UQAC
Université du Québec
à Chicoutimi

AMÉLIORATION DE LA QUALITÉ DES IMAGES MÉDICALES À L'AIDE DES MODÈLES DE SUPER-RÉSOLUTION POUR UN MEILLEUR DIAGNOSTIC

PAR GILDAS AIMÉ SEDOU FOFE

MÉMOIRE PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À CHICOUTIMI COMME EXIGENCE PARTIELLE EN VUE DE L'OBTENTION DU GRADE DE MAÎTRE ÈS SCIENCES EN INFORMATIQUE

QUÉBEC, CANADA

RÉSUMÉ

L'imagerie médicale joue un rôle vital dans les diagnostics et les soins aux patients. Cependant, l'accès à des équipements de haute qualité reste souvent limité en raison de coûts élevés et de contraintes logistiques, en particulier dans des régions éloignées des centres urbains. Face à ces défis, la restauration des images médicales provenant d'équipements moins performants émerge comme une solution prometteuse. Cette recherche s'inscrit dans cette perspective en évaluant divers modèles de restauration d'images.

Nous avons commencé par étudier les modèles de super-résolution existants basés sur les GAN tels que SRGAN, BSRGAN, RANK-SRGAN, et SIR-SRGAN, en comparant leurs résultats selon des mesures telles que le PSNR, le SSIM, le LPIPS, le HaarPSI, le Clip-IQA, la taille du fichier généré et le temps d'exécution. Parmi ces modèles, SIR-SRGAN a offert en moyenne les meilleurs résultats. Cependant, les images générées par SIR-SRGAN présentaient des zones floues. Pour pallier cette limitation, nous proposons notre propre architecture, SIR-SRGAN-ResNEXT, une amélioration de SIR-SRGAN.

SIR-SRGAN-ResNEXT conserve le mécanisme de classement par auto-interpolation proposé par SIR-SRGAN, permettant au générateur de se concentrer sur les différences entre l'image reconstruite et l'image originale pour optimiser la qualité de la reconstruction. En outre, plusieurs modifications ont été apportées, notamment le remplacement du discriminateur basé sur l'architecture du PatchGAN par un discriminateur basé sur l'architecture U-Net. L'architecture du générateur a également été modifiée pour adopter une architecture basée sur ResNeXT, avec l'ajout de couches d'attention pour améliorer l'analyse des caractéristiques des images par le discriminateur.

Bien que SIR-SRGAN-ResNEXT offre de meilleurs résultats que SIR-SRGAN, il n'était pas entièrement satisfaisant par rapport à nos attentes de qualité d'image. Nous avons alors exploré des méthodes de restauration d'images basées sur les transformateurs. Les architectures basées sur des transformateurs, telles que SWINIR et SWIN2SR, ont surpassé celles basées sur les GAN en générant des textures d'image plus claires.

Ces modèles sont très efficaces pour restaurer les images grâce à leur mécanisme d'attention, qui divise l'image en plusieurs régions afin que le modèle puisse se concentrer sur différentes parties de l'image séparément. Cependant, la complexité de ce mécanisme d'attention est quadratique, ce qui affecte le coût d'utilisation de ces modèles sur des images de grande taille.

Dans ce travail, nous présentons le modèle Flatten-SwinIR pour la super-résolution d'images, une version améliorée et optimisée de SwinIR. Nous avons modifié le mécanisme "Window Attention" de SwinIR en le remplaçant par un mécanisme appelé "Flatten Attention", un mécanisme d'attention à complexité linéaire. Ce dernier offre un meilleur temps d'exécution tout en améliorant la qualité des images générées.

Les expériences ont été menées sur dix ensembles de données d'images médicales et de référence générale. Cette recherche éclaire la voie vers une utilisation plus répandue et plus efficace de la restauration des images médicales, contribuant ainsi à l'amélioration globale des soins de santé à l'échelle mondiale.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ABRÉVIATIONS

SRGAN Super Resolution Generative Adversarial Networks
GAN Generative Adversarial Network
SIR-SRGAN Super-Resolution Generative Adversarial Network with Self-Interpolation Ranker
ViT Vision Transformers
SWIN Shifted windows
ResNet Residual Network
CNN Convolutional neural network
IRM Imagerie par résonance magnétique
RSTB Residual Swin Transformer blocks
PSNR Peak Signal to Noise Ratio
SSIM Structural Similarity Index Measure
LPIPS Learned Perceptual Image Patch Similarity
HaarPSI Haar wavelet-based perceptual similarity index
SRCNN Super-Resolution Convolutional Neural Network
MSE Mean Squared Error
SRResNet Super-Resolution Residual Network
VGG Visual Geometry Group
TV Total Variation
BSRGAN Blind Super-Resolution Generative Adversarial Network
CLIP-IQA Contrastive Language-Image Pre-training - Image Quality Assessment
DWPT Discrete Wavelet Packet Transform
PDL Patch Distance Loss
SwinIR Swin Transformer for Image Restoration
FRTB Flatten Residual Transformer Block
FSTL Flatten Swin Transformer Layer
NLSA Non-Local Sparse Attention
HAN Holistic Attention Network
RCAN Residual Channel Attention Network
NLA Non-Local Attention
LSH Locality Sensitive Hashing
CSAM channel-spatial attention module
LAM layer attention module
DWC Depthwise Convolution

ReLU Rectified Linear Unit
MLP Multi-Layer Perceptron
GELU Gaussian Error Linear Unit
LN Layer Normalization
EDSR enhanced deep super-resolution network

DÉDICACE

Je dédie ce travail à ma famille. Pour votre amour indéfectible et votre soutien constant. Ce mémoire est le fruit de notre engagement mutuel. Sachez que j'en suis très reconnaissant.

Merci pour tout !

REMERCIEMENTS

En premier lieu, j'adresse de sincères remerciements à ma directrice de recherche, Professeur Haïfa Nakouri, pour sa patience, ses remarques, ses conseils, sa disponibilité et sa bienveillance.
J'exprime également ma sincère gratitude envers le Professeur Bob-Antoine J. Ménélas, dont l'encadrement et le soutien logistique ont grandement contribué à la réussite de ce projet.

Je tiens à remercier tous les membres du personnel du département d'informatique et de mathématiques de l'UQAC pour leur assistance, directe et indirecte, durant mes études à l'UQAC.

À ma famille, à mes amis et spécialement à Marie Noelle Mekontso dont les prières et les encouragements m'ont permis de surmonter tous les obstacles.
Enfin, que toutes les personnes ayant contribué à la réalisation de ce travail soient assurées de ma profonde reconnaissance.

AVANT-PROPOS

L'achèvement de ce mémoire marque un moment significatif dans mon parcours académique, jalonné par ma passion pour l'intelligence artificielle et son potentiel d'impact dans le domaine médical. Dans ma quête d'un sujet de recherche à fort impact, j'ai eu la chance de collaborer avec le professeur Haifa Nakouri sur le sujet de la super-résolution avec les réseaux antagonistes génératifs (GAN).

Notre première exploration a conduit au développement d'un GAN amélioré appelé SIR-SRGAN-RESNEXT qui ameliore la qualité des images et donc les tests ont donné de bons résultats sur les images médicales. Cependant, notre ambition était de repousser les limites et d'explorer d'autres architectures d'amélioration de la qualité des images.

Ainsi est né notre deuxième modèle, Flatten-SwinIR, résultat d'une collaboration intensive et de recherches approfondies. Notre objectif est de simplifier le travail des professionnels de la santé à travers le monde, en offrant des solutions d'imagerie médicale de haute qualité même dans des environnements dépourvus d'équipements sophistiqués.

CHAPITRE I

INTRODUCTION

1.1 CONTEXTE

Certains établissements de santé, en raison de contraintes financières et de leur situation géographique éloignée, font face à des difficultés d'accès à des équipements médicaux de haute qualité. Les coûts d'installation élevés et la voluminosité des dispositifs d'imagerie médicale constituent des obstacles majeurs, particulièrement dans les régions éloignées des centres urbains. En réponse à cette problématique, de nombreux professionnels de la santé adoptent des solutions plus légères, mobiles et économiques, se tournant vers des dispositifs médicaux mobiles abordables.

Par exemple, dans le domaine de l'ophtalmologie, des dispositifs tels que le Peek Retina et D-EYE sont utilisés (Yusuf *et al.*, 2022). En dentisterie, l'utilisation d'endoscopes, comme la Dental Camera, permet des examens plus accessibles. Pour l'échographie, des échographes portatifs sont privilégiés (Hunt *et al.*, 2021), tandis que des dermatoscopes mobiles tels que le MoleScope sont employés en dermatologie (Mitchell *et al.*, 2021). Ces alternatives offrent des solutions mobiles et économiques, bien que la qualité des images obtenues ne puisse rivaliser avec celle d'équipements sophistiqués dans des hôpitaux modernes.

Afin de surmonter cette limitation de qualité des images médicales, notre proposition s'oriente vers l'intégration des méthodes de vision par ordinateur à travers des modèles de super-résolution. Ces modèles, appliqués aux images capturées par ces dispositifs alternatifs, peuvent améliorer leur résolution. Cette approche permet aux médecins de faire des diagnostics plus précis, malgré les contraintes liées aux équipements disponibles, en exploitant le potentiel des

modèles de super-résolution pour augmenter la qualité des images médicales obtenues dans des contextes où l'accès à des équipements sophistiqués est limité.

1.2 PROBLÉMATIQUE

Au sein de la littérature, une multiplicité de solutions ont été développées basées sur le modèle de réseau de neurones convolutifs (CNN), à savoir SRCNN (Dong *et al.*, 2014a), FSRCNN (Dong *et al.*, 2016), EDSR (Lim *et al.*, 2017). Cependant, ces modèles ne produisent pas des images de haute qualité lorsque l'image d'entrée est fortement dégradée. Pour résoudre ce problème, des modèles basés sur des réseaux antagonistes génératifs encore appelés GAN (Goodfellow *et al.*, 2014) tels que Super-Resolution SRGAN (Ledig *et al.*, 2017), ESRGAN (Wang *et al.*, 2019) ont été introduits. Cependant, ces solutions produisent des images avec des zones floues. Des modèles comme SIR-SRGAN (Huang *et al.*, 2021) et RankSRGAN (Zhang *et al.*, 2019) ont été créés pour atténuer ce problème, en utilisant des classeurs de rang. BSRGAN (Zhang *et al.*, 2021a) quant à lui intègre dans son processus d'entrainement, un module de dégradation plus complexe pour dégrader le jeu de données et s'entrainer avec des jeux de données très dégradés.

Parallèlement, d'autres chercheurs ont tenté de relever les défis de la restauration d'images en utilisant des mécanismes d'attention complexes (Bahdanau *et al.*, 2015). Notamment, le modèle HAN (Niu *et al.*, 2020) utilise plusieurs modules de couche d'attention pour restaurer une image en tenant compte des interdépendances entre différentes régions de l'image. Dai *et al.* (2019) a proposé le modèle SAN qui est un CNN avec plusieurs connexions résiduelles (He *et al.*, 2016a) utilisant un mécanisme d'attention appelé SOCA (Dai *et al.*, 2019) et des opérations Non Locales (Wang *et al.*, 2018) capables d'extraires les relations entre les caractéristiques d'une image. NLSA (Mei *et al.*, 2021a) est un modèle basé sur l'architecture EDSR (Lim *et al.*, 2017) qui utilise le mécanisme d'attention Non Locale (Mei *et al.*, 2020) et

le hachage à sensitivité locale (Datar *et al.*, 2004) pour exploiter la similarité entre les petits motifs d'une image pour améliorer sa représentation. Zhang *et al.* (2018b) ont proposé RCAN, une architecture très profonde utilisant plusieurs couches d'attention.

D'autre part, le modèle Vision Transformer (ViT) (Dosovitskiy *et al.*, 2020) a ouvert la voie à des modèles innovants de restauration d'images. L'un de ces modèles est SwinIR (Liang *et al.*, 2021b), un modèle de restauration d'images qui utilise Swin Transformer (Liu *et al.*, 2021), qui est un transformateur de vision derivé de ViT. SwinIR produit des images de haute qualité par rapport aux modèles de super-résolution d'images mentionnés précédemment. Cependant, l'entrainement de SwinIR n'est pas très stable et son coût de calcul est très élevé en raison de la complexité quadratique de son transformateur (Kitaev *et al.*, 2020). Plus tard, Swin2SR (Conde *et al.*, 2022) a été introduit comme une amélioration de SwinIR qui utilise Swin Transformer V2 (Liu *et al.*, 2022), qui est un transformateur plus stable avec un coût de calcul inférieur par rapport à Swin Transformer. Malgré ses performances impressionnantes en termes de qualité d'image, Swin2SR présente également une complexité quadratique qui entraîne des coûts de calcul élevés.

La problématique centrale de cette recherche réside dans la compréhension des architectures de super-résolution les mieux adaptées aux exigences particulières des images médicales. Face à une multitude d'options, il devient essentiel de déterminer les architectures les plus pertinentes. Et comment pouvons-nous les améliorer ?

1.3 OBJECTIFS

Pour ce travail nous nous sommes fixés les objectifs suivant :

- Décrire les concepts autour de la super-résolution et son importance dans l'imagerie médicale.

- Déterminer les meilleurs catégories d'architectures de modèles de super-résolution et décrire les modèles les plus innovants de chaque catégorie.
- Proposer des modèles de super-resolution ameliorés par rapport aux modèles existants. Ces modèles doivent générer des images de meilleur qualité tout en ayant un temps d'éxécution court.
- Entrainer nos modèles et les comparer aux modèles existants sur la base de mesures bien définies.

1.4 SOLUTION PROPOSÉE

Nous proposons deux modèles de super-résolution comme solutions pour palier aux problèmes de qualité d'images des modèles de super-résolution. Le premier modèle se nomme SIR-SRGAN-ResNeXt, une amélioration du modèle SIR-SRGAN, qui s'appuie sur l'architecture des réseaux génératifs antagonistes (GAN). SIR-SRGAN-ResNeXt se distingue par son générateur basé sur l'architecture ResNeXt. L'intégration de ResNeXt permet au modèle de mieux capturer les relations spatiales et hiérarchiques entre les pixels, et de générer des images plus réalistes.

Le modèle SIR-SRGAN-ResNeXt utilise également un discriminateur U-Net, qui, contrairement aux discriminateurs classiques, est capable d'extraire simultanément des caractéristiques globales et locales de l'image. Notre deuxième solution est un modèle qui s'appuie sur une approche différente à savoir les transformateurs de vision, des architectures récemment développées pour la vision par ordinateur. Ce modèle se nomme Flatten-SwinIR, il se distingue par son utilisation d'une architecture de transformateur de vision appelée Flatten Swin Transformer Layer (FSTL), qui exploite un mécanisme d'attention linéaire appelé Flatten Attention. Ce mécanisme offre une complexité computationnelle linéaire, ce qui permet de réduire consi-

dérablement le coût du calcul par rapport à l'attention quadratique des modèles du même type tels que SwinIR et Swin2SR, tout en augmentant la qualité des images générées.

1.5 RÉSULTATS ET CONTRIBUTIONS

Les deux modèles SIR-SRGAN-ResNeXT et Flatten-SwinIR permettent d'obtenir des images de meilleure qualité à partir d'images basse résolution. Ainsi ils peuvent être utiliser pour améliorer le diagnotic à partir d'appareils médicales moins performants.

SIR-SRGAN-ResNeXt, grâce à l'intégration de l'architecture ResNeXt et du discriminateur U-Net, affiche des scores sur des mesures tels que PSNR (Peak Signal-to-Noise Ratio) et LPIPS (Learned Perceptual Image Patch Similarity) significativement plus élevés que les autres modèles de type GAN, notamment sur les ensembles de données médicales comme Messidor-2 et Breakhis-400x. Par exemple, sur Messidor-2, le modèle atteint un PSNR de 42.079 (+1,197 par rapport à SIR-SRGAN) et un LPIPS de 0.0121 (+0,0039 par rapport à SIR-SRGAN), démontrant ainsi sa capacité à restituer des images de fond d'œil avec un niveau de détail et de précision élevé. Flatten-SwinIR, quant à lui, fait nettement mieux avec un PSNR de 44.3906 (+0,2652 par rapport à Swin2SR) et un LPIPS de 0.0094 (+0.001 par rapport à Swin2SR). Flatten-SwinIR donne de telles performance tout en ayant un temps d'éxécution très court. Par exemple sur le meme jeu de donnée Messidor-2 nous avons un temps d'éxécution de 90 secondes. Alors que Swin2SR donne un temps d'éxécution plus long de 152 secondes.

Flatten-SwinIR permet aussi un débruitage des images. La capacité de ces modèles à améliorer la qualité des images médicales peut etre utiliser dans des appareils médicals afin d'accroître la précision des diagnostics, de faciliter la planification des traitements et d'ouvrir de nouvelles voies de recherche dans le domaine de la santé.

1.6 ORGANISATION

Cette section présente la méthodologie adoptée pour structurer ce mémoire.

- **Chapitre I : Introduction**

  Ce chapitre introduit l'importance de la super-résolution dans l'imagerie médicale. Il met en lumière les difficultés d'accès à des équipements médicaux de haute qualité dans certains établissements de santé. Le chapitre souligne l'importance des méthodes de vision par ordinateur, en particulier les modèles de super-résolution et décrit le cadre de notre recherche.

- **Chapitre II : Revue de la littérature**

  Dans ce chapitre nous présentons un état de l'art de la super-résolution d'images, en se concentrant sur les modèles les plus pertinents pour l'imagerie médicale. Tout d'abord nous définissons les concepts fondamentaux autour de l'image numérique, la vision par ordinateur et la super-résolution, puis nous explorons les différentes techniques d'imagerie médicale. Le chapitre détaille ensuite les architectures de différents modèles de super-résolution, notamment les modèles basés sur les GANs, les couches d'attention personnalisées et les transformateurs de vision. Il examine les forces et les faiblesses de chaque approche et discute des mesures d'évaluation les plus utilisées pour mesurer la qualité des images super-résolues.

- **Chapitre III : Architecture des modèles proposés**

  Ce chapitre se concentre sur l'architecture détaillée des deux modèles de super-résolution proposés : SIR-SRGAN-ResNeXt et Flatten-SwinIR. Nous décrivons en détail les

composants de ces modèles, leur principaux atouts, leur fonctions de perte. Nous présentons l'amélioration que leur architecture apporte par rapport à leurs concurrents.

– **Chapitre IV : Expérimentation et résultats**

Ce chapitre présente les résultats des expériences menées sur les deux modèles de super-résolution proposés, SIR-SRGAN-ResNeXt et Flatten-SwinIR. Les performances de ces modèles sont comparées à celles d'autres architectures de pointe sur différents ensembles de données d'images médicales et générales. L'analyse des résultats et des études d'ablation met en évidence le potentiel de nos deux modèles pour la super-résolution d'images, en particulier sur des images médicales.

CHAPITRE II
REVUE DE LA LITTÉRATURE

2.1 DÉFINITIONS

2.1.1 IMAGE NUMÉRIQUE

Une image est un concept très large qui est une représentation du monde. Une image est une représentation visuelle d'un objet, d'une scène, d'une idée, d'une personne ou de tout autre élément perceptible par la vue. Elle peut être capturée, créée ou stockée à l'aide de divers outils technologiques.

On distingue deux catégories d'images à savoir les images analogiques et les images numériques.

- IMAGE ANALOGIQUE : C'est une représentation physique, continue et d'un objet ou d'une scène. Elle est créée par la capture et le stockage de la lumière sur un support physique sensible à la lumière. Exemples : Photographies argentiques, films cinématographiques, peintures, dessins.
- IMAGE NUMÉRIQUE : C'est une représentation numérique d'un sujet, créée par la conversion de la lumière en données numériques. Contrairement à une image analogique, l'image numérique est constituée d'un nombre fini de points, appelés pixels. Chaque pixel est associé à une valeur numérique qui représente sa couleur et sa luminosité. Exemples : Images prises avec des appareils photo numériques, images scannées, images générées par ordinateur.

Les images numériques peuvent être classées en plusieurs types selon leur utilisation. Parmi les principaux types nous avons :

- IMAGE MATRICIELLE : Elle est aussi appelée image « bitmap », Elle est constituée d'une grille de pixels. Chaque pixel est une case qui contient une couleur codée par un nombre. En grossissant une image matricielle on perd de la qualité.
- IMAGE VECTORIELLE : Les images vectorielles sont composées de formes géométriques définies par des formules mathématiques. Un grossissement de ce type d'image n'affecte pas la qualité de l'image car les formes sont recalculées sans perdre de qualité.

Les images produites par les appareils photographiques sont des images matricielles, ce type d'image permet de capturer les details fins et les différentes couleurs du monde réel. Contrairement aux images vectorielles qui sont faites de lignes droites et courbes, les images vectorielles ne peuvent pas représenter de façon réaliste les images du monde réel.

2.1.2 CARACTÉRISTIQUES D'UNE IMAGE MATRICIELLE

Toute image matricielle possède les caractéristiques suivantes :

- PIXEL : C'est l'abréviation en anglais de « **pic**ture **el**ement ». C'est l'élément de base de l'image. L'ensemble des pixels contenu dans une grille à deux dimensions (largeur et hauteur) constitue l'image.
- RÉSOLUTION : C'est le nombre de pixels par unité de pouce (1 pouce = 2.54 centimètres) dans l'image. Elle représente la densité des pixels. Plus il y a de pixels par unité de longueur plus la résolution est élevée.
  Elle est exprimée en « PPP » (points par pouce) ou DPI (dots per inch).
- TAILLE EN PIXELS : Encore appelé définition de l'image, il représente le nombre de pixels qui composent l'image. Il est égal au produit du nombre de pixels dans la longueur,

par le nombre de pixels dans la largeur de l'image.

Exemple : une image dont la définition est 2000×1800 correspond à une image de 2000 pixels en largeur et 1800 pixels en hauteur.

- PROFONDEUR DE BIT : Encore appelé profondeur de couleur, c'est le nombre de bits utilisés pour représenter un pixel dans l'image.

  les images bitonales (noir et blanc) utilisent 1 bit par pixel, chaque bit représentant soit le noir soit le blanc. Les images en niveaux de gris ont 256 couleurs (ou nuances de gris) et nécessitent 8 bits par pixel. Les images à milliers de couleurs utilisent 16 bits par pixel, offrant une palette plus étendue de couleurs. Les images en millions de couleurs sont les plus couramment rencontrées, utilisent 24 bits par pixel, permettant de représenter environ 16,7 millions de couleurs différentes. Les images plus colorées nécessitant une plus grande profondeur de bits pour chaque pixel.

- TAILLE DU FICHIER (POIDS) : C'est le produit entre le nombre de pixels de l'image et le poids d'un pixel. Le poids d'un pixel est la quantité d'octet sur lequel il est codé.

- LUMINANCE : Encore appelée moyenne, elle représente l'intensité de la lumière émise ou réfléchie par l'image telle que perçue par l'œil humain. Elle se calcule en faissant la moyenne des pixels de l'image après avoir convertir l'image en niveau de gris. Sa formule est donnée par l'équation 2.1.

$$\text{Luminance} = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \tag{2.1}$$

  Où :

  - N et M sont respectivement le nombre de pixels en largeur et en hauteur de l'image.
  - $f(x,y)$ est la valeur de luminance du pixel à la position (x,y).

La luminance ou moyenne d'une image donne une indication globale de la luminosité de l'image. Si sa valeur est faible (proche de 0) alors l'image est globalement sombre. Cela signifie que la majorité des pixels ont des valeurs de gris faibles (proche du noir). Si la luminance est autour de 128 alors l'image a un mélange équilibré de zones sombres et claires. Par contre si luminance est elevée (proche de 255), l'image est globalement claire. La majorité des pixels ont des valeurs de gris élevées (proches du blanc).

- CONTRASTE : elle représente la différence entre les niveaux de luminance (ou de couleur) dans différentes parties de l'image. Il indique à quel point les parties sombres et claires de l'image sont distinctes.

$$\text{Contraste} = \sqrt{\frac{1}{MN}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\left(f(x,y)-\text{Moy}\right)^2} \tag{2.2}$$

où Moy est la luminance de l'image

Un contraste élevé signifie que l'image a des différences marquées entre les zones claires et sombres. Si il est faible alors les différences entre les zones claires et sombres sont moins prononcées.

2.1.3 VISION PAR ORDINATEUR

La vision par ordinateur est un domaine de l'intelligence artificielle qui vise à permettre aux ordinateurs d'interpréter et comprendre le monde visuel de manière similaire à la perception humaine. Elle consiste à extraire des informations pertinentes à partir d'images numériques ou de vidéos et à les utiliser pour prendre des décisions, faire des recommendations ou automatiser des tâches. La vision par ordinateur utilise plusieurs techniques et algorithmes pour analyser, interpréter, transformer le contenu visuel.

Parmis les applications de la vision par ordinarteur nous avons la reconnaissance d'objets qui

permet d'identifier des objets dans des images, la reconnaissance faciale, la reconnaissance optique de caractères.

SUPER-RÉSOLUTION

La super-résolution désigne l'ensemble des techniques dont l'objectif est d'améliorer la qualité des images en augmentant leur résolution. C'est un processus de reconstruction d'image qui vise à générer une image de haute résolution (HR) à partir d'une image de basse résolution (LR) en ajoutant des données dans l'image pour créer une image plus détaillée et plus nette (Dong *et al.*, 2015).

La super-résolution trouve des applications dans de nombreux domaines dont voici les principaux :

- SURVEILLANCE ET SÉCURITÉ : La super-résolution permet l'amélioration de la résolution des images et séquences de vidéo de surveillance pour identifier les visages, des plaques d'immatriculation et d'autres détails importants dans les enregistrements de sécurité (Nasrollahi & Moeslund, 2014).
- IMAGERIE MÉDICALE : La super-résolution permet l'amélioration de la résolution des images médicales pour une meilleure détection et diagnostic des maladies (De Bruijne *et al.*, 2021).
- IMAGERIE SATELLITAIRE : La super-résolution permet l'amélioration des images satellitaires pour bien visualiser les surfaces terrestres, la végétation et les structures urbaines (Salvetti *et al.*, 2020).
- PHOTOGRAPHIE : La super-résolution permet l'amélioration de la qualité des images prises avec des appareils photo numériques ou des smartphones.

- ART ET CULTURE : La super-résolution permet la restauration d'Œuvres d'art ou de livres anciens. Elle pernet aussi l'amélioration des images de fouilles archéologiques pour une analyse plus précise des artefacts (Gatys *et al.*, 2016).

Les premières méthodes de super-résolution d'image étaient des méthodes de traitement d'image classiques. Cela signifie que nous utilisons une transformation qui est appliquée sur chaque pixel de l'image , telles que les méthodes d'interpolation-restauration (Yang & Huang, 2017). Cependant, ces méthodes fournissent des images trop lisses qui n'approchent pas de la cible. Les meilleures méthodes de super-résolution ont commencé à émerger avec l'introduction de la vision par ordinateur dans ce domaine de recherche. Ainsi, Dong *et al.* (2014b) a développé SRCNN qui est l'un des premiers réseaux de neuronnes convolutives profonds Krizhevsky *et al.* (2012) qui transforment une image à basse résolution en une image haute résolution. Cela marque une avancée majeure dans la recherche sur la super-résolution.

2.1.4 IMAGERIE MÉDICALE

L'imagerie medicale designe l'ensemble des techniques utilisant des technologies pour visualiser des différents tissus ou organes du corps humain, afin de réaliser un diagnostic ou une intervention médicale (Ma *et al.*, 2021). Les techniques d'imagerie médicale comprennent des tests non invasifs qui permettent aux médecins de diagnostiquer des blessures et des maladies sans avoir à introduire un appareil dans le corps humain.

