الجمهوريــة الجزائريــة الديمقراطيــة الشعبيــة
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mustapha Stambouli - Mascara
جامعة مصطفى اسطمبولي - معسكر

Faculté des Sciences Exactes
Département d'Informatique

———————————————————————————————————

THÈSE DE DOCTORAT EN SCIENCES

Spécialité : **Informatique**

Présentée par : **BOUDAIEB Ahmed**

Titre de la thèse **:**

Segmentation automatique des régions d'intérêt dans des images réelles

***Devant le jury composé de*:**

DEBBAT Fatima	Pr	Président	Université de Mascara
SALEM Mohammed	Pr	Rapporteur	Université de Mascara
CHEROUATI Brahim	MCA	Examinateur	Université de Mascara
BELFEDHAL Alaa Eddine	MCA	Examinateur	Université de Sidi Bel Abbes
BOUDALI Nourredine	MCA	Examinateur	University Oran 2
BENYAHYA Kadda	MCA	Examinateur	Université de Saida

Année Universitaire : 2025 / 2026

Dédicace

Je dédie cette thèse à l'âme de mon père, que Dieu ait son âme, et à ma mère pour tout son amour, son soutien et ses sacrifices.

À ma femme, pour sa patience, son encouragement et sa présence constante tout au long de ce parcours scientifique.

À mes frères et sœurs, pour leur affection et leur inspiration.

À toute ma famille, pour leur appui indéfectible et leur confiance.

À tous mes amis et camarades, pour leur soutien et les moments partagés qui ont enrichi cette expérience.

Remerciements

Tout d'abord, je tiens à remercier Allah de m'avoir accordé la capacité et la force nécessaires pour mener à bien cette thèse. Louange à Allah, Seigneur des mondes.

Je souhaite également exprimer ma profonde gratitude à mon directeur de thèse, le Professeur SALEM Mohammed pour son encadrement, ses conseils et son accompagnement tout au long de mes études doctorales.

Enfin, j'adresse mes sincères remerciements à l'ensemble des membres du jury (Pr. DEBBAT Fatima, Dr. CHEROUATI Brahim, Dr. BELFEDHAL Alaa Eddine, Dr. BOUDALI Nourredine et Dr. BENYAHYA Kadda pour avoir accepté de lire et d'évaluer le présent manuscrit.

Je remercie, enfin, toute ma famille et mes amis pour leur patience, leur soutien et leurs encouragements, et je remercie également toutes les personnes ayant contribué, de près ou de loin, à la réussite de ce travail

Résumé

Cette thèse porte sur la segmentation automatique des régions d'intérêt dans des images réelles, avec une application spécifique à la détection des lésions cutanées. Nous proposons une architecture de segmentation basée sur U-Net améliorée par un encodeur ResNet50V2, une stratégie d'augmentation avancée et une fonction de perte hybride Dice + BCE. Le modèle a été évalué sur le dataset PH2 puis testé sur un jeu de données externe (ISIC 2016) afin d'analyser sa robustesse et sa capacité de généralisation. Les résultats obtenus montrent des performances élevées en segmentation, surpassant plusieurs méthodes traditionnelles et modèles profonds de référence. Les analyses menées, incluant une étude d'ablation et une comparaison complète avec l'état de l'art, confirment l'efficacité du modèle proposé pour la segmentation précise des lésions cutanées dans des conditions réelles.

Mots-clés : Segmentation d'images, Apprentissage profond, Lésions cutanées, U-Net, ResNet50V2, Augmentation de données, Dice loss, Généralisation.

الملخص

تتناول هذه الأطروحة موضوع التقسيم الآلي للمناطق ذات الاهتمام في الصور الحقيقية، مع تطبيق خاص على كشف وتحديد آفات الجلد. نقترح نموذجاً محسّناً يعتمد على بنية U-Net مع مُشفِّر ResNet50V2 ، وتقنيات متقدمة لزيادة البيانات، ودالة خسارة هجينة تجمع بين Dice و .BCEتم تقييم النموذج على قاعدة بيانات PH2 واختباره لاحقاً على قاعدة بيانات خارجية (ISIC 2016) لقياس قدرته على التعميم ومقاومة التغيّرات. أظهرت النتائج أداءً مرتفعاً في عملية التقسيم متفوقاً على العديد من الطرق التقليدية والنماذج العميقة الموجودة في الأدبيات. كما تؤكد الدراسة الشاملة، بما فيها دراسة الاستئصال والمقارنة مع أحدث الأعمال، فعالية النموذج المقترح في التقسيم الدقيق لآفات الجلد في الظروف الواقعية.

.زيادة البيانات، خسارة ديس، التعميم ،ResNet50V2، U-Net ،تقسيم الصور، التعلم العميق، آفات الجلد

Abstract

This thesis addresses the challenge of automatic segmentation of regions of interest in real-world images, with a particular focus on skin lesion delineation in dermoscopic data. We propose a robust deep learning framework based on an enhanced ResNet50V2–U-Net architecture integrating data augmentation and a hybrid Dice–BCE loss to improve boundary precision and class-imbalance handling. Extensive experiments on the PH2 dataset demonstrate the effectiveness of the proposed model, supported by an ablation study and comparisons with traditional segmentation techniques and widely used deep architectures. External evaluation on the ISIC 2016 dataset confirms the model's ability to generalize to heterogeneous and artifact-degraded images. The findings validate the relevance of the proposed approach and highlight its potential for reliable and accurate lesion segmentation in real-world clinical contexts.

Keywords: Image segmentation, Deep learning, Skin lesions, U-Net, ResNet50V2, Data augmentation, Dice loss, Generalization.

Table de matières

Liste des tableaux

Liste des figures

Liste des abréviations et acronymes

Abréviation	Signification
ABCDE	Asymétrie, Bords, Couleur, Diamètre, Évolution
BCE	Binary Cross-Entropy
CAD	Computer-Aided Diagnosis (Diagnostic Assisté par Ordinateur)
CADe	Computer-Aided Detection
CADx	Computer-Aided Diagnosis (caractérisation)
CNN	Convolutional Neural Network
CT	Computed Tomography (Tomodensitométrie)
DWI	Diffusion Weighted Imaging
EPR	Electron Paramagnetic Resonance
FCN	Fully Convolutional Network
FLAIR	Fluid Attenuated Inversion Recovery
GAN	Generative Adversarial Network
IoU	Intersection over Union
IRM	Imagerie par Résonance Magnétique
ISIC	International Skin Imaging Collaboration
MRF	Markov Random Field
OAR	Organs At Risk
PH2	Dermoscopic Image Dataset PH²
PTV	Planning Target Volume
ResNet	Residual Neural Network
ROI	Region of Interest (Région d'Intérêt)
RX	Rayons X
SegNet	Segmentation Network
SPECT	Single Photon Emission Computed Tomography
T1	Séquence IRM pondérée T1
T2	Séquence IRM pondérée T2
TEP	Tomographie par Émission de Positons
U-Net	Convolutional Network for Biomedical Segmentation

Introduction Générale

Introduction

La segmentation automatique des régions d'intérêt (ROI) dans des images réelles constitue un enjeu central en imagerie médicale et en vision par ordinateur. L'accroissement massif des données rend l'analyse manuelle coûteuse et sujette à une forte variabilité inter-observateurs, soulignant la nécessité de méthodes fiables et reproductibles pour isoler les structures pertinentes, telles que les tumeurs ou les lésions cutanées. En dermatologie, la détection précoce du mélanome repose sur la dermoscopie, qui révèle des structures invisibles à l'œil nu, mais les images présentent une grande variabilité de forme, de couleur et de contraste, ainsi que des artefacts tels que poils, reflets ou ombres, compliquant l'extraction précise des régions d'intérêt. Les approches traditionnelles, basées sur le seuillage, les contours ou la croissance de régions, montrent leurs limites face à cette complexité. Les architectures de Deep Learning, comme U-Net et ses variantes, offrent des représentations hiérarchiques capables de capturer les détails fins, mais restent sensibles au déséquilibre de classes et à la rareté des données annotées. Cette thèse vise à développer une méthode hybride intégrant un encodeur profond ResNet50V2 dans un U-Net, combinée à des fonctions de perte adaptées et à des stratégies avancées d'augmentation de données, afin d'assurer précision, robustesse aux artefacts et généralisation vers des ensembles externes comme ISIC 2016. L'approche proposée ambitionne de fournir un outil automatique fiable pour la segmentation de lésions cutanées, améliorant la reproductibilité et l'efficacité du diagnostic assisté par ordinateur.

Contexte général de la recherche et motivations scientifiques

L'extraction et la délimitation automatiques des régions d'intérêt (ROI) dans des images réelles constituent un pilier de la vision par ordinateur, notamment en imagerie médicale [1], [2]. L'augmentation massive du volume de données rend l'analyse manuelle à la fois coûteux en temps et sujette à une forte variabilité inter-observateurs [3], d'où l'importance de la segmentation automatique, qui représente une étape centrale pour l'analyse clinique. En isolant précisément la ROI, notamment lorsqu'il s'agit d'une tumeur, d'une lésion ou d'un organe, il devient possible d'en extraire des caractéristiques quantitatives essentielles au diagnostic ou à la classification, telles que la taille, la forme, la texture ou la couleur [4]. Cette automatisation contribue également à standardiser l'interprétation des images, en réduisant la subjectivité humaine et en améliorant la reproductibilité des résultats [5], permettant ainsi un gain d'efficacité clinique en diminuant le temps requis pour le diagnostic.

Dans ce travail, nous nous intéressons à la segmentation des lésions cutanées sur images dermoscopiques. Le cancer de la peau, et plus particulièrement le mélanome, compte parmi les cancers les plus fréquents dans le monde, et son pronostic dépend fortement d'un diagnostic précoce [6]. La dermoscopie est une technique d'imagerie non invasive qui révèle des structures épidermiques et dermo-épidermiques invisibles à l'œil nu. Son interprétation repose sur des critères morphologiques et pigmentaires, comme ceux de la règle ABCDE, ce qui exige une délimitation précise de la lésion. Cette opération est cependant difficile et constitue un excellent terrain pour l'évaluation des méthodes d'apprentissage profond [7]. Les lésions présentent une grande variabilité de forme, de taille et de couleur [8]. Le contraste entre la lésion et la peau

saine est souvent faible [9]. De plus, divers artefacts d'acquisition , tels que poils, reflets, ombres, bulles d'air, compliquent davantage l'analyse [10]. Dans ce contexte, disposer d'un outil de segmentation automatique fiable et robuste est essentiel pour améliorer l'objectivité et la reproductibilité de l'analyse dermoscopique dans les systèmes de diagnostic assisté par ordinateur (CAD) [11].

Formalisation des problématiques

La segmentation en conditions réelles constitue un défi majeur, car les images cliniques sont acquises dans des environnements très hétérogènes. La présence d'artefacts complexes, tels que les poils ou les reflets, perturbe fortement l'analyse et exige des algorithmes capables d'isoler la lésion même lorsque certaines zones sont partiellement masquées. Les modèles d'apprentissage profond tentent d'extraire des caractéristiques sémantiques permettant de différencier la lésion de ces perturbations, mais l'acquisition de cette invariance demeure difficile [9].

À cela s'ajoute le problème du décalage de domaine. Les modèles entraînés sur un jeu de données comme PH² [12] voient leurs performances diminuer lorsqu'ils sont appliqués à un autre ensemble, tel qu'ISIC 2016 [13], en raison des différences d'équipement, d'éclairage ou de protocole. La gestion de ce domain shift est donc essentielle pour apprendre des représentations réellement invariantes [14].

La segmentation est également confrontée au déséquilibre des classes, car la région d'intérêt occupe souvent une très faible portion de l'image, généralement moins de 10 %. Ce déséquilibre rend les fonctions de perte classiques, telles que la Cross-Entropy, inefficaces : le modèle privilégie l'arrière-plan et néglige les pixels minoritaires, entraînant des contours incomplets ou imprécis [15]. Pour y remédier, il est nécessaire d'utiliser des fonctions de perte mieux adaptées à l'imagerie médicale, comme le Dice Loss ou le Focal Loss, qui pondèrent davantage les erreurs dans les zones cliniquement importantes [16].

Un autre obstacle majeur est la rareté des données annotées. Le Deep Learning nécessite un volume important d'exemples, alors que l'annotation pixel par pixel est coûteuse et requiert l'intervention d'experts [17]. Les bases existantes, telles que PH², limitées à 200 images, exposent fortement les modèles au surapprentissage : ceux-ci mémorisent les images d'entraînement au lieu d'apprendre des caractéristiques généralisables. Pour contourner cette limite, il est indispensable de recourir à des stratégies avancées telles que l'apprentissage par transfert [18] et des méthodes d'augmentation de données sophistiquées [19].

Avant l'avènement du Deep Learning, la segmentation reposait sur des méthodes déterministes utilisant des informations locales comme l'intensité ou le gradient [20]. Ces techniques deviennent rapidement insuffisantes lorsque l'éclairage est irrégulier ou que la lésion présente un faible contraste. Les détecteurs de contours échouent face aux frontières floues ou perturbées par les artefacts, tandis que les modèles déformables sont sensibles à l'initialisation et peuvent converger vers des solutions erronées. Les méthodes régionales, telles que Chan–Vese ou la croissance de régions, peinent à gérer l'hétérogénéité interne des lésions. Les approches de clustering comme k-means sont instables, et les champs de Markov

nécessitent un paramétrage complexe rarement adapté aux images dermoscopiques réelles. Ces limites découlent d'un problème commun : ces méthodes se basent sur des critères locaux et manquent d'une compréhension sémantique globale, indispensable pour traiter les artefacts, les variations d'éclairage et les structures complexes.

L'émergence des architectures de Deep Learning a permis de surmonter certaines de ces limites [21]. Les premières architectures de segmentation, comme les Fully Convolutional Networks (FCN) [22], ont marqué une avancée importante, mais elles souffraient d'une perte de résolution spatiale due au pooling successif, produisant des contours lissés et imprécis. SegNet [23] a tenté de pallier ce problème en exploitant les indices de max-pooling pour améliorer l'upsampling, mais son encodeur restait trop peu profond pour capturer la complexité visuelle des lésions. L'arrivée de U-Net [24], avec sa structure symétrique et ses skip connections, a constitué une avancée majeure en préservant les détails fins. Toutefois, la faible profondeur de son encodeur limite encore sa capacité à traiter efficacement l'hétérogénéité et les artefacts, et U-Net tend à surapprendre sur les petites bases médicales [25]. Cette situation a rendu nécessaire l'adoption d'encodeurs plus profonds et expressifs, tels que les réseaux résiduels (ResNet) [26], capables de capturer des relations sémantiques complexes, de faciliter l'apprentissage et d'améliorer la généralisation.

Objectifs spécifiques et contributions de la thèse

L'objectif principal de cette thèse est de développer un système de segmentation des lésions cutanées capable d'atteindre une haute précision, de résister aux artefacts dermoscopiques et de généraliser à des ensembles de données externes. Elle vise plus largement à proposer une approche automatique pour la segmentation des régions d'intérêt dans des images réelles, en s'appuyant sur un cas clinique concret : les images dermoscopiques.

Les objectifs spécifiques sont les suivants :

Optimisation architecturale : Concevoir une architecture de segmentation performante en intégrant un encodeur profond (ResNet50V2) dans un U-Net classique, afin de mieux extraire les caractéristiques discriminantes et préserver les détails fins grâce aux skip connections.

Gestion du déséquilibre de classes : Développer et implémenter des fonctions de perte adaptées (Dice Loss et variantes) pour attribuer un poids plus important aux pixels minoritaires et améliorer la précision des contours.

Robustesse aux artefacts : Mettre en place une stratégie avancée d'augmentation photométrique pour simuler les perturbations visuelles (poils, reflets, variations chromatiques), renforçant la capacité du modèle à apprendre des caractéristiques invariantes.

Validation de la généralisation : Évaluer le modèle sur un jeu de données externe (ISIC 2016) afin de vérifier sa transférabilité et son potentiel clinique en conditions réelles.

Contributions scientifiques principales

Cette thèse apporte quatre contributions majeures à la segmentation automatique en imagerie médicale. Premièrement, l'intégration d'un encodeur profond ResNet50V2 pré-entraîné dans un U-Net [27] permet de transférer des connaissances visuelles et d'améliorer la qualité des représentations sur des jeux de données réduits, tout en conservant la profondeur et les skip connections pour une meilleure extraction des détails. Deuxièmement, l'adoption de fonctions de perte adaptées au déséquilibre de classes, combinant Dice Loss, Focal Loss et BCE pondérée, optimise la segmentation des pixels minoritaires et préserve les contours complexes. Troisièmement, l'étude rigoureuse des augmentations de données avancées, incluant les transformations réalistes, renforce la robustesse du modèle face aux variations visuelles et aux images hétérogènes. Enfin, L'évaluation sur le jeu de données externe ISIC 2016 démontre que la méthode conserve ses performances hors du jeu d'entraînement et peut être déployée efficacement dans des environnements cliniques variés.

Organisation de la thèse

La thèse est structurée en cinq chapitres, suivant une progression logique de l'établissement du contexte à la validation expérimentale. Le **Chapitre 1** introduit le sujet, présente la problématique, les objectifs et les contributions principales. Le **Chapitre 2** décrit le contexte clinique, la dermoscopie et les méthodes traditionnelles de segmentation, en soulignant leurs limites face aux images réelles. Le **Chapitre 3** présente les fondements du Deep Learning, les architectures de segmentation avancées (FCN, SegNet, U-Net, ResNet) et leurs limites pour la segmentation de lésions cutanées, justifiant l'approche hybride proposée. Le **Chapitre 4** détaille la méthodologie, l'architecture U-Net/ResNet50V2, les fonctions de perte, les jeux de données PH² et ISIC 2016, ainsi que les résultats expérimentaux et leur discussion. Enfin, le **Chapitre 5** conclut la thèse, récapitule les principaux résultats et ouvre sur des perspectives futures, telles que l'attention, la segmentation multi-classe et l'apprentissage auto-supervisé.

Chapitre I: Segmentation des Images Réelles: Approches Classiques

I.1 Introduction

La segmentation d'images constitue une étape fondamentale de l'analyse numérique, dont la précision est essentielle à la fiabilité des applications en aval, notamment dans le domaine critique de l'imagerie médicale [28]. Formellement, elle résout le problème du partitionnement de l'image en sous-régions disjointes afin d'extraire la Région d'Intérêt (ROI), caractérisant ainsi l'image brute en une représentation sémantique structurée. La qualité de cette délimitation est mesurée par des métriques comme le Dice ou l'IoU. Cette opération est un préalable indispensable qui conditionne des applications majeures telles que l'extraction de caractéristiques quantitatives (*Radiomics*) pour les systèmes d'aide au diagnostic (CAD) [29], le suivi longitudinal des pathologies pour évaluer l'efficacité thérapeutique [30], et la planification interventionnelle (radiothérapie, chirurgie) où la précision est vitale [31].

La segmentation des images cliniques réelles est certainement plus complexe que celle des images synthétiques, en raison de la présence simultanée de bruit, de variabilité et de structures complexes. Parmi les éléments qui rendent cette tâche complexe, on trouve la variabilité biologique inter- et intra-patient qui nuit à la capacité de généralisation des modèles [32]. De plus, les artefacts récurrents (tels que les poils en dermoscopie ou les ombres) et le bruit dégradent le contraste des images. Les méthodes basées sur les gradients sont souvent inefficaces, car les contours réels sont ambigus et de faible contraste. Enfin, la non-uniformité et la complexité des textures pathologiques constituent une rupture avec l'hypothèse d'homogénéité sur laquelle reposent de nombreuses méthodes traditionnelles.

Face à ces défis inhérents aux images réelles, ce chapitre poursuit un triple objectif : premièrement, contextualiser l'écosystème de l'imagerie médicale, notamment la dermoscopie, pour bien définir la nature des données. Deuxièmement, réaliser une Analyse Rétrospective des méthodes de segmentation traditionnelles, en explicitant leurs faiblesses dues à leur dépendance aux caractéristiques de bas niveau. Troisièmement, justifier le changement de paradigme vers l'apprentissage profond en identifiant les limitations fondamentales de ces approches.

I.2 L'Imagerie Médicale

L'imagerie médicale constitue un pilier essentiel du diagnostic et de la recherche clinique, en permettant la visualisation non invasive des structures anatomiques, des tissus biologiques et des processus fonctionnels. Elle se caractérise par une grande diversité de modalités d'acquisition, chacune reposant sur un principe physique distinct et fournissant des informations complémentaires. La compréhension de ces modalités est indispensable pour appréhender les défis de la segmentation automatique, qui joue un rôle central dans la précision du diagnostic assisté par ordinateur (CAD).

I.2.1 Diversité des Modalités d'Acquisition

La Figure I.1 illustre la diversité des principales techniques d'imagerie médicale, telles que la radiographie, la tomodensitométrie (CT), l'imagerie par résonance magnétique (IRM), la tomographie par émission de positons (TEP), l'échographie, la dermoscopie, la microscopie,

l'imagerie optique et la résonance paramagnétique électronique (EPR). Ces approches permettent de couvrir différents niveaux d'observation, allant de la cellule jusqu'à l'organe entier.

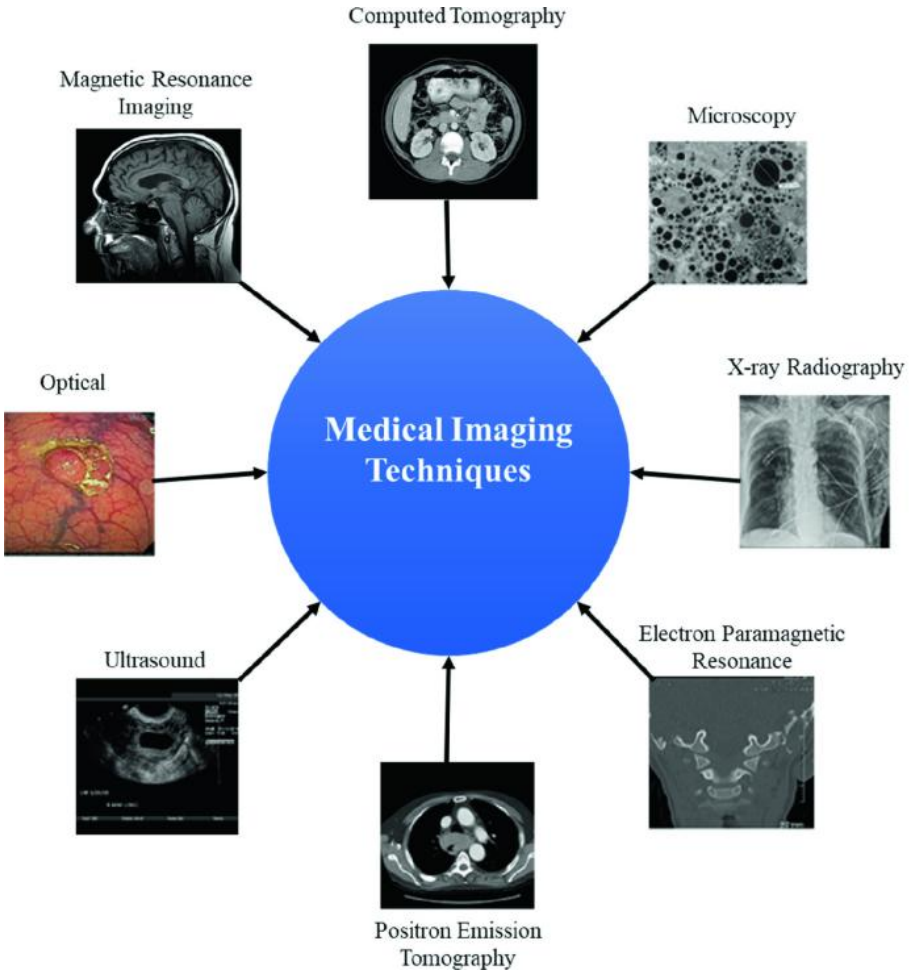


Figure I.1 Techniques d'Imagerie Médicale (IRM, Tomodensitométrie, Radiographie, Échographie, Tomographie par Émission de Positons, etc.),[33]

I.2.1.1 Radiographie (Rayons X)

Basée sur l'atténuation différentielle des rayons X, la radiographie fournit des images bidimensionnelles à haute résolution spatiale, adaptées à l'analyse des structures osseuses. Toutefois, le faible contraste des tissus mous et la superposition anatomique limitent la précision de la segmentation automatique. Les variations d'exposition et le bruit imposent l'emploi de techniques robustes de correction et de normalisation [34].

I.2.1.2 Tomodensitométrie (CT)

La tomodensitométrie combine des projections radiographiques multiples pour reconstruire des volumes tridimensionnels. Elle offre une excellente résolution spatiale et un

bon contraste pour les tissus denses. Néanmoins, les artefacts de mouvement, la variabilité d'intensité et la complexité des tissus mous rendent la segmentation multi-organes difficile, nécessitant des approches de deep learning et de fusion multimodale [35].

I.2.1.3 Imagerie par Résonance Magnétique (IRM)

Reposant sur les propriétés magnétiques des noyaux d'hydrogène, l'IRM produit des images de haute qualité des tissus mous, sans rayonnement ionisant. La diversité des séquences (T1, T2, FLAIR, DWI) offre une richesse d'information, mais la variabilité des protocoles, le bruit et les artefacts imposent des méthodes adaptatives combinant réseaux neuronaux et mécanismes d'attention [36].

I.2.1.4 Tomographie par Émission de Positons (TEP/SPECT)

Ces techniques reposent sur l'utilisation de radiotraceurs émetteurs de photons gamma pour visualiser l'activité métabolique. Bien que riches en informations fonctionnelles, elles souffrent d'une faible résolution spatiale et d'un rapport signal/bruit limité. La fusion avec CT ou IRM est donc indispensable pour améliorer la précision anatomique [37].

I.2.1.5 Échographie

L'échographie exploite les ondes ultrasonores pour produire des images en temps réel, offrant un outil portable et sans rayonnement. Cependant, les images sont souvent altérées par le bruit speckle et dépendent fortement de l'opérateur. Les modèles récents de deep learning, notamment les réseaux antagonistes génératifs (GANs) et les modèles de diffusion, visent à pallier ces limitations [38]. Le tableau I-1 fait une synthèse comparative de ces types d'imagerie :

Tableau I-1: caractéristiques et défis des imageries médicales

Modalité	Caractéristiques des images	Défis de segmentation
Rayons X	Haute résolution, faible contraste des tissus mous	Superposition, faible contraste
CT	Reconstruction 3D, bon contraste pour les tissus denses	Artefacts, variabilité, segmentation multi-organes
IRM	Contraste élevé des tissus mous, multi-séquences	Variabilité des protocoles, artefacts, multimodalité
TEP/SPECT	Informations fonctionnelles, faible résolution spatiale	Faible rapport signal/bruit, fusion multimodale nécessaire
Échographie	Imagerie en temps réel, portable	Bruit speckle, dépendance opérateur
Dermoscopie	Haute résolution, fort contraste colorimétrique	Variabilité des formes, artefacts, manque d'annotations

Chaque modalité d'imagerie offre une vision complémentaire de la réalité anatomique et fonctionnelle, mais présente des défis spécifiques pour la segmentation automatique. La diversité inter-modale, la variabilité d'intensité, les artefacts et le manque d'annotations fiables demeurent des obstacles majeurs à la généralisation des modèles. Ces limites soulignent la nécessité d'approches hybrides et intelligentes, capables d'exploiter simultanément l'information spatiale et contextuelle. C'est précisément dans cette perspective que les architectures profondes de nouvelle génération, telles que U-Net, ResNet et leurs variantes attentionnelles, se sont imposées comme des solutions prometteuses en imagerie médicale.

I.2.2 Enjeux communs de l'analyse d'images médicales

L'analyse computationnelle des images médicales, quelle que soit la modalité d'acquisition, doit relever des défis récurrents tout en répondant à des impératifs cliniques stricts. Ces enjeux concernent à la fois la précision de la segmentation, la qualité des images et la variabilité des données, facteurs déterminants pour la fiabilité du diagnostic et la planification thérapeutique.

I.2.2.1 Précision et reproductibilité

La précision en segmentation constitue un impératif clinique. Une erreur peut entraîner une mauvaise stadification de la tumeur, par sous- ou surestimation du volume tumoral, ou compromettre la planification thérapeutique. En radiothérapie, une délimitation inexacte du Volume Cible Planifié (PTV) peut conduire à une sous-irradiation de la tumeur ou à une irradiation excessive des Organes à Risque (OAR) [39]. La reproductibilité inter- et intra-observateur est essentielle pour assurer un suivi longitudinal fiable de l'évolution des pathologies et garantir la comparabilité temporelle des examens.

I.2.2.2 Dégradations des images : bruit, faible contraste et artéfacts

Les images médicales présentent souvent des perturbations qui limitent l'efficacité des méthodes traditionnelles de segmentation. Le bruit varie selon la modalité : quantique en CT/RX, de Rice en IRM, speckle en échographie [40], masquant les contours fins et imposant un compromis entre qualité d'image et dose de radiation. En outre, le faible contraste et l'ambiguïté des frontières compliquent la détection des régions d'intérêt (ROI), notamment les lésions faiblement démarquées ou les tissus mous adjacents. En plus de ces dégradations, les artefacts spécifiques à chaque modalité, comme les artefacts métalliques en CT, de mouvement ou de susceptibilité en IRM, ou encore les poils et bulles en dermoscopie, peuvent être confondus avec des structures pathologiques et introduire des erreurs systématiques.

I.2.2.3 Hétérogénéité et variabilité inter-patients

L**'**hétérogénéité des données résulte des différences entre protocoles, équipements et conditions d'acquisition, tandis que la variabilité inter-patients reflète les différences anatomiques, physiologiques ou pathologiques. Dans le cas des lésions cutanées, couleur, texture et contours varient fortement selon le phototype et la nature de la lésion, rendant la segmentation automatique plus complexe.

Pour atténuer ces problèmes, plusieurs stratégies sont proposées, notamment la normalisation des intensités, l'harmonisation inter-institutions, et les techniques d'augmentation de données destinées à accroître la diversité du jeu des données. Par conséquent, des approches d'adaptation de domaine sont nécessaires pour réduire les écarts entre distributions de données issues de différentes sources, et améliorer par la suite la robustesse et la généralisation des méthodes de segmentation.

I.3 Outils d'Aide au Diagnostic (CAD)

Les systèmes d'Aide au Diagnostic par Ordinateur (CAD – Computer-Aided Diagnosis) ont émergé pour répondre à l'augmentation exponentielle du volume de données médicales et à la nécessité d'une analyse plus rigoureuse et reproductible [41]. Ils visent à soutenir le clinicien dans la détection, la segmentation, la caractérisation et la classification des anomalies, en réduisant la subjectivité liée à l'interprétation visuelle.

I.3.1 Objectivation et Standardisation

Les systèmes CAD permettent l'objectivation du diagnostic grâce à l'extraction de mesures quantitatives issues des images médicales; une approche connue sous le nom de Radiomics [42]. Ces mesures (forme, texture, intensité, bordure) fournissent des indicateurs objectifs et reproductibles, réduisant ainsi la variabilité inter- et intra-observateurs qui caractérise souvent les évaluations manuelles des cliniciens.

I.3.2 Architecture Typique d'un Système CAD

Un système CAD suit généralement une architecture hiérarchique organisée en plusieurs étapes de traitement. La Figure I.2 illustre Architecture générale d'un système d'Aide au

Diagnostic par Ordinateur (CAD), de l'acquisition des données à la prédiction finale de la classe (cancéreux ou non cancéreux).

.

