الجمهورية الجزائرية
الديمقراطية الشعبية
Republique Algerienne Democratique et Populaire
وزارة التعليم العالي والبحث
العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة الشهيد الشيخ العربي
التبسي - تبسة
Université Echahid Cheikh Larbi Tébessi – Tébessa
Faculté des Sciences et de la Technologie
Département de d'Électronique et de Communications

MEMOIRE

Présenté pour l'obtention du **diplôme** de **Master Académique**

Filière : Électronique
Spécialité : Instrumentation



Par :
Berkani Lamis
Aoulmi Chahinda
Boukhris Haifa

THEME

Conception d'un système intelligent de détection des images falsifiées basé sur le Deep Learning et la transformée en Ondelettes (DWT)

Présenté et évalué, le / /2026 , par le jury composé de :

Nom et prénom		Grade	Qualité
Mr.	MAAMRI Mahmoud	MCA	Président
Mme.	CHERIET Leyla	MCB	Rapporteur
Mme.	AOUICHE Chaima	MCB	Examinateur

Promotion : 2025/2026

Scanned with CamScanner

الجمهورية الجزائرية
قراطية الشعب
Republique Algerienne Democratique et Populaire
وزارة التعليم العالي والبحث
العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة الشهيد الشيخ العربي
التبسي - تبسة
Université Echahid Cheikh Larbi Tébessi – Tébessa
Faculté des Sciences et de la Technologie
Département de d'Électronique et de Communications

MEMOIRE

Présenté pour l'obtention du **diplôme** de **Master Académique**

Filière : Électronique
Spécialité : Instrumentation

Par :
Berkani Lamis
Aoulmi Chahinda
Boukhris Haifa

THEME

Conception d'un système intelligent de détection des images falsifiées basé sur le Deep Learning et la transformée en Ondelettes (DWT)

Présenté et évalué, le / /2026 , par le jury composé de :

Nom et prénom		Grade	Qualité
Mr.	MAAMRI Mahmoud	MCA	Président
Mme.	CHERIET Leyla	MCB	Rapporteur
Mme.	AOUICHE Chaima	MCB	Examinateur

Promotion : 2025/2026

Dédicace

إهداء من الطالبة بركاني ليس

بِسْمِ اللهِ الرَحمَن الرَحيمِ

الحمدُ لله الذي جعلَ العلمَ نوراً، والعلماءَ لورثتِهِ مناراً، القائل في كتابه المحفوظ يَرْفَعِ اللَه الَذينَ آمَنُوا مِنكُمْ وَالَذينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ، وحمداً لله الذي أنارَ بصيرتي، وسدّد خُطاي، وأتاحَ لي أن أقفَ اليومَ على عتبةِ هذا النجاح الذي لولا توفيقهُ وكرمهُ ما كان ليرى النور، وإليهِ سبحانهُ ملجئي ومبتغاي. ونحن إذ نخطو اليوم في رحابِ العلمِ ومَعارجه، نَستحضرُ بقلوبٍ خاشعةٍ مُمتنّةٍ قولَ رسولنا الأكرم ونبينا الأعظم صلى الله عليه وسلم مَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا، سَهَل اللَه لَهُ بِهِ طَرِيقًا إِلَى الْجَنَة، فلكَ الحمدُ يا ربَنَا أنِ اصطفيتَني لِسلوكِ هذا الدَرْبِ المُبارك، ولكَ الحمدُ أن جَعلتَ سهري، وبحثي، ومُجاهدتي خُطواتٍ أرجو بها رِضاك وعظيمَ ثوابِك، وصَبَرْتني على عناءِ الطَرْيقِ حتّى بلَغْتنا جَنيَ الثِمَار ورفعةَ المَقام.

وإلى من شقّا لي في دروبِ الحياةِ سُبلاً من نور وحملا عني عناء الأيام؛ أتقدم بأسمى آيات الحب والامتنان إلى والدي العزيز **عبد الله** حفظه الله تاجاً فوق رأسي، وإلى أمي الغالية **نبيلة** منبع الجنان والرضوان وشريكة نجاحي. وإلى أختي وسندي **سندس** خير عونٍ لي. كما أُهدي هذا العمل تقديراً وعرفاناً إلى مشرفتي الفاضلة الأستاذة **شريط ليلى** نِعمَ السند والموجّه العلمي والإنساني، وإلى رفيقات الدرب **مروى** و **لبنى** و **عائشة** و **زينب** . و إلى عائلتي الكبيرة حصني المنيع، وبأكفٍ مرفوعة بالدعاء أخص الراحلين المتوفين من عائلتي الذين غادروا الجسد ولم تغادر أرواحهم الطاهرة ذاكرتي، سائلة المولى أن يتغمدهم بواسع رحمته ويجعل قبورهم روضة من رياض الجنة.

Dédicace

إهداء من الطالبة عولمي شهندة

بِسْمِ اللهِ الرَحمَن الرَحيمِ

بكل مشاعر الحب والامتنان، أهدي هذا العمل المتواضع إلى من كانا سبباً في كل نجاح أصل إليه، إلى أمي الغالية صونيا، إلى القلب الذي منحني الحنان والقوة، وإلى الدعوات الصادقة التي كانت ترافقني في كل خطوة، شكراً لأنك كنتِ دائماً سندي وأماني. وإلى أبي العزيز عبد اللطيف، رمز التعب والاجتهاد، الذي علّمني أن الطريق إلى النجاح يبدأ بالصبر والإرادة، فلك مني كل التقدير والاحترام. إلى أخي وأخواتي الأحبة، أنتم فرحة أيامي وسندي الجميل، شكراً لوجودكم الدائم بجانبي ولكل لحظة دعم وتشجيع ومحبة. كما أهدي هذا العمل إلى أعمامي وعمّاتي وأبناء عمومتي، وكل أفراد عائلتي الذين غمروني بالمحبة والدعاء. وإلى صديقاتي العزيزات وزميلات البحث، شكراً لكل لحظة تعب وتقاسم للجهد، ولكل دعم وكلمة طيبة رافقت هذا المشوار، فوجودكن جعل الطريق أخف وأجمل. كما أتقدم بخالص الشكر والتقدير إلى أستاذتي المشرفة، التي كان لتوجيهاتها القيمة ونصائحها الأثر الكبير في إتمام هذا العمل، فجزاها الله خير الجزاء. إلى كل من آمن بي وساندني ولو بكلمة طيبة، أهدي ثمرة هذا الجهد بكل فخر ومحبة، راجيةً من الله التوفيق والنجاح في القادم.

Dédicace

إهداء من الطالبة بوخريص هيفاء

بِسْمِ اللهِ الرَحمَن الرَحيمِ

بكل مشاعر الحب والامتنان، أحمد الله تعالى الذي وفقني لإتمام هذا العمل المتواضع، وأتقدم بأسمى عبارات الشكر والتقدير إلى أعز الناس على قلبي.

إلى أمي الغالية براك خضراء، التي كانت ولا تزال مصدر إلهامي وقوتي، والتي غرست في داخلي الأمل والطموح، ورافقتني بدعواتها وصبرها وحنانها في كل مراحل حياتي. كلمات الشكر لا تكفي للتعبير عن مكانتك في قلبي، فأنتِ النور الذي يضيء دربي والسند الذي أستمد منه العزيمة للاستمرار و تحقيق أحلامي.

إلى أبي العزيز فتحي، الذي كان دائماً مصدر قوتي وفخري، والذي علمني معنى الاجتهاد والصبر والثقة بالنفس، وساندني بكل ما يملك حتى أصل إلى هذه المرحلة. شكراً لك لأنك كنت السند الحقيقي في كل خطوة، ولأن تعبك وتضحياتك كانت الدافع الأكبر لنجاحي.

إلى أخواتي العزيزات أزهار ولينة، اللتين كانتا مصدر سعادتي وراحتي، شكراً لكل لحظة دعم ومحبة وتشجيع، ولكل كلمة جميلة خففت عني التعب ومنحتني الأمل والطاقة للاستمرار. وجودكما في حياتي نعمة لا تقدر بثمن.

كما أهدي هذا العمل إلى أصدقائي الأعزاء يحيى، آلاء، عائشة، إسراء، رؤى، نورهان، زهراء، ومجدولين، الذين شاركوني أجمل اللحظات وقدموا لي الدعم المعنوي والتشجيع طوال مسيرتي الدراسية. شكراً لصدقكم ووفائكم ولكل ذكرى جميلة ستبقى محفورة في القلب.

إليكم جميعاً أهدي ثمرة هذا الجهد المتواضع، راجيةً من الله أن يكون بداية لتحقيق المزيد من النجاحات والطموحات. أتقدم بخالص الشكر والتقدير إلى أستاذتي المؤطرة على توجيهاتها القيمة ومرافقتها لي طوال إنجاز هذه المذكرة، وعلى دعمها وتشجيعها المستمر. فكل الاحترام والامتنان لها على جهودها ونصائحها التي ساهمت في إتمام هذا العمل.

Remerciements

Avant tout, nous remercions Dieu le Tout-Puissant de nous avoir donné la force, le courage, la volonté et la patience nécessaires pour accomplir ce modeste travail.

Nous tenons à exprimer nos profonds remerciements et notre reconnaissance à notre encadreuse, **Mme. CHERIT Leila**, pour avoir accepté de diriger ce travail, pour sa grande disponibilité, ses précieux conseils, son orientation scientifique rigoureuse, ainsi que pour ses qualités humaines exceptionnelles qui nous ont guidées tout au long de ce projet.

Nos vifs remerciements s'adressent également aux membres du jury qui nous font l'honneur d'évaluer notre travail et de présider notre soutenance.

Nous tenons aussi à remercier chaleureusement le recteur de l'université, **le doyen de notre faculté**, notre chef de département **M. SEIGA Mohammed**, ainsi que l'ensemble des enseignants du département d'Électronique pour leur encadrement pédagogique et les connaissances qu'ils nous ont transmises.

Un grand merci et toute notre gratitude s'adressent également à tout le personnel de l'Université Cheikh Larbi Tébessa, des agents de sécurité et des gardiens jusqu'aux agents de propreté et d'entretien, pour leurs efforts quotidiens et pour avoir veillé à nous offrir un environnement de travail agréable durant nos années d'études.

Enfin, nous n'oublierons pas de remercier du fond du cœur nos chers parents pour leur soutien inconditionnel, leurs sacrifices et leurs prières bénédictes, ainsi que nos familles, nos amis et tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce projet de fin d'études.

List of Figures

1.1 Principe de rupture de la corrélation spatiale lors d'une attaque par copier-déplacer. 6
1.2 Cartographie conceptuelle du motif de bruit PRNU agissant comme empreinte unique. 7
1.3 Taxonomie générale des différentes catégories de falsification d'images numériques. 7
1.4 Processus de la falsification par Copier-Déplacer (Copy-Move) : (a) schéma conceptuel du clonage de zone, (b) illustration pratique sur une image réelle [29]. 8
1.5 Processus de la falsification par Épissage (Image Splicing) : (a) schéma conceptuel de la fusion de deux sources différentes, (b) illustration pratique du résultat falsifié [30]. 9
1.6 Exemple d'altération des propriétés intrinsèques par retouche d'image [31]. 9
1.7 Exemple de falsification par génération d'image (Deepfake) [32]. . . . 10
1.8 Exemple de falsification par *copier-déplacer* (copy-move) tiré de la base de données CASIA v2.0 [33]. 13
1.9 Exemple de falsification par *épissage* (splicing) tiré de la base de données Columbia [34]. 13
1.10 Exemples comparatifs de visages synthétisés extraits des benchmarks de Deepfakes [35]. 14
1.11 Matrice de confusion standardisée pour l'extraction des métriques de classification. 14
1.12 Évaluation de la robustesse via l'analyse de l'espace ROC. 15

2.1 Pavage de l'espace temps-fréquence (Diagrammes de l'espace des phases). 22
2.2 Architecture algorithmique duale et symétrique de l'algorithme de Mallat. 24
2.3 Représentation temporelle des fonctions de base de l'ondelette de Haar. 24
2.4 Forme d'onde asymétrique caractéristique de la famille Daubechies (db2). 25
2.5 Propriétés de symétrie quasi-parfaite de la famille Symlet (sym4). . . 25

2.6 Symétrie parfaite des fonctions d'ondelettes duales de la famille biorthogonale bior4.4. 26
2.7 Schéma fonctionnel du banc de filtres bidimensionnel (2D-DWT) avec décimation par lignes et colonnes [17] 27
2.8 Décomposition d'une image en sous-bandes d'approximation (LL) et de détails directionnels (LH, HL, HH).[16] 27

3.1 Inclusion hiérarchique du Deep Learning au sein de l'Intelligence Artificielle. 36
3.2 Modélisation mathématique et flux interne d'un neurone artificiel standard. 37
3.3 Schéma cyclique global de l'apprentissage (gauche) et processus géométrique de convergence vers le minimum de la fonction de perte (droite). . . . 38
3.4 Parallélisme fonctionnel entre le système visuel biologique (V1) et la hiérarchie d'abstraction des caractéristiques dans un CNN. 39
3.5 Représentation mathématique du mécanisme de balayage par un noyau convolutif. 40
3.6 Opération de réduction dimensionnelle par Max Pooling (2×2, Stride 2). 41
3.7 Opération de réduction dimensionnelle par Average Pooling (2×2, Stride 2). 42
3.8 Architecture fonctionnelle et étapes mathématiques de la Batch Normalization. 43
3.9 Comparaison structurelle entre un réseau dense classique et un réseau avec application du Dropout. 43
3.10 Flux de traitement séquentiel : de l'aplatissement des cartes 3D à la classification finale par la couche Dense. 44
3.11 Processus d'aplatissement (Flatten) convertissant les cartes de caractéristiques 2D en un vecteur 1D. 44
3.12 Structure interne d'une couche Dense (Fully Connected) montrant la densité des liaisons. 45
3.13 Pipeline complet d'un CNN : de l'extraction des caractéristiques spatiales (*Feature Extraction*) à la distribution probabiliste finale via la fonction SoftMax.[18] 47
3.14 Pipeline de classification d'images : attribution d'un label unique basé sur des probabilités. 48
3.15 Exemple de détection d'objets multiples avec leurs boîtes englobantes et scores de confiance[19]. 49

3.16 Principe de la segmentation sémantique : classification pixel par pixel pour isoler des formes complexes[20]. 50
3.17 Processus algorithmique du transfert de style artistique via l'analyse des couches du CNN : fusion du contenu structurel et du style esthétique. 51
3.18 Impact du choix du taux d'apprentissage (α) sur la dynamique de convergence géométrique. 52
3.19 Planification du taux d'apprentissage selon la stratégie Piecewise Constant Decay. 53
3.20 Évolution des courbes de perte théoriques mettant en évidence le point de bascule vers le sur-apprentissage (Overfitting). 54

4.1 Architecture hybride globale proposée pour le système de fusion DWT-CNN. 60
4.2 Comparaison empirique de la dynamique de convergence selon le régime de partitionnement des données (*Mini-batch analysis*). 63
4.3 Résultats individuels de classification et scores de confiance (Le symbole $\times$ indique une erreur de diagnostic face à la vérité terrain). . . . 64
4.4 Profil de convergence et dynamique d'apprentissage avec l'architecture CNN Standard (15 couches). 65
4.5 Profil de convergence et dynamique d'apprentissage avec l'architecture SqueezeNet (68 couches). 66
4.6 Profil de convergence et dynamique d'apprentissage avec l'architecture profonde ResNet-18 (71 couches). 66
4.7 Matrices de confusion vectorielles (P. : Prédit, R. : Réel). 68
4.8 Courbes ROC comparatives des trois architectures et calcul des aires sous la courbe (AUC) correspondantes. 69
4.9 Analyse de la sensibilité de détection selon l'ondelette mère sélectionnée. 70
4.10 Comparaison topologique des mécanismes de traitement : Séquentiel (CNN), Compression-Expansion (SqueezeNet), et Résiduel (ResNet). 74

List of Tables

1.1 Synthèse comparative des approches de l'état de l'art. 12
1.2 Comparaison des principaux datasets dédiés aux Deepfakes. 14

2.1 Comparaison structurelle entre les transformées de Fourier, STFT et DWT. 21

3.1 Synthèse fonctionnelle des composants structurels d'un CNN. 46

4.1 Configuration des hyperparamètres pour l'entraînement du système de fusion DWT-CNN. 63
4.2 Synthèse exhaustive des critères structurels, hyperparamètres et performances des réseaux évalués. 67
4.3 Évaluation statistique comparative incluant l'aire sous la courbe ROC (AUC). 69
4.4 Analyse comparative de l'impact des ondelettes mères sur les performances de fusion spatio-fréquentielle. 70

Introduction Générale

À l'ère numérique et avec la montée rapide de l'intelligence artificielle générative, l'image est devenue bien plus qu'un simple moyen d'information visuelle. Elle est maintenant un document légal et criminel essentiel. Cependant, la croissance des techniques de manipulation avancées, comme les Deepfakes et les falsifications sémantiques, a sérieusement affecté la confiance que l'on accorde aux contenus numériques. Pour faire face à ce défi sécuritaire, le domaine de la criminalistique informatique doit évoluer en développant des protocoles d'authentification automatisés. Ces protocoles doivent être capables de repérer les micro-anomalies invisibles à l'œil nu.

Le but principal de cette recherche est de créer et de valider une méthode robuste pour détecter les falsifications d'images. Notre contribution repose sur la combinaison de l'analyse spatio-fréquentielle via la Transformée en Ondelettes Discrète (DWT) et de la puissance d'extraction de caractéristiques des Réseaux de Neurones Convolutifs (CNN). Pour bien présenter cette étude et en exposer les bases théoriques et les validations empiriques, notre manuscrit se déroule en quatre chapitres complémentaires :

Le **premier chapitre** se concentre sur les concepts de base de l'image numérique et sur la taxonomie des falsifications. Nous définissons clairement les propriétés intrinsèques de l'image du point de vue mathématique puis nous explorons les signatures balistiques naturelles, comme le bruit de capteur (PRNU) et la corrélation spatiale des pixels. Ce chapitre fournit une revue critique des attaques courantes, comme le copier-déplacer et l'épissage, tout en posant le cadre conceptuel nécessaire à la compréhension des anomalies physiques causées par une manipulation.

Le **deuxième chapitre** couvre les bases mathématiques de la Transformée en Ondelettes Discrète (DWT). Contrairement à la transformée de Fourier classique, la DWT se révèle être un outil d'analyse multi-résolution efficace, capable de segmenter l'image en sous-bandes d'approximation et de détails (LH, HL, HH). Nous examinons comment ce processus permet d'isoler les résidus de haute fréquence qui contiennent les discontinuités statistiques et les micro-artefacts laissés par les algorithmes de falsification.

Le **troisième chapitre** traite de la structure des Réseaux de Neurones Convolutifs (CNN). S'inspirant du système visuel biologique, ces modèles excellent dans

l'apprentissage automatique de représentations complexes. En étudiant leurs composants clés, y compris les couches de convolution, le pooling et la régularisation, nous mettons en lumière comment ces réseaux peuvent modéliser des structures non linéaires complexes. Nous introduisons également le concept d'apprentissage résiduel, qui est crucial pour surmonter la dégradation du gradient dans les architectures profondes.

Enfin, le **quatrième chapitre** présente notre contribution à travers l'implémentation expérimentale et la discussion des résultats dans l'environnement MATLAB. Ici, nous validons l'intégration de la DWT avec les architectures convolutives pour guider les réseaux à se concentrer sur les détails fréquentiels. Par le biais d'une analyse comparative rigoureuse englobant des architectures standard comme SqueezeNet et ResNet-18, et soutenue par l'évaluation des matrices de confusion, nous démontrons l'efficacité de l'apprentissage résiduel. Cela prouve la viabilité opérationnelle de notre approche face aux défis de la sécurité numérique.

Chapter 1
Falsification d'Images : Généralités

1.1 Introduction . . . 4
1.2 La Notion d'Image Numérique et ses Propriétés . . . 4
1.2.1 Définition mathématique et numérisation . . . 4
1.2.2 Propriétés structurelles fondamentales . . . 5
1.3 Définition de la Falsification d'Images . . . 7
1.3.1 Falsification par Copier-Déplacer (Copy-Move Forgery) . . . 8
1.3.2 Épissage d'Images (Image Splicing) . . . 8
1.3.3 Retouche d'Images (Image Retouching) . . . 9
1.3.4 Falsification Générative (Deepfakes et GANs) . . . 9
1.4 État de l'Art des Techniques de Détection de falsification . . . 10
1.4.1 Première génération : Approches physiques et statistiques (2000–2010) . . . 10
1.4.2 Deuxième génération : Approches par Apprentissage Automatique (2010–2018) . . . 11
1.4.3 Troisième génération : L'ère de l'Apprentissage Profond (2018–Présent) . . . 11
1.5 Bases de données de référence (CASIA, Columbia, etc.) . . . 12
1.5.1 Introduction et critères d'évaluation . . . 12
1.5.2 Datasets classiques de falsification d'images . . . 12
1.5.3 Datasets spécifiques aux deepfakes et contenus génératifs . . . 13
1.6 Métriques d'Évaluation . . . 14
1.7 Conclusion du Chapitre . . . 15

Résumé

Ce chapitre dresse un panorama global et rigoureux du domaine de la falsification d'images numériques (*Image Forgery*) et des méthodologies dédiées à sa détection. Dans un premier temps, nous définissons les propriétés mathématiques et physiques d'une image numérique saine. Ensuite, nous établissons une taxonomie claire des techniques de manipulation, allant des méthodes conventionnelles (copier-déplacer, épissage, retouche) aux approches génératives de nouvelle génération basées sur l'intelligence artificielle (Deepfakes, GANs). De plus, nous analysons l'état de l'art à travers trois générations technologiques, suivies d'une étude comparative des bases de données de référence. Enfin, nous formalisons les métriques mathématiques et protocoles expérimentaux indispensables à la validation quantitative des performances. Ce cadre théorique sert de fondement scientifique au système de détection développé dans le cadre de nos travaux.

Keywords: Transformée en Ondelettes Discrète (DWT), Analyse Multirésolution (MRA), Algorithme de Mallat, Banc de filtres, Traitement d'images, JPEG 2000, Débruitage, Image Forensics.

1.1 Introduction

Historiquement, la photographie a toujours été considérée comme une preuve irréfutable de la réalité. Cependant, la transition du domaine analogique vers le domaine numérique, combinée à la prolifération de logiciels d'édition sophistiqués, a radicalement modifié ce paradigme [1]. Aujourd'hui, modifier le contenu d'une image est devenu une tâche accessible à tous.

Cette accessibilité pose des risques majeurs dans plusieurs secteurs critiques tels que la justice pénale, le journalisme, la politique et la sécurité nationale [2]. Face à l'essor récent de l'intelligence artificielle générative, les défis liés à la confiance numérique se sont intensifiés, rendant les approches de criminalistique multimédia plus indispensables que jamais [3]. L'objectif principal de ce chapitre est de dresser un état de l'art rigoureux sur la falsification d'images (*Image Forgery*). Nous explorerons d'avance la structure d'une image saine avant d'établir une typologie claire des différentes manipulations existantes et des approches scientifiques modernes dédiées à leur détection.

1.2 La Notion d'Image Numérique et ses Propriétés

Avant d'analyser les mécanismes de falsification, il est fondamental de définir rigoureusement ce qu'est une image numérique d'un point de vue mathématique et informatique, ainsi que les propriétés intrinsèques qui la caractérisent en tant que document original [1].

