

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Master

Filière : Électronique
Spécialité : Systèmes des télécommunications

Présenté par

OUADAH HADJER

et

YETTOU CHERIFA KHAOULA

Implémentation de la Super-Résolution d'images à l'aide du GAN

Proposé par : Mr. NAMANE ABDERRAHMANE

Mr. AOUINANE BILAL

Année Universitaire 2023-2024

Remerciements

Au nom d'Allah, le Tout Miséricordieux, le Très Miséricordieux.

Tout d'abord, nous exprimons notre gratitude la plus profonde à Allah, le Tout-Puissant, pour nous avoir accordé la force, la guidance et la sagesse tout au long de notre parcours de recherche et de la rédaction de notre thèse.

Nous tenons à exprimer notre profonde reconnaissance envers notre promoteur, le Mr. Abderrahmane Namane et Mr Aouinane Bilal, pour son précieux encadrement, son soutien et son expertise approfondie ont joué un rôle essentiel dans l'orientation et la qualité de cette recherche. Nous sommes sincèrement reconnaissants de sa direction, qui a été déterminante dans notre croissance intellectuelle et la réussite de cette thèse.

Nous sommes redevables envers nos familles bien-aimées pour leur amour inconditionnel, leur encouragement et leur compréhension. Leur soutien constant, leur motivation et leurs sacrifices ont été les pierres angulaires de nos études. À nos parents, qui sont nos piliers de force, nous leur sommes éternellement reconnaissants pour leur confiance en nous et leurs prières incessantes.

Nous souhaitons exprimer notre sincère appréciation aux membres éminents du jury pour leur précieux temps, leur évaluation méticuleuse et leurs commentaires constructifs. Leur expertise et leurs suggestions avisées ont grandement enrichi ce travail et contribué à sa qualité globale. Nous sommes honorés d'avoir eu l'opportunité de présenter nos recherches devant des personnalités aussi distinguées.

Dédicace

Je dédie ce modeste travail

À

Mon chère papa que dieu te protège **ALI**.

Mon adorable mère **Naima**.

A mes 2 chères sœurs **Tinhinane** et **Ikram** et ma petite nièce **Daline**

A ma grande mère adorée **Zahwa**

Je remercie toute ma famille et mes proches pour leurs encouragements et leur soutien moral.

A mon binôme **Hadjer** pour sa patience et sa compréhension toute au long de ce projet

A tous mes amies

Un profond respect et un remerciement particulier pour : les promoteurs **Mr. Namane abderrahmane** et **Mr. Aouinane bilal**, mes enseignants, et tous ceux qui m'ont aidé à préparer ce mémoire.

Cherifa Khaoula.

Dédicace

Je dédie ce modeste travail à ceux qui, quels que soient les termes utilisés, je n'arriverais jamais à leur exprimer ma sincère reconnaissance. A met tout mon respect :

A la femme qui tiennent une place immense dans mon cœur, merci pour l'encouragement et surtout Les sacrifices, mon adorable mère **djaouida**

mon cher père **mustapha** à ma sœur **khadidja** et **mes frères**, vous avez toujours été pour moi d'une aide très précieuse. Je vous remercie pour soutenir tout au long de mes études que Dieu les protège et leur offre la chance et le bonheur.

A tous mes amis particulièrement, merci pour leur soutien et leur encouragement.

Sans oublier mon binôme **khaoula cherifa** pour son soutien moral sa patience et sa compréhension toute au long de ce projet, du plus profond de mon cœur je vous dis merci ma sœur d'être toujours là pour moi.

A tous mes enseignants de Département d'Électronique université de-Blida Un profond respect **Mr. NAMANE ABDERRAHMANE** et **Mr.AOUINANE BILAL**

et

un remerciement particulier pour mes collègues de travail Pour les aider et les soutenir dans les moments difficiles.

Le permet de remercier les personnes qui ont soutenu et encouragé le rédacteur tout au long de son travail.

Enfin, je dédie ce travail à tous ceux qui me connaissent de près ou de loin.

Hadjer

ملخص

الهدف الأساسي من المذكرة هو تطوير وتقييم طريقة لتحسين دقة الصور باستخدام الشبكات التوليدية المتضادة (Generative Adversarial Networks, GAN). تهدف هذه الطريقة إلى تحسين جودة ودقة الصور منخفضة الدقة من خلال توليد تفاصيل دقيقة وواقعية، مما يتفوق على تقنيات الاستيفاء التقليدية.

كلمات المفاتيح: التعلم العميق، الشبكات التوليدية التنافسية، تحسين الدقة الفائقة للصور.

Résumé

L'objectif principal de ce mémoire est de développer et d'évaluer une méthode de super-résolution d'images basée sur les réseaux antagonistes génératifs (Generative Adversarial Networks, GAN). Cette méthode vise à améliorer la qualité et la résolution des images à faible résolution en générant des détails fins et réalistes, surpassant ainsi les techniques traditionnelles d'interpolation.

Mots clés : Super-résolution d'images, Réseaux antagonistes génératifs (GANs), Apprentissage profond.

Abstract

The objective of implementing Super Resolution imaging using Generative Adversarial Networks (GANs) is to realistically enhance the quality and resolution of low-resolution images. Specifically, it involves transforming images with low pixel density into high quality images by adding fine and realistic details, which traditional super resolution methods often fail to capture.

Keywords: Image super resolution, Generative Adversarial Networks (GANs), Deep learning

Liste des acronymes et abréviations

SR	Super Resolution
HR	Haute Resolution
LR	Low Resolution
RVB	Rouge Vert Bleu
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
IA	Intelligence Artificielle
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
LSTM	Long Short-Term Memory
JPEG	Joint Photographic Experts Group
MLP	Multi-Layer Perceptron
SGD	Stochastic Gradient Descent
ReLU	Rectified Linear Unit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
SRGAN	Super-Resolution Generative Adversarial Network
SRCNN	Super-Resolution Convolutional Neural Network
CIFAR-100	Canadian Institute For Advanced Research - 100 classes
VGGNet	Visual Geometry Group Network
SKFF	Selective Kernel Feature Fusion
SRB	Super-Resolution Block
HWAB	Half Wavelet Attention Block
DAU	Dual Attention Unit
IDWT	Inverse Discrete Wavelet Transform
GAP	Global Average Pooling
RaGAN	Relativistic Average GAN
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
MSE	Mean Squared Error
PIQE	Perception based Image Quality Evaluator
NIQE	Natural Image Quality Evaluator
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator

Table des matières

Remerciements	
Dédicace	
Résumé	
Liste des acronymes et abréviations	
Table des matières	
Liste des figures	
Liste des tableaux	
Introduction générale.....	1
Chapitre 1 : Généralités sur la super résolution	
1.1 Introduction	3
1.2 Traitement d'images	3
1.2.1 Définition d'image	3
1.2.2 Image numérique.....	3
1.2.3 Image en couleurs (rvb).....	4
1.3 Super résolution.....	4
1.3.1 Définition	4
1.3.2 Raisons derrière la super résolution	5
1.3.3 Domaines d'application.....	5
1.3.4 Techniques de la super résolution	7
1.4 Transformée en ondelettes	11
1.4.1 Définition	11
1.4.2 Théorie de la transformée en ondelettes.....	12
1.4.3 Types de transformée en ondelettes	13
1.4.4 Applications de la transformée en ondelettes.....	14
1.5 Intelligence artificielle.....	14
1.6 Apprentissage automatique	15
1.7 Réseaux de neurones	15
1.7.1 Définition	15
1.7.2 Importance dans la super résolution.....	16
1.7.3 Différents types de réseaux de neurones	17
1.7.4 Réseaux de neurones multicouches (MLP).....	17
1.7.5 Apprentissage profond	20

1.7.6 Couches clés d'un CNN.....	21
1.7.7. Etudes de cas et exemples	24
1.8 Réseaux antagonistes génératifs	26
1.8.1 Présentation de la structure du gan.....	26
1.8.2 Super résolution avec generative adversarial network	28
1.9 Conclusion.....	30

Chapitre 2 : Réseaux antagoniste génératif

2.1 Introduction	32
2.2 Méthodologie	32
2.2.1 Générateur	32
2.2.2 Discriminateur	39
2.2.3 Fonctions de perte (loss functions).....	40
2.3 Conclusion.....	41

Chapitre 3 : Simulation de la super résolution

3.1 Introduction	43
3.2 Outils utilisés.....	43
3.3 Datasets	44
3.4 Détails de l'implémentation	45
3.5 Métriques d'évaluation de la qualité de l'image	45
3.5.1 Métriques de qualité avec référence.....	45
3.5.2 Métriques de qualité sans référence	46
3.6 Résultats et discussion.....	47
3.6.1 Comparaison qualitative	47
3.6.2 Comparaison quantitative	51
3.7 Conclusion.....	51
Conclusion générale	53
Bibliographie	55

Liste des figures

N°	Figure	Page
01	Figure 1.1. : Image numérique à niveaux de gris	4
02	Figure 1.2 : Image numérique en couleurs (RVB)	4
03	Figure 1.3 : Basse résolution et haute résolution d'image	5
04	Figure 1.4 : Utilisation de la Super-Résolution pour l'Amélioration des Images de Surveillance	6
05	Figure 1.5 : Utilisation de la Super-Résolution pour l'Amélioration des Images Médicales	6
06	Figure 1.6 : Utilisation de la super résolution pour l'amélioration des images satellitaires	7
07	Figure 1.7 : Utilisation de la super résolution en photographie numérique	7
08	Figure 1.8 : Classification de l'approche de la super résolution d'image	8
09	Figure 1.9 : Méthode du plus proche voisin (Nearest Neighbor)	8
10	Figure 1.10 : Interpolation bi-cubique	9
11	Figure 1.11 : Méthodes fréquentielles	10
12	Figure 1.12 : Transformée en ondelettes (wavelet transform)	11
13	Figure 1.13 : Spectres d'absorption et (B) spectres CWT des composés M (-) et N (-) et de leur mélange (-)	13
14	Figure 1.14: Transformée en ondelettes discrètes.	13
15	Figure 1.15 : Relations hiérarchiques entre différentes sous-catégories du domaine de l'intelligence artificielle (IA)	15
16	Figure 1.16: Vue simplifiée d'un réseau artificiel de neurones	16
17	Figure 1.17 : Réseaux de Neurones Multicouches (MLP) [26]	18
18	Figure 1.18 : Application des MLP dans la Super Résolution « CIFAR-100 »	20
19	Figure 1.19 : Schéma d'un Réseau de neurones à convolution (CNN)	21
20	Figure 1.20 : Représentation d'une couche convolutive	22
21	Figure 1.21 : La couche d'activation ReLU (Rectified Linear Unit)	22
22	Figure 1.22 : La couche d'activation Sigmoid Tanh et ReLU	22
23	Figure 1.23 : Processus de max-pooling	23
24	Figure 1.24 : Les couches entièrement connectées	23
25	Figure 1.25 : Modèle de réseau neuronal par Dropout	24

26	Figure 1.26: AlexNet architecture	25
27	Figure 1.27: VGGNet architecture	25
28	Figure 1.28: YOLO architecture	26
29	Figure 1.29 : Schéma des réseaux antagonistes génératifs (GAN)	27
30	Figure 1.30 : Comportement du réseau avant convergence	27
31	Figure 1.31 : Résultat après convergence	28
32	Figure 1.32: Architecture du SRGAN.	28
33	Figure 1.33 : SRCNN architecture	29
34	Figure 2.1 : Schéma synoptique simplifié du SRGAN	32
35	Figure 2.2 : Architecture M-Net	33
36	Figure 2.3 : Architecture du stage 1 (Agrandissement de l'image)	33
37	Figure 2.4 : Architecture du stage 2 (Restauration de l'image)	34
38	Figure 2.5: Super resolution block (SRB)	36
39	Figure 2.6: Selective kernel feature fusion (SKFF)	37
40	Figure 2.7: Half Wavelet Attention Block (HWAB)	38
41	Figure 2.8: Dual attention unit (DAU)	38
42	Figure 2.9: Discriminateur Markovien (PatchGAN)	39
43	Figure 2.10: Architecture du discriminateur Markovien (PatchGAN)	40
44	Figure 3.1 : Résultats obtenus (stage 1)	47
45	Figure 3.2 : Résultats obtenus (stage 2)	48
46	Figure 3.3 : Comparaison qualitative des résultats obtenus avec la super-résolution par l'interpolation bicubique.	49
47	Figure 3.4 : Comparaison qualitative des résultats obtenus avec la super-résolution par l'interpolation bilinéaire.	50

Liste des tableaux

N°	Tableau	Page
01	Tableau 3.1 : Comparaison quantitative des résultats sur des ensembles de données. (+) indique que le score le plus élevé est meilleur, (-) indique le score le plus bas est meilleur. Les deux meilleurs résultats sont en rouge et bleu respectivement	51

Introduction générale

Introduction générale

L'émergence de l'intelligence artificielle (IA) et son importance croissante comme technologie transformatrice dans divers domaines ont conduit à des avancées significatives dans le traitement des images. La super-résolution d'image (SRI), qui consiste à améliorer la qualité des images basse résolution (LR) pour obtenir des images haute résolution (HR), est une application particulièrement prometteuse. Elle trouve des applications dans la surveillance, la médecine, la photographie et bien d'autres secteurs.

Dans ce mémoire, nous explorons l'utilisation des réseaux adversaires génératifs (GAN) pour la super-résolution d'image. Les GAN, composés d'un générateur et d'un discriminateur, permettent de générer des images HR à partir d'images LR en utilisant un processus de compétition entre les deux réseaux. Ce processus, introduit par Ian Goodfellow en 2014, a révolutionné la manière dont les images peuvent être améliorées, surpassant souvent les méthodes traditionnelles.

Notre objectif principal est de développer et d'évaluer un modèle de super résolution basé sur les GAN. Nous explorons diverses architectures et techniques pour améliorer les performances du modèle. Nous utilisons des mesures de qualité d'image telles que le PSNR, le SSIM et d'autres indicateurs pour évaluer l'efficacité du modèle proposé.

Le mémoire est structuré en trois chapitres suivis d'une conclusion générale :

- ❖ Chapitre 1: Ce chapitre fournit un aperçu des techniques de super-résolution d'image, en expliquant les concepts de base et les différentes méthodes utilisées dans ce domaine.
- ❖ Chapitre 2: Ce chapitre aborde les aspects théoriques des GAN, y compris leur architecture, leur fonctionnement, et les raisons pour lesquelles ils sont efficaces pour la super résolution d'image.
- ❖ Chapitre 3: Le troisième chapitre se concentre sur la mise en œuvre pratique de notre modèle. Il décrit les étapes de développement, et présente les résultats obtenus

Chapitre 1 : Généralités sur la super-résolution

1.1 Introduction

La super résolution joue un rôle crucial dans de nombreux domaines en améliorant la résolution spatiale des images, permettant d'obtenir des détails fins à partir d'images de résolution plus basse. Cette technique est essentielle pour des applications telles que l'imagerie médicale, l'imagerie satellite, la surveillance et la photographie numérique, où la qualité et le niveau de détail des images sont primordiaux. Dans le monde actuel, l'image est une source importante d'information, utilisée sous diverses formes telles que photos, graphiques, données 2D et 3D, et résonance magnétique (MR).

Le traitement d'images, qui inclut des techniques de super-résolution, est utilisé pour extraire des informations fiables des images. Les technologies récentes se concentrent sur l'amélioration, la restauration, la compression, l'extraction, la transformation et la super-résolution d'images dans les systèmes de vision industrielle. Avec l'évolution des technologies de l'information, la super-résolution a fait de grands progrès, devenant une classe importante de techniques de traitement d'images et de vision par ordinateur. La super-résolution d'image (SR) consiste à reconstruire des images haute résolution (HR) à partir d'images basse résolution (LR) [1].

1.2 Traitement d'images

1.2.1 Définition d'image

Une image est un visuel d'un objet, d'une personne, d'un animal ou d'un autre objet qui peut être saisi par l'œil humain à l'aide de différentes techniques : dessin, conception, photographie, vidéo, etc.....

1.2.2 Image numérique

L'image numérique se présente sous la forme d'un tableau à deux dimensions contenant des valeurs entières. Chaque élément de ce tableau, appelé pixel, est codé par un nombre binaire de 8 bits pour les images en niveaux de gris (voir figure 1.1), ou par trois nombres binaires de 24 bits, représentant les teintes de rouge, vert et bleu (codage RVB) pour les images en couleur (voir figure 1.1) [1].

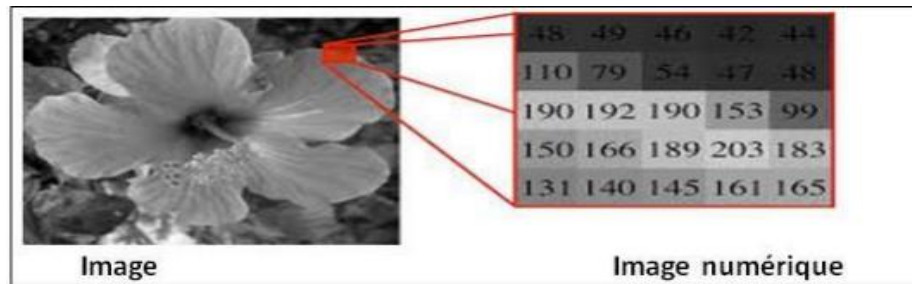


Figure 1.1 : Image numérique à niveaux de gris

1.2.3 Image en couleurs (RVB)

Dans une image couleur, chaque pixel possède une couleur décrite par la quantité de ces 3 composantes : rouge(R), vert(V) et bleu (B), chacune de ces couleurs est codée sur l'intervalle $[0, 255]$ (voir figure 1.2). La combinaison de ces trois couleurs donne un point lumineux (un pixel) d'une certaine couleur. Donc le système RVB est une des façons de décrire une couleur en informatique. Par exemple : Le triplet $\{255, 255, 255\}$ donnera du blanc, $\{255, 0, 0\}$ un rouge pur, $\{100, 100, 100\}$ un gris, etc. Le premier nombre donne la composante rouge, le deuxième la composante verte et le dernier la composante bleue.

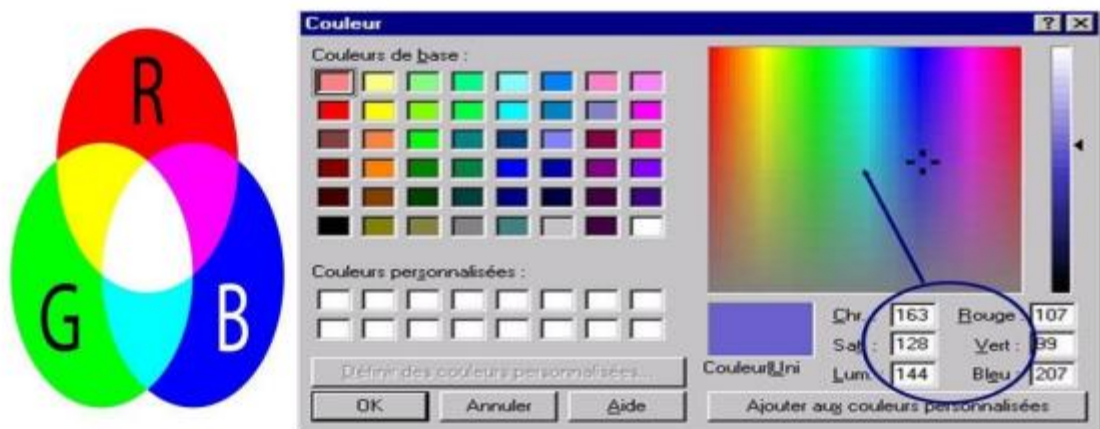


Figure 1.2 : Image numérique en couleurs (RVB)

1.3 Super résolution

1.3.1 Définition

La super résolution désigne une série de méthodes d'imagerie qui permettent de convertir une image ou une vidéo à faible résolution en image de résolution plus élevée. Le thème a été évoqué pour la première fois au milieu des années 80 et le mot «super résolution» lui-même est apparu vers 1990. Toutefois, cette idée s'est développée au fil du temps, il a gagné en popularité

et est devenu particulièrement pratique aujourd'hui. Même si plusieurs méthodes ont été mises en place, elles partagent généralement la même base, à savoir la régénération des pixels manquants à partir d'une image de faible résolution grâce à des algorithmes spécifiques [2].