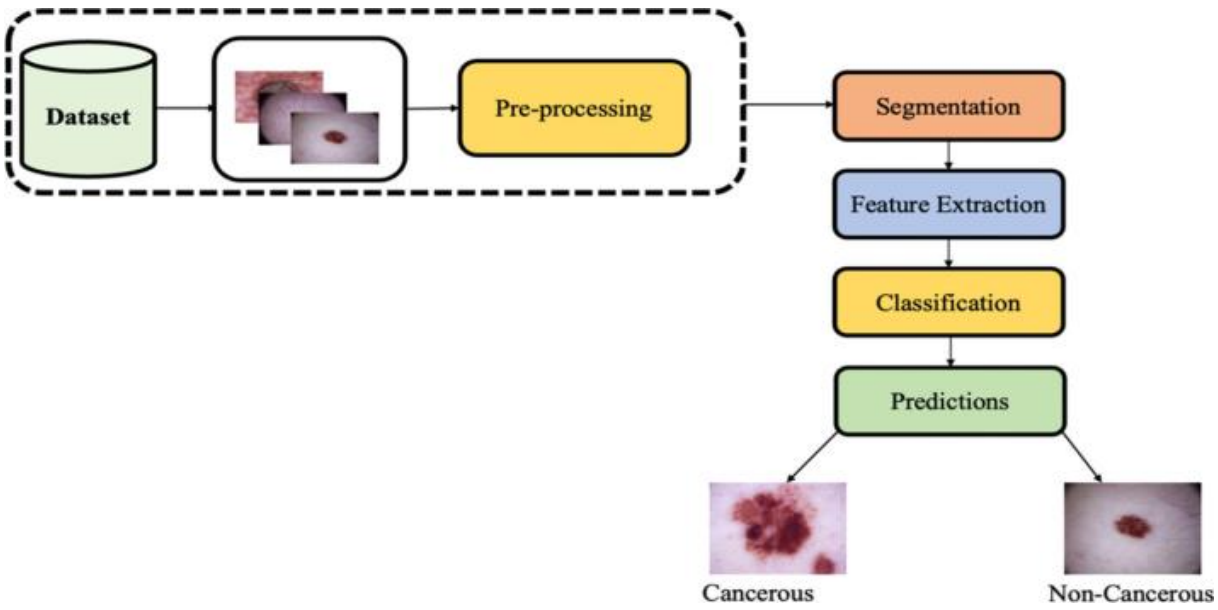


Figure I.2 Étapes générales impliquées dans un système de diagnostic assisté par ordinateur (CAD) [43]

Un système CAD suit généralement une architecture hiérarchique organisée en plusieurs étapes de traitement :

✓ **Acquisition et Prétraitement des Données** : Le processus débute par la collecte d'images provenant d'une base de données (dermoscopiques, radiologiques, etc.). Le prétraitement vise à améliorer la qualité de l'image en supprimant le bruit, en corrigeant l'illumination et en normalisant les contrastes pour faciliter les étapes ultérieures.

✓ **Segmentation** : Cette étape isole la Région d'Intérêt (ROI), correspondant par exemple à une lésion cutanée, un organe ou une tumeur. Elle constitue une phase cruciale permettant de délimiter précisément la zone pathologique à analyser.

✓ **Extraction de Caractéristiques (Feature Extraction)** : Les caractéristiques géométriques, texturales et statistiques sont extraites à partir de la région segmentée. Ces biomarqueurs d'imagerie traduisent la complexité morphologique et radiométrique de la lésion sous forme de descripteurs numériques.

✓ **Classification** : À l'aide de techniques d'apprentissage automatique ou profond, le système attribue chaque ROI à une catégorie diagnostique (par exemple, cancéreuse ou non cancéreuse). Cette étape transforme les caractéristiques extraites en inférences cliniques automatisées.

✓ **Prédiction et Aide à la Décision** : Enfin, le module de prédiction présente au clinicien un résultat assisté sous forme d'évaluation quantitative (score de probabilité, étiquette de

classe). Le système agit ainsi comme un lecteur d'appui intégré au raisonnement clinique. Son assistance intervient à différents niveaux : il peut seconder la phase de détection (CADe) en identifiant automatiquement les régions d'intérêt, et/ou la phase de caractérisation (CADx) en fournissant une estimation du risque pathologique. Cette intégration a pour finalité de consolider le diagnostic en apportant une analyse objective complémentaire, sans se substituer à l'expertise médicale finale, comme l'a démontré l'étude prospective de McKinney et al. en mammographie de dépistage [44]

I.3.3 Assistance Clinique

Les systèmes CAD s'imposent ainsi comme de véritables partenaires du clinicien, en automatisant les tâches répétitives et en accélérant l'interprétation d'examens complexes. Ils contribuent à une réduction du temps d'analyse dans les environnements à forte charge de travail et à une amélioration de la détection précoce des pathologies.

La présence des défis structurels, tels que le volume croissant des données, la complexité des textures et la variabilité inter-patients, ont mis en évidence les limites des approches classiques de segmentation et de classification basées sur des caractéristiques définies manuellement. C'est pourquoi l'évolution naturelle des systèmes d'aide au diagnostic (CAD) s'est orientée vers les méthodes d'apprentissage profond (Deep Learning), capables d'apprendre automatiquement des représentations hiérarchiques et sémantiques directement à partir des données brutes. Ces méthodes offrent ainsi une robustesse et une précision supérieures face à la diversité et l'hétérogénéité des images réelles.

I.4 Dermoscopie et Lésions Cutanées

Les avancées récentes en apprentissage profond appliqué à la segmentation des lésions cutanées reposent largement sur la qualité, la complexité et la diversité des images dermoscopiques. Cette section présente le contexte clinique du mélanome, les fondements techniques de la dermoscopie, ainsi que les principaux défis rencontrés lors de la segmentation automatique de ces images. Elle se conclut par une présentation des bases de données de référence utilisées pour l'évaluation des algorithmes de segmentation.

I.4.1 Enjeux cliniques : le mélanome et l'importance du diagnostic précoce

Le mélanome cutané représente l'un des cancers les plus agressifs de la peau et constitue un enjeu majeur de santé publique à l'échelle mondiale [45]. Bien qu'il ne concerne qu'environ cinq pour cent des cancers cutanés, il est responsable de la majorité des décès d'origine dermatologique en raison de son fort potentiel métastatique. Son incidence a doublé tous les dix à vingt ans dans les populations à peau claire [46], principalement en raison de l'exposition croissante aux rayonnements ultraviolets et de comportements liés au bronzage.

Le diagnostic précoce demeure déterminant pour le pronostic du patient. La survie à cinq ans dépasse quatre-vingt-quinze pour cent pour les mélanomes de faible épaisseur, alors qu'elle chute à moins de vingt pour cent pour les formes métastatiques [47]. L'indice de Breslow, qui mesure la profondeur d'invasion tumorale, reste le principal critère pronostique. Traditionnellement, le diagnostic repose sur l'examen visuel du dermatologue à partir de

critères cliniques standardisés. Parmi ces derniers, la règle ABCDE (Asymétrie, Bords irréguliers, Couleur variable, Diamètre supérieur à six millimètres et Évolution de la lésion) est l'un des outils les plus répandus pour identifier les lésions suspectes [48], [49]. Elle permet une première estimation du risque de malignité à partir de caractéristiques visuelles simples observables sur les images dermoscopiques.

La règle ABCDE est souvent complétée par la liste de contrôle à sept points (Seven-Point Checklist) [50], qui hiérarchise les critères cliniques en distinguant les critères majeurs, tels que les réseaux pigmentaires atypiques, les voiles bleus-blanchâtres ou les zones de régression, et les critères mineurs, comme les points, les globules ou les structures irrégulières. Ces outils ont significativement amélioré la sensibilité du dépistage et demeurent des références dans la pratique dermatologique contemporaine.

Cependant, malgré leur efficacité, ces approches reposent fortement sur l'expérience du praticien et introduisent une part de subjectivité dans l'interprétation. La perception des couleurs, la délimitation des contours ou la reconnaissance des structures dermoscopiques peuvent varier selon le dermatologue, son niveau d'expertise et les conditions d'acquisition des images. Cette variabilité inter-observateur peut réduire la fiabilité du diagnostic, en particulier pour les lésions atypiques, de petite taille ou faiblement pigmentées. La dermoscopie s'est ainsi imposée comme un outil essentiel dans la détection précoce du mélanome. En offrant une visualisation plus fine des structures sous-épidermiques, elle améliore la sensibilité et la spécificité du diagnostic par rapport à l'examen à l'œil nu [51].

I.4.2 La dermoscopie : technique d'imagerie de surface et caractéristiques des images

La dermoscopie ou microscopie de surface, est une technique d'imagerie non invasive qui permet d'observer les structures épidermiques et dermiques superficielles. Deux principaux modes d'acquisition sont employés en dermoscopie, comme illustré à la Figure I.3. Le premier, appelé dermoscopie à immersion, consiste à appliquer un liquide d'interface tel que l'huile, le gel ou l'alcool entre la plaque de contact et la peau. Ce liquide réduit la réflexion lumineuse à la surface cutanée et améliore la transmission de la lumière dans les couches épidermiques et dermiques, permettant ainsi une meilleure visualisation des structures pigmentaires internes de la lésion. Bien que cette méthode offre une excellente qualité d'image, elle nécessite un contact direct avec la peau et des précautions d'hygiène rigoureuses entre chaque observation.

Le second mode, la dermoscopie à polarisation croisée, repose sur l'utilisation de filtres polarisants orientés perpendiculairement sur la source lumineuse et sur le détecteur optique. Cette configuration élimine les reflets de surface sans recourir à un liquide d'immersion, tout en laissant passer la lumière diffusée en profondeur. Elle permet ainsi une acquisition rapide et non invasive, adaptée aux systèmes numériques modernes.

Ces deux approches se distinguent principalement par leur mécanisme de suppression des reflets : l'immersion utilise un milieu liquide pour homogénéiser la surface cutanée**,** tandis que la polarisation croisée filtre optiquement la lumière réfléchie**,** améliorant la clarté et le contraste des images dermoscopiques.

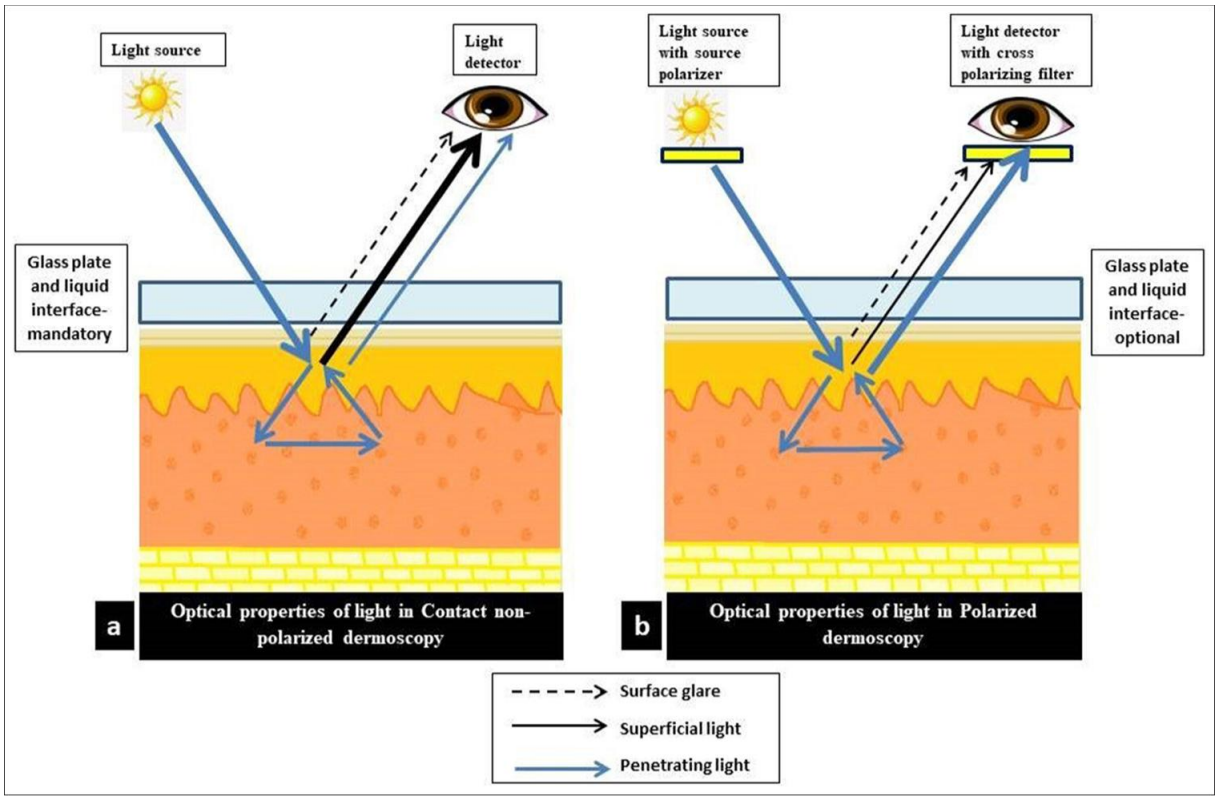


Figure I.3 Propriétés optiques de la lumière en dermoscopie. (a) Dermoscopie à immersion (b) Dermoscopie à polarisation croisée[52]

La lumière traverse un milieu liquide et pénètre dans la peau après réduction des reflets de surface ; les filtres polarisants éliminent les reflets sans contact direct, permettant une observation rapide et non invasive des structures sous-cutanées.

Les images dermoscopiques se distinguent par plusieurs caractéristiques :

- Une haute résolution spatiale, essentielle pour la visualisation des structures pigmentaires fines telles que les réseaux, les globules ou les stries ;
- Une richesse chromatique marquée par des nuances de brun, noir, rouge, bleu-gris et blanc, correspondant à la profondeur de la mélanine et à la vascularisation ;
- Une géométrie particulière, souvent circulaire et sujette à un effet de vignettage, nécessitant des précautions spécifiques lors du prétraitement algorithmique.

I.4.3 Défis spécifiques des images dermoscopiques

Malgré leurs atouts diagnostiques, les images dermoscopiques posent plusieurs défis techniques majeurs pour la segmentation automatique, notamment en raison de leur variabilité visuelle, de la présence d'artéfacts et du déséquilibre des classes au niveau des pixels.

I.4.3.1 Défis intrinsèques : faible contraste et bordures floues

Les lésions cutanées présentent fréquemment un faible contraste par rapport à la peau saine, en particulier dans le cas des mélanomes peu pigmentés ou observés sur des phototypes foncés. Les contours flous ou irréguliers, souvent caractéristiques du mélanome, compliquent

considérablement la détection algorithmique. Par ailleurs, l'hétérogénéité intra-lésionnelle, marquée par des zones de régression ou des variations locales de couleur et de texture, contredit l'hypothèse d'homogénéité régionale sur laquelle reposent de nombreuses approches de segmentation traditionnelles. La variabilité inter-lésionnelle, observée entre naevi, kératoses, carcinomes et mélanomes, renforce la nécessité de modèles robustes capables de généraliser à des cas très divers.

I.4.4 Artéfacts perturbateurs : poils, bulles d'air et reflets

Les artéfacts sont une source importante d'erreurs. Les poils, fréquemment présents, peuvent masquer partiellement la lésion ou être confondus avec des structures pathologiques [53]. Les bulles d'air, visibles sous forme de zones claires circulaires lors de la dermoscopie à immersion, et les reflets spéculaires, produisant des taches saturées, perturbent la lecture chromatique. De plus, les variations d'illumination et les structures naturelles de la peau, comme les ridules, les pores ou les vaisseaux, introduisent une hétérogénéité supplémentaire qu'il est essentiel de corriger lors du prétraitement des images (Figure I.4).

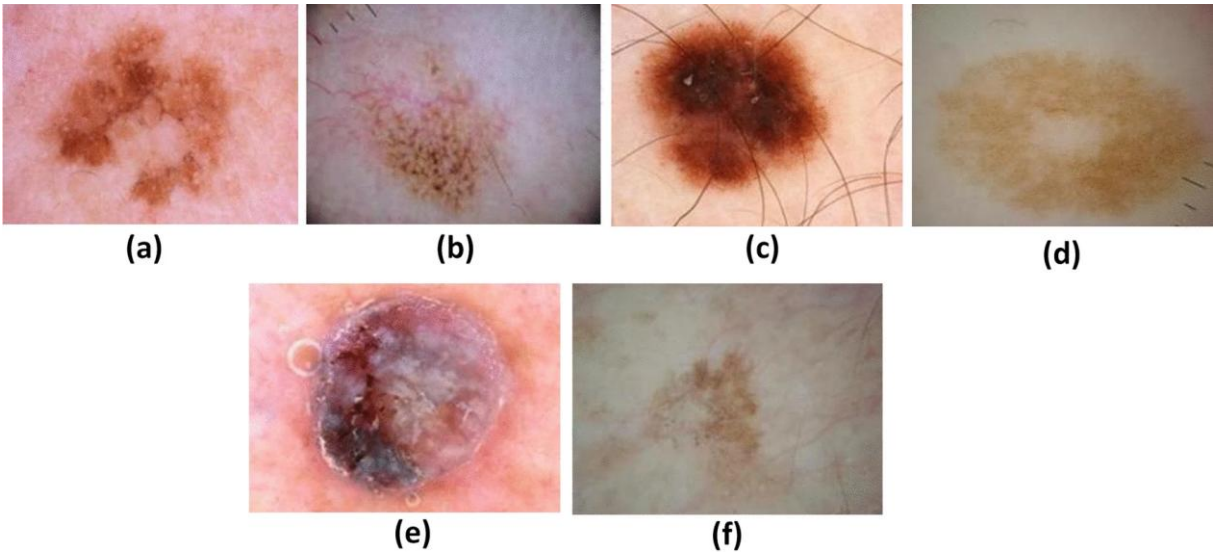


Figure I.4 Cas de lésions cutanées problématiques. a. Contours irréguliers, b. vaisseaux sanguins, c. lignes capillaires, d. illumination colorée, e. bulles, f. faible contraste

I.4.4.1 Déséquilibre des classes au niveau des pixels

Le déséquilibre entre la classe du fond et celle de la lésion constitue un autre problème majeur. Dans la plupart des images dermoscopiques, la lésion occupe moins de vingt pour cent de la surface totale. Ce rapport, souvent inférieur à un pour dix, conduit les modèles d'apprentissage à privilégier la classe dominante, correspondant à la peau saine. Ce biais se traduit par une sous-segmentation des lésions, notamment pour les petites régions ou celles à faible contraste, et par des métriques d'évaluation trompeuses, telles que l'accuracy, qui surestime artificiellement les performances.

I.5 Méthodes Traditionnelles de Segmentation d'Image

Les méthodes traditionnelles de segmentation peuvent être globalement classées en plusieurs catégories majeures, chacune reposant sur des principes mathématiques distincts pour isoler les Régions d'Intérêt.

I.5.1 Segmentation Basée sur les Techniques de Seuillage (Thresholding)

Le seuillage est l'une des techniques de segmentation les plus fondamentales et rapides, reposant sur l'hypothèse que la Région d'Intérêt (ROI) peut être séparée du fond (*background*) par une valeur de coupure unique dans l'espace des intensités ou des couleurs. Ces méthodes sont non contextuelles, car l'étiquette d'un pixel est déterminée indépendamment des pixels voisins.

I.5.1.1 Seuillage Binaire

Soit I l'image source, et soit $f(p)$ la valeur d'une caractéristique (intensité, composante de couleur, etc.) au pixel p. La segmentation par seuillage binaire (séparation en deux classes : $K = 2$) est une opération $\mathcal{S}$ qui attribue une étiquette $L(p) \in \{0,1\}$ à p en fonction d'un seuil prédéfini τ :

$$L(p) = \mathcal{S}(f(p)) = \begin{cases} 1 & \text{si } f(p) \geq \tau \quad \text{(Classe de la ROI)} \\ 0 & \text{si } f(p) < \tau \quad \text{(Classe du fond)} \end{cases} \tag{I.1}$$

Le défi principal du seuillage réside dans le choix d'un seuil τ qui minimise le risque de fausse classification. Deux approches principales existent :

1) **Seuillage Global** : Un seul seuil τ est appliqué à l'ensemble de l'image. Le seuil est souvent déterminé à partir de l'analyse de l'histogramme de l'image. La Méthode d'Otsu (1979) [54] technique est très utilisée pour calculer le seuil τ^* qui maximise la variance inter-classes (σ^2_{inter}) de l'histogramme de l'image C. Si l'on note $\omega_0(\tau)$ et $\omega_1(\tau)$ les probabilités des classes 0 et 1, et $\mu_0(\tau)$ et $\mu_1(\tau)$ leurs moyennes respectives, τ^* est trouvé par :

$$\tau^* = \underset{\tau}{\text{argmax}}\{\omega_0(\tau)\omega_1(\tau)[\mu_0(\tau) - \mu_1(\tau)]^2\} \tag{I.2}$$

Cette méthode est optimale si et seulement si l'histogramme est bimodal et que la variance intra-classe est faible.

2) **Seuillage Adaptatif (Local) :** Un seuil $\tau(p)$ est calculé pour chaque pixel p en se basant uniquement sur les statistiques (moyenne, médiane, etc.) des pixels dans un voisinage local $\mathcal{N}_p$.

$$L(p) = \begin{cases} 1 & \text{si } f(p) \geq \tau(p) \\ 0 & \text{si } f(p) < \tau(p) \end{cases} \quad \text{où } \tau(p) = \text{Fonction}(\{f(q) \mid q \in \mathcal{N}_p\}) \tag{I.3}$$

Cette approche est indispensable lorsque l'image présente des variations d'éclairement importantes, car elle permet de s'adapter aux changements locaux de contraste.

I.5.1.2 Limitations et Rôle Actuel

Les méthodes de seuillage, bien que rapides, sont intrinsèquement limitées face aux images réelles (cf. Section 1.2) :

- ✓ **Faible Robustesse aux Contours Flous :** Elles ne gèrent pas bien les gradients progressifs d'intensité aux frontières (ambiguïté de localisation).
- ✓ **Sensibilité aux Artéfacts :** Le bruit local ou les artéfacts (comme les poils) peuvent fausser la détermination du seuil optimal ou introduire des discontinuités dans le masque binaire.
- ✓ **Dépendance à la Bimodalité :** Le succès du seuillage global dépend fortement d'une bonne séparation statistique des deux classes.

En conséquence, le seuillage est aujourd'hui principalement utilisé comme une étape de pré-traitement pour initialiser des modèles plus complexes (tels que les modèles actifs ou les MRF) ou comme une étape de filtrage rapide pour générer un masque binaire grossier qui sera raffiné par la suite.

I.5.2 Segmentation Basée sur la Détection de Contours Classiques

Les méthodes basées sur la détection de contours (ou bords) segmentent une image en identifiant les points où l'intensité (ou la couleur) de l'image subit un changement brusque et significatif. Ces discontinuités correspondent aux frontières d'objets ou de régions. Ces méthodes reposent sur l'utilisation d'opérateurs différentiels (dérivées spatiales) pour mettre en évidence les gradients.

I.5.2.1 Principe du Gradient

Le principe fondamental est de calculer le vecteur gradient $\nabla I(x, y)$ de l'image I au pixel (x, y), où I est considérée comme une fonction bidimensionnelle de l'intensité.

Le gradient est défini par :

$$\nabla I(x, y) = \begin{pmatrix} G_x \\ G_y \end{pmatrix} = \begin{pmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{pmatrix} \tag{I.4}$$

La magnitude du gradient $M(x, y)$ indique la force du changement d'intensité et est utilisée pour détecter les bords :

$$M(x, y) = \| \nabla I(x, y) \| = \sqrt{G_x^2 + G_y^2} \tag{I.5}$$

Les opérateurs classiques (Sobel, Prewitt, Roberts) sont des filtres de convolution discrète qui approximent ces dérivées partielles.

Opérateur de Sobel (Exemple de G_x) : Un filtre 3×3 est appliqué pour estimer G_x (bord vertical) :

$$G_x \approx \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * I \quad \text{(I.6)}$$

Un bord est finalement détecté lorsque $M(x, y)$ dépasse un seuil prédéfini τ.

I.5.2.2 Opérateur de Canny : Approche Optimale

L'opérateur de Canny, développé par John Canny (1986) [55], est la méthode de détection de bords la plus influente et la plus utilisée en vision par ordinateur. Il est basé sur trois critères d'optimalité :

- ✓ **Bonne Détection** : Faible taux de faux positifs (bruit) et de faux négatifs (bords réels manqués).
- ✓ **Bonne Localisation** : La position des bords détectés doit être la plus proche possible des bords réels.
- ✓ **Réponse Unique** : Un seul point par bord réel (éviter la détection multiple).

L'algorithme Canny procède en plusieurs étapes formalisées :

1) **Lissage (Réduction du Bruit)** : Application d'un filtre Gaussien G_σ à l'image I : $I_{\text{lissée}} = G_\sigma * I$.
2) **Calcul du Gradient** : Calcul de la magnitude M et de l'orientation Θ du gradient sur $I_{\text{lissée}}$.
3) **Suppression des Non-Maxima (Non-Maximum Suppression - NMS)** : Ce processus affine les bords en s'assurant que seuls les maximas locaux du gradient (dans la direction perpendiculaire au bord) sont conservés, satisfaisant le critère de réponse unique.
4) **Seuillage par Hystérésis (Hysteresis Thresholding)** : Application de deux seuils, τ_{haut} (pour la détection des bords forts) et τ_{bas} (pour la connexion des bords faibles), ce qui assure la continuité du contour (critère de bonne détection).

I.5.2.3 Opérateurs Basés sur le Laplacien (Dérivée Seconde)

Les opérateurs basés sur le Laplacien (comme le Laplacien de Gaussienne - LoG) utilisent la dérivée seconde pour détecter les bords, qui correspondent aux passages par zéro (*zero-crossings*) de l'opérateur. Le Laplacien d'une image I est défini comme la trace de la matrice Hessienne :

$$\nabla^2 I(x, y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad \text{(I.7)}$$

Le LoG combine lissage et détection des bords en un seul filtre :

$$\text{LoG}(x, y) = \nabla^2 (G_\sigma * I) \quad \text{(I.8)}$$

Les points où le LoG change de signe (passage par zéro) indiquent l'emplacement précis d'un bord. L'avantage est la précision de la localisation, mais l'inconvénient est une sensibilité extrême au bruit (bien que partiellement corrigée par le filtre Gaussien initial).

Bien que fondamentales, ces méthodes sont fortement compromises en imagerie médicale :

✓ **Faible Contraste** : Les contours faibles des lésions cutanées (flou intrinsèque) ne produisent pas de gradients suffisamment forts pour dépasser le seuil τ.

✓ **Bruit/Artefacts** : Le bruit et les artefacts génèrent de faux gradients locaux (faux positifs), qui peuvent être confondus avec des bords réels.

✓ **Absence de Fermeture** : Ces méthodes ne garantissent pas que les bords détectés formeront un contour fermé et continu (nécessaire pour la segmentation de la ROI), nécessitant une post-segmentation supplémentaire.

I.5.3 Segmentation Basée sur la Région

Les méthodes de segmentation basées sur la région exploitent l'hypothèse de l'homogénéité locale pour regrouper les pixels adjacents partageant des caractéristiques similaires (intensité, couleur, texture). Contrairement au seuillage, ces approches sont contextuelles, car elles tiennent compte de la relation spatiale entre les pixels.

La segmentation basée sur la région vise à produire un partitionnement $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_K$ de l'image I tel que deux conditions principales soient satisfaites :

✓ **Homogénéité Interne (Prédicat $\mathcal{P}$)** : Chaque région $\mathcal{R}_k$ doit être homogène selon un prédicat $\mathcal{P}$ donné (ex. : la variance de l'intensité est inférieure à un seuil ϵ).

$$\forall k \in \{1, \dots, K\}, \quad \mathcal{P}(\mathcal{R}_k) = \text{Vrai} \tag{I.9}$$

✓ **Hétérogénéité Externe (Arrêt)** : L'union de deux régions adjacentes $\mathcal{R}_i$ et $\mathcal{R}_j$ doit être hétérogène.

$$\forall i \neq j, \quad \mathcal{P}(\mathcal{R}_i \cup \mathcal{R}_j) = \text{Faux} \tag{I.10}$$

I.5.3.1 Croissance de Région (Region Growing)

La croissance de région est une approche ascendante (*bottom-up*) qui construit les régions en partant de points initiaux (*seeds*) selon le processus :

- Sélection d'un pixel initial (graine) $p_0 \in I$.
- Initialisation de la région $\mathcal{R}$ avec p_0.
- Tant que la condition d'homogénéité est satisfaite, les pixels voisins de $\mathcal{R}$ qui respectent le critère de similarité sont ajoutés à $\mathcal{R}$.

Pour un pixel candidat p voisin de la région $\mathcal{R}$, le critère de similarité repose sur la comparaison de la valeur de sa caractéristique $f(p)$ avec la moyenne $\mu_{\mathcal{R}}$ de la région existante :

$$p \in \mathcal{R} \cup \{p\} \quad \text{si} \quad |f(p) - \mu_{\mathcal{R}}| \leq \delta \tag{I.11}$$

où δ est un seuil de tolérance.

La performance dépend fortement du choix initial des graines et du seuil δ.

I.5.3.2 Division et Fusion (Split and Merge)

Cette technique combine les approches descendante (*top-down*) et ascendante, permettant de corriger les erreurs de division initiale et d'obtenir des régions plus uniformes et cohérentes.

- ✓ **Division (Split) :**

L'image I est récursivement divisée (souvent via un quadtree) en sous-régions $\mathcal{R}_{sub}$ jusqu'à ce que chaque sous-région satisfasse le prédicat d'homogénéité $\mathcal{P}$.

$$\text{Si } \mathcal{P}(\mathcal{R}) = \text{Faux}, \quad \text{Diviser } \mathcal{R} \text{ en } \mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4. \tag{I.12}$$

- ✓ **Fusion (Merge) :**

Les régions adjacentes $\mathcal{R}_i$ et $\mathcal{R}_j$ sont fusionnées si leur union satisfait la condition d'homogénéité (ou si leur différence de moyenne est inférieure à un seuil).

$$\text{Fusionner } \mathcal{R}_i \text{ et } \mathcal{R}_j \quad \text{si} \quad \mathcal{P}(\mathcal{R}_i \cup \mathcal{R}_j) = \text{Vrai}. \tag{I.13}$$

Les méthodes basées sur la région ont historiquement permis de produire des frontières plus fermées et des segments plus cohérents que le seuillage. Cependant, elles sont très sensibles aux problématiques des images réelles :

- ✓ Elles échouent lorsque l'hétérogénéité intra-lésionnelle (pigmentation variable) est forte, conduisant à une sur-segmentation (une lésion est décomposée en plusieurs sous-régions).
- ✓ Elles peuvent souffrir de fuites de frontière (*boundary leakage*) si le contraste local entre la lésion et la peau est très faible ou interrompu par des artefacts.

I.5.4 Segmentation Basée sur le Partitionnement (Clustering)

Le partitionnement (clustering) est une technique de segmentation non supervisée qui regroupe les pixels d'une image en classes (clusters) en fonction de leur similarité dans un espace de caractéristiques (couleur, intensité, texture, coordonnées spatiales). Contrairement au seuillage qui opère sur une seule dimension, le clustering opère généralement dans un espace multidimensionnel. Le processus de segmentation consiste à attribuer une étiquette de cluster à chaque pixel, convertissant ainsi un regroupement de caractéristiques en une partition spatiale de l'image.

I.5.4.1 Méthode K-means

Le modèle de clustering par partitionnement le plus répandu est le K-means. Il vise à diviser l'ensemble des vecteurs de caractéristiques des pixels $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N\}$ en K partitions

$\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K$ en minimisant l'inertie intra-cluster, c'est-à-dire la Somme des Carrés des Erreurs (SSE).

La fonction d'objectif J à minimiser est donnée par :

$$J(\mathcal{C}, \mathbf{M}) = \sum_{k=1}^{K} \sum_{\mathbf{f}_i \in \mathcal{C}_k} \| \mathbf{f}_i - \mathbf{m}_k \|^2 \quad \text{(I.14)}$$

où :

N est le nombre total de pixels

K est le nombre prédéfini de clusters (segments)

$\mathbf{m}_k$ est le centroïde (moyenne) du cluster $\mathcal{C}_k$

$\|\cdot\|^2$ est la distance euclidienne au carré (mesure de similarité)

L'algorithme converge itérativement vers une solution locale en recalculant les centroïdes et en réattribuant les pixels à leur centroïde le plus proche. Le résultat de la segmentation est une carte d'étiquettes $L(p_i)$ où $L(p_i) = k$ si le vecteur de caractéristiques du pixel p_i appartient au cluster $\mathcal{C}_k$.

I.5.4.2 Clustering Flou (Fuzzy C-means - FCM)

Pour gérer l'ambiguïté inhérente aux images médicales (faible contraste et frontières progressives), le C-means flou (FCM), introduit par Dunn (1973) [56], est souvent préféré. Au lieu d'une partition dure (appartenance totale à un seul cluster), le FCM assigne un degré d'appartenance $u_{ik} \in [0,1]$ à chaque pixel p_i pour chaque cluster $\mathcal{C}_k$.