Ces techniques d'imagerie médicale ont permit une avancèe significative dans la médécine avec de nombreuses applications dans le diagnostic des maladies graves telque les pathologies du myocarde, des cancers, des troubles neurologiques, rétinopathie, des maladies cardiaques, des fractures osseuses et d'autres conditions médicales graves.

Hussain *et al.* (2022) présentent comme principales techniques d'imagerie médicale les techniques suivantes :

– La radiographie

La radiographie fonctionne en utilisant des rayons X pour créer des images de l'intérieur du corps. Les rayons X sont des ondes électromagnétiques à haute énergie capables de traverser les tissus mous comme la peau et les muscles, mais sont partiellement absorbés par les structures plus denses telles que les os. Lors d'une radiographie, une machine envoie un faisceau de rayons X à travers la zone du corps à examiner. Une plaque ou un détecteur placé de l'autre côté capte les rayons X après qu'ils aient traversé le corps, formant une image en fonction de la quantité de rayons absorbés par les différentes structures internes.

Les images obtenues sont illustrées par la figure 2.1. On observe des contrastes où les os apparaissent en blanc en raison de leur densité élevée qui absorbe plus de rayons X, tandis que les tissus mous apparaissent en nuances de gris ou noir. Cette technique permet aux médecins de visualiser et de diagnostiquer des fractures, des infections, des anomalies osseuses, et d'autres conditions médicales internes. La radiographie est une procédure rapide, non invasive et couramment utilisée dans les hôpitaux et les cliniques pour diverses applications diagnostiques.

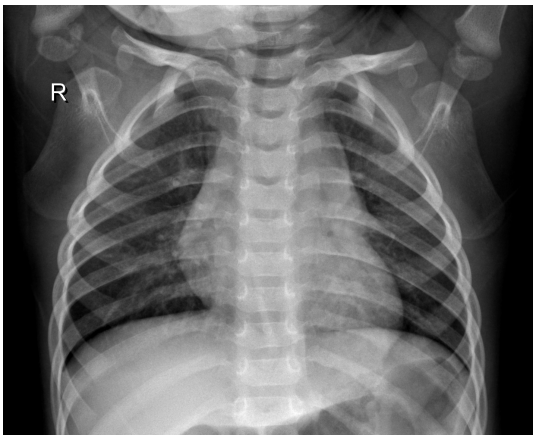

FIGURE 2.1 : Image de radiographie pulmonaire tiré du jeu de données Chest X-Ray (Wang *et al.*, 2017).

– LA TOMODENSITOMÉTRIE (TDM)

La tomodensitométrie (TDM), utilise des rayons X pour créer des images détaillées en coupe transversale de l'intérieur du corps. Le patient est placé sur une table qui glisse à l'intérieur d'un anneau appelé gantry, lequel contient un tube à rayons X et des détecteurs. Le tube à rayons X tourne autour du patient, émettant des faisceaux de rayons X à travers le corps sous différents angles. Les rayons X traversent le corps et sont absorbés différemment par les divers tissus et structures internes, créant ainsi des variations dans l'intensité des rayons détectés de l'autre côté du corps. Les détecteurs capturent ces variations et envoient les données à un ordinateur, qui reconstruit les données en images tomographiques. Ces images peuvent être visualisées en deux dimensions (coupe transversale) ou en trois dimensions pour fournir des vues précises et détaillées des organes, des os, des vaisseaux sanguins et des tissus mous. La TDM est particulièrement utile pour diagnostiquer et surveiller des conditions médicales complexes telles que les tumeurs, les lésions internes, les infections et les maladies cardiovasculaires, car elle offre une résolution et une clarté supérieures à celles des radiographies.

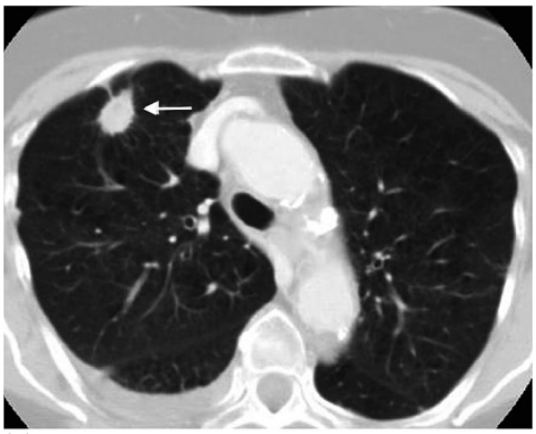

FIGURE 2.2 : Image de tomodensitométrie (TDM) thoracique tiré du Jeu de données LIDC-IDRI (Armato III *et al.*, 2011).

– LA PHOTOGRAPHIE RÉTINIENNE

La photographie rétinienne, ou imagerie rétinienne, est une technique qui permet de capturer des images numériques haute résolution et en couleur de la rétine, du nerf optique et des vaisseaux sanguins situés à l'arrière de l'œil. Utilisant des lasers à faible puissance, cette technologie projette de la lumière à travers la pupille jusqu'à la rétine, où elle forme des images détaillées recueillies par une machine spécialisée. Ces images permettent aux ophtalmologistes d'examiner minutieusement la rétine et de détecter précocement diverses maladies oculaires telles que la rétinopathie diabétique, le glaucome, la dégénérescence maculaire liée à l'âge, le décollement de la rétine et même certains types de cancers comme le mélanome rétinien.

FIGURE 2.3 : Image du fond de l'oeil tiré du jeu de données Messidor-2 (Decencière *et al.*, 2014).

– L'IMAGERIE PAR RÉSONANCE MAGNÉTIQUE (IRM)

L'imagerie par résonance magnétique (IRM) est une technique qui utilise des champs magnétiques et des ondes radio pour produire des images détaillées de l'intérieur du corps. Lors d'un examen IRM, un aimant supraconducteur crée un champ magnétique fort, ce qui aligne les protons des atomes d'hydrogène présents dans les tissus corporels. Ensuite, des impulsions de radiofréquence sont émises, perturbant cet alignement. Lorsque les protons reviennent à leur état initial, ils émettent des signaux qui sont captés par des détecteurs et transformés en images par un ordinateur. Ces images offrent une excellente résolution des tissus mous, ce qui permet de visualiser les structures internes sans avoir recours à des rayonnements ionisants, comme ceux utilisés en tomodensitométrie (TDM). L'IRM est particulièrement utile pour diagnostiquer des affections telles que les tumeurs, les infections, les lésions des ligaments et des tendons, ainsi que les maladies du cerveau et de la moelle épinière. Elle permet également d'étudier la fonction cérébrale, le flux sanguin et la diffusion des molécules d'eau dans les tissus. La figure 2.4 présente une image IRM d'un cerveau humain.

FIGURE 2.4 : Image par résonance magnétique (IRM) d'un cerveau humain tiré du jeu de données BraTS (Ghaffari *et al.*, 2019).

– LA MICROSCOPIE OPTIQUE

La microscopie optique est une technique d'imagerie d'objets qui est utilisé dans l'imagerie médicale essentiellement pour visualiser des structures biologiques à une échelle microscopique. La microscopie optique est largement employée en histopathologie pour créer des images histopathologiques, la figure 2.5 présente une image histopathologique de cancer du sein. Dans ce processus, des échantillons de tissus prélevés sont fixés, coupés en sections fines, puis colorés pour mettre en évidence les différentes structures cellulaires et tissulaires. Ces sections sont ensuite placées sous un microscope optique, où la lumière passe à travers les échantillons. Les différentes structures cellulaires absorbent ou réfléchissent la lumière différemment, ce qui crée des contrastes visuels permettant aux pathologistes de visualiser et d'analyser les tissus. Les images résul-

tantes sont examinées pour diagnostiquer des maladies, évaluer leur gravité et guider les décisions thérapeutiques.

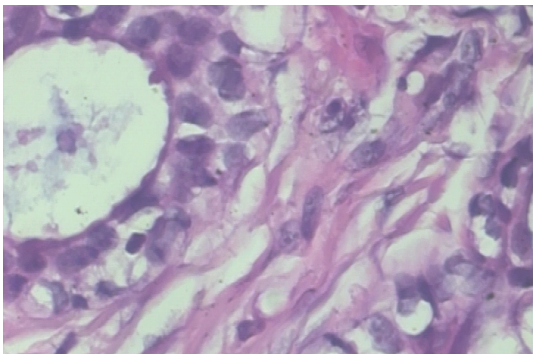

FIGURE 2.5 : Image histopathologique du cancer du sein tiré du Jeu de donnée BreaKHis 400X (Spanhol *et al.*, 2016).

– L'ÉCHOGRAPHIE DIAGNOSTIQUE

L'échographie médicale, est une technique d'imagerie qui utilise des ondes sonores de haute fréquence pour produire des images des structures internes du corps. Le composant clé de ce système est le transducteur à ultrasons, qui émet des ondes sonores et capte les échos réfléchis par les différents tissus du corps. Ces échos sont ensuite transformés en images visualisées sur un écran. L'échographie est largement utilisée dans divers domaines médicaux en raison de son absence de rayonnement ionisant, de son coût relativement faible, et de sa portabilité. Elle permet de visualiser des organes tels que les reins, le cœur, le foie, et les vaisseaux sanguins, ainsi que de surveiller les grossesses, détectant les anomalies fœtales et les conditions telles que le placenta praevia et les grossesses multiples. La figure 2.6 présente une image d'échographie d'un foi.

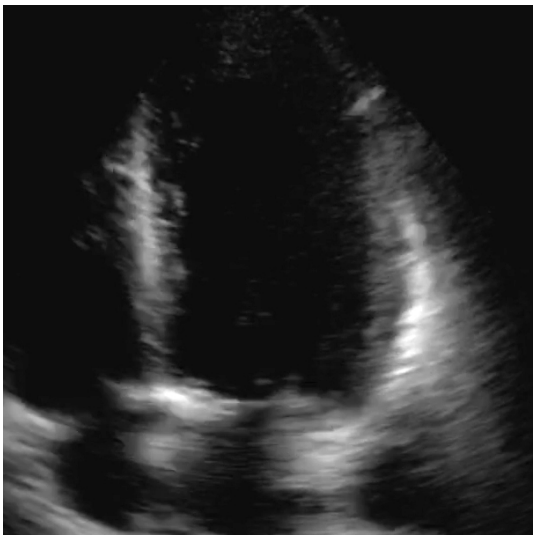

FIGURE 2.6 : Image d'échographie d'un foi tiré du jeu de données US-4 (Chen *et al.*, 2021).

2.1.5 DISPOSITIFS PORTABLES POUR L'IMAGERIE MÉDICALE

Les technologies d'imagerie médicale ont considérablement évolué pour inclure une gamme variée d'appareils exploitant des formes spécifiques d'énergie. Les rayons X sont largement utilisés dans les appareils de radiologie pour produire des images détaillées des structures internes du corps. Les scanners IRM exploitent les ondes radio pour générer des images précises des organes et des tissus. Les échographes utilisent des ultrasons pour produire des images en temps réel, facilitant ainsi le diagnostic médical. Ces machines sont souvent lourdes et peu accessibles dans les régions éloignées des centres urbains , comme l'appareil d' IRM illustré dans la figure 2.7. Pour pallier ce problème, on observe une utilisation croissante d'appareils de contournement d'imagerie médicale, également connus sous le nom de dispositifs portables d'imagerie médicale Hunt *et al.* (2021), offrant ainsi une solution plus accessible et pratique pour les patients dans divers contextes cliniques.

FIGURE 2.7 : Un appareil d'IRM Philips, au sein de l'hôpital universitaire de Sahlgrenska en Suède .

De « Philips MRI in Sahlgrenska Universitetsjukhuset, Gothenburg, Sweden », par Jan Ainali, 2008 (https ://commons.wikimedia.org/wiki/File :MRI-Philips.JPG). CC BY 3.0.

– Dispositifs portables d'imagerie médicale

Dans de nombreuses régions du monde, en particulier dans les zones éloignées des centres urbains et dans les pays en développement, l'accès à des équipements médicaux sophistiqués est limité en raison de leur coût élevé et de leur poids (car très lourd). Les équipements médicaux conventionnels, tels que les scanners d'IRM, les machines à rayons X, et les échographes, nécessitent souvent des infrastructures coûteuses et un personnel spécialisé pour leur utilisation et leur maintenance (Gonçalves-Bradley *et al.*, 2020).

Cela crée des disparités significatives dans l'accès aux soins de santé de qualité, car de nombreuses régions éloignées ne disposent pas des ressources nécessaires pour acquérir et entretenir ces équipements.

Cependant, l'émergence d'appareils de contournement, tels que les dispositifs médicaux basés sur smartphone ou les dispositifs portables, a révolutionné la prestation des soins de santé dans ces régions mal desservies. Ces appareils, souvent abordables et faciles à transporter, offrent des solutions innovantes pour le diagnostic, la surveillance et le traitement des patients (Ma *et al.*, 2021).

Par exemple, des dispositifs portables de surveillance de la glycémie permettent aux patients diabétiques de surveiller leur glycémie à domicile, réduisant ainsi la nécessité de visites fréquentes à l'hôpital.

Un exemple frappant est l'utilisation croissante d'échographes portables dans les régions rurales et éloignées. Ces appareils compacts permettent aux professionnels de la santé de réaliser des échographies sur le terrain, facilitant ainsi le diagnostic précoce des maladies et des complications, notamment dans les cas d'urgence.

De plus, les échographes portables sont souvent utilisés lors de missions médicales dans des zones reculées ou lors de catastrophes naturelles, fournissant une aide médicale vitale là où les infrastructures médicales traditionnelles sont absentes ou endommagées.

Nous avons aussi des appareils de contournement médicaux basés sur smartphones qui exploitent les capacités photographiques et les capteurs des smartphones pour diverses applications médicales, allant de la surveillance de la santé au diagnostic, en passant par le suivi des traitements et les interventions chirurgicales.

Par exemple, l'appareil Volk VIVA illustré par la figure 2.8 est un outil léger permettant de dianostiquer le fond de l'oeil . Ou encore l'outil Peek Retina, présenté à la figure 2.9, bien qu'il ne soit plus en vente actuellement, il a démontré qu'il pouvait diagnostiquer des cas de rétinopathie diabétique sans avoir besoin d'un équipement complexe, comme

l'a montré une étude menée dans un hôpital en Ouganda (Yusuf *et al.*, 2022).

Nous avons aussi des microscopes adaptables aux smartphones qui permettent des analyses biologiques ex vivo, les dermatoscopes facilitent les diagnostics dermatologiques à distance, et les endoscopes mini-invasifs offrent des examens internes portables et économiques. Des dispositifs de photothérapie basés sur un smartphone permettent de prendre des images de zones précises du corps de manière non invasive.

Ces technologies, portables et connectées, démocratisent l'accès aux soins, permettent le partage instantané des données médicales pour des consultations à distance, et sont facilement utilisables grâce aux interfaces intuitives des smartphones.

Cependant, ils présentent des limites techniques par rapport aux équipements médicaux spécialisés et dépendent de la variabilité des modèles de smartphones et des mises à jour technologiques continues.

Malgré ces défis, les dispositifs basés sur smartphones constituent une avancée majeure pour améliorer les soins de santé, particulièrement dans les régions à ressources limitées, et favorisent la télémédecine.

FIGURE 2.8 : Image de l'appareil Volk VIVA.

De « Portable Fundus Camera », par Volk Optical Inc., 2024 (https ://www.volk.com/pages/custom-product), © 2024 par Volk Optical Inc. Reproduit avec permission.

FIGURE 2.9 : Image d'un Peek Retina adapté pour smartphone

De « Peek Retina sales closing on 28 September, » par Peek Vision Ltd. © 2024 par Peek Vision Ltd. Reproduit avec permission.

– Importance de la super-résolution dans l'imagerie medicale

Les appareils de contournement basés sur smartphones pour l'imagerie médicale font face à plusieurs limitations que la super-résolution peut aider à surmonter. La résolution des capteurs d'image des smartphones, bien qu'avancée, est souvent inférieure à celle des équipements médicaux professionnels, les objectifs des smartphones, plus petits que ceux des microscopes spécialisés, capturent moins de détails, de plus, les conditions d'éclairage sous-optimales et les capteurs plus petits peuvent introduire du bruit et des artefacts.

Les techniques de super-résolution peuvent réduire ces artefacts, revéler les détails fins des images obtenues par ces appareils, améliorant ainsi la qualité des images. En outre, pour rester portables et économiques, ces appareils ne peuvent pas utiliser de composants optiques sophistiqués, mais la super-résolution permet d'obtenir des images de haute qualité avec des composants plus simples et moins coûteux.

En somme, la super-résolution améliore la performance des appareils de contournement en augmentant la résolution et la qualité des images, atténuant les effets des limitations optiques, et rendant les diagnostics plus précis et fiables, tout en maintenant la portabilité et l'accessibilité économique des dispositifs médicaux de contournements.

La super-résolution rend les diagnostics par imagerie médicale plus accessibles et économiques, soutient la télémédecine en permettant le partage d'images détaillées pour des consultations à distance, et contribue à la recherche biomédicale en permettant de constituer des bases de données d'images de haute qualité pour la recherche.

2.2 MODÈLES DE SUPER-RÉSOLUTION

Un modèle de super-résolution est un algorithme d'apprentissage automatique qui augmente la résolution d'une image basse résolution, en utilisant des techniques avancées telles

que les réseaux neuronaux convolutifs (CNN), les réseaux antagonistes génératifs Goodfellow *et al.* (2014) et bien d'autres techniques d'apprentissage automatique.

Ces modèles visent à reconstruire des détails fins et réalistes dans les images super-résolues. Les premiers modèles de surper-résolution comme SRCNN Dong *et al.* (2014b) se concentrent uniquement sur la minimisation de l'erreur quadratique moyenne (MSE), les nouveaux modèles tiennent aussi compte de la similarité perceptuelle à savoir la manière dont l'humain perçoit l'image super-résolue.

Dans cette partie, nous explorons les modèles les plus avancés dans ce domaine de la vision par ordinateur, en mettant l'accent sur les approches architecturales de conception les plus pertinentes tout en examinant en détail leurs principes, leurs performances et leurs défis.

2.2.1 MODÈLES BASÉS SUR L'ARCHITECTURE GAN

– SRGAN

Le modèle SRGAN (Super-résolution Generative Adversarial Network) Ledig *et al.* (2017) est construit en utilisant une architecture CNN appelée SRResNet qui utilise des connexions résiduelles. L'ajout de connexions résiduelles He *et al.* (2016b) est une technique qui facilite l'apprentissage des caractéristiques de l'image à faible résolution. Les blocs résiduels de l'architecture SRResNet permettent d'apprendre des représentations plus complexes tout en facilitant la propagation du gradient pendant la formation.

En utilisant les caractéristiques extraites par un modèle VGG pré-entraîné Simonyan & Zisserman (2014) pour calculer la perte de contenu, des niveaux plus élevés d'informations perceptives telles que la structure et la texture sont pris en compte, ce qui permet d'obtenir des résultats visuels plus réalistes et plus détaillés.

Cependant SRGAN ne génère pas de bonnes images lorsque l'image basse résolution contient trop de bruits, dans ce cas le modèle a tendance à augmenter le bruit. Pour

résoudre ce problème d'autres solutions ont emergé.

La formule de la fonction de perte du générateur du SRGAN est donnée par l'équation 2.3.

$$Perte_{SRGAN} = l_{Content} + \alpha \cdot l_{adv}^{Gen} + \beta \cdot l_{TV} \tag{2.3}$$

$$l_{Content} = l_{SR}^{VGG/i,j} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(\phi_{i,j}(I_{HR})^{(x,y)} - \phi_{i,j}(Gen(I_{LR}))^{(x,y)} \right)^2 \tag{2.4}$$

$$l_{adv}^{Gen} = -\sum_{n=1}^{N} \log Disc(Gen(I_{LR}^{(n)})) \tag{2.5}$$

$$l_{TV}(x) = \frac{1}{N} \left(\sum_{i=1}^{H-1} \sum_{j=1}^{W} (x_{i+1,j} - x_{i,j})^2 + \sum_{i=1}^{H} \sum_{j=1}^{W-1} (x_{i,j+1} - x_{i,j})^2 \right) \tag{2.6}$$

Où $l_{Content}$ est la perte de contenu et l_{adv}^{Gen} est la perte adversariale du Générateur.

l_{SR} est l'image haute résolution de référence, I_{LR} est l'image basse résolution.

$\phi_{i,j}$ représente la carte de caractéristiques obtenue à partir du j-ième filtre après la i-ième couche de maxpooling du réseau VGG pré-entraîné.

l_{TV} est la perte de variation totale, c'est une mesure de la variation totale dans une image, elle est utilisée pour encourager la régularité spatiale dans les images générées par des réseaux de neurones. $x_{i,j}$ représente la valeur du pixel à la position (i, j) de l'image x de hauteur H et largeur W. N etant le nombre total de pixel.

– BSRGAN

Zhang et al. proposent BSRGAN (Zhang *et al.*, 2021a), un modèle de super-résolution a été entrainé sur des images plus dégradées que ceux du SRGAN. En effet BSRGAN

inclu dans son algorithme un mecanisme de degradation qui lui permet de s'entrainer sur les images très dégradées.

Ce mecanisme utilise un flou aléatoirement, des opérations de sous-échantillonnage, et des dégradations de bruit. Le flou est particulièrement approximé par deux convolutions avec des noyaux gaussiens isotropes et anisotropes.

Le sous-échantillonnage est effectué par des interpolations telles que le plus proche voisin, l'interpolation bilinéaire et la convolution cubique, le bruit est un bruit gaussien. Ainsi, le BSRGAN donne de meilleurs résultats que le SRGAN sur des images présentant de nombreuses dégradations visuelles, en particulier sur des images de visages humains. Cependant les images générées par BSRGAN sont trop lisses et s'éloignent enormément de l'image originale (image cible). BSRGAN utilise la même fonction de perte que le SRGAN.

– RANKSRGAN

RankSRGAN (Zhang *et al.*, 2021b) est un modèle introduit par Zhang et al. qui utilise une classeur de type CNN appelée « Ranker » pour classer les images en fonction de leur qualité perceptuelle. Le classeur est entraîné à guider le générateur dans la direction des mesures en utilisant une fonction appelée la perte de contenu de rang ou perte de classement. Ce qui permet d'obtenir des images moins floues et plus détaillées par rapport au SRGAN.

Ce classeur agit comme un second discriminateur, en donnant un score à chaque image générée et en forçant le générateur à orienter son entraînement pour obtenir de meilleurs scores. RankSRGAN génére des images de qualité supérieure à SRGAN et ayant une bonne qualité visuelle.

Cependant RankSRGAN a du mal à bien représenter les hautes fréquences (bordures d'objets dans les images) dans une image très floue. Pour une telle image RankSRGAN

donne certe une image ayant une bonne qualité visuelle mais qui s'èloigne de l'image de d'origine. La fonction de perte du générateur de RankSRGAN est donné par l'équation 2.7.

$$Perte_{\text{RankSRGAN}} = L_{\text{G}} + \lambda \cdot L_{\text{Rank}} \tag{2.7}$$

$$L_{\text{Rank}} = \text{sigmoid}(R(G(x))) \tag{2.8}$$

Où, L_{G} est la perte du générateur comme dans SRGAN, L_{Rank} est la perte de classement, introduite pour améliorer la qualité perceptuelle des images générées.
λ est un hyperparamètre de pondération qui balance l'importance des deux termes de la perte.
$R(G(x)$ le score de classement de l'image donnée par le classeur. Un score de classement plus bas indique une meilleure qualité perceptuelle.

– SIR-SRGAN

Le SIR-SRGAN (Huang *et al.*, 2021) développé par Huang est un modèle de super-résolution d'image qui vise à améliorer la qualité de la reconstruction en se concentrant sur les différences intrinsèques entre l'image reconstruite et l'image originale. SIR-SRGAN s'inspire de Rank-SRGAN, mais apporte plusieurs modifications significatives axées sur trois axes particuliers :

- **Architecture :** Il a deux classeurs de rang ("Rankers") au lieu d'un, un classeur standard et un classeur de caractéristiques.
- **Training :** Les classeurs sont entraînés simultanément avec le générateur, contrairement à RankSRGAN, où un seul est pré-entraîné avant d'être utilisé pour entraîner

le GAN.

SIR-SRGAN utilise une interpolation d'images, au lieu d'envoyer uniquement l'image générée par le générateur aux classeurs pendant la formation, le modèle envoie plusieurs variantes possibles des images générées. Ces variantes sont obtenues en interpolant les pixels de l'image super-résolution (SR) générée avec l'image haute résolution (HR) originale. Ces interpolations sont destinées à enrichir les données d'apprentissage en fournissant différentes variantes de l'image super-résolue avec différents niveaux d'interpolation, contribuant ainsi à améliorer la robustesse et la généralisation des classeurs, et par conséquent celle du générateur grâce à la rétropropagation des gradients.

- **Fonction de perte :** SIR-SRGAN ajoute une nouvelle composante, la perte de distance de patchs (PDL), présentée dans l'équation 2.10, au calcul de la fonction de perte globale du générateur.

$$Perte_{\text{SIR-SRGAN}} = \phi \cdot L_{\text{adv}} + \beta \cdot L_{\text{content}} + \gamma \cdot L_{\text{rank\_pixel}} + \delta \cdot L_{\text{rank\_feature}} + \tau \cdot L_{\text{PDL}} \tag{2.9}$$

$$Perte_{PDL} = L_{\text{PDL}} = PDL(G(L_R), H_R) + PDL(DWPT(G(L_R)), DWPT(H_R)) \tag{2.10}$$

Où nous avons tout comme dans SRGAN, L_{adv} est la perte adversariale du générateur, L_{content} la perte de contenu.

$L_{\text{rank\_pixel}}$ et $L_{\text{rank\_feature}}$ représentent respectivement les pertes des classeurs standard et de caractéristiques.

L_R représente les images à basse résolution, H_R représente les images à haute

résolution d'origine et *DWPT* représente la transformée en paquets d'ondelettes discrètes de Haar.

La fonction *PDL* décrite dans l'article SIR-SRGAN (Huang *et al.*, 2021) est une fonction qui découpe les caractéristiques à haute fréquence d'une image en parcelles 4×4 et considère chaque parcelle comme un vecteur. Elle mesure la distance entre les parcelles à haute fréquence de l'image, en utilisant la similarité cosinusoïdale pour calculer la distance entre les parcelles.

– SRGAN-RESNEXT

Juhong *et al.* (2023) proposent le SRGAN-ResNeXt, une variante du SRGAN qui utilise l'architecture ResNeXt (Xie *et al.*, 2017) au lieu de ResNet (He *et al.*, 2016b) pour son générateur. ResNeXt introduit le concept de « cardinalité », qui représente le nombre de chemins de transformations parallèles dans le modèle ResNeXt .

L'augmentation de la cardinalité permet d'améliorer les performances sans accroître excessivement la complexité du modèle (profondeur). Contrairement à ResNet, qui utilise des blocs résiduels uniques, ResNeXt utilise des blocs résiduels parallèles où plusieurs transformations sont effectuées en parallèle.

Cela permet de capturer des caractéristiques plus riches et d'améliorer la capacité du modèle sans augmenter de manière significative la complexité du modèle et d'éviter les problèmes connexes tels que le surajustement ou l'explosion des gradients .

Les résultats sur des images médicales histopathologiques du cancer du sein indiquent que SRGAN-ResNeXt, avec la même fonction de perte que SRGAN offre une amélioration significative de la qualité de l'image par rapport à SRGAN qui utilise un générateur basé sur l'architecture ResNet.