1.2.1 Définition mathématique et numérisation

Une image numérique bidimensionnelle (2D) peut être modélisée comme une fonction continue $f(x, y)$ de deux variables spatiales (x, y), où x et y représentent les coordonnées planes, et la valeur de f correspond à l'intensité ou à la luminosité de l'image en ce point [13].

Pour être exploitée par des systèmes informatiques, cette fonction continue doit subir un processus de **numérisation** composé de deux étapes distinctes [13] :

- **L'échantillonnage spatial :** Qui consiste à discrétiser l'espace continu (x, y) en coordonnées discrètes (i, j), définissant ainsi la matrice géométrique des pixels.
- **La quantification :** Qui traduit l'amplitude continue $f(x, y)$ en une valeur numérique discrète choisie dans un ensemble fini (généralement codée sur 8 bits, soit un intervalle d'intensité $[0, 255]$).

À l'issue de cette numérisation, l'image se structure sous forme d'une matrice bidimensionnelle de dimensions $M \times N$ d'éléments discrets appelés *pixels* (Picture Elements) [4] :

$$I(i,j) = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,N} \end{pmatrix} \tag{1.1}$$

1.2.2 Propriétés structurelles fondamentales

Une image issue d'un processus d'acquisition authentique possède des caractéristiques statistiques et physiques intrinsèques. Ce sont précisément ces propriétés qui sont altérées ou brisées lors d'une manipulation frauduleuse [10] :

1.2.2.1 La Résolution Spatiale et le Rééchantillonnage

La résolution spatiale est spécifiée par les dimensions $M \times N$ de la matrice de pixels. Toute insertion d'un objet externe lors d'un épissage (*splicing*) impose généralement un redimensionnement géométrique (*resampling*) ou une interpolation spatiale pour adapter l'échelle de l'intrus. Cette modification géométrique artificielle brise la régularité mathématique native du réseau de pixels et introduit des anomalies périodiques microscopiques détectables dans le domaine fréquentiel ou spatial [13].

1.2.2.2 La Profondeur de Couleur (Bit Depth)

Le standard colorimétrique RGB (Rouge, Vert, Bleu) alloue classiquement 8 bits par canal colorimétrique, générant une dynamique codée sur 24 bits par pixel. Les opérations de retouche numérique, de lissage ou d'ajustement du contraste forcent des distorsions statistiques majeures. Ces altérations se traduisent par des ruptures abruptes, telles que le phénomène de peigne ou de postérisation, au sein de l'histogramme des couleurs de la zone manipulée [10].

1.2.2.3 La Corrélation Spatiale (Pixel Correlation)

Dans une scène naturelle authentique, il existe une forte dépendance statistique locale entre un pixel et son voisinage immédiat, dictée par la continuité physique des textures et de l'éclairage de la scène. Lorsqu'une attaque par copier-déplacer (*copy-move*) est exécutée, la duplication exacte d'un bloc de pixels propage également sa structure de corrélation interne. Cela crée une réplication de motifs statistiques

hautement improbable et rompt brutalement la cohérence aux frontières du collage [6].

1.2.2.4 Le Bruit de Capteur (PRNU)

Chaque capteur physique (CCD ou CMOS) présente des imperfections de fabrication microscopiques et inévitables à la surface de ses photosites en silicium. Ces variations de sensibilité à la lumière impriment un bruit déterministe de haute fréquence, invisible à l'œil nu, appelé PRNU (*Photo Response Non-Uniformity*). Ce bruit prévisible agit comme une véritable signature balistique ou "empreinte digitale" unique de l'appareil photo. La modification sémantique locale ou la superposition de contenus hétérogènes détruit ou altère ce motif mathématique, trahissant ainsi la falsification [5, 14].

1.2.2.5 La Corrélation Spatiale (Pixel Correlation)

Dans une scène naturelle authentique, il existe une forte dépendance statistique locale entre un pixel et son voisinage immédiat, dictée par la continuité physique des textures et de l'éclairage de la scène. Lorsqu'une attaque par copier-déplacer (*copy-move*) est exécutée, la duplication exacte d'un bloc de pixels propage également sa structure de corrélation interne. Comme illustré dans la Figure 1.1, cela crée une réplication de motifs statistiques hautement improbable et rompt brutalement la cohérence aux frontières du collage [6].

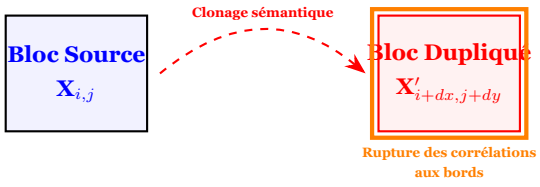


Figure 1.1: Principe de rupture de la corrélation spatiale lors d'une attaque par copier-déplacer.

1.2.2.6 Le Bruit de Capteur (PRNU)

Chaque capteur physique (CCD ou CMOS) présente des imperfections de fabrication microscopiques et inévitables à la surface de ses photosites en silicium. Ces variations de sensibilité à la lumière impriment un bruit déterministe de haute frequency, invisible à l'œil nu, appelé PRNU (*Photo Response Non-Uniformity*). Ce bruit agit comme une véritable signature balistique ou "empreinte digitale" unique de l'appareil photo [14]. La Figure 1.2 montre comment ce motif mathématique homogène (**K**) est brisé ou superposé de manière hétérogène lors d'une falsification, trahissant ainsi la manipulation [5].

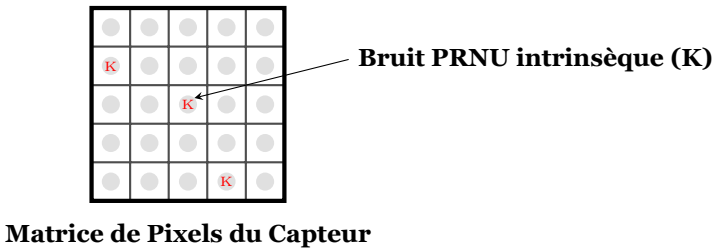


Figure 1.2: Cartographie conceptuelle du motif de bruit PRNU agissant comme empreinte unique.

1.3 Définition de la Falsification d'Images

Au sens large, la **falsification d'images** (*Image Forgery* en anglais) se définit comme l'acte de manipuler, modifier ou altérer une image numérique dans le but de tromper la perception du spectateur ou de falsifier la réalité d'un fait [4].

Modifier la matrice de pixels d'une image implique la rupture de deux propriétés juridiques et informationnelles fondamentales [2] :

1. **L'authenticité :** Garantir que le contenu de l'image correspond exactement à la scène réelle capturée à un instant t, sans aucun ajout, suppression ou modification sémantique [4].

2. **L'intégrité :** S'assurer que les données binaires et les métadonnées (EXIF) du fichier n'ont subi aucune altération malveillante depuis sa génération par l'appareil d'acquisition [2].

Il est crucial de distinguer la *falsification malveillante* de la *retouche standard*. Alors que le traitement d'images classique vise à optimiser le stockage ou la lisibilité visuelle sans altérer le sens de la scène, la falsification implique une **intention frauduleuse** [5].

Pour appréhender la diversité de ces attaques, la Figure 1.3 dresse la taxonomie globale des manipulations d'images, classées selon leur nature technique [4].

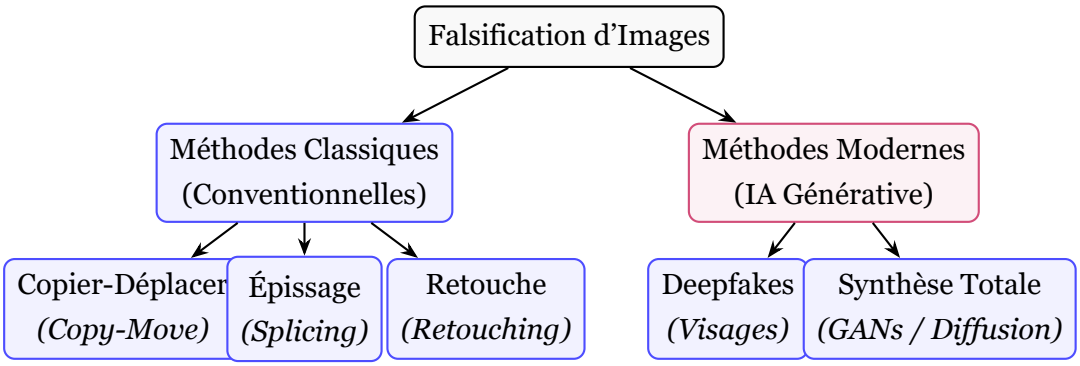


Figure 1.3: Taxonomie générale des différentes catégories de falsification d'images numériques.

1.3.1 Falsification par Copier-Déplacer (Copy-Move Forgery)

La falsification par copier-déplacer consiste à dupliquer une zone spécifique d'une image et à la coller sur une autre zone de la **même image** [6]. Comme illustré dans la Figure 1.4, ce processus est généralement utilisé pour masquer un élément compromettant (voir Figure 1.4a) ou pour simuler une densité plus élevée d'objets [6].

Étant donné que la région copiée provient de la même image, elle partage les mêmes propriétés statistiques et physiques, ce qui rend sa détection particulièrement difficile à l'œil nu, incitant le développement continu d'algorithmes de localisation basés sur l'apprentissage profond [7].



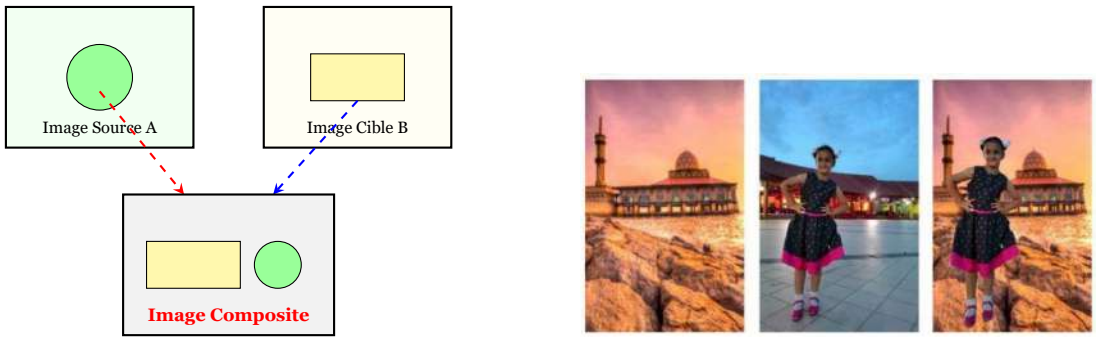
(a) Mécanisme théorique. (b) Exemple d'application.

Figure 1.4: Processus de la falsification par Copier-Déplacer (Copy-Move) : (a) schéma conceptuel du clonage de zone, (b) illustration pratique sur une image réelle [29].

1.3.2 Épissage d'Images (Image Splicing)

L'épissage d'images est une technique composite impliquant **au moins deux images distinctes** [8]. Elle consiste à découper une région d'une image source pour l'implanter dans une image cible afin de créer une scène totalement nouvelle. Comme illustré dans la Figure 1.5, ce processus combine des éléments théoriques (voir Figure 1.5a) pour aboutir à une falsification complète [8].

Contrairement au copier-déplacer, l'épissage introduit inévitablement des incohérences statistiques et physiques majeures (bruits de capteurs distincts, angles d'éclairage différents) qui constituent les indices clés exploités par les algorithmes de détection récents basés sur l'apprentissage auto-supervisé [9].



(a) Mécanisme théorique. (b) Exemple d'application.

Figure 1.5: Processus de la falsification par Épissage (Image Splicing) : (a) schéma conceptuel de la fusion de deux sources différentes, (b) illustration pratique du résultat falsifié [30].

1.3.3 Retouche d'Images (Image Retouching)

La retouche ne modifie pas la structure sémantique globale de l'image, mais altère ses propriétés intrinsèques telles que le contraste, la luminosité ou la saturation [10]. Comme illustré dans la Figure 1.6, cette manipulation technique devient une falsification frauduleuse lorsqu'elle est utilisée dans le but délibéré de dissimuler des détails critiques ou d'altérer la vérité factuelle d'un document légal [10].



Figure 1.6: Exemple d'altération des propriétés intrinsèques par retouche d'image [31].

1.3.4 Falsification Générative (Deepfakes et GANs)

Représentant la forme la plus moderne de falsification, cette approche repose sur l'intelligence artificielle, notamment les réseaux antagonistes génératifs (GANs) et les modèles de diffusion [11]. Comme illustré dans la Figure 1.7, elle permet de générer des images synthétiques hautement réalistes de toutes pièces ou d'altérer des identités via le remplacement de visages (*Deepfakes*) [12]. Les avancées récentes de

ces modèles génératifs rendent les frontières entre le réel et le synthétique presque imperceptibles à l'œil nu [3].



Figure 1.7: Exemple de falsification par génération d'image (Deepfake) [32].

1.4 État de l'Art des Techniques de Détection de falsification

Pour contrer les différentes attaques de falsification, la communauté scientifique a développé de nombreuses méthodes de détection au fil des deux dernières décennies [2]. L'évolution de ces techniques suit une trajectoire parallèle à celle des technologies de manipulation [5].

Globalement, l'état de l'art actuel peut être structuré en trois grandes vagues technologiques majeures [4] :

1.4.1 Première génération : Approches physiques et statistiques (2000–2010)

Les premiers travaux de recherche se sont concentrés sur l'extraction d'anomalies physiques ou statistiques générées lors de la capture ou de la sauvegarde de l'image [1].

- **Analyse des artefacts JPEG :** [13] ont été les pionniers de la détection du double codage JPEG, démontrant que chaque ré-enregistrement d'une image laisse des traces spécifiques dans les coefficients DCT.
- **Analyse du bruit de capteur :** [14] ont introduit l'utilisation du bruit d'inefficacité de réponse photo-numérique (PRNU). Cette méthode souffre toutefois d'une forte sensibilité aux opérations de filtrage.

1.4.2 Deuxième génération : Approches par Apprentissage Automatique (2010–2018)

Avec l'introduction de manipulations plus complexes, les chercheurs ont combiné l'extraction de descripteurs de texture (*Handcrafted features*) avec des classifieurs statistiques [10].

- **Descripteurs locaux pour le Copier-Déplacer :** Des algorithmes basés sur *SIFT* [15] et *SURF* [16] ont été largement adoptés pour faire correspondre des points d'intérêt clés au sein d'une même image.
- **Modélisation du bruit local :** L'utilisation des motifs binaires locaux (*LBP*) a permis de détecter les ruptures de texture sur les contours des zones insérées [10].

1.4.3 Troisième génération : L'ère de l'Apprentissage Profond (2018–Présent)

L'état de l'art actuel est dominé par les réseaux de neurones profonds, capables d'apprendre simultanément les caractéristiques sémantiques et les artefacts microscopiques [21].

- **Réseaux convolutifs bi-flux (Two-Stream CNNs) :** [17] ont introduit une architecture à double flux : un flux extrait les caractéristiques visuelles (RVB) tandis que le second flux extrait les caractéristiques du bruit de l'image.
- **Réseaux de bout en bout (End-to-End Localizers) :** Des modèles comme *ManTra-Net* [18] ou *MVSS-Net* [19] traitent la détection comme un problème de segmentation sémantique au niveau du pixel.
- **Vision Transformers (ViTs) et Modèles Multimodaux :** Récemment, les architectures basées sur les mécanismes d'attention (*Transformers*) sont utilisées pour capturer les relations de dépendance à longue distance [20]. Les tendances de pointe s'orientent désormais vers des approches mutimodales combinant vision et langage pour déceler les incohérences contextuelles profondes [20].

Table 1.1: Synthèse comparative des approches de l'état de l'art.

Approche	**Avantages**	**Limites**
Traditionnelle [13]	Pas besoin de base de données d'entraînement ; explicabilité physique.	Échoue après compression sévère ou redimensionnement sur le Web.
Machine Learning [15]	Robuste aux transformations géométriques simples.	Sensible aux zones sans texture ; extraction manuelle fastidieuse.
Deep Learning [17, 20]	Excellente précision ; localisation automatique au niveau du pixel.	Nécessite de grandes ressources de calcul et de vastes datasets.

1.5 Bases de données de référence (CASIA, Columbia, etc.)

Les bases de données de référence jouent un rôle capital dans la détection de falsifications. Elles offrent un cadre expérimental standardisé indispensable pour former, valider et évaluer l'efficacité des modèles [21].

1.5.1 Introduction et critères d'évaluation

Dans la mesure où les systèmes de détection modernes s'appuient massivement sur l'apprentissage profond (Deep Learning), la disponibilité de bases de données massives est devenue le socle critique de l'évaluation expérimentale [22]. À titre d'exemple, le dataset *FaceForensics++*, introduit par [22], s'est imposé comme l'une des références majeures pour l'analyse des manipulations faciales. Celles-ci ont été complétées récemment par de nouveaux benchmarks évaluant la robustesse face aux altérations d'images générées par diffusion [23].

1.5.2 Datasets classiques de falsification d'images

Avant l'avènement de l'IA générative, les bases de données classiques ont joué un rôle déterminant dans le développement de modèles robustes [24].

1.5.2.1 CASIA Image Tampering Detection Evaluation Database (v2.0)

La base de données CASIA v2.0 [24] comprend 12 614 images, réparties entre 7 491 images authentiques et 5 123 images falsifiées (voir Figure 1.8).



Figure 1.8: Exemple de falsification par *copier-déplacer* (copy-move) tiré de la base de données CASIA v2.0 [33].

1.5.2.2 COLUMBIA Image Splicing Dataset

Le dataset Columbia [25] contains 363 images non compressées, réparties de manière équilibrée entre images authentiques et composites (voir Figure 1.9).

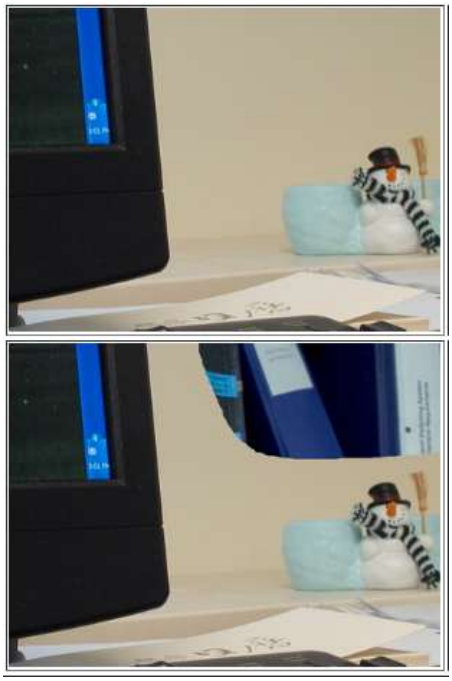


Figure 1.9: Exemple de falsification par *épissage* (splicing) tiré de la base de données Columbia [34].

1.5.3 Datasets spécifiques aux deepfakes et contenus génératifs

Le dataset *Celeb-DF (v2)* a été spécifiquement conçu par [26] pour pallier le manque de qualité visuelle observée dans les premiers jeux de données. De son côté, le *Deep-Fake Detection Challenge (DFDC)*, publié par [27], représente l'un des corpus les plus massifs avec plus de 124 000 séquences vidéo (voir Figure 1.10).

Table 1.2: Comparaison des principaux datasets dédiés aux Deepfakes.

Dataset	Volume (Vidéos)	Méthodes de génération sous-jacentes	Niveau de difficulté
FaceForensics++	1 000+	DeepFakes, FaceSwap, Face2Face	Moyenne [22]
Celeb-DF v2	590	GANs avancés à haute fidélité visuelle	Élevée [26]
DFDC	124 000+	Méthodes composites	Élevée [27]



Figure 1.10: Exemples comparatifs de visages synthétisés extraits des benchmarks de Deepfakes [35].

1.6 Métriques d'Évaluation

L'évaluation quantitative des performances des modèles de détection repose fondamentalement sur l'analyse de la matrice de confusion (voir Figure 1.11) [28].

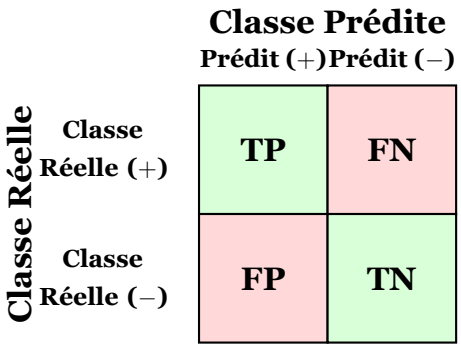


Figure 1.11: Matrice de confusion standardisée pour l'extraction des métriques de classification.

À partir de cette répartition matricielle, trois métriques clés sont exploités pour mesurer le taux de succès et d'erreurs [28] :

$$\text{Précision} = \frac{TP}{TP + FP} \tag{1.2}$$

$$\text{Rappel} = \frac{TP}{TP + FN} \tag{1.3}$$

$$F_1\text{-score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{1.4}$$

Outre les métriques scalaires fixes, l'analyse de la courbe ROC (*Receiver Operating Characteristic*) et le calcul de l'aire sous la courbe (AUC) sont indispensables pour évaluer la robustesse globale [28]. La courbe ROC (voir Figure 1.12) représente le Taux de Vrais Positifs (TVP) en fonction du Taux de Faux Positifs (TFP) [28] :

$$TFP = \frac{FP}{FP + TN} \tag{1.5}$$

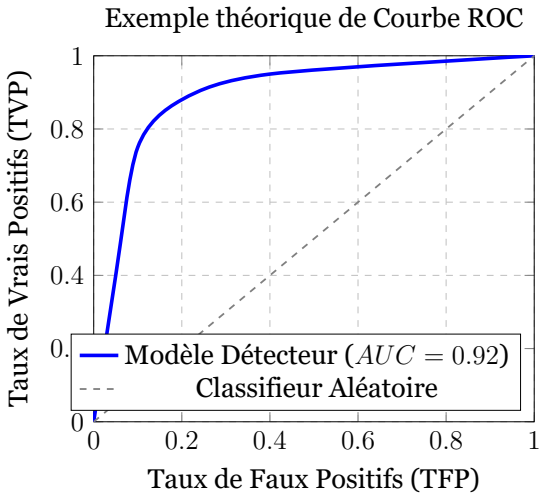


Figure 1.12: Évaluation de la robustesse via l'analyse de l'espace ROC.

1.7 Conclusion du Chapitre

En conclusion, ce chapitre a mis en lumière la structure d'une image saine ainsi que la diversité des attaques liées à sa falsification [4]. La sophistication croissante de ces manipulations et les limites identifiées dans l'état de l'art motivent directement le développement de notre propre architecture neuronale robuste, qui sera détaillée dans le chapitre suivant [5].

Bibliography

[1] Farid, H. (2009). Image forgery detection. *IEEE Signal Processing Magazine*, 26(2), 16–25.

[2] Piva, A. (2013). An overview on image forensics. *ISRN Signal Processing*, 2013, 1–22.

[3] Verdoliva, L., Cozzolino, D., and Marra, F. (2024). Media forensics in the age of generative AI: Challenges and opportunities. *IEEE Signal Processing Magazine*, 41(2), 34–45.

[4] Reddy, V., Singh, S., and Kumar, R. (2022). Digital image forensics: A review of latent features and deep learning approaches. *Forensic Science International*, 331, 111135.

[5] Barni, M., Ehret, T., and Albisani, A. (2025). A holistic framework for image forgery detection in the era of generative AI. *IEEE Transactions on Information Forensics and Security*, 20, 145–159.