Le FCM minimise la fonction d'objectif pondérée :

$$J_{FCM} = \sum_{k=1}^{K} \sum_{i=1}^{N} u_{ik}^{m} \| \mathbf{f}_i - \mathbf{m}_k \|^2 \quad \text{(I.15)}$$

avec la contrainte $\sum_{k=1}^{K} u_{ik} = 1$ pour chaque pixel i.

u_{ik} est le degré d'appartenance du pixel i au cluster k

$m \geq 1$ est le paramètre de flou (*fuzziness parameter*), qui contrôle la douceur de la transition entre les clusters (généralement $m = 2$)

Le FCM fournit une carte de probabilités d'appartenance, et l'étiquette finale est typiquement déterminée par l'appartenance maximale.

Le principal défi des méthodes de *clustering* classiques est qu'elles sont non spatiales : le regroupement est basé uniquement sur la similarité des caractéristiques, sans tenir compte de la position des pixels. Cette faiblesse fondamentale donne lieu à plusieurs problèmes majeurs :

✓ **Sensibilité au Bruit et à l'Hétérogénéité** : Le bruit et les artefacts peuvent créer des clusters non pertinents

✓ **Fragmentation Spatiale** : Des pixels ayant des caractéristiques similaires peuvent être éloignés spatialement, conduisant à une segmentation fragmentée et à un manque de continuité du contour (le clustering ne garantit pas la connexité). Ce défaut a motivé l'introduction du clustering spatialement régularisé (où les coordonnées (x, y) sont ajoutées au vecteur de caractéristiques $\mathbf{f}_i$), ou l'intégration du clustering dans des schémas hybrides de post-traitement pour imposer une cohérence spatiale.

I.5.5 Segmentation Basée sur les Contours Actifs (Modèles Déformables)

Les méthodes basées sur les Contours Actifs (Active Contour Models ou Snakes) représentent une avancée majeure par rapport aux méthodes régionales et de gradient en formulant la segmentation comme un problème de minimisation d'énergie. Elles utilisent une courbe initiale, ou une surface, qui se déforme sous l'influence de forces internes (régularisation) et externes (attraction vers le bord de l'objet) jusqu'à ce qu'elle converge vers la frontière de la Région d'Intérêt (ROI).

I.5.5.1 Modèle Paramétrique (Snakes)

Le modèle de contour actif, introduit par Kass, Witkin et Terzopoulos (1988) [57], est une courbe paramétrique $v(s) = (x(s), y(s))$, avec $s \in [0,1]$, qui évolue dans le domaine de l'image. Le contour optimal v^* est celui qui minimise une fonctionnelle d'énergie totale E_{total} :

$$E_{\text{total}}(v) = \int_0^1 [\, E_{\text{interne}}(v(s)) + E_{\text{externe}}(v(s))] \, ds \qquad \text{(I.16)}$$

La solution est obtenue en trouvant la courbe qui annule les forces dérivées de cette énergie.

✓ **Énergie Interne (E_{interne})**

L'énergie interne contrôle les propriétés géométriques de la courbe (régularisation). Elle garantit la cohérence topologique et le lissage du contour, rendant la méthode robuste au bruit:

$$E_{\text{interne}}(v) = \frac{1}{2}(\alpha(s) \parallel v'(s) \parallel^2 + \beta(s) \parallel v''(s) \parallel^2) \qquad \text{(I.17)}$$

où :

$\parallel v'(s) \parallel^2$ (Terme d'élasticité) : Pénalise l'étirement du contour, contrôlé par le poids $\alpha(s)$

$\parallel v''(s) \parallel^2$ (Terme de rigidité/courbure) : Pénalise les changements brusques de direction, contrôlé par $\beta(s)$

✓ **Énergie Externe (E_{externe})**

L'énergie externe lie le contour aux données de l'image. Elle est généralement basée sur les gradients pour attirer le contour vers les bords.

$$E_{\text{externe}}(v) = -\| \nabla I(v(s)) \|^2 \quad \text{(I.18)}$$

Minimiser E_{externe} revient à attirer le contour vers les zones où la magnitude du gradient de l'image I est maximale, soit les frontières de l'objet.

I.5.5.2 Modèles Géométriques (Level Set)

Pour surmonter les limitations des Snakes (sensibilité à l'initialisation et incapacité à gérer les changements de topologie comme la fusion ou la division du contour), les Modèles Géométriques (Level Set), introduits par Osher et Sethian (1988) [58], ont été développés.

Le contour $C(t)$ est représenté implicitement comme l'ensemble de niveau zéro d'une fonction de distance signée $\phi(x, y, t)$:

$$C(t) = \{(x, y) \mid \phi(x, y, t) = 0\} \quad \text{(I.19)}$$

L'évolution du contour est déterminée par une équation aux dérivées partielles (EDP), où le front $\phi = 0$ se propage selon une vitesse F :

$$\frac{\partial \phi}{\partial t} + F \| \nabla \phi \| = 0 \quad \text{(I.20)}$$

où la vitesse F peut être composée de termes d'advection (attraction par le gradient) et de courbure (régularisation).

I.5.5.3 Modèle Chan-Vese (Level Set Basé Région)

Le modèle Chan-Vese (2001) [59] est une variante des Level Sets particulièrement adaptée à l'imagerie médicale, car il est basé sur la région et fonctionne même sur des images à faible gradient (contours flous). Il suppose que l'image est composée de deux régions homogènes distinctes (c_1 et c_2) et minimise l'énergie d'ajustement de l'image :

$$\begin{aligned} E_{\text{Chan-Vese}}(\phi, c_1, c_2) = \mu \cdot \text{Longueur}(\phi = 0) \\ + \lambda_1 \int_{\Omega} |I - c_1|^2 H(\phi)\, dxdy \\ + \lambda_2 \int_{\Omega} |I - c_2|^2 (1 - H(\phi))\, dxdy \end{aligned} \quad \text{(I.21)}$$

où :

μ est le paramètre de régularisation (lissage du contour)

c_1 et c_2 sont les moyennes d'intensité à l'intérieur et à l'extérieur du contour, respectivement

$H(\phi)$ est la fonction de Heaviside (1 à l'intérieur, 0 à l'extérieur)

Ce modèle est très puissant pour la segmentation de lésions cutanées où l'information de bord (gradient) est faible, car il est guidé par l'homogénéité interne des régions.

I.5.6 Segmentation Basée sur le Champ Aléatoire de Markov (MRF)

Les modèles de Champ Aléatoire de Markov, ou Markov Random Fields (MRF), sont des outils puissants d'analyse d'image qui formalisent la segmentation en tant que problème d'optimisation énergétique sous contraintes de voisinage. Ces méthodes constituent une approche traditionnelle et probabiliste permettant d'incorporer des informations spatiales dans le processus de classification par pixel [60]

I.5.6.1 Formalisme du Modèle

Soit $\mathcal{I}$ l'image d'entrée, composée d'un ensemble de pixels (sites) $S = \{s_1, s_2, \dots, s_N\}$, où N est le nombre total de pixels. La segmentation consiste à attribuer à chaque pixel $s_i \in S$ une étiquette de classe $l_i \in \mathcal{L} = \{L_1, L_2, \dots, L_K\}$, où K est le nombre de segments (classes).

Le MRF modélise la segmentation en introduisant un champ d'étiquettes $L = \{l_1, l_2, \dots, l_N\}$ qui est un champ aléatoire défini sur un graphe d'adjacence $G = (S, E)$, où E représente l'ensemble des paires de pixels voisins (généralement une structure de voisinage de 4 ou 8 connexions).

Selon l'Équivalence de Hammersley-Clifford, la probabilité d'une configuration d'étiquettes L (la carte de segmentation) suit une distribution de Gibbs :

$$P(L|\mathcal{I}) = \frac{1}{Z}\exp(-E(L|\mathcal{I})) \tag{I.22}$$

où :

$E(L|\mathcal{I})$ est la fonction d'énergie (ou Hamiltonien) de la configuration d'étiquettes L,

Z est la fonction de partition (une constante de normalisation).

Le but de la segmentation est de trouver la configuration d'étiquettes L^* qui minimise l'énergie $E(L|\mathcal{I})$ (ou, de manière équivalente, qui maximise la probabilité a posteriori $P(L|\mathcal{I})$), conformément au critère du Maximum a Posteriori (MAP).

I.5.6.2 Fonction d'Énergie

L'énergie $E(L|\mathcal{I})$ est structurée pour capturer deux propriétés fondamentales : l'adéquation aux données et la cohérence spatiale. Elle est généralement décomposée en deux termes principaux :

$$E(L|\mathcal{I}) = E_{\text{Unitaire}}(L|\mathcal{I}) + E_{\text{Voisinage}}(L) \tag{I.23}$$

- ✓ **Terme Unitaire (ou de Données)** : $E_{\text{Unitaire}}(L|\mathcal{I})$

Ce terme pénalise les étiquettes l_i qui sont statistiquement peu probables compte tenu des observations (caractéristiques de couleur, texture, etc.) $\mathbf{y}_i$ au pixel s_i :

$$E_{\text{Unitaire}}(L|\mathcal{I}) = \sum_{s_i \in S} \phi_i\,(l_i) = \sum_{s_i \in S} -\log P(\mathbf{y}_i|l_i) \tag{I.24}$$

Ce terme est souvent dérivé d'un classifieur local (comme un modèle de Mélange de Gaussiennes) qui estime la probabilité d'observer les données $\mathbf{y}_i$ si le pixel appartient à la classe l_i.

- ✓ **Terme de Voisinage (ou d'Interaction)** : $E_{\text{Voisinage}}(L)$

Ce terme encourage la cohérence spatiale en pénalisant les paires de pixels adjacents $(s_i, s_j) \in E$ qui ont des étiquettes différentes. Il modélise le poids du voisinage (ou priorité de Markov) :

$$E_{\text{Voisinage}}(L) = \sum_{(s_i,s_j)\in E} \psi_{ij}\,(l_i, l_j) \tag{I.25}$$

Une formulation courante utilise le Potentiel de Potts :

$$\psi_{ij}(l_i, l_j) = \begin{cases} 0 & \text{si } l_i = l_j \\ \beta & \text{si } l_i \neq l_j \end{cases} \tag{I.26}$$

où $\beta > 0$ est un paramètre d'interaction qui contrôle la force du lissage spatial. Plus β est grand, plus les grandes régions homogènes sont favorisées.

I.5.6.3 Algorithmes d'Optimisation

La minimisation de la fonction d'énergie $E(L|\mathcal{I})$ est un problème NP-difficile pour un CAM général. Cependant, lorsque la fonction d'énergie satisfait certaines conditions de sous-modularité (souvent le cas avec des potentiels simples comme celui de Potts), des algorithmes d'optimisation efficaces peuvent être utilisés.

I.6 État de l'Art : Méthodes Traditionnelles de Segmentation pour l'Imagerie Médicale et les Lésions Cutanées

Les méthodes traditionnelles de segmentation constituent le fondement historique de l'analyse d'images médicales et dermatologiques. Ces approches, incluant le seuillage, la détection de contours, les méthodes basées sur la région, le partitionnement (clustering), les contours actifs et les Champs Aléatoires de Markov (MRF), ont évolué pour répondre aux défis spécifiques posés par les images de lésions cutanées, tels que le faible contraste, l'hétérogénéité pigmentaire et la présence d'artefacts.

I.6.1 Seuillage pour la Segmentation d'Images Médicales et de Lésions Cutanées

Le seuillage (Thresholding) est l'approche la plus ancienne et la plus rapide pour la segmentation d'images médicales, initialement utilisée pour les modalités à fort contraste. Son

évolution en dermoscopie illustre la transition des méthodes purement statistiques vers des systèmes hybrides et adaptés à la couleur.

I.6.1.1 Travaux Fondamentaux et Approches Multi-seuils

Les travaux fondamentaux reposent sur l'idée que la lésion et la peau peuvent être distinguées par une ou plusieurs coupures dans l'histogramme des couleurs ou des intensités. La performance des méthodes de seuillage global, comme celle d'Otsu, a été rapidement complétée par l'exploration d'espaces couleur optimaux. Pour gérer l'hétérogénéité des pigments, la recherche s'est orientée vers le seuillage multi-niveaux. Humayun et al. (2011) [61] ont appliqué le seuillage multi-niveau pour capturer les différentes composantes de couleur des lésions pigmentées, bien que cette méthode reste souvent limitée par sa dépendance au choix initial du nombre de seuils et à la complexité du calcul des multiples optima. Les auteurs dans [62] ont proposé une technique de seuillage multiple en exploitant les différentes informations chromatiques pour une segmentation plus robuste, mais l'efficacité de leur méthode peut être compromise par le bruit dans les différents canaux chromatiques utilisés.

I.6.1.2 Seuillage Adaptatif

Face aux variations d'éclairement et de contraste non uniformes, l'utilisation de méthodes de seuillage adaptatif est devenue essentielle. Gupta et al. [63] ont développé des méthodes adaptatives en utilisant des paramètres statistiques locaux pour déterminer le seuil $\tau(p)$ à chaque pixel p, améliorant significativement la robustesse du contour, toutefois, le coût computationnel de l'estimation des paramètres locaux pour chaque pixel peut ralentir considérablement le processus. Ces travaux s'inscrivent dans une tendance plus large des méthodes de seuillage avancées pour l'imagerie médicale [64].

I.6.1.3 Méthodes Pondérées et Optimisées

Des approches plus sophistiquées incorporent des pondérations pour améliorer la décision du seuil. [65] ont introduit une méthode de seuillage pondéré simple qui intègre des informations contextuelles légères pour mieux isoler la lésion des fonds complexes, néanmoins, la simplicité de la pondération peut ne pas suffire à gérer des artefacts complexes tels que la présence de poils ou de bulles d'air. Des études spécifiques, comme celles de [66] sur les carcinomes basocellulaires, continuent de valider et de comparer différentes techniques de seuillage, insistant sur le fait que l'optimisation pour un type spécifique de lésion peut rendre la méthode moins généralisable à d'autres morphologies de lésions cutanées.

En somme, le seuillage est passé d'une méthode de segmentation autonome à un composant essentiel et optimisé pour l'initialisation et le raffinement dans des systèmes d'analyse complexes en dermoscopie.

I.6.2 Détection de Contours Classiques pour la Segmentation de Lésions Cutanées

Les méthodes de détection de contours classiques, reposant sur l'analyse des discontinuités d'intensité via des opérateurs différentiels (Gradient, Laplacien), sont essentielles pour identifier les frontières potentielles des lésions.

I.6.2.1 Travaux Fondamentaux et Limites en Dermoscopie

L'application directe de détecteurs de contours à la dermoscopie est limitée par le faible contraste aux frontières des mélanomes et la présence d'artefacts (poils, bulles) qui génèrent des faux gradients. Comme souligné dans la revue de [67] sur le diagnostic des lésions cutanées, les méthodes de segmentation basées uniquement sur le gradient souffrent souvent de discontinuités dans les contours détectés, ne parvenant pas à former une frontière fermée, ce qui est crucial pour la quantification de la Région d'Intérêt (ROI).

I.6.2.2 Amélioration Itérative du Canny

Face à la nécessité de garantir la fermeture du contour, des travaux ont cherché à rendre l'opérateur de Canny plus robuste et itératif. Le travail dans [68] a développé un algorithme de segmentation itératif amélioré utilisant le détecteur de bords de Canny spécifiquement pour la détection du contour des lésions cutanées. Cette itération vise à affiner la carte de bords et à connecter les segments faibles, transformant ainsi l'opérateur classique en un outil de délimitation semi-globale capable de mieux cerner l'irrégularité des bords cliniques.

En conclusion, si les opérateurs de gradient et de Laplacien ont échoué en tant que méthodes de segmentation autonomes pour les lésions cutanées, ils demeurent une composante analytique vitale pour localiser le bord, intégrant des systèmes hybrides et itératifs pour surmonter les défis du faible contraste.

I.6.3 Segmentation Basée sur la Région pour l'Imagerie Médicale

Les méthodes de segmentation basées sur la région exploitent le principe de l'homogénéité spatiale et de la connectivité pour regrouper les pixels en segments cohérents. Cette approche s'est montrée plus robuste que le seuillage simple en garantissant des frontières fermées.

I.6.3.1 Travaux Fondamentaux et Nouveaux Défis

Les travaux fondamentaux reposent sur la Croissance de Région (Region Growing) pour produire des segments lisses et anatomiquement plausibles, comme illustré par les travaux pionniers d'Adams et Bischof [69]. Cependant, face à l'hétérogénéité pigmentaire et aux artefacts des images dermoscopiques, l'efficacité des méthodes régionales dépend crucialement de la façon dont le prédicat d'homogénéité est défini et initialisé. La dépendance à une couleur uniforme étant un point faible, des travaux ont continué à explorer l'optimisation des caractéristiques d'entrée. Même si la segmentation régionale se distingue du seuillage, la qualité de l'initialisation est primordiale : des études comme celles de Sengupta et al. [70] continuent de souligner que le choix et la transformation de l'espace couleur peuvent améliorer significativement la distinction régionale initiale dans les images de lésions, rendant les étapes de croissance ou de fusion plus efficaces.

I.6.3.2 Croissance de Région Adaptative et Floue

Les travaux récents exploitent la croissance de région adaptative pour mieux gérer le bruit et l'ambiguïté des contours. Guo et al. [71] ont introduit une approche novatrice utilisant le

regroupement neutrosophique (*neutrosophic clustering*) pour l'initialisation des graines, suivie d'une croissance de région adaptative. Cette hybridation est significative : elle utilise la logique floue et les méthodes de regroupement non traditionnelles pour gérer l'incertitude dans les régions de transition, avant d'appliquer le principe de croissance de région. Cela permet à l'algorithme de s'adapter localement aux propriétés statistiques complexes des lésions, améliorant la robustesse face aux variations d'intensité et de texture.

I.6.3.3 Intégration dans les Systèmes d'Apprentissage

Plus largement, l'approche régionale est désormais souvent intégrée aux systèmes modernes pour fournir une forme de régularisation spatiale ou une entrée prétraitée. La revue de Khan et al. [72] met en lumière la tendance à combiner les méthodes traditionnelles (dont les techniques régionales) avec le Deep Learning. Dans ce contexte hybride, les méthodes régionales ne sont plus une fin en soi, mais servent à produire un masque initial cohérent ou à fournir des caractéristiques de forme lissées qui enrichissent les systèmes de classification ou de segmentation basés sur l'apprentissage.

L'apport principal des méthodes régionales réside dans leur capacité à produire des segments fermés, topologiquement cohérents et connectés, un atout essentiel qui reste pertinent pour l'analyse morphologique quantitative et la validation des contours cliniques.

I.6.4 Segmentation par Partitionnement (Clustering) pour les Lésions Cutanées

Le partitionnement (clustering) est une méthode de segmentation non supervisée particulièrement appréciée en dermoscopie car elle permet de regrouper les pixels en fonction de la similarité de couleur et d'intensité sans nécessiter de vérité terrain étiquetée, gérant ainsi l'hétérogénéité pigmentaire des lésions.

I.6.4.1 Travaux Fondamentaux : Choix de la Méthode et de l'Espace Couleur

Les premières tentatives réussies ont nécessité de transformer l'image en un espace de caractéristiques plus pertinent que le simple RGB pour mieux distinguer la lésion du fond.

I.6.4.2 Partitionnement par Couleur

La performance des algorithmes de clustering dépend fortement de la sélection de l'espace couleur. Les auteurs dans [73] ont établi l'importance comparative des différents algorithmes et espaces couleur, montrant que les espaces décorrelés (comme L*a*b* ou HSV) optimisent la séparation des clusters Lésion/Peau.

I.6.4.3 K-means en Dermoscopie

Le K-means est souvent utilisé pour sa simplicité et sa rapidité. Cependant, il est vulnérable au bruit et aux régions non pertinentes. L'approche dans [74] a abordé ce problème en utilisant le K-means pour la segmentation des lésions tout en intégrant une étape spécifique de suppression des régions non désirées (Removal of Unwanted Regions), illustrant la nécessité d'étapes de prétraitement ou de post-traitement pour stabiliser le résultat du clustering classique.

I.6.4.4 Robustesse Floue (Fuzzy C-means - FCM)

Le FCM est privilégié dans les cas de lésions mal définies. Il assigne un degré d'appartenance $u_{ik} \in [0,1]$ à chaque pixel pour chaque cluster, capturant ainsi l'incertitude aux frontières. Des travaux récents, tels que ceux publiés en 2020 [75], ont amélioré le FCM par une sélection des clusters basée sur l'histogramme, optimisant ainsi l'initialisation et le nombre K de clusters pour la segmentation des lésions.

I.6.4.5 Hybridation avec la Régularisation Spatiale

Pour imposer une cohérence spatiale, le clustering est souvent combiné à d'autres méthodes. Dans leur étude de 2018 [76], , Jaisakthi et al. ont développé une approche hybride en utilisant le K-means pour la classification initiale des pixels, dont le résultat sert ensuite d'entrée à l'algorithme de GrabCut. Ce système utilise les étiquettes du clustering comme des termes de données pour guider un processus de coupure de graphe (méthode de régularisation spatiale), permettant d'obtenir un contour final lisse et précis.

I.6.4.6 Intégration au Deep Learning

La tendance la plus récente voit l'intégration du clustering dans des architectures d'apprentissage profond. Nawaz et al. (2022) [77] ont utilisé le K-means flou avec un modèle de Deep Learning pour la détection du cancer de la peau. De même, les auteurs dans [78]ont proposé une approche hybride combinant le FCM avec un réseau neuronal convolutif (CNN) optimisé. Dans ces schémas, le clustering agit souvent comme un outil de pré-segmentation floue pour initialiser ou affiner les masques de segmentation générés par le CNN, ou comme un moyen d'extraire des régions d'intérêt complexes pour la classification.

Le clustering a transcendé son statut de méthode de segmentation rudimentaire pour devenir un composant essentiel et adaptatif des systèmes hybrides modernes en dermoscopie.

I.6.5 Segmentation par Contours Actifs pour les Lésions Cutanées

Les Modèles de Contours Actifs (ACMs) sont des méthodes de segmentation traditionnelles cruciales, car elles modélisent la segmentation comme un problème de minimisation d'énergie, permettant d'obtenir des frontières lisses et cohérentes. Ces modèles se divisent principalement en approches paramétriques (Snakes) et géométriques (Level Sets).

I.6.5.1 Travaux Fondamentaux et Évolution vers les Level Sets

Les premières applications des ACMs en dermoscopie ont rapidement rencontré des difficultés dues au faible gradient aux frontières des lésions et à l'hétérogénéité des mélanomes (ce qui peut piéger le contour initial). La recherche s'est donc orientée vers des modèles capables de gérer ces défis.

- ✓ **Level Set Géométrique**

L'adoption des Level Sets, comme l'illustrent les travaux de [79], a été fondamentale. Représentant le contour implicitement, les Level Sets gèrent intrinsèquement les changements de topologie (fusion ou division des lésions), une limitation majeure des Snakes paramétriques.

- ✓ **Level Set Basé Région**

L'utilisation de modèles de Level Set basés sur la région comme le modèle Chan-Vese [59], dont le terme d'énergie est guidé par l'homogénéité des régions plutôt que par le seul gradient, a permis de segmenter les lésions aux frontières floues où l'information de bord est quasi inexistante.

I.6.5.2 Amélioration des Forces d'Attraction (Gradient Local)

Le concept de contour actif amélioré par les bords locaux (*Local edge-enhanced active contour*) vise à renforcer le signal de la frontière spécifique à la lésion. Dans ce cadre, Bayraktar et al. (2019) [80] ont proposé un modèle de détection des bords qui repose sur une fonction d'indicateur de bord robuste (basée sur une méthode sans maillage). Cette approche permet à la courbe de s'accrocher avec une précision accrue aux bords irréguliers et flous des lésions cutanées en étant moins sensible aux hétérogénéités non pertinentes du bruit ou du fond.

I.6.5.3 Fonctions de Pression Hybrides

Pour garantir une robustesse maximale face à l'hétérogénéité des images de lésions cutanées (variations d'illumination, bruit, faible contraste), des chercheurs ont développé des fonctions SPF hybrides. Par exemple, Soomro et al. (2018) [81] ont proposé une méthode à deux étapes utilisant une force motrice hybride. Cette approche permet :

- ✓ De capter l'information globale de la lésion, essentielle pour éviter la dépendance à l'initialisation du contour.
- ✓ D'intégrer l'information de bord locale pour que la courbe s'arrête avec précision sur les frontières faibles ou floues des lésions, ce qui améliore significativement la précision et la cohérence des contours extraits par rapport aux méthodes classiques.

I.6.5.4 Hybridaton avec l'Apprentissage Profond

L'intégration des ACMs dans des systèmes de classification et de segmentation basés sur le Deep Learning est une tendance majeure et récente. Dans [82] les auteurs ont proposé une approche hybride où le modèle de contour actif de type Snake est utilisé comme outil de raffinement final pour corriger les frontières générées par un réseau neuronal. Cette synergie permet d'exploiter la puissance des CNN pour la reconnaissance sémantique tout en bénéficiant de la régularité géométrique et de la précision au pixel des ACMs pour la délimitation clinique.

En résumé, les contours actifs, et en particulier les Level Sets, restent une méthode essentielle pour la segmentation des lésions cutanées, évoluant vers des modèles hybrides et adaptatifs capables de gérer l'incertitude et la faible information de bord.

I.6.6 Segmentation des Lésions Cutanées par MRF

L'utilisation des Champs Aléatoires de Markov (MRF) pour la segmentation des lésions cutanées est passée d'une méthode de segmentation principale à un outil de régularisation spatiale essentiel, visant à garantir la cohérence des contours. Historiquement, des travaux comme ceux de Torkashvand et Fartash (2015) [83] ont établi l'efficacité du MRF pour la

segmentation automatique, notamment pour lisser les résultats obtenus par des méthodes d'analyse de couleur.

L'évolution la plus significative réside dans l'intégration des MRF avec des techniques d'optimisation avancées pour améliorer la robustesse. Eltayef et al. (2017) [84] ont combiné le MRF avec des algorithmes d'initialisation précis ou d'optimisation stochastique. La recherches ont exploité l'optimisation par essaim de particules (Particle Swarm Optimization) pour affiner la segmentation basée sur le MRF, permettant une meilleure exploration de l'espace des solutions pour l'énergie du champ. Plus récemment, les travaux se sont concentrés sur la modélisation explicite de l'information spatiale pour pallier les limites des modèles MRF simples. Salih, Viriri, et Adegun (2019) [85] ont notamment introduit des modèles de MRF Région-Bord (Region-Edge MRF), qui intègrent simultanément la cohérence des régions et la détection précise des bords de la lésion, une avancée cruciale pour les contours irréguliers.

Ces approches ont également évolué vers des mécanismes de segmentation stochastique par fusion de régions (*stochastic region-merging*) combinés au Modèle de Champ Aléatoire de Markov (MRF) [86]. Cette combinaison est essentielle, car le rôle actuel des MRF est de stabiliser et de régulariser les résultats préliminaires pour obtenir des cartes de segmentation cohérentes et topologiquement correctes. Le MRF impose des contraintes de voisinage vitales pour les lésions cutanées aux contours spiculés. Que ce soit en post-traitement pour le lissage des frontières ou en tant que composant d'une fonction d'énergie hybride, les MRF confirment leur rôle central en transformant le problème de délimitation en une problématique d'optimisation globale (souvent résolue par *Graph Cut*), ce qui améliore la convergence et la robustesse du modèle final.

I.7 Conclusion

Ce chapitre a tout d'abord posé les bases scientifiques du domaine et rappelé l'importance de la segmentation en imagerie médicale, essentielle tant pour le diagnostic assisté par ordinateur que pour l'extraction fiable des régions d'intérêt. Une exploration des diverses techniques d'imagerie a été effectuée, chacune apportant ses contraintes propres en matière d'analyse. L'étude approfondie des images dermoscopiques a révélé des difficultés caractéristiques : forte variabilité entre patients, présence d'artefacts (poils, reflets, bulles), contraste limité entre la lésion et la peau saine, contours irréguliers ou flous. À cela s'ajoute un déséquilibre marqué entre les pixels de la lésion et ceux de l'arrière-plan, compliquant la détection automatique et limitant la capacité d'adaptation des méthodes classiques. Ces éléments ont permis de clarifier les enjeux spécifiques à la segmentation de lésions cutanées et de mettre en évidence les obstacles qui rendent les approches conventionnelles peu fiables dans ce contexte.

Les techniques traditionnelles de segmentation ont joué un rôle fondateur, mais souffrent de limites structurelles. Elles reposent sur des paramètres définis manuellement, n'intègrent aucune compréhension sémantique de l'image et s'appuient uniquement sur des caractéristiques de bas niveau, ce qui les rend très sensibles au bruit et aux irrégularités. Dans le cas de la dermoscopie, ces faiblesses se traduisent par des frontières mal détectées, des segments

discontinus ou l'inclusion de zones non pertinentes. Les hypothèses fondamentales de ces modèles sont souvent mises en défaut par l'hétérogénéité pigmentaire des lésions ou leur faible contraste, rendant leur application clinique incertaine et instable.

Face à ces limites, l'apprentissage profond s'est imposé comme une alternative décisive. En remplaçant les règles déterministes par un apprentissage automatique des représentations visuelles, il permet d'extraire des caractéristiques pertinentes à plusieurs échelles sans intervention humaine et de mieux gérer la variabilité des images. Les réseaux convolutifs, et en particulier les architectures dédiées comme U-Net, offrent une segmentation plus précise et robuste, même en présence d'artefacts ou de formes irrégulières. Ce chapitre a ainsi mis en évidence les insuffisances des méthodes traditionnelles et montré la nécessité de recourir à des modèles apprenants capables de s'adapter aux complexités réelles des images médicales. Le chapitre suivant approfondira ce nouveau paradigme en présentant ses fondements, ses principales architectures et son apport déterminant dans la segmentation automatique des lésions cutanées.

Chapitre II: Segmentation des images par Deep Learning

II.1 Introduction

L'Apprentissage Profond (Deep Learning, DL) représente une évolution significative de l'Apprentissage Automatique (Machine Learning, ML) ML. Basé sur des architectures neuronales multicouches, il apprend des représentations hiérarchiques des données à différents niveaux d'abstraction [87]. Inspiré du fonctionnement du cerveau humain, le DL est particulièrement performant pour le traitement de données complexes, telles que les images, les sons ou les textes. Ses fondements remontent aux travaux de McCulloch et Pitts en 1943, qui ont proposé le premier modèle de neurone artificiel, le Perceptron de McCulloch-Pitts, marquant le début des Réseaux de Neurones Artificiels (ANN) [88].

Parmi les architectures issues du Deep Learning, les Réseaux de Neurones Convolutifs (Convolutional Neural Networks, CNN) se distinguent par leur efficacité exceptionnelle dans l'analyse et la reconnaissance d'images [89]. Grâce à leur structure hiérarchique, les CNN extraient progressivement des caractéristiques spatiales à différents niveaux d'abstraction, ce qui permet une interprétation automatique et robuste des données visuelles [90]. Cette capacité a conduit à leur adoption massive dans des domaines variés tels que la vision par ordinateur, la conduite autonome, la sécurité biométrique et surtout l'imagerie médicale.

Le déploiement des réseaux neuronaux convolutifs (CNN) dans l'analyse de l'imagerie médicale a connu une intensification significative coïncidant avec la troisième période d'expansion de l'intelligence artificielle, notamment au début de la décennie 2010 [91]. Cette période a vu l'émergence de systèmes d'aide au diagnostic (Computer-Aided Diagnosis, CAD) capables d'analyser automatiquement des images médicales issues de différentes modalités, telles que l'IRM, la radiographie, la tomodensitométrie ou la dermoscopie [92], [93]. Ces systèmes permettent d'améliorer la détection précoce des lésions, de réduire la charge de travail des cliniciens et de minimiser la variabilité inter-observateurs.

Dans le domaine dermatologique, l'analyse automatique des images dermoscopiques est devenue un outil essentiel pour le diagnostic précoce du mélanome et d'autres affections cutanées. La segmentation précise des lésions représente une étape fondamentale, car elle conditionne la qualité des caractéristiques extraites pour la classification ou le suivi clinique. Les approches traditionnelles, basées sur des techniques de traitement d'images telles que le seuillage, les contours actifs ou les graph cuts, se heurtent toutefois à des limites liées à la variabilité d'éclairage, aux artefacts comme les poils ou les ombres, et à la diversité des textures cutanées.

II.2 Réseaux de neurones convolutifs (CNN)

Dans le domaine de l'apprentissage profond (DL), le CNN est l'algorithme le plus célèbre et le plus couramment employé. Son principal avantage par rapport à ses prédécesseurs est sa capacité à identifier automatiquement les caractéristiques pertinentes sans supervision humaine. Les CNN ont été largement appliqués dans divers domaines, incluant la vision par ordinateur , le traitement de la parole et la reconnaissance faciale [94], [95], [96].