1.3.2 Raisons derrière la super résolution

La super-résolution (SR) vise principalement à produire une image de résolution plus élevée à partir d'images de résolution plus faibles. Les images haute résolution présentent une densité de pixels élevée, ce qui permet de capturer davantage de détails sur la scène initiale. Cette richesse en détails est cruciale pour de nombreuses applications de vision par ordinateur, qui nécessitent une résolution élevée afin d'améliorer les performances dans la reconnaissance des formes et l'analyse des images [3].

La résolution élevée joue un rôle crucial dans l'imagerie médicale pour le diagnostic. Un zoom sur une zone d'intérêt spécifique dans l'image est indispensable pour de nombreuses applications, telles que les applications de surveillance, médico-légales et d'imagerie par satellite (voir figure 1.3).



Figure 1.3: Basse résolution et haute résolution d'image

1.3.3 Domaines d'application

La super résolution a de nombreuses applications dans divers domaines en raison de sa capacité à améliorer la qualité visuelle et la clarté des images. Voici les principaux domaines d'application de la super-résolution :

- Domaine de surveillance : La super résolution est utilisée pour détecter, identifier et effectuer une reconnaissance faciale sur des images basse résolution obtenues à partir de caméras de sécurité (voir figure 1.4). Elle permet d'extraire plus de détails des images de surveillance pour améliorer les performances des systèmes de détection et d'identification [4].

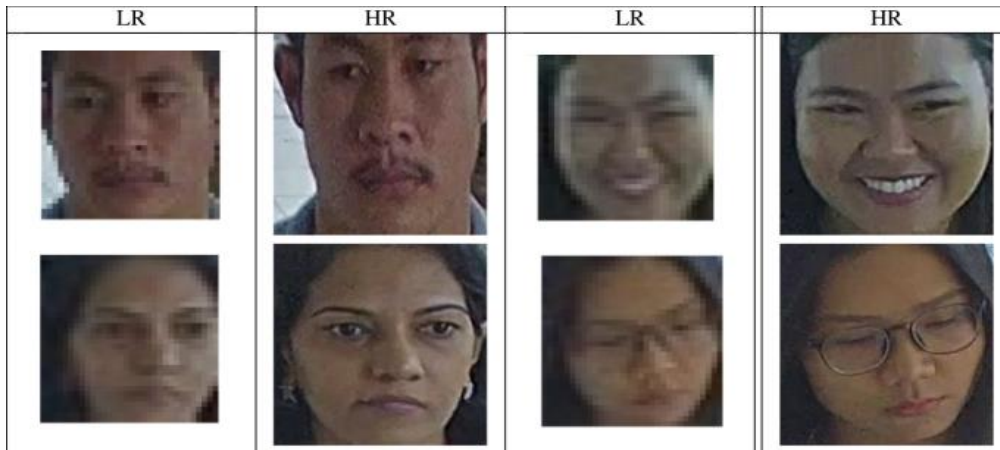


Figure 1.4: Utilisation de la Super-Résolution pour l'Amélioration des Images de Surveillance

-Domaine médical : En imagerie médicale, la super-résolution aide à générer des images IRM haute résolution à partir d'images autrement à faible résolution (voir figure 1.5). Cela permet de résoudre les défis liés à l'acquisition d'images IRM haute résolution en termes de durée d'acquisition, de couverture spatiale et de rapport signal sur bruit. Une résolution élevée est cruciale pour le diagnostic médical [4].

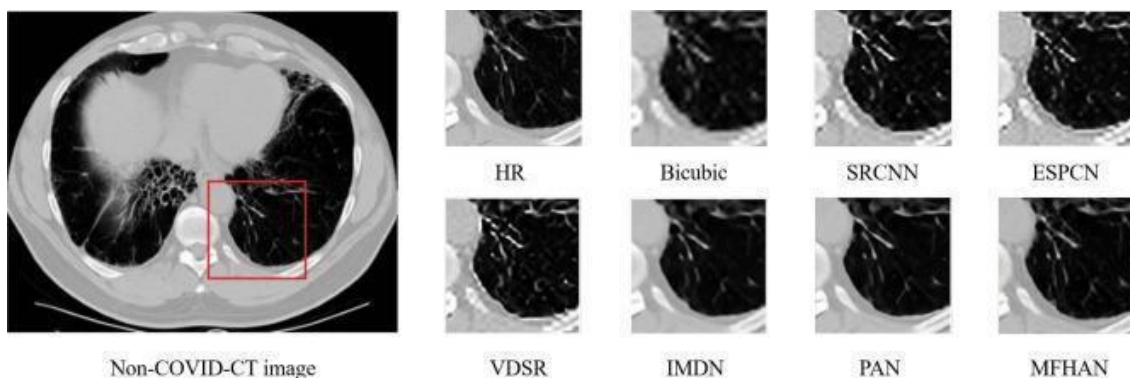


Figure 1.5: Utilisation de la Super-Résolution pour l'Amélioration des Images Médicales

-Domaine satellitaire : la super résolution est appliquée à l'imagerie satellitaire pour obtenir plus de détails des images basse résolution (voir figure 1.6). Un zoom sur des zones d'intérêt spécifiques est souvent nécessaire dans ce domaine, et la super-résolution permet d'améliorer la qualité des images pour des applications telles que la surveillance environnementale et la cartographie [4].



Figure 1.6: Utilisation de la super résolution pour l'amélioration des images satellitaires.

-Domaine de photographie numérique : la super-résolution est utilisée pour améliorer la qualité des images, en particulier lors d'un zoom numérique (voir figure 1.7). Elle permet de générer des images haute résolution à partir d'images basse résolution, offrant plus de détails et de netteté [4].

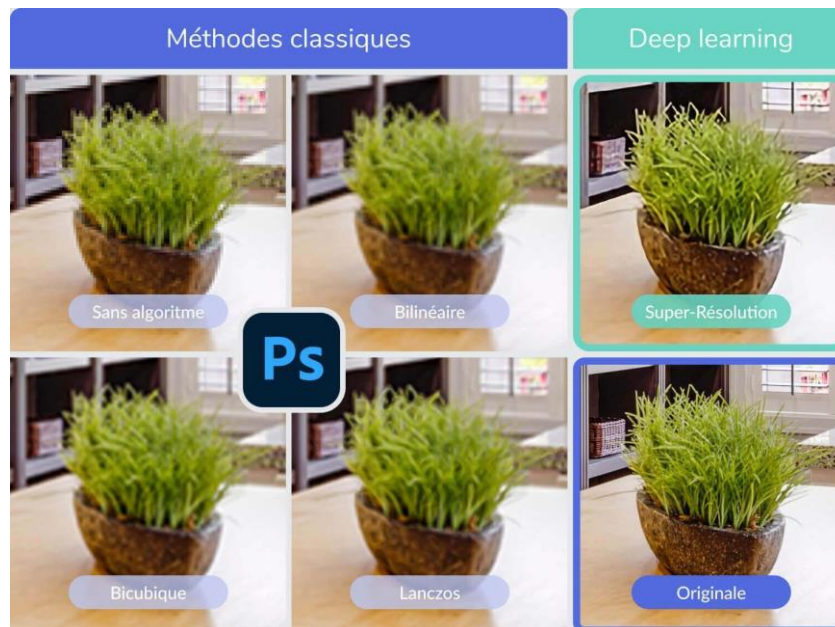


Figure 1.7: Utilisation de la super résolution en photographie numérique

1.3.4 Techniques de la super résolution

La super-résolution d'image (SRI) est une technique visant à augmenter la résolution d'une image en améliorant la qualité et la clarté des détails. Les approches de super-résolution peuvent être classées en plusieurs catégories basées sur leurs méthodes et techniques (voir figure 1.8) [5].

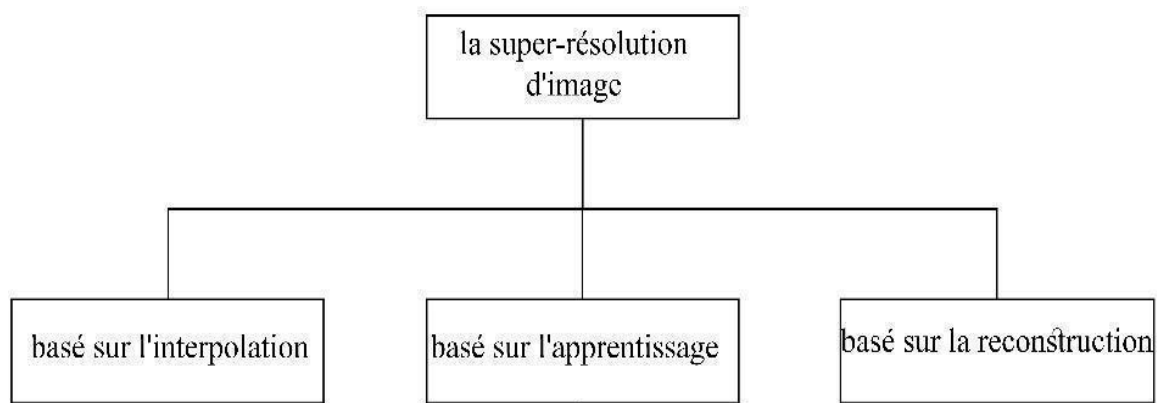


Figure 1.8 : Classification de l'approche de la super résolution d'image [5].

a - Méthode d'interpolation

L'interpolation est l'une des méthodes les plus simples et les plus couramment utilisées pour la super-résolution. Elle consiste à estimer les valeurs des pixels manquants en utilisant les pixels voisins [5].

- Méthode de plus proche voisin :** La méthode de plus proche voisin (en anglais ; Nearest Neighbor) repose sur un principe simple : chaque pixel de l'image haute résolution est attribué la valeur du pixel le plus proche de l'image basse résolution (voir figure 1.9). Cette méthode est appréciée pour sa simplicité et sa rapidité. Cependant, elle présente des limites notables, notamment la production d'images avec des artefacts en forme de marche d'escalier et une qualité inférieure. Malgré ces inconvénients, elle est utilisée dans des contextes où la rapidité est plus importante que la qualité, comme dans certaines applications en temps réel [1,5].

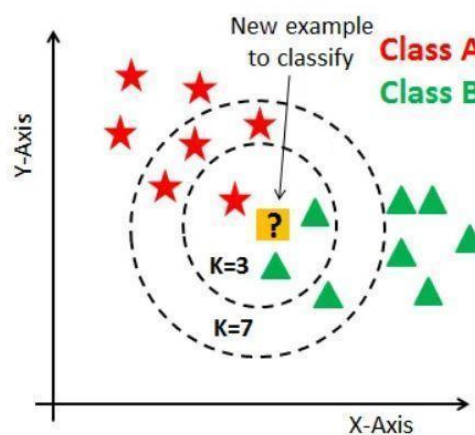


Figure 1.9: Méthode du plus proche voisin (Nearest Neighbor)

• **Interpolation Bi-Cubique** : L'interpolation bi-cubique utilise une fonction cubique pour interpoler les pixels, en prenant en compte 16 pixels voisins. Cette méthode produit des images plus nettes et avec moins d'artefacts que la méthode du plus proche voisin (voir figure 1.10). Cependant, elle est plus complexe et plus lente. L'interpolation bi-cubique est couramment utilisée dans les logiciels de traitement d'images pour des tâches de redimensionnement où la qualité est essentielle [1].

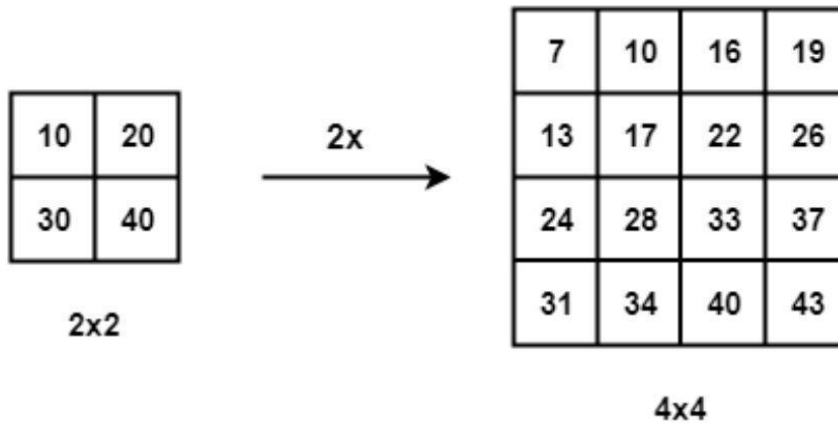
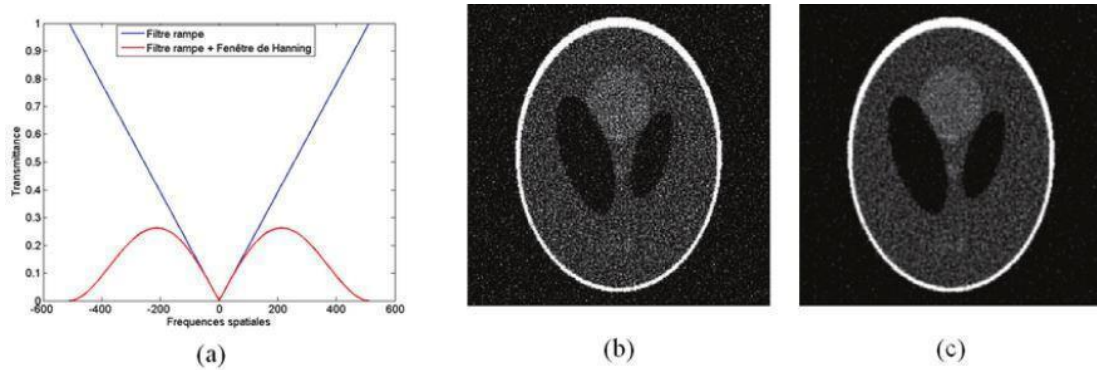


Figure 1.10: Interpolation bi-cubique .

b - Méthode de reconstruction

En plus des méthodes d'interpolation, d'autres techniques classiques sont utilisées pour la super-résolution [1].

• **Méthodes fréquentielles** : Les méthodes fréquentes utilisent la transformation de Fourier pour manipuler les coefficients de fréquence afin d'améliorer la résolution des images (voir figure 1.11). Cette approche repose sur l'analyse des fréquences spatiales de l'image. Elle est particulièrement utile dans des domaines nécessitant une analyse fréquentielle précise, tels que l'imagerie médicale et satellitaire [6-7].



Réponse fréquentielle du filtre rampe et du filtre apodisé (a). Les deux autres figures représentent la reconstruction par FBP d'un sinogramme bruité, traité avec un filtre rampe (b) et un filtre rampe apodisé par une fenêtre de Hanning (c).

Figure 1.11: Méthodes fréquentielles [8].

- **Approches par projection arrière :** Les approches par projection arrière corrigent les erreurs de projection en ajustant itérativement l'image reconstruite pour qu'elle soit conforme aux projections observées. Cela implique souvent des itérations multiples où l'image est progressivement affinée. Ces méthodes sont utilisées dans des applications de reconstruction d'image où plusieurs vues de la même scène sont disponibles, comme en tomographie ou en imagerie 3D [6-7].

c- Méthodes basées sur l'apprentissage

L'apprentissage est une méthode essentielle pour la super-résolution, et elle comprend plusieurs sous-méthodes.

- **Méthode expositive :** La méthode expositive est basée sur un apprentissage vertical, où le formateur détient toutes les connaissances et les transmet aux apprenants. Un exemple typique de cette méthode est le cours magistral à l'université, où les élèves prennent des notes tandis que l'enseignant dispense des connaissances [9-10].
- **Méthode démonstrative :** La méthode démonstrative repose sur une approche descendante de l'enseignement, où l'enseignant démontre les procédures et les stagiaires les reproduisent. Cette méthode est souvent utilisée pour la visualisation et la mémorisation des actions du formateur, suivie de leur reproduction par les apprenants [9,11].
- **Méthode active :** La méthode active met l'action au cœur de l'apprentissage, favorisant la construction des connaissances par l'apprenant. Des exemples typiques incluent les études de cas, les projets de groupe et les jeux de rôle, qui encouragent la réflexion et l'engagement actif des apprenants [9,11].

- **Méthode expérientielle** : La méthode expérientielle se concentre sur l'apprentissage dans des conditions réelles, où l'action et l'erreur sont des éléments clés pour acquérir des compétences. Cette méthode implique la pratique et des expériences concrètes, permettant aux apprenants d'améliorer leurs compétences grâce à des situations réelles [10-11].
- **Méthode heuristique** : La méthode heuristique fait appel à la créativité et à l'imagination des apprenants, les plaçant au cœur de l'action de formation. Elle utilise des techniques comme le brainstorming et la création de scénarios pour favoriser l'autonomie et la collaboration parmi les apprenants [10-11].

1.4 Transformée en ondelettes

La super résolution consiste à améliorer la résolution d'une image basse résolution en générant une image de plus haute résolution. La transformée en ondelettes est utilisée dans ce contexte pour analyser les détails de l'image à différentes échelles et améliorer la précision de la reconstruction. En décomposant l'image en composantes de haute et basse fréquence, la transformée en ondelettes permet d'isoler les détails fins et de les traiter séparément, ce qui est essentiel pour une reconstruction précise [12].

1.4.1 Définition

La transformée en ondelettes est une technique mathématique utilisée pour décomposer une fonction ou un signal en composantes de différentes fréquences, chacune localisée dans le temps (voir figure 1.12). Contrairement à la transformée de Fourier, qui fournit une analyse en fréquence globale, la transformée en ondelettes permet une analyse à la fois en temps et en fréquence, ce qui est particulièrement utile pour les signaux non stationnaires. Les ondelettes sont des fonctions oscillantes de durée finie qui peuvent être utilisées pour analyser les variations locales d'un signal [12].

