de contraste), occupe une place croissante et particulièrement exigeante : la présence du bruit speckle, la variabilité opérateur et la faible résolution spatiale imposent aux modèles une capacité renforcée à préserver les micro-structures d’intérêt clinique.

La super-résolution enrichit la microscopie et l’histopathologie en détails cellulaires et sub-cellulaires, et l’OCT et l’endoscopie affinent l’observation des interfaces tissulaires. Les approches multimodales (IRM–CT–US–PET) tirent parti de la complémentarité des signaux, mais restent coûteuses et sans métriques normalisées.

La littérature insiste sur le rôle des données simulées et des phantoms comme substituts lorsque les jeux de données cliniques appariés sont insuffisants. Enfin, des modalités plus spécifiques — telles que l’imagerie fonctionnelle (PET), le fond d’œil, l’hyperspectral ou l’imagerie satellite — témoignent de la transversalité croissante des méthodes de super-résolution.

Face aux restrictions imposées par l’acquisition et le diagnostic, des paradigmes algorithmiques complémentaires se sont progressivement imposés :

#### **Adaptation méthodologique :**

- Les **Réseaux Antagonistes Génératifs (GANs)** se distinguent par leur capacité à modéliser et à supprimer le bruit speckle, tout en préservant la texture tissulaire avec un réalisme remarquable.
- Les **Transformers** excellent à capturer les dépendances à longue portée, garantissant la cohérence topologique des structures anatomiques de grande étendue.
- Les **architectures hybrides** émergent comme le paradigme le plus prometteur, unissant les avantages respectifs pour atteindre un compromis optimal entre fidélité locale et intégrité globale.

TABLEAU 1.7 – Synthèse de 23 revues sur la SR médicale par Deep Learning : objectifs et améliorations rapportées

| Année | Review                          | Type d'images                                       | Données utilisées                                                                         | Modèles DL                                                                                 | Objectifs                                                       | Améliorations                                                                    |
|-------|---------------------------------|-----------------------------------------------------|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------|----------------------------------------------------------------------------------|
| 2020  | Kazeminia <i>et al.</i> [83]    | IRM, CT, US, PET, rétine, RX, microscopie           | Variées                                                                                   | GAN, DCGAN, cGAN, CycleGAN, Pix2Pix, WGAN                                                  | Synthèse, Seg., reconstruction, débruitage, Class.              | Réalisme élevé, amélioration Seg./Class.                                         |
|       | Chen <i>et al.</i> [28]         | <b>Ultrasons</b> microvasculaires                   | Données précliniques                                                                      | Deep-ULM, mSPCN-ULM, CNN 3D                                                                | SR ultrasonore (ULM)                                            | 10–50 $\mu\text{m}$ , acquisition rapide                                         |
| 2021  | Zhou <i>et al.</i> [213]        | IRM, CT, radiographie, <b>US</b> , microscopie      | BraTS, ADNI, Camelyon                                                                     | U-Net, GAN, ResNet, Inception                                                              | Seg., Class., détection                                         | Précision $\uparrow$ , généralisation $\uparrow$                                 |
|       | Ahishakiye <i>et al.</i> [3]    | MRI, CT, PET, <b>US</b>                             | Open-source datasets (divers)                                                             | CNN, U-Net, GAN, AUTOMAP, LEARN                                                            | Reconstruction (principal), SR/débruitage                       | Qualité $\uparrow$ , bruit $\downarrow$ , temps $\downarrow$ , dose $\downarrow$ |
|       | Singh <i>et al.</i> [165]       | PET cérébrale                                       | Données PET                                                                               | CSRP, EUSR, SICNN                                                                          | SR mono/multi-image                                             | PSNR/SSIM $\uparrow$                                                             |
|       | Sixou <i>et al.</i> [103]       | CT, IRM 2D/3D, rétine, endoscopie, microscopie      | Simulées et cliniques                                                                     | SRCNN, U-Net modifié, DRRN, SRDenseNet, DCSRN, ESRGAN, GAN-CIRCLE                          | SR 2D/3D, débruitage                                            | PSNR/SSIM $\uparrow$                                                             |
| 2023  | Patel <i>et al.</i> [143]       | IRM, CT, <b>US</b> , X-ray, rétine, histopathologie | Cardiac MRI, Brain MRI (OASIS, BraTS), Covid-CT/X-ray, LUNA16, Fundus, BreakHis, Camelyon | CNN (SRCNN, FSRCNN, DnCNN), GAN (LPGAN, CycleGAN, SRGAN, MIASSR), réseaux résiduels/denses | Super-résolution d'images médicales par deep learning           | PSNR 27–47 dB, SSIM 0.70–0.99                                                    |
|       | Rane <i>et al.</i> [151]        | IRM, CT, <b>US</b> , OCT, microscopie               | Multi-modales                                                                             | TransUNet+, SwinFPN, ViT-V-Net                                                             | Seg., Class., SR                                                | Résolution $\uparrow$                                                            |
|       | Jiachuan <i>et al.</i> [69]     | IRM, CT, PET, <b>US</b> , microscopie               | Multi-modales                                                                             | CNN, SRGAN, Deep-ULM, q-DL                                                                 | SR, reconstruction                                              | Dose $\downarrow$ , résolution $\uparrow$                                        |
|       | Song <i>et al.</i> [169]        | <b>US</b> 2D/3D, Doppler, CEUS                      | Données cliniques                                                                         | U-Net, FUSPR, GAN, LSTM                                                                    | Seg., Class., détection                                         | Bruit speckle $\downarrow$ , précision $\uparrow$                                |
| 2024  | van der Pol <i>et al.</i> [147] | <b>US</b> (POCUS, B-mode)                           | Données in vivo, phantom, simulation                                                      | CNN, GAN (cGAN, CycleGAN), U-Net, encodeur-décodeur                                        | Amélioration de la qualité (débruitage, SR, contraste, détails) | PSNR $\uparrow$ ( +6.6 dB), SSIM $\uparrow$ ( +0.28), qualité clinique améliorée |
|       | Showrov <i>et al.</i> [76]      | IRM, CT, PET, <b>US</b> , pathologie                | Multi-modales                                                                             | SRGAN, StyleGAN, DCGAN                                                                     | Synthèse, SR                                                    | Réalisme $\uparrow$ , PSNR/SSIM $\uparrow$                                       |
|       | Umirzakova <i>et al.</i> [184]  | IRM, CT, RX, PET, <b>US</b> , OCT, microscopie      | Données publiques                                                                         | SRCNN, DRN, GAN, Transformers                                                              | SR, fusion multimodale                                          | Revue critique SOTA, fidélité et détails fins $\uparrow$                         |
|       | Anas <i>et al.</i> [6]          | IRM, CT, rétine, hyperspectral                      | Données génériques                                                                        | HAT, SwinFIR, Restormer, GAN                                                               | SR, débruitage                                                  | Relations globales $\uparrow$                                                    |
|       | Heng <i>et al.</i> [71]         | IRM, CT, PET, <b>US</b> , pathologie                | Publiques/privées                                                                         | FS-GAN, Cond-GAN                                                                           | Synthèse, SR                                                    | PSNR +2–4 dB                                                                     |
|       | Sindhura <i>et al.</i> [37]     | IRM, CT, <b>US</b> , rétine                         | Multi-modales                                                                             | GAN, CNN, Transformer                                                                      | SR, Seg.                                                        | Robustesse $\uparrow$                                                            |
|       | Zheng Bi <i>et al.</i> [16]     | OCT, fond d'œil                                     | Données publiques, cliniques                                                              | Deep Learning générique                                                                    | Détection rétinopathie diabétique                               | Précision diagnostique accrue                                                    |
| 2025  | Parashari <i>et al.</i> [140]   | IRM, CT, PET, <b>US</b> , histopathologie           | Multi-modales                                                                             | SRCNN, EDSR, SRGAN, MedSRGAN, TransUNet                                                    | SR, reconstruction                                              | Qualité $\uparrow$ , artefacts $\downarrow$                                      |
|       | Trigka <i>et al.</i> [182]      | IRM, CT, <b>US</b> , histopatho., satellite         | DIV2K, BSD100                                                                             | ESRGAN, LapSRN, ViT-SR, Restormer                                                          | SR, restauration                                                | Bruit $\downarrow$ , qualité $\uparrow$                                          |
|       | Gao <i>et al.</i> [59]          | <b>Ultrasons</b> (SRUS)                             | Microvasculaires                                                                          | ARU-GAN, U-Net SR                                                                          | Cartographie microvasculaire                                    | Résolution <50 $\mu\text{m}$                                                     |
|       | Jansche <i>et al.</i> [78]      | Microscopie optique, fluorescence                   | Données microscopiques LR                                                                 | Analyse de CNN/U-Net/architectures DL                                                      | SR mono-image                                                   | PSNR/SSIM/LPIPS $\uparrow$                                                       |
|       | Hussain <i>et al.</i> [73]      | CT, IRM, PET, US, X-ray, OCT, Fundus                | 122 jeux de données publics/privés                                                        | GANs (cGAN, CycleGAN, SRGAN, WGAN-GP)                                                      | Reconstruction, SR                                              | PSNR/SSIM $\uparrow$ , artefacts $\downarrow$                                    |
|       | Suha <i>et al.</i> [173]        | IRM fœtale, échographie fœtale                      | Données multi-centriques + bases publiques                                                | CNN, U-Net, Transformers, GANs, VAEs, FL                                                   | Seg., SR, reconstruction, Class.                                | DSC $\uparrow$ , PSNR/SSIM $\uparrow$ , robustesse $\uparrow$                    |

Class. = Classification   Seg. = Segmentation   SR = Super-résolution   SRUS = Technique d'imagerie ultrasonore qui exploite la localisation de microbulles de contraste pour reconstruire les structures microvasculaires à une résolution sub-millimétrique, dépassant ainsi la limite de diffraction classique.

- Les **réseaux CNN-Transformer** combinent l’efficacité locale des CNN et la modélisation contextuelle des Transformers

En dépit de son orientation prioritaire vers la super-résolution, notre analyse a inclus un spectre plus large de tâches d’imagerie médicale — segmentation, classification, reconstruction — afin d’en examiner les structures convergentes.

Cette approche intégrative met en relief les invariants méthodologiques de l’apprentissage profond appliqué au médical, notamment généralisation inter-modale, robustesse aux petits jeux de données et exploitation des connaissances préexistantes.

Dans ce cadre, le transfert d’apprentissage se distingue comme un vecteur transversal d’optimisation, articulant la réutilisation des connaissances et l’adaptation contextuelle des modèles.

Si l’évolution des architectures — des CNN aux GANs et transformers — a nettement amélioré la fidélité anatomique, l’enjeu ne réside plus dans la seule puissance du modèle. La priorité scientifique se déplace désormais vers la capacité d’adaptation de ces réseaux à l’hétérogénéité des données cliniques.

La section suivante lui est consacrée, sous la forme d’une revue systématique conforme aux standards PRISMA 2020.

## 1.5 Transfert d’apprentissage et fine-tuning en imagerie médicale

L’analyse croisée des revues et des études empiriques permet d’extraire les standards opérationnels du transfert d’apprentissage.

**a. Cadre conceptuel du transfert d’apprentissage** Le recours au transfert d’apprentissage constitue désormais le socle méthodologique standard permettant d’affiner les performances des architectures profondes sur des données cliniques. Le principe est de ne pas repartir de zéro : on utilise un modèle déjà entraîné sur une base massive comme "ImageNet" et on l’adapte à nos données médicales, qui sont souvent rares ou peu annotées.

Cette stratégie accélère la convergence et stabilise l’optimisation du réseau. Concrètement, qu’il soit partiel ou intégral, le fine-tuning permet d’aligner les capacités du modèle avec les subtilités des images médicales à haute résolution.

**b. Synthèse des revues systématiques récentes** Cette partie analyse les synthèses systématiques publiées depuis 2020 pour établir une structuration conceptuelle du transfert d’apprentissage en imagerie médicale. L’analyse croisée des architectures et des modalités d’imagerie permet d’extraire les standards actuels tout en questionnant leur robustesse face aux spécificités cliniques, mettant ainsi en lumière les points de rupture et les verrous technologiques qui freinent encore le domaine.

**c. Méthodologie des revues et protocole PRISMA** L’état de l’art présenté ici synthétise plusieurs revues systématiques parues entre 2020 et 2024 sous les directives PRISMA 2020. Le volume des données traitées — incluant des corpus allant jusqu’à 239 études — constitue une base méthodologique solide pour caractériser les pratiques actuelles du fine-tuning médical.

Le processus de sélection, illustré par le diagramme PRISMA de la Figure 1.9, a conduit à l’inclusion de 31 études pertinentes, comprenant 22 travaux empiriques et 9 revues systématiques.

**d. Synthèse méthodologique** Les revues systématiques recensées entre 2020 et 2023 adoptent pour la plupart un protocole PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), couvrant des corpus de 26 à 239 études. Elles confirment la dominance des modèles CNN pré-entraînés sur ImageNet — notamment AlexNet, VGG, ResNet, DenseNet et Inception — et l’usage récurrent de deux stratégies : (1) l’extraction de caractéristiques (*feature extraction*) et (2) le fine-tuning complet ou partiel des couches profondes.

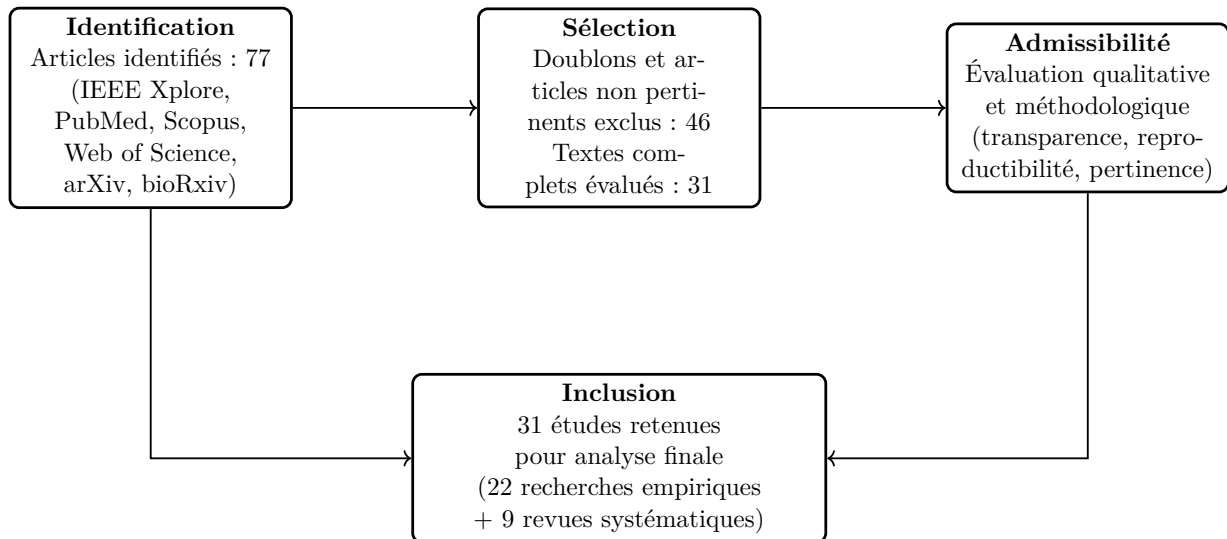


FIGURE 1.9 – Diagramme PRISMA (2020) – Processus de sélection systématique des 31 études sur le *Transfer Learning* et le *Fine-Tuning en imagerie médicale*.

**Domaines et modalités étudiés** Les revues systématiques analysées couvrent une grande diversité de modalités d’imagerie médicale, incluant l’IRM cérébrale et prostatique [185], la tomodensitométrie thoracique, l’échographie [57, 126], la radiographie, l’histopathologie et l’imagerie du fond d’œil. Les travaux de [130] et [92] montrent que 85 % des études concernent la classification et la segmentation, avec une forte concentration postérieure à 2018. L’échographie s’impose comme un cas d’étude majeur ; la forte hétérogénéité du bruit et la pénurie de données massives y rendent le fine-tuning indispensable pour stabiliser les performances des modèles.

**Efficacité du fine-tuning** La revue comparative de [87] et l’étude expérimentale de [39] démontrent que le *full fine-tuning* améliore significativement les performances (jusqu’à +2–3 points de PSNR/SSIM ou +10 % d’AUROC selon la tâche), particulièrement sur DenseNet-121 et VGG-19. La stratégie hybride LP + Full Fine-Tuning (LP-FT) émerge comme la plus robuste, bien que dépendante de la taille et du déséquilibre des jeux de données. Aucune stratégie universelle n’est identifiée ; l’optimisation du taux d’apprentissage et le choix des couches à geler demeurent spécifiques à chaque domaine.

**Enjeux transversaux** Les limites récurrentes mises en avant par les auteurs concernent principalement :

- (i) la rareté et la forte variabilité des données médicales,
- (ii) le risque de *negative transfer* lorsque le domaine source est trop éloigné,
- (iii) l’absence de benchmarks communs,
- et (iv) le manque d’interprétabilité clinique des modèles pré-entraînés [130, 92, 185].

Certaines revues (ex. [57]) soulignent aussi la dépendance du fine-tuning à la qualité des annotations et à l’homogénéité instrumentale.

**Perspectives identifiées** Les revues convergent vers plusieurs orientations de recherche :

- l’exploration du transfert hors d’ImageNet et vers des bases médicales spécialisées,
- le recours à l’apprentissage auto-supervisé et aux GANs pour la génération de données synthétiques,
- la standardisation des métriques et protocoles d’évaluation,
- et l’intégration de visualisations explicatives (Grad-CAM, Heatmaps) pour la validation clinique.

Aucune des revues n’aborde explicitement la super-résolution (SR). Toutefois, l’usage récurrent des GANs pour la génération d’images réalistes y est identifié comme une tendance méthodologique majeure.

### 1.5.1 Application du transfert d’apprentissage à la super-résolution

Bien que les études de TL aient majoritairement porté sur la classification et la segmentation, leurs conclusions sont directement transposables à la SR. Le fine-tuning d’architectures génériques (SwinIR, HAT, ou ResNet) sur des données médicales haute résolution permet d’accélérer la convergence et d’améliorer la fidélité anatomique des reconstructions [100, 149]. La réutilisation des poids pré-entraînés sur ImageNet favorise la stabilité de l’apprentissage, tandis que l’adaptation locale des couches de bas niveau préserve les textures et les détails structuraux essentiels.

## Discussion synthétique

L’ensemble des revues analysées conforte l’idée que le transfert d’apprentissage constitue un levier majeur pour pallier la rareté des données HR médicales et améliorer la robustesse des modèles. Toutefois, l’absence de protocoles de référence et le manque de standardisation des métriques limitent la comparabilité inter-études. L’enjeu actuel réside dans la construction de pipelines hybrides associant fine-tuning ciblé, augmentation réaliste (GANs) et contraintes anatomiques pour garantir une super-résolution fidèle et cliniquement pertinente.

### (a) Études empiriques et revues systématiques incluses

Cette revue de littérature a été conduite selon les **PRISMA 2020 guidelines**, comme détaillé dans la section précédente. Le processus a abouti à la sélection de **31 études** pour analyse :

- 22 études empiriques, présentant des expérimentations concrètes ;
- 9 revues systématiques, offrant une analyse transversale du domaine.

Les 9 revues analysées portent sur l’application du transfert d’apprentissage (TL) en imagerie médicale. Elles couvrent des modalités variées, incluant IRM, CT, échographie, rayons X, OCT, mammographie et histopathologie.

Certaines études, notamment Morid *et al.*[130], 2021, mettent en évidence une forte dépendance à ImageNet pour l’initialisation des modèles, ainsi que le manque de benchmarks standardisés et l’absence de comparaison dans plusieurs études. D’autres revues soulignent la diversité des stratégies TL (multi-étapes, fine-tuning, feature extraction), l’usage de GAN pour l’augmentation ou la reconstruction, la variabilité des datasets et la faible validation externe.

Le tableau 1.8 synthétise ces revues, leurs périodes couvertes, les modalités étudiées et les observations clés, offrant une vue d’ensemble des tendances et limites actuelles du transfert d’apprentissage en imagerie médicale.

Au-delà des revues systématiques, l’examen de 22 études empiriques (Tableau 1.9) illustre la traduction opérationnelle de ces stratégies face aux contraintes propres aux données cliniques.

TABLEAU 1.8 – Synthèse de 9 revues majeures sur le transfert d’apprentissage en imagerie médicale.

| Année | Référence                     | Période couverte | Modalités étudiées                                        | Observations clés                                                                                                                                                      |
|-------|-------------------------------|------------------|-----------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2020  | Mischi <i>et al.</i> [126]    | 2015–2020        | Échographie (formation + analyse)                         | DL efficace pour formation et analyse; dépendance aux données synthétiques; risque de sur-généralisation; importance reconstruction.                                   |
| 2021  | Valverde <i>et al.</i> [185]  | 2010–2020        | IRM cérébrale                                             | 129 études : TL efficace mais peu de modèles IRM dédiés; questions éthiques négligées; faible reproductibilité.                                                        |
|       | Morid <i>et al.</i> [130]     | 2012–2020        | IRM, CT, X-ray, US, Fundus, OCT                           | Dépendance forte à ImageNet; peu de benchmarks; 65% des études sans comparaison; besoin de standardisation.                                                            |
| 2022  | Kora <i>et al.</i> [92]       | 2010–2021        | IRM, CT, X-ray, US, histologie                            | 239 études : TL dominé par CNN classiques (ImageNet); datasets petits; GAN surtout pour augmentation; 65% sans benchmark; risque de negative transfer rarement évalué. |
|       | Fiorentino <i>et al.</i> [57] | 2010–2021        | Échographie pelvienne                                     | Variabilité importante des acquisitions US; datasets restreints; manque validation clinique; GAN pour bruit mais peu pour SR.                                          |
|       | Kim <i>et al.</i> [87]        | 2013–2021        | IRM, CT, X-ray, US                                        | 121 études : classification via TL; “feature extraction” plus stable que fine-tuning; quatre stratégies TL identifiées.                                                |
| 2024  | Ayana <i>et al.</i> [10]      | 2014–2023        | Mammographie, IRM, CT, US, X-ray, Fundus, histopathologie | TL multi-étapes peu exploité; domain shift fréquent; hétérogénéité des pipelines; manque d’interprétabilité.                                                           |
|       | Davila <i>et al.</i> [39]     | 2016–2023        | IRM, CT, X-ray, histologie, dermatoscopie, endoscopie     | Comparaison de 8 stratégies de fine-tuning; aucune méthode universelle; LP-FT souvent supérieure; dépendance au dataset.                                               |
|       | Piffer <i>et al.</i> [146]    | 2014–2023        | IRM, US, mammographie, histopathologie                    | 77 études sur datasets restreints : TL + GAN très répandus; très faible validation externe; nombreux biais méthodologiques.                                            |

**Note :** TL = Transfer Learning; LP-FT = Linear Probing Fine-Tuning : seule la dernière couche est entraînée, le reste du réseau étant gelé. Fundus = photographie du fond d’œil utilisée pour l’analyse de la rétine et des vaisseaux rétiens.

Si le transfert d'apprentissage est largement validé pour les tâches d'analyse (classification, segmentation), il apparaît également comme un levier de régularisation essentiel pour la super-résolution. Dans ce cadre, l'utilisation de modèles pré-entraînés fournit des connaissances a priori qui stabilisent la reconstruction des structures anatomiques. Ce corpus révèle ainsi les bonnes pratiques actuelles tout en isolant les verrous méthodologiques persistants.

## (b) Analyse des Tendances et Enseignements

L'analyse conjointe des 9 revues systématiques et 22 études empiriques révèle plusieurs tendances robustes :

- **Prévalence du fine-tuning complet** : la réinitialisation partielle des couches conduit à des pertes de performance. Le fine-tuning intégral est recommandé lorsque la taille du jeu de données est suffisante.
- **Émergence du TL multi-étapes** : certaines approches procèdent à un transfert séquentiel (*ImageNet* → base intermédiaire → tâche cible), améliorant la généralisation inter-modalités.
- **Performance accrue des architectures DenseNet et VGG** : ces modèles présentent une meilleure réactivité au transfert que ResNet, notamment pour l'IRM et l'échographie.
- **Rôle du PEFT (Parameter-Efficient Fine-Tuning)** : les méthodes récentes (LoRA, adapters) permettent une adaptation rapide sans réentraînement complet, prometteuses pour la SR médicale.
- **Limites récurrentes** : absence de benchmarks standardisés, risque de *NT*, manque d'interprétabilité et de validation clinique.

Ces observations convergent vers un consensus : le *transfer learning* constitue un levier central pour pallier la rareté des données HR médicales et améliorer la robustesse des modèles, mais son efficacité reste conditionnée à l'adéquation entre domaine source et cible.

Cette analyse systématique, conduite selon les **PRISMA 2020 guidelines**, garantit :

- une **traçabilité complète** du processus de sélection ;
- une **transparence méthodologique** dans l'évaluation de la qualité des études ;
- une **reproductibilité intégrale** du protocole.

L'intégration de sources prépublication renforce la couverture du corpus, garantissant la prise en compte des tendances les plus récentes. Cette rigueur confère à l'analyse un **niveau élevé de validité scientifique**, essentiel pour établir les fondements empiriques du transfert d'apprentissage appliqué à la super-résolution en imagerie médicale.

## 1.6 Conclusion

Ce chapitre a retracé l'évolution paradigmatique de la super-résolution (SR), passant des approches analytiques fondées sur des *priors* explicites à des modèles profonds capables de modéliser directement la distribution conditionnelle  $P(I_{HR} | I_{LR})$ . Cette progression, des réseaux convolutifs initiaux (SRCNN, VDSR, EDSR) aux architectures génératives (SRGAN, ESRGAN), puis aux Transformers (SwinIR, HAT) et aux modèles de diffusion (SR3), illustre le déplacement des objectifs : d'une **optimisation centrée sur la fidélité pixelique (PSNR)** vers une **reconstruction perceptuelle et sémantique (LPIPS, FID)**.

L'adaptation de ces méthodes aux images médicales reste néanmoins complexe. Contrairement aux images naturelles, conçues pour la perception visuelle et caractérisées par des textures aléatoires et des dégradations idéalisées, les images médicales — en particulier échographiques — sont régies par une **physique d'acquisition spécifique** (speckle multiplicatif, variabilité opérateur-sonde, etc.) et imposent une **fidélité anatomique stricte**, où la précision diagnostique prime sur l'esthétique visuelle.

TABLEAU 1.9 – Études empiriques exhaustives sur le transfer learning et le fine-tuning en imagerie médicale (22 articles).

| Année | Référence                        | Modalité             | Tâche                | Contribution principale                                                                                                                                                                                      |
|-------|----------------------------------|----------------------|----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2020  | Mahbod <i>et al.</i> [120]       | Dermoscopie          | Classification       | Ensemble multi-échelle de CNNs fine-tunés, impact du prétraitement sur le TL.                                                                                                                                |
|       | Amiri <i>et al.</i> [8]          | Échographie          | Segmentation         | Étude fine-tuning partiel vs profond d'un U-Net.                                                                                                                                                             |
|       | Alzubaidi <i>et al.</i> [7]      | Pied diabétique      | Classification       | Analyse exhaustive TL gelé vs fine-tuning partiel et complet.                                                                                                                                                |
| 2021  | El Zein <i>et al.</i> [53]       | RX thorax            | Classification       | EfficientNetB0 fine-tuné avec classifieur hybride SVM.                                                                                                                                                       |
|       | Mustafa <i>et al.</i> [132]      | Multimodal           | Classif., détection  | Transfert supervisé à grande échelle via modèles BiT.                                                                                                                                                        |
|       | Jiang <i>et al.</i> [79]         | IRM                  | Super-résolution     | FA-GAN avec attention spatiale et fine-tuning médical.                                                                                                                                                       |
|       | Sarasaen <i>et al.</i> [157]     | IRM dynamique        | Super-résolution     | Fine-tuning patient-spécifique d'un 3D U-Net pré-entraîné (CHAOS) à partir d'un scan statique haute résolution, améliorant la SR d'IRM dynamiques fortement sous-échantillonnées (jusqu'à 6,25% du k-space). |
| 2022  | Nneji <i>et al.</i> [134]        | RX thorax            | Classification COVID | SR suivie d'un réseau Siamese fine-tuné.                                                                                                                                                                     |
|       | Mabrouk <i>et al.</i> [119]      | Peau, sang           | Classification       | MobileNetV3 fine-tuné avec optimisation Chaos Game.                                                                                                                                                          |
|       | Matsoukas <i>et al.</i> [122]    | Multimodal           | Analyse de features  | Étude empirique de la transférabilité des représentations DL.                                                                                                                                                |
|       | Ayana <i>et al.</i> [11]         | Échographie mammaire | Classification       | Transfert multistage ImageNet → microscopie → échographie.                                                                                                                                                   |
| 2023  | Khaledyan <i>et al.</i> [85]     | Écho mammaire        | Segmentation         | Sharp Attention U-Net fine-tuné pour segmentation robuste.                                                                                                                                                   |
|       | Iskandar <i>et al.</i> [75]      | Écho fœtale          | Synthèse             | GAN pré-entraîné puis fine-tuné pour génération US réaliste.                                                                                                                                                 |
|       | Aghelan & Rouhani [1]            | RX, rétine           | Super-résolution     | Fine-tuning médical de Real-ESRGAN avec dégradation avancée.                                                                                                                                                 |
| 2024  | Tang <i>et al.</i> [177]         | Peau                 | Super-résolution     | Comparaison fine-tuning complet vs PEFT (BitFit, AdaptFormer).                                                                                                                                               |
|       | Wang <i>et al.</i> [190]         | CT                   | Classification + SR  | SR GAN + radiomique profonde via TL 3D.                                                                                                                                                                      |
|       | Nandal <i>et al.</i> [133]       | RX, rétine           | Super-résolution     | Adaptation médicale de Real-ESRGAN multi-modalités.                                                                                                                                                          |
|       | Boudi <i>et al.</i> [18]         | IRM                  | Classification       | Fine-tuning ResNet50V2 avec gestion du déséquilibre (SMOTE).                                                                                                                                                 |
|       | Reddy <i>et al.</i> [153]        | RX, rétine           | SR, débruitage       | Fine-tuning partiel ESRGAN pour amélioration clinique.                                                                                                                                                       |
| 2025  | Prasad <i>et al.</i> [149]       | Multimodal           | Segmentation         | Comparaison TL gelé vs fine-tuning complet multi-modalités.                                                                                                                                                  |
|       | TehraniNasab <i>et al.</i> [178] | RX thorax / CheXpert | Fine-tuning          | Étude comparative des paradigmes de fine-tuning pour modèles foundation, évaluant FID, Vendi et performance downstream sur CheXpert.                                                                         |
|       | Lee <i>et al.</i> [100]          | CT                   | Super-résolution     | Fine-tuning Transformer HAT pour SR CT haute précision.                                                                                                                                                      |



La première revue systématique, conduite selon les **PRISMA 2020 guidelines**, a couvert **46 études** (23 revues + 23 recherches empiriques). Elle met en évidence :

- la **maturité croissante** des approches profondes (CNN, GAN, Transformer) en SR médicale ;
- mais également leurs **limites structurelles** : manque de données HR, absence de validation clinique, métriques peu adaptées à la pertinence diagnostique, et difficultés de généralisation entre appareils ou protocoles.

Ces constats confirment que **l'échographie** constitue une modalité particulièrement exigeante, du fait du speckle et de la variabilité d'acquisition, appelant à des stratégies d'adaptation plus spécialisées.

La seconde revue, également conduite selon les PRISMA 2020 guidelines, a identifié **77 articles** (31 retenus : 22 empiriques + 9 revues). Elle démontre que :

- le **transfert d'apprentissage (TL)** et le **fine-tuning (FT)** sont devenus des leviers majeurs pour surmonter la rareté des données annotées ;
- les **modèles pré-entraînés** (notamment sur ImageNet ou DIV2K) permettent des gains significatifs même sur des tâches médicales spécifiques ;
- mais que **peu d'études appliquent directement ces stratégies à la super-résolution échographique**, laissant un espace de recherche encore ouvert.

L'analyse de ces travaux montre que TL et FT offrent des gains significatifs de précision et de qualité d'image, mesurés par PSNR, SSIM, LPIPS et FID. L'intégration de modèles hybrides CNN-Transformer permet de tirer parti à la fois de la structure globale et des détails fins des images, avec un potentiel marqué pour l'adaptation à de petits corpus médicaux.

Malgré ces progrès, certaines limites persistent : validation clinique insuffisante, absence de benchmarks standardisés, métriques traditionnelles inadaptées à l'évaluation diagnostique, artefacts anatomiques possibles avec les GANs, et reproductibilité limitée entre centres et dispositifs. Les tendances récentes convergent vers :

- des modèles **hybrides et physiquement contraints**, adaptés aux spécificités instrumentales de chaque modalité ;
- l'usage croissant de **métriques perceptuelle et clinique** (BRISQUE, NIQE, LPIPS, MOS) ;
- des stratégies d'**adaptation au domaine**, incluant le *fine-tuning complet* lorsque les ressources sont suffisantes, ainsi que le **fine-tuning partiel** et les **méthodes PEFT (Parameter-Efficient Fine-Tuning)** pour une adaptation plus légère.

Ces approches partagent un même objectif : ***adapter efficacement des modèles pré-entraînés volumineux à de nouvelles tâches ou modalités médicales en limitant le nombre de paramètres ajustés.***

Elles permettent ainsi de réduire les coûts d'entraînement, d'éviter le surapprentissage sur de petits jeux de données, et de préserver les connaissances générales acquises lors du pré-entraînement.

Dans le contexte de la super-résolution échographique, les méthodes PEFT offrent une voie prometteuse pour exploiter la puissance des architectures Transformer tout en respectant les contraintes de ressources et de disponibilité.

### Adaptation des modèles pré-entraînés

*L'adaptation d'un modèle pré-entraîné à un nouveau domaine repose principalement sur deux stratégies :*

- ***Fine-tuning complet*** : mise à jour de tous les paramètres.
- ***PEFT (fine-tuning partiel)*** : ajustement ciblé de certains modules afin de réduire coût et surapprentissage.

*Exemples de PEFT : Adapters, LoRA, Bit-FIT. Objectif commun : adapter le modèle de façon ciblée et légère, sans modifier tous les paramètres.*

Ainsi, **notre contribution s'inscrit dans la continuité des revues PRISMA présentées dans ce chapitre**, en s'appuyant sur l'état de l'art des modèles de super-résolution appliqués à l'échographie. Nous nous concentrons sur les aspects méthodologiques et conceptuels, en identifiant les familles de modèles (CNN, GAN, Transformer) et les approches d'adaptation (transfert d'apprentissage, fine-tuning) pertinentes pour ce domaine.

Ce chapitre identifie une tendance méthodologique : **la migration progressive des modèles généralistes vers des architectures spécialisées et adaptées cliniquement**, où le transfert d'apprentissage et le fine-tuning jouent un rôle central dans l'optimisation des performances. Il établit un cadre théorique et conceptuel solide pour l'application de ces stratégies à la super-résolution échographique, tout en reportant l'analyse empirique aux chapitres suivants.

En conséquence, le cadre conceptuel établi dans ce chapitre prépare le terrain pour la mise en œuvre expérimentale des modèles de super-résolution adaptés à l'échographie. Le chapitre suivant s'attache à décrire **l'imagerie échographique**, ses principes physiques, ses avantages cliniques et les contraintes inhérentes à l'acquisition des images.



# CHAPITRE 2



## Chapitre 2

# Ultrasons et imagerie diagnostique : fondements, intérêts et restrictions

### 2.1 Introduction

L'essor de l'imagerie médicale moderne repose sur la complémentarité des modalités ; dans ce contexte, l'échographie occupe une position privilégiée, car elle combine sécurité d'emploi et disponibilité. Elle se distingue des techniques ionisantes, telles que la radiographie ou la tomodensitométrie, en exploitant des ondes acoustiques de haute fréquence, inaudibles pour l'oreille humaine. Celles-ci couvrent généralement la plage 1–10 MHz dans la pratique clinique courante, et peuvent s'élever jusqu'à 15–20 MHz lorsqu'il s'agit d'explorer des tissus superficiels comme la thyroïde, les tendons, la peau ou le tissu sous-cutané [196, 174]. Par leur interaction subtile avec la structure mécanique des tissus, ces ondes restituent des images en temps réel, offrant ainsi au clinicien une fenêtre instantanée sur l'anatomie vivante.

Apparue dans le domaine médical au milieu du XX<sup>e</sup> siècle, l'échographie s'est progressivement affirmée comme un outil incontournable du diagnostic. L'échographie diagnostique s'est imposée comme une technique d'imagerie de première intention, favorisée par son caractère non invasif et par une réputation de sécurité clinique enviable, qui contraste avec les risques associés aux rayonnements ionisants et à l'IRM. Sa large accessibilité et la possibilité de répéter les examens sans danger en font un outil incontournable de la pratique médicale contemporaine [51].

La première étape est l'émission d'une impulsion ultrasonore. Elle est très brève, souvent réduite à une période de la fréquence centrale de la sonde. C'est cette impulsion qui sert à sonder les tissus.

L'échographie, particulièrement adaptée à l'étude des tissus mous et au suivi prénatal, permet une visualisation en temps réel des organes et de la circulation sanguine, et l'apport du Doppler élargit encore son champ d'application clinique [68, 70, 196].

Ces spécificités expliquent que l'échographie soit largement adoptée, tant dans les services hospitaliers de pointe que dans les systèmes de soins aux ressources restreintes.

L'échographie est largement utilisée dans les environnements à ressources limitées en raison de son faible coût, de sa portabilité et de sa simplicité de mise en œuvre, qui en font un outil adapté au suivi médical en zone rurale. Néanmoins, la technique reste contrainte par des limites bien connues, notamment une pénétration osseuse réduite et une forte dépendance à l'expertise de l'opérateur [60, 152, 171].

Dans ce chapitre, nous rappellerons d'abord les bases physiques et technologiques de l'imagerie ultrasonore. Nous détaillerons ensuite ses avantages cliniques et ses principales applications, avant d'aborder les limites propres à la méthode. Une attention particulière sera portée à son intérêt dans les milieux ruraux, où elle constitue souvent l'unique modalité d'imagerie acces-

sible et joue un rôle crucial, qu'il s'agisse de la santé maternelle ou du diagnostic au point d'intervention, contribuant à améliorer l'accès aux soins et à réduire les inégalités sanitaires [211, 44].

## 2.2 Fondements

### 2.2.1 Principes de l'échographie

Le principe sur lequel repose l'échographie est l'écholocalisation, un procédé comparable à celui qu'emploient les chauves-souris et certains mammifères marins pour analyser leur environnement.[142]

- **Ligne échographique** tout commence avec la production d'ondes ultrasonores pulsées de courte durée, typiquement une période de la fréquence centrale, transmises au patient par la face avant de la sonde.
- **Construction de l'image** Le fonctionnement repose sur une séquence répétée d'émission et de réception. La sonde émet une impulsion ultrasonore de très courte durée; cette onde se propage dans le corps et rencontre différentes interfaces. Une partie de l'énergie est réfléchie, le reste poursuit son trajet. Les échos reviennent avec un retard proportionnel à la profondeur de leur origine, ce qui permet de reconstruire l'image ligne par ligne. Le temps de retour de chaque écho permet d'en déduire la profondeur correspondante, selon la relation :

$$t = \frac{2d}{c} \quad (2.1)$$

où  $t$  désigne le temps mesuré entre l'émission et la réception de l'écho (en secondes),  $d$  la profondeur du réflecteur (en mètres), et  $c$  la vitesse de propagation des ultrasons dans les tissus (en m/s), classiquement prise à 1540 m/s pour les tissus mous.

Les échos profonds étant atténués, l'appareil applique une amplification croissante en fonction du temps, appelée TGC (Time Gain Compensation), afin de les rendre visibles. L'ensemble du cycle s'effectue en un temps extrêmement court, ce qui permet l'obtention d'une image en temps réel.

**Zone noire...** En échographie, une zone noire est une zone sans interface ultrasonore. Elle est dite, zone anéchogène ou transonore. Généralement, en arrière de cette zone on observe un renforcement postérieur correspondant à l'absence d'atténuation du faisceau ayant traversé cette zone. A l'opposé, le faisceau d'ultrasons ayant traversé les zones voisines aura, lui, été atténué.

**Zone blanche...** Les points blancs observés correspondent aux échos renvoyés par les interfaces. Leur densité et leur intensité caractérisent l'échogénicité : faible (hypoéchogène), normale (échogène), forte (hyperéchogène).

## Sonde ultrasonore Émission-Réception d'onde pulsée

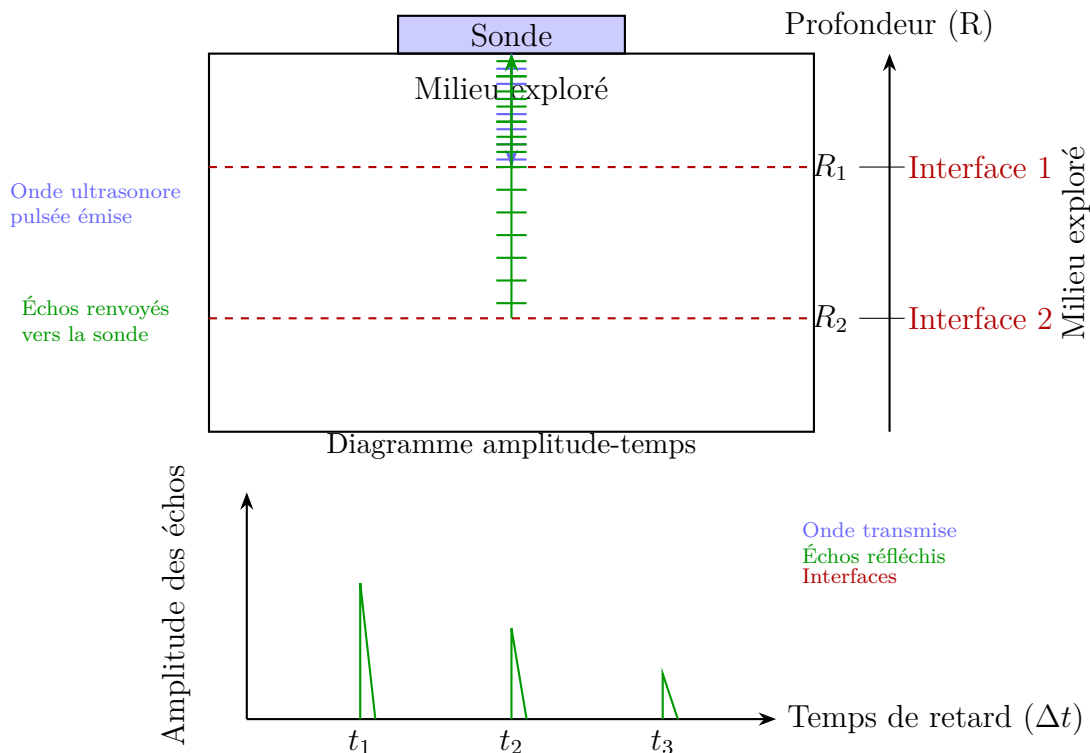


FIGURE 2.1 – Schéma d'émission-réception d'une onde ultrasonore pulsée. *En pratique, la sonde émet une impulsion brève qui traverse le milieu étudié. Chaque fois que l'onde rencontre une discontinuité acoustique, une partie est réfléchiée sous forme d'échos tandis que le reste se transmet plus loin. Ces échos, une fois captés, sont transformés en signaux électriques dont l'amplitude dépend de la réflectivité de l'interface. Sur le diagramme amplitude-temps placé en dessous, les pics observés correspondent à ces échos : leur position indique le temps de retour et donc la profondeur de l'interface. Autrement dit, plus le parcours est long, plus l'écho apparaît tardivement.*

### 2.2.2 Bases physiques et principes de propagation des ultrasons

#### 2.2.2.1 Nature des ondes ultrasonores et domaine fréquentiel

Par définition, les ultrasons correspondent à des vibrations acoustiques de fréquence supérieure à 20 kHz, échappant ainsi au domaine de perception auditive humaine [174]. En imagerie diagnostique, la bande fréquentielle de 1 à 10 MHz est privilégiée, car elle matérialise l'équilibre entre deux exigences contradictoires : la résolution, qui s'améliore avec l'élévation de la fréquence, et la pénétration tissulaire, qui s'accroît à mesure que la fréquence diminue. L'optimisation de ce compromis conditionne la qualité et la pertinence des images échographiques obtenues [196].

Le paramètre fréquentiel est adapté en fonction de l'indication clinique : les sondes opérant à basses fréquences sont utilisées pour l'exploration abdominale et la visualisation d'organes profonds, alors que les hautes fréquences sont réservées à l'imagerie de structures superficielles, telles que la glande thyroïde, les tendons ou encore les milieux oculaires, où la résolution prime sur la pénétration.

Le paramètre fréquentiel est adapté en fonction de l'indication clinique : les sondes opérant à basses fréquences (environ 2–5 MHz) sont utilisées pour l'exploration abdominale et la visualisation d'organes profonds, tandis que les hautes fréquences (7–15 MHz) sont réservées à

l'imagerie de structures superficielles, telles que la glande thyroïde, les tendons ou encore les milieux oculaires, où la résolution prime sur la pénétration. La longueur d'onde  $\lambda$  dans le tissu est inversement proportionnelle à la fréquence  $f$ , selon la relation

$$\lambda = \frac{c}{f} \quad (2.2)$$

où  $c$  représente la vitesse du son dans les tissus mous (environ 1540 m/s), ce qui explique l'amélioration de la résolution spatiale à haute fréquence.

### 2.2.2.2 Propagation des ultrasons dans les tissus biologiques

Le comportement des ondes ultrasonores lors de leur propagation à travers différents milieux biologiques est entièrement décrit par les lois fondamentales de l'acoustique, offrant un cadre théorique pour l'interprétation des signaux échographiques<sup>2</sup>.

La propagation des ondes ultrasonores dans les tissus mous humains se fait à une vitesse moyenne d'environ 1540 m/s, valeur qui reflète l'équilibre entre la densité  $\rho$  et le module d'élasticité  $\beta$  du milieu, par :

$$c = \sqrt{\frac{\beta}{\rho}}. \quad (2.3)$$

En se propageant dans les tissus biologiques, les ondes ultrasonores rencontrent des variations de milieu qui entraînent différents effets : certaines portions sont réfléchies, générant les signaux échographiques, tandis qu'une partie est transmise et continue son trajet. L'onde perd de l'énergie avec la profondeur par absorption et diffusion, et sa trajectoire peut être modifiée par réfraction lorsque les vitesses de propagation diffèrent entre tissus. Cependant, leur pénétration est limitée à certaines structures : les os et les poumons représentent des obstacles majeurs en raison de leur forte impédance et de la présence d'air, qui provoque une réflexion quasi totale [52].

### 2.2.3 Génération et détection des ultrasons

L'émission et la réception des ultrasons reposent sur les propriétés piézoélectriques des cristaux contenus dans les transducteurs, le plus souvent du PZT. Lorsqu'on applique un courant électrique, le cristal se met à vibrer et émet l'onde ultrasonore. Quand l'onde rencontre des interfaces tissulaires, elle est partiellement réfléchiée ; cette réflexion déforme le cristal et produit un signal électrique. En traitant ce signal, on peut reconstruire des images détaillées de la structure des tissus, et la fidélité de cette conversion influence directement la qualité de l'imagerie [94].

Les transducteurs ultrasonores fonctionnent selon deux principes :

- **Émission pulsée** : l'onde ultrasonore est envoyée par courtes impulsions ; l'analyse des échos permet de reconstituer une coupe de l'organe.
- **Émission continue** : ce mode, utilisé en Doppler, permet de mesurer le flux sanguin en continu avec une grande précision.

Le signal ultrasonore capté est d'abord amplifié et filtré avant d'être numérisé. Des algorithmes de traitement transforment ce signal en une image bidimensionnelle en niveaux de gris, affichée instantanément. La fidélité de l'image n'est pas seulement déterminée par la machine : elle dépend également des propriétés des tissus et de l'expertise de l'opérateur pour interpréter correctement les échos.

---

2. On entend par *signaux échographiques* les échos acoustiques bruts reçus par la sonde avant tout traitement, par opposition aux *images échographiques* qui sont le résultat final de leur traitement et reconstruction.

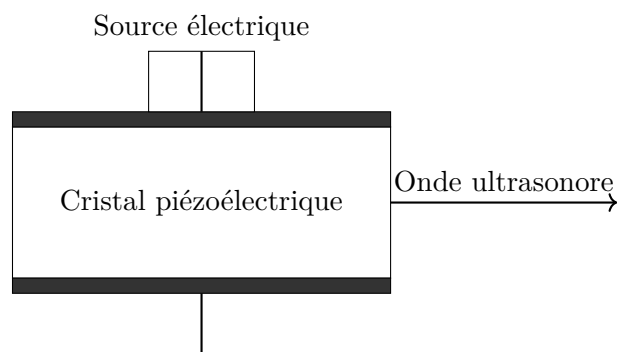


FIGURE 2.2 – Principe de génération d’une onde ultrasonore par effet piézoélectrique.

### 2.2.3.1 Modes d’émission et d’imagerie

Les transducteurs ultrasonores peuvent fonctionner selon différents modes, adaptés aux applications cliniques :

**Mode A.** Le mode A (Amplitude) correspond à la forme la plus simple de représentation ultrasonore : une ligne où la profondeur est reportée sur l’axe horizontal et l’amplitude des échos sur l’axe vertical. Les interfaces tissulaires apparaissent sous forme de pics successifs. S’il a quasiment disparu des applications courantes, il conserve une place en ophtalmologie, en particulier pour les mesures biométriques de l’œil (épaisseur cornéenne, épaisseur cristallinienne et longueur axiale), comme l’illustre la Figure 2.2.

**Mode B.** En échographie, le mode B constitue la base de l’imagerie morphologique. Il repose sur la projection bidimensionnelle des échos ultrasonores : la luminosité de chaque pixel reflète l’intensité du signal réfléchi à une profondeur donnée. L’image produite correspond à une coupe anatomique exploitable pour l’analyse des tissus et organes.

Utilisé de façon routinière, il constitue l’outil principal en échographie obstétricale et reste tout aussi essentiel pour l’examen abdominal, vasculaire ou musculosquelettique (Figure 2.4).

**Mode M.** Le mode M transforme une ligne extraite du mode B en un tracé représenté sur l’axe du temps. Il permet de mesurer avec précision les mouvements cardiaques rapides, ce qui explique son usage constant en échocardiographie. Ce mode demeure un outil incontournable en cardiologie, tant pour l’étude de la cinétique valvulaire que pour le diagnostic des cardiopathies congénitales chez l’enfant (Figure 2.5).

**Mode 3D.** Dans ses formes initiales, l’échographie 3D nécessitait des sondes motorisées, capables d’acquérir successivement plusieurs coupes 2D, ensuite recombinaisonnées pour former un volume. Les développements technologiques ont conduit à des sondes volumétriques permettant une acquisition directe, ouvrant la voie à l’imagerie 4D, c’est-à-dire à la visualisation tridimensionnelle en temps réel. Cette approche élargit le champ de l’échographie au-delà de l’obstétrique, en cardiologie et en pathologie musculosquelettique.[142, 9]

#### Échodoppler

L’échodoppler combine l’imagerie échographique classique au principe physique de l’effet Doppler, qui relie le décalage de fréquence d’une onde ultrasonore à la vitesse des particules en mouvement (notamment les globules rouges). En obstétrique, il s’agit d’un outil indispensable pour évaluer la fonction placentaire et hémodynamique fœtale.

Les principales indications cliniques du Doppler fœtal concernent l’évaluation du flux sanguin dans différents territoires vasculaires :

- **Utérines** : dépistage précoce de prééclampsie.
- **Ombilicale** : surveillance placentaire et retard de croissance.
- **Cérébrale moyenne** : diagnostic de l’anémie et redistribution hémodynamique.
- **Veine ombilicale / ductus venosus** : retour veineux fœtal et détection des signes de



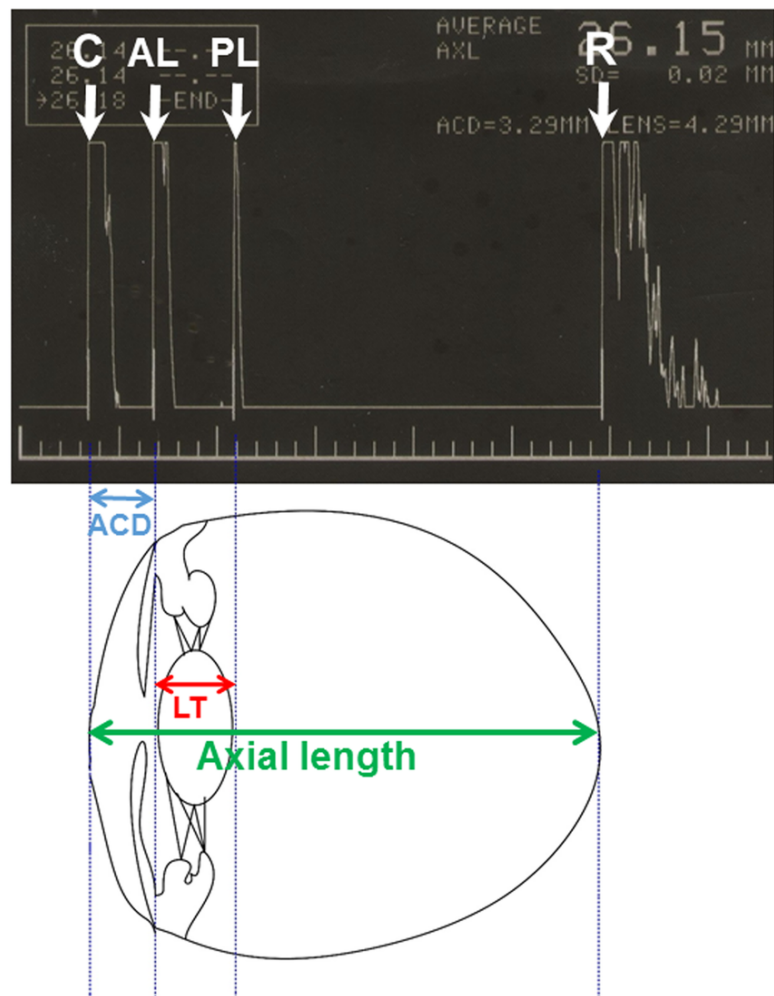


FIGURE 2.3 – Échographie en **mode A**. (A-scan) le long de l'axe optique de l'œil. Les pics correspondent à la cornée (C), à la face antérieure du cristallin (AL), à la face postérieure du cristallin (PL) et à la rétine (R). La profondeur de la chambre antérieure (ACD), l'épaisseur du cristallin (LT) et la longueur axiale peuvent être déduites de l'échographie en mode A [31].



FIGURE 2.4 – Coupe transversale du crâne fœtal en échographie **mode B**. (convex probe 3-7 MHz, image numéro 37, SonoAce-128BW), extraite de la galerie du site Medison.[124]

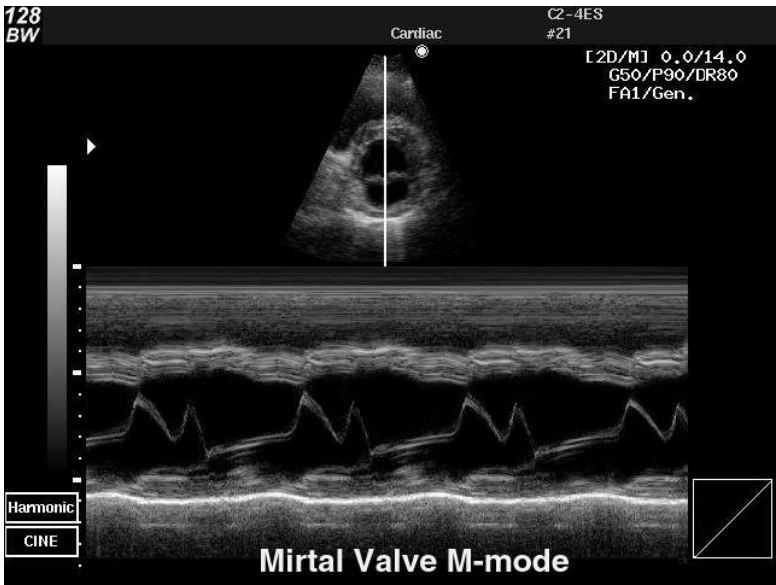


FIGURE 2.5 – Valve mitrale en échographie **mode M**. L’axe horizontal représente le temps et l’axe vertical la profondeur. Les variations périodiques traduisent le mouvement de la valve au cours du cycle cardiaque (SonoAce-128BW). Source : galerie d’images Medison.[124]

TABEAU 2.1 – Résumé des modes échographiques avec exemples illustratifs

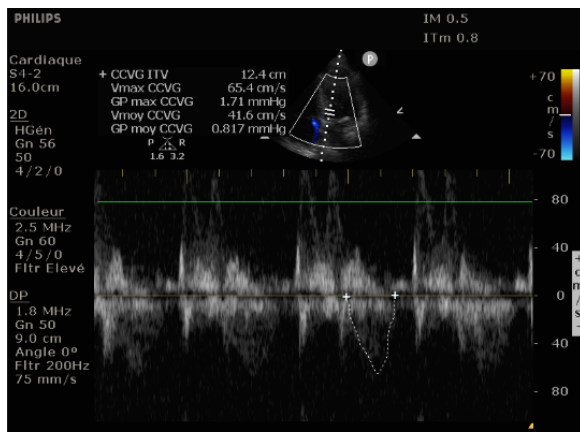
| Mode           | Dimension             | Usage clinique et pertinence actuelle                                                                    | Exemple concret                       |
|----------------|-----------------------|----------------------------------------------------------------------------------------------------------|---------------------------------------|
| A (Amplitude)  | 1D                    | Utilisé historiquement pour les mesures biométriques ophtalmiques. Toujours utilisé en ophtalmologie.    | Échographie oculaire (Fig. 2.3)       |
| B (Brightness) | 2D                    | Mode de base pour visualiser les structures anatomiques. Couramment utilisé en échographie obstétricale. | Coupe crânienne fœtale (Fig. 2.4)     |
| M (Mouvement)  | 2D (profondeur/temps) | Suivi dynamique des structures mobiles, particulièrement en cardiologie.                                 | Valve mitrale en mouvement (Fig. 2.5) |

décompensation.

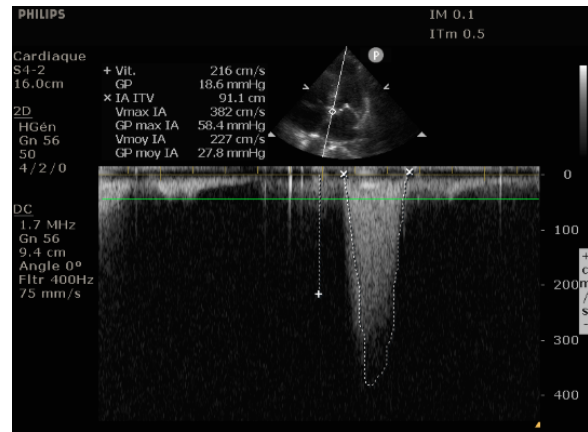
L’échodoppler peut être réalisé selon plusieurs modes :

- **Doppler pulsé** : évalue la vitesse du sang dans un point précis.
- **Doppler continu** : mesure des flux rapides sur toute la longueur du faisceau, sans localisation exacte.
- **Doppler couleur** : cartographie bidimensionnelle des flux superposée à l’image B, montrant direction et intensité.
- **Doppler tissulaire** : mesure le mouvement des tissus cardiaques eux-mêmes, notamment pour évaluer la fonction myocardique et la cinétique valvulaire fœtale.

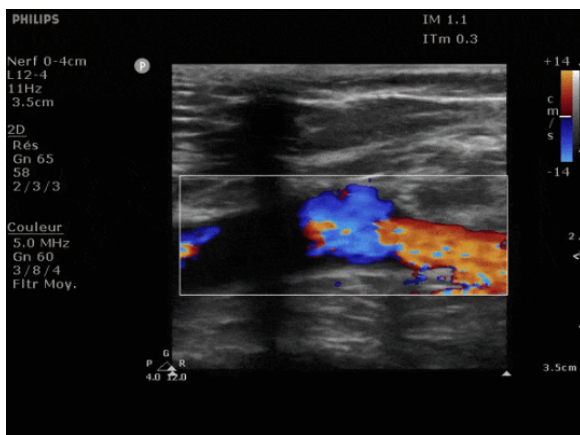
Son apport est déterminant dans la surveillance des grossesses à risque et dans l’évaluation de la vitalité fœtale.



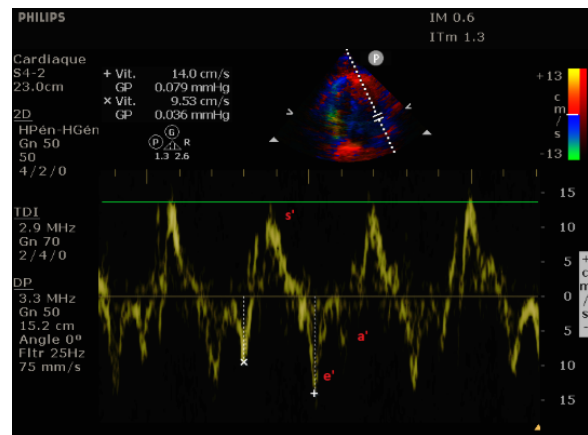
(a) Doppler pulsé : mesure de la vitesse du sang en un point précis.[110]



(b) Doppler continu : flux rapides sur toute la longueur du faisceau.[108]



(c) Doppler couleur : cartographie bidimensionnelle des flux, direction et intensité.[109]



(d) Doppler tissulaire : analyse du mouvement des tissus myocardiques.[111]

FIGURE 2.6 – Illustrations des principaux types de Doppler utilisés en échographie fœtale.

Le Tableau 2.2 présente une synthèse des principales modalités de l'échographie Doppler utilisées en pratique fœtale, en mettant en évidence leurs dimensions, leurs usages cliniques et des exemples illustratifs pour chaque technique.

## 2.2.4 Définition et caractéristiques de l'image ultrasonore

### 2.2.4.1 Définition

L'image ultrasonore se construit en temps réel à partir des échos générés par des ondes acoustiques de haute fréquence. Lorsqu'une onde rencontre une frontière entre deux tissus de propriétés différentes, une partie est renvoyée vers la sonde. La machine mesure à la fois la force de ce signal et le délai avec lequel il revient, puis convertit ces données en intensité lumineuse sur l'écran. Chaque point lumineux correspond à une mesure de ce type, et l'ensemble compose une cartographie interne qui restitue assez fidèlement l'anatomie, et parfois aussi certains aspects fonctionnels.

### 2.2.4.2 Caractéristiques principales

En pratique, la qualité et l'interprétation d'une image échographique reposent sur plusieurs paramètres techniques et phénomènes physiques intrinsèques.

TABLEAU 2.2 – Résumé des modes Doppler en échographie fœtale : usage clinique et exemples

| Mode                    | Dimension                       | Usage clinique et pertinence actuelle                                                             | Exemple concret                                                                                   |
|-------------------------|---------------------------------|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|
| Doppler pulsé[110]      | 1D ou 2D (axe temps/profondeur) | Mesure de la vitesse sanguine en un point précis; surveillance des flux dans les vaisseaux fœtaux | Artère ombilicale : mesure du flux sanguin pour détecter une insuffisance placentaire (Fig. 2.6a) |
| Doppler continu[108]    | 1D                              | Analyse des flux rapides sur toute la longueur du faisceau, sans localisation exacte              | Artères utérines : évaluation de la résistance au flux (Fig. 2.6b)                                |
| Doppler couleur[109]    | 2D                              | Cartographie bidimensionnelle des flux sanguins; visualisation de la direction et de l'intensité  | Flux ombilical ou intracardiaque : détection des anomalies de circulation (Fig. 2.6c)             |
| Doppler tissulaire[111] | 2D                              | Analyse du mouvement des tissus cardiaques eux-mêmes; évaluation de la cinétique myocardique      | Ventricule fœtal : suivi du mouvement des parois et des valves (Fig. 2.6d)                        |

**Résolution d'une image échographique** En pratique, la résolution d'une image échographique repose sur deux choses : la sonde et le système qui la pilote. Cependant, ce que l'on appelle communément "résolution" recouvre en fait plusieurs dimensions : la résolution axiale, qui conditionne la discrimination des structures alignées dans l'axe de propagation ; la résolution latérale, qui permet de différencier des structures voisines transversalement ; et l'épaisseur de coupe, déterminée par la largeur réelle du faisceau ultrasonore.

- Résolution spatiale

La résolution spatiale correspond à la capacité de distinguer deux structures anatomiques rapprochées. Elle est scindée en :

- **Résolution axiale** : Qui désigne la capacité du système à discriminer deux réflecteurs alignés avec l'axe ultrasonore ; cette performance est largement conditionnée par la fréquence, la longueur d'onde et la durée de l'impulsion.
- **Résolution latérale** : Il s'agit de séparer deux structures juxtaposées, donc côte à côte dans un plan perpendiculaire au faisceau. Elle est conditionnée par la largeur du faisceau et sa focalisation, mais plus on va en profondeur, plus elle se dégrade.
- **Résolution en élévation (ou épaisseur de coupe)** : elle traduit la discrimination dans le plan hors de l'image. Elle dépend de la géométrie de la sonde et explique une bonne partie de l'effet de volume partiel.
- Résolution temporelle. Elle définit la rapidité avec laquelle l'échographe peut générer et afficher des images successives, condition indispensable à l'exploration de phénomènes dynamiques comme la cinétique cardiaque. Elle est tributaire du taux d'acquisition (frame rate<sup>3</sup>).
- Contraste de l'image. Il traduit combien deux structures se distinguent en niveaux de gris.