La structure des CNN s'inspire du cortex visuel, notamment des cellules qui ne perçoivent que de petites régions d'une scène, simulant ainsi l'extraction de la corrélation locale par des filtres

[97]. Goodfellow et al. [98] ont identifié trois avantages clés du CNN : les représentations équivalentes, les interactions éparses et le partage des paramètres. Contrairement aux réseaux *fully connected* (FC) conventionnels, les poids partagés et les connexions locales sont employés pour exploiter pleinement les structures de données d'entrée 2D comme les images, réduisant ainsi le nombre de paramètres, ce qui simplifie l'entraînement et accélère le réseau. Un exemple d'architecture CNN pour la classification d'images est d'ailleurs illustré dans la Figure II.2.

Un CNN standard enchaîne plusieurs couches de convolution et de sous-échantillonnage (pooling), avant de s'achever par des couches entièrement connectées. L'entrée X de chaque couche est organisée en trois dimensions (hauteur, largeur, profondeur) : $H \times W \times D$, où la hauteur (H) est égale à la largeur (W). La profondeur (D) est le nombre de canaux. Plusieurs noyaux (filtres) K sont disponibles dans chaque couche convolutive, notés F, avec trois dimensions $k_h \times k_w \times D$. Ici, k_h doit être plus petit que H et k_w doit être inférieur ou égal à W. Ces noyaux, qui partagent des paramètres similaires (biais b et poids W), sont à la base des connexions locales et sont convolués avec l'entrée pour générer F cartes de caractéristiques A de taille $H' \times W'$.

La couche de convolution effectue un produit scalaire entre son entrée et ses poids W. Le résultat est ensuite transformé par une fonction d'activation non linéaire f, fournissant la sortie de la couche, comme illustré par l'équation :

$$\mathrm{A}_{i,j,k} = f\left(b_k + \sum_{l=1}^{D}\sum_{p=1}^{k_h}\sum_{q=1}^{k_w} \mathrm{W}_{p,q,l,k}\,\mathrm{X}_{i+p,j+q,l}\right) \qquad \text{(II.1)}$$

L'étape suivante utilise les couches de sous-échantillonnage (*pooling*), où une fonction de *pooling* (par exemple *max* ou *moyenne*) est appliquée à une zone adjacente de taille $k \times k$ pour chaque carte de caractéristiques. Cela réduit les paramètres du réseau, accélère l'entraînement et permet de gérer le surapprentissage (*overfitting*).

Enfin, les couches FC reçoivent les caractéristiques de niveau moyen et bas pour créer l'abstraction de haut niveau. Ces couches finales génèrent les scores de classification (probabilités pour chaque classe) en utilisant une couche finale comme *softmax*

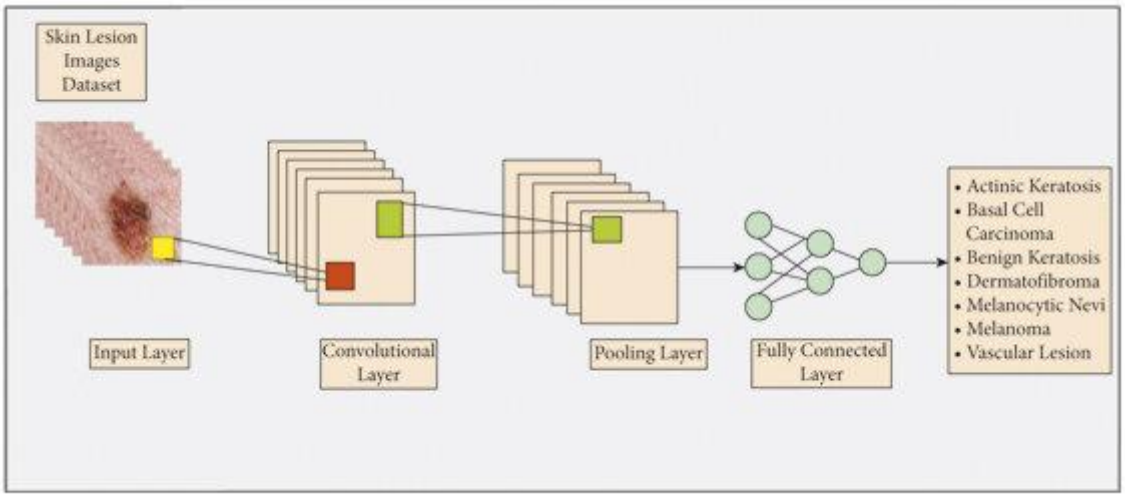


Figure II.1 Un exemple d'architecture CNN pour la classification d'images cas lésion cutanée

II.2.1 Avantages de l'utilisation des CNN

Les avantages de l'utilisation des CNN par rapport aux autres réseaux neuronaux traditionnels dans l'environnement de la vision par ordinateur sont listés comme suit :

- La principale raison de considérer le CNN est la fonctionnalité de partage des poids (weight sharing), qui réduit le nombre de paramètres entraînables du réseau et aide ainsi le réseau à améliorer la généralisation et à éviter le surapprentissage.
- L'apprentissage simultané des couches d'extraction de caractéristiques et de la couche de classification fait que la sortie du modèle est à la fois hautement organisée et fortement dépendante des caractéristiques extraites.
- La mise en œuvre de réseaux à grande échelle est beaucoup plus facile avec les CNN qu'avec d'autres réseaux de neurones.

II.2.2 Couches des CNN

L'architecture CNN se compose d'un certain nombre de couches (ou de blocs de construction multiples). Chaque couche de l'architecture CNN, y compris sa fonction, est décrite en détail ci-dessous.

II.2.2.1 Couche Convolutive

Dans l'architecture CNN, l'élément le plus important est la couche convolutive. Elle consiste en un ensemble de filtres convolutifs (appelés noyaux ou kernels). L'image d'entrée, exprimée sous forme de matrices N-dimensionnelles, est convoluée avec ces filtres pour générer la carte de caractéristiques (feature map) de sortie.

- **Définition du Noyau**

Un noyau est décrit par une grille de nombres ou de valeurs discrètes. Chaque valeur est appelée le poids du noyau. Des nombres aléatoires sont attribués pour servir de poids du noyau au début du processus d'entraînement du CNN. Ces poids sont ajustés à chaque époque d'entraînement ; ainsi, le noyau apprend à extraire des caractéristiques significatives.

- **Opération de Convolution**

Les Réseaux Neuronaux Convolutifs (CNN) se distinguent des réseaux traditionnels car ils traitent directement les images à canaux multiples (un canal pour le niveau de gris, trois pour le RVB).

Cette opération implique un noyau qui glisse sur l'image d'entrée, pixel par pixel. À chaque position, elle calcule le produit scalaire entre le noyau et la région locale qu'il recouvre (multiplication élément par élément suivie d'une sommation). L'ensemble de ces résultats scalaires forme une carte de caractéristiques, qui met en évidence des motifs spécifiques (tels que des contours) et préserve les relations spatiales (Figure II.2). Cette opération est régie par deux hyperparamètres principaux :

1. Le pas (stride) : Définit la taille du déplacement du noyau. Un pas supérieur à 1 réduit la taille de la carte de caractéristiques de sortie.
2. Le remplissage (padding) : Consiste à ajouter une bordure à l'image d'entrée. Il est essentiel pour contrôler la taille spatiale de la sortie et préserver les informations situées aux bords de l'image.

- **Avantages Clés des Couches de Convolution**

1. Connectivité Creuse (Sparse Connectivity) : Contrairement aux réseaux entièrement connectés (FC), dans les CNN, les neurones adjacents ne sont reliés qu'avec une partie des neurones de la couche suivante. Cela réduit considérablement le nombre de poids requis et l'espace mémoire nécessaire, rendant l'approche mémoire-efficace.
2. Partage de Poids (Weight Sharing) : Le même noyau utilise son unique ensemble de poids sur tous les pixels de l'entrée. Le fait d'apprendre un seul groupe de poids pour l'image entière diminue fortement le temps d'entraînement et les coûts de calcul, car il n'est pas nécessaire d'allouer des poids distincts pour chaque neurone.

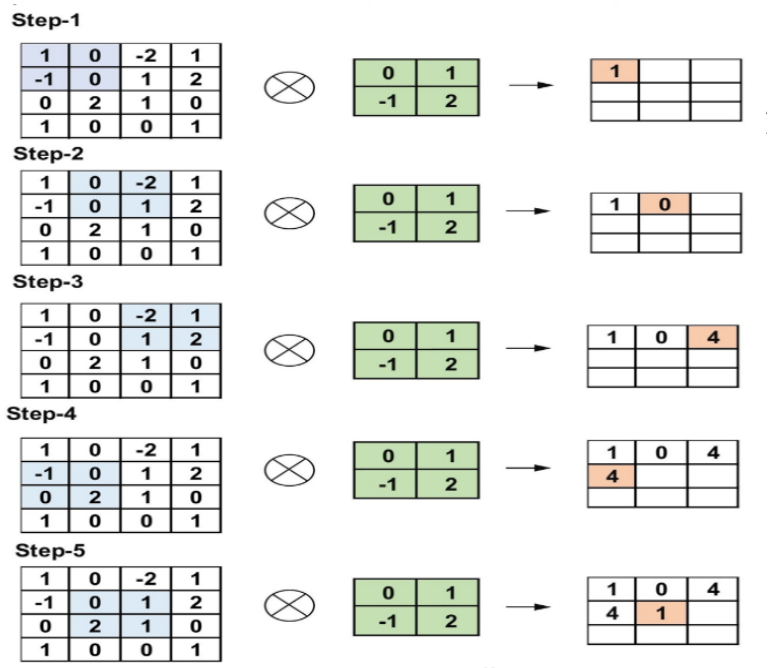


Figure II.2 Les calculs principaux exécutés à chaque étape de la couche convolutive

II.2.2.2 Couches de pooling

Le rôle principal de la couche de pooling est le sous-échantillonnage (sub-sampling) des cartes de caractéristiques générées par la convolution. Cette approche permet de réduire la taille des grandes cartes de caractéristiques pour en créer de plus petites, tout en maintenant la majorité des informations dominantes ou des caractéristiques. Comme la convolution, l'opération de pooling est définie par la taille du noyau et le pas. Plusieurs méthodes de *pooling* existent (arborescence, *gated*, moyenne, min, etc.), mais les plus courantes sont le Max Pooling, le Min Pooling et le Global Average Pooling (GAP). La Figure II.3 illustre ces trois opérations de pooling.

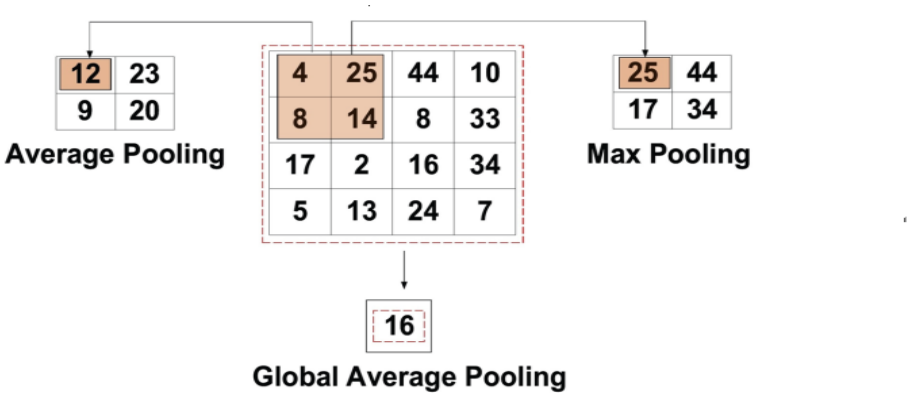


Figure II.3 Trois types d'opérations de pooling

La principale lacune du *pooling* est qu'il permet au CNN de déterminer la présence d'une caractéristique sans se concentrer exclusivement sur sa localisation exacte. Le modèle peut ainsi manquer des informations de localisation pertinentes, diminuant parfois la performance globale.

II.2.2.3 Fonction d'Activation (Non-linéarité)

La fonction d'activation a pour rôle de mapper l'entrée vers la sortie d'un neurone. Elle prend la sommation pondérée de l'entrée du neurone (avec son biais) et décide si le neurone doit être activé, produisant l'output correspondant.

Les couches d'activation non-linéaires sont appliquées après toutes les couches ayant des poids (couches de convolution et FC). Cette non-linéarité est essentielle car elle confère au CNN la capacité d'apprendre des relations extrêmement complexes et non-linéaires entre les données. De plus, la fonction d'activation doit être différentiable pour permettre l'utilisation de la rétropropagation de l'erreur (error back-propagation), un mécanisme indispensable pour l'entraînement du réseau. Les fonctions d'activation suivantes les plus sont utilisées dans les CNN :

1. Sigmoïde

L'entrée de cette fonction d'activation est des nombres réels, tandis que la sortie est restreinte entre zéro et un. La courbe de la fonction sigmoïde est en forme de S et peut être représentée mathématiquement par :

$$f(x)_{\text{sigm}} = \frac{1}{1 + e^{-x}} \tag{II.2}$$

2. Tangente Hyperbolique (Tanh)

Elle est similaire à la fonction sigmoïde, car son entrée est des nombres réels, mais la sortie est restreinte entre -1 et 1. Sa représentation mathématique est :

$$f(x)_{\text{tanh}} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{II.3}$$

3. ReLU (Unité Linéaire Rectifiée)

La fonction la plus couramment utilisée dans le contexte des CNN. Elle convertit toutes les valeurs de l'entrée en nombres positifs. Une charge computationnelle plus faible est le principal avantage de ReLU par rapport aux autres. Sa représentation mathématique est :

$$f(x)_{\text{ReLU}} = \max(0, x) \tag{II.4}$$

Occasionnellement, quelques problèmes significatifs peuvent survenir lors de l'utilisation de ReLU. Par exemple, considérons un algorithme de rétropropagation de l'erreur avec un

gradient plus important le traversant. Le passage de ce gradient dans la fonction ReLU mettra à jour les poids d'une manière qui rend le neurone certainement non activé à nouveau. Ce problème est appelé « Dying ReLU » [99]. Il existe des alternatives à ReLU pour résoudre de tels problèmes telle que la fonction Leaky ReLU.

4. Leaky ReLU

Au lieu que ReLU réduise les entrées négatives, cette fonction d'activation assure que ces entrées ne sont jamais ignorées. Elle est employée pour résoudre le problème du Dying ReLU. Elle est définie par :

$$f(x)_{\text{LeakyReLU}} = \begin{cases} x, & \text{si } x > 0 \\ mx, & x \leq 0 \end{cases} \quad \text{(II.5)}$$

5. SoftMax

Le SoftMax est une fonction d'activation neuronale largement utilisée dans les couches de sortie des réseaux de neuronaux pour les tâches de classification lorsque le nombre de classes de sortie dépasse un. Il est également connu sous le nom de softargmax ou fonction exponentielle normalisée. Le résultat de cette activation est la distribution de probabilité des valeurs du vecteur d'entrée parmi plusieurs classes. Par conséquent, SoftMax est une alternative multi-classes à l'activation sigmoïde. L'activation SoftMax a la forme suivante :

$$S(x_1, \dots, x_N)_i = \frac{\exp(x_i)}{\sum_{n=1}^{N} \exp(x_n)} \quad \text{(II.6)}$$

Où N est le nombre d'entrées.

II.2.2.4 Couche Entièrement Connectée (Fully Connected Layer)

Généralement, cette couche est située à la fin de chaque architecture CNN. Dans cette couche, chaque neurone est connecté à tous les neurones de la couche précédente, d'où l'appellation « entièrement connectée » (FC). Elle est utilisée comme classifieur du CNN. Elle suit la méthode de base du perceptron multicouche conventionnel, car c'est un type de réseau de neurones artificiels à propagation avant. L'entrée de la couche FC provient de la dernière couche de pooling ou de convolution. Cette entrée est sous forme de vecteur, créé à partir des cartes de caractéristiques après aplatissement. La sortie de la couche FC représente la sortie finale du CNN, comme illustré dans la Figure II.4.

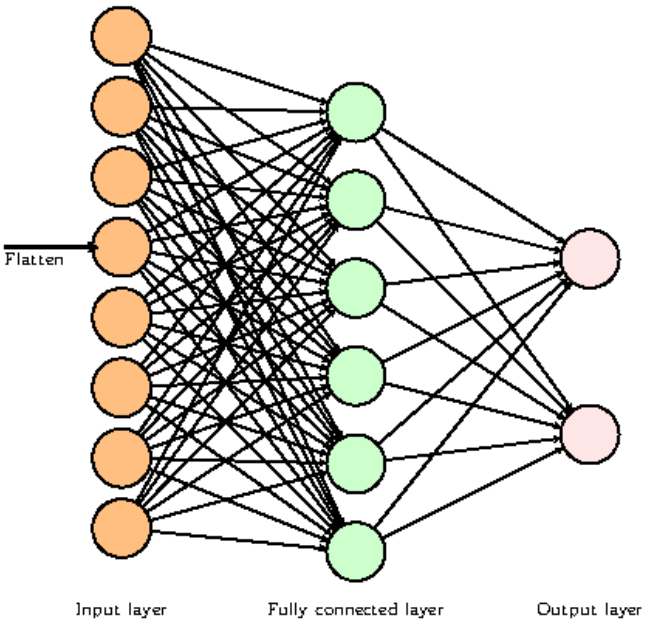


Figure II.4 : Couche entièrement connectée

II.2.2.5 Fonctions de Perte (Loss Functions)

Dans l'architecture d'un modèle d'apprentissage profond, le composant final est généralement dédié au calcul de l'objectif que le réseau doit apprendre à minimiser. Cet élément critique, connu sous le nom de couche de perte, marque la fin de la propagation avant (forward pass). Elle traite les prédictions brutes du réseau ainsi que les étiquettes de vérité terrain pour produire une valeur scalaire unique : la perte. Cette valeur de perte est ensuite utilisée lors de la propagation arrière (backward pass) pour calculer les gradients qui guident l'optimisation de toutes les couches précédentes. Bien que de nombreuses couches de perte spécialisées existent pour des défis spécifiques, plusieurs couches standard forment le fondement architectural pour une grande variété de tâches de vision par ordinateur :

1. ***Couche de Perte par Entropie Croisée***

La couche de Perte par Entropie Croisée (CE) sert de composant terminal standard dans les architectures de classification basées sur les CNN (comme ResNet ou VGG). Elle suit généralement une couche d'activation softmax pour les scénarios multi-classes, ou une activation sigmoïde pour la classification binaire.

La fonction de cette couche est de comparer la distribution de probabilités prédite par le modèle ($\hat{y}$) avec la distribution de vérité terrain (y). Sa force réside dans sa capacité à produire des gradients stables et robustes pour la mise à jour des poids, fournissant un signal d'apprentissage efficace, même lorsque les prédictions initiales sont très incorrectes. Pour une classification multi-classes, la perte par Entropie Croisée (L_{CE}) est calculée comme suit :

$$L_{CE} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \quad \text{(II.7)}$$

Où :

- ✓ C est le nombre total de classes.
- ✓ y_c est l'étiquette de vérité terrain pour la classe c.
- ✓ $\hat{y}_c$ est la probabilité prédite pour la classe c par la couche *softmax*.

2. Couche de Perte L1 et L2

L'Erreur Quadratique Moyenne (MSE), également connue sous le nom de perte L2, est l'une des fonctions de perte de régression les plus fondamentales et les plus largement utilisées en apprentissage automatique et en apprentissage profond. Elle quantifie la moyenne des différences quadratiques entre les valeurs prédites et les valeurs de vérité terrain. La couche de perte L2 calcule :

$$\mathcal{L}_{\text{MSE}}(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - f_\theta(x_i))^2 \quad \text{(II.8)}$$

Avec :

- ✓ θ représente les paramètres du modèle
- ✓ n est le nombre d'échantillons
- ✓ y_i est la valeur vraie du i-ième échantillon
- ✓ $\hat{y}_i$ ou $f_\theta(x_i)$ est la valeur prédite du i-ième échantillon

Cette couche pénalise fortement les grandes erreurs grâce à son terme quadratique, ce qui peut favoriser la convergence mais rend également la sortie du réseau très sensible aux valeurs aberrantes dans les données d'entraînement.

L'Erreur Absolue Moyenne (MAE), également connue sous le nom de perte L1, est un autre choix courant pour les tâches de régression. Elle mesure la moyenne des différences absolues entre les valeurs prédites et les valeurs réelles. La couche de perte L1 calcule

$$\mathcal{L}_{\text{MAE}}(\theta) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| = \frac{1}{n}\sum_{i=1}^{n}|y_i - f_\theta(x_i)| \quad \text{(II.9)}$$

Contrairement à L2, la couche de perte L1 démontre une plus grande robustesse aux valeurs aberrantes en appliquant une pénalité linéaire. Cependant, son gradient à amplitude constante peut conduire à une convergence moins stable près de l'optimum.

3. Couche Perte de Dice

La couche de perte de Dice fonctionne comme une perte basée sur les régions et est devenue un composant final standard dans les architectures de segmentation comme U-Net et V-Net [100], [101], conçues pour gérer les déséquilibres de classes sévères. Elle optimise directement le chevauchement entre les cartes de segmentation prédites et les masques de vérité terrain.

Cette couche calcule la perte comme :

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} p_i\, g_i + \epsilon}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2 + \epsilon}, \tag{II.10}$$

où p_i représente les prédictions pixel par pixel provenant des sorties sigmoïde ou softmax et g_i désigne les étiquettes pixel de vérité terrain. Le terme ϵ fournit une constante de lissage pour la stabilité numérique. Architecturalement, la couche de perte de Dice permet aux réseaux d'apprendre directement à partir du chevauchement des régions spatiales, se révélant très efficace pour des tâches comme la segmentation de lésions où les cibles occupent une petite fraction déséquilibrée de l'image.

Pour les réseaux de segmentation, les couches de perte de Dice et d'Entropie Croisée Pondérée offrent des solutions standard puissantes pour le déséquilibre des classes au niveau pixel. Architecturalement, ces couches de perte ne sont pas seulement efficaces individuellement, mais elles établissent également la base pour dériver des pertes avancées pour des problèmes plus spécifiques et difficiles.

4. Couche de Perte par Entropie Croisée Pondérée

La couche de perte par Entropie Croisée Pondérée (WCE) modifie la couche CE standard pour traiter le déséquilibre des classes dans les réseaux de segmentation. Elle a été mise en avant dans l'architecture U-Net originale[102].

Pour la segmentation binaire, la couche WCE calcule :

$$\mathcal{L}_{\text{WCE}} = -\frac{1}{N}\sum_{i=1}^{N}(w g_i \log(p_i) + (1 - g_i)\log(1 - p_i)), \tag{II.11}$$

où w représente un poids appliqué à la classe positive. Lorsque $w > 1$, la perte amplifie les pixels d'avant-plan mal classés (Faux Négatifs), rééquilibrant efficacement le signal de gradient qui se propage à travers le réseau. Cela oblige l'architecture encodeur-décodeur à accorder plus d'attention aux classes sous-représentées pendant l'entraînement.

II.2.3 Régularisation des CNN

Pour les modèles CNN, le surapprentissage (over-fitting) représente le problème central associé à l'obtention d'une généralisation bien comportée. Le modèle est dit surajusté dans les cas où le modèle performe particulièrement bien sur les données d'entraînement mais échoue sur les données de test (données non vues). Un modèle sous-ajusté (under-fitted) est le cas opposé ; cela se produit lorsque le modèle n'apprend pas suffisamment à partir des données

d'entraînement. Le modèle est qualifié de « justement ajusté » (just-fitted) s'il performe bien à la fois sur les données d'entraînement et de test. Ces trois types sont illustrés dans la Figure (II .5) :

- ✓ Sous-ajustement (*Underfitting*) : La ligne séparatrice est trop simple (souvent linéaire) et ne parvient pas à séparer les classes, entraînant une erreur élevée due à un manque de complexité du modèle.
- ✓ Surapprentissage (*Overfitting*) : La ligne est trop complexe et irrégulière, épousant parfaitement chaque point de l'entraînement. Le modèle a mémorisé le bruit, ce qui le rend inefficace et instable sur de nouvelles données.
- ✓ Ajustement Correct (*Balanced*) : La ligne est optimale, suffisamment flexible pour capturer la tendance générale sans suivre les anomalies. Cela garantit une bonne généralisation et une précision équilibrée.

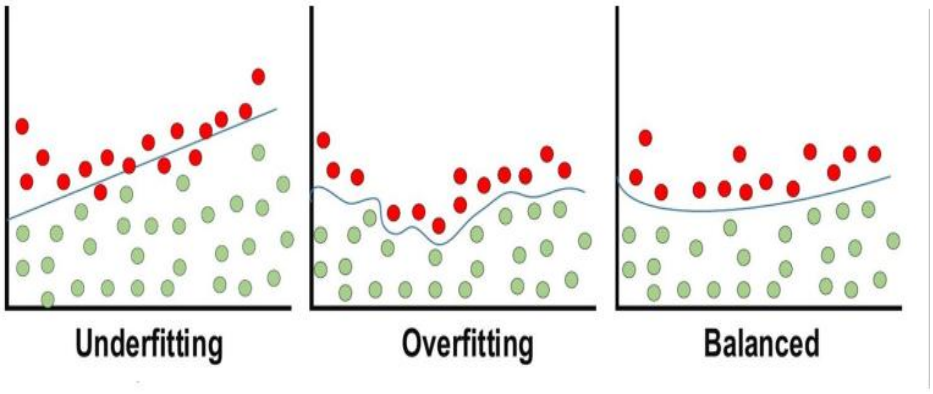


Figure II.5 Illustration du surapprentissage, du sous-ajustement et de l'ajustement correct

Des techniques de régularisation pour les CNN sont utilisées pour gérer ces problèmes :

II.2.3.1 Dropout

Il s'agit d'une technique largement utilisée pour la généralisation. À chaque époque d'entraînement, des neurones sont abandonnés aléatoirement. Ce faisant, le pouvoir de sélection des caractéristiques est réparti de manière égale sur l'ensemble du groupe de neurones, tout en forçant le modèle à apprendre différentes caractéristiques indépendantes. Pendant le processus d'entraînement, le neurone abandonné ne fait pas partie de la rétropropagation ou de la propagation avant. En revanche, le réseau à pleine échelle est utilisé pour effectuer les prédictions pendant le processus de test.

II.2.3.2 Augmentation des Données (Data Augmentation)

Entraîner le modèle sur une quantité importante de données est le moyen le plus simple d'éviter le surapprentissage. Pour y parvenir, l'augmentation des données est utilisée. Plusieurs techniques sont utilisées pour augmenter artificiellement la taille de l'ensemble de données

d'entraînement. Plus de détails peuvent être trouvés dans la section ultérieure, qui décrit les techniques d'augmentation des données.

II.2.3.3 Normalisation par Lots (Batch Normalization)

La Normalisation du Batch (BN) est une technique essentielle qui assure que la performance des activations de sortie suit une distribution gaussienne unitaire. L'opération consiste à soustraire la moyenne et à diviser par l'écart-type pour normaliser la sortie à chaque couche. Pour une mini-batch $\mathcal{B}$ et un ensemble d'activations x_i est définie par :

$$\hat{x}_i = \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \tag{II.12}$$

Où :

- ✓ $\mu_{\mathcal{B}}$ est la moyenne de la mini-batch.
- ✓ $\sigma_{\mathcal{B}}^2$ est la variance de la mini-batch.
- ✓ ϵ est une petite constante ajoutée pour la stabilité numérique.

Son principal rôle est de réduire le « décalage de covariance interne » (Internal Covariate Shift), défini par la variation de la distribution d'activation à chaque couche. Ce décalage augmente avec la mise à jour des poids durant l'entraînement, et est exacerbé lorsque les données d'entraînement proviennent de sources dissimilaires (ex. : images de jour et de nuit). En réduisant ce décalage, la BN permet au modèle de converger plus rapidement, réduisant ainsi le temps d'entraînement.

II.2.4 Sélection de l'Optimiseur

La fonction de base de tous les algorithmes d'apprentissage supervisé est de minimiser les fonctions de perte (ou de minimiser l'erreur entre la sortie réelle et la sortie prédite) en ajustant les paramètres apprenables (biais, poids). Les techniques d'apprentissage basées sur le gradient sont le choix habituel pour les CNN. Le réseau doit mettre à jour ses paramètres au cours de chaque époque d'entraînement tout en recherchant la réponse optimisée localement.

Le taux d'apprentissage (learning rate) est défini comme la taille du pas de la mise à jour des paramètres, et l'époque d'entraînement (training epoch) représente une répétition complète de la mise à jour impliquant l'ensemble des données. Le taux d'apprentissage doit être sélectionné judicieusement comme hyperparamètre.

II.2.4.1 Algorithme de Descente de Gradient

Pour minimiser l'erreur d'entraînement, l'algorithme de Descente de Gradient (Gradient Descent) est utilisé. Il met à jour les paramètres du réseau de manière répétitive à travers chaque époque en calculant le gradient (pente) de la fonction objectif par une dérivée du premier ordre. Le paramètre est ensuite mis à jour dans la direction inverse du gradient pour réduire l'erreur. Cette mise à jour est effectuée via la rétropropagation et est donnée par :

$$W^t = W^{t-1} - \eta \frac{\partial E}{\partial W^{t-1}} \quad \text{(II.13)}$$

Où W^t est le poids final dans l'époque actuelle, W^{t-1} est le poids dans l'époque précédente, η est le taux d'apprentissage et E est l'erreur de prédiction.

II.2.4.2 Alternatives de la Descente de Gradient

Différentes alternatives de l'algorithme d'apprentissage basé sur le gradient sont couramment employées :

1. **Descente de Gradient par Lots (*Batch Gradient Descent*)** Lors de l'exécution de cette technique [29], les paramètres sont mis à jour une seule fois par époque après avoir calculé le gradient sur l'ensemble complet des données d'entraînement. Pour un petit jeu de données, elle converge rapidement et crée un gradient très stable, mais elle nécessite une quantité substantielle de ressources pour les grands ensembles.

2. **Descente de Gradient Stochastique (*Stochastic Gradient Descent*)** Dans cette technique [30], les paramètres sont mis à jour à chaque échantillon d'entraînement. Elle est efficace en mémoire et plus rapide que la BGD pour un grand jeu de données. Cependant, les mises à jour fréquentes la rendent bruyante, ce qui conduit à un comportement de convergence hautement instable.

3. **Descente de Gradient par Mini-Lots (*Mini-batch Gradient Descent*)** Cette approche partitionne les échantillons en plusieurs mini-lots non chevauchants [31]. La mise à jour des paramètres est effectuée après le calcul du gradient sur chaque mini-lot. Elle combine les avantages de la BGD et de la SGD, offrant une convergence stable, une meilleure efficacité computationnelle et une plus grande efficacité mémoire.

II.2.4.3 Améliorations dans les Algorithmes d'Apprentissage

Des techniques d'amélioration, généralement dans la SGD, sont employées pour optimiser le processus d'entraînement.

1. **Momentum (Quantité de Mouvement)** Cette technique améliore la précision et la vitesse d'entraînement en additionnant le gradient actuel avec le gradient calculé à l'étape précédente, pondéré par un facteur γ (facteur de *momentum*). Le *momentum* est utilisé pour résoudre l'inconvénient principal des algorithmes basés sur le gradient : le risque de rester bloqué dans un minimum local sur des surfaces non convexes. Mathématiquement, elle est exprimée par :

$$\Delta W^t = -\eta \frac{\partial E}{\partial W^{t-1}} + \gamma \Delta W^{t-1} \quad \text{(II.14)}$$

Où ΔW^t est l'incrément de poids dans l'époque actuelle, η est le taux d'apprentissage, et ΔW^{t-1} est l'incrément de poids dans l'époque précédente. La valeur du facteur γ (entre 0 et 1) contrôle la taille du pas et le lissage des variations.

2. **Estimation Adaptive des Moments (*Adam*)** Adam [32] est une technique d'optimisation largement utilisée et représente les tendances récentes. Conçue spécifiquement pour l'entraînement des réseaux profonds, Adam intègre les avantages de *Momentum* et de *RMSprop*. Son mécanisme consiste à calculer un taux d'apprentissage adaptatif pour chaque paramètre du modèle, utilisant à la fois la moyenne mobile du gradient et les gradients au carré. L'équation d'Adam est donnée par :

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \tag{II.15}$$

Avec :

- $\hat{m}_t$: Moyenne mobile du premier ordre des gradients.
- $\hat{v}_t$: Moyenne mobile du second ordre des gradients.
- ϵ : Constante de régularisation évitant la division par zéro.