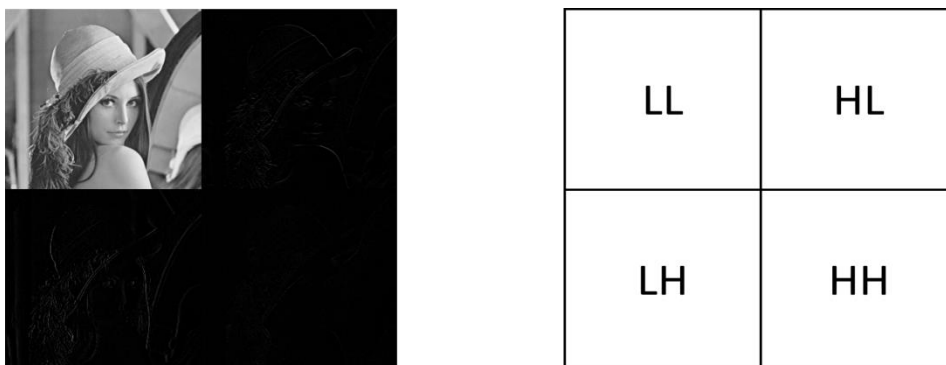


Figure 1.12 : Transformée en ondelettes (Wavelet Transform)

1.4.2 Théorie de la transformée en ondelettes

Transformée en ondelettes est une transformation continue qui permet de représenter un signal comme une somme pondérée de petites ondelettes translatées et dilatées. Elle est utilisée pour analyser les signaux non stationnaires et les signaux à fréquence variable [12]. La transformée en ondelettes est définie par la formule suivante :

$$g(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \overline{\psi_{a,b}(t)} dt$$

où :

- $\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$ est l'ondelette fille,
- a est l'échelle,
- b est la position,
- $x(t)$ est le signal,
- $\psi(t)$ est l'ondelette mère

a . Concept de Base

La transformée en ondelette est un outil mathématique utilisé pour décomposer un signal en différents composants à différentes échelles. Contrairement à la transformée de Fourier, qui décompose un signal en une série de sinusoides infiniment étendues, la transformée en ondelette utilise des fonctions localisées dans le temps et la fréquence appelées ondelettes. Cela permet d'obtenir des informations à la fois temporelles et fréquentielles de manière plus précise, ce qui est particulièrement utile pour l'analyse des signaux non stationnaires [12].

b . Ondelettes mères et ondelettes filles

Le concept central de la transformée en ondelettes repose sur les ondes mères et filles:

- **Ondelette Mère (Mother Wavelet)** : L'ondelette mère est une fonction de base qui est dilatée et translatée pour analyser le signal. Elle doit être bien localisée en temps et avoir une oscillation limitée [13].
- **Ondelette Fille (Daughter Wavelet)** : Les ondelettes filles sont générées à partir de l'ondelette mère par dilatation (changement d'échelle) et translation (déplacement dans le temps) [13]. Les ondelettes filles sont définies par la formule :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

où a est le facteur d'échelle et b est le facteur de translation.

1.4.3 Types de transformée en ondelettes

a . Transformée en ondelettes continue (CWT)

La CWT est utilisée pour obtenir une représentation redondante d'un signal à différentes échelles et translations. Elle offre une grande flexibilité et précision pour l'analyse des signaux, mais elle est souvent coûteuse en termes de calcul (voir figure 1.13) [14].

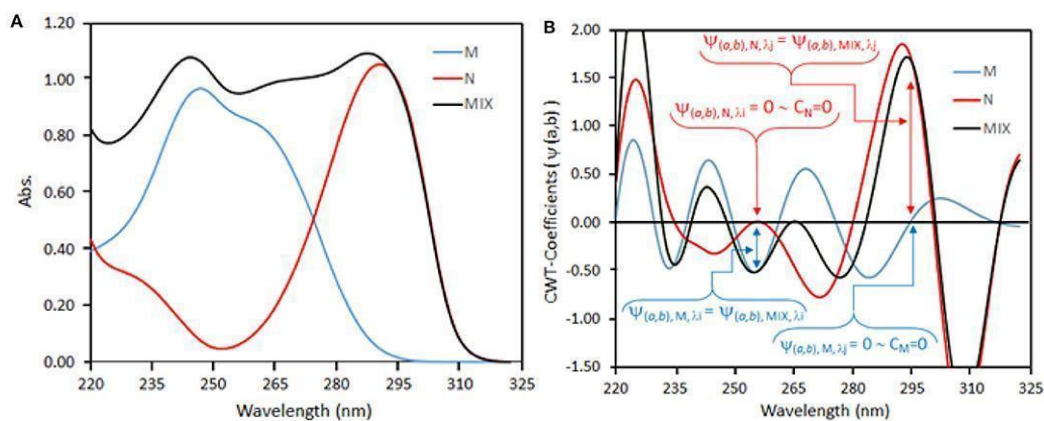


Figure 1.13 : Spectres d'absorption et (B) spectres CWT des composés M (–) et N (–) et de leur mélange (–)

b . Transformation en ondelettes discrètes (DWT)

Transformation en ondelettes discrètes est une transformation qui décompose un signal donné en plusieurs ensembles, où chaque ensemble est une série temporelle de coefficients décrivant l'évolution temporelle du signal dans la bande de fréquences correspondante (voir figure 1.14) [13].

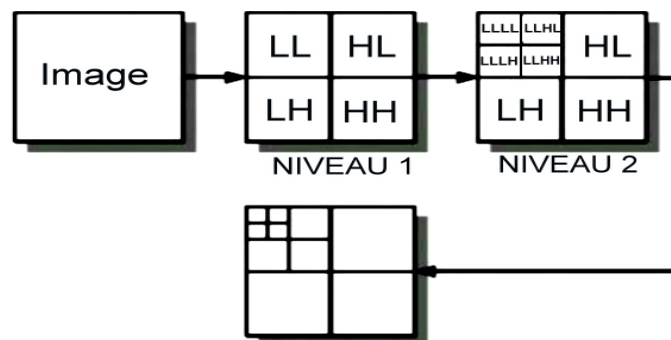


Figure 1.14: Transformée en ondelettes discrètes.

La DWT est largement utilisée dans la compression d'images, notamment dans la norme JPEG-2000. Elle permet une compression efficace en décomposant l'image en sous-bandes de fréquences [15].

1.4.4 Applications de la transformée en ondelettes

a . Compression d'Images

La compression d'images est une application importante de la transformée en ondelette. Cette méthode permet de réduire la taille de l'image en décomposant l'image en composantes fréquentielles et spatiales. Les coefficients d'ondelette sont quantifiés et codés pour réduire la taille de l'image. Le format JPEG2000, qui utilise la transformée en ondelette, offre une meilleure compression que le format JPEG traditionnel. Cette méthode est particulièrement efficace pour les images contenant des détails fins et des textures complexes [16].

b . Détection et débruitage de signaux

La détection et le débruitage de signaux sont deux applications clés de la transformée en ondelette. Les ondelettes sont utilisées pour détecter les contours des objets dans une image. Cette application est particulièrement utile pour la reconnaissance des empreintes digitales. Les ondelettes sont également efficaces pour le débruitage de signaux en distinguant le bruit présent partout sur une image de l'information réelle. Cette méthode est particulièrement utile pour améliorer la qualité des images numériques [17].

c . Analyse multi-résolution

L'analyse multi-résolution est une application fondamentale de la transformée en ondelette. Cette méthode permet d'analyser un signal à différentes échelles et résolutions, ce qui est particulièrement utile pour l'analyse de signaux non stationnaires. La transformée en ondelettes offre la possibilité d'analyser un signal simultanément dans le domaine du temps et celui des fréquences, ce qui est impossible avec la transformée de Fourier. Cette méthode est particulièrement utile pour l'analyse de signaux complexes et non stationnaires [17].

1.5 Intelligence artificielle

L'intelligence artificielle (IA) est un processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique. Son but est de permettre à des ordinateurs de penser et d'agir comme des êtres humains (voir figure 1.15) [18].

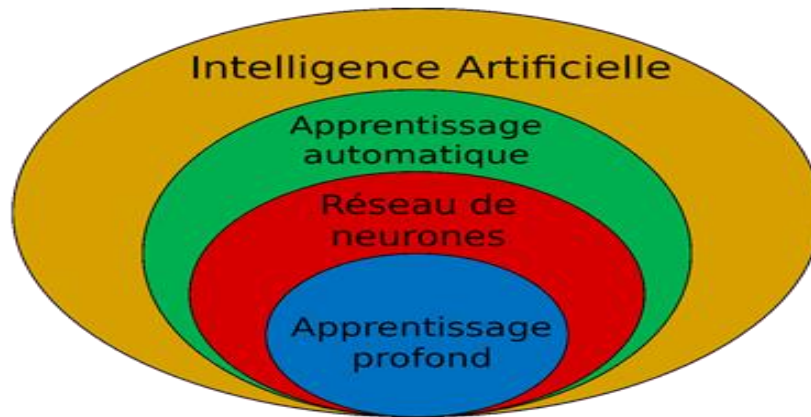


Figure 1.15 : Relations hiérarchiques entre différentes sous-catégories du domaine de l'intelligence artificielle (IA)

1.6 Apprentissage automatique

L'apprentissage automatique ou Machine Learning (ML) est un sous-ensemble de l'intelligence artificielle qui se concentre sur le développement d'algorithmes et de modèles statistiques capables de faire des prédictions ou de classer des données sur la base des informations apprises à partir des données d'apprentissage. C'est-à-dire d'effectuer des tâches sans être explicitement programmé [19].

Le ML est de plus en plus populaire en raison de ses nombreux avantages. La disponibilité accrue de données et la capacité de calcul croissante ont permis de développer des logiciels "intelligents" capables de résoudre des problèmes complexes et variés. Grâce à l'apprentissage automatique, les ordinateurs sont capables d'apprendre en permanence et d'améliorer leur performance, parfois même au-delà des capacités humaines [19].

1.7 Réseaux de neurones

Les premiers réseaux de neurones ont été développés par Yann LeCun en 1988. Il s'agissait d'un réseau utilisé pour le traitement d'image et la détection d'objets. Yann LeCun a introduit le concept des réseaux de neurones convolutifs (CNNs), qui sont particulièrement efficaces pour analyser les données visuelles [20].

1.7.1 Définition

Les réseaux de neurones artificiels sont des systèmes de calcul inspirés du fonctionnement du cerveau humain (voir figure 1.16). Composés de nœuds interconnectés appelés neurones, ils apprennent à partir de données et prennent des décisions basées sur ces apprentissages. Ces

neurones sont organisés en différentes couches, incluant une couche d'entrée pour recevoir les informations, une ou plusieurs couches cachées pour effectuer les calculs, et une couche de sortie pour fournir le résultat final, ces réseaux transmettent les résultats entre neurones successifs. Largement utilisés en intelligence artificielle pour des tâches complexes telles que la reconnaissance d'images, le traitement du langage naturel et la prédiction, ils sont capables d'apprendre à partir d'exemples, de détecter des modèles et des tendances dans les données, et de généraliser pour faire des prédictions précises [20] [21] [22].

Il existe différents types de réseaux de neurones qui diffèrent par leur topologie, leurs fonctions d'agrégation et de seuillage, leurs algorithmes d'apprentissage, etc... [21]. Ils représentent une avancée majeure qui a révolutionné de nombreux domaines comme la santé, la finance et la technologie [21].

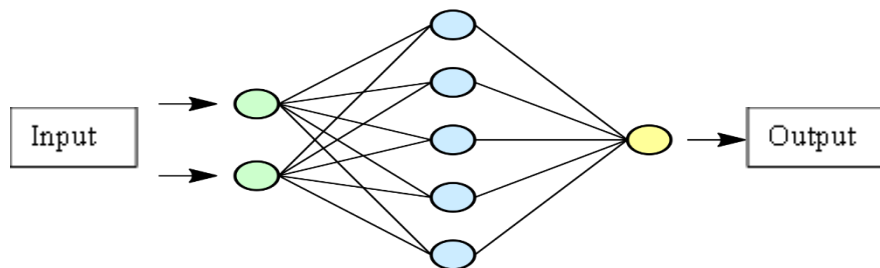


Figure 1.16: Vue simplifiée d'un réseau artificiel de neurones

1.7.2 Importance dans la super résolution

Les réseaux de neurones sont largement utilisés dans la super résolution d'images pour plusieurs raisons clés :

a- Capacité d'apprentissage complexe : Les réseaux de neurones sont capables d'apprendre des fonctions extrêmement complexes à partir d'exemples, permettant de reconstruire des détails fins et d'améliorer la qualité des images de manière précise [20,24].

b- Généralisation à partir d'exemples : Après entraînement sur des paires d'images basse et haute résolution, les réseaux de neurones peuvent généraliser pour reconstruire des images haute résolution à partir de nouvelles images basse résolution [20,23].

c- Automatisation de l'apprentissage : Contrairement aux méthodes traditionnelles, les réseaux de neurones apprennent automatiquement les filtres optimaux pour la super-résolution à partir des données, sans nécessiter de conception manuelle [17,23].

d- Performances élevées : Les meilleures méthodes de super-résolution actuelles sont toutes basées sur les réseaux de neurones, surpassant les approches classiques en termes de qualité d'image reconstruite [20,25].

e- Flexibilité et adaptabilité : Les architectures de réseaux de neurones peuvent être adaptées et améliorées pour la super-résolution, par exemple en ajoutant des blocs résiduels, des réseaux adversaires, ou en intégrant des cycles [23,25].

1.7.3 Différents types de réseaux de neurones

Les réseaux de neurones sont généralement classés en fonction du nombre d'épaisseurs qui séparent l'entrée de données de la sortie du résultat, chacun étant conçu pour résoudre des tâches spécifiques en fonction de leur structure et de leur fonctionnement. Voici quelques illustrations de réseaux neuronaux fréquemment employés :

- **Réseaux de neurones multicouches (MLP):** connus sous le nom de réseaux de neurones à propagation avant.
- **Réseaux de neurones convolutifs (CNN):** Ces réseaux sont plus utilisés pour le traitement des images et de la vision par ordinateur. Elles contiennent des couches de convolution qui filtrent les intrants pour en extraire les caractéristiques significatives, suivies de couches de regroupement qui réduisent la taille de la représentation [1].
- **Réseaux de neurones récurrents (RNN):** Ces réseaux ont pour but de traiter des données séquentielles.
- **Réseaux de neurones générateurs adverses (GAN):** Ces réseaux sont communément utilisés pour produire des images, des vidéos ou des sons réalistes à partir de données brutes [26].
- **Réseaux de neurones auto-encodeurs (AE):** Ces réseaux sont utilisés pour la compression et la reconstruction de données. Ils sont constitués de deux parties : un encodeur qui transforme les données en une représentation compressée, et un décodeur qui transforme la représentation compressée en une reconstruction de l'entrée originale [8].
- **Réseaux de neurones récurrents à mémoire à court terme (LSTM) :** Ces réseaux sont une forme de RNN qui a la capacité de conserver des informations sur une longue période de temps [5].

1.7.4 Réseaux de neurones multicouches (MLP)

Les réseaux de neurones multicouches (MLP) sont un type de modèle de réseau de neurones artificiel qui se compose d'une ou plusieurs couches de neurones interconnectées [27-29] (voir figure 1.17). Ils sont utilisés pour résoudre des problèmes de prédiction de classification et de régression sur des données tabulaires [27-28].

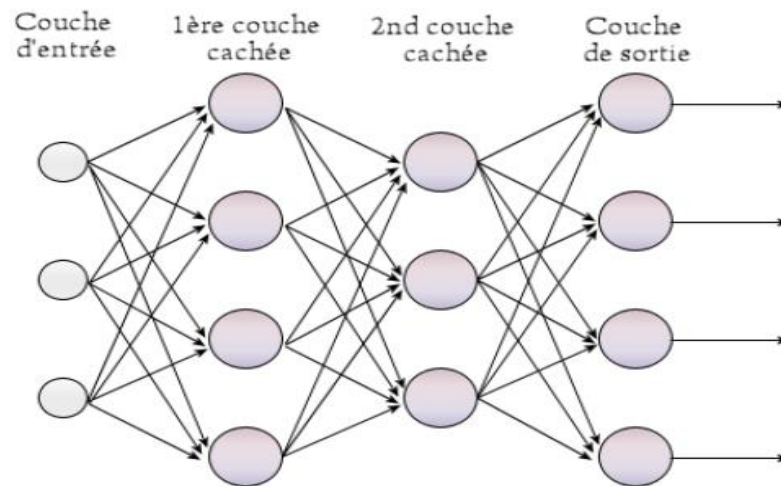


Figure 1.17 : Réseaux de Neurons Multicouches (MLP) [28].

a . Structure et fonctionnement des MLP

Les perceptrons multicouches (MLP) sont des réseaux neuronaux artificiels organisés en plusieurs couches. Voici la structure des MLP :

• Couches d'un MLP

- **Couche d'entrée :** Reçoit les entrées du modèle, avec un nombre de neurones correspondant aux caractéristiques de l'entrée.
- **Couches cachées :** Entre la couche d'entrée et la couche de sortie, ces couches extraient des caractéristiques et transforment les données.
- **Couche de sortie :** Produit la sortie du modèle, avec un nombre de neurones adapté à la tâche (par exemple, un neurone pour la classification binaire).

• Neurones et connexions

Chaque neurone est connecté à tous les neurones des couches précédentes et suivantes, avec des connexions pondérées. Les neurones appliquent des fonctions d'activation non linéaires comme la sigmoïde, la tangente hyperbolique (Tanh) ou la ReLU.

• Fonctionnement des MLP

- **Propagation avant :** Les données sont propagées couche par couche, calculant la somme pondérée des entrées et les passant par la fonction d'activation.
- **Erreur et fonction de coût :** La sortie est comparée à la sortie attendue avec une fonction de coût comme l'erreur quadratique moyenne (MSE) ou l'entropie croisée.

- **Rétro-propagation** : Algorithme clé pour ajuster les poids, calculant les gradients de la fonction de coût et mettant à jour les poids pour minimiser l'erreur.

Les MLP sont puissants pour des tâches complexes grâce à leur capacité à apprendre des représentations hiérarchiques des données et à ajuster les poids par rétro-propagation pour améliorer les prédictions [29-32].

- **Algorithmes de formation**

Les algorithmes de formation des perceptrons multicouches (MLP) sont essentiels pour ajuster les poids du réseau afin de minimiser l'erreur de prédiction. Le processus d'apprentissage est généralement supervisé et repose sur des algorithmes d'optimisation qui utilisent le gradient de la fonction de coût. Voici les principaux algorithmes de formation utilisés pour les MLP [33-35]:

- **Descente de gradient (Gradient Descent)** : La descente de gradient est l'algorithme de base pour l'optimisation des MLP [36]. Il ajuste les poids du réseau en suivant la direction opposée du gradient de la fonction de coût par rapport aux poids :

$$w_{ij} = w_{ij} - \eta \frac{\partial J}{\partial w_{ij}}$$

où η est le taux d'apprentissage, w_{ij} est le poids entre les neurones i et j , et J est la fonction de coût.

- **Type d'ajustement de poids synaptiques** :

- **Batch gradient descent** : Utilise l'ensemble complet des données pour calculer le gradient. Cela peut être lent pour les grands ensembles de données [36].
- **Stochastic gradient descent (SGD)** : Utilise une seule instance de données à la fois pour mettre à jour les poids. Cela introduit une certaine variabilité dans les mises à jour, ce qui peut aider à échapper aux minima locaux [36].
- **Mini-batch gradient descent** : Une combinaison des deux précédents, utilisant un petit lot (mini-batch) de données pour chaque mise à jour, offrant un bon compromis entre la stabilité et la vitesse [36].

- **Application des MLP dans la super résolution**

Les Multi-Layer Perceptrons (MLP) sont utilisés dans divers domaines de l'apprentissage automatique, y compris la super résolution d'images. Une étude a présenté un classificateur

hybride MLP-CNN pour la classification d'images à très haute résolution obtenues par télédétection [36].

De plus, le modèle MLP-Mixer est une architecture compétitive pour la vision par ordinateur qui se base exclusivement sur des MLP, offrant des performances comparables à des modèles de pointe en classification d'images [36].

Enfin, des modèles modernes basés sur des MLP, tels que le MLP-Mixer, le FNet et le gMLP, ont été implémentés pour la classification d'images, montrant des résultats prometteurs sur des ensembles de données comme CIFAR-100 (voir figure 1.18) [36].