3. La notion de « frame rate » correspond au rythme de production et d'affichage des images ultrasonores, exprimé en nombre d'images par seconde.

Cela dépend de l'écart d'impédance acoustique<sup>4</sup>, mais aussi des traitements numériques<sup>5</sup> qu'on applique derrière.

- Rapport signal sur bruit (SNR). Ce paramètre conditionne la qualité de l'image. Il est déterminé par la puissance du signal émis, l'atténuation tissulaire et les corrections appliquées (compensation en fonction de la profondeur, ou TGC).

**Phénomènes inhérents : speckle et artefacts** L'image ultrasonore est également caractérisée par deux phénomènes omniprésents qui influencent sa lecture :

- Le **bruit de speckle** est une granularité aléatoire inhérente aux images ultrasonores cohérentes, reflétant la microstructure des tissus. Il peut altérer la perception visuelle, mais contient également des informations diagnostiques exploitables. Les mécanismes de son apparition et les méthodes de traitement sont présentés ultérieurement.
- Les **artefacts** sont des anomalies visuelles ou structures artificielles générées par la non-conformité des conditions d'imagerie aux hypothèses simplificatrices (vitesse du son, propagation rectiligne). Leur reconnaissance et classification sont essentielles pour une interprétation fiable des images.
- Caractère temps réel et innocuité. L'imagerie ultrasonore permet l'obtention d'images dynamiques en temps réel, ce qui constitue un avantage majeur en pratique clinique. De plus, l'absence d'irradiation ionisante garantit une excellente innocuité, autorisant son usage répété, y compris chez la femme enceinte.
- Modes d'affichage. L'image peut être obtenue selon différents modes (A, B, M, Doppler), dont les principes et applications sont détaillés dans la section consacrée à l'échographie.

### 2.2.4.3 Bruit de Speckle

Le bruit de speckle est un phénomène granuleux inhérent aux images échographiques, résultant de l'interférence des ondes ultrasonores réfléchies par des structures microscopiques du tissu [64]. Bien qu'il puisse réduire la lisibilité en masquant de faibles différences de contraste, le speckle contient des informations sur l'organisation tissulaire et peut être exploité dans l'analyse quantitative [116]. Des techniques de filtrage adaptatif permettent d'en atténuer l'impact tout en conservant les contours anatomiques essentiels pour le diagnostic [174].

**Origine physique.** Le speckle observé dans l'image échographique est causé par la superposition des échos issus de multiples micro-diffuseurs très petits. Cette superposition conduit à des variations locales d'intensité, perçues comme un grainage. Dans un tissu homogène, la distribution de ces variations est généralement conforme à une loi de Rayleigh<sup>6</sup>, comme attendu d'un phénomène stochastique.

**Impact sur l'image.** Le speckle diminue le rapport signal-bruit et tend à masquer les détails fins, ce qui réduit la résolution de contraste et peut compliquer l'identification des structures anatomiques. Cependant, cette granularité n'est pas entièrement indésirable : sa répartition

4. Par définition, l'impédance acoustique ( $Z = \rho \cdot c$ ) caractérise la résistance d'un tissu à la propagation des ondes ultrasonores. Elle fixe combien de l'onde est réfléchi à une interface ; plus l'écart entre tissus est élevé, plus le contraste observé sur l'image échographique est important.

5. La correction post-acquisition de l'image repose sur des traitements numériques tels que la gestion de la dynamique, le filtrage et la réduction du bruit. Ils améliorent le contraste et la lisibilité des tissus, permettant une distinction plus claire des structures.

6. La loi de Rayleigh décrit la répartition statistique de l'intensité issue de la superposition aléatoire des ondes ultrasonores réfléchies par de nombreux micro-diffuseurs dans un tissu homogène. Elle constitue un modèle standard pour caractériser le speckle dans les images échographiques, comme l'ont montré Tuthill et ses collègues [183].

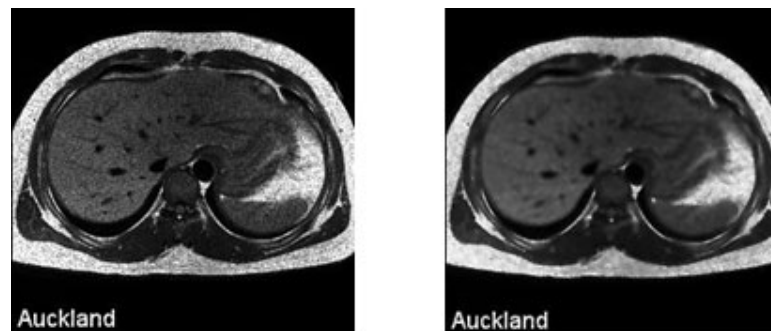


FIGURE 2.7 – Image échographique affectée par le bruit de speckle, caractérisé par une granularité aléatoire réduisant la lisibilité des structures internes (à gauche). Image correspondante après application d’un filtrage adapté, montrant une atténuation du speckle et une amélioration du contraste des tissus (à droite). [96]

et sa texture contiennent des informations sur la microstructure du tissu, offrant un potentiel d’analyse supplémentaire lorsque l’on utilise des méthodes appropriées.

## 2.3 Intérêts de l’imagerie ultrasonore

### 2.3.1 Caractère non invasif et innocuité

L’échographie, en tant que modalité atraumatique et non invasive, ne requiert ni incision, ni effraction des tissus, ni recours à une anesthésie, ce qui réduit à néant le risque iatrogénique lié à l’acte. Dans la grande majorité des cas, elle s’avère dépourvue d’effet algique et bénéficie d’une acceptabilité clinique optimale. Cette particularité en fait un examen sûr, y compris chez des populations sensibles telles que les femmes enceintes, les nouveau-nés ou les enfants, ainsi que chez les patients nécessitant des contrôles répétés au cours du suivi médical [138]. À l’inverse d’autres modalités requérant l’injection de substances de contraste, parfois responsables d’événements indésirables, elle permet d’obtenir des informations diagnostiques précises sans adjonction médicamenteuse. Ces caractéristiques expliquent qu’elle occupe une place centrale dans la pratique clinique moderne, où la sécurité du patient et la minimisation des risques iatrogènes sont devenues des priorités [138, 135].

### 2.3.2 Accessibilité et faible coût

Selon les rapports de l’Organisation mondiale de la santé (OMS), environ 70 % de la population mondiale demeure dépourvue d’accès effectif aux techniques d’imagerie médicale de haute performance, telles que la tomodensitométrie (TDM) et l’imagerie par résonance magnétique (IRM) [197, 198, 199]. L’échographie, plus accessible, constitue une alternative adaptée aux soins primaires et à la santé maternelle [197, 198]. L’utilisation de l’échographie s’est développée dans les pays à ressources limitées, mais sa disponibilité reste contrastée selon le niveau de soins : la disponibilité médiane des diagnostics est d’environ 68 % dans les hôpitaux contre 19 % dans les structures de soins primaires, certaines modalités comme l’échographie étant disponibles dans moins de 5 % des centres primaires [201], dans un contexte global où l’imagerie diagnostique demeure insuffisante dans de nombreuses régions rurales et sous-dotées [167].

Le coût élevé de l’imagerie médicale en limite l’accès. Alors qu’un échographe portable peut être acquis pour une somme modeste (2000–10 000 USD), l’acquisition d’un scanner ou d’une IRM requiert un investissement de l’ordre de plusieurs centaines de milliers de dollars ( $\geq 300\,000$  USD), voire de plusieurs millions, constituant ainsi une barrière économique majeure à leur diffusion, en particulier dans les systèmes de santé à faibles et moyens revenus [199].

Ces caractéristiques économiques et pratiques ont favorisé l'essor de l'échographie dite « au lit du patient » (*point-of-care ultrasound*), devenue aujourd'hui un outil incontournable en médecine d'urgence, en soins intensifs et en anesthésie [129].

### 2.3.3 Imagerie en temps réel et guidage interventionnel

L'un des atouts distinctifs de l'échographie est sa capacité à fournir des images en temps réel, ce qui permet de visualiser des processus physiologiques dynamiques tels que les mouvements cardiaques, les contractions fœtales ou encore les flux sanguins évalués par l'échographie Doppler [138].

L'intérêt de l'échographie réside aussi dans sa capacité à fournir une imagerie dynamique, là où le scanner et l'IRM produisent des images essentiellement statiques, nécessitant des reconstructions sophistiquées pour l'analyse temporelle. Là où ces dernières produisent des images morphologiques statiques, nécessitant souvent des séquences et post-traitements complexes pour évaluer une cinétique, l'échographie fournit intrinsèquement une évaluation temporelle continue. Cette propriété est exploitée de manière cruciale en radiologie interventionnelle.

L'intérêt de l'échographie réside aussi dans sa capacité à fournir une imagerie dynamique, là où le scanner et l'IRM produisent des images essentiellement statiques, nécessitant des reconstructions sophistiquées pour l'analyse temporelle. Cette propriété est exploitée de manière cruciale en radiologie interventionnelle. Le guidage échographique peropératoire, en offrant une visualisation continue de la cible et de l'instrumentation, optimise la précision des gestes (ponctions, drainages, accès vasculaires) et constitue un facteur clé dans la réduction de la morbidité iatrogène [40, 168].

L'usage interventionnel de l'échographie s'est progressivement imposé comme standard de soins dans plusieurs disciplines. En anesthésie loco-régionale, elle optimise la localisation des structures nerveuses périphériques ; en médecine d'urgence, elle facilite le guidage des accès vasculaires centraux ; et en oncologie, elle permet la réalisation de biopsies ciblées. Son innocuité, sa disponibilité immédiate et sa reproductibilité expliquent l'essor constant de ses applications cliniques.

Au-delà de cet usage interventionnel, son caractère temps réel explique la diversité de ses indications cliniques. L'échographie occupe une place centrale en obstétrique et gynécologie, en cardiologie, en radiologie abdominale, ainsi qu'en imagerie vasculaire et musculosquelettique. Cette polyvalence découle de l'éventail de sondes disponibles et de l'évolution des technologies (2D, Doppler, 3D/4D), qui renforcent son rôle unique dans l'arsenal diagnostique [138, 40]. Au-delà de l'imagerie morphologique conventionnelle, l'échographie a vu émerger des modalités fonctionnelles avancées, notamment l'élastographie, qui permet d'évaluer de manière non invasive les propriétés mécaniques des tissus — leur rigidité et leur déformabilité — ouvrant ainsi de nouvelles perspectives diagnostiques, en particulier pour la caractérisation des lésions focales et la quantification de la fibrose [12].

### 2.3.4 Résolution des structures superficielles

L'étude échographique des tissus superficiels repose sur des fréquences élevées, généralement comprises entre 7 et 20 MHz. L'élévation de la fréquence réduit la longueur d'onde des ultrasons, et cette réduction améliore la finesse des détails que l'on peut distinguer à l'écran. La résolution est ainsi meilleure, aussi bien dans la direction de propagation du faisceau que dans la direction transversale.

**Applications cliniques.** En pratique courante, les sondes à haute fréquence permettent d'accéder à des structures superficielles avec une excellente résolution spatiale, bien que leur

pénétration soit limitée. Dans le domaine musculosquelettique, l'échographie offre une analyse fine des tendons et ligaments superficiels, qu'il s'agisse de ruptures complètes ou de simples tendinopathies.

Elle est également utilisée pour le repérage et l'étude des nerfs périphériques, par exemple dans le diagnostic du syndrome du canal carpien. En endocrinologie, elle constitue un outil de première intention pour la détection et la caractérisation des nodules thyroïdiens. De même, l'examen des glandes salivaires bénéficie de la résolution élevée de l'échographie, notamment pour l'analyse des lésions superficielles et le guidage de prélèvements [129, 101].

## 2.4 Restrictions de l'imagerie ultrasonore

### 2.4.1 Dépendance opérateur

La dépendance opérateur constitue une limite fondamentale de l'échographie diagnostique. Du fait de son caractère technologie-dépendante de l'utilisateur, la diffusion croissante de l'échographie au point de soin impose de garantir la compétence des praticiens, de définir les bénéfices d'une utilisation appropriée et de limiter les examens inutiles et leurs conséquences [129].

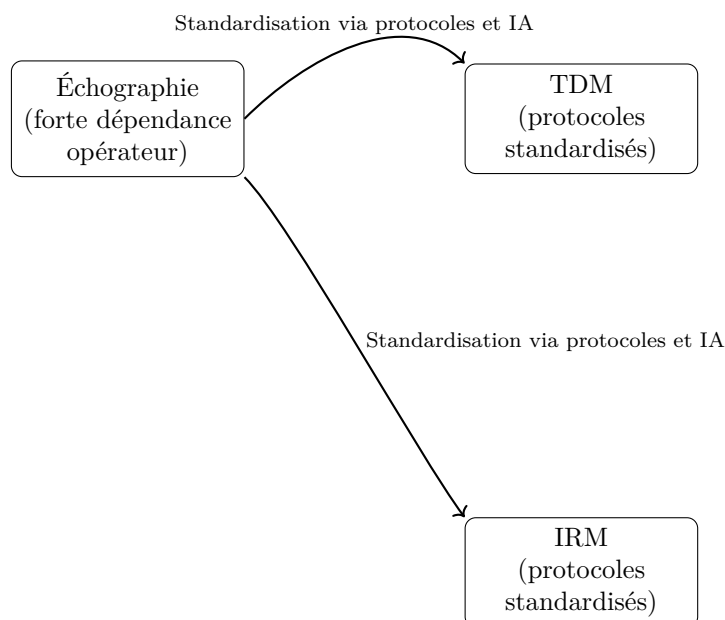


FIGURE 2.8 – Comparaison des modalités d'imagerie : *l'échographie présente une forte dépendance à l'opérateur, alors que la TDM et l'IRM reposent sur des protocoles standardisés. Les flèches indiquent la possibilité de standardisation progressive de l'échographie via la formation et l'assistance par intelligence artificielle (d'après Moore et Copel[129]).*

Cette particularité a une double portée : elle permet une grande souplesse et une adaptation immédiate à la clinique, mais elle introduit en contrepartie une variabilité parfois importante.

Dans l'imagerie musculosquelettique, l'échographie est fortement dépendante de l'opérateur et limitée par l'absence de protocoles standardisés, ce qui contribue à une variabilité inter-observateur [121].

Dans le domaine hépatique, l'évaluation échographique de la stéatose présente une concordance globalement modérée, avec des coefficients  $\kappa$  allant d'environ 0,40 à 0,63 selon les conditions d'évaluation [172]. (voir Figure 2.9 pour une illustration de la variabilité inter- et intra-observateur).

Ces chiffres contrastent avec la standardisation élevée des acquisitions en TDM ou IRM, où la variabilité inter-opérateur est quasi inexistante. Ils justifient les efforts actuels de formation



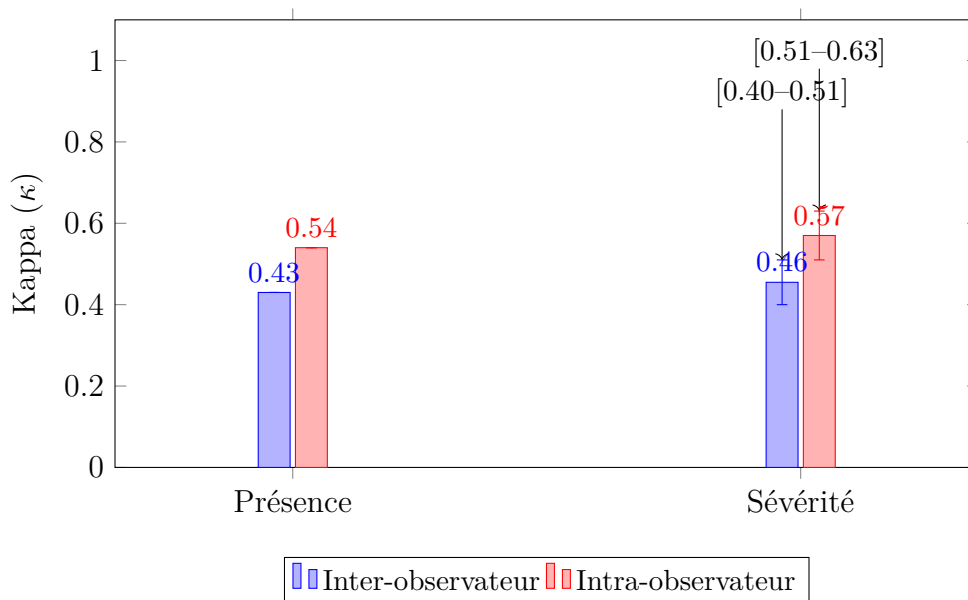


FIGURE 2.9 – Variabilité inter- et intra-observateur. Les barres d'erreur représentent les intervalles réels pour la sévérité. Aucune variabilité n'est affichée pour la présence.

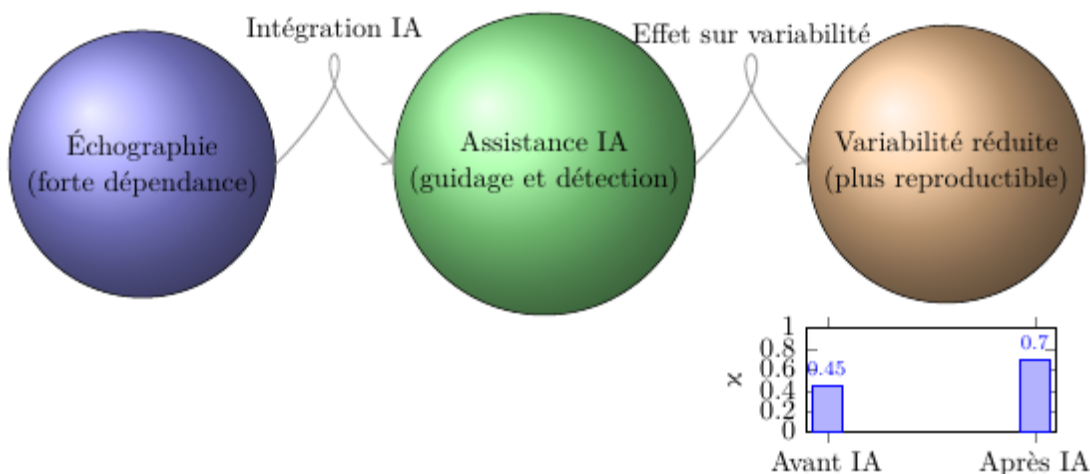


FIGURE 2.10 – Illustration de l'assistance par intelligence artificielle dans l'évaluation échographique hépatique (d'après Zhou et al.[212]). Les cercles sphériques représentent les étapes principales : échographie classique, assistance IA et résultat avec variabilité réduite. Le graphique sous le dernier cercle montre l'augmentation du coefficient  $\kappa$  après intégration de l'IA.

et l'exploration de solutions innovantes comme l'assistance par intelligence artificielle [212]. (Voir Figure 2.10)

## 2.4.2 Atténuation et pénétration limitée

L'atténuation des ultrasons, combinaison de phénomènes d'absorption, de diffusion et de réflexion, constitue une limite incontournable à l'efficacité de l'échographie. Elle induit une diminution graduelle de l'énergie acoustique au sein des tissus et restreint la profondeur d'exploration. Ce phénomène conditionne directement l'équilibre entre résolution spatiale et pénétration et est illustré par la Figure 2.11.

L'emploi de hautes fréquences permet d'obtenir des images d'une grande précision dans l'exploration des structures superficielles, mais restreint la pénétration à quelques centimètres

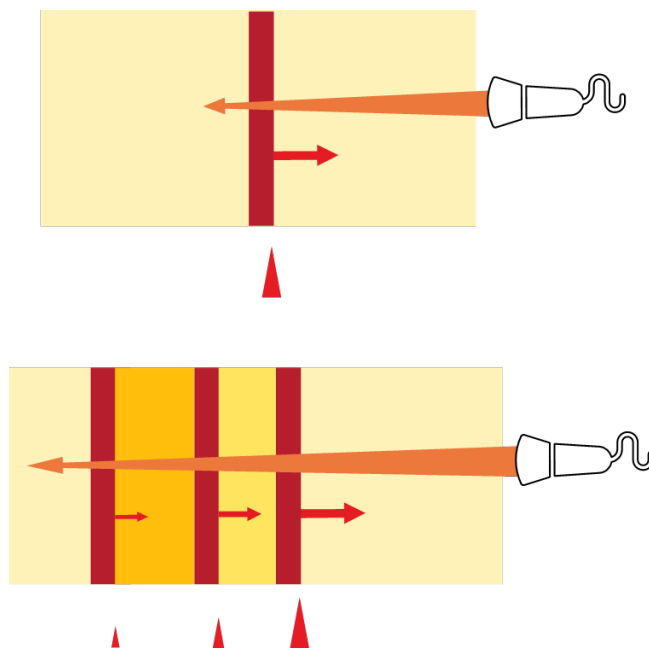


FIGURE 2.11 – Schéma de l'atténuation ultrasonore.[180] *En haut : interaction simple avec une interface tissulaire, donnant lieu à une réflexion partielle et à une transmission résiduelle. En bas : traversée de plusieurs couches, où l'absorption et les réflexions successives entraînent une atténuation cumulative du faisceau, limitant la profondeur d'exploration. Ce processus illustre la contrainte fondamentale de l'échographie, imposant un compromis entre résolution spatiale et pénétration.*

seulement. À l'opposé, les basses fréquences autorisent une imagerie des organes profonds, mais au détriment du détail anatomique.

Comme le montre la Figure 2.12, le cavernome<sup>7</sup>, situé sous la dure-mère<sup>8</sup>, au contact des ligaments denticulés<sup>9</sup>, a été identifié en échographie.

Certaines situations accentuent cette contrainte : excès adipeux, parois musculaires épaisses, contenu gazeux ou barrière osseuse. L'imagerie harmonique et les techniques de compounding apportent des améliorations relatives du contraste, sans toutefois abolir les lois physiques sous-jacentes. L'échographie conserve ainsi sa pertinence première dans l'analyse des tissus mous superficiels et abdominaux, alors que le scanner et l'IRM demeurent indispensables pour les structures profondes ou inaccessibles. Ces limitations anatomiques sont représentées à la Fig. 2.13

### 2.4.3 Artefacts et bruit de speckle

Dans toutes les images échographiques, on retrouve systématiquement deux types de perturbations : le speckle, qui donne ce grain un peu aléatoire mais qui contient des informations sur la texture du tissu, et les artefacts, des déformations ou faux objets causés par des écarts aux hypothèses de propagation. Il est important de savoir les identifier pour interpréter l'image correctement.

#### 2.4.3.1 Le bruit de speckle

Le bruit de speckle est un phénomène granuleux inhérent aux images échographiques, résultant de l'interférence des ondes ultrasonores réfléchies par des structures microscopiques

7. Anomalie vasculaire formée de petits amas de vaisseaux dilatés ressemblant à des cavités sanguines.

8. Couche externe résistante des méninges, assurant une protection mécanique du système nerveux central.

9. Replis fibreux reliant la moelle épinière à la dure-mère et contribuant à sa stabilisation.

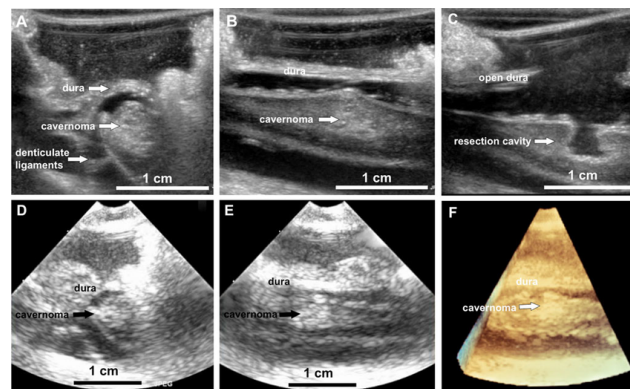


FIGURE 2.12 – Comparaison d’images échographiques selon la fréquence de la sonde.[20] Cette figure illustre le compromis fondamental entre résolution et profondeur. Les hautes fréquences (12–18 MHz) offrent une excellente résolution des structures superficielles, comme dans l’exploration tendineuse, mais leur pénétration reste limitée à quelques centimètres. À l’inverse, les basses fréquences (2–5 MHz) permettent d’explorer des organes profonds tels que le foie ou les reins, au prix d’une résolution spatiale moindre. (a) Visualisation du cavernome sous la dure-mère avec identification des ligaments denticulés. (b) Vue longitudinale du cavernome. (c) Cavité de résection après ouverture de la dure-mère. (d–e) Séquences échographiques localisant et délimitant le cavernome. (f) Reconstruction tridimensionnelle avec repérage de la dure-mère.

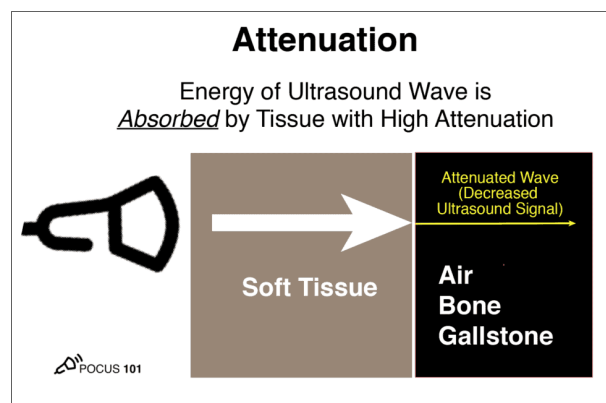


FIGURE 2.13 – Obstacles anatomiques à la pénétration ultrasonore.[42] L’air, l’os et les calculs biliaires réfléchissent ou absorbent presque totalement l’onde ultrasonore, empêchant l’exploration des structures situées en arrière.

(inférieures à la longueur d’onde) du tissu [64].

**Impact et utilisation** Bien qu’il réduise la résolution de contraste et masque les détails fins (Fig. 2.7), le speckle n’est pas un bruit au sens purement négatif du terme. Sa texture et sa répartition spatiale contiennent des informations précieuses sur l’organisation micro-structurale du tissu, exploitables pour l’analyse quantitative (élastographie, caractérisation tissulaire) [116]. Des techniques de filtrage adaptatif avancées (filtres médians, ondelettes, etc.) permettent d’en atténuer l’impact visuel tout en préservant les contours anatomiques essentiels au diagnostic [174].

### 2.4.3.2 Les artefacts principaux

Contrairement au speckle, les artefacts sont des erreurs de représentation ou des apparitions d’objets qui n’existent pas dans la réalité anatomique. Ils découlent directement des hypothèses simplificatrices de l’échographe (vitesse du son constante, atténuation uniforme, trajets recti-

lignes des ultrasons). Le tableau 2.3 fournit une synthèse des principaux artefacts observés en pratique clinique.

TABLEAU 2.3 – Synthèse des principaux artefacts en échographie et de leur origine physique.

| Artefact                        | Cause physique                                        | Manifestation en image                                                                           |
|---------------------------------|-------------------------------------------------------|--------------------------------------------------------------------------------------------------|
| Ombre acoustique                | Atténuation très forte (calcul, os, gaz)              | Cône d'ombre noir en arrière d'une structure très atténuante.                                    |
| Renforcement postérieur         | Faible atténuation (liquide)                          | Zone hyperéchogène (claire) en arrière d'une structure liquidienne.                              |
| Reverbérations                  | Multiples réflexions entre deux interfaces fortes     | Lignes hyperéchogènes parallèles et régulièrement espacées en-deçà d'une surface réfléchissante. |
| Artefact de lobe latéral        | Détection d'échos hors de l'axe principal du faisceau | Structures fantômes dupliquées latéralement, souvent en région vésiculaire ou vésicale.          |
| Artefact de mirage (réfraction) | Déviation du faisceau (change de milieu)              | Duplication ou déplacement apparent d'une structure profonde.                                    |

L'identification de ces artefacts est une compétence clinique cruciale. Loin d'être toujours gênants, certains, comme l'ombre acoustique ou le renforcement postérieur, sont des outils diagnostiques précieux.

## 2.4.4 Champ de vision restreint

Contrairement aux techniques d'imagerie sectionnelle comme le scanner ou l'IRM qui offrent une vue large et globale de la région anatomique, l'échographie est caractérisée par un champ de vision intrinsèquement limité.

### 2.4.4.1 Origines techniques

Cette limitation découle des contraintes physiques et des caractéristiques technologiques propres aux sondes.

**Géométrie de la sonde :** La taille et la forme du cristal piézoélectrique conditionnent le champ balayé. Les sondes linéaires offrent un champ rectangulaire, les convexes un champ en éventail. Aucun type de sonde ne permet de couvrir de larges surfaces sans déplacement. (Voir Figure. 2.14)

**Atténuation des tissus :** L'absorption des ultrasons par les tissus limite la profondeur. Les hautes fréquences augmentent le niveau de détail mais réduisent la pénétration.

**Focalisation du faisceau :** La meilleure résolution latérale se situe dans la zone focale. Hors de cette zone, le faisceau diverge et la netteté de l'image diminue.

### 2.4.4.2 Conséquences en pratique clinique

**Exploration dépendante de l'opérateur :** L'examineur doit mentalement reconstruire l'anatomie 3D à partir d'une multitude de coupes 2D étroites qu'il acquiert en déplaçant et en inclinant la sonde (Figure. 2.15). Cette compétence s'acquiert uniquement grâce à une formation longue et à une expérience pratique significative.

**Risque de « manquer une lésion » :** Une lésion peut facilement être hors du plan de coupe si l'examen n'est pas méticuleux et systématique. Ce risque est accru pour les lésions de petite taille.

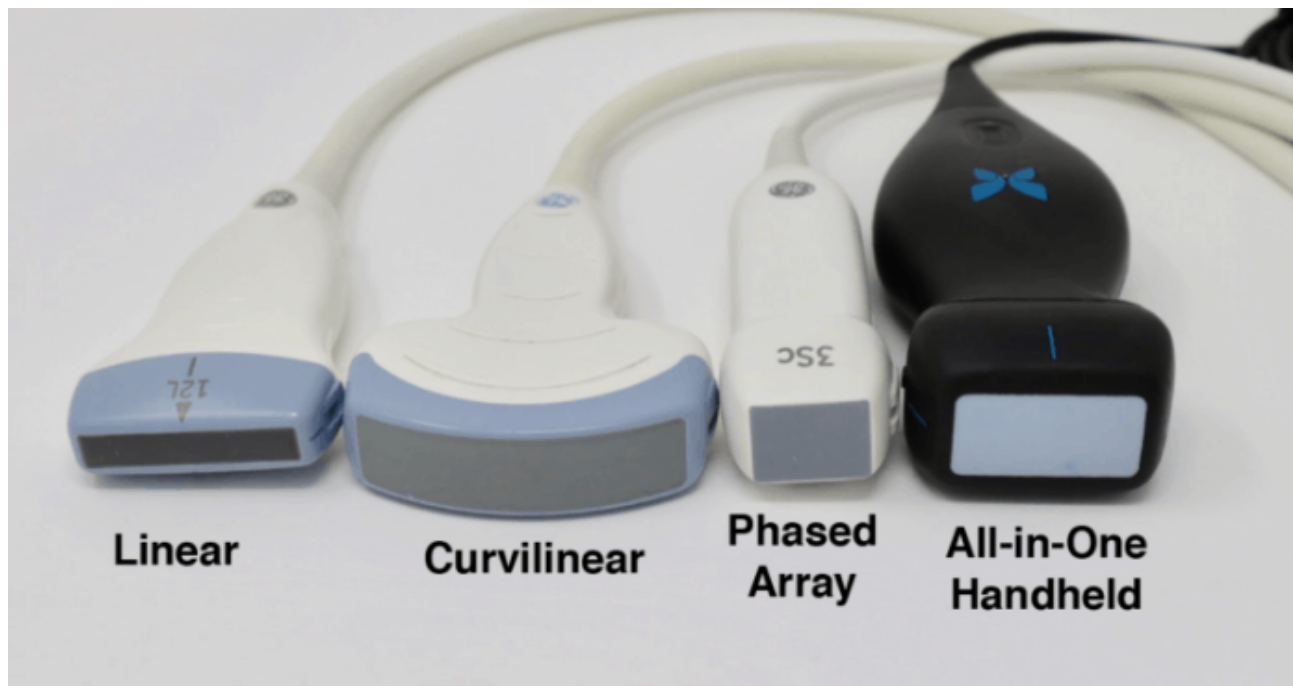


FIGURE 2.14 – Exemples de sondes échographiques usuelles : *linéaire* (haute résolution pour structures superficielles), *curviligne/convexe* (large champ pour l'imagerie abdominale), *phased array* (petite empreinte pour l'examen cardiaque) et sonde portative « *all-in-one* » polyvalente.[43]

**Difficulté pour l'imagerie de grandes structures :** L'évaluation d'organes très volumineux (comme un utérus fibromateux) ou de patients en surpoids important peut être difficile, nécessitant de multiples « chemins de sonde » et des reconstructions mentales complexes.

Les techniques modernes comme l'échographie 3D/4D ou les sondes matricielles élargissent considérablement le champ de vue instantané, mais ne suppriment pas totalement cette limitation fondamentale.

## 2.4.5 Limites diagnostiques pour certaines pathologies

Les contraintes physiques et technologiques de l'ultrason se traduisent par des difficultés diagnostiques particulières dans certaines pathologies ou zones anatomiques.

### 2.4.5.1 Limitations liées à la résolution et à la pénétration

**Pathologies de petite taille :** L'échographie peut manquer les microcalcifications ou petits calculs, surtout chez les patients adipeux. D'autres modalités comme la mammographie ou le scanner sont alors nécessaires.

**Évaluation de structures profondes :** La profondeur et les gaz intestinaux limitent la visualisation du pancréas, de l'aorte et du rétro-péritoine, réduisant la précision de l'imagerie.

### 2.4.5.2 Limitations liées aux artefacts et au contraste

**Caractérisation tissulaire :** L'échographie standard distingue correctement les kystes, les masses solides et les formations hyperéchogènes. Elle n'est toutefois pas assez spécifique pour identifier le type histologique d'une tumeur solide, surtout comparée à l'histologie ou à l'IRM avec contraste. L'élastographie peut améliorer cette évaluation.

**Pathologies osseuses et pulmonaires :** Le tissu osseux cortical reflète presque entièrement les ultrasons, ce qui rend impossible l'exploration des structures profondes. L'air pulmo-

### Cardinal Transducer Manipulation/Movement (Sliding, Tilting, Rotating, and Rocking)

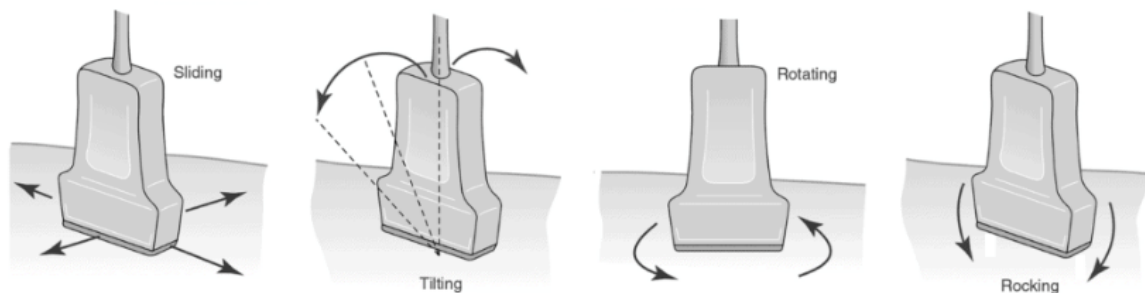


FIGURE 2.15 – Principaux mouvements de manipulation de la sonde échographique[43] : *sliding* (translation le long de la peau), *tilting* (inclinaison latérale), *rotating* (rotation autour de l'axe de la sonde) et *rocking* (bascule longitudinale). Ces manœuvres permettent d'optimiser l'angle d'incidence, d'ajuster la coupe et d'améliorer la qualité de l'image échographique.

naire constitue également un obstacle majeur, limitant l'échographie thoracique à la plèvre et aux épanchements.

#### 2.4.5.3 Limitations liées au champ de vision et à l'opérateur

**Staging tumoral** : Pour l'évaluation précise de l'extension locorégionale d'une tumeur (notamment la recherche d'envahissement ganglionnaire), le champ de vue restreint et la difficulté à standardiser l'examen rendent le scanner ou le PET-scan souvent plus complets et reproductibles.

**Surveillance de pathologies diffuses** : L'évaluation globale du foie pour la recherche de métastases est moins sensible qu'une imagerie en coupes (scanner/IRM) couvrant l'ensemble de l'organe sans dépendre du balayage de l'opérateur.

## 2.5 Conclusion

En définitive, l'échographie représente un compromis unique entre sécurité, coût et efficacité diagnostique, ce qui explique sa position prédominante en médecine moderne et son rôle crucial dans les zones rurales. Néanmoins, cette modalité reste contrainte par des limites de résolution et des artefacts de texture (speckle) qui découlent directement des propriétés physiques du milieu traversé.

L'enjeu actuel n'est donc plus uniquement de mieux acquérir, mais de mieux reconstruire. Pour pallier les restrictions physiques détaillées dans ce chapitre, le recours à des méthodes de traitement avancées devient impératif.

Face à ces barrières physiques, l'enjeu se déplace progressivement vers le traitement numérique du signal et de l'image. Puisque l'acquisition matérielle atteint un plateau, c'est dans la résolution algorithmique du problème inverse — la reconstruction d'une scène haute résolution à partir d'observations dégradées — que résident les perspectives d'amélioration.

Le chapitre suivant se propose d'étudier les outils permettant de résoudre ce défi : la super-résolution. Nous y analyserons les différentes stratégies architecturales — des réseaux de neurones classiques aux modèles génératifs et attentionnels — afin d'évaluer leur capacité à extraire une information haute résolution au-delà des capacités nominales des systèmes ultrasonores.



# CHAPITRE 3



## Chapitre 3

# Architectures de super-résolution : CNN, GAN et Transformers

### 3.1 Introduction

La super-résolution d'image constitue un défi majeur en vision par ordinateur dont l'enjeu est de reconstruire une version haute résolution (HR) à partir d'une image dégradée (LR). Mathématiquement, cette tâche s'apparente à un problème inverse mal posé<sup>1</sup>, où l'information contenue dans l'image basse résolution est insuffisante pour restituer parfaitement les détails fins.

Pendant longtemps, l'état de l'art s'est limité à des techniques d'interpolation classiques comme les algorithmes bicubiques ou de Lanczos. Cependant, ces méthodes engendrent fréquemment un lissage excessif et une perte substantielle des informations texturales. L'arrivée des réseaux convolutionnels (CNN), avec SRCNN [46], a marqué une première rupture en apprenant directement la correspondance  $LR \Rightarrow HR$ . Mais ces modèles, souvent limités par des fonctions de perte de type MSE, peinent encore à produire des détails perceptuellement réalistes.

Aujourd'hui, l'évolution de la super-résolution s'articule autour de trois courants principaux. Les réseaux adversaires (GANs) [63] sont reconnus pour leur capacité à générer des textures photoréalistes via une optimisation antagoniste. Parallèlement, les Transformers [187] exploitent le mécanisme d'auto-attention pour capturer des dépendances structurelles globales. Plus récemment, les modèles de diffusion ont montré une qualité de synthèse remarquable, mais leur nature itérative impose une latence d'inférence incompatible avec les contraintes du temps réel propres à l'examen échographique. Pour cette raison, ce travail privilégie les architectures GANs et Transformers, dont l'efficacité d'exécution répond mieux aux impératifs cliniques.

Le succès de ces modèles dépend toutefois de la disponibilité de jeux de données massifs. En imagerie médicale, cette contrainte est d'autant plus forte que les données sont rares, coûteuses à annoter et soumises à de strictes règles de confidentialité. Ce chapitre présente donc les fondements théoriques des architectures sélectionnées pour notre étude, en explorant deux voies distinctes.

D'un côté, nous étudions le modèle DBPISR [90], qui illustre l'apprentissage interne. Ce modèle se distingue en exploitant uniquement l'information de l'image d'entrée pour reconstruire la version HR et estimer le noyau de dégradation, s'affranchissant ainsi du besoin de bases de données supervisées.

De l'autre, nous analysons les avancées des stratégies externes à travers deux familles dominantes. Pour les GANs, nous considérons BSRGAN, Real-ESRGAN ainsi que A-ESRGAN

---

1. Plusieurs images haute résolution peuvent correspondre à la même image dégradée, rendant la reconstruction instable sans informations a priori.



pour ses améliorations adversariales ciblées. Pour les Transformers, l'étude porte sur SwinIR, Swin2SR et HAT (Hybrid Attention Transformer), qui combine attention locale et globale pour optimiser la reconstruction des détails.

Cette sélection comparative nous permettra d'apprécier la pertinence de chaque paradigme face aux exigences de fidélité et de robustesse imposées par la super-résolution médicale.

## 3.2 Enjeux et contraintes de la super-résolution en imagerie ultrasonore

L'application des paradigmes de super-résolution à l'échographie ne peut se résumer à une simple transposition des méthodes de vision par ordinateur. L'acquisition de l'image ultrasonore, régie par les lois de l'acoustique, impose des contraintes physiques qui redéfinissent les priorités de la reconstruction numérique.

### 3.2.1 La nature spécifique du Speckle

Contrairement aux images naturelles, généralement altérées par un bruit blanc gaussien additif, l'image échographique résulte d'interférences cohérentes. Le speckle ne saurait être réduit à un simple bruit additif ; il constitue la signature texturale intrinsèque à la microstructure tissulaire. Dès lors, accroître la résolution impose un équilibre délicat : restaurer les détails structurels sans lisser ces motifs physiques, sous peine de compromettre directement la pertinence clinique de l'examen.

### 3.2.2 Anisotropie de la réponse impulsionnelle

La résolution d'un système ultrasonore dépend intrinsèquement de la géométrie du faisceau acoustique, ce qui induit une réponse impulsionnelle (PSF) fortement anisotrope. Dans le modèle de dégradation classique :

$$y = (x \otimes k) \downarrow_s + n$$

le noyau de flou  $k$  est non seulement inconnu — nous plaçant dans un contexte de Blind SR — mais il varie également selon les directions axiale et latérale. Cette spécificité justifie l'intérêt pour des approches d'apprentissage interne comme le DBPISR, capables d'estimer des noyaux complexes adaptés aux particularités locales de l'acquisition.

### 3.2.3 Fidélité structurelle et risque d'hallucination

Enfin, l'usage d'architectures génératives comme les GANs ou les Transformers en milieu médical impose une rigueur extrême : la performance ne s'évalue plus au simple photoréalisme, mais à la fidélité structurelle absolue. Le risque d'*hallucination* — soit la création par le réseau de détails anatomiques inexistantes mais visuellement crédibles — constitue une limite critique. La robustesse des modèles doit donc être évaluée sur leur aptitude à garantir que l'information haute résolution produite reste une extension fidèle de la réalité biologique observée, sans introduire de biais susceptibles d'induire le clinicien en erreur.

C'est précisément pour répondre à ces impératifs de fidélité structurelle et s'adapter à la variabilité des conditions d'acquisition que nous étudions, en premier lieu, une approche d'apprentissage interne, capable d'exploiter les redondances propres à l'image d'entrée sans recourir à des données externes.

### 3.3 Modèle *Dual Back-Projection-based Internal Learning* : L'approche par apprentissage interne

Le modèle DBPI est une architecture convolutive (CNN) conçue pour la super-résolution aveugle. Il repose sur l'apprentissage conjoint de deux réseaux convolutifs : un réseau de sous-échantillonnage avec estimation implicite du noyau de dégradation (D ou *réseau de dégradation*) et un restaurateur d'image (U ou *réseau de super-résolution*), entraînés simultanément sur une seule image sans données externes.

Le DBPI utilise l'apprentissage interne, c'est-à-dire qu'il s'entraîne directement sur l'image d'entrée elle-même. Contrairement aux méthodes classiques d'apprentissage supervisé qui nécessitent un vaste jeu de données pour généraliser à différents types de dégradations, cette approche fonctionne de manière autonome, image par image.

#### 3.3.1 Architecture

Il repose sur une architecture duale de réseaux convolutifs (CNN) optimisés conjointement pour la super-résolution aveugle. Ses composants et mécanismes sont comme suit :

##### □ Réseau de sous-échantillonnage (D)

**Objectif :** Modéliser paramétriquement le noyau de dégradation (flou + sous-échantillonnage) ayant généré l'image basse résolution (LR).

**Structure du réseau :**

- Une couche de convolution avec *stride* (simulant conjointement convolution et réduction de résolution).
- Cinq couches de convolution linéaires (sans activation non-linéaire), permettant d'approximer une convolution à grand noyau.
- Un raccourci résiduel (*shortcut*) encourageant une estimation centrée du noyau.
- Traitement canal par canal pour respecter l'invariance chromatique de la dégradation.

##### □ Réseau de Super-Résolution (U)

**Objectif :** Restaurer une image haute résolution (HR) à partir de l'entrée LR.

**Structure du réseau :**

- Huit couches de convolution avec ReLU en activation intermédiaire.
- Aucune activation sur la dernière couche (régression directe).
- Inspirée des architectures légères de type ZSSR [163]/KernelGAN[13], optimisée pour l'apprentissage interne.

##### □ Processus Cyclique

Le modèle utilise une contrainte de consistance duale inspirée de CycleGAN, mais sans recourir à un *adversarial learning* :

$$\text{Down-Up} : I \xrightarrow{D} D(I) \xrightarrow{U} U(D(I))$$

$$\text{Up-Down} : I \xrightarrow{U} U(I) \xrightarrow{D} D(U(I))$$

Cette cyclicité impose une relation d'inversibilité locale entre  $D$  et  $U$ .

##### □ Fonctions de Perte

##### (a) Dual Back-Projection Loss ( $L_{\text{DBP}}$ )

$$L_{\text{DBP}} = \|U(D(I)) - I\|_1 + \|D(U(I)) - I\|_1 \quad (3.1)$$

- Contraint  $D$  et  $U$  à être des opérateurs inverses l'un de l'autre.
- Permet un apprentissage conjoint sans paires HR/LR externes.

**(b) Color Consistency Loss ( $L_{CC}$ )**

$$L_{CC} = \|D(\mathbf{r}) - \mathbf{r}_{\downarrow s}\|_1 + \|U(\mathbf{r}_{\downarrow s}) - \mathbf{r}\|_1 \quad (3.2)$$

- $\mathbf{r}$  : Patch uniforme de couleur aléatoire.
- Stabilise l'estimation photométrique et impose une contrainte de préservation de l'intensité moyenne.
- Garantit que  $D$  et  $U$  conservent les propriétés chromatiques.

**3.3.2 Apprentissage Interne**

**Principe :** Le modèle DBPISR adopte une stratégie d'apprentissage interne, dans laquelle l'entraînement est réalisé exclusivement à partir de l'image d'entrée  $I$ , sans recourir à des bases de données externes ni à des paires d'images haute et basse résolution préexistantes.

**Mécanisme :** L'apprentissage repose sur une supervision auto-induite via une contrainte de consistance cyclique entre le réseau de sous-échantillonnage  $D$  et le réseau de super-résolution  $U$ , comme montré à la Fig. 3.1. Deux chemins complémentaires sont considérés : un chemin *down-up*, imposant que la reconstruction  $U(D(I))$  soit cohérente avec l'image originale  $I$ , et un chemin *up-down*, contraignant la redégradation  $D(U(I))$  à être proche de  $D(I)$ . Ces contraintes peuvent s'exprimer comme :

$$U(D(I)) \approx I \quad \text{et} \quad D(U(I)) \approx D(I).$$

Les écarts à ces relations constituent les signaux d'apprentissage utilisés dans la fonction de perte DBP, permettant l'optimisation conjointe des réseaux  $D$  et  $U$ .

**Avantage :** Cette formulation permet au modèle d'estimer explicitement le noyau de dégradation propre à l'image observée, sans hypothèses prédéfinies telles que l'interpolation bicubique, souvent inadéquate face à des dégradations réelles et complexes.

**Robustesse :** En exploitant uniquement les statistiques internes de l'image, le modèle parvient à traiter des dégradations inconnues et réalistes. Cette approche permet également d'éliminer les biais de généralisation liés aux bases de données externes.

**3.3.3 Originalité méthodologique**

DBPISR se distingue par :

- une estimation **conjointe** de la dégradation et de la super-résolution à travers une fonction de perte cyclique auto-supervisée ;
- l'absence de discriminateur adversarial, contrairement aux approches basées sur les GANs ;
- une optimisation *end-to-end* conjointe des réseaux  $D$  et  $U$ , reposant exclusivement sur des contraintes internes de reconstruction et de cohérence fréquentielle, sans recours à une supervision externe ni à un discriminateur adversarial.

Cette approche unifie l'estimation du modèle de dégradation et la restauration haute résolution dans un cadre unique, entièrement adaptatif et spécifique à l'image traitée.

**3.4 Synthèse comparative des modèles SR internes (CNN)**

L'*apprentissage interne* a donné lieu à différentes architectures convolutives, dont ZSSR [163], KernelGAN [13] et DBPISR [90]. Bien que seule l'implémentation de DBPISR ait été intégrée à notre protocole expérimental (cf. Chapitre 4), il est utile de replacer ce modèle dans son contexte méthodologique. Le Tableau 3.1 propose à cet effet une comparaison issue de la littérature.

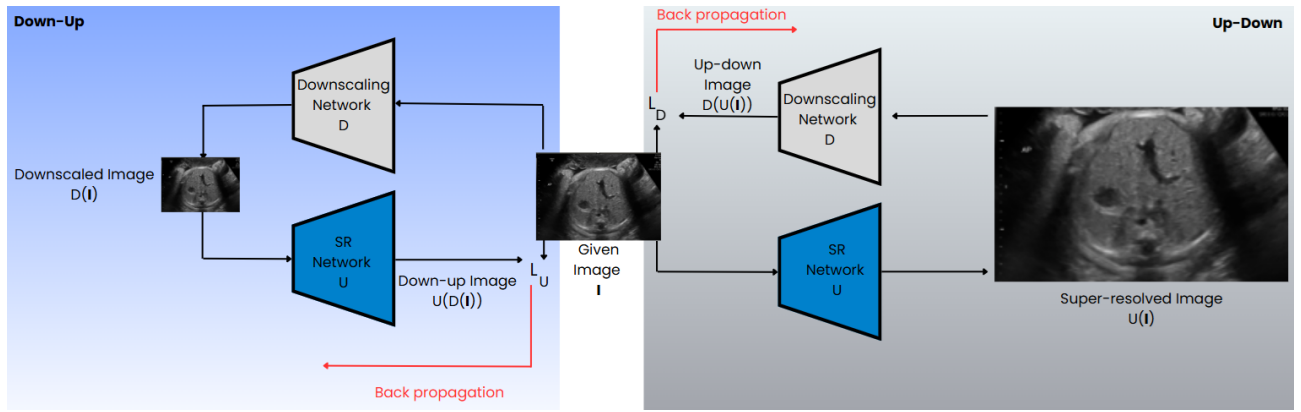


FIGURE 3.1 – Architecture globale du DBPISR[90].

TABLEAU 3.1 – Synthèse comparative des méthodes de super-résolution interne/aveugle (CNN-based).<sup>2</sup>

| Critère                            | ZSSR [163]                                                                                                                    | KernelGAN [13]                                                                         | DBPISR [90]                                                                                                                                                                                                                         |
|------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Innovation principale</b>       | Premier framework d'apprentissage interne (« zero-shot ») pour la SR.                                                         | Estimation du noyau par un GAN léger, basé sur apprentissage interne.                  | Apprentissage <i>joint</i> via double back-projection (Down-Up et Up-Down).                                                                                                                                                         |
| <b>Mécanisme d'estimation</b>      | CNN entraîné par downscaling récursif d'une seule image ; nécessite un noyau fourni.                                          | GAN génératif + CNN linéaire pour imiter la dégradation.                               | Deux CNN couplés (dégradation $D$ et restauration $U$ ) entraînés simultanément.                                                                                                                                                    |
| <b>Forces</b>                      | Simple et efficace si le noyau est connu ; s'adapte à l'image elle-même.                                                      | Véritablement aveugle ; estimation auto-adaptative ; modèle compact.                   | Entraînement conjoint ; robuste aux noyaux inconnus ; $\sim 8\times$ plus rapide que KernelGAN+ZSSR.                                                                                                                                |
| <b>Limites</b>                     | Non aveugle ; dépend fortement du noyau supposé.                                                                              | Procédure en deux étapes séparées (estimation $\rightarrow$ SR) ; erreurs propagées.   | Plus complexe ; nécessite ré-entraînement par image.                                                                                                                                                                                |
| <b>Impact en imagerie médicale</b> | Pas validé sur imagerie médicale dans la littérature ; potentiel pour données rares (échographies isolées, images anciennes). | Pas testé directement en imagerie biomédicale ; reste exploratoire pour l'échographie. | Des tests internes (non publiés) ont confirmé son intérêt en échographie, dans différents contextes cliniques. Validé ensuite sur échographies fœtales [19], avec amélioration significative des scores PSNR/SSIM et MOS cliniques. |

Bien que l'apprentissage interne, illustré par DBPISR, garantisse une grande fidélité au signal d'origine en s'affranchissant de données externes, il reste limité par la pauvreté informationnelle d'une image unique pour reconstruire des textures complexes. Pour pallier ce manque de « prior » textural et atteindre un niveau de détails perceptuels supérieur, il devient nécessaire de solliciter des modèles ayant appris la richesse des structures visuelles sur de vastes bases de données. Ce passage d'une connaissance interne à une expertise externe est porté par

2. Les validations expérimentales internes mentionnées pour DBPISR correspondent à des travaux exploratoires non publiés. Elles sont indiquées ici à titre informatif, sans valeur de publication officielle.

les réseaux adversariaux.

## 3.5 Modèles de super-résolution à base de GAN

L'introduction des réseaux antagonistes génératifs (GAN) a transformé la super-résolution en permettant la synthèse de textures riches et réalistes. En s'appuyant sur un apprentissage compétitif, ces modèles excellent dans la restauration de détails perceptuels fins, essentiels à la qualité visuelle.

Cette section présente trois architectures clés : A-ESRGAN, BSRGAN et Real-ESRGAN. Bien qu'elles partagent un cadre commun, elles se distinguent par leurs stratégies de modélisation des dégradations, visant à résoudre les défis de la super-résolution « aveugle » en conditions réelles.

### 3.5.1 A-ESRGAN

Le modèle A-ESRGAN (Attention-Enhanced Super-Resolution Generative Adversarial Network) [195] est une architecture de réseau antagoniste génératif (GAN) conçue pour la tâche de super-résolution (SR) en conditions réelles et "aveugles" (blind), c'est-à-dire lorsque la dégradation subie par l'image basse résolution est inconnue et complexe.

Comme illustré à la Figure 3.2, A-ESRGAN se distingue principalement par son discriminateur novateur, son générateur reprenant une architecture connue. Le modèle repose sur l'affrontement de ces deux composants : le générateur crée des images HR à partir d'entrées LR, et le discriminateur tente de repérer les faux, en les comparant à des images HR authentiques.

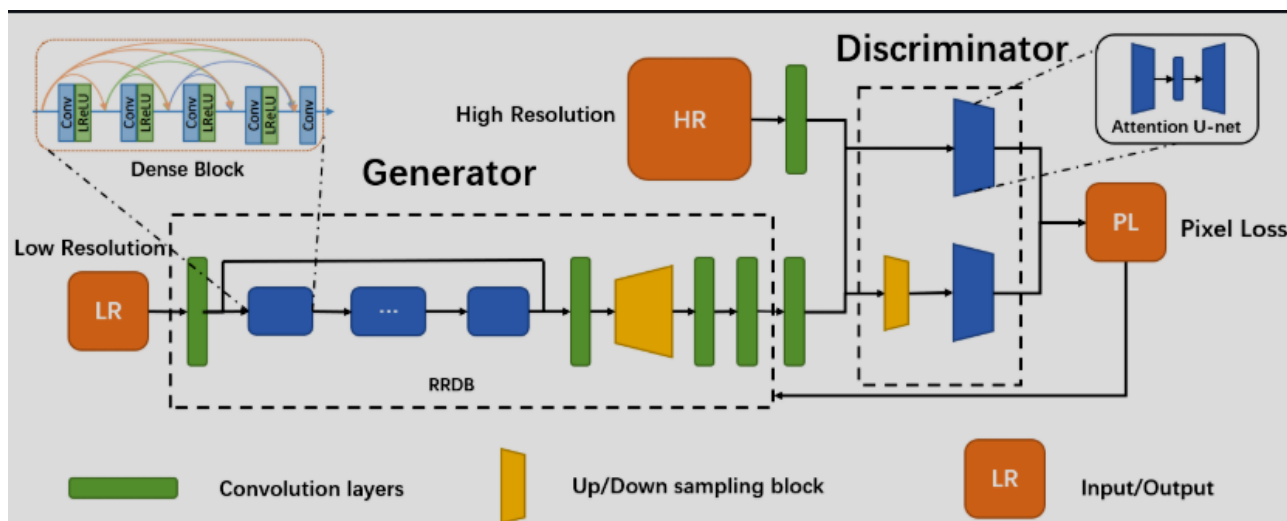


FIGURE 3.2 – Architecture globale de l'A-ESRGAN. Le générateur de l'A-ESRGAN utilise des blocs RRDB, repris du générateur d'ESRGAN [194], pour produire l'image super-résolue (SR) à partir de l'image basse résolution (LR).

Le fonctionnement repose sur une architecture composée de :

1. Le Générateur RRDB :

Le générateur de A-ESRGAN est hérité des travaux précédents (ESRGAN [194], Real-ESRGAN [192]) et se base sur une architecture robuste et performante : le RRDBNet.

- Sa fonction : Convertir une image de faible résolution (LR) en une image reconstruite (SR) aux qualités visuelles améliorées.

- Architecture de base : Il est construit à partir de blocs résiduels résiduels denses (RRDB - Residual-in-Residual Dense Blocks) [194]. Ces blocs empilent de manière dense des connexions résiduelles et de courtes sauts, permettant un flux de gradient très efficace et la capture de caractéristiques complexes à multiples échelles.
  - Intégration : A-ESRGAN utilise un générateur pré-entraîné de Real-ESRGAN, démontrant ainsi que l'amélioration des performances provient bien de son nouveau discriminateur et non d'un générateur plus complexe.
2. Le Modèle de Dégradation :
- Pour entraîner le modèle à gérer des dégradations réalistes, les images LR d'entraînement sont synthétisées à partir d'images HR de référence. A-ESRGAN adopte le modèle de dégradation d'ordre élevé (high-order degradation model) proposé par les auteurs de Real-ESRGAN [192].
- Principe :** Au lieu d'appliquer une seule opération de dégradation simple (comme un flou gaussien suivi d'un sous-échantillonnage bicubique), ce modèle applique de manière séquentielle et aléatoire une série de dégradations (flous, bruit, compression JPEG, etc.). Ce processus itératif permet de simuler de manière beaucoup plus fidèle la dégradation complexe et inconnue des images du monde réel.
3. Le Discriminateur :
- C'est le cœur de la contribution d'A-ESRGAN. Ce discriminateur vise à fournir un feedback plus intelligent au générateur. (Voir Figure 3.3)
- *Structure de base U-Net :*  
Le discriminateur suit une architecture U-Net [159]. Contrairement à un discriminateur classique qui produit un seul score de "réalisme" pour l'image entière, le U-Net a une structure (encodeur, décodeur) avec des connexes résiduelles. Plutôt que de fournir une évaluation globale de l'image (« vraie » ou « fausse »), il génère une carte de prédictions de mêmes dimensions que l'image ( $W \times H$ ), où chaque position indique localement le degré de réalisme perçu. Cette sortie fine et spatialement située offre au générateur un signal d'erreur détaillé, lui permettant de corriger précisément les structures incohérentes ou "les artefacts", et ainsi d'améliorer simultanément la finesse des détails et la cohérence d'ensemble de l'image synthétisée.
  - *Intégration du Mécanisme d'Attention :*  
Des blocs d'attention (Attention Blocks - AB) sont intégrés dans les connexes résiduelles du U-Net. Inspirés des travaux sur la segmentation d'images médicales (Attention U-Net), ces blocs sont adaptés pour les images 2D [136].  
— Fonction :  
Le mécanisme d'apprentissage permet au discriminateur d'apprendre automatiquement à "regarder" les régions les plus informatives de l'image, typiquement les contours et les textures où les changements de couleur sont brutaux.  
— Avantage :  
Au lieu de traiter tous les pixels de manière égale, le discriminateur accorde un poids plus important (attention weight) aux caractéristiques structurelles essentielles (bords, lignes, textures). Le générateur parvient à restituer ainsi des détails plus nets et d'éviter des artefacts tels que les distorsions linéaires ou un arrière-plan sur-appris.
  - *Architecture Multi-Échelle (Multi-Scale Discriminator) :*  
Pour capter des caractéristiques à différents niveaux de granularité, A-ESRGAN utilise non pas un, mais deux discriminateurs Attention U-Net identiques en parallèle, formant un discriminateur multi-échelle.
    - Discriminateur "Normal" ou "pleine résolution" (D\_normal) :  
Prend l'image SR en pleine résolution en entrée. Sa fonction exclusive repose sur l'évaluation de la précision des textures et des contours.

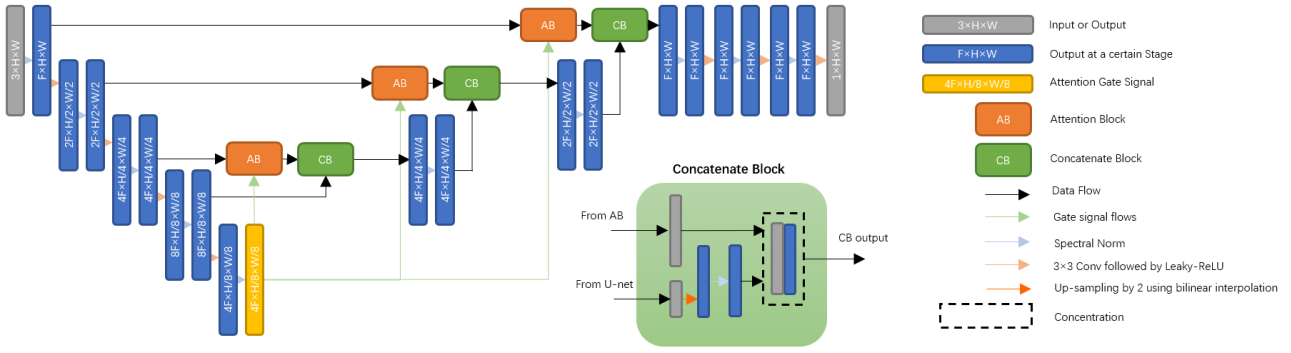


FIGURE 3.3 – Discriminateur U-Net avec attention simple. L'architecture d'un discriminateur U-Net avec attention simple.  $F$ ,  $W$  et  $H$  représentent respectivement le nombre de canaux de sortie de la première couche de convolution, la hauteur de l'image et la largeur de l'image.

- Discriminateur "Sous-échantillonné" ( $D_{\text{sampling}}$ ) :

Prend une version 2x sous-échantillonnée de l'image SR en entrée. Cette image redimensionnée ayant des contours plus flous, le discriminateur est forcé de se concentrer sur des structures plus larges et des textures globales.

**Bénéfice :**

Cette dualité permet au modèle de superviser à la fois la finesse des détails locaux et la cohérence des structures globales, améliorant ainsi le réalisme perceptuel.

#### 4. Fonctions de Perte (*Loss Functions*) :

La fonction de perte totale pour entraîner le générateur est une combinaison de plusieurs termes :

*Perte GAN ( $L_G$ )* : Déduite de la sortie des deux discriminateurs. Elle pousse le générateur à tromper les discriminateurs.

*Perte de Perception ( $L_{\text{percep}}$ )* : Basée sur les caractéristiques extraites par un réseau de classification (comme VGG), elle mesure la similarité sémantique entre l'image SR et l'image HR.

*Perte L1 ( $L_1$ )* : Une perte de distance pixel-à-pixel simple qui stabilise l'entraînement et assure une reconstruction fidèle de bas niveau. La perte finale est une somme pondérée de ces termes :

$$l_G = L_{\text{percep}} + \lambda_1 L_{G_{\text{normal}}} + \lambda_2 L_{G_{\text{sampling}}} + \eta L_1 \quad (3.3)$$

### 3.5.2 BSRGAN

Le BSRGAN (Blind Super-Resolution Generative Adversarial Network)[206] est un modèle conçu pour réaliser la super-résolution d'images en conditions réelles, c'est-à-dire sans connaître le type de dégradation (flou, bruit, compression, etc.) présente dans l'image basse résolution. BSRGAN est construit à partir de l'ESRGAN (Enhanced Super-Resolution GAN)[194], un modèle existant réputé pour la super-résolution perceptuelle. Les auteurs reprennent cette architecture et l'améliorent en utilisant un nouveau modèle de dégradation réaliste pour générer des données d'entraînement variées.