2.2.2 MODÈLES BASÉS SUR DES COUCHES D'ATTENTION PERSONNALISÉES

– RCAN

RCAN (Zhang *et al.*, 2018b) comporte quatre parties principales : l'extraction de caractéristiques peu profondes, l'extraction de caractéristiques profondes (nommé RIR), avec des couches résiduelles, le module de mise à l'échelle et la partie de restauration de l'image.

Dans un premier temps, une couche convolutive est utilisée pour extraire les caractéristiques superficielles de l'image basse résolution. Ensuite, ces caractéristiques sont soumises à une extraction profonde par le biais de la partie RIR, qui comprend plusieurs groupes résiduels.

Cette partie utilise un mécanisme d'attention qui analyse canaux contenant des informations pertinentes pour la reconstruction. En se concentrant sur les canaux informatifs, le mécanisme d'attention permet au réseau de se concentrer sur les aspects les plus importants de l'image,

Après l'extraction des caractéristiques profondes, un module de mise à l'échelle est utilisé pour augmenter la résolution spatiale des caractéristiques. Enfin, les caractéristiques mises à l'échelle sont reconstruites en une image haute résolution à l'aide d'une couche de restauration convolutive.

Cette architecture permet au RCAN d'atteindre une grande profondeur et d'apprendre de manière adaptative des caractéristiques utiles de l'image et améloire les performances de super-résolution de l'image.

La fonction de perte du RCAN est la perte L1 encore appelée la fonction de perte de la moyenne de la valeur absolue.

– HAN

HAN (Niu *et al.*, 2020), est un modèle de super-résolution mettant en œuvre deux modules d'attention : le module d'attention par couche nommé LAM et le module d'attention canal-spatial nommé CSAM.

Le modèle HAN se compose de plusieurs parties. Tout d'abord, il extrait les caractéristiques d'une image d'entrée à faible résolution à l'aide d'une couche convolutive. Ces caractéristiques sont ensuite progressivement améliorées par un ensemble de groupes de couches résiduelles, qui constituent le cœur du réseau.

LAM vise à prendre en compte les corrélations entre les différentes couches d'entités extraites par les groupes résiduels.

Concrètement, il apprend une matrice de corrélation entre ces couches, ce qui permet au réseau d'accorder plus de poids aux couches d'entités informatives tout en éliminant les couches redondantes.

De cette manière, ce module contribue à améliorer la représentation des caractéristiques en tenant compte des dépendances entre les couches hiérarchiques.

Le module CSAM s'attaque à un problème courant dans les mécanismes d'attention existants en explorant les corrélations non seulement entre les canaux, mais aussi entre les positions spatiales. En utilisant la convolution 3D, ce module génère une carte d'attention en combinant les informations spatiales et les caractéristiques des canaux.

Cette approche permet d'extraire des représentations puissantes pour décrire les interdépendances spatiales et entre les canaux, améliorant ainsi la qualité de la reconstruction.

Ce modèle utilise la même fonction de perte que RCAN.

– NLSA

NLSA (Mei *et al.*, 2021b) est un modèle d'amélioration de la qualité des images qui utilise une forme spécifique d'attention appelée attention non locale (NLA), qui permet au modèle d'incorporer des informations provenant de différentes parties de l'image

pour reconstruire les détails les plus fins.

Pour rendre cette opération d'attention plus efficace, NLSA utilise un module appelé « Attention Bucket », qui permet de regrouper les parties similaires de l'image. L'attention non locale utilise une technique de hachage appelée « Locality Sensitive Hashing » (LSH), qui partitionne l'espace d'entrée en seaux de hachage contenant des caractéristiques similaires.

Les seaux de hachage permettent de grouper les pixels de caractéristiques similaires, ce qui optimise le processus d'attention en se concentrant uniquement sur les zones pertinentes.

De plus, pour atténuer les effets des seaux déséquilibrés, une permutation des éléments est effectuée avant de les diviser en tronçons de taille fixe, permettant une exécution parallèle plus efficace. Le modèle peut également effectuer plusieurs tours de LSH pour augmenter la robustesse de l'attention. Sa fonction de perte est l'erreur quadratique moyenne

2.2.3 MODELES BASÉS SUR L'ARCHITECTURE VISION TRANSFORMER

Dans le domaine de la vision par ordinateur, une nouvelle forme d'architecture a émergé sous le nom de transformateurs de vision (couramment appelé "Vision Transformer").
Elle offre une nouvelle perspective sur la manière de traiter les tâches complexes de vision par ordinateur. Inspirés par le succès des transformateurs (Vaswani *et al.*, 2017) dans le traitement du langage naturel, les Transformateurs de Vision (Dosovitskiy *et al.*, 2020) ont eu un grand succès car ils ont un mécanisme d'attention qui permet de capturer les dépendances des caractéristiques des images ainsi que les caractéristiques globales de façon plus optimale.
Ce mécanisme surpasse ceux des architectures comme HAN ou NLSA dans diverses tâches de vision.

– VISION TRANSFORMER

L'introduction des transformateurs (Vaswani *et al.*, 2017) dans la vision par ordinateur a marqué une avancée significative. Initialement conçu pour le traitement du langage naturel (NLP), il a d'abord été utilisé dans le modèle Vision Transformer (ViT) (Dosovitskiy *et al.*, 2020), un modèle de classification d'images. ViT divise une image en un ensemble de parcelles sur lesquelles un mécanisme d'auto-attention basé sur la fonction SoftMax est appliqué.

Lorsque Vit prend en entrée une image il la subdivise en une séquence de N parcelles (appelées patchs) d'image de taille D. L'entrée devient un vecteur X de dimension de $N \times D$, où N est la longueur de la séquence (nombre de patchs) et D est la dimension de chaque patch.

Le mécanisme d'auto-attention appliqué à X est basé sur la formule présentée dans l'équation 2.11.

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{D}}\right) \cdot V \tag{2.11}$$

Où la requête Q, la clé K et la valeur V sont des matrices obtenues par projection linéaire de X ($Q = XW_Q, K = XW_K, V = XW_V$). Avec W_Q, W_K et W_V les matrices de poids apprises.

ViT a ouvert la voie à des modèles de restauration d'image innovants, notamment le modèle SwinIR.

– SWINIR

Proposé par Liang *et al.* (2021a), SwinIR est un modèle de restauration d'image qui utilise une architecture nommé Swin Transformer (Liu *et al.*, 2021), qui est un module de transformateurs utilisant le mécanisme d'attention de ViT. Le modèle SwinIR comprend

trois modules principaux : l'extraction de caractéristiques peu profondes, l'extraction de caractéristiques profondes et le module de restauration d'images.

Le module d'extraction des caractéristiques peu profondes utilise une couche convolutionnelle 3x3 pour extraire les caractéristiques peu profondes de l'image d'entrée de faible qualité. Les caractéristiques peu profondes sont transmises directement au module de reconstruction afin de conserver les informations de basse fréquence.

Le module d'extraction des caractéristiques profondes est principalement composé de blocs résiduels appelés " Residual Swin Transformer Block" (RSTB). Chaque RSTB utilise plusieurs couches de Swin Transformer pour extraire les dépendances entre les caractértiques de l'image. Une couche convolutive est ajoutée à la fin de chaque bloc pour améliorer les caractéristiques, et une connexion résiduelle permet l'agrégation de ces caractéristiques.

Le module de restauration d'image combine les caractéristiques superficielles et profondes pour la reconstruction finale de l'image de haute qualité. Pour la tâche de super-résolution, le module utilise une couche convolutive pour effectuer le suréchantillonnage des caractéristiques.

SwinIR tire parti du mécanisme d'auto-attention des transformateurs pour capturer les interactions globales dans l'image tout en conservant la capacité de convolution pour le traitement local. La fonction de perte utilisée est la perte L1 pixel, qui mesure la différence absolue entre les valeurs de pixels prédites et les valeurs de pixels réelles dans les images originale de haute qualité. L'attention est mesurée par l'équation 2.12.

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{D}} + B\right) \cdot V \tag{2.12}$$

Où S est une matrice de biais apprise lors de l'entrainement qui permet au modèle de

mieux capturer les relations spatiales entre les caractéristiques.

– SWIN2SR

Swin2SR (Conde *et al.*, 2022), une évolution du modèle SwinIR, se distingue par son mécanisme d'attention amélioré. Contrairement à SwinIR, Swin2SR utilise dans mécanisme d'attention, un produit scalaire en cosinus au lieu d'un produit scalaire simple, cela permet d'atténuer les valeurs extrêmes des cartes de poids d'attention, améliorant ainsi la stabilité de l'entraînement. Ces cartes d'attention sont générées par les couches d'attention du modèle. Elles indiquent quelles parties de l'entrée, le modèle devrait se concentrer pour la prise de décision sur une tâche donnée.
Dans le mécanisme d'attention de Swin2SR, la normalisation est faite de façon post-résiduelle, ce qui atténue l'instabilité en cas d'augmentation de la capacité du modèle. L'attention est mesurée par l'équation 2.13.

$$Attention(Q,K,V) = \text{Softmax}\left(\frac{\cos(Q,K)}{\tau} + B\right)V \tag{2.13}$$

Où Q est la matrice de requête, K est la matrice des clés, V est la matrice des valeurs. τ est un scalaire apprenable, non partagé entre les couches, qui est utilisé pour diminuer l'échelle des valeurs du produit scalaire cosinus.

2.3 MESURES D'ÉVALUATION DES MODÈLES DE SUPER-RÉSOLUTION

2.3.1 LES METRIQUES D'ÉVALUATION DE LA QUALITÉ DE L'IMAGE

Un modèle de super-résolution prend une image en entrée de faible résolution et donne une image en sortie de haute résolution. L'évaluation de la performance d'un tel modèle passe par l'évalution de la qualite des images genérées. Pour cela nous avons recensé les plus

pertinentes mesures d'évalutation de la qualité d'une image.

Il existe deux catégories de mesures pour évaluer la qualité d'une image : les mesures avec référence et les mesures sans référence. Les mesures avec référence utilisent une image de référence (l'image originale) pour évaluer l'image super-résolue. Elles nécessitent dont deux paramètres en entrée : l'image originale et l'image super-résolue. En revanche, les mesures sans référence n'utilisent que l'image super-résolue et emploient des algorithmes pour déterminer un score de qualité. L' objectif principal des mesures sans référence est d'évaluer la qualité perçue par l'œil humain.

Pour évaluer la qualité des images produites par un modèle de super-résolution, les travaux de recherche utilisent généralement 02 mesures avec référence qui sont le PSNR et le SSIM. À ces deux mesures, nous avons ajouté d'autres mesures avec référence qui sont LPIPS et HaarPSI, ainsi qu'une mesure sans référence, ClipIQA.

- **PSNR (Peak Signal-to-Noise Ratio)**

  Le PSNR (Kim, 1988) est une mesure quantitative utilisée pour évaluer la qualité d'une image reconstruite par rapport à une image originale. Exprimé en décibels (dB), le PSNR compare le maximum possible de puissance d'un signal à la puissance du bruit qui affecte la qualité de sa représentation.

  La formule du PSNR donnée en 2.14, inclut le calcul de l'erreur quadratique moyenne (MSE) entre l'image originale et l'image reconstruite. Plus la MSE est faible, c'est-à-dire plus les différences entre les deux images sont petites, plus le PSNR est élevé. Cela suggère que l'image reconstruite ressemble étroitement à l'image originale.

  Toutefois, il est important de noter que le PSNR ne prend pas en compte les aspects perceptuels de la vision humaine et se concentre uniquement sur une comparaison numérique directe des pixels, c'est pour cela qu'il ne peut pas etre utiliser comme seul indicateur de qualité d'image.

Le PSNR est généralement exprimé en termes d'échelle logarithmique de décibels. La mesure PSNR est toujours supérieure à 0. Plus la valeur est élevée, meilleure est la performance.

$$\text{PSNR}(I,J) = 10 \cdot \log_{10}\left(\frac{\max(I^2)}{\text{MSE}(I,J)}\right) \tag{2.14}$$

Dans cette formule, I et J représentent les deux images à comparer, et MSE représente l'erreur quadratique moyenne entre les deux images.

– **SSIM (Structural SIMilarity)**

SSIM (Wang *et al.*, 2004) est une mesure d'évaluation de la qualité d'une image avec reférence tout comme le PSNT. Elle compare les pixels entre l'image de référence (image originale) et une image à évaluer (image super-résolue).

Contrairement au PSNR, qui se base uniquement sur des différences de valeur des pixels et ne prend pas en compte la perception visuelle humaine, le SSIM intègre des aspects perceptuels tels que la luminance, le contraste et la structure locale des images, ce qui le rend plus pertinent pour évaluer la qualité visuelle perçue.

Cependant, SSIM présente certaines limites, il peut produire des résultats inattendus dans certains cas, comme par exemple des images avec des valeurs trés faibles de luminance ou contraste (Nilsson & Akenine-Möller, 2020).

La forumule du SSIM est donné par l'équation 2.15, sa valeur se situe dans l'intervalle [0,1]. Plus la valeur est élevée, meilleure est la performance.

$$\text{SSIM}(x,y) = \left(\frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)}\right) \tag{2.15}$$

Dans cette formule :

μ_A et μ_B représentent les moyennes locales des blocs autour des pixels dans les images A et B, respectivement.

σ_A^2 et σ_B^2 sont les variances locales des blocs autour des pixels dans les images A et B, respectivement.

σ_{AB} est la covariance locale entre les blocs autour des pixels dans les images A et B.

C_1 et C_2 sont des constantes définies par l'utilisateur pour stabiliser la division dans le cas où le dénominateur deviendrait trop petit.

– **LPIPS (Learned Perceptual Image Patch Similarity)**

La mesure LPIPS (Zhang *et al.*, 2018a) est une mesure d'évaluation de la qualité d'une image avec reférence, elle compare l'image super-résolue de l'image originale en se basant sur des caractéristiques extraites d'un réseau de neurones profond.

Contrairement au PSNR et au SSIM, qui se concentrent principalement sur les caractéristiques bas-niveau des images, la LPIPS prend en compte des caractéristiques de haut-niveau, capturant ainsi des informations plus complexes sur la perception humaine. Elle utilise des réseaux de neurones pré-entrainés comme le VGG-16 ou VGG-19 pour extraire les caractéristiques des images, puis calcule la distance entre ces caractéristiques. Ses avantages par rapport au PSNR et au SSIM résident dans sa capacité à capturer des informations de haut-niveau sur la similarité perceptuelle, ce qui la rend plus adaptée pour des tâches où la perception humaine est cruciale, comme la génération d'images ou le transfert de style.

La formule de la LPIPS est donné par l'équation 2.16, la mesure LPIPS couvre l'intervalle [0,1]. Plus la valeur est faible, plus le modèle d'amélioration de l'image est performant.

$$\text{LPIPS}(x,y) = \sum_{i=1}^{N} \lambda_i ||\phi_i(x) - \phi_i(y)||_2 \tag{2.16}$$

Où x et y sont les images à comparer, $\phi_i()$ est la fonction de mapping du réseau de neurones à la couche i qui extrait les caractéristiques, et λ_i sont des poids optionnels pour chaque couche.

– **HaarPSI (Haar Wavelet-Based Perceptual Similarity Index)**

La mesure HaarPSI (Reisenhofer *et al.*, 2018a) est une mesure d'évaluation d'image avec référence , qui évalue la qualité visuelle en se basant sur la décomposition en ondelettes de Haar.

Contrairement aux mesures PSNR et SSIM qui se concentrent principalement sur la comparaison des valeurs de pixel, HaarPSI évalue les similitudes locales entre deux images en utilisant des coefficients obtenus à partir de la décomposition en ondelettes de Haar. Elle permet de détecter les distorsions entre deux images et d'évaluer les similitudes locales entre les images.

Ce qui le rend plus proche de la perception humaine. En comparaison avec LPIPS qui est une mesure d'apprentissage profond, HaarPSI permet d'avoir un résultat comparable à LPSI tout en gardant une simplicité computationnelle. Ce qui le rend plus rapide à calculer et plus facile à interpréter.

La valeur HaarPSI se situe dans l'intervalle [0,1]. Plus la valeur est élevée, meilleure est la performance du modèle. Sa formule est donné par l'équation 2.17

$$\text{HaarPSI}(f_1, f_2) = l_\alpha^{-1}\left(\frac{\sum_x \sum_{k=1}^{2} \text{HS}(k)_{f1,f2}[x] \cdot \text{W}(k)_{f1,f2}[x]}{\sum_x \sum_{k=1}^{2} \text{W}(k)_{f1,f2}[x]}\right)^2 \tag{2.17}$$

f_1 f_2 représentent l'image originale et l'image super-résolue converties en niveau de gris,

$l_\alpha^{-1}(\cdot)$ est la fonction inverse de la fonction logistique avec le paramètre α.

$\mathrm{HS}(k)_{f1,f2}[x]$ est la mesure de similarité locale basée sur les coefficients d'ondelette de Haar pour les filtres de haute fréquence.

$\mathrm{W}(k)_{f1,f2}[x]$ est la carte de pondération dérivée de la réponse du filtre d'ondelette de Haar à basse fréquence.

– **CLIP-IQA (Contrastive Language-Image Pre-training - Image Quality Assessment**

La mesure CLIP-IQA (Wang *et al.*, 2022) est une mesure sans référence. Cela veut dire qu'elle prend en parametre une seule image à savoir l'image à évaluer (image super-résolue). Elle utilise le modèle CLIP (Contrastive Language-Image Pretraining model) developpé par l'organisme OpenAI, qui convertit le texte en images. CLIP-IQA adopte une approche holistique en exploitant la richesse des représentations visuelles et linguistiques pré-entraînées dans CLIP.

Cette approche permet à CLIP-IQA de capturer non seulement les aspects tangibles de la qualité des images (luminance, Contraste), mais aussi les perceptions abstraites telles que l'esthétique et les émotions associées à une image. En utilisant des paires de prompts antonymes, CLIP-IQA parvient à réduire l'ambiguïté linguistique lors de l'évaluation, ce qui permet d'obtenir des scores de qualité plus précis et significatifs.

CLIP-IQA évalue à la fois les aspects tangibles et intangibles de la qualité visuelle. Elle donne un score à l'image . Si le score est élevé, meilleure est la qualité de l'image. La valeur de ClipIQA se situe dans l'intervalle [0,1].

2.3.2 AUTRES MESURES

Nous pouvons évaluer les modèles sur d'autre critères autre que la qualité de l'image à savoir :

– **Temps d'exécution**

Le temps d'exécution d'un modèle de super-résolution est une mesure cruciale pour les méthodes de super-résolution, particulièrement dans le contexte de l'imagerie médicale. Tout d'abord, dans les environnements cliniques, les médecins et les radiologues ont souvent besoin de résultats en temps réel ou quasi-temps réel pour prendre des décisions rapides et éclairées sur les soins aux patients.

Afin d'éviter les retards qui pourraient compromettre le traitement des patients. De plus, les hôpitaux et les centres de diagnostic disposent souvent de grandes quantités de données d'imagerie à traiter quotidiennement.

– **Importance de la taille du fichier de sortie**

La taille du fichier de sortie est une autre mesure essentielle dans les modèles de super-résolution. Des fichiers de sortie plus volumineux peuvent contenir plus de détails et offrir une meilleure qualité d'image, ce qui est crucial pour des diagnostics précis. Cependant, ces fichiers peuvent également poser des défis logistiques significatifs.

Par exemple, les infrastructures hospitalières doivent être capables de stocker et de gérer de grandes quantités de données, ce qui peut nécessiter des investissements coûteux en matériel et en systèmes de gestion de l'information.

De plus, les fichiers volumineux peuvent ralentir les systèmes de réseau, rendant le partage plus lent. Un compromis doit donc être trouvé entre la qualité d'image requise et la taille des fichiers.

2.4 CONCLUSION

L'imagerie médicale est un domaine en constante évolution qui s'est transformé au fil des ans grâce à l'essor de technologies innovantes qui permettent de visualiser l'intérieur du corps humain avec précision .

De l'imagerie par résonance magnétique (IRM) qui révèle les détails complexes des organes et

des tissus, à la microscopie optique qui explore le monde microscopique des cellules, chaque technique apporte des contributions uniques au diagnostic et au traitement des maladies.
Cependant, l'accès à ces technologies de pointe reste un défi majeur, particulièrement dans les régions mal desservies.
Les équipements médicaux conventionnels, souvent lourds, coûteux et exigeant des infrastructures spécialisées, ne sont pas facilement disponibles pour tous. C'est dans ce contexte que l'émergence d'appareils de contournement d'imagerie médicale, également appelés dispositifs portables, constitue une véritable révolution dans le domaine de la santé.
Ces appareils, souvent basés sur des smartphones ou conçus pour être portables, offrent une solution accessible et pratique pour les patients dans divers contextes cliniques. L'un des défis auxquels sont confrontés ces dispositifs portables est la résolution des images qu'ils produisent.
La super-résolution, s'avère être une solution efficace pour pallier ce problème. Grâce à des algorithmes d'intelligence artificielle, la super-résolution permet d'améliorer la qualité des images obtenues, révélant des détails qui permettent d'obtenir des diagnostics plus précis et fiables.
Dans ce chapitre, nous avons exploré les principales approches architecturales de modèles de super-résolution et les principes sous-jacents qui distinguent ces modèles.
Les modèles basés sur les réseaux génératifs antagonistes (GANs) comme SRGAN, BSRGAN, RankSRGAN, SIR-SRGAN et SRGAN-ResNeXT exploitent la puissance des réseaux neuronaux convolutifs (CNNs) pour générer des images de haute résolution. Ils s'appuient sur des techniques comme les connexions résiduelles, la perte de contenu basée sur VGG et des classeurs de rang pour améliorer la qualité et le réalisme des images reconstruites.
D'autres approches, comme celles utilisant principalement des mécanismes d'attention à savoir, RCAN, HAN et NLSA, se concentrent sur l'extraction et l'exploitation des informations pertinentes des images.

Enfin, les modèles basés sur les transformateurs de vision, tels que SwinIR et Swin2SR, peuvent capturer les dépendances à long terme des caractéristiques de l'image et obtenir des résultats encore plus performants.

L'évaluation de ces modèles de super-résolution s'effectue à travers des mesures de la qualité des images reconstruites, le coût de calcul du modéle et la taille des fichiers générées.

CHAPITRE III
ARCHITECTURE DES MODÈLES PROPOSÉES

3.1 SIR-SRGAN-RESNEXT

3.1.1 ARCHITECTURE GLOBALE

SIR-SRGAN-ResNeXt est le premier modèle de super-résolution que nous proposons. Il est une amélioration du modèle SIR-SRGAN.

Son architecture est illustrée par la figure 3.1 s'inspire de celle du SIR-SRGAN. Elle repose sur l'utilisation de deux classeurs appelés "rankers". Il s'agit du classeur standard (encore appelé Stander Ranker) et du classeur de caractéristiques (encore appelé Feature Ranker) ; ils sont utilisés en conjonction avec un générateur et un discriminateur.

Pendant l'entraînement, nous utilisons le même processus d'interpolation d'image utilisé dans SIR-SRGAN. C'est à dire nous interpolons les images super-résolues (SR) avec les images originales de haute résolution (HR).

Le classeur standard est conçu pour trier ces images interpolées, en apprenant les différences intrinsèques entre les images générées et haute résolution. Simultanément, le classeur des caractéristiques utilise des caractéristiques de haut niveau extraites grâce au modèle VGG19 pour évaluer ces images.

Ces classeurs sont d'une grande importance puisqu'ils guident le processus de génération pour améliorer la qualité perçue des images super-résolues.

Ensuite, le discriminateur évalue la probabilité qu'une image soit générée ou pas, tandis que le générateur cherche continuellement à créer des images qui trompent le discriminateur. Cette architecture incorpore également la perte de distance de patchs (Patch Distance Loss), une mesure dérivée du SIR-SRGAN, pour bien représenter les hautes fréquences.

Les mises à jour introduites dans cette architecture apportent des améliorations significatives au processus de super-résolution.

- **Générateur :** Le générateur standard basé sur ResNet a été remplacé par un générateur basé sur ResNeXt (Xie *et al.*, 2017), offrant une capacité accrue pour apprendre des caractéristiques plus complexes et générer des images super-résolues de meilleure qualité.
- **Discriminateur :** le discriminateur conventionnel a été remplacé par une architecture U-Net (Ronneberger *et al.*, 2015) , favorisant une évaluation plus robuste des images générées et contribuant à stabiliser l'entrainement du modèle. L'ajout de couches d'attention au niveau du discriminateur là où les caractéristiques à grande échelle sont extraites, améliore la capacité du modèle à se concentrer sur les régions clés pendant le processus de super-résolution.

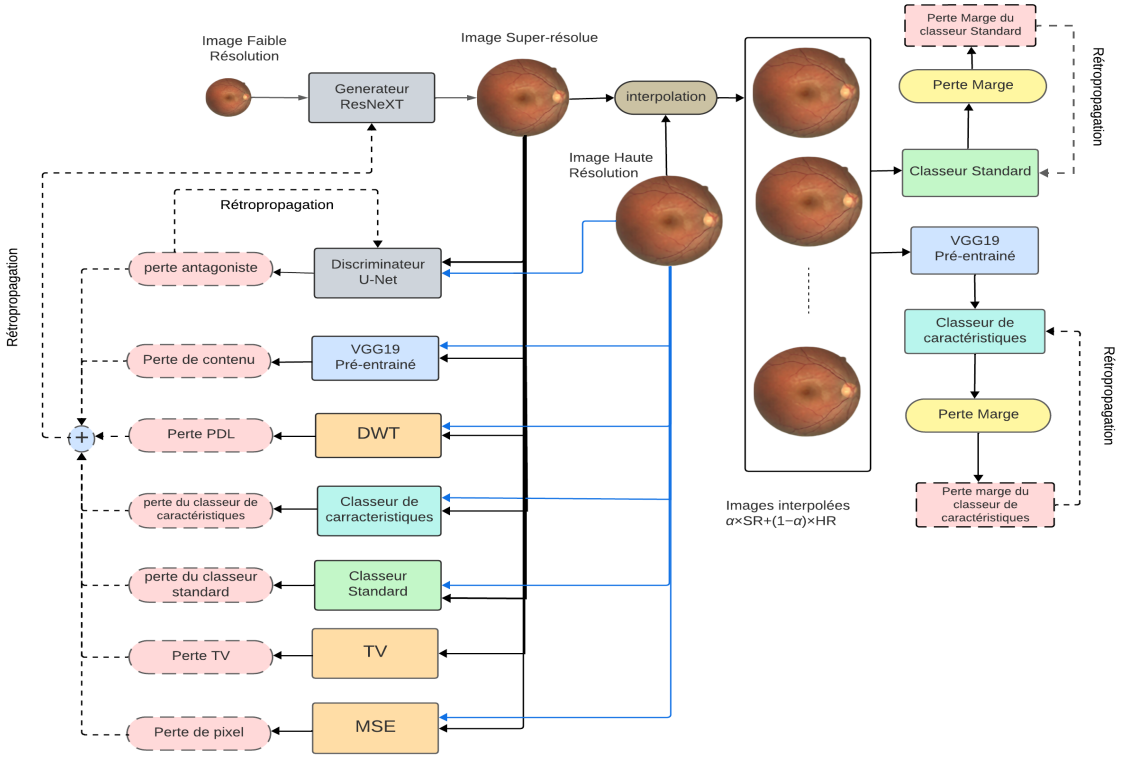

FIGURE 3.1 : Architecture du modèle SIR-SRGAN-ResNeXt

© Gildas Aimé Sedou Fofe

3.1.2 GENERATEUR

Le générateur de notre modèle est illustré par la figure 3.2, il utilise l'architecture ResNeXt pour apprendre des représentations profondes et complexes de l'image.