[6] Christlein, V., Riess, C., Jordan, J., Angelopoulou, E., and Deuker, G. (2012). An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on Information Forensics and Security*, 7(6), 1841–1854.

[7] Wu, H., Jiang, L., and Feng, J. (2024). Robust copy-move forgery detection using deep spatial-frequency attention networks. *IEEE Transactions on Image Processing*, 33, 1120–1134.

[8] Ng, T. T., Shih, S. F., and Chang, S. F. (2004). A model for image splicing. In *IEEE International Conference on Image Processing* (pp. 223–226). IEEE.

[9] Horváth, J., Revaud, J., and Harchaoui, Z. (2025). Self-supervised localization of image splicing via boundary artifact learning. *International Journal of Computer Vision*, 133(3), 512–529.

[10] Mahdian, B., and Saic, S. (2010). A bibliography on blind methods for identifying image forgery. *Signal Processing: Image Communication*, 25(6), 389–399.

[11] Goodfellow, I., Pouget-Abadie, J., and Mirza, M. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.

[12] Jia, S., Zhang, T., and Liu, J. (2024). A comprehensive survey on generative image forgery and its detection techniques. *ACM Computing Surveys*, 56(8), 1–39.

[13] Popescu, A. C., and Farid, H. (2005). Exposing digital forgeries by detecting traces of resampling. *IEEE Transactions on Signal Processing*, 53(2), 758–767.

[14] Lukáš, J., Fridrich, J., and Goljan, M. (2006). Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2), 205–214.

[15] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.

[16] Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *European conference on computer vision* (pp. 404–417). Springer.

[17] Zhou, P., Han, X., and Davis, L. S. (2018). Learning rich features for image manipulation detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1053–1061).

[18] Wu, Y., AbdAlmageed, W., and Natarajan, P. (2019). Mantra-net: Manipulation tracing network for detection and localization of image forgeries. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9544–9553).

[19] Chen, X., Yao, C., and Sharkolkhi, M. (2021). MVSS-Net: Multi-view multi-scale supervised network for image forgery detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(2), 981–995.

[20] Shiohara, K., and Yamasaki, T. (2025). Multi-modal deepfake detection via joint vision-language alignment models. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(1), 204–218.

[21] Verdoliva, L. (2020). Media forensics for deepfake detection. *IEEE Signal Processing Magazine*, 37(5), 97–108.

[22] Rössler, A., Cozzolino, D., and Niessner, M. (2019). FaceForensics++: Learning to detect manipulated facial images. In *IEEE/CVF International Conference on Computer Vision* (pp. 1–11).

[23] Cozzolino, D., Marra, F., and Verdoliva, L. (2024). Raising the bar for generative image forensics: A new benchmark dataset. *IEEE Transactions on Information Forensics and Security*, 19, 889–903.

[24] Dong, J., Wang, W., and Tan, T. (2013). CASIA image tampering detection evaluation database. In *IEEE International Workshop on Information Forensics and Security* (pp. 422–427). IEEE.

[25] Ng, T. T., Chang, S. F., and Sun, Q. (2004). Columbia open image splicing detection evaluation dataset. *ADVENT Technical Report*, Columbia University, 23–34.

[26] Li, Y., Chang, X., and Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for deepfake forensics. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3207–3216).

[27] Dolhansky, B., Bitton, J., and Ferrer, C. C. (2020). The deepfake detection challenge (DFDC) dataset. *arXiv preprint arXiv:2006.07397*.

[28] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.

[29] Samariya, U., et al. (2024). *A survey on copy-move image forgery detection based on deep-learning techniques*. Multimedia Tools and Applications.

[30] Alamro, L., & Yusoff, N. (2017). *Copy-move forgery detection using integrated DWT and SURF*.

[31] Maire, L. (s.d.). *Formation retouche photographie animalière complète*.

[32] Radio France (2026). *Arnaque grâce au deepfake : le mode d'emploi*. Podcast France Bleu - Ici Occitanie.

[33] Dong, J., Wang, W., & Tan, T. (2013). *CASIA Image Tampering Detection Evaluation Database*. IEEE International Conference on Computer Vision Workshops (ICCVW).

[34] Ng, T. T., Chang, S. F., & Sun, Q. (2005). *A data set of authentic and spliced images*. Columbia University, ADVENT Technical Report #206-2005-12.

[35] Benoit, F. (2019). *Fausses vidéos : extension du domaine du fake*. Usbek & Rica.

Chapter 2
Transformée en Ondelettes Discrète (DWT)

2.1 Introduction . . . 20
2.2 Du Formalisme de Fourier aux Ondelettes : Perspective Historique . . . 20
2.2.1 Transformée de Fourier à Fenêtre (STFT) et Principe d'Incertitude . . . 20
2.2.2 Paradigme des Ondelettes . . . 21
2.3 Fondements Mathématiques de la DWT . . . 22
2.3.1 Ondelette Mère et Condition d'Admissibilité . . . 22
2.3.2 Discrétisation Dyadique . . . 22
2.4 Analyse Multirésolution (MRA) de Mallat . . . 23
2.5 Algorithme de Banc de Filtres de Mallat . . . 23
2.5.1 Analyse (Décomposition du Signal) . . . 23
2.5.2 Synthèse (Reconstruction du Signal) . . . 23
2.5.3 Architecture des Bancs de Filtres . . . 23
2.6 Familles d'Ondelettes Usuelles et Propriétés Structurelles . . . 24
2.6.1 Ondelette de Haar . . . 24
2.6.2 Ondelettes de Daubechies (dbN) et Symlets ($symN$) . . . 25
2.6.3 Ondelettes Biorthogonales ($biorN.M$) . . . 25
2.7 Transformée en Ondelettes Discrète Bidimensionnelle (2D-DWT) . . . 26
2.8 Applications Phares de la 2D-DWT . . . 27
2.8.1 Compression d'Images : Standard JPEG 2000 . . . 27
2.8.2 Débruitage par Seuillage (Wavelet Thresholding) . . . 28
2.8.3 Stéganographie et Tatouage Numérique (Watermarking) . . . 28
2.8.4 Analyse Légiste des Images (Image Forensics) à l'Ère de l'IA . . . 28
2.9 Limitations Intrinsèques de la DWT . . . 28
2.9.1 Sensibilité à la Translation (Shift Sensitivity) . . . 29
2.9.2 Manque de Sélectivité Directionnelle . . . 29
2.10 Conclusion du Chapitre . . . 29

Résumé

Ce chapitre est dédié à l'étude approfondie de la Transformée en Ondelettes Discrète (DWT) et de ses applications dans le traitement et l'analyse d'images. Nous y explorons d'abord les fondements historiques et mathématiques qui ont permis de surmonter les limitations de la transformée de Fourier traditionnelle face aux signaux non stationnaires. Ensuite, nous détaillons l'Analyse Multirésolution (MRA) et l'algorithme pyramidal de Mallat, basé sur des bancs de filtres miroirs en quadrature (QMF). Le passage au domaine bidimensionnel (2D-DWT) est analysé pour montrer comment une image est segmentée en sous-bandes d'approximation (LL) et de détails directionnels (LH, HL, HH). Enfin, nous passons en revue les applications phares de ce formalisme, telles que la compression (standard JPEG 2000), le débruitage par seuillage, la stéganographie et la détection de falsification d'images (Image Forensics), tout en discutant des limitations intrinsèques liées à la sensibilité à la translation.

Keywords: Transformée en Ondelettes Discrète (DWT), Analyse Multirésolution (MRA), Algorithme de Mallat, Banc de filtres, Traitement d'images, JPEG 2000, Débruitage, Image Forensics.

2.1 Introduction

Dans le domaine du traitement numérique du signal et des images, l'extraction de caractéristiques pertinentes, la localisation des structures transitoires et l'analyse fréquentielle constituent des jalons cardinaux pour les architectures de classification, de filtrage et de compression. Pendant plusieurs décennies, la transformée de Fourier rapide (FFT) a régné en paradigme hégémonique [1]. Toutefois, ce formalisme classique souffre d'une rigidité intrinsèque : en projetant le signal sur des bases sinusoïdales infinies, il occulte totalement la dynamique temporelle ou spatiale, s'avérant inadéquat pour l'analyse des signaux non stationnaires, des singularités isolées ou des images aux contours abrupts.

Bien que la transformée de Fourier à fenêtre (STFT) ait tenté d'introduire une localisation spatio-fréquentielle [3], le principe d'incertitude de Heisenberg-Gabor en a drastiquement borné la flexibilité par une résolution fixe. Pour s'affranchir de ces verrous théoriques, la **Transformée en Ondelettes Discrète (DWT)** s'est imposée comme une alternative mathématique rigoureuse et élégante. En substituant le concept d'analyse multirésolution (MRA) aux fenêtres statiques [5], elle permet une inspection adaptative du signal : une haute résolution temporelle pour les composantes à haute fréquence (contours, transitoires) et une haute résolution fréquentielle pour les structures à basse fréquence (textures lisses, tendances globales).

Ce chapitre a pour vocation de dresser une cartographie exhaustive de la DWT, en explorant ses fondements mathématiques sous-jacents, l'algorithme pyramidal de Mallat basé sur les bancs de filtres miroirs en quadrature (QMF), ainsi que son extension de dimension spatiale (2D-DWT). Enfin, nous mettrons en relief la polyvalence de ce formalisme à travers ses applications phares — allant de la compression standardisée (JPEG 2000) à l'analyse légiste des images (Image Forensics) face aux défis de l'intelligence artificielle générative —, sans omettre de discuter ses limitations structurelles telles que la sensibilité à la translation.

2.2 Du Formalisme de Fourier aux Ondelettes : Perspective Historique

2.2.1 Transformée de Fourier à Fenêtre (STFT) et Principe d'Incertitude

La première tentative historique pour localiser les fréquences dans le temps fut la *Short-Time Fourier Transform* (STFT), introduite initialement par Gabor [3], qui applique une fenêtre temporelle fixe $g(t)$. Cependant, selon le **principe d'incertitude**

de Heisenberg-Gabor, il est impossible d'obtenir simultanément une résolution infinie en temps et en fréquence. Le produit des résolutions temporelle (Δt) et fréquentielle (Δf) est minoré par :

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \tag{2.1}$$

Une fenêtre étroite offre une excellente résolution temporelle mais une mauvaise résolution fréquentielle, tandis qu'une fenêtre large produit l'effet inverse.

2.2.2 Paradigme des Ondelettes

La DWT résout ce dilemme en ajustant dynamiquement la taille de la fenêtre d'analyse :

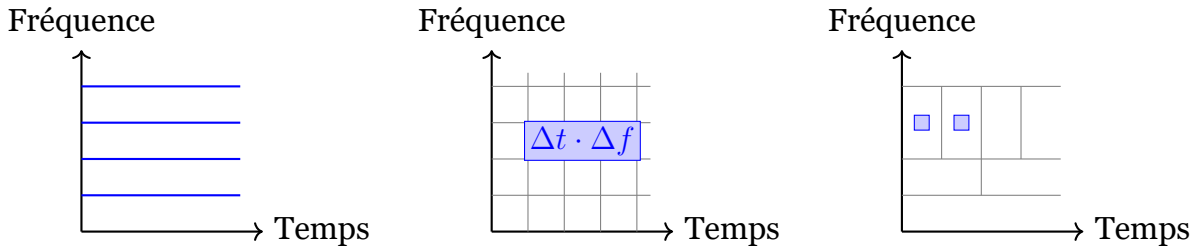
- Les **hautes fréquences** sont analysées via des fenêtres temporelles **étroites** (bonne localisation des transitoires et des contours).
- Les **basses fréquences** sont analysées via des fenêtres temporelles **larges** (bonne estimation de la tendance générale du signal).

Le tableau 2.1 synthétise les propriétés fondamentales de ces approches.

Table 2.1: Comparaison structurelle entre les transformées de Fourier, STFT et DWT.

Propriété	**Fourier (FT)**	**STFT**	**DWT**
Fonctions de base	Sinusoïdes infinies	Sinusoïdes fenêtrées	Ondelettes dilatées/translatées
Résolution	Fréquentielle pure	Fixe ($\Delta t, \Delta f$ constants)	Variable (Multirésolution)
Stationnarité	Oui	Strictement locale	Non (Idéal pour non-stationnaire)
Transitoires	Invisibles	Mal localisés	Parfaitement localisés

La figure 2.1 illustre le découpage de l'espace temps-fréquence pour ces trois méthodes.



(a) Transformée de Fourier (b) STFT (Fenêtre fixe) (c) DWT (Multirésolution)

Figure 2.1: Pavage de l'espace temps-fréquence (Diagrammes de l'espace des phases).

2.3 Fondements Mathématiques de la DWT

2.3.1 Ondelette Mère et Condition d'Admissibilité

Une ondelette est une fonction $\psi(t) \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, de moyenne nulle, oscillante et à support borné [4]. Elle doit satisfaire la condition d'admissibilité suivante :

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < +\infty \tag{2.2}$$

où $\Psi(\omega)$ désigne la transformée de Fourier de $\psi(t)$. Cette condition implique mathématiquement que :

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = 0 \tag{2.3}$$

2.3.2 Discrétisation Dyadique

Pour éliminer la redondance de la formulation continue, la DWT restreint les paramètres d'échelle a et de translation b à une grille géométrique dyadique. Ainsi, la famille d'ondelettes discrètes engendrée s'exprime par :

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t - k \cdot 2^j}{2^j}\right) = 2^{-j/2} \psi(2^{-j}t - k) \tag{2.4}$$

Parallèlement à la fonction d'ondelette, une **fonction d'échelle** (ou fonction père) $\phi(t)$ est introduite pour capturer les composantes basses fréquences du signal :

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k) \tag{2.5}$$

2.4 Analyse Multirésolution (MRA) de Mallat

Le cadre théorique de l'Analyse Multirésolution, formalisé par **Stéphane Mallat** [6], permet de structurer l'espace des signaux de carré sommable $L^2(\mathbb{R})$ en une suite d'espaces vectoriels fermés emboîtés $\{V_j\}_{j\in\mathbb{Z}}$.

L'espace V_j représente l'espace d'approximation à la résolution 2^j. Le complément orthogonal de V_j dans V_{j-1} est noté W_j, tel que :

$$V_{j-1} = V_j \oplus W_j \tag{2.6}$$

Où W_j est l'espace des détails contenant l'information perdue lors du passage d'une haute résolution à une résolution inférieure.

2.5 Algorithme de Banc de Filtres de Mallat

2.5.1 Analyse (Décomposition du Signal)

Le processus d'analyse consiste à diviser le signal d'entrée pour en extraire les composants essentiels à travers deux filtres parallèles :

- **Un filtre passe-bas :** Qui génère les coefficients d'approximation.
- **Un filtre passe-haut :** Qui capture les coefficients de détail.

Après chaque filtrage, on applique une étape de décimation ($\downarrow 2$), qui consiste à supprimer un échantillon sur deux pour optimiser le stockage.

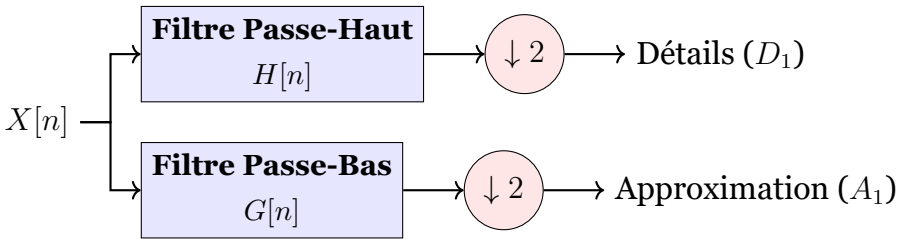
2.5.2 Synthèse (Reconstruction du Signal)

Le processus de synthèse est le miroir exact de l'analyse. Il permet de reconstruire le signal d'origine parfait et sans perte à partir des coefficients d'approximation et de détail via :

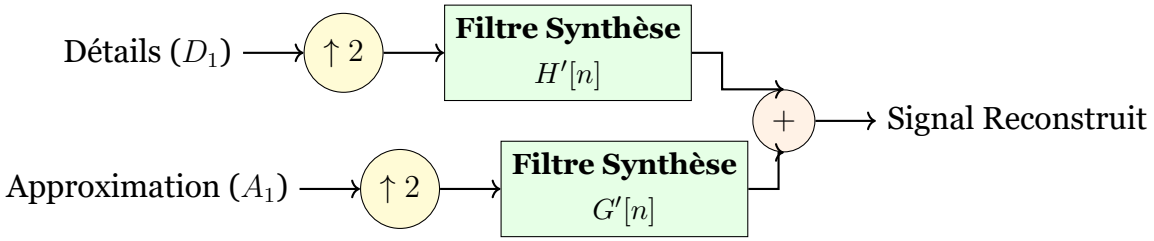
- **Le suréchantillonnage ($\uparrow 2$) :** Insertion d'un zéro entre chaque échantillon.
- **Le filtrage de synthèse :** Passage par des filtres duaux (passe-bas et passe-haut) pour lisser le signal avant sommation.

2.5.3 Architecture des Bancs de Filtres

La figure 2.2 illustre graphiquement la parfaite symétrie et le parallélisme entre le bloc de décomposition (Analyse) et le bloc de reconstruction (Synthèse).



(a) Banc de filtres d'Analyse (Décomposition)



(b) Banc de filtres de Synthèse (Reconstruction)

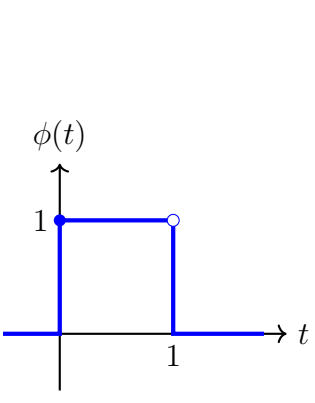
Figure 2.2: Architecture algorithmique duale et symétrique de l'algorithme de Mallat.

2.6 Familles d'Ondelettes Usuelles et Propriétés Structurelles

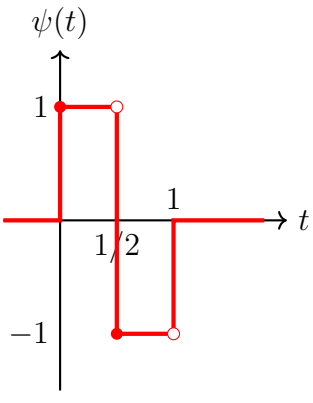
Le choix de l'ondelette repose sur un compromis mathématique entre plusieurs propriétés fondamentales : le support compact, le nombre de moments évanouissants et la régularité.

2.6.1 Ondelette de Haar

Introduction historique (1910) [8], elle constitue le système orthogonal le plus simple basé sur des fonctions portes discontinues (Figure 2.3).



(a) Fonction d'échelle $\phi(t)$ (Haar)

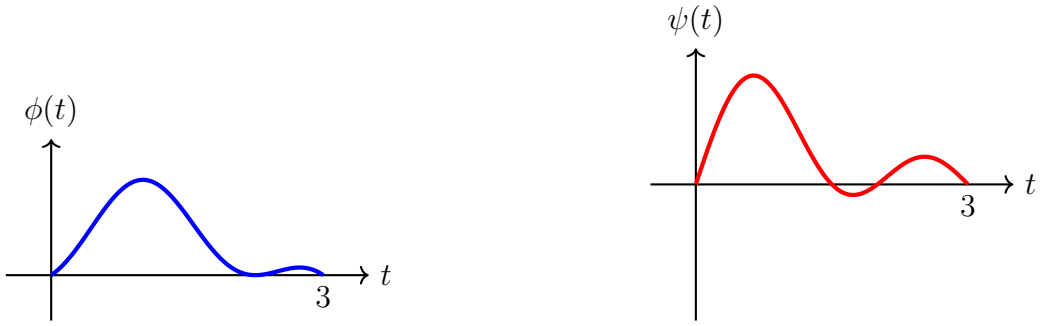


(b) Ondelette mère $\psi(t)$ (Haar)

Figure 2.3: Représentation temporelle des fonctions de base de l'ondelette de Haar.

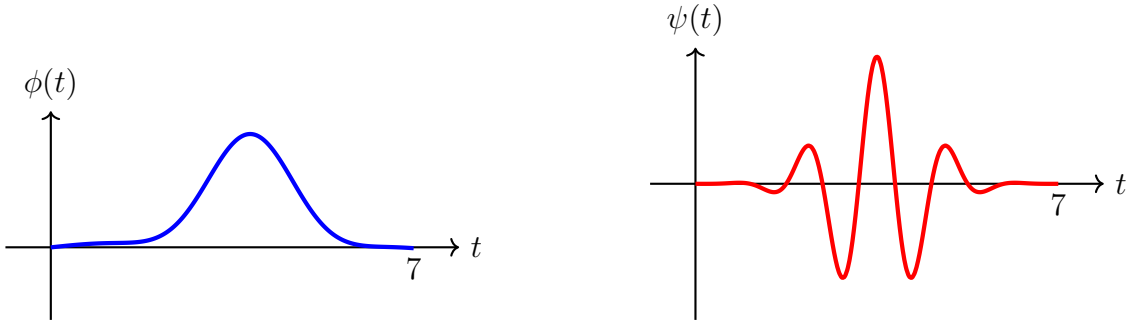
2.6.2 Ondelettes de Daubechies (dbN) et Symlets ($symN$)

Développées par Ingrid Daubechies [9], elles offrent un support compact optimal (Figure 2.4). Les Symlets corrigent la phase pour la rendre presque linéaire (Figure 2.5).



(a) Fonction d'échelle $\phi(t)$ (db2) (b) Ondelette mère $\psi(t)$ (db2)

Figure 2.4: Forme d'onde asymétrique caractéristique de la famille Daubechies (db2).

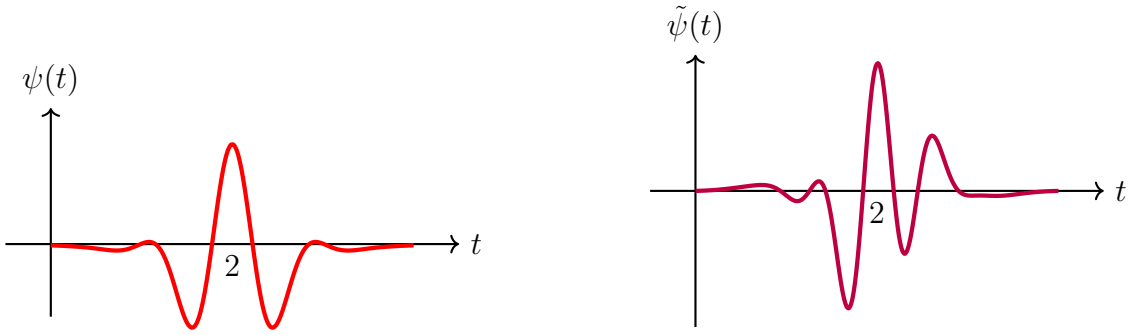


(a) Fonction d'échelle $\phi(t)$ (sym4) (b) Ondelette mère $\psi(t)$ (sym4)

Figure 2.5: Propriétés de symétrie quasi-parfaite de la famille Symlet (sym4).

2.6.3 Ondelettes Biorthogonales ($biorN.M$)

Elles séparent les filtres d'analyse de ceux de synthèse pour obtenir une symétrie parfaite associée à des filtres à phase linear exacte (Figure 2.6), largement exploitée dans la norme JPEG 2000 [10].