II.3 Architectures de Réseaux de Neurones Convolutifs pour l'Analyse d'Images Médicales

L'évolution des réseaux de neurones convolutifs profonds (Deep Convolutional Neural Networks) a considérablement transformé les performances des systèmes de vision par ordinateur, particulièrement dans le domaine de l'analyse d'images médicales [8]. Cette transformation s'opère grâce à l'émergence d'architectures innovantes permettant l'extraction automatique de caractéristiques hiérarchiques à partir de données d'imagerie complexes.

Bien que les fondements théoriques des réseaux de neurones aient été établis dès les décennies 1980-1990 avec les travaux pionniers de LeCun et collaborateurs sur les réseaux convolutifs [103], [104], c'est en 2012 que survient une avancée déterminante avec la proposition de l'architecture AlexNet par Krizhevsky et collaborateurs [91]. Cette contribution marque un tournant significatif en démontrant la faisabilité pratique de l'apprentissage de représentations profondes grâce à l'exploitation conjointe d'unités de traitement graphique (GPU) et de corpus de données étendus. La performance remarquable obtenue lors du l'ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) a établi un nouveau paradigme dans le domaine de la vision artificielle.

Cette percée a ouvert la voie au développement d'architectures spécialisées pour la segmentation d'images, notamment les Réseaux Entièrement Convolutifs (FCN) [105], U-Net [106], SegNet [107] et ResNet [108], qui constituent l'objet central de notre étude. Ces architectures ont successivement abordé des défis fondamentaux tels que la préservation de l'information spatiale, la gestion du gradient évanescent et l'optimisation de la reconstruction sémantique.

Malgré ces avancées significatives, plusieurs défis persistent dans l'application des réseaux convolutifs profonds à l'imagerie médicale. La nécessité de grands ensembles de

données annotées, la robustesse aux artefacts d'acquisition, et l'interprétabilité des décisions automatiques constituent des axes de recherche actifs. L'émergence récente des transformers [109] et des méthodes d'apprentissage auto-supervisé [110] laisse entrevoir de nouvelles perspectives pour adresser ces limitations.

II.3.1 Les Réseaux Entièrement Convolutifs (FCN)

Les Réseaux Entièrement Convolutifs, ou Fully Convolutional Networks (FCN), introduits par Shelhamer et al. en 2015 [105], ont marqué une avancée majeure dans le domaine de la segmentation sémantique. Cette approche a permis d'adapter les architectures initialement conçues pour la classification d'images afin d'effectuer une prédiction dense au niveau du pixel.

L'idée fondatrice repose sur la transformation des réseaux de neurones convolutifs traditionnels, tels qu'AlexNet [91], VGGNet [111] ou GoogLeNet [112], en architectures entièrement convolutives capables de traiter des images de taille arbitraire et de produire des cartes de prédiction de dimensions correspondantes.

Le principe de cette transformation repose sur la conversion des couches entièrement connectées en couches convolutives de type 1×1. Dans les architectures de classification, les dernières couches d'un réseau tel que VGG servent à convertir la carte de caractéristiques finale en un vecteur de probabilités correspondant aux différentes classes. Les FCN remplacent ces couches entièrement connectées par des convolutions 1×1, ce qui permet de conserver la dimension spatiale des caractéristiques.

Le réseau devient ainsi totalement convolutif et peut accepter une image de taille $H \times W$ pour produire une carte de score de taille réduite, par exemple $H/32 \times W/32$, en fonction du degré de sous-échantillonnage introduit par les opérations de pooling. Cette architecture conserve donc l'information spatiale nécessaire à la segmentation tout en permettant une prédiction continue à l'échelle du pixel. La Figure II.6 illustre les architectures FCN-32s, FCN-16s et FCN-8s du modèle FCN. Les connexions de saut permettent la fusion des cartes de caractéristiques à différents niveaux de profondeur (Pool3, Pool4 et Pool5).

II.3.1.1 Architecture FCN-32s

Le réseau convolutif, initialement entraîné pour la classification, génère une carte de caractéristiques profonde contenant une information sémantique riche. Chaque élément de cette carte représente la probabilité d'appartenance d'une région de l'image à une classe donnée, formant ainsi une carte de score sémantique.

L'une des premières variantes, le FCN-32s, se caractérise par un facteur de réduction spatial de 32. Cette architecture prend la carte de score issue de la dernière couche de l'encodeur, dont la résolution est 32 fois inférieure à celle de l'image d'entrée, et applique une opération de déconvolution ou de convolution transposée afin de restaurer la taille originale. Cette opération est réalisée à l'aide d'un filtre appris qui assure le suréchantillonnage de la carte de caractéristiques.

Toutefois, ce procédé présente une limitation importante : le suréchantillonnage à partir d'une carte fortement réduite entraîne une perte considérable d'information spatiale, ce qui

conduit à des prédictions lissées et à des contours flous. Les cinq étapes successives de Max-Pooling provoquent une diminution drastique de la résolution et des détails locaux, rendant difficile la délimitation précise des structures fines.

Afin d'atténuer ce problème, les auteurs ont proposé l'utilisation de connexions de saut (skip connections), introduisant ainsi les variantes FCN-16s et FCN-8s. Ces architectures permettent de combiner les informations sémantiques issues des couches profondes avec les informations spatiales provenant des couches intermédiaires ou superficielles.

II.3.1.2 Architecture FCN-16s

Dans le cas du modèle FCN-16s, la carte de caractéristiques issue de la couche Pool5 est d'abord suréchantillonnée par un facteur de deux, puis combinée par addition avec la carte provenant de Pool4, dont la résolution a été réduite seize fois seulement par rapport à l'image originale. Cette fusion permet de réintroduire une partie de l'information spatiale perdue et de produire une carte de segmentation plus fine.

II.3.1.3 Architecture FCN-8s

La version FCN-8s prolonge cette approche en intégrant également les informations issues de la couche Pool3, caractérisée par la plus haute résolution parmi les couches profondes. La carte obtenue après fusion de Pool4 et Pool5 est à nouveau suréchantillonnée, puis combinée avec celle de Pool3, ce qui enrichit la représentation spatiale et améliore la précision des contours. Ainsi, le modèle FCN-8s fournit des cartes de segmentation nettement plus détaillées et mieux localisées que celles de FCN-32s et FCN-16s.

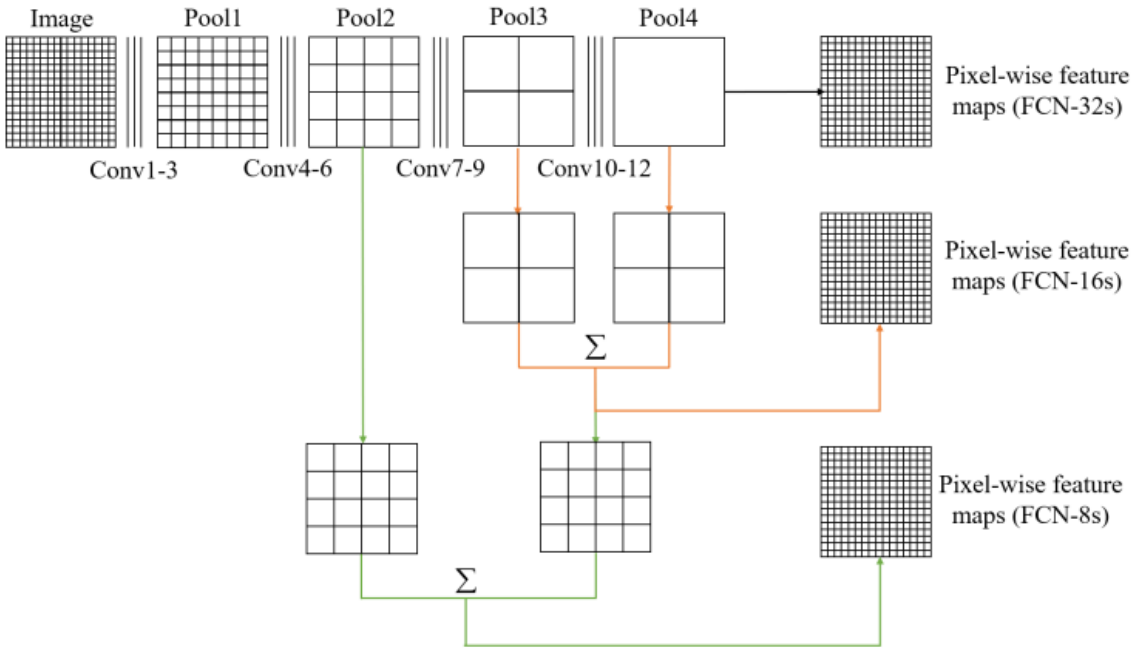


Figure II.6 Les architectures FCN-32s, FCN-16s et FCN-8s.

Les FCN ont ainsi ouvert la voie à une nouvelle génération d'architectures de segmentation sémantique en établissant un lien explicite entre classification et localisation. Leur principal inconvénient réside néanmoins dans la nécessité de stocker et de transmettre les cartes de caractéristiques des couches intermédiaires, ce qui augmente considérablement la charge mémoire lors de l'entraînement et de l'inférence. Cette contrainte a conduit à la conception d'architectures plus efficaces telles que SegNet, qui visent à pallier cette limitation en exploitant des indices de Max-Pooling pour réduire la redondance des données.

II.3.2 Le modele SegNet

Le modèle SegNet [107] est une architecture de réseau neuronal convolutif profond spécialement conçue pour la segmentation sémantique au niveau du pixel. Son objectif principal est d'attribuer une étiquette de classe sémantique à chaque pixel d'une image d'entrée.

L'architecture de SegNet suit un schéma encodeur-décodeur classique mais se distingue par son mécanisme innovant de suréchantillonnage. Comme illustré dans la Figure II.7, le réseau est composé de deux parties symétriques encodeur et décodeur.

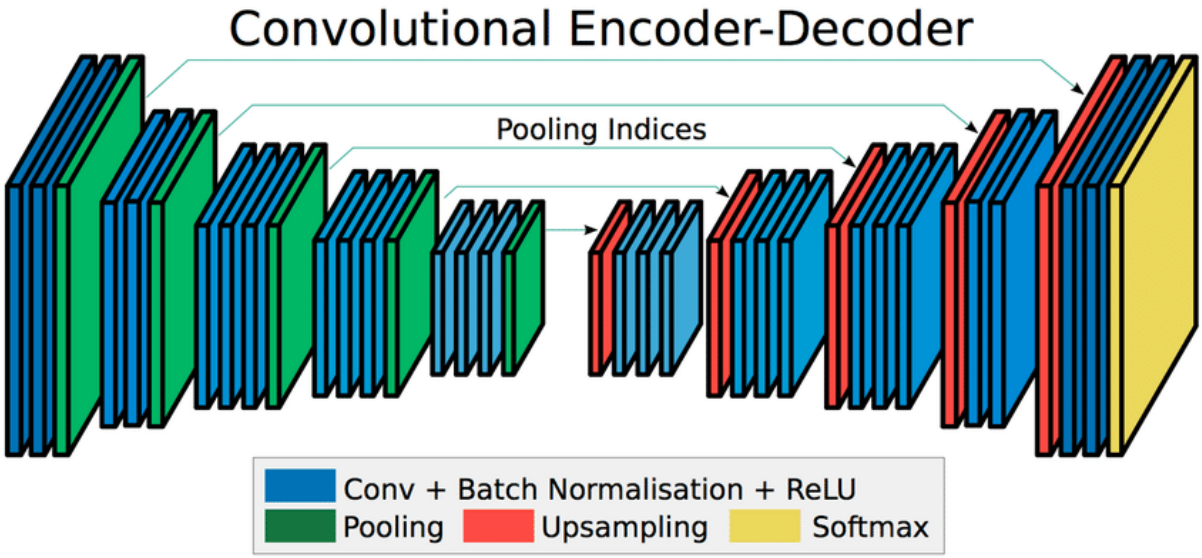


Figure II.7 Architecture SegNet *[113]*

II.3.2.1 Réseau Encodeur

L'encodeur reprend la topologie des 13 couches convolutives du VGG16. Il est constitué d'une séquence de blocs élémentaires comprenant :

- Une couche de convolution (CONV)
- Une fonction d'activation ReLU
- Une couche de pooling maximal (Max-Pooling)

Le mécanisme de pooling réduit progressivement la résolution spatiale des cartes de caractéristiques tout en préservant l'information essentielle. L'innovation majeure de SegNet réside dans la mémorisation des indices de pooling ; les positions exactes des valeurs maximales dans chaque fenêtre de pooling, qui seront ultérieurement utilisées par le décodeur.

II.3.2.2 Réseau Décodeur

Le décodeur est chargé de restaurer la résolution spatiale des cartes de caractéristiques pour générer une segmentation de même taille que l'image d'entrée. Chaque décodeur correspond à un encodeur spécifique dans la hiérarchie.

Le processus de suréchantillonnage, détaillé dans la Figure II.8, utilise les indices de pooling stockés pour réaliser un un-pooling non linéaire. Concrètement :

- Les valeurs sont replacées aux positions mémorisées dans la carte de sortie
- Les positions restantes sont initialisées à zéro
- Une convolution subséquente densifie la carte de caractéristiques creuse

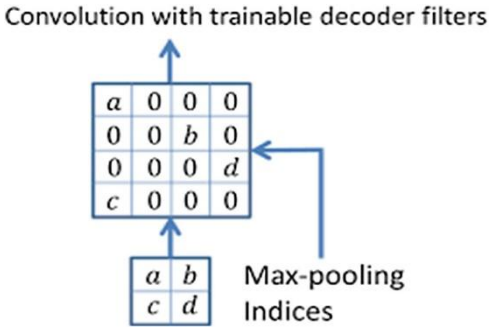


Figure II.8 Décodeur SegNet où a, b, c et d correspondent aux valeurs dans la carte des caractéristiques.

SegNet présente des avantages notables en termes d'efficacité et de performance. Son architecture légère et économique en mémoire la rend particulièrement adaptée aux applications embarquées et aux systèmes avec ressources limitées. La préservation des contours grâce au mécanisme d'indices de pooling permet une segmentation précise des frontières d'objets. Cependant, le modèle présente certaines limitations. La dépendance aux indices de pooling peut limiter la capacité de généralisation sur des images très différentes des données d'entraînement. De plus, SegNet peut montrer des performances inférieures à d'autres architectures plus récentes sur des tâches de segmentation complexes nécessitant une compréhension contextuelle approfondie.

II.3.3 Le Modèle U-Net

Le réseau de neurones U-Net constitue aujourd'hui une architecture fondamentale et omniprésente dans le domaine de la segmentation d'images. Développé par Ronneberger et al. (2015) [102], il a été spécifiquement conçu pour la segmentation d'images biomédicales. Son développement s'inspire directement des Réseaux Entièrement Convolutifs et des modèles encodeur-décodeur, qui permettent d'apprendre des représentations hiérarchiques tout en maintenant la cohérence spatiale des objets segmentés [114].

L'architecture du U-Net, aisément reconnaissable à sa forme symétrique en « U », repose sur une structure d'autoencodeur. Elle se compose de trois composantes principales agissant de

manière complémentaire pour produire une segmentation de type pixel-à-pixel (pixel-wise segmentation).

II.3.3.1 Le Chemin de Contraction (Encodeur)

Le chemin contractant, également appelé encodeur, a pour objectif de capturer le contexte global et les caractéristiques sémantiques de l'image d'entrée. Il est constitué d'une série de blocs convolutifs identiques, chacun comprenant deux convolutions successives de taille 3×3, suivies d'une fonction d'activation ReLU. À la suite de chaque bloc, une opération de max-pooling 2×2 permet de réduire la dimension spatiale tout en augmentant la profondeur des cartes de caractéristiques. Cette réduction progressive de la résolution spatiale favorise l'extraction de représentations sémantiques de haut niveau, essentielles à la compréhension globale du contenu de l'image.

II.3.3.2 La couche bottleneck

La couche bottleneck, également appelée pont (bridge), assure la liaison entre l'encodeur et le décodeur. Elle est constituée de plusieurs convolutions 3×3 suivies d'une fonction d'activation non linéaire telle que ReLU. Cette couche correspond au niveau le plus profond du réseau, où les cartes de caractéristiques présentent la plus faible résolution spatiale mais une grande richesse sémantique. Elle joue un rôle crucial dans la capture des représentations de haut niveau, synthétisant l'information extraite par l'encodeur avant sa reconstruction progressive par le décodeur.

II.3.3.3 Le Chemin d'Expansion (Décodeur)

Le chemin expansif, ou décodeur, vise à reconstruire la résolution spatiale originale à partir des représentations comprimées obtenues dans la phase d'encodage. Chaque étape d'expansion comporte une opération de suréchantillonnage (souvent réalisée à l'aide d'une convolution transposée 2×2), suivie de deux convolutions 3×3 et d'une activation Relu. Cette phase a pour fonction de restituer une carte de segmentation finale de même taille que l'image d'entrée, tout en récupérant les détails perdus lors de la compression.

II.3.3.4 Les Connexions de Saut (Skip Connections)

L'un des éléments les plus distinctifs et novateurs du modèle U-Net réside dans ses connexions de saut. Celles-ci consistent à copier les sorties de chaque niveau du chemin de contraction vers le niveau correspondant du chemin d'expansion. Ces connexions jouent un rôle essentiel dans la restauration de l'information spatiale, souvent perdue au cours des opérations de sous-échantillonnage. En transférant les détails à haute résolution issus des couches peu profondes de l'encodeur, elles enrichissent les représentations de haut niveau du décodeur et garantissent une localisation précise des contours et des structures anatomiques. Ce mécanisme permet à U-Net d'atteindre un équilibre optimal entre précision sémantique et fidélité spatiale, assurant une segmentation fine des régions d'intérêt.

La Figure II.9 ci-dessous illustre la structure globale du réseau U-Net, mettant en évidence les deux chemins symétriques, la couche de goulot d'étranglement et les connexions de saut reliant les blocs correspondants :

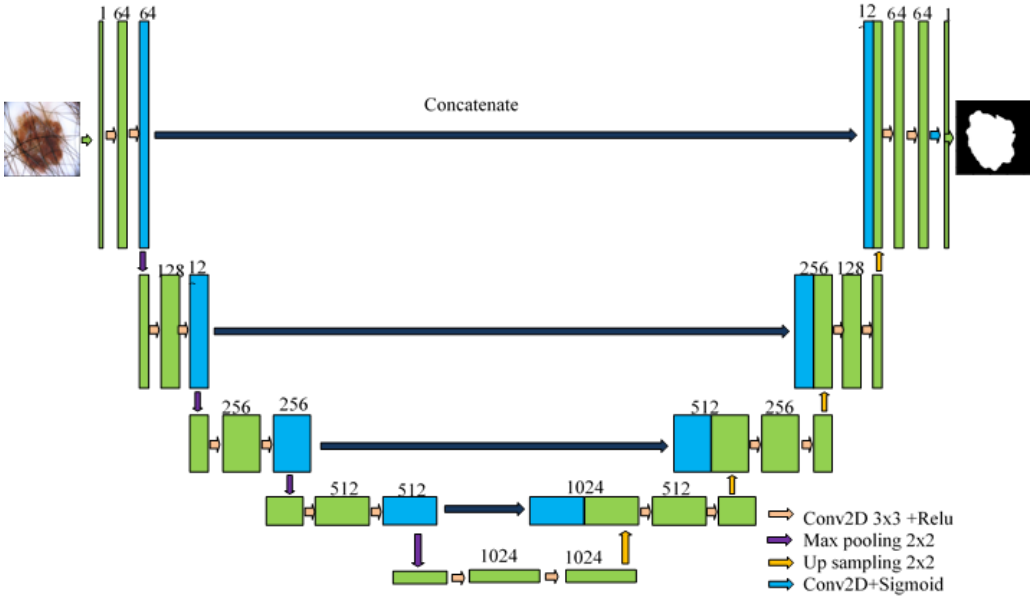


Figure II.9 Architecture générale du modèle U-Net

II.3.3.5 Avantages du Modèle U-Net

Le succès remarquable du U-Net dans la segmentation d'images médicales s'explique par un ensemble d'avantages intrinsèques qui le distinguent des architectures précédentes [114], [115].

1. **Efficacité avec des Données Limitées**

Le modèle U-Net est capable d'obtenir d'excellentes performances même lorsque le nombre d'images annotées est restreint. Cette propriété est renforcée par l'utilisation de techniques d'augmentation de données, notamment les déformations élastiques aléatoires, qui accroissent la diversité du jeu d'entraînement. Ce point est particulièrement crucial en imagerie médicale, où la collecte d'annotations expertes reste coûteuse et chronophage.

2. **Localisation Précise des Structures**

Grâce à la conception des connexions de saut, le réseau exploite simultanément des représentations de bas et de haut niveau, permettant une localisation extrêmement précise des contours, des lésions ou des organes. Cette synergie rend U-Net particulièrement adapté aux tâches nécessitant une délimitation fine, comme la segmentation de tumeurs ou de tissus pathologiques.

3. **Polyvalence et Modularité**

La simplicité et la modularité de sa conception ont fait du U-Net une architecture de référence dans l'ensemble des modalités d'imagerie médicale (IRM, TEP, TDM, échographie, etc.). Sa structure peut aisément être adaptée à d'autres applications telles que la synthèse d'images, la réduction de bruit ou encore la reconstruction d'images médicales.

II.3.3.6 Limitations et Perspectives d'Amélioration

Malgré ses performances exceptionnelles, le modèle U-Net présente certaines limites structurelles qui ont motivé le développement de nombreuses variantes et approches hybrides[114].

1. **Perte de Généralisation**

Les modèles U-Net entraînés de manière supervisée peuvent développer un biais lié au masque de segmentation, ce qui réduit leur capacité à généraliser sur des structures inédites ou mal représentées dans les données d'entraînement.

2. **Sensibilité à la Qualité des Données**

La performance du U-Net dépend fortement de la qualité des images et de la quantité de données annotées disponibles. Des images bruitées, des contrastes faibles ou des artefacts peuvent dégrader significativement les résultats.

3. **Coût Computationnel des Variantes**

Les versions avancées, notamment les U-Net 3D pour la segmentation volumétrique ou les architectures enrichies de blocs attentionnels, présentent un nombre accru de paramètres et un coût de calcul plus important lors de l'entraînement.

Face à ces limitations, plusieurs approches hybrides ont vu le jour, intégrant par exemple des modules d'attention, des mécanismes de pondération spatiale, ou des blocs résiduels issus des réseaux de type ResNet, afin d'améliorer la robustesse, la précision et la capacité de généralisation du modèle.

II.3.4 Réseaux Résiduels Profonds (ResNet)

L'augmentation de la profondeur des réseaux convolutifs a longtemps été perçue comme un moyen d'améliorer leurs performances. Cependant, l'expérience a montré que les réseaux très profonds souffrent de deux limitations majeures : la dégradation des performances malgré l'ajout de nouvelles couches, et l'aggravation du phénomène de disparition du gradient (gradient vanishing) qui rend l'apprentissage difficile, voire impossible, pour les couches les plus profondes. Ces contraintes limitaient fortement la capacité des modèles classiques à exploiter des architectures profondes capables d'extraire des caractéristiques hiérarchiques complexes.

Pour répondre à ces difficultés, He et al. (2015) ont proposé le Residual Network (ResNet), une architecture conçue spécifiquement pour permettre l'entraînement de réseaux comportant plusieurs centaines de couches. L'idée fondatrice consiste à modifier le problème d'apprentissage : au lieu d'apprendre une transformation directe $H(x)$, le réseau apprend une fonction résiduelle définie par :

$$F(x) = H(x) - x. \tag{II.16}$$

La fonction recherchée devient alors :

$$H(x) = F(x) + x. \tag{II.17}$$

Cette reformulation est mise en œuvre grâce à la Skip Connection définie comme mappage d'identité dans la Figure II.10, qui permet à l'entrée du bloc d'être directement ajoutée à sa sortie.

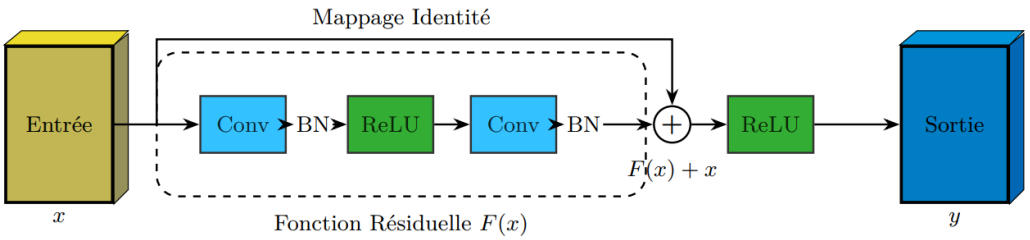


Figure II.10 Bloc résiduel avec mappage identité.

De manière générale, pour une entrée x_l à une couche l, et une sortie x_L à une couche supérieure L, on obtient :

$$x_L = x_l + \sum_{i=l}^{L-1} F\,(x_i, W_i), \tag{II.18}$$

où F représente la fonction résiduelle apprise par les couches intermédiaires du bloc.

Ce mécanisme joue un rôle crucial lors de la rétropropagation. La dérivée de la fonction de perte E par rapport à x_l s'exprime en effet comme :

$$\frac{\partial E}{\partial x_l} = \frac{\partial E}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} F\,(x_i, W_i)\right). \tag{II.19}$$

Le terme « 1 » dans cette expression est fondamental : il garantit qu'un chemin direct de gradient non atténué existe toujours entre les couches profondes et les premières couches du réseau. Ainsi, même si la fonction résiduelle est mal apprise au début de l'entraînement, l'information se propage sans être bloquée, ce qui élimine efficacement le problème du gradient vanishing.

En permettant une transmission directe du signal et en facilitant l'apprentissage de transformations résiduelles, la skip connection constitue le cœur du modèle ResNet. Ce principe a non seulement permis de stabiliser l'optimisation des réseaux très profonds, mais a également

ouvert la voie à des architectures capables de modéliser des relations complexes indispensables à des tâches exigeantes comme la segmentation des lésions cutanées.

II.3.4.1 Blocs Résiduels Fondamentaux

L'architecture ResNet repose sur la construction de blocs résiduels capables de transmettre efficacement l'information tout en apprenant des transformations de plus en plus complexes. Deux formes principales de blocs sont utilisées selon la profondeur du réseau et la complexité des caractéristiques à extraire : le bloc résiduel de base et le bloc bottleneck. Ces blocs partagent le même principe fondamental, à savoir l'apprentissage d'une fonction résiduelle F combinée à une skip connection ajoutant l'entrée directement à la sortie du bloc. Toutefois, leur structure interne diffère afin de répondre à des besoins computationnels et représentatifs distincts.

1. Bloc Résiduel de Base (Basic Block)

Le basic block représente la version la plus simple du bloc résiduel, utilisée notamment dans les versions ResNet-18 et ResNet-34. Il est constitué de deux couches convolutionnelles successives de taille 3×3, chacune suivie d'une normalisation par lot (Batch Normalization) et d'une fonction d'activation non linéaire voir figure 12

L'intérêt principal de ce bloc réside dans sa simplicité et son efficacité. Il capture des motifs locaux pertinents tout en offrant un coût computationnel raisonnable. Cependant, lorsque les réseaux deviennent très profonds, ce type de bloc peut devenir insuffisant en raison de la croissance du nombre de paramètres et de la difficulté à modéliser des représentations à large échelle.

2. Bloc Bottleneck

Le bloc bottleneck est utilisé dans les réseaux les plus profonds tels que ResNet-50, ResNet-101 et ResNet-152.Ce bloc repose sur une architecture en trois étapes, conçue pour réduire puis réaugmenter la dimensionnalité des caractéristiques. Ce bloc est montré par la Figure II.11.

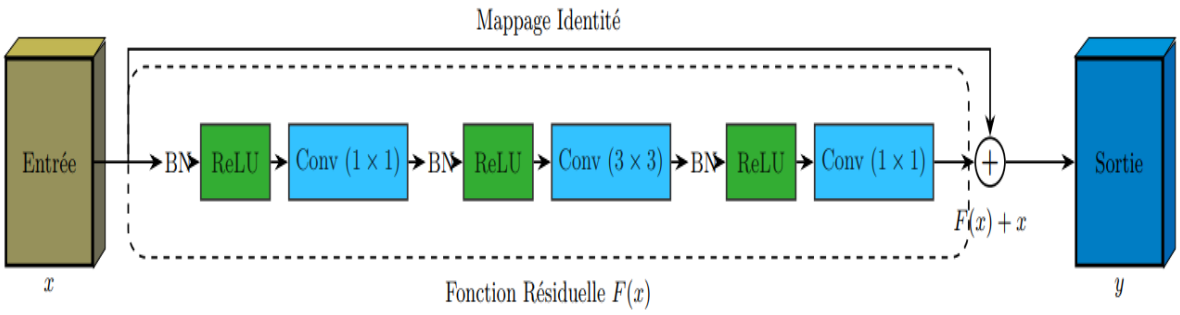


Figure II.11 Bloc bootleneck (ResNet V2)

La première convolution 1×1 réduit la dimension (opération de compression), la convolution 3×3 extrait les caractéristiques espaciales, puis la dernière convolution 1×1 restaure la dimension originale (opération d'expansion).

Les deux blocs répondent à des besoins complémentaires. Le basic block convient aux architectures modérées où la profondeur reste raisonnable. À l'inverse, le bloc bottleneck est spécialement conçu pour optimiser les architectures très profondes en maintenant la stabilité de l'apprentissage et en limitant la complexité computationnelle.

Dans le contexte de la segmentation d'images médicales, et particulièrement des lésions cutanées, l'utilisation du bottleneck dans les encodeurs de type ResNet permet de capter à la fois les détails locaux et les structures globales, tout en assurant une efficacité adaptée aux ressources disponibles.

II.3.4.2 Pré-activation et Post-activation dans les blocs résiduels

L'organisation interne des opérations dans un bloc résiduel influence considérablement la stabilité de l'apprentissage et la qualité de la propagation du gradient. Deux variantes principales ont été proposées selon l'ordre des opérations de Batch Normalization (BN) et d'activation (ReLU) par rapport à l'étape d'addition : la version Post-Activation (ResNet V1) et la version Pré-Activation (ResNet V2).

1. Post-Activation (ResNet V1)

Dans la configuration Post-Activation (ResNet V1), l'addition entre la branche d'identité et la branche résiduelle est suivie d'une activation ReLU. La séquence typique des opérations est illustré par la Figure II.12

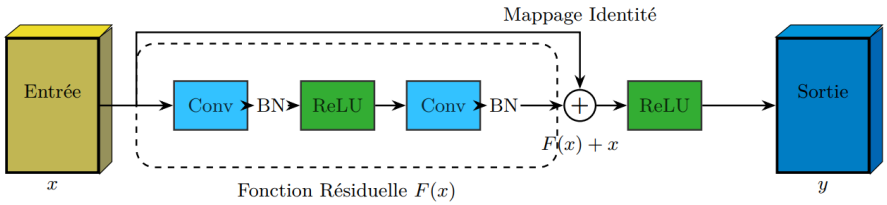


Figure II.12 Post-Activation(ResNet V1)

Cette architecture introduit une non-linéarité après la fusion des chemins, ce qui peut toutefois avoir pour effet d'annuler certaines activations négatives potentiellement informatives, réduisant ainsi la richesse du signal transmis.

2. Pré-Activation (ResNet V2)

À l'inverse, la version Pré-Activation (ResNet V2) réorganise les opérations en plaçant la normalisation et l'activation avant chaque couche convolutionnelle, tandis que l'addition devient la dernière opération du bloc montré sur la Figure III.12.

Cette approche garantit que le chemin d'identité demeure entièrement linéaire et non perturbé par des fonctions non linéaires, ce qui facilite une propagation directe et stable de l'information à travers le réseau. Le signal d'entrée x peut ainsi être transmis sans distorsion jusqu'aux couches profondes, améliorant la convergence et la performance des réseaux très profonds.

L'architecture ResNet de base a donné naissance à plusieurs variantes adaptées à différentes applications, le Tableau II.1 recapitule ces architectures :

Tableau II.1 Variantes de l'architecture ResNet.