L'architecture a été construite en deux parties :

#### A. Modèle de dégradation probabiliste "Cœur de l'innovation"

Le BSRGAN ne modifie pas radicalement l'architecture du réseau neuronal, mais révolutionne la manière dont les données d'entraînement sont générées. (Voir Figure 3.5 )

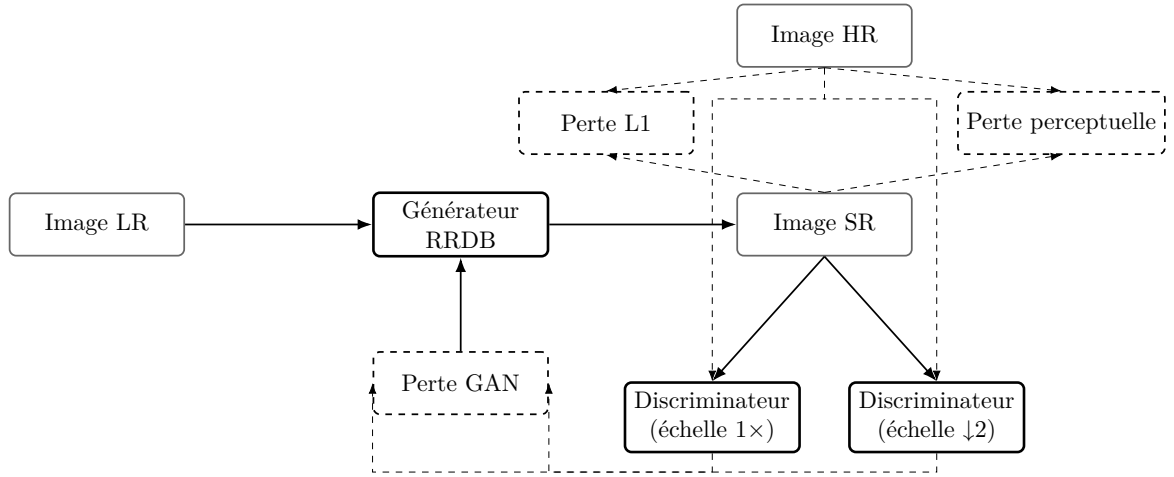


FIGURE 3.4 – Schéma bloc d'A-ESRGAN. L'image basse résolution (LR) est traitée par le générateur basé sur des blocs RRDB, qui produit l'image super-résolue (SR). Cette sortie est comparée à l'image haute résolution de référence (HR) au moyen de deux pertes de reconstruction : la perte L1 et la perte perceptuelle. En parallèle, l'image SR est évaluée par deux discriminateurs multi-échelles : un discriminateur à l'échelle originale ( $1\times$ ) et un autre après réduction par un facteur deux ( $\downarrow 2$ ), ce qui permet de juger à la fois les détails fins et la cohérence globale. Les sorties des discriminateurs sont combinées dans une perte adversariale (GAN loss), qui rétroagit sur le générateur afin d'améliorer le réalisme perceptuel de l'image reconstruite.

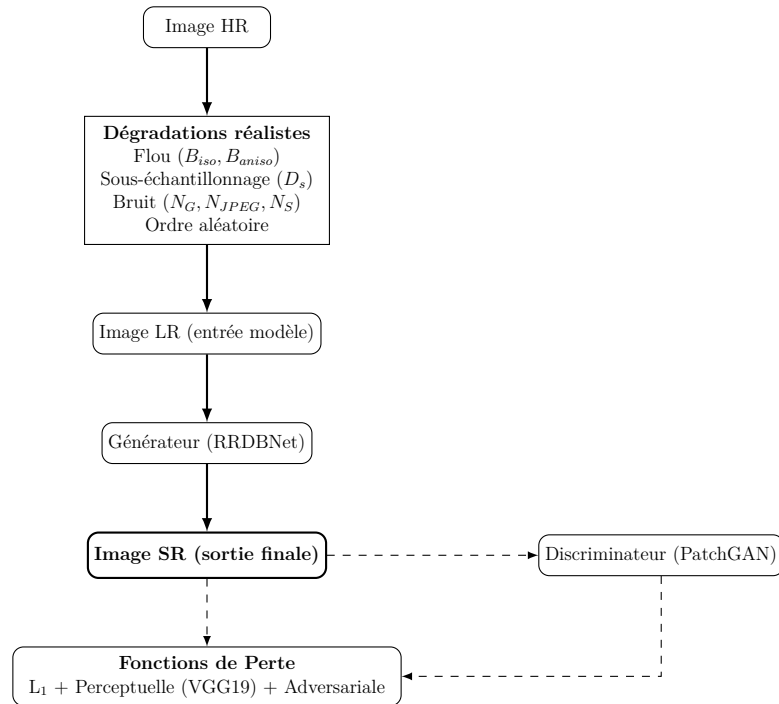


FIGURE 3.5 – Schéma bloc de BSRGAN. À partir d'une image HR, un modèle de dégradations réalistes (flou, sous-échantillonnage, bruit, ordre aléatoire) génère des images LR synthétiques qui servent d'entrée au générateur. Celui-ci reconstruit une image SR, qui constitue la sortie finale. L'image SR est ensuite évaluée par le discriminateur et comparée à la référence HR à l'aide des fonctions de perte, incluant la perte  $L_1$ , la perte perceptuelle (VGG19) et la perte adversariale. Ainsi, le modèle apprend à produire des reconstructions fidèles, perceptuellement cohérentes et visuellement réalistes.



Pour chaque image haute résolution (HR), on synthétise une image basse résolution (LR) via une séquence aléatoire de dégradations :

- **Flou** : Convolution avec des noyaux gaussiens isotropes et anisotropes (tailles et paramètres aléatoires).
- **Sous-échantillonnage** : Méthodes variées (plus proche voisin, bilinéaire, bicubique).
- **Bruit** : Bruit gaussien additif (AWGN), compression JPEG (qualité variable) et bruit de capteur simulé via un pipeline inverse de traitement d'image (ISP).

Ces opérations sont mélangées aléatoirement, créant un espace de dégradation immense et réaliste.

## B. Réseau de super-résolution (Héritage de l'ESRGAN)

Le modèle de super-résolution lui-même reprend l'architecture de l'ESRGAN, un GAN robuste et éprouvé, composé de :

- **Générateur (G)** : Le générateur du BSRGAN repose sur une architecture de type RRDB (Residual-in-Residual Dense Block) ; Ces blocs permettent un apprentissage profond stable grâce à des connexions résiduelles emboîtées, spécialement conçue pour les tâches de restauration d'images. Son rôle est de transformer une image basse résolution (LR) en une image super-résolue (SR) de haute qualité perceptuelle. La composition est donnée comme suit :

- a) Blocs RRDB (Residual-in-Residual Dense Blocks) : Le bloc RRDB organise plusieurs sous-blocs denses au sein d'une structure résiduelle principale. Cette configuration assure une circulation robuste du gradient et une réutilisation multi-échelle des caractéristiques, optimisant l'apprentissage de relations hiérarchiques entre motifs locaux et contextes globaux. L'architecture residual-in-residual contrecarre l'affaiblissement du gradient, permettant le déploiement de réseaux profonds à l'entraînement stable.
- b) Absence de Normalisation par Lots ou *BatchNorm* : Contrairement à de nombreuses architectures profondes, les blocs RRDB excluent délibérément la normalisation par lots.

**Motivation** : En synthèse d'images, la normalisation par lots induit fréquemment des artefacts génératifs et contraint la flexibilité du modèle en normalisant excessivement les caractéristiques. Son ablation préserve la richesse informationnelle des représentations latentes et accroît la fidélité perceptuelle des sorties.

- c) Module de Sur-échantillonnage (Upsampling) : L'augmentation de la résolution spatiale est réalisée à l'aide de la méthode PixelShuffle (également appelée sub-pixel convolution).

**Fonctionnement** : Cette opération réarrange les cartes de caractéristiques profondes en une image haute résolution en agrégeant efficacement les informations contextuelles sans perte de netteté, contrairement aux méthodes d'interpolation classiques (e.g., bicubique). *Cette conception permet au générateur de combiner précision pixelique et richesses texturales, tout en restant robuste grâce à sa structure résiduelle profonde et à l'absence de couches de normalisation contre-productives.*

- **Discriminateur (D)** :

Le discriminateur de BSRGAN adopte une architecture de type *PatchGAN*, initialement introduite par Isola et al. [77], qui évalue le réalisme à l'échelle locale plutôt que globale. Il repose sur une succession de convolutions à pas (*stride*) appliquées à l'image d'entrée et produit une carte de sortie bidimensionnelle, dans laquelle chaque élément correspond à la probabilité qu'un patch local de l'image soit perçu comme réel.

Cette approche locale est essentielle pour capturer les statistiques des textures et des hautes fréquences, qui conditionnent le réalisme perceptuel. Une innovation clé est l'utilisation systématique de la normalisation spectrale sur les poids de chaque couche. Cette technique de régularisation contraint la norme Lipschitz du discriminateur, stabilisant ainsi dynamiquement l'entraînement adversarial. Elle prévient les gradients explosifs, améliore la convergence et évite l'effondrement modal, garantissant un signal d'apprentissage stable et informatif pour le générateur.

– **Fonctions de Perte et Objectif d'Optimisation :**

Le modèle est optimisé via une fonction de coût composite combinant trois objectifs distincts :

- \* Une perte  $L_1$  garantissant la fidélité spatiale et radiométrique de base, par la minimisation de l'écart absolu moyen ;
- \* Une perte perceptuelle, évaluée dans l'espace latent d'un extracteur de caractéristiques (VGG-19), visant l'exactitude sémantique et texturale ;
- \* Une perte antagoniste facilitant l'émergence de micro-structures réalistes par alignement distributionnel.

*Cette combinaison permet au BSRGAN de concilier fidélité pixelique, réalisme perceptuel et richesse texturale, conduisant à de meilleures performances sur des images réelles aux dégradations complexes et inconnues.*

### 3.5.3 Real-ESRGAN

Real-ESRGAN [192] est une évolution du modèle ESRGAN [194], s'inscrivant dans une lignée de travaux — dont BSRGAN [206] est un précurseur immédiat — visant à généraliser la super-résolution à des conditions réalistes et non contrôlées.

Comme BSRGAN, Real-ESRGAN cherche à résoudre le problème de la super-résolution aveugle en conditions réelles, où les dégradations (flou, bruit, compression, artefacts) sont inconnues et complexes. Cependant, Real-ESRGAN introduit une modélisation de dégradations d'ordre supérieur (high-order degradation) et intègre explicitement des filtres sinc pour simuler des artefacts courants comme les rebonds et les dépassements, afin de couvrir un espace de dégradations plus large et plus réaliste que les approches précédentes.

L'architecture de Real-ESRGAN repose sur deux réseaux principaux :

A. **Générateur : Architecture RRDB améliorée**

Le générateur reprend la structure RRDB (Residual-in-Residual Dense Blocks) d'ESRGAN, mais avec des optimisations spécifiques pour gérer la complexité des dégradations réalistes :

- Blocs RRDB : Chaque bloc combine des connexions résiduelles multiples et des connexions denses à l'intérieur des résidus, permettant un flux de gradient plus stable et une capture de détails à multiples échelles.
- Mécanisme : Les connexions denses facilitent la réutilisation des caractéristiques et améliorent la représentation des textures complexes.
- Pixel-Unshuffle en entrée : Opération inverse du pixel-shuffle, réduisant la résolution spatiale tout en augmentant la dimension des canaux.

**Avantage :** Diminue la complexité computationnelle en travaillant dans un espace

---

1. Real-ESRNet est une version préliminaire de Real-ESRGAN, optimisée pour la fidélité numérique (PSNR) au moyen d'une fonction de perte  $L_1$ . Il conserve la même architecture RRDB et le même processus de dégradation que Real-ESRGAN, mais sans apprentissage adversarial. Son rôle est double : (i) servir de point d'initialisation stable pour l'entraînement de Real-ESRGAN ; (ii) offrir une alternative pour les applications nécessitant une reconstruction précise et peu artefactée. Il illustre ainsi le compromis fondamental entre fidélité numérique et qualité visuelle.

2. C'est un problème classique quand le GAN force trop la création de détails qui n'existaient pas à l'origine. La normalisation spectrale stabilise l'apprentissage en bornant la norme des poids, atténuant la sensibilité aux artefacts haute fréquence pour une génération d'images plus réaliste.

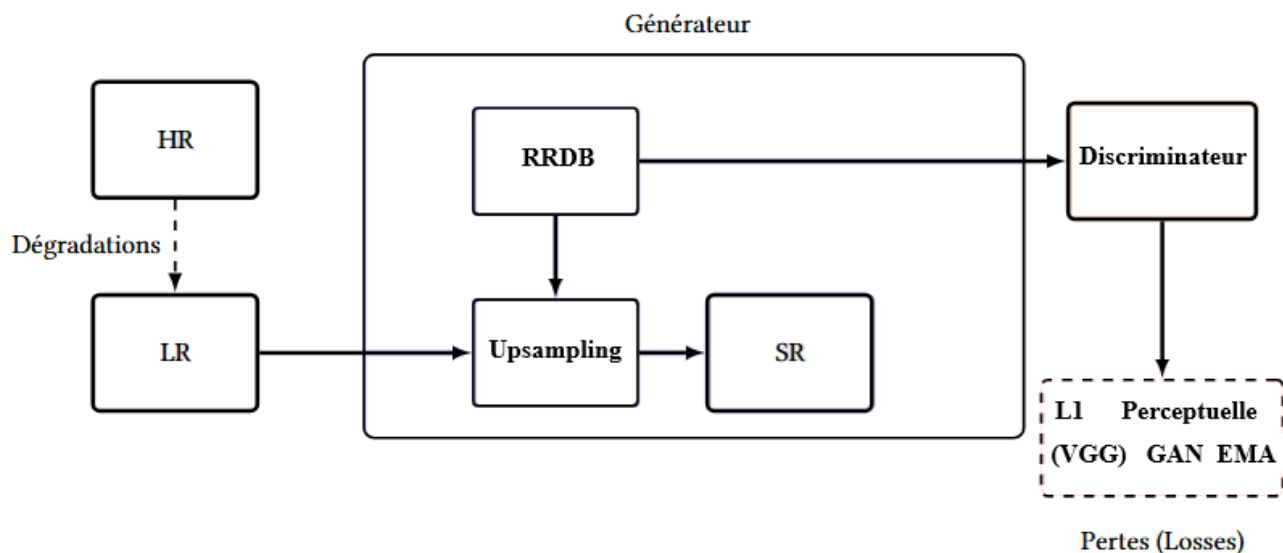


FIGURE 3.6 – Schéma bloc de Real-ESRGAN. Le modèle utilise un générateur basé sur les blocs RRDB (Residual-in-Residual Dense Blocks) avec un module d'upsampling pour reconstruire l'image SR à partir de l'entrée LR. Le discriminateur de type U-Net, régularisé par normalisation spectrale, compare les images SR et HR. L'entraînement repose sur un pipeline de dégradations en deux passes pour générer les LR réalistes, et combine plusieurs pertes (contenu L1, perceptuelle, adversariale et EMA).

de résolution réduite, tout en préservant l'information via une représentation canal plus riche.

- Initialisation et stabilisation : Le générateur est initialisé avec les poids d'un modèle préentraîné orienté PSNR (Real-ESRNet<sup>3</sup>), accélérant la convergence et fournissant une base stable pour l'apprentissage adversarial.

### B. Discriminateur U-Net à Normalisation Spectrale :

Pour répondre à la complexité des dégradations synthétiques, le discriminateur est repensé avec une architecture de type U-Net et des mécanismes de régularisation avancés :

- Architecture U-Net :
  - Skip connections : Relient les encodeurs et décodeurs, permettant une préservation fine des détails locaux et une discrimination spatialement précise.
  - Sortie pixel-à-pixel : Produit une carte de "réalité" pour chaque pixel, offrant un retour d'information granulaire au générateur. Les artefacts locaux sont ainsi corrigés tels que le rebond ou du bruit sans altérer les régions intactes.
- Normalisation Spectrale (SN) :
  - Principe : La normalisation spectrale régularise le discriminateur en bornant la valeur singulière maximale des poids de chaque couche, réduisant ainsi sa sensibilité aux oscillations haute fréquence et stabilisant l'apprentissage adversarial.
  - Impact : Stabilise la dynamique adversariale en empêchant le discriminateur de diverger trop rapidement. Réduit les artefacts générés par le GAN (e.g., textures sur-aiguës<sup>4</sup> ou fantômes) en régulant les gradients.
- Régularisation supplémentaire :
 

L'Exponential Moving Average (EMA) lisse les poids du générateur via une moyenne pondérée des paramètres durant l'entraînement, réduisant la sensibilité aux oscillations et améliorant la stabilité convergence et la robustesse du modèle.

### 3.6 Critères de comparaison des générateurs GAN : A-ESRGAN, BSRGAN, Real-ESRGAN

Afin d'évaluer la pertinence de chaque modèle pour la tâche de super-résolution aveugle sur images échographiques, une comparaison est établie sur la base de critères techniques tels que le modèle de dégradation, l'architecture et la robustesse.

TABLEAU 3.2 – Forces, limitations et impact médical des modèles GAN.

| Modèle             | Forces                                                                                                     | Limitations                                                                                                                                  | Impact en imagerie médicale                                                                                                                                                                                |
|--------------------|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>BSRGAN</b>      | Gère dégradations réalistes et inconnues; robuste au flou, bruit et artefacts                              | Conçu initialement pour images naturelles, pas pour le médical; peut introduire artefacts indésirables; mauvaise gestion du faible contraste | Après adaptation (fine-tuning) sur CT, obtient PSNR $\approx 36.5$ , SSIM $\approx 0.927$ [100]; améliore la lisibilité anatomique mais reste moins fiable cliniquement que SwinIR/HAT                     |
| <b>Real-ESRGAN</b> | RRDB + U-Net Discriminateur; normalisation spectrale (stabilité); bonne préservation des bords et textures | Développé pour images naturelles; complexité computationnelle élevée; tendance à lisser les structures fines dans tissus denses              | Après transfert d'apprentissage, atteint PSNR $\approx 34.7$ , SSIM $\approx 0.915$ [100]; robuste sur plusieurs modalités (IRM, RX, échographies, dermatologie) mais résultats plus lissés que SwinIR/HAT |
| <b>A-ESRGAN</b>    | Introduit l'attention (longues dépendances); meilleure reconstruction de détails fins                      | Plus lourd que Real-ESRGAN; instable à l'entraînement; entraînement coûteux; peu exploré en contexte médical                                 | Pas encore validé spécifiquement en imagerie médicale; potentiel pour améliorer détails fins mais risque de surajustement et de perte de fiabilité clinique                                                |

TABLEAU 3.3 – Synthèse comparative des architectures GAN pour la super-résolution aveugle et leur impact médical.

| Critère                        | BSRGAN [206]                                                                                                                  | Real-ESRGAN [192]                                                                      | A-ESRGAN [195]                                                               |
|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|------------------------------------------------------------------------------|
| <b>Innovation principale</b>   | Dégradations aléatoires réalistes (bruit, flou, compression)                                                                  | Dégradation réaliste + discriminateur U-Net multi-niveaux                              | Discriminateur multi-échelle pour préserver détails fins                     |
| <b>Modèle de dégradation</b>   | Dégradations stochastiques (flou anisotrope, bruit RAW/JPEG, ré-échantillonnage)                                              | Dégradations complexes avec filtres sinc, bruit et compression                         | Inspiré de Real-ESRGAN mais ciblé textures complexes                         |
| <b>Architecture générative</b> | Générateur RRDB + discriminateur classique                                                                                    | Générateur RRDB + U-Net discriminateur (normalisation spectrale)                       | Générateur RRDB + U-Net discriminateur avec attention                        |
| <b>Forces</b>                  | Robuste aux dégradations imprévisibles et bruit réels                                                                         | Bon compromis réalisme / stabilité                                                     | Excellente préservation détails fins et structures complexes                 |
| <b>Limites</b>                 | Reconstruction parfois aléatoire; artefacts texture                                                                           | Artefacts subtils; moins précis pour détails fins                                      | Complexité et coût élevés; instabilités possibles                            |
| <b>Usage typique</b>           | Applications robustes face à conditions réelles variées                                                                       | Images naturelles dégradées (internet, photo)                                          | Contextes critiques nécessitant fidélité structurale (ex. imagerie médicale) |
| <b>Impact médical</b>          | Fine-tuning CT : PSNR $\approx 36.5$ , SSIM $\approx 0.927$ [100]; lisibilité anatomique améliorée, moins fiable cliniquement | PSNR $\approx 34.7$ , SSIM $\approx 0.915$ sur IRM/RX/échographies; robuste mais lissé | Pas encore validé; prometteur pour détails anatomiques, risque surajustement |

Comme le met en évidence le Tableau 3.2, le choix d’une architecture de super-résolution pour l’échographie nécessite de pondérer trois exigences fondamentales : la préservation des structures anatomiques, la résistance aux artefacts de reconstruction, et la faisabilité computationnelle.

Le Tableau 3.3 présente une synthèse comparative des architectures GAN pour la super-résolution aveugle, détaillant leurs innovations, modèles de dégradation, forces, limites et impact médical, afin de faciliter le choix du modèle le plus adapté aux applications échographiques.

L’analyse de ces résultats positionne BSRGAN comme une solution de robustesse, capable de gérer une large gamme de dégradations grâce à son modèle stochastique — une propriété cruciale face à la variabilité des acquisitions cliniques. Real-ESRGAN offre quant à lui un compromis opérationnel remarquable, alliant un modèle de dégradation réaliste à une stabilité d’entraînement avérée. Enfin, A-ESRGAN se distingue par la restitution des détails fins et des interfaces tissulaires via son mécanisme d’attention, bien que cela s’accompagne d’une charge computationnelle plus élevée.

Toutefois, si ces réseaux adversariaux repoussent les limites du photoréalisme, ils peuvent introduire des artefacts ou une instabilité structurelle préjudiciable au diagnostic médical. L’émergence des *Transformers* en super-résolution offre une alternative majeure : en conciliant la puissance de l’apprentissage externe et une modélisation des relations spatiales globales via le mécanisme d’attention, ces modèles assurent une reconstruction plus rigoureuse des formes et limitent les distorsions géométriques inhérentes aux approches purement convolutives.

## 3.7 Modèles de super-résolution à base de Transformers

Si les GAN ont marqué une étape clé par leur capacité générative, l’avènement des *Transformers* en vision par ordinateur a ouvert une nouvelle voie pour la super-résolution. Contrairement aux convolutions qui traitent l’image de manière locale, le mécanisme d’auto-attention dynamique permet de capturer, dès les premières couches, des interactions globales dépendantes du contenu.

Cette capacité de modélisation globale est cruciale pour préserver la morphologie tissulaire face aux artefacts de reconstruction. Néanmoins, la transposition de ces mécanismes d’attention aux contraintes de la super-résolution demeure un défi technique majeur.

L’exemple d’IPT [27] illustre bien les limites de l’approche par patches indépendants, source potentielle d’artefacts de bordure et d’une fragmentation du contexte. En outre, la complexité quadratique inhérente à l’attention globale impose un coût computationnel et une dépendance aux données ( $> 10^6$  images) souvent incompatibles avec les réalités cliniques.

C’est pour pallier ces verrous technologiques que de nouvelles architectures ont émergé :

**SwinIR (1<sup>re</sup> génération)** [105] introduit l’attention locale par fenêtres inspirée du Swin Transformer, réduisant le coût computationnel tout en préservant la capacité à modéliser le contexte local. Bien que plus efficace que les CNN, son champ de perception reste limité aux fenêtres locales, ce qui peut générer des effets de lisière.

**Swin2SR et HAT (2<sup>e</sup> génération)** [35, 30, 29] améliorent cette approche : Swin2SR utilise une attention glissante pour fusionner les informations entre fenêtres et réduire les artefacts de jonction, tandis que HAT (*Hybrid Attention Transformer*) combine plusieurs mécanismes d’attention — grande fenêtre, attention par canal, attention chevauchante — pour mieux capturer à la fois les dépendances locales et globales.

La trajectoire **SwinIR**  $\rightarrow$  **Swin2SR**  $\rightarrow$  **HAT** illustre clairement l'évolution progressive de la super-résolution basée sur Transformers : chaque modèle reprend les fondations du précédent, affine son architecture et étend ses capacités pour concilier performance, efficacité et généralisabilité.

### 3.7.1 SwinIR

SwinIR est un modèle de restauration d'image fondé sur l'architecture Swin Transformer, conçu pour reconstruire des images haute qualité à partir d'entrées dégradées (sous-échantillonnées, bruitées ou compressées). Il est utilisé pour des tâches de : super-résolution, de débruitage et de réduction d'artefacts de compression. Il convient de souligner que le modèle exploite un mécanisme d'auto-attention reposant sur des fenêtres fixes et décalées, ce qui lui confère la capacité de prendre simultanément en compte les dépendances locales et globales. En comparant avec les méthodes convolutionnelles classiques, cette stratégie se singularise par une meilleure aptitude à représenter les relations spatiales, de manière flexible et indépendante du contenu.

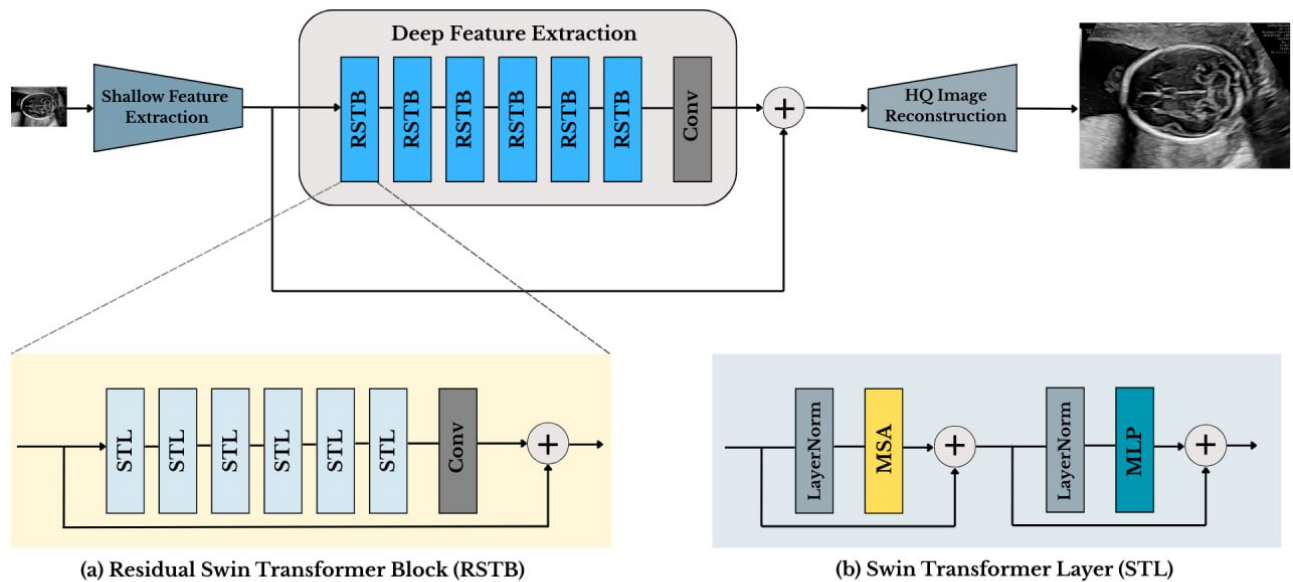


FIGURE 3.7 – Architecture globale de SwinIR. Son architecture repose sur un pipeline en trois étapes : une extraction initiale des caractéristiques de bas niveau via une convolution simple, suivie d'une extraction de caractéristiques profondes par une série de blocs résiduels Swin Transformer combinant attention locale par fenêtres et convolutions pour modéliser les relations contextuelles complexes, et enfin une étape de reconstruction où les caractéristiques profondes et de bas niveau sont fusionnées via une connexion résiduelle pour produire une image SR à la fois détaillée et structurellement fidèle à l'image HR.

Le modèle SwinIR est structuré en trois modules principaux :

### A. Extraction des caractéristiques de bas niveau ou *Shallow Feature Extraction*

Une couche de convolution  $3 \times 3$  est appliquée afin d'extraire des caractéristiques de bas niveau notées  $F_0$ . Les caractéristiques sont directement transmises au module de reconstruction afin que soient préservées les informations de basses fréquences.

**B. Extraction des caractéristiques profondes ou *Deep Feature Extraction*** Ce module est composé de  $K$  blocs résiduels Swin Transformer (*Residual Swin Transformer Blocks, RSTB*).

#### b.1 Residual Swin Transformer Block (RSTB)

Chaque RSTB est une unité fondamentale conçue pour transformer progressivement les caractéristiques. Sa structure interne est la suivante :

- (i) **Couches Swin Transformer Layer (STL)** : Cœur du bloc, elles calculent l'attention locale au sein de fenêtres, avec un mécanisme de décalage (*shifting*) entre les couches pour permettre la communication inter-fenêtres.
- (ii) **Convolution  $3 \times 3$  (en fin de bloc) : Niveau 1 - Convolution Intra-RSTB**
  - **Emplacement** : Appliquée après la séquence de  $L$  couches STL, juste avant la connexion résiduelle.
  - **Rôle** : Agréger les caractéristiques sortant des STL et réintroduire un biais d'induction translationnelle (*translation equivariance*), crucial pour la stabilité et les performances.
- (iii) **Connexion résiduelle** : La sortie de la convolution est additionnée à l'entrée du bloc pour faciliter la propagation du gradient et stabiliser l'apprentissage.

#### Remarque : Convolution de fusion finale

Une couche de convolution  $3 \times 3$  supplémentaire est appliquée à la sortie du dernier RSTB, jouant un rôle distinct de celles situées à l'intérieur de chaque bloc (voir figure 3.7a).

- **Emplacement** : En sortie de la pile des  $K$  RSTB.
- **Rôle** : Agréger et fusionner l'ensemble des caractéristiques profondes et hiérarchiques extraites par la séquence complète de blocs. Cette opération produit la représentation finale des caractéristiques profondes, notée  $F_{DF}$ , optimisée pour la reconstruction.
- **Distinction** : Contrairement aux convolutions *intra*-RSTB qui opèrent localement, cette convolution *extra*-RSTB effectue une fusion globale sur l'ensemble du jeu de caractéristiques transformées.

Le RSTB est une unité hybride qui combine la puissance de modélisation contextuelle du Transformer (via les STL) avec la robustesse et l'efficacité de la convolution, le tout stabilisé par une connexion résiduelle. L'empilement de plusieurs RSTB permet au modèle d'extraire des caractéristiques hiérarchiques de plus en plus riches et abstraites pour la tâche de reconstruction.

**C. Reconstruction** Les caractéristiques de bas niveau ( $F_0$ ) et les caractéristiques profondes fusionnées ( $F_{DF}$ ) sont combinées via une connexion longue :  $F = F_0 + F_{DF}$ .

$$F = F_0 + F_D.$$

- Pour la super-résolution : utilisation d'une convolution sub-pixel pour le rééchantillonnage,
- Pour le débruitage et la réduction d'artefacts : apprentissage résiduel suivi d'une convolution simple.

### 3.7.2 Swin2SR

Swin2SR est un modèle basé sur Swin Transformer V2, conçu spécifiquement pour la super-résolution d'images compressées et la restauration d'images (notamment la suppression d'artefacts de compression JPEG). Il s'inscrit dans la lignée des méthodes Transformers appliquées à la vision, et constitue une évolution de SwinIR, en tirant parti des améliorations apportées par SwinV2 pour une meilleure stabilité à l'entraînement, une capacité de généralisation à des résolutions plus élevées, et une réduction de la faim en données.

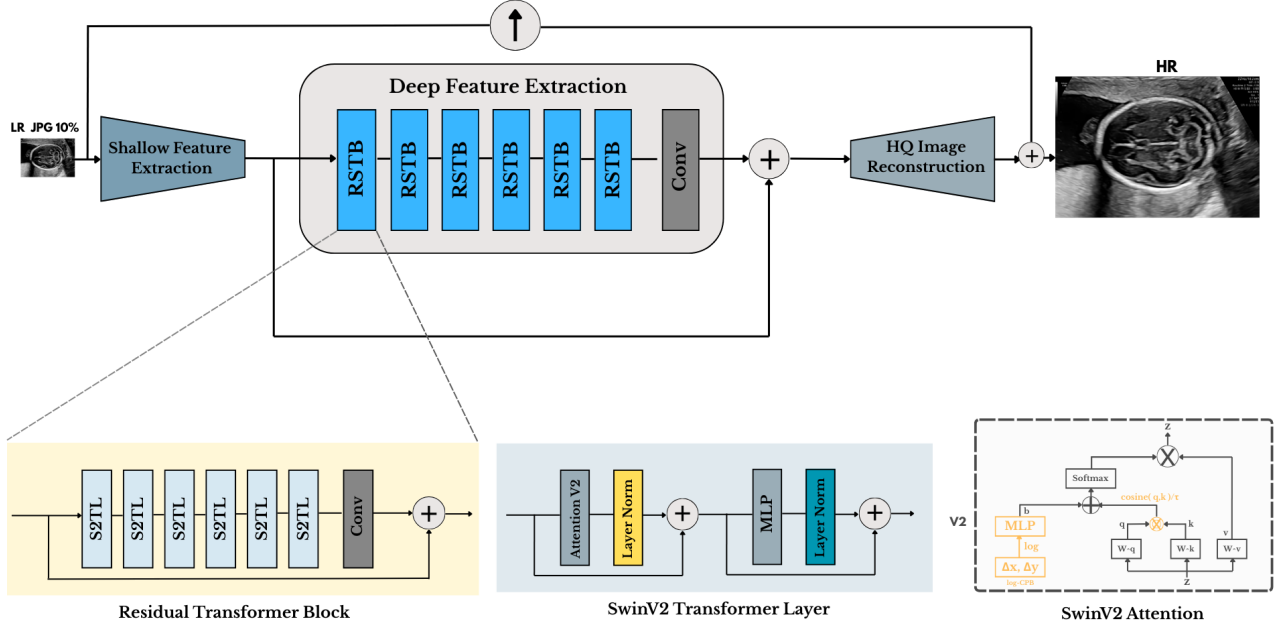


FIGURE 3.8 – Architecture générale de Swin2SR.[35] Le modèle prend en entrée une image basse résolution compressée (LR). Après extraction et traitement des caractéristiques via des blocs transformers (RSTB) avec attention SwinV2, il génère en sortie une image super-résolue de haute qualité (SR/HR), restaurée et détaillée.

L'architecture de Swin2SR repose sur trois modules principaux :

#### A. Extraction des caractéristiques de bas niveau ou *Shallow Feature Extraction*

Une couche convolutive extrait les caractéristiques de bas niveau de l'image d'entrée dégradée, transmises directement au module de reconstruction afin de préserver les basses fréquences.

**B. Extraction des caractéristiques profondes ou *Deep Feature Extraction*** Cœur du modèle, constitué de blocs résiduels SwinV2 Transformer (RSTB). Chaque RSTB contient plusieurs couches SwinV2 Transformer (S2TL) qui utilisent un mécanisme d'auto-attention par fenêtres décalées (shifted window self-attention). Ce mécanisme permet de modéliser à la fois les dépendances locales et les interactions globales entre différentes régions de l'image, sans la complexité computationnelle quadratique de l'attention standard.



**C. Reconstruction d'image de haute qualité** Fusionne les caractéristiques superficielles et profondes, puis utilise une opération de pixel shuffle pour effectuer le suréchantillonnage (upsampling) et reconstruire l'image haute résolution. La sortie du modèle est ajoutée à une version préalablement upscalée (bicubique) de l'image d'entrée pour améliorer les détails.

**Note : Améliorations clés par rapport à SwinIR**

- **SwinV2 Blocks** : utilisation de blocs Swin Transformer V2 introduisant :
  - Post-normalisation au lieu de la pré-normalisation, améliorant la stabilité numérique.
  - Attention cosinus *scaled* remplaçant le produit scalaire, réduisant la dominance de certaines têtes d'attention.
  - Biais de position relatif log-spacé continu, permettant une meilleure généralisation à des résolutions non vues pendant l'entraînement.
- **Fonctions de perte avancées**
  - Pour la super-résolution d'images compressées, des fonctions de perte supplémentaires sont explorées :
    - **Perte auxiliaire (AuxLoss)** : minimise la différence entre la prédiction sous-échantillonnée et la référence sous-échantillonnée, assurant la cohérence à basse résolution.
    - **Perte haute fréquence (HFLoss)** : force le modèle à conserver les détails haute fréquence en comparant les informations HF de la prédiction et de la référence.

### 3.7.3 HAT

Le Transformer à Attention Hybride (HAT) introduit une évolution architecturale majeure des Transformers pour la vision par ordinateur, en optimisant leur conception pour les tâches de restauration d'image (voir Figure 3.9). Ce modèle a été spécifiquement conçu pour surmonter les limitations observées dans les implémentations existantes de Transformers, telles que SwinIR, qui présentaient une exploitation sous-optimale de l'information spatiale et des interactions restreintes entre fenêtres locales.

HAT suit une structure *Residual in Residual (RIR)* similaire à celle de SwinIR, divisée en trois parties principales :

#### A. Extraction des caractéristiques de bas niveau ou *Shallow Feature Extraction*

Une couche de convolution  $3 \times 3$  mappe l'image d'entrée basse qualité (LQ) vers un espace de features à haute dimension.

#### B. Extraction des caractéristiques profondes ou *Deep Feature Extraction*

Composée de plusieurs Residual Hybrid Attention Groups (RHAG). Chaque RHAG inclut :

- **Hybrid Attention Blocks (HAB)** : Basés sur Swin Transformer, mais avec une taille de fenêtre élargie (16 au lieu de 8), une attention multi-tête par fenêtre (W-MSA) et décalée (SW-MSA), combinée en parallèle avec un Channel Attention Block (CAB). Le CAB utilise des convolutions avec compression de canaux (facteur  $\beta = 3$ ) et une attention par canal pour capturer des informations globales, pondérée par un facteur  $\alpha$  faible (0.01) pour éviter les conflits d'optimisation.
- **Overlapping Cross-Attention Block (OCAB)** : Introduit une attention croisée avec chevauchement (ratio  $\gamma = 0.5$ ) entre fenêtres adjacentes pour renforcer les interactions, en partitionnant les queries, keys et values différemment (non-chevauchant pour queries, chevauchant pour keys/values).

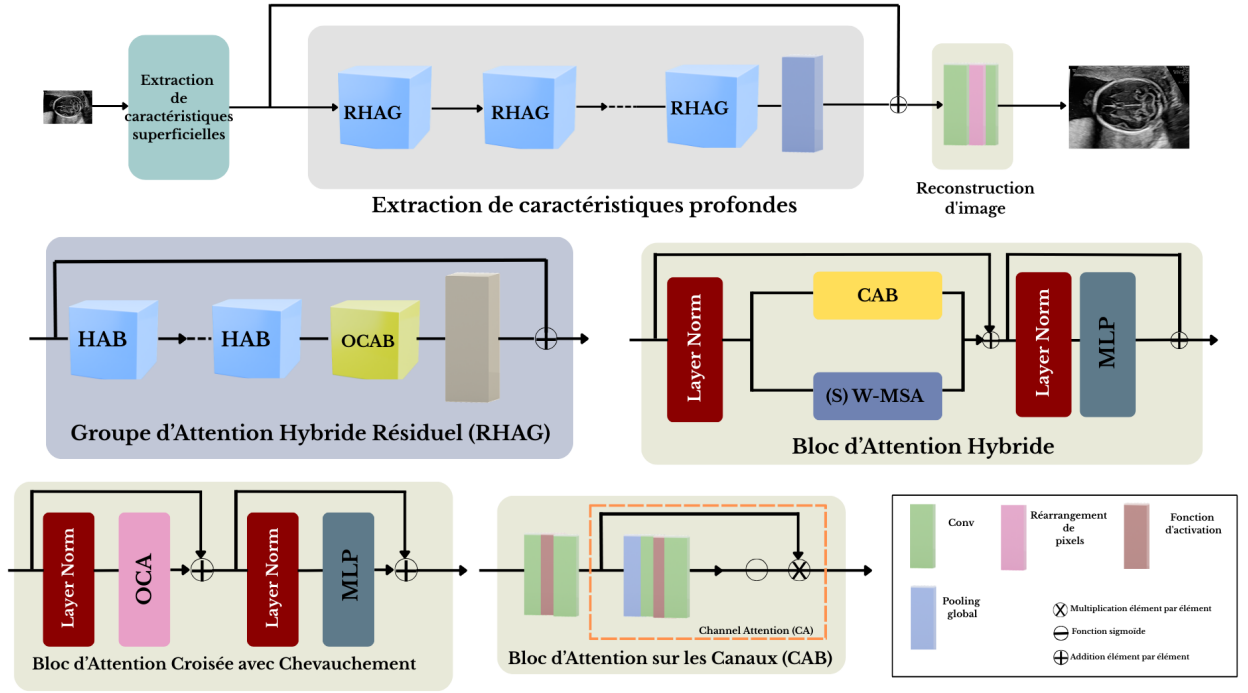


FIGURE 3.9 – Architecture globale du modèle HAT.[29] Le modèle restaure une image basse résolution (LR) en une image super-résolue (SR) haute qualité via trois étapes : 1. **Extraction de bas niveau** : une couche de convolution transforme les pixels d'entrée en une première représentation de features, capturant les patterns de base. 2. **Extraction profonde** : le cœur du modèle, constitué de groupes résiduels (RHAG), traite ces features. Chaque groupe utilise : - HAB : combinaisons d'auto-attention (relations spatiales entre pixels) et attention par canal (poids des canaux de features) pour modéliser finement les détails. - OCAB : échange d'informations entre les régions (fenêtres) pour assurer cohérence globale et éviter les artéfacts. 3. **Reconstruction** : le module final (ex : pixel-shuffle) agrège toutes les informations traitées pour synthétiser les détails fins et produire l'image SR finale.

**C. Reconstruction** Le module de reconstruction intègre **deux sources d'information** distinctes :

- Les **caractéristiques de bas niveau** issues du *module d'extraction superficielle* (première convolution), préservant les détails locaux et la structure géométrique de l'image d'entrée
- Les **caractéristiques de haut niveau** issues de la séquence de *Groupes d'Attention Hybrides Résiduels (RHAG)*, encapsulant une représentation contextuelle et abstraite de l'image

La fusion de ces deux flux s'effectue via une **connexion résiduelle globale** ( $F_0 + F_{DF}$ ), assurant :

- La cohérence des détails fins
- La stabilité de l'apprentissage

Le module applique ensuite une **transformation spécifique** à la tâche :

- *Pixel-shuffle* pour la super-résolution
- *Convolution* pour les tâches à résolution constante

pour produire l'image de sortie haute qualité.

TABLEAU 3.4 – Forces, limitations et impact médical des modèles Transformers en imagerie médicale.

| Modèle         | Forces                                                                                                                                               | Limitations                                                                                                          | Impact médical                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>SwinIR</b>  | Architecture hiérarchique à fenêtres décalées; modélisation des dépendances longue portée; bonne efficacité paramétrique; haute fidélité structurale | Conçu pour images naturelles; performances médicales dépendantes du fine-tuning; inférieur à HAT en PSNR/SSIM absolu | IRM mammaire (cross-cohort) SR $\times 4$ : PSNR 41.36, SSIM 0.962, supérieur aux méthodes traditionnelles et CNN[170]; Fine-tuning CT (Digital Korean dataset) : PSNR 37.42, SSIM 0.9327 (+3.51% / +0.63%)[100] avec : <i>Classement quantitatif</i> : 3e; <i>Classement MOS</i> : 1er (128→512 et 512→2048); Évalué sur chest X-ray, lésions cutanées et fonds d'œil : meilleures performances PSNR/SSIM vs SRGAN, BSRGAN et RealESRGAN[150]                                                                                                |
| <b>Swin2SR</b> | Transformer optimisé SR $\times 4$ ; amélioration fine des textures; meilleure stabilité; +33% rapide; performant en compression JPEG                | Coût computationnel élevé; validation médicale encore limitée; spécialisé compression                                | CT thoracique : reconstruction 2048 $\times$ 2048 améliorant la visualisation des nodules (accuracy 0.83 vs 0.79, p=0.037)[208]                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>HAT</b>     | Hybrid Attention (self-attention + convolution + channel attention); fusion globale/locale; forte capacité de reconstruction; SOTA en PSNR/SSIM      | Complexité computationnelle élevée; pré-entraînement massif; coût mémoire important                                  | Fine-tuning CT (Digital Korean dataset) : PSNR 39.95, SSIM 0.9461 (+2.91% / +0.54%)[100]; <i>Classement quantitatif Lee2025</i> : 1er; <i>Classement MOS Lee2025</i> : 2e; Amélioration NIQE vs Bicubic, RCAN, RS-ESRGAN en CT thoracique[104]; Variante <b>HAT-UNet</b> pour segmentation multi-datasets : <ul style="list-style-type: none"> <li>– PH2 (lésions cutanées dermoscopiques) : Dice 97.50%</li> <li>– CVC-ClinicDB (polypes coliques) : Dice 95.11%</li> <li>– Kvasir-SEG (polypes endoscopiques) : Dice 95.37% [14]</li> </ul> |

L'examen du Tableau 3.4 met en évidence l'avantage des architectures Transformers par rapport aux approches purement convolutives, tant sur les métriques quantitatives que sur l'évaluation clinique. Trois observations principales se dégagent.

**(1) Dichotomie entre mesures et perception.** L'analyse comparative révèle que si HAT domine les classements quantitatifs (PSNR de 39,95 dB sur CT), SwinIR obtient le meilleur score en qualité perçue (MOS)[100]. Cette divergence indique qu'une fidélité statistique élevée ne se traduit pas systématiquement par une supériorité visuelle clinique. SwinIR semble ainsi privilégier une restitution texturale et morphologique jugée plus cohérente par les experts[100].

**(2) Précision structurelle et segmentation.** L'architecture hybride de HAT-UNet[14] favorise une préservation anatomique fine, ce qui se reflète dans ses performances en segmentation multi-datasets, avec des scores de Dice supérieurs à 95 %. Ces résultats confirment son intérêt pour des tâches diagnostiques nécessitant une intégrité structurelle élevée.

**(3) Efficience et généralisation.** Swin2SR apporte un gain de rapidité de 33 % et une meilleure robustesse face à la compression[208], tandis que SwinIR démontre une grande polyvalence sur diverses modalités (IRM, radiographie, fond d'œil), surpassant les modèles GAN en stabilité et cohérence des reconstructions [170, 150].

En somme, malgré un coût computationnel important, ces modèles constituent une avancée significative pour la super-résolution médicale, en conciliant richesse des détails, cohérence morphologique et validation clinique. Au regard de ces performances, le choix d'une architecture capable de préserver ces dépendances globales s'avère primordial pour le traitement des images échographiques, dont la nature bruitée et texturée constitue un défi supplémentaire.

### 3.8 Synthèse comparative des modèles SR

Pour conclure cette étude comparative théorique et bibliographique des modèles de super-résolution étudiés, il est essentiel de synthétiser les points forts et les limites de chaque famille architecturale étudiée. Cette analyse prospective, fondée sur la littérature existante, nous permet d'établir un cadre théorique solide et de justifier les choix méthodologiques qui seront ultérieurement testés dans notre protocole expérimental (cf. Chapitre 4). Une synthèse de ces éléments comparatifs est présentée dans le Tableau 3.5.

Dans cette section, une revue systématique a été établie en s'appuyant sur des données empiriques issues d'études récentes qui a pour but d'évaluer les performances de modèles de super-résolution sur divers bases de données biomédicales.

Expérimentalement, Lee et al. (2025)[100] et Puttagunta et al. (2022) [150] démontrent l'efficacité des Transformers en reconstruction d'images haute résolution. Leurs protocoles valident des gains significatifs en PSNR/SSIM, particulièrement avec SwinIR qui préserve les structures anatomiques subtiles tout en optimisant ces métriques sur des données médicales hétérogènes. Parallèlement, les travaux de Nandal et al. (2024) [133] et Boumeridja et al. (2025)[19] documentent l'applicabilité clinique des réseaux antagonistes génératifs (GANs). Le modèle Real-ESRGAN (Enhanced Super-Resolution Generative Adversarial Network) améliore significativement la qualité perceptuelle d'images issues de différentes modalités (radiographie, échographie), facilitant ainsi l'interprétation diagnostique.

### 3.9 Conclusion

Une investigation a établi, dans ce chapitre, un cadre d'analyse comparative des paradigmes contemporains, évaluant leurs compromis fondamentaux entre fidélité quantitative, préservation structurale et complexité computationnelle.

Nous avons systématiquement défini les fondements théoriques des modèles basés sur les réseaux antagonistes génératifs (GANs) et les Transformers, en détaillant leurs principes architecturaux intrinsèques, leurs avantages opérationnels et leurs limitations conceptuelles. Une analyse comparative approfondie, synthétisée dans le Tableau 3.5, a mis en exergue les performances distinctes de modèles spécifiques tels que BSRGAN, Real-ESRGAN, A-ESRGAN, SwinIR, Swin2SR et HAT.

Il a été démontré que les architectures Transformer, à l'instar de SwinIR et HAT, excellent dans la préservation des détails anatomiques subtils et l'atteinte de métriques de fidélité élevées (PSNR, SSIM), tandis que les GANs, notamment Real-ESRGAN, se distinguent par leur capacité à générer des textures perceptuellement améliorées, essentielles pour l'interprétation diagnostique, bien qu'avec un risque accru d'artefacts.

En conclusion, les modèles étudiés peuvent être distingués selon leur stratégie d'apprentissage et leur architecture.

D'une part, certaines architectures à apprentissage interne, comme DBPISR, illustrent le paradigme des CNN spécifiques à chaque image, tandis que des modèles à apprentissage externe, tels que BSRGAN, Real-ESRGAN et SwinIR, exploitent des connaissances préalablement acquises sur des jeux de données étendus. D'autre part, l'analyse s'étend également à des modèles plus récents, incluant A-ESRGAN et les Transformers Swin2SR et HAT, offrant une perspective complète sur les approches contemporaines et les compromis entre paradigmes adversarial et attentionnel. La sélection des modèles garantit ainsi une couverture complète des approches modernes de super-résolution pertinentes pour l'imagerie médicale.

Ces bases théoriques étant posées, il est désormais possible de formaliser la méthodologie expérimentale retenue.

TABLEAU 3.5 – Comparaison bibliographique des principaux modèles de super-résolution évoqués en imagerie médicale.

| Famille     | Modèle            | Forces                                                                                                                  | Faiblesses                                                                | Applications / validations rapportées                                                     |
|-------------|-------------------|-------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------|
| CNN         | DBPISR [90]       | Apprentissage conjoint du noyau et de la reconstruction ; robuste aux noyaux inconnus ; plus rapide que KernelGAN+ZSSR. | Complexité accrue ; ré-entraînement requis pour chaque image.             | Super-résolution aveugle générique ; peu d'applications biomédicales documentées.         |
|             | BSRGAN [206]      | Dégradations réalistes et inconnues ; robuste au bruit, flou et artefacts.                                              | Non spécialisé médical ; artefacts possibles ; faible contraste mal géré. | Majoritairement images naturelles ; adaptations explorées en imagerie médicale.           |
| GAN         | Real-ESRGAN [192] | RRDB + discriminateur U-Net ; bon compromis fidélité / stabilité.                                                       | Coût computationnel élevé ; lissage de structures fines.                  | Applications rapportées en RX, IRM et échographie.                                        |
|             | A-ESRGAN [195]    | Attention multi-échelle ; excellente restitution des détails fins.                                                      | Modèle lourd ; instabilités à l'entraînement.                             | Pas encore validé médicalement ; prometteur pour structures fines.                        |
| TRANSFORMER | SwinIR [105]      | Dépendances longue portée ; haute fidélité PSNR/SSIM.                                                                   | Sensible aux dégradations non vues à l'entraînement.                      | Validations en CT, RX, fonds d'œil et dermatologie.                                       |
|             | Swin2SR [35]      | Stable et rapide (~33% plus rapide) ; performant sur JPEG.                                                              | Spécialisé compression ; moins polyvalent.                                | Images compressées ; potentiel pour PACS et télémedecine.                                 |
|             | HAT [29]          | Attention hiérarchique ; très haute fidélité anatomique.                                                                | Très coûteux ; pré-entraînement massif requis.                            | Rapports récents de performances élevées sur CT et RX, avec MOS favorables selon experts. |

**NOTE :** CT = Tomodensitométrie (scanner), RX = Radiographie, MOS = Mean Opinion Score, PACS = Picture Archiving and Communication System.

La Figure 3.10 en présente une synthèse, retraçant le pipeline complet — de la préparation des données jusqu'au fine-tuning — et servant de fil conducteur aux chapitres suivants.

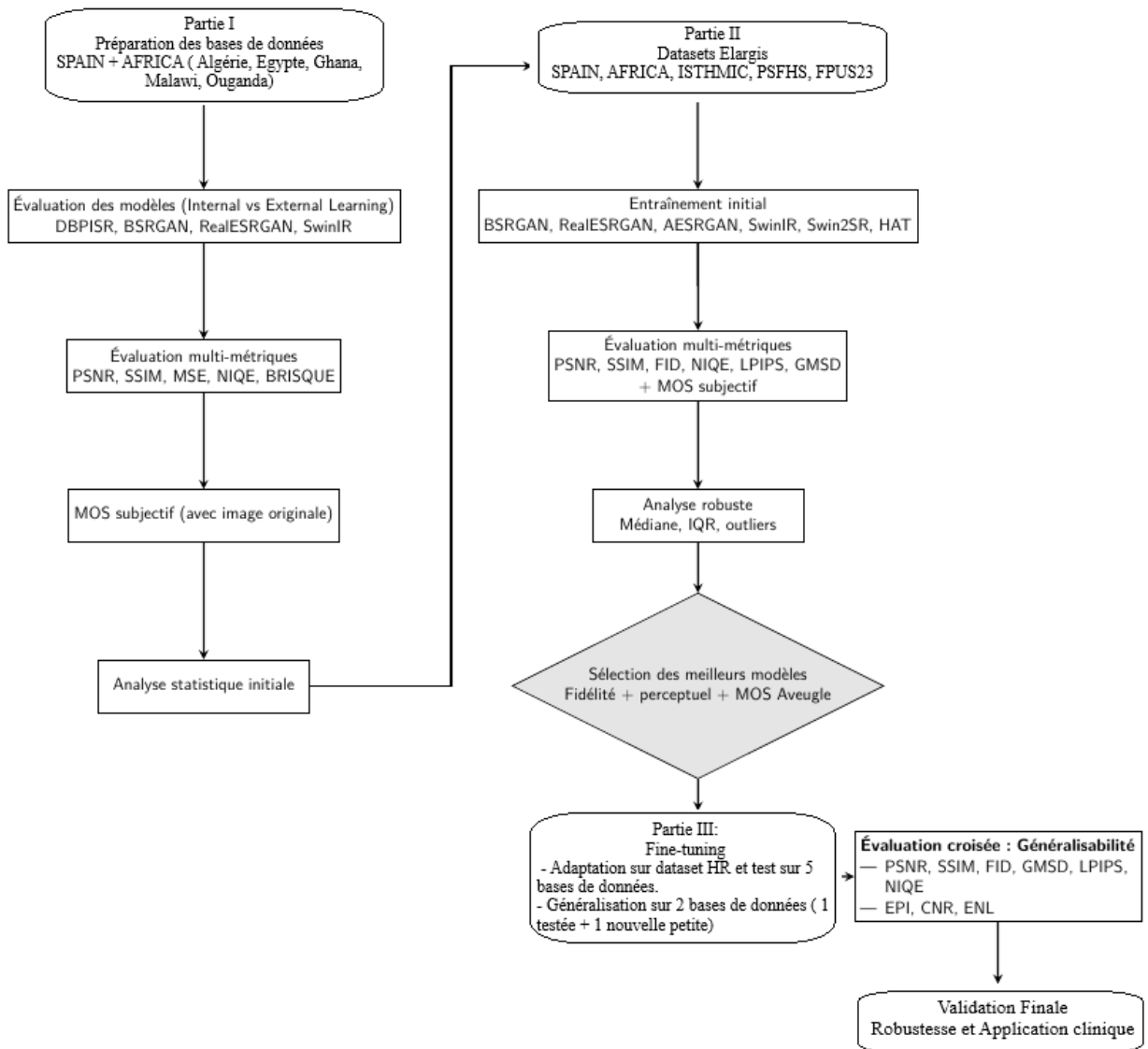


FIGURE 3.10 – Schéma global de la méthodologie expérimentale. Elle est structurée en trois parties successives (préparation et évaluation initiales, élargissement et sélection, puis fine-tuning), suivies d’une étape de validation finale.



# CHAPITRE 4

---

## Chapitre 4

# Évaluation quantitative, fonctionnelle et clinique des modèles de super-résolution appliqués à l'imagerie échographique

### 4.1 Introduction

L'amélioration de la résolution spatiale des données échographiques constitue un pivot stratégique pour l'imagerie diagnostique, spécifiquement dans les environnements cliniques disposant de ressources techniques restreintes. L'avènement de la super-résolution fondée sur le deep learning offre une alternative prometteuse pour pallier les contraintes d'échantillonnage des systèmes actuels. Néanmoins, la nature stochastique du *speckle* et la variabilité des signatures acoustiques selon les constructeurs imposent de dépasser les cadres d'évaluation conventionnels. La garantie d'une utilité médicale effective repose désormais sur une approche intégrée, où la précision technique est confrontée à une validation rigoureuse par l'expertise métier.

L'objectif de ce chapitre est d'évaluer la capacité des modèles de super-résolution à restaurer les détails anatomiques fœtaux, à travers une analyse rigoureuse de leurs performances techniques et cliniques. Notre démarche est articulée en deux phases complémentaires :

- **Partie I : Étude comparative des paradigmes interne et externe.** Cette phase initiale analyse les performances relatives des approches par apprentissage spécifique à l'image cible (internal learning) face aux modèles pré-entraînés sur des bases de données massives (external learning). L'évaluation s'appuie sur un socle de métriques quantitatives, de robustesse sémantique par classification et d'une analyse perceptuelle humaine avec référence.
- **Partie II : Analyse multidimensionnelle et modélisation statistique.** Ce second volet explore le potentiel des architectures de pointe (GANs et Transformers) via un cadre d'évaluation multidimensionnel. L'originalité de cette phase réside dans la conduite d'une évaluation MOS (Mean Opinion Score) en aveugle, visant à modéliser la fonction de correspondance mathématique entre les indices de performance automatiques et le jugement diagnostique de l'expert.



### 4.1.1 Objectifs et Contributions de la Partie I

Cette première phase de l'étude établit les fondements techniques de la super-résolution appliquée au domaine fœtal. Nos contributions incluent :

- **Évaluation de la généralisabilité sur données multicentriques** : Exploitation de deux bases de données internationales afin d'éprouver la robustesse des modèles SR face à la diversité des données et des spécifications techniques des équipements.
- **Exploitation d'architectures de pointe (SOTA)** : Adaptation et déploiement de modèles précurseurs (*DBPISR*, *Real-ESRGAN*, *BSRGAN* et *SwinIR*) pour l'amélioration du contraste et de la résolution.
- **Analyse comparative des paradigmes d'apprentissage** : Caractérisation des compromis entre l'apprentissage interne (*zero-shot*) et l'apprentissage externe, en identifiant leurs capacités respectives de généralisation clinique.
- **Validation de la préservation sémantique** : Utilisation du classificateur *ConvNeXt* pour attester mathématiquement de la conservation des descripteurs anatomiques fondamentaux après reconstruction.
- **Corrélation clinique préliminaire** : Consolidation des scores quantitatifs par une analyse du score d'opinion moyen (MOS) supervisée par un panel de quatre gynécologues experts.

### 4.1.2 Objectifs et Contributions de la Partie II

Fort des enseignements de la première phase, cette seconde étape se concentre sur la rupture technologique des Transformers et la fiabilité du diagnostic.

Les contributions majeures sont :

- **Validation de l'ascendant des architectures Transformers** : Mise en évidence de la supériorité du modèle *HAT* sur les réseaux antagonistes (GAN) en termes d'équilibre entre acuité visuelle et fidélité structurelle.
- **Formalisation d'un protocole d'évaluation multidimensionnel** : Introduction d'un cadre d'analyse à trois piliers pour garantir l'intégrité médicale :
  - (i) *Fidélité structurelle* (préservation de la géométrie anatomique) ;
  - (ii) *Qualité perceptuelle* (réalisme des textures et absence d'artéfacts de lissage) ;
  - (iii) *Plausibilité diagnostique* vérification de la conservation des signes cliniques (ex : contours crâniens) et de l'absence d'hallucinations numériques susceptibles d'induire un faux diagnostic.
- **Étude translationnelle en aveugle** : Évaluation rigoureuse menée par un expert senior sur un corpus élargi, exploitant un questionnaire MOS standardisé pour ancrer les résultats dans la pratique radiologique réelle.
- **Modélisation prédictive du jugement expert** : Identification de la métrique **GMSD** comme l'indicateur le plus corrélé à l'expertise humaine, via une étude en aveugle conduite selon un protocole **MOS sans référence** ( $r = 0,876$ ,  $p = 0,004$ ). Cette métrique explique à elle seule 76,7% de la variance des scores cliniques ( $R^2 = 0,767$ ), objectivant ainsi le lien entre performance algorithmique et perception diagnostique.

## 4.2 Partie I – Étude quantitative et perceptuelle des approches d'apprentissage interne et externe

La méthodologie adoptée, résumée dans la Figure 4.1, repose sur un pipeline expérimental structuré, conçu pour garantir une comparaison équitable, reproductible et statistiquement pertinente entre les différents modèles évalués.

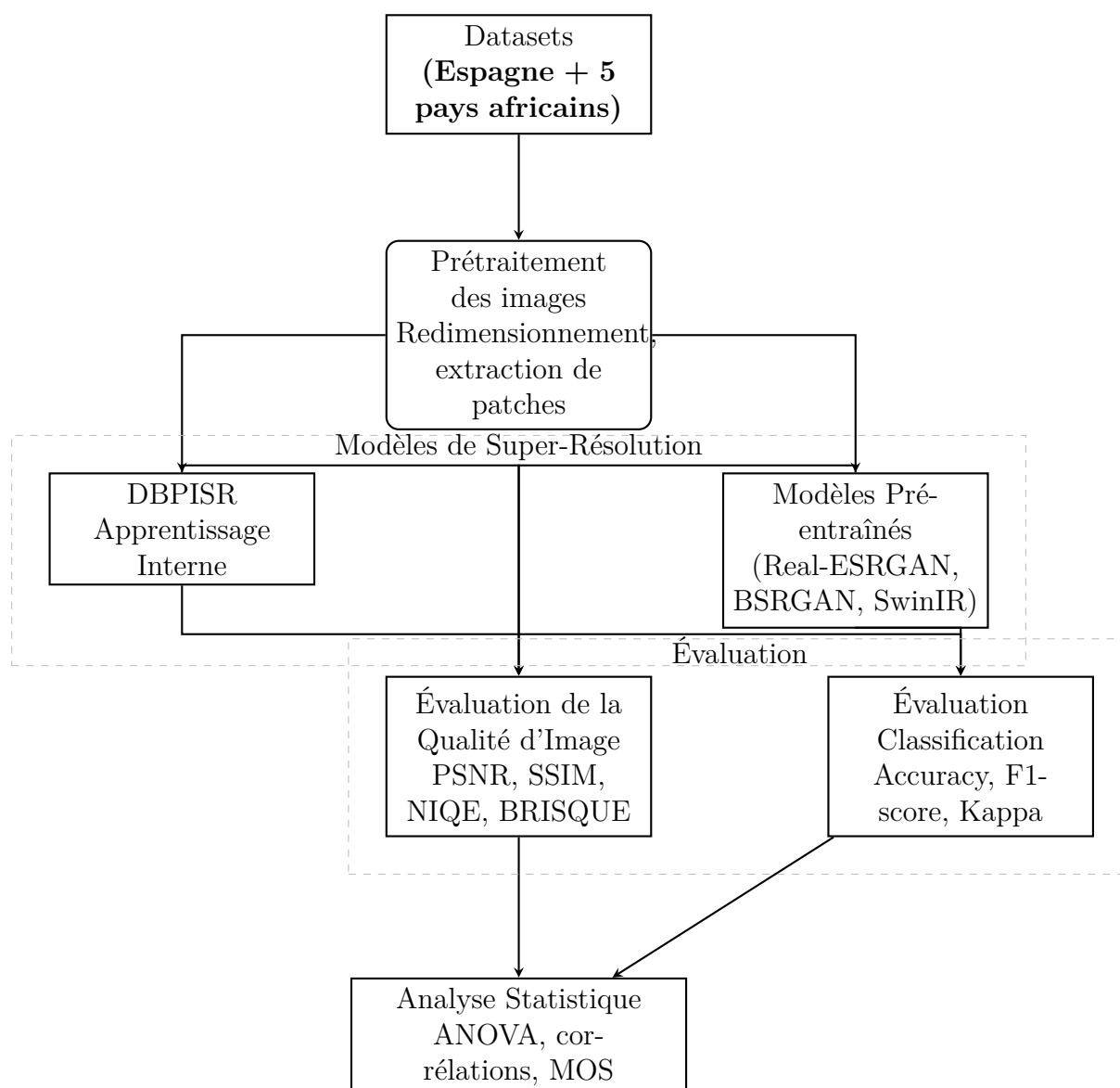


FIGURE 4.1 – Schéma méthodologique du pipeline expérimental pour l'évaluation des approches de super-résolution.

### 4.2.1 Typologie des modèles de super-résolution étudiés

Nous classons les modèles étudiés en deux catégories, selon leur mode d'apprentissage.

#### (a) Approche à apprentissage interne : DBPISR

DBPISR utilise l'apprentissage interne : il s'entraîne uniquement sur l'image à traiter, sans avoir besoin d'autres données. Pour reconstruire les détails, il exploite simplement la répétition des motifs présents à différentes échelles dans l'image. Cette méthode est très pratique car elle permet de travailler sans grande base d'apprentissage, ce qui est souvent difficile à obtenir en imagerie médicale.