ResNeXt est une extension de l'architecture ResNet, C'est une architecture conçue pour améliorer les performances des réseaux convolutifs tout en contrôlant la complexité. Elle se distingue par l'introduction de la notion de cardinalité, qui représente le nombre de chemins de transformations parallèles qui permettent une plus grande expressivité du modèle, l'objectif étant de permettre au modèle d'apprendre plus de caractéristiques sans affecter la stabilité de l'apprentissage du modèle.

Un bloc de base ResNeXt (ResNeXtBlock) est formé de plusieurs chemins parallèles (cardinalité) de convolutions.

Chaque chemin effectue une convolution 1x1, suivie d'une convolution 3x3 (qui est en fait une convolution en groupe). Les sorties de ces chemins parallèles sont concaténées et ajoutées à l'entrée initiale du bloc pour former l'opération résiduelle.

Cette structure de son bloc de base permet au générateur ResNeXt d'avoir une plus grande précision que le générateur standard base sur ResNet. Car le modèle peut capturer des caractéristiques plus diversifiées de l'image (Xie *et al.*, 2017).

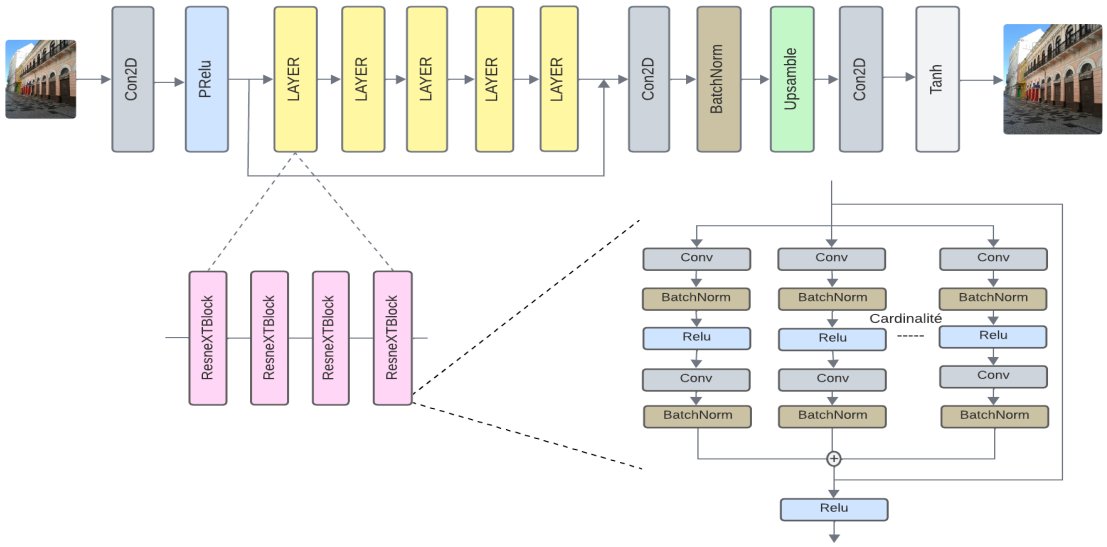

FIGURE 3.2 : Generateur du modèle SIR-SRGAN-ResNeXt

© Gildas Aimé Sedou Fofe

3.1.3 DISCRIMINATEUR

Schonfeld *et al.* (2020) ont démontré que les discriminateurs U-Net, comparés aux discriminateurs classiques, extraient simultanément des caractéristiques globales et locales,

fournissant un retour d'information spatial cohérent au générateur. En revanche, le discriminateur classique utilisé dans SIR-SRGAN, extrait soit des caractéristiques locales, soit des caractéristiques globales. Le discriminateur U-Net fournit un retour plus prononcé au générateur qu'un discriminateur classique.

Le discriminateur utilisé dans SIR-SRGAN-ResNeXt, est illustré par la figure 3.3, il s'inspire de l'architecture U-Net avec une normalisation spectrale (Miyato *et al.*, 2018) . Il prend une image en entrée et effectue plusieurs couches de convolution pour extraire des caractéristiques à différentes échelles.

L'architecture du discriminateur comprend plusieurs couches de résolution descendante qui réduisent progressivement la résolution spatiale de l'image, suivies de couches de résolution ascendante qui restaurent la résolution spatiale. Des connexions résiduelles sont utilisées pour relier les couches de résolution descendante et ascendante, ce qui favorise un flux d'informations efficace dans tout le réseau.

En outre, des mécanismes d'attention sont incorporés pour permettre au modèle d'optimiser l'extraction des caractéristiques de l'image au niveau de ces couches afin d'augmenter la capacité du discriminateur à discerner les details fins dans les images pendant le processus d'apprentissage. Le modèle utilise Leaky ReLU (Maas *et al.*, 2013), comme fonction d'activation pour introduire la non-linéarité. La normalisation spectrale (Miyato *et al.*, 2018) est appliquée aux couches pour stabiliser l'apprentissage du modèle. Au final, le discriminateur produit une seule valeur de sortie, indiquant la probabilité que l'image d'entrée soit générée ou pas.

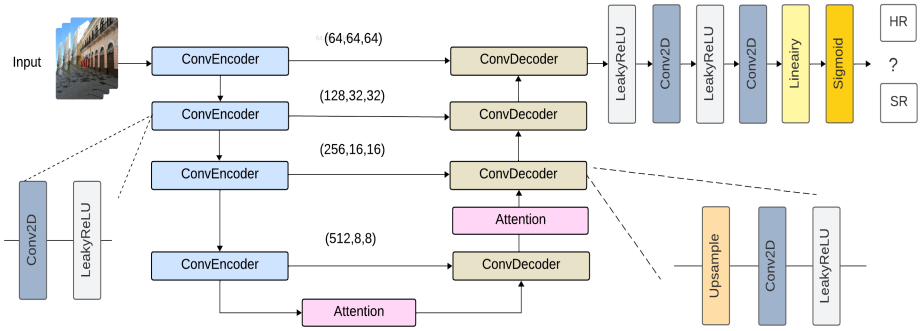

FIGURE 3.3 : Discriminateur du model SIR-SRGAN-ResNeXt

© Gildas Aimé Sedou Fofe

3.1.4 FONCTION DE PERTE

La fonction de perte globale du modèle proposé est donnée par l'équation suivante 3.1 :

$$Perte_{SIR-SRGAN-ResNeXt} = \alpha \cdot l_{TVLoss} + \beta \cdot l_{Content} + \gamma \cdot l_{Adv} + \lambda \cdot l_{Rank} + \theta \cdot l_{Pixel} + \varepsilon \cdot l_{PDL} \tag{3.1}$$

- **Perte adversariale (l_{Adv}) :** La perte adversariale aide à guider le générateur pour qu'il produise des images impossibles à distinguer des images réelles cibles du point de vue du discriminateur.
- **La perte de contenu ($l_{Content}$) :** mesure la différence entre les caractéristiques de l'image générée par le générateur et de l'image réelle cible dans un réseau pré-entraîné (VGG-19). L'objectif est de s'assurer que les caractéristiques perceptives de l'image générée sont similaires à celles de l'image haute résolution.
- **La perte de classement (l_{Rank}) :** La perte de classement représente la perte obtenue par le classeur standard (Stander Ranker) plus celle du classeur de caractéristiques (Feature Ranker). La perte de classement mesure la divergence entre les scores obtenus par les

classeurs sur le classement de l'image super-résolue (SR) et de l'image originale haute résolution (HR) en termes de pixels et de caractéristiques.

- **La perte PDL (l_{PDL}) :** La perte PDL est la perte introduite dans le document SIR-SRGAN. Cette perte utilise la transformée en ondelettes discrète (DWT), une transformation capable de capturer des informations à différentes échelles et fréquences dans les images. Cette composante permet au modèle d'être plus robuste aux variations spatiales à différentes résolutions, ce qui peut améliorer la qualité des images super-résolues générées.
- **La perte de pixels (l_{Pixel}) :** la perte de pixels mesure la différence d'erreur quadratique moyenne entre les valeurs des pixels de l'image super-résolue et de l'image à haute résolution. Cela favorise la similarité au niveau des pixels entre les images générées et les images réelles.
- **Perte de Variation Totale (l_{TVLoss}) :** la variation totale qui mesure la quantité totale de variation dans une image. Elle favorise la génération d'images avec des transitions plus douces et moins de variations abruptes entre les pixels voisins. Sa formule est donnée par l'équation 2.6.

L'utilisation des coefficients α, β, γ, λ, θ, et ε dans l'équation 3.1 permet de pondérer l'impact de chaque terme, offrant ainsi la flexibilité d'ajuster la contribution relative de chaque perte dans le processus d'apprentissage.

Pour identifier les paramètres optimaux de notre fonction de perte, nous utilisons l'optimisation bayésienne (Snoek *et al.*, 2012) afin d' estimer des intervalles de paramètres prometteurs, puis on met en œuvre la méthode de recherche en grille (Kohavi *et al.*, 1995) pour sélectionner les meilleurs paramètres. La performance des paramètres a été mesurée à l'aide du PSNR et du SSIM après 100 époques d'entrainement du modèle. Finalement, les résultats finaux

sont obtenus avec les valeurs optimales suivantes : α =2 $\times$ 10^{-8}, β =8 $\times$ 10^{-3}, γ=15 $\times$ 10^{-4}, λ =3 $\times$ 10^{-2}, θ = 1, et ε =5 $\times$ 10^{-3}.

3.2 FLATTEN-SWINIR

3.2.1 ARCHITECTURE GLOBALE

Après avoir explorer des modèles de super-résolution basés sur l'architecture des transformateurs de vision. Notre investigation nous a conduit au développement d'un second modèle nommé Flatten-SwinIR.

Le modèle Flatten-SwinIR se compose de trois modules principaux, comme illustré dans la Figure 3.4 : extraction de caractéristiques peu profondes, extraction de caractéristiques profondes, et le module de reconstruction d'images de haute qualité.

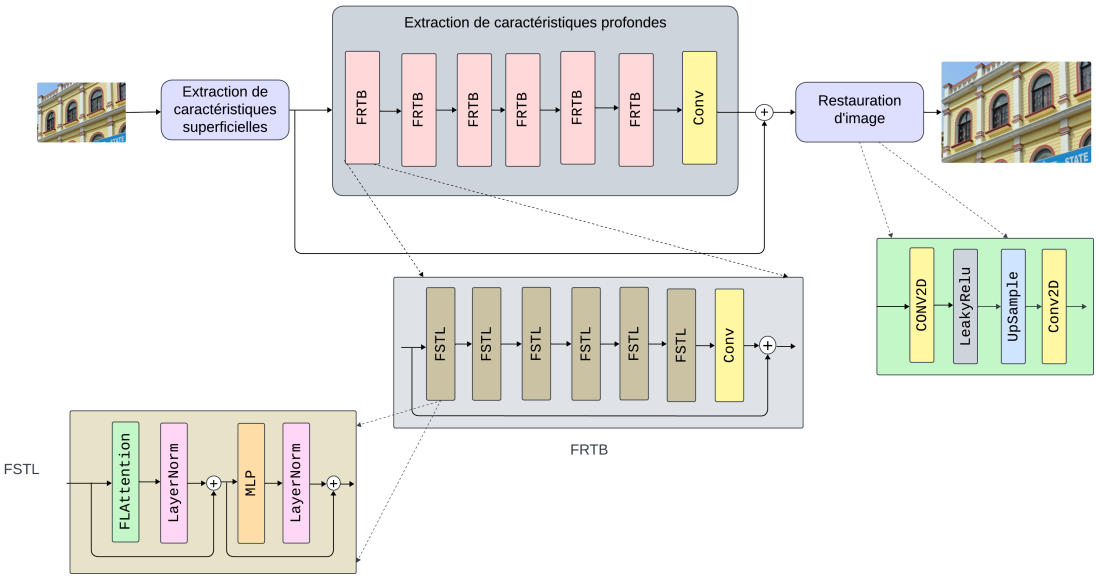

FIGURE 3.4 : Architecture du modèle Flatten-SwinIR

© Gildas Aimé Sedou Fofe

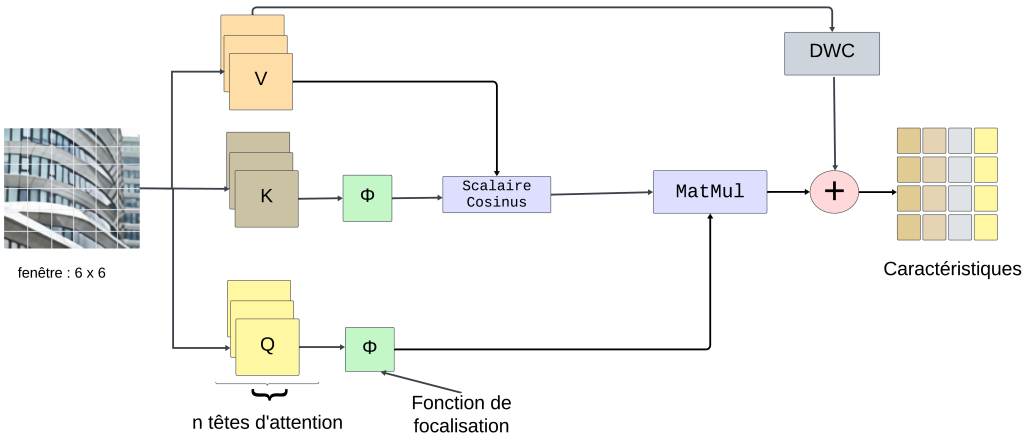

FIGURE 3.5 : Architecture du module Flatten Attention.

© Gildas Aimé Sedou Fofe

3.2.2 LE MODULE D'EXTRACTION DE CARACTÉRISTIQUES SUPERFICIELLES

L'objectif principal de ce module est de capturer les caractéristiques de bas niveau. Ce module se compose d'une seule couche, qui est une couche convolutive 3×3, utilisée pour capturer les caractéristiques locales de l'image et générer sa carte de caractéristiques peu profonde contenant des informations de bas niveau telles que les bords, les textures et d'autres détails locaux, qui sont utiles pour la restauration de l'image.

3.2.3 LE MODULE D'EXTRACTION DE CARACTÉRISTIQUES PROFONDES

Les caractéristiques peu profondes sont ensuite utilisées comme entrée dans le module d'extraction de caractéristiques profondes, qui se compose de blocs de couches résiduelles basées sur l'architecture d'un transformateur de vision appelé FSTL (Flatten Swin Transformer Layer). Nous appelons ces blocks, des blocs FRTB.

Chaque bloc FRTB permet d'extraire des caractéristiques plus complexes que les caracté-

ristiques superficielles, comme les structures hiérarchiques entre les caractéristiques, les interdépendances entre les pixels et les liaisons non linéaires entre les régions de l'image.

LES BLOCKS FRTB

Chaque block FRTB est un regroupement de transformateurs de vision nommé FSTL. Les blocks FRTB contiennemt des connexions résiduelles cela contribue à la stabilité de l'apprentissage en facilitant le flux d'informations à travers les blocs. Les caractéristiques profondes sont obtenues en utilisant tous les blocs FRTB de façon successif, où chaque bloc génère des caractéristiques intermédiaires $(Caract1, Caract2, \ldots, CaractK)$. La sortie finale du module est la caractéristique profonde, qui représente une représentation riche et complexe de l'image d'entrée.

LE MODULE FSTL (FLATTEN SWIN TRANSFORMER LAYER)

La transformateur de vision FSTL est le composant clé du modèle. Elle est utilisée pour effectuer des opérations d'auto-attention sur les données d'entrée.
Dans ce processus, l'image est divisée en plusieurs fragments (ou patchs) appelées fenêtres d'image locales. Chaque fragment est traité indépendamment en utilisant le mécanisme d'auto-attention appelé Flatten Attention, dont l'architecture est illustrée à la Figure 3.5. Ce mécanisme permet au modèle de se concentrer sur les relations à courte portée entre les fragments de l'image. Chaque fragment est aplati en un vecteur (on passe le fragment à travers une fonction de projection linéaire). Si un fragment a une taille de $M \times M$, alors le vecteur résultant aura une taille de M^2. Une fois les fragments extraits, ils sont projetés dans les espaces de requête (Q), clé (K) et valeur (V). Cette projection est réalisée en appliquant des transformations linéaires à chaque fragment d'image. Chaque fragment est alors traité comme

un jeton individuel, similaire aux mots dans le traitement de texte avec les transformateurs traditionnels.

Dans le modèle SwinIR, le mécanisme d'attention (appelé Windows Attention) a une complexité quadratique ; dans Flatten-SwinIR nous utilisons le mécanisme Flatten Attention, dont la complexité est linéaire. L'architecture du mécanisme d'Attention est représentée à la figure 3.5.

FLATTEN ATTENTION

Flatten Attention est un mécanisme d'attention inspiré par le mécanisme Focused Linear Attention décrit dans les travaux de Han *et al.* (2023). Nous avons apporté de légères modifications à ce mécanisme d'attention pour l'adapter à notre modèle de restauration d'image.

Les matrices Q, K et V sont obtenues en multipliant l'entrée X ($X \in \mathbb{R}^{N \times d}$, où N est le nombre de jetons ou de fragments d'images et d est la dimension de la projection lineaire appliqué sur X) par des matrices de poids W_Q, W_K, $W_V \in \mathbb{R}^{d \times d}$ (les valeurs des poids sont obtenues pendant l'apprentissage, leur valeur initiale est donnée aléatoirement).

L'attention utilisée par SwinIR et Swin2SR est une équation de la forme suivante :

$$Attention_{Quadrac}(Q,K,V) = \phi(QK^T)V + B(V) \tag{3.2}$$

Où $Q,K,V \in \mathbb{R}^{N \times d}$, $QK^T \in \mathbb{R}^{N \times N}$ et $\phi()$ est une fonction d'activation.

B est un biais qui aide à capturer les relations entre les éléments spatiaux de l'image.

Le produit entre une matrice de dimension $(N \times N)$ et une matrice de dimension $(N \times d)$ nécessite des opérations de l'ordre $O(N \times N \times d)$.

$Attention_{Quadrac}(Q,K,V)$ a une complexité par rapport au nombre de jetons N égale à $O(N^2d)$, il s'agit de la forme d'attention utilisé dans SwinIR et Swin2SR.

Cependant le mécanisme d'attention utilisé par Flatten Attention est un mécanisme d'attention linéaire car il est de la forme :

$$Attention_{Linear}(Q,K,V) = \phi(Q)(\phi(K^T)V) + h(V) \tag{3.3}$$

Où h est une fonction de stabilisation.
Dans l'attention linéaire, nous changeons l'ordre de multiplication pour réduire la complexité. Ainsi nous avons $\phi(Q) \in \mathbb{R}^{N\times d}$, et $\phi(K)^T V \in \mathbb{R}^{d\times d}$.
Le produit entre une matrice de dimension $(N \times d)$ et une matrice de dimension $(d \times d)$ nécessite des opérations $(N \times d \times d)$. Il utilise une complexité de $O(Nd^2)$.

Flatten Attention est un mécanisme d'attention linéaire qui ajuste l'organisation des requêtes et clés (Q, K, V) tout en permettant au modèle d'améliorer sa capacité d'extraction des caractéristiques. Afin que les paires de caractéristiques (Q, K, V) qui sont sensiblement égales soient proches en distance, tandis que les paires différentes seront plus éloignées. Cela rend les poids d'attention plus discriminants, permettant au modèle de se concentrer davantage sur les caractéristiques importantes et ainsi d'augmenter son efficacité.
Enfin, Flatten Attention utilise une couche de convolution en profondeur (DWC) (Han *et al.*, 2023) appliquée à la matrice des valeurs V. Cette couche aide à maintenir un rang élevé de la matrice d'attention linéaire, assurant une diversité des caractéristiques dans la sortie. En effet, plus le rang de la matrice de sortie de Flatten Attention est élevé, plus le modèle aura une information diverse sur l'image. Cependant, les mécanismes d'attention linéaire produisent des matrices de sortie avec des rangs faibles. La DWC permet au Flatten Attention de mitiger la perte de diversité des caractéristiques causée par l'attention linéaire, améliorant ainsi

l'expressivité du modèle tout en maintenant une complexité computationnelle raisonnable. L'opération à l'intérieur du Flatten Attention peut être représentée comme l'Équation 3.4 :

$$\text{Output} = \phi(Q)(\phi(K)^T V) + \text{DWC}(V) \tag{3.4}$$

- **Nombre de jetons N :** correspond au nombre de fragments en lesquels une image est divisée.
- **Dimension des caractéristiques d :** C'est la taille de chaque vecteur de jetons après la projection linéaire de chaque fragment.
- **Matrices Q, K, V :** $Q = XW_Q$, $K = XW_K$, $V = XW_V$ où X $\in \mathbb{R}^{N\times d}$ (valeur d'entrée)
- **Poids :** Les poids W_Q W_K W_V $\in \mathbb{R}^{d\times d}$ dans un module d'attention sont obtenus par l'apprentissage du modèle pendant l'entraînement.
- **Fonction de focalisation** $\phi()$ **:** les opérations $\phi(Q)$ et $\phi(K)$ représentent des transformations (fonction de concentration) appliquées à ces matrices pour capturer les relations entre les données.

  $\phi(Q) \in \mathbb{R}^{N\times d}$, $\phi(K)^T V \in \mathbb{R}^{d\times d}$, $Output \in \mathbb{R}^{N\times d^2}$

La complexité du mécanisme Flatten Attention est linéaire par rapport au nombre de jetons N car dans ce mécanisme nous évitons le produit scalaire QK^T (présent dans le mécanisme d'attention de SwinIR et Swin2SR). Au lieu de cela, nous faisons d'abord le produit scalaire : $\phi_p(K)^T V$, ce qui fait passer la complexité de $O(N^2 d)$ à $O(Nd^2)$.

Le mécanisme Flatten Attention a une complexité linéaire par rapport à N et une complexité quadratique par rapport à d.

Cependant, entre les variables N et d, celle qui nous intéresse le plus est N, car dans l'implémentation de notre modèle ainsi que dans celles de SwinIR et Swin2SR, la variable d est fixée à une valeur constante. Ainsi, si la taille de l'image augmente et que d (la taille d'un fragment) reste fixe, N augmente. Par conséquent, en réduisant la complexité de $O(N^2 d)$ à $O(Nd^2)$, nous

réduisons le coût de calcul de notre modèle pour des images de grande taille, contrairement à SwinIR et Swin2SR, dont le coût augmente.

PRODUIT SCALAIRE PAR COSINUS

Dans le calcul de l'auto-attention dans SwinIR, un simple produit scalaire est utilisé dans le produit entre les vecteurs de requête (Q) et de clé (K). Les auteurs de Swin2SR (Conde *et al.*, 2022), ont trouvé que lorsque cette approche est utilisée dans des modèles larges ou très profonds, les cartes d'attention apprises de certains blocs sont souvent dominées par quelques paires de pixels. Pour atténuer ce problème, ils utilisent un produit scalaire par cosinus (Liu *et al.*, 2022). Nous utilisons également cette approche comme illustré dans la Figure 3.5.
Ainsi le mécanisme "Flatten attention" que nous avons developpé est une implémentation du mécanisme "Focused Linear Attention" proposé par (Han *et al.*, 2023). Dans lequel nous avons utilisé un produit scalaire par cosinus.

FONCTION DE FOCALISATION

La fonction de focalisation est une fonction utilisée dans le cadre du mécanisme d'attention pour ajuster la direction des caractéristiques des requêtes et des clés, améliorant ainsi la capacité des modules d'attention linéaire à se concentrer sur des caractéristiques spécifiques. Le principal objectif de la fonction de focalisation est d'accentuer les similitudes et de minimiser les différences entre les caractéristiques des requêtes et des clés, permettant à l'attention linéaire de se concentrer sur les informations pertinentes dans les données d'entrée.
La fonction de focalisation comporte deux composants principaux :

- **Fonction d'activation ReLU :** Avant d'appliquer la fonction de focalisation, les caractéristiques des requêtes et des clés sont d'abord passées à travers une fonction

d'activation ReLU (Nair & Hinton, 2010). Cette fonction d'activation ReLU garantit que les valeurs des caractéristiques restent non négatives, ce qui est important pour garantir la validité des calculs ultérieurs.

- **Fonction de mappage f_p :** La fonction de mappage est la partie principale de la Fonction de focalisation. Elle est conçue pour ajuster la direction des caractéristiques des requêtes et des clés afin de les rendre plus similaires lorsque les caractéristiques sont pertinentes et moins similaires lorsque les caractéristiques sont différentes.

Nous avons utilisé la fonction focalisastion : $\phi_p(x) = f_p(\text{ReLU}(x))$ dont la formulation est donnée par l'Équation 3.5 :

$$f_p(x) = \frac{\|x\|}{\|x^{**p}\|} x^{**p} \tag{3.5}$$

Où x représente les caractéristiques des requêtes ou des clés et x^{**p} représente les caractéristiques des requêtes ou des clés élevées à la puissance p (facteur de focalisation). p est un paramètre contrôlant le degré d'ajustement effectué par la fonction f_p. $\|x\|$ représente la norme des caractéristiques.

L'effet de la fonction de focalisation est de pousser chaque vecteur de caractéristiques vers son axe le plus proche, réduisant les distances entre les caractéristiques similaires et augmentant les distances entre les caractéristiques différentes. En ajustant le paramètre p, nous pouvons contrôler la force de cet effet et, par conséquent, la capacité de l'attention linéaire à se focaliser.

PERCEPTRON MULTI-COUCHE

Une fois que le mécanisme d'attention a été appliquée, les caractéristiques résultantes sont passées à travers un perceptron multicouche (MLP). Le MLP se compose de deux couches entièrement connectées avec une fonction d'activation GELU (Hendrycks & Gimpel, 2016) entre elles . Cette structure permet au réseau d'apprendre à travers des transformations non linéaires, capturant les caractéristiques les plus riches parmi celles déjà extraites.

POST-NORMALISATION ET CONNEXIONS RÉSIDUELLES

Dans SwinIR, une normalisation est appliquée avant la couche MLP et la couche Attention pour stabiliser l'apprentissage en normalisant les activations. Cependant, les expériences de (Liu *et al.*, 2022) montrent que cette méthode atteint ses limites lorsque le modèle est plus complexe.

Flatten-SwinIR utilise une normalisation post-linéaire (après la couche MLP et la couche d'attention) comme dans Swin2SR. Cet approche stabilise l'apprentissage des modèles plus larges, stabilise les activations et améliore la convergence de l'apprentissage en réduisant le risque de saturation des activations.

Nous utisons aussi une connexion résiduelle, ce qui signifie que les caractéristiques d'entrée sont ajoutées aux sorties de la couche de normalisation. Cette connexion résiduelle évite les problèmes de disparition du gradient et facilite l'entraînement en profondeur. Le processus décrit peut être résumé par l'équation d'affectation suivante 3.6 :

$$X = \mathrm{LN}(\mathrm{MLP}(X)) + X. \tag{3.6}$$

Où LN représente la normalisation lineaire.

3.2.4 MODULE DE RESTAURATION

Le rôle du module de restauration encore appelé module de reconstruction d'image est de créer une version de haute qualité d'une image à partir de ses entrées, à savoir les caractéristiques superficielles et profondes. Nous utilisons des couches de convolution ainsi qu'une couche d'opération de suréchantillonnage pour augmenter la taille de l'image dans ce module.