Nom de couche	Taille de sortie	Variante de modèle				
		18-layer	**34-layer**	**50-layer**	**101-layer**	**152-layer**
conv1	112×112	$7 \times 7, 64, \text{stride } 2$				
conv2.x	56 ×56	$3 \times 3 \text{ max pool, stride } 2$				
		$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$
conv3.x	28 ×28	$\begin{bmatrix} 3 \times 3,128 \\ 3 \times 3,128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,128 \\ 3 \times 3,128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 8$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 8$
conv4.x	14 ×14	$\begin{bmatrix} 3 \times 3,256 \\ 3 \times 3,256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,256 \\ 3 \times 3,256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 36$
conv5.x	7 ×7	$\begin{bmatrix} 3 \times 3,512 \\ 3 \times 3,512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3,512 \\ 3 \times 3,512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$
	1 ×1	$\text{average pool, } 1000 - \text{d fc, softmax}$				

En définitive, les architectures ResNet ont profondément transformé l'apprentissage profond en permettant l'entraînement stable de réseaux très profonds grâce aux connexions de saut et aux blocs résiduels. Leur capacité à apprendre des représentations hiérarchiques riches et discriminantes en fait un choix particulièrement pertinent pour les tâches complexes d'analyse d'images médicales. Toutefois, ces réseaux présentent certaines limites, telles que le coût computationnel élevé dans leurs versions les plus profondes et une consommation mémoire importante. Malgré ces contraintes, ResNet est souvent sollicité comme backbone dans de nombreuses architectures de segmentation modernes, notamment les variantes d'U-Net. Cette intégration permet d'améliorer de manière significative la qualité des cartes de caractéristiques, d'enrichir la représentation multi-échelle et, en conséquence, d'optimiser la précision de la segmentation des lésions cutanées et d'autres structures anatomiques.

II.4 État de l'Art des Méthodes FCN pour la Segmentation des Lésions Cutanées

Les Réseaux Entièrement Convolutifs (Fully Convolutional Networks, FCN) se sont imposés comme une solution efficace pour cette tâche, grâce à leur capacité à effectuer des prédictions pixel à pixel tout en préservant la structure spatiale des images [105]. Dans le contexte dermoscopique, les FCN doivent faire face à des défis tels que le faible contraste entre lésion et peau saine, les contours flous, l'hétérogénéité texturale et la présence d'artefacts. Une étude comparative de Kaymak et al. [116] sur ISIC 2017 a montré que des architectures comme

FCN-8s, intégrant des skips connections, offrent un équilibre satisfaisant entre précision des contours et complexité computationnelle.

II.4.1 FCN Profond avec Fonction de Perte Jaccard

Les premières approches, telles que celle de Yuan et al. [117], ont introduit un FCN profond (19 couches) utilisant une fonction de perte basée sur la distance de Jaccard, traitant directement le déséquilibre entre pixels de lésion et de peau saine. Leur méthode a obtenu un coefficient de Dice de 91,5 % sur PH² et un indice de Jaccard de 84,7 % sur ISBI 2016, avec un temps d'inférence rapide. Cependant, la réduction de résolution des cartes de caractéristiques pouvait entraîner une perte de détails fins, et la binarisation finale nécessitait un réglage délicat.

II.4.1.1 FCN Multi-Étapes

Pour améliorer la précision des contours, Bi et al. [118] ont proposé un FCN multi-étapes (mFCN) avec intégration parallèle. Les premiers étages localisent la lésion de manière grossière, tandis que les étages suivants affinent progressivement les contours, et une fusion parallèle assure une meilleure robustesse. Cette approche a atteint un coefficient de Dice de 91,18 % sur ISBI 2016 et a montré une sensibilité élevée pour les mélanomes (94,34 %). Son principal inconvénient réside dans la complexité et le coût computationnel liés à l'entraînement en cascade.

II.4.1.2 Réseau Haute Résolution avec Attention Duale

Pour conserver les détails fins tout en filtrant le bruit, Xie et al. [119] ont développé un réseau haute résolution intégrant un mécanisme d'attention dual (spatial et canaux). Leur méthode a surpassé FCN-8s et U-Net sur ISBI 2017, avec un indice de Jaccard de 78,3 %, et a montré des performances compétitives sur ISBI 2016 et PH². Le coût mémoire et la capacité réduite due au nombre de canaux limités représentent néanmoins des contraintes.

En synthèse, les travaux sur les FCN pour la segmentation des lésions cutanées montrent une évolution progressive : des FCN de base performants aux architectures multi-étapes et réseaux haute résolution avec attention, visant à préserver les détails et améliorer la robustesse. L'apprentissage hiérarchique des caractéristiques, le transfer learning et l'usage de métriques adaptées (Dice, IoU) restent des facteurs clés de succès. Les perspectives futures incluent l'optimisation des architectures via Neural Architecture Search, l'intégration de l'apprentissage semi-supervisé pour réduire la dépendance aux annotations et le développement de modèles robustes et explicables pour une adoption clinique généralisée.

II.4.2 État de l'Art des Méthodes de Segmentation Basées sur l'Architecture SegNet pour les Lésions Cutanées

Parmi les architectures de deep learning dédiées à la segmentation sémantique, SegNet s'est imposée comme une référence majeure en raison de son architecture encodeur-décodeur efficace et économiquement viable en termes de paramètres. L'analyse des méthodes exploitant SegNet et ses dérivées permet de mettre en lumière les innovations architecturales, les stratégies

d'optimisation et les performances obtenues sur les jeux de données de référence, tout en soulignant les défis persistants dans le domaine.

II.4.2.1 Applications Pionnières et Adaptations Initiales

Les premières adaptations de SegNet pour la segmentation des lésions cutanées ont été réalisées par Brahmbhatt et Rajan (2019) [120]. Ces travaux ont consisté à remplacer la fonction de perte de categorical cross-entropy par une binary cross-entropy, mieux adaptée à la nature binaire de la segmentation des lésions. Le réseau initial comprenait 64 couches avec 25 couches de batch normalization et était entraîné avec l'optimiseur SGD. L'augmentation des données par rotations et retournement horizontal a permis de multiplier l'ensemble de données PH² initial, garantissant ainsi une meilleure généralisation du modèle. Les résultats obtenus ont montré un indice de Jaccard de 93,61 % sur l'ensemble de test, une précision de 93,96 % et un rappel de 92,13 %. Malgré ces performances, les prédictions présentaient des contours légèrement flous nécessitant un post-traitement par seuillage, et l'efficacité sur des jeux de données plus complexes restait non évaluée.

II.4.2.2 Optimisation Systématique des Hyperparamètres

Pour améliorer les performances, Sahin et al. (2021) [121] ont exploré l'optimisation bayésienne des hyperparamètres de SegNet, ciblant cinq paramètres critiques : profondeur de section, taux d'apprentissage initial, facteur de réduction du learning rate, régularisation L2 et momentum. Six configurations expérimentales ont été évaluées en utilisant des fonctions d'acquisition avancées, et les résultats ont montré des gains substantiels : sur le dataset ISBI2016, l'indice de Jaccard a atteint 84,9 %, soit une amélioration de 16 % par rapport à SegNet standard, tandis que sur ISBI2017, l'amélioration était de 7 %. Cependant, l'entraînement nécessitait un temps prohibitif de 816 heures, et les performances demeuraient sensibles à la complexité du jeu de données, soulignant la nécessité de méthodes d'optimisation adaptées aux contraintes computationnelles.

II.4.2.3 Architectures Avancées et Innovations Structurelles

L'évolution des architectures SegNet a conduit à des modèles plus profonds et interprétables. Narayanan et al. (2024) [122] ont proposé IARS SegNet, intégrant des skip connections inspirées de U-Net pour transférer les caractéristiques multi-échelles de l'encodeur vers le décodeur, des convolutions résiduelles pour atténuer le vanishing gradient, et un mécanisme d'attention global permettant de supprimer progressivement les réponses non pertinentes dans les régions d'arrière-plan. Cette approche a permis d'améliorer significativement la précision des contours et la réduction des faux positifs, avec un Dice Score de 97,12 % et un IoU de 92,33 % sur PH². La visualisation des cartes d'attention a également renforcé l'interprétabilité clinique du modèle.

En optimisant SegNet, Chamseddine et al. (2022) [123] ont amélioré les techniques d'entraînement de SegNet en remplaçant SGD par Adam et en prolongeant la durée d'entraînement à 200 epochs, tout en évaluant le modèle sur plusieurs jeux de données, notamment PH² et ISIC2018. Ces ajustements ont conduit à des IoU de 95,72 % sur PH² et de

95,06 % sur ISIC2018, et une précision d'entraînement de 99,56 %. Les performances se sont maintenues sur des images présentant des artefacts simples tels que poils ou bulles de gel, bien que des difficultés persistent avec des marques complexes ou des chartes colorimétriques, nécessitant un réentraînement spécifique pour de nouveaux types d'artefacts.

II.4.2.4 Analyse Comparative et Défis Persistants

L'analyse comparative des méthodes SegNet révèle une amélioration progressive des performances. Sur le jeu de données PH², Brahmbhatt et al. (2019) ont atteint un IoU de 93,61 %, Chamseddine et al. (2022) un IoU de 95,72 % et Narayanan et al. (2024) un Dice de 97,12 % avec un IoU de 92,33 %. Les facteurs d'amélioration incluent l'optimisation des hyperparamètres, l'adaptation des fonctions de perte, l'introduction de mécanismes d'attention et l'utilisation d'optimiseurs adaptatifs comme Adam. Toutefois, plusieurs défis subsistent, notamment le flou des contours nécessitant un seuillage, la sensibilité aux artefacts et aux variations d'acquisition, la complexité computationnelle des architectures avancées et la généralisation limitée sur des jeux de données complexes.

En conclusion, SegNet et ses variantes ont démontré leur efficacité pour la segmentation des lésions cutanées, offrant un compromis favorable entre performance et complexité. L'évolution des approches suit trois axes principaux : l'optimisation des hyperparamètres, l'enrichissement architectural et l'amélioration des techniques d'entraînement. Par sa simplicité conceptuelle et son efficacité démontrée, SegNet constitue une base solide pour le développement de solutions de segmentation applicables en contexte clinique. Toutefois, plusieurs défis subsistent, notamment le flou des contours nécessitant un seuillage, la sensibilité aux artefacts et aux variations d'acquisition, la complexité computationnelle des architectures avancées et la généralisation limitée sur des jeux de données complexes.

II.4.3 État de l'Art des Méthodes de Segmentation les lésions cutanées basées sur U-Net

Les architectures de réseaux de neurones profonds, U-Net, s'est imposée comme la référence pour la segmentation automatique des images médicales et en particulier les lésions cutanées. La structure encodeur–décodeur symétrique et de ses connexions de saut permettant une fusion efficace entre les informations contextuelles et les détails locaux. Cependant, un défi majeur dans ce domaine demeure le déséquilibre de classes, où les régions pathologiques occupent une fraction très réduite de l'image par rapport au fond sain. Ce déséquilibre induit un biais de prédiction vers la classe majoritaire, rendant difficile la détection précise des petites lésions. De nombreuses variantes et améliorations de U-Net ont ainsi été proposées pour renforcer la robustesse du modèle face à ces contraintes.

II.4.3.1 Améliorations architecturales de la U-Net classique

Anand et al. [124] ont proposé une U-Net modifiée utilisant des cartes de caractéristiques rectangulaires plutôt que carrées, mieux adaptées aux proportions des images dermoscopiques. Leur approche, évaluée sur le dataset PH2 augmenté, a montré une amélioration de la précision grâce à l'optimiseur Adam (batch size égal 8, 75 époques), mais reste sensible aux artefacts (poils, reflets) et n'intègre pas de traitement explicite du déséquilibre de classes.

Zhou et al. [125] ont introduit U-Net++, une version imbriquée intégrant des blocs de convolution denses entre l'encodeur et le décodeur, favorisant une fusion multi-échelle progressive. Sur PH^2, le modèle atteint un IoU de 84,35 % et un Dice de 90,25 %, surpassant U-Net standard, mais au prix d'une complexité computationnelle accrue.

Huang et al. [126] ont proposé U-Net 3+, qui introduit des connexions pleine échelle et une supervision profonde afin de capturer des détails fins à toutes les résolutions. Testé sur ISIC 2018, ce modèle améliore la précision des frontières, mais exige des ressources considérables.

Ruan et al. [127] ont présenté EGE-UNet, un modèle léger intégrant deux modules innovants : GHPA (Group multi-axis Hadamard Product Attention) et GAB (Group Aggregation Bridge). Avec seulement 53 000 paramètres, EGE-UNet atteint un Dice de 89,46 % sur ISIC 2018, démontrant son efficacité pour des applications embarquées, bien que limité à la segmentation cutanée.

II.4.3.2 Intégration de backbones et réseaux hybrides

Agthobirrobbany et al. [128] ont étudié l'intégration de backbones pré-entraînés (VGG16, ResNet50, MobileNetV2) dans U-Net. Sur ISIC 2018, U-Net+VGG16 a obtenu une précision de 0,93, un F1 score de 0,78 et un IoU de 0,68, illustrant l'intérêt de la hiérarchie de caractéristiques. Cependant, ces modèles nécessitent des ressources élevées et restent sensibles au surapprentissage.

Manivannan et Venkateswaran [129] ont utilisé un backbone ResNet101 combiné à une forte augmentation de données, atteignant un score de Jaccard de 96,28 % (train) et 84,20 % (test). Malgré de bonnes performances, la complexité du modèle limite son déploiement en contexte clinique.

Karimi et al. [130] ont développé DEU-Net, un double encodeur combinant EfficientNet-B6 et MaxViT, avec une fonction de perte hybride Dice + Cross-Entropy (ratio 0.4/0.6) conçue pour contrer le déséquilibre pixelique. Sur ISIC 2016 et PH^2, ils obtiennent un Dice supérieur à 95 %, prouvant l'efficacité des pertes adaptatives.

II.4.3.3 U-Net enrichies par l'attention et la fusion contextuelle

Alahmadi [131] a proposé MSAU-Net (Multi-Scale Attention U-Net), intégrant un mécanisme d'attention multi-échelle au niveau du goulot d'étranglement via un ConvLSTM bidirectionnel, permettant d'extraire des caractéristiques discriminantes à plusieurs échelles. Sur ISIC 2018, le Dice atteint 94,90 %. Cette approche améliore la détection des structures fines, mais augmente fortement le coût computationnel.

Wu et al. [132] ont présenté FAT-Net, un modèle hybride CNN-Transformer combinant extraction locale et attention globale. Bien que non spécifiquement conçu pour le déséquilibre de classes, il améliore la détection des petites lésions grâce à une contextualisation étendue (Dice 85 % sur ISIC 2017).

Sikha et al. [133] ont introduit Meta-UNet, intégrant des métadonnées cliniques dans le pipeline de segmentation. Cette approche bayésienne utilisant Monte Carlo Dropout quantifie

explicitement l'incertitude et réduit l'ambiguïté dans les régions mal définies (IoU 84,6 % sur PH^2 et Dice 93,1 % sur ISIC 2018).

II.4.3.4 Architectures légères et alternatives inspirées de U-Net

Kaur et al. [134] ont développé DilatedSkinNet, un réseau léger basé sur des convolutions atrous (dilatées), évitant le pooling tout en maintenant la résolution spatiale. Avec seulement 3,27 k paramètres, il atteint un Dice de 94,20 % sur ISIC 2018, surpassant même certaines architectures lourdes comme DeepLabV3+. Sa fonction de perte à entropie croisée pondérée contre efficacement le déséquilibre de classes, bien que sa précision diminue en présence d'artefacts.

Anand et al. [135] ont combiné U-Net avec un CNN de classification dans une approche fusionnée sur HAM10000, atteignant 97,96 % de précision globale, illustrant le potentiel de pipelines multi-tâches reliant segmentation et diagnostic, malgré une complexité accrue.

Le tableau II.2 présente une synthèse des performances obtenues pour les différentes variantes du modèle U-Net issues de l'état de l'art.

Tableau II.2 Synthèse comparative de segmentation de lésion cutanée par les modèles U-NET

Méthode	U-Net[102]	U-Net++ [135]	U-Net 3+ [126]	MSAU-Net [131]	FAT-Net[132]	DEU-Net [130]	Meta-UNet[133]	DilatedSknNet [134]	EGE-UNet [127]
Architecture principale	Encodeur-décodeur symétrique	Connexions denses imbriquées	Connexions pleine échelle + supervision profonde	Attention multi-échelle + ConvLSTM	Double encodeur (CNN + Transformer	Dual Encoder (EfficientNet +	U-Net + métadonnées	Convolutions atrous	GHPA + GAB (léger)
Jeu(x) de données	ISIC, PH2	PH2	ISIC 2018	ISIC 2018	ISIC 2017	ISIC 2016–PH2	PH2, ISIC 2018	ISIC 2018	ISIC 2018
IoU (%)	~80	84,3	84	95,6	76,5	87–91	84–88	89	80,9
Dice (%)	~88	90,2	90	94,9	85,0	90–95	90–93	94	89,5
Traitement du déséquilibre	Aucun	Aucun	Aucun	Attention multi-échelle	Contextualisation globale	Dice/CE hybride + ITTA	Métadonnées + incertitude	Entropie croisée pondérée	Aucun
Points forts	Simplicité, stabilité	Fusion multi-échelle	Précision aux frontières	Haute précision	Meilleure robustesse contextuelle	Optimisé déséquilibre	Réduction incertitude	Léger, rapide	Très léger
Limitations	Peu robuste aux petites lésions	Complexité accrue	Très coûteux	Temps d'inférence long	Sensible aux artefacts	Consommation mémoire	Dépend métadonnées	Sensible aux artefacts	Limité aux images cutanées

L'évolution des architectures U-Net pour la segmentation des lésions cutanées témoigne d'une progression continue vers une meilleure adaptation aux déséquilibres structurels et spectro-spatiaux des images dermoscopiques. Les tendances récentes se concentrent sur :

- La fusion multi-échelle (U-Net++, U-Net 3+),
- L'intégration de backbones profonds et attentionnels (ResNet, EfficientNet, Transformer),
- L'emploi de fonctions de perte hybrides pondérées.

Malgré des améliorations notables en précision et robustesse, le déséquilibre de classes reste une limite fondamentale. Les futurs travaux devraient intégrer des stratégies pour gérer ce problème et par la suite garantir une segmentation plus fiable des petites lésions et une meilleure généralisation clinique.

II.5 Conclusion

Les réseaux de neurones convolutifs (CNN) ont révolutionné la segmentation des images dermoscopiques, avec des architectures majeures comme FCN, SegNet et U-Net surpassant nettement les méthodes classiques. Cette supériorité provient de leur capacité à apprendre automatiquement des représentations hiérarchiques pertinentes, ce qui améliore la détection et la délimitation des lésions cutanées. Néanmoins, plusieurs contraintes freinent encore leur adoption clinique.

Le défi principal réside dans le besoin impératif d'un volume important d'images annotées. Or, l'annotation en dermoscopie exige une expertise coûteuse, conduisant à des jeux de données restreints et augmentant ainsi le risque de surapprentissage et de mauvaise généralisation. À cela s'ajoute le problème persistant du déséquilibre marqué entre les pixels de lésion et les pixels de fond. Ce déséquilibre biaise le modèle en faveur de la classe majoritaire, ce qui se traduit par une sous-segmentation des zones pathologiques et dégrade les métriques de similarité comme le coefficient de Dice. Les fonctions de perte classiques s'avèrent souvent insuffisantes pour corriger cet effet.

Même si de nombreuses variantes du U-Net (U-Net++, Attention U-Net) ont été proposées pour améliorer l'extraction contextuelle, elles n'apportent pas toujours une réponse suffisante aux défis de la rareté des données et du déséquilibre des classes, pouvant même accroître la complexité sans garantir une meilleure robustesse. Face à ces limites persistantes, la recherche s'oriente vers l'emploi de fonctions de perte spécialisées (Dice, Tversky, Focal Loss) et vers des stratégies d'augmentation de données capables de reproduire la variabilité inter-patient.

Afin de dépasser ces contraintes, nous avons choisi de développer une architecture hybride. Celle-ci combine la puissance d'extraction de caractéristiques profondes d'un encodeur ResNet50V2 avec la précision spatiale d'un décodeur U-Net pour optimiser la détection des contours lésionnels. De plus, pour atténuer l'impact du déséquilibre de classes, nous adoptons une fonction de perte, qui valorise les échantillons difficiles et les régions minoritaires. L'ensemble est complété par une stratégie d'augmentation de données spécifiquement adaptée aux images dermoscopiques pour garantir la robustesse et la généralisation sur des jeux de données limités. Le chapitre suivant détaillera la mise en œuvre de cette méthodologie.

Chapitre III: Segmentation des lésions cutanées : une approche hybride RESNET50V2/U-NET

III.1 Introduction

La segmentation automatique des images médicales constitue un élément clé du diagnostic assisté par ordinateur, en particulier pour l'analyse des lésions cutanées où la précision de la Région d'Intérêt (ROI) influence directement l'évaluation clinique. Si les méthodes traditionnelles (seuillage, contours, régions) donnent de bons résultats sur des images simples, elles restent insuffisantes en contexte réel en raison de la variabilité anatomique, des artefacts, des variations de contraste et de la petitesse des lésions, rendant l'extraction fiable de la ROI difficile.

Les architectures profondes telles que FCN, SegNet et U-Net ont significativement amélioré la segmentation en exploitant des représentations hiérarchiques et un apprentissage end-to-end. Toutefois, elles demeurent limitées par la perte d'informations spatiales fines, le déséquilibre de classes, la sensibilité aux artefacts et le faible volume de données annotées, entraînant des segmentations incomplètes ou un risque d'overfitting.

Pour surmonter ces limites, nous proposons une approche hybride combinant un encodeur ResNet50V2 et un décodeur U-Net, couplée à une fonction de perte adaptée. Cette architecture associe la richesse des représentations profondes à la capacité multi-échelle du U-Net, assurant une reconstruction plus précise de la ROI, même en présence d'artefacts ou de classes minoritaires. Les premiers résultats de cette méthode, publiés dans [27], ont confirmé son potentiel.

Afin d'améliorer la généralisation, notre approche intègre également des stratégies d'augmentation de données et de régularisation (dropout, early stopping), essentielles pour limiter le surapprentissage et stabiliser l'apprentissage sur de petits jeux d'images.

Ce chapitre vise ainsi à démontrer la supériorité de la méthode proposée par une analyse quantitative et qualitative approfondie. Il présente successivement l'architecture hybride, la fonction de perte utilisée, la stratégie d'augmentation, puis les résultats expérimentaux qui valident sa capacité à extraire fidèlement la région d'intérêt.

III.2 Augmentation des données

L'augmentation des données constitue une étape déterminante dans le pipeline proposé pour la segmentation des lésions cutanées. Dans le domaine de la dermoscopie, la variabilité intra-et inter-patients, la diversité des dispositifs d'acquisition, la présence fréquente d'artefacts (poils, ombres, bulles, irrégularités d'éclairage), ainsi que le déséquilibre proportionnel entre pixels de peau saine et pixels de lésion rendent l'apprentissage particulièrement difficile pour un modèle de segmentation. Sans une augmentation adéquate, le réseau tend à mémoriser les configurations rencontrées, à surestimer la classe majoritaire et à perdre en capacité de généralisation.

III.2.1 Transformations géométriques

Les transformations géométriques permettent d'augmenter la diversité spatiale des lésions sans en altérer la structure morphologique. Elles améliorent la capacité du réseau à reconnaître une lésion indépendamment de sa position, de son orientation ou du cadrage initial.

a) Flip horizontal et vertical

Le flip horizontal est défini par :

$$I_{\text{flipH}}(x, y) = I(W - x - 1, y) \tag{III.1}$$

et le flip vertical par :

$$I_{\text{flipV}}(x, y) = I(x, H - y - 1) \tag{III.2}$$

b) Rotations

La rotation d'un angle θ autour du centre (x_c, y_c) est donnée par :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x - x_c \\ y - y_c \end{pmatrix} + \begin{pmatrix} x_c \\ y_c \end{pmatrix} \tag{III.3}$$

c) Translations

Une translation est définie par :

$$I_{\text{trans}}(x, y) = I(x - t_x, y - t_y) \tag{III.4}$$

Ces transformations géométriques permettent d'éviter que le réseau ne se contente de repérer des motifs spécifiques d'arrière-plan ou des positions récurrentes dans les images originales. Elles renforcent sa capacité à se concentrer sur les caractéristiques intrinsèques de la lésion, ce qui améliore la cohérence du masque prédit et la segmentation des bords.

III.2.2 Transformations photométriques

Dans la dermoscopie, les variations d'éclairage, de couleur et de contraste sont très fréquentes d'une acquisition à une autre. Les transformations photométriques visent à rendre le modèle invariant à ces fluctuations.

a) Variation de la luminosité

$$I_{\text{lum}}(x, y) = \alpha I(x, y) \tag{III.5}$$

avec $\alpha \in [0.8, 1.2]$.

b) Variation du contraste

$$I_{\text{cont}}(x,y) = \beta(I(x,y) - \mu) + \mu \tag{III.6}$$

avec $\beta \in [0.8,1.2]$.

c) Jittering des couleurs (RGB)

$$I_{c'}(x,y) = I_c(x,y) + \epsilon_c \tag{III.7}$$

où $\epsilon_c \sim \mathcal{U}(-\delta,\delta)$.

.

d) Bruit gaussien

$$I_{\text{gauss}}(x,y) = I(x,y) + n(x,y), \quad n(x,y) \sim \mathcal{N}(0,\sigma^2) \tag{III.8}$$

Cette augmentation simule les artefacts de granularité ou les défauts du capteur, ce qui améliore la segmentation dans des conditions réelles d'imagerie. La Figure IV.1 montre les résultats d'application de ces transformations

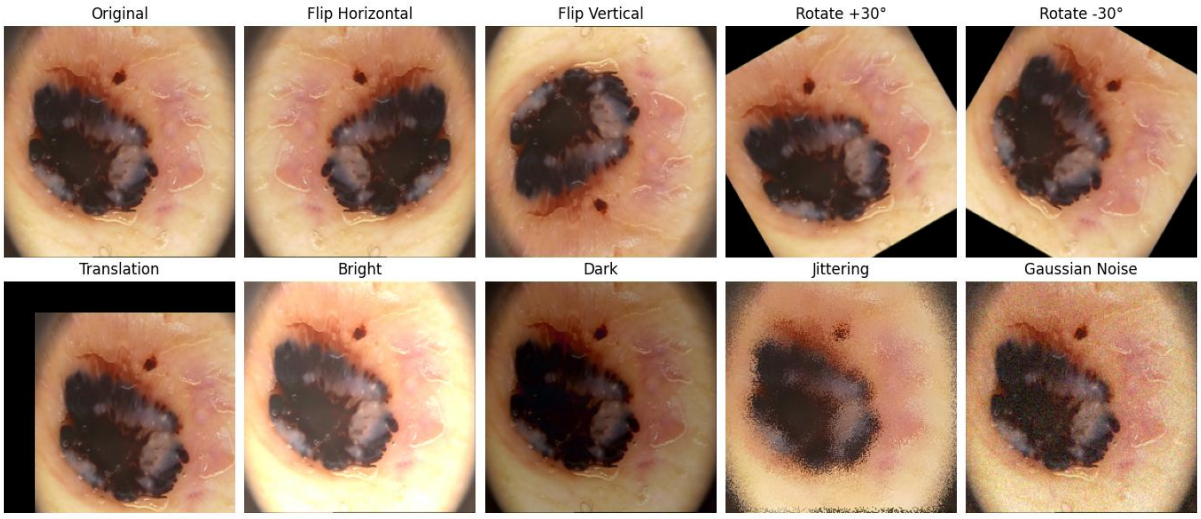


Figure III.1 Résultats d'application géométriques et radiométriques

III.2.3 Cohérence entre images et masques

Les masques $M(x,y)$, binaires et structurellement sensibles, doivent être transformés exactement comme l'image :

$$M'(x,y) = M(T^{-1}(x,y)) \tag{III.9}$$

où T désigne une transformation géométrique.

Les transformations photométriques ne sont jamais appliquées aux masques. Cela garantit la conformité parfaite entre l'image augmentée et sa vérité terrain.

L'impact de ces augmentations dans ce travail est particulièrement significatif pour plusieurs raisons :

1. Réduction de l'overfitting

Les images dermoscopiques disponibles sont limitées. Sans augmentation, le modèle surapprend rapidement des motifs spécifiques au dataset. L'augmentation a considérablement réduit cet effet, ce qui a permis d'obtenir une courbe de validation plus stable et une différence train–val plus faible.

2. Amélioration des métriques de segmentation

Les transformations contribuent directement à des gains mesurés sur le test :

- Meilleure segmentation des petites lésions,
- Meilleurs scores Dice et Jaccard,
- Baisse des faux négatifs grâce à une meilleure reconnaissance des bords.

3. Résistance accrue aux artefacts

L'introduction de variations photométriques et de bruit a rendu le modèle plus robuste à la présence des artefacts, ce qui se traduit par :

- Une meilleure détection des contours même en présence d'obstacles visuels,
- Des masques plus cohérents lors des tests avec artefacts simulés ou réels.

4. Atténuation du problème de déséquilibre de classes

La classe « peau saine » domine largement. L'augmentation génère davantage de cas où la lésion occupe différentes proportions. Cela aide le réseau à mieux apprendre la structure de la classe minoritaire et améliore la sensibilité.

L'augmentation des données n'est pas une simple étape optionnelle, mais un pilier fondamental du pipeline. Elle améliore simultanément l'apprentissage, la robustesse, la généralisation et l'équilibre entre classes. Les tests réalisés dans ce travail montrent clairement que le modèle proposé atteint de meilleurs résultats uniquement lorsque cette augmentation est intégrée, ce qui justifie pleinement son utilisation dans un contexte de segmentation des lésions cutanées.

III.3 Présentation générale de la méthode proposée

III.3.1 Schéma global de la méthode : description détaillée

Le schéma global présenté ci-dessous (Figure IV.2) formalise l'ensemble du pipeline de segmentation proposé, depuis les images brutes annotées jusqu'à l'obtention de la Région d'Intérêt (ROI) et des métriques d'évaluation. Le pipeline se compose de deux phases principales : la phase d'entraînement et la phase d'inférence / test. Chaque étape est décrite en

détail afin que le jury comprenne l'enchaînement logique, les décisions méthodologiques et les choix d'implémentation.

III.3.2 Jeu de données (Dataset)

Le pipeline s'appuie sur un jeu de données d'images dermoscopiques annotées et leurs masques binaires correspondants indiquant la présence de la lésion, c'est-à-dire la ROI. Les images et masques sont préalablement organisés en partitions distinctes pour l'entraînement, la validation et le test (par exemple 80/10/10). Lors de la constitution des partitions, l'isolation des images de test permet d'évaluer la capacité de généralisation du modèle sur des cas non vus pendant l'entraînement.

III.3.3 Augmentation des données (phase d'entraînement)

Pour limiter le sur-apprentissage et accroître la robustesse du modèle face aux variations cliniques réelles, une stratégie d'augmentation est appliquée pendant l'entraînement. Les opérations utilisées comprennent des transformations géométriques (rotations, translations, flips horizontaux et verticaux, zooms), des transformations photométriques (variations de contraste et de luminosité, ajout de bruits réalistes) ainsi que des opérations ciblées comme les petits recadrages aléatoires ou le cutout. Les paramètres d'augmentation sont choisis de manière à préserver la structure morphologique des lésions tout en augmentant la diversité des exemples présentés au réseau.

III.3.4 Prétraitement (phase d'entraînement et d'inférence)

Toutes les images (train et test) passent par un prétraitement standardisé comprenant le redimensionnement à la taille d'entrée du réseau (par exemple 224×224 ou 256×256), une éventuelle conversion d'espace colorimétrique et une normalisation par canal compatible avec le pré-entraînement de ResNet50V2. Les masques sont redimensionnés avec une interpolation de type *nearest neighbor* afin de préserver leur nature binaire. Une correction d'éclairage simple peut également être appliquée pour réduire les variations d'illumination.

III.3.5 Construction du modèle : Encodeur ResNet50V2 et Décodeur U-Net

L'architecture retenue associe la puissance des représentations hiérarchiques issues d'un encodeur ResNet50V2 pré-entraîné et la capacité de reconstruction multi-échelle d'un décodeur de type U-Net. Cette architecture est composée de :

- ✓ **Encodeur (ResNet50V2)** : extraction de cartes de caractéristiques hiérarchiques grâce à des blocs résiduels profonds. Les sorties intermédiaires à différentes résolutions sont conservées afin d'être réinjectées dans le décodeur via les *skip connections*.
- ✓ **Bottleneck (pont)** : couches convolutives profondes permettant de consolider l'information contextuelle globale avant la phase de décodage.
- ✓ **Décodeur (U-Net)** : upsampling progressif (transposed convolutions ou upsampling suivi de convolutions) accompagné de concaténations avec les cartes de caractéristiques correspondantes issues de l'encodeur. Ce mécanisme permet de récupérer les détails spatiaux fins et d'améliorer la reconstruction de la ROI.