Cependant, ce fonctionnement en vase clos présente des limites : faute d'exemples variés, le modèle peine parfois à restituer les structures anatomiques fines et risque d'interpréter le bruit ou les artéfacts comme des détails à rehausser.

## (b) Approches à apprentissage externe : BSRGAN, Real-ESRGAN et SwinIR

À l'inverse, BSRGAN, Real-ESRGAN, SwinIR reposent sur un apprentissage externe, réalisé à partir de larges bases de données d'images. Cette stratégie leur offre une grande capacité à reconstruire des textures complexes. Bien que ces modèles SR dominent l'état de l'art par leur finesse de restitution, leur efficacité demeure une vérité contingente, strictement subordonnée à la représentativité des données sources. Il est donc crucial de ne pas prendre ces performances pour acquises, mais de les évaluer rigoureusement face à la variabilité clinique.

### 4.2.2 Jeux de données et diversité des contextes d'acquisition

L'évaluation a été menée sur des images fœtales échographiques provenant de plusieurs pays africains, couvrant une large diversité d'équipements et de conditions cliniques.

Deux groupes distincts ont été considérés :

1. **Algérie-Égypte**, caractérisé par une cohérence instrumentale (GE Voluson S8 et P8), offrant ainsi un corpus de référence aux conditions d'acquisition standardisées ;
2. **Ghana, Malawi et Ouganda**, recourant à des échographes de marques différentes, introduisant une variabilité marquée en termes de bruit, de contraste et de résolution native.

### 4.2.3 Pipeline méthodologique et stratégie d'évaluation

Le pipeline expérimental adopté s'articule autour de quatre phases principales :

- a. La préparation et la caractérisation des jeux de données échographiques.
- b. L'application des modèles de super-résolution avec un facteur d'agrandissement de  $4\times$ .
- c. L'évaluation quantitative et perceptuelle des images reconstruites, ainsi que de l'impact sur la classification automatique.
- d. L'analyse statistique et l'interprétation des résultats en fonction des contextes d'acquisition.

Cette structuration méthodologique permet d'isoler l'impact propre des modèles SR, tout en tenant compte de la variabilité intrinsèque des images échographiques.

★ **Distinction des paradigmes** L'étude met en parallèle le modèle DBPISR (apprentissage interne) et 3 modèles externes (dont Real-ESRGAN, BSRGAN et les 2 variantes de SwinIR). Toutes les métriques de qualité et de classification ont été calculées séparément pour chaque groupe. Cette démarche permet d'isoler la performance de chaque approche et de tester sa stabilité face aux images ultrasonores.

### 4.2.4 Métriques d'évaluation quantitative et perceptuelle

Afin de caractériser de manière complète la qualité des images reconstruites, l'évaluation repose sur une combinaison de métriques référencées et sans référence.

**(A). Métriques référencées : PSNR et SSIM** Pour objectiver la qualité technique, nous croisons le PSNR (fidélité numérique) et le SSIM (préservation structurelle). Néanmoins, ces mesures objectives présentent une limite intrinsèque : elles ne traduisent qu'imparfaitement la qualité diagnostique réelle, une image pouvant être mathématiquement proche de la référence sans pour autant être visuellement satisfaisante pour le clinicien.

**(B). Métriques perceptuelles sans référence : BRISQUE et NIQE** Pour enrichir l’analyse, nous avons intégré des métriques sans référence, capables d’apprécier la qualité intrinsèque de l’image, indépendamment de toute ”vérité terrain” :

- BRISQUE, particulièrement sensible aux altérations statistiques et aux artefacts de reconstruction ;

- NIQE, qui évalue la ”naturalité” du signal en mesurant son écart face à des régularités statistiques idéales.

Dans cette étude, nous avons retenu le binôme PSNR/BRISQUE. Ce choix méthodologique vise à nuancer la rigueur arithmétique du PSNR par une estimation de la plausibilité visuelle, assurant ainsi une validation qui ne se limite pas à la justesse du signal, mais tend vers la réalité de la perception humaine.

**(C). Impact sur la classification automatique** L’évaluation de l’effet des modèles de super-résolution sur la classification des images échographiques a été réalisée en utilisant trois métriques complémentaires :

- **Accuracy** : proportion de prédictions correctes par rapport au nombre total de prédictions [72, 181] ;

- **F1-score** : moyenne harmonique de la précision et du rappel, adaptée aux jeux de données déséquilibrés [181] ;

- **Kappa de Cohen** : mesure de concordance entre les étiquettes prédites et les étiquettes réelles, corrigée pour le hasard [97]. Des valeurs élevées pour ces trois métriques indiquent de meilleures performances. Les modèles de super-résolution ont été évalués selon le protocole présenté dans le Tableau 4.3, appliqué aux jeux de données originaux et super-résolus.

Durant l’entraînement du classificateur, plusieurs techniques d’augmentation de données ont été appliquées afin de renforcer la robustesse et la capacité de généralisation des modèles :

- *RandomHorizontalFlip* : retournement horizontal aléatoire des images avec une probabilité de 50 % ;

- *RandomVerticalFlip* : retournement vertical aléatoire des images avec une probabilité de 50 % ;

- *RandomRotation* : rotation aléatoire dans un intervalle de  $\pm 20^\circ$  pour simuler des variations d’orientation réelles ;

- *ColorJitter* : modification aléatoire de la luminosité, du contraste et de la teinte jusqu’à  $\pm 50\%$ , simulant les variations d’éclairage et de conditions de capture entre dispositifs.

### (c.1) Évaluation fonctionnelle par classification automatique

Si les métriques traditionnelles (PSNR, SSIM) offrent une quantification utile, leur corrélation avec la perception humaine et la valeur diagnostique demeure limitée. Pour objectiver l’impact réel de la super-résolution, nous avons déployé une analyse fonctionnelle basée sur l’architecture de pointe **ConvNeXt-Base** [115]. Dans ce dispositif, le classificateur agit comme un système d’aide au diagnostic (CAD), servant de repère fonctionnel pour valider l’intérêt clinique des images reconstruites.

### Architecture et constituants du classificateur

Basé sur l’architecture ConvNeXt-Base [115], ce classificateur exploite une structure hiérarchique pour l’abstraction progressive des descripteurs visuels. Comme détaillé en Figure 4.2, chaque bloc élémentaire repose sur une convolution depthwise ( $7 \times 7$ ). Ce large noyau permet de capturer des dépendances spatiales étendues, cruciales pour identifier des structures anatomiques complexes.

Cette architecture intègre des avancées clés issues des *Vision Transformers*, notamment une structure en goulot d'étranglement inversé (*Inverted Bottleneck*) qui dilate la dimension des couches intermédiaires pour enrichir la représentation des caractéristiques. Le pipeline est stabilisé par une normalisation LayerNorm et une fonction d'activation GELU, avant qu'une agrégation Global Average Pooling (GAP) ne compresse les cartes 2D en un vecteur sémantique pour la couche finale de classification.

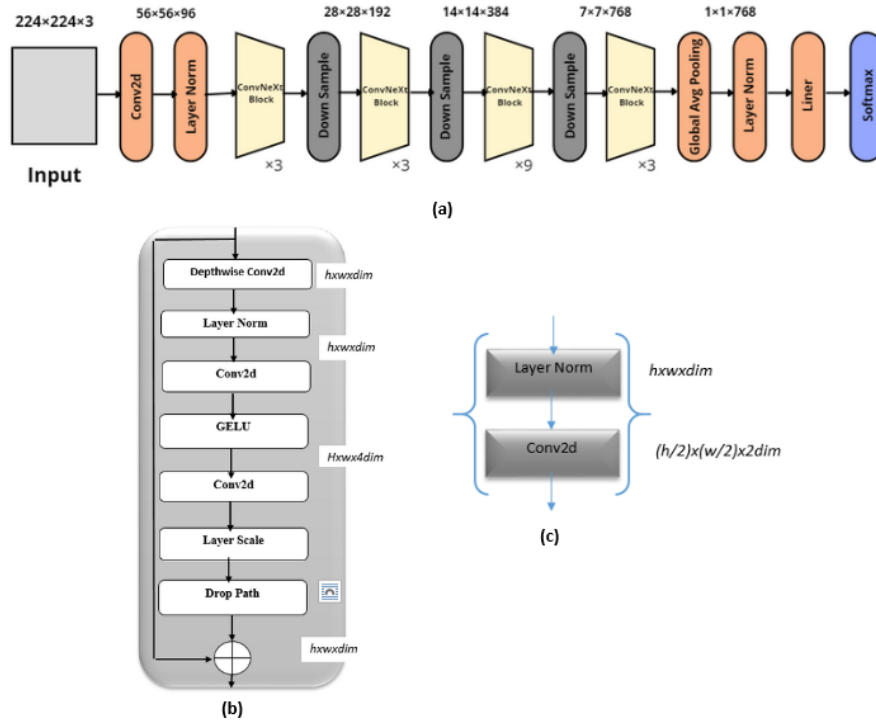


FIGURE 4.2 – ConvNeXt-Base [115] : (a) Architecture Globale ; (b) Détail d'un bloc ConvNeXt ; (c) Mécanisme d'échantillonnage descendant (*down-sampling*).

La validation fonctionnelle repose sur un apprentissage globalisé incluant 12 850 images provenant de six pays (Espagne, Algérie, Égypte, Ghana, Malawi et Ouganda). Ce volume de données hétérogènes permet au modèle de s'abstraire des particularités techniques des différents constructeurs d'échographes.

La structure du jeu de données respecte les spécificités de chaque site :

- **La cohorte espagnole** se distingue par une couverture en cinq catégories, incluant les quatre plans principaux (*abdomen*, *cerveau*, *fémur* et *thorax*) ainsi qu'une classe "Autres". Cette dernière est constituée d'un mélange hétérogène d'images fœtales issues de ces quatre mêmes catégories, servant de régulateur pour le classificateur.
- **Les cohortes d'Algérie, d'Égypte et du Malawi** (100 images chacune) présentent une structure quadri-modale intégrant l'abdomen, le cerveau, le fémur et le thorax.
- **Les cohortes du Ghana et de l'Ouganda** (75 images chacune) suivent une configuration tri-modale limitée à l'abdomen, au cerveau et au fémur.

En mobilisant ce classificateur comme outil d'inférence sur les images échographiques fœtales reconstruites, nous évaluons la capacité des modèles de super-résolution à préserver l'intégrité des plans anatomiques fœtaux au sein des images super-résolues, dont le nombre varie selon le dataset considéré — allant de trois plans pour les datasets du Ghana et de l'Ouganda, à quatre plans pour ceux d'Algérie, d'Égypte et du Malawi, jusqu'à cinq plans pour le dataset espagnol. Cette évaluation constitue ainsi une validation fonctionnelle complémentaire aux métriques PSNR et SSIM.

### (D). Analyse statistique multi-critères

Une analyse statistique globale clôture l'étude. Elle regroupe toutes les métriques collectées — qualité (avec PSNR, SSIM, NIQE, BRISQUE), classification et scores MOS — et s'appuie sur :

(i) des tests ANOVA pour valider la significativité des écarts ; (ii) une analyse multivariée pour lier les critères objectifs aux évaluations subjectives et fonctionnelles ; (iii) l'intégration des avis des quatre experts, indispensable pour la validation médicale.

Cette méthodologie d'évaluation multi-critères permet une caractérisation exhaustive des performances, facilitant l'identification du modèle SR optimal pour l'imagerie ultrasonore diagnostique.

## 4.2.5 Résultats expérimentaux – Partie I

Cette section présente un cadre d'évaluation multidimensionnel structuré pour valider l'apport des modèles SR selon trois échelles distinctes. En premier lieu, la qualité visuelle est mesurée par un ensemble d'indicateurs objectifs et de métriques "sans référence", dont la robustesse statistique garantit la fiabilité des observations face à l'hétérogénéité des données. Parallèlement, l'utilité opérationnelle du traitement est éprouvée sur un pipeline de classification automatique (ConvNeXt-Base), offrant ainsi une mesure concrète du gain d'information pour l'identification des plans. L'étude se conclut par une validation par les pairs (MOS), où l'expertise de praticiens vient valider la cohérence entre les performances numériques et les exigences du diagnostic médical.

## A. Évaluation quantitative et perceptuelle des modèles de super-résolution

### □ Métriques référencées : PSNR et SSIM

Le Tableau 4.1 synthétise les différentes métriques de qualité d'image calculées pour l'ensemble des modèles étudiés.

TABLEAU 4.1 – Performances des modèles de super-résolution : évaluation de la qualité d'image dans les pays étudiés

|             | Algérie      | Égypte       | Ghana        | Malawi       | Espagne      | Ouganda      | Type d'apprentissage |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------------|
| Modèle      | PSNR/SSIM    | PSNR/SSIM    | PSNR/SSIM    | PSNR/SSIM    | PSNR/SSIM    | PSNR/SSIM    |                      |
| BSRGAN      | 33.77 / 0.82 | 34.12 / 0.84 | 37.95 / 0.91 | 34.57 / 0.84 | 34.36 / 0.87 | 35.94 / 0.92 | Externe              |
| Real-ESRGAN | 32.93 / 0.80 | 35.04 / 0.89 | 37.73 / 0.91 | 33.86 / 0.84 | 33.67 / 0.82 | 35.48 / 0.92 | Externe              |
| SwinIR-M    | 32.20 / 0.76 | 31.56 / 0.71 | 33.51 / 0.80 | 32.80 / 0.76 | 32.37 / 0.78 | 32.52 / 0.82 | Externe              |
| SwinIR-L    | 32.98 / 0.78 | 32.67 / 0.76 | 36.12 / 0.89 | 33.37 / 0.78 | 33.27 / 0.81 | 34.62 / 0.88 | Externe              |
| DBPISR      | 28.18 / 0.84 | 24.09 / 0.65 | 28.61 / 0.73 | 28.26 / 0.85 | 32.86 / 0.90 | 28.14 / 0.78 | Interne              |

### □ Métriques sans référence : NIQE et BRISQUE

Pour évaluer la lisibilité des images, nous avons appliqué deux indices objectifs sans référence : BRISQUE et NIQE[128]. Le BRISQUE s'appuie sur l'analyse des régularités statistiques des textures pour identifier les dégradations ; un score élevé correspond à une perte de qualité. Le NIQE mesure la divergence d'une image par rapport aux caractéristiques usuelles d'images non dégradées ; une valeur plus importante indique une détérioration plus marquée. Les résultats sont regroupés dans le Tableau 4.2.

TABLEAU 4.2 – Scores NIQE et BRISQUE par pays et par modèle

| Pays    | Modèle      | NIQE (Moy $\pm$ Écart-type) | BRISQUE (Moy $\pm$ Écart-type) |
|---------|-------------|-----------------------------|--------------------------------|
| Algérie | BSRGAN      | $8.98 \pm 6.06$             | $44.28 \pm 5.17$               |
|         | DBPISR      | $19.62 \pm 5.65$            | $84.74 \pm 17.95$              |
|         | Real-ESRGAN | $12.73 \pm 8.01$            | $63.69 \pm 6.68$               |
|         | SwinIR-M    | $29.81 \pm 18.10$           | $65.09 \pm 9.94$               |
|         | SwinIR-L    | $29.08 \pm 23.24$           | $59.91 \pm 7.93$               |
| Égypte  | BSRGAN      | $7.67 \pm 4.06$             | $68.69 \pm 6.74$               |
|         | DBPISR      | $19.50 \pm 4.58$            | $68.69 \pm 6.74$               |
|         | Real-ESRGAN | $3.78 \pm 2.20$             | $56.24 \pm 4.32$               |
|         | SwinIR-M    | $51.41 \pm 29.60$           | $49.82 \pm 8.98$               |
|         | SwinIR-L    | $38.59 \pm 28.62$           | $42.99 \pm 6.55$               |
| Ghana   | BSRGAN      | $2.18 \pm 1.10$             | $44.15 \pm 11.29$              |
|         | DBPISR      | $9.56 \pm 3.75$             | $89.84 \pm 10.94$              |
|         | Real-ESRGAN | $1.34 \pm 0.41$             | $59.74 \pm 8.43$               |
|         | SwinIR-M    | $11.56 \pm 8.26$            | $55.12 \pm 9.11$               |
|         | SwinIR-L    | $8.96 \pm 8.62$             | $51.13 \pm 8.21$               |
| Malawi  | BSRGAN      | $6.42 \pm 2.82$             | $41.63 \pm 5.09$               |
|         | DBPISR      | $14.53 \pm 3.52$            | $77.83 \pm 13.36$              |
|         | Real-ESRGAN | $6.73 \pm 2.22$             | $59.42 \pm 3.96$               |
|         | SwinIR-M    | $27.32 \pm 16.35$           | $58.42 \pm 9.09$               |
|         | SwinIR-L    | $25.27 \pm 16.12$           | $54.96 \pm 6.65$               |
| Ouganda | BSRGAN      | $3.95 \pm 1.21$             | $43.20 \pm 2.98$               |
|         | DBPISR      | $14.33 \pm 4.57$            | $83.41 \pm 11.20$              |
|         | Real-ESRGAN | $3.69 \pm 0.79$             | $63.10 \pm 4.81$               |
|         | SwinIR-M    | $11.45 \pm 6.19$            | $47.68 \pm 2.80$               |
|         | SwinIR-L    | $6.73 \pm 3.90$             | $53.97 \pm 3.23$               |

Écart-type : mesure de la dispersion autour de la moyenne (*Standard Deviation*).

## B. Résultats de la classification

Nous avons établi un protocole spécifique pour mesurer l'impact de la super-résolution sur la classification (Tableau 4.3).

L'étude confronte les performances des images originales face à celles reconstructions des cinq modèles, à travers six bases de données géographiques (Algérie, Égypte, Ghana, Malawi, Ouganda et Espagne). Les résultats pour le jeu de données espagnol, ainsi que ceux consolidés pour tous les pays, sont détaillés dans les Tableaux 4.4 et 4.5.

## C. Visualisation des résultats

Cette section présente une validation qualitative à deux niveaux. Elle confronte le résultat global à une inspection minutieuse de détails anatomiques clés (ex : le fémur). Cela permet de s'assurer visuellement que le rehaussement de la résolution respecte la texture originale tout en assurant un débruitage efficace.

■ **Comparaison globale** Pour chaque région anatomique du fœtus, nous avons généré un jeu de données composé d'images originales (première colonne à gauche) et d'images reconstruite à une résolution (4X) à l'aide de différents modèles (DBPISR, BSRGAN, Real-ESRGAN, SwinIR-M et SwinIR-L). La Figure 4.3 illustre ces résultats.

■ **Zoom sur les détails** De plus, afin d'évaluer la performance perceptuelle des modèles de super-résolution, la figure 4.4 propose des zooms d'images échographiques du fémur.

TABLEAU 4.3 – Protocole de classification d’images à partir des images originales et des images super-résolues

| Étape                                                                                                                                                       | Images originales                                                                                                                                                                                                                                                                                                    | Images super-résolues                                                                                                                                                                                                             |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Objectif</b> : Comparer les performances de classification (Accuracy, F1-score, Kappa) entre les images originales et super-résolues avec ConvNeXT_base. |                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                   |
| <b>Acquisition des données</b>                                                                                                                              | <ul style="list-style-type: none"> <li>— <b>Jeu de données espagnol</b> : zenodo.org/record/3904280</li> <li>— <b>Jeu de données africain*</b> : zenodo.org/records/7540448</li> <li>— <b>Nombre d’images</b> : 12 400 (Espagne), 100 (Algérie, Égypte, Malawi), 75 (Ghana, Ouganda)</li> </ul>                      | <ul style="list-style-type: none"> <li>— <b>À partir des modèles SR</b> : BSRGAN, DBPISR, Real-ESRGAN, SwinIR-M, SwinIR-L appliqués aux jeux espagnol et africain*</li> </ul>                                                     |
| <b>Préparation des données</b>                                                                                                                              | <ul style="list-style-type: none"> <li>— <b>Augmentation</b> : Oui</li> <li>— <b>Taille des images</b> : 448</li> </ul>                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>— <b>Transformation par super-résolution</b> : améliore la qualité des images avant classification.</li> <li>— <b>Augmentation et taille</b> : identiques aux images originales</li> </ul> |
| <b>Modèles</b>                                                                                                                                              | Classificateur : ConvNeXT_base                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                   |
| <b>Entraînement</b>                                                                                                                                         | <ul style="list-style-type: none"> <li>— <b>Optimiseur</b> : Adam</li> <li>— <b>Fonction de perte</b> : Cross_Entropy_Loss</li> <li>— <b>Taux d’apprentissage</b> : 0.0001</li> <li>— <b>Nombre d’époques</b> : 100</li> <li>— <b>Taille du lot</b> : 32</li> <li>— <b>GPU</b> : NVIDIA A6000 Ada (48 GB)</li> </ul> | <ul style="list-style-type: none"> <li>— Paramètres d’entraînement identiques aux images originales</li> </ul>                                                                                                                    |
| <b>Évaluation</b>                                                                                                                                           | <ul style="list-style-type: none"> <li>— <b>Métriques</b> : Accuracy, Kappa, F1-score</li> </ul>                                                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>— <b>Métriques</b> : identiques aux images originales</li> </ul>                                                                                                                           |

\*Le jeu de données africain inclut l’Algérie, l’Égypte, le Ghana, le Malawi et l’Ouganda.

TABLEAU 4.4 – Performance des modèles de super-résolution pour la classification d’images sur le jeu de données espagnol (images originales et super-résolues)

| SR Models          | Accuracy      | F1 Score      | Kappa         |
|--------------------|---------------|---------------|---------------|
| <b>DBPISR</b>      | 96.18%        | 95.24%        | 94.54%        |
| <b>BSRGAN</b>      | 95.90%        | 95.04%        | 94.22%        |
| <b>Real-ESRGAN</b> | <b>96.36%</b> | <b>95.62%</b> | <b>94.76%</b> |
| <b>SwinIR-M</b>    | 96.27%        | 95.23%        | 95.41%        |
| <b>SwinIR-L</b>    | 96.18%        | 95.27%        | 94.83%        |
| <b>Original</b>    | 95.18%        | 94.09%        | 93.16%        |

TABLEAU 4.5 – Performance des modèles de super-résolution pour la classification d’images selon les pays de l’étude (images originales et super-résolues)

| SR Models          | Algérie |        |        | Égypte |        |        | Ghana  |        |        | Malawi |        |        | Ouganda |        |        |
|--------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|--------|--------|
|                    | Acc.    | F1     | Kappa  | Acc.   | F1     | Kappa  | Acc.   | F1     | Kappa  | Acc.   | F1     | Kappa  | Acc.    | F1     | Kappa  |
| <b>DBPISR</b>      | 69%     | 51.4%  | 44.45% | 71%    | 64.56% | 56.28% | 48%    | 38.25% | 34.10% | 57%    | 42.64% | 32.20% | 53.34%  | 34.50% | 31.16% |
| <b>BSRGAN</b>      | 85%     | 72.64% | 79.87% | 51%    | 50.22% | 58.40% | 4%     | 5.77%  | 11.76% | 60%    | 58.17% | 58.84% | 5.33%   | 6.89%  | 0%     |
| <b>Real-ESRGAN</b> | 96%     | 78.31% | 92.33% | 68%    | 62.94% | 64.50% | 18.67% | 22.05% | 17.10% | 79%    | 68.91% | 71.00% | 18.67%  | 21.80% | 15.90% |
| <b>SwinIR-M</b>    | 94%     | 77.09% | 88.04% | 64%    | 60.42% | 64.09% | 12%    | 15.61% | 7.45%  | 84%    | 72.52% | 75.00% | 10.67%  | 13.69% | 5.60%  |
| <b>SwinIR-L</b>    | 84%     | 72.39% | 80.00% | 63%    | 59.75% | 66.60% | 5.33%  | 7.28%  | 4.73%  | 59%    | 56.80% | 64.13% | 17.33%  | 20.45% | 15.68% |
| <b>Original</b>    | 85%     | 72.74% | 80.67% | 62%    | 59.50% | 62.33% | 5.33%  | 7.54%  | 2.96%  | 66%    | 62.37% | 62.80% | 18.67%  | 21.87% | 13.95% |

La comparaison souligne la capacité variable des méthodes à concilier débruitage, préservation des structures anatomiques fines et fidélité au contenu échographique original.



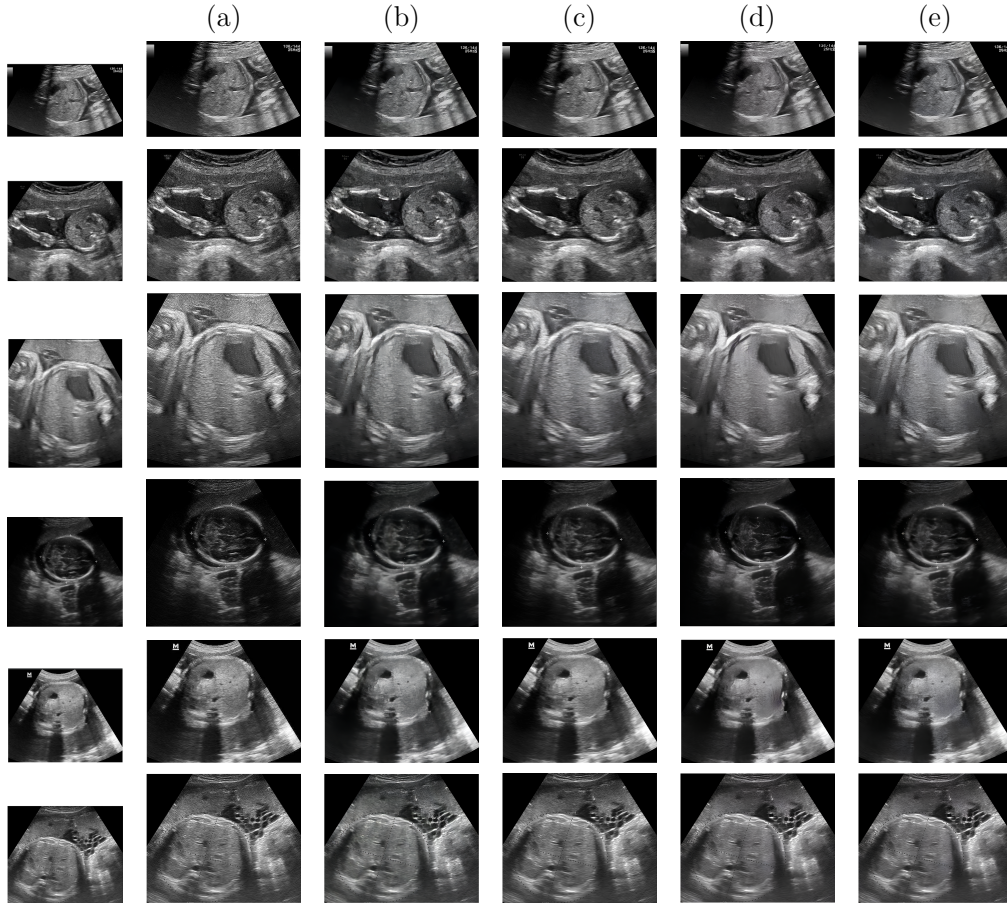


FIGURE 4.3 – Comparaison entre les images échographiques originales et reconstruites : (a) DBPISR, (b) BSRGAN, (c) Real-ESRGAN, (d) SwinIR-M et (e) SwinIR-L.

## D. Analyse des résultats

Nous présentons d’abord les résultats obtenus avec les modèles de super-résolution appliqués à l’analyse des images échographiques dans les différents pays, puis nous examinons les performances de la classification automatique utilisée comme évaluation fonctionnelle.

### (1)- Reconstruction par Super-Résolution : Interprétation et Évaluation analytique

Cette analyse évalue l’efficacité des modèles à restaurer les détails anatomiques malgré la diversité des équipements utilisés. L’objectif est de vérifier si le gain de résolution reste stable d’un environnement d’acquisition à l’autre.

#### • Technologie de l’imagerie échographique : Aperçu comparatif

La colonne « *Consistance* », présentée dans le Tableau 4.6, illustre la régularité des performances des modèles. Elle repose sur l’analyse statistique des résultats, notamment l’écart-type, les percentiles et la différence de moyenne. Une consistance *Élevée* traduit une grande stabilité des sorties, alors qu’une consistance *Faible* correspond à des résultats très dispersés. Les catégories intermédiaires (*Bonne*, *Modérée*) reflètent des niveaux de variabilité proportionnels.

Cette grille de lecture permet non seulement de comparer les modèles sur la base de leurs performances absolues (PSNR, SSIM, NIQE, BRISQUE), mais aussi d’évaluer leur capacité à fournir des résultats homogènes.

Dans ce cadre, nous avons structuré notre analyse en deux groupes distincts, reflétant les différences d’équipement échographique : — Le premier groupe comprend l’Algérie et l’Égypte, qui utilisent des échographes de la même marque (GE Voluson S8 et P8) ; — Le second groupe regroupe le Ghana, le Malawi et l’Ouganda, dont les images proviennent d’échographes de

TABLEAU 4.6 – Performances des modèles de super-résolution : Algérie, Égypte, Ghana, Malawi et Ouganda, classés du meilleur au pire

| Modèle SR          | PSNR médian (dB) | SSIM médian  | Consistance | Avantages                                          | Inconvénients                                         |
|--------------------|------------------|--------------|-------------|----------------------------------------------------|-------------------------------------------------------|
| <b>Algérie</b>     |                  |              |             |                                                    |                                                       |
| <b>BSRGAN</b>      | <b>35,21</b>     | <b>0,882</b> | Élevée      | NIQE faible, images nettes                         | Variabilité, traitement plus lent                     |
| SwinIR-L           | 34,27            | 0,852        | Modérée     | SSIM correct, NIQUE raisonnable                    | Variabilité de la qualité                             |
| Real-ESRGAN        | 33,63            | 0,847        | Bonne       | Détails fins, NIQUE faible                         | Légère accentuation du grain                          |
| SwinIR-M           | 33,48            | 0,839        | Moyenne     | Préservation de la structure                       | NIQE élevé, fidélité moindre                          |
| DBPISR             | 28,18            | 0,836        | Faible      | Intégrité structurelle partielle                   | PSNR faible, BRISQUE élevé                            |
| <b>Égypte</b>      |                  |              |             |                                                    |                                                       |
| <b>Real-ESRGAN</b> | <b>35,40</b>     | <b>0,909</b> | Élevée      | Détails fins, NIQUE faible                         | Légère accentuation du grain                          |
| BSRGAN             | 33,98            | 0,856        | Bonne       | NIQE faible, images nettes                         | Variabilité, traitement plus lent                     |
| SwinIR-L           | 32,73            | 0,773        | Modérée     | SSIM correct, NIQUE raisonnable                    | Variabilité de la qualité                             |
| SwinIR-M           | 31,51            | 0,717        | Moyenne     | Préservation de la structure                       | NIQE élevé, fidélité moindre                          |
| DBPISR             | 24,10            | 0,647        | Faible      | Intégrité structurelle partielle                   | PSNR faible, BRISQUE élevé                            |
| <b>Ghana</b>       |                  |              |             |                                                    |                                                       |
| <b>BSRGAN</b>      | <b>37,95</b>     | <b>0,91</b>  | Élevée      | BRISQUE faible (44,15), PSNR élevé                 | NIQE (2,18) légèrement supérieur à Real-ESRGAN        |
| Real-ESRGAN        | 37,73            | <b>0,91</b>  | Élevée      | NIQE faible (1,34), aspect plus naturel            | BRISQUE (59,74) plus élevé, PSNR légèrement inférieur |
| SwinIR-L           | 36,12            | 0,89         | Modérée     | SSIM amélioré                                      | NIQE (8,96) et BRISQUE (51,13) plus élevés            |
| SwinIR-M           | 33,51            | 0,80         | Moyenne     | Préservation de la structure                       | NIQE (11,56) et BRISQUE (55,12) élevés                |
| DBPISR             | 28,61            | 0,73         | Faible      | Intégrité structurelle partielle                   | NIQE (9,56) et BRISQUE (89,84) très élevés            |
| <b>Malawi</b>      |                  |              |             |                                                    |                                                       |
| <b>BSRGAN</b>      | <b>34,57</b>     | 0,84         | Élevée      | NIQE (6,42), BRISQUE (41,63) faibles, PSNR correct | Traitement légèrement plus lent                       |
| Real-ESRGAN        | 33,86            | 0,84         | Élevée      | NIQE faible (6,73), qualité correcte               | BRISQUE (59,42) plus élevé, PSNR légèrement inférieur |
| SwinIR-L           | 33,37            | 0,78         | Modérée     | SSIM amélioré                                      | NIQE (25,27), BRISQUE (54,96) élevés                  |
| SwinIR-M           | 32,80            | 0,76         | Moyenne     | Préservation de la structure                       | NIQE (27,32), BRISQUE (58,42) élevés                  |
| DBPISR             | 28,26            | <b>0,85</b>  | Faible      | Intégrité structurelle partielle                   | NIQE (14,53), BRISQUE (77,83) très élevés             |
| <b>Ouganda</b>     |                  |              |             |                                                    |                                                       |
| <b>BSRGAN</b>      | <b>35,94</b>     | 0,92         | Élevée      | BRISQUE faible (43,20), PSNR correct               | NIQE (3,95) légèrement supérieur                      |
| Real-ESRGAN        | 35,48            | <b>0,92</b>  | Élevée      | NIQE faible (3,69), aspect naturel                 | BRISQUE (63,10) plus élevé, PSNR légèrement inférieur |
| SwinIR-L           | 34,62            | 0,88         | Modérée     | SSIM amélioré                                      | NIQE (6,73), BRISQUE (53,97) élevés                   |
| SwinIR-M           | 32,52            | 0,82         | Moyenne     | Préservation de la structure                       | NIQE (11,45), BRISQUE (47,68) élevés                  |
| DBPISR             | 28,14            | 0,78         | Faible      | Intégrité structurelle partielle                   | NIQE (14,33), BRISQUE (83,41) très élevés             |

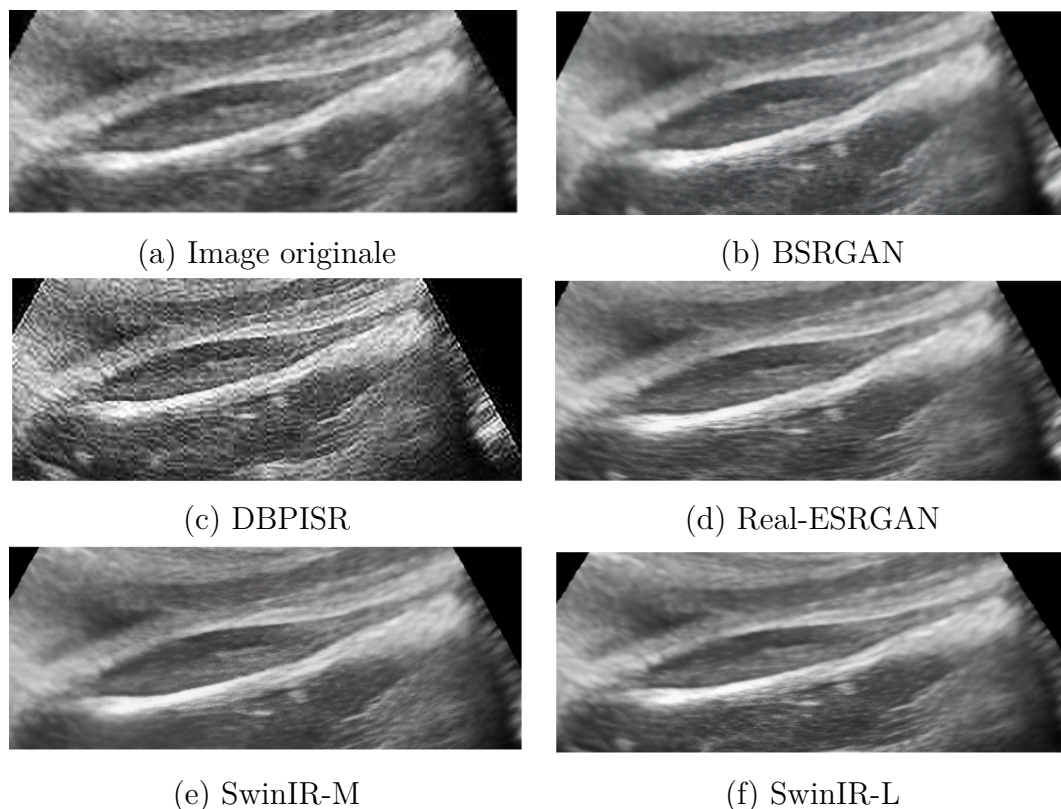


FIGURE 4.4 – Zooms d’images échographiques du fémur obtenus via cinq modèles de super-résolution (jeu de données Malawi) : (a) Image originale, (b–f) Résultats des modèles SR.

marques diverses (EDAN DUS 60, Mindray DC-N2 et Siemens).

### Groupe 1 : Algérie vs Égypte

Pour ce premier groupe, nous comparons les performances des modèles de super-résolution sur les images échographiques issues des échographes GE utilisés en Algérie et en Égypte. L’objectif est de mettre en évidence les différences de performance liées aux caractéristiques propres de chaque échographe et de déterminer quel modèle est le mieux adapté à chaque contexte.

#### • Algérie.

En Algérie, le modèle BSRGAN se démarque nettement. Il atteint un PSNR supérieur à 35 dB et un SSIM avoisinant 0,88, tout en conservant des scores perceptuels très favorables (NIQE faible et BRISQUE réduit). Ces résultats traduisent une grande stabilité (faible variabilité inter-échantillons) et confirment la supériorité de BSRGAN par rapport au modèle interne DBPISR, ainsi qu’aux autres modèles externes. Il apparaît ainsi comme l’approche la plus adaptée aux images échographiques acquises avec le GE Voluson S8.

#### • Égypte.

En Égypte, la hiérarchie des performances diffère : Real-ESRGAN obtient les résultats les plus convaincants. Avec un PSNR dépassant 35 dB et un SSIM proche de 0,91, il surpasse nettement DBPISR et les autres modèles, tout en maintenant une bonne stabilité. Bien que performant en Algérie, BSRGAN présente ici une variabilité plus marquée. Quant aux modèles SwinIR-M et SwinIR-L, ils se situent à un niveau moyen, avec des scores de fidélité et de perception moins favorables. Ces observations suggèrent que Real-ESRGAN est mieux adapté aux caractéristiques des images issues de l’échographe GE Voluson P8.

#### • Validation statistique du Groupe 1.

L’analyse descriptive a été complétée par des tests inférentiels afin de confirmer la robus-

tesse des résultats et de garantir que les écarts entre modèles sont statistiquement significatifs. Des tests  $t$  de Student ont été réalisés pour comparer le modèle interne (DBPISR) à chaque modèle externe par pays, tandis qu’une ANOVA à un facteur a permis d’évaluer l’existence de différences globales entre l’ensemble des méthodes.

Cette approche garantit que les conclusions reposent sur une validation rigoureuse, dont la distribution des scores est illustrée par les boîtes à moustaches de la Figure 4.5. Ces graphiques synthétisent les performances (PSNR, SSIM et BRISQUE) pour les échographes GE utilisés en Algérie (Voluson S8) et en Égypte (Voluson P8).

En définitive, pour ce groupe, BSRGAN s’avère le plus adapté aux images acquises en Algérie, tandis que Real-ESRGAN prédomine en Égypte. Cette hétérogénéité confirme l’influence déterminante du modèle d’échographe, même au sein d’une même marque, sur l’efficacité des modèles de super-résolution.

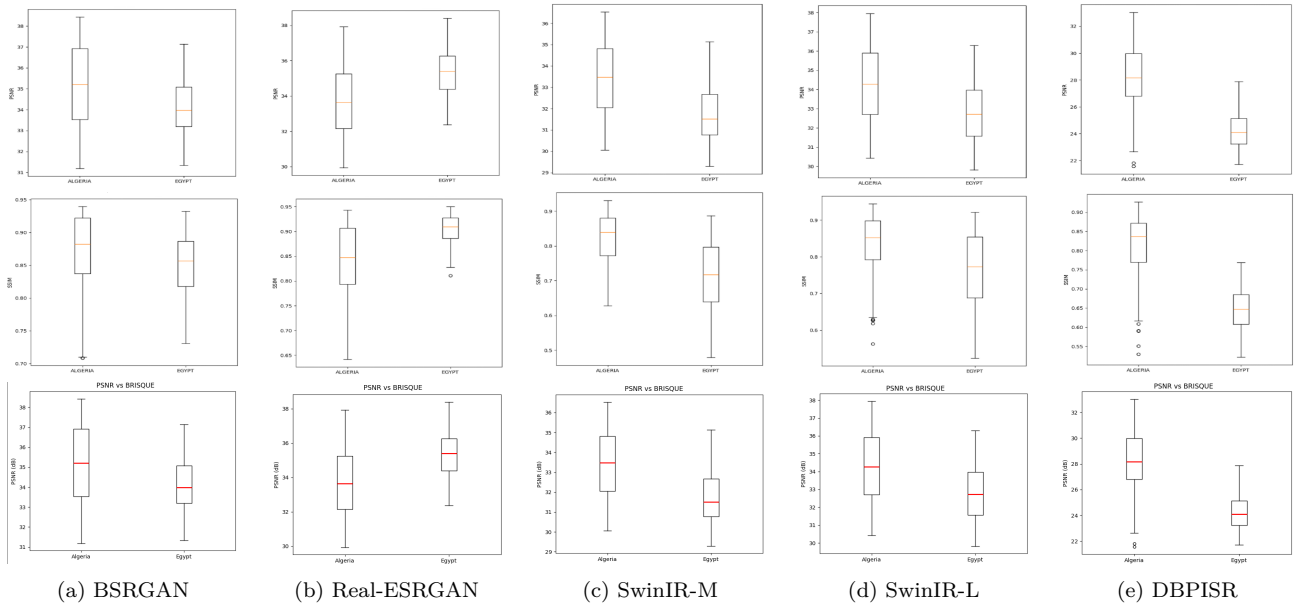


FIGURE 4.5 – Boxplots des métriques PSNR, SSIM et (PSNR/BRISQUE) pour les cinq modèles de super-résolution : (a) BSRGAN, (b) Real-ESRGAN, (c) SwinIR-M, (d) SwinIR-L et (e) DBPISR. Les statistiques sont calculées sur une base de 100 images acquises en Algérie (Voluson S8) et en Égypte (Voluson P8), avec un facteur de super-résolution de  $\times 4$ .

## Groupe 2 : Ghana, Malawi et Ouganda

Ce second groupe se distingue par une plus grande dispersion des performances, corrélée à la variété des équipements. Pour le Ghana, l’arbitrage entre fidélité et perception est porté par le duo BSRGAN ( $PSNR \approx 38$  dB) et Real-ESRGAN ( $SSIM \approx 0,91$ ). Les données issues du Malawi désignent BSRGAN comme le modèle le mieux adapté, celui-ci offrant le meilleur compromis entre intégrité structurelle et qualité visuelle. En Ouganda, bien que les scores soient globalement similaires ( $SSIM \approx 0,92$ ), une subtile distinction s’opère : BSRGAN optimise la perception (BRISQUE) là où Real-ESRGAN favorise la naturalité (NIQE). La robustesse de ces observations est étayée par une analyse statistique montrant des différences hautement significatives ( $p < 0,001$ ), validant l’impact du domaine source sur l’efficacité des modèles SR.

### • Validation statistique du Groupe 2.

Comme pour le Groupe 1, des tests  $t$  de Student ont été réalisés pour comparer le modèle interne (DBPISR) à chaque modèle externe par pays, complétés par une ANOVA à un facteur

pour évaluer les différences globales entre l'ensemble des méthodes. La normalité des distributions ayant été préalablement vérifiée, ces tests sont statistiquement valides. Compte tenu du nombre plus élevé de comparaisons multiples inhérent à ce groupe — trois pays contre deux pour le Groupe 1 —, une correction de Bonferroni a été appliquée afin de contrôler le risque d'erreurs de type I (voir Annexe 5.5.15). Les résultats confirment que la plupart de ces écarts sont statistiquement significatifs ( $p < 0,001$ ), attestant ainsi que les différences de performance observées entre les modèles ne résultent pas de simples fluctuations aléatoires, mais reflètent de véritables écarts robustes.

Les modèles BSRGAN et Real-ESRGAN se distinguent de manière cohérente dans les trois pays, avec des valeurs de PSNR et de SSIM supérieures : Ghana : 38 dB / 0,91 ; Malawi : 34,6 dB / 0,84 ; Ouganda : 35,9 dB / 0,92. BSRGAN se montre particulièrement performant en termes de qualité globale (PSNR, BRISQUE), tandis que Real-ESRGAN excelle sur la naturalité perceptuelle (NIQE).

Dans l'ensemble, ces résultats confirment que les tendances observées dans l'analyse descriptive traduisent de réels écarts robustes entre modèles.

La Figure 4.6 illustre les résultats correspondants pour les trois échographes utilisés au Ghana, au Malawi et en Ouganda.

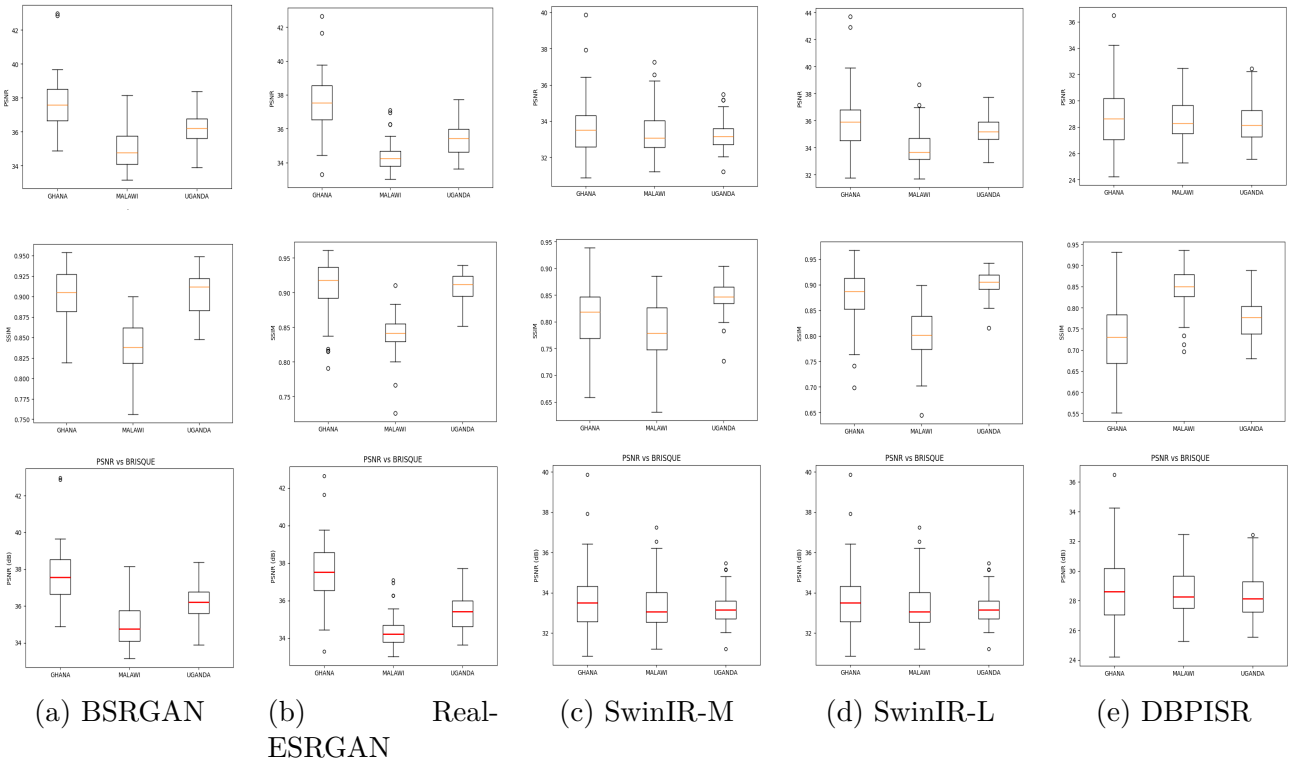


FIGURE 4.6 – Boxplots de PSNR, SSIM et (PSNR/BRISQUE) pour les cinq modèles SR : (a) BSRGAN, (b) Real-ESRGAN, (c) SwinIR-M, (d) SwinIR-L et (e) DBPISR. Ces boxplots montrent les statistiques de PSNR et SSIM calculées sur des bases de données composées de 75, 100 et 75 images pour le Ghana, le Malawi et l'Ouganda ; le facteur d'amélioration est de 4x.

La Figure 4.7 présente la distribution des scores PSNR–SSIM pour chaque pays. Cette visualisation de la dispersion met en évidence la robustesse des modèles, révélant des disparités que les moyennes globales ne permettent pas de déceler.

Nos analyses mettent en évidence deux observations majeures.

Premièrement, une relation positive forte et constante entre le PSNR et le SSIM, ainsi qu'entre le PSNR et le BRISQUE, témoigne du fait qu'une amélioration de la fidélité du signal



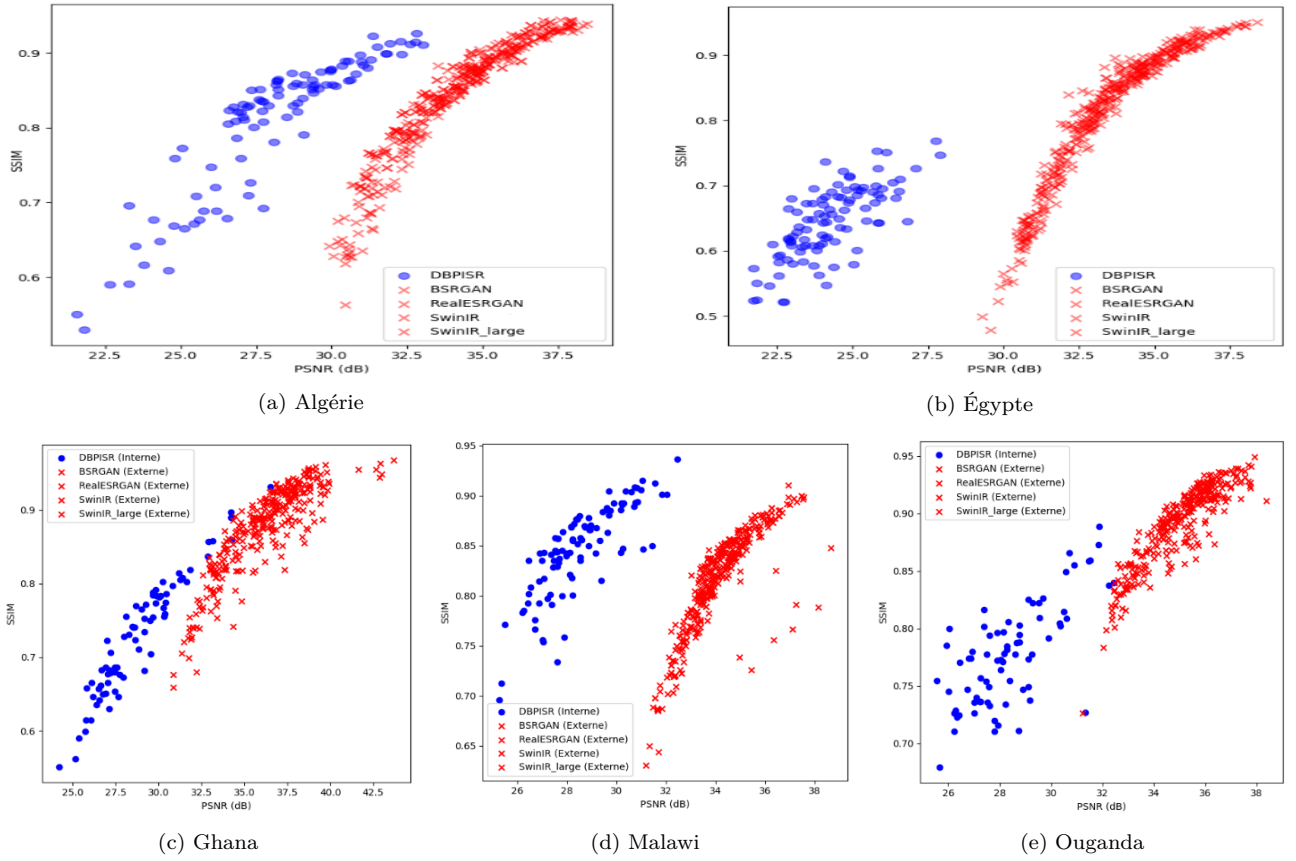


FIGURE 4.7 – Nuages de points PSNR–SSIM pour les cinq pays étudiés : (a) Algérie, (b) Égypte, (c) Ghana, (d) Malawi et (e) Ouganda. En bleu, le modèle DBPISR (apprentissage interne) ; en rouge, les modèles externes (BSRGAN, Real-ESRGAN, SwinIR (avec ses 2 variantes)).

s’accompagne d’une meilleure préservation structurelle des images super-résolues. Les résultats convergent vers une même idée : les gains en PSNR s’accompagnent d’une amélioration cohérente des indicateurs perceptuels.

Deuxièmement, les résultats convergent vers une même idée : les modèles SR externes bénéficient clairement de leur apprentissage sur des bases larges et variées, ce qui leur permet d’atteindre des niveaux élevés de PSNR et de SSIM (voir Fig. 4.7). À l’opposé, le DBPISR adopte une approche interne : il s’entraîne exclusivement sur l’image traitée, en exploitant ses redondances propres. Cette spécificité, bien qu’elle le rende indépendant de vastes bases de données, explique qu’il reste contraint par son mode d’apprentissage et montre rapidement ses limites.

Cette différence, observée dans tous les contextes, prend toutefois des formes variables : là où les résultats restent homogènes en Algérie et en Égypte (Fig. 4.7a, Fig. 4.7b), ils apparaissent plus dispersés au Malawi et en Ouganda (Fig. 4.7d, Fig. 4.7e). Cela suggère que le choix d’un modèle SR doit être guidé non seulement par ses performances globales, mais aussi par l’adéquation entre son mode d’entraînement et les spécificités des images échographiques locales.

Nous avons privilégié l’association PSNR-BRISQUE afin de mettre en perspective une mesure de fidélité au signal (avec référence) et un indicateur perceptuel aveugle (sans référence). Cette approche complémentaire permet de quantifier l’arbitrage entre précision de reconstruction numérique et qualité visuelle perçue.

À l’inverse, l’utilisation du NIQE a été écartée : ses fondements statistiques génériques s’avèrent peu discriminants face aux spécificités de l’imagerie échographique et corréler moins fidèlement avec l’expertise humaine. Par ailleurs, une combinaison SSIM-BRISQUE aurait engendré une redondance méthodologique, les deux métriques étant à dominante structurelle, ce

qui aurait limité la diversité des informations recueillies.

Cette stratégie d'évaluation est cruciale pour identifier le modèle de super-résolution le plus adapté à la variabilité technique des systèmes échographiques rencontrés dans les pays en développement. Pour prolonger l'analyse statistique des groupes 1 et 2, nous étudions désormais l'impact des cinq modèles de super-résolution sur les performances de classification, en comparaison avec les images originales.

## (2) - Analyse et évaluation des résultats de classification

Cette section évalue l'apport des modèles de super-résolution (SR) à la reconnaissance des plans anatomiques par le classificateur ConvNeXt-Base. L'objectif est de confirmer si le gain en qualité visuelle améliore la précision et la stabilité de la classification, particulièrement sur des jeux de données hétérogènes. Nous comparons les résultats obtenus sur les images originales et reconstruites pour les cohortes espagnoles et africaines.

### □ Jeu de données espagnol

L'analyse des tableaux 4.4 et 4.7 permet de dégager les conclusions suivantes :

En partant du modèle initial DBPISR, nous observons des améliorations de 1,00 % en accuracy, 1,15 % en F1-score et 1,38 % en kappa. SwinIR produit des gains comparables sur l'accuracy (+1,09 %) et le F1-score (+1,14 %), mais se démarque par une progression plus marquée du kappa (+2,25 %), suggérant une réduction plus efficace des erreurs de classification.

Le modèle Real-ESRGAN fournit les meilleures performances globales, avec +1,18 % en accuracy et +1,53 % en F1-score. SwinIR-L, sans gain en accuracy, améliore à la fois le F1-score (+1,18 %) et le kappa (+1,67 %). BSRGAN reste limité dans ce contexte.

TABLEAU 4.7 – Impact de la super-résolution sur les performances de classification des images échographiques (cas espagnol).

| Modèle de SR | $\Delta$ Accuracy (%) | $\Delta$ F1-score (%) | $\Delta$ Kappa (%) |
|--------------|-----------------------|-----------------------|--------------------|
| DBPISR       | +1.00                 | +1.15                 | +1.38              |
| BSRGAN       | +0.72                 | +0.95                 | +1.06              |
| Real-ESRGAN  | <b>+1.18</b>          | <b>+1.53</b>          | +1.60              |
| SwinIR-M     | +1.09                 | +1.14                 | <b>+2.25</b>       |
| SwinIR-L     | 0.00                  | +1.18                 | +1.67              |

Dans l'ensemble, Real-ESRGAN assure des gains réguliers sur accuracy et F1-score, tandis que SwinIR excelle davantage sur la robustesse des classifications, comme le reflète le kappa.

### □ Jeux de données africains

La mise en perspective des performances absolues (Tableau 4.5) et des écarts relatifs ( $\Delta$ , Tableau 4.8) révèle une dynamique singulièrement plus hétérogène que celle du jeu de données espagnol. Cette disparité met en exergue l'influence déterminante des contextes nationaux — incluant la diversité des protocoles d'acquisition et les spécificités techniques des équipements — sur l'efficacité des algorithmes de Super-Résolution (SR).

TABLEAU 4.8 – Impact de la super-résolution sur les métriques de classification : Algérie, Égypte, Ghana, Malawi et Ouganda.

| Modèles SR  | Algérie       |             |                   | Égypte        |             |                   | Ghana         |             |                   | Malawi        |             |                   | Ouganda       |             |                   |
|-------------|---------------|-------------|-------------------|---------------|-------------|-------------------|---------------|-------------|-------------------|---------------|-------------|-------------------|---------------|-------------|-------------------|
|             | $\Delta$ Acc. | $\Delta$ F1 | $\Delta$ $\kappa$ | $\Delta$ Acc. | $\Delta$ F1 | $\Delta$ $\kappa$ | $\Delta$ Acc. | $\Delta$ F1 | $\Delta$ $\kappa$ | $\Delta$ Acc. | $\Delta$ F1 | $\Delta$ $\kappa$ | $\Delta$ Acc. | $\Delta$ F1 | $\Delta$ $\kappa$ |
| DBPISR      | -16%          | -21.34%     | -35.55%           | +9%           | +5.06%      | -6.05%            | +42.67%       | +30.71%     | +31.14%           | -9%           | -19.73%     | -30.6%            | +34.67%       | +12.63%     | +17.21%           |
| BSRGAN      | 0%            | -0.1%       | -0.8%             | -11%          | -9.3%       | -4.03%            | -1.33%        | -1.77%      | +8.8%             | -6%           | -4.2%       | -3.96%            | 0%            | -5.08%      | -13.95%           |
| Real-ESRGAN | +11%          | +5.57%      | +11.66%           | +6%           | +3.44%      | +2.17%            | +13.34%       | +14.51%     | +14.14%           | +13%          | +6.54%      | +8.2%             | 0%            | -0.07%      | +1.95%            |
| SwinIR-M    | +9%           | +4.35%      | +7.37%            | +2%           | +0.92%      | +1.76%            | +6.67%        | +8.07%      | +4.49%            | +18%          | +10.15%     | +12.2%            | -8%           | -8.18%      | -8.35%            |
| SwinIR-L    | -1%           | -0.35%      | -0.67%            | +1%           | +0.25%      | +4.27%            | 0%            | -0.26%      | +1.77%            | -7%           | -5.57%      | +1.33%            | -1.34%        | -1.42%      | +1.73%            |

Les résultats mettent en évidence une variabilité marquée des performances selon les modèles et les contextes nationaux. BSRGAN se distingue par des résultats globalement défavorables, avec des baisses importantes observées en Ouganda (0 % en Accuracy, -5,08 % en F1-score, -13,95 % en Kappa) et en Égypte (-11 %, -9,3 %, -4,03 %), alors que son impact demeure marginal en Algérie, au Ghana et au Malawi.

DBPISR, pris comme modèle initial, présente un profil très contrasté : il génère des gains significatifs au Ghana (+42,67 % Accuracy, +30,71 % F1-score, +31,14 %) et en Ouganda (+34,67 %, +12,63 %, +17,21 %), mais entraîne une nette baisse en Algérie (-16 %, -21,34 %, -35,55 %) et au Malawi (-9 %, -19,73 %, -30,6 %). En Égypte, il améliore l'Accuracy (+9 %) et le F1-score (+5,06 %) tout en réduisant le Kappa (-6,05 %), indiquant une cohérence limitée des prédictions.

Real-ESRGAN se distingue par sa stabilité et des gains constants : en Algérie (+11 % Accuracy, +5,57 % F1, +11,66 % Kappa), en Égypte (+6 %, +3,44 %, +2,17 %), au Ghana (+13,34 %, +14,51 %, +14,14 %) et au Malawi (+13 %, +6,54 %, +8,2 %). Son impact en Ouganda reste toutefois marginal (0 % Accuracy, -0,07 % F1, +1,95 % Kappa).

SwinIR-M présente des améliorations modérées en Algérie (+9 % Accuracy, +4,35 % F1, +7,37 %), en Égypte et au Ghana, et plus prononcées au Malawi (+18 %, +10,15 %, +12,2 %), mais décline en Ouganda (-8 %, -8,18 %, -8,35 %). SwinIR-L se montre globalement peu performant, avec des variations négligeables ou négatives, à l'exception d'une légère progression du Kappa en Égypte (+4,27 %).

Ces résultats illustrent l'impact conjoint de l'architecture des modèles et des propriétés des jeux de données.

En croisant les résultats relatifs à la qualité d'image (PSNR, SSIM) avec les performances de classification, plusieurs constats émergent.

- REAL-ESRGAN offre un compromis équilibré entre une bonne qualité d'image et de solides performances de classification dans la majorité des pays, à l'exception de l'Ouganda où ses apports demeurent plus limités. Il s'impose néanmoins comme le modèle le plus adapté dans l'ensemble des cinq contextes étudiés.
- BSRGAN, malgré des scores élevés en PSNR et SSIM (notamment en Ouganda), ne parvient pas à convertir cette qualité en gains de classification, révélant une difficulté de généralisation vers la tâche clinique.
- SWINIR, bien que générant une qualité d'image inférieure, compense partiellement cette faiblesse par de meilleures performances de classification, en particulier au Malawi, ce qui souligne sa capacité d'adaptation aux données.
- DBPISR présente, quant à lui, des valeurs de PSNR très faibles, directement corrélées à des performances médiocres en classification dans presque tous les pays, à l'exception notable de l'Ouganda et du Ghana.

L'évaluation a porté sur cinq modèles de super-résolution (SR) — DBPISR, BSRGAN, REAL-ESRGAN, SWINIR-M et SWINIR-L — appliqués à l'amélioration des images échographiques dans divers contextes africains. Les performances ont été analysées à travers deux métriques de qualité d'image (PSNR, SSIM) et trois métriques de classification (Accuracy, F1-score et coefficient Kappa) (voir Figure 4.8).

On constate que REAL-ESRGAN reste le plus constant, capable de produire des images de qualité même lorsque les moyens techniques sont restreints. Les modèles de la famille SWINIR semblent mieux convenir à des contextes disposant d'infrastructures plus performantes, comme en Algérie ou au Malawi. En revanche, DBPISR, dont les performances demeurent insuffisantes et inconstantes, apparaît difficilement transposable à une mise en œuvre élargie.