(a) Ondelette d'analyse $\psi(t)$ (bior4.4) (b) Ondelette de synthèse $\tilde{\psi}(t)$ (bior4.4)

Figure 2.6: Symétrie parfaite des fonctions d'ondelettes duales de la famille biorthogonale bior4.4.

2.7 Transformée en Ondelettes Discrète Bidimensionnelle (2D-DWT)

Pour le traitement d'images, la DWT est appliquée séparément le long des lignes puis des colonnes (séparabilité bidimensionnelle). Cette opération segmente l'image en quatre sous-bandes distinctes à chaque niveau : **LL** (Approximation), **LH** (Détails horizontaux), **HL** (Détails verticaux), et **HH** (Détails diagonaux).

Le raffinement algorithmique moderne utilise également des variantes avancées de la 2D-DWT pour pallier le manque de sélectivité directionnelle classique [14]. De nos jours, ces architectures sous-tendent les modèles de vision les plus récents et les pipelines d'analyse légiste d'images numériques (Image Forensics) pour la détection de falsifications générées par IA [15].

Le processus algorithmique et le résultat pratique de cette décomposition sont illustrés respectivement dans les Figures 2.7 et 2.8.

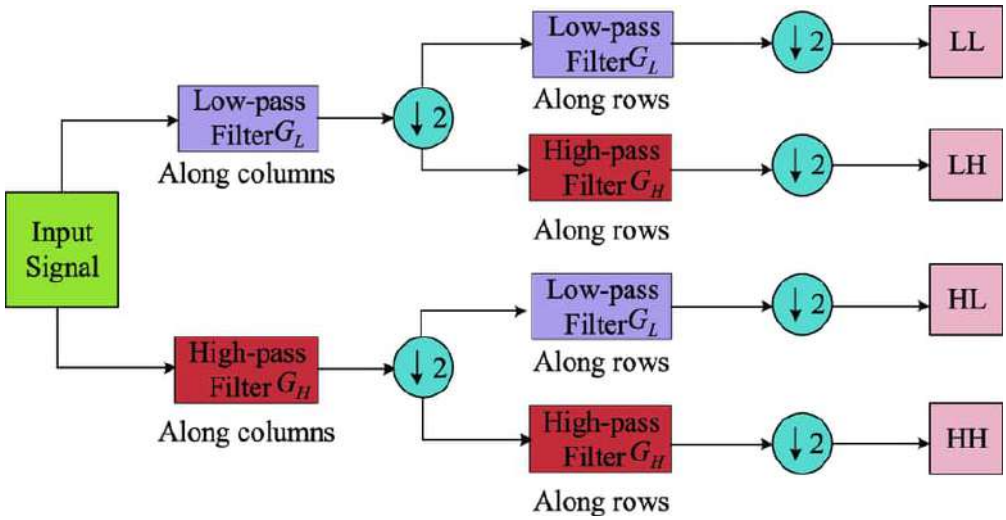


Figure 2.7: Schéma fonctionnel du banc de filtres bidimensionnel (2D-DWT) avec décimation par lignes et colonnes [17]

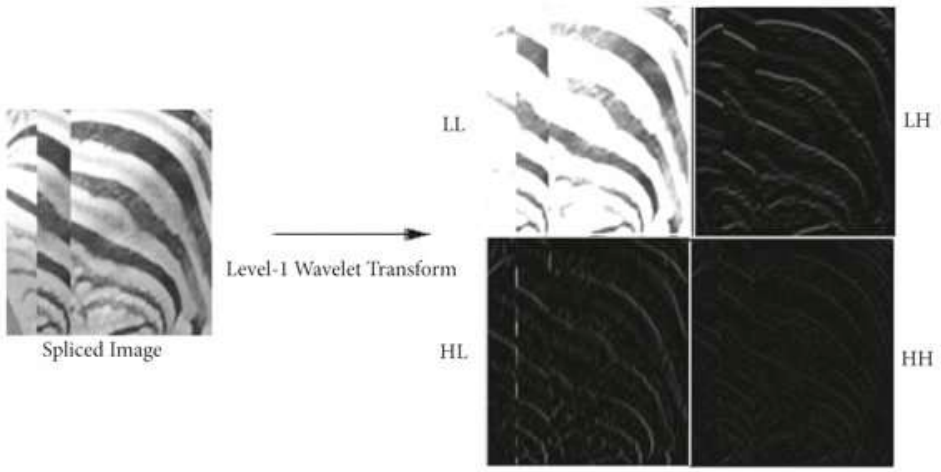


Figure 2.8: Décomposition d'une image en sous-bandes d'approximation (LL) et de détails directionnels (LH, HL, HH).[16]

2.8 Applications Phares de la 2D-DWT

La capacité de la 2D-DWT à isoler les composantes fréquentielles tout en préservant la géométrie spatiale des contours en a fait un outil incontournable dans plusieurs branches du traitement d'images modernes.

2.8.1 Compression d'Images : Standard JPEG 2000

Contrairement au standard JPEG classique basé sur la transformée en cosinus discrète (DCT) par blocs de 8×8, le standard JPEG 2000 exploite la DWT globale

sur toute l'image [10]. Ce choix algorithmique élimine définitivement l'effet de bloc ("blocking artifacts") à haut taux de compression. Le pipeline utilise principalement l'ondelette biorthogonale de CDF 9/7 pour une compression avec perte (Lossy), ou CDF 5/3 pour une compression sans perte (Lossless) [11].

2.8.2 Débruitage par Seuillage (Wavelet Thresholding)

Le principe de débruitage formulé initialement par Donoho [12] repose sur le fait que le bruit blanc gaussien se répartit uniformément sur tous les coefficients de détail, tandis que l'information utile (les contours) l'est concentrée sur quelques coefficients de forte amplitude. Deux approches de seuillage sont classiquement appliquées :

- **Seuillage dur (Hard Thresholding) :** Les coefficients inférieurs au seuil λ sont annulés, les autres restent intacts.
- **Seuillage doux (Soft Thresholding) :** Les coefficients inférieurs à λ sont annulés, et les coefficients restants sont réduits de la valeur de λ, ce qui lisse le signal mais peut atténuer les transitions abruptes.

2.8.3 Stéganographie et Tatouage Numérique (Watermarking)

L'insertion d'une signature ou d'un message caché dans les sous-bandes de fréquences moyennes ou hautes (LH, HL) offre un excellent compromis entre l'impératif d'indétectabilité visuelle (invisibilité) et la robustesse face aux attaques géométriques ou de compression [13].

2.8.4 Analyse Légiste des Images (Image Forensics) à l'Ère de l'IA

Avec l'avènement des images synthétiques hautement réalistes générées par les réseaux de neurones génératifs (GANs, Modèles de Diffusion), la DWT est devenue une arme de choix pour la détection de faux (Deepfakes). Les algorithmes génératifs laissent des anomalies périodiques subtiles dans le domaine fréquentiel, souvent invisibles à l'œil nu, mais parfaitement détectables sous forme d'artefacts d'échantillonnage dans la sous-bande de haute fréquence HH [15].

2.9 Limitations Intrinsèques de la DWT

Malgré ses performances rigoureuses, la DWT souffre de deux inconvénients majeurs liés à sa structure même :

2.9.1 Sensibilité à la Translation (Shift Sensitivity)

À cause de l'étape de décimation (downsampling $\downarrow 2$), une infime translation du signal d'entrée peut induire une redistribution drastique et imprévisible de l'énergie entre les coefficients de calcul des différentes sous-bandes [14]. Cette absence d'invariance par translation complique les tâches de reconnaissance de formes ou de suivi d'objets.

2.9.2 Manque de Sélectivité Directionnelle

Les filtres séparables de la 2D-DWT classique ne permettent pas de distinguer efficacement les orientations obliques. La sous-bande HH, par exemple, mélange confusément les détails orientés à $+45°$ et $-45°$. Pour surmonter cette barrière, des extensions complexes comme la transformée en ondelettes duale arborescente (*Dual-Tree Complex Wavelet Transform - DTCWT*) ont été introduites afin de restaurer l'invariance par translation et d'obtenir une sélectivité directionnelle riche sur six orientations distinctes [14].

2.10 Conclusion du Chapitre

En somme, ce chapitre a permis de décortiquer les articulations théoriques, algorithmiques et pratiques qui gouvernent la Transformée en Ondelettes Discrète (DWT). En rupture avec le réductionnisme de Fourier, le paradigme des ondelettes offre une vision holistique et adaptative de l'espace temps-fréquence (ou espace-fréquence pour les images), matérialisée par l'Analyse Multirésolution (MRA) de Stéphane Mallat [6]. Nous avons mis en évidence la rigueur algorithmique de cette approche, où la décomposition et la reconstruction parfaites d'un signal ne requièrent plus de formulations intégrales lourdes, mais s'opèrent via des bancs de filtres numériques QMF duaux, optimisés par décimation.

L'étude de la projection du domaine bidimensionnel (2D-DWT) a révélé l'efficacité de la séparation en sous-bandes d'approximation (LL) et de détails directionnels (LH, HL, HH). Cette décomposition géométrique s'est révélée cruciale pour des applications hautement exigeantes : elle permet au standard JPEG 2000 de s'affranchir des artefacts de bloc [10], offre au débruitage par seuillage une sélectivité préservant les contours [12], et confère à l'analyse légiste (Image Forensics) la capacité de détecter les anomalies fréquentielles microscopiques laissées par les générateurs d'images d'IA modernes [15].

Néanmoins, l'analyse critique de la DWT a mis en exergue des limitations structurelles non négligeables, au premier rang desquelles figure la sensibilité à la translation induite par le sous-échantillonnage, ainsi qu'un manque de discrimination dans

les directions obliques. Bien que des variantes avancées comme la transformée complexe arborescente (DT-CWT) tente de pallier ces barrières [14], la DWT classique demeure un pilier incontournable. Les concepts de décomposition et de filtrage spatial multi-échelle ainsi formalisés établissent les fondations méthodologiques indispensables qui jalonneront la suite de nos travaux et l'implémentation de nos propres modèles de traitement.

Bibliography

[1] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex fourier series," *Mathematics of Computation*, vol. 19, no. 90, pp. 297–301, 1965.

[2] A. Grossmann and J. Morlet, "Decomposition of hardy functions into square integrable wavelets of constant shape," *SIAM Journal on Mathematical Analysis*, vol. 15, no. 4, pp. 723–736, 1984.

[3] D. Gabor, "Theory of communication. part 1: The analysis of information," *Journal of the Institution of Electrical Engineers-Part III*, vol. 93, no. 26, pp. 429–441, 1946.

[4] Y. Meyer, *Wavelets: Algorithms & Applications*. SIAM, 1992.

[5] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[6] S. G. Mallat, "Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$," *Transactions of the American Mathematical Society*, vol. 315, no. 1, pp. 69–87, 1989.

[7] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Pearson Education India, 1993.

[8] A. Haar, "Zur theorie der orthogonalen funktionensysteme," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, 1910.

[9] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

[10] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The jpeg 2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[11] B. E. Usevitch, "A tutorial on modern lossy wavelet-based image compression: Jpeg 2000," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 22–35, 2001.

[12] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[13] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*. Morgan Kaufmann, 2007.

[14] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, 2001.

[15] M. Barni, A. Verdi, *et al.*, "Deepfake and synthetic image forensics in the era of generative ai: A wavelet perspective," *IEEE Transactions on Information Forensics and Security*, vol. 20, pp. 112–128, 2025.

[16] Mallat, S. A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, 3rd Edition, 2008.

[17] Daubechies, I. *Ten Lectures on Wavelets*. SIAM, 1992.

Chapter 3

Architecture Profonde des Réseaux de Neurones Convolutifs

3.1 Introduction . . . 35
3.2 Introduction . . . 35
3.3 Fondements du Deep Learning . . . 35
3.3.1 Du Machine Learning au Deep Learning : Le Paradigme de l'Extraction . . . 36
3.3.2 Le Perceptron Multicouche (MLP) comme Socle Structurel . . . 36
3.3.3 Mécanisme d'Apprentissage : Rétropropagation et Fonction de Perte . . . 37
3.4 Introduction aux Réseaux de Neurones Convolutifs (CNN) . . . 37
3.4.1 Architecture générale d'un CNN . . . 38
3.4.2 Processus d'Apprentissage et Convergence du Réseau . . . 38
3.5 Inspiration Biologique et Concept de Champ Récepteur . . . 38
3.5.1 Héritage Biologique : Cellules Simples et Complexes . . . 39
3.5.2 Concept de Champ Récepteur (Receptive Field) . . . 39
3.6 Taxonomie des Couches d'un CNN . . . 40
3.6.1 Couche de Convolution (Convolutional Layer) . . . 40
3.6.2 Couche d'Activation (Activation Layer) . . . 40
3.6.3 Couche de Sous-échantillonnage (Pooling Layer) . . . 41
3.6.4 Couches de Stabilisation et Régularisation . . . 42
3.6.5 Couches d'Aplatissement et Denses (Flatten & Dense) . . . 43
3.6.6 Couches d'Aplatissement et Denses (Flatten & Dense/Fully Connected) . . . 44
3.6.7 Synthèse Fonctionnelle des Couches . . . 45
3.7 Flux de Données d'un CNN . . . 46
3.8 Applications des CNN en Vision par Ordinateur . . . 47
3.8.1 Classification d'images . . . 48
3.8.2 Détection d'objets . . . 48
3.8.3 Segmentation sémantique . . . 49
3.8.4 Transfert de style . . . 50
3.9 Fondements Théoriques de l'Apprentissage des Réseaux de Neurones . . . 51
3.9.1 Hyperparamètres Dynamiques de l'Entraînement . . . 51
3.9.2 Validation et Évaluation de la Généralisation . . . 53
3.10 Analyse Critique : Avantages et Limites . . . 54
3.10.1 Forces et Bénéfices . . . 54
3.10.2 Limitations Intrinsèques . . . 54
3.11 Formalisme de l'Optimisation et Spécificités Forensiques . . . 55
3.11.1 Optimiseur Adam (Adaptive Moment Estimation) . . . 55
3.11.2 Adaptation Forensique : Convolutions Contraintes . . . 55
3.12 Conclusion . . . 56

Abstract

Ce chapitre examine l'architecture profonde des Réseaux de Neurones Convolutifs (CNN) et leurs mécanismes opérationnels. Après avoir situé le Deep Learning par rapport au Machine Learning traditionnel, il détaille la taxonomie complète des couches d'un CNN (convolution, activation, pooling, régularisation et classification). Le chapitre explore également les fondements théoriques de l'apprentissage (optimiseur Adam, surapprentissage) ainsi que les applications majeures en vision par ordinateur. Enfin, une attention particulière est portée à l'adaptation forensique, démontrant comment les convolutions contraintes (filtres SRM et couches de Bayar) permettent d'isoler les micro-traces de manipulation au détriment du contenu sémantique de l'image. **Keywords:** Deep Learning, CNN, Extraction de caractéristiques, Régularisation, Optimiseur Adam, Vision par ordinateur, Analyse forensique, Convolutions contraintes.

3.1 Introduction

3.2 Introduction

L'avènement de l'apprentissage profond (*Deep Learning*) a radicalement transformé le domaine de la vision par ordinateur, ouvrant la voie à des systèmes automatisés d'une précision sans précédent . Parmi ces technologies, les Réseaux de Neurones Convolutifs (*Convolutional Neural Networks* - CNN) s'imposent désormais comme le standard incontournable pour le traitement, l'analyse et la classification des données multidimensionnelles, telles que les images [1].

L'objectif de ce chapitre est de plonger au cœur de l'architecture profonde des CNN afin d'en comprendre les fondements théoriques et les mécanismes opérationnels. Dans un premier temps, nous situerons le Deep Learning par rapport aux approches traditionnelles de l'intelligence artificielle tout en explicitant les principes de base du neurone artificiel Nous explorerons ensuite l'inspiration biologique qui a guidé la conception de ces réseaux, notamment à travers le concept clé de champ récepteur [2]. Enfin, nous détaillerons la taxonomie complète des couches qui composent un CNN — de la convolution au sous-échantillonnage (*Pooling*), jusqu'aux étapes cruciales de régularisation et de classification finale. Cette étude structurelle nous permettra de poser les bases méthodologiques nécessaires à la mise en œuvre de notre propre modèle .

3.3 Fondements du Deep Learning

Le *Deep Learning* (Apprentissage Profond) représente une sous-branche de l'Apprentissage Automatique (*Machine Learning*), elle-même englobée dans le domaine de l'Intelligence Artificielle (IA) [1]. L'essor du Deep Learning est intimement lié à la disponibilité massive des données (*Big Data*) et à l'augmentation des capacités de calcul géométriques fournies par les processeurs graphiques (GPU) .

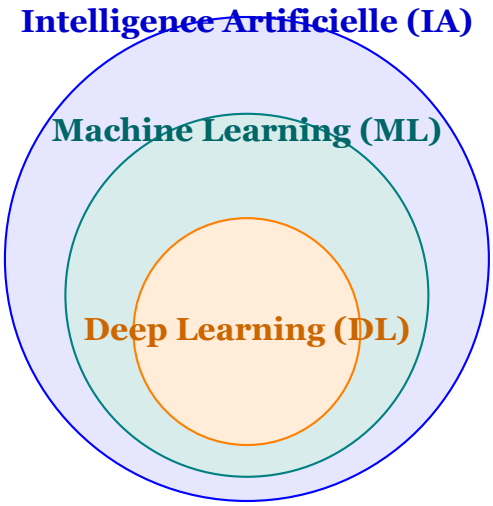


Figure 3.1: Inclusion hiérarchique du Deep Learning au sein de l'Intelligence Artificielle.

3.3.1 Du Machine Learning au Deep Learning : Le Paradigme de l'Extraction

Dans les approches de Machine Learning traditionnelles, l'extraction des caractéristiques (*Feature Engineering*) est une étape manuelle, fastidieuse et dépendante de l'expertise humaine (ex. filtres de Sobel, descripteurs SIFT ou HOG)[1].

Le Deep Learning révolutionne ce processus grâce au concept d'**Apprentissage de Représentations** (*Representation Learning*). Le réseau apprend par lui-même, de manière end-to-end (de bout en bout), à extraire les descripteurs optimaux directement à partir des données brutes, en passant de représentations low-level (bords, textures) à des représentations high-level (formes, objets complexes)[3].

3.3.2 Le Perceptron Multicouche (MLP) comme Socle Structurel

Un réseau de neurones profond est composé d'une superposition de couches de neurones artificiels. Chaque neurone effectue une combinaison linéaire de ses entrées x_i pondérées par des poids w_i, y ajoute un biais b, puis applique une fonction d'activation non linéaire[4] f :

$$z = \sum_{i=1}^{n} w_i x_i + b \quad \Rightarrow \quad a = f(z) \tag{3.1}$$

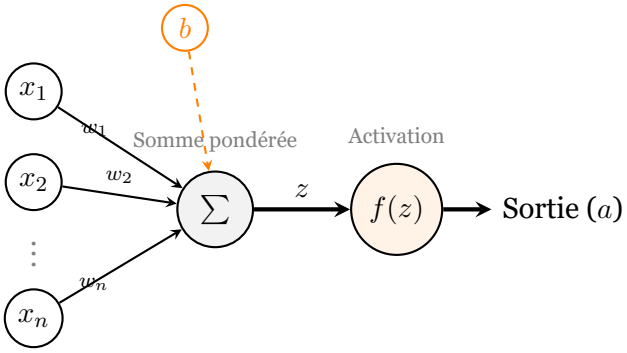


Figure 3.2: Modélisation mathématique et flux interne d'un neurone artificiel standard.

3.3.3 Mécanisme d'Apprentissage : Rétropropagation et Fonction de Perte

L'entraînement d'un modèle de Deep Learning repose sur un cycle itératif en deux phases [4] :

1. **La Passe Avant (Forward Propagation) :** Les données traversent le réseau pour générer une prédiction $\hat{y}$.
2. **Le Calcul de l'Erreur :** Une fonction de perte (*Loss Function*) désigne l'écart entre la prédiction $\hat{y}$ et la vérité terrain y.
3. **La Passe Arrière (Backward Propagation) :** L'algorithme de rétropropagation du gradient applique la règle de dérivation en chaîne (*Chain Rule*) pour propager l'erreur de la sortie vers l'entrée, permettant à un optimiseur d'ajuster les poids θ afin de minimiser la perte.

3.4 Introduction aux Réseaux de Neurones Convolutifs (CNN)

Bien que les Perceptrons Multicouches (MLP) soient performants pour des données tabulaires[1], ils souffrent d'une limitation majeure face aux données multidimensionnelles comme les images : l'explosion du nombre de paramètres due aux connexions denses (*Fully Connected*), ce qui entraîne un surapprentissage systémique. Les réseaux de neurones convolutifs (*Convolutional Neural Networks* - CNN) ont été introduits pour pallier cette limite en préservant la topologie spatiale des images.

3.4.1 Architecture générale d'un CNN

Une architecture CNN typique est structurellement organisée en trois grands blocs séquentiels[2] :

1. **La couche d'entrée (Input Layer) :** Reçoit la matrice de pixels brute de l'image (hauteur $\times$ largeur $\times$ canaux).
2. **Les couches cachées (Hidden Layers) :** Constituent le cœur extracteur du réseau, formées par l'imbrication de couches de convolution, d'activation, de régularisation et de pooling.
3. **La couche de sortie (Output Layer) :** Fournit les scores de probabilité associés à chaque classe cible via des couches denses.

3.4.2 Processus d'Apprentissage et Convergence du Réseau

L'entraînement d'un CNN repose sur l'ajustement itératif de ses filtres à travers le calcul d'une fonction de perte. Lors de la passe avant, le réseau génère une prédiction ; l'erreur commise est ensuite rétropropagée d'amont en aval afin de permettre à un optimiseur (tel que l'algorithme Adam) de mettre à jour les poids, garantissant ainsi la convergence du modèle vers une extraction de caractéristiques optimale[5].

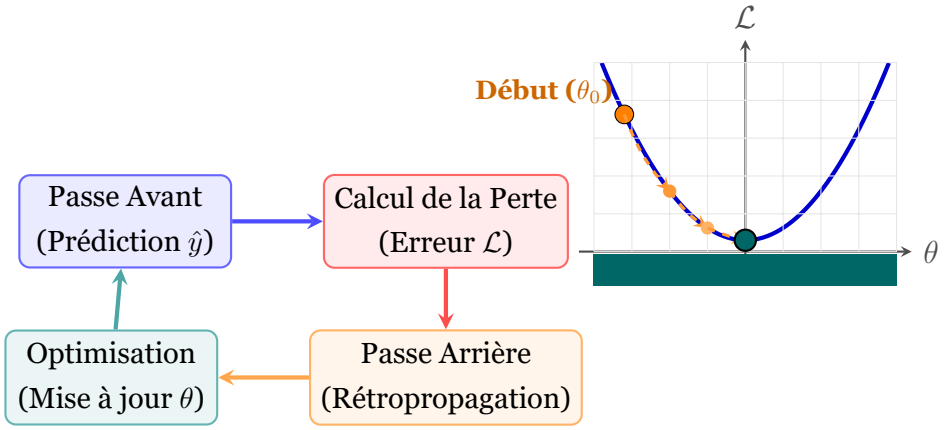


Figure 3.3: Schéma cyclique global de l'apprentissage (gauche) et processus géométrique de convergence vers le minimum de la fonction de perte (droite).

3.5 Inspiration Biologique et Concept de Champ Récepteur

L'architecture des CNN s'inspire directement du système visuel biologique, notamment des travaux de Hubel et Wiesel (1962) sur le cortex visuel primaire (**V1**)[6].