3.2.5 FONCTION DE PERTE

La fonction de perte utilisée pour entraîner Flatten-SwinIR est la perte de pixel L1 encore appelé l'erreur absolue moyenne ou MAE.

$$L_1(I_{HR}, I_{SR}) = \frac{1}{N}\sum_{i=1}^{N} \left| I_{HR}^{i} - I_{SR}^{i} \right| \tag{3.7}$$

Cette perte est utilisé car elle est moins sensible aux valeurs aberrantes et favorise la préservation des détails et des contours dans les images. Le modèle est ainsi encouragé à produire des valeurs de pixel proches de celles de l'image haute résolution réelle. Cela contribue à préserver les détails locaux et les contours.

3.3 CONCLUSION

Dans ce chapitre nous avons présenté l'architecture des deux modèles de super-résolution proposés, SIR-SRGAN-ResNeXt et Flatten-SwinIR, destinés à améliorer la qualité des images médicales. SIR-SRGAN-ResNeXt est une amélioration du modèle SIR-SRGAN, il utilise un générateur basé sur ResNeXt pour une meilleure capacité d'apprentissage des caractéristiques, un discriminateur U-Net pour une évaluation plus robuste, et une fonction de perte qui est une

combinaison de plusieurs autres fonctions de perte afin d'optimiser la qualité de la perception visuelle des images super-résolues.

Flatten-SwinIR, quant à lui, utilise une architecture de transformateur de vision avec un mécanisme d'attention linéaire appelé Flatten Attention, conçu pour réduire la complexité computationnelle. Ce modèle est composé de trois modules principaux : un module d'extraction de caractéristiques superficielles, un autre pour les caractéristiques profondes, et un module de reconstruction d'images.

CHAPITRE IV

EXPÉRIMENTATIONS ET RÉSULTATS

4.1 SIR-SRGAN-RESNEXT

Dans cette section, nous présentons les jeux de données utilisés et la stratégie de validation pour entraîner et tester le modèle, puis nous discutons de la performance du SIR-SRGAN-ResNeXt proposé. Nous évaluons d'abord le modèle proposé en termes de performance haute résolution en utilisant les mesures PSNR, SSIM, LPIPS, HaarPSI et ClipIQA, puis en termes de taille du fichier d'image de sortie et de coût de calcul.

4.1.1 PRÉPARATION DES ENSEMBLES DE DONNÉES

Les expériences sont menées sur cinq ensembles d'images de référence médicales et générales, et une comparaison avec les modèles de pointe de super-résolution baées sur l'architecture GAN, à savoir SRGAN, RankSRGAN, BSRGAN, SRGAN-ResNeXt, et SIR-SRGAN, est effectuée.

Pour entraîner et évaluer le modèle proposé, nous avons considéré six ensembles de données : deux pour l'entraînement (DIV2K [1] et Flickr2K [2]) et les quatre autres pour les tests (BSD100 [3], Messidor-2 [4], URBAN100 [5] et Breakhis-400x [6]). Les détails et caractéristiques de chaque ensemble de données sont présentés dans le Tableau 4.1. Pour augmenter l'ensemble de données

1. https://www.kaggle.com/datasets/joe1995/div2k-dataset
2. https://www.kaggle.com/datasets/daehoyang/flickr2k
3. https://www.kaggle.com/datasets/asilva1691/bsd100
4. https://www.adcis.net/en/third-party/messidor2/
5. https://www.kaggle.com/datasets/harshraone/urban100
6. https://www.kaggle.com/datasets/forderation/breakhis-400x

d'entraînement, chaque image est partitionnée en fragments (patchs) de taille 192×192. Et pour chaque époque, un échantillon est extrait aléatoirement de l'ensemble d'entraînement et utilisé pour entraîner le modèle.

Les images de l'ensemble de données URBAN100 sont trop grandes pour être testées sur notre ordinateur, nous avons donc également redimensionné ces images comme indiqué dans le Tableau 4.1. Les ensembles de données DIV2K et Flickr2K sont utilisés exclusivement pour entraîner le modèle. Avec ces deux ensembles de données, nous avons 3550 images utilisées pour l'entraînement.

Pour améliorer la robustesse de notre modèle et éviter le surapprentissage, nous avons employé une stratégie de validation croisée à K-Blocs (encore appelé K-Fold Validation) avec 10 blocs. Ainsi, pour chaque époque, nous avons 355 images pour la validation et 3195 images pour l'entraînement.

Après avoir appliqué le mécanisme d'augmentation des données (chaque image d'entraînement est divisée en fragments de taille 192×192), nous avons un total de $223,650$ images d'entraînement. Cette approche nous permet de maximiser l'utilisation de nos ressources et d'optimiser le potentiel d'apprentissage de notre modèle de super-résolution.

Le modèle proposé est conçu pour améliorer la résolution des images. Pour cela, nous avons évalué ses performances sur les mesures les plus pertinentes pour mesurer la qualité et la résolution des images, à savoir ClipIQA Wang *et al.* (2022), LPIPS Zhang *et al.* (2018a), HaarPSI Reisenhofer *et al.* (2018b), PSNR Kim (1988), et SSIM Wang *et al.* (2004). De plus, l'espace de stockage des images est un aspect essentiel, nous avons donc également exploré la taille des fichiers de sortie des images améliorées. Enfin, nous analysons le temps de calcul que chaque modèle nécessite pour traiter chacun des quatre ensembles de données d'images de test.

TABLEAU 4.1 : Jeu de données d'entrainement et de test.

Jeu de données	Description	Taille après réduction (Pixel)	Nombre de fichiers	Taille Originale (Pixel)	Usage
DIV2K Agustsson & Timofte (2017)	Images à haute résolution	2040×1400	128×128	900	Entrainement
Flickr2K (Timofte *et al.*, 2017)	Images à haute résolution	2040×1356	192×192	2650	Entrainement
BSD100 Arbeláez *et al.* (2011)	Image de type générale	480×320	—	100	Test
Messidor-2 Decencière *et al.* (2014)	Images de rétinopathie diabétique	512×512	—	250	Test
URBAN100 Huang *et al.* (2015)	Image de type générale	984×796	245×198	100	Test
Breakhis-400x Spanhol *et al.* (2016)	Image de cellules cancéreuses	700×460	—	176	Test

4.1.2 RÉSULTATS

Toutes les expériences ont été menées sur un ordinateur équipé d'un GPU NVidia A100SXM4 et de 16GB de mémoire. Nous avons entraîné notre modèle sur 1000 époques avec une taille de lot de 12 images.

Les résultats visuels sur les quatre ensembles de test sont illustrés dans les figures 4.1 et 4.2. Visuellement, tous les modèles renvoient des images super-résolues assez similaires, sauf BSRGAN qui génère des images significativement plus lisses. Les gros plans d'images générées sur les quatre ensembles de test montrent que SIR-SRGAN-ResNeXt semble générer des images haute résolution avec une bonne définition des bords et des textures.
Un aperçu plus approfondi des résultats est donné dans le Tableau 4.2 présentant les performances des différents modèles sur les quatre ensembles de test et basé sur les mesures d'évaluation. Nous mettons en évidence les meilleures mesures en **gras** et les deuxièmes meilleures en ***gras-italique***.

TABLEAU 4.2 : Comparaison des performances de l'amélioration d'image sur les quatre ensembles de données de test.

Model	PSNR	SSIM	LPIPS	HaarPSI	ClipIQA	Output File Size (MB)	Time (s)
BSD100 data set							
SRGAN	27.517	***0.9087***	0.0677	0.8874	0.3605	***21.0***	14
BSRGAN	24.932	0.8296	0.2358	0.8514	***0.5699***	**16.2**	15
SRGAN-ResNeXt	27.686	0.9077	***0.0675***	0.8900	0.3502	21.2	***12***
RANK-SRGAN	25.054	0.8582	0.0788	0.8484	**0.7248**	29.7	**9**
SIR-SRGAN	***27.705***	0.9084	***0.0675***	***0.8907***	0.3683	21.1	13
SIR-SRGAN-ResNeXt	**27.816**	**0.9099**	**0.0660**	**0.8929**	0.4533	***21.0***	***12***
URBAN100 data set							
SRGAN	24.156	0.8690	0.1071	0.8637	0.4066	***7.77***	**5**
BSRGAN	21.216	0.7378	0.2746	0.7937	0.4883	**6.77**	12
SRGAN-ResNeXt	***24.353***	***0.8715***	***0.0951***	***0.8696***	0.4678	7.90	***6***
RANK-SRGAN	22.804	0.8420	**0.0932**	0.8315	**0.6002**	10.3	***6***
SIR-SRGAN	24.327	0.8704	0.0997	0.8687	0.4506	7.81	**5**
SIR-SRGAN-ResNeXt	**24.500**	**0.8746**	0.0980	**0.8730**	***0.5132***	7.79	***6***
Messidor-2 data set							
SRGAN	40.730	0.9788	***0.0131***	0.9888	0.4142	***58.8***	56
BSRGAN	35.001	0.9266	0.1169	0.9682	**0.6588**	**34.2**	***46***
SRGAN_ResNeXt	40.764	***0.9806***	0.0144	0.9886	0.3562	60,0	53
RANKSRGAN	39.457	0.9700	0.0167	0.9851	0.4941	82.6	**17**
SIR_SRGAN	***40.882***	0.9802	0.0160	***0.9896***	***0.5310***	59.4	54
SIR-SRGAN-ResNeXt	**42.079**	**0.9827**	**0.0121**	**0.9918**	0.4358	60.2	56
Breakhis-400x data set							
SRGAN	35.6487	0.9879	0.0122	0.9628	0.0980	71.5	50
BSRGAN	28.7088	0.9492	0.0970	0.9434	**0.4320**	**66.5**	***43***
SRGAN-ResNeXt	36.4377	0.9884	0.0114	0.9769	0.0860	71.1	49
RANK-SRGAN	33.6930	0.9788	0.0206	0.9499	***0.2090***	93,1	**16**
SIR-SRGAN	***36.8844***	***0.9886***	***0.0110***	***0.9793***	0.1010	***70.4***	49
SIR-SRGAN-ResNeXt	**37.2361**	**0.9891**	**0.0101**	**0.9822**	0.1169	***70.4***	48

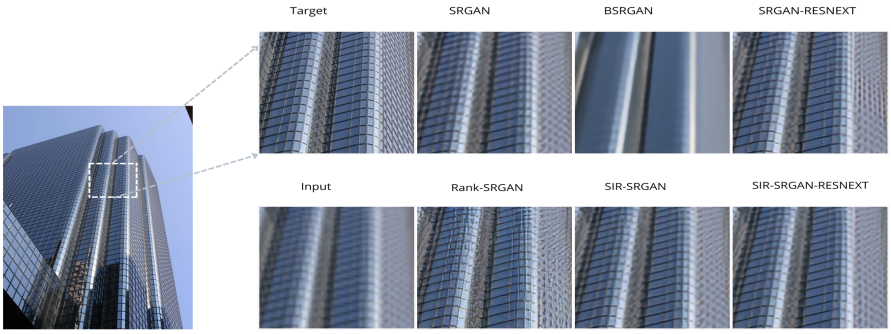

FIGURE 4.1 : Comparaison visuelle des différents modèles de super-résolution (×4) sur des inages de l'ensemble de données URBAN100.

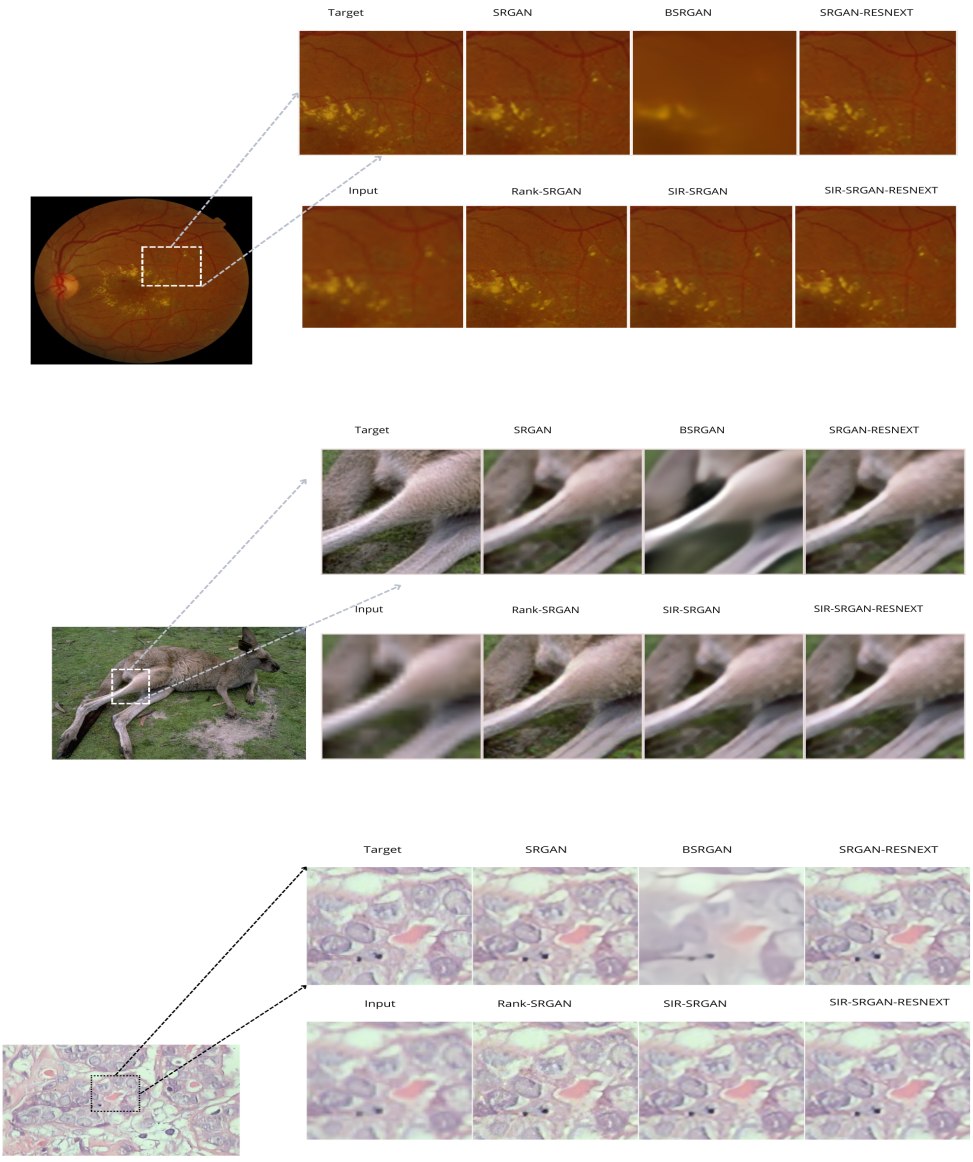

FIGURE 4.2 : Comparaison visuelle des différents modèles de super-résolution (×4) sur des images des ensembles de données Messidor, BSD1100 et Breakhis-400x.

Le modèle SIR-SRGAN-ResNeXt démontre des performances supérieures sur plusieurs mesures. Pour l'ensemble de données BSD100, il atteint les meilleurs scores de PSNR (27.816) et SSIM (0.9099), et la plus faible valeur de LPIPS (0.0660), indiquant une meilleure

qualité perceptuelle. Sur URBAN100, il obtient le meilleur PSNR (24.500) et SSIM (0.8746), surpassant les autres modèles en termes de qualité d'image. Pour Messidor-2, il enregistre un PSNR de 42.079 et un SSIM de 0.9827, les plus élevés parmi tous les modèles testés, ainsi qu'une LPIPS de 0.0121 et une HaarPSI de 0.9918, confirmant son excellence en termes de qualité visuelle. Enfin, sur Breakhis-400x, il maintient sa domination avec un PSNR de 37.2361, un SSIM de 0.9891, et une LPIPS de 0.0101.

De plus, il parvient à équilibrer la taille des fichiers de sortie et le temps de traitement, ce qui en fait une solution robuste et efficace pour la super-résolution d'images.

Le jeu de données URBAN100 est particulièrement difficile car il contient des images avec de multiples hautes fréquences, à savoir des objets avec des bordures. Le modèle SIR-SRGAN-ResNeXt obtient les meilleures performances sur la plupart des mesures basées sur les pixels avec une taille de fichier de sortie relativement faible.

Le jeu de données Breakhis-400x est un ensemble de scans de cancer où les images englobent de nombreuses formes avec des bords visibles (par exemple, des cellules). De plus, Messidor-2, un ensemble de données composé d'images de fond d'œil utilisées pour la prédiction précoce de la rétinopathie diabétique, montre l'existence de veines sanguines très délicates, cruciales à mettre en évidence mais non visibles à l'œil nu. Sur les deux ensembles de données médicales, notre modèle donne les meilleures mesures sauf pour ClipIQA. L'image n'est pas considérée comme assez esthétique du point de vue visuel humain, mais elle est la plus proche de l'image réelle selon les mesures basées sur les pixels.

Pour la plupart, le modèle proposé SIR-SRGAN-ResNeXt se distingue de manière significative, offrant de très bonnes performances avec une taille de fichier de sortie acceptable. Cela en fait un choix particulièrement bon pour les applications traitant de grandes quantités de données, où la conservation de l'espace de stockage est cruciale. Les résultats garantissent également la supériorité du SIR-SRGAN-ResNeXt en termes de mesures basées sur les pixels PSNR, SSIM, LPIPS et HaarPSI. Ces mesures témoignent de la qualité des images générées par le

modèle, soulignant sa capacité à améliorer la résolution tout en maintenant une taille de fichier modeste et un temps d'éxécution semblable à ceux du SRGAN et SIR-SRGAN.

Comparativement, le modèle BSRGAN a produit des images excessivement lisses. Cela lui donne une bonne mesure CLI-IQA mais au détriment des performances en PSNR, SSIM, LPIPS et HaarPSI. Les auteurs de l'article sur BSRGAN ont axé leur application sur les visages humains et les images en mouvement, ce qui limite l'efficacité du modèle sur les images médicales.

En ce qui concerne Rank-SRGAN, bien qu'il ait obtenu un score élevé en ClipIQA en raison de la réduction du flou considérable dans les images, il a été observé que cette amélioration était accompagnée d'une tordure des lignes dans l'image, ce qui fait en sorte que l'image générée s'éloigne de la cible (Target). Cette déformatiom peut être préjudiciable dans les applications médicales nécessitant une analyse minutieuse.

Le SIR-SRGAN-ResNeXt se distingue également par son compromis optimal entre une taille de fichier raisonnable et de très bonnes performances concernant les diverses mesures. Cette performance s'est avérée particulièrement remarquable lors des tests sur des images Breakhis-400x, sur les tranches d'images de cancer.

En conséquence, nous suggérons le SIR-SRGAN-ResNeXt comme une solution prometteuse pour les applications de super-résolution, combinant efficacité de stockage, temps de calcul raisonnable et haute qualité de l'image, en particulier pour des images médicales cruciales et méticuleuses telles les images histopathologiques et les images du fond de l'œil.

4.1.3 ÉTUDE D'ABLATION

Dans l'étude d'ablation réalisée, des paramètres clés et des composants ont été expérimentés pendant un maximum de 150 époques (encore appelé epochs) pour observer leur

impact sur les performances du modèle SIR-SRGAN-ResNeXt proposé. L'étude a évalué les mesures de qualité de l'image générée et la stabilité du modèle pendant l'entraînement.

- **Stabilité du discriminateur**

  La première expérience a été menée pour trouver le meilleur discriminateur pour stabiliser l'apprentissage du GAN. L'expérience a été réalisée en utilisant trois discriminateurs différents, U-Net, PatchGAN et ViT. Ils ont été évaluée sur le jeu de données Urban100. Les résultats de la Figure 4.4 ont montré que les mesures PSNR et SSIM étaient plus stables lorsque le discriminateur U-Net était utilisé par rapport aux deux autres discriminateurs.

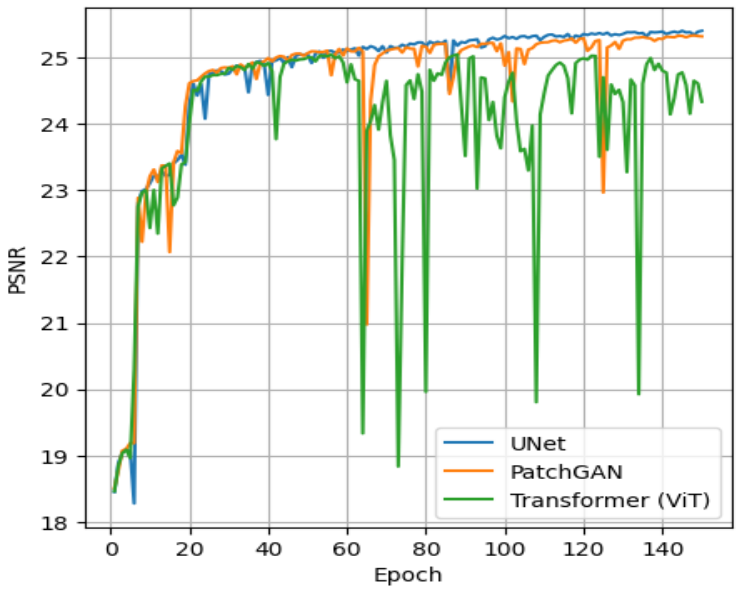

FIGURE 4.3 : Evolution de la stabilité des mesures PSNR durant l'entrainement avec différents discriminateurs.

© Gildas Aimé Sedou Fofe

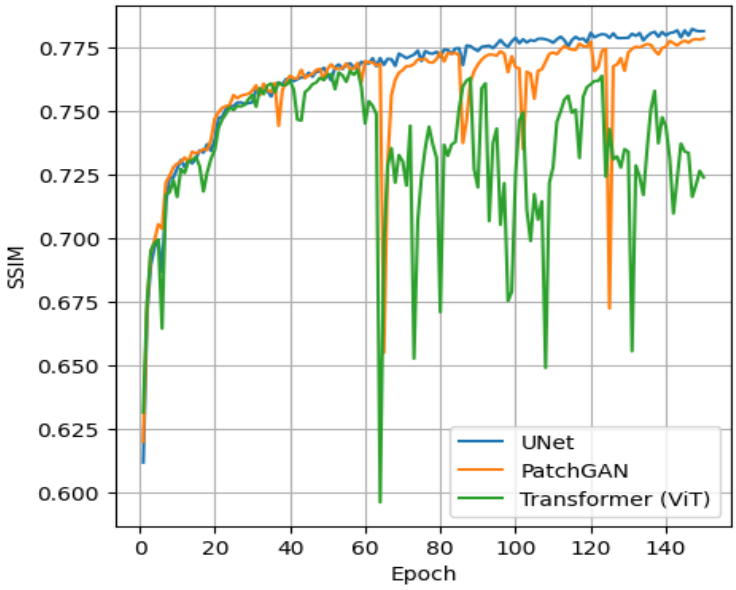

FIGURE 4.4 : Evolution de la stabilité des mesures SSIM durant l'entrainement avec différents discriminateurs.

- **Nombre de couches d'attention**

  Ensuite, nous avons évalué l'impact du nombre de couches d'attention dans le discriminateur U-Net, sur la qualité de l'image avec le jeu de données Urban100. L'étude a testé jusqu'à 3 couches d'attention et a constaté, comme le montre la figure 4.5, que PSNR et HaarPSI augmentent avec le nombre de couches d'attention, mais nous avons les meilleurs valeurs du SSIM et LPIPS lorsqu'on a 2 couches.

  Pour cela nous avons choisi 2 couches d'attention pour le discrininateur de notre modèle.

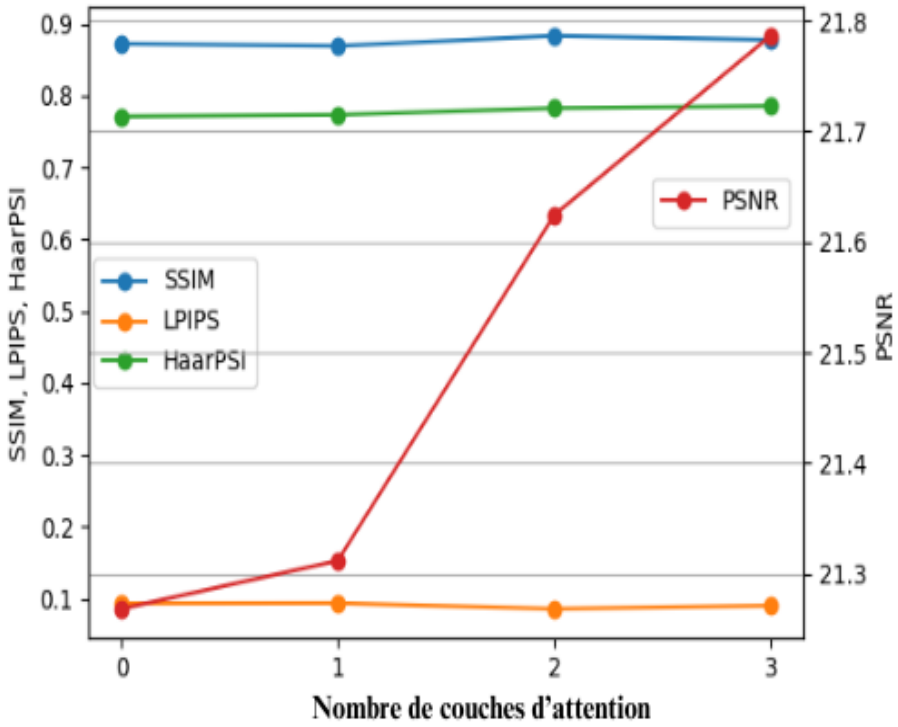

FIGURE 4.5 : Evolution des mesures au regard du nombre de couches d'attention.

© Gildas Aimé Sedou Fofe

- **Cardinalité**

  Le modèle ResNeXt effectue ses transformations sur des blocs de couches en parrallèle, le nombre de blocks en parrallèle s'appelle cardinalité comme indiqué dans la figure 3.2.

  La figure 4.6 montre qu'une cardinalité de 64 donne les meilleures valeurs PSNR et SSIM, tandis qu'une cardinalité de 16 donne également de bons résultats. Cependant, une cardinalité trop basse ou trop élevée ne donne pas de bons résultats. Bien qu'une cardinalité de 64 soit idéale, elle nécessite un GPU avec au moins 32 Go de mémoire pour l'entraînement et une cardinalité de 16 nécessite 16 Go de mémoire. Pour une comparaison équitable de notre modèle avec ses concurrents SRGAN, SIR-SRGAN, nous avons opté pour une cardinalité de 16. Afin que notre modèle ne soit pas trop large

et que son nombre de paramètres soit à peu près semblable à ceux de ses concurrents pour une comparaison juste. De plus nous avions un accès limité aux GPU de 32 Go de mémoire.