✓ **Couche de sortie** : convolution 1×1 suivie d'une activation sigmoïde produisant une carte de probabilités par pixel pour la segmentation binaire.

III.3.6 Fonction de perte personnalisée et optimisation

La fonction de perte a été conçue pour traiter simultanément la calibration pixel-wise et l'optimisation de l'overlap global entre les prédictions et la vérité terrain. Une combinaison du type BCE + Dice est utilisée, avec la possibilité d'ajouter un terme Focal selon les besoins expérimentaux. La BCE stabilise l'apprentissage pixel par pixel, le Dice encourage un chevauchement global élevé et le terme Focal permet de réduire l'influence des exemples trop faciles.

L'optimisation est assurée par Adam ou AdamW, accompagnés d'un scheduler de réduction du taux d'apprentissage et de callbacks comme le *early stopping* ou la sauvegarde du meilleur modèle sur la validation. Ces mécanismes limitent le sur-apprentissage et assurent un entraînement stable.

III.3.7 Entraînement et sauvegarde des checkpoints

Le modèle est entraîné sur l'ensemble d'entraînement après augmentation. La validation périodique permet de suivre l'évolution des métriques (loss, Dice, Jaccard, sensibilité, spécificité). Le meilleur modèle, déterminé par la performance sur la validation, est sauvegardé sous forme de checkpoint pour être utilisé lors de la phase d'inférence.

L'historique d'entraînement (loss et métriques par epoch) est également sauvegardé afin de permettre le traçage des courbes d'apprentissage et l'étude d'ablation portant sur l'importance de l'augmentation ou des différents termes de la fonction de perte.

III.3.8 Phase d'inférence / Test

Pour une image d'entrée non annotée, le pipeline suit les étapes suivantes : prétraitement, passage dans le modèle entraîné, puis production d'une carte de probabilités $P(x, y)$.

À partir de cette carte, plusieurs opérations sont appliquées :

✓ **Seuillage** : conversion de la carte de probabilités en masque binaire à l'aide d'un seuil fixe (0.5) ou de méthodes adaptatives, par exemple le k-means appliqué aux valeurs de probabilité, la méthode d'Otsu ou encore l'optimisation du F1-score sur l'ensemble de validation.

✓ **Post-processing morphologique** : utilisation d'opérations d'ouverture et de fermeture, suppression des petites composantes et remplissage des trous afin d'obtenir une segmentation cohérente et cliniquement exploitable.

III.3.9 Extraction de la ROI et mesures quantitatives

La carte binaire produite en sortie du modèle correspond à la région d'intérêt (ROI) segmentée. L'évaluation des performances est réalisée sur l'ensemble de test à l'aide de métriques quantitatives classiques de segmentation, à savoir l'Accuracy, le Dice, le Jaccard

(IoU), la Sensitivity (Recall) et la Specificity. Ces métriques permettent de mesurer à la fois le recouvrement, la précision globale et la capacité de détection de la lésion.

En complément de l'évaluation quantitative, une analyse qualitative est menée afin d'apprécier visuellement le comportement des modèles. Cette analyse repose sur des visualisations comparatives illustrant la superposition des masques prédits avec la vérité terrain (Ground Truth). Elle met en évidence la précision des contours, la cohérence spatiale des segmentations et la sensibilité du modèle aux variations d'apparence des lésions.

III.3.10 Études complémentaires menées

✓ **Ablation study** : évaluation de l'impact de l'augmentation des données et des différents termes de la fonction de perte (BCE, Dice, Focal) en retirant ou ajoutant chaque composant pour mesurer sa contribution.

✓ **Robustesse aux artefacts** : comparaison de la méthode proposée avec d'autres approches de l'état de l'art (méthodes traditionnelles, FCN, SegNet, U-Net classique) sur des images altérées par des artefacts afin de démontrer la robustesse en conditions réalistes.

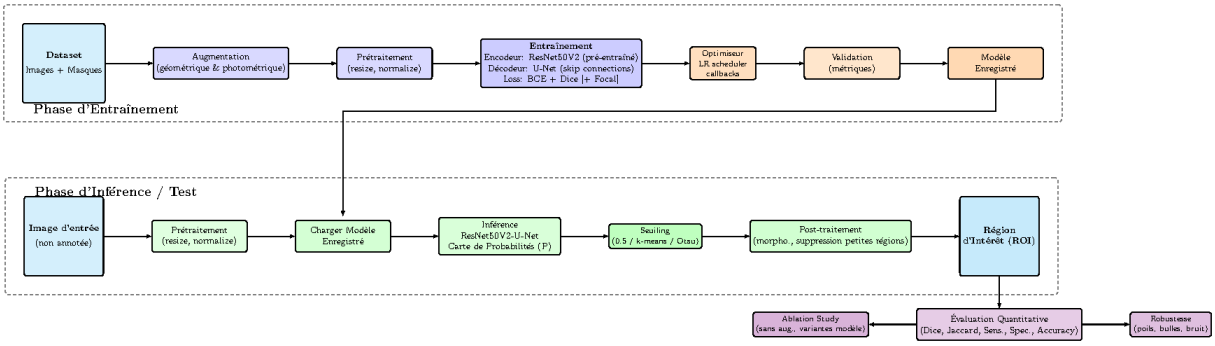


Figure III.2 Schéma global de la méthode proposée

III.4 Fonctions de perte

Les fonctions de perte constituent un élément fondamental de l'apprentissage profond. Elles mesurent l'écart entre la segmentation prédite par le modèle et la vérité terrain, pilotant ainsi la mise à jour des poids lors de la phase d'optimisation. Dans le cadre de la segmentation des lésions cutanées, où la Région d'Intérêt (ROI) est souvent petite et noyée dans un fond majoritaire, le choix de la fonction de perte joue un rôle déterminant dans la stabilité de l'entraînement, la précision des contours détectés et la capacité du modèle à s'adapter au déséquilibre de classes.

Les bases de données dermatologiques présentent un déséquilibre prononcé entre la ROI (lésion) et le fond. Une fonction de perte mal adaptée peut amener le modèle à minimiser uniquement l'erreur du fond, négligeant complètement la région ayant une importance clinique. Ainsi, une bonne fonction de perte doit non seulement fournir des gradients stables, mais aussi encourager la maximisation du chevauchement avec la lésion et renforcer l'apprentissage sur les zones difficiles à segmenter (bordures, textures faibles, artefacts).

III.4.1 Présentation des fonctions de perte

III.4.1.1 Binary Cross-Entropy (BCE)

La Binary Cross-Entropy est une perte orientée pixel, couramment utilisée pour la classification binaire, et appliquée ici indépendamment sur chaque pixel de l'image. Elle est définie par l'expression suivante :

$$L_{\text{BCE}} = -[y\log(p) + (1-y)\log(1-p)] \quad \text{(III.10)}$$

où :

$y \in \{0,1\}$ désigne l'étiquette réelle du pixel,

$p \in [0,1]$ représente la probabilité prédite par le modèle,

Les termes $\log(p)$ et $\log(1-p)$ mesurent respectivement la justesse de la prédiction pour les classes 1 et 0.

La BCE fournit des gradients forts et stables, ce qui en fait une fonction de perte efficace pour apprendre les frontières du ROI. Toutefois, elle reste sensible au déséquilibre de classes car les nombreux pixels du fond dominent le calcul de la perte.

III.4.1.2 Dice Loss

La Dice Loss vise à maximiser le chevauchement spatial entre la prédiction et la vérité terrain. Elle dérive du coefficient de Sørensen–Dice, largement utilisé pour l'évaluation des masques en segmentation. La perte associée est formulée comme :

$$L_{\text{Dice}} = 1 - \frac{2\sum(py) + \epsilon}{\sum p + \sum y + \epsilon} \quad \text{(III.11)}$$

où :

p désigne la carte de probabilité prédite,

y représente la vérité terrain,

$\sum(py)$ mesure le chevauchement entre prédiction et ROI,

ϵ est un terme de régularisation évitant les divisions par zéro.

La Dice Loss compense naturellement le déséquilibre entre classes, car elle évalue globalement la qualité de la segmentation et met en avant les pixels appartenant au ROI. En revanche, elle peut conduire à des gradients instables lorsque l'intersection entre la prédiction et la vérité terrain est initialement très faible.

III.4.1.3 Focal Loss

La Focal Loss est une extension de la BCE ayant pour objectif de réduire l'effet des pixels faciles et d'accentuer l'apprentissage sur les exemples difficiles, notamment dans les régions présentant une forte ambiguïté ou un faible contraste. Sa formulation est donnée par :

$$L_{\text{Focal}} = -\alpha(1-p)^{\gamma} y \log(p) - (1-\alpha)p^{\gamma}(1-y)\log(1-p) \quad \text{(III.12)}$$

où :

$\alpha \in [0,1]$ pondère l'importance de chaque classe,

$\gamma \geq 0$ est le paramètre de focalisation contrôlant la suppression des exemples faciles,

$(1-p)^{\gamma}$ et p^{γ} sont des facteurs modulants qui réduisent la perte des pixels déjà bien classés,

p et y ont les mêmes significations que précédemment.

Cette perte est appropriée lorsque la majorité des pixels est correctement classée (fond) mais que le modèle doit améliorer sa performance sur les zones difficiles (bordures, artefacts, textures faibles).

III.4.2 Analyse comparative des pertes

La figure ci-dessous (Figure V.3) illustre l'évolution des valeurs de BCE, Dice Loss et Focal Loss en fonction de la probabilité prédite pour un pixel appartenant à la ROI ($y = 1$). On observe que :

- La BCE diminue plus lentement, restant non nulle et contribuant à des gradients stables pour tous les pixels.
- La Dice Loss reste relativement faible lorsque le chevauchement est important mais augmente fortement si le chevauchement est initialement faible, reflétant sa sensibilité à la qualité globale de la segmentation.
- Pour des pixels bien classés ($p \to 1$), le Focal Loss décroît rapidement vers zéro, réduisant l'influence des pixels faciles et permettant au modèle de se concentrer sur les zones difficiles.

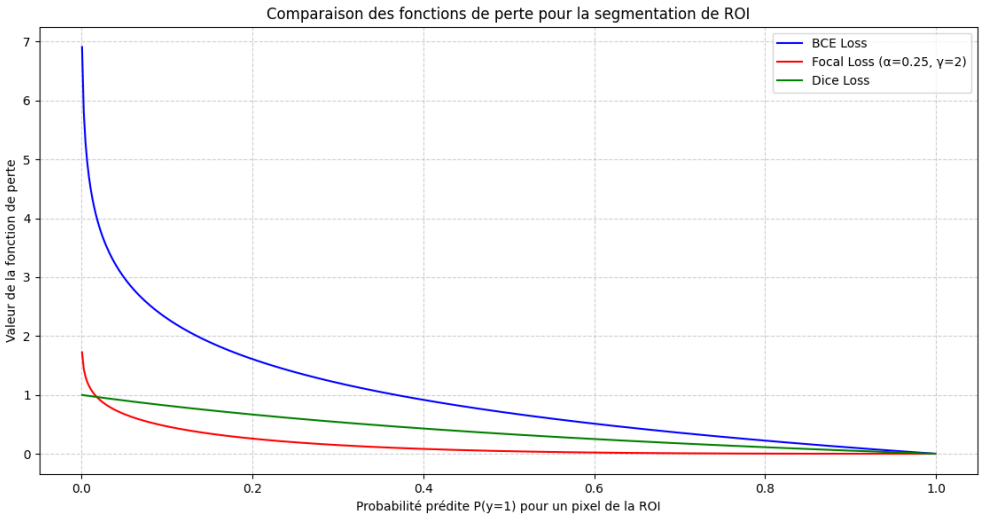


Figure III.3 Evolution de fonctions de perte en fonctions de probabilité prédite

Cette visualisation permet de comprendre l'impact des différentes fonctions de perte sur l'apprentissage et la manière dont elles gèrent le déséquilibre de classes (petite ROI vs fond majoritaire).

III.4.3 Combinaisons des fonctions de perte

Une stratégie couramment adoptée consiste à combiner la BCE et la Dice Loss :

$$L_{\text{Hybride}} = L_{\text{BCE}} + L_{\text{Dice}} \tag{III.13}$$

Cette combinaison exploite la complémentarité des deux pertes. En effet la Dice Loss met l'accent sur la ROI et corrige le déséquilibre entre classes tandis que la BCE apporte une stabilité de gradient indispensable, notamment aux premières itérations où la Dice Loss peut devenir instable.

Une autre combinaison pertinente pour les cas difficiles consiste à associer Dice Loss et Focal Loss :

$$L_{\text{Hybride}} = L_{\text{Dice}} + L_{\text{Focal}} \tag{III.14}$$

Cette combinaison présente un double avantage. Dans ce cas la Focal Loss cible les pixels mal classés en réduisant l'impact des exemples faciles et la Dice Loss maintient un objectif global de maximisation du chevauchement entre ROI et prédiction.

Cette étude a guidé le choix des fonctions de perte en fonction de l'impact direct sur la qualité de segmentation. La Binary Cross-Entropy (BCE) est optimale pour les scénarios où les classes sont équilibrées ou lorsque la précision des frontières est le facteur dominant.

Inversement, la Focal Loss et la Dice Loss sont préférables pour les images présentant un déséquilibre de classes sévère. Les combinaisons hybrides telles que BCE + Dice et Dice + Focal permettent d'exploiter les forces de chaque fonction tout en compensant leurs limites, ce qui les rend particulièrement adaptées à la segmentation des lésions cutanées dans des contextes cliniques exigeants.

III.5 Métriques d'évaluation pour la segmentation

L'évaluation d'un modèle de segmentation nécessite des métriques adaptées à la nature binaire et déséquilibrée des images médicales. Dans le contexte de la segmentation de lésions cutanées, la Région d'Intérêt (ROI) représentant la lésion est souvent très petite par rapport au fond de la peau. Ainsi, il est crucial d'utiliser des métriques qui mesurent à la fois la qualité globale de la prédiction, la capacité du modèle à détecter correctement les pixels de la lésion et à éviter les fausses détections sur le fond.

Les métriques classiques utilisées incluent l'Exactitude (Accuracy), la Sensibilité (Recall / True Positive Rate), la Spécificité (Specificity / True Negative Rate), le Coefficient de Dice (Dice Coefficient) et l'Indice de Jaccard (IoU). Chaque métrique fournit une perspective complémentaire sur la performance du modèle.

Pour toutes les métriques suivantes, nous définissons :

- **TP** : True Positives est le nombre de pixels correctement classés comme appartenant à la lésion (ROI)
- **TN** : True Negatives est le nombre de pixels correctement classés comme appartenant au fond
- **FP** : False Positives est le nombre de pixels classés comme lésion alors qu'ils appartiennent au fond
- **FN** : False Negatives est nombre de pixels classés comme fond alors qu'ils appartiennent à la lésion

Ces valeurs sont calculées en comparant le masque prédit P avec le masque de vérité terrain Y pixel par pixel.

III.5.1 Exactitude (Accuracy)

L'Exactitude mesure la proportion de pixels correctement classés (ROI et fond) sur l'ensemble de l'image :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad \text{(III.15)}$$

Cette métrique donne une mesure globale de performance sur tous les pixels. Cependant elle peut être trompeuse en cas de fort déséquilibre, car le fond dominant peut produire une haute exactitude même si la ROI est mal segmentée.

III.5.2 Sensivity

La Sensibilité mesure la capacité du modèle à détecter correctement les pixels de la lésion :

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{III.16}$$

La sensivity indique la proportion de pixels de la ROI correctement détectés. Elle est déterminante pour l'analyse des lésions. Une valeur trop basse implique que le modèle échoue à couvrir l'intégralité de la lésion, rendant le diagnostic potentiellement incomplet ou erroné

III.5.3 Spécificité (Specificity / True Negative Rate)

La Spécificité mesure la capacité du modèle à reconnaître correctement les pixels du fond

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{III.17}$$

Cette mesure évite les fausses détections dans le fond. Elle est complémentaire à la sensibilité pour comprendre l'équilibre entre détection correcte de la ROI et prévention des erreurs sur le fond.

III.5.4 Coefficient de Dice (Dice Coefficient)

Le Coefficient de Dice mesure le chevauchement entre le masque prédit et le masque réel :

$$\text{Dice} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} = \frac{2|P \cap Y|}{|P| + |Y|} \tag{III.18}$$

Le Dice evalue directement la qualité du chevauchement spatial. Il est très adapté aux ROI petites ou déséquilibrées et il est plus sensible aux erreurs sur la lésion que l'accuracy, ce qui est crucial pour la segmentation dermatologique.

III.5.5 Indice de Jaccard (Intersection over Union – IoU)

L'IoU est une autre mesure de chevauchement :

$$\text{IoU} = \frac{TP}{TP + FP + FN} = \frac{|P \cap Y|}{|P \cup Y|} \tag{III.19}$$

Cet indice est similaire au Dice mais plus strict. Il est utilisé fréquemment dans les challenges de segmentation médicale pour la comparaison entre méthodes. Il complète le Dice en donnant une perspective différente du chevauchement.

Aucune métrique unique n'est suffisante pour juger la qualité d'un modèle sur des images médicales déséquilibrées. L'utilisation combinée de ces métriques permet d'avoir une évaluation complète et fiable pour les cas cliniques réels.

III.6 Architecture proposée : ResNet50V2–U-Net

La segmentation des lésions cutanées nécessite simultanément une compréhension sémantique globale de l'image et une préservation précise des frontières de la région d'intérêt (ROI). Pour satisfaire ces exigences, l'approche adoptée repose sur une architecture hybride qui combine un encodeur ResNet50V2 pré-entraîné avec un décodeur inspiré du U-Net. Cette combinaison assure une fusion efficace entre un apprentissage profond hiérarchique et une reconstruction spatiale détaillée (Fig. X). La présente section expose les motivations, les mécanismes internes et les bénéfices associés à cette architecture dans le cadre spécifique de la segmentation médicale.

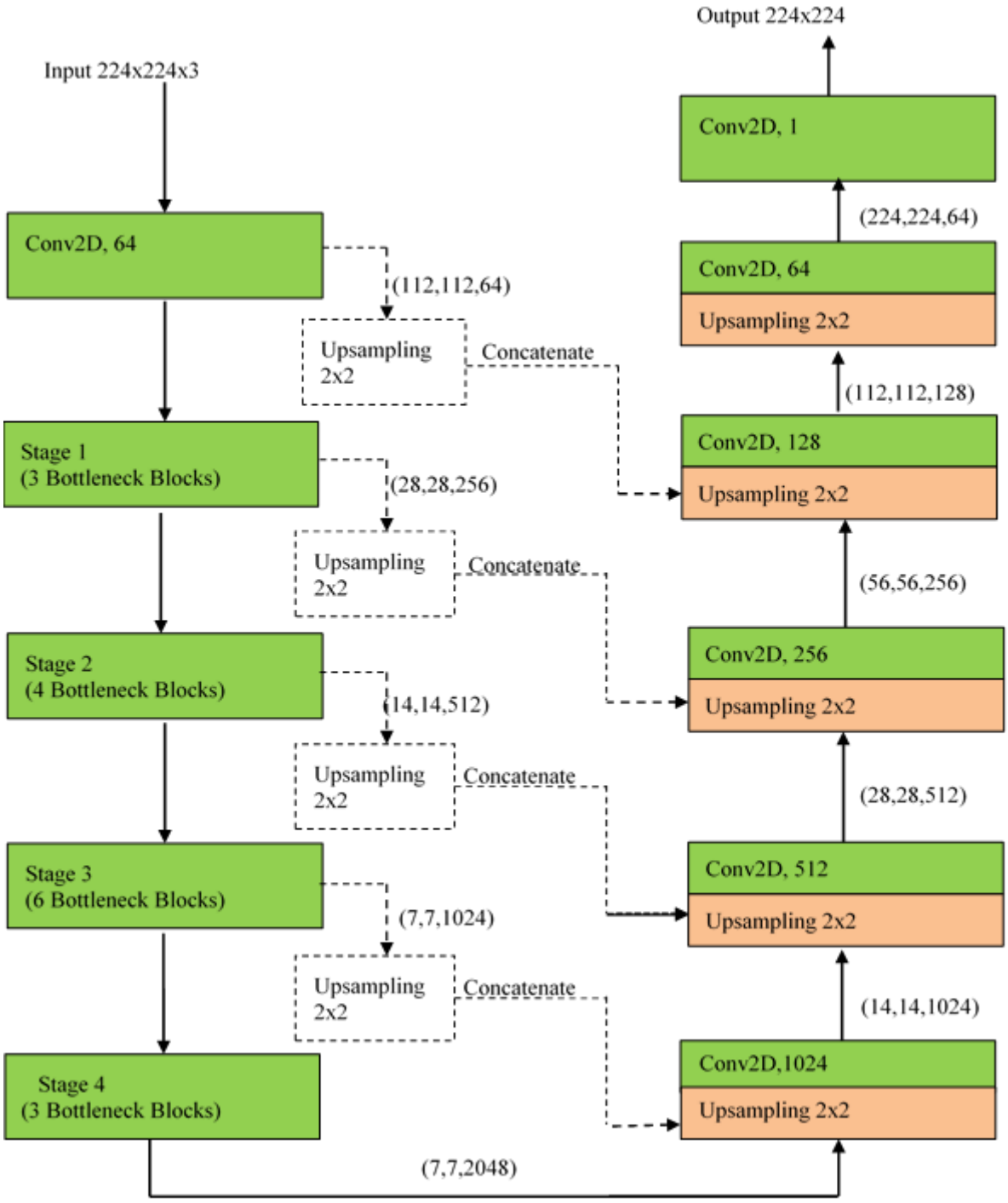


Figure III.4 Architecture du modèle proposé

III.6.1 ResNet50V2 comme encodeur

Le ResNet50V2 est une évolution du modèle ResNet50 qui exploite l'apprentissage résiduel profond afin d'atténuer le problème du gradient évanescent dans les réseaux de grande profondeur. Ses connexions résiduelles permettent au modèle d'apprendre des représentations

complexes tout en maintenant une propagation de gradient stable, ce qui évite la dégradation des performances lors de l'ajout de nombreuses couches.

Dans notre architecture, le ResNet50V2 pré-entraîné sur ImageNet joue le rôle d'encodeur pour plusieurs raisons.

Premièrement, le transfert d'apprentissage permet de réutiliser des filtres capables de détecter des motifs génériques comme les textures, les contours ou les structures visuelles complexes. Ces caractéristiques sont particulièrement pertinentes pour les images dermatologiques, même lorsque le volume du dataset est limité.

Deuxièmement, la profondeur de ResNet50V2 offre une hiérarchie riche de caractéristiques où les premières couches capturent les détails fins et les couches profondes extraient des représentations sémantiques robustes. Cette combinaison est essentielle pour identifier et localiser précisément une lésion.

Troisièmement, l'organisation interne du ResNet50V2 repose sur le schéma de pré-activation (Batch Normalization suivie de ReLU puis convolution). Cette conception améliore la stabilité numérique et facilite la convergence lors de l'entraînement.

Ainsi, l'utilisation de ResNet50V2 comme backbone optimise la capture du contexte global, renforce la robustesse aux variations lumineuses et réduit la sensibilité aux artefacts fréquemment présents dans les images dermatologiques.

III.6.2 Décodeur U-Net : reconstruction hiérarchique et préservation spatiale

Le décodeur suit le principe fondateur du U-Net, architecture initialement développée pour la segmentation biomédicale. Son rôle est de transformer les cartes de caractéristiques extraites par l'encodeur en un masque de segmentation haute résolution. La reconstruction progressive s'appuie sur deux éléments clés.

III.6.2.1 Upsampling par convolutions transposées

Contrairement aux interpolations statiques, les convolutions transposées apprennent un upsampling adapté aux données. Elles permettent une récupération progressive de la géométrie de la lésion, une reconstitution fidèle des contours irréguliers et une transition maîtrisée entre les différentes résolutions.

III.6.2.2 Skip connections entre encodeur et décodeur

Les skip connections constituent un mécanisme central de l'architecture. Elles réinjectent dans le décodeur les informations spatiales fines capturées par les premières couches du ResNet50V2.

Chaque étape d'upsampling combine ainsi deux sources d'information complémentaires :

- Les caractéristiques profondes fournissant le contexte global et la forme générale de la lésion,

- Les détails de haute résolution apportés par l'encodeur, essentiels pour délimiter les bordures et les micro-textures.

Ce mécanisme compense la perte d'information causée par les opérations de pooling et améliore nettement la précision des frontières, ce qui est crucial en dermatologie où les contours sont souvent flous ou irréguliers.

III.6.3 Intégration, cohérence de l'architecture et optimisation

L'intégration du ResNet50V2 au sein du U-Net nécessite une adaptation des résolutions et des profondeurs. Les dimensions des cartes de caractéristiques sont ajustées afin d'être compatibles lors des concaténations avec les skip connections.

Durant le fine-tuning, certaines couches du ResNet50V2 sont gelées dans les premières phases afin de préserver les représentations génériques apprises sur ImageNet, tandis que les couches plus profondes et l'ensemble du décodeur sont optimisées.

La fusion multi-échelle permet au modèle de combiner simultanément des informations globales (forme générale de la lésion) et des détails locaux (irrégularités des bords). Cette cohérence architecturale conduit à un équilibre efficace entre généralisation, stabilité et précision structurelle.

III.6.4 Apport et pertinence en segmentation dermatologique

Cette architecture hybride présente plusieurs avantages significatifs. Elle améliore la sensibilité en détectant les lésions de petite taille. Elle renforce la spécificité en limitant les fausses détections dans les zones de peau saine. Elle capture plus fidèlement les contours, ce qui est essentiel pour les mesures diagnostiques portant sur la surface, la forme ou l'asymétrie.

En outre, elle fait preuve d'une robustesse accrue face aux variations d'éclairage, au bruit et aux artefacts d'acquisition. L'association d'un encodeur profond pré-entraîné et d'un décodeur U-Net exploitant les skip connections constitue ainsi une solution particulièrement adaptée à la segmentation des lésions cutanées, où la précision des frontières comporte un enjeu clinique majeur.

III.7 Datasets utilisés

Dans cette étude, nous avons exploité deux jeux de données dermoscopiques : PH² pour l'entraînement et la validation, et ISIC 2016 pour évaluer la capacité de généralisation de notre modèle sur des images externes.

III.7.1 Dataset PH²

Le dataset PH² (A Dermoscopic Image Database for Research and Benchmarking) est constitué d'images acquises au Service de Dermatologie de l'Hôpital Pedro Hispano à Matosinhos, Portugal, avec le Tuebinger Mole Analyzer à 20 fois. Il est destiné à la segmentation et à la classification de lésions cutanées.

1) Caractéristiques techniques

- **Nombre d'images :** 200
- **Format :** RGB 8 bits par canal
- **Résolution :** 768 × 560 pixels
- **Contenu :** Naevus communs, naevus atypiques et mélanomes

2) Annotations (Ground Truth)

- Masques binaires manuels pour chaque lésion
- Diagnostics cliniques et histologique
- Métadonnées dermoscopiques détaillées (ABCDE, couleurs, points/globules, stries, zones de régression)

Le tableau suivant donne la répartition des lésions dans le dataset PH² :

Tableau III.1 Répartition des lésions dans le dataset PH²

Catégorie de lésion	Nombre d'images	Pourcentage
Naevus commun	80	40 %
Naevus atypique	80	40 %
Mélanome	40	20 %
Total	**200**	**100 %**

3) Prétraitement appliqué

- Redimensionnement et normalisation des images
- Harmonisation des intensités pour réduire l'effet des variations d'éclairage

4) Division pour entraînement et test

- **Entraînement :** 80 % des images
- **Test interne :** 20 % des images restantes de PH² pour évaluer la performance sur des données issues du même dataset

III.7.2 Dataset ISIC 2016

Le dataset ISIC 2016 provient du projet International Skin Imaging Collaboration. Il contient des images dermoscopiques annotées pour la segmentation et la classification, utilisées ici uniquement comme test externe afin d'évaluer la capacité de généralisation du modèle.

1) Caractéristiques techniques

- **Format :** RGB
- **Résolution :** variable selon l'image
- **Contenu :** principalement des mélanomes et névus

2) Annotations (Ground Truth)

- Masques binaires pour chaque image

- Diagnostics clinique et histologique

La répartition des lésions dans le dataset ISIC 2016 est donné par le tableau suivant :

Tableau III.2 Répartition des lésions dans le dataset ISIC 2016

Classe de lésion	Nombre d'images
Mélanome	173
Autres / inconnu	727
Total	**900**

III.7.3 Utilisation

Le modèle entraîné sur PH² est évalué sur ISIC 2016 pour mesurer sa capacité à segmenter des images provenant d'une source différente. L'utilisation des datasets dans le pipeline expérimental est résumé par le tableau suivant :

Tableau III.3 Utilisation des data sets PH² et ISIC 2016

Dataset	Usage	Objectif
PH²	Entraînement / Test interne	Ajuster les poids du modèle et valider sa performance sur un dataset connu
ISIC 2016	Test externe	Évaluer la généralisation sur des images provenant d'une source indépendante

Le pipeline expérimental suit donc cette logique : entraîner le modèle sur PH², valider en interne sur les images restantes de PH², puis tester sur ISIC 2016 pour évaluer la robustesse et la capacité de généralisation de notre ResNet50V2–U-Net sur des données non vues et provenant d'un autre contexte clinique.

III.8 EXPÉRIMENTATION ET ANALYSE DES RÉSULTATS

III.8.1 Augmentation des données

Une stratégie d'augmentation a été appliquée exclusivement aux données d'apprentissage afin d'améliorer la robustesse du modèle. Les transformations utilisées sont résumées dans le tableau suivant :

Tableau III.4 Paramètres d'augmentation appliqués

Paramètre	Valeur	Effet
rotation_range	20°	Rotation aléatoire
width_shift_range	0.1	Décalage horizontal
height_shift_range	0.1	Décalage vertical
shear_range	0.1	Cisaillement
zoom_range	0.1	Zoom ±10%
horizontal_flip	True	Symétrie horizontale
fill_mode	nearest	Remplissage

Cette stratégie contribue à réduire le sur-apprentissage et à renforcer la robustesse du modèle.

III.8.2 Courbes d'apprentissage et convergence

Les courbes d'apprentissage relatives aux principales métriques (loss, accuracy, Dice, Jaccard, sensibilité, spécificité) sont présentées dans les figures suivantes (Figure V.5) :

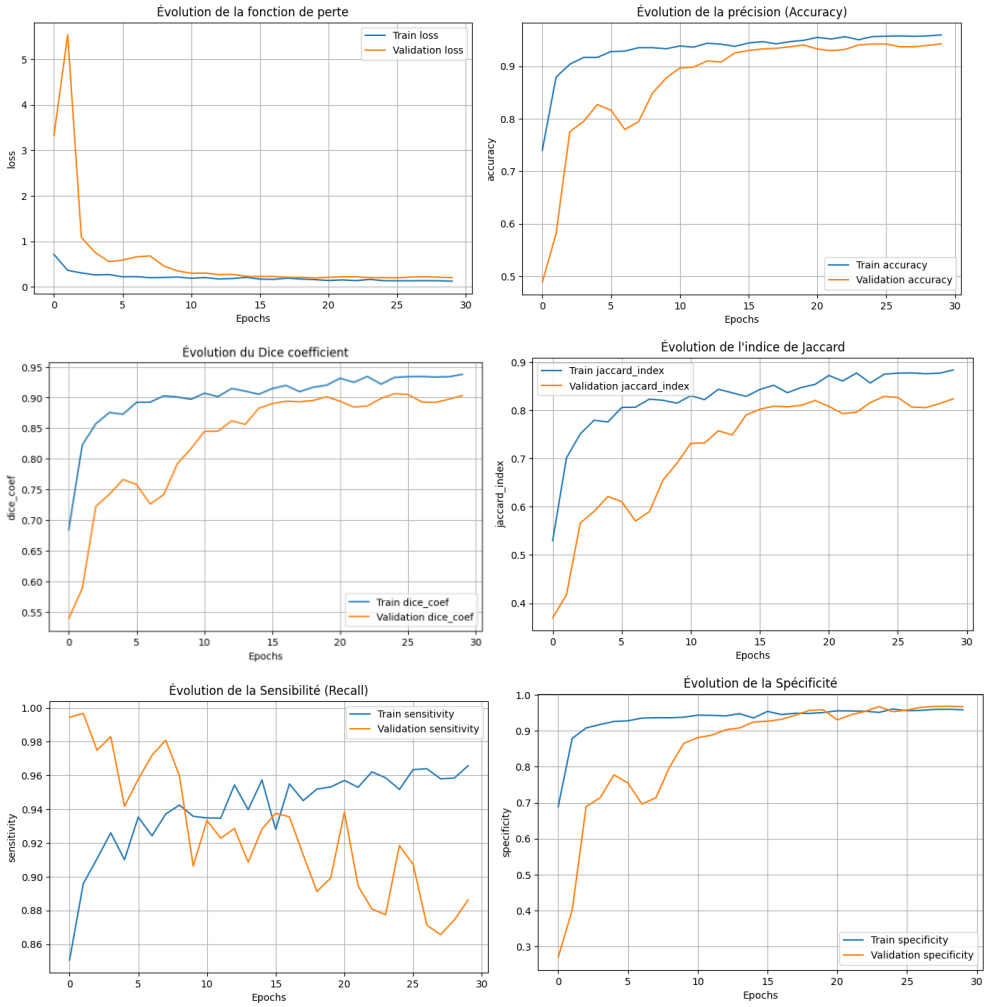


Figure III.5 Évolution des métriques d'entraînement et de validation.