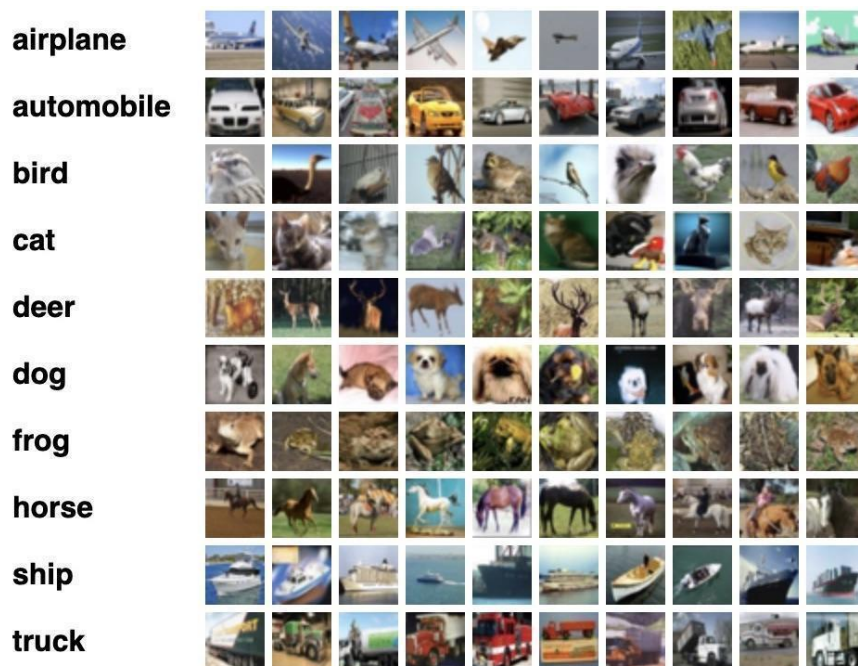


Figure 1 .18 : Application des MLP dans la Super Résolution « CIFAR-100 »

1.7.5 Apprentissage profond

L'apprentissage profond ou « Deep learning » est une branche de l'intelligence artificielle dérivée de l'apprentissage automatique ou « machine learning ». Contrairement à la programmation qui consiste à exécuter des règles préétablies, ce type d'apprentissage permet aux machines d'apprendre de manière autonome. En fait, les machines sont capables de s'entraîner sur des données, de reconnaître des modèles et de prendre des décisions basées sur ces données sans intervention humaine.

Apprentissage profond ou réseaux de neurones convolutifs, également appelés CNN (Convolutional Neural Networks) ou ConvNet constituent une classe de réseaux de neurones

profonds qui ont été élaborés à l'origine pour la reconnaissance d'images. Ce sont des Modèles de programmation puissants qui permettent de résoudre des problèmes de vision informatique tels que : la classification d'images, la détection d'objets, la segmentation sémantique, la reconnaissance faciale, etc. Parmi les plus anciens CNN nous pouvons citer Le-Net introduit par Le-Cun en 1998 (voir figure 1.19) [37].

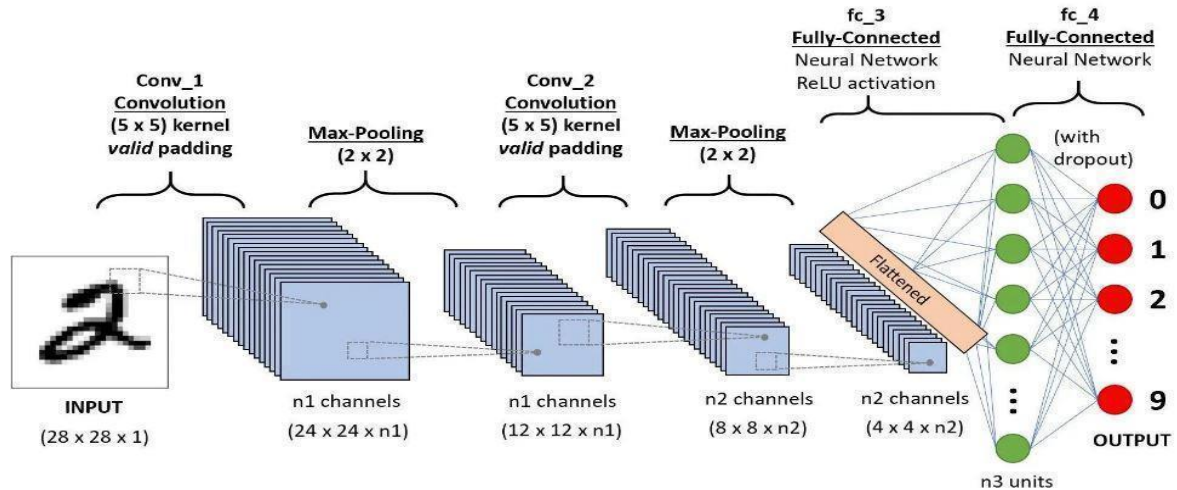


Figure 1. 19 : Schéma d'un Réseau de neurones à convolution (CNN)

1.7.6 Couches clés d'un CNN

Dans les prochaines sections, nous examinerons plus en détail les deux parties précédentes. En général, un réseau neuronal convolutif est constitué de plusieurs couches, chacune ayant sa propre fonctionnalité. Les couches clés d'un CNN sont les suivantes :

a. . Couches de convolution : C'est la première couche d'un CNN qui joue un rôle principal pour son fonctionnement, cette couche utilise un ensemble de filtres pour extraire la caractéristique (features) de l'image d'entrée (input image) comme l'indique la figure 1.20. Chaque filtre est une petite matrice de poids qui parcourt l'image d'entrée, en effectuant un produit de convolution entre les entrées et ces filtres et au même temps une carte d'activation est produite et elles sont modulées par une fonction d'activation, généralement une unité linéaire rectifiée (communément appelée Relu). La taille et le nombre de ces filtres sont définis à priori par la couche de convolution. Elles fonctionnent indépendamment des différentes profondeurs du réseau et ne requièrent pas de poids à entraîner .

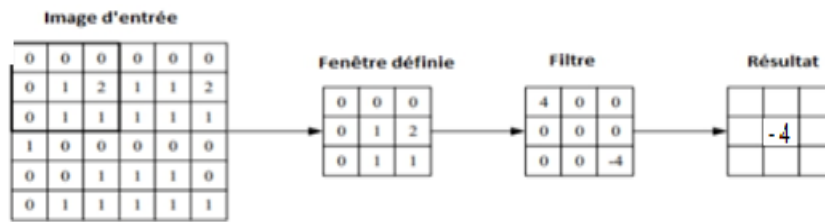


Figure 1.20 : Représentation d'une couche convolutive [17]

b. . Couches d'activation : Relu (Rectified linear Unit) Introduit de la non-linéarité dans le réseau, ce qui permet de modéliser des relations plus complexes comme l'indique la figure 1.21 [39-40].

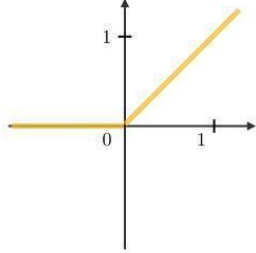
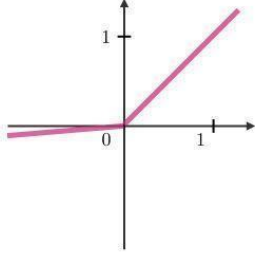
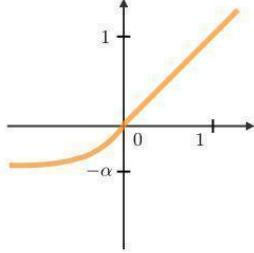
ReLU	Leaky ReLU	ELU
$g(z) = \max(0, z)$	$g(z) = \max(\epsilon z, z)$ with $\epsilon \ll 1$	$g(z) = \max(\alpha(e^z - 1), z)$ with $\alpha \ll 1$
		
<ul style="list-style-type: none"> Complexités non-linéaires interprétables d'un point de vue biologique 	<ul style="list-style-type: none"> Répond au problème de <i>dying ReLU</i> 	<ul style="list-style-type: none"> Dérivable partout

Figure 1.21 : La couche d'activation ReLU (Rectified Linear Unit) [39].

Autres : Sigmoid, Tanh, mais ReLU est la plus couramment utilisée dans les CNN en raison de son efficacité comme l'indique la figure 1.22.

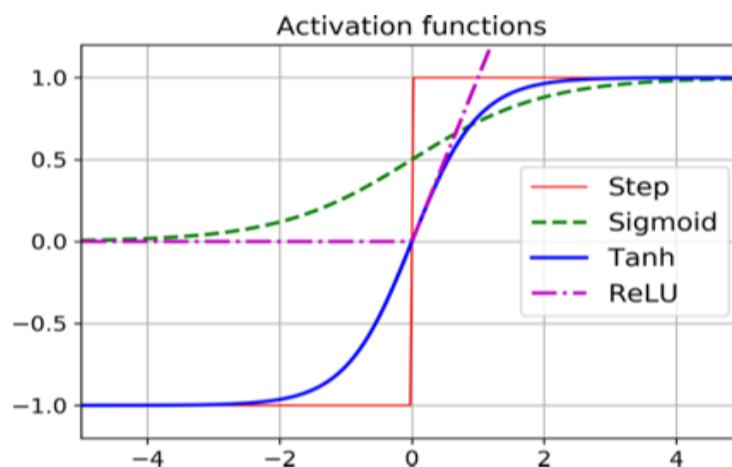
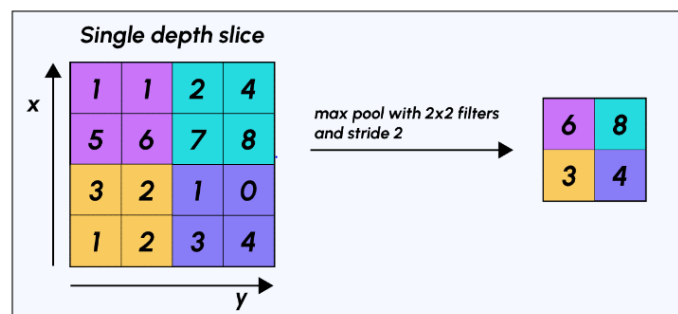


Figure 1.22 : La couche d'activation Sigmoid, Tanh et ReLU [39].

c. Couches de regroupements (pooling) : Cette couche a pour objectif de diminuer la taille de la carte de caractéristique produite. Le regroupement maximal est la technique la plus utilisée pour diminuer la taille d'une image (voir figure 1.23). Elle consiste à réduire une fenêtre donnée à une valeur unique en sélectionnant la plus grande valeur parmi les éléments de la fenêtre.

Le CNN est une des différentes formes de perceptron multicouche. Il peut avoir plus d'une couche de convolution, ce qui rend le réseau très profond avec moins de paramètres. Le CNN est très efficace pour la reconnaissance d'images et l'identification de différents modèles d'images [40].



Processus de Max-Pooling

Figure 1.23 : Processus de max-pooling [40].

d. Flattenning : c'est un processus de transformation d'un tableau multidimensionnel (comme une image) en un tableau unidimensionnel. C'est une étape cruciale souvent utilisée lors de la transition des couches convolutionnelles vers les couches entièrement connectées au sein d'un réseau de neurones.

e. Couches complètement connectées (fully connected layers) : Les couches entièrement connectées (fully connected layers) : c'est la dernière couche du CNN. Elle utilise les caractéristiques extraites des couches précédentes pour effectuer une classification (voir figure 1.24). Les neurones de ces couches sont tous reliés aux neurones des cartes d'activation précédentes [35-38].

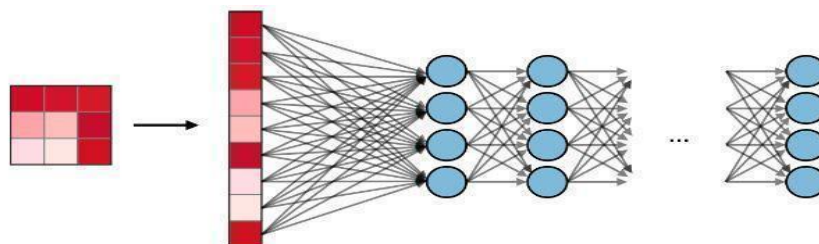


Figure 1.24: Les couches entièrement connectées [39].

Afin d'améliorer les performances d'un CNN, il existe d'autres types de couches qui peuvent être utilisées. Parmi celles-ci, on peut citer :

f. Couches de chute « Dropout »: Le Dropout est l'un des types de couches utilisés pour éviter le sur apprentissage (en anglais ; overfitting) du réseau. Cette technique a été introduite par Srivastava et al [22]. Le principe de cette technique est de désactiver aléatoirement un certain pourcentage de neurones lors de l'entraînement du réseau. Cela permet d'éviter les dépendances excessives entre les neurones et force le réseau à apprendre des caractéristiques plus robustes et plus générales. En désactivant les neurones de façon aléatoire, le stalling agit comme une forme de régularisation et réduit le risque de surentraînement en rendant le réseau plus résistant aux petits changements dans les données d'entraînement (Voir figure 1.25).

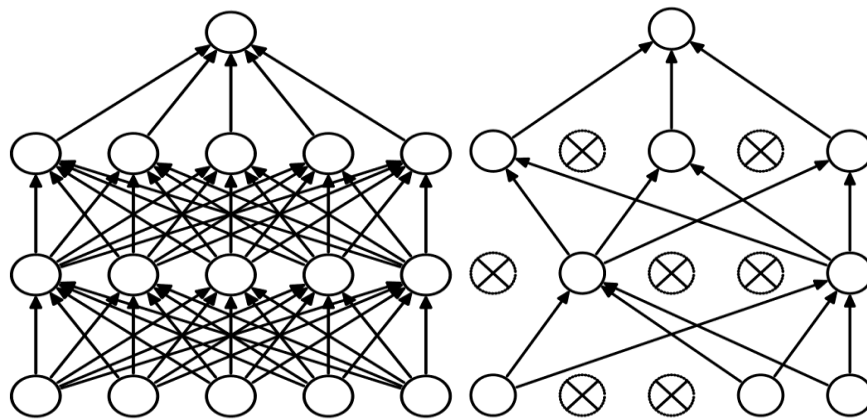


Figure 1.25: Modèle de réseau neuronal par Dropout [22].

1.7.7 Études de cas et exemples

a . Reconnaissance d'Images avec AlexNet

AlexNet, un CNN profond développé par Alex Krizhevsky, Ilya Sutskever et Geoffrey Hinton, a remporté la compétition ImageNet Large Scale Visual Recognition Challenge (ILSVRC) en 2012, réduisant considérablement le taux d'erreur par rapport aux approches précédentes (voir figure 1.26). L'utilisation de plusieurs couches convolutionnelles, de ReLU comme fonction d'activation, de dropout pour la régularisation et de GPU pour l'accélération du calcul a permis à AlexNet d'atteindre une précision inégalée à l'époque, ouvrant la voie à des architectures plus profondes et plus complexes [41-42].

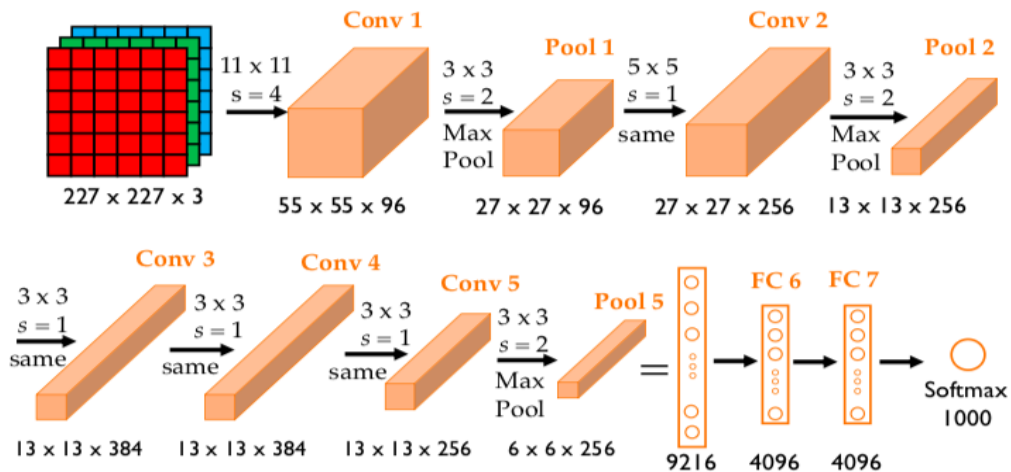


Figure 1.26 : AlexNet architecture

b . Classification d'Images avec VGGNet

VGGNet, développé par l'Université d'Oxford, utilise des couches convolutionnelles très profondes (jusqu'à 19 couches) avec des filtres de petite taille (3×3). VGGNet a amélioré la précision de la classification des images en approfondissant le réseau, ce qui permet une meilleure capture des caractéristiques complexes des images. Il a été largement adopté pour diverses tâches de vision par ordinateur (voir figure 1.27) [43].

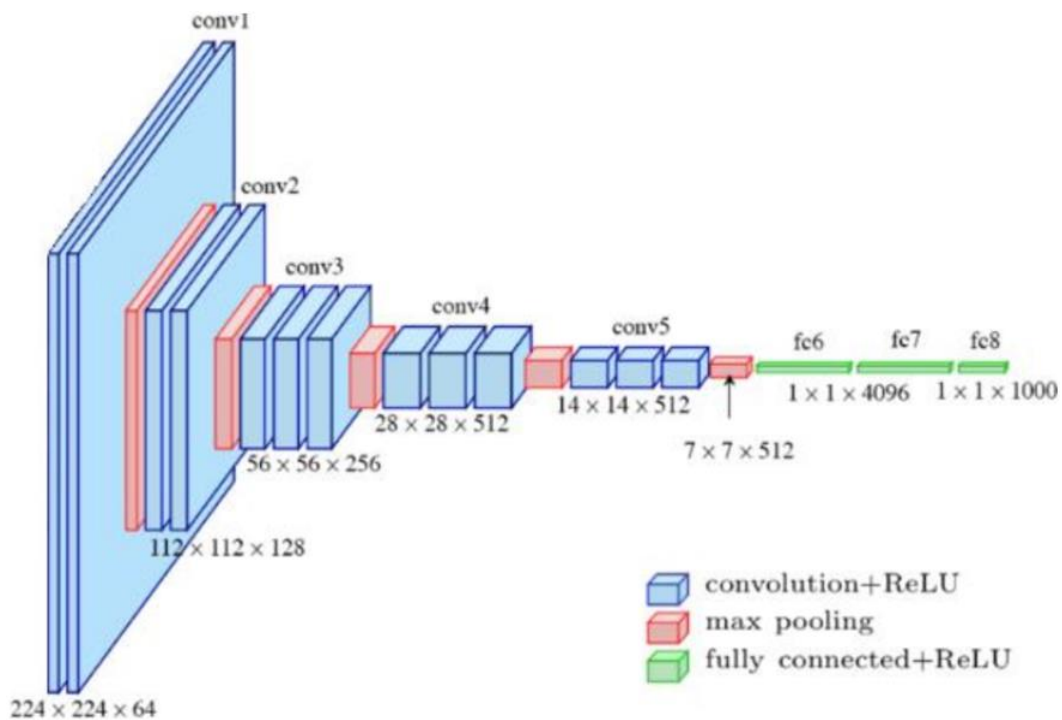


Figure 1.27 : VGGNet architecture

c . Détection d'objets avec YOLO (You Only Look Once)

YOLO, développé par Joseph Redmon et ses collègues, est une approche CNN qui traite la détection d'objets comme un problème de régression unique, prédisant directement les bounding boxes et les probabilités de classe à partir d'une seule évaluation du réseau (voir figure 1.28). YOLO est capable de traiter des images à une vitesse incroyablement rapide tout en maintenant une précision élevée, ce qui le rend idéal pour les applications en temps réel [44].