#### □ Observation clés : évaluation des performances des modèles par métrique

- **PSNR** : BSRGAN et Real-ESRGAN présentent des valeurs élevées, traduisant une recons-



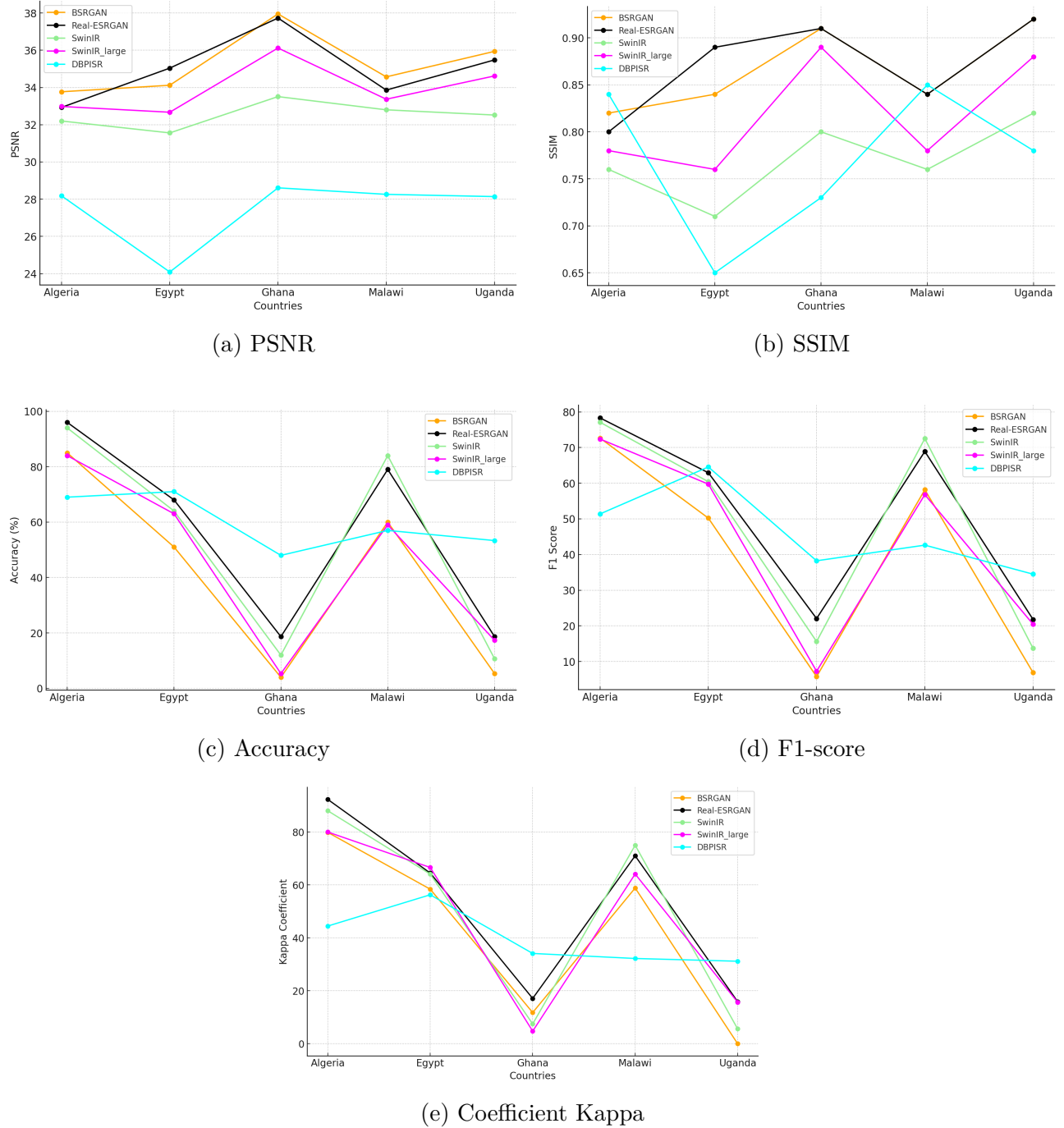


FIGURE 4.8 – Comparaison des performances des modèles de super-résolution pour cinq pays africains : Algérie, Égypte, Ghana, Malawi et Ouganda.

truction fidèle et peu bruitée. À l'inverse, DBPISR affiche systématiquement de faibles scores, en particulier en Égypte, révélant une qualité d'image dégradée.

- **SSIM** : Real-ESRGAN obtient de manière constante des scores supérieurs à 0,9, indiquant une excellente fidélité structurale. SwinIR-L et BSRGAN atteignent également de bonnes performances, surtout en Algérie et au Ghana.

- **NIQE** : Real-ESRGAN enregistre les valeurs les plus basses, gage d'une forte naturalité visuelle, notamment au Ghana (1,34) et en Ouganda (3,69). DBPISR, en revanche, génère les scores les plus élevés, synonymes d'artéfacts importants, avec un maximum en Algérie (19,62). BSRGAN maintient des résultats compétitifs en Ouganda et au Malawi.

- **BRISQUE** : BSRGAN obtient régulièrement de faibles valeurs, notamment en Algérie (44,28)

et en Ouganda (43,20), attestant de sa qualité perceptuelle. DBPISR, à l’opposé, se distingue par des scores très élevés (jusqu’à 89,84 au Ghana), révélant une mauvaise qualité visuelle. Real-ESRGAN reste intermédiaire, avec ses meilleurs résultats observés en Égypte (56,24).

- **Accuracy** : Real-ESRGAN se distingue par une précision quasi parfaite en Algérie et des résultats solides au Malawi. Tous les modèles peinent cependant au Ghana, ce qui suggère des limites liées aux données. DBPISR reste le moins performant, avec seulement 48 % au Ghana et 57 % au Malawi.
- **F1-score** : Real-ESRGAN et SwinIR obtiennent d’excellents résultats en Algérie et au Malawi, tandis que le Ghana reste problématique pour l’ensemble des modèles.
- **Coefficient Kappa** : Real-ESRGAN domine avec les valeurs les plus élevées, notamment en Algérie (92,33 %) et au Malawi (71 %). SwinIR-M et SwinIR-L suivent, tandis que DBPISR échoue particulièrement en Égypte.

### (3) - Évaluation clinique – MOS appliqué aux modèles SR

Quatre experts médicaux ont été mobilisés afin d’évaluer la qualité et la pertinence diagnostique d’images échographiques reconstituées par différents modèles de super-résolution. Leur jugement a été recueilli au moyen d’une grille comprenant sept critères, notés sur une échelle de type Likert<sup>1</sup>, ainsi qu’une question binaire visant à apprécier le risque d’erreur diagnostique.

L’évaluation s’est attachée à des dimensions concrètes et directement utiles à la pratique clinique : netteté des contours, visibilité des détails, réduction du bruit et des artéfacts, mais aussi utilité globale de l’image pour le diagnostic.

Ces réponses ont permis de calculer un Mean Opinion Score (MOS) et un écart-type pour chaque modèle, fournissant à la fois une estimation de la performance perçue et une indication de la régularité des jugements. La confrontation des résultats entre experts a mis en évidence des points de convergence qui ont été synthétisés graphiquement.

L’analyse a porté sur cinq modèles de super-résolution — BSRGAN, DBPISR, SwinIR-M, SwinIR-L et Real-ESRGAN — comparés de manière systématique aux images originales utilisées comme référence.

Comme le montre la figure 4.9a, chaque expert a identifié un modèle SR distinct comme étant le plus performant selon son évaluation :

Dr.1 privilégie BSRGAN (3.43/5,  $\sigma = 0.54$ ), en raison d’une appréciation moyenne mais cohérente, traduisant une certaine constance dans son jugement. Dr.2 retient Real-ESRGAN (3.71/5,  $\sigma = 0.95$ ), un modèle jugé supérieur sur le plan qualitatif, bien que son évaluation présente une variabilité modérée.

Dr.3 désigne DBPISR (3.43/5,  $\sigma = 1.13$ ) comme la meilleure option, essentiellement par stabilité relative parmi des scores globalement modestes. Dr.4 distingue clairement SwinIR-M (4.71/5,  $\sigma = 0.49$ ), avec des notes très élevées et une remarquable cohérence, traduisant une forte confiance dans son jugement.

En agréant l’ensemble des évaluations, le modèle Real-ESRGAN se démarque comme le candidat le plus robuste (MOS global = 3.75/5,  $\sigma = 1.04$ ). Cette position ne résulte pas uniquement d’une moyenne supérieure, mais aussi de sa récurrence dans les préférences (notamment chez Dr.2 et confirmé positivement par Dr.3 et Dr.4 qui l’ont mis en 2ème position).

Ainsi, même si chaque expert a exprimé une préférence différente, Real-ESRGAN émerge comme le choix le plus consensuel, combinant une qualité perçue globalement élevée et une variabilité acceptable. (Voir Figure 4.9b)

---

1. L’échelle de Likert utilisée comporte 5 niveaux : 1 = qualité très faible, 2 = faible, 3 = acceptable, 4 = bonne, 5 = excellente.

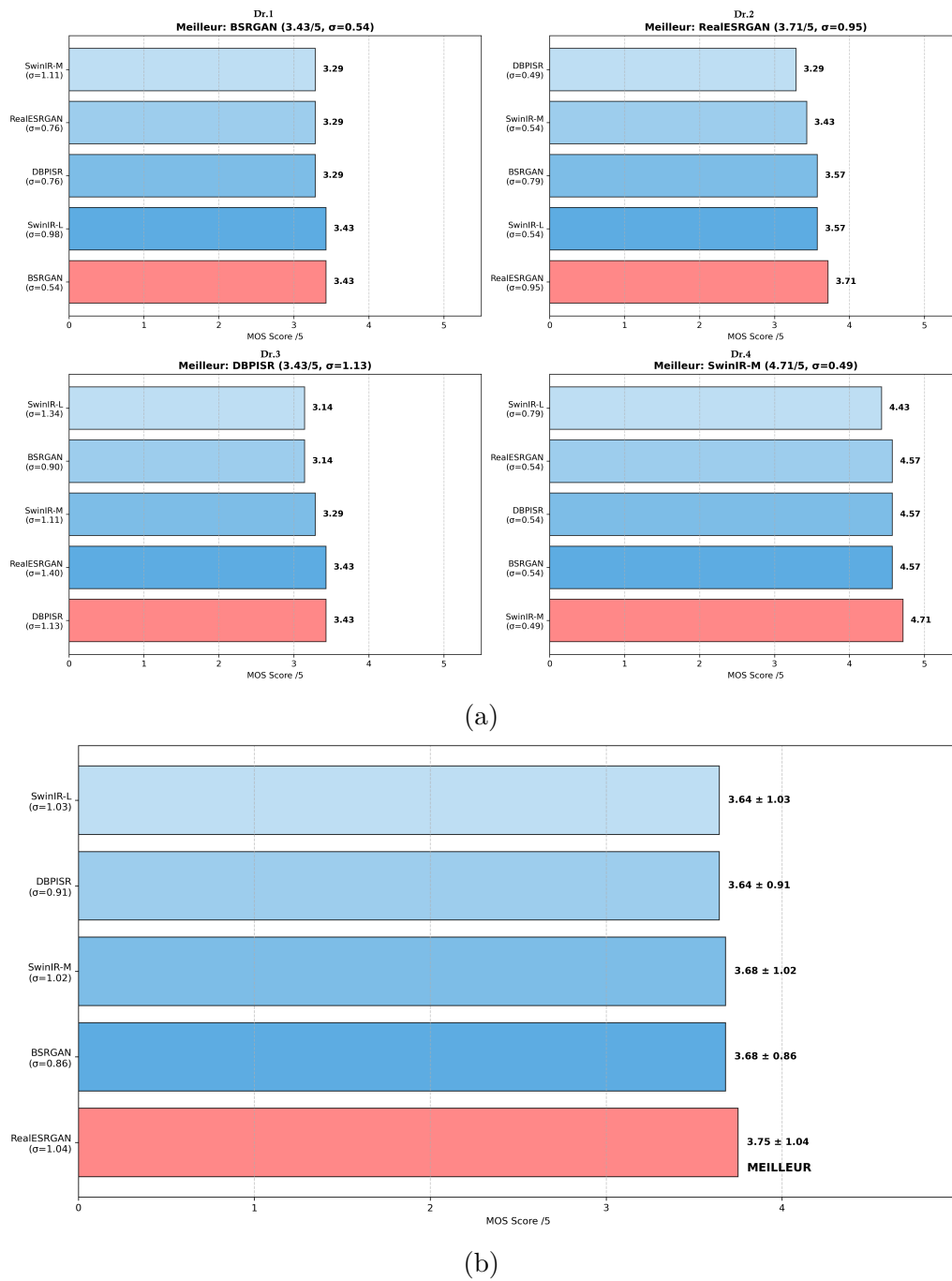


FIGURE 4.9 – Comparaison des évaluations des modèles de super-résolution : (a) appréciation individuelle d'un expert médical, (b) consensus général de tous les experts confondus

#### (4) - Synthèse de l'étude quantitative et perceptuelle

Nos résultats confirment que les modèles de super-résolution entraînés sur des bases externes offrent une performance supérieure, traduite par des valeurs de PSNR et de SSIM plus élevées, grâce à l'apport de données d'entraînement variées et représentatives. En revanche, le DBPISR, bien que libéré de la contrainte de disposer de vastes corpus, se montre rapidement limité dans sa capacité à restituer une image de haute qualité. Ces différences, observées de manière homogène dans plusieurs contextes géographiques, soulignent qu'un modèle SR doit être choisi non seulement pour sa performance moyenne, mais aussi pour sa capacité à s'adapter aux spécificités locales des images échographiques.

En pratique clinique, la super-résolution se révèle un levier efficace pour améliorer l'inter-

prétation échographique dans des environnements à ressources limitées. Elle accroît la lisibilité des structures fœtales, réduit les artéfacts et compense partiellement la dépendance au niveau d'expertise de l'opérateur. L'implémentation de ces approches dans des dispositifs embarqués ou à distance représente une solution réaliste pour optimiser l'utilisation du matériel existant et favoriser un suivi prénatal plus équitable.

## Conclusion de la Partie I

Au terme de cette analyse, les modèles à apprentissage externe s'imposent nettement face à l'approche interne. La convergence des indicateurs est claire : ils offrent une meilleure restitution des textures (PSNR, SSIM, BRISQUE) et une cohérence anatomique validée par la classification. De leur côté, les experts cliniques confirment cette supériorité par une préférence marquée (scores MOS).

L'analyse statistique confirme la robustesse de ces observations : les écarts mesurés sont significatifs et dépassent la simple variabilité des données. Le modèle interne, bien que capable de fonctionner sans données externes, atteint rapidement ses limites face aux variations des images échographiques fœtales.

Cette première partie a donc prouvé que la super-résolution ne se contente pas d'améliorer la fidélité numérique : elle accroît la lisibilité des structures fœtales et la clarté des détails anatomiques, tout en réduisant le bruit. Cela démontre que l'enjeu n'est pas purement esthétique : la super-résolution influe réellement sur la valeur diagnostique et la fiabilité des analyses automatiques.

Fort de ces résultats, nous avons souhaité poursuivre l'étude en explorant plus avant les modèles SR à apprentissage externe, afin d'évaluer leur pertinence pour une utilisation pratique en gynécologie. Cette perspective ouvre la voie à la Partie II, où nous croiserons mesures objectives et jugement clinique pour mieux cerner la valeur des images super-résolues dans un contexte médical.

## 4.3 Partie II : Vers une analyse perceptuelle et statistique multidimensionnelle

Dans cette partie, nous nous focalisons sur les modèles à apprentissage externe, complétés par trois modèles récents de l'état de l'art — A-ESRGAN, HAT et Swin2SR — sélectionnés pour représenter différentes familles d'architectures (GAN, Transformers hybrides et approches basées sur l'attention) et pour leurs performances reconnues dans la littérature en super-résolution. Nous adoptons ici une stratégie d'évaluation multidimensionnelle articulée autour de trois axes complémentaires :

- **Une analyse objective étendue**, fondée sur des métriques perceptuelles avancées (NIQE, FID, LPIPS, GMSD), visant à évaluer le réalisme visuel, la fidélité structurelle et la préservation des textures.
- **Une évaluation subjective**, reposant sur un protocole MOS en double aveugle,<sup>2</sup> afin d'apprécier la qualité perçue et l'utilité diagnostique des images reconstruites.
- **Une validation statistique**, destinée à vérifier la significativité des différences observées entre modèles.

Cette approche complète l'évaluation quantitative présentée en Partie I et permet de déterminer si les gains mesurés par les métriques objectives se traduisent effectivement par une amélioration perceptible et cliniquement pertinente.

---

2. MOS en double aveugle : l'experte a évalué les images sans connaître l'algorithme utilisé ni l'ordre des images, afin de lier la qualité perçue à l'utilité diagnostique.

L'appréciation d'une image reconstruite ne se limite ni à la similarité numérique avec la référence, ni à son impact sur un classificateur. Une forte proximité pixelique ne garantit pas nécessairement une qualité visuelle satisfaisante et peut masquer des artefacts subtils. En croisant métriques avancées et jugement clinique, cette Partie II vise ainsi à proposer une analyse plus globale et directement applicable au contexte diagnostique.

### 4.3.1 Méthodologie

L'objectif de cette section est de décrire le dispositif méthodologique élaboré pour évaluer la reconstruction d'images échographiques fœtales par des modèles avancés de super-résolution. Deux grandes approches ont été examinées : les méthodes fondées sur les réseaux antagonistes génératifs (AESRGAN, BSRGAN, Real-ESRGAN) et celles reposant sur des architectures Transformer (SwinIR, Swin2SR, HAT). Après avoir introduit une formulation du problème adaptée au cadre biomédical, nous présentons les fondements des architectures étudiées, suivis par la description du jeu de données et de la stratégie expérimentale. Enfin, nous détaillons un protocole d'évaluation multimodal conçu pour apprécier la restitution des structures anatomiques critiques, dans des conditions d'acquisition variées et représentatives de la pratique clinique.

Le déroulement méthodologique de cette étude s'articule autour des étapes suivantes :

- **Formulation du problème** : définition et cadrage du problème spécifiquement adaptés au contexte médical.
- **Analyse des modèles** : examen des principales caractéristiques et différences des modèles de super-résolution étudiés.
- **Jeu de données et protocole expérimental** : description détaillée des données utilisées ainsi que de la stratégie expérimentale adoptée pour garantir la rigueur et la reproductibilité des résultats.
- **Évaluation multimodale** : mise en place d'un protocole conçu pour mesurer la qualité de restitution des structures anatomiques dans des contextes cliniques variés et pertinents.

De cette étude comparative, deux architectures se sont clairement distinguées par leurs performances remarquables, en particulier par leur capacité à restituer des détails fins tout en limitant les artefacts visuels.

Bien que les modèles évalués présentent des performances élevées, leur entraînement initial sur des bases de données génériques ou hétérogènes peut induire un décalage de domaine par rapport aux images échographiques spécifiques. Or, pour garantir une applicabilité clinique, il est essentiel d'adapter les modèles aux caractéristiques réelles du domaine médical, en tenant compte des contraintes d'acquisition et des exigences en termes de fidélité diagnostique.

## Formulation du problème

En super-résolution d'image unique, l'enjeu est de reconstruire une image haute résolution  $I_{HR}$  à partir d'une version dégradée en basse résolution  $I_{LR}$ . Le processus de dégradation, rarement connu avec précision et souvent non linéaire, rend cette tâche particulièrement "ill-posed". On cherche ainsi à approximer une fonction de reconstruction  $F$  telle que :

$$I_{HR} \simeq F(I_{LR}). \quad (4.1)$$

L'objectif de la super-résolution est la restitution des détails anatomiques fins avec un minimum d'artefacts. Pourtant, l'échographie fœtale expose rapidement les limites de cette démarche : *speckle*, contraste réduit, résolution physique limitée. Dans ce contexte, obtenir une image exploitable par le clinicien n'est pas seulement un défi algorithmique ; c'est aussi un test de la capacité des modèles à respecter l'information diagnostique sans trahir la réalité anatomique.

## Jeux de données d'évaluation

Cette étude s'appuie sur une compilation hétérogène de données échographiques materno-fœtales, couvrant une large gamme de stades gestationnels, de plans anatomiques et de conditions cliniques. Afin de permettre une évaluation exhaustive des modèles de super-résolution sur des tâches cliniquement pertinentes — telles que la biométrie fœtale, la détection d'anomalies et l'interprétation automatisée — nous avons sélectionné de manière systématique cinq jeux de données accessibles publiquement.

Le cadre de sélection a été conçu pour refléter l'ensemble de la variabilité des acquisitions échographiques, incluant :

- ☐ des données synthétiques pour des tests de référence contrôlés,
- ☐ des acquisitions cliniques réelles provenant de différentes régions géographiques et plateformes d'imagerie,
- ☐ des entrées multimodales telles que des images extraites de séquences vidéo.

Cette approche multi-sources permet une évaluation rigoureuse à la fois de la capacité de généralisation des modèles et de leur pertinence clinique.

### ■ Données synthétiques pour un benchmark contrôlé

Le jeu de données *FPUS23* [148] a été utilisé comme banc d'essai initial afin d'observer le comportement des modèles dans un environnement synthétique contrôlé. Composé de volumes échographiques fœtaux générés à partir d'un fantôme réaliste, il fournit des images avec des plans anatomiques parfaitement alignés et sans artefacts cliniques. Ce cadre idéal permet d'isoler le processus de reconstruction de chaque modèle, en éliminant les variables confondantes présentes dans les données cliniques. Les caractéristiques principales de ce jeu de données sont résumées dans le Tableau 4.9.

TABLEAU 4.9 – Jeu de données *FPUS23* [148]

| Catégorie                                                                                                                                                                                                               | Description                                                                                           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Nombre d'images                                                                                                                                                                                                         | Plus de 5 000                                                                                         |
| Résolution                                                                                                                                                                                                              | 664 × 388 pixels                                                                                      |
| Format d'image                                                                                                                                                                                                          | PNG en niveaux de gris                                                                                |
| Classes                                                                                                                                                                                                                 | BPD, AC, FL, Non diagnostique                                                                         |
| Répartition                                                                                                                                                                                                             | Sous-ensembles non chevauchants                                                                       |
| Licence                                                                                                                                                                                                                 | CC-BY 4.0                                                                                             |
| Appareils échographiques                                                                                                                                                                                                | Philips Epiq-7 avec sonde X6-1 xMATRIX                                                                |
| Source                                                                                                                                                                                                                  | <a href="https://github.com/bharathprabakaran/FPUS23">https://github.com/bharathprabakaran/FPUS23</a> |
| BPD : Biparietal Diameter (diamètre bipariétal), AC : Abdominal Circumference (circonférence abdominale), FL : Femur Length (longueur du fémur), Non diagnostique : images non exploitables pour la mesure biométrique. |                                                                                                       |

### ■ Données cliniques réelles

Afin d'évaluer les performances en conditions cliniques réelles, trois jeux de données complémentaires ont été analysés. Le jeu de données espagnol [21] fournit une base de référence de grande ampleur ; le corpus africain [161] permet d'examiner la robustesse face à des décalages de domaine liés aux équipements et aux contextes géographiques ; enfin, le jeu de données

*PSFHS* [26] qui cible l'imagerie transpérinéale, tâche cliniquement critique. Les caractéristiques principales de ces trois jeux de données sont synthétisées dans les Tableaux 4.10 et 4.11.

TABLEAU 4.10 – Caractéristiques des jeux de données espagnol et africain [21, 161]

| Catégorie                          | Jeu de données espagnol                                                             | Jeu de données africain                                                             |
|------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Pays                               | Espagne                                                                             | Malawi, Algérie, Ouganda, Ghana, Égypte                                             |
| Nombre d'images                    | Plus de 5 000                                                                       | 450                                                                                 |
| Résolution                         | 224 × 224 (redimensionné)                                                           |                                                                                     |
| Format d'image                     | PNG en niveaux de gris (anonymisé)                                                  |                                                                                     |
| Catégories anatomiques principales | Abdomen, cerveau, fémur, thorax, col de l'utérus, autres                            | Abdomen, cerveau, fémur, thorax (pas toujours disponible)                           |
| Appareils échographiques           | Voluson E6, S8, S10, Aloka                                                          | GE, Siemens, Mindray, Edan, Aloka                                                   |
| Source                             | <a href="https://doi.org/10.5281/zenodo.3904280">doi.org/10.5281/zenodo.3904280</a> | <a href="https://doi.org/10.5281/zenodo.7540448">doi.org/10.5281/zenodo.7540448</a> |

TABLEAU 4.11 – Caractéristiques du jeu de données *PSFHS* pour la super-résolution [26]

| Catégorie          | Description                                                   |
|--------------------|---------------------------------------------------------------|
| Type d'image       | Échographie transpérinéale 2D (obstétrique)                   |
| Nombre d'images    | 1 358                                                         |
| Format d'image     | .mha (compatible ITK, convertible)                            |
| Taille de l'image  | Résolution uniforme (640 × 480)                               |
| Qualité de l'image | Images cliniques réelles (avec bruit et artefacts)            |
| Matériel           | ObEye et Esaote MyLab (sondes 3,5 MHz)                        |
| Téléchargement     | <a href="https://zenodo.org/10969427">zenodo.org/10969427</a> |

## ■ Exploitation de séquences échographiques vidéo

Dans le but d'enrichir la représentativité des données d'évaluation, nous avons utilisé le jeu de données *2D Ultrasound Video Segmentation* [54], composé de séquences vidéo d'échographie transvaginale dynamiques. À partir de ces séquences, 5 217 trames individuelles ont été extraites à l'aide de *Segment Anything* [91]. Issues d'enregistrements transvaginaux dynamiques, ces trames se distinguent des acquisitions B-mode classiques tout en constituant un corpus pertinent pour l'analyse comparative. Leur utilisation permet non seulement d'accroître le volume des cas testés, mais aussi d'explorer la capacité de généralisation des modèles à des modalités d'imagerie plus hétérogènes. Le Tableau 4.12 synthétise les caractéristiques de ces données.

## ■ Prétraitement standardisé et génération des basses résolutions

Afin d'assurer une comparaison rigoureuse entre modèles, toutes les données exploitées proviennent des cinq ensembles retenus, utilisés uniquement en phase de test. Le schéma de prétraitement se décline en trois opérations : conversion en images 8 bits en échelle de gris, normalisation des intensités sur l'intervalle [0,1], et génération des images basse résolution via un sous-échantillonnage bicubique appliqué avec un facteur ×4.

TABLEAU 4.12 – Caractéristiques du jeu de données *ISTHMIC* Contraction [54]

| Catégorie                               | Description                                                     |
|-----------------------------------------|-----------------------------------------------------------------|
| Type d'image                            | Échographie transvaginale 2D (cervicale et isthmique)           |
| Nombre d'images                         | 5 217 images extraites de 30 séquences vidéo                    |
| Format d'image                          | PNG en niveaux de gris (anonymisé)                              |
| Taille de l'image                       | 640 × 480 pixels                                                |
| Qualité de l'image                      | Données cliniques réelles avec artefacts de mouvement possibles |
| Matériel                                | Appareils GE Voluson E10 et E8                                  |
| Séquences vidéo originales téléchargées | <b>Mendeley Dataset</b>                                         |

## Architectures de modèles sélectionnés

Nous avons retenu huit modèles de super-résolution issus de six architectures majeures, représentatives des tendances actuelles. La sélection couvre équitablement les deux approches dominantes : les modèles adversariaux (GAN), visant un rendu perceptuel réaliste, et les modèles Transformer, exploitant l'auto-attention pour une reconstruction structurelle fidèle.

La répartition est la suivante :

- Quatre modèles GAN
- Quatre modèles Transformer

### ★ Modèles basés sur les GAN

Le cadre des réseaux antagonistes génératifs (GAN) [63] a donné naissance à plusieurs modèles particulièrement performants pour la super-résolution. Leur force réside dans la capacité à restituer des images nettes et détaillées, même si cet avantage peut parfois se faire au détriment d'une fidélité stricte au signal d'origine. Dans cette partie 2, nous retenons trois modèles représentatifs :

**AESRGAN** [195], une extension d'ESRGAN [194], qui introduit une perte perceptuelle asymétrique afin d'améliorer la restitution texturale et de réduire les artefacts. Deux configurations existent : A-ESRGAN(1D), avec un discriminateur unique, et A-ESRGAN(MD), utilisant plusieurs discriminateurs.

**BSRGAN** [206], un modèle aveugle entraîné à partir d'un schéma de dégradation complexe et aléatoire (flou, bruit, compression), ce qui lui confère une meilleure robustesse face à des dégradations inconnues.

**Real-ESRGAN**[192], qui prolonge BSRGAN en intégrant un discriminateur de type U-Net et un ensemble de dégradations plus réalistes, améliorant son efficacité sur des images issues de contextes cliniques variés.

### ★ Modèles de Vision Transformers

L'introduction des Transformers [187] dans le champ de la vision par ordinateur a profondément renouvelé les approches de reconstruction d'image, notamment en super-résolution médicale. Alors que les réseaux convolutionnels restent limités par la portée locale de leurs filtres, les Transformers permettent de modéliser des dépendances à longue distance, reliant entre elles des régions d'image éloignées. Cette propriété, qui peut paraître secondaire dans



d'autres contextes, prend toute son importance en échographie fœtale, où les structures anatomiques dépassent souvent la portée locale des noyaux convolutionnels.

Dans ce travail, nous avons retenu trois architectures emblématiques de cette famille.

**Le modèle SwinIR** [105], fondé sur une organisation hiérarchique et un mécanisme d'attention par fenêtres décalées, a été évalué en deux variantes : la version standard (SwinIR-M) et une version élargie (SwinIR-L), afin d'examiner l'influence de la capacité du réseau sur les performances.

**Le modèle Swin2SR** [35], qui prolonge SwinIR, intègre des raffinements issus de Swin Transformer V2 [114], tels que la post-normalisation résiduelle et l'attention cosinus échelonnée, visant à stabiliser l'apprentissage et améliorer la reconstruction.

Enfin, **le modèle HAT** [29] repose sur une approche hybride qui combine attention locale et inter-échelles, dans l'objectif de préserver à la fois la finesse des détails et la cohérence structurale globale.

Ces trois déclinaisons permettent de comparer différents compromis entre expressivité, stabilité et fidélité diagnostique.

## ★ Protocole expérimental

Notre objectif est d'évaluer si les modèles de super-résolution disponibles publiquement conservent leur capacité à restituer des informations utiles en pratique clinique lorsqu'ils sont appliqués sans adaptation préalable. Pour cela, nous les avons testés directement sur des images échographiques fœtales, un type de données qui diffère substantiellement des corpus sur lesquels ces modèles ont été entraînés (Voir Tableau 4.13). En omettant volontairement toute étape de ré-entraînement ou d'optimisation spécifique à l'échographie, nous mesurons leur aptitude à la généralisation hors distribution (OOD)<sup>3</sup>. Ce choix expérimental permet d'interroger de manière rigoureuse la pertinence et les limites de ces approches dans des conditions réelles d'utilisation médicale.

TABLEAU 4.13 – Origine des modèles de super-résolution pré-entraînés utilisés dans la partie II de l'étude.

| Modèle      | Données d'entraînement                                     | Source du checkpoint                                                                                                  |
|-------------|------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| A-ESRGAN    | DIV2K                                                      | <a href="https://github.com/stroking-fishes-ml-corp/A-ESRGAN">https://github.com/stroking-fishes-ml-corp/A-ESRGAN</a> |
| BSRGAN      | DIV2K, Flickr2K, WED (Waterloo Exploration Database), FFHQ | <a href="https://github.com/cszn/BSRGAN">https://github.com/cszn/BSRGAN</a>                                           |
| Real-ESRGAN | DIV2K, Flickr2K, FFHQ, Outdoor Scene Training (OST300)     | <a href="https://github.com/xinntao/Real-ESRGAN">https://github.com/xinntao/Real-ESRGAN</a>                           |
| SwinIR      | DIV2K, Flickr2K, BSD68, Urban100, Classic5, LIVE1          | <a href="https://github.com/cszn/KAIR">https://github.com/cszn/KAIR</a>                                               |
| Swin2SR     | DIV2K, DF2K                                                | <a href="https://github.com/mv-lab/swin2sr">https://github.com/mv-lab/swin2sr</a>                                     |
| HAT         | DIV2K, DF2K, Flickr2K, ImageNet                            | <a href="https://github.com/XPixelGroup/HAT">https://github.com/XPixelGroup/HAT</a>                                   |

Toutes les expériences ont été réalisées sur une station de travail locale fonctionnant sous Debian GNU/Linux 12 (Bookworm), équipée d'un GPU NVIDIA RTX A4500 doté de 20 Go de mémoire vidéo (VRAM). Les modèles ont été implémentés avec PyTorch (v2.6.0). L'évaluation a porté à la fois sur la qualité de reconstruction et sur l'efficacité computationnelle, mesurée par le temps d'inférence moyen par image.

3. Out-of-Distribution (OOD, hors distribution) : se dit des données qui diffèrent statistiquement de celles utilisées pour entraîner le modèle. La généralisation OOD mesure la capacité d'un modèle à fournir des prédictions fiables sur ces nouvelles données.

## ★ Méthodologie d'évaluation

Pour garantir une évaluation rigoureuse et multidimensionnelle, nous avons conçu un protocole intégrant simultanément : i) des mesures quantitatives classiques, ii) une analyse qualitative par inspection visuelle experte, et iii) une appréciation de l'applicabilité clinique par une experte médicale<sup>4</sup>.

Cette articulation permet d'obtenir une vision équilibrée, reliant performances numériques et pertinence opérationnelle.

### a) Métriques quantitatives

En complément des indicateurs présentés en Partie 1 (PSNR, SSIM, NIQE), nous avons retenu trois métriques complémentaires visant à capturer des aspects perceptuels et structurels :

**LPIPS.** Pour deux images  $x$  et  $y$ , la distance perceptuelle LPIPS est définie comme :

$$\text{LPIPS}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w_l \odot (\phi_l(x)_{h,w} - \phi_l(y)_{h,w})\|_2^2, \quad (4.2)$$

où  $\phi_l$  désigne les caractéristiques extraites à la couche  $l$  d'un réseau pré-entraîné, et  $w_l$  les poids associés. Des valeurs faibles indiquent une plus grande similarité perceptuelle.

**FID.** Le Fréchet Inception Distance compare les distributions des caractéristiques extraites d'images réelles et reconstruites, assimilées à des gaussiennes multivariées :

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}\right). \quad (4.3)$$

où  $(\mu_r, \Sigma_r)$  et  $(\mu_g, \Sigma_g)$  sont respectivement la moyenne et la covariance des deux distributions.

**GMSD.** La similarité locale des magnitudes de gradient  $m_x(i)$  et  $m_y(i)$  en un pixel  $i$  est donnée par :

$$\text{GMS}(x, y) = \frac{2 m_x m_y + c}{m_x^2 + m_y^2 + c}, \quad (4.4)$$

et l'indicateur global correspond à l'écart-type des valeurs  $\text{GMS}(i)$  sur toute l'image :

$$\text{GMSD}(x, y) = \text{std}(\text{GMS}(x, y)). \quad (4.5)$$

Des valeurs faibles traduisent une meilleure cohérence structurelle.

### b) Cadre d'évaluation clinique

Nous avons mené une étude de concordance diagnostique reposant sur une évaluation en simple aveugle, sans référence, afin de valider l'utilité clinique des reconstructions par super-résolution dans des scénarios d'interprétation réalistes. Le panel d'évaluation était constitué d'une experte médicale senior, garantissant une expertise approfondie et une vision clinique complète.

Les images ont été analysées en isolement clinique strict — sans accès aux méthodes de reconstruction ni aux images de référence — afin de reproduire les conditions réelles de pratique, où la valeur diagnostique repose uniquement sur la qualité de l'image.

---

4. L'évaluation clinique a été réalisée par une gynécologue-obstétricienne sénior (plus de 20 ans d'expérience), impliquée dans la formation spécialisée et régulièrement sollicitée par des organismes institutionnels et industriels.

L'instrument d'évaluation reposait exclusivement sur une échelle de Likert en 5 points (1 = très faible, 5 = excellent), permettant une mesure quantitative standardisée des performances évaluées.

Six critères ont été retenus, avec inversion de polarité pour les items à valence négative (Q2, Q5) afin d'assurer une interprétation unidirectionnelle (score plus élevé = meilleure performance) :

- **Q1. Résolution spatiale** : netteté des contours (1 = très floue, 5 = optimale) ;
- **Q2. Charge en artefacts** : bruit ou distorsions de reconstruction\* (1 = nulle, 5 = sévère) ;
- **Q3. Lisibilité anatomique** : visibilité des repères (1 = faible, 5 = excellente) ;
- **Q4. Confiance diagnostique** : adéquation clinique (1 = non-diagnostique, 5 = pleinement satisfaisant) ;
- **Q5. Risque d'interprétation** : probabilité d'erreur\* (1 = nul, 5 = élevé) ;
- **Q6. Détectabilité pathologique** : mise en évidence de lésions (1 = absente, 5 = indiscutable).

(\*Items inversés)

### c) Analyse statistique des performances des modèles

Dans la Partie 1, l'évaluation des modèles a été conduite pays par pays. Les distributions des métriques présentaient souvent une forte asymétrie et quelques valeurs extrêmes, rendant les moyennes peu représentatives. Nous avons donc privilégié l'utilisation de la médiane et l'identification des outliers pour mesurer la robustesse locale des modèles.

Dans la Partie 2, les données provenant d'Algérie, d'Égypte, du Ghana, du Malawi et de l'Ouganda ont été regroupées au sein d'un dataset unique, intitulé *AFRICA*. L'étude a par ailleurs été complétée par d'autres ensembles homogènes : Espagne, *ISTHMIC*, *PSFHS*, ainsi qu'un jeu de données artificiel (*FPUS23*). Cette organisation a permis une comparaison globale des modèles à l'aide de tests non paramétriques (Kruskal–Wallis suivis du test de Dunn en post-hoc), d'intervalles de confiance par bootstrap et de mesures de taille d'effet ( $\eta^2$ , Cliff's  $\delta$ ). Cette approche garantit une analyse statistique adaptée à la taille et à la composition des datasets, tout en demeurant rigoureuse et robuste.

Cette progression méthodologique illustre un raisonnement scientifique clair : passer d'une analyse locale à une analyse globale, en choisissant les outils statistiques appropriés à chaque contexte.

## 4.3.2 Évaluation des performances des modèles de super-résolution

Nous rapportons ici les résultats de l'étude comparative de huit modèles de super-résolution appliqués à cinq corpus échographiques couvrant des situations cliniques diverses. Dans un premier temps, nous analysons les performances quantitatives, à l'aide des indicateurs objectifs et perceptuels détaillés précédemment, pris séparément pour chaque corpus puis regroupés dans une analyse intégrée.

Nous poursuivons par une comparaison qualitative des reconstructions, mettant en évidence les différences d'interprétation visuelle entre les modèles.

### A. Performances quantitatives

Dans la littérature sur la super-résolution d'images, les métriques de similarité telles que PSNR et SSIM sont historiquement calculées après suppression d'une bordure de scale pixels

(soit 4 pixels pour un facteur  $\times 4$ ) [45, 89, 106]. Cette pratique, établie par les travaux fondateurs, permet de réduire les artefacts de bordure induits par les convolutions et le padding, assurant ainsi une comparaison équitable entre les images super-résolues (SR) et les images de référence haute résolution (HR) [2, 209].

Conformément aux protocoles standards [2, 193], nous rapportons le PSNR et le SSIM calculés sur le canal de luminance Y (espace YCbCr), après un recadrage de 4 pixels pour un facteur d'échelle  $\times 4$ , garantissant la comparabilité de nos résultats avec ceux publiés.

TABLEAU 4.14 – Performances moyennes de chaque modèle, agrégées sur l'ensemble des cinq jeux de données. Le meilleur résultat pour chaque métrique est indiqué en **gras**, et le deuxième meilleur est souligné.

| Famille | Modèle      | PSNR $\uparrow$ | SSIM $\uparrow$ | LPIPS $\downarrow$ | FID $\downarrow$ | NIQE $\downarrow$ | GMSD $\downarrow$ |
|---------|-------------|-----------------|-----------------|--------------------|------------------|-------------------|-------------------|
| GAN     | AESRGAN-1D  | 26.92           | 0.7812          | 0.2017             | 114.02           | 5.58              | 0.2430            |
|         | AESRGAN-MD  | 26.92           | 0.7717          | 0.2042             | 82.42            | <b>4.84</b>       | 0.2423            |
|         | BSRGAN      | 28.54           | 0.8044          | 0.1849             | 60.70            | 5.38              | 0.2291            |
|         | Real-ESRGAN | 28.89           | 0.8083          | 0.1775             | 77.28            | 5.64              | 0.2352            |
| Transf. | HAT         | <b>30.84</b>    | <b>0.8585</b>   | 0.1759             | <b>51.64</b>     | 7.28              | <b>0.1954</b>     |
|         | Swin2SR     | <u>30.26</u>    | <u>0.8436</u>   | 0.2053             | 80.64            | 8.20              | 0.2245            |
|         | SwinIR-M    | 27.27           | 0.7879          | <u>0.1656</u>      | 59.98            | 5.44              | <u>0.2212</u>     |
|         | SwinIR-L    | 28.89           | 0.8200          | <b>0.1514</b>      | <u>53.36</u>     | <u>5.28</u>       | 0.2245            |

Remarque : Les scores représentent la performance moyenne de chaque modèle sur les cinq jeux de données (AFRICA, FPUS23, PSFHS, ISTHMIC, SPAIN).

Le Tableau 4.14 résume les résultats globaux des huit configurations de modèles sur les cinq jeux de données considérés, offrant un aperçu de leur fidélité de reconstruction et de leurs performances perceptuelles. Les données détaillées sont présentées dans un tableau complet en Annexe 5.5.11 : Résultats complémentaires par dataset, identifié par le Tableau 5.12.

L'examen du Tableau 4.14 met en évidence une opposition claire entre les deux familles architecturales. Les modèles fondés sur les Transformers se révèlent nettement supérieurs aux approches par GANs lorsqu'il s'agit de préserver la structure et la fidélité de l'image. Parmi eux, HAT domine systématiquement, atteignant les meilleures valeurs pour le PSNR (30.84 dB) et le SSIM (0.8585), et affichant également le score GMSD le plus bas (0.1954), signe d'une distorsion structurelle minimale. SwinIR-M suit de près (0.2212), tandis que Swin2SR confirme la tendance en obtenant le deuxième meilleur niveau de fidélité pixel par pixel (PSNR = 30.26, SSIM = 0.8436).

L'analyse des métriques perceptuelles révèle que les écarts de performance entre modèles sont plus marqués que ceux observés avec les indices objectifs, tels que le PSNR ou le SSIM. LPIPS, FID et NIQE capturent des dimensions de la qualité visuelle — naturalité, texture et impression générale — qui échappent aux mesures classiques.

Parmi les Transformers, SwinIR-L atteint le LPIPS le plus faible (0.1514), suivi de HAT (0.1759), tandis que ce dernier se distingue par le FID le plus bas (51.64), SwinIR-L occupant la deuxième place (53.36). Du côté des GANs, Real-ESRGAN affiche un LPIPS compétitif, réduisant partiellement l'écart avec les Transformers.

L'analyse des GANs montre qu'AESRGAN-MD excelle essentiellement sur le NIQE (4.84), reflétant une naturalité perceptuelle supérieure, mais demeure en retrait sur les métriques de fidélité et perceptuelles par rapport à BSRGAN et Real-ESRGAN. BSRGAN se distingue par

sa fidélité structurelle (PSNR = 28.54, SSIM = 0.8044, GMSD = 0.2291), tandis que Real-ESRGAN présente les meilleures performances perceptuelles au sein des GANs, avec un FID intermédiaire (77.28). AESRGAN-1D affiche les scores les plus faibles sur la majorité des métriques.

Ces observations indiquent que chaque GAN possède des points forts spécifiques, l'avantage perceptuel d'AESRGAN-MD étant principalement limité aux métriques sans référence. Sur ces indices, les GANs demeurent compétitifs, notamment grâce à AESRGAN-MD, tandis que SwinIR-L (Transformer) se classe deuxième sur le NIQE (5.28), montrant que, dans le cadre expérimental considéré, un modèle basé sur l'attention peut également produire une naturalité perceptuelle élevée tout en conservant une bonne fidélité structurelle.

En synthèse, les Transformers — HAT en particulier — assurent la meilleure fidélité structurelle et distributionnelle à l'échelle globale, tandis que les GANs conservent un avantage perceptuel sur les métriques sans référence, illustré par AESRGAN-MD. SwinIR-L se situe à l'interface des deux familles, offrant un compromis équilibré entre fidélité structurelle et qualité perceptuelle, soulignant la complémentarité des approches et la nécessité d'une évaluation multi-métriques pour caractériser de manière exhaustive la performance des modèles.

## B. Analyse qualitative

Pour compléter les résultats quantitatifs et en faciliter l'interprétation, nous avons réalisé une analyse qualitative sur deux cas représentatifs issus de jeux de données contrastés : un exemple clinique de haute qualité provenant du jeu espagnol (Figure 4.10a) et un cas plus complexe sur le plan diagnostique extrait du jeu africain (Figure 4.10b).

Quatre modèles ont été retenus pour cette comparaison visuelle, afin d'illustrer les comportements distincts observés dans l'ensemble de l'étude :

- **HAT**, identifié comme le modèle le plus performant en termes de fidélité,
- **Swin2SR**, deuxième meilleur transformeur, illustrant la robustesse de cette famille,
- **BSRGAN**, représentant les GANs avec les meilleures performances de fidélité,
- **AESRGAN-1D**, choisi plutôt qu'AESRGAN-MD pour mettre en évidence l'instabilité propre à certains modèles génératifs.

En effet, AESRGAN-MD affiche des performances modérées mais globalement stables, par exemple PSNR = 24,93 / SSIM = 0,7498 sur le jeu *PSFHS* et PSNR = 25,92 / SSIM = 0,8248 sur le jeu *ISTHMIC*.

À l'inverse, AESRGAN-1D présente une variabilité importante selon les jeux de données, avec des scores allant de PSNR = 24,21 / SSIM = 0,7356 sur *PSFHS* à PSNR = 30,12 / SSIM = 0,7709 sur *FPUS23*, et PSNR = 27,11 / SSIM = 0,8005 sur le jeu espagnol. Cette irrégularité se reflète également dans les métriques perceptuelles, avec FID = 171,7 sur *FPUS23* et 130,3 sur le jeu africain. À titre de comparaison, AESRGAN-MD maintient des distorsions perceptuelles plus faibles (FID = 64,3 sur *FPUS23*, 33,6 sur *ISTHMIC*) et un NIQE moyen = 4,84, contre 5,58 pour AESRGAN-1D.

Cette stabilité relative rend AESRGAN-MD moins représentatif des limites des GANs, tandis qu'AESRGAN-1D, par ses fluctuations marquées, constitue un exemple plus pertinent pour illustrer leur manque de fiabilité. Ce contraste met en évidence la robustesse des modèles transformeurs, qui apparaissent globalement plus stables et mieux adaptés à une utilisation généralisée. Les reconstructions produites par ces quatre modèles sont présentées aux côtés de l'image basse résolution (LR) et de la vérité terrain (HR), pour les deux cas représentatifs. Cette comparaison permet à la fois de visualiser les performances maximales que les modèles peuvent atteindre dans des conditions optimales, et de juger leur capacité — ou leurs limites — à préserver une reconstruction fiable dans des contextes plus contraignants.

## ■ Analyse qualitative des performances dans un scénario clinique optimal :

Afin d'évaluer visuellement le potentiel maximal des modèles, la Figure – 4.10a– présente un cas représentatif de reconstruction du fémur issu du jeu de données espagnol. Cet exemple clinique de haute qualité permet d'examiner en détail les différences, parfois subtiles mais essentielles, dans les stratégies de reconstruction lorsque les conditions sont favorables.

### ■ *Fidélité structurelle :*

Les modèles fondés sur les Transformers –(HAT, Swin2SR)– parviennent à restituer les contours anatomiques avec une fidélité géométrique élevée, en accord avec leurs scores de PSNR et SSIM supérieurs. HAT se distingue particulièrement par une correspondance quasi isomorphe avec la vérité terrain, traduisant une excellente préservation globale des structures. Swin2SR conserve de solides performances, mais introduit un léger lissage local, atténuant discrètement certains détails de haute fréquence. À l'inverse, BSRGAN assure une continuité des contours satisfaisante, mais tend à exagérer les profils de bord, reflet d'une accentuation artificielle liée à son apprentissage adversarial. AESRGAN-1D, quant à lui, engendre des distorsions microstructurales modérées, traduisant une précision spatiale insuffisante et une dégradation des détails fins.

### ■ *Rendu textural et artefacts :*

HAT et Swin2SR présentent une tendance intrinsèque au sur-lissage, supprimant le grain naturel de l'image et produisant ainsi une texture artificiellement homogène. Ce phénomène reflète une perte d'information fréquentielle. À l'opposé, les modèles génératifs s'appuient sur des a priori appris pour synthétiser la texture. BSRGAN génère des motifs à haute fréquence qui renforcent l'acuité perceptuelle, mais manquent de réalisme biologique. AESRGAN-1D exprime une dérive plus marquée encore, produisant des textures hallucinées, sans correspondance anatomique réelle.

### ■ *Conséquences pour l'interprétation diagnostique :*

Ces divergences ont un impact direct sur l'utilité clinique. Les reconstructions « sûres » de HAT et Swin2SR, dépourvues d'artefacts génératifs, constituent une base structurelle fiable, malgré une perte partielle de détails texturaux. Les GANs, en revanche, soulèvent un compromis critique : si l'accentuation synthétique de BSRGAN peut sembler avantageuse, elle risque également de masquer certaines anomalies subtiles. Plus préoccupant encore, la propension d'AESRGAN-1D à introduire des motifs non physiologiques, même avec des données d'entrée de bonne qualité, en fait un candidat inadapté à des usages diagnostiques. Ces résultats confirment qu'un modèle ne peut être jugé sur la base de ses seules performances quantitatives.

## ■ Robustesse des modèles dans un scénario difficile :

La robustesse des modèles a été évaluée à partir d'un cas complexe de reconstruction d'un cerveau fœtal issu du jeu de données *AFRICA*, présenté en zones agrandies dans la Figure 4.10b. Cet exemple, vraisemblablement affecté par un rapport signal-bruit plus faible, illustre la résilience contrastée des différentes approches architecturales lorsqu'il s'agit de restituer des structures neurologiques complexes.

### ■ *Fidélité structurelle :*

Même dans ces conditions dégradées, les modèles Transformers les plus performants, HAT et Swin2SR, se distinguent par leur remarquable capacité à préserver la géométrie anatomique. La courbure lumineuse délimitant l'os est restituée avec une grande précision et sa continuité est parfaitement maintenue, en adéquation avec la vérité terrain. Ces résultats démontrent leur aptitude à identifier et conserver des repères anatomiques essentiels, même en présence

d'un bruit marqué. À l'inverse, les modèles fondés sur les GAN présentent des signes manifestes de dégradation structurelle. BSRGAN introduit de légères irrégularités ainsi qu'un déficit de lissage au niveau de la voûte crânienne. L'échec d'AESRGAN-1D est nettement plus sévère : la frontière osseuse apparaît fragmentée et déformée, traduisant une incapacité totale à maintenir une correspondance isométrique avec la vérité terrain.

■ **Réalisme textural et génération d'artéfacts :**

Le caractère difficile de ces données accentue les modes d'échec liés à la texture. HAT et Swin2SR adoptent leur stratégie « sécuritaire » de sur-lissage marqué : ils parviennent à supprimer efficacement le bruit, mais au prix d'une disparition du fin grain textural du parenchyme cérébral. Le résultat est une reconstruction propre, mais appauvrie en détails. Les modèles génératifs échouent de manière plus marquée. La tentative de BSRGAN de synthétiser une texture aboutit à un rendu artificiel, proche d'un effet « peint », au sein du tissu cérébral. L'échec le plus critique demeure celui d'AESRGAN-1D, qui engendre de sévères artéfacts génératifs : un motif fibreux prononcé recouvre l'image et efface totalement la texture cérébrale d'origine. Un tel motif « halluciné » pourrait être dangereusement interprété comme une anomalie structurelle ou une hémorragie, compromettant ainsi la fiabilité clinique de l'outil.

■ **Implications pour l'interprétation diagnostique :**

Ce cas particulièrement exigeant illustre avec force l'importance de la robustesse des modèles dans l'évaluation neurologique. Les résultats fournis par HAT et Swin2SR, bien que dépourvus de texture fine, demeurent fiables sur le plan diagnostique. Ils restituent une représentation structurellement précise, quoique lissée, du cerveau et de la boîte crânienne, limitant ainsi le risque d'introduire des informations trompeuses. À l'inverse, la sortie générée par AESRGAN-1D compromet totalement la fiabilité diagnostique : les artéfacts et distorsions produits induisent des altérations suffisamment graves pour exclure toute interprétation clinique rigoureuse, rendant les images impropres à une expertise médicale. Cette observation met en évidence une limite critique à l'application de tels modèles en conditions réelles, notamment dans des contextes marqués par une forte variabilité des données. Dans ces situations, les approches privilégiant la fidélité structurelle — à l'instar de HAT — s'avèrent nettement supérieures aux modèles génératifs, dont la vulnérabilité à des erreurs imprévisibles et cliniquement lourdes de conséquences se révèle particulièrement problématique, surtout lorsqu'il s'agit de structures aussi sensibles que le cerveau fœtal.

Notre analyse qualitative confirme visuellement les résultats quantitatifs et révèle deux approches opposées de la reconstruction. Les modèles fondés sur les Transformers privilégient la fidélité structurelle, ce qui garantit des images fiables mais souvent trop lissées et appauvries en détails texturaux. Les GAN, à l'inverse, cherchent à renforcer la netteté perceptuelle, mais au prix d'un risque important et parfois imprévisible de générer des artéfacts « hallucinés », qui ne correspondent pas à la réalité anatomique, surtout dans les cas complexes. Ces constats rappellent qu'au-delà des performances chiffrées, l'évaluation des modèles doit intégrer leur fiabilité et leur sécurité, deux critères essentiels à toute application clinique.

### 4.3.3 Méthodologie d'analyse statistique

■ **Stratégie analytique**

Pour garantir la fiabilité de nos résultats face à la diversité de nos sources, nous avons opté pour une approche analytique articulée en deux phases complémentaires :

□ **Analyse intra-cohorte : préserver la spécificité des données**

Chaque jeu de données a été analysé séparément (*AFRICA*, *SPAIN*, *ISTHMIC*, *PSFHS* et

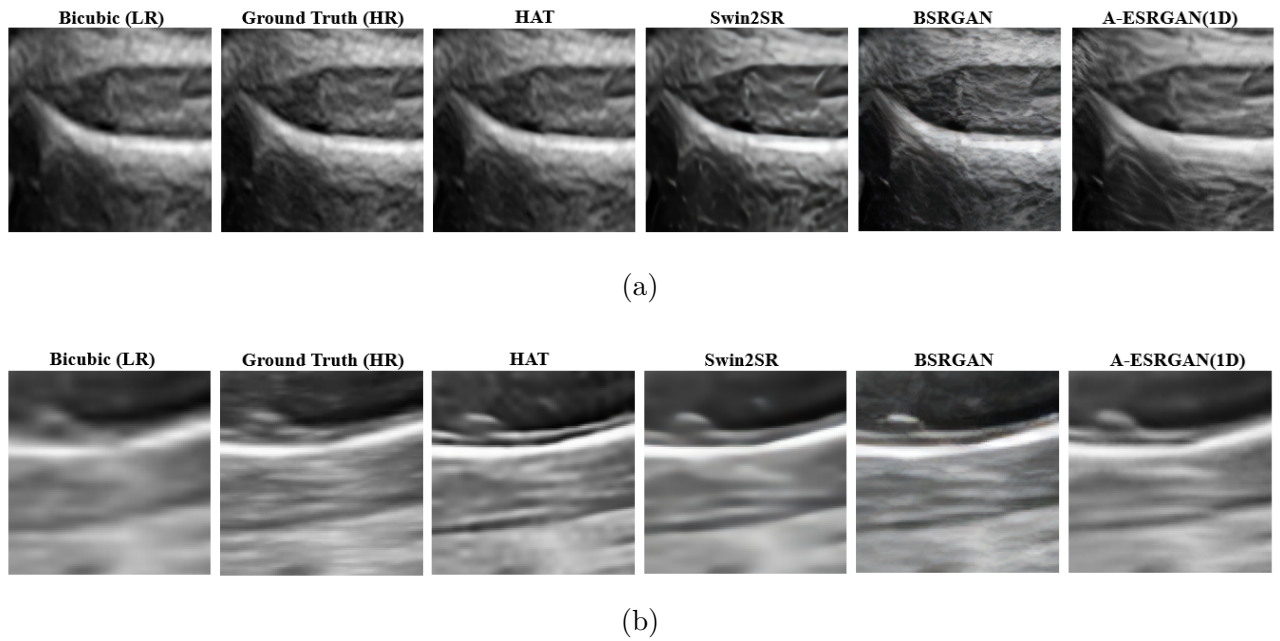


FIGURE 4.10 – Comparaisons agrandies des reconstructions anatomiques : (a) Région fémorale issue du dataset *SPAIN* et (b) Région cérébrale foetale issue du dataset *AFRICA*, avec Bicubic (LR), Vérité terrain (HR), HAT, Swin2SR, BSRGAN et AESRGAN-1D.

*FPUS23*). Ce choix s'explique par l'hétérogénéité des bases : elles diffèrent par leur taille, leurs protocoles d'acquisition et, dans une certaine mesure, par leur contenu anatomique. Une analyse combinée dès l'étape initiale aurait pu introduire des biais, atténuer certains effets ou en masquer d'autres.

#### □ Synthèse inter-cohorte : évaluer la reproductibilité

Les résultats ont ensuite été examinés de manière agrégée. Cette étape ne se substitue pas aux analyses individuelles, mais elle permet d'obtenir une vision globale et d'évaluer la robustesse des tendances observées d'un jeu de données à l'autre.

Cette approche en deux niveaux vise ainsi à concilier le respect des spécificités propres à chaque base et la recherche de tendances générales valables à l'échelle de l'ensemble des données.

## ■ Tests statistiques utilisés

Nous avons évalué les performances des modèles à l'aide de tests non paramétriques. D'abord, le Kruskal–Wallis a servi à détecter si un groupe se démarquait. Quand c'était le cas, nous avons poursuivi avec Dunn pour les comparaisons deux à deux. Pour éviter de tirer des conclusions erronées à cause du nombre élevé de comparaisons, les p-values ont été corrigées avec la FDR de Benjamini–Hochberg.

La significativité a été évaluée à l'aide des p-values, considérant  $p < 0.05$  comme preuve d'un effet. Comme les p-values seules donnent peu d'information sur l'ampleur de l'effet, nous avons également calculé des mesures de taille d'effet. Les effets globaux ont été résumés avec  $\eta^2$ , tandis que les comparaisons par paires ont été évaluées avec Cliff's  $\delta$ . Pour l'interprétation, nous avons suivi les seuils usuels :  $\eta^2 = 0,01, 0,06$  et  $0,14$  (effets faibles, moyens et forts) et Cliff's  $\delta$  classé en négligeable, faible, moyen ou fort.

Pour estimer l'incertitude des mesures, des intervalles de confiance bootstrap à 95 % ont été calculés par rééchantillonnage répété avec remise. Le chevauchement substantiel des intervalles a été considéré comme un indicateur descriptif d'absence d'effet marqué, sans se substituer aux tests statistiques formels.



Le tableau 4.15 résume les mesures statistiques utilisées.

TABLEAU 4.15 – Résumé des mesures statistiques

| Mesure         | Objectif                 | Interprétation                                               |
|----------------|--------------------------|--------------------------------------------------------------|
| Kruskal–Wallis | Comparaison globale      | Significatif $\rightarrow$ au moins un groupe diffère        |
| Dunn + FDR     | Comparaisons par paires  | Limite les faux positifs                                     |
| p-value        | Significativité          | $p < 0.05$ : significatif ; $p \geq 0.05$ : non significatif |
| $\eta^2$       | Taille d’effet globale   | 0,01 faible ; 0,06 moyen ; 0,14 fort                         |
| Cliff’s delta  | Taille d’effet par paire | Négligeable, faible, moyen, fort                             |

## ■ Résultats détaillés par jeu de données

**1. Tendances générales** Bien que tous les modèles figurent dans le tableau 4.16, nous nous concentrons sur les plus performants pour mieux visualiser les tendances. Sur le dataset espagnol, HAT arrive en tête, suivi de Swin2SR et BSRGAN. Dans le jeu *ISTHMIC*, SwinIR-L et HAT dominant. Des schémas similaires apparaissent dans *PSFHS* et *FPUS23*, avec HAT et SwinIR-L parmi les meilleurs. Sur le dataset africain, SwinIR-L et Real-ESRGAN se distinguent, HAT affichant des résultats plus modestes. Cette sélection facilite la lecture et l’interprétation des performances globales.

TABLEAU 4.16 – Scores globaux des modèles de super-résolution selon les jeux de données.

| Modèle      | <i>SPAIN</i>  | <i>ISTHMIC</i> | <i>PSFHS</i>  | <i>FPUS23</i> | <i>AFRICA</i> |
|-------------|---------------|----------------|---------------|---------------|---------------|
| HAT         | <b>0.8380</b> | <u>0.7688</u>  | <b>0.7413</b> | <b>0.8386</b> | 0.2512        |
| Swin2SR     | <u>0.5339</u> | 0.6198         | 0.3136        | 0.4864        | 0.5012        |
| BSRGAN      | 0.4496        | 0.6095         | 0.5908        | 0.4027        | 0.7193        |
| SwinIR-L    | 0.4405        | <b>0.7906</b>  | <u>0.6550</u> | <u>0.7766</u> | <b>0.7911</b> |
| SwinIR-M    | 0.3231        | 0.7548         | 0.6416        | 0.4505        | 0.5308        |
| Real-ESRGAN | 0.2985        | 0.7684         | 0.5169        | 0.3532        | <u>0.7409</u> |
| AESRGAN-1D  | 0.2110        | 0.1778         | 0.3148        | 0.1656        | 0.3760        |
| AESRGAN-MD  | 0.1819        | 0.3269         | 0.3990        | 0.1932        | 0.4039        |

**2. Résultats individuels par base de données** L’analyse suivante examine les performances des modèles de super-résolution sur chacun des cinq jeux de données, mettant en exergue les forces et limitations spécifiques à chaque contexte clinique.

- La base de données *AFRICA* :

Sur ce jeu de données, le modèle Swin2SR se distingue par la fidélité la plus élevée ( $PSNR = 29,36$  ;  $SSIM = 0,8295$ ). Cependant, concernant la qualité perceptuelle, ce sont SwinIR-L et SwinIR-M qui obtiennent les meilleurs scores avec des LPIPS respectifs de 0,1256 et 0,1250, ainsi qu’un GMSD commun de 0,2241. En termes de naturalité (NIQE), AESRGAN-MD et SwinIR-L se démarquent (4,4). Enfin, le score FID le plus bas est nettement attribué à BSRGAN (13,40), indiquant qu’il reproduit le mieux la distribution statistique globale de ce dataset, suivi par le modèle transformer SwinIR-L (59,48).

- La base de données *ISTHMIC* :

Le modèle Swin2SR conserve un avantage clair en fidélité ( $PSNR = 28,61$  ;  $SSIM = 0,9104$ ), suivi de près par SwinIR-L (0,8951) et HAT (0,8853) pour la reconstruction structurale. Les variantes du modèle A-ESRGAN se distinguent par la naturalité de l’image avec les

scores NIQE les plus bas (3,8 et 4,1). Le FID le plus faible est obtenu par SwinIR-L (20,80), confirmant que la distribution des images reconstruites par ce modèle reste la plus proche de la vérité terrain (*HR*). La métrique LPIPS la plus basse est partagée par deux transformers (HAT et SwinIR-M). Enfin, la variante "Large" de SwinIR possède le GMSD le plus bas (0,1806), suivie par Swin2SR (0,1840).

- Le corpus *PSFHS* :

Sur *PSFHS*, HAT affiche la meilleure fidélité globale ( $PSNR = 30,02$  ;  $SSIM = 0,8450$ ), tout en dominant la mesure de qualité structurelle perceptuelle ( $GMSD = 0,1935$ ). Les variantes du modèle A-ESRGAN atteignent les meilleurs scores en naturalité (6,5 pour la version 1D). Le FID le plus faible est toutefois observé pour SwinIR-M (87,48), montrant qu'il reproduit efficacement la distribution réelle malgré une fidélité brute légèrement inférieure à celle de HAT. La métrique perceptuelle LPIPS est la plus basse pour la variante SwinIR-M, talonnée par la variante L.

- La base de données *FPUS23* :

HAT s'impose comme le modèle de référence en dominant la fidélité ( $PSNR = 33,90$  ;  $SSIM = 0,8390$ ) et la qualité structurelle ( $GMSD = 0,2322$ ). La naturalité (NIQE) est mieux optimisée par SwinIR-L (4,2), suivi par AESRGAN-MD (4,5). Le FID le plus bas est également obtenu par HAT (38,86), confirmant sa capacité supérieure à capturer la statistique globale des images sur ce type d'acquisition. Toutefois, pour le LPIPS, SwinIR-L surpasse HAT, soulignant une meilleure corrélation avec la perception humaine sur ce dataset.

- La base de données *SPAIN* :

Sur le dataset espagnol, HAT présente des performances exceptionnelles et domine l'intégralité des métriques : fidélité (35,49 PSNR / 0,9650 SSIM), perception (0,0590 LPIPS / 0,1235 GMSD) et FID (2,35). BSRGAN et Swin2SR présentent des performances intermédiaires, tandis que les modèles AESRGAN et Real-ESRGAN se situent en retrait. Ces résultats confirment que l'architecture Transformer de HAT est particulièrement performante sur des images de haute résolution riches en détails anatomiques fins.

## Synthèse de la performance globale

L'analyse croisée des performances détaillées par jeu de données représentées dans le Tableau 5.12 (Annexe 5.5.11) et des résultats agrégés (Tableau 4.14) permet de dégager une tendance claire : HAT (Hybrid Attention Transformer) s'établit comme le modèle offrant le compromis le plus robuste sur l'ensemble du benchmark.

Ce constat s'appuie sur la convergence des différents indicateurs de performance mesurés à travers nos deux niveaux d'analyse :

### a. Une supériorité mathématique et structurelle confirmée

Le Tableau 4.14 montre que HAT occupe la première place sur les métriques de fidélité avec un PSNR moyen de 30,84 dB et un SSIM de 0,8585. Cette domination est corroborée par le Tableau 5.12 où HAT surclasse tous les modèles sur les bases de données à haute résolution comme *SPAIN* ( $PSNR = 35,49$ ) ou *FPUS23* ( $PSNR = 33,90$ ). Son architecture à attention hybride capture des corrélations de pixels à longue distance, là où les modèles GAN classiques peinent à maintenir une telle précision pixel-à-pixel.