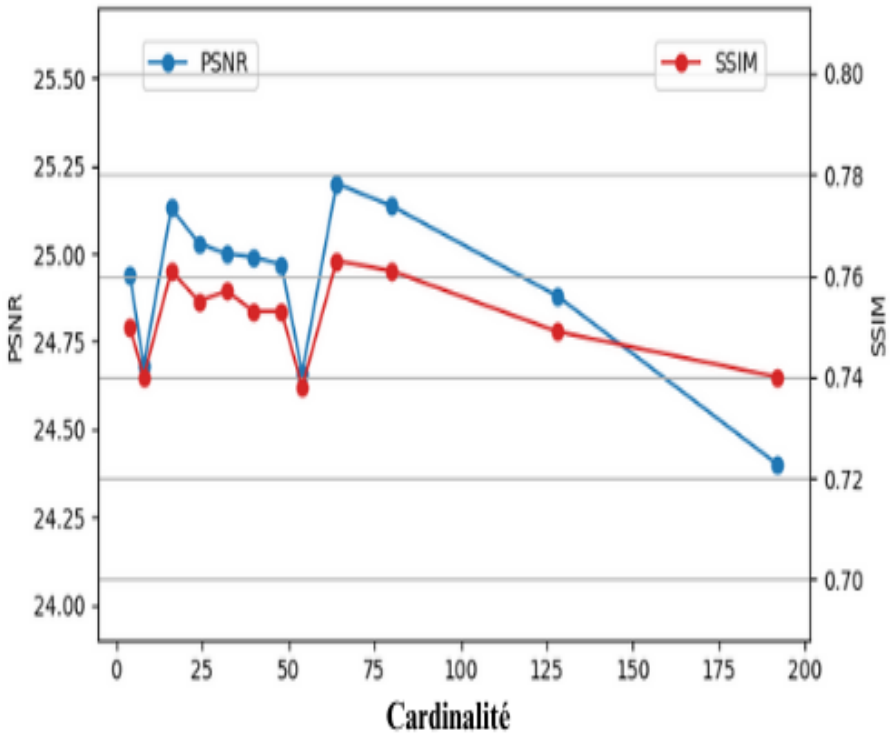

FIGURE 4.6 : Evolution des mesures au regard de la cardinalité

© Gildas Aimé Sedou Fofe

- **Nombre de blocs ResNeXtBlocks**

Nous avons également étudié l'influence de la variation du nombre de ResNeXtBlocks sur la qualité des images générées. Différentes versions du modèle ont été entraînées sur 100 époques en variant ce paramètre. Et il a été observé dans la Figure 4.7 que PSNR et SSIM étaient meilleurs lorsque le nombre de blocs était de 8, 16 ou 32. Un diviseur de 64 est requis pour ce nombre de blocs, car le nombre de canaux d'entrée et de sortie

des ResNeXtBlocks est de 64. Nous avons choisi un nombre de blocs égal à 8 pour les mêmes raisons que l'étude de la cardinalité.

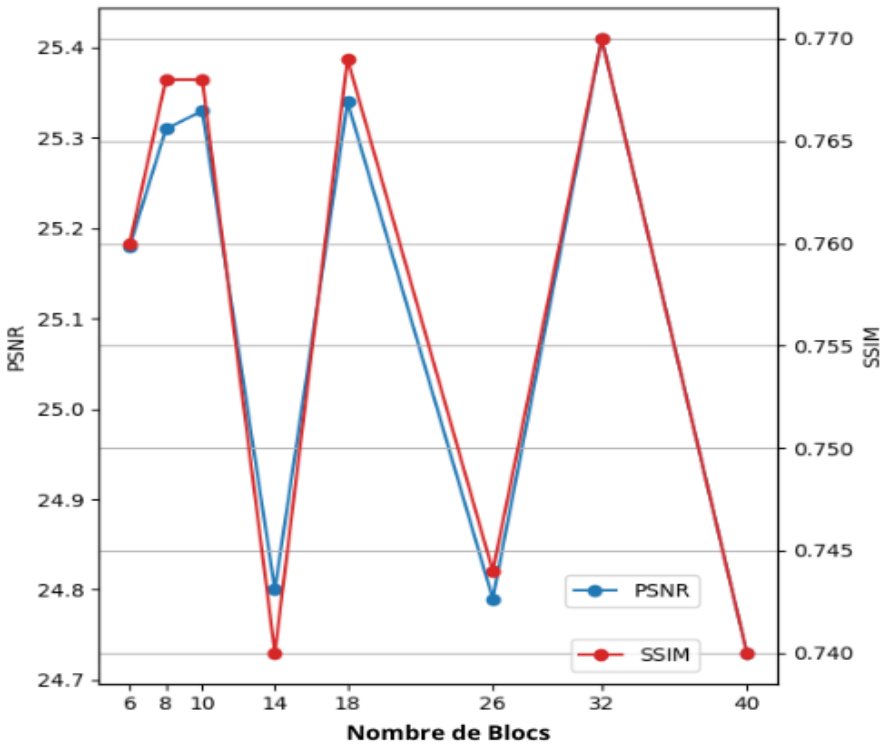

FIGURE 4.7 : Evolution des mesures au regard du nombre de blocs.

© Gildas Aimé Sedou Fofe

- **Taille du lot**

Enfin, nous avons examiné l'entraînement d'un modèle sur 80 époques en modifiant la taille du lot. Le jeu de données de validation Div2k a été utilisé pour effectuer les tests pendant l'entraînement.

La figure 4.18 montre de bonnes valeurs PSNR et SSIM avec des tailles de lot plus petites (par exemple la taille 4). Cependant, une très petite taille de lot entraîne un temps d'entraînement plus long. Par conséquent, une taille de lot de 12 a été choisie pour converger plus rapidement vers un modèle optimal.

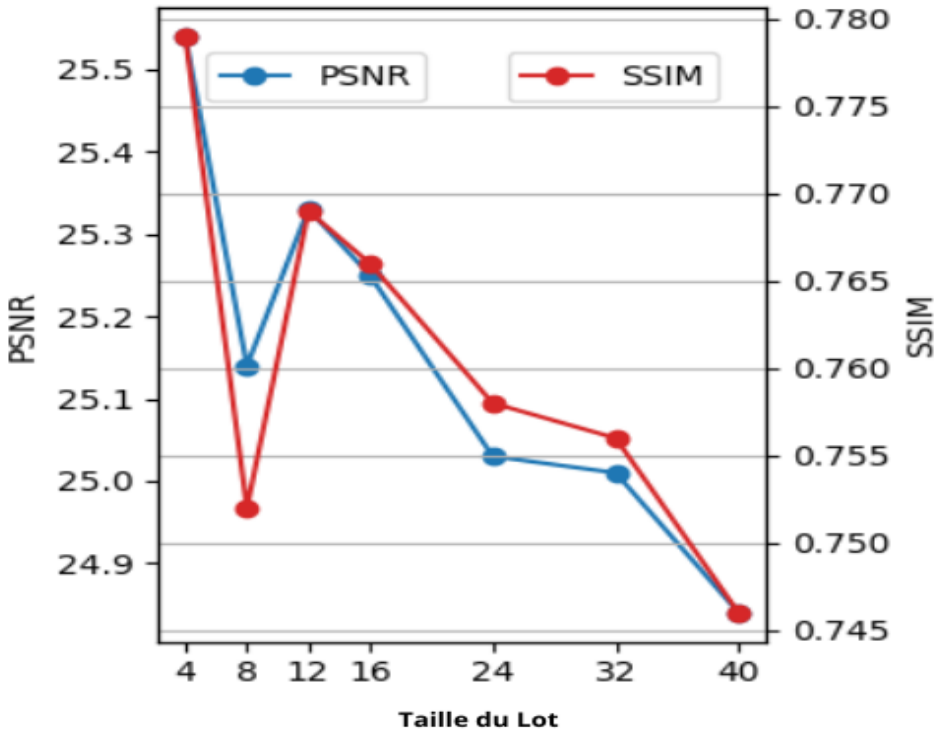

FIGURE 4.8 : Evolution des mesures au regard de la taille du lot.

© Gildas Aimé Sedou Fofe

4.2 FLATTEN-SWINIR

4.2.1 PRÉPARATION DES ENSEMBLES DE DONNÉES

Les expériences sont menées sur dix ensembles de données constitués d'images de référence médicales et générales, et une comparaison avec les modèles de pointe de super-résolution basées sur l'architecture GAN, le mécanisme d'attention et les transformateurs de vision.

Sur les dix ensembles de données, nous utilisons deux pour l'entraînement (Flickr2K [1] DIV2K [2]) et les huit restants pour les tests (BSD100 [3], Messidor-2 [4], Breakhis-400x [5], URBAN100 [6], CBSD68 [7], BSD68 [8], Kodak24 [9], Set12 [10]).

Les détails et les caractéristiques de chaque ensemble de données sont présentés dans le Tableau 4.3. Pour augmenter l'ensemble de données d'entraînement, chaque image est partitionnée en fragments (patchs) de taille 192×192.

Les transformateurs de vision se caractérisent par leur architecture complexe et la présence d'un grand nombre de paramètres. L'utilisation de la parallélisation distribuée devient cruciale pour accélérer l'entraînement de ces modèles sur de grands ensembles de données.

Ainsi, les modèles basés sur les transformateurs de visions ont été entraînés en utilisant la technique DistributedDataParallel (Li *et al.*, 2020) de la librairie PyTorch.

Les ensembles de données DIV2K et Flickr2K sont utilisés exclusivement pour l'entraînement du modèle. Avec ces deux ensembles de données, nous disposons de 3550 images utilisées pour l'entraînement. Nous avons utilisé une stratégie de Validation croisée à k blocs, avec 10 blocs.

Ainsi, pour chaque époque, nous avons 355 images pour la validation et 3195 images pour l'entraînement. Après avoir appliqué le mécanisme d'augmentation des données (chaque

1. https://www.kaggle.com/datasets/daehoyang/flickr2k
2. https://www.kaggle.com/datasets/joe1995/div2k-dataset
3. https://www.kaggle.com/datasets/asilva1691/bsd100
4. https://www.adcis.net/en/third-party/messidor2/
5. https://www.kaggle.com/datasets/forderation/breakhis-400x
6. https://www.kaggle.com/datasets/harshraone/urban100
7. https://www.kaggle.com/datasets/tarekmebrouk/cbsd68
8. https://github.com/smartboy110/denoising-datasets/tree/main/BSD68
9. https://www.kaggle.com/datasets/sherylmehta/kodak-dataset
10. https://https://www.kaggle.com/datasets/leweihua/set12-231008

image d'entraînement est divisée en fragments de taille 192×192), nous avons aussi un total de $223,650$ images d'entraînement.

TABLEAU 4.3 : Ensembles de données utilisés.

Ensemble de données	Description	Taille d'origine (pixels)	Taille réduite (pixels)	Taille de l'échantillon	Utilisation
Flickr2K (Timofte *et al.*, 2017)	Images haute résolution	2040×1356	192×192	2650	Entraînement
DIV2K (Agustsson & Timofte, 2017)	Images haute résolution	2040×1400	192×192	900	Entraînement
BSD100 (Arbeláez *et al.*, 2011)	Images de type générale	480×320	—	100	Test
Messidor-2 (Decencière *et al.*, 2014)	Images de rétinopathie diabétique	512×512	—	250	Test
Breakhis-400x (Spanhol *et al.*, 2016)	Images de cellules cancéreuses	700×460	—	176	Test
URBAN100 (Huang *et al.*, 2015)	Images de type générale	984×796	245×198	100	Test
CBSD68 (Arbeláez *et al.*, 2011)	Images de type générale	481×321	—	68	Test
BSD68 (Arbeláez *et al.*, 2011)	Images en niveaux de gris	481×321	—	68	Test
Kodak24 (Franzen, 1999)	Images de type générale	256×256	—	24	Test
Set12 (Sun *et al.*, 2008)	Images en niveaux de gris	512×512	—	12	Test

4.2.2 RÉSULTATS

Flatten-SwinIR, ses concurrents SwinIR, Swin2SR et les modèles basés sur un mécanisme d'attention personnalisé, ont une architecture plus complexe que les GANs précédents, à cause de cela ils utilisent beaucoup plus de mémoire. Alors nous les avons entrainé sur un ordinateur avec 04 GPU NVidia A100SXM4 de 16 Go de mémoire chacun. Nous utilisons la technique DistributedDataParallel (Li *et al.*, 2020) de la librairie PyTorch pour répartir les données sur les GPUs lors de l'entrainement. Nous avons aussi entraîné ces modèles sur 1000 époques avec une taille de lot de 16 images.

PERFORMANCES EN SUPER-RÉSOLUTION

Les résultats visuels globaux des performances de super-résolution sur quatre des ensembles de test sont présentés dans les figures 4.9 à 4.11 . La performance visuelle montrée dans la figure 4.9 démontre que Flatten-SwinIR produit des bords nets très similaires à ceux de l'image cible. Les modèles basés sur GAN et sur un mécanisme d'attention personnalisé

conduisent cependant à une moins bonne qualité visuelle, en particulier aux bords des objets dans l'image.

Bien que RankSRGAN génère des images moins floues, ses images de sortie ont des bords dégradés qui ne ressemblent pas à ceux de l'image originale, et seul Swin2SR offre une performance d'amélioration visuellement proche de celle de Flatten-SwinIR.

La figure 4.12 montre que les méthodes basées sur GAN donnent des résultats avec un flou substantiel, sauf pour BSRGAN, qui produit une image excessivement lisse et très éloignée de l'image cible. Les résultats de RankSRGAN, cependant, sont visuellement similaires à ceux des méthodes basées sur les couches d'attention, SwinIR, Swin2SR, et Flatten-SwinIR. Finalement, Flatten-SwinIR et Swin2SR sont plus performants pour accentuer les petites terminaisons nerveuses de l'œil.

La figure 4.10 permet de voir les résultats des modèles à partir du jeu de données Breakhis-400x. Parmi les méthodes discutées, seules RankSRGAN et les modèles utilisant les couches d'attention et les transformers génèrent des images avec un flou minimal. Les résultats de ces méthodes sont visuellement similaires. Cependant, NLSA, Swin2SR, et Flatten-SwinIR améliorent efficacement la distinction des bords des cellules cancéreuses.

Néanmoins, la figure 4.11 montre que, à l'exception de BSRGAN, les modèles basés sur GAN surpassent les modèles de transformateurs de vision et ceux basés sur l'attention. Nous observons la performance particulière de RankSRGAN dans le défloutage et la mise en valeur des détails du pelage.

Une analyse approfondie des résultats de super-résolution est donnée dans le tableau 2, montrant les performances des différents modèles sur les ensembles de données Breakhis-400x, BSD100, Messidor-2, et URBAN100, et basées sur les mesures d'évaluation. Nous soulignons les meilleures mesures en gras et les secondes meilleures en italique gras.

Les résultats montrent clairement que BSRGAN possède une caractéristique particulière qui le distingue des autres modèles, à savoir une taille de fichier de sortie très petite sur des

images de type général et médical. Cette réduction de la taille de fichier est attribuée au lissage excessif des images effectué par BSRGAN pour réduire le bruit et les détails fins dans l'image. Cependant, cette stratégie conduit à une augmentation significative du score ClipIQA sur les ensembles de données d'images médicales (Messidor-2, Breakhis-400x), qui mesure la qualité visuelle des images basée sur la perception humaine.

Ce résultat montre que le lissage excessif effectué par BSRGAN peut rendre les images médicales visuellement très nettes, mais au détriment de la fidélité à l'image originale et de la préservation des détails.

Cette tendance est confirmée par l'analyse des mesures PSNR, SSIM, LPIPS et HaarPSI, où BSRGAN obtient des scores inférieurs à la plupart des autres modèles. Ces mesures évaluent mieux la fidélité de la reconstruction par rapport à l'original.

Le fait que BSRGAN obtienne de mauvais résultats sur ces mesures suggère que le lissage excessif détruit des informations pertinentes dans les images, les éloignant ainsi de la cible souhaitée. Par exemple, sur l'ensemble de données Breakhis-400x, BSRGAN obtient un PSNR de 28.7088, un SSIM de 0.9492, un LPIPS de 0.097 et un HaarPSI de 0.9434, tandis que Flatten-SwinIR obtient respectivement 38.471, 0.9907, 0.0088 et 0.9865, soutenant un compromis entre la taille du fichier et la qualité de l'image, alors que BSRGAN privilégie la clarté visuelle par rapport à la réplication exacte de l'image.

Le Tableau 4.4 montre que le modèle Flatten-SwinIR se distingue considérablement par ses performances sur les différentes mesures d'évaluation. Sur l'ensemble de données Breakhis-400x, composé d'images microscopiques de tissus cancéreux, Flatten-SwinIR atteint le PSNR le plus élevé (38.471), le SSIM le plus élevé (0.9907), le LPIPS le plus bas (0.0088), le HaarPSI le plus élevé (0.9865) et le ClipIQA le plus élevé (0.403). Ces indicateurs mettent en évidence la capacité de Flatten-SwinIR à produire des images améliorées de haute qualité avec une fidélité remarquable à l'original. De plus, la performance de Flatten-SwinIR est également remarquable sur les ensembles de données URBAN100, Messidor-2 et BSD100. Dans

l'ensemble, Flatten-SwinIR se classe parmi les meilleurs modèles, démontrant sa robustesse et sa généralisation à différents types d'images.

D'autre part, les résultats montrent également un écart significatif dans le temps d'exécution entre les différents modèles, les modèles basés sur GAN ont généralement des temps d'exécution plus courts que les autres. Cela peut s'expliquer par le fait que les méthodes basées sur GAN utilisent principalement des couches convolutionnelles, qui sont moins coûteuses en termes de temps de calcul.

En revanche, les modèles utilisant des couches d'attention personnalisées, comme NLSA et HAN, ont les temps d'exécution les plus longs. Cette observation découle du fait que ces modèles empilent résiduellement plusieurs couches d'attention complexes, ce qui nécessite plus de temps de calcul.

De plus, bien que les modèles basés sur les transformateurs de vision prennent plus de temps à exécuter, ils améliorent significativement la qualité des images. Dans l'ensemble, les modèles basés sur les transformateurs tendent à générer des images de meilleure qualité par rapport aux autres modèles.

Une caractéristique particulièrement remarquable du modèle Flatten-SwinIR est sa vitesse d'exécution significativement plus rapide par rapport aux autres modèles basés sur les transformateurs vision et l'attention. Malgré la taille légèrement plus grande de ses images générées par rapport au modèle NLSA, Flatten-SwinIR affiche des temps d'exécution nettement inférieurs à ceux de ses concurrents.

Par exemple, sur l'ensemble de données BSD100, Flatten-SwinIR ne nécessite que 25 secondes, tandis que d'autres modèles peuvent prendre jusqu'à 60 secondes. Cette efficacité temporelle est cruciale pour les applications en temps réel ou le traitement rapide d'un grand volume d'images. De plus, Flatten-SwinIR est plus rapide que SwinIR et Swin2SR tout en les surpassant toujours sur pratiquement toutes les mesures.

Par conséquent, les résultats prouvent que Flatten-SwinIR est une excellente option pour

l'amélioration des images, offrant un excellent équilibre entre une amélioration de haute qualité et une efficacité computationnelle par rapport aux modèles de référence, le plaçant à l'avant-garde dans ce domaine de recherche.

TABLEAU 4.4 : Résultats des différents modèles dans le cas de la super-resolution.

Model	PSNR	SSIM	LPIPS	HaarPSI	ClipIQA	Output File Size (MB)	Time (s)
Jeu de données BSD100							
SRGAN	27.5170	0.9087	0.0677	0.8874	0.360	21.0	14
BSRGAN	24.9323	0.8296	0.2358	0.8514	0.569	**16.2**	15
RANKSRGAN	25.0541	0.8582	0.0788	0.8484	0.724	29.7	**9**
SRGAN_ResNeXT	27.6868	0.9077	0.0675	0.8900	0.350	21.2	***12***
SIR_SRGAN	27.7053	0.9084	0.0675	0.8907	0.368	21.1	13
HAN	27.9119	0.9117	0.0669	0.8949	**0.773**	19.3	46
NLSA	28.0320	0.9146	0.0658	0.8981	0.761	***18.8***	60
SwinIR	28.1557	0.9160	***0.0642***	0.8999	0.758	22.4	31
Swin2SR	***28.1662***	***0.9165***	0.0644	***0.9001***	0.744	22.6	35
Flatten-SwinIR	**28.2355**	**0.9175**	**0.0632**	**0.9011**	***0.771***	22.2	25
Jeu de données URBAN100							
SRGAN	24.1567	0.8690	0.1071	0.8637	0.406	7.77	**5**
BSRGAN	21.2160	0.7378	0.2746	0.7937	0.488	**6.77**	12
RANKSRGAN	22.8048	0.8420	***0.0932***	0.8315	0.600	10.3	***6***
SRGAN_ResNeXT	24.3530	0.8715	0.0951	0.8696	0.467	7.90	***6***
SIR_SRGAN	24.3273	0.8704	0.0997	0.8687	0.450	7.81	**5**
HAN	24.4784	0.8718	0.1080	0.8727	**0.641**	7.46	38
NLSA	24.7803	0.8812	0.1018	0.8778	0.616	***7.29***	45
SwinIR	25.0650	0.8859	0.0981	0.8811	0.618	8.53	16
Swin2SR	***25.1809***	***0.8892***	0.0964	***0.8824***	0.621	8.58	19
Flatten-SwinIR	**25.2164**	**0.8907**	**0.0930**	**0.8830**	***0.624***	8.47	12
Jeu de données Messidor-2							
SRGAN	40.7307	0.9788	0.0131	0.9888	0.414	58.8	56
BSRGAN	35.0012	0.9266	0.1169	0.9682	**0.658**	**34.2**	***46***
RANKSRGAN	39.4577	0.9700	0.0167	0.9851	0.494	82.6	**17**
SRGAN_ResNeXT	40.7646	0.9806	0.0144	0.9886	0.356	60.0	53
SIR_SRGAN	40.8820	0.9802	0.0160	0.9896	***0.531***	59.4	54
HAN	42.5156	0.9787	0.0131	0.9907	0.523	56.4	137
NLSA	43.7979	0.9856	0.0112	0.9936	0.512	***48.8***	184
SwinIR	44.0802	***0.9869***	***0.0101***	0.9944	0.505	53.1	108
Swin2SR	***44.1254***	0.9868	0.0104	***0.9945***	0.472	55.3	152
Flatten-SwinIR	**44.3906**	**0.9875**	**0.0094**	**0.9948**	0.474	52.5	90
Jeu de données Breakhis-400x							
SRGAN	35.6487	0.9879	0.0122	0.9628	0.098	71.5	50
BSRGAN	28.7088	0.9492	0.0970	0.9434	**0.432**	66.5	*43*
RANKSRGAN	33.6937	0.9788	0.0206	0.9499	0.209	93,1	**16**
SRGAN_ResNeXT	36.4377	0.9884	0.0114	0.9769	0.086	71.1	49
SIR_SRGAN	36.8844	0.9886	0.0110	0.9793	0.101	70.4	49
HAN	38.1220	0.9904	0.0092	0.9854	0.379	***65.3***	105
NLSA	37.9950	0.9903	0.0092	0.9856	0.361	**64.1**	159
SwinIR	38.2641	0.9904	0.0091	0.9858	0.396	79.4	88
Swin2SR	***38.3202***	***0.9905***	***0.0090***	***0.9860***	0.323	80.1	104
Flatten-SwinIR	**38.4714**	**0.9907**	**0.0088**	**0.9865**	***0.403***	79.3	73

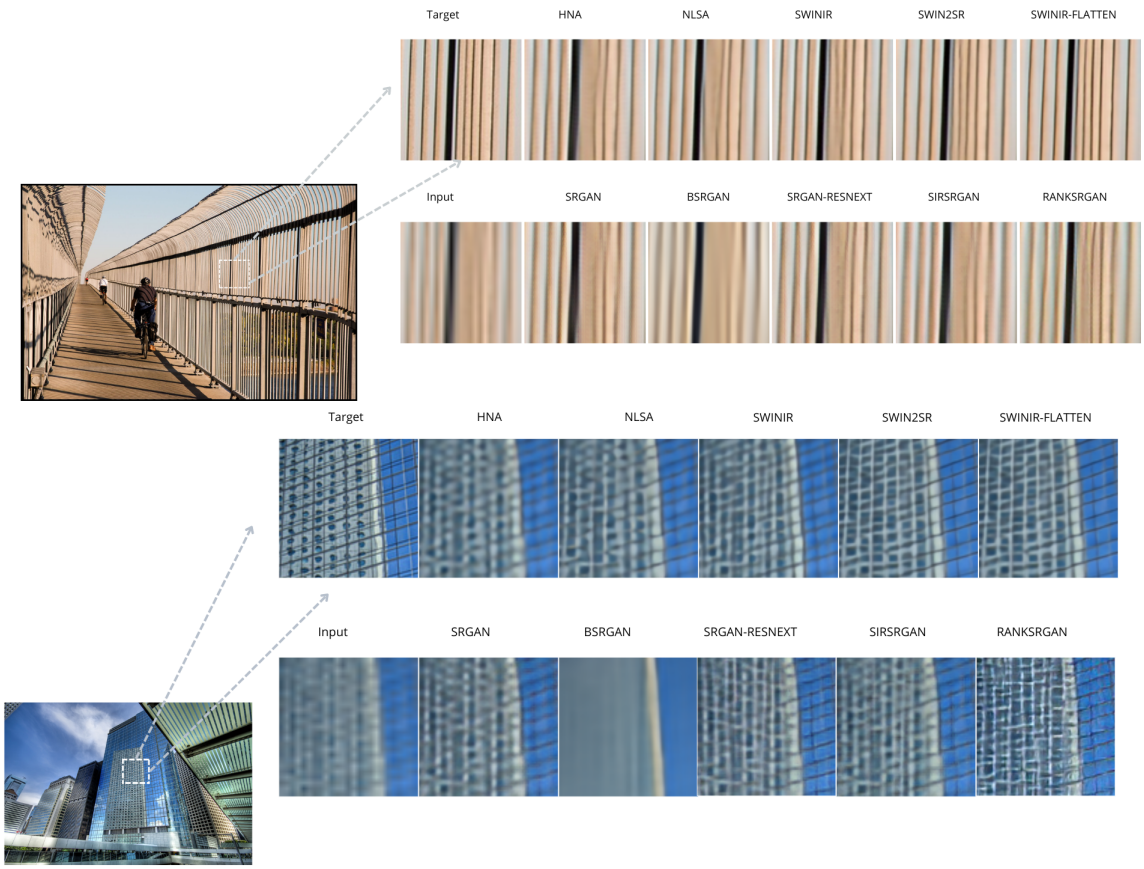

FIGURE 4.9 : Comparaison visuelle des différents modèles de super-résolution (×4) sur une image de l'ensemble de données URBAN100.

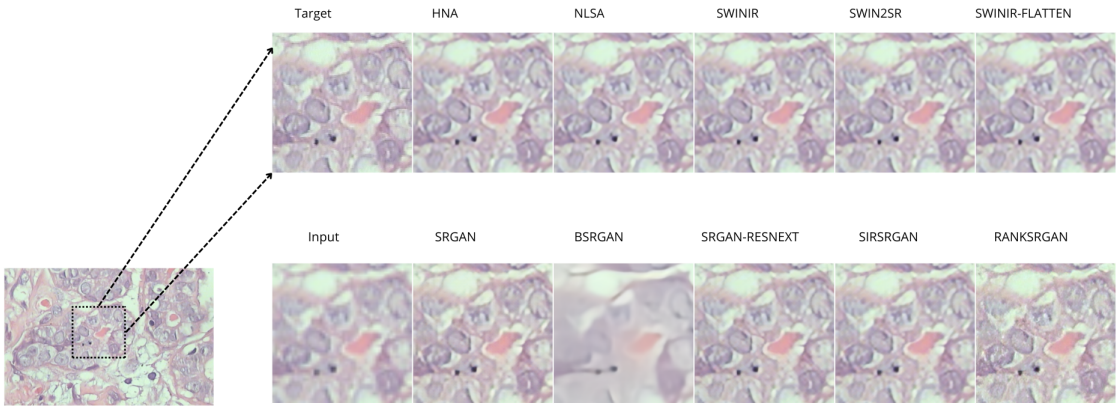

FIGURE 4.10 : Comparaison visuelle des différents modèles de super-résolution (×4) sur une image de l'ensemble de données Breakhis-400x .