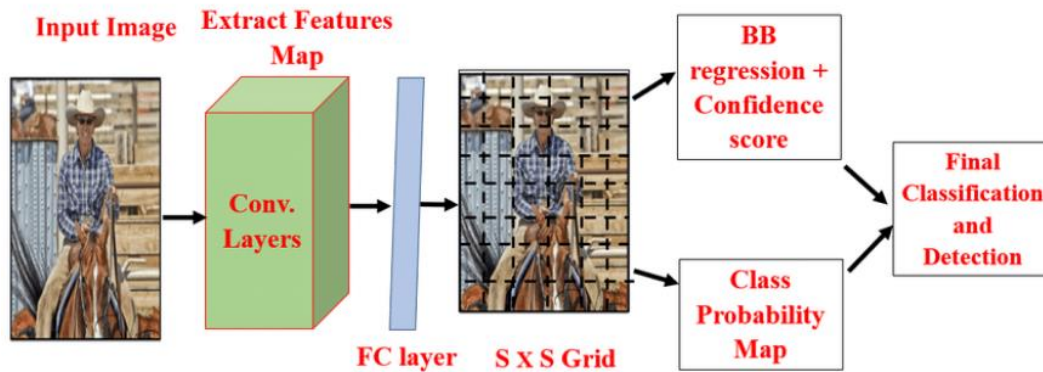


Figure 1.28 : YOLO architecture

1.8 Réseaux antagonistes génératifs

Les réseaux antagonistes génératifs apportent une valeur ajoutée significative à la super-résolution en générant des détails réalistes et en améliorant la qualité visuelle des images haute résolution. L'intégration de GAN dans des systèmes de super-résolution représente une avancée majeure dans le domaine de l'amélioration de la qualité des images [44].

1.8.1 Présentation de la structure du gan

Les réseaux antagonistes génératifs (GAN) sont constitués de deux réseaux de neurones ; un générateur et un discriminateur (voir figure 1.29). Le générateur crée des images synthétiques à partir de bruit aléatoire, tandis que le discriminateur essaie de distinguer les images synthétiques des images réelles. Les deux réseaux s'entraînent de manière compétitive, ce qui améliore progressivement la qualité des images générées [44].

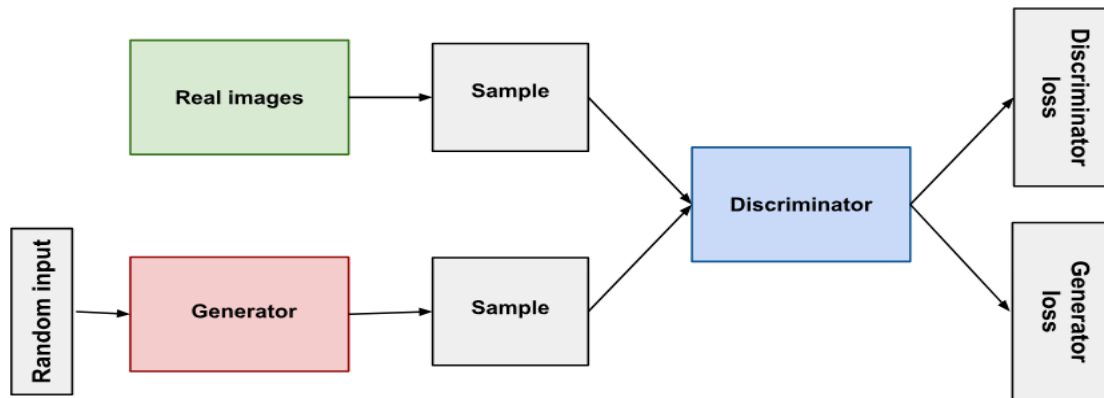
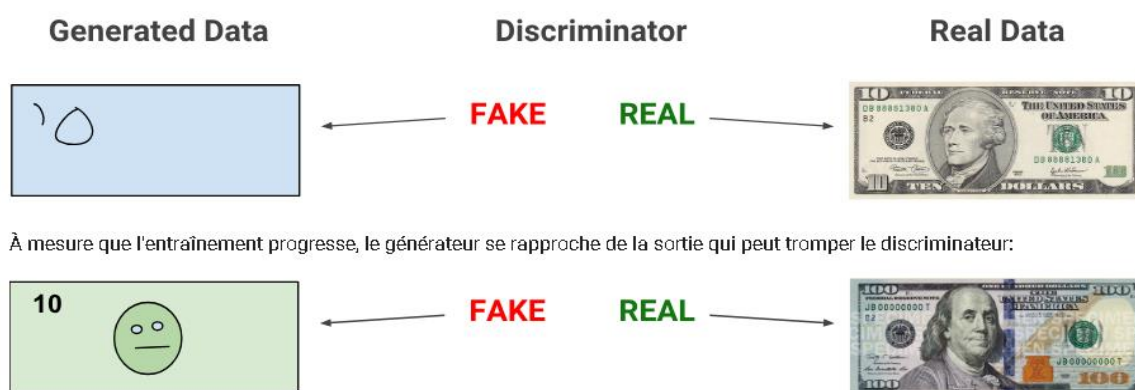


Figure 1.29 : Schéma des réseaux antagonistes génératifs (GAN)

Un réseau antagoniste génératif (GAN) se compose de deux parties:

- Le **générateur** apprend à générer des données plausibles. Les instances générées deviennent des exemples d'entraînement négatif pour le discriminateur [44].
- Le **discriminateur** apprend à distinguer les données fictives des données réelles. Le discriminateur pénalise le générateur pour avoir produit des résultats impossibles à comprendre [44].

Lorsque l'entraînement commence, le générateur génère des données manifestement factices, et le discriminateur apprend rapidement à dire qu'il s'agit de fausses données (voir figure 1.30) [45].



À mesure que l'entraînement progresse, le générateur se rapproche de la sortie qui peut tromper le discriminateur:

Figure 1.30 : Comportement du réseau avant convergence.

Enfin, si l'entraînement du générateur fonctionne bien, le discriminateur s'aggrave pour faire la différence entre le vrai et le faux. Il commence à classer les fausses données comme étant réelles et sa précision diminue (voir figure 1.31) [45].



Figure 1.31 : Résultat après convergence.

1.8.2 Super résolution avec generative adversarial network

Le GAN est utilisé dans la super résolution selon la figure 1.39. Le système est basé sur les deux points suivants :

- **Formation adversarielle:** SRGAN combine la super résolution avec l'apprentissage génératif adversarial (GAN), où un générateur tente de produire des images haute résolution indiscernables des images réelles, tandis qu'un discriminateur tente de les distinguer (voir figure 1.32) [46].
- **Loss function perceptuelle:** Il utilise une fonction de perte perceptuelle qui incorpore à la fois la différence de contenu et de style entre les images générées et les images réelles [46].

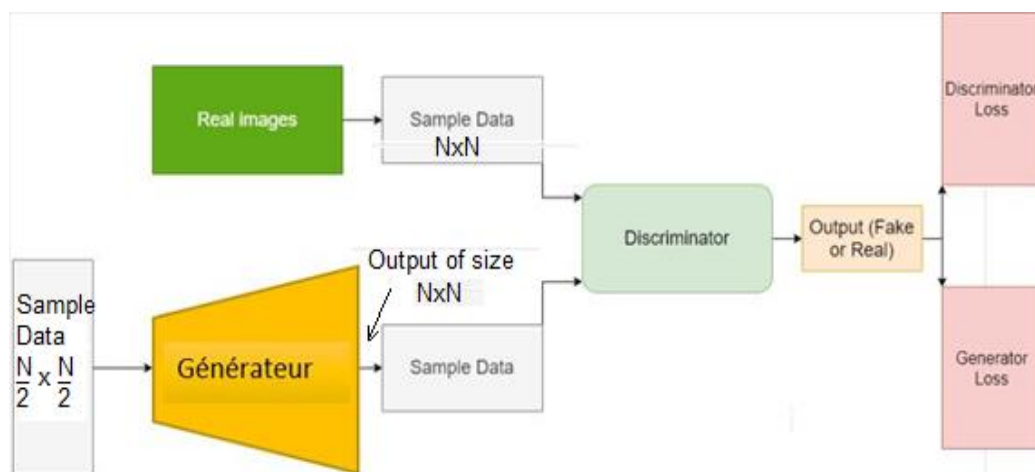


Figure 1.32: Architecture du SRGAN.

a . **Modèles basés sur les ondelettes pour la super résolution :** Les modèles basés sur les ondelettes utilisent la décomposition en ondelettes pour séparer les composantes de haute et basse fréquence d'une image. Les détails fins (hautes fréquences) et les structures globales

(basses fréquences) sont traités séparément pour améliorer la qualité de l'image reconstruite. Des algorithmes tels que la décomposition en ondelettes discrètes (DWT) et la transformée en ondelettes à paquet (WP) sont couramment utilisés [17].

b . Approches hybrides combinant CNN et transformée en ondelettes : Les approches hybrides combinent les réseaux de neurones convolutifs (CNN) avec la transformée en ondelettes pour tirer parti des avantages des deux techniques. Les CNN sont utilisés pour apprendre des caractéristiques complexes et non linéaires, tandis que la transformée en ondelette aide à préserver les détails à différentes échelles. Ces modèles hybrides peuvent offrir des performances supérieures en termes de qualité d'image et de robustesse [17].

c . Autres méthodes de fusion d'ondelettes pour la super résolution : D'autres méthodes de fusion d'ondelettes incluent l'utilisation de différentes stratégies de fusion pour combiner les informations de plusieurs échelles. Par exemple, la fusion de coefficients d'ondelettes issus de différentes images sources peut améliorer la qualité de la super résolution en intégrant des détails complémentaires [17].

d . Super résolution d'Images avec SRCNN : SRCNN (Super Resolution Convolutional Neural Network), développé par Chao Dong et ses collègues, utilise des CNN pour apprendre directement le mapping des images basse résolution vers haute résolution (voir figure 1.33). SRCNN a démontré une performance supérieure par rapport aux méthodes traditionnelles de super-résolution, offrant une meilleure qualité d'image avec des détails plus fins et des artefacts réduits [47].

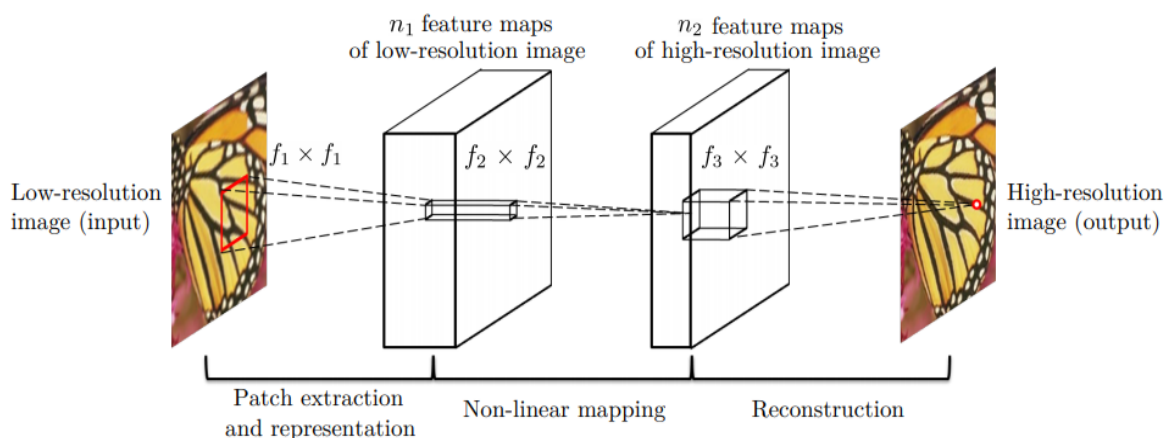


Figure 1.33 : SRCNN architecture

1.9 Conclusion

La super-résolution d'images par CNN est un domaine en pleine évolution avec un fort potentiel pour améliorer la qualité des images dans de nombreux domaines d'application. Les recherches continues et les progrès technologiques permettront d'obtenir des résolutions encore plus élevées, des détails plus fins et des images plus réalistes, ouvrant la voie à de nouvelles applications innovantes

Chapitre 2 :
Réseaux antagonistes
génératifs

2.1 Introduction

Ces dernières années, la demande croissante d'images haute résolution dans des domaines tels que l'imagerie médicale, l'imagerie satellitaire et les applications multimédia a conduit au développement de techniques avancées de super-résolution d'image (SR). Les méthodes traditionnelles, basées sur l'interpolation, sont souvent insuffisantes pour récupérer les détails fins et les textures. Pour cela, nous proposons une méthode basée sur l'apprentissage profond, en particulier les réseaux antagonistes génératifs (GAN : Generative Adversarial Neural network; en anglais). Cette méthode repose sur l'architecture U-Net [48], qui a montré de bons résultats dans plusieurs domaines, et permet de générer des images de haute qualité et réalistes. De plus, nous utilisons la transformée en ondelettes (WT : Wavelet Transform en anglais) pour l'analyse multi-résolution, ce qui facilite l'extraction et la reconstruction améliorées des caractéristiques. (Mettre un schéma synoptique simplifié)

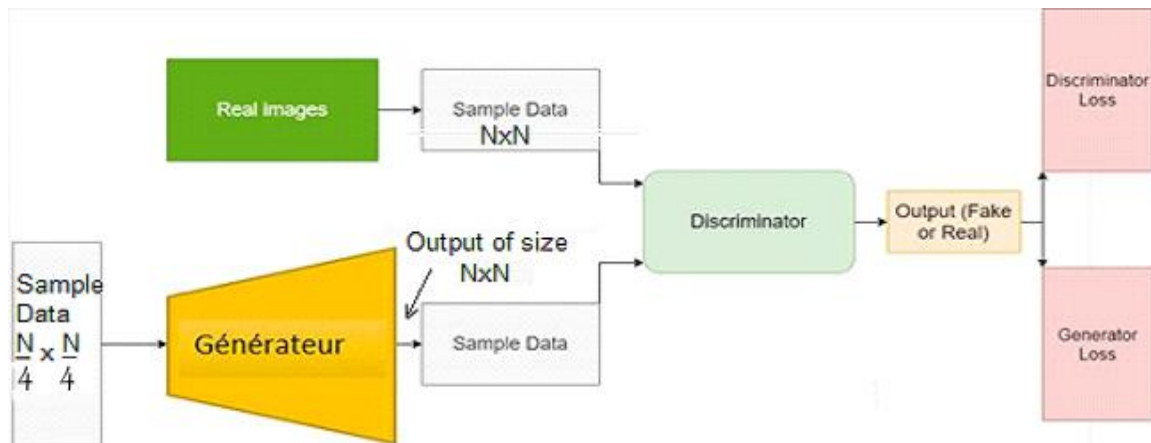


Figure 2.1: schéma synoptique simplifié du SRGAN.

2.2 Méthodologie

2.2.1 Générateur

Nous proposons un réseau à deux étapes, conçu pour générer des images de haute qualité et restaurer les images générées qui sont dégradées par divers facteurs tels que le flou. Les générateurs des deux étapes sont basés sur l'architecture M-Net [49], illustrée dans la Figure 2.1, qui elle-même repose sur l'architecture U-Net, afin d'apprendre et de reconstruire efficacement les détails fins et les textures.

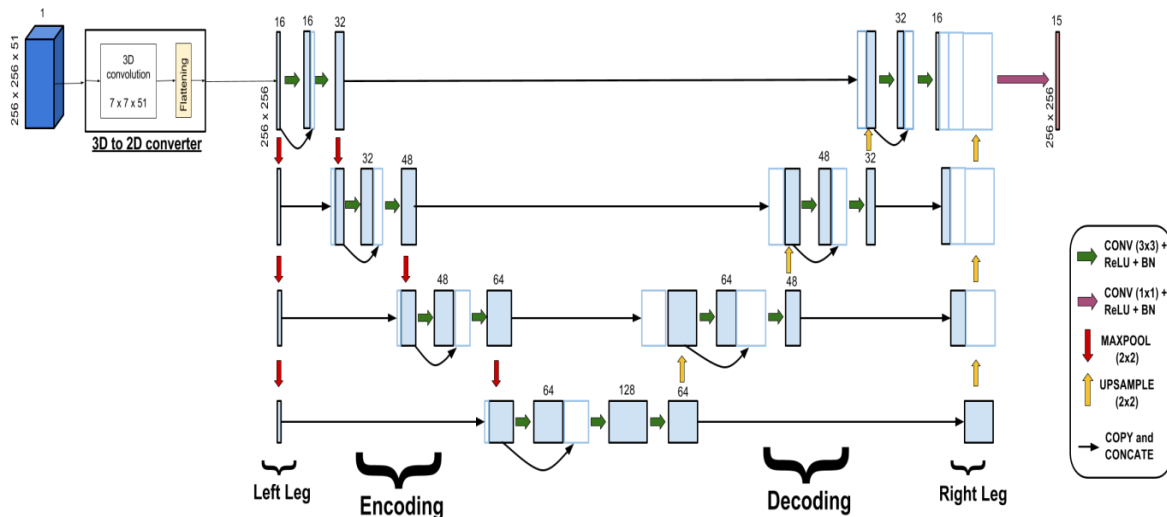


Figure 2.2: Architecture M-Net

Dans la première étape, nous nous concentrons sur la génération d'images haute résolution à partir d'entrées basse résolution. La deuxième étape de notre réseau est conçue pour restaurer l'image haute résolution générée à la première étape, en appliquant un traitement supplémentaire pour éliminer les dégradations et améliorer la qualité globale de l'image. Pour le processus de restauration, au lieu d'utiliser la concaténation pour fusionner les caractéristiques à différentes résolutions, nous adoptons un bloc de fusion des caractéristiques (SKFF : Selective Kernel Feature Fusion) pour combiner efficacement ces caractéristiques [50-51]. Les architectures de la première et la deuxième étape sont illustrées dans les Figures 2.2 et 2.3 respectivement.

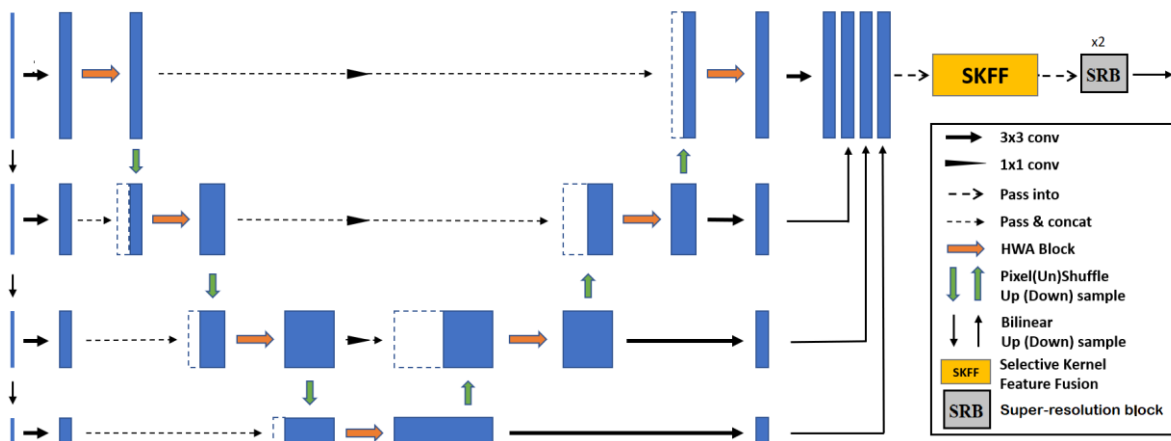


Figure 2.3: Architecture du stage 1 (Agrandissement de l'image)

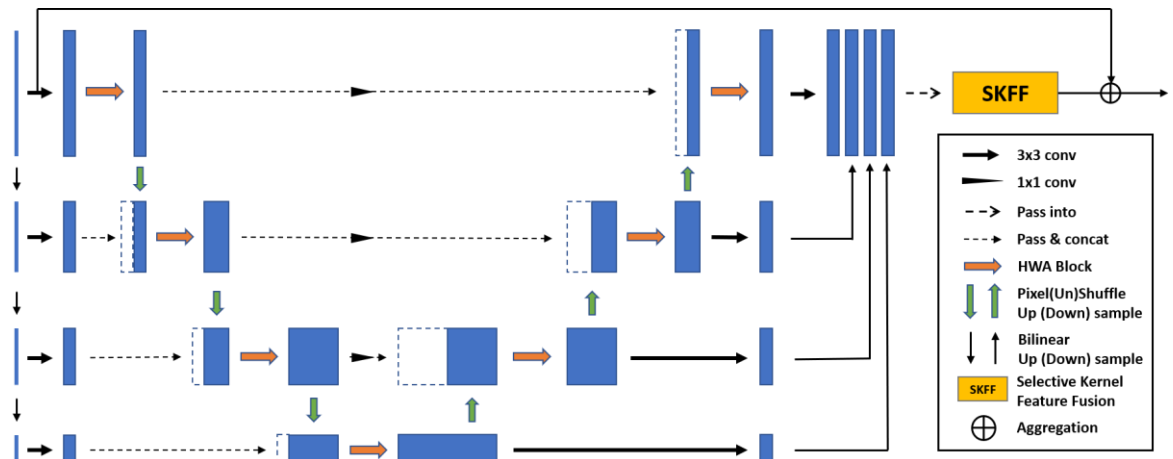


Figure 2.4: Architecture du stage 2 (Restauration de l'image)

a. Principales caractéristiques de l'architecture M-Net

L'architecture U-Net se compose d'une structure encodeur-décodeur avec des connexions de saut symétriques qui relient les couches correspondantes dans l'encodeur et le décodeur, ce qui est crucial pour reconstruire avec précision des images haute résolution [48]. L'encodeur compresse l'image d'entrée en une représentation latente, capturant les caractéristiques essentielles, tandis que le décodeur reconstruit l'image haute résolution à partir de cette représentation.