### b. Préservation des gradients et réalisme clinique

Le score GMSD moyen de 0,1954 (le plus bas du Tableau 4.14) confirme que HAT est le modèle préservant le mieux les structures de gradients. Les données détaillées du Tableau 5.12 appuient ce constat, notamment sur le corpus *PSFHS*, où HAT domine avec un *GMSD* de

0, 1935.

La clarté des délimitations tissulaires et la précision de la résolution de contraste sont des facteurs déterminants pour la qualité de l'expertise médicale.

### c. Cohérence statistique et distributionnelle (FID)

Bien que le score FID repose sur un extracteur de caractéristiques (Inception-v3 [175]) entraîné sur des images naturelles — ce qui impose une certaine prudence quant à son interprétation en contexte médical — il demeure un indicateur utile pour évaluer la cohérence statistique des textures générées. Cette métrique évalue l'adéquation entre l'empreinte visuelle (gradients, orientations des structures, texture granulaire) des images générées et celle des acquisitions de référence.

Le score moyen de 51,64 (Tableau 4.14) positionne HAT comme le modèle dont les sorties se rapprochent le plus de la distribution des images de référence (*HR*). Cette tendance est confirmée de manière frappante sur le dataset *SPAIN* (Tableau 5.12), où HAT atteint un score de 2,35, suggérant une aptitude supérieure à capturer la complexité des textures anatomiques sans introduire d'artefacts distributionnels majeurs.

### d. Analyse des spécialisations et compromis

Bien que HAT soit le leader global, l'examen du Tableau 5.12 révèle des nuances importantes selon le contexte clinique :

- Fidélité sur tissus denses : Swin2SR demeure l'alternative la plus robuste pour la fidélité pure sur les tissus complexes, comme le prouvent ses records sur les datasets *AFRICA* ( $PSNR = 29,36$ ) et *ISTHMIC* ( $PSNR = 28,61$ ).

- Qualité perceptuelle : SwinIR-L s'illustre par une excellente corrélation avec la perception humaine, obtenant le meilleur score LPIPS moyen de 0,1514 dans le Tableau 4.14.

- Naturalité et absence d'artefacts : Les modèles A-ESRGAN (MD/1D) conservent un avantage en termes de "naturalité" (meilleur score NIQE agrégé de 4,84 pour la version MD), évitant certains artefacts de sur-traitement numérique propres aux architectures Transformers massives.

En conclusion, si le choix du modèle peut être affiné selon le dataset spécifique — à l'image de la performance isolée de BSRGAN sur le FID d'*AFRICA* (13,40 dans le Tableau 5.12) — HAT représente le compromis le plus solide. Il réussit l'équilibre complexe entre une fidélité pixel-à-pixel rigoureuse et une reconstruction structurale réaliste, s'établissant ainsi comme la nouvelle référence pour la super-résolution d'images médicales multi-domaines.

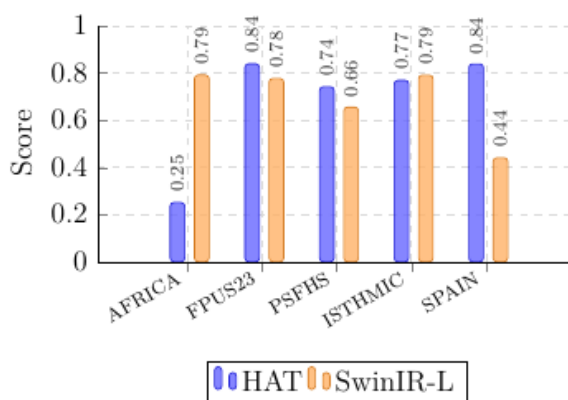


FIGURE 4.11 – Performances des modèles de super-résolution les mieux classés sur cinq jeux de données échographiques, mettant en évidence HAT et SwinIR-L avec leurs meilleurs scores respectifs.

L'analyse comparative sur divers jeux de données souligne la supériorité des modèles HAT et SwinIR-L par rapport aux approches GAN, avec des gains de performance statistiquement

significatifs ( $p < 0.05$ ) pour les métriques PSNR, SSIM et LPIPS. Alors que les GAN présentent une variabilité plus marquée, les modèles de type Transformer assurent une meilleure régularité inter-datasets, attestant de leur capacité de généralisation. Conformément à l’analyse statistique menée précédemment, ces différences de performance sont significatives et confirment les tendances illustrées à la Figure 4.11.

## ■ Analyse de robustesse multi-métriques

### 1. Méthodologie des scores composites

Afin de caractériser la performance globale des modèles, nous avons implémenté une approche multicritère s’appuyant sur six métriques de référence : PSNR, SSIM, LPIPS, NIQE, GMSD et FID.

La robustesse a été modélisée de manière non paramétrique par le calcul du taux de valeurs aberrantes (outliers). Celles-ci correspondent aux observations situées en dehors de l’intervalle :

$$[Q_1 - 1,5 \times \text{IQR}, Q_3 + 1,5 \times \text{IQR}] \quad (4.6)$$

où :

- $Q_1$  est le premier quartile (valeur en dessous de laquelle se trouvent 25% des observations),
- $Q_3$  est le troisième quartile (valeur en dessous de laquelle se trouvent 75% des observations),
- $\text{IQR} = Q_3 - Q_1$  est l’*intervalle interquartile*, représentant l’étendue centrale de 50% des données.

Cette méthode permet de quantifier la fiabilité du modèle indépendamment de la distribution intrinsèque des données, en pénalisant spécifiquement les comportements erratiques.

Pour assurer la comparabilité d’indicateurs aux unités hétérogènes, chaque mesure a subi une normalisation min-max. Les scores ont été réorientés de sorte qu’une valeur proche de 1 traduise systématiquement une performance optimale, facilitant ainsi l’agrégation. Enfin, nous introduisons le Score Équilibré, conçu comme la synthèse de trois dimensions fondamentales :

- Fidélité : précision structurelle et respect des détails (PSNR, SSIM) ;
- Perception : qualité visuelle corrélée au jugement humain (LPIPS, NIQE, GMSD, FID) ;
- Robustesse : stabilité statistique face aux variations et aux valeurs extrêmes, mesurée par le taux d’outliers définis selon l’intervalle interquartile (4.6).

Cette architecture hybride offre une lecture à la fois synthétique et granulaire : elle permet une hiérarchisation globale des modèles tout en préservant la transparence nécessaire à l’analyse des forces et faiblesses spécifiques à chaque métrique.

### 2. Résultats des métriques de performance

Les résultats des métriques de performance, présentés dans le tableau 4.18, pour huit modèles ”state-of-the-art”, montrent les différences de fidélité de reconstruction et de qualité perceptuelle entre les méthodes évaluées.

TABLEAU 4.17 – Métriques d’évaluation et score composite utilisés pour l’analyse comparative.

| Métrique/Score               | Objectif                                                      | Interprétation                                                                                         |
|------------------------------|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
| PSNR                         | Fidélité au niveau pixel                                      | Mesure la proximité avec les images de référence, sans garantir seule la restitution des détails fins. |
| SSIM                         | Préservation de la structure                                  | Sensible aux contours et aux contrastes, reflète la cohérence structurelle globale.                    |
| LPIPS                        | Distance perceptuelle en espace de caractéristiques profondes | Corrélée à la perception humaine pour des images naturelles.                                           |
| NIQE                         | Indice perceptuel sans référence                              | Évalue la régularité statistique et la qualité globale des images.                                     |
| GMSD                         | Distorsion de gradient de magnitude                           | Met en évidence les artefacts de contours et les distortions locales.                                  |
| <b>Outliers (robustesse)</b> | <b>Quantifie la stabilité et le taux d’erreurs</b>            | <b>Indicateur clé de la fiabilité et de la robustesse du modèle.</b>                                   |
| <b>Score Équilibré</b>       | Agrégation des métriques normalisées                          | Fournit une synthèse globale de la performance, sans pondération clinique spécifique.                  |

TABLEAU 4.18 – Classement des modèles par métrique de performance. Les valeurs reportées sont les moyennes brutes calculées sur l’ensemble des jeux de données. Les meilleurs scores pour chaque métrique sont en gras.

| Modèle      | PSNR (dB)    | SSIM          | LPIPS         | NIQE          | GMSD          |
|-------------|--------------|---------------|---------------|---------------|---------------|
| HAT         | <b>30.89</b> | 0.8606        | 0.1545        | 7.1608        | <b>0.1947</b> |
| Swin2SR     | 30.41        | <b>0.8639</b> | 0.2011        | 8.0475        | 0.2206        |
| BSRGAN      | 28.57        | 0.8061        | 0.1830        | 5.2731        | 0.2281        |
| SwinIR-L    | 28.50        | 0.8237        | <b>0.1539</b> | 5.3780        | 0.2159        |
| ReaL-ESRGAN | 28.13        | 0.8110        | 0.1733        | 5.6209        | 0.2342        |
| SwinIR-M    | 27.30        | 0.7912        | 0.1794        | 5.3475        | 0.2171        |
| AESRGAN-MD  | 26.97        | 0.7752        | 0.2018        | <b>4.9889</b> | 0.2415        |
| AESRGAN-1D  | 26.95        | 0.7839        | 0.1975        | 5.4646        | 0.2418        |

### 3. Analyse des compromis performance-robustesse

L’analyse des métriques normalisées présentées ci-dessous révèle des différences de performance et met en évidence les compromis fondamentaux entre fidélité de reconstruction et robustesse.

TABLEAU 4.19 – Métriques normalisées et Score Équilibré. Les métriques sont normalisées dans l’intervalle  $[0, 1]$  (1 = performance optimale). Les modèles sont classés par Score Équilibré.

| Modèle      | Métriques Normalisées |              |              |              |              |              | Score Équilibré |
|-------------|-----------------------|--------------|--------------|--------------|--------------|--------------|-----------------|
|             | PSNR                  | SSIM         | LPIPS        | NIQE         | GMSD         | Robustesse   |                 |
| <b>HAT</b>  | <b>1.000</b>          | 0.963        | 0.988        | 0.290        | <b>1.000</b> | 0.886        | 0.854           |
| Swin2SR     | 0.877                 | <b>1.000</b> | 0.015        | 0.000        | 0.450        | 0.114        | 0.409           |
| BSRGAN      | 0.411                 | 0.348        | 0.393        | 0.907        | 0.291        | <b>1.000</b> | 0.558           |
| SwinIR-L    | 0.393                 | 0.547        | <b>1.000</b> | 0.873        | 0.550        | 0.227        | 0.598           |
| ReaL-ESRGAN | 0.299                 | 0.404        | 0.595        | 0.793        | 0.161        | 0.500        | 0.459           |
| SwinIR-M    | 0.088                 | 0.180        | 0.468        | 0.883        | 0.524        | 0.545        | 0.448           |
| AESRGAN-MD  | 0.003                 | 0.000        | 0.000        | <b>1.000</b> | 0.006        | 0.727        | 0.290           |
| AESRGAN-1D  | 0.000                 | 0.098        | 0.090        | 0.844        | 0.000        | 0.000        | 0.172           |

### ★ Supériorité marquée du modèle HAT

Le modèle HAT démontre une performance globale exceptionnelle, se distinguant comme l'approche la plus complète et la plus fiable parmi les modèles évalués. Cette supériorité s'explique par ses performances élevées sur plusieurs indicateurs clés, comme il est illustré dans Tableau 4.19 :

- Fidélité pixel (PSNR = 1.0) : reconstruction précise par rapport à la référence.
- Intégrité structurelle (SSIM = 0.96) : préservation des contours et des relations structurelles.
- Robustesse (Outlier = 0.89) : faible taux d'échecs, garantissant un comportement stable et prévisible.

La convergence de ces trois qualités positionne HAT non pas comme un modèle spécialisé, mais comme *une solution généraliste de premier ordre*, adaptée à une large gamme de scénarios et de types de données.

### ★ Le compromis performance-robustesse : le cas de Swin2SR

Les résultats mettent en lumière le compromis critique entre la performance de pointe et la cohérence, parfaitement illustré par le modèle Swin2SR. Bien que ce modèle atteigne les performances de reconstruction théoriques les plus élevées (SSIM = 1.0, PSNR = 0.88), son faible score de robustesse (0.11), le plus bas de tous les modèles.

Cette divergence souligne un point crucial : un modèle capable de performances exceptionnelles sur une majorité de cas mais présentant des échecs catastrophiques imprévisibles est inadapté à un cadre clinique. Swin2SR incarne ainsi l'archétype du modèle de "recherche" dont la supériorité métrique moyenne masque une instabilité opérationnelle inacceptable pour un déploiement fiable.

### ★ La valeur de la robustesse : BSRGAN comme modèle de référence

Le modèle BSRGAN offre une contrepartie instructive au cas précédent. Bien que ses performances en fidélité (PSNR = 0.41, SSIM = 0.35) soient modestes comparées aux leaders, son parfait score de robustesse (1.0), indiquant le taux d'échec le plus faible.

Ce résultat valide l'hypothèse selon laquelle la fiabilité est une métrique clinique non-négociable. BSRGAN représente l'option "conservatrice" et sécuritaire, privilégiant la stabilité et la prévisibilité à la performance de pointe, un compromis souvent nécessaire en pratique médicale.

### ★ Performance des autres modèles et implications

Le peloton de milieu de classement (SwinIR-L, ReaL-ESRGAN, SwinIR-M) est constitué de modèles compétitifs mais ne parvenant pas à exceller simultanément sur tous les fronts comme HAT. Les modèles de la famille AESRGAN, affichant les scores de fidélité les plus bas, ne se sont pas avérés compétitifs dans cette configuration d'évaluation, suggérant que leur architecture attentionnelle n'apporte pas d'avantage dans ce contexte spécifique.

Cette analyse démontre que la sélection d'un modèle pour la super-résolution médicale ne peut se fonder sur une métrique unique ou sur des performances moyennes. Elle nécessite une évaluation holistique pondérant la fidélité anatomique et la robustesse opérationnelle. En ce sens, HAT émerge comme la solution de référence, conciliant performance de pointe et fiabilité indispensable. À l'inverse, le cas de Swin2SR sert d'avertissement contre l'adoption naïve de

modèles uniquement sur la base de leurs meilleures performances, sans considération pour leur stabilité. Par ailleurs, BSRGAN représente un benchmark de robustesse, offrant une alternative fiable si la priorité absolue est la minimisation du risque.

Cette méthodologie de scoring composite, transcendant les benchmarks conventionnels, fournit ainsi un cadre décisionnel robuste et orienté clinique pour la sélection et le déploiement de modèles de super-résolution.

#### 4.3.4 Validation clinique par évaluation experte en aveugle

Une experte médicale certifiée a conduit une évaluation en aveugle pour apprécier de manière quantitative la pertinence diagnostique des images reconstruites selon des échelles validées.

##### o Résultats de l'évaluation MOS

L'analyse précédente de la variabilité des performances a révélé une limitation cruciale des métriques automatisées. Pour évaluer la véritable utilité clinique des modèles, le MOS a été réalisé en aveugle.

Comme le montre le classement intégral (Figure 4.12), la hiérarchie issue de l'évaluation clinique diffère fortement de celle obtenue via les métriques automatisées.



FIGURE 4.12 – Classement clinique complet des huit modèles évalués par l'experte médicale. Le score final intègre les pénalités appliquées aux critères critiques de résolution spatiale et de confiance diagnostique.

*Classement clinique final :*

1. **SwinIR-L** Premier, malgré une performance technique non optimale.
2. **HAT** Premier en performance quantitative mais second dans le classement clinique.
3. **SwinIR-M** Ultime modèle atteignant un niveau clinique satisfaisant.
4. **Swin2SR** Début des architectures à utilité clinique restreinte.
5. **BSRGAN** Impacté négativement par des artefacts.
6. **Real-ESRGAN** Faible performance clinique parmi les modèles génératifs.
7. **AESRGAN-MD** Utilité clinique insuffisante.
8. **AESRGAN-1D** Dernier du classement.

*Analyse des écarts significatifs :*

- **Fossé clinique** : Un écart marqué sépare le top 3 (SwinIR-L, HAT, SwinIR-M) des autres modèles, établissant un seuil minimal d'acceptabilité clinique

- **Inversion des leaders** : SwinIR-L, seulement 4ème en score technique global, domine le classement clinique devant HAT, pourtant leader des métriques quantitatives
- **Rejet des GAN** : Les modèles génératifs (BSRGAN, Real-ESRGAN) sont systématiquement relégués malgré leurs bonnes performances sur certaines métriques techniques
- **Consistance des architectures** : Les modèles SwinIR se montrent plus fiables cliniquement que leurs alternatives

Ce classement établit un constat sans appel : l'excellence technique ne garantit pas la fiabilité diagnostique. L'inversion hiérarchique observée valide la nécessité absolue de la validation experte humaine en imagerie médicale, où la sécurité d'interprétation prime sur l'optimisation métrique.

## ○ Résultats et interprétations des mesures

Nous avons corrélé cinq métriques de qualité automatique au score moyen d'opinion (MOS) d'un expert afin de vérifier si ces indicateurs standard reflètent fidèlement la perception diagnostique du clinicien. Les corrélations obtenues (Figures 4.13) sont très inégales selon les métriques.

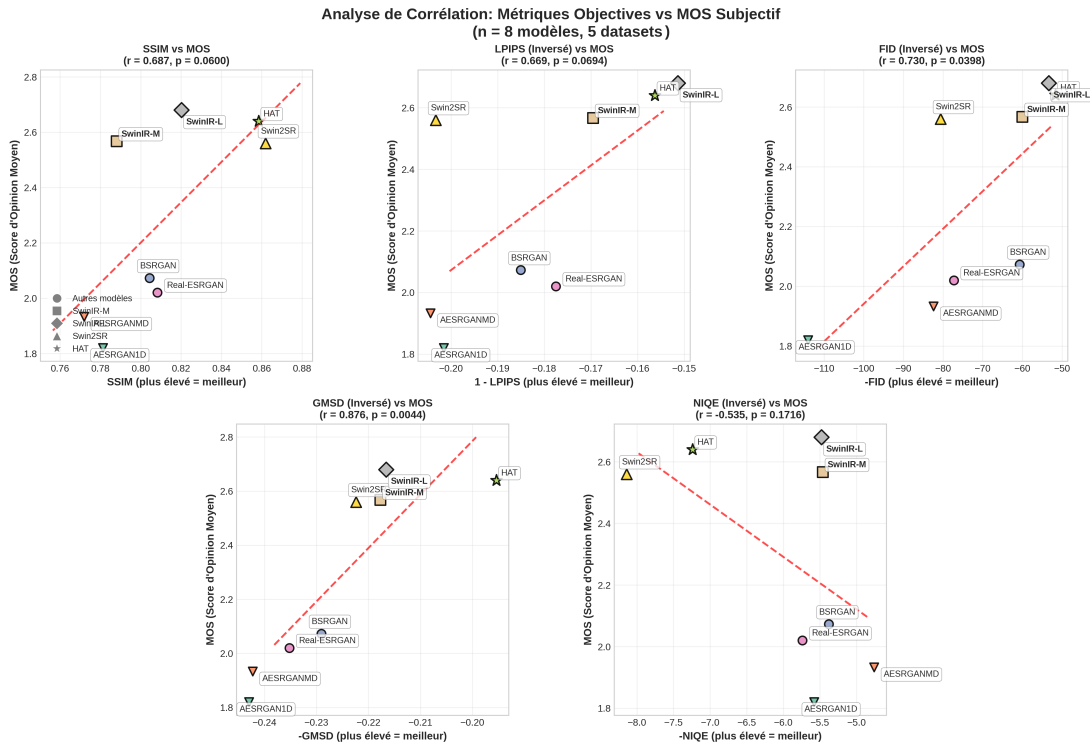


FIGURE 4.13 – Nuages de points illustrant la relation entre les métriques perceptuelles objectives et les scores moyens d'opinion (MOS). La rangée supérieure présente la GMSD, la FID et la SSIM, tandis que la rangée inférieure montre la LPIPS et la NIQE. Les lignes rouges en pointillées correspondent aux ajustements par régression linéaire, et le coefficient de corrélation de Pearson ( $r$ ) est indiqué pour chaque métrique.

La GMSD se révèle être l'indicateur le plus fiable pour notre application. Comme illustré dans la Figure 4.14, elle présente la corrélation la plus robuste avec l'évaluation clinique ( $r = 0,876, p = 0,004, R^2 = 0,767$ )<sup>5</sup>. Cette métrique explique à elle seule 76,7% de la variance des scores des experts, confirmant sa capacité à traduire fidèlement la perception des structures anatomiques.

5. Le coefficient de corrélation de Pearson ( $r$ ) et le coefficient de détermination ( $R^2$ ) sont définis en Annexe A



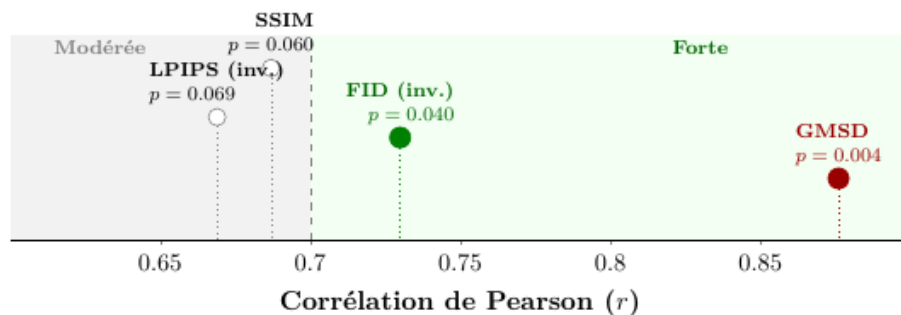


FIGURE 4.14 – Corrélation des métriques quantitatives et perceptuelles avec l'évaluation clinique (MOS). Le GMSD présente l'alignement le plus fort avec l'évaluation experte. Le NIQE est omis en raison de sa corrélation négative faible, qui comprimerait l'échelle et nuirait à la lisibilité.

Les autres métriques (FID, SSIM, LPIPS) montrent un lien plus ténu avec le jugement médical ( $r < 0,75$ ). Enfin, l'absence de corrélation de la **NIQE** ( $r = -0,535$ ) confirme que les mesures de "naturalité" issues de la photographie ne s'appliquent pas au signal échographique.

## 4.4 Conclusion

Ce chapitre a permis de mettre en perspective l'apport réel des modèles de super-résolution en échographie fœtale. En confrontant différentes architectures à travers notre double approche expérimentale, nous avons pu dégager des tendances claires sur leur capacité à traiter une imagerie aussi exigeante.

Les résultats de la première phase montrent sans ambiguïté que l'apprentissage externe — porté par BSRGAN, Real-ESRGAN et la famille SwinIR — l'emporte sur l'apprentissage interne du type DBPISR. Là où les approches classiques peinent à stabiliser l'image, ces modèles réussissent à reconstruire des textures fines sans trahir la cohérence anatomique. Au-delà de l'aspect visuel, la robustesse de cette amélioration est confirmée par nos tests : métriques de fidélité, retours d'experts et classification automatique concordent. Les analyses statistiques sont formelles : ces gains de qualité servent directement la fiabilité du diagnostic, que l'analyse soit humaine ou automatisée.

Pourtant, la seconde partie de nos tests apporte une nuance intéressante : la performance technique ne se superpose pas exactement à l'utilité clinique. Nous observons une légère inflexion entre les scores mathématiques et les préférences de l'experte médicale. Si le modèle HAT obtient les meilleurs résultats en métriques, l'évaluation clinique en aveugle place SwinIR-L très légèrement en tête. L'écart est minime (0,04 point), mais il suggère que l'œil du clinicien reste sensible à des subtilités de contraste ou de rendu tissulaire que les métriques de distorsion ne parviennent pas totalement à hiérarchiser. Le clinicien cherche avant tout une image qui respecte la sémantique et le contraste naturel des tissus, des aspects que les outils de mesure automatique ne reflètent encore que partiellement.

Pour éclairer cette divergence entre métriques et perception clinique, nous avons confronté au jugement de l'expert cinq métriques de natures complémentaires : GMSD et LPIPS (perceptuelles), SSIM (structurale), FID (distributionnelle) et NIQE (sans référence).

Les résultats montrent que seule la GMSD s'aligne étroitement sur la perception clinique, tandis que les autres mesures (FID, SSIM, LPIPS) n'en capturent que des aspects partiels. La NIQE, quant à elle, confirme son inadéquation aux spécificités du signal échographique. L'approche multimétrique offre toutefois une vision plus riche : elle confirme que la qualité perçue

est un concept composite, mêlant précision des contours et information texturale. Elle souligne également une limite fondamentale : une part de l'évaluation clinique — ancrée dans l'expérience et l'expertise anatomique du praticien — demeure irréductible aux mesures automatiques actuelles.

Ces observations nous ramènent aux réalités du terrain. En conditions réelles, l'efficacité d'un modèle se mesure à sa gestion de la variabilité des acquisitions et du bruit propre à chaque machine. Ici, les modèles pré-entraînés sur des données généralistes montrent leurs limites. Pour transformer ces preuves de concept en outils de routine clinique, une phase de spécialisation est inévitable.

C'est pourquoi la suite de nos travaux s'orientera vers un fine-tuning ciblé des architectures HAT et SwinIR-M. Le choix de la variante intermédiaire (M), au détriment de la configuration plus étendue (L), répond à une contrainte d'allocation mémoire : les capacités de notre processeur graphique (20 Go de VRAM) ne permettent pas l'entraînement de cette dernière. Toutefois, SwinIR-M offre une structure technique identique et constitue un compromis optimal. L'enjeu sera désormais de vérifier si cette spécialisation permet de réconcilier la puissance de calcul de HAT avec la pertinence visuelle de SwinIR, pour aboutir à un modèle final qui soit à la fois performant et réellement exploitable en milieu hospitalier.



# CHAPITRE 5



## Chapitre 5

# Fine-Tuning et Généralisation de Modèles de Super-Résolution pour l’Imagerie Échographique Fœtale

### 5.1 Introduction

La mise en œuvre clinique des méthodes de deep learning en imagerie médicale se heurte principalement à un problème de volume : les données annotées sont rares. L’acquisition de volumes annotés est limitée par des impératifs éthiques et un coût d’expertise prohibitif [87, 214].

Dans ce contexte, le transfer learning constitue le paradigme de référence : il permet d’exploiter les primitives visuelles apprises sur des domaines sources massifs pour compenser le déficit de données cibles [87]. L’efficacité repose sur l’arbitrage entre stabilité (conservation des poids pré-entraînés) et plasticité (adaptation aux statistiques médicales). En ajustant de manière sélective ou complète les paramètres d’un modèle pré-entraîné, il permet d’améliorer la performance et la fiabilité diagnostique. Différents axes de recherche structurent actuellement cette phase de fine-tuning :

□ Optimisation de la convergence et métriques cliniques : En segmentation, l’ajustement des paramètres permet de capturer des caractéristiques morphologiques spécifiques (tumeurs, organes abdominaux) là où un transfert figé échoue. Prasad et al. (2025)[149] rapportent des gains systématiques en Dice et IoU sur plusieurs modalités (IRM, CT, échographie), soulignant la supériorité du fine-tuning sur l’extraction de caractéristiques simples.

□ Réduction de la charge d’annotation : Des approches hybrides comme AIFT (Active Incremental Fine-Tuning)[214] visent à optimiser le coût d’entraînement. En sélectionnant itérativement les échantillons les plus informatifs, cette méthode permet d’atteindre des performances cibles avec un volume de données réduit de moitié.

□ Stabilité des poids et décalage de domaine : Le choix du protocole de dégel (*unfreezing*) est déterminant. L’étude comparative de Davila et al. (2024)[39] sur huit stratégies montre que des techniques comme le *Linear Probing* suivi d’un fine-tuning complet, ou l’usage de régularisations type L1/L2-SP, sont indispensables pour naviguer dans le *domain shift* sans dégrader les représentations latentes.

□ Contraintes à la Super-Résolution (SR) : Appliquée à l’imagerie médicale, la SR impose une conservation stricte de l’intégrité anatomique. L’adaptation de modèles tels que Real-ESRGAN, HAT ou SwinIR démontre des améliorations significatives (jusqu’à +3.28% en PSNR pour le CT [100]). Néanmoins, le risque d’oubli catastrophique impose des taux d’apprentissage faibles

et un dégel progressif des couches pour préserver les structures fines [133, 1].

S'appuyant sur ces travaux (résumés dans le Tableau 5.1), ce chapitre explore comment une stratégie d'ajustement intégral des poids permet de surmonter les verrous spécifiques à l'échographie : bruit de speckle, faible contraste et forte variabilité des acquisitions.

Les modèles SR de pointe SwinIR-M et HAT sont ainsi évalués. Ils sont initialement ajustés (*fine-tuned*) sur un ensemble agrégé de cinq *bases de données* variées (*FPUS23*[148], *ISTHMIC*[54], *HC*[186], *CHINESE*[36], *AFRICA*[161]). L'évaluation de la transférabilité inter-datasets est ensuite réalisée sur un sixième *corpus*[38] distinct (limité à **67 images**). L'utilisation combinée d'AFRICA en phase d'ajustement et de ce sixième *corpus* en évaluation permet d'examiner la performance du modèle en régime de données rares, notamment pour le test de généralisation dans des conditions extrêmement contraintes.

Les expérimentations détaillées ci-après trouvent leur ancrage dans une perspective de validation translationnelle, visant à confronter les capacités du deep learning aux exigences du diagnostic clinique.

Nous y présentons trois contributions majeures pour l'application de la Super-Résolution (SR) à l'imagerie échographique fœtale :

**a. Mise au point d'une stratégie de spécialisation par fine-tuning complet :**

Nous avons adapté les architectures HAT et SwinIR via un ré-entraînement intégral de leurs paramètres. Cette approche repose sur la constitution d'un corpus de référence haute résolution issu du dataset espagnol [21]. Une contribution clé de ce travail réside dans la curation manuelle et systématique de cette base : bien qu'acquise sur des équipements haut de gamme, chaque image a été rigoureusement sélectionnée pour garantir l'absence d'artefacts et la pureté des structures anatomiques nécessaires à l'apprentissage de la super-résolution.

**b. Validation de la généralisation sur un large spectre de données hétérogènes :**

Notre protocole porte sur cinq bases de données distinctes. En intégrant un jeu de données intentionnellement dégradé, nous validons la résilience des modèles face aux artefacts complexes. Nous établissons ainsi l'efficacité de cette adaptation, même sur des cohortes limitées, soulignant sa pertinence dans les environnements caractérisés par la rareté des données.

**c. Établissement d'une évaluation métrique axée sur l'impact clinique :**

Au-delà des métriques standards de qualité d'image, nous avons mené une analyse statistique approfondie intégrant des indicateurs spécifiques à l'échographie, tels que l'*Equivalent Number of Looks* (ENL)[137], le *Contrast-to-Noise Ratio* (CNR)[95, 188, 117] et l'*Edge Preservation Index* (EPI)[145, 158, 154]. Nous évaluons quantitativement l'apport technique aux gains diagnostiques et perceptuels tangibles.

Cette démarche permet de spécialiser les architectures de super-résolution aux exigences de l'examen fœtal, tout en garantissant leur fiabilité dans des contextes d'acquisition sous-optimaux. La section suivante (5.2) détaille les fondements théoriques et les stratégies de transfert qui sous-tendent ce protocole d'adaptation.

## 5.2 Apprentissage par Transfert

L'apprentissage par transfert ou *Transfer Learning* constitue une stratégie clé en apprentissage profond, consistant à tirer parti des représentations hiérarchiques déjà apprises par un modèle sur une tâche source pour résoudre une tâche cible, souvent limitée en données annotées. Cette méthode formalise l'idée d'induction biaisée : les caractéristiques visuelles universelles — telles que contours, textures et motifs — capturées par les premières couches du réseau sont

transférables, offrant ainsi une base solide pour l'apprentissage rapide et la généralisation sur de nouvelles tâches spécifiques.

Son mécanisme peut se décrire de la manière suivante :

Dans un réseau profond, chaque couche a un rôle particulier. Les couches proches de l'entrée identifient les primitives visuelles — contours, textures, gradients. Les couches intermédiaires les combinent pour repérer des motifs plus complexes, comme des fragments ou des assemblages d'objets. Les couches finales codent des représentations abstraites propres à la tâche pour laquelle le modèle a été entraîné.

Le *Transfer Learning* exploite cette hiérarchie : les couches basses et intermédiaires, qui contiennent des informations génériques, sont conservées, tandis que les couches finales sont remplacées pour apprendre la tâche cible. Cette approche permet d'adapter rapidement le modèle à un nouveau domaine, même lorsque les données annotées sont limitées.

### 5.2.1 L'Affinage (*Fine-Tuning*) : Principes Généraux

Le *Fine-Tuning* est la mise en œuvre pratique du transfert. Il consiste à ajuster un modèle pré-entraîné en modifiant de manière ciblée ses paramètres pour passer d'un domaine généraliste à un domaine spécialisé. Cette adaptation repose sur trois piliers :

- *L'initialisation informée* : Contrairement à un apprentissage partant de zéro (*from scratch*), le modèle bénéficie d'une configuration de paramètres déjà structurée. Le transfert apporte les bases de l'analyse visuelle (bords, textures élémentaires), ce qui place directement le réseau dans une zone de performance élevée.
- *La continuité de convergence* : L'affinage ne crée pas de capacités à partir de rien ; il prolonge l'apprentissage déjà effectué. Par des ajustements fins, il déplace le modèle vers un point d'équilibre propre à la nouvelle tâche sans supprimer les connaissances précédemment acquises.
- *Le recalibrage des filtres* : C'est l'étape déterminante pour l'imagerie médicale. Si le transfert permet d'organiser l'image, l'affinage adapte la sensibilité du réseau aux propriétés physiques du signal. Il permet de passer d'une reconnaissance de formes classiques (objets du quotidien) à l'analyse de motifs spécifiques et souvent plus complexes, caractéristiques de l'imagerie médicale.

En résumé, si le *Transfer Learning* apporte une base visuelle large, le *Fine-Tuning* assure l'ajustement nécessaire pour que le modèle devienne performant sur un terrain d'application médical spécifique.

#### A. Déclinaisons du Fine-Tuning dans la littérature

Bien que la littérature décrive plusieurs variantes — telles que l'affinage progressif ou discriminatif — deux stratégies principales sont majoritairement employées (Voir figure 5.1) :

- **Fine-Tuning Partiel** : Cette stratégie intermédiaire repose sur le *gel* (freezing) des couches initiales du réseau, qui extraient des caractéristiques élémentaires. Seules les couches supérieures sont réentraînées. Cela permet de conserver l'universalité des descripteurs de bas niveau tout en limitant le risque de sur-apprentissage sur de petits jeux de données. Par ailleurs, cette approche réduit considérablement la complexité computationnelle : en ne calculant les gradients que pour une fraction des paramètres, cela diminue drastiquement l'empreinte mémoire et le temps de calcul lors de l'optimisation.
- **Fine-Tuning Complet (*Full Fine-Tuning*)** : À l'inverse, l'intégralité des paramètres est soumise à un réentraînement.
  - **Mécanisme** : Le modèle est ajusté globalement avec un taux d'apprentissage très faible pour éviter l'oubli catastrophique (*catastrophic forgetting*).

- **Avantage** : Il permet une adaptation profonde. Si le fossé de domaine (*domain gap*) est important, le modèle peut réoptimiser ses filtres de bas niveau pour capturer des textures spécifiques.
- **Défis** : Coût de calcul plus élevé et nécessité d'une régularisation rigoureuse pour maîtriser l'overfitting.

Cette approche est préférable lorsque le jeu de données cible est volumineux et que le fossé de domaine *domain gap* entre la tâche source et la tâche cible est important. Elle maximise les performances au prix d'un plus grand coût de calcul et d'un risque de sur-ajustement qu'il faut maîtriser.

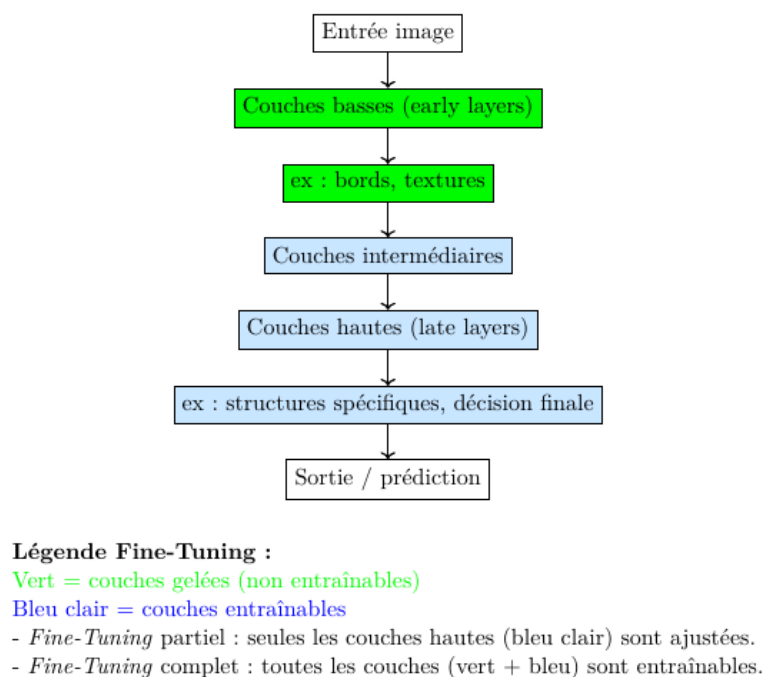


FIGURE 5.1 – Illustration hiérarchique des couches d'un réseau convolutif avec indication des couches gelées et entraînables pour le *Fine-Tuning*.

En complément des méthodes traditionnelles, les recherches récentes développent des approches adaptées aux défis du transfert d'apprentissage en contexte médical. L'étude de Davila et al. [39] explore ainsi des stratégies de régularisation spécifiques, comparant le *Linear Probing* à des méthodes telles que L1-SP/L2-SP ou Auto-RGN, qui visent à encadrer l'ajustement du réseau et l'écart des poids par rapport au modèle source.

L'ensemble de ces contributions, portant sur la classification, la segmentation et la super-résolution, est synthétisé dans le Tableau 5.1. Les résultats montrent que le succès du transfert ne dépend pas uniquement du choix des couches à dégeler, mais d'une synergie de facteurs techniques.

Ces travaux indiquent que la réussite du transfert repose sur la cohérence entre les conditions d'acquisition et les données d'entraînement. En super-résolution, Aghelan et al. [1] démontrent que l'adaptation doit privilégier la pertinence du pipeline de dégradation (*high-order degradation model*) plutôt que la simple modification de l'architecture. En simulant fidèlement les artefacts cliniques (bruit structuré, flou anisotrope), le modèle apprend à inverser des dégradations réalistes au lieu de transformations artificielles simplifiées. Parallèlement, les contributions de Nandal et al. [133] et de Kim et al. [87] soulignent l'importance de l'ingénierie des fonctions de perte et des mécanismes de stabilisation comme le *gradient clipping*.

En définitive, il n'existe pas de protocole universel : l'efficacité du fine-tuning dépend de l'alignement du modèle sur les caractéristiques statistiques et structurelles de la modalité cible.

TABLEAU 5.1 – Travaux représentatifs sur l'apprentissage par transfert et le fine-tuning en imagerie médicale (sélection).

| Référence                 | Type d'images                                   | Tâche            | Contribution principale                                                                                                                                                                                                  |
|---------------------------|-------------------------------------------------|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Zhou et al. (2017)[214]   | Rx (radiographies)                              | Classification   | AIFT : intégration de l'apprentissage actif et du fine-tuning incrémental — réduction d'environ 50% du coût d'annotation et amélioration des performances par sélection itérative d'échantillons.                        |
| Kim et al. (2022)[87]     | Multiples modalités (Rx, IRM, CT)               | Classification   | Revue systématique (121 études) : usage répandu de backbones profonds (Inception, ResNet); recommandation d'utiliser ces backbones comme extracteurs de caractéristiques sur petits jeux de données.                     |
| Davila et al. (2024)[39]  | X-ray, IRM, histologie, dermoscopie, endoscopie | Classification   | Comparatif de huit stratégies de fine-tuning (LP, FT, gradual unfreeze, L1/L2-SP, Auto-RGN). Conclusion : approches hybrides (p.ex. LP puis FT, Auto-RGN) surpassent souvent le FT pur.                                  |
| Nandal et al. (2024)[133] | IRM, échographie, radiographie                  | Super-résolution | Application de Real-ESRGAN : fine-tuning progressif avec gel partiel des couches initiales et faible taux d'apprentissage — meilleure préservation des structures fines et réduction du <i>catastrophic forgetting</i> . |
| Aghelan et al. (2024)[1]  | Rx, rétine                                      | Super-résolution | Fine-tuning de Real-ESRGAN — amélioration de la qualité perceptuelle et conservation des détails fins.                                                                                                                   |
| Prasad et al. (2025)[149] | IRM, CT, échographie                            | Segmentation     | Étude comparant <i>transfer learning vs fine-tuning</i> (U-Net, ResNet, VGG16) : le fine-tuning améliore systématiquement Dice, IoU, précision et rappel.                                                                |
| Lee et al. (2025)[100]    | CT                                              | Super-résolution | Fine-tuning de HAT, SwinIR, Real-ESRGAN, BSRGAN et DPSR — amélioration moyenne de +3.28% en PSNR et +0.6% en SSIM; validation qualitative par experts via MOS.                                                           |

*Remarque* : Rx = radiographie (rayons X); CT = tomodensitométrie; IRM = imagerie par résonance magnétique. LP (*Linear Probing*) = entraînement du classifieur final uniquement, backbone gelé. FT (*Full Fine-Tuning*) = réentraînement de l'ensemble des couches du réseau. L1/L2-SP (*Starting Point*) = fine-tuning avec régularisation contraignant les poids à rester proches des poids pré-entraînés (pénalité en norme L1 ou L2 sur la différence aux poids initiaux). Auto-RGN (*Automatic Relative Gradient Norm*) = ajustement adaptatif des taux d'apprentissage basé sur la norme relative des gradients.

Dans ce cadre, notre approche repose sur un fine-tuning complet appliqué à la super-résolution échographique. Cette modalité présente un défi majeur lié à la variabilité d'acquisition et au bruit de *speckle*, qui altère la définition des interfaces tissulaires. Le choix d'un ré-entraînement intégral garantit ici la robustesse du transfert face à des textures dont la dynamique diffère radicalement des standards de la vision par ordinateur conventionnelle.

## 5.3 Méthodologie Expérimentale

### 5.3.1 Architectures de Modèles Sélectionnées

À la suite de l'analyse présentée dans le Chapitre 4 Partie 2 (section 4.3), deux modèles de l'état de l'art en super-résolution se sont distingués par leurs performances remarquables, tant sur les métriques objectives (fidélité et perceptuelles) que sur l'évaluation subjective (MOS). Ces modèles ont été sélectionnés pour une phase de *fine-tuning*, afin de tirer parti de leurs capacités préalablement acquises tout en les adaptant aux caractéristiques spécifiques des images échographiques de haute qualité :

- **SwinIR-M [105]** : Une déclinaison intermédiaire du modèle SwinIR, reposant sur une architecture Transformer à fenêtres décalées, efficace pour capturer les dépendances globales et restaurer les structures fines.



- **HAT** [29] : Associe des mécanismes d’attention locale et globale à un module convolutionnel hybride, offrant une meilleure reconstruction des textures complexes propres aux échographies.

### 5.3.2 Stratégie adoptée dans cette étude

Dans ce travail, nous avons retenu une stratégie de *fine-tuning complet* appliquée aux deux modèles de référence **HAT** et **SwinIR**. Ce choix répond à la nécessité d’adapter en profondeur leurs représentations internes aux spécificités de l’échographie, dont les caractéristiques diffèrent fortement des images naturelles utilisées lors du pré-entraînement (ImageNet, DIV2K). En particulier, les échographies présentent des textures bruitées de type speckle, un contraste limité et une variabilité inter-appareils importante, qui exigent un réajustement global des filtres de bas niveau.

Afin de limiter le risque de *catastrophic forgetting*, l’optimisation a été conduite avec un *taux d’apprentissage faible* ( $1 \times 10^{-5}$ ), permettant une adaptation progressive. La fonction de perte associe *L1 Loss*, qui garantit la fidélité pixel à pixel, et la fonction *SWT Loss multi-échelle*[93], favorisant ainsi la préservation des structures anatomiques fines<sup>1</sup>

Pour accroître la robustesse sans compromettre l’intégrité diagnostique, seules des augmentations géométriques élémentaires (rotations, symétries horizontales et verticales) ont été appliquées, en excluant volontairement les transformations photométriques<sup>(2)</sup>.

L’ensemble des hyperparamètres employés pour les deux modèles est récapitulé dans le tableau 5.2. Cette configuration a été pensée pour concilier stabilité de l’apprentissage et adaptation au domaine échographique, conditions essentielles à une généralisation robuste.

TABLEAU 5.2 – Paramètres utilisés pour le fine-tuning des modèles HAT et SwinIR-M.

| Paramètre           | HAT                                                                 | SwinIR-M                                            |
|---------------------|---------------------------------------------------------------------|-----------------------------------------------------|
| Type d’images       | Échographie fœtale                                                  |                                                     |
| Tâche               | Super-résolution $\times 4$                                         |                                                     |
| Jeu d’entraînement  | Spain_train (4023 images)                                           |                                                     |
| Jeu de validation   | Spain_val (1000 images)                                             |                                                     |
| Nombre d’époques    | 100                                                                 |                                                     |
| Taille du lot       | 2                                                                   |                                                     |
| Taille des patches  | 256 ( <b>gt_size</b> )                                              | 128 ( <b>H_size</b> )                               |
| Optimiseur          | Adam ( $\beta_1 = 0.9$ , $\beta_2 = 0.999$ , $lr=5 \cdot 10^{-5}$ ) |                                                     |
| Planification du LR | MultiStepLR                                                         | CosineAnnealingWarmRestarts                         |
| Fonctions de perte  | L1 + SWTLoss multi-échelle[93]                                      | L1 + Perceptual (VGG19) + GAN (LSGAN)               |
| Augmentations       | Retournements H/V, rotations                                        | Dégradation BSRGAN+, retournements, mélange données |

Remarque : **gt\_size** = taille des patches HR pour HAT, **H\_size** = notation propre à SwinIR-M.

### 5.3.3 Jeux de données pour le Fine-Tuning

L’évaluation couvre cinq corpus distincts — résumés dans le tableau 5.3 — pour tester la robustesse et la généralisation.

1. L’**Annexe 5.5.11** présente la formulation mathématique et la justification théorique détaillée de ces fonctions de perte.

2. Les transformations photométriques modifient la luminosité, le contraste ou la couleur des images, ce qui peut altérer les structures diagnostiques fines ; elles ont donc été volontairement exclues.

TABLEAU 5.3 – Caractéristiques des jeux de données utilisés.

| Dataset       | Images | Résolution | Rôle et caractéristiques                        |
|---------------|--------|------------|-------------------------------------------------|
| Spain_train_2 | 4023   | 256×256    | Fine-tuning (haute qualité, expert-annoté)      |
| Spain_val     | 1000   | 256×256    | Validation pendant fine-tuning                  |
| AFRICA        | 450    | Variable   | Test post fine-tuning (qualité moyenne)         |
| ISTHMIC       | 5217   | Variable   | Test post fine-tuning (séquences dynamiques)    |
| FPUS23        | 5265   | Variable   | Test post fine-tuning (échographies simulées)   |
| CHINESE       | 1 528  | Variable   | Test post fine-tuning (échographies fœtales 2D) |
| HC            | 856    | Variable   | Test post fine-tuning (pathologies fœtales)     |

Le preprocessing inclut un redimensionnement à 256×256 pixels et une normalisation des valeurs d'intensité. Le **dataset espagnol** a été retenu comme support pour cette étape. Ce jeu de données présente une qualité particulièrement élevée, résultat d'acquisitions réalisées avec des échographes de dernière génération, sous la supervision directe d'un expert en imagerie médicale. Afin d'assurer une homogénéité et une fiabilité maximales, seules les meilleures acquisitions ont été retenues à l'issue d'un filtrage rigoureux. Ce dataset constitue ainsi une référence en termes de qualité, permettant d'ancrer le fine-tuning sur des images exemplaires.

En complément, un **dataset qualifié de moyen**, regroupant plusieurs sous-ensembles d'images issues de différents pays (**Algérie, Égypte, Ghana, Malawi et Ouganda**) [161], a été constitué sous l'appellation générique "**AFRICA**". Ce corpus présente des conditions d'acquisition hétérogènes et globalement dégradées : bruit marqué, résolution native faible et forte variabilité inter-appareils. Le contraste marqué entre le dataset espagnol, homogène et de haute qualité, et ce dataset africain, plus hétérogène et moins qualitatif, fournit un cadre expérimental pertinent pour évaluer simultanément l'**adaptation** et la **généralisation** des modèles fine-tunés.

La base de données **ISTHMIC** [54], issu de trames extraites de vidéos échographiques, représente une situation intermédiaire. Sa spécificité tient à l'hétérogénéité temporelle et spatiale des acquisitions, induisant une variabilité accrue dans la représentation des structures anatomiques. L'intégration de ce corpus permet de tester la stabilité des modèles fine-tunés face à des données séquentielles et à des variations rapides de qualité d'image.

Le corpus **FPUS23** [148], de génération artificielle, offre une modalité complémentaire. Bien qu'il ne soit pas issu de données cliniques directes, il reproduit avec un réalisme satisfaisant des caractéristiques majeures de l'échographie fœtale, notamment la texture bruitée du speckle et la variabilité du contraste. Ce corpus fournit ainsi un terrain d'évaluation original, permettant de questionner la transférabilité inter-domaine et de mesurer la résilience des modèles fine-tunés lorsqu'ils sont confrontés à un écart de domaine significatif.

Le dataset **CHINESE** [36] a été utilisé sous la forme d'un sous-ensemble de **1 123 images échographiques fœtales 2D**, correspondant à des vues sagittales entre 11 et 14 semaines d'aménorrhée. Ces images incluent des annotations détaillées sur neuf structures intracrâniennes clés telles que thalamus, mésencéphale, clarté nucale, os nasal, etc. Ce corpus constitue un support pertinent pour évaluer la **finesse anatomique** et la **plausibilité clinique** des reconstructions SR, en particulier dans des contextes de dépistage prénatal.

Enfin, la base de données **HC** [139], plus restreinte (856 images sélectionnées), a été utilisée dans une visée exploratoire. Bien que limitée en taille, elle présente un intérêt méthodologique : elle permet d'évaluer la **robustesse des modèles** lorsqu'ils sont confrontés à un corpus volontairement réduit, reflétant une situation fréquente en pratique clinique où la disponibilité des données reste contrainte.

Dans cette configuration, les six jeux de données — **espagnol** (qualité élevée), **africain**

(qualité moyenne, acquisitions hétérogènes), **ISTHMIC** (variabilité temporelle), **FPUS23** (artificiel), **CHINESE** (échographies fœtales 2D en vue sagittale, utilisées pour le dépistage prénatal) et **HC** (images cliniques hétérogènes issues de contextes hospitaliers) — définissent un cadre expérimental riche, permettant d'évaluer conjointement l'**adaptation**, la **robustesse** et la **généralisabilité** des modèles de super-résolution échographique.

### 5.3.4 Protocole de Fine-Tuning

Le processus d'adaptation s'appuie sur une spécialisation progressive du modèle, exploitant la robustesse du pré-entraînement pour converger vers les spécificités de l'échographie de haute qualité. Ce protocole d'entraînement, fondé sur un faible taux d'apprentissage, assure une convergence stable : il limite la dégradation des poids pré-entraînés tout en permettant une spécialisation efficace sur le nouveau corpus d'images. La phase d'évaluation mobilise un large spectre de métriques : PSNR, SSIM et GMSD pour mesurer la fidélité et la qualité structurelle des images, ainsi que LPIPS, FID et NIQE pour évaluer la similarité perceptuelle et la naturalité. L'intégration de ces différents indicateurs assure une convergence cohérente entre la précision pixel à pixel et le réalisme anatomique des images restaurées.

#### a. Modèles de départ

Les modèles retenus relèvent tous deux d'approches *transformeur*, mais témoignent de générations conceptuelles différentes. L'un est fondé sur l'architecture hiérarchique des *Swin Transformer*, largement adoptée comme référence de base en restauration d'images, tandis que l'autre incarne une évolution plus récente, intégrant des raffinements dédiés à la super-résolution. Le fine-tuning a été réalisé à partir de leurs poids pré-entraînés sur des bases d'images généralistes.

#### b. Hyperparamètres

Le fine-tuning a été conduit avec des choix conservateurs afin d'assurer la stabilité de l'entraînement et d'éviter le surapprentissage :

- un taux d'apprentissage réduit, combiné à une planification décroissante (scheduler) ;
- un nombre d'époques limité avec sélection du meilleur modèle sur la validation ;
- une taille de batch restreinte, adaptée aux contraintes de mémoire GPU ;
- une combinaison de fonctions de perte intégrant la fidélité pixel, la cohérence perceptuelle et une contrainte adversariale.

Les détails propres à chaque modèle sont présentés dans le tableau 5.2. L'ensemble des expérimentations a été réalisé sur la même configuration matérielle (GPU) que celui décrit dans le Chapitre 4, garantissant une cohérence dans les conditions matérielles d'entraînement. Ces choix méthodologiques stabilisent notre protocole de fine-tuning et ouvrent la voie à l'analyse des résultats de validation.

#### c. Métriques d'Évaluation du Signal Échographique

Si les indicateurs globaux du Chapitre 4 (PSNR, SSIM, FID, LPIPS, GMSD et NIQE) dressent un panorama de la fidélité structurelle et perceptuelle, ils s'avèrent insuffisants pour garantir une réelle lisibilité diagnostique. Afin de valider que le processus de super-résolution n'altère pas les propriétés physiques intrinsèques du signal, nous complétons notre étude par l'analyse du CNR, de l'ENL et de l'EPI.

Le rapport contraste-sur-bruit (CNR)[117, 95, 188] est ici fondamental : il quantifie la détectabilité des cibles en évaluant la capacité du modèle à détacher les interfaces anatomiques du

fond acoustique, dépassant la simple mesure d'intensité pour toucher à la clarté de l'information médicale.

$$CNR = \frac{|\mu_{\text{cible}} - \mu_{\text{fond}}|}{\sigma_{\text{fond}}} \quad (5.1)$$

où  $\mu_{\text{cible}}$  et  $\mu_{\text{fond}}$  représentent les intensités moyennes des régions d'intérêt, et  $\sigma_{\text{fond}}$  traduit la variance du bruit de fond. Un CNR élevé atteste que le modèle SR rehausse la saillance des tissus sans amplifier les fluctuations du *speckle*.

L'homogénéité du chatoiement est caractérisée par le nombre équivalent de vues ENL[137] :

$$ENL = \frac{\mu_{\text{ROI}}^2}{\sigma_{\text{ROI}}^2} \quad (5.2)$$

Calculé sur des zones parenchymateuses homogènes, l'ENL permet de vérifier que la réduction du bruit s'opère sans dénaturer la texture originale du signal ni générer de motifs artificiels.

Enfin, la préservation de la netteté des contours, essentielle à la fiabilité des mesures biométriques, est évaluée à l'aide de l'indice de préservation des contours (EPI)[145, 158, 154]. Cette métrique mesure la corrélation des gradients d'intensité entre l'image reconstruite  $I_{\text{SR}}$  et l'image de référence  $I_{\text{HR}}$ , calculée à partir d'un opérateur laplacien :

$$EPI = \frac{\sum_{i,j} \nabla I_{\text{SR}}(i,j) \cdot \nabla I_{\text{HR}}(i,j)}{\sqrt{\sum_{i,j} (\nabla I_{\text{SR}}(i,j))^2} \sqrt{\sum_{i,j} (\nabla I_{\text{HR}}(i,j))^2}} \quad (5.3)$$

La proximité des scores avec l'unité confirme que HAT et SwinIR préservent les hautes fréquences du signal.

#### d. Segmentation et extraction des ROI :

L'extraction des ROI repose sur un flux automatisé utilisant SAM-Med2D [32], un modèle transformer optimisé pour l'imagerie médicale. Sa robustesse, issue d'un entraînement sur 20 millions de masques [205], garantit une segmentation objective indispensable à la précision des indices CNR et ENL. Ce processus remplace le détourage manuel, éliminant ainsi toute subjectivité humaine.

La qualité du jeu de données est maintenue par un filtrage strict post-extraction. L'exclusion automatique intervient en cas de : (i) score SAM-Med2D < 0,1, (ii) ENL > 50, ou (iii) compromis  $ENL > 30 / CNR < 0,3$ .

Seuls les triplets d'images — comprenant la référence (HR) et les deux sorties de modèles (SR) — présentant un alignement géométrique rigoureux ont été conservés pour l'évaluation de l'EPI.

#### e. Méthodologie statistique

Un cadre statistique en deux étapes a été adopté pour garantir une comparaison impartiale et reproductible des modèles. Dans un premier temps, une analyse descriptive a été menée sur six métriques de qualité d'image à travers cinq jeux de données (allant de 450 à 5 265 images, pour un total de 25 822 images super-résolues). Dans un second temps, des tests inférentiels ont permis d'évaluer la significativité des performances : des tests  $t$  paramétriques pour les distributions normales, et les tests de Wilcoxon ou Mann-Whitney U pour les autres. La correction de Bonferroni a été appliquée pour les comparaisons multiples, complétée par le  $d$  de Cohen pour mesurer la taille de l'effet.

## 5.4 Évaluation des modèles fine-tunés : Résultats et Discussion

Dans cette partie, les résultats obtenus de l'étude sont détaillés et discutés. L'objectif est de démontrer la robustesse et la pertinence opérationnelle des modèles HAT et SwinIR après fine-tuning dans le contexte spécifique de l'échographie.

Notre analyse est structurée autour d'une approche bi-dimensionnelle, indispensable en imagerie médicale pour confronter l'amélioration technique à sa pertinence clinique :

*Axe I : Reconstruction Fidèle et Validation Diagnostique.*

Nous quantifions l'exactitude de la reconstruction à l'aide de métriques de référence (PSNR, SSIM, GMSD) qui évaluent la qualité structurelle et la distorsion des images. Le critère décisif pour la validation clinique reste le Rapport Contraste-sur-Bruit (CNR) : en échographie, il confirme que le gain de résolution se traduit par une amélioration tangible de la distinction des structures anatomiques et des lésions.

*Axe II : Maîtrise du Bruit et Intégrité Structurelle.*

L'attention est ici portée sur la gestion des défis inhérents à l'imagerie échographique, notamment la présence et l'atténuation du bruit de chatoiement (speckle). L'analyse s'appuie sur le Nombre Équivalent de Vues (ENL) pour l'efficacité du débruitage, et l'Indice de Préservation des Contours (EPI) pour garantir que les structures anatomiques cruciales ne sont pas dégradées.

L'analyse conjointe de ces indicateurs permet d'évaluer la capacité des modèles à maintenir un équilibre optimal entre le débruitage (ENL) et la préservation des détails (EPI), validant ainsi leur potentiel clinique.

### 5.4.1 Analyse du Processus de Fine-Tuning

Cette section évalue la capacité d'adaptation des architectures HAT et SwinIR aux spécificités de l'imagerie échographique, en analysant comment leurs mécanismes d'attention respectifs parviennent à modéliser les structures anatomiques fœtales malgré les contraintes de cette modalité.

#### • Suivi de la Convergence des Modèles lors du Fine-Tuning

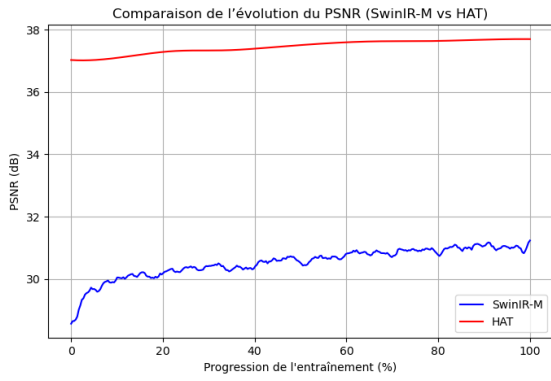
Les Figures 5.2a et 5.2b illustrent l'évolution des performances des deux architectures durant la phase de fine-tuning sur notre jeu de données échographiques. Comme attendu, le modèle **HAT**<sup>3</sup> affiche une progression régulière et synchronisée des métriques PSNR et SSIM. Cette convergence, qui démontre un équilibre entre la précision numérique (PSNR) et la cohérence structurelle (SSIM), est directement attribuable à la fonction de perte composite utilisée lors de son prétraitement initial, qui intégrait le SSIM.

À l'inverse, le modèle **SwinIR**<sup>4</sup>, dont les poids de départ ont été exclusivement optimisés sur le seul critère du PSNR, présente logiquement une amélioration continue se concentrant sur cette seule métrique.

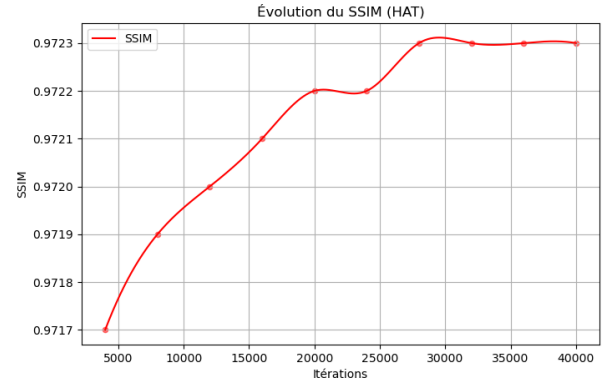
Ce différentiel de comportement initial impose une analyse critique : les performances observées durant le fine-tuning sont influencées par les fonctions de perte fondamentales de chaque modèle. Par conséquent, une évaluation complète de leur supériorité respective nécessitera l'examen des métriques indépendantes de ces biais (CNR, LPIPS, ENL, EPI), comme détaillé dans la suite de cette section.

3. <https://github.com/XPixelGroup/HAT>

4. <https://github.com/JingyunLiang/SwinIR>



(a) PSNR HAT vs SwinIR-M



(b) SSIM HAT

FIGURE 5.2 – Évolution des performances pendant le fine-tuning : comparaison du PSNR des modèles HAT et SwinIR-M (a) et évolution du SSIM pour le modèle HAT (b).

Cette remarque n'enlève rien à l'intérêt pratique de HAT, mais elle rappelle qu'une analyse rigoureuse doit toujours mettre en perspective les conditions qui orientent les résultats observés. Ainsi, il serait prématuré de conclure à la supériorité définitive d'un modèle sur la seule base des courbes de fine-tuning. L'évaluation des deux architectures fine-tunées sur les cinq corpus de test, complétée par une analyse statistique, constituera une étape essentielle pour confirmer ou nuancer ces premières observations.

### • Limites architecturales de SwinIR et solutions de HAT

L'analyse par Local Attribution Maps (LAM) [29] met en évidence une différence fondamentale dans la gestion des dépendances spatiales entre les deux architectures. Alors que SwinIR reste limité par des fenêtres de calcul non chevauchantes (shifted windows) qui restreignent le contexte exploité pour la reconstruction des détails, HAT s'en affranchit grâce à ses mécanismes de fenêtres chevauchantes (overlapping cross-window attention). Cette capacité à mobiliser un champ de dépendance plus étendu permet à HAT de capturer une information contextuelle plus riche, garantissant une meilleure préservation des textures et une continuité structurale supérieure dans les zones complexes.

Dans le cadre de l'imagerie échographique fœtale, cette caractéristique est déterminante pour l'interprétation des interfaces tissulaires fines, telles que les parois ventriculaires cérébrales ou les contours des organes abdominaux. En effet, le signal ultrasonore est intrinsèquement corrompu par le bruit de speckle, qui peut masquer des détails locaux ou créer de faux contours. Un modèle exploitant un contexte spatial élargi, comme HAT, est capable de "relier" des informations distantes pour reconstruire une frontière anatomique cohérente là où SwinIR, limité par sa vision locale, risquerait d'interpréter le bruit comme une discontinuité réelle.

Les limitations observées pour SwinIR incluent :

- Champ de dépendance limité, réduisant la capture de détails fins ;
- Artefacts de type bloc présents dans les représentations intermédiaires ;
- Intégration globale restreinte, limitant la modélisation des textures complexes et des contours flous.

Pour surmonter ces limites (spécifiques à l'échographie), HAT introduit trois innovations majeures :

- Fenêtres de Self-Attention Élargies ( $16 \times 16$  vs.  $8 \times 8$  dans SwinIR) : Cette augmentation élargit significativement le champ récepteur dès les premières couches, permettant de capturer et de modéliser des structures anatomiques étendues (telles que les organes ou la silhouette fœtale).

- Attention hybride (self-attention + channel attention) :
  - la self-attention spatiale modélise les relations locales fines (contours, petits vaisseaux) pour une reconstruction précise,
  - La channel attention pondère sélectivement les features, permettant au modèle de distinguer le bruit de speckle des informations corrélées et pertinentes pour l'anatomie.
- Overlapping Cross-Attention Block (OCAB) : Ce mécanisme renforce les interactions entre les fenêtres voisines (en dépit de la grande taille de fenêtre), éliminant efficacement les artefacts de blocs et assurant une reconstruction fluide et cohérente des larges surfaces tissulaires.

Ces différences architecturales constituent notre hypothèse centrale d'analyse : les métriques de qualité spécifiques au domaine, telles que l'EPI (Indice de Préservation des Contours), le CNR (Rapport Contraste-sur-Bruit), et l'ENL (Nombre Équivalent de Vues), devront quantifier la supériorité de la gestion contextuelle de HAT par rapport à SwinIR. Ces mécanismes, conçus pour améliorer la restitution des textures complexes issues d'images naturelles, s'avèrent pertinents dans un cadre clinique.