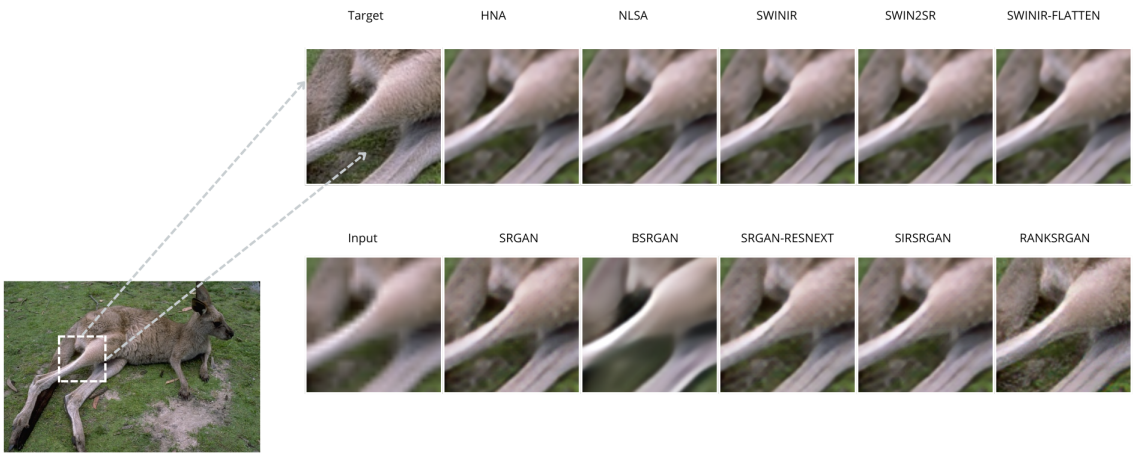

FIGURE 4.11 : Comparaison visuelle des différents modèles de super-résolution (×4) sur une image de l'ensemble de données BSD100

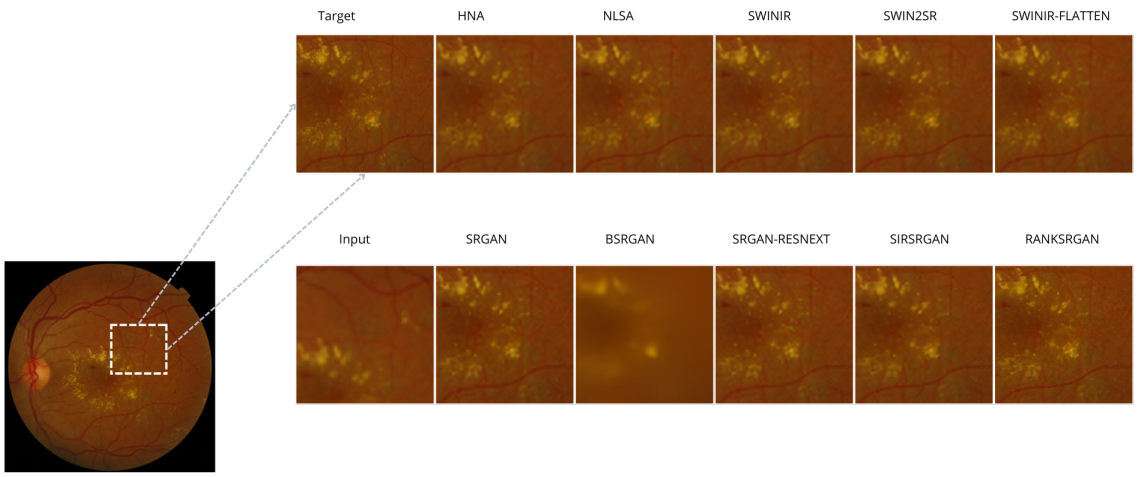

FIGURE 4.12 : Comparaison visuelle des différents modèles de super-résolution (×4) sur une image de l'ensemble de données Messidor.

PERFORMANCE EN DÉBRUITAGE

L'objectif de cette expérience est de tester la capacité de notre modèle à éliminer le bruit des images. Les résultats de débruitage sont comparés à ceux des principaux concurrents, à savoir SwinIR et Swin2SR.

Les modèles basés sur les GAN et les modules personnalisés d'attention ont été créés spécifiquement pour la super-résolution. Donc, nous ne les avons pas expérimenté dans ces scénarios de débruitage. Les résultats sont illustrés dans le tableau 4.5 et le tableau 4.6.

TABLEAU 4.5 : Résultats des différents modèles dans le cas du débruitage des images en couleur

Model	PSNR	SSIM	LPIPS	HaarPSI	ClipIQA	Output File Size (MB)	Time (s)
Jeu de données Kodak24							
SwinIR	***38.3842***	***0.9871***	**0.0114**	***0.9866***	***0.8604***	**14.4**	***324***
Swin2SR	37.9005	0.9860	0.0107	0.9853	0.8258	***15.2***	346
Flatten-SwinIR	**38.4314**	**0.9875**	***0.0115***	**0.9868**	**0.8607**	**14.4**	**239**
Jeu de données URBAN100							
SwinIR	*37.6327*	*0.9950*	**0.0044**	*0.9892*	*0.7910*	*13,6*	*2720*
Swin2SR	36.8617	0.9942	***0.0050***	0.9873	0.7370	14,2	3001
Flatten-SwinIR	**37.7851**	**0.9953**	**0.0044**	**0.9895**	**0.7917**	**13.5**	**1890**
Jeu de données CBSD68							
SwinIR	***36.2949***	***0.9830***	***0.0162***	***0.9807***	**0.8572**	**17.7**	***357***
Swin2SR	35.9399	0.9823	0.0157	0.9793	0.8398	***18.4***	396
Flatten-SwinIR	**36.3198**	**0.9835**	**0.0163**	**0.9808**	***0.8563***	**17.7**	**255**
Jeu de données Messidor-2							
SwinIR	***42.0018***	**0.9772**	0.0222	***0.9915***	0.4034	**51,6**	2388
Swin2SR	41.2161	0.9732	***0.0178***	0.9897	**0.6038**	59,4	***2225***
Flatten-SwinIR	**42.0781**	**0.9832**	**0.0221**	**0.9925**	***0.4178***	**52,9**	**1589**
Jeu de données Breakhis-400x							
SwinIR	***38.1419***	***0.9910***	**0.0098**	***0.9890***	***0.2830***	***78,0***	***2024***
Swin2SR	37.6475	0.9901	0.0103	0.9882	**0.2846**	***81.1***	2271
Flatten-SwinIR	**38.1835**	**0.9912**	***0.0098***	**0.9892**	0.2733	**77.4**	**1443**

TABLEAU 4.6 : Résultats des différents modèles dans le cas du débruitage des ensembles de données en niveaux de gris.

Model	PSNR	SSIM	LPIPS	HaarPSI	ClipIQA	Output File Size (MB)	Time (s)
SET12 data set							
SwinIR	***36.2532***	***0.9830***	***0.0264***	**0.9787**	0.8076	**0.985**	***57***
Swin2SR	35.9594	0.9799	0.0283	0.9771	***0.8458***	***0.995***	65
Flatten-SwinIR	**36.2707**	**0.9836**	**0.0226**	***0.9784***	**0.8762**	0.999	**43**
BSD68 data set							
SwinIR	***33.9864***	***0.9680***	0.0372	***0.9655***	0.8413	**5.39**	356
Swin2SR	33.7746	0.9667	***0.0366***	0.9642	**0.8695**	***5.40***	***347***
Flatten-SwinIR	**33.4000**	**0.9685**	**0.0334**	**0.9658**	***0.8672***	5.53	**257**

FIGURE 4.13 : Comparaison visuelle des méthodes de débruitage d'images couleur (niveau de bruit 20) sur une images provenant de l'ensemble de données URBAN100.

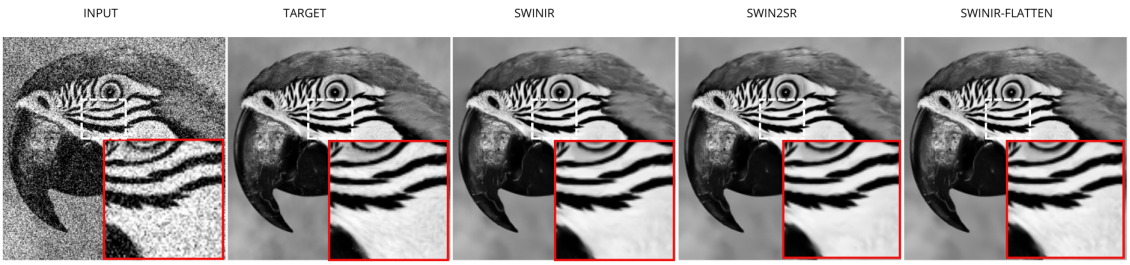

FIGURE 4.14 : Comparaison visuelle des méthodes de débruitage d'images en niveaux de gris (niveau de bruit 50) sur une image provenant de l'ensemble de données Set12.

La Figure 4.14 montre que SwinIR, Swin2SR et Flatten-SwinIR éliminent impeccablement le bruit de l'image en niveaux de gris de l'ensemble de données Set12. Cependant,

SwinIR et Swin2SR suppriment également les petites taches grises sur le plumage blanc de l'oiseau, tandis que Flatten-SwinIR les récupère, produisant une image reconstruite qui est aussi proche que possible de l'original. Une observation similaire peut être faite pour la Figure 4.13, où Flatten-SwinIR, SwinIR et Swin2SR démontrent une forte performance de débruitage sur une image en couleur de l'ensemble de données URBAN100, Swin2SR surpassent les deux autres modèles en récupérant une image moins floue et mieux définie.

Les résultats de débruitage sur les ensembles de données en couleur et en niveaux de gris sont respectivement présentés dans les Tableaux 4.5 et 4.6. Les résultats du Tableau 4.5 montrent une tendance similaire sur tous les ensembles de données. Par exemple, avec les images de Kodak24, le modèle SwinIR obtient un PSNR de 38.3842 et un SSIM de 0.9871.

Cependant, son temps d'exécution moyen est relativement élevé, nécessitant 324 secondes. Swin2SR performe légèrement mieux, avec un PSNR de 37.9005 et un SSIM de 0.9860, mais son temps d'exécution est également assez élevé, atteignant 346 secondes. Flatten-SwinIR, en revanche, se distingue non seulement en offrant une performance comparable voire supérieure en termes de qualité d'image, avec un PSNR de 38.4314 et un SSIM de 0.9875, mais aussi en réduisant significativement le temps d'exécution moyen, prenant 239 secondes. Cet avantage du temps d'exécution plus faible de Flatten-SwinIR tout en maintenant une haute performance de débruitage d'image souligne son efficacité et sa pertinence pour les applications nécessitant un traitement rapide des images en couleur ou en niveaux de gris.

Sur l'ensemble de données en niveaux de gris SET12, le Tableau 4.6 montre que Flatten-SwinIR livre des résultats remarquables avec un temps d'exécution de 43 secondes. Cette performance est significativement meilleure que celle de SwinIR, qui prend en moyenne 57 secondes, et bien plus rapide que Swin2SR, qui prend 65 secondes. Sur l'ensemble de données en couleur URBAN100, Flatten-SwinIR prend beaucoup plus de temps (1890 secondes) mais reste remarquablement plus rapide que SwinIR et Swin2SR qui dépassent 2700 secondes. Cela

fait de Flatten-SwinIR un choix optimal pour le débruitage d'images en couleur et en niveaux de gris, confirmant sa capacité à restaurer efficacement les images dans des délais compétitifs.

4.2.3 ETUDE DE L'ABLATION

Nous avons réalisé une étude d'ablation où les principaux paramètres ont été testés sur un maximum de 60 000 itérations pour observer leur impact sur les performances du modèle. Les tests de cette étude ont été faite sur l'ensemble de données URBAN100.

- **Taille du patch**

  Nous avons évalué l'impact de la taille du fragment (encore appelé patch) d par lequel on subdivise l'image sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la Figure 4.15, montrent une amélioration de la qualité de l'image lorsque la taille du patch est augmentée.

  Cependant, entre une taille de patch de 128 et 256, nous avons une évolution constante. Une taille de patch de 256 ou plus nécessite une carte graphique avec plus de 16 Go de mémoire. Nous avons choisi une taille de 192 pour entraîner le modèle Flatten-SwinIR. Si nous avions choisi une taille de 384, nous n'aurions pas pu entraîner les modèles concurrents HAN, NLSA ou Swin2SR en raison des limites de mémoire.

  Ainsi, pour faire une comparaison équitable, nous avons développé le modèle Flatten-SwinIR avec la même taille de patch que ses concurrents.

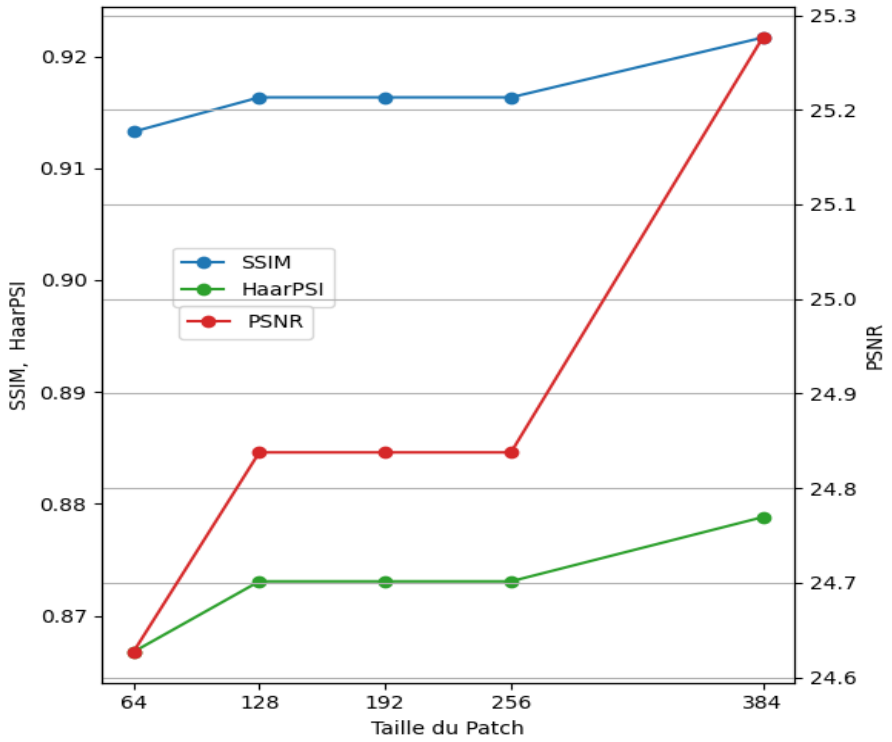

FIGURE 4.15 : Evolution des mesures au regard de la taille du patch.

© Gildas Aimé Sedou Fofe

- **Taille de la fenêtre**

Nous avons évalué l'impact de la taille de la fenêtre sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.16, montrent une amélioration de la qualité de l'image lorsque la taille de la fenêtre augmente. Cependant, nous observons une légère diminution du PSNR à taille de la fenêtre $= 16$ et un pic du PSNR et du HaarPSI à taille de la fenêtre $= 12$ et celui du SSIM à taille de la fenêtre $= 8$. Pour les mêmes raisons de comparaison equitable que précédement. Nous avons développé notre modèle avec une taille de la fenêtre $= 8$.

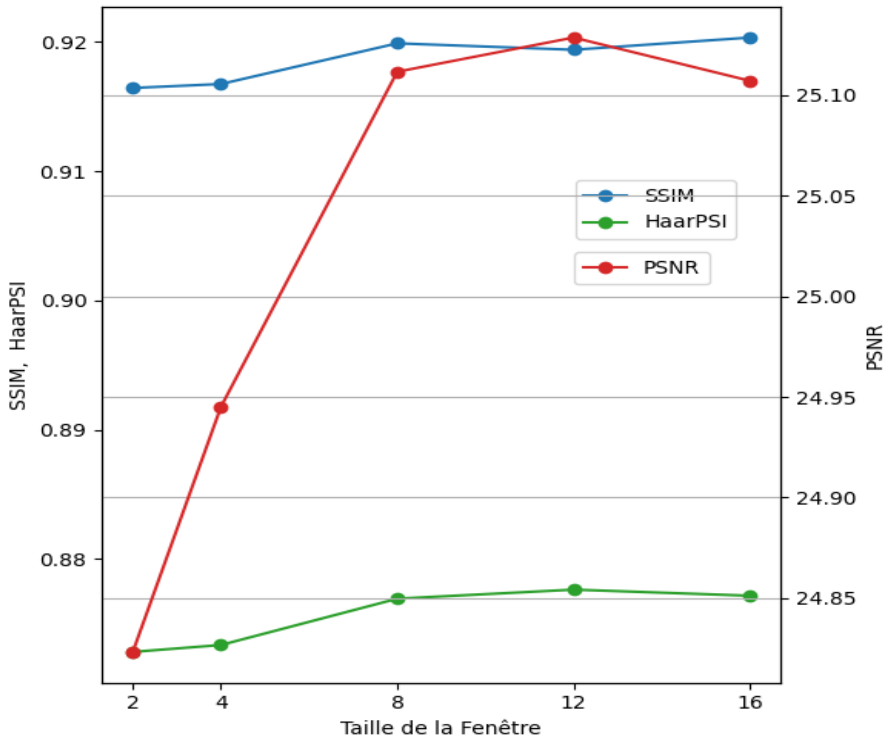

FIGURE 4.16 : Evolution des mesures au regard de la Fenêtre

© Gildas Aimé Sedou Fofe

- **Facteur de focalisation**

Nous avons évalué l'impact du facteur de focalisation sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.17, montrent que les meilleures mesures sont obtenues avec un facteur de focalisation $= 3$.

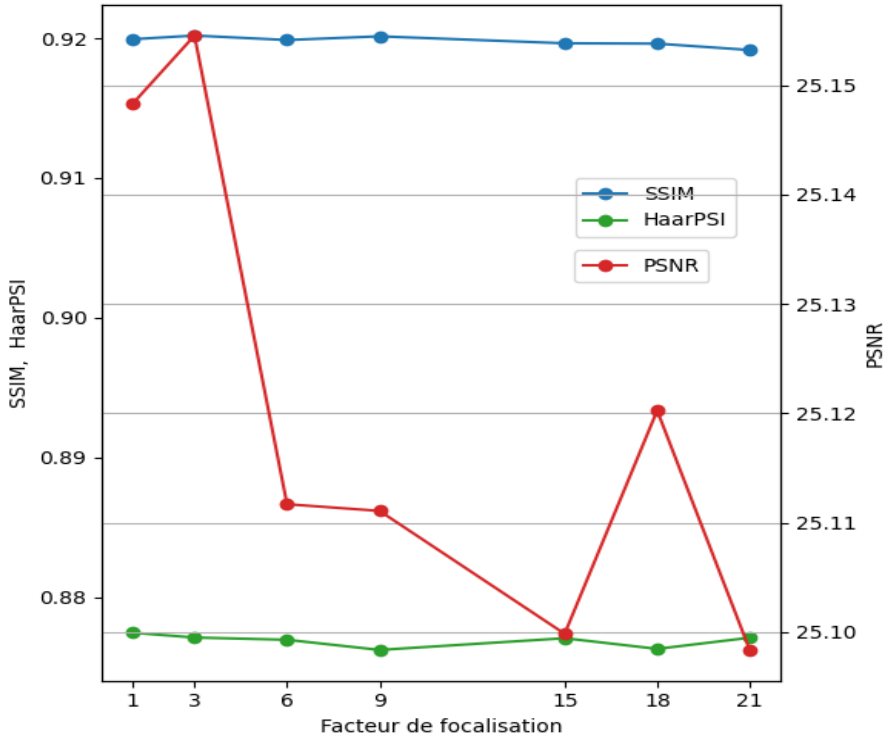

FIGURE 4.17 : Evolution des mesures au regard du facteur de focalisation

© Gildas Aimé Sedou Fofe

- **Taille du lot**

Nous avons évalué l'impact de la taille du lot sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.18, montrent que la qualité de l'image augmente avec une augmentation de la taille du lot. Cependant, nous avons opté pour une taille de lot de 16 pour les mêmes raisons de comparaison équitable qu'avec l'ablation study de la taille du patch.

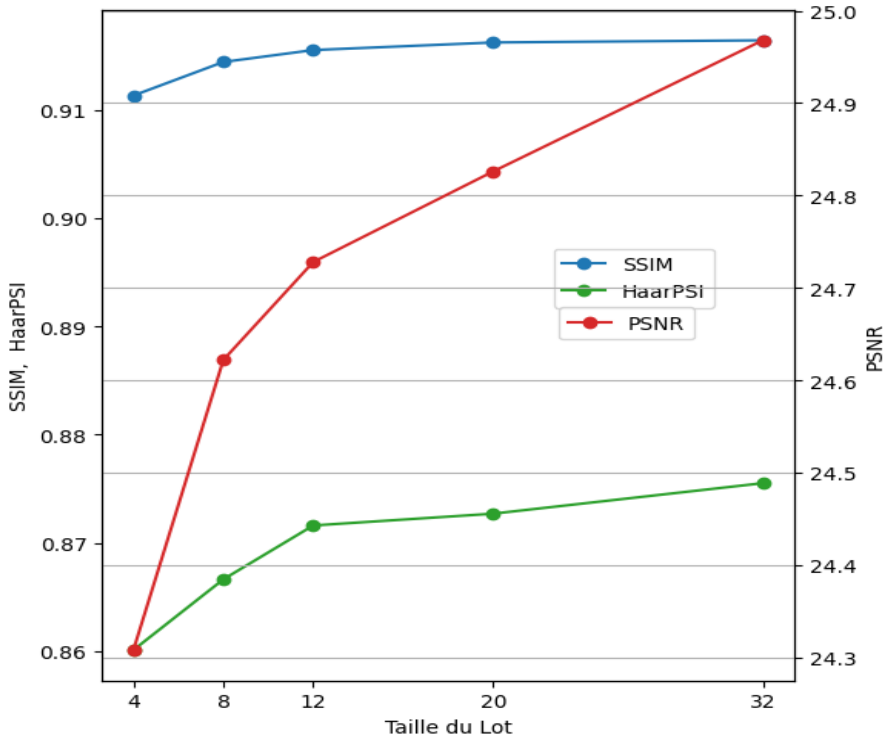

FIGURE 4.18 : Evolution des mesures au regard de la taille du lot.

© Gildas Aimé Sedou Fofe

- **Nombre de couches dans un bloc FRTB**

Nous avons évalué l'impact du nombre de couches dans un bloc FRTB sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.19, montrent que la qualité de l'image augmente avec l'augmentation du nombre de couches. Nous avons opté pour 6 couches pour une équité de comparaison avec les modèles concurrents.

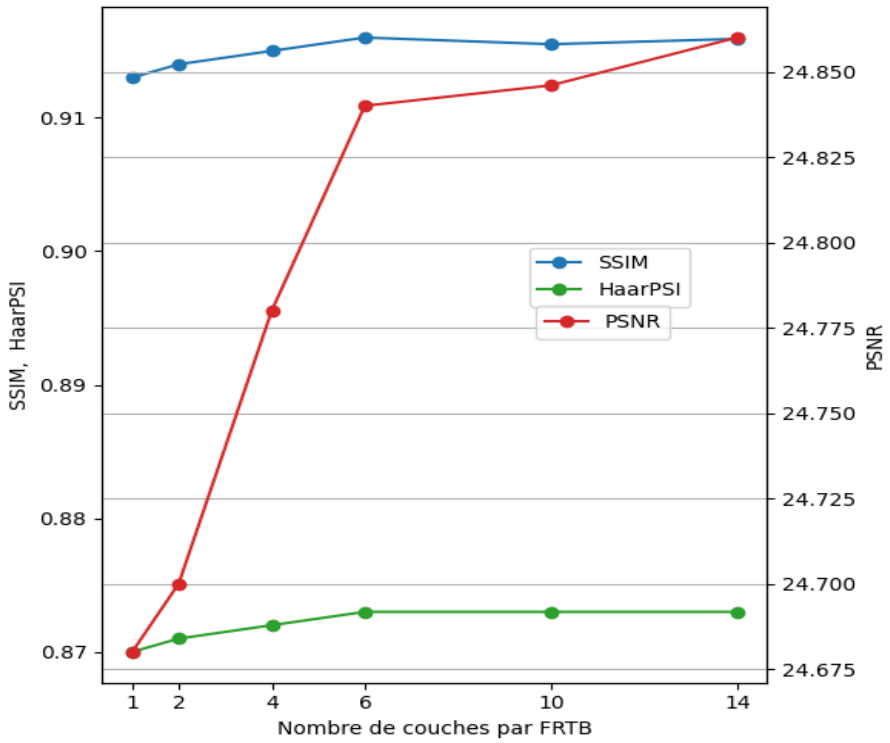

FIGURE 4.19 : Evolution des mesures au regard du nombre de couches dans un bloc FRTB

© Gildas Aimé Sedou Fofe

- **Nombre de blocs FRTB**

  Nous avons évalué l'impact du nombre de blocs FRTB sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.20, montrent qu'il y a une amélioration croissante de la qualité de l'image avec l'augmentation du nombre de blocs. Les scores SSIM et HaarPSI montrent une tendance constante entre 4 et 8. Nous avons choisi un nombre de blocs égal à 6 pour une équité de comparaison avec les modèles concurrents.

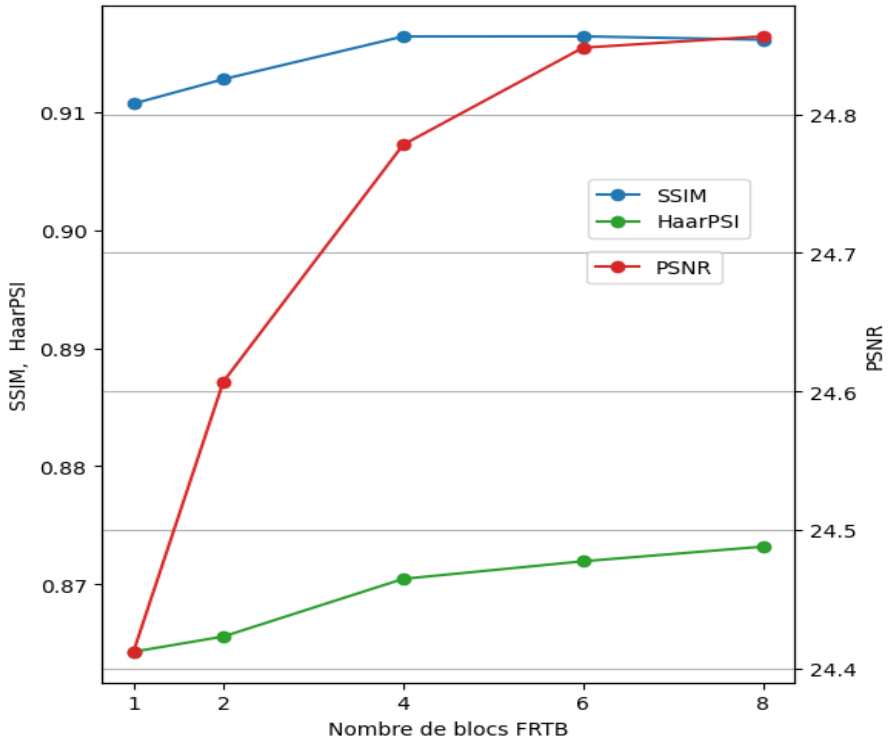

FIGURE 4.20 : Evolution des mesures au regard du nombre de blocs FRTB.