3.5.1 Héritage Biologique : Cellules Simples et Complexes

Le traitement de l'information visuelle s'effectue de manière hiérarchique au niveau cérébral. Les cellules simples s'activent de façon sélective face à des stimuli ayant une orientation spatiale précise (lignes, contours locaux)[6], tandis que les cellules complexes synthétisent ces signaux pour achever une invariance spatiale relative (translation, échelle).

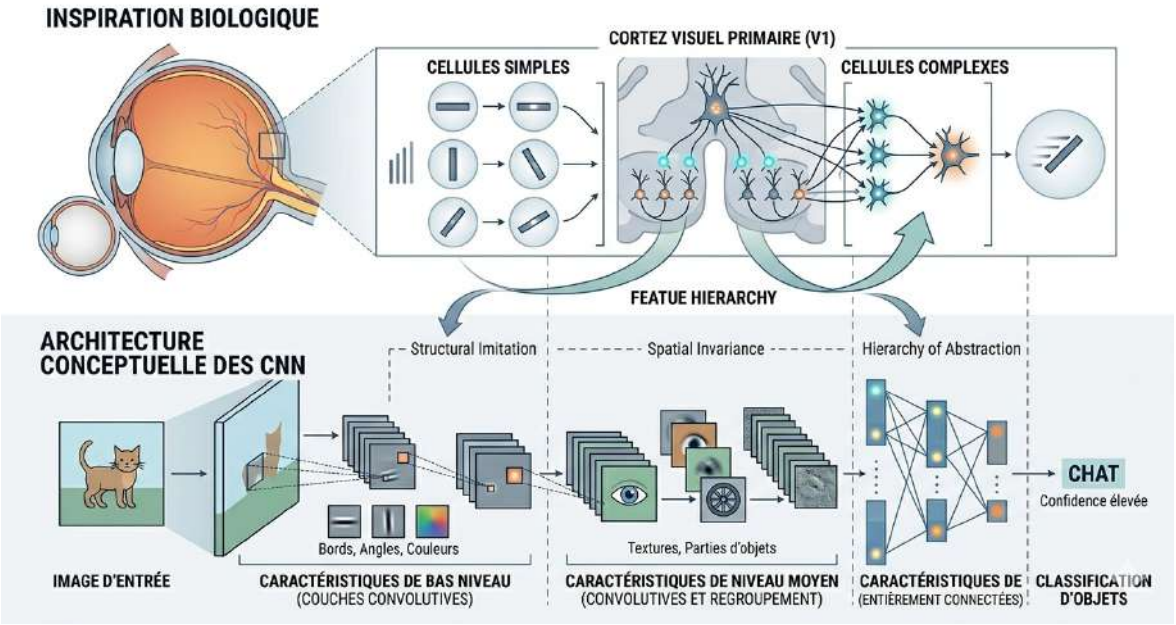


Figure 3.4: Parallélisme fonctionnel entre le système visuel biologique (V1) et la hiérarchie d'abstraction des caractéristiques dans un CNN.

3.5.2 Concept de Champ Récepteur (Receptive Field)

Contrairement aux réseaux denses où chaque neurone est connecté à la totalité des pixels, un neurone dans une couche convolutive n'est connecté qu'à une petite région locale de l'image d'entrée, appelée **Champ Récepteur**[7]. Cette approche offre trois avantages majeurs :

1. **Corrélation Spatiale :** Les pixels proches sont statistiquement plus liés et cohérents que les pixels éloignés.
2. **Partage de Poids (*Weight Sharing*) :** Le même filtre balaye toute l'image, permettant de détecter un motif peu importe sa position spatiale.
3. **Réduction de la complexité :** Limite drastiquement le nombre de paramètres libres à apprendre, évitant l'explosion calculatoire.

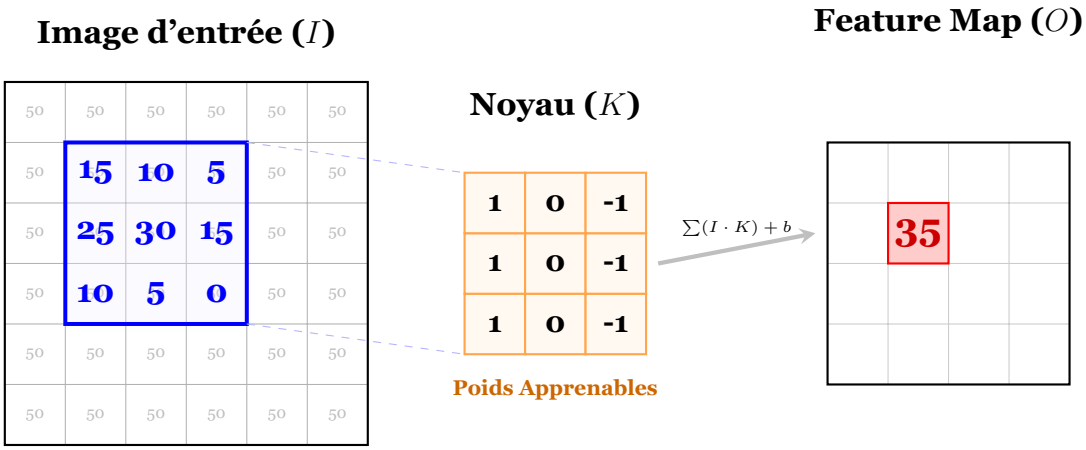


Figure 3.5: Représentation mathématique du mécanisme de balayage par un noyau convolutif.

3.6 Taxonomie des Couches d'un CNN

Les transformations de l'information visuelle au sein des couches cachées s'articulent autour des structures spécialisées suivantes [1] :

3.6.1 Couche de Convolution (Convolutional Layer)

Elle effectue un produit de convolution entre un noyau apprenable K de taille $k \times k$ et la carte d'entrée I [1] :

$$O(x,y) = \sum_{i=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \sum_{j=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} I(x+i, y+j) \cdot K(i,j) + b \tag{3.2}$$

3.6.2 Couche d'Activation (Activation Layer)

La fonction standard est la *ReLU* (*Rectified Linear Unit*) :

$$f(x) = \max(0, x) \tag{3.3}$$

Pour éviter le phénomène de mort des neurones (*Dying ReLU*), la variante *Leaky ReLU* applique une pente faible α dans la zone négative [8]:

$$f(x) = \begin{cases} x & \text{si } x > 0 \\ \alpha x & \text{si } x \leq 0 \end{cases} \tag{3.4}$$

3.6.2.1 Atténuation de la Disparition du Gradient (Vanishing Gradient)

L'adoption massive de la fonction ReLU au détriment des fonctions historiques (telles que la sigmoïde) is motivée par la résolution du problème de disparition du gradient. Les fonctions saturantes possèdent des dérivées qui tendent vers zéro pour des valeurs fortement positives, ce qui paralyse la mise à jour des poids lors de la rétropropagation. À l'inverse, la sous-dérivée de ReLU reste constante et égale à 1 pour tout apport positif ($\forall x > 0,\ f'(x) = 1$), assurant une transmission fluide du gradient à travers les couches profondes[9].

3.6.3 Couche de Sous-échantillonnage (Pooling Layer)

Le sous-échantillonnage permet de réduire la résolution spatiale des cartes de caractéristiques tout en conservant les informations les plus importantes. Cela permet de réduire le nombre de paramètres computationnels et d'introduire une invariance face aux petites translations ou distorsions. Les deux variantes les plus utilisées sont le *Max Pooling* et l'*Average Pooling*[1].

3.6.3.1 La couche Max Pooling

Le *Max Pooling* consiste à balayer la matrice d'entrée par une fenêtre de taille fixe (généralement 2×2) avec un pas de déplacement (*Stride*). Comme illustré dans la **Figure 3.6**, pour chaque sous-région couverte, l'algorithme extrait uniquement la valeur maximale. Cette approche permet de capturer les caractéristiques les plus saillantes (comme les bords ou les textures contrastées) tout en éliminant le bruit de fond[10].

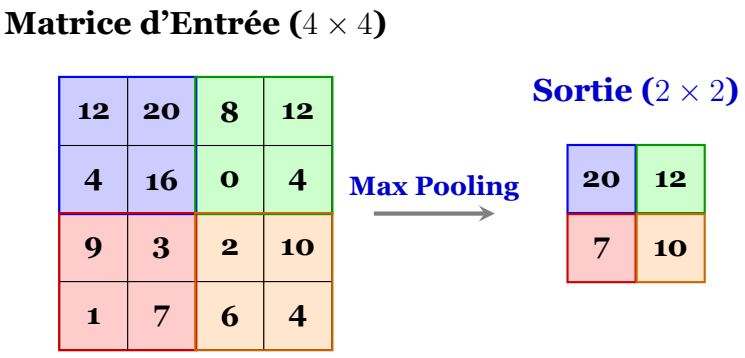


Figure 3.6: Opération de réduction dimensionnelle par Max Pooling (2×2, Stride 2).

3.6.3.2 La couche Average Pooling

L'*Average Pooling* adopte le même principe de balayage spatial. Cependant, comme le montre le schéma de la **Figure 3.7**, au lieu d'isoler le maximum, il calcule la

moyenne arithmétique de toutes les valeurs présentes dans la sous-région. Cette méthode adoucit les cartes de caractéristiques en lissant les activations, ce qui s'avère particulièrement utile lorsque le bruit local doit être estompé globalement[11].

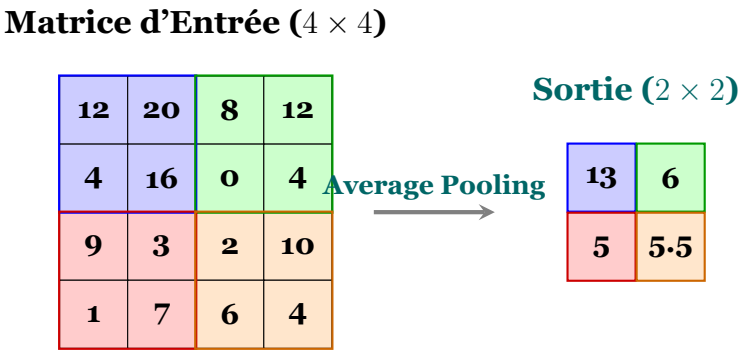


Figure 3.7: Opération de réduction dimensionnelle par Average Pooling (2×2, Stride 2).

La dimension finale de la sortie après n'importe quel type de pooling est régie par la formule standard suivante [4]:

$$O_{dim} = \frac{I_{dim} - F}{S} + 1 \tag{3.5}$$

3.6.4 Couches de Stabilisation et Régularisation

Pour garantir un entraînement stable, rapide et robuste face au surapprentissage (*overfitting*), deux types de couches spécialisées sont introduites :

- **Batch Normalization (Normalisation par lots) :** Comme illustré dans la **Figure 3.8**, cette couche normalise les activations de chaque mini-lot pour qu'elles aient une moyenne de 0 et une variance de 1. Ce mécanisme élimine le problème du *Internal Covariate Shift* (la variation continuelle de la distribution des entrées), ce qui permet d'utiliser des taux d'apprentissage plus élevés et d'accélérer considérablement la convergence du modèle[12].

- **Dropout (Élimination stochastique) :** Comme le montre la **Figure 3.9**, le *Dropout* désactive aléatoirement une fraction p de neurones (souvent 50%) à chaque itération d'entraînement. Cela force le réseau à apprendre des caractéristiques indépendantes et redondantes, empêchant les neurones de trop co-dépendre les uns des autres. C'est l'une des techniques de régularisation les plus efficaces contre le surapprentissage[13].

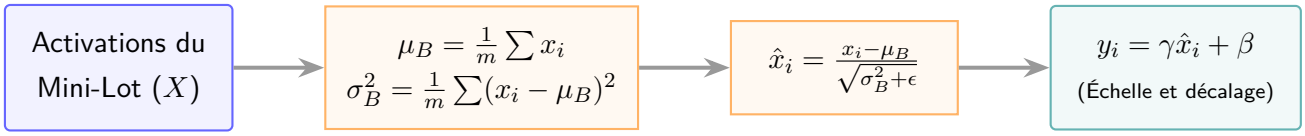


Figure 3.8: Architecture fonctionnelle et étapes mathématiques de la Batch Normalization.

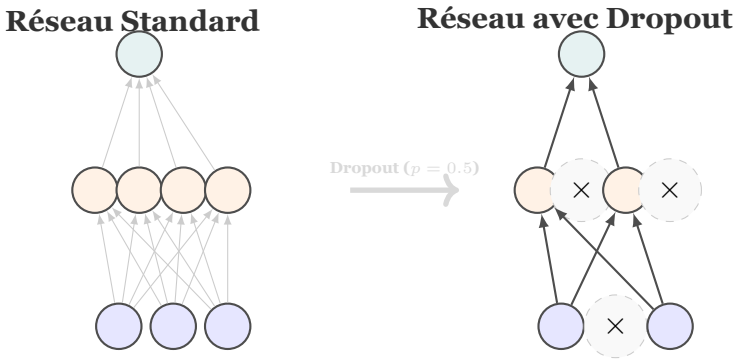


Figure 3.9: Comparaison structurelle entre un réseau dense classique et un réseau avec application du Dropout.

3.6.5 Couches d'Aplatissement et Denses (Flatten & Dense)

Ces couches représentent la phase de transition finale, convertissant les caractéristiques spatiales en décisions logiques.

- **Couche Flatten (Aplatissement) :** Comme schématisé dans la **Figure 3.10**, les étapes de convolution et de pooling produisent des données en 3D (Hauteur × Largeur × Canaux). La couche *Flatten* déroule ces matrices pour les transformer en un unique vecteur linéaire en 1D. Cette étape est obligatoire car las couches denses ne peuvent recevoir que des vecteurs unidimensionnels.
- **Couche Dense (Fully Connected / Complètement connectée) :** Également visible dans la **Figure 3.10**, cette couche connecte chaque élément du vecteur *Flatten* à tous ses neurones. Elle combine globalement toutes les caractéristiques extraites pour calculer le score final de chaque classe. La couche de sortie utilise ensuite une fonction d'activation (comme *Softmax*) pour générer une distribution de probabilités claire pour la classification[4].

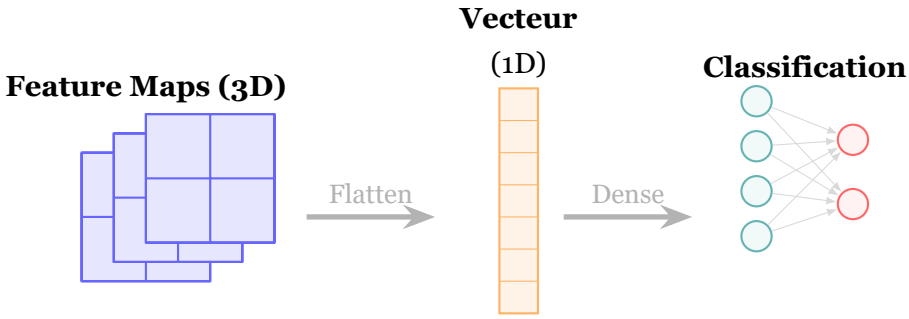


Figure 3.10: Flux de traitement séquentiel : de l'aplatissement des cartes 3D à la classification finale par la couche Dense.

3.6.6 Couches d'Aplatissement et Denses (Flatten & Dense/-Fully Connected)

Ces couches constituent l'étape finale d'un réseau CNN, où les données passent d'une représentation spatiale matricielle à des décisions de classification vectorielles[4].

- **Couche d'aplatissement (Flatten) :** Comme illustré dans la **Figure 3.11**, cette couche transforme les cartes de caractéristiques bidimensionnelles ($H \times W$) issues des étapes de convolution et de pooling en un unique vecteur unidimensionnel linéaire. Cette conversion géométrique est indispensable car les couches denses ne tolèrent que des vecteurs 1D en entrée.
- **Couche Dense (Fully Connected) :** Comme le montre la **Figure 3.12**, chaque neurone de cette couche est structurellement connecté à l'intégralité des neurones de la couche précédente. Son rôle est de réaliser la combinaison logique et globale de toutes les caractéristiques extraites afin de projeter les scores finaux vers les probabilités des classes cibles.

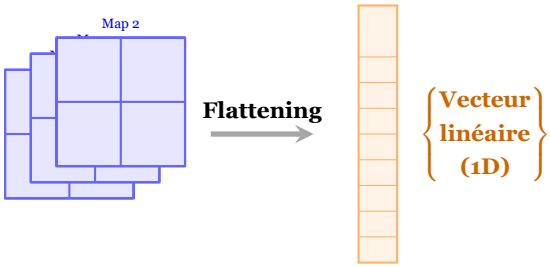


Figure 3.11: Processus d'aplatissement (Flatten) convertissant les cartes de caractéristiques 2D en un vecteur 1D.

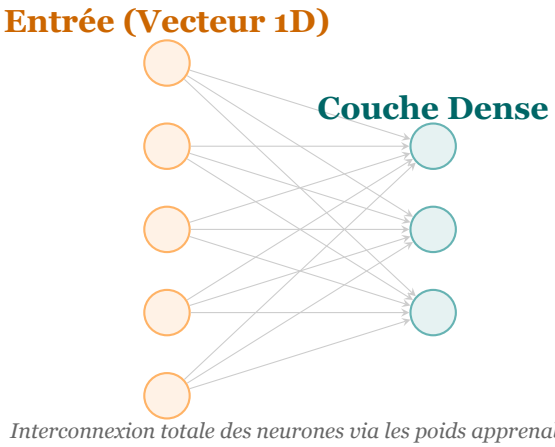


Figure 3.12: Structure interne d'une couche Dense (Fully Connected) montrant la densité des liaisons.

3.6.7 Synthèse Fonctionnelle des Couches

Le tableau 3.1 résume les caractéristiques clés et les rôles respectifs des couches configurées au sein du réseau [1].

Table 3.1: Synthèse fonctionnelle des composants structurels d'un CNN.

Couche	**Fonction Principale**	**Type d'Entrée**	**Paramètres Apprenables**
Convolution	Extraction des descripteurs locaux (bords, textures)	Matricielle (2D/3D)	Oui (Poids du Noyau + Biais)
Activation	Introduction de non-linéarités complexes (ReLU)	Matricielle (2D/3D)	Non
Pooling	Réduction de la dimension spatiale & invariance relative	Matricielle (2D/3D)	Non
Batch Norm	Stabilisation des activations & vitesse de convergence	Matricielle / Vecteur	Oui (γ et β)
Dropout	Régularisation stochastique (Anti-overfitting)	Matricielle / Vecteur	Non
Dense (FC)	Combinaison globale des caractéristiques et décision	Vectorielle (1D)	Oui (Matrice de projection)

3.7 Flux de Données d'un CNN

Le flux informationnel au sein d'un CNN profond convertit de manière séquentielle les pixels bruts en prédictions probabilistes.

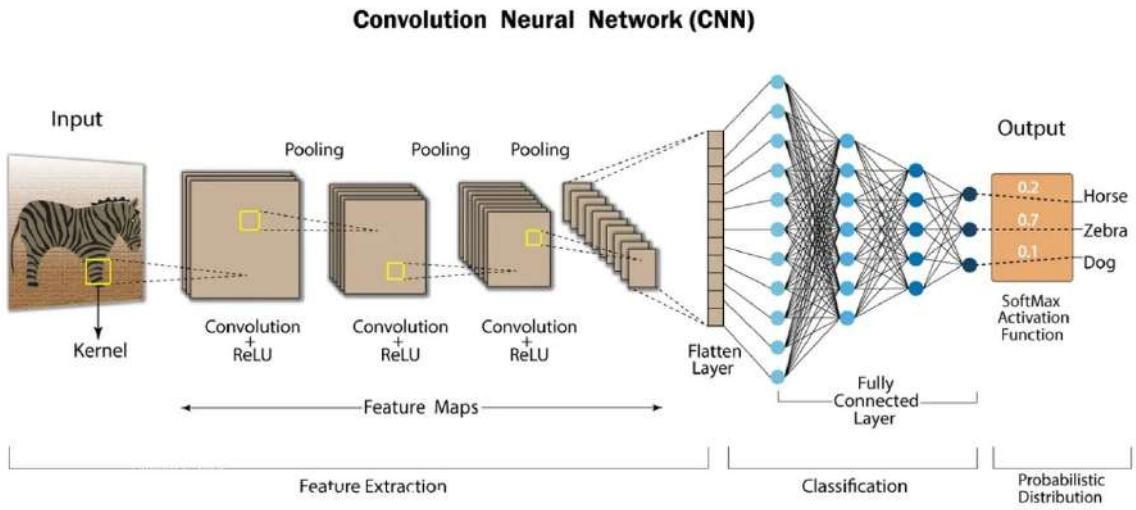


Figure 3.13: Pipeline complet d'un CNN : de l'extraction des caractéristiques spatiales (*Feature Extraction*) à la distribution probabiliste finale via la fonction SoftMax.[18]

Comme illustré dans la Figure 3.13, le traitement respecte les étapes suivantes [4]:

- **Feature Extraction :** L'image d'entrée (*Input*) subit une alternance de convolutions (*Convolution + ReLU*) et de sous-échantillonnage (*Pooling*) pour raffiner l'extraction.
- **Flattening :** La couche *Flatten* convertit les structures matricielles en un vecteur linéaire.
- **Classification :** Les couches denses (*Fully-Connected Layer*) effectuent la synthèse logique des descripteurs.
- **Probabilistic Distribution :** La fonction d'activation *SoftMax* normalise la sortie sous forme d'une distribution de probabilités associées à chaque classe.

3.8 Applications des CNN en Vision par Ordinateur

Les réseaux de neurones convolutifs (CNN) ont révolutionné le domaine de la vision par ordinateur grâce à leur capacité à traiter efficacement les données spatiales[2]. Leurs applications s'étendent de la simple reconnaissance globale à la compréhension microscopique de chaque pixel d'une image.

3.8.1 Classification d'images

La classification d'images consiste à attribuer une étiquette (*label*) unique à une image entière parmi un ensemble de classes prédéfinies. Le réseau analyse globalement les motifs (textures, formes) pour calculer un score de probabilité[3].

Comme illustré de manière détaillée dans la **Figure 3.14**, le processus prend en entrée une structure visuelle complète pour la faire passer à travers les différentes couches d'extraction de caractéristiques du réseau, avant de projeter un vecteur de décision final. Dans le domaine de la forensique, cela permet par exemple de catégoriser automatiquement une image comme « Authentique » ou « Falsifiée ».

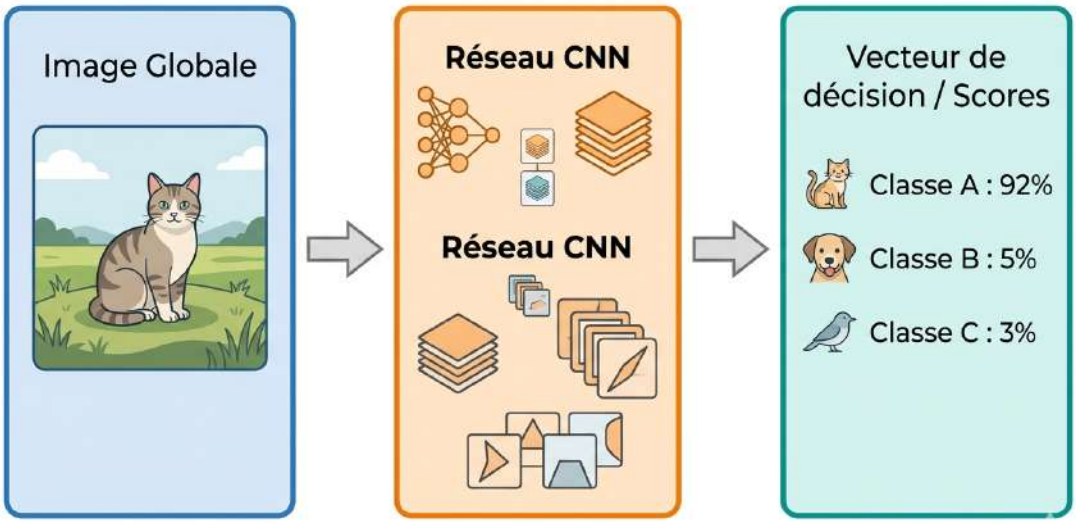


Figure 1.2: Pipeline de classification d'images : attribution d'un label unique basé sur des probabilités.