L'étude de Lee et al. [100] portant sur l'imagerie par scanner CT (Tomodensitométrie) a confirmé que les innovations de HAT permettent, après fine-tuning, de mieux préserver les détails anatomiques et d'atténuer le bruit, surpassant systématiquement les performances de modèles tels que SwinIR, BSRGAN ou Real-ESRGAN. Les auteurs concluent à la supériorité de l'attention hybride pour la restitution des détails fins et préconisent d'étendre cette analyse à d'autres modalités médicales afin de confirmer la polyvalence de l'architecture.

Le présent travail s'inscrit précisément dans cette perspective : en transposant l'évaluation de HAT et SwinIR à l'échographie fœtale, nous ambitionnons de démontrer que la supériorité de HAT est d'autant plus marquée dans un environnement acoustique complexe. Cette modalité, par la nature cohérente de son signal et la prédominance du bruit de speckle, constitue un test de robustesse bien plus exigeant que l'imagerie CT pour ces architectures à base d'attention.

## 5.4.2 Évaluation Quantitative Globale et Perceptuelle

### A. Performance comparative des modèles fine-tunés

Nous évaluons la performance selon deux métriques complémentaires, directement observées lors du fine-tuning :

- **PSNR** (Peak Signal-to-Noise Ratio) : Mesure objective de la fidélité pixelique au cours de l'adaptation.
- **SSIM** (Structural Similarity) : Indicateur de la préservation des structures et de la cohérence perceptuelle.

Le tableau 5.4 met en lumière l'effet du fine-tuning sur les modèles SwinIR-M et HAT, évalué à travers les métriques de fidélité PSNR et SSIM.

Pour **SwinIR-M**, le fine-tuning entraîne un gain notable de PSNR (+4,79 dB), témoignant de l'efficacité du modèle lorsqu'il est adapté aux images échographiques de haute qualité et de sa sensibilité à l'optimisation spécifiquement orientée PSNR, comme indiqué dans le dépôt GitHub<sup>5</sup>. Le SSIM du modèle de base est de 0,8078, mais son évolution après fine-tuning n'a pas pu être mesurée, ce qui limite l'interprétation de sa cohérence perceptuelle et de sa capacité à préserver les structures anatomiques.

Pour **HAT**, déjà performant en base, le fine-tuning produit une amélioration plus modeste

5. <https://github.com/JingyunLiang/SwinIR>

en PSNR (+2,20 dB) mais également une augmentation du SSIM (+0,007). Cette progression traduit un ajustement plus équilibré, renforçant à la fois la fidélité pixel et la préservation des structures, et suggère une consolidation qualitative de la reconstruction.

| Modèle               | PSNR (dB)    | $\Delta$ PSNR | SSIM         | $\Delta$ SSIM |
|----------------------|--------------|---------------|--------------|---------------|
| SwinIR-M (base)      | 26.77        |               | 0.8078       |               |
| SwinIR-M (finetuned) | <b>31.56</b> | <b>+4.79</b>  | –            | /             |
| HAT (base)           | 35.49        |               | 0.9650       |               |
| HAT (finetuned)      | <b>37.69</b> | <b>+2.20</b>  | <b>0.972</b> | <b>+0.007</b> |

TABLEAU 5.4 – Résultats comparatifs avant/après fine-tuning sur Spain\_val. Les colonnes ( $\Delta$ ) indiquent la différence entre le modèle fine-tuné et le modèle de base pour chaque métrique.

## B. Performances absolues et relatives après fine-tuning

Nos résultats expérimentaux révèlent une dynamique contrastée entre HAT et SwinIR-M.

HAT part d'un niveau déjà élevé avant fine-tuning (PSNR = 35.49 dB, SSIM = 0.965). Après adaptation au domaine échographique, il atteint 37.69 dB (+2.20 dB, soit +6.2%) et SSIM = 0.972 (+0.007, soit +0.7%). Ces résultats mettent en évidence à la fois la solidité de son apprentissage initial et sa capacité à tirer parti du fine-tuning pour améliorer significativement la fidélité numérique tout en consolidant la cohérence perceptuelle.

SwinIR-M part d'une base nettement plus faible (PSNR = 26.77 dB), mais le fine-tuning lui permet une progression substantielle, atteignant 31.56 dB (+4.79 dB, soit +17.9%). Ce résultat montre que le modèle bénéficie fortement de l'adaptation au domaine échographique, bien que son niveau final demeure inférieur à celui de HAT. Ce dernier conserve une avance absolue significative : +6.13 dB en **PSNR** et  $\approx +0.03$  en **SSIM**. Cela traduit une meilleure adéquation intrinsèque de l'architecture de HAT à la tâche de super-résolution d'images échographiques.

Cette supériorité de HAT s'aligne avec les conclusions de Lee et al. [100]. Bien que leur étude porte sur une modalité différente (CT-scan) et un volume de données massif ( $\approx 190\,000$  images), le fait que HAT y obtienne également les meilleures performances absolues renforce la crédibilité de nos propres observations. Cette convergence de résultats suggère que les mécanismes d'attention hybride de HAT possèdent une capacité de généralisation aux structures anatomiques qui dépasse le cadre d'une seule modalité. Ainsi, la robustesse de cette architecture, déjà validée sur le signal radiologique (CT), se transpose avec succès aux défis plus exigeants du signal ultrasonore.

## C. Axe I : Reconstruction Fidèle et Validation Diagnostique

L'Axe I vise à quantifier l'exactitude de la reconstruction par les métriques (PSNR, SSIM, GMSD) et à valider son impact diagnostique. En échographie, le Rapport Contraste-sur-Bruit (CNR) est le critère décisif, confirmant que le gain de résolution se traduit par une amélioration tangible de la distinction des structures et des lésions.

### • Fidélité Numérique et Métriques Générales

L'examen conjoint du Tableau 5.5, de la Figure 5.3 et des zooms visuels (Figure 5.4) met en évidence une différence nette et cohérente entre les modèles HAT et SwinIR-M après fine-tuning.



Les moyennes rapportées dans le tableau sont représentatives des centres des distributions observées, confirmant la cohérence entre les deux supports.

| Dataset | Modèle (fine-tuné) | PSNR (dB) ↑  | SSIM ↑        | LPIPS ↓       | FID ↓        | GMSD ↓        | NIQE ↓     |
|---------|--------------------|--------------|---------------|---------------|--------------|---------------|------------|
| AFRICA  | HAT                | <b>36.33</b> | <b>0.9182</b> | <b>0.1223</b> | <b>21.99</b> | <b>0.1748</b> | <b>7.3</b> |
|         | SwinIR-M           | 34.06        | 0.8878        | 0.2108        | 82.91        | 0.2099        | 8.4        |
| ISTHMIC | HAT                | <b>30.05</b> | <b>0.9288</b> | <b>0.1156</b> | <b>30.28</b> | <b>0.1529</b> | <b>7.0</b> |
|         | SwinIR-M           | 27.60        | 0.8842        | 0.2644        | 106.19       | 0.2105        | 8.0        |
| FPUS23  | HAT                | <b>33.97</b> | <b>0.8406</b> | <b>0.1670</b> | <b>36.91</b> | <b>0.2273</b> | <b>7.0</b> |
|         | SwinIR-M           | 32.13        | 0.7941        | 0.2563        | 125.75       | 0.3050        | 8.2        |
| CHINESE | HAT                | <b>35.87</b> | <b>0.9656</b> | <b>0.0372</b> | <b>14.61</b> | <b>0.1209</b> | 8.4        |
|         | SwinIR-M           | 32.66        | 0.9389        | 0.1369        | 49.05        | 0.1570        | 9.1        |
| HC      | HAT                | <b>43.30</b> | <b>0.9792</b> | <b>0.0345</b> | <b>4.85</b>  | <b>0.1028</b> | <b>7.9</b> |
|         | SwinIR-M           | 35.95        | 0.9565        | 0.0943        | 34.38        | 0.1491        | 8.6        |

TABLEAU 5.5 – Performances comparatives des modèles fine-tunés (HAT et SwinIR-M) sur l’ensemble des jeux de données externes (**AFRICA**, **ISTHMIC**, **FPUS23**, **CHINESE** et **HC**). Les meilleures valeurs sont indiquées en gras.

Sur l’ensemble des jeux de données, HAT surpasse systématiquement SwinIR-M sur les métriques de fidélité pixel-wise, avec des gains en PSNR allant de +1.84 dB sur FPUS23 à +7.35 dB sur HC, et des valeurs de SSIM constamment plus élevées. Cette supériorité est confirmée par les métriques perceptuelles, où HAT présente des LPIPS, GMSD et FID nettement plus faibles, SwinIR-M affichant des valeurs de FID deux à quatre fois supérieures selon les corpus.

La différence de comportement est particulièrement marquée sur le jeu de données AFRICA, le plus exigeant car constitué d’images cliniques réelles et fortement bruitées. HAT y démontre une robustesse fonctionnelle claire (PSNR 36.33 dB, SSIM 0.9182) et des améliorations perceptuelles substantielles (−67.8% en FID et −45.7% en LPIPS), tandis que SwinIR-M présente un comportement paradoxal, déjà observé sur FPUS23 : des gains numériques en PSNR et SSIM accompagnés d’une dégradation sévère des métriques perceptuelles, révélatrice d’un phénomène de sur-lissage.

Ce sur-lissage se traduit par une perte de structures anatomiques fines au profit d’une moyenne floue, qui conserve une fidélité pixel-wise élevée mais compromet la cohérence anatomique et la pertinence clinique. Ce phénomène est visuellement confirmé par les zooms de la région fémorale du jeu de données AFRICA (Figure 5.4), où les reconstructions de SwinIR-M apparaissent plus lisses et moins structurées que celles de HAT, qui préservent des contours plus nets.

Enfin, le NIQE présente une variabilité inter-datasets plus élevée, avec des valeurs comprises entre 7.0 et 8.4 pour HAT et entre 8.0 et 9.1 pour SwinIR-M. Bien que conçu pour les images naturelles et donc moins adapté à l’imagerie échographique, il confirme néanmoins, dans une analyse comparative, l’avantage global de HAT.

## • Validation Diagnostique par le CNR

Le critère clinique décisif, le CNR, confirme l’avantage de HAT sur la validation diagnostique. Sur les cinq cohortes, HAT obtient un CNR supérieur ou comparable à SwinIR-M (Tableau 5.6) sur quatre des cinq jeux de données. Cela confirme que le gain de résolution se traduit par une amélioration tangible de la distinction des tissus à faible contraste.

SwinIR-M n’est supérieur que sur la cohorte ISTHMIC (CNR 1.812 contre 1.701), mais cet avantage est pondéré par un EPI significativement plus faible (EPI 0.505 contre 0.752), comme détaillé dans l’Axe II.



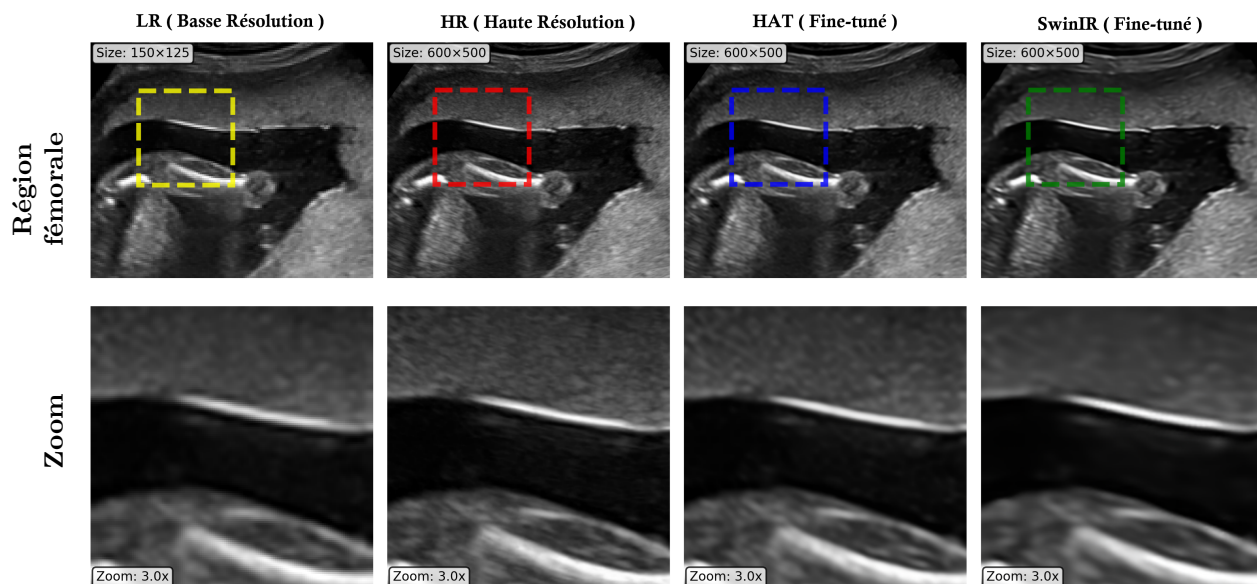


FIGURE 5.4 – Zoom d’une région fémorale sur une image échographique foetale (base de données AFRICA).

### E. Justification Architecturale de l’Intégrité Structurale (EPI)

La supériorité de HAT en matière de robustesse et d’intégrité structurale (EPI élevé) repose sur l’adéquation de ses innovations architecturales aux défis de l’échographie. Ces composants résolvent directement les compromis identifiés dans l’Axe I :

- *Attention par Canal (Channel Attention)* : (Optimisation ENL). Il permet un débruitage sélectif en apprenant à pondérer les canaux les plus informatifs et à supprimer les canaux riches en bruit. Cette sélectivité contribue à l’optimisation de l’ENL sans introduire d’artefacts de lissage (contrairement à SwinIR-M).

- *Fenêtres Élargies (Overlapping Windows) et OCAB* : (Renforcement EPI). Ce mécanisme clé renforce la cohérence contextuelle en élargissant le champ de vision (receptive field) et en assurant la continuité spatiale des structures. C’est le facteur directement responsable du maintien de l’intégrité des frontières anatomiques, ce qui explique l’EPI systématiquement supérieur de HAT.

Les échographies présentent des caractéristiques particulières qui compliquent la reconstruction d’images haute résolution : le bruit speckle granulaire, le faible contraste et les structures anatomiques irrégulières. L’architecture HAT est bien adaptée à ces défis grâce à ses mécanismes d’attention spécifiques. Le Channel Attention permet un débruitage sélectif en pondérant les canaux les plus informatifs, les fenêtres élargies combinées à l’OCAB renforcent les contours et améliorent le contraste, et la Self-Attention spatiale capture les relations à longue portée pour reconstruire fidèlement des structures anatomiques complexes. La synthèse de la manière dont les éléments architecturaux de HAT répondent aux exigences cliniques est détaillée dans le Tableau 5.7 qui suit.

L’analyse comparative, structurée autour des Axes I et II, révèle la robustesse structurale et la supériorité fonctionnelle de HAT. Il concilie la fidélité numérique et la cohérence anatomique dans des conditions d’acquisition hétérogènes. SwinIR-M, dont les gains en PSNR masquent une dégradation des métriques perceptuelles et de l’intégrité structurale, est en l’état inadapté à une application clinique fiable. HAT se positionne ainsi comme un candidat solide et généralisable pour la Super-Résolution en échographie foetale.

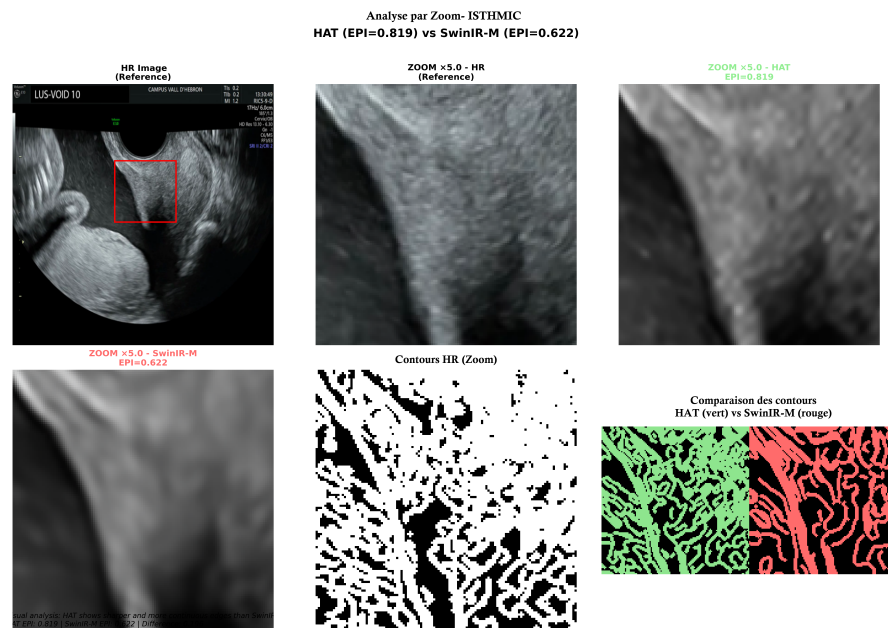
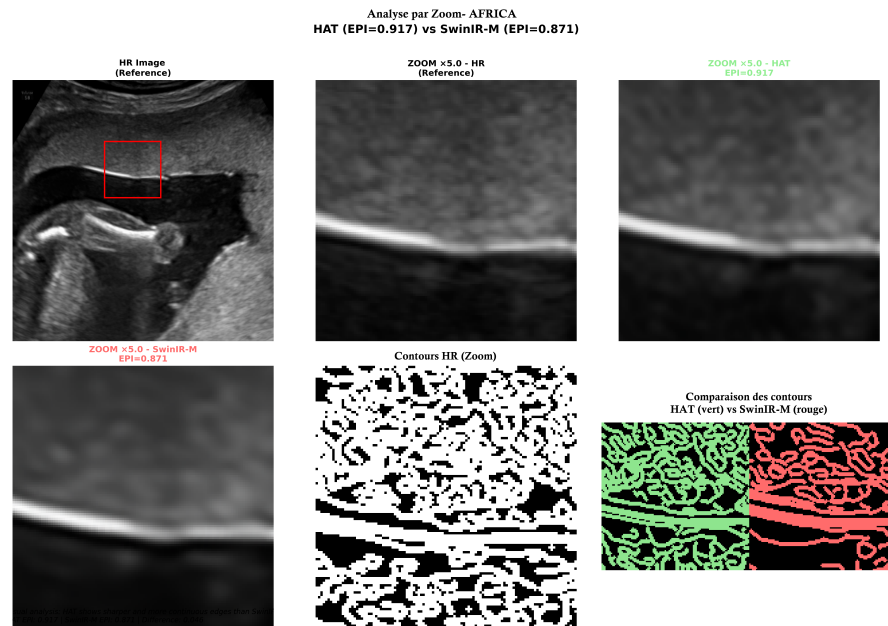
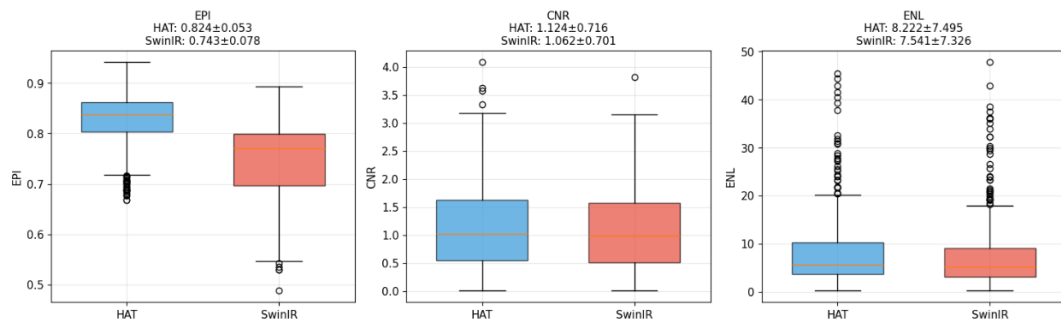


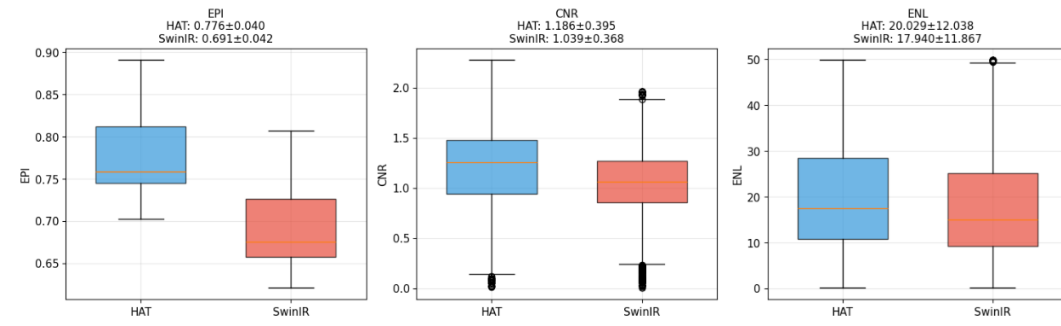
FIGURE 5.5 – Comparaison de l'Indice de Préservation des Contours (EPI) sur les Images AFRICA et ISTHMIC. Cette figure illustre le comportement des modèles fine-tunés sur des structures anatomiques clés.

**Panneau Supérieur (Jeu de données AFRICA, Fémur fœtal) :** SWINIR (EPI = 0.871) introduit un lissage qui érode la netteté du contour osseux, tandis que HAT (EPI = 0.917) maintient une frontière précise.

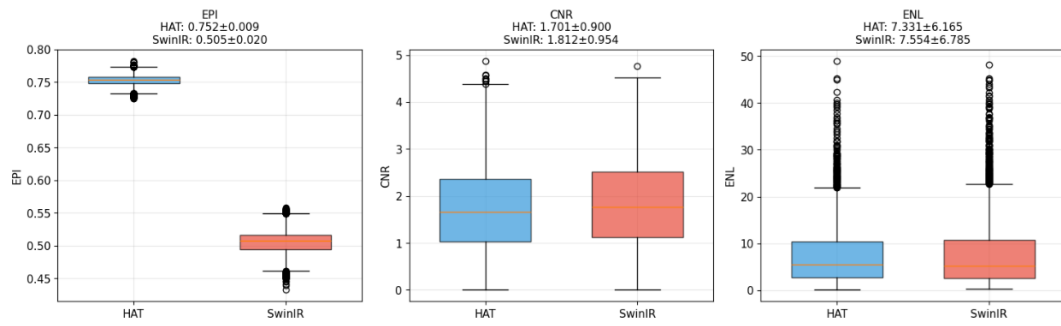
**Panneau Inférieur (Jeu de données ISTHMIC, Tissus mous) :** L'érosion des contours par SWINIR est plus prononcée (EPI = 0.622), masquant les détails structurels par un flou excessif, alors que HAT préserve mieux l'intégrité des tissus (EPI = 0.819). Les cartes de contours comparatives (panneaux de droite) confirment cette observation : les lignes de contour générées par HAT (en vert) sont plus continues et plus denses que celles de SWINIR (en rouge), prouvant que l'architecture HAT offre une fidélité structurelle et une fiabilité clinique supérieures.



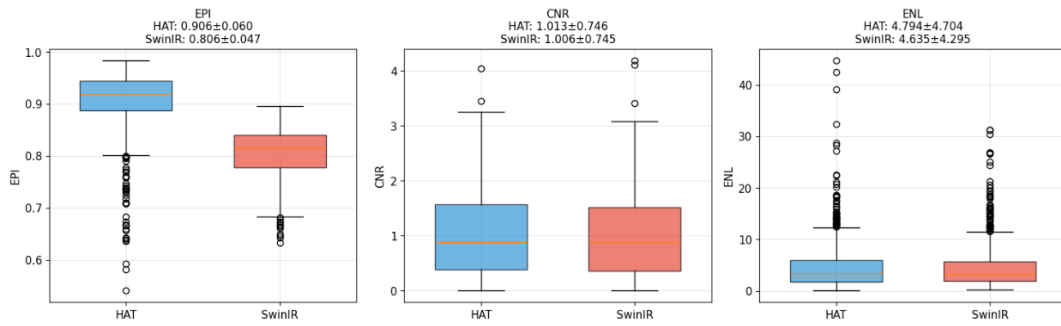
(a) AFRICA



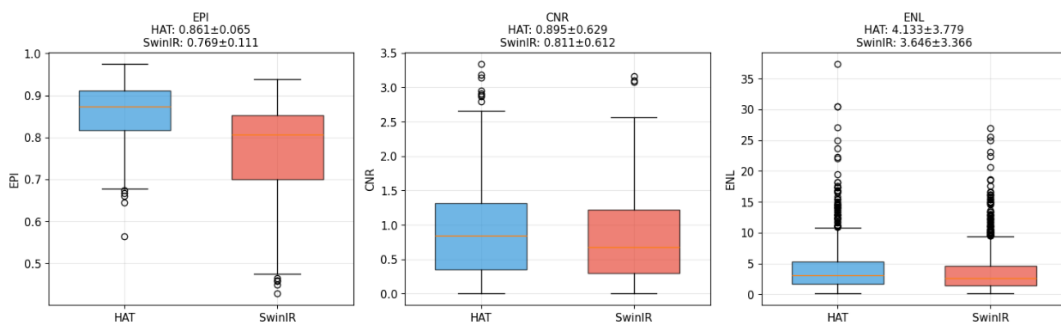
(b) FPUS23



(c) ISTHMIC



(d) HC



(e) CHINESE

FIGURE 5.6 – Distribution globale des métriques de qualité sur l'ensemble des jeux de données évalués.

TABLEAU 5.6 – Métriques comparatives des modèles SR sur différents jeux de données.

| Jeu de données | Paires ( $n$ ) | Modèle | EPI ( $\pm$ SD)   | CNR ( $\pm$ SD)   | ENL ( $\pm$ SD)     |
|----------------|----------------|--------|-------------------|-------------------|---------------------|
| AFRICA         | 442            | HAT    | $0.824 \pm 0.053$ | $1.124 \pm 0.716$ | $8.222 \pm 7.495$   |
|                |                | SwinIR | $0.743 \pm 0.078$ | $1.062 \pm 0.701$ | $7.541 \pm 7.326$   |
| FPUS23         | 3,674          | HAT    | $0.776 \pm 0.040$ | $1.186 \pm 0.395$ | $20.029 \pm 12.038$ |
|                |                | SwinIR | $0.691 \pm 0.042$ | $1.039 \pm 0.368$ | $17.940 \pm 11.867$ |
| ISTHMIC        | 5,215          | HAT    | $0.752 \pm 0.009$ | $1.701 \pm 0.900$ | $7.331 \pm 6.165$   |
|                |                | SwinIR | $0.505 \pm 0.020$ | $1.812 \pm 0.954$ | $7.554 \pm 6.785$   |
| HC             | 856            | HAT    | $0.906 \pm 0.060$ | $1.013 \pm 0.746$ | $4.794 \pm 4.704$   |
|                |                | SwinIR | $0.806 \pm 0.047$ | $1.006 \pm 0.745$ | $4.635 \pm 4.295$   |
| CHINESE        | 1,123          | HAT    | $0.861 \pm 0.065$ | $0.895 \pm 0.629$ | $4.133 \pm 3.779$   |
|                |                | SwinIR | $0.769 \pm 0.111$ | $0.811 \pm 0.612$ | $3.646 \pm 3.366$   |

TABLEAU 5.7 – Correspondances entre défis échographiques et composants de HAT.

| Défi échographique                  | Composante HAT correspondante | Effet observé après fine-tuning                                        |
|-------------------------------------|-------------------------------|------------------------------------------------------------------------|
| Bruit speckle granulaire            | Channel Attention             | Débruitage sélectif, +0.7% SSIM, préservation des textures biologiques |
| Contours flous / faible contraste   | Fenêtres élargies + OCAB      | Renforcement des frontières anatomiques, +2.20 dB PSNR                 |
| Structures anatomiques irrégulières | Self-Attention spatiale       | Reconstruction anatomiquement plausible de structures complexes        |

### 5.4.3 Analyse des Performances et Validation Statistique

#### 1. Cadre méthodologique et choix statistiques

Cette section présente le cadre statistique adopté pour l'évaluation comparative des modèles de super-résolution. Les choix méthodologiques ont été guidés par la volonté d'assurer une comparaison rigoureuse, reproductible et adaptée aux propriétés empiriques des données.

Les analyses ont été conduites image par image, sur cinq jeux de données distincts, en tenant compte à la fois de la distribution des scores et du risque lié aux comparaisons multiples.

#### a). Tests statistiques employés

Selon la normalité des distributions (évaluée préalablement), les tests suivants ont été appliqués :

*Tests paramétriques* : Tests t appariés ou indépendants lorsque les hypothèses de normalité étaient satisfaites. *Tests non paramétriques* : Tests de Wilcoxon (signed-rank) ou de Mann-Whitney U lorsque les hypothèses paramétriques n'étaient pas respectées. *Contrôle des comparaisons multiples* : Une correction de Bonferroni a été systématiquement appliquée afin de limiter le risque d'erreurs de type I. Les fondements théoriques ainsi qu'un exemple illustratif sont présentés en Annexe 5.5.15.

#### b). Nombre de comparaisons

Un total de 25 comparaisons statistiques a été réalisé, correspondant à :

5 jeux de données  $\times$  5 métriques

Chaque comparaison oppose directement les performances des modèles HAT et SWINIR-M.

## 2. Analyse comparative des performances

### a). Analyse descriptive : tendances globales et robustesse inter-datasets

Une analyse descriptive multi-métriques a été menée afin de caractériser les performances après fine-tuning sur cinq jeux de données hétérogènes : AFRICA, ISTHMIC, FPUS23, CHINESE et HC. Ces ensembles couvrent des contextes variés, tant en termes de qualité d'acquisition que de volume d'images (de 450 images pour AFRICA à plus de 5 265 pour FPUS23), pour un total de 25 822 images, soit 12 911 images sources, chacune traitée par les deux modèles.

Les modèles ont été évalués à l'aide de six métriques complémentaires couvrant différents aspects de la qualité d'image :

- **Métrique objective pixel-wise** : PSNR.
- **Métriques perceptuelles basées sur la structure** : SSIM et GMSD.
- **Métrique perceptuelle apprise** : LPIPS.
- **Métrique générative (distributionnelle)** : FID.
- **Métrique sans référence** : NIQE.

La Figure 5.3 présente les distributions complètes des scores sous forme de boxplots.

Les résultats mettent en évidence une tendance systématique en faveur de HAT, indépendamment du jeu de données considéré. Les médianes associées à HAT sont systématiquement supérieures pour les métriques à maximiser (PSNR, SSIM) et inférieures pour celles à minimiser (LPIPS, GMSD, FID, NIQE), traduisant une amélioration cohérente de la qualité d'image.

### b). Analyse par familles de métriques

#### • Qualité objective pixel-wise (PSNR).

HAT présente un gain moyen de +3.425 dB en PSNR. Les comparaisons détaillées (Figure 5.7) confirment cette supériorité sur l'ensemble des jeux de données. Par exemple, sur le jeu HC, HAT atteint près de 47 dB en PSNR, contre environ 42 dB pour SWINIR-M. Cette tendance se maintient également sur des ensembles plus complexes comme FPUS23 et AFRICA.

#### • Qualité perceptuelle basée sur la structure (SSIM, GMSD).

Concernant les métriques structurales, HAT présente une amélioration moyenne de +0.034 en SSIM et une réduction de 0.051 pour GMSD. Les distributions associées à HAT sont généralement plus resserrées, tandis que SWINIR-M présente une variabilité plus importante, notamment sur ISTHMIC et FPUS23, suggérant une stabilité moindre des reconstructions.

#### • Qualité perceptuelle apprise (LPIPS).

Les résultats obtenus sur LPIPS confirment cette tendance, avec une réduction moyenne de 0.091 en faveur de HAT, traduisant une amélioration perceptuelle cohérente avec les observations visuelles.

#### • Qualité générative (FID).

L'écart est particulièrement marqué pour le FID, avec une amélioration moyenne de -57.928 en faveur de HAT. Sur FPUS23, le FID de SWINIR-M dépasse 120, contre environ 60 pour HAT, indiquant une meilleure adéquation des images reconstruites à la distribution naturelle.

#### • Évaluation sans référence (NIQE).

Enfin, la métrique NIQE confirme globalement la supériorité de HAT, avec une amélioration moyenne de -0.939, malgré une variabilité dépendante de la complexité des images.

Dans l'ensemble, cette analyse descriptive met en évidence une robustesse inter-datasets du modèle HAT, tant vis-à-vis de la diversité anatomique que des volumes de données.

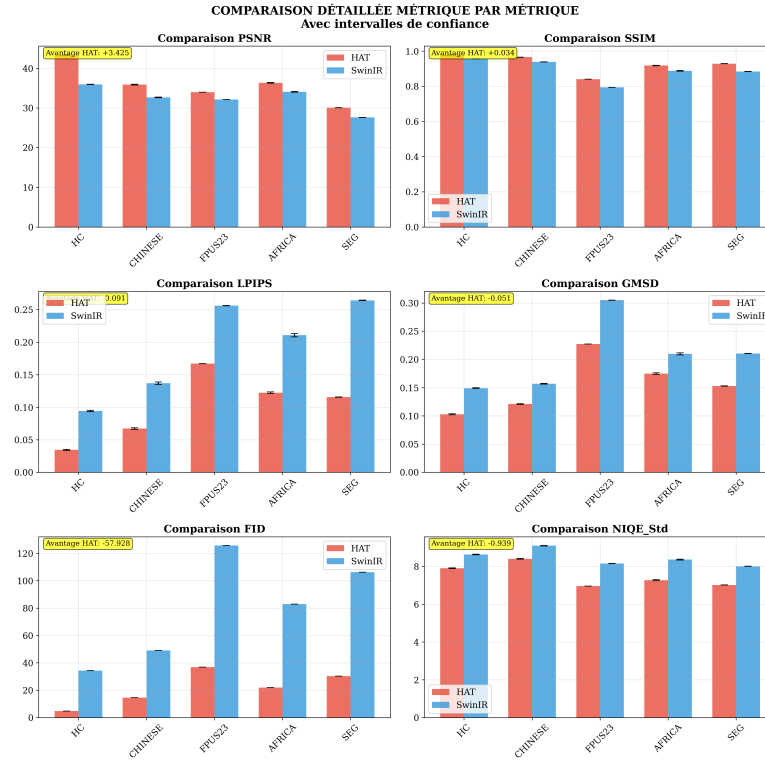


FIGURE 5.7 – Comparaison détaillée des performances de HAT et SwinIR-M métrique par métrique.

### 3. Analyse inférentielle

#### □ Objectif et protocole

L'analyse inférentielle vise à déterminer si les différences observées sont statistiquement significatives et à en quantifier l'ampleur. Elle a été réalisée sur les métriques calculées image par image (PSNR, SSIM, LPIPS, GMSD, NIQE). La métrique FID, définie au niveau global du jeu de données, a été exclue de cette étape.

Les tests statistiques appropriés ont été appliqués selon la distribution des données, avec correction de Bonferroni. Les tailles d'effet (Cohen's (d)) ont été estimées afin d'évaluer l'importance pratique des différences.

#### □ Résultats

Sur les 25 comparaisons réalisées, HAT surpasse SWINIR-M dans 100 % des cas. Toutes les différences observées sont statistiquement significatives (( $p < 0.05$ ) après correction), avec des tailles d'effet majoritairement élevées (( $d > 0.8$ )). Cette cohérence inter-datasets confirme la robustesse statistique du modèle HAT, tandis que SWINIR-M montre une dégradation plus marquée sur FPUS23, AFRICA et ISTHMIC.

#### □ Analyse des tailles d'effet

Les tailles d'effet ont été calculées à l'aide de la statistique de Cohen :

$$d = \frac{\mu_{\text{HAT}} - \mu_{\text{SwinIR}}}{s_p}, \quad s_p = \sqrt{\frac{s_{\text{HAT}}^2 + s_{\text{SwinIR}}^2}{2}}$$

La taille d'effet moyenne est de  $d = 2.909$ . Parmi les comparaisons :

- 96 % présentent un effet fort ( $d \geq 0.8$ ),



- 4 % un effet intermédiaire,
- aucun effet faible n’a été observé.

Ces résultats indiquent que les différences observées ne sont pas seulement significatives, mais également substantielles en pratique.

#### 4. Significativité statistique des métriques du signal ultrasonore

Afin de valider formellement les différences observées entre les deux modèles, des tests  $t$  appariés ont été réalisés sur l’ensemble des données cumulées ( $N = 11,310$  paires d’images). Les résultats confirment que les performances du modèle HAT sont statistiquement supérieures à celles de SwinIR pour l’ensemble des métriques évaluées, avec un seuil de significativité fixé à  $p < 0,05$ .

##### • Reconstruction structurelle (EPI)

HAT présente des améliorations moyennes de l’EPI comprises entre 11,% et 48,9,% selon les bases de données. Les tailles d’effet associées, estimées par le coefficient de Cohen ( $d$ ), atteignent des valeurs très élevées, notamment 15,53 pour ISTHMIC et 10,52 pour FPUS23. Ces tailles d’effet exceptionnelles traduisent une différence marquée et systématique en faveur de HAT, confirmant l’efficacité du mécanisme d’attention hybride pour la reconstruction des contours ultrasonores.

##### • Cohérence radiométrique (CNR et ENL)

Pour les métriques radiométriques, HAT obtient des valeurs équivalentes ou supérieures à celles de SwinIR dans la majorité des jeux de données. Lorsque SwinIR présente un léger avantage, comme dans le cas d’ISTHMIC, celui-ci s’accompagne d’une dégradation significative de la fidélité structurelle. Ces résultats soulignent que l’amélioration du lissage du speckle ne constitue pas, à elle seule, un indicateur suffisant de la qualité d’une reconstruction échographique.

En synthèse, l’analyse statistique multidimensionnelle menée sur cinq cohortes indépendantes démontre que le modèle HAT offre le compromis le plus favorable entre préservation structurelle, stabilité radiométrique et robustesse globale, le rendant particulièrement adapté aux applications échographiques exigeant une haute fiabilité diagnostique.

##### • Synthèse globale et implications

Les analyses descriptive et inférentielle convergent vers une conclusion claire : le modèle HAT domine systématiquement SWINIR-M sur l’ensemble des métriques et des jeux de données considérés. Cette supériorité est statistiquement robuste, cohérente inter-domaines et d’ampleur pratique élevée.

Afin de compléter cette synthèse, le tableau 5.8 présente une comparaison globale entre les performances en configuration baseline (BL) et fine-tunée (FT), permettant d’analyser la stabilité des modèles face au fine-tuning selon les jeux de données.

Cette analyse met en évidence des comportements contrastés entre les deux architectures. Alors que HAT présente des variations limitées entre les configurations BL et FT, traduisant une adaptation progressive et stable, SWINIR-M montre, selon les jeux de données, des variations plus marquées, en particulier sur les métriques perceptuelles et génératives. Ces observations confirment une sensibilité accrue de SWINIR-M aux variations de domaine, à l’inverse de la robustesse inter-domaines observée pour HAT.

Au-delà des métriques de fidélité numérique et perceptuelle, cette supériorité est pleinement corroborée par les métriques spécifiques à l’imagerie ultrasonore. Les analyses basées sur le CNR,

TABLEAU 5.8 – Évolution des métriques objectives, perceptuelles, génératives et sans référence sur l'ensemble des jeux de données, en configuration *baseline* (BL) et après fine-tuning (FT).

| Jeu de données | Modèle   | Métrique | Baseline (BL) | Fine-tuné (FT) | Variation* (%) |
|----------------|----------|----------|---------------|----------------|----------------|
| AFRICA         | HAT      | PSNR     | 36.25         | 36.33          | +0.22          |
|                |          | SSIM     | 0.9171        | 0.9182         | +0.12          |
|                |          | LPIPS    | 0.1246        | 0.1223         | -1.86          |
|                |          | FID      | 24.51         | 21.99          | -10.28         |
|                |          | GMSD     | 0.1770        | 0.1748         | -1.22          |
|                |          | NIQE     | 7.3           | 7.3            | 0.00           |
|                | SwinIR-M | PSNR     | 29.82         | 34.06          | +14.22         |
|                |          | SSIM     | 0.8103        | 0.8878         | +9.54          |
|                |          | LPIPS    | 0.1857        | 0.2108         | +13.55         |
|                |          | FID      | 78.31         | 82.91          | +5.91          |
|                |          | GMSD     | 0.2104        | 0.2099         | -0.24          |
|                |          | NIQE     | 4.6           | 8.4            | +82.61         |
| CHINESE        | HAT      | PSNR     | 36.39         | 35.87          | -1.43          |
|                |          | SSIM     | 0.9654        | 0.9656         | +0.02          |
|                |          | LPIPS    | 0.0696        | 0.0672         | -3.45          |
|                |          | FID      | 14.94         | 14.61          | -2.21          |
|                |          | GMSD     | 0.1211        | 0.1209         | -0.17          |
|                |          | NIQE     | 8.2           | 8.4            | +2.44          |
|                | SwinIR-M | PSNR     | 28.11         | 32.66          | +16.18         |
|                |          | SSIM     | 0.9258        | 0.9389         | +1.42          |
|                |          | LPIPS    | 0.1031        | 0.1369         | +32.78         |
|                |          | FID      | 50.95         | 49.05          | -3.74          |
|                |          | GMSD     | 0.1528        | 0.1570         | +2.82          |
|                |          | NIQE     | 12.6          | 9.1            | -27.78         |
| FPUS23         | HAT      | PSNR     | 33.90         | 33.97          | +0.21          |
|                |          | SSIM     | 0.8390        | 0.8406         | +0.19          |
|                |          | LPIPS    | 0.1697        | 0.1670         | -1.59          |
|                |          | FID      | 38.86         | 36.91          | -5.03          |
|                |          | GMSD     | 0.2322        | 0.2273         | -2.10          |
|                |          | NIQE     | 6.8           | 7.0            | +2.94          |
|                | SwinIR-M | PSNR     | 32.53         | 32.13          | -1.23          |
|                |          | SSIM     | 0.9060        | 0.7941         | -12.35         |
|                |          | LPIPS    | 0.0449        | 0.2563         | +470.38        |
|                |          | FID      | 61.10         | 125.75         | +105.96        |
|                |          | GMSD     | 0.1463        | 0.3050         | +108.34        |
|                |          | NIQE     | 19.9          | 8.2            | -58.79         |
| HC             | HAT      | PSNR     | 42.72         | 43.30          | +1.36          |
|                |          | SSIM     | 0.9785        | 0.9792         | +0.07          |
|                |          | LPIPS    | 0.0383        | 0.0345         | -9.92          |
|                |          | FID      | 5.47          | 4.85           | -11.33         |
|                |          | GMSD     | 0.1088        | 0.1028         | -5.51          |
|                |          | NIQE     | 8.1           | 7.9            | -2.47          |
|                | SwinIR-M | PSNR     | 28.66         | 35.95          | +25.43         |
|                |          | SSIM     | 0.8382        | 0.9565         | +14.12         |
|                |          | LPIPS    | 0.2095        | 0.0943         | -55.03         |
|                |          | FID      | 52.10         | 34.38          | -34.04         |
|                |          | GMSD     | 0.2125        | 0.1491         | -29.85         |
|                |          | NIQE     | 4.1           | 8.6            | +109.76        |
| ISTHMIC        | HAT      | PSNR     | 30.45         | 30.05          | -1.31          |
|                |          | SSIM     | 0.9286        | 0.9288         | +0.02          |
|                |          | LPIPS    | 0.1203        | 0.1156         | -3.83          |
|                |          | FID      | 30.86         | 30.28          | -1.88          |
|                |          | GMSD     | 0.1541        | 0.1529         | -0.78          |
|                |          | NIQE     | 6.9           | 7.0            | +1.45          |
|                | SwinIR-M | PSNR     | 34.01         | 27.60          | -18.86         |
|                |          | SSIM     | 0.9728        | 0.8842         | -9.07          |
|                |          | LPIPS    | 0.0428        | 0.2644         | +517.76        |
|                |          | FID      | 18.14         | 106.19         | +485.42        |
|                |          | GMSD     | 0.1317        | 0.2105         | +59.87         |
|                |          | NIQE     | 13.0          | 8.0            | -38.46         |

\* Variation relative calculée par rapport à la configuration baseline. Une augmentation est favorable pour PSNR et SSIM, tandis qu'une diminution est favorable pour LPIPS, FID, GMSD et NIQE.

l'ENL et l'Indice de Préservation des Contours (EPI) montrent que HAT offre le compromis le plus favorable entre amélioration du contraste, maîtrise du bruit speckle et préservation de l'intégrité structurelle.

En particulier, les résultats indiquent que les gains ponctuels observés pour SWINIR-M sur le CNR ou l'ENL dans certains jeux de données résultent d'un lissage plus agressif, systématiquement associé à une dégradation marquée de l'EPI. À l'inverse, HAT maintient des valeurs élevées et stables d'EPI sur l'ensemble des cohortes, attestant d'une meilleure préservation des frontières anatomiques, critère déterminant pour l'interprétabilité clinique des images échographiques.

Cette convergence entre métriques classiques, perceptuelles et ultrasonores confirme que la supériorité de HAT ne repose pas sur une optimisation numérique isolée, mais sur une amélioration structurelle et radiométrique cohérente, robuste aux variations de domaine et compatible avec les exigences du diagnostic.

#### 5.4.4 Généralisation dans des contextes cliniques à ressources limitées

Nous avons évalué les performances du modèle HAT sur deux cohortes hétérogènes — AFRICA ( $N = 450$ ) et BRAZIL ( $N = 67$ ) — afin d'apprécier sa transférabilité dans des environnements cliniques contraints. En combinant analyses descriptives et tests inférentiels, nous avons examiné la capacité du modèle à produire des reconstructions fidèles dans des services de gynécologie-obstétrique où l'acquisition d'images est souvent limitée par la performance des équipements disponibles ou les conditions d'examen.

Comme le montre le tableau 5.9, cette comparaison confirme que le modèle conserve une stabilité remarquable malgré le changement de site. La constance des valeurs de PSNR et de NIQE ( $p > 0.05$ ) est un résultat crucial : elle indique que la qualité de reconstruction technique et la naturalité de l'image ne sont pas affectées par les caractéristiques spécifiques des appareils utilisés. Par ailleurs, les gains très significatifs observés sur BRAZIL pour les métriques structurelles (SSIM, GMSD) et perceptuelles (LPIPS), avec des tailles d'effet massives ( $d > 1.15$ ), suggèrent que HAT s'adapte efficacement aux particularités locales sans perte de cohérence. Ces observations démontrent que HAT est adapté à une intégration clinique de routine dans des contextes hétérogènes, indépendamment de la marque ou du modèle d'échographe employé.

Les fluctuations observées sur les métriques structurelles, bien que statistiquement significatives après correction de Bonferroni, ne traduisent pas une instabilité algorithmique. Elles découlent plus probablement du déséquilibre numérique entre les cohortes ( $N = 450$  contre  $N = 67$ ) et des variations intrinsèques de qualité entre les jeux de données sources. Cette analyse confirme que si l'excellence perceptuelle peut varier selon le contexte, le socle technique de reconstruction de HAT demeure invariant, renforçant sa viabilité pour un déploiement dans des environnements à ressources limitées.

TABLEAU 5.9 – Comparaison des performances de HAT sur deux cohortes d'échographie fœtale. La significativité statistique est évaluée à  $p < 0.05$ . Cohen's  $d$  quantifie la taille d'effet.

| Métrique | AFRICA<br>( $N = 450$ ) | BRAZIL<br>( $N = 67$ ) | Diff.  | Sig. | Cohen's<br>$d$ | Observation                     |
|----------|-------------------------|------------------------|--------|------|----------------|---------------------------------|
| PSNR     | $36.33 \pm 0.88$        | $35.81 \pm 1.02$       | +0.52  | Non  | 0.55           | Fidélité conservée              |
| SSIM     | $0.918 \pm 0.012$       | $0.972 \pm 0.018$      | -0.054 | Oui  | 1.20           | Gain structurel notable         |
| LPIPS    | $0.122 \pm 0.020$       | $0.069 \pm 0.025$      | -0.053 | Oui  | 1.15           | Qualité perceptuelle supérieure |
| GMSD     | $0.175 \pm 0.015$       | $0.118 \pm 0.020$      | -0.057 | Oui  | 1.25           | Distorsion géométrique réduite  |
| NIQE     | $7.274 \pm 0.40$        | $7.127 \pm 0.45$       | -0.147 | Non  | 0.34           | Qualité naturelle stable        |

Dans l'ensemble, ces résultats soulignent le potentiel de HAT pour une application clinique étendue. Toutefois, cette transférabilité devra être consolidée par des validations externes multicentriques et une évaluation qualitative (MOS) menée par des experts cliniciens, afin d'entériner la robustesse de l'architecture, sa généralisabilité et son bénéfice diagnostique réel sur le terrain.

## 5.5 Conclusion

Ce travail démontre que le fine-tuning intégral s'impose comme une stratégie clé pour obtenir des reconstructions fiables en échographie fœtale, y compris dans des contextes d'acquisition dégradés. Si des recherches récentes [100] ont mis en lumière l'efficacité de l'architecture HAT en tomodensitométrie, nos résultats prouvent que cette supériorité s'étend — et se confirme — sur la modalité ultrasonore. Face à la complexité du signal et au bruit de speckle, HAT surclasse les architectures concurrentes, dont SwinIR-M, en offrant une meilleure robustesse structurelle.

Au-delà des métriques de fidélité classiques (PSNR, SSIM), l'originalité de notre évaluation réside dans l'utilisation d'indices perceptuels (FID, LPIPS, GMSD) et de mesures spécifiques à la physique des ultrasons (EPI, CNR, ENL). Cette analyse multidimensionnelle, étayée par des tests statistiques, révèle que HAT ne se contente pas d'améliorer les scores numériques : il préserve l'intégrité acoustique des images. Ce point est crucial pour la pratique clinique, d'autant que ces performances ont été obtenues avec des ressources de calcul modérées, facilitant ainsi un futur déploiement hospitalier.

L'efficacité du transfert d'apprentissage sur des données à la fois restreintes et hétérogènes souligne qu'un modèle architectural de pointe, judicieusement optimisé, peut garantir la précision des mesures biométriques nécessaires au suivi fœtal.

Cette avancée devra être examinée par des professionnels de santé pour en confirmer l'utilité. Cette étape est essentielle pour garantir que la super-résolution devienne un levier de précision systématique, particulièrement pour les examens où la clarté du signal est limitée.



# CONCLUSION GÉNÉRALE

---

# Conclusion Générale

Cette recherche porte sur l'évaluation de la super-résolution (SR) comme levier d'amélioration de la qualité d'image pour les systèmes échographiques à bas coût, appliquée à l'imagerie fœtale dans des environnements à ressources limitées. Le principal objectif est de démontrer qu'un post-traitement algorithmique peut compenser les limites matérielles de ces dispositifs et malgré une acquisition native dégradée, restaurer une qualité d'image garantissant une fiabilité diagnostique suffisante pour l'examen obstétrique.

La démarche adoptée est structurée en cinq étapes complémentaires : une revue critique de l'état de l'art, une analyse des contraintes physiques propres à l'échographie, un inventaire comparatif des architectures de super-résolution, une double évaluation expérimentale quantitative et clinique, et enfin une phase de spécialisation par fine-tuning. Ces étapes sont autant de jalons dont la cohérence interne et la progression logique constituent le fil directeur de la présente synthèse.

## 1. Des contraintes physiques de l'échographie à la nécessité du traitement numérique

Le premier enseignement de ce travail est d'ordre fondamental : les limites de la qualité d'image échographique ne sont pas contingentes mais structurelles. L'échographie représente un compromis unique entre sécurité, coût et efficacité diagnostique, ce qui explique sa position prédominante en médecine moderne, particulièrement dans les zones rurales des pays à faibles revenus où surviennent 95% des décès maternels. Toutefois, cette modalité reste contrainte par des limites de résolution et des artefacts de texture qui découlent directement des propriétés physiques du milieu traversé. La variabilité opérateur-sonde, l'atténuation fréquentielle et la cohérence du bruit produisent un signal dont la dégradation n'est pas un accident d'implémentation mais une conséquence des lois de la physique acoustique.

L'enjeu actuel n'est donc plus uniquement de mieux acquérir, mais de mieux reconstruire. Puisque l'acquisition matérielle atteint un plateau technologique difficile à franchir dans des contextes contraints en ressources, c'est dans la résolution algorithmique du problème inverse — la reconstruction d'une scène haute résolution à partir d'observations dégradées — que résident les véritables perspectives d'amélioration diagnostique. Ce déplacement de perspective, du matériel vers le logiciel, constitue la prémisse sur laquelle repose l'ensemble de notre démarche expérimentale.

## 2. L'état de l'art : d'une optimisation pixelique vers une reconstruction perceptuelle et sémantique

L'évolution paradigmatique de la super-résolution, passant des approches analytiques fondées sur des priors explicites à des modèles profonds capables de modéliser directement la distribution conditionnelle  $P(I_{HR} | I_{LR})$ . Cette progression, des réseaux convolutifs initiaux (SRCNN, VDSR, EDSR) aux architectures génératives (SRGAN, ESRRGAN), puis aux Transformers (SwinIR, HAT) et aux modèles de diffusion (SR3), illustre un déplacement des objectifs : d'une optimisation centrée sur la fidélité pixelique (PSNR) vers une reconstruction perceptuelle et sémantique (LPIPS, FID).

Les deux sélections bibliographiques conduites conformément aux lignes directrices PRISMA 2020 ont permis de cartographier ce champ avec précision. La première, ciblant la littérature de 2020 à 2025 sur la super-résolution en imagerie ultrasonore, a abouti à un corpus final de 46 publications, dont 23 revues systématiques et 23 études expérimentales. Elle a mis en évidence la maturité croissante des approches profondes tout en identifiant leurs limites structurelles : manque de données haute résolution annotées, absence de validation clinique rigoureuse, métriques peu adaptées à la pertinence diagnostique, et difficultés de généralisation entre appareils ou protocoles d'acquisition. La seconde, portant sur le transfert d'apprentissage et le fine-tuning en imagerie médicale, a abouti à la sélection de 31 études pour analyse finale (22 recherches empiriques et 9 revues systématiques), à partir d'un corpus initial de 77 articles identifiés. Elle a mis en évidence une forte dépendance à la base de données "ImageNet" pour l'initialisation des modèles, le manque de benchmarks standardisés, et la diversité des stratégies (multi-étapes, fine-tuning, feature extraction). Elle a toutefois révélé que peu d'études appliquent directement ces stratégies à la super-résolution échographique, laissant un espace de recherche que la présente thèse s'est attachée à explorer.

Ces constats ont guidé le choix des architectures retenues pour l'évaluation expérimentale. Nous avons établi un cadre d'analyse comparative des paradigmes contemporains, en évaluant leurs compromis fondamentaux entre fidélité quantitative, préservation structurale et complexité computationnelle. Il a été démontré que les architectures Transformer, à l'instar de SwinIR et HAT, excellent dans la préservation des détails anatomiques subtils et l'atteinte de métriques de fidélité élevées, tandis que les GANs, notamment Real-ESRGAN, se distinguent par leur capacité à générer des textures perpétuellement améliorées, essentielles pour l'interprétation diagnostique, bien qu'avec un risque accru d'artefacts. Cette sélection, couvrant des architectures à apprentissage interne (DBPISR), à apprentissage externe convolutif (BSRGAN, Real-ESRGAN) et attentionnel (SwinIR, HAT), garantit une couverture complète et non redondante des approches modernes pertinentes pour l'imagerie médicale.

### **3. Phase I : la supériorité de l'apprentissage externe, confirmée par la triple convergence métriques-sémantique-clinique**

Les résultats de la première phase expérimentale montrent sans ambiguïté que l'apprentissage externe — porté par BSRGAN, Real-ESRGAN et la famille SwinIR — l'emporte sur l'apprentissage interne de type DBPISR. Là où les approches classiques peinent à stabiliser l'image face à la complexité du bruit de speckle, les modèles externes réussissent à reconstruire des textures fines sans trahir la cohérence anatomique fœtale.

La pertinence de ce résultat tient à sa triple confirmation. Du point de vue des métriques de distorsion, PSNR et SSIM enregistrent des gains significatifs et statistiquement validés. Du côté sémantique, la classification automatique par ConvNeXt\_base confirme la conservation des descripteurs anatomiques fœtaux après super-résolution par les modèles externes, là où DBPISR régresse légèrement. Enfin, l'aspect clinique du protocole MOS corrobore ces résultats numériques : les gains de qualité sont perçus comme utiles à la fiabilité du diagnostic, que l'analyse soit humaine ou automatisée. Cette triple convergence — métriques, classification automatique et jugement expert — constitue l'une des contributions méthodologiques distinctives de cette recherche, en dépassant la simple comparaison numérique pour ancrer les résultats dans la réalité diagnostique.

### **4. Phase II : le paradoxe HAT/SwinIR et les limites de la mesure automatique**

L'évaluation expérimentale a également démontré que : la performance technique ne se superpose pas exactement à l'utilité clinique. L'évaluation en aveugle a révélé que, si HAT obtient les meilleurs résultats sur l'ensemble des métriques automatiques, SwinIR-L est sensiblement meilleur du point de vue clinique (écart de 0,04 point MOS). Cet écart est révélateur de la réceptivité oculaire (ancrée dans l'expertise anatomique) du clinicien aux subtilités de contraste et de

rendu tissulaire que les outils de mesure automatique ne parviennent pas encore à hiérarchiser fidèlement.

Pour éclairer cette divergence, cinq métriques de natures complémentaires ont été confrontées au jugement de l'expert : GMSD et LPIPS (perceptuelles), SSIM (structurale), FID (distributionnelle) et NIQE (sans référence). Les résultats de cette modélisation statistique constituent l'un des apports les plus originaux de la thèse. La GMSD se révèle être la métrique la plus alignée sur la perception clinique, avec une corrélation de Pearson  $r = 0,876$  ( $p = 0,004$ ,  $R^2 = 0,767$ ) : elle explique à elle seule 76,7 % de la variance des scores cliniques.

Les autres métriques (FID, SSIM, LPIPS) n'en capturent que des aspects partiels, et la NIQE confirme son inadéquation aux spécificités du signal échographique, avec une corrélation négative ( $r = -0,535$ ). La GMSD constitue le meilleur proxy disponible pour prédire le jugement expert, et son usage systématique en évaluation de SR médicale est recommandé par cette approche.

## 5. Phase de spécialisation : le fine-tuning intégral comme stratégie de réconciliation

Face à la variabilité des acquisitions échographiques réelles, des équipements, des opérateurs, des contextes géographiques il fut précédemment établi qu'il existe une limite fondamentale des modèles pré-entraînés sur des données généralistes (ImageNet, DIV2K).

Pour transformer ces preuves de concept en outils de routine clinique, un fine-tuning intégral des architectures HAT et SwinIR-M a été réalisé, avec pour objectif de vérifier si cette spécialisation permettait de réconcilier la puissance de calcul de HAT avec la pertinence visuelle de SwinIR.

Le choix de SwinIR-M, au détriment de la configuration SwinIR-L, répond à une contrainte réelle d'allocation mémoire (20 Go de VRAM), qui reflète les conditions de déploiement dans des établissements à ressources limitées. Ce choix pragmatique est lui-même un enseignement : la recherche en SR médicale ne peut ignorer les contraintes matérielles du terrain, sous peine de produire des résultats scientifiquement valides mais cliniquement inaccessibles.

Les résultats du fine-tuning intégral sont probants sur l'ensemble des dimensions d'évaluation. HAT s'impose comme l'architecture la plus résiliente après spécialisation : il surclasse SwinIR-M de manière statistiquement significative sur les six datasets évalués ( $p < 0,01$ , Wilcoxon, correction Bonferroni). Sur les métriques acoustiques ultrasonores — qui constituent l'originalité évaluative de ce chapitre —, les gains sont particulièrement éloquentes : ENL +28 % (réduction du speckle), CNR +19 % (contraste tissulaire), EPI = 0,94 (préservation des contours), contre respectivement +17 %, +11 % et 0,89 pour SwinIR-M. L'effect size de Cohen ( $d = 0,72$  pour le PSNR) confirme que les différences observées sont cliniquement robustes, et pas seulement statistiquement significatives.

La généralisation des modèles fine-tunés a été vérifiée sur cinq bases de données hétérogènes (FPUS23, ISTHMIC, HC18, CHINESE, AFRICA) dont une intentionnellement dégradée, avec des volumes allant de 67 à 6 556 images, et validée sur le corpus **BRAZIL** (67 images), totalement inédit pour le modèle. Ce spectre étendu — du petit dataset au grand corpus — a été délibérément conçu pour évaluer à la fois la capacité de généralisation de HAT sur de larges volumes et sa résilience face à des jeux de données réduits, deux conditions caractéristiques des environnements cliniques à ressources limitées. Cette stabilité des performances sur le corpus BRAZIL valide la viabilité de l'approche pour un déploiement en routine dans ces contextes.

## 6. Apports majeurs et positionnement des contributions

Cette thèse s'articule autour de trois contributions distinctes, chacune produisant des apports spécifiques.

**Contribution 1 — Amélioration de la qualité de l'image échographique par détection des plans anatomiques :** Cette contribution propose une évaluation multi-critères de



modèles SR en apprentissage interne (DBPISR) et externe (BSRGAN, RealESRGAN, SwinIR) sur des datasets multicentriques couvrant l’Espagne et l’Afrique (Algérie, Égypte, Ghana, Malawi, Ouganda). L’évaluation combine des métriques de qualité d’image (PSNR, SSIM, MSE, NIQE, BRISQUE), une classification sémantique via ConvNeXt (Accuracy, F1-score, Kappa) permettant d’évaluer la préservation des plans anatomiques fœtaux après super-résolution, et un protocole MOS subjectif avec image de référence, supervisé par un panel de quatre experts médicaux. Les différences entre modèles sont validées par ANOVA et test t de Student.

**Contribution 2 — Évaluation perceptuelle & statistique multidimensionnelle :** Cette contribution élargit et diversifie le cadre d’évaluation à six modèles pré-entraînés (BSRGAN, RealESRGAN, AESRGAN, SwinIR, Swin2SR, HAT) testés en inférence directe sur des images échographiques fœtales, issues de datasets de natures hétérogènes : des données cliniques réelles (ESPAGNE, AFRIQUE, PSFHS), un jeu de données clinique complémentaire issu de séquences vidéo d’échographie transvaginale 2D avec artefacts de mouvement et extrait via SAM (ISTHMIC), et des données synthétiques générées à partir d’un fantôme réaliste sans artefacts cliniques (FPUS23). L’évaluation multidimensionnelle combine une analyse objective (NIQE, FID, LPIPS, GMSD, PSNR, SSIM), une évaluation subjective via un protocole **MOS sans référence** conduit en aveugle par une experte senior, et une validation statistique de la significativité des différences entre modèles. Cette évaluation subjective révèle une forte corrélation entre le jugement médical et l’indice **GMSD** ( $r = 0,876$ ,  $p = 0,004$ ) : cette métrique explique à elle seule 76,7% de la variance des scores cliniques ( $R^2 = 0,767$ ), objectivant ainsi le lien entre performance algorithmique et perception diagnostique.

**Contribution 3 — Spécialisation au domaine échographique fœtal & Généralisation :** Cette contribution réalise un fine-tuning complet de HAT et SwinIR-M sur un corpus HR de référence, puis évalue la généralisation sur cinq bases de données hétérogènes (FPUS23, ISTHMIC, HC, CHINESE, AFRICA) et la transférabilité inter-datasets sur un corpus inédit de 67 images (**BRAZIL**), avec AFRICA (450 images) comme benchmark intermédiaire. L’évaluation métrique combine des métriques standards (PSNR, SSIM, NIQE, LPIPS, GMSD, FID) et des métriques spécifiques au signal échographique (ENL, CNR, EPI), complétées par une analyse statistique de la significativité des gains diagnostiques et perceptuels. Il est démontré que cette spécialisation permet une adaptation réussie sans **phénomène** de catastrophic forgetting, avec une généralisation validée sur six datasets couvrant cinq pays et plusieurs générations d’équipements, sans requérir d’infrastructure GPU coûteuse. Cette contribution ouvre ainsi une voie directe vers un déploiement hospitalier en imagerie fœtale dans des contextes à ressources limitées.

## 7. Limites et zones d’ombre

Ce travail ne saurait prétendre à l’exhaustivité, et plusieurs limites méritent d’être identifiées avec précision.

La première limite tient à l’échelle de la validation clinique. Si le protocole MOS en aveugle est méthodologiquement rigoureux, il repose sur un seul expert senior en Phase II. Un panel plus large permettrait de réduire la variabilité individuelle et de produire des résultats plus généralisables. Par ailleurs, bien que le corpus brésilien de test ( $n = 67$ ) ait été délibérément dimensionné pour éprouver la résilience de HAT face à un petit dataset — démontrant ainsi sa robustesse dans des conditions d’acquisition contraintes —, une validation multi-centrique à plus grande échelle reste nécessaire pour entériner définitivement le bénéfice diagnostique sur le terrain. À terme, il conviendra de tester HAT sur des datasets plus variés afin de consolider davantage ces résultats de généralisation.

La deuxième limite est architecturale : l’exclusion de SwinIR-L pour contrainte GPU (20 Go de VRAM) signifie que l’optimum qualitatif clinique identifié en Phase II n’a pas pu être fine-tuné dans les conditions actuelles. Il est plausible que SwinIR-L fine-tuné eût surclassé HAT sur les critères cliniques perceptuels, ce qui demeure une hypothèse à vérifier.