L'architecture M-Net est un type d'architecture de réseau neuronal conçu principalement pour la segmentation d'images médicales. Il est une extension de l'architecture U-Net, largement utilisée à des fins similaires. Il introduit plusieurs modifications pour améliorer les performances. M-Net comprend principalement quatre voies de filtres 2D : deux voies principales d'encodage et de décodage et deux voies latérales, qui fournissent une supervision profonde à l'architecture. Chaque voie comprend quatre étapes.

❖ **Encodeur** : L'encodeur se compose de plusieurs couches convolutives suivies de couches de pooling, similaire aux réseaux neuronaux convolutifs (CNN) standards. Il extrait des caractéristiques de haut niveau des images d'entrée. Dans le chemin d'encodage, chaque étape comporte une cascade de deux blocs convolutifs suivis d'un max-pooling de 2x2 pour réduire la taille d'entrée de moitié et permettre au réseau d'apprendre des informations contextuelles. Chaque bloc contient des filtres de convolution 2D de taille 3x3 suivis d'une fonction d'activation ReLU et d'une normalisation par lots. Dans la cascade de filtres de convolution, une connexion de saut est introduite pour permettre au réseau d'apprendre de meilleures caractéristiques.

- ❖ **Décodeur** : La partie décodeur se compose de couches de sur échantillonnage (upsampling) suivies de couches convolutives, visant à reconstruire les caractéristiques à partir des caractéristiques de haut niveau. Les sorties de différentes échelles sont combinées pour former une des caractéristiques unifiées. La couche de décodage est identique aux couches d'encodage, à une exception près : le max-pooling est remplacé par une couche de sur échantillonnage pour doubler la taille d'entrée et récupérer une image de sortie de la taille d'origine.
- ❖ **Connexions de saut (Skip connections)** : M-Net est comme U-Net, elle utilise des connexions de saut entre les couches correspondantes de l'encodeur et du décodeur. Cela aide à préserver les informations spatiales qui pourraient se perdre dans les couches plus profondes et à assurer que le réseau dispose des informations suffisantes pour dériver un étiquetage précis des détails fins de l'image sans post-traitement.
- ❖ **Pieds (Legs) gauche et droit** : Le Pied gauche opère sur une taille d'entrée d'image (HxW) avec quatre couches de max-pooling de taille 2x2, et les sorties sont données en entrée aux couches d'encodage correspondantes. Le Pied droite sur-échantillonne (upsamples) la sortie de chacune des couches de décodage à la taille d'origine (HxW). Enfin, la couche de décodage et la sortie de la leg droite sont traitées par une couche de convolution 1x1 suivie d'une fonction d'activation ReLU et d'une normalisation par lots.

b. Architecture M-Net améliorée

L'architecture M-Net rencontre deux problèmes majeurs. Premièrement, pour l'encodeur, elle utilise le max-pooling comme méthode de sous-échantillonnage à la fois dans les chemins de caractéristiques gatepost et U-Net. Étant donné que les tâches de restauration sont des tâches de vision pixel par pixel, il est inapproprié d'utiliser le sous-échantillonnage qui entraînera une perte sérieuse d'informations spatiales. Deuxièmement, pour le décodeur, en raison de l'ajout de chemins de caractéristiques gatepost supplémentaires, le processus de reconstruction (décodeur) doit fusionner de nombreuses caractéristiques, ce qui augmente la charge pour l'ensemble du réseau [51].

Pour nous concentrer sur ces deux problèmes, nous avons proposé une architecture avec deux améliorations. Premièrement, nous utilisons le sous-échantillonnage par pixel un-shuffle dans le chemin U-Net et le sous-échantillonnage bilinéaire pour le chemin gatepost (Pieds), ce qui rend les caractéristiques en cascade plus diversifiées. Deuxièmement, nous utilisons une

méthode de fusion de caractéristiques différente, SKFF [50], pour résumer l'information dans le décodeur (processus de reconstruction). De plus, dans la première étape de notre réseau, nous utilisons un bloc de double super-résolution pour augmenter la taille de l'image par un facteur de 4.

➤ **Super résolution block (SRB)**

Comme illustré dans la figure 2.2, ce bloc est utilisé pour augmenter la taille de l'image par un facteur de deux afin de réduire l'introduction d'artefacts et la dégradation de la qualité de l'image. Nous avons utilisé deux blocs pour augmenter la taille de l'image par un facteur de quatre. Ce bloc, illustré dans la Figure 2.4, se compose de filtres de convolution 2D de taille 3x3 suivis d'un pixel shuffle pour agrandir l'image avec un facteur de deux, et d'une fonction d'activation ReLU.

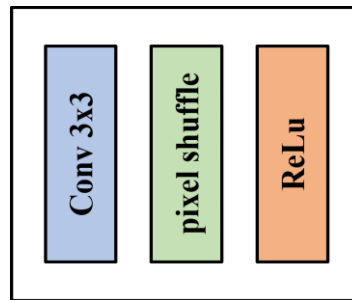


Figure 2.5: Super resolution block (SRB)

➤ **Bloc de fusion des caractéristiques (SKFF)**

Comme illustré dans la Figure 2.5, le module de fusion de caractéristiques à noyau sélectif (SKFF) ajuste dynamiquement les champs récepteurs en utilisant les procédures FUSE et SELECT. Les informations de différents flux de résolution sont combinées avec l'opérateur FUSE pour fournir des descripteurs de caractéristiques globales [52]. Ensuite, l'opérateur SELECT utilise ces descripteurs pour recalibrer les cartes de caractéristiques des différents flux, qui sont ensuite agrégées.

-Fusion (Fuse) : Le module SKFF reçoit des entrées de quatre flux de convolution parallèles portant des informations à différentes échelles. Nous combinons d'abord ces caractéristiques multi-échelles en utilisant une somme par élément comme suit :

$$L = L1 + L2 + L3 + L4 \quad (2.1)$$

Ensuite, nous appliquons un pooling moyen global (GAP) sur les dimensions spatiales de $L \in \mathbb{R}^{H \times W \times C}$ pour calculer les statistiques par canal $s \in \mathbb{R}^{1 \times 1 \times C}$. Ensuite, nous appliquons une

couche de convolution pour réduire le nombre de canaux afin de générer une représentation compacte des caractéristiques $z \in \mathbb{R}^{1 \times 1 \times r}$, où $r = C/8$. Enfin, le vecteur de caractéristiques z passe à travers quatre couches de convolution parallèles (une pour chaque flux de résolution) pour générer 4 descripteurs de caractéristiques v_1, v_2, v_3 et v_4 , chacun ayant des dimensions $1 \times 1 \times C$ [50].

-Sélectionner (Select) : Cet opérateur applique la fonction softmax à v_1, v_2, v_3 et v_4 , ce qui produit des activations d'attention s_1, s_2, s_3 et s_4 que nous utilisons pour recalibrer de manière adaptative les cartes de caractéristiques multi-échelles L_1, L_2, L_3 et L_4 respectivement [50].

Le processus global de recalibrage et d'agrégation des caractéristiques est défini comme suit :

$$U = s_1 \times L_1 + s_2 \times L_2 + s_3 \times L_3 + s_4 \times L_4 \tag{2.2}$$

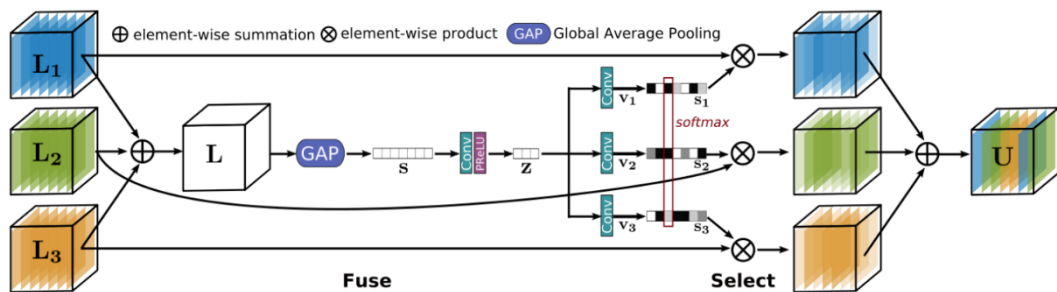


Figure 2.6: Selective kernel feature fusion (SKFF)

➤ **Block d'attention à demi ondelette (HWAB : Half Wavelet Attention Block)**

Chaque niveau de l'encodeur et du décodeur de M-Net est remplacé par HWAB [51], comme illustré dans les figures 2.2 et 2.3. Ce bloc incorpore l'extraction de caractéristiques à la fois dans le domaine spatial et fréquentiel. Il s'agit d'un module spécialisé conçu pour améliorer l'extraction et la représentation des caractéristiques dans les réseaux neuronaux en utilisant des transformations par ondelettes et des mécanismes d'attention. HWAB commence par recevoir des entrées à partir de plusieurs flux de résolution. Ces flux contiennent des caractéristiques à différentes échelles, capturant des niveaux diversifiés de détails et de contexte. Comme illustré dans la Figure 2.6, chaque caractéristique d'entrée est divisée en deux parties égales : $f_{identity}$ et f_l le long du canal d'entrée.

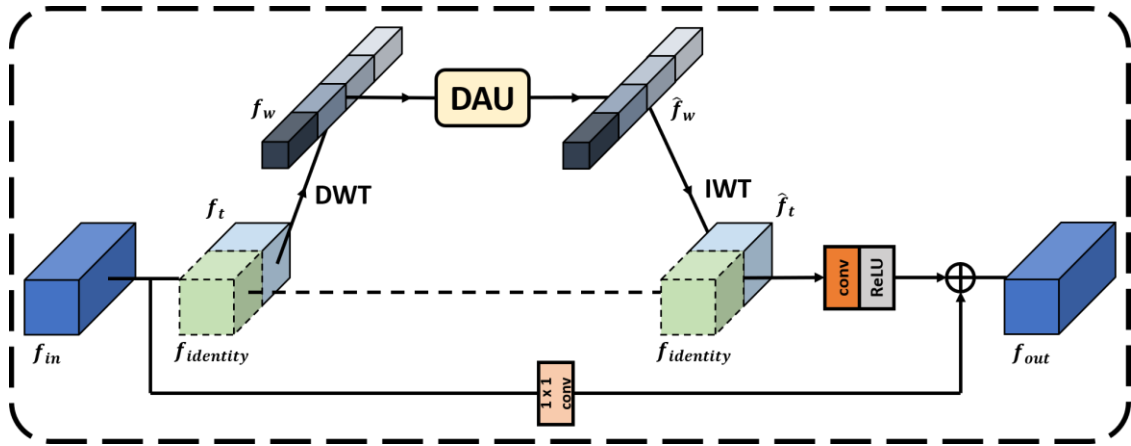


Figure 2.7: Half Wavelet Attention Block (HWAB)

Le but principal de diviser les caractéristiques d'entrée est de réduire la complexité de calcul tout en préservant l'information de contexte [51], où $f_{identity}$ conserve les caractéristiques normales du domaine, tandis que l'autre partie des caractéristiques f_t subit une transformation en ondelettes discrètes (DWT) pour obtenir les caractéristiques du domaine des ondelettes. La carte f_t passe par une DWT pour la décomposer en quatre sous-bandes de résolution inférieure. Cette étape capture les composantes hautes fréquences et basse fréquence des caractéristiques d'entrée, ce qui est crucial pour la représentation détaillée de l'image. Les sous-bandes sont alimentées dans l'Unité d'Attention Double (DAU), illustrée dans la figure 2.7, pour se concentrer sur les informations pertinentes des différents canaux de caractéristiques et sur l'information spatiale importante [50].

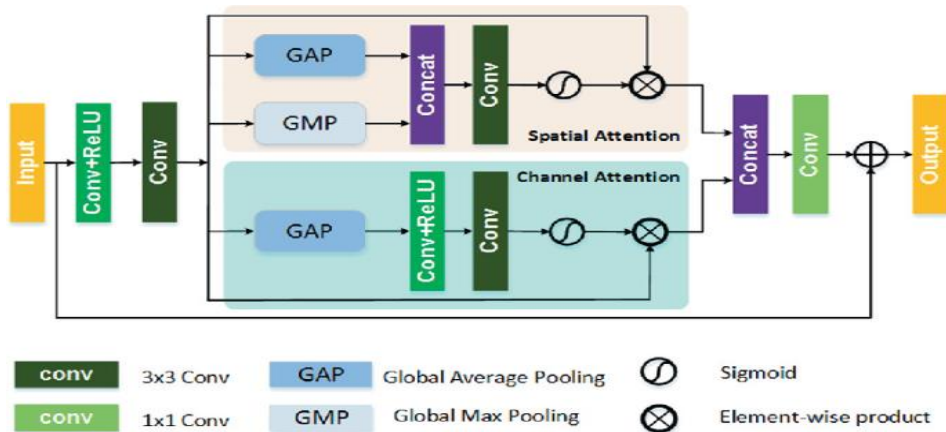


Figure 2.8: Dual attention unit (DAU)

Le block DAU utilise une attention sur les canaux et une attention spatiale pour extraire des caractéristiques, puis les fusionne ensemble pour obtenir une information contextuelle enrichie. Après le traitement par le DAU, les caractéristiques sont reconstruites à leur résolution

d'origine à l'aide de la Transformée en Ondelettes Discrète Inverse (IDWT). Ensuite, les caractéristiques de sortie et f_{identity} sont concaténées et passées à une couche de convolution 3×3 suivie d'une fonction d'activation ReLU. Ces caractéristiques sont ensuite ajoutées aux caractéristiques d'entrée qui ont passé par une convolution 1×1 pour obtenir des caractéristiques résiduelles, donnant ainsi les caractéristiques de sortie finales f_{out} .

2.2.2 Discriminateur

Le discriminateur dans un GAN joue un rôle crucial en agissant essentiellement comme un classifieur, chargé de distinguer les images réelles des images générées par le générateur [53]. Dans notre approche, nous avons adopté un discriminateur Markovien (PatchGAN) [54], qui est un réseau de neurones convolutif profond spécialement conçu pour cette tâche. Contrairement au modèle GAN standard qui classifie l'image entière comme réelle ou fausse, comme illustrée dans la figure 2.8, le PatchGAN opère en classifiant des patches individuels ($N \times N$) de l'image d'entrée comme réels ou faux [54].

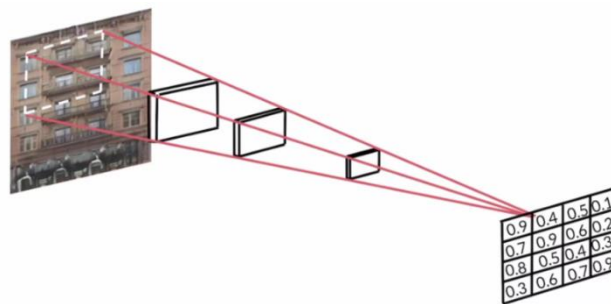


Figure 2.9: Discriminateur Markovien (PatchGAN)

Cette méthode permet d'imposer des contraintes plus strictes favorisant la préservation des détails nets à haute fréquence. De plus, le PatchGAN présente l'avantage d'avoir moins de paramètres et d'être plus rapide que les approches classifiant l'image entière [54]. Comme illustré dans la Figure 2.9, le discriminateur se compose de cinq blocs convolutifs. Chaque bloc est constitué de filtres de convolution 2D de taille 4×4 avec un pas de 2×2 , suivis d'une couche de normalisation par lots (batch normalization) et d'une fonction d'activation Leaky ReLU, à l'exception des premier et dernier bloc. Le premier bloc est composé de filtres de convolution 2D de taille 3×3 avec un pas de 2×2 , suivis d'une fonction d'activation Leaky ReLU. Le dernier bloc est composé de filtres de convolution 2D de taille 4×4 avec un pas de 1×1 , suivis d'une couche de normalisation par lots et d'une fonction d'activation Leaky ReLU. La sortie du discriminateur est une matrice 2D de taille $20 \times 20 \times 1$.

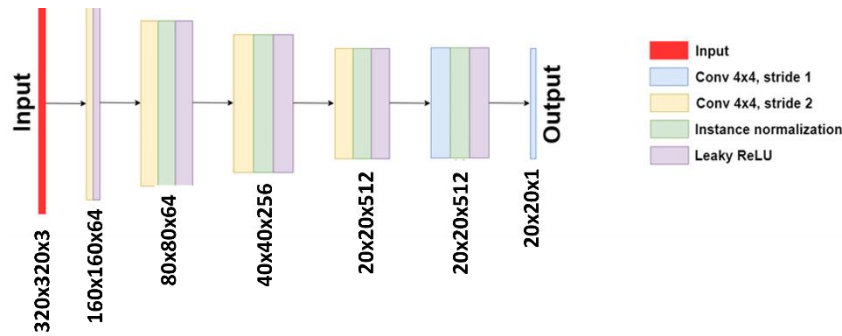


Figure 2.10: Architecture du discriminateur Markovien (PatchGAN)

2.2.3 Fonctions de perte (loss functions)

Nous utilisons certaines fonctions de perte pertinentes et appropriées pour notre approche. Nous rappelons ces fonctions de perte ci-dessous.

❖ **Perte du générateur et du discriminateur** : En plus de la structure améliorée du générateur, nous améliorons également le discriminateur basé sur le GAN moyen relativiste (RaGAN). Contrairement au discriminateur standard, qui estime la probabilité que les données réelles soient en moyenne plus réalistes que les données générées [55], les fonctions de perte pour le discriminateur D et le générateur G sont exprimées comme suit :

$$L_D^{RaGAN} = \mathbb{E}_{x_r \sim \mathbb{P}} [f_1 (C(x_r) - \mathbb{E}_{x_f \sim \mathbb{Q}} C(x_f))] + \mathbb{E}_{x_f \sim \mathbb{Q}} [f_2 (C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}} C(x_r))] \quad (2.3)$$

$$L_G^{RaGAN} = \mathbb{E}_{x_r \sim \mathbb{P}} [g_1 (C(x_r) - \mathbb{E}_{x_f \sim \mathbb{Q}} C(x_f))] + \mathbb{E}_{x_f \sim \mathbb{Q}} [g_2 (C(x_f) - \mathbb{E}_{x_r \sim \mathbb{P}} C(x_r))] \quad (2.4)$$

Où f_1, f_2, g_1 et g_2 sont des fonctions scalaire-à-scalaire. x_r, x_f sont respectivement les données réelles et générées.