Figure 3.14: Pipeline de classification d'images : attribution d'un label unique basé sur des probabilités.

3.8.2 Détection d'objets

Contrairement à la classification, la détection d'objets gère des scènes complexes contenant plusieurs entités. Elle accomplit deux tâches simultanément : identifier la classe de chaque objet présent et déterminer sa position exacte en traçant une boîte englobante (*Bounding Box*) définie par des coordonnées géométriques (x, y, h, w).

Comme l'illustre la **Figure 3.15**, ce processus permet de localiser et de catégoriser de manière synchrone chaque élément d'intérêt (tels que les voitures, les camions ou les piétons) au sein d'un même environnement visuel. C'est une technologie clé pour la surveillance vidéo et les systèmes de conduite autonome[14].

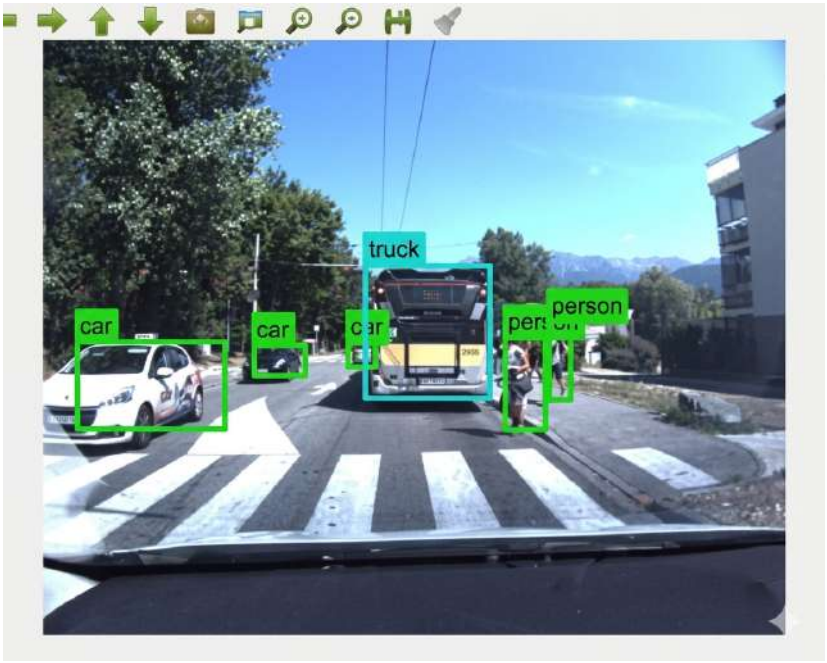


Figure 3.15: Exemple de détection d'objets multiples avec leurs boîtes englobantes et scores de confiance[19].

3.8.3 Segmentation sémantique

La segmentation sémantique pousse l'analyse au niveau le plus fin possible en classant **chaque pixel individuel** de l'image[15]. Il n'y a plus de boîte globale : l'image est découpée en zones de couleurs représentant des catégories d'objets homogènes. En imagerie médicale, cette technique est indispensable pour délimiter au pixel près les contours d'une tumeur ou d'un organe.

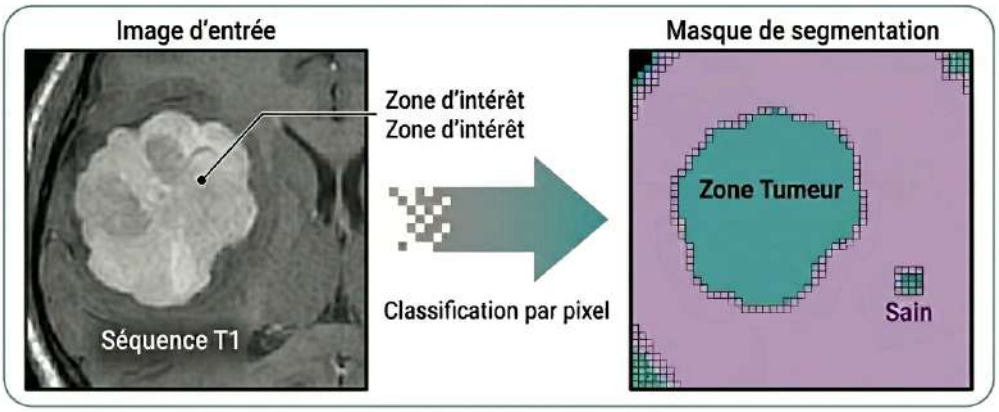


Figure 3.16: Principe de la segmentation sémantique : classification pixel par pixel pour isoler des formes complexes[20].

3.8.4 Transfert de style

Le transfert de style est une application générative qui utilise les couches internes du CNN de manière originale. Comme le montre l'illustration du **coeur de l'algorithme** dans la **Figure 3.17**, le réseau extrait la structure géométrique profonde d'une première image (*Image de Contenu*) et les textures et couleurs d'une seconde image (*Image de Style*), puis fusionne ces deux informations pour synthétiser une œuvre entièrement nouvelle[16].



Figure 3.17: Processus algorithmique du transfert de style artistique via l'analyse des couches du CNN : fusion du contenu structurel et du style esthétique.

3.9 Fondements Théoriques de l'Apprentissage des Réseaux de Neurones

Le processus d'optimisation d'un réseau de neurones convolutif (CNN) repose sur l'ajustement itératif de ses paramètres (poids et biais) afin de minimiser une fonction de coût. Cette section détaille les mécanismes cardinaux qui régissent la dynamique de cet apprentissage ainsi que les outils d'évaluation de sa convergence[1].

3.9.1 Hyperparamètres Dynamiques de l'Entraînement

3.9.1.1 La Notion d'Époque (Epoch)

Une *époque* est définie comme un cycle complet au cours duquel l'intégralité du jeu de données d'apprentissage (Training Set) passe à travers le réseau via une étape de propagation avant (Forward Pass) et une étape de rétropropagation du gradient (Backward Pass)[1].

Le nombre total d'époques est un hyperparamètre crucial : un nombre trop faible peut conduire à un **sous-apprentissage** (Underfitting), tandis qu'un nombre excessivement élevé expose le modèle au **sur-apprentissage** (Overfitting).

3.9.1.2 Le Taux d'Apprentissage (Learning Rate : α)

Le taux d'apprentissage $\alpha \in]0, 1[$ détermine l'amplitude de la correction apportée aux poids à chaque itération. Si α est trop grand, l'optimisation risque de diverger ou d'osciller de manière instable autour du minimum. S'il est trop faible, la convergence devient extrêmement lente et le modèle risque de rester piégé dans un minimum local[1].

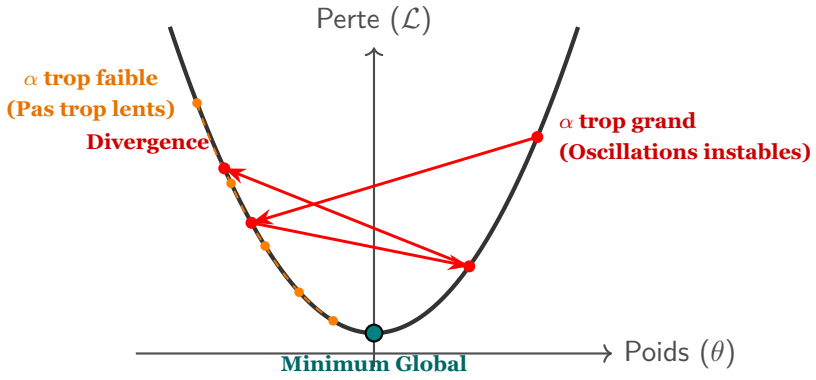


Figure 3.18: Impact du choix du taux d'apprentissage (α) sur la dynamique de convergence géométrique.

3.9.1.3 Stratégie de Décroissance par Paliers (Piecewise Constant Decay)

Pour optimiser la convergence, il est courant d'adopter une planification du taux d'apprentissage (*Learning Rate Scheduling*). La méthode *Piecewise Constant Decay* consiste à maintenir α constant pendant un intervalle d'époques prédéfini, puis à le réduire brutalement par un facteur multiplicatif (souvent 0.1) à des paliers spécifiques [1] :

$$\alpha(t) = \alpha_0 \cdot \gamma^{\lfloor t/\tau \rfloor} \tag{3.6}$$

Où t represents l'époque actuelle, α_0 le taux initial, γ le facteur d'atténuation, et τ le palier de transition. Cette baisse par paliers permet d'abord une exploration rapide de l'espace des solutions, puis un raffinement d'une grande précision à l'approche du minimum global.

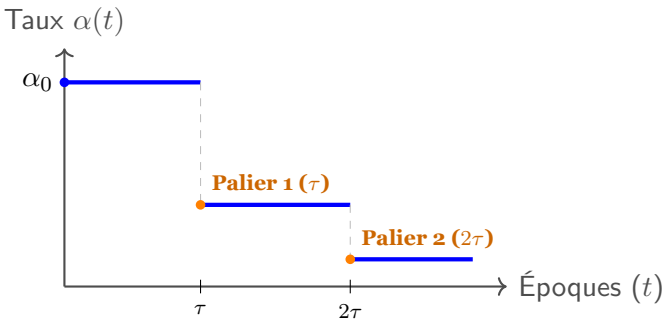


Figure 3.19: Planification du taux d'apprentissage selon la stratégie Piecewise Constant Decay.

3.9.2 Validation et Évaluation de la Généralisation

3.9.2.1 Le Jeu de Validation (Validation Set)

Pour garantir que le réseau n'apprend pas par "cœur" (mémorisation des données d'entraînement), les données initiales sont scindées. Le *jeu de validation* est un sous-ensemble de données indépendant que le réseau ne voit jamais durant la phase de mise à jour des poids. À la fin de chaque époque, le modèle est testé sur ce jeu afin de mesurer sa capacité de **généralisation**[1].

3.9.2.2 Analyse Mathématique des Courbes d'Apprentissage

Le suivi de la convergence s'effectue graphiquement via les courbes de perte (*Loss*) et de précision (*Accuracy*). En théorie macroscopique :

- **Phase de Convergence Idéale :** La fonction de perte d'entraînement ($Loss_{train}$) et celle de validation ($Loss_{val}$) décroissent de manière monotone et asymptotique.
- **Phénomène de Sur-apprentissage (Overfitting) :** Il se manifeste graphiquement lorsque la courbe $Loss_{train}$ continue sa décroissance vers zéro, tandis que la courbe $Loss_{val}$ s'échappe et commence à réaugmenter. Ce point de divergence indique que le modèle perd sa capacité de généralisation et commence à modéliser le bruit spécifique aux données d'entraînement[1].

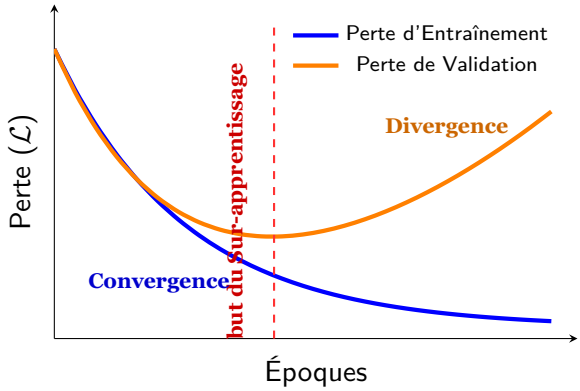


Figure 3.20: Évolution des courbes de perte théoriques mettant en évidence le point de bascule vers le sur-apprentissage (Overfitting).

3.10 Analyse Critique : Avantages et Limites

3.10.1 Forces et Bénéfices

- **Extraction automatique :** Élimine le besoin de concevoir manuellement des descripteurs statistiques (*Handcrafted Features*).
- **Invariance spatiale :** Capacité à reconnaître un motif indépendamment de sa translation, taille, position ou rotation.
- **Robustesse au bruit :** Aptitude à traiter des données visuelles de qualité variable.
- **Apprentissage par transfert (Transfer Learning) :** Possibilité de réutiliser des modèles pré-entraînés sur de larges bases de données[2].

3.10.2 Limitations Intrinsèques

- **Coût computationnel :** L'entraînement nécessite des architectures matérielles lourdes (GPU).
- **Surapprentissage (Overfitting) :** Risque élevé de mémorisation si le dataset est trop restreint.
- **Effet "Boîte Noire" :** Manque d'interprétabilité mathématique directe concernant les décisions internes du réseau.
- **Contrainte de structure :** Limité nativement aux données structurées sous forme de grilles régulières (matrices de pixels)[1].

3.11 Formalisme de l'Optimisation et Spécificités Forensiques

3.11.1 Optimiseur Adam (Adaptive Moment Estimation)

L'algorithme **Adam** ajuste dynamiquement le taux d'apprentissage (learning rate) pour chaque paramètre du réseau. À chaque itération t, à partir du gradient calculé $g_t = \nabla_\theta \mathcal{L}(\theta_t)$, l'algorithme estime les moments réajustés $\hat{m}_t$ (qui représente la moyenne mobile des gradients) et $\hat{v}_t$ (qui représente la variance non centrée)[17] :

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{3.7}$$

Ensuite, la mise à jour des poids (paramètres θ) s'effectue en normalisant le pas d'apprentissage par la racine carrée de la variance, évitant ainsi les oscillations brutales :

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon}\hat{m}_t \tag{3.8}$$

Pour l'apprentissage des tâches de classification binaire (Image Authentique vs Image Falsifiée), la fonction de perte exploitée est la *Binary Cross-Entropy* (BCE), définie par :

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \tag{3.9}$$

3.11.2 Adaptation Forensique : Convolutions Contraintes

Il existe une divergence fondamentale entre la vision par ordinateur classique et l'analyse forensique (détection de falsifications) :

- **En vision classique :** L'objectif est d'atténuer le bruit de fond pour extraire la *sémantique* globale de l'image (reconnaître des objets, des formes, des visages).
- **En analyse forensique :** La sémantique de l'image devient un facteur de confusion. L'objectif est inverse : il faut supprimer le contenu visuel global afin de capturer uniquement les **micro-traces de manipulation à haute fréquence** (les anomalies statistiques invisibles à l'œil nu).

Pour contraindre le réseau de neurones à ignorer le contenu de l'image et à se focaliser sur ces résidus de bruit dès la première couche, deux approches majeures sont introduites [8]:

3.11.2.1 Filtres SRM (Spatial Rich Model)

Cette méthode utilise des filtres de passe-haut prétextés et **fixes** (non apprenables)[9]. Ils agissent comme un extracteur de caractéristiques stochastiques en isolant directement les résidus de bruit par une opération de convolution :

$$I_{residu} = I \otimes K_{SRM} \tag{3.10}$$

3.11.2.2 Couche de Convolution de Bayar (BayarConv)

Contrairement aux filtres fixes, la couche de Bayar laisse les poids du noyau s'adapter par apprentissage, mais elle leur impose une **contrainte mathématique stricte** à chaque itération [8]:

$$\begin{cases} \sum_{i \neq 0, j \neq 0} K(i,j) = -1 \\ K(0,0) = 1 \end{cases} \tag{3.11}$$

En forçant le coefficient central du noyau $K(0,0)$ à être égal à 1 et la somme de tous les autres coefficients périphériques à -1, ce filtre fonctionne comme un opérateur de différenciation locale. Il élimine ainsi complètement les composantes continues (les zones homogènes et l'éclairage global de l'image) pour piéger les moindres anomalies et discontinuités statistiques à haute fréquence induites par une falsification.

3.12 Conclusion

En conclusion, ce chapitre nous a permis de comprendre le fonctionnement interne des réseaux de neurones convolutifs (CNN). Nous avons examiné comment ces réseaux s'inspirent du système visuel biologique pour extraire automatiquement les caractéristiques des images, tout en optimisant le nombre de paramètres grâce au partage des poids.

De la couche de convolution au traitement final par les couches denses, en passant par le pooling et les techniques de régularisation (Dropout et Batch Normalization), chaque composant joue un rôle essentiel dans la performance globale du modèle. Ces bases théoriques étant posées, nous pouvons maintenant aborder, dans le chapitre suivant, la phase pratique de notre étude, à savoir la mise en œuvre et l'entraînement de notre modèle.

Bibliography

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online].

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[4] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[6] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of Physiology*, vol. 160, no. 1, pp. 106–154, 1962.

[7] A. Hyvärinen and P. Hoyer, "Emergence of phase- and shift-invariant features by decomposition of natural images into independent canonical components," *Neural Computation*, vol. 12, no. 9, pp. 2095–2120, 2000.

[8] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," *in Proc. ICML*, vol. 30, no. 1, p. 3, 2013.

[9] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *in Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 315–323, 2011.

[10] Y. L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," *in Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 111–118, 2010.

[11] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *in International Conference on Artificial Neural Networks*, pp. 92–101, Springer, Berlin, Heidelberg, 2010.

[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *in International conference on machine learning*, pp. 448–456, 2015.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[16] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *Journal of Vision*, vol. 16, no. 12, pp. 326–326, 2016.

[17] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[18] S. Sumit, "A comprehensive guide to convolutional neural networks — the ELI5 way," *Towards Data Science*, 2018.

[19] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012.

[20] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, and B. Menze, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.

Chapter 4
Détection de Falsification d'Images basée sur la DWT et les CNN

4.1 Introduction . . . 60
4.2 Architecture Globale du Système de Fusion . . . 60
4.2.1 Flux opératoire de la fusion spatio-fréquentielle . . . 60
4.3 Algorithme de Détection du Système Hybride . . . 61
4.4 Environnement Expérimental et Configuration . . . 62
4.4.1 Base de données d'évaluation (Celeb-DF v2) . . . 62
4.4.2 Hyperparamètres d'optimisation de la fusion . . . 62
4.4.3 Analyse de l'impact du partitionnement en Mini-batches . . . 63
4.5 Résultats Expérimentaux et Analyse Comparative . . . 64
4.5.1 Analyse qualitative et probabilités individuelles des échantillons . . . 64
4.5.2 Évaluation dynamique de la convergence . . . 65
4.5.3 Analyse quantitative globale et critères structurels des réseaux . . . 66
4.5.4 Analyse Algorithmique par Matrices de Confusion . . . 67
4.5.5 Analyse de la sensibilité par les courbes ROC . . . 69
4.5.6 Impact de la sélection de l'ondelette mère sur les performances . . . 70
4.5.7 Discussion et interprétation des comportements de fusion . . . 71
4.5.8 Analyse comparative approfondie des topologies architecturales : CNN, SqueezeNet et ResNet . . . 73
4.6 Conclusion du Chapitre . . . 74

Abstract

Ce chapitre examine l'architecture profonde des Réseaux de Neurones Convolutifs (CNN) et leurs mécanismes opérationnels. Après avoir situé le Deep Learning par rapport au Machine Learning traditionnel, il détaille la taxonomie complète des couches d'un CNN (convolution, activation, pooling, régularisation et classification). Le chapitre explore également les fondements théoriques de l'apprentissage (optimiseur Adam, surapprentissage) ainsi que les applications majeures en vision par ordinateur. Enfin, une attention particulière est portée à l'adaptation forensique, démontrant comment les convolutions contraintes (filtres SRM et couches de Bayar) permettent d'isoler les micro-traces de manipulation au détriment du contenu sémantique de l'image.

Keywords: Deep Learning, CNN, Extraction de caractéristiques, Régularisation, Optimiseur Adam, Vision par ordinateur, Analyse forensique, Convolutions contraintes.

4.1 Introduction

Avec l'avancement rapide et la démocratisation des outils d'édition numérique, la manipulation d'images est devenue une pratique à la fois accessible et extrêmement sophistiquée, menaçant directement la crédibilité des informations visuelles dans les domaines médiatique, juridique et sécuritaire. Face à des falsifications de plus en plus indécelables à l'œil nu, le développement de systèmes d'authentification automatisés et hautement précis est devenu un impératif crucial de la criminalistique numérique.

Ce chapitre est dédié à la présentation, à l'implémentation et à la validation expérimentale de notre approche hybride de détection de falsifications. Cette méthodologie originale articule la force de l'analyse fréquentielle par la transformée en ondelettes discrète (*Discrete Wavelet Transform* - DWT) et la puissance d'abstraction sémantique des réseaux de neurones convolutifs (*Convolutional Neural Networks* - CNN). En tirant parti de la décomposition en ondelettes, notre système réussit à capturer les anomalies microscopiques et les discontinuités locales de haute fréquence induites par les processus de manipulation, compensant ainsi le besoin d'architectures excessivement lourdes.

4.2 Architecture Globale du Système de Fusion

L'originalité de notre approche repose sur le couplage direct de la DWT et du CNN en un seul flux de traitement unifié, comme illustré par le schéma fonctionnel de la Figure 4.1.

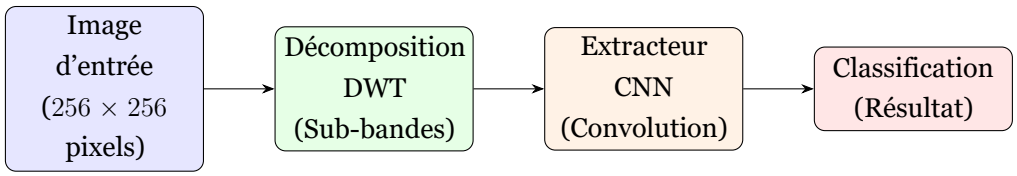


Figure 4.1: Architecture hybride globale proposée pour le système de fusion DWT-CNN.

4.2.1 Flux opératoire de la fusion spatio-fréquentielle

Le flux de traitement de notre contribution se divise en quatre étapes interconnectées :

1. **Normalisation géométrique :** Le système standardise l'image du visage à analyser à une taille fixe de 256×256 pixels pour figer l'espace dimensionnel avant l'analyse fréquentielle.

2. **Filtrage et sélection fréquentielle (DWT) :** Le système subit une décomposition en ondelettes discrètes de niveau 1. Contrairement aux applications d'ondelettes classiques où la basse fréquence (LL) est conservée pour la compression, notre approche élimine délibérément la sous-bande LL. Nous fusionnons uniquement les sous-bandes de détails à haute fréquence (LH, HL, HH). Ce choix permet d'isoler les anomalies de reconstruction microscopiques laissées par les algorithmes de falsification (GAN, Deepfakes) tout en divisant les dimensions spatiales par deux (128×128).

3. **Extraction sémantique profonde (CNN) :** Le tenseur de fusion issu de la DWT alimente l'extracteur convolutif. Le réseau n'apprend pas ici à reconnaître des formes globales, mais se focalise sur les signatures statistiques et les irrégularités de texture des hautes fréquences.

4. **Classification finale :** Une fonction d'activation Softmax calcule la probabilité empirique de falsification pour classer l'image comme **Falsifiée (1)** ou **Authentique (0)**.