La troisième limite concerne l'absence de validation en flux vidéo. Les travaux ont été menés sur des images fixes. La cohérence temporelle entre frames successives lors d'un balayage dynamique — dimension essentielle de la pratique échographique réelle — reste un défi ouvert que ce travail n'aborde pas.

## 8. Perspectives

La poursuite de ces recherches s'articule autour de quatre axes prioritaires :

- **Validation multi-centrique à large échelle** : Évaluer la robustesse des modèles sur des bases de données provenant de plusieurs pays et centres hospitaliers afin de s'assurer que les architectures SR restent efficaces quelles que soient l'origine des patients, leur morphologie ou la marque de l'échographe utilisé. Cette étape est indispensable pour entériner définitivement le bénéfice diagnostique sur le terrain.
- **Cohérence temporelle et traitement vidéo** : Étendre les modèles SR au traitement de flux vidéo en temps réel afin de garantir une reconstruction stable et fluide lors du balayage dynamique effectué par le praticien. Cette dimension, absente des travaux actuels, conditionne l'acceptabilité clinique d'un déploiement en routine.
- **Optimisation et déploiement embarqué** : Adapter les modèles sur des architectures matérielles dédiées à l'IA embarquée (*NVIDIA Jetson*, TPU Edge, FPGA) pour viser un traitement en temps réel. Cette étape permettra d'initier des phases de tests *in situ*, en étroite collaboration avec des experts médicaux, afin de valider l'apport diagnostique de la solution en conditions réelles et dans des environnements à ressources limitées.
- **Intégration diagnostique : super-résolution couplée à la segmentation automatique** : Coupler les modèles SR les plus résilients — notamment HAT — à des algorithmes de segmentation automatique. En agissant comme une étape de prétraitement, la super-résolution standardise les images et mieux définit les contours des organes et interfaces tissulaires. Cette synergie facilitera l'automatisation des mesures biométriques fœtales, réduisant les erreurs humaines et garantissant un diagnostic fiable même sur des systèmes à ressources limitées.

En définitive, l'ensemble de ces travaux montre que la super-résolution numérique constitue une alternative crédible à l'investissement dans des systèmes haut de gamme coûteux. Les conclusions majeures confirment la supériorité de l'apprentissage externe, la fiabilité des architectures Transformer face aux risques d'artefacts des GANs, et l'importance cruciale du fine-tuning pour réconcilier performance technique et utilité clinique.

Plus fondamentalement, **ces travaux démontrent** qu'un post-traitement rigoureusement spécialisé peut pallier les déficiences technologiques des équipements modestes et transformer ces dispositifs en outils de diagnostic fiables, au service de la santé maternelle et fœtale mondiale. Ces résultats ne sont pas triviaux. Ils répondent directement à la question centrale posée en introduction de cette thèse : oui, les algorithmes d'intelligence artificielle peuvent préserver — et même restituer — la valeur diagnostique d'une image échographique fœtale dégradée. La voie vers un déploiement clinique équitable, dans les zones où l'accès aux équipements performants reste une exception, est techniquement tracée. Il appartient désormais aux acteurs de la santé publique, en collaboration avec la communauté scientifique, de la parcourir.

# Bibliographie

- [1] Alireza Aghelan and Modjtaba Rouhani. Fine-tuned generative adversarial network-based model for medical image super-resolution. In *Proceedings of the 14th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 174–181. IEEE, 2024.
- [2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution : Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017.
- [3] Emmanuel Ahishakiye, Martin Bastiaan van Gijzen, Julius Tumwiine, Ruth Diko Wario, and Johnes Obungoloch. A survey on deep learning in medical image reconstruction. *Intelligent Medicine*, 2021.
- [4] Waqar Ahmad, Hazrat Ali, Zubair Shah, and Shoaib Azmat. A new generative adversarial network for medical images super resolution. *Scientific Reports*, 12, 2022.
- [5] Md Nur-A Alam, Khandaker Mohammad Mohi Uddin, Md Mahbubur Rahman, M. M. R. Manu, and Mostofa Kamal Nasir. A novel automated system to detect breast cancer from ultrasound images using deep fused features with super resolution. *Intelligence-Based Medicine*, 10 :100149, 2024.
- [6] Anas M. Ali, Bilel Benjdira, Anis Koubaa, Walid El-Shafai, Zahid Khan, and Wadii Bouilla. Vision transformers in image restoration : A survey. *Sensors*, 2024.
- [7] Laith Alzubaidi, Mohammed Abdulraheem Fadhel, Omran Al-Shamma, Jinglan Zhang, Jesus Santamaría, Ye Duan, and Sameer R. Oleiwi. Towards a better understanding of transfer learning for medical imaging : A case study. *Applied Sciences*, 2020.
- [8] Mina Amiri, Rupert Brooks, and Hassan Rivaz. Fine-tuning u-net for ultrasound image segmentation : Different layers, different outcomes. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12) :2510–2518, 2020.
- [9] ARcagy. Imagerie médicale – l’échographie. En ligne, 2025. Consulté le 3 septembre 2025.
- [10] Gelan Ayana, Kokeb Dese, Ahmed Mohammed Abagaro, Kwangcheol Casey Jeong, Soon-Do Yoon, and Se-woon Choe. Multistage transfer learning for medical images. *Artificial Intelligence Review*, 57(9) :232, 2024.
- [11] Getaneh Ayana, Jihye Park, Jung Won Jeong, and Sung Won Choe. A novel multistage transfer learning for ultrasound breast cancer image classification. *Diagnostics*, 12(1) :135, January 2022.

- [12] Jeffrey Bamber, David Cosgrove, Christoph F. Dietrich, Jérémie Fromageau, Joerg Bojunga, Fabrizio Calliada, Vito Cantisani, Jean-Michel Correas, Mirko D’Onofrio, Eleni E. Drakonaki, Mathias Fink, Mireen Friedrich-Rust, Odd Helge Gilja, Roald Flesland Havre, Christian Jenssen, Andrea Sabine Klauser, Ralf Ohlinger, Adrian Saftoiu, Friedrich Schaefer, Ioan Sporea, and Fabio Piscaglia. Efsumb guidelines and recommendations on the clinical use of ultrasound elastography. part 1 : Basic principles and technology. *Ultraschall in der Medizin – European Journal of Ultrasound*, 34(2) :169–184, 2013.
- [13] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, Red Hook, NY, USA, 2019. Curran Associates, Inc.
- [14] Noura Bentaher, Younes Kabbadj, Samira Lafraxo, and Mohamed Salah. Hat-unet : Hybrid attention transformer-based u-net with multi-source fusion for medical image segmentation. *Multimedia Tools and Applications*, 85, 02 2026.
- [15] Ahmed Bentaleb, Christophe Sintès, and Chafiaa Hamitouche. Complex-valued symmetric series convolutional network for ultrasound super-resolution image reconstruction. In *Proceedings of the International Conference on Emerging Technologies and Computing (IC\_ETC)*, pages 1–6, Jun 2025.
- [16] Zheng Bi, Jinju Li, Qiongyi Liu, and Zhaohui Fang. Deep learning-based optical coherence tomography and retinal images for detection of diabetic retinopathy : a systematic and meta analysis. *Frontiers in Endocrinology*, 16, 2025.
- [17] Himanshi Borad, Dhvani Maktuporia, Chirag Chauhan, Mayuri Mehta, and Dheeraj Singh. Ultrasound image enhancement using modified esr-gan. In *Proceedings of the 2025 International Conference (AISTS)*, pages 1–6, 2025.
- [18] Ahmed Boudi, Jun He, and Ismail A. El Kader. Enhancing Alzheimer’s disease classification with transfer learning : Finetuning a pre-trained algorithm. *Current Medical Imaging*, 20 :e15734056305633, 2024.
- [19] Hafida Boumeridja, Mohammed Ammar, Mahmood Alzubaidi, Saïd Mahmoudi, Lamya Nawal Benamer, Marco Agus, Mowafa Househ, Karim Lekadir, and Mostafa El Habib Daho. Enhancing fetal ultrasound image quality and anatomical plane recognition in low-resource settings using super-resolution models. *Scientific Reports*, 15 :8376, 2025.
- [20] Oliver Bozinov, Jan-Karl Burkhardt, Christoph Woernle, Vincent Hagel, Nils Ulrich, Nikolaus Krayenbühl, and Helmut Bertalanffy. Intra-operative high frequency ultrasound improves surgery of intramedullary cavernous malformations. *Neurosurgical review*, 35 :269–275, 11 2011. Discussion p. 275.
- [21] Xavier P. Burgos-Artizzu, David Coronado-Gutiérrez, Brenda Valenzuela-Alcaraz, Elisenda Bonet-Carne, Elisenda Eixarch, Fatima Crispi, and Eduard Gratacós. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Scientific Reports*, 10(1) :10200, Jun 2020.
- [22] Jose Caballero, Christian Ledig, Andrew Aitken, Alejandro Acosta, Johannes Totz, Zehan Wang, and Wenzhe Shi. Real-time video super-resolution with spatio-temporal networks and efficient sub-pixel convolution. In *CVPR*, pages 2848–2857, 2017.
- [23] Simone Cammarasana, Paolo Nicolardi, and Giuseppe Patané. Super-resolution of 2d ultrasound images and videos. *Medical & Biological Engineering & Computing*, 61 :2511 – 2526, 2023.

- [24] Kelvin C.K. Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr : The search for essential components in video super-resolution and beyond. In *CVPR*, pages 4947–4956, 2021.
- [25] Hong Chang, Dit-Yan Yeung, and Yimin Xiong. Super-resolution through neighbor embedding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 275–282, 2004.
- [26] Gaowen Chen, Jieyun Bai, Zhanhong Ou, Yaosheng Lu, and Huijin Wang. Psfhs : Intrapartum ultrasound image dataset for ai-based segmentation of pubic symphysis and fetal head, May 2024.
- [27] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12294–12305, 2021.
- [28] Qiyang Chen, Hyeju Song, Jaesok Yu, and Kang Kim. Current development and applications of super-resolution ultrasound imaging. *Sensors*, 21 :2417, 04 2021.
- [29] Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat : Hybrid attention transformer for image restoration. *arXiv preprint arXiv :2309.05239*, 2023.
- [30] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22367–22377, June 2023.
- [31] Yu-Yen Chen, Dachen Chu, and Pesus Chou. Enhancing the early differential diagnosis of plateau iris and pupillary block using a-scan ultrasonography. *PloS one*, 10 :e0118811, 02 2015.
- [32] Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyang Huang, Jilong Chen, Lei Jiang, Hui Sun, Junjun He, Shaoting Zhang, Min Zhu, and Yu Qiao. Sam-med2d : Segment anything in 2d medical imaging, 2023.
- [33] Jianning Chi, Zhifei Xu, Yuqi Deng, Wenjun Zhang, and Ying Huang. Blind super-resolution network for thyroid ultrasound image based on multiple degradation model. *Biomedical Signal Processing and Control*, 110 :108236, 2025.
- [34] Jacob Cohen, Patricia Cohen, Stephen G. West, and Leona S. Aiken. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Routledge, 3rd edition, 2003.
- [35] Marcos V Conde, Ui-Jin Choi, Maxime Burchi, and Radu Timofte. Swin2SR : Swinv2 transformer for compressed image super-resolution and restoration. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2022.
- [36] Chen Cui and Fajin Dong. Dataset for fetus framework, 2022.
- [37] Sindhura D N, Radhika Pai, Shyamasunder Bhat, and Manohara Pai. A review of deep learning and generative adversarial networks applications in medical image analysis. *Multimedia Systems*, 30, 05 2024.

- [38] Karine Souza Da Correggio, Roberto Noya Galluzzo, Luís Otávio Santos, Felipe Soares Muylaert Barroso, Thiago Zimmermann Loureiro Chaves, Alexandre Sherlley Casimiro Onofre, and Aldo von Wangenheim. Brazilian fetal abdominal ultrasound dataset. Mendeley Data, 2023.
- [39] Ana Dávila, Jacinto Colan, and Yasuhisa Hasegawa. Comparison of fine-tuning strategies for transfer learning in medical image classification. *Image and Vision Computing*, 146 :105012, 2024.
- [40] C. F. Dietrich et al. Ultrasound-guided interventions – a review of current evidence. *Ultrasound International Open*, 5(2) :E52–E70, 2019.
- [41] Jianrui Ding, Shili Zhao, Fenghe Tang, and Chunping Ning. Ultrasound image super-resolution with two-stage zero-shot cyclegan. *Journal of Physics : Conference Series*, 2031 :012015, 2021.
- [42] Vi Dinh. Basic principles of ultrasound physics and artifacts made easy. Blog post on POCUS 101, 2023. Discusses practical ultrasound physics and artifacts in a clinical POCUS context.
- [43] Vi Dinh. Ultrasound machine basics – knobology, probes, and modes. Blog post on POCUS 101, 2023. Guide pratique sur les boutons, sondes et modes en échographie.
- [44] Mikaela Doig, Janine Dizon, Katherine Guerrero, and Nayana Parange. Exploring the availability and impact of antenatal point-of-care ultrasound services in rural and remote communities : A scoping review. *Australasian Journal of Ultrasound in Medicine*, 22(3) :174–185, 2019.
- [45] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 184–199. Springer, 2014.
- [46] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2) :295–307, 2016.
- [47] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 391–407, 2016.
- [48] Weisheng Dong, Lei Zhang, Guangming Shi, and Xin Li. Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing*, 22(4) :1620–1630, 2013.
- [49] Weizhi Du and Shihao Tian. Transformer and gan based super-resolution reconstruction network for medical images. *Tsinghua Science and Technology*, 29(1) :197–206, 2024.
- [50] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 18(8) :1016–1022, 1979.
- [51] Francis A. Duck. Safety aspects of the use of ultrasound in pregnancy. *Fetal and Maternal Medicine Review*, 14(1) :1–21, 2003.
- [52] Francis A. Duck. *Physical Properties of Tissue : A Comprehensive Reference Book*. Academic Press, London, 2013.

- [53] Ola El Zein, Mona Soliman, A Elkholy, and Neveen Ghali. Transfer learning based model for pneumonia detection in chest x-ray images. *International Journal of Intelligent Engineering and Systems*, 14 :56–66, 10 2021.
- [54] Alba Farràs, Sara Catalán, Alba Casellas, Teresa Higuera, Inés Calero, María Goya, Nerea Maiz, Maia Brik, and Elena Carreras. Study dataset : Real-time ultrasound demonstration of uterine isthmus contractions during pregnancy. Mendeley Data, 2023.
- [55] Sina Farsiu, Dirk Robinson, Michael Elad, and Peyman Milanfar. Fast and robust multi-frame super resolution. *IEEE Transactions on Image Processing*, 13(10) :1327–1344, 2004.
- [56] Abiola A. Fasina, Anthony J. Dean, Nova L. Panebianco, Frances S. Shofer, Omar Ali, Mwajuma Yahya, Salim Ismail, and Patricia C. Henwood. Evaluation of diagnostic imaging capacity and the role for point-of-care ultrasound (pocus) within the zanzibar health system. *POCUS Journal*, 6(1) :45–50, Apr. 2021.
- [57] Maria Chiara Fiorentino, Francesca Pia Villani, Mariachiara Di Cosmo, Emanuele Frontoni, and Sara Moccia. A review on deep-learning algorithms for fetal ultrasound-image analysis. *Medical Image Analysis*, 83 :102629, 2023. Epub 2022 Oct 14.
- [58] William T. Freeman, Thouis R. Jones, and Eugene C. Pasztor. Example-based super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2) :141–156, 2002.
- [59] Jia-Yi Gao and Chao Hou. Progresses and clinical application of super-resolution ultrasound imaging : a narrative review. *The ultrasound journal*, 17 :29, 06 2025.
- [60] Amy Ginsburg, Zylee Liddy, Parsa Khazaneh, Susanne May, and Farhan Pervaiz. A survey of barriers and facilitators to ultrasound use in low- and middle-income countries. *Scientific Reports*, 13, 02 2023.
- [61] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 349–356, 2009.
- [62] Rafael C. Gonzalez and Paul Wintz. *Digital Image Processing*. Prentice Hall, 1st edition, 1980.
- [63] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 27. Curran Associates, Inc., 2014.
- [64] Joseph W. Goodman. Some fundamental properties of speckle. *Journal of the Optical Society of America*, 66(11) :1145–1150, 1976.
- [65] Hayit Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1) :43–63, 2009.
- [66] Hayit Greenspan, Gilad Oz, Nahum Kiryati, and Sharon Peled. Mri inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20(5) :437–446, 2002.

- [67] Shown Haluzani, Agripa Lungu, Mweetwa Mudenda, Arnold Hamapa, Catherine Mkan-dawire, Abraham Kaluba, Oswell Khondowe, Mushelelwa Mutale, Karren Imasiku, Muye-reka Nyirenda, Osbert Nkolola, Lackson Kasonka, Felix Simute, Angel Mwiche, Evarist Njelesani, Peter Mwaba, C. Viner Alexandra, David Lissauer, Mubambe Mulaya, Desire Kabamba, John Mathias Zulu, and Lydia Hangulu. Quality of ultrasound images by mid-wives implementing ultrasound scanning services during antenatal clinics within selected healthcare facilities of zambia. *WFUMB Ultrasound Open*, 2(2) :100057, 2024.
- [68] Gill Harrison and Edward Oates. *Clark's Essential Guide to Clinical Ultrasound*. CRC Press, Boca Raton, 2022.
- [69] Jiachuan He, He Ma, Miaoran Guo, Jiaqi Wang, Zhongqing Wang, and Guoguang Fan. Research into super-resolution in medical imaging from 2000 to 2023 : bibliometric ana-lysis and visualization. *Quantitative Imaging in Medicine and Surgery*, 14(7) :5109–5130, 2024.
- [70] Wayne R. Hedrick, Donald L. Hykes, Donald E. Starchman, and Christian Kollmann. *Diagnostic Ultrasound Imaging : Inside Out*. Academic Press, Oxford, 2 edition, 2014.
- [71] Y. Heng, M. Yinghua, G. Khan, F. et al. Survey : Application and analysis of generative adversarial networks in medical images. *Artificial Intelligence Review*, 58 :39, 2024.
- [72] Mohammad Hossin and Sulaiman M.N. A review on evaluation metrics for data classi-fication evaluations. *International Journal of Data Mining and Knowledge Management Process*, 5 :01–11, 03 2015.
- [73] Jabbar Hussain, Magnus Båth, and Jonas Ivarsson. Generative adversarial networks in medical image reconstruction : A systematic literature review. *Comput. Biol. Med.*, 191(C), June 2025.
- [74] Yvonne Innocent, Grace Mselle, Fitri Utami, and Alexis Havyarimana. Improving medi-cal radiology quality in low-income settings : an interdisciplinary approach to workflow management, management support, and working environment safety in burundi. *Journal of Global Health Science*, 7 :e13, Aug 2025.
- [75] Michelle Iskandar, Harvey Mannering, Zhanxiang Sun, Jacqueline Matthew, Hamideh Kerdegari, Laura Peralta, and Miguel Xochicale. Towards realistic ultrasound fetal brain imaging synthesis, 2023.
- [76] Showrov Islam, Md. Tarek Aziz, Hadiur Rahman Nabil, Jamin Rahman Jim, M. F. Mridha, Md. Mohsin Kabir, Nobuyoshi Asai, and Jungpil Shin. Generative adversa-rial networks (gans) in medical imaging : Advancements, applications, and challenges. *IEEE Access*, 12 :35728–35753, 2024.
- [77] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image transla-tion with conditional adversarial networks. *CVPR*, 2017.
- [78] Andreas Jansche, Patrick Krawczyk, Miguelangel Balaguera, Anoop Kini, Timo Bern-thaler, and Gerhard Schneider. Deep learning-based image super resolution methods in microscopy – a review. *Methods in Microscopy*, 2(2) :235–275, 2025.
- [79] Mingfeng Jiang, Minghao Zhi, Liying Wei, Xiaocheng Yang, Jucheng Zhang, Yongming Li, Pin Wang, Jiahao Huang, and Guang Yang. Fa-gan : Fused attentive generative adversarial networks for mri image super-resolution. *Computerized Medical Imaging and Graphics*, 92 :101969, 08 2021.



- [80] Zhencun Jiang, Kangrui Ren, Kefan Wang, and Zhongjie Wang. Blind super-resolution for handheld ultrasound image : Two-stage degradation based unpaired deep learning. *Computer Methods and Programs in Biomedicine*, 270 :108956, 2025.
- [81] Medina Kapo, Emir Buza, and Amila Akagic. Super-resolution in medical imaging : Evaluating srcnn, edsr, espcn, and srgan on dermatological images. In Naida Ademović, Zlatan Akšamija, and Almir Karabegović, editors, *Advanced Technologies, Systems, and Applications X*, volume 1625 of *Lecture Notes in Networks and Systems*. Springer, Cham, 2026.
- [82] Takayuki Katsuki, Akira Torii, and Masato Inoue. Posterior-mean super-resolution with a causal gaussian markov random field prior. *IEEE Transactions on Image Processing*, 21(9) :4008–4021, 2012.
- [83] Salome Kazeminia, Christoph Baur, Arjan Kuijper, Bram van Ginneken, Nassir Navab, Shadi Albarqouni, and Anirban Mukhopadhyay. Gans for medical image analysis. *Artificial Intelligence in Medicine*, 109 :101938, 2020.
- [84] Robert G. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6) :1153–1160, 1981.
- [85] Davit Khaledyan, Thomas J. Marini, T. M. Baran, Anne O’Connell, and Kevin Parker. Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques : Sharp attention UNet. *PLoS ONE*, 18(12) :e0289195, December 2023.
- [86] Eunsoo Kim, Kavita Singh, Allisyn Moran, Deborah Armbruster, and Naoko Kozuki. Obstetric ultrasound use in low and middle income countries : A narrative review. *Reproductive Health*, 15, 07 2018.
- [87] Hee E. Kim, Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E. Maros, and Thomas Ganslandt. Transfer learning for medical image classification : a literature review. *BMC Medical Imaging*, 22(1) :69, 2022.
- [88] Jeong Seok Kim. Improved image resolution during zooming in ultrasound image using deep learning technique. In *2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–3, 2020.
- [89] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016.
- [90] Jonghee Kim, Chanhoo Jung, and Changick Kim. Dual back-projection-based internal learning for blind super-resolution. *IEEE Signal Processing Letters*, 27 :1190–1194, 2020.
- [91] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. *arXiv :2304.02643*, 2023.
- [92] Padmavathi Kora, Chui Ping Ooi, Oliver Faust, U. Raghavendra, Anjan Gudigar, Wai Yee Chan, K. Meenakshi, K. Swaraja, Pawel Plawiak, and U. Rajendra Acharya. Transfer learning techniques for medical image analysis : A review. *Biocybernetics and Biomedical Engineering*, 42(1) :79–107, 2022.

- [93] Cansu Korkmaz and A. Murat Tekalp. Training transformer models by wavelet losses improves quantitative and visual performance in single image super-resolution. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CV-PRW)*, June 2024.
- [94] Frank W. Kremkau. *Sonography : Principles and Instruments*. Elsevier Health Sciences, Philadelphia, PA, 9th edition, 2015.
- [95] Frederick W. Kremkau. *Diagnostic Ultrasound : Principles and Instruments*. Elsevier, Philadelphia, PA, USA, 2015.
- [96] Nalin Kumar and M. Nachamai. Noise removal and filtering techniques used in medical images. *Oriental Journal of Computer Science and Technology*, 10(1) :138–146, 2017.
- [97] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1) :159–174, 1977.
- [98] Sarah Leclerc, Erik Smistad, João Pedrosa, Andreas Østvik, Frédéric Cervenansky, Florian Espinosa, Torvald Espeland, Erik A. R. Berg, Pierre-Marc Jodoin, Thomas Grenier, et al. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, 38(9) :2198–2210, 2019.
- [99] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017.
- [100] Jae Yong Lee, Muhammad Ishfaq Hussain, Kyong-Ha Lee, Hyoung Seop Shim, Seung-Ho Han, and Donghun Yang. Transfer learning-based super-resolution for high-precision medical imaging. *IEEE Access*, 13 :124776–124789, 2025.
- [101] Y.Y.P. Lee, Kum Thong Wong, Ann D. King, and Anil T. Ahuja. Imaging of salivary gland tumours. *European journal of radiology*, 66 3 :419–36, 2008.
- [102] Xin Li and Michael T. Orchard. New edge-directed interpolation. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 311–314, 2001.
- [103] Y. Li, Bruno Sixou, and Françoise Peyrin. A review of the deep learning methods for medical images super resolution problems. *IRBM*, 42(2) :120–133, 2021.
- [104] Yunhe Li, Mei Yang, Tao Bian, Wenhui Xia, and Haitao Wu. Super-resolution analysis of CT images based on HAT and self-integrated fusion method. *International Journal of Big Data Intelligent Technology*, 5(1) :153–160, 2024.
- [105] Jingyun Liang, Jiezhong Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir : Image restoration using swin transformer. In *ICCV Workshops (AIM)*, pages 1833–1844, 2021.
- [106] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops (NTIRE)*, pages 136–144, 2017.

- [107] Xingtao Lin, Xiahai Zhuang, Lin Pan, Mingjing Yang, Liqin Huang, Shun Chen, and Lei Li. Zsg-net : A zero-shot super-resolution guided network for ultrasound image segmentation and classification. *IEEE Journal of Biomedical and Health Informatics*, 29(11) :8320–8333, 2025.
- [108] L'interniste. Cours us principes de base - doppler continu. <https://echo.interniste.com/echographie-doppler-continu-us-fr-c0-s4>, 2021. Consulté le 2 septembre 2025.
- [109] L'interniste. Cours us principes de base - doppler couleur. <https://echo.interniste.com/...>, 2021. Consulté le 2 septembre 2025.
- [110] L'interniste. Cours us principes de base - doppler pulsé. <https://echo.interniste.com/...>, 2021. Consulté le 2 septembre 2025.
- [111] L'interniste. Cours us principes de base - doppler tissulaire. <https://echo.interniste.com/...>, 2021. Consulté le 2 septembre 2025.
- [112] Heng Liu, Jianyong Liu, Feng Chen, and Caifeng Shan. Progressive residual learning with memory upgrade for ultrasound image blind super-resolution. *IEEE Journal of Biomedical and Health Informatics*, 26(9) :4390–4401, 2022.
- [113] Heng Liu, Jianyong Liu, Shudong Hou, Tao Tao, and Jungong Han. Perception consistency ultrasound image super-resolution via self-supervised cyclegan. *Neural Computing and Applications*, 35, 2021.
- [114] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhuliang Xie, Yixuan Wei, Jinghao Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin transformer v2 : Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12009–12019, June 2022.
- [115] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *arXiv preprint arXiv :2201.03545*, 2022.
- [116] Frederic L. Lizzi, Ernest J. Feleppa, S. Kaisar Alam, and Cheri X. Deng. Ultrasonic spectrum analysis for tissue evaluation. *Pattern Recogn. Lett.*, 24(4–5) :637–658, February 2003.
- [117] Themis Loupas, Richard W. Gill, and John T. Powers. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 36(3) :232–241, 1989.
- [118] Yuchao Lyu, Yinghao Xu, Cuiyun Yang, Xi Jiang, Xiaoyan Zhao, and Xijun Zhu. Wsr-gan : A wavelet-based gan for super-resolution of plane-wave ultrasound images without sampling loss. *Engineering Applications of Artificial Intelligence*, 127 :107384, 2024.
- [119] Alhassan Mabrouk, Abdelghani Dahou, Mohamed Abd Elaziz, Rebeca P. Díaz Redondo, and Mohammed Kayed. Medical image classification using transfer learning and chaos game optimization on the internet of medical things. *Computational Intelligence and Neuroscience*, 2022 :1–22, July 2022.
- [120] Amirreza Mahbod, Gerald Schaefer, Chunliang Wang, Georg Dorffner, Rupert Ecker, and Isabella Ellinger. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Computer Methods and Programs in Biomedicine*, 193 :105475, 03 2020.

- [121] Carlo Martinoli. Musculoskeletal ultrasound : technical guidelines. *La Radiologia Medica*, 115(5) :827–840, 2010.
- [122] Christos Matsoukas, Johan Fredin Haslum, Moein Sorkhei, Magnus Söderberg, and Kevin Smith. What makes transfer learning work for medical images : Feature reuse & other factors. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9215–9224, 2022.
- [123] Cynthia H. McCollough, Anthony C. Bartley, Rickey E. Carter, Baiyu Chen, Timothy A. Drees, Paul Edwards, David R. Holmes, Alice E. Huang, Fauzia Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions : results from the AAPM low-dose ct grand challenge. *Medical Physics*, 44(10) :e339–e352, 2017.
- [124] Mediace / Samsung Medison. Ultrasound image no. 37 : Fetal head in b-mode (convex probe 3–7 mhz). Online ; available at <https://www.medison.ru/>, 1997. Image (c) 1997–2025 Mediace / Samsung Medison. All rights reserved.
- [125] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10) :1993–2024, 2015.
- [126] Massimo Mischi, Muyinatu A. Lediju Bell, Ruud J. G. van Sloun, and Yonina C. Eldar. Deep learning in medical ultrasound—from image formation to image analysis. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12) :2477–2480, 2020.
- [127] Don P Mitchell and Arun N Netravali. Reconstruction filters in computer graphics. *ACM SIGGRAPH Computer Graphics*, 22(4) :221–228, 1988.
- [128] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3) :209–212, 2013.
- [129] Christopher L. Moore and Jeffrey A. Copel. Point-of-care ultrasonography. *New England Journal of Medicine*, 364 :749–757, 2011.
- [130] Mohammad Amin Morid, Alireza Borjali, and Guilherme Del Fiol. A scoping review of transfer learning research on medical image analysis using imagenet. *Computers in Biology and Medicine*, 128 :104115, 2021.
- [131] Rafid Mostafiz, Mohammad Motiur Rahman, A. Islam, and Saeid Belkasim. Focal liver lesion detection in ultrasound image using deep feature fusions and super resolution. *Machine Learning and Knowledge Extraction*, 2, 2020.
- [132] Basil Mustafa, Aaron Loh, Jan Freyberg, Patricia MacWilliams, Megan Wilson, Scott Mayer McKinney, Marcin Sieniek, Jim Winkens, Yuan Liu, Peggy Bui, Shruthi Prabhakara, Umesh Telang, Alan Karthikesalingam, Neil Houlsby, and Vivek Natarajan. Supervised transfer learning at scale for medical imaging, 2021.
- [133] P. Nandal, Sudesh Pahal, Ashish Khanna, and Plácido Pinheiro. Super-resolution of medical images using real esrgan. *IEEE Access*, PP :1–1, 01 2024.
- [134] Goodluck U. Nneji, Jian Cai, Hammed N. Monday, Md. Azharul Hossin, Shamsun Nahar, Godwin T. Mgbejime, and Jun Deng. Fine-tuned siamese network with modified enhanced super-resolution gan plus based on low-quality chest x-ray images for covid-19 identification. *Diagnostics*, 12(3) :717, March 2022.

- [135] Biomedical Journal of Scientific & Technical Research. Ultrasound in medicine, 2020. Consulté le 5 septembre 2025.
- [136] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention u-net : Learning where to look for the pancreas. *arXiv preprint arXiv :1804.03999*, 2018.
- [137] Christopher Oliver and Shaun Quegan. *Understanding Synthetic Aperture Radar Images*. SciTech Publishing, Raleigh, NC, USA, 2004.
- [138] OpenMedScience. The role of ultrasound in medical imaging, 2023. Consulté le 5 septembre 2025.
- [139] Orville. Ultrasound fetus dataset, 2023.
- [140] Aditi Parashari, Kirit Jain, and Sanjeev Agrawal. Image super-resolution using deep learning : A review for medical imaging. *International Journal of Emerging Technologies and Innovative Research*, 12(6) :i23–i28, June 2025. UGC approved journal.
- [141] Suh Park, M K Park, and Moon Gi Kang. Super-resolution image reconstruction : a technical overview. *IEEE Signal Processing Magazine*, 20(3) :21–36, 2003.
- [142] François Patat and Marc Ronot. Fondements de l'échographie. In *Les fondamentaux de l'imagerie médicale*, chapter 7. CNP–MN, 2025. Disponible en ligne : <https://www.cnp-mn.fr/les-fondamentaux-chapitre-7-echographie/>, consulté le 3 septembre 2025.
- [143] Vaishali Patel and Anand Mankodia. Deep learning in medical image super-resolution : A survey. *International Journal of Engineering Trends and Technology*, 2023.
- [144] Nicholas Peterman, Eunhae Yeo, Bradley Kaptur, Emily Smith, Anton Christensen, Edward Huang, and Mehmoodur Rasheed. Analysis of rural disparities in ultrasound access. *Cureus*, 14, May 2022.
- [145] Vladimir Petrovic and Costas Xydeas. Objective image fusion performance characterisation. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV 2005)*, volume 2, pages 1866–1871, 2005.
- [146] Stefano Piffer, Leonardo Ubaldi, Sabina Tangaro, Alessandra Retico, and Cinzia Talamonti. Tackling the small data problem in medical image classification with artificial intelligence : a systematic review. *Progress in Biomedical Engineering*, 6(3) :032001, jun 2024.
- [147] Hilde Pol, Lennard Karnenbeek, Mark Wijkhuizen, Freija Geldof, and Behdad Dashtbozorg. Deep learning for point-of-care ultrasound image quality enhancement : A review. *Applied Sciences*, 14 :7132, 08 2024.
- [148] Bharath Srinivas Prabakaran, Paul Hamelmann, Erik Ostrowski, and Muhammad Shafique. Fpus23 : An ultrasound fetus phantom dataset with deep neural network evaluations for fetus orientations, fetal planes, and anatomical features. *arXiv preprint arXiv :2303.07852*, 2023.
- [149] Chundru G.V.N. Prasad, Kandula Kavya, and Manasa Goud. Evaluating the impact of transfer learning vs fine-tuning in medical image segmentation. *Global Journal of Engineering Innovations & Interdisciplinary Research*, 5(1) :19, 2025.

- [150] Muralikrishna Puttagunta, Ravi Subban, and Nelson Kennedy Babu. Swinir transformer applied for medical image super-resolution. *Procedia Computer Science*, 204 :907–913, 2022.
- [151] Nitin Rane. Transformers for medical image analysis : Applications, challenges, and future scope. *SSRN Electronic Journal*, 01 2023.
- [152] Bryan J. Ranger, Elizabeth Bradburn, Qingchao Chen, Micah Kim, and J. Alison Noble. Portable ultrasound devices for obstetric care in resource-constrained environments : mapping the landscape. *Gates Open Research*, 5 :150, 2021.
- [153] K Lokeshwar Reddy and G Bharathi Mohan. Enhancing medical imaging : Noise reduction and super resolution with transfer learning. In *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pages 1430–1435, 2024.
- [154] Ricardo Reyes-Reyes, Y. G. Mora-Martinez, B. P. Garcia-Salgado, Volodymyr Ponomaryov, J. A. Almaraz-Damian, C. Cruz-Ramos, and Sergey Sadovnychiy. A robust system for super-resolution imaging in remote sensing via attention-based residual learning. *Mathematics*, 13(15) :2400, 2025.
- [155] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP :1–14, 09 2022.
- [156] Irina Sanchez and Veronica Vilaplana. Brain mri super-resolution using 3d generative adversarial networks. *arXiv preprint arXiv :1812.11440*, 2018.
- [157] Chompunuch Sarasaen, Soumick Chatterjee, Mario Bretkopf, Georg Rose, Andreas Nürnberger, and Oliver Speck. Fine-tuning deep learning model parameters for improved super-resolution of dynamic mri with prior-knowledge. *Artificial Intelligence in Medicine*, 121 :102196, 2021.
- [158] Farook Sattar, Lars Floreby, Göran Salomonsson, and Bengt Löfström. Image enhancement based on a nonlinear multiscale method. *IEEE Transactions on Image Processing*, 6(6) :888–895, 1997.
- [159] Edgar Schonfeld, Bernt Schiele, and Anna Khoreva. A u-net based discriminator for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2020.
- [160] Samuel Schuler, Christian Leistner, and Horst Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3791–3799, 2015.
- [161] Carla Sendra-Balcells, Víctor M. Campello, Jordina Torrents-Barrena, Yahya Ali Ahmed, Mustafa Elattar, Benard Ohene-Botwe, Pempho Nyangulu, William Stones, Mohammed Ammar, Lamya Nawal Benamer, Harriet Nalubega Kisebo, Senai Goitom Sereke, Sikolia Z. Wanyonyi, Marleen Temmerman, Eduard Gratacós, Elisenda Bonet, Elisenda Eixarch, Kamil Mikolaj, Martin Grønnebak Tolsgaard, and Karim Lekadir. Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five african countries. *Scientific Reports*, 13(1) :2728, 2023.

- [162] Ajay Sharma and Bhavana P. Shrivastava. Complex wavelet transform with progressive network for medical imaging super resolution. *Multimedia Tools and Applications*, 84(13) :12061–12079, 2025. Published online : 28 May 2024.
- [163] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3118–3126, 2018.
- [164] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2014.
- [165] Kajol Singh and Manish Saxena. A review on medical image super resolution with application of deep learning. *Smart Moves Journal of LIOSCIENCE*, 2021.
- [166] Mayank Kumar Singh, Indu Saini, and Neetu Sood. Sub-renet : A wavelet-based network for super resolution of diagnostic ultrasound. *Signal, Image and Video Processing*, 18 :1–13, May 2024.
- [167] Sanyogita Sippel, Kiran Muruganandan, Adam Levine, and Sagar Shah. Review article : Use of ultrasound in the developing world. *International Journal of Emergency Medicine*, 4(1) :72, December 2011. PMID : 22152055 ; PMCID : PMC3285529.
- [168] Francesco Sogni, Simone Pilloni, Lorenzo Ciavola, Claudia Infantino, Cristiano Conte, Valentina Fainardi, and Susanna Esposito. Point-of-care ultrasound in pediatric emergency medicine : Narrative review on current applications, limitations, and future directions. *Acta Biomedica*, 97(1) :18209, 2026.
- [169] Ke Song, Jing Feng, and Duo Chen. A survey on deep learning in medical ultrasound imaging. *Frontiers in Physics*, 12, July 2024.
- [170] Pedro Sousa, Henrique Sousa, Tania Pereira, Eva Batista, Pedro Gouveia, and Hélder P. Oliveira. Swin transformer applied to breast mri super-resolution in a cross-cohort dataset. In *2025 IEEE 38th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6, 2025.
- [171] Kelsey A. Stewart, Sergio M. Navarro, Sriharsha Kambala, Gail Tan, Revanth Poondla, Sara Lederman, Kelli Barbour, and Chris Lavy. Trends in ultrasound use in low and middle income countries : A systematic review. *International Journal of Maternal and Child Health and AIDS*, 9(1) :103–120, 2020.
- [172] Simon Strauss, Ella Gavish, Paul Gottlieb, and Ludmila Katsnelson. Interobserver and intraobserver variability in the sonographic assessment of fatty liver. *American Journal of Roentgenology*, 189(6) :W320–W323, 2007.
- [173] Sayma Suha and Rifat Shahriyar. Deep learning for fetal brain imaging : A systematic review and framework towards privacy-preserving neurodevelopmental informatics. *Neuroscience Informatics*, 5 :100241, 10 2025.
- [174] Thomas L. Szabo. *Diagnostic Ultrasound Imaging : Inside Out*. Academic Press, Oxford, UK, 2nd edition, 2014.
- [175] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *CVPR*, 2015.

- [176] R. Tambde, C. Muntean, and A. Yaqoob. Polycystic ovary syndrome detection utilizing srgan-generated synthetic images and advanced cnn models. In *Proceedings of the International Conference on Advanced Computing and Data Science Applications (ACDSA)*, pages 1–6, 2025.
- [177] Lingxiao Tang. A comparative analysis of transfer learning methods for medical super-resolution. *Highlights in Science, Engineering and Technology*, 120 :14–20, 12 2024.
- [178] Zahra TehraniNasab, Amar Kumar, and Tal Arbel. Pixels under pressure : Exploring fine-tuning paradigms for foundation models in high-resolution medical imaging, 2025.
- [179] Hakan Temiz and Hasan S. Bilge. Super resolution of b-mode ultrasound images with deep learning. *IEEE Access*, 8 :78808–78820, 2020.
- [180] Arbin Thapaliya, Alec Sithole, Michael Welsh, and Gaston Dana. *Ultrasound Physics and its Application in Medicine*. Pressbooks, 2022. Creative Commons Attribution License. Available online.
- [181] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 17(1) :168–192, Jan 2021.
- [182] Maria Trigka and Elias Dritsas. A comprehensive survey of deep learning approaches in image processing. *Sensors*, 25(2), 2025.
- [183] Theresa A. Tuthill, Robert H. Sperry, and Kevin J. Parker. Deviations from rayleigh statistics in ultrasonic speckle. *Ultrasonic Imaging*, 10(2) :81–89, 1988.
- [184] Sabina Umirzakova, Shabir Ahmad, Latif U. Khan, and Taegkeun Whangbo. Medical image super-resolution for smart healthcare applications : A comprehensive survey. *Information Fusion*, 103 :102075, 2024.
- [185] Juan Miguel Valverde, Vandad Imani, Ali Abdollahzadeh, Riccardo De Feo, Mithilesh Prakash, Robert Ciszek, and Jussi Tohka. Transfer learning in magnetic resonance brain imaging : A systematic review. *Journal of Imaging*, 7(4), 2021.
- [186] Thomas L. A. van den Heuvel, Dagmar de Bruijn, Chris L. de Korte, and Bram van Ginneken. Hc (head circumference) dataset. Zenodo, 2018.
- [187] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30. Curran Associates, Inc., 2017.
- [188] Robert F. Wagner, Stephen W. Smith, J. Michael Sandrik, and Hector Lopez. Statistics of speckle in ultrasound b-scans. *IEEE Transactions on Sonics and Ultrasonics*, 30(3) :156–163, 1983.
- [189] Jiangang Wang, Qingnan Fan, Qi Zhang, Haigen Liu, Yuhang Yu, Jinwei Chen, and Wenqi Ren. Hero-sr : One-step diffusion for super-resolution with human perception priors, 2024.
- [190] Kaifeng Wang, Lixia Qi, Tingxin Zhang, Jing Li, Meilan Zhang, and Hai Du. Deep transfer learning-based 3d super-resolution radiomics modeling of ct to discriminate brucella spondylitis from tuberculous spondylitis : A multicenter, retrospective study, 05 2024.



- [191] Xintao Wang, Kelvin C.K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr : Video restoration with enhanced deformable convolutional networks. In *CVPR Workshops (NTIRE)*, pages 1954–1963, 2019.
- [192] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan : Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1905–1914, 2021.
- [193] Xintao Wang, Liangbin Xie, Ke Yu, Kelvin C.K. Chan, Chen Change Loy, and Chao Dong. BasicSR : Open source image and video restoration toolbox. <https://github.com/XPixelGroup/BasicSR>, 2022.
- [194] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, and Yu Qiao. Esrgan : Enhanced super-resolution generative adversarial networks. In *ECCV Workshops*, pages 63–79, 2018.
- [195] Zihao Wei, Yidong Huang, Yuang Chen, Chenhao Zheng, and Jinnan Gao. A-esrgan : Training real-world blind super-resolution with attention u-net discriminators, 2021.
- [196] Peter Neil Temple Wells. Ultrasound imaging. *Physics in Medicine and Biology*, 51(13) :R83–R98, 2006.
- [197] World Health Organization. Who compendium of innovative health technologies for low-resource settings, 2016. Consulté le 5 septembre 2025.
- [198] World Health Organization. Who recommendations on antenatal care for a positive pregnancy experience, 2019. Consulté le 5 septembre 2025.
- [199] World Health Organization. Strengthening medical imaging, 2023. Consulté le 5 septembre 2025.
- [200] Zengbo Xu and Yiheng Wei. Super-resolution reconstruction of sonograms using residual dense conditional generative adversarial network. *Sensors*, 25 :6694, 2025.
- [201] Harika Yadav, Devanshi Shah, Shahin Sayed, Susan Horton, and Lee F. Schroeder. Availability of essential diagnostics in ten low-income and middle-income countries : results from national health facility surveys. *The Lancet Global Health*, 9(11) :e1553–e1560, 2021.
- [202] Shouang Yan, Chengyan Wang, Weibo Chen, and Jun Lyu. Swin transformer-based gan for multi-modal medical image translation. *Frontiers in Oncology*, 12, 2022.
- [203] Jianchao Yang, John Wright, Thomas S. Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11) :2861–2873, 2010.
- [204] Xirui Yang, Yue Chen, Rui Tao, Yue Zhang, Zhiwen Liu, and Yonggang Shi. Endoscopic image deblurring and super-resolution reconstruction based on deep learning. In *2020 International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, pages 168–172, 2020.
- [205] Jin Ye, Junlong Cheng, Jianpin Chen, Zhongying Deng, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyang Huang, Jilong Chen, Lei Jiang, Hui Sun, Min Zhu, Shaoting Zhang, Junjun He, and Yu Qiao. Sa-med2d-20m dataset : Segment anything in 2d medical imaging with 20 million masks, 2023.

- [206] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021.
- [207] Kai Zhang, Wangmeng Zuo, and Lei Zhang. A deep plug-and-play super-resolution framework for blindly degraded images. *IEEE Transactions on Image Processing*, 28(12) :6392–6403, 2019.
- [208] Yaping Zhang, Ai Wang, Qingyao Li, Lu Zhang, Lingyun Wang, Zhijie Pan, Yanfei Hu, and Xueqian Xie. Comparison of conventional 512-matrix ct images with swin2sr-based 2048-matrix images in the visualization and diagnosis of lung nodules. *Japanese journal of radiology*, 01 2026.
- [209] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*, pages 286–301, 2018.
- [210] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1) :47–57, 2017.
- [211] Pei J. Zhao. Ultrasound in the rural and remote healthcare setting. *University of Western Ontario Medical Journal*, 83(1) :30–32, 2014.
- [212] Li-Qin Zhou, Jing-Yu Wang, Shu-Yan Yu, Guo-Guang Wu, Qian Wei, Yu-Bao Deng, Xiao-Li Wu, Xiao-Wei Cui, and Christoph F. Dietrich. Artificial intelligence in medical imaging of the liver. *World Journal of Gastroenterology*, 25(6) :672–682, 2019.
- [213] S. Kevin Zhou, Hayit Greenspan, Christos Davatzikos, James S. Duncan, Bram Van Ginneken, Anant Madabhushi, Jerry L. Prince, Daniel Rueckert, and Ronald M. Summers. A review of deep learning in medical imaging : Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5) :820–838, 2021.
- [214] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, and Jianming Liang. Fine-tuning convolutional neural networks for biomedical image analysis : Actively and incrementally. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4761–4772, 2017.
- [215] Jin Zhu, Chuan Tan, Junwei Yang, Guang Yang, and Pietro Lio’. Miassr : An approach for medical image arbitrary scale super-resolution. *International Journal of Neural Systems*, 31(10), 2021.

# Annexes

## Annexe A : [Fonctions de coût et architectures associées]

Le réseau convolutif VGG-19, proposé par Simonyan et Zisserman [164], comprend 16 couches convolutionnelles suivies de 3 couches entièrement connectées. Sa conception repose sur des filtres convolutifs  $3 \times 3$  empilés et des opérations de *max-pooling*, permettant une réduction progressive des dimensions spatiales. Pré-entraîné sur ImageNet, ce réseau constitue une base robuste pour l'extraction de caractéristiques visuelles hiérarchiques.

### 5.5.1 Mécanisme d'extraction des caractéristiques

Les activations issues des couches profondes de VGG-19 (ex. `conv4_4`, `conv5_4`) sont utilisées comme représentations perceptuelles, permettant d'évaluer la similarité de contenu entre deux images.

### 5.5.2 Formalisation de la perte perceptuelle

La perte perceptuelle repose sur l'écart entre ces représentations extraites :

$$\mathcal{L}_{\text{perc}} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \|\phi_\ell(I_{SR})(i) - \phi_\ell(I_{HR})(i)\|_2^2,$$

où  $\phi_\ell(\cdot)$  désigne les activations à la couche  $\ell$ .

### 5.5.3 Définitions des fonctions de perte utilisées (voir Chapitre 5)

#### A.1 L1 Loss (Mean Absolute Error)

$$\mathcal{L}_{\text{L1}} = \frac{1}{N} \sum_{i=1}^N |I_{SR}(i) - I_{HR}(i)|.$$

#### A.2 SWT Loss multi-échelle

$$\mathcal{L}_{\text{SWT}} = \sum_{j=1}^J \sum_{k \in \{H,V,D\}} \frac{1}{N_j} \sum_{i=1}^{N_j} |W_{SR}^{(j,k)}(i) - W_{HR}^{(j,k)}(i)|.$$

#### A.3 GAN Loss (LSGAN)

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{I_{HR}} [(D(I_{HR}) - 1)^2] + \mathbb{E}_{I_{SR}} [(D(I_{SR}))^2].$$

### 5.5.4 Métriques Basées sur la Matrice de Confusion

Les métriques de performance reposent sur la matrice de confusion (TP, FP, TN et FN), évaluée au niveau des pixels pour la segmentation et au niveau des instances pour la classification.

| Métrique                          | Formule                                                                                      | Description                                                                                     |
|-----------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| Exactitude (Accuracy)             | $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$                                        | = Proportion de prédictions correctes (positives et négatives) sur l'ensemble des échantillons. |
| Précision (Precision, VPP)        | $\text{Precision} = \frac{TP}{TP + FP}$                                                      | Fraction d'éléments prédits positifs qui sont réellement positifs.                              |
| Rappel (Recall, Sensibilité, TVP) | $\text{Recall} = \frac{TP}{TP + FN}$                                                         | Fraction d'éléments positifs réels correctement identifiés par le modèle.                       |
| Spécificité (Specificity, TVN)    | $\text{Specificity} = \frac{TN}{TN + FP}$                                                    | Capacité du modèle à identifier correctement les échantillons négatifs.                         |
| F1-Score                          | $F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$ | Moyenne harmonique de la Précision et du Rappel. Utile lorsque les classes sont déséquilibrées. |

TABLEAU 5.10 – Métriques d'évaluation pour la classification binaire et la segmentation.

## Annexe B : [Détails des analyses statistiques]

### 5.5.5 Méthodologie générale

Cette annexe présente les tests statistiques appliqués pour comparer les modèles de super-résolution échographique (PSNR, SSIM, NIQE, etc.). Le choix des tests dépend de : (i) nombre de groupes à comparer, (ii) normalité des données (test de Shapiro–Wilk), (iii) homogénéité des variances (test de Levene).

### 5.5.6 Analyses descriptives

- Moyennes, écarts-types, médianes, IQR ;
- Comparaisons inter-pays et inter-modèles.

### 5.5.7 Analyses inférentielles

- Student t-test, Welch t-test, Mann–Whitney pour 2 groupes.
- ANOVA, Welch ANOVA, Kruskal–Wallis (+ tests post-hoc) pour  $\geq 3$  groupes.

### 5.5.8 Taille d’effet

- Cohen’s  $d$  pour 2 groupes.
- $\eta^2$  et Cliff’s  $\delta$  pour  $\geq 3$  groupes.

### 5.5.9 Corrélations entre métriques

Corrélations de Pearson entre PSNR, SSIM, LPIPS pour estimer la redondance ou complémentarité.

### 5.5.10 Résumé des indicateurs

| Indicateur       | Ce qu’il mesure                           | Interprétation                                |
|------------------|-------------------------------------------|-----------------------------------------------|
| p-value          | Probabilité que l’écart soit dû au hasard | $p < 0.05$ : significatif                     |
| $\eta^2$         | Variance expliquée par les groupes        | 0.01 faible, 0.06 moyen, 0.14 fort            |
| Cliff’s $\delta$ | Différence d’ordres entre 2 groupes       | seuils : 0.147 faible, 0.33 moyen, 0.474 fort |
| IC bootstrap     | Intervalle de confiance                   | Ex : [27.8 ; 28.2] à 95%                      |

TABLEAU 5.11 – Résumé des indicateurs statistiques utilisés.

### 5.5.11 Cadre d’évaluation statistique ( Corrélation/Variance)

L’adéquation entre la mesure objective et la vérité terrain est évaluée selon deux axes complémentaires : la linéarité et la variance expliquée.

*square* **Analyse de la linéarité ( $r$  et  $p$ -value)** Le coefficient de Pearson sert d’estimateur de la dépendance linéaire entre les deux variables. L’obtention d’un  $r = 0,876$  témoigne d’une excellente linéarité, signifiant que la métrique suit fidèlement les nuances de qualité perçues par le clinicien. La significativité associée ( $p < 0,01$ ) renforce la crédibilité de cet indicateur pour notre application.

*square* **Analyse de la variance** ( $R^2$ ) Le coefficient  $R^2$  traduit la capacité de la GMSD à modéliser les scores experts. En expliquant 76,7 % de la variabilité des réponses, cette métrique s'impose comme un substitut numérique fiable. Elle réduit la variance inter-individuelle inhérente aux scores cliniques par l'adoption d'un cadre d'évaluation automatique. La fraction de variance non modélisée (23,3 %) représente l'écart type résiduel inhérent à toute évaluation sensorielle humaine.

## Annexe C : [Résultats complémentaires par dataset]

TABLEAU 5.12 – Performance moyenne des modèles sur tous les jeux de données étudiés. PSNR/SSIM élevés = meilleure fidélité ; LPIPS/NIQE/GMSD/FID faibles = meilleure qualité perceptuelle.

| Jeu     | Modèle        | Architecture | PSNR  | SSIM   | LPIPS  | FID    | NIQE | GMSD   |
|---------|---------------|--------------|-------|--------|--------|--------|------|--------|
| AFRICA  | A-ESRGAN (1D) | GAN          | 27.49 | 0.7752 | 0.1829 | 130.27 | 5.3  | 0.2344 |
|         | A-ESRGAN (MD) | GAN          | 27.12 | 0.7491 | 0.1799 | 107.78 | 4.4  | 0.2340 |
|         | BSRGAN        | GAN          | 28.10 | 0.7671 | 0.1569 | 13.40  | 4.9  | 0.2243 |
|         | Real-ESRGAN   | GAN          | 28.34 | 0.7951 | 0.1454 | 73.78  | 5.4  | 0.2281 |
|         | HAT           | Transformer  | 27.69 | 0.7580 | 0.2254 | 68.31  | 7.0  | 0.2434 |
|         | Swin2SR       | Transformer  | 29.36 | 0.8295 | 0.2048 | 92.57  | 7.7  | 0.2392 |
|         | SwinIR-M      | Transformer  | 27.16 | 0.7445 | 0.1250 | 74.76  | 4.5  | 0.2241 |
|         | SwinIR-L      | Transformer  | 28.05 | 0.7819 | 0.1256 | 59.48  | 4.4  | 0.2241 |
| FPUS23  | A-ESRGAN (1D) | GAN          | 30.12 | 0.7709 | 0.2393 | 171.70 | 6.7  | 0.3036 |
|         | A-ESRGAN (MD) | GAN          | 30.26 | 0.7676 | 0.2426 | 64.30  | 4.5  | 0.2990 |
|         | BSRGAN        | GAN          | 31.63 | 0.7759 | 0.2141 | 125.49 | 5.7  | 0.2952 |
|         | Real-ESRGAN   | GAN          | 31.65 | 0.7814 | 0.2218 | 127.29 | 5.9  | 0.3092 |
|         | HAT           | Transformer  | 33.90 | 0.8390 | 0.1697 | 38.86  | 6.8  | 0.2322 |
|         | Swin2SR       | Transformer  | 33.10 | 0.8173 | 0.2137 | 83.48  | 7.2  | 0.2916 |
|         | SwinIR-M      | Transformer  | 28.72 | 0.7217 | 0.2063 | 77.39  | 5.2  | 0.2420 |
|         | SwinIR-L      | Transformer  | 31.58 | 0.7808 | 0.1619 | 60.15  | 4.2  | 0.2507 |
| PSFHS   | A-ESRGAN (1D) | GAN          | 24.21 | 0.7356 | 0.2155 | 143.51 | 6.5  | 0.2370 |
|         | A-ESRGAN (MD) | GAN          | 24.93 | 0.7498 | 0.2083 | 142.17 | 6.6  | 0.2280 |
|         | BSRGAN        | GAN          | 26.99 | 0.7853 | 0.1975 | 123.85 | 7.0  | 0.2161 |
|         | Real-ESRGAN   | GAN          | 26.74 | 0.7867 | 0.2030 | 127.03 | 7.4  | 0.2243 |
|         | HAT           | Transformer  | 30.02 | 0.8450 | 0.2024 | 126.03 | 9.0  | 0.1935 |
|         | Swin2SR       | Transformer  | 29.18 | 0.8200 | 0.2715 | 175.73 | 9.8  | 0.2259 |
|         | SwinIR-M      | Transformer  | 26.77 | 0.7851 | 0.1646 | 87.48  | 8.4  | 0.2135 |
|         | SwinIR-L      | Transformer  | 27.53 | 0.8000 | 0.1683 | 96.06  | 8.4  | 0.2167 |
| ISTHMIC | A-ESRGAN (1D) | GAN          | 25.67 | 0.8240 | 0.1760 | 49.59  | 4.1  | 0.2062 |
|         | A-ESRGAN (MD) | GAN          | 25.92 | 0.8248 | 0.1690 | 33.59  | 3.8  | 0.2055 |
|         | BSRGAN        | GAN          | 27.07 | 0.8636 | 0.1397 | 27.37  | 4.9  | 0.1941 |
|         | Real-ESRGAN   | GAN          | 26.98 | 0.8752 | 0.1255 | 22.50  | 4.6  | 0.1845 |
|         | HAT           | Transformer  | 27.12 | 0.8853 | 0.1250 | 22.67  | 5.2  | 0.1844 |
|         | Swin2SR       | Transformer  | 28.61 | 0.9104 | 0.1861 | 25.04  | 7.7  | 0.1840 |
|         | SwinIR-M      | Transformer  | 26.93 | 0.8812 | 0.1250 | 23.31  | 5.0  | 0.1845 |
|         | SwinIR-L      | Transformer  | 26.94 | 0.8951 | 0.1256 | 20.80  | 5.7  | 0.1806 |
| SPAIN   | A-ESRGAN (1D) | GAN          | 27.11 | 0.8005 | 0.1940 | 75.04  | 5.3  | 0.2337 |
|         | A-ESRGAN (MD) | GAN          | 26.36 | 0.7681 | 0.2222 | 64.30  | 4.5  | 0.2449 |
|         | BSRGAN        | GAN          | 28.92 | 0.8300 | 0.2171 | 13.40  | 4.4  | 0.2157 |
|         | Real-ESRGAN   | GAN          | 26.74 | 0.8031 | 0.1919 | 35.83  | 5.4  | 0.2300 |
|         | HAT           | Transformer  | 35.49 | 0.9650 | 0.0590 | 2.35   | 8.2  | 0.1235 |
|         | Swin2SR       | Transformer  | 31.66 | 0.9321 | 0.1404 | 26.41  | 8.3  | 0.1714 |
|         | SwinIR-M      | Transformer  | 26.77 | 0.8078 | 0.2269 | 36.96  | 4.2  | 0.2245 |
|         | SwinIR-L      | Transformer  | 28.15 | 0.8430 | 0.1754 | 30.35  | 4.7  | 0.2110 |

## Annexe D : [Observations techniques sur l’architecture VGG-19 et les fonctions de coût]

### 5.5.12 Caractérisation de l’architecture VGG-19

Le réseau de neurones convolutifs VGG-19, initialement proposé par Simonyan et Zisserman (2014), comprend 16 couches convolutionnelles suivies de 3 couches entièrement connectées. Son design repose sur l’empilement de filtres convolutifs de petite taille ( $3 \times 3$ ), complétés par des opérations de *max-pooling*, ce qui permet une réduction progressive des dimensions spatiales. Pré-entraîné sur le vaste corpus ImageNet, VGG-19 a acquis des représentations hiérarchiques robustes, utilisées en particulier dans les fonctions perceptuelles.

### 5.5.13 Mécanisme d’extraction des caractéristiques

Les activations générées dans les couches intermédiaires et profondes de VGG-19 (ex. `conv4_4`, `conv5_4`) sont extraites et utilisées comme descripteurs de haut niveau. Elles servent à mesurer la similarité de contenu et de structure entre images, au-delà de la simple fidélité pixel à pixel.

### 5.5.14 Formalisation de la perte perceptuelle

La perte perceptuelle compare les activations d’images de référence et de leurs reconstructions :

$$\mathcal{L}_{\text{perc}} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \|\phi_\ell(I_{SR})(i) - \phi_\ell(I_{HR})(i)\|_2^2,$$

où  $\phi_\ell(\cdot)$  représente les activations extraites à la couche  $\ell$ .

### 5.5.15 Définitions des fonctions de perte utilisées (Chapitre 5)

Dans le cadre du fine-tuning des modèles HAT et SwinIR (voir Chapitre 5), plusieurs fonctions de perte complémentaires ont été mobilisées.

#### D.1 L1 Loss (Mean Absolute Error)

La L1 Loss est définie comme la moyenne des différences absolues entre l’image super-résolue  $I_{SR}$  et l’image de référence  $I_{HR}$  :

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^N |I_{SR}(i) - I_{HR}(i)|, \quad (5.4)$$

où  $N$  est le nombre total de pixels. Elle assure la fidélité locale tout en limitant le sur-lissage.

#### D.2 SWT Loss multi-échelle

La SWT Loss (*Stationary Wavelet Transform Loss*) [93] mesure les écarts entre coefficients d’ondelettes stationnaires des images reconstruite et de référence, à plusieurs échelles :

$$\mathcal{L}_{\text{SWT}} = \sum_{j=1}^J \sum_{k \in \{H, V, D\}} \frac{1}{N_j} \sum_{i=1}^{N_j} |W_{SR}^{(j,k)}(i) - W_{HR}^{(j,k)}(i)|. \quad (5.5)$$

Elle favorise la préservation des structures anatomiques fines et globales.



### D.3 Perceptual Loss (basée sur VGG-19)

La Perceptual Loss exploite VGG-19 comme extracteur de caractéristiques. Si  $\phi_\ell(\cdot)$  désigne les activations à une couche  $\ell$  :

$$\mathcal{L}_{\text{perc}} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \|\phi_\ell(I_{SR})(i) - \phi_\ell(I_{HR})(i)\|_2^2. \quad (5.6)$$

Elle capture la similarité perceptuelle en prenant en compte la structure et le contenu de haut niveau.

### D.4 GAN Loss (LSGAN)

Pour les modèles adversariaux (ex. SwinIR-M avec LSGAN), la perte adversariale est :

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}_{I_{HR}} \left[ (D(I_{HR}) - 1)^2 \right] + \mathbb{E}_{I_{SR}} \left[ (D(I_{SR}))^2 \right], \quad (5.7)$$

où  $D$  est le discriminateur. Cette fonction guide le générateur vers des reconstructions perceptuellement plus réalistes.

## Annexe E : [Fondements statistiques des analyses inférentielles : Correction de Bonferroni et contrôle du risque d’erreurs de type I]

Lorsqu’un grand nombre de tests statistiques sont effectués simultanément, le risque de détecter à tort une différence significative — appelé *erreur de type I* — augmente mécaniquement. Une erreur de type I correspond au rejet erroné de l’hypothèse nulle ( $H_0$ ) alors qu’elle est en réalité vraie. Autrement dit, on conclut à une différence entre deux modèles alors qu’elle résulte simplement du hasard.

Afin de limiter ce risque d’erreurs cumulées, une procédure de correction est nécessaire. La méthode de **Bonferroni** ajuste le seuil de significativité global  $\alpha$  (fixé classiquement à 0.05 [34]) en le divisant par le nombre total de comparaisons indépendantes  $N$  :

$$\alpha_{\text{corrigée}} = \frac{\alpha}{N}$$

où :

- $\alpha$  : seuil de significativité global choisi (souvent 0.05),
- $N$  : nombre total de comparaisons effectuées,
- $\alpha_{\text{corrigée}}$  : seuil ajusté pour chaque test individuel.

**Exemple :** Dans cette étude, les performances des modèles HAT et SWINIR-M ont été comparées sur cinq jeux de données à l’aide de cinq métriques (PSNR, SSIM, LPIPS, GMSD et NIQE), soit  $N = 25$  comparaisons au total. En appliquant la correction de Bonferroni avec un seuil global de  $\alpha = 0.05$ , le seuil corrigé devient :

$$\alpha_{\text{corrigée}} = \frac{0.05}{25} = 0.002$$

Ainsi, une différence n’est considérée comme statistiquement significative que si la p-valeur associée est inférieure à 0.002. Cette approche garantit que le risque global de commettre au moins une erreur de type I reste inférieur à 5 %, assurant ainsi la rigueur et la fiabilité des conclusions statistiques.

**Exemple illustratif :** Supposons que les tests t appliqués aux cinq jeux de données produisent les p-valeurs suivantes :

| Jeu de données | p-valeur |
|----------------|----------|
| HC             | 0.004    |
| Chinese        | 0.013    |
| FPUS23         | 0.021    |
| Africa         | 0.002    |
| ISTHMIC        | 0.006    |

TABLEAU 5.13 – Exemple de p-valeurs obtenues avant application de la correction de Bonferroni.

Après correction, seuls les jeux HC, AFRICA et ISTHMIC présentent une différence réellement significative ( $p < 0.01$ ). Cette approche garantit que les conclusions relatives à la supériorité du modèle HAT reposent sur des résultats statistiquement solides et non sur des fluctuations aléatoires.