❖ **Perte de Charbonnier** [51] : La perte de Charbonnier est une fonction de perte utilisée principalement dans les tâches de traitement d'image. Elle est définie comme suit :

$$I_{Char} = \sqrt{(x_r + x_f)^2 + \epsilon^2} \quad (2.5)$$

Où x_r, x_f sont respectivement les données réelles et générées respectivement. La constante ϵ est fixée à 10^{-3} .

❖ **Fonction de Perte Perceptuelle** [56]: Nous définissons une perte perceptuelle comme la distance euclidienne entre les représentations des caractéristiques VGG19 des données réelles et générées. Elle est définie comme suit :

$$I_{per} = \|\varphi(x_r) - \varphi(x_f)\|_2^2 \quad (2.6)$$

Où ϕ est les représentations des caractéristiques extraites par le VGG19.

2.3 Conclusion

Dans ce chapitre, nous avons présenté l'architecture de notre approche de super résolution d'image basée sur des réseaux antagonistes génératifs (GAN), en utilisant l'architecture M Net pour le générateur et en intégrant des techniques avancées comme la transformée en ondelette. En exploitant la structure améliorée du M Net avec des blocs spécifiques tels que le bloc fusion des caractéristiques (SKFF) et le bloc d'attention à demi ondelette (HWAB), notre modèle parvient à capturer et à restaurer des détails fins et des textures complexes. Nous avons également discuté de l'architecture utilisée pour le discriminateur, à savoir le PatchGAN, ainsi que des différentes fonctions de perte employées pour entraîner notre modèle, telles que la perte de Charbonnier et la perte perceptuelle. Ces fonctions de perte renforcent la performance de notre modèle en termes de fidélité et de réalisme des images reconstruites

Chapitre 3 :
Simulation de la
super-résolution

3.1 Introduction

Dans ce chapitre, nous abordons l'évaluation pratique de notre système de super-résolution d'images à travers des simulations détaillées et l'analyse des résultats obtenus. Nous présentons d'abord les outils et le Framework utilisés, ainsi que les bases de données exploitées pour l'entraînement et la validation de nos modèles.

L'évaluation de la qualité des images obtenues repose sur des métriques de qualité avec référence et sans référence. Les métriques avec référence, telles que le MSE, le PSNR et le SSIM, permettent une comparaison objective entre l'image reconstruite et l'image originale haute résolution. Les métriques sans référence, telles que NIQE, PIQE et BRISQUE, évaluent la qualité perçue de l'image sans nécessiter de comparaison directe avec une image de référence, apportant ainsi une perspective complémentaire sur les performances de notre système.

Nous effectuons également une comparaison entre les méthodes traditionnelles de super-résolution d'images et notre approche. Cette comparaison, accompagnée d'analyses pour illustrer les améliorations en termes de qualité visuelle et de mesures objectives, permet de mettre en lumière les gains de performance apportés par notre méthode, ainsi que les défis techniques et les limitations rencontrées par les méthodes classiques.

3.2 Outils utilisées

➤ **Langage de programmation Python** : Python est un langage de programmation puissant et polyvalent, largement utilisé pour diverses applications, notamment le traitement d'images, l'analyse de données, l'apprentissage automatique et le développement web. Il a été créé par Guido van Rossum, un informaticien néerlandais, qui a commencé à travailler sur Python à la fin des années 1980 en tant que successeur du langage de programmation ABC. La première version de Python, Python 0.9.0, est sortie en 1991 [57]. Python est caractérisé par plusieurs atouts majeurs [57] :

-**Portabilité** : Python est un langage portable, ce qui signifie qu'il peut s'exécuter sur une large gamme de plateformes sans nécessiter de modifications majeures du code.

-**Utilisation commerciale gratuite** : Python est un logiciel libre et open-source, ce qui signifie qu'il peut être utilisé et distribué gratuitement, y compris à des fins commerciales.

-**Syntaxe simple et lisible** : Python est réputé pour sa syntaxe simple et intuitive, ce qui le rend facile à apprendre et à utiliser, même pour les débutants. Il utilise des mots-clés anglais et une

structure de code indentée qui ressemble au langage naturel, rendant le code plus lisible et compréhensible par rapport à d'autres langages de programmation.

-Gestion des types de données évolués : Python offre des types de données dynamiques et évolutifs, tels que les listes et les dictionnaires, qui permettent de stocker et de manipuler des données de manière flexible. Cette capacité est essentielle pour le traitement d'images et l'analyse de données complexes.

-Multithreading (optionnel) : Python prend en charge le multi threading, ce qui permet d'exécuter plusieurs tâches simultanément au sein d'un même programme. Cette fonctionnalité est particulièrement utile pour les applications nécessitant une exécution parallèle, telles que les simulations et les traitements de données à grande échelle.

➤ **Framework PyTorch** : PyTorch est un Framework open-source puissant et dynamique pour l'apprentissage profond, qui a gagné en importance dans les domaines de l'intelligence artificielle et de l'apprentissage automatique. Développé par le laboratoire de recherche en intelligence artificielle de Facebook (FAIR), PyTorch offre une plateforme flexible et intuitive pour la construction et l'entraînement des réseaux neuronaux. Il est largement acclamé pour son graphe computationnel dynamique, qui permet des ajustements faciles des modèles en cours d'exécution, ce qui en fait un excellent choix tant pour les expérimentations de recherche que pour les déploiements en production [58].

3.3 Datasets

Dans notre projet, nous avons utilisé deux bases de données (datasets) pour entraîner notre modèle. Pour la première étape, nous avons utilisé le dataset Flickr2K pour la super-résolution, et le dataset DIV2K pour la deuxième étape, dédiée à la restauration et à la déconvolution.

❖ **Flickr2K** [59]: le dataset Flickr2K est une collection de haute qualité d'images haute résolution, utilisée principalement dans le domaine de la super-résolution d'images et d'autres tâches de vision par ordinateur. Ce dataset est largement utilisé pour entraîner et évaluer des modèles d'apprentissage profond pour des applications telles que la restauration d'images et la génération de contenu haute résolution. Ce dataset contient 2650 paires d'images haute résolution avec leurs versions basse résolution, couvrant différentes scènes et différents types de dégradations.

❖ **DIV2K**: le dataset DIV2K (Diverse 2K Resolution) est une collection d'images de haute résolution utilisée principalement pour les tâches de super-résolution d'images [60]. Il a été

créé dans le cadre des compétitions de la communauté de vision par ordinateur, notamment le Challenge on Image Super-Resolution (NTIRE) organisé à la conférence CVPR. Le dataset contient 1000 paires d'images haute résolution avec leurs versions basse résolution, parmi lesquelles 800 sont destinées à l'entraînement, 100 pour la validation, et 100 pour les tests. Les images couvrent une large gamme de scènes et de sujets, incluant des paysages naturels, des bâtiments, des objets, et des scènes urbaines.

3.4 Détails de l'implémentation

Nous avons implémenté nos architectures en utilisant le framework PyTorch sur NVIDIA Quadro RTX 6000. Nos modèles sont entraînés à partir de zéro pendant 100 époques avec un taux d'apprentissage (learning rate) de $1e-4$. Nous utilisons l'optimiseur Adam (Adam optimizer) et la taille de lot (batch size) est fixée à 10. Pendant la phase d'entraînement, nous découpons aléatoirement les images à une taille de 80×80 pour le 1^{er} stage et de 320×320 pour le 2^{ème} stage.

3.5 Métriques d'évaluation de la qualité de l'image

Les métriques de qualité d'image peuvent être divisées en deux grandes catégories : les métriques Full-Reference (FR) et les métriques No-Reference (NR). Chacune de ces catégories mesure différents aspects de la qualité d'une image, en fonction de la disponibilité de l'image de référence.

3.5.1 Métriques de qualité avec référence

- **Erreur quadratique moyenne (MSE)** : L'erreur quadratique moyenne mesure la moyenne des différences quadratiques entre les valeurs réelles et idéales des pixels [61]. Cette métrique est simple à calculer mais peut ne pas correspondre parfaitement à la perception humaine de la qualité.

$$\text{MSE} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (f(i, j) - g(i, j))^2 \quad (3.1)$$

- $f(i, j)$: l'image originale de dimension $M \times N$.
- $g(i, j)$: l'image restaurée de dimension $M \times N$.

- **Peak Signal-to-Noise Ratio (PSNR)** : le PSNR est dérivé de l'erreur quadratique moyenne et indique le rapport entre l'intensité maximale des pixels et la puissance de la distorsion [61].

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (3.2)$$

Où MAX est la valeur maximale possible d'un pixel de l'image.

- **Indice de similarité structurelle (SSIM)** : la métrique SSIM combine la structure locale de l'image, la luminance et le contraste en un seul score de qualité locale. Dans cette métrique, les structures représentent les motifs d'intensité de pixels, notamment parmi les pixels voisins, après normalisation de la luminance et du contraste [61].

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.3)$$

- μ_x et μ_y sont les moyennes des images x et y.
- σ_x^2 et σ_y^2 sont les variances des images x et y.
- σ_{xy} La covariance entre x et y.
- $C_1 = (k_1L)^2$ et $C_2 = (k_2L)^2$ sont deux constantes de stabilisation pour éviter la division par zéro, où L est la plage dynamique des valeurs de pixel et k_1 et k_2 sont des constantes par défaut (typiquement $k_1=0.01$ et $k_2=0.03$).

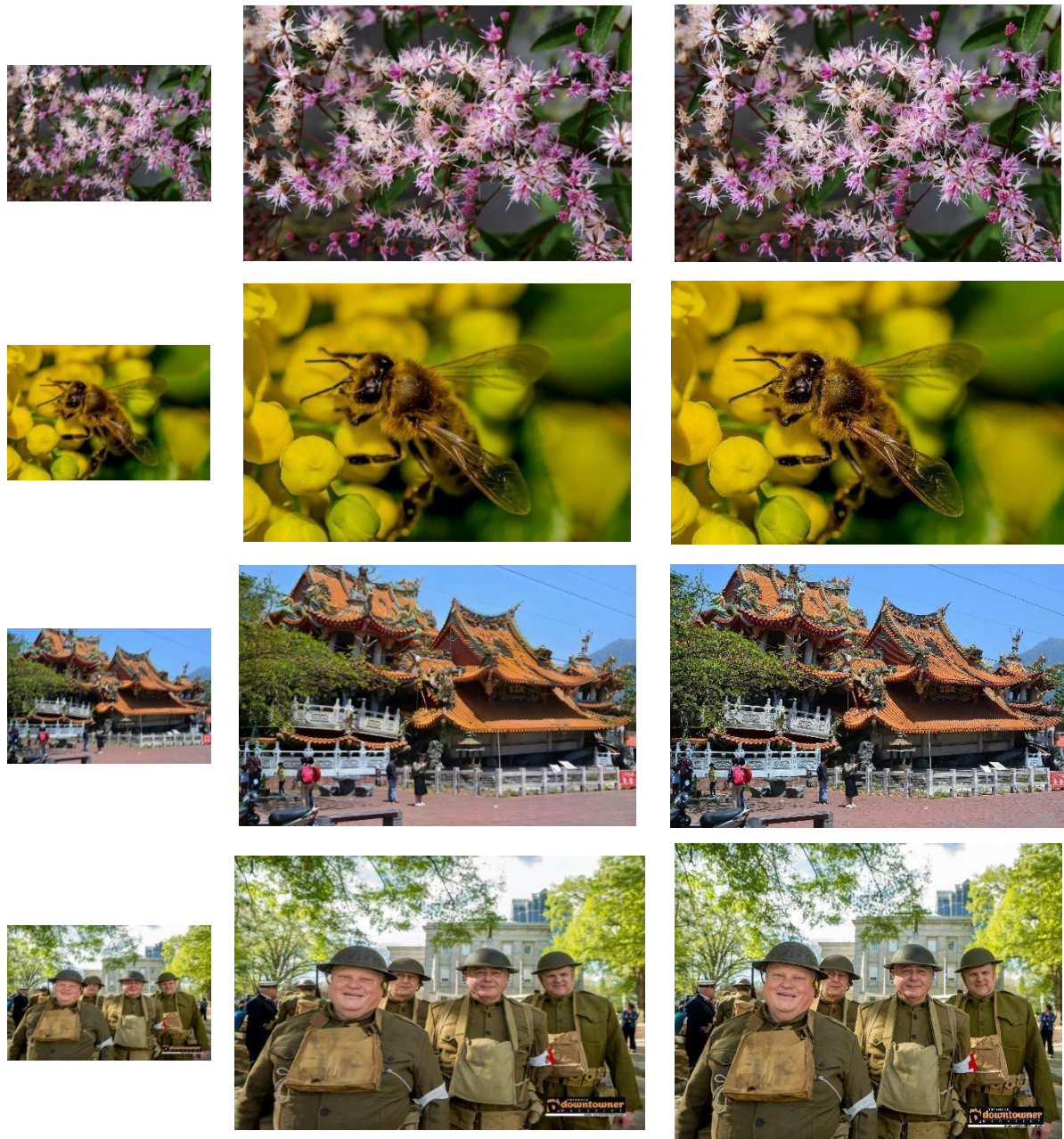
3.5.2 Métriques de qualité sans référence

- **Évaluateur de qualité spatiale d'image aveugle/sans référence (BRISQUE)**: Un modèle BRISQUE est entraînée sur une base de données d'images avec des distorsions connues, et BRISQUE est limité à évaluer la qualité des images présentant le même type de distorsion [61].
- **Évaluateur de qualité d'image naturelle (NIQE)** : un modèle NIQE soit entraîné sur une base de données d'images impeccables, NIQE peut mesurer la qualité d'images avec des distorsions arbitraires [61].
- **Évaluateur de qualité d'image basé sur la perception (PIQE)**: L'algorithme PIQE est insensible aux opinions et non supervisé, ce qui signifie qu'il ne nécessite pas de modèle entraîné [61]. PIQE peut mesurer la qualité d'images avec des distorsions arbitraires et, dans la plupart des cas, produit des résultats similaires à NIQE. PIQE estime la distorsion par blocs et mesure la variance locale des blocs perceptiblement distordus pour calculer le score de qualité.

3.6 Résultats et discussion

3.6.1 Comparaison qualitative

Nous avons réalisé une évaluation visuelle des performances de nos modèles de super-résolution d'images, basée sur des critères visuels et perceptuels pour évaluer la qualité des images reconstruites. Cette analyse nous permet de comparer notre méthode avec les approches traditionnelles de super-résolution d'images.

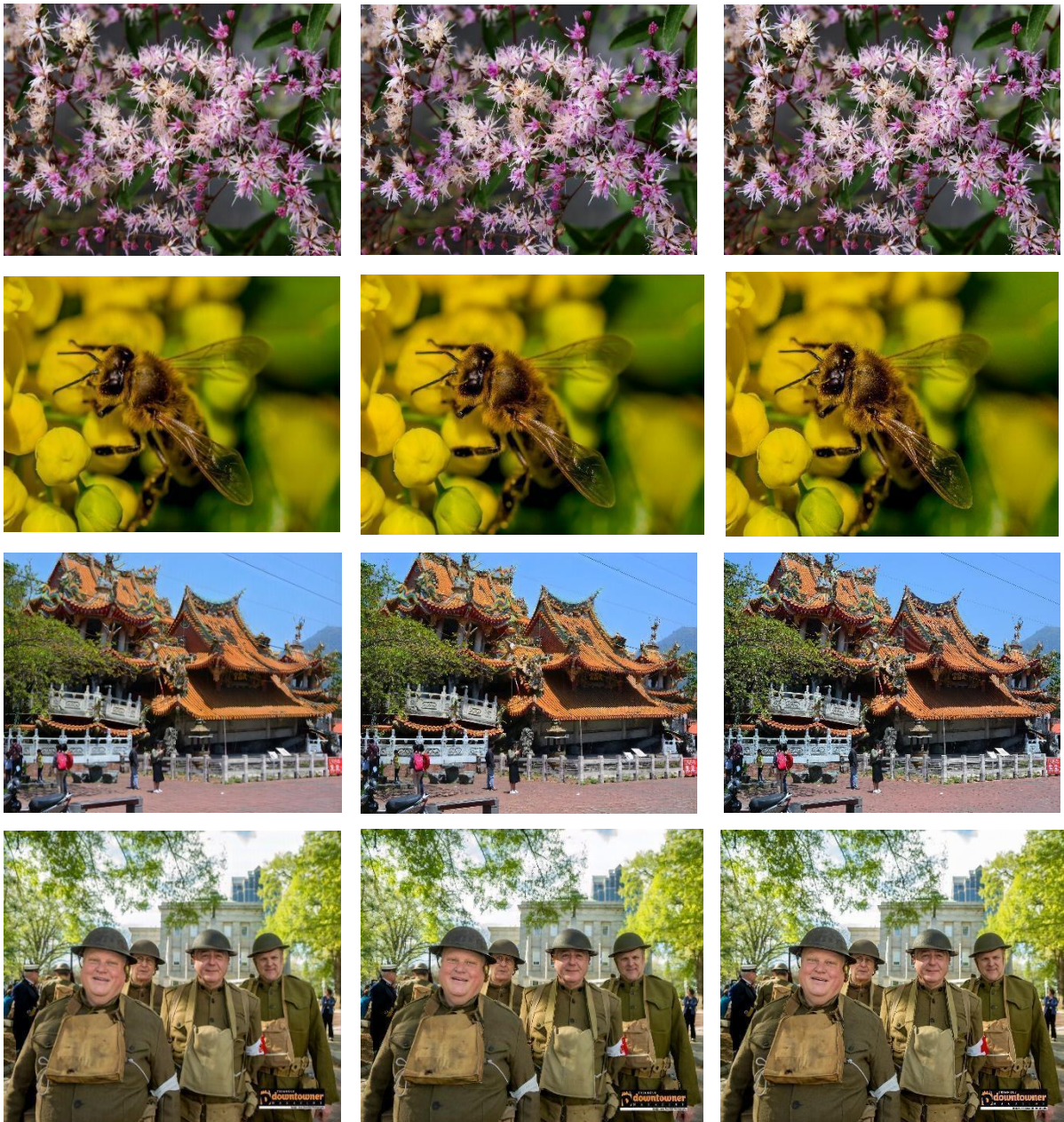


a) Image basse résolution

b) Image super-résolution (stage 1)