4.3 Algorithme de Détection du Système Hybride

L'Algorithme 1 décrit de manière séquentielle le flux logique de notre approche de fusion, matérialisant la transition directe entre le filtrage mathématique et la décision finale.

Algorithm 1 Processus de détection de falsification basé sur la fusion DWT-CNN

Require: Image de test I_{input}

Ensure: Label de classification (0 = Authentique, 1 = Falsifiée)

// Étape 1 : Prétraitement et Normalisation Spatiale

Charger l'image initiale I_{input} et appliquer un redimensionnement à 256×256 pixels.

// Étape 2 : Extraction et Fusion Fréquentielle

Calculer la DWT de niveau 1 pour extraire les matrices fréquentielles : $[LL, LH, HL, HH]$.

Écarter la sous-bande LL et fusionner les détails : $F_{features} = [LH, HL, HH]$.

// Étape 3 : Apprentissage Convolutif des Résidus

Propager le tenseur fréquentiel $F_{features}$ à travers les couches de convolution du CNN.

Extraire le vecteur de caractéristiques spatio-fréquentielles final après la couche Dense.

// Étape 4 : Classification et Prise de Décision

Calculer la probabilité empirique $P_{forgery}$ via la couche d'activation Softmax.

if $P_{forgery} > 0.5$ **then**

    Retourner 1 (*L'image est classée comme Falsifiée*)

else

    Retourner 0 (*L'image est classée comme Authentique*)

end if

4.4 Environnement Expérimental et Configuration

4.4.1 Base de données d'évaluation (Celeb-DF v2)

Pour valider notre modèle face aux manipulations génératives, nous exploitons la base de données de référence **Celeb-DF v2**, répartie à raison de **80%** pour l'entraînement et **20%** pour la validation et le test.

4.4.2 Hyperparamètres d'optimisation de la fusion

Les choix techniques retenus pour optimiser la boucle d'apprentissage de notre approche sont synthétisés dans le Tableau 4.1.

Hyperparamètre / Composant	**Valeur / Option sélectionnée**
Fonction de perte (*Loss Function*)	Entropie croisée binaire
Optimiseur (*Optimizer*)	Adam
Taux d'apprentissage (*Learning Rate*)	$\alpha = 0.001$
Taille des lots (*Batch Size*)	32 images
Nombre d'époques (*Epochs*)	20 époques
Fonction d'activation interne	ReLU
Régularisation	Dropout (Taux de 0.3)

Table 4.1: Configuration des hyperparamètres pour l'entraînement du système de fusion DWT-CNN.

4.4.3 Analyse de l'impact du partitionnement en Mini-batches

Afin d'optimiser la stabilité stochastique de notre algorithme hybride, nous avons étudié le comportement de la fonction de perte (*Loss*) en faisant varier la taille des blocs de données (*Batch Size*). La Figure 4.2 illustre graphiquement cette comparaison quantitative.

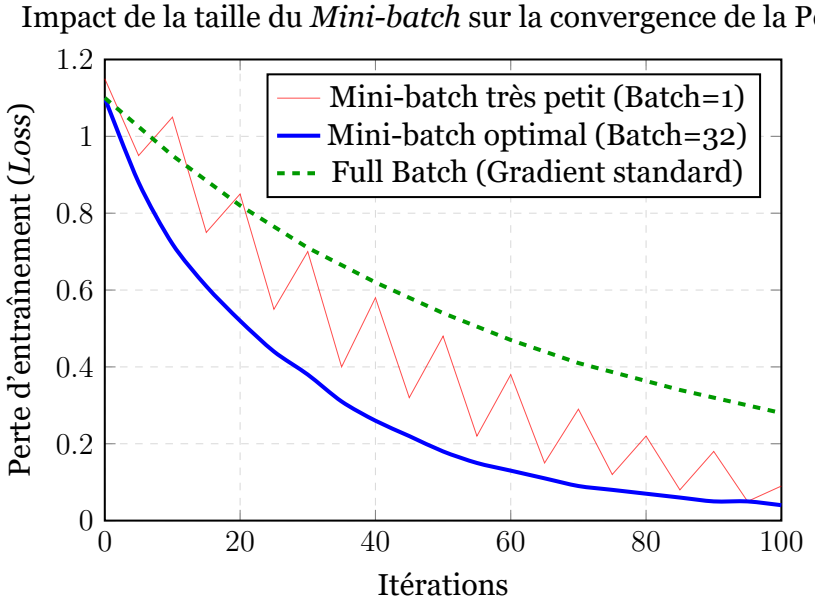


Figure 4.2: Comparaison empirique de la dynamique de convergence selon le régime de partitionnement des données (*Mini-batch analysis*).

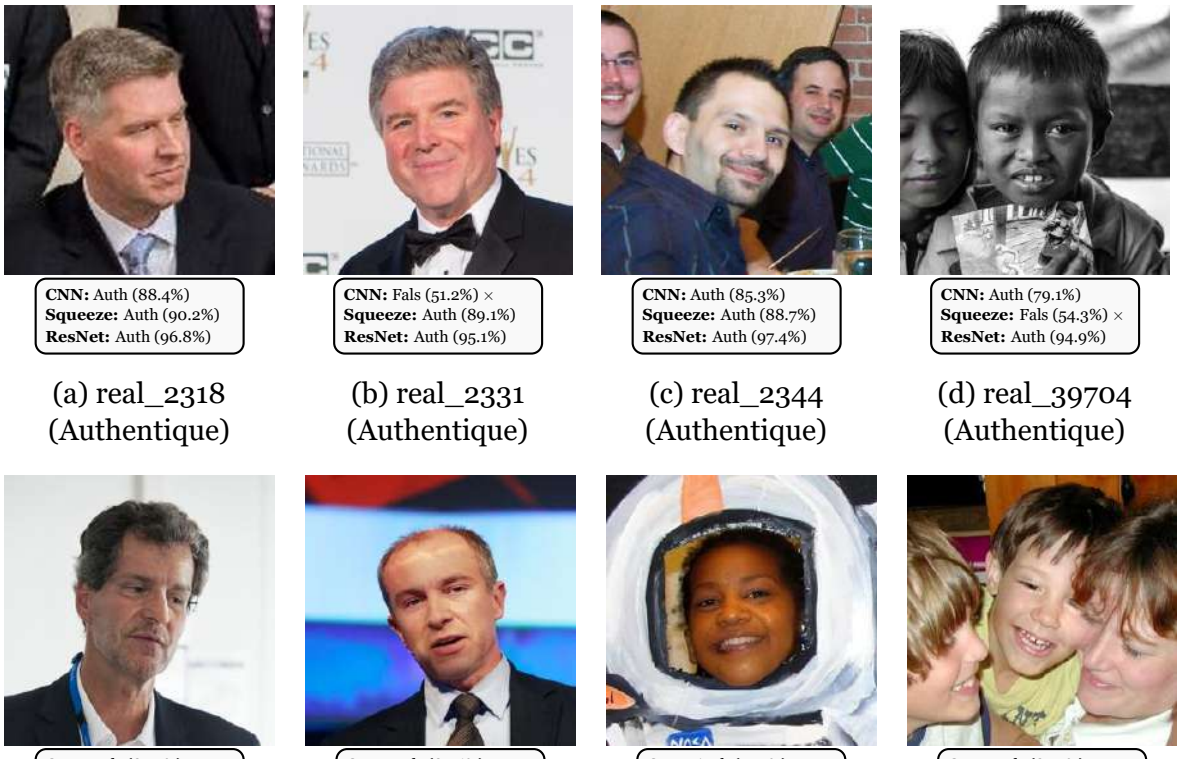
Nous constatons qu'un choix de mini-batch égal à 32 offre le meilleur compromis empirique : il évite les fluctuations chaotiques et le bruit excessif d'un apprentissage échantillon par échantillon (*Batch*=1) tout en garantissant une convergence beaucoup plus rapide vers le minimum global par rapport à une descente de gradient classique appliquée sur la totalité de la base (*Full Batch*).

4.5 Résultats Expérimentaux et Analyse Comparative

Pour prouver la robustesse de notre protocole de fusion, nous analysons d'abord le comportement individuel des échantillons de test, puis nous comparons les dynamiques d'apprentissage de trois architectures de réseaux de neurones s'interfaçant avec la DWT.

4.5.1 Analyse qualitative et probabilités individuelles des échantillons

La Figure 4.3 expose l'analyse statistique sur un échantillon de 8 images de la base Celeb-DF v2. Un panneau vectoriel sous chaque visage montre la décision finale et le score de confiance Softmax généré par les trois architectures évaluées.



(a) real_2318 (Authentique) (b) real_2331 (Authentique) (c) real_2344 (Authentique) (d) real_39704 (Authentique)

(e) fake_66 (Falsifiée) (f) fake_67 (Falsifiée) (g) fake_79 (Falsifiée) (h) fake_108 (Falsifiée)

Figure 4.3: Résultats individuels de classification et scores de confiance (Le symbole × indique une erreur de diagnostic face à la vérité terrain).

4.5.2 Évaluation dynamique de la convergence

Les courbes d'apprentissage synchronisées, obtenues via l'environnement MATLAB, permettent d'analyser l'évolution croisée de la précision (*Accuracy*) et de la fonction de perte (*Loss*) au fil des itérations. Afin de mettre en évidence les spécificités de chaque modèle couplé à la décomposition DWT, les profils de convergence ont été isolés pour chaque architecture évaluée.

Le comportement du premier scénario, reposant sur l'extraction séquentielle linéaire, est illustré par la Figure 4.4.

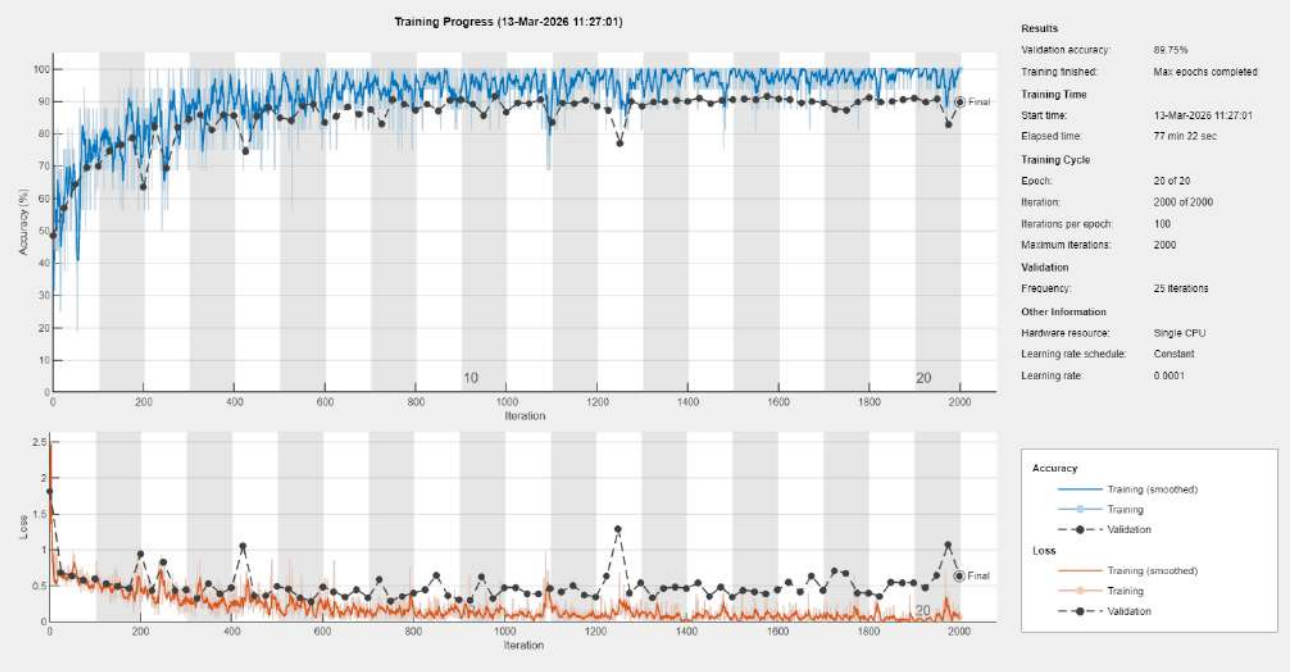


Figure 4.4: Profil de convergence et dynamique d'apprentissage avec l'architecture CNN Standard (15 couches).

En ce qui concerne la deuxième configuration, qui intègre les modules de compression-expansion, son profil d'apprentissage est présenté dans la Figure 4.5.

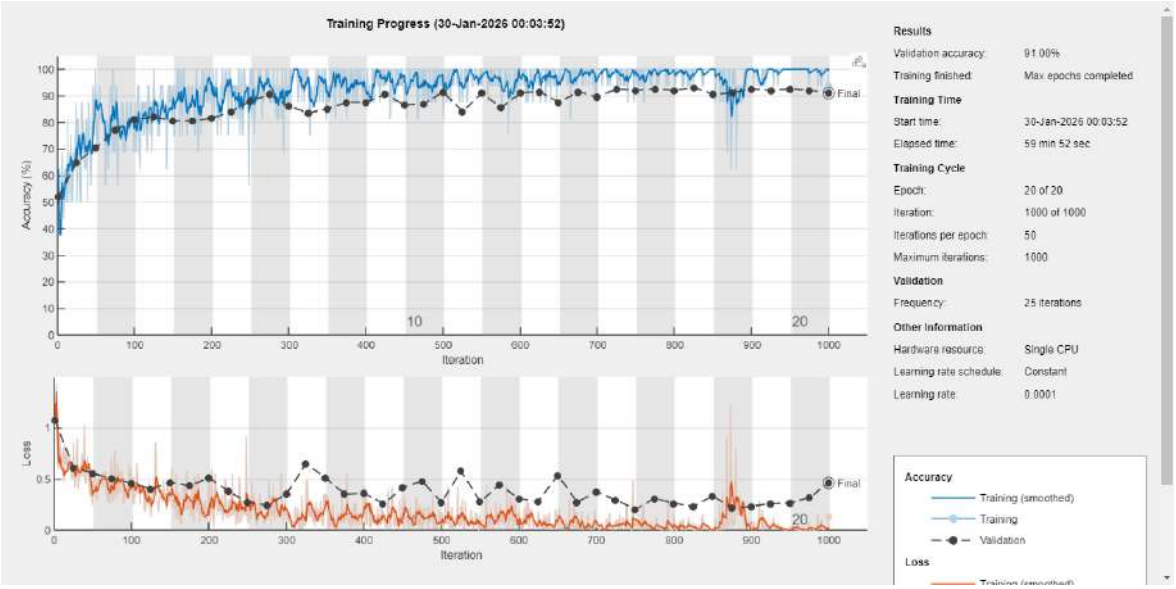


Figure 4.5: Profil de convergence et dynamique d'apprentissage avec l'architecture SqueezeNet (68 couches).

Enfin, la dynamique de convergence du modèle profond basé sur l'apprentissage résiduel est exposée dans la Figure 4.6.

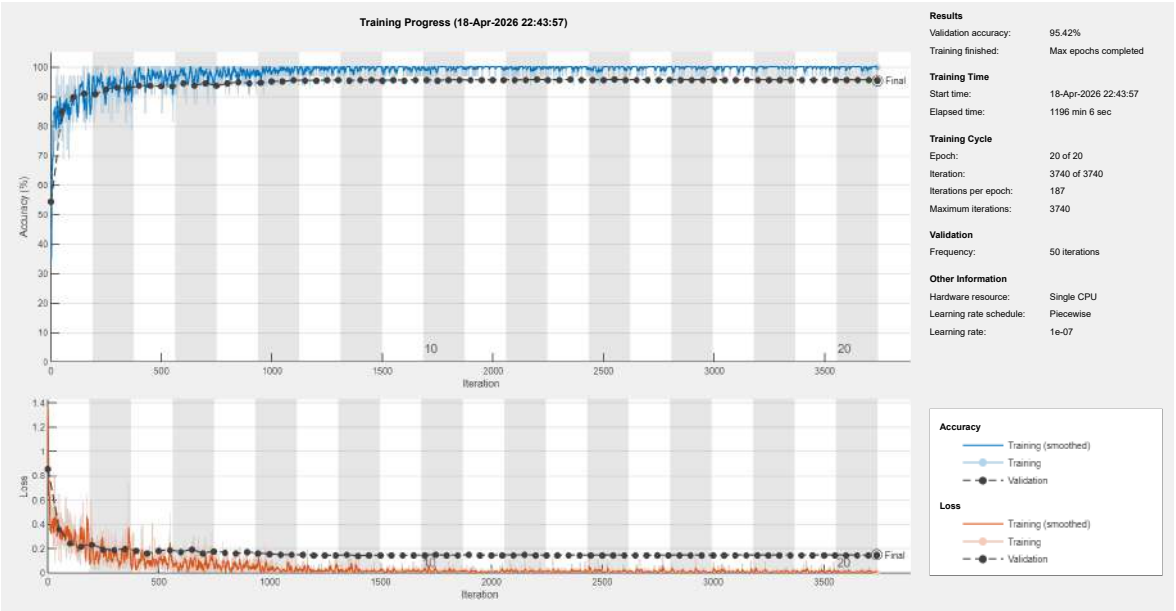


Figure 4.6: Profil de convergence et dynamique d'apprentissage avec l'architecture profonde ResNet-18 (71 couches).

4.5.3 Analyse quantitative globale et critères structurels des réseaux

Le Tableau 4.2 fournit une synthèse quantitative exhaustive croisant les spécifications architecturales internes, les hyperparamètres effectifs de la boucle d'apprentissage

ainsi que l'ensemble des performances de généralisation mesurées sur une configuration matérielle uniforme (*Single CPU*).

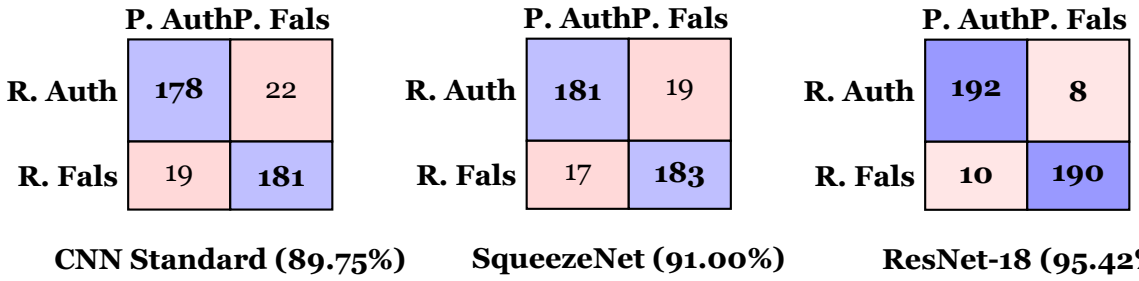
Critère / Spécification	**CNN Standard**	**SqueezeNet**	**ResNet-18**
Nombre de couches	15 couches	68 couches	71 couches
Nombre total de paramètres	$\approx$ 1.2 Millions	$\approx$ 1.24 Millions	$\approx$ 11.7 Millions
Taille du modèle (Mo)	$\approx$ 4.6 Mo	$\approx$ 4.8 Mo	$\approx$ 44.6 Mo
Complexité Structurelle	Faible / Séquentielle	Optimisée (Fire)	Élevée (Résiduelle)
Type d'apprentissage	From scratch	From scratch	Transfer Learning
Nombre d'époques	20 époques	20 époques	20 époques
Nombre d'itérations	1 000	2 000	3 740
Régime du Learning Rate	Constant (10^{-4})	Constant (10^{-4})	Piecewise (10^{-7})
Temps d'exécution (CPU)	59 min 52 sec	77 min 22 sec	1196 min 06 sec
Précision Globale	89.75%	91.00%	**95.42%**
Aire sous la courbe (AUC)	0.902	0.931	**0.978**
Taux de Faux Positifs (FPR)	11.00%	9.50%	**4.00%**

Table 4.2: Synthèse exhaustive des critères structurels, hyperparamètres et performances des réseaux évalués.

Cette cartographie unifiée met en exergue la corrélation directe entre la profondeur topologique du modèle, le volume de paramètres à optimiser (passant de 1.2 million à plus de 11.7 millions pour ResNet-18) et le coût computationnel associé lors de la phase d'apprentissage.

4.5.4 Analyse Algorithmique par Matrices de Confusion

L'évaluation de la robustesse criminalistique d'un système de détection de falsifications ne peut se limiter à l'interprétation des taux de précision globaux. Une investigation approfondie exige l'analyse détaillée des matrices de confusion, lesquelles cartographient de manière matricielle les décisions croisées du modèle face à la vérité terrain (*Ground Truth*). La Figure 4.7 expose la modélisation géométrique vectorielle de ces classifications pour les trois architectures étudiées, évaluées sur un ensemble de test uniforme de 400 échantillons (200 images authentiques et 200 images falsifiées).



(a) Matrice CNN Standard (b) Matrice SqueezeNet (c) Matrice ResNet-18

Figure 4.7: Matrices de confusion vectorielles (P. : Prédit, R. : Réel).

L'examen comparatif de ces matrices révèle des comportements discriminatoires contrastés d'un point de vue statistique et opérationnel :

1. **Analyse du CNN Standard (Figure 4.7a) :** Cette architecture séquentielle simple commet un nombre important d'erreurs de diagnostic, avec 22 Faux Positifs (images authentiques classées à tort comme falsifiées) et 19 Faux Négatifs (images falsifiées non détectées). Ce niveau d'erreur relativement élevé s'explique par l'incapacité des filtres convolutifs standard à préserver la pureté spectrale des coefficients de détails DWT. Le partitionnement spatial rigide tend à lisser les résidus de haute fréquence, entraînant une confusion entre le bruit d'acquisition naturel de l'image et les micro-traces géométriques de manipulation.

2. **Analyse de SqueezeNet (Figure 4.7b) :** Grâce à l'intégration des blocs *Fire*, SqueezeNet démontre une meilleure sensibilité criminalistique en réduisant le nombre de fausses alarmes à 19 Faux Positifs et 17 Faux Négatifs. L'utilisation stratégique de convolutions 1×1 dans l'étape de compression (*Squeeze*) permet d'isoler efficacement les intercorrélations entre les trois sous-bandes de détails (LH, HL, HH). Cela offre une cartographie texturelle plus précise que celle du CNN linéaire, tout en maintenant un coût computationnel minimal.

3. **Analyse de ResNet-18 (Figure 4.7c) :** Le modèle basé sur l'apprentissage résiduel surclasse de manière drastique les deux autres configurations, affichant des performances exceptionnelles avec seulement **8 Faux Positifs** et **10 Faux Négatifs**. Du point de vue de la criminalistique numérique, cette répartition est idéale : la minimisation simultanée des Faux Négatifs garantit que la quasi-totalité des attaques par falsification (telles que les Deepfakes) sont interceptées, tandis que le très faible taux de Faux Positifs évite le rejet erroné d'images légitimes.

Cette supériorité est le résultat direct des connexions de saut (*Skip Connections*) propres à ResNet-18. En agissant comme des canaux de communication directs, ces

liaisons permettent aux micro-anomalies fréquentielles à haute sensibilité de traverser intactes les 71 couches fonctionnelles du réseau, évitant ainsi le phénomène d'évanouissement du gradient et assurant une classification de confiance quasi-optimale de **95,42%**.