© Gildas Aimé Sedou Fofe

- **Nombre de couches de convolution**

Nous avons évalué l'impact du nombre de couches de convolution dans le module d'extraction des caractéristiques peu profondes sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la Figure 4.21, montrent que les résultats sont meilleurs avec 3 convolutions.

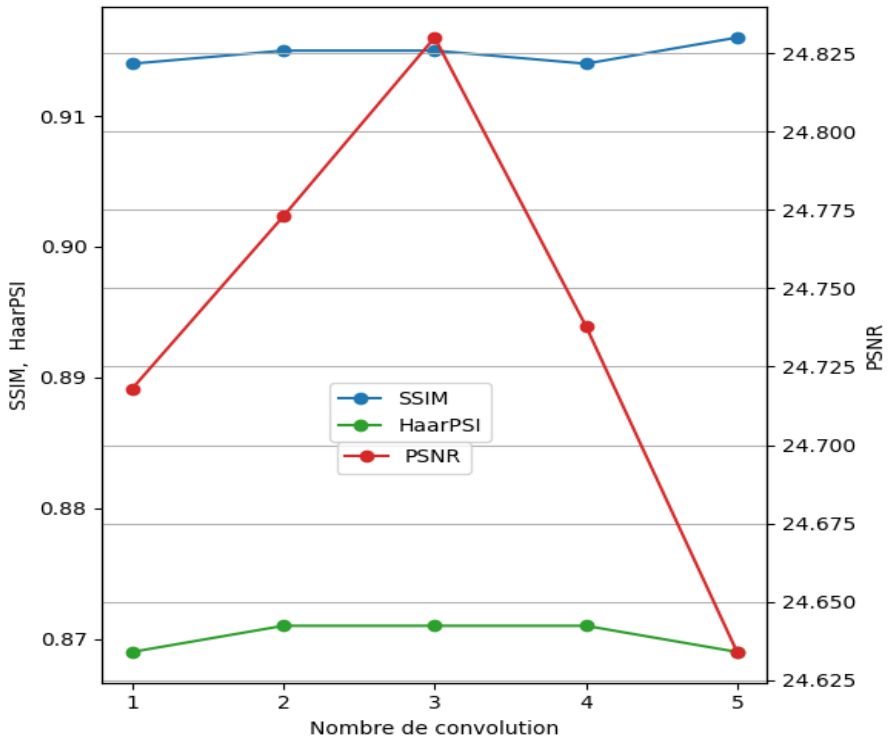

FIGURE 4.21 : Evolution des mesures au regard au regard du nombre de convolutions dans le module d'extraction des caractéristiques superfichielles .

© Gildas Aimé Sedou Fofe

- **Nombre de têtes d'attention**

  Nous avons évalué l'impact du nombre de têtes d'attention dans le mécanisme d'attention Flatten Attention, sur la qualité des images de super-résolution (échelle $\times 4$). Les expériences, illustrées dans la figure 4.22, montrent un pic de performance à 10. Nous avons choisi cette valeur pour notre modèle.

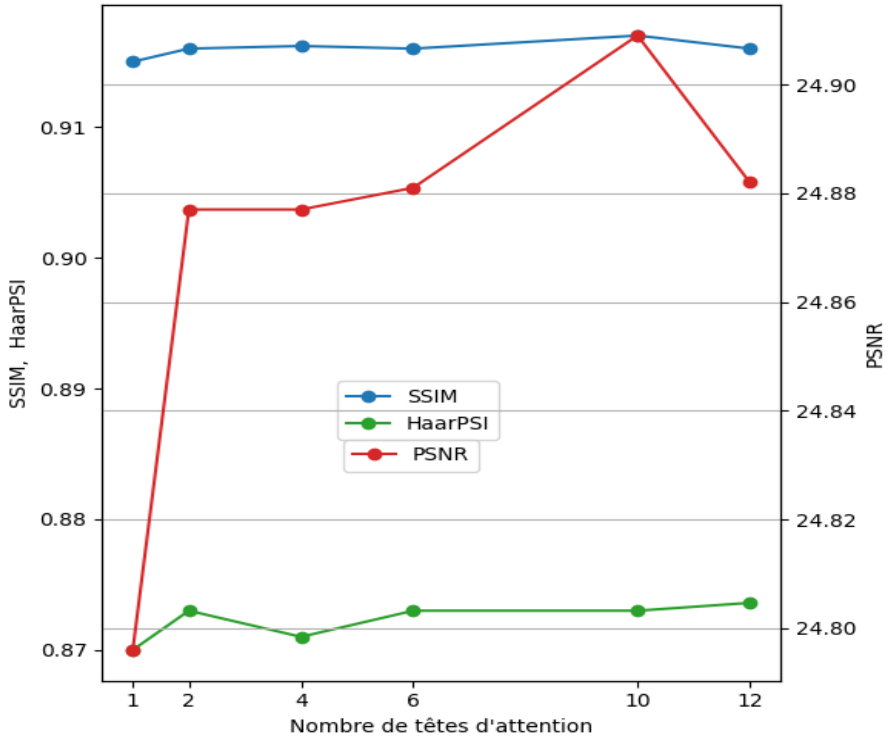

FIGURE 4.22 : Evolution des mesures au regard du nombre de têtes d'attention

© Gildas Aimé Sedou Fofe

4.3 CONCLUSION

Ce chapitre explore les résultats des expériences menées sur les deux modèles de super-résolution proposés, SIR-SRGAN-ResNeXt et Flatten-SwinIR. Les résultats ont montré que SIR-SRGAN-ResNeXt, grâce à son générateur ResNeXt et son discriminateur U-Net avec des mécanismes de normalisation spectrale et d'attention, surpasse ses concurrents GANs en termes de PSNR, SSIM, LPIPS et HaarPSI sur des ensembles de données d'images médicales comme Messidor-2 et Breakhis-400x.

L'étude d'ablation a révélé l'impact des différents composants sur la performance du modèle. D'autre part, Flatten-SwinIR, avec son architecture de transformateur de vision FSTL intégre le mécanisme d'attention Flatten Attention. Ce mécanisme lui permet d'avoir des résultats

supérieures à ceux des modèles de type GAN et même les modèles utilisant aussi les transformateurs de vision comme SwiniR et Swin2SR.

Flatten-SwinIR offre également des temps d'exécution significativement plus courts par rapport à SwinIR, Swin2SR, ce qui est crucial pour des applications en temps réel.

Une étude d'ablation a été faite afin de mettre en lumière les paramètres clés du modèle Flatten-SwinIR.

CONCLUSION

Dans ce mémoire, nous avons exploré et développé des modèles de super-résolution d'images afin d'améliorer la qualité des images générées par des équipements médicaux. Nous avons introduit deux modèles novateurs : le modèle SIR-SRGAN-ResNeXt, une amélioration du SIR-SRGAN et le modèle de restauration d'images Flatten-SwinIR, une amélioration du modèle SwinIR utilisant un mécanisme d'attention nommé Flatten Attention.

Le modèle SIR-SRGAN-ResNeXt utilise deux classeurs de rang (encore appelés "Rankers"), un générateur basé sur ResNeXt et un discriminateur basé sur l'architecture U-Net. Le générateur ResNeXt permet l'apprentissage de caractéristiques plus complexes et la génération d'images super-résolues de meilleure qualité.

Nous ajoutons au discriminateur U-Net une normalisation spectrale et des couches d'attention, pour améliorer la capacité du modèle à évaluer les images générées, contribuant ainsi à une meilleure discrimination entre les images reconstruites et les images originales. La fonction de perte, intégrant divers éléments avec des coefficients soigneusement ajustés, met l'accent sur la similarité perceptive et la précision au niveau des pixels.

Grâce à ces améliorations SIR-SRGAN-ResNeXt surpasse plusieurs autres modèles GAN de super-résolutions sur divers metriques de qualité d'image. Mais la qualité visuelle de ses images n'etait pas satisfaisant alors nous avons exploré de nouveaux types d'architectures de modèles de super-résolution, notamment les architectures basés sur les transformateurs de vision.

Le modèle Flatten-SwinIR est basé sur ce nouveau type d'architecture et se positionne comme une solution efficace face aux défis de la restauration d'images. Il utilise un transformateur de vision nommé "Flatten Transformer", qui est doté d'un mécanisme d'auto-attention ayant une complexité computationnelle linéaire par rapport à la complexité quadratique de ses concurrents SwinIR, et Swin2SR.

Les tests effectués sur divers ensembles de données, montrent que Flatten-SwinIR obtient des résultats supérieurs sur les mesures telles que PSNR, SSIM, LPIPS et HaarPSI avec un temps d'éxecution meilleur que SwinIR, Swin2SR et d'autres modèles basés sur des mécanismes d'attention complexes.

PESPECTIVES

Les résultats obtenus au cours de cette recherche ouvrent la voie à plusieurs perspectives d'amélioration, visant à renforcer l'efficacité des modèles de super-résolution d'images médicales.

1. **Architecture GAN avec Flatten-SwinIR comme générateur :** Une pespective d'amelioration serait de créer un GAN avec Flatten-SwinIR comme Générateur. Nous avons dans nos travaux tester de telles architectures GAN, avec un générateur basé sur Flatten-SwinIR et un distriminateur basé sur U-Net. Mais les résutats ne sont pas aussi meilleurs qu'un modèle Flatten-SwinIR tout seul. Une orientation de recherche serait d'élaborer ce nouveau style de GAN.
2. **Architecture en Serie Flatten-SwinIR et SIR-SRGAN-ResneXT :** Une autre direction prometteuse serait de combiner les forces des modèles Flatten-SwinIR et SIR-SRGAN-ResNeXt en une architecture en série. L'idée serait d'utiliser Flatten-SwinIR pour effectuer une première phase de restauration d'image, en tirant parti de ses capacités de traitement efficace et de haute qualité, puis de passer les images ainsi améliorées à SIR-SRGAN-ResNeXt pour une super-résolution finale. Cette approche pourrait potentiellement surmonter les limitations individuelles de chaque modèle et fournir des résultats encore plus impressionnants.
3. **Optimisation des Hyperparamètres et Réduction de la Complexité Computationnelle :** Une troisième perspective d'amélioration serait d'optimiser davantage les hy-

perparamètres des deux modèles et de réduire leur complexité computationnelle. Cela pourrait inclure des optimisations algorithmiques et l'exploration de nouvelles architectures de réseau qui maintiennent ou améliorent la qualité de l'image tout en réduisant les exigences en temps de calcul et en ressources matérielles.

En somme, les travaux réalisés dans ce mémoire jettent les bases pour des recherches futures riches et variées dans le domaine de la super-résolution d'images, avec des implications particulièrement prometteuses pour l'amélioration des techniques d'imagerie médicale et, par conséquent, pour la qualité des soins aux patients.

BIBLIOGRAPHIE

Agustsson, E. & Timofte, R. (2017). Ntire 2017 challenge on single image super-resolution : Dataset and study. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135.

Arbeláez, P., Maire, M., Fowlkes, C. & Malik, J. (2011). Contour Detection and Hierarchical Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(5), 898–916.

Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A. *et al.* (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI) : a completed reference database of lung nodules on CT scans. *Medical physics*, *38*(2), 915–931.

Bahdanau, D., Cho, K. & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. Dans Y. Bengio & Y. LeCun (Éds.). *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Repéré à http://arxiv.org/abs/1409.0473

Chen, Y., Zhang, C., Liu, L., Feng, C., Dong, C., Luo, Y. & Wan, X. (2021). USCL : pretraining deep ultrasound image diagnosis model through video contrastive representation learning. Dans *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021 : 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pp. 627–637. Springer.

Conde, M. V., Choi, U.-J., Burchi, M. & Timofte, R. (2022). *Swin2SR : SwinV2 Transformer for Compressed Image Super-Resolution and Restoration*.

Dai, T., Cai, J., Zhang, Y., Xia, S.-T. & Zhang, L. (2019). Second-order attention network for single image super-resolution. Dans *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11065–11074.

Datar, M., Immorlica, N., Indyk, P. & Mirrokni, V. S. (2004). Locality-sensitive hashing scheme based on p-stable distributions. Dans *Proceedings of the twentieth annual symposium on Computational geometry*, pp. 253–262.

De Bruijne, M., Cattin, P. C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y. & Essert, C. (2021). *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021 : 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III*, Vol. 12903. Springer Nature.

Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J.-R., Massin, P., Erginay, A., Charton, B. & Klein, J.-C. (2014). Feedback on a Publicy Distributed Image Database : The Messidor Database. *Image Analysis & Stereology*, pp. 231–234.

Dong, C., Loy, C. C., He, K., Tang, Xiaoou, e.-D., Pajdla, T., Schiele, B. & Tuytelaars, T. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. Dans *Computer Vision – ECCV 2014*, pp. 184–199., Cham. Springer International Publishing.

Dong, C., Loy, C. C., He, K. & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. Dans *Computer Vision–ECCV 2014 : 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pp. 184–199. Springer.

Dong, C., Loy, C. C., He, K. & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), 295–307.

Dong, C., Loy, C. C., Tang, Xiaoou, e.-B., Matas, J., Sebe, N. & Welling, M. (2016). Accelerating the Super-Resolution Convolutional Neural Network. Dans *Computer Vision – ECCV 2016*, pp. 391–407., Cham. Springer International Publishing.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. *et al.* (2020). An image is worth 16x16 words : Transformers for image recognition at scale. *arXiv preprint arXiv :2010.11929*.

Franzen, R. (1999). Kodak lossless true color image suite. *source : http ://r0k. us/graphics/-kodak*, *4*(2), 9.

Gatys, L. A., Ecker, A. S. & Bethge, M. (2016). Image style transfer using convolutional neural networks. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423.

Ghaffari, M., Sowmya, A. & Oliver, R. (2019). Automated brain tumor segmentation using multimodal brain scans : a survey based on models submitted to the BraTS 2012–2018 challenges. *IEEE reviews in biomedical engineering*, *13*, 156–168.

Gonçalves-Bradley, D. C., Maria, A. R. J., Ricci-Cabello, I., Villanueva, G., Fønhus, M. S., Glenton, C., Lewin, S., Henschke, N., Buckley, B. S., Mehl, G. L. *et al.* (2020). Mobile technologies to support healthcare provider to healthcare provider communication and management of care. *Cochrane Database of Systematic Reviews*, (8).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Han, D., Pan, X., Han, Y., Song, S. & Huang, G. (2023). *FLatten Transformer : Vision Transformer using Focused Linear Attention*.

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep Residual Learning for Image Recognition. Dans *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778. doi: 10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hendrycks, D. & Gimpel, K. (2016). Gaussian error linear units (gelus). *arXiv preprint arXiv :1606.08415*.

Huang, J.-B., Singh, A. & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5197–5206.

Huang, J.-H., Wang, H.-K. & Liao, Z.-W. (2021). SIR-SRGAN : Super-Resolution Generative Adversarial Networks with Self-Interpolation Ranker. Dans *BMVC*, p. 52.

Hunt, B., Ruiz, A. J. & Pogue, B. W. (2021). Smartphone-based imaging systems for medical applications : a critical review. *Journal of Biomedical Optics*, *26*(4), 040902–

040902.

Hussain, S., Mubeen, I., Ullah, N., Shah, S. S. U. D., Khan, B. A., Zahoor, M., Ullah, R., Khan, F. A., Sultan, M. A. *et al.* (2022). Modern diagnostic imaging technique applications and risk factors in the medical field : a review. *BioMed research international*, *2022*.

Juhong, A., Li, B., Yao, C.-Y., Yang, C.-W., Agnew, D. W., Lei, Y. L., Huang, X., Piyawattanametha, W. & Qiu, Z. (2023). Super-resolution and segmentation deep learning for breast cancer histopathology image analysis. *Biomed. Opt. Express*, *14*(1), 18–36.

Kim, T. (1988). New finite state vector quantizers for images. Dans *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 1180–1183.

Kitaev, N., Kaiser, Ł. & Levskaya, A. (2020). Reformer : The efficient transformer. *arXiv preprint arXiv :2001.04451*.

Kohavi, R. *et al.* (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Dans *Ijcai*, Vol. 14, pp. 1137–1145. Montreal, Canada.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Dans *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 105–114. doi: 10.1109/CVPR.2017.19

Li, S., Zhao, Y., Varma, R., Salpekar, O., Noordhuis, P., Li, T., Paszke, A., Smith, J., Vaughan, B., Damania, P. *et al.* (2020). Pytorch distributed : Experiences on accelerating data parallel training. *arXiv preprint arXiv :2006.15704*.

Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L. V. & Timofte, R. (2021). SwinIR : Image Restoration Using Swin Transformer. Dans *IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2021, Montreal, BC, Canada, October 11-17, 2021*, pp. 1833–1844. Repéré à https://doi.org/10.1109/ICCVW54120.2021.00210

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L. & Timofte, R. (2021). SwinIR : Image Restoration Using Swin Transformer. Dans *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844. doi: 10.1109/ICCVW54120.2021.00210

Lim, B., Son, S., Kim, H., Nah, S. & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. Dans *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140. doi: 10.1109/CVPRW.2017.151

Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L. *et al.* (2022). Swin transformer v2 : Scaling up capacity and resolution. Dans *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12009–12019.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. (2021). Swin transformer : Hierarchical vision transformer using shifted windows. Dans *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022.

Ma, Y., Liu, K., Xiong, H., Fang, P., Li, X., Chen, Y., Yan, Z., Zhou, Z. & Liu, C. (2021). Medical image super-resolution using a relativistic average generative adversarial network. *Nuclear Instruments and Methods in Physics Research Section A : Accelerators, Spectrometers, Detectors and Associated Equipment*, *992*, 165053.

Maas, A. L., Hannun, A. Y., Ng, A. Y. *et al.* (2013). Rectifier nonlinearities improve neural network acoustic models. Dans *Proc. icml*, Vol. 30, p. 3. Atlanta, GA.

Mei, Y., Fan, Y. & Zhou, Y. (2021). Image Super-Resolution with Non-Local Sparse Attention. Dans *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3516–3525. doi: 10.1109/CVPR46437.2021.00352

Mei, Y., Fan, Y. & Zhou, Y. (2021). Image super-resolution with non-local sparse attention. Dans *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3517–3526.

Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S. & Shi, H. (2020). Image Super-Resolution With Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining. Dans *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5689–5698. doi: 10.1109/CVPR42600.2020.00573

Mitchell, C., Oakford, M. & Murray, C. (2021). Medical education in the COVID-19 era : a remote dermatology attachment A. Cummin, C. Christie, 2 A. Fityan, H. Lotery.

Miyato, T., Kataoka, T., Koyama, M. & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv :1802.05957*.

Nair, V. & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. Dans *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, p. 807–814., Madison, WI, USA. Omnipress.

Nasrollahi, K. & Moeslund, T. B. (2014). Super-resolution : a comprehensive survey. *Machine vision and applications*, *25*, 1423–1468.

Nilsson, J. & Akenine-Möller, T. (2020). Understanding ssim. *arXiv preprint arXiv :2006.13846*.

Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, Haifeng, e. A., Bischof, H., Brox, T. & Frahm, J.-M. (2020). Single Image Super-Resolution via a Holistic Attention Network. Dans *Computer Vision – ECCV 2020*, pp. 191–207. Springer International Publishing.

Reisenhofer, R., Bosse, S., Kutyniok, G. & Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing : Image Communication*, *61*, 33–43.

Reisenhofer, R., Bosse, S., Kutyniok, G. & Wiegand, T. (2018). A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing : Image Communication*, p. 33–43. doi: 10.1016/j.image.2017.11.001

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. Dans *Medical image computing and computer-assisted intervention–MICCAI 2015 : 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241. Springer.

Salvetti, F., Mazzia, V., Khaliq, A. & Chiaberge, M. (2020). Multi-image super resolution of remotely sensed images using residual attention deep neural networks. *Remote Sensing*,

12(14), 2207.

Schonfeld, E., Schiele, B. & Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks. Dans *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8207–8216.

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*.

Snoek, J., Larochelle, H. & Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. Dans F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Éds.). *Advances in Neural Information Processing Systems*, Vol. 25. Curran Associates, Inc.

Spanhol, F. A., Oliveira, L. S., Petitjean, C. & Heutte, L. (2016). A Dataset for Breast Cancer Histopathological Image Classification. *IEEE Transactions on Biomedical Engineering*, *63*(7), 1455–1462.

Sun, J., Xu, Z. & Shum, H.-Y. (2008). Image super-resolution using gradient profile prior. Dans *2008 IEEE conference on computer vision and pattern recognition*, pp. 1–8. IEEE.

Timofte, R., Agustsson, E., Van Gool, L., Yang, M.-H. & Zhang, L. (2017, July). NTIRE 2017 Challenge on Single Image Super-Resolution : Methods and Results. Dans *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u. & Polosukhin, I. (2017). Attention is All you Need. Dans I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Éds.). *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. Repéré à https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wang, J., Chan, K. C. K. & Loy, C. C. (2022). *Exploring CLIP for Assessing the Look and Feel of Images*.

Wang, X., Girshick, R., Gupta, A. & He, K. (2018, jun). Non-local Neural Net-

works. Dans *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7794–7803., Los Alamitos, CA, USA. IEEE Computer Society. doi: 10.1109/CVPR.2018.00813

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. & Summers, R. M. (2017). Chestx-ray8 : Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y. & Loy, C. C. (2019). ESRGAN : Enhanced Super-Resolution Generative Adversarial Networks. Dans L. Leal-Taixé & S. Roth (Éds.). *Computer Vision – ECCV 2018 Workshops*, pp. 63–79., Cham. Springer International Publishing.

Wang, Z., Bovik, A., Sheikh, H. & Simoncelli, E. (2004). Image quality assessment : from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. (2017). Aggregated residual transformations for deep neural networks. Dans *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1492–1500.

Yang, J. & Huang, T. (2017). Image super-resolution : Historical overview and future challenges. Dans *Super-resolution imaging* pp. 1–34. CRC Press.

Yusuf, A. M., Lusobya, R. C., Mukisa, J., Batte, C., Nakanjako, D. & Juliet-Sengeri, O. (2022). Validity of smartphone-based retinal photography (PEEK-retina) compared to the standard ophthalmic fundus camera in diagnosing diabetic retinopathy in Uganda : A cross-sectional study. *Plos one*, *17*(9), e0273633.

Zhang, K., Liang, J., Van Gool, L. & Timofte, R. (2021). Designing a practical degradation model for deep blind image super-resolution. Dans *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4791–4800.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. (2018). The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. Dans *CVPR*.

Zhang, W., Liu, Y., Dong, C. & Qiao, Y. (2019). RankSRGAN : Generative Adversarial Networks With Ranker for Image Super-Resolution. Dans *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3096–3105. doi: 10.1109/ICCV.2019.00319

Zhang, W., Liu, Y., Dong, C. & Qiao, Y. (2021). *RankSRGAN : Super Resolution Generative Adversarial Networks with Learning to Rank*.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Yun, e. V., Hebert, M., Sminchisescu, C. & Weiss, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks. Dans *Computer Vision – ECCV 2018*, pp. 294–310., Cham. Springer International Publishing.

APPENDICE A
STRUTURE DES FICHIERS

A.1 JEU DE DONNÉES : BSD100

Le Jeu de données BSD100 (Berkeley Segmentation Dataset) est une collection d'images utilisée pour la segmentation et l'évaluation des algorithmes de traitement d'images. Il est contient 100 images au format JPEG de diverses scènes, il permet d'évaluer les algorithmes.

A.2 JEU DE DONNÉES : BREAKHIS

Le Jeu de données BreakHis (Breast Cancer Histopathological Image Classification) est une collection d'images histopathologiques de tissus mammaires humains, conçue pour faciliter la classification des tumeurs bénignes et malignes à différents niveaux de grossissement : 40X, 100X, 200X, et 400X. La structure des fichiers est organisée de manière hiérarchique pour une utilisation efficace.

Le dossier principal "BreakHis" contient des sous-dossiers correspondant à chaque niveau de grossissement. Chaque sous-dossier de grossissement (par exemple, 40X, 100X, 200X, 400X) est divisé en deux dossiers principaux : "benign" pour les tumeurs bénignes et "malignant" pour les tumeurs malignes. À l'intérieur de ces dossiers "benign" et "malignant", les images sont encore subdivisées en sous-types spécifiques de tumeurs.

Pour les tumeurs bénignes, les sous-types incluent Adenosis, Fibroadenoma, PhyllodesTumor, et TubularAdenoma. Pour les tumeurs malignes, les sous-types sont DuctalCarcinoma, LobularCarcinoma, MucinousCarcinoma, et PapillaryCarcinoma. Chaque sous-dossier de sous-type contient des images histopathologiques au format JPEG ou PNG, et des fichiers texte ou CSV fournissant des annotations et des métadonnées supplémentaires.

A.3 JEU DE DONNÉES : MESSIDOR-2

Le jeu de donnée Messidor-02 est contient un dossier principal avec plusieurs sous-dossiers. Le dossier "images" contient 1744 images du fond d'œil au format JPEG, numérotées de 1 à 1744. Le dossier "annotations" contient un fichier CSV qui fournit des informations supplémentaires sur le diagnostic fait au patient.

A.4 JEU DE DONNÉES : URBAN100

Le jeu de données Urban100 est un ensemble d'images de haute-résolution de scènes urbaines. Il contient 100 images de haute résolution et trois versions à basse résolution pour chaque image, correspondant aux facteurs d'échelle x2, x3 et x4. La structure du jeu de données est hiérarchique, avec des dossiers pour chaque image et des sous-dossiers pour chaque niveau de mise à l'échelle.

A.5 JEU DE DONNÉES : KODAK24

Le jeu de données Kodak24 est une collection de 24 images au format PNG. Sa structure est simple, avec un seul dossier principal "Kodak24" contenant les 24 images, nommées de "kodim01.png" à "kodim24.png".

A.6 JEU DE DONNÉES : CBSD68

Le jeu de données CBSD68 est un ensemble de 68 images de référence pour l'évaluation des algorithmes de détection des bords. Il est composé d'images de diverses sources, telles que des photographies, des images médicales et des images synthétiques.

A.7 JEU DE DONNÉES : BSD68

BSD68 est un ensemble de donnée constitué de 68 images prises dans le dataset BSD100. Ces images ont été converties en niveaux de gris.

A.8 JEU DE DONNÉES : SET12

Le dataset Set12 est un ensemble de données couramment utilisé pour l'évaluation des algorithmes de débruitage d'images. Il contient 12 images, chacune étant en niveaux de gris.

A.9 JEU DE DONNÉES : DIV2K

Le dataset DIV2K (DIVerse 2K resolution) est une collection de 900 images haute résolution utilisée pour la formation et l'évaluation des algorithmes de super-résolution d'images. Il est structuré en deux dossiers principaux : 'HR' pour les images haute résolution et 'LR' pour les versions basse résolution des mêmes images, avec des échelles de dégradation de 'X2', 'X3', 'X4', et 'X8'. Chaque image dans le dossier 'LR' est nommée en correspondance avec son équivalent haute résolution dans le dossier 'HR', mais avec un suffixe indiquant le facteur d'échelle (par exemple, '0001x2.png').

A.10 JEU DE DONNÉES : FLICKR2K

Le dataset Flickr2K est une collection d'images haute résolution utilisée pour la formation et l'évaluation des algorithmes de super-résolution d'images, similaire au dataset DIV2K. Il contient 2650 images, provenant de la plateforme Flickr, et est structuré de manière à inclure des versions haute résolution et basse résolution des mêmes images. Ces images sont utilisées pour améliorer et tester les performances des algorithmes de super-résolution en fournissant des données variées et de haute qualité.