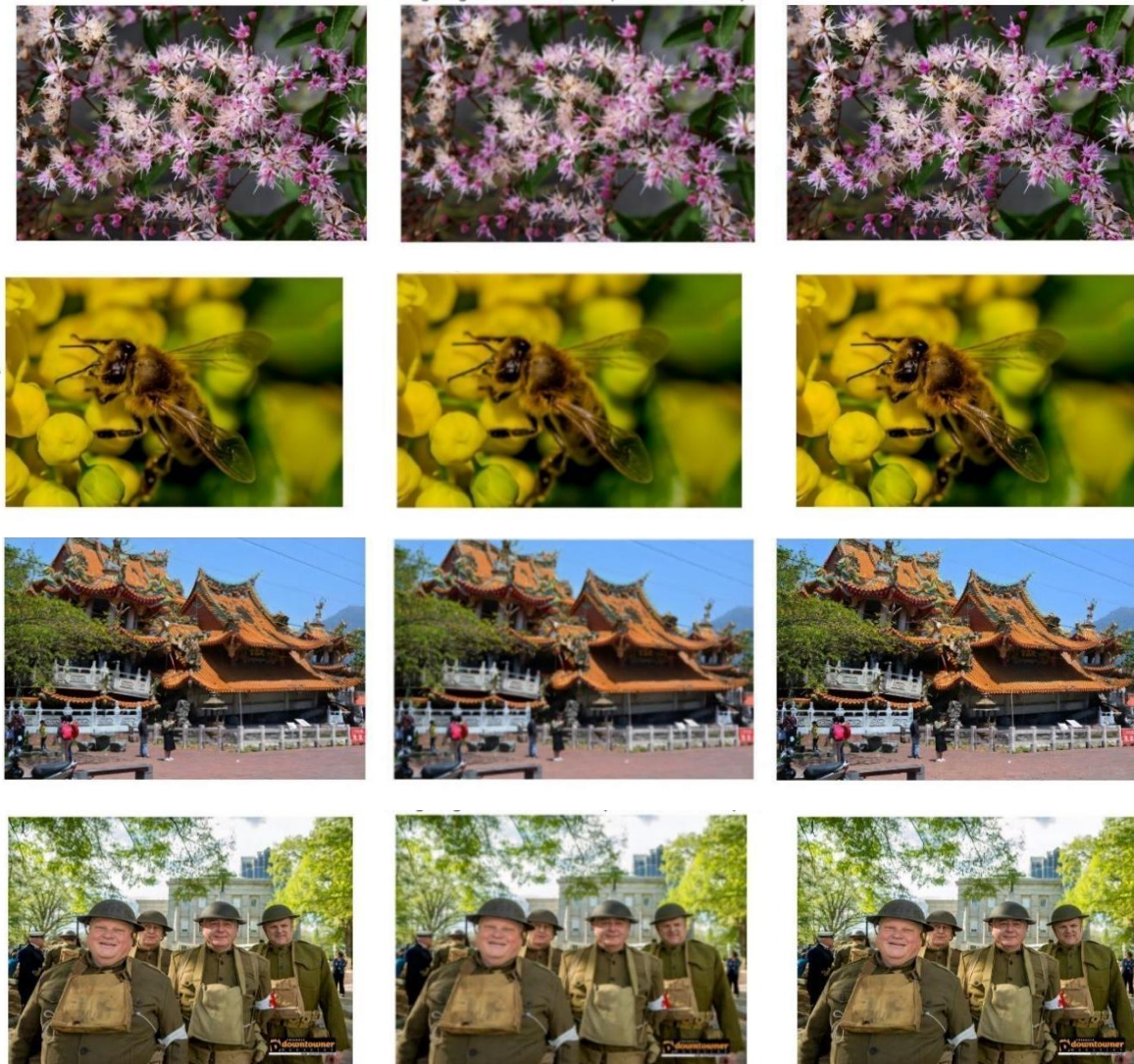
c) Image réelle (Ground-Truth)

Figure 3.1 : Résultats obtenus (stage 1)

a) Image super-résolution
(stage 1)b) Image super-résolution
(stage 2)c) Image réelle (Ground-
Truth)**Figure 3.2 :** Résultats obtenus (stage 2)

Come illustré dans les Figures 3.1 et 3.2, les résultats obtenus grâce à notre approche démontrent une capacité remarquable à restaurer la netteté et la clarté des détails fins. Les contours des objets sont nettement définis et les textures fines, telles que les motifs et les structures des surfaces, sont mieux préservées. De plus, nos modèles montrent une excellente capacité à maintenir la fidélité des couleurs, préservant ainsi les nuances subtiles et évitant les distorsions et artefacts de couleur.

Notre méthode est comparée avec deux approches traditionnelles : l'interpolation bilinéaire et l'interpolation bicubique. Les résultats obtenus sont illustrés dans les Figures 3.3 et 3.4.

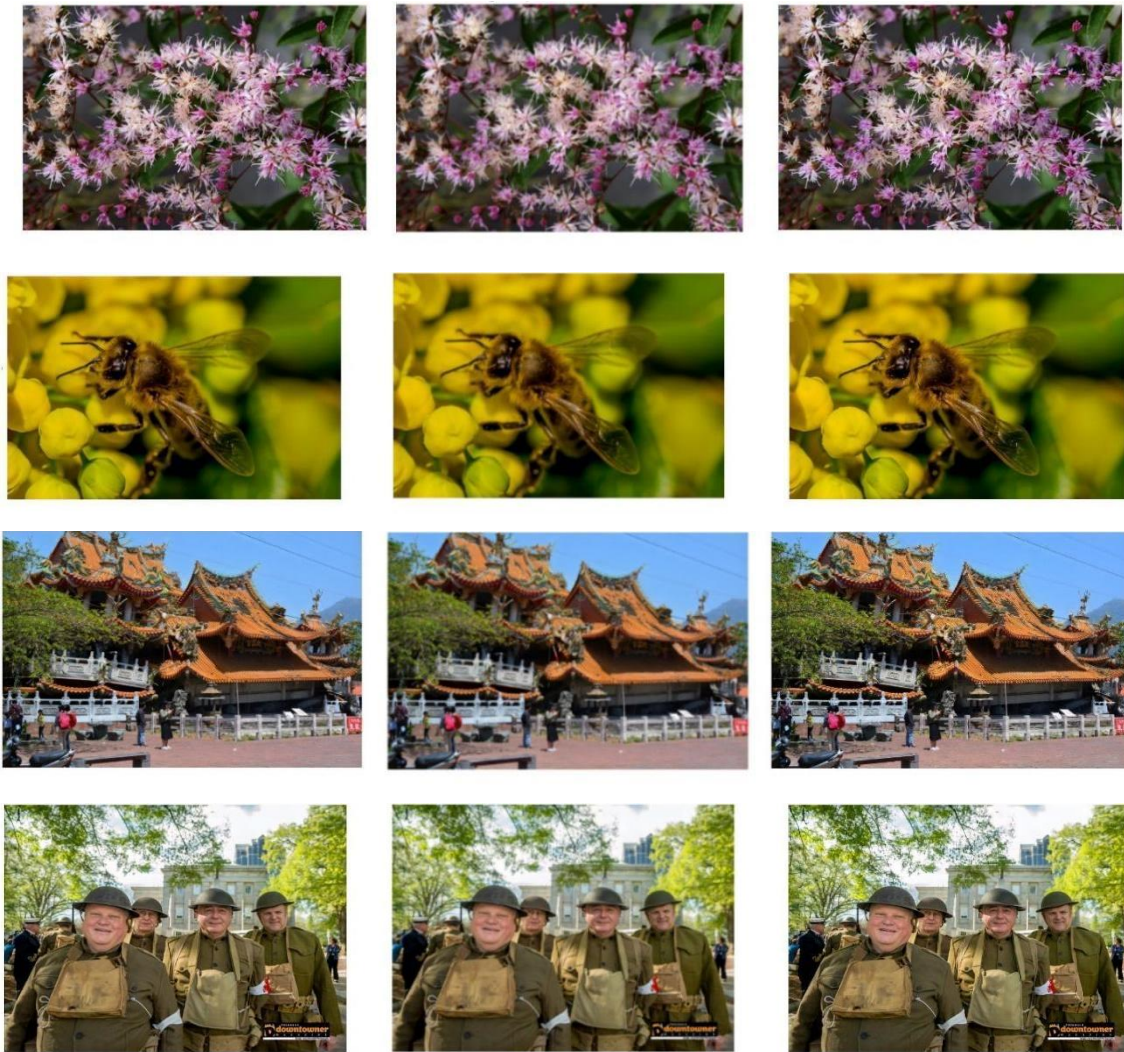


a) Image réelle (Ground-Truth)

b) Interpolation bicubique

c) Image super-résolution (stage 2)

Figure 3.3 : Comparaison qualitative des résultats obtenus avec la super-résolution par l'interpolation bicubique.



a) Image réelle
(Ground-Truth)

b) Interpolation bilinéaire

c) Image super-résolution
(stage 2)

Figure 3.4 : Comparaison qualitative des résultats obtenus avec la super-résolution par l'interpolation bilinéaire.

Comme illustré dans les Figures 3.3 et 3.4, les résultats montrent les limitations notables en termes de qualité des images de super-résolution obtenues par les méthodes traditionnelles. Les contours des objets apparaissent flous et les détails fins sont souvent perdus. Les textures sont moins définies et la netteté globale de l'image est compromise.

En revanche, les résultats obtenus avec notre méthode montrent une nette supériorité par rapport aux deux approches traditionnelles. Nos modèles basés sur l'apprentissage profond offrent une restauration exceptionnelle des détails fins et des contours nettement définis. Les textures sont recréées avec une grande précision, ce qui contribue à une qualité visuelle globale bien supérieure.

3.6.2 Comparaison quantitative

Tableau 3.1 : Comparaison quantitative des résultats sur des ensembles de données. (+) indique que le score le plus élevé est meilleur, (-) indique le score le plus bas est meilleur. Les deux meilleurs résultats sont en rouge et bleu respectivement.

Méthode	MSE-	PSNR+	SSIM+	NIQE-	PIQE-	BRISQUE-
Interpolation bilinéaire	0.0311	21.3300	0.7734	5.4014	89.3628	61.5190
Interpolation bicubique	0.0313	21.3106	0.7735	5.2706	82.7253	62.7925
Méthode proposée	0.0157	24.7745	0.8550	3.9187	12.3062	25.1142

Les résultats illustrés dans le tableau 3.1 montrent que notre méthode surpasse les approches traditionnelles sur toutes les mesures évaluées qui sont une meilleure précision et meilleure fidélité et similarité structurelle et perçues comme plus naturelles et proches de l'original visuellement.

Ces résultats soulignent l'efficacité supérieure des approches basées sur l'apprentissage profond, telles que celle que nous avons mise en œuvre, par rapport aux méthodes traditionnelles d'interpolation. L'utilisation de réseaux neuronaux permet de capturer des caractéristiques complexes et de réaliser des ajustements fins, conduisant à une amélioration significative de la qualité visuelle des images super-résolues.

3.7 Conclusion

Dans ce chapitre, nous avons présenté les outils utilisés pour explorer et évaluer nos modèles de super résolution d'images. Notre évaluation des performances a été réalisée à l'aide de comparaisons qualitatives et quantitatives en utilisant des métriques de qualité avec référence et sans référence, mettant en lumière l'efficacité de notre modèle dans le domaine de la super résolution d'images. Les résultats expérimentaux obtenus sur des ensembles de données ont clairement démontré que notre approche surpasse les méthodes traditionnelles en termes de qualité de super résolution d'images. Notre méthode parvient à restaurer les images basses résolution avec une fidélité et des détails comparables à ceux des autres méthodes, se rapprochant significativement des images réelles de haute qualité.

Conclusion générale

Conclusion générale

Dans ce mémoire, nous avons présenté le développement et l'évaluation d'un modèle de super-résolution d'image basé sur les réseaux antagonistes génératifs (GAN). Notre approche a montré que les GAN peuvent significativement améliorer la qualité des images basse résolution, offrant des résultats surpassant ceux des méthodes traditionnelles.

Nous avons utilisé des architectures de GAN avancées et exploré diverses techniques pour optimiser les performances du modèle. Les résultats obtenus, évalués à l'aide de mesures de qualité telles que le PSNR et le SSIM, montrent une amélioration notable de la clarté, de la netteté et du réalisme des images reconstruites.

Cependant, plusieurs défis restent à relever, notamment en termes de réduction du temps de calcul et de l'optimisation des architectures de réseaux pour des applications spécifiques. Les futures recherches pourraient se concentrer sur l'intégration de techniques de transfert de style et d'autres approches avancées pour améliorer encore la qualité et l'efficacité des modèles de super-résolution.

En conclusion, ce travail apporte une contribution significative au domaine de la super-résolution d'image, démontrant le potentiel des GAN pour améliorer la qualité des images. Les résultats de cette recherche ouvrent la voie à de nouvelles applications et innovations dans divers domaines nécessitant des images hautes résolution.

Bibliographie

Bibliographie

- [1] <https://fr.wikipedia.org/wiki/Super-résolution> (accessed June 11, 2024).
- [2] <https://www.shimmercat.com/image-api-feature/superresolution.html> (accessed 11/6/2024)
- [3] [Super_Resolution_CVonline.pdf](#) (ed.ac.uk) (accessed 11/6/ 2024).
- [4] <https://ar5iv.labs.arxiv.org/html/2102.03113> (accessed 7/6/2024).
- [5] <https://www.mdpi.com/2227-7390/8/11/1907/htm> (accessed June 15, 2024).
- [6] <http://dspace.univbatna.dz/bitstream/123456789/764/1/sce%20Yahiaoui%20mhamed%20aid.pdf> (accessed 7/6/2024).
- [7] <https://theses.hal.science/tel-02080798v1/document> (accessed June 7, 2024).
- [8] https://www.researchgate.net/figure/1-Reponse-frequentielle-du-filtre-rampe-et-du-filtre-apodise-a-Les-deux-autres_fig37_278632533 (accessed 15/6/2024).
- [9] <https://www.protegeraltersecourir.fr/blog/les-differentes-methodes-d-apprentissage.html> (accessed 11/6/2024).
- [10] <https://www.digiforma.com/guide-of/methodes-outils-pedagogiques/> (accessed 12/6/2024)
- [11] <https://www.formagora.fr/actualites/les-differentes-methodes-pedagogiques-de-la-formation/> (accessed 12/6/2024).
- [12] <https://www.sciencedirect.com/topics/computer-science/wavelet-transforms> (accessed 12/6/2024).
- [13] Wang, Y., Liu, Z., & Tan, M. (2020). Hybrid Deep Learning Model for Image Classification Using Wavelet Transform. *Neurocomputing*, 379, 190-198.
- [14] <https://imasolia.com/amelioration-dimages-de-la-super-resolution-avec-du-deep-learning/> (accessed 11/6/2024).
- [15] Taubman, D., & Marcellin, M. W. (2002). *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers.
- [16] Sayood, K. (2017). *Introduction to Data Compression* (5th ed.). Morgan Kaufmann Publishers.
- [17] Tang, J., & Tao, J. (2018). Applications of Wavelet Transform in Signal Processing. *Journal of Applied Signal Processing*, 2018, 567-578
- [18] <https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/> (accessed 12/6/2024).
- [19] <https://deeplylearning.fr/cours-pratiques-deep-learning/segmentation-semantique-dimages/> (accessed 19/6/2024).

- [20] http://www.mickaeltits.be/super_resolution/ (accessed 17/6/2024).
- [21] <https://www.salesforce.com/fr/resources/definition/reseau-de-neurones/> (accessed 17/6/2024).
- [22] <https://www.futura-sciences.com/tech/definitions/informatique-reseau-neuronal-601/> (accessed 17/6/2024).
- [23] <https://www.cnil.fr/fr/definition/reseau-de-neurones-artificiels-artificial-neural-network> (accessed 17/6/2024).
- [24] <https://www.frcneurodon.org/informer-sur-la-recherche/actus/les-grandes-avancees-technologiques-en-neurosciences/> (accessed 18/6/2024).
- [25] https://www.medecinesciences.org/en/articles/medsci/full_html/2020/06/msc200039/msc200039.html (accessed 17/6/2024).
- [26] https://www.researchgate.net/figure/Bicubic-Interpolation_fig2_365560492 (accessed 12/6/2024).
- [27] <https://neuroconnection.eu/quand-utiliser-les-reseaux-de-neurones-mlp-cnn-et-rnn/> (accessed 18/6/2024).
- [28] <https://kongakura.fr/article/Le-perceptron-multicouches> (accessed 18/6/2024).
- [29] https://www.iro.umontreal.ca/~pift6266/H12/html/mlp_fr.html (accessed 18/6/2024).
- [30] https://rtavenar.github.io/deep_book/fr/content/fr/mlp.html (accessed 18/6/2024).
- [31] <https://apiacoa.org/publications/teaching/nn/MLP.pdf> (accessed 18/6/2024).
- [32] https://fr.wikipedia.org/wiki/Perceptron_multicouche (accessed 18/6/2024).
- [33] Simon Haykin, "Neural Networks and Learning Machines", McMaster University Hamilton, Ontario, Canada, 3rd edition.
- [34] LeCun, Yann, et al. "Gradient-based learning applied to document recognition" *Proceedings of the IEEE* 86.11 (1998): 2278-2324.
- [35] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [36] Heaton, Jeff. "Ian goodfellow, yoshua bengio, and aaron courville: *Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618.*" *Genetic programming and evolvable machines* 19.1 (2018): 305-307.
- [37] <https://www.aquiladata.fr/insights/classification-dimages-et-detection-dobjets-par-cnn/> (accessed 18/6/2024).
- [38] 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.

- [39] <https://stanford.edu/~shervine/1/fr/teaching/cs-230/pense-bete-reseaux-neurones-convolutionnels> (accessed 18/6/2024).
- [40] <https://datascientest.com/convolutional-neural-network> (accessed 18/6/2024).
- [41] https://members.loria.fr/EKerrien/files/data/CETS8AH_DeepLearningIntro_TD.pdf (accessed 18/6/2024).
- [42] <https://dspace.univguelma.dz/jspui/bitstream/123456789/4224/1/PFE2019%20Djabali%20Lounis%20%28Tchi%20drive2%29%20%281%29.pdf> (accessed 19/6/2024).
- [43] <https://fr.mathworks.com/discovery/convolutional-neural-network.html> (accessed 19/6/2024).
- [44] <https://blog.octo.com/you-only-look-once-un-reseau-de-neurones-pour-la-detection-dobjets> (accessed 19/6/2024).
- [45] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 391-407). Springer International Publishing.
- [46] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
- [47] <https://phototrend.fr/2021/03/nouveautes-photoshop-agrandissement-sans-perte-avec-super-resolution-et-version-mac-m1/> (accessed 19/6/2024).
- [48] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “Unet: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [49] Raghav Mehta and Jayanthi Sivaswamy, “M-net: A convolutional neural network for deep brain structure segmentation,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 437–440.
- [50] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “Learning enriched features for real image restoration and enhancement,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*. Springer, 2020, pp. 492–511.

- [51] Fan, Chi-Mao, Tsung-Jung Liu, and Kuan-Hsien Liu. "Half wavelet attention on M-Net+ for low-light image enhancement." 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022.
- [52] Das Mou, Trisha, et al. "Multi-Range Sequential Learning Based Dark Image Enhancement with Color Upgradation." Applied Sciences 13.2 (2023): 1034.
- [53] Brownlee, Jason. Generative adversarial networks with python: deep learning generative models for image synthesis and image translation. Machine Learning Mastery, 2019.
- [54] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [55] Jolicoeur-Martineau, Alexia. "The relativistic discriminator: a key element missing from standard GAN." arXiv preprint arXiv: 1807.00734 (2018).
- [56] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- [57] https://www.w3schools.com/python/python_intro.asp (accessed 22/6/2024).
- [58] https://medium.com/@Sunil_Kumawat/exploring-pytorch-from-fundamentals-to-neural-network-construction (accessed 22/6/2024).
- [59] Wang, Yingqian, et al. "Flickr1024: A large-scale dataset for stereo image super-resolution." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.
- [60] Agustsson, Eirikur, and Radu Timofte. "Ntire 2017 challenge on single image super-resolution: Dataset and study." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017.
- [61] https://fr.mathworks.com/help/images/image-quality-metrics.html#mw_2ebd06de-46dc-4394-9fd9-9c2ad8e84e10 (accessed 22/6/2024).