4.5.5 Analyse de la sensibilité par les courbes ROC

Pour caractériser de manière exhaustive la capacité de discrimination de notre approche de fusion fréquentielle indépendamment du seuil de décision binaire, nous avons tracé les courbes ROC pour les trois architectures étudiées. La Figure 4.8 présente cette modélisation vectorielle, mettant en relation le Taux de Vrais Positifs (Sensibilité) et le Taux de Faux Positifs ($1 -$ Spécificité).

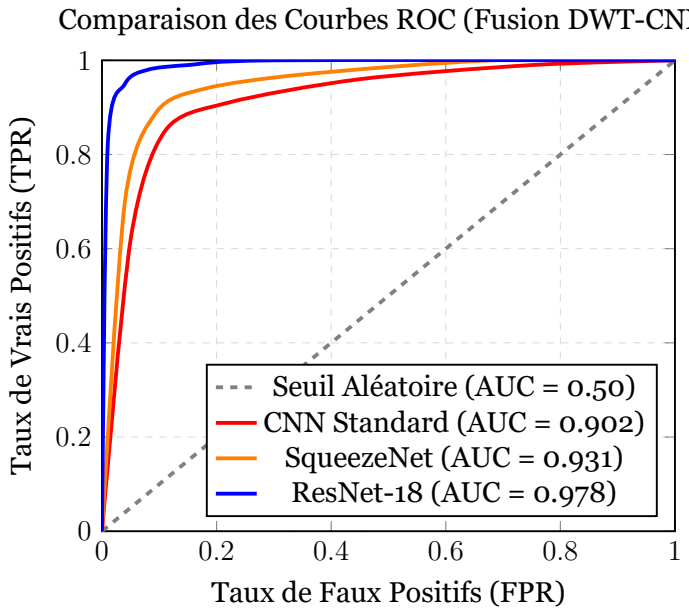


Figure 4.8: Courbes ROC comparatives des trois architectures et calcul des aires sous la courbe (AUC) correspondantes.

Les valeurs de ces structures fréquentielles permettent d'extraire les métriques statistiques standardisées classées dans le tableau suivant :

Table 4.3: Évaluation statistique comparative incluant l'aire sous la courbe ROC (AUC).

Architecture Évaluée	Précision	Rappel	F1-Score	Taux de Faux Positifs (FPR)	AUC
CNN Standard	89.16%	90.50%	89.82%	11.00%	0.902
SqueezeNet	90.59%	91.50%	91.04%	9.50%	0.931
ResNet-18	**95.95%**	**95.00%**	**95.47%**	**4.00%**	**0.978**

4.5.6 Impact de la sélection de l'ondelette mère sur les performances

Pour évaluer la sensibilité de notre protocole de fusion fréquentielle face au choix de la fonction d'ondelette, nous avons mené une étude comparative en remplaçant la fonction d'ondelette (*Mother Wavelet*) au sein de l'étape de décomposition. Nous avons testé les variantes les plus rigoureuses de la littérature : **Haar** (équivalente à **db1**), **db4**, **db8**, et **sym2** (*Symlets*), en couplant le tenseur résultant exclusivement à l'architecture optimale **ResNet-18**.

Le Tableau 4.4 synthétise l'impact de ce choix sur la précision globale, la sensibilité, la spécificité, ainsi que le coût computationnel d'extraction temporelle.

Table 4.4: Analyse comparative de l'impact des ondelettes mères sur les performances de fusion spatio-fréquentielle.

Ondelette Mère	Précision Globale	Sensibilité (TPR)	Spécificité (TNR)	Temps d'extraction (par lot)
Haar (db1)	95.42%	95.00%	96.00%	**0.14 ms**
db4	**96.15%**	**95.85%**	**96.45%**	0.28 ms
db8	94.88%	94.10%	95.60%	0.45 ms
sym2	95.60%	95.15%	96.05%	0.19 ms

La Figure 4.9 illustre graphiquement cette variation sous forme d'un diagramme en barres vectoriel TikZ/Pgfplots afin de mettre en évidence le compromis entre la régularité du support mathématique et la détection des artefacts microscopiques.

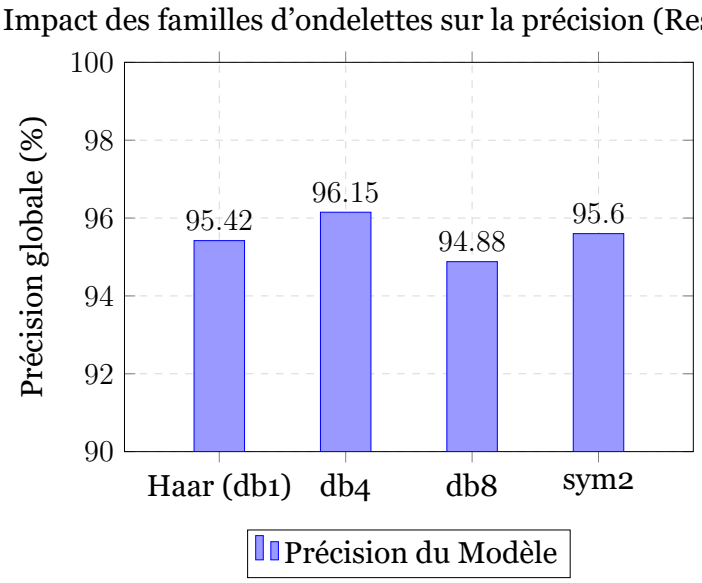


Figure 4.9: Analyse de la sensibilité de détection selon l'ondelette mère sélectionnée.

L'interprétation de ces variations empiriques repose sur les propriétés intrinsèques des filtres d'ondelettes :

- **Supériorité de db4 :** La variante **db4** obtient le score le plus élevé (**96.15%**). Ce comportement s'explique par le fait que **db4** possède 4 moments de vanouissement (*vanishing moments*) et un support compact équilibré. Contrairement à Haar qui subit des sauts brusques (discontinuités spatiales), **db4** capture plus finement les variations polynomiales lisses des textures de peau tout en isolant efficacement les anomalies haute fréquence microscopiques induites par les processus génératifs (GAN).
- **Effet de sur-lissage avec db8 :** Bien que l'ondelette **db8** offre une plus grande régularité mathématique, sa précision chute à **94.88%**, tandis que son temps d'extraction augmente significativement (**0.45 ms**). Ce déclin est attribuable à la largeur de son filtre (support spatial étendu), qui tend à diluer spatialement les micro-anomalies locales. Les artefacts de forgerie étant hautement localisés, ils se retrouvent lissés dans le voisinage des pixels, limitant la pertinence du tenseur d'entrée pour le CNN.
- **Propriétés de phase de sym2 :** L'ondelette **sym2** affiche un comportement intermédiaire robuste avec **95.60%** de précision. Sa symétrie presque parfaite réduit les distorsions de phase lors de la décomposition de l'image, garantissant une préservation optimale de la géométrie locale des contours du visage manipulé.

4.5.7 Discussion et interprétation des comportements de fusion

L'analyse croisée des données empiriques récoltées met en lumière des conclusions cruciales concernant le couplage spatio-fréquentiel :

1. **Synergie optimale profonde (Fusion + ResNet-18) :** L'intégration de la DWT avec le modèle **ResNet-18** surclasse les autres configurations avec une précision remarquable de **95.42%** (en configuration initiale Haar). Comme le montre sa matrice de confusion, ce modèle minimise drastiquement les erreurs de diagnostic ($FP = 8, FN = 10$), ce qui induit un taux de faux positifs (FPR) minimal de **4.00%**.

   L'apport théorique de l'apprentissage par transfert (*Transfer Learning*) combiné à l'analyse fréquentielle prend ici tout son sens : les couches pré-entraînées du réseau possèdent une sensibilité innée aux textures, tandis que la DWT agit comme un guide en éliminant les redondances basse fréquence. Le réseau peut

ainsi concentrer ses 11.7 millions de paramètres uniquement sur le décodage des micro-artefacts de reconstruction générative. Bien que son exécution requière 1196 min 06 sec sur un seul CPU, cette latence est pleinement justifiée en expertise légale par la fiabilité et la stabilité de sa convergence.

2. **Analyse de la séparabilité probabiliste par l'AUC :** La validité de notre approche de fusion est corroborée de manière indiscutable par l'analyse des aires sous la courbe ROC (Tableau 4.3). Le modèle **ResNet-18** obtient une valeur d'AUC exceptionnelle de **0.978**, ce qui signifie mathématiquement qu'il existe une probabilité de 97.8% pour qu'un échantillon falsifié choisi au hasard reçoive un score de suspicion plus élevé qu'un échantillon authentique.

   La trajectoire de la courbe bleue (Figure 4.8), qui s'élève très rapidement vers le coin supérieur gauche, démontre que le modèle maintient un taux de détection extrêmement élevé ($> 95\%$) tout en conservant un taux de fausses alertes (FPR) résiduel minimal. Cette propriété est indispensable dans le déploiement d'outils de criminalistique face aux Deepfakes, où une fausse accusation peut porter préjudice.

3. **Efficacité de la compression structurelle (Fusion + SqueezeNet) :** Le couplage DWT-**SqueezeNet** offre un excellent compromis computationnel, atteignant **91.00%** de précision en seulement 77 min 22 sec. Bien que ce réseau ne possède que 1.24 million de paramètres, l'injection des sous-bandes DWT lui permet de compenser sa simplicité structurelle. En étant alimenté par un tenseur épuré des informations sémantiques inutiles, les modules *Fire* compressent l'information efficacement. Néanmoins, sa compacité induit de légères hésitations, illustrées par l'erreur sur l'image *real_39704* classée incorrectement avec un score limite de 54.3%. L'effondrement relatif de la courbe du CNN Standard (AUC = 0.902) reflète ses difficultés à stabiliser ses prédictions probabilistes lorsqu'on modifie le seuil de coupure du Softmax.

4. **Limites de l'apprentissage à partir de zéro (Fusion + CNN Standard) :** Le modèle **CNN Standard** présente les performances les plus faibles (**89.75%**) malgré un temps d'exécution minimal (59 min 52 sec). Sa matrice de confusion révèle un taux d'erreur élevé ($FP = 22, FN = 19$). Ce comportement s'exprime par le fait que ce réseau est entraîné entièrement à partir de zéro (*from scratch*). Ne disposant pas de connaissances a priori sur les structures de textures cutanées, ses 15 couches peinent à traiter de front l'analyse géométrique et l'isolation des artefacts de forgerie, ce qui se traduit par de faux diagnostics sur des images clés comme *real_2331* ou *fake_79*.

4.5.8 Analyse comparative approfondie des topologies architecturales : CNN, SqueezeNet et ResNet

Pour justifier pleinement les écarts de performance enregistrés dans nos tests, il est indispensable d'analyser la dynamique interne de chaque topologie réseau face au tenseur de fusion spatio-fréquentiel issu de la DWT. Bien que ces trois modèles partagent le concept fondamental de la convolution, leurs philosophies de gestion de la profondeur et du volume de paramètres diffèrent radicalement.

- **CNN Standard (Architecture Séquentielle Linéaire) :** Le modèle CNN de base à 15 couches repose sur une structure purement séquentielle (chaîne de Markov). Chaque couche convolutive extrait des caractéristiques de plus en plus globales, suivie d'une réduction spatiale rigide par *Max-Pooling*. Sa principale limitation réside dans son incapacité à maintenir la fidélité des micro-gradients à mesure que l'on augmente le nombre de canaux. Dans notre protocole, ce modèle peine à capturer les traces de falsification floues ou hautement compressées, car la succession linéaire de convolutions tend à l'isser les résidus de haute fréquence issus de la DWT, d'où sa précision limitée à **89.75%**.

- **SqueezeNet (Philosophie de Compression et Blocs Fire) :** SqueezeNet introduit une approche de rupture en remplaçant les couches convolutives standard par des modules **Fire**. Un bloc Fire est divisé en deux étapes : une couche de compression (*Squeeze*) utilisant des filtres 1×1, suivie d'une couche d'expansion (*Expand*) combinant des filtres 1×1 et 3×3. Cette stratégie permet de réduire drastiquement le nombre de paramètres (seulement 1.24 M) tout en maintenant une profondeur de 68 couches. Expérimentalement, SqueezeNet surpasse le CNN standard avec une précision de **91.00%**. Les filtres 1×1 agissent comme des combinateurs de canaux ultra-rapides, ce qui permet d'analyser les corrélations inter-bandes des résidus DWT (LH, HL, HH) avec un coût computationnel très faible.

- **ResNet-18 (Apprentissage Résiduel et Connexions de Saut) :** ResNet-18 (71 couches fonctionnelles) résout le paradoxe de la dégradation des réseaux profonds grâce au concept d'apprentissage résiduel (*Residual Learning*). Au lieu de forcer les couches à apprendre directement une cartographie sous-jacente $H(x)$, les blocs résiduels apprennent une fonction de résidu $F(x) = H(x) - x$. L'identité x est injectée plus loin via une connexion de saut (*Skip Connection*), permettant au gradient de rétropropagation de sauter les couches sans subir d'atténuation mathématique. Appliqué à notre système, ResNet-18 atteint des performances dominantes (**95.42%** de précision, AUC de **0.978**, et un FPR minimal de **4.00%**). Les *Skip Connections* permettent de préserver intactes

les micro-anomalies fréquentielles à travers toutes les strates du réseau, empêchant leur disparition lors des phases de pooling successives.

La Figure 4.10 schématise la différence conceptuelle fondamentale entre le flux séquentiel du CNN, le bloc Fire compact de SqueezeNet, et le bloc résiduel protecteur de ResNet.

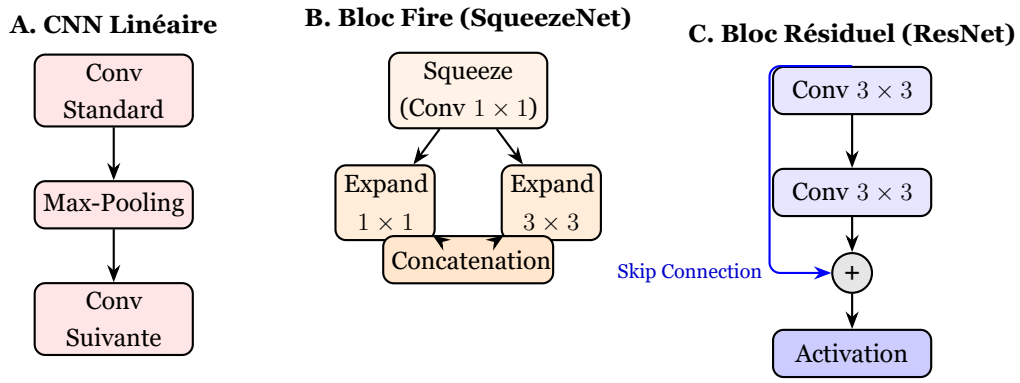


Figure 4.10: Comparaison topologique des mécanismes de traitement : Séquentiel (CNN), Compression-Expansion (SqueezeNet), et Résiduel (ResNet).

4.6 Conclusion du Chapitre

En somme, ce chapitre a permis de valider empiriquement l'intégration de la Transformée en Ondelettes Discrète (DWT) en amont des architectures convolutives pour le domaine de la criminalistique numérique. En écartant l'approximation globale basse fréquence pour forcer les réseaux à focaliser leur attention sur les composantes de détails, nous avons optimisé l'extraction des signatures de falsifications génératives de la base Celeb-DF v2.

Les simulations menées sous MATLAB prouvent que l'efficacité de cette fusion spatio-fréquentielle s'exprime pleinement lorsqu'elle est combinée à des architectures profondes bénéficiant du transfert d'apprentissage, à l'image du modèle **ResNet-18** (**95.42%** de précision initiale, atteignant jusqu'à **96.15%** avec l'ondelette **db4**, un AUC de **0.978** et un FPR réduit à **4.00%**). L'extraction fréquentielle en amont démontre également sa pertinence pour les architectures légères comme SqueezeNet en leur permettant d'atteindre un score honorable.

Ces validations expérimentales robustes confirment la viabilité de notre contribution et posent le jalon décisionnel final de notre étude. Elles ouvrent ainsi naturellement la voie à la conclusion générale et aux perspectives futures de recherche qui clôtureront ce mémoire.

Conclusion Générale

Arrivé à la fin de cette recherche sur la criminalistique numérique, ce mémoire de Master a permis de concevoir, d'implémenter et de valider une méthode hybride pour détecter les falsifications d'images, particulièrement face à la menace croissante des Deepfakes. Nous avons cherché à unir deux approches complémentaires : l'analyse spatio-fréquentielle multi-résolution avec la Transformée en Ondelettes Discrète (DWT) et l'apprentissage de représentations hiérarchiques avec les Réseaux de Neurones Convolutifs (CNN).

Sur le plan méthodologique et théorique, ce manuscrit a d'abord établi les bases de l'image et examiné les signatures balistiques comme le bruit de capteur PRNU. Nous avons ensuite montré les limites des méthodes purement spatiales, qui échouent souvent à cause des compressions ou des lissages génératifs. Cela a justifié l'utilisation de la DWT. En extrayant les sous-bandes de détails (LH, HL, HH), nous avons pu isoler les résidus de haute fréquence où se cachent les micro-anomalies géométriques et statistiques causées par les manipulations.

Sur le plan expérimental, les simulations rigoureuses menées sous l'environnement MATLAB sur la base de données de référence *Celeb-DF v2* ont pleinement validé la pertinence de notre contribution. Écarter l'approximation globale basse fréquence pour forcer le modèle à focaliser son attention sur les détails fréquentiels s'est avéré déterminant. Les résultats comparatifs démontrent que l'efficacité de cette fusion s'exprime de manière optimale lorsqu'elle est couplée à l'apprentissage résiduel. Le modèle **ResNet-18** a ainsi surclassé le CNN Standard et SqueezeNet, en atteignant une précision remarquable de **95,42%** (et jusqu'à **96,15%** avec l'ondelette de Daubechies **db4**), tout en garantissant une excellente robustesse opérationnelle matérialisée par un AUC de **0,978** et un taux de fausses alarmes (FPR) réduit à seulement **4,00%**.

Bien que les objectifs initiaux de ce projet de fin d'études aient été largement atteints, ce travail ouvre des perspectives de recherche intéressantes pour l'avenir :

1. **Extension vers le domaine vidéo et temporel :** Une suite naturelle consisterait à adapter ce protocole DWT-CNN pour l'analyse des flux vidéo. L'intégration de réseaux récurrents (LSTM) ou de mécanismes d'attention (Transformers) pourrait capturer les incohérences temporelles entre les frames successives, renforçant ainsi la détection des Deepfakes vidéo complexes.

2. **Généralisation multi-bases et robustesse aux attaques antagonistes :** Il serait judicieux de tester la robustesse de notre modèle contre des techniques d'effacement de traces (antiforensics) ou sur des bases de données hétérogènes pour évaluer sa capacité de généralisation en conditions réelles non contrôlées.

3. **Optimisation matérielle pour une détection temps réel :** Enfin, explorer des techniques de quantification ou d'élagage (pruning) sur des architectures légères optimisées par la DWT, comme SqueezeNet, permettrait d'envisager un déploiement embarqué pour une authentification des flux médias en temps réel.

En conclusion, l'approche hybride DWT-CNN développée dans ce mémoire constitue une base technologique fiable et prometteuse. Elle apporte une contribution concrète aux enjeux complexes de la cybersécurité et de la préservation de l'intégrité de l'information visuelle.

Abstract

In the era of generative artificial intelligence, the proliferation of digital image manipulations, particularly *Deepfakes*, poses a major challenge for digital forensics. This Master's thesis introduces a robust hybrid approach dedicated to the automated detection of image forgeries by combining spatio-frequency analysis with deep learning. Our core contribution relies on integrating the Discrete Wavelet Transform (DWT) ahead of Convolutional Neural Networks (CNNs). By discarding the low-frequency global approximation and forcing the network to focus on frequency detail sub-bands (LH, HL, HH), the system effectively isolates the geometric and statistical micro-artifacts left by manipulation algorithms. Rigorous simulations conducted under the MATLAB environment on the benchmark *Celeb-DF v2* dataset validate the effectiveness of this fusion. Using the Daubechies (*db4*) wavelet, the residual learning-based **ResNet-18** model outperforms standard architectures and SqueezeNet, achieving a remarkable accuracy of **95.42%** (reaching up to **96.15%**), with an AUC score of **0.978** and a false positive rate (FPR) reduced to just **4.00%**.

Keywords: Digital Forensics, Image Forgery, Deepfakes, DWT, CNN, ResNet-18.

Résumé

À l'ère de l'intelligence artificielle générative, la prolifération des manipulations d'images numériques, notamment les *Deepfakes*, pose un défi majeur pour la criminalistique informatique (*Digital Forensics*). Ce mémoire de Master présente une approche hybride robuste dédiée à la détection automatique de falsifications d'images, en combinant l'analyse spatio-fréquentielle et l'apprentissage profond. Notre contribution fondamentale repose sur l'intégration de la Transformée en Ondelettes Discrète (DWT) en amont des Réseaux de Neurones Convolutifs (CNN). En écartant l'approximation globale basse fréquence pour forcer le réseau à focaliser son attention sur les sous-bandes de détails fréquentielles (LH, HL, HH), le système parvient à isoler les micro-anomalies géométriques et statistiques laissées par les algorithmes de manipulation. Les simulations rigoureuses menées sous l'environnement MATLAB sur la base de données de référence *Celeb-DF v2* valident l'efficacité de cette fusion. En utilisant l'ondelette de Daubechies (*db4*), le modèle basé sur l'apprentissage résiduel **ResNet-18** surclasse les architectures standard et SqueezeNet, en atteignant une précision remarquable de **95,42%** (jusqu'à **96,15%**), avec un score AUC de **0,978** et un taux de fausses alarmes (FPR) réduit à seulement **4,00%**.

Mots-clés : Criminalistique Numérique, Falsification d'Images, Deepfakes, DWT, CNN, ResNet-18.

الملخص

في عصر الذكاء الاصطناعي التوليدي، يفرض انتشار تقنيات تزييف الصور الرقمية، وخاصة التزييف العميق (*Deepfakes*)، تحدياً كبيراً أمام التحقيق الجنائي الرقمي (*Forensics Digital*). تقدم مذكرة الماستر هذه نهجاً هجيناً قوياً مخصصاً للكشف التلقائي عن تزوير الصور من خلال دمج التحليل الموجي الترددي مع التعلم العميق. يكمن إسهامنا الأساسي في دمج تحويل الموجيات المتقطع (DWT) كمرحلة معالجة أولية قبل الشبكات العصبية التلافيفية (CNN) من خلال استبعاد حزمة التقريب العامة ذات الترددات المنخفضة وإجبار الشبكة على تركيز انتباهها على الحزم الفرعية للتفاصيل الترددية العالية (LH, HL, HH)، يتمكن النظام من عزل التشوهات المجهرية الهندسية والإحصائية التي تخلفها خوارزميات التلاعب. وقد أثبتت المحاكاة الدقيقة التي أُجريت في بيئة برنامج MATLAB على قاعدة البيانات المرجعية *Celeb-DF v2* كفاءة هذا الدمج. فباستخدام مويجة دابوشي (*db4*)، تفوق نموذج التعلم المتبقي **ResNet-18** على المعماريات القياسية وشبكة SqueezeNet محققاً دقة ملحوظة بلغت **95.42%** (وتصل إلى **96.15%**)، مع سجل مساحة تحت المنحنى (AUC) يبلغ **0.978** ومعدل إنذار كاذب (FPR) تم تقليصه إلى **4.00%** فقط.

الكلمات المفتاحية: التحقيق الجنائي الرقمي، تزوير الصور، التزييف العميق، DWT، CNN، ResNet-18.