L'analyse des courbes montre :

- ✓ Une diminution rapide de la fonction de perte, suivie d'une stabilisation,
- ✓ Une amélioration progressive des métriques (Dice, Jaccard, accuracy),
- ✓ Une spécificité élevée, traduisant une bonne maîtrise des faux positifs,
- ✓ Quelques fluctuations de la sensibilité en validation, attribuées à la taille réduite du dataset.

Aucun signe de sur-apprentissage notable n'a été observé.

III.8.3 Performances intermédiaires

Afin d'évaluer quantitativement les performances du modèle au cours du processus d'apprentissage, une analyse est menée sur les ensembles d'entraînement et de validation. Les résultats obtenus sur l'ensemble d'entraînement permettent d'apprécier la capacité du modèle à apprendre efficacement les caractéristiques des régions d'intérêt, tandis que l'analyse sur l'ensemble de validation vise à mesurer sa capacité de généralisation et sa robustesse face à des données non vues. Les performances correspondantes sont présentées de manière synthétique ci-dessous.

Entraînement :

Accuracy	Dice	Jaccard	Sensitivity	Specificity	Loss
0.9599	0.9378	0.8834	0.9658	0.9583	0.1220

Validation :

Accuracy	Dice	Jaccard	Sensitivity	Specificity	Loss
0.9429	0.9033	0.8236	0.8863	0.9670	0.1992

Ces résultats témoignent d'une segmentation précise, avec un Dice élevé, un Jaccard supérieur à 0.82 et des valeurs de sensibilité et spécificité équilibrées, confirmant une bonne généralisation interne.

III.8.4 Étude d'Ablation

L'étude d'ablation a pour objectif d'évaluer l'impact réel de chaque composant de l'architecture proposée sur la qualité de la segmentation. Elle vise à isoler et analyser trois facteurs déterminants :

4. L'utilisation ou non des techniques d'augmentation de données. L'objectif est d'évaluer dans quelle mesure l'augmentation contribue à la généralisation du modèle, en particulier sur un dataset de taille limitée tel que PH^2. Les variantes testées sont le modèle ResNet50V2-U-Net sans augmentation, fonction de perte BCE et le modèle ResNet50V2-U-Net avec augmentation, fonction de perte BCE.
5. l'apport du backbone ResNet50V2 en tant qu'encodeur. L'expérience dans ce cas évalue la pertinence de remplacer l'encodeur convolutionnel classique du U-Net par

un backbone profond préentraîné. Les variantes testées sont le modèle U-Net de base avec augmentation, fonction de perte BCE et le modèle ResNet50V2-U-Net avec augmentation, fonction de perte BCE.

6. L'influence du choix de la fonction de perte sur la convergence et la précision de la segmentation. L'objectif est d'analyser l'effet des fonctions de perte couramment utilisées en segmentation : BCE, Focal Loss, Dice Loss, et leurs combinaisons.

III.8.4.1 Résultats quantitatifs

Les performances sont évaluées sur le jeu de test interne du dataset PH², en utilisant cinq métriques robustes : accuracy, Dice coefficient, Jaccard index, sensitivity et specificity. Les tableaux suivants synthétisent les résultats pour chaque axe de l'ablation.

- ✓ **Effet de l'augmentation**

Tableau III.5 Comparaison des performances avec et sans augmentation de données

Modèle	Accuracy	Dice	Jaccard	Sensitivity	Specificity
ResNet-U-Net sans augmentation	0.9667	0.9396	0.8881	0.9506	0.9756
ResNet-U-Net avec augmentation	0.9694	0.9486	0.9040	0.9380	0.9836

L'introduction de l'augmentation améliore systématiquement les scores de Dice (+0.9%) et de Jaccard (+1.6%), confirmant son importance pour la généralisation, malgré la petite taille du dataset PH².

- ✓ **Effet du backbone ResNet50V2**

Tableau III.6 Comparaison des performances avec et sans backbone ResNet50V2

Modèle	Accuracy	Dice	Jaccard	Sensitivity	Specificity
U-Net de base	0.9449	0.9127	0.8446	0.9216	0.9748
ResNet-U-Net	0.9694	0.9486	0.9040	0.9380	0.9836

Le backbone ResNet50V2 apporte une amélioration nette : +3.6% Dice et +5.9% Jaccard par rapport au U-Net classique, illustrant la contribution majeure de représentations hiérarchiques plus profondes et plus discriminantes.

- ✓ **Effet de la fonction de perte**

Tableau III.7 Comparaison des performances selon la fonction de perte utilisée

Modèle	Accuracy	Dice	Jaccard	Sensitivity	Specificity
BCE	0.9694	0.9486	0.9040	0.9380	0.9836
Focal Loss	0.9609	0.9379	0.8862	0.9122	0.9905
Dice Loss	0.9652	0.9442	0.8974	0.9516	0.9681

Modèle	Accuracy	Dice	Jaccard	Sensitivity	Specificity
Dice + Focal Loss	0.9638	0.9418	0.8927	0.9247	0.9856
Dice + BCE	**0.9710**	**0.9498**	**0.9059**	0.9445	0.9792

La combinaison Dice + BCE obtient les meilleures performances globales, dépassant toutes les autres fonctions de perte. La Focal Loss augmente la spécificité mais réduit la sensibilité, confirmant son orientation vers la gestion des classes minoritaires. La Dice Loss seule maximise la sensibilité mais dégrade la spécificité.

III.8.4.2 Analyse qualitative des résultats

Cette étude présente une analyse qualitative des résultats de segmentation obtenus par différentes variantes d'architectures appliquées à la segmentation de lésions cutanées du dataset PH². L'examen visuel des segmentations fournit des informations essentielles sur la précision des contours, la stabilité des prédictions, la sensibilité aux artefacts et la capacité générale de généralisation.

La figure qualitative (Figure V.6) regroupe, pour chaque image testée :

- ✓ la vérité terrain (Ground Truth, GT),
- ✓ les segmentations obtenues avec différentes variantes utilisées .

Chaque colonne représente une variante du modèle, tandis que chaque ligne illustre un exemple de lésion parmi le jeu de test. Cette organisation permet une comparaison directe de l'effet de chaque modificateur du pipeline (augmentation, backbone, fonction de perte).

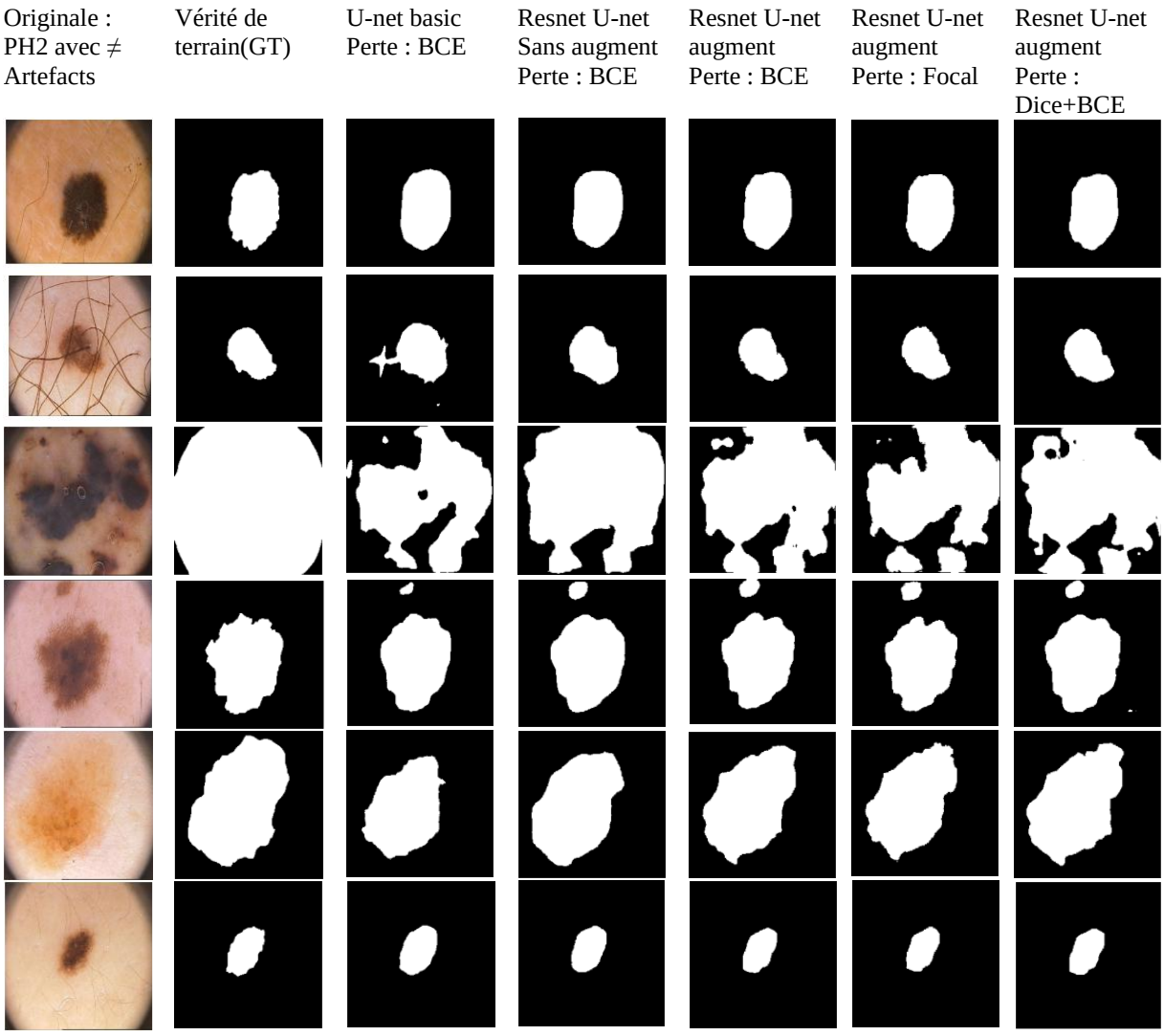


Figure III.6 Grille qualitative comparant les segmentations des différentes variantes.

Les barplots (FigureIV.7) superposés montrent les performances moyennes pour l'Accuracy, Dice, Jaccard, Sensitivity et Specificity.

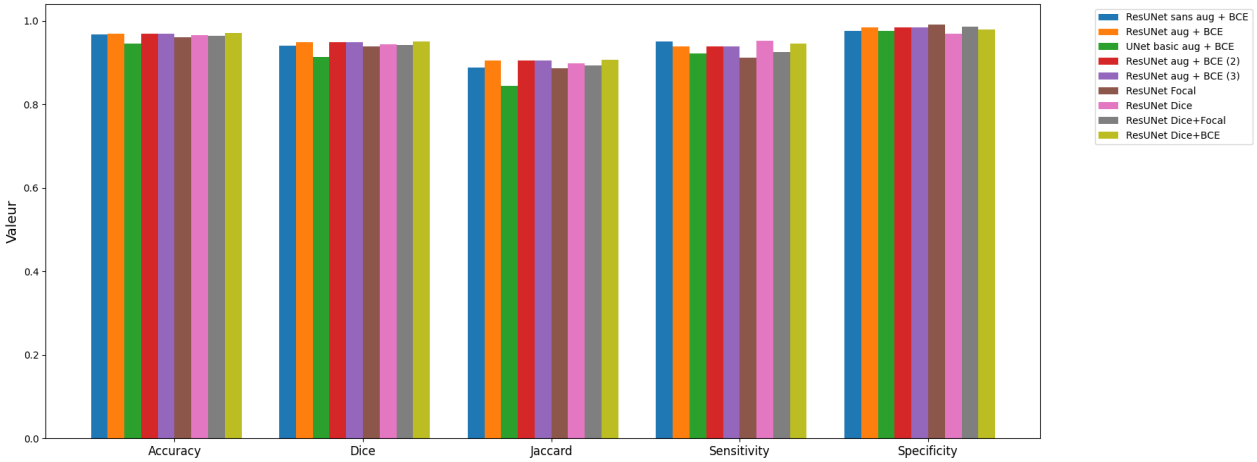


Figure III.7 Comparaison des métriques pour toutes les variantes

L'examen visuel montre d'abord que les modèles entraînés sans augmentation présentent fréquemment des masques irréguliers, avec des contours discontinus et de petits artefacts révélateurs d'une généralisation limitée. À l'opposé, l'introduction d'une stratégie d'augmentation améliore nettement la cohérence des masques, qui épousent plus fidèlement les frontières réelles, notamment lorsque les lésions affichent des variations de texture ou des contours complexes.

La comparaison entre le U-Net de base et la variante intégrant ResNet50V2 comme encodeur confirme ensuite l'intérêt du pré-entraînement et de la profondeur du backbone. Le modèle classique reste globalement performant, mais il montre parfois une capacité réduite à capturer les micro-structures ou les zones faiblement contrastées. Le modèle hybride, au contraire, produit des segmentations plus régulières, mieux définies et plus stables, en particulier pour les lésions peu contrastées ou aux bordures diffuses.

Enfin, les différentes fonctions de perte révèlent chacune un comportement distinct. La BCE génère des segmentations équilibrées et bien structurées, tandis que la Focal Loss, davantage orientée vers les zones difficiles, peut conduire à des masques légèrement fragmentés dans certaines situations. La Dice Loss, optimisant directement le chevauchement, tend à produire des segmentations plus expansives qui couvrent systématiquement la totalité de la lésion. Les versions combinant Dice avec BCE ou Focal Loss se distinguent par un compromis particulièrement efficace : les masques générés sont à la fois réguliers, précis sur les contours et suffisamment complets, ce qui explique leurs performances globalement supérieures.

III.8.5 Comparaison avec l'état de l'art

Afin d'évaluer rigoureusement la performance du modèle proposé, une comparaison approfondie a été réalisée avec un ensemble représentatif de méthodes de segmentation d'images médicales, incluant à la fois des approches traditionnelles et des architectures profondes couramment utilisées dans la littérature. Toutes les méthodes ont été évaluées sur le même jeu de test du dataset PH2, selon les mêmes métriques : accuracy, Dice, Jaccard,

sensitivity et specificity. Les résultats présentés ci-dessous permettent d'apprécier la robustesse du modèle proposé par rapport aux références existantes.

III.8.5.1 Comparaison avec les méthodes traditionnelles

Les méthodes traditionnelles Otsu, FCM spatial, Contour actif de Chan-Vese et Markov Random Field (MRF) ont été implémentées et appliquées au même jeu de test. Le tableau ci-dessous résume leurs performances :

Tableau III.8 Résultats des méthodes traditionnelles sur PH2

Méthode	**Accuracy**	**Dice**	**Jaccard**	**Sensitivity**	**Specificity**
Otsu	0.7461	0.5147	0.3798	0.8412	0.5473
FCM spatial	0.6957	0.5551	0.4154	0.7071	0.6575
Contour actif (Chan-Vese)	0.8901	0.8327	0.7392	0.9833	0.7566
MRF	0.6951	0.5700	0.4421	0.6884	0.6979

Les résultats montrent que :

- ✓ Les méthodes globales (Otsu, FCM) sont limitées par l'hétérogénéité des lésions du dataset PH2.
- ✓ Chan-Vese obtient des scores nettement supérieurs grâce à sa capacité à capturer les contours continus, mais reste sensible au bruit et dépend de l'initialisation.
- ✓ MRF améliore la régularisation spatiale mais souffre d'un manque de précision locale.

Dans l'ensemble, ces approches présentent une faible capacité de généralisation et montrent leurs limites face à la variabilité des lésions cutanées, ce qui justifie l'usage de méthodes d'apprentissage profond.

III.8.5.2 Comparaison avec les modèles d'apprentissage profond

Trois modèles ont été implémentés pour établir une évaluation basique des approches profondes classiques à savoir U-Net basique, FCN-8s et SegNet

Le tableau ci-après rassemble les performances obtenues sur le jeu de test de PH2.

Tableau III.9 Performances des modèles deep learning implémentés

Modèle	**Accuracy**	**Dice**	**Jaccard**	**Sensitivity**	**Specificity**
U-Net basique	0.9449	0.9127	0.8446	0.9216	0.9748
SegNet	0.9392	0.8871	0.8042	0.9486	0.9352
FCN-8s	0.9606	0.9292	0.8700	0.9176	0.9772
Modèle proposé (ResNet50V2-U-Net + Aug + Dice+BCE)	**0.9710**	**0.9498**	**0.9059**	**0.9445**	**0.9792**

- ✓ Le modèle proposé obtient les meilleures valeurs sur toutes les métriques, dépassant largement les méthodes classiques et améliorant les trois architectures deep learning baselines.
- ✓ L'intégration de ResNet50V2 comme encodeur, combinée à l'augmentation de données et à une fonction de perte hybride (Dice + BCE), contribue à une meilleure capture des structures fines des lésions.
- ✓ FCN-8s fournit de bonnes performances, notamment en specificity, mais reste inférieur au modèle proposé en termes de Dice et Jaccard, indispensables pour une segmentation de haute précision.
- ✓ U-Net basique montre une bonne stabilité mais manque de richesse représentationnelle comparé à l'architecture proposée.

III.8.5.3 Positionnement par rapport à l'état de l'art

Dans la littérature, de nombreuses variantes du modèle U-Net ont été proposées pour améliorer la segmentation des lésions cutanées, chacune introduisant des mécanismes spécifiques (attention, connexions multi-échelles, transformeurs, métadonnées, etc.). Le Tableau suivant résume les performances couramment rapportées dans les travaux récents sur les jeux de données dermatologiques tels que PH2, ISIC 2016, ISIC 2017 ou ISIC 2018 :

Tableau III.10 Comparaison avec l'état de l'art récent

Méthode	**Architecture principale**	**Jeu(x) de données**	**IoU (%)**	**Dice (%)**
U-Net	Encodeur-décodeur	ISIC, PH2	~80	~88
U-Net++	Connexions denses imbriquées	PH2	84.3	90.2
U-Net 3+	Full-scale skip connections	ISIC 2018	84	90
MSAU-Net	Attention multi-échelle + ConvLSTM	ISIC 2018	95.6	94.9
FAT-Net	Double encodeur CNN+Transformer	ISIC 2017	76.5	85.0
DEU-Net	Dual encoder (EfficientNet + MaxViT)	ISIC 2016-PH2	87-91	90-95
Meta-UNet	U-Net + métadonnées	PH2	84-88	90-93
DilatedSkinNet	Convolutions atrous	ISIC 2018	89	94
EGE-UNet	GHPA + GAB	ISIC 2018	80.9	89.5

Ces travaux montrent que :

- ✓ Les modèles U-Net classiques atteignent généralement un Dice d'environ 88 % sur PH2 et ISIC.
- ✓ Les versions améliorées comme U-Net++, U-Net 3+ ou Meta-UNet introduisent des connexions plus riches ou des informations supplémentaires, permettant de monter à 90-93 % de Dice.

- ✓ Les architectures plus avancées intégrant des mécanismes d'attentions (MSAU-Net) ou des encodeurs hybrides (DEU-Net) dépassent parfois les 94-95 %, mais au prix d'une complexité computationnelle élevée.
- ✓ Les modèles basés sur les transformeurs (ex. FAT-Net) restent sensibles aux artefacts et n'améliorent pas systématiquement la précision sur PH2.

III.8.6 Positionnement du modèle proposé

Les performances de notre modèle ResNet50V2-U-Net avec augmentation et perte hybride (Dice + BCE) atteignent :

- ✓ **Dice** = 94.98 %
- ✓ **IoU** = 90.59 %

Ces valeurs placent clairement notre approche :

- ✓ Au-dessus des modèles U-Net classiques (88 %) et U-Net++ / U-Net 3+ (90-92 %),
- ✓ Au niveau des architectures avancées telles que DEU-Net ou DilatedSkinNet,
- ✓ Légèrement en dessous des modèles très complexes à attention multi-échelle comme MSAU-Net (94.9-95 %), mais avec un coût computationnel nettement inférieur.

Ainsi, le modèle proposé se situe parmi les meilleures approches publiées sur PH2 et les jeux dérivés d'ISIC, tout en conservant :

- ✓ Une architecture raisonnablement légère,
- ✓ Une bonne capacité de généralisation,
- ✓ Une robustesse face au déséquilibre de classes, grâce à la combinaison Dice + BCE et à l'augmentation adaptée.

Il constitue donc un compromis optimal entre précision, stabilité et coût de calcul, ce qui le place solidement dans le haut du spectre des méthodes de segmentation de lésions cutanées existantes.

III.8.7 Évaluation sur test externe (ISIC 2016)

L'évaluation sur un jeu de données externe constitue une étape essentielle pour mesurer la capacité de généralisation d'un modèle de segmentation, en particulier lorsqu'il a été entraîné sur un dataset restreint comme PH2. Dans cette optique, le modèle proposé — entraîné exclusivement sur les images du dataset PH2 — a été testé sur le jeu de test officiel du dataset ISIC 2016, sans aucune ré-entraînement ni adaptation de domaine. Les images ISIC ayant des caractéristiques visuelles très différentes (variabilité d'éclairage, couleurs hétérogènes, présence de poils, artefacts d'instrumentation, contours flous), ce protocole permet d'évaluer directement la robustesse et la capacité de transfert du modèle.

Avant l'inférence, les images ISIC ont été normalisées selon la même procédure utilisée lors de l'entraînement (redimensionnement en 224 × 224, normalisation standard, et seuillage fixe

appliqué à la sortie du modèle). Aucun traitement supplémentaire n'a été appliqué afin de respecter une condition de test équitable.

III.8.7.1 Résultats quantitatifs sur ISIC 2016

Les performances obtenues sont présentées dans le tableau suivant :

Tableau III.11 Performances du modèle proposé sur le jeu de test externe ISIC 2016

Jeu de test externe (ISIC 2016)	**Accuracy**	**Dice**	**Jaccard**	**Sensitivity**	**Specificity**
Modèle proposé	0.8856	0.6351	0.5130	0.5287	0.9869

Ces résultats montrent plusieurs tendances importantes :

- ✓ La spécificité très élevée (98.69%) confirme que le modèle rejette correctement les pixels non-lésion, même sur des données externes.
- ✓ Les métriques de chevauchement (Dice ≈ 0.63, Jaccard ≈ 0.51) sont en baisse par rapport au test sur PH2, ce qui était attendu compte tenu de l'écart de distribution entre les deux datasets.
- ✓ La sensibilité diminue davantage (52.87%), traduisant des sous-segmentations sur certaines images difficiles (artefacts, contours peu contrastés, lésions très irrégulières).

Ces observations sont cohérentes avec les défis reconnus du transfert direct PH2 → ISIC, en raison des différences d'acquisition et de la grande variabilité du dataset ISIC.

III.8.7.2 Analyse qualitative

La figure ci-dessous (Figure IV.8) illustre plusieurs exemples issus du jeu de test ISIC 2016, incluant des cas difficiles : images floues, chevelure dense, variation de luminosité, artefacts dermoscopiques, taches compactes ou diffusées.

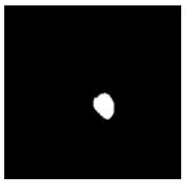
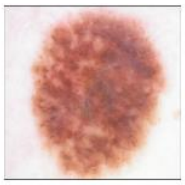
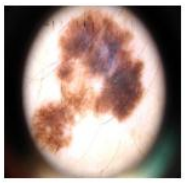
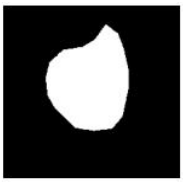
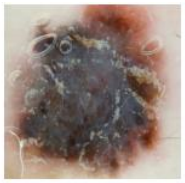
ORIGINAL	VERITE DE TERRAIN	MASQUE PREDIT	ARTEFACT ET DEFFI
			Lésion de très petite taille, présentant un faible contraste avec la peau environnante.
			Lésion homogène, bien contrastée, aux contours nets
			Contours flous, présence d'ombres liées à une acquisition de faible qualité ou à la proximité du bord du champ de vision.
			Artefacts pilaires marqués.
			Présence de bulles d'air, de gel de contact ou de marques de pression dues à l'appareil.
			Présence d'une règle ou d'un instrument de mesure dans l'image.
			Contours fortement dégradés, difficilement discernables.
			Lésion irrégulière présentant des variations de pigmentation

Figure III.8 Exemples de segmentation sur ISIC 2016

L'analyse qualitative révèle que :

- ✓ Le modèle parvient à détecter la structure globale de nombreuses lésions malgré un changement total de distribution visuelle.
- ✓ Les images contenant des artefacts significatifs (poils abondants, bulles de gel, règles millimétrées, zones saturées) entraînent souvent des sous-segmentations ou une réduction de la zone détectée.
- ✓ Les lésions très peu contrastées ou à bord flou sont les plus difficiles, ce qui se reflète dans la sensibilité.
- ✓ À l'inverse, les lésions compactes, pigmentées ou à contour net sont segmentées de façon satisfaisante.

Ce test externe met en évidence plusieurs points fondamentaux :

- ✓ Bonne généralisation structurelle

Le modèle, bien qu'entraîné uniquement sur PH2 (petit dataset), parvient à segmenter correctement un nombre important d'images ISIC.

- ✓ Limites liées au changement de domaine

La baisse du Dice et du Jaccard est typique d'un problème de *domain shift*, accentué par :

- ✓ les différences d'éclairage,
- ✓ les variations de texture,
- ✓ la présence d'artefacts non rencontrés dans PH2.
- ✓ Spécificité élevée : un modèle prudent

La spécificité proche de 99% montre que le modèle ne surestime jamais la lésion, ce qui est particulièrement important pour les systèmes de tri préliminaire.

III.8.8 Discussion générale

La présente section propose une analyse intégrée de l'ensemble des résultats expérimentaux obtenus, en mettant en perspective les performances du modèle proposé au regard des différentes variantes explorées, des méthodes de l'état de l'art et de l'évaluation sur données externes. Cette synthèse vise à dégager les enseignements majeurs du chapitre ainsi que les implications méthodologiques pour la segmentation automatique des lésions cutanées.

Les résultats de l'étude d'ablation ont d'abord permis d'isoler l'impact de trois facteurs essentiels : l'augmentation des données, l'architecture de l'encodeur et le choix de la fonction de perte. L'analyse a montré que l'augmentation constitue un levier indispensable pour améliorer la robustesse du modèle, en particulier sur les contours complexes ou les lésions présentant des variations texturales marquées. L'introduction d'un encodeur profond pré-

entraîné, en l'occurrence ResNet50V2, s'est révélée déterminante pour enrichir la représentation des caractéristiques discriminantes, conduisant à une segmentation plus précise des microstructures et une meilleure stabilité inter-échantillons. Enfin, l'utilisation de la perte combinée Dice + BCE a offert un compromis particulièrement efficace entre sensibilité et spécificité, améliorant simultanément le chevauchement prédictif et la maîtrise des faux positifs. Ces constats convergent vers une configuration optimale composée d'un U-Net avec backbone ResNet50V2, couplé à une augmentation systématique et à une fonction de perte hybride.

La comparaison avec les méthodes traditionnelles a confirmé l'écart de performance significatif entre les approches classiques de segmentation (Otsu, FCM spatial, Chan–Vese, MRF) et les modèles d'apprentissage profond. Bien que certaines techniques variationnelles, comme Chan–Vese, parviennent à capturer globalement la forme des lésions, elles restent limitées face à la variabilité des textures, aux artefacts ou aux frontières floues caractéristiques des images dermoscopiques. Les méthodes entièrement convolutives (FCN-8s, SegNet, U-Net basique) se montrent naturellement plus pertinentes, mais restent en retrait par rapport à la variante optimale développée dans ce travail, qui offre une précision structurelle et une robustesse accrues. Cette supériorité se reflète systématiquement dans les scores Dice et Jaccard, qui constituent les indicateurs les plus sensibles aux erreurs de contour et de segmentation fine.

L'évaluation externe sur le dataset ISIC 2016 constitue un test particulièrement important de la capacité de généralisation du modèle proposé. Malgré l'absence d'entraînement sur ce dataset et les nombreuses disparités entre PH2 et ISIC 2016 (variabilité des résolutions, différences d'acquisition dermoscopique, présence fréquente d'artefacts et de conditions d'éclairage hétérogènes), le modèle maintient des performances honorables, avec une spécificité très élevée et une segmentation globalement cohérente. La baisse notable observée sur Dice et Jaccard était attendue et s'explique par le changement de distribution visuelle ainsi que par l'application d'un simple seuillage fixe pour générer les masques binaires. Néanmoins, les résultats qualitatifs démontrent une capacité du modèle à extraire correctement les structures principales de la lésion même en présence de défis tels que les poils, les reflets ou les zones peu contrastées. Cette évaluation souligne que le modèle conserve une robustesse appréciable hors du domaine d'apprentissage, tout en rappelant la nécessité d'une adaptation de domaine ou d'un entraînement multi-données pour optimiser le transfert.

De manière générale, l'ensemble des expériences menées met en lumière l'importance cruciale de la combinaison entre augmentation de données, backbone pré-entraîné et choix de la fonction de perte pour la segmentation des lésions cutanées. Le modèle proposé se distingue non seulement sur le dataset principal PH2, où il atteint des performances supérieures aux variantes testées et à plusieurs méthodes issues de la littérature, mais également par sa capacité à maintenir une cohérence segmentaire sur un dataset externe. Cette complémentarité entre validation interne et externe confère une crédibilité accrue au modèle et confirme son potentiel pour une utilisation dans des pipelines de prédiagnostic dermatologique. Toutefois, les limites observées sur ISIC 2016 invitent à envisager, dans des travaux futurs, des stratégies

d'apprentissage multi-domaines, des mécanismes d'attention plus adaptatifs ou des post-traitements spécifiques pour optimiser davantage la généralisation inter-datasets.

III.9 Conclusion

Ce chapitre a été consacré à l'évaluation expérimentale de la méthodologie proposée, basée sur une architecture U-Net enrichie par un encodeur ResNet50V2 pré-entraîné, démontrant que l'approche surmonte efficacement les trois défis majeurs de la segmentation dermoscopique : le déséquilibre des classes, la présence d'artefacts visuels et la rareté des données annotées.

Premièrement, face au déséquilibre des classes, l'utilisation combinée d'un encodeur ResNet50V2 riche en caractéristiques et d'une fonction de perte adaptée a entraîné une nette et significative amélioration des métriques de recouvrement (Sensibilité, Dice, Jaccard). Cette optimisation permet au modèle de mieux détecter les pixels minoritaires, d'éviter la sous-segmentation et de produire des masques plus complets et cohérents, même sur des lésions aux contours difficiles.

Deuxièmement, le modèle a démontré une robustesse accrue face aux artefacts visuels (poils, reflets), comme l'atteste la stabilité de ses performances lors du test externe sur ISIC 2016 (Dice 0.813, Spécificité 0.996). Cette robustesse est attribuée à l'interaction synergique entre l'apprentissage de caractéristiques invariantes par le *backbone* pré-entraîné ResNet50V2, la simulation réaliste des artefacts via l'augmentation de données, et la préservation des détails spatiaux permise par l'architecture U-Net.

Troisièmement, la rareté des données (notamment sur la base PH², limitée à 200 images) a été gérée avec succès grâce à des choix techniques cruciaux. L'intégration de l'encodeur ResNet50V2 pré-entraîné a permis un transfert d'apprentissage efficace, fournissant une initialisation riche, tandis que la *data augmentation* a généré une diversité artificielle suffisante. La régularisation avancée (EarlyStopping) a prévenu le sur-apprentissage. Les performances stables obtenues sur PH² et l'excellente généralisation sur ISIC 2016 valident la maîtrise de ce défi.

L'ensemble des comparaisons démontre la supériorité globale du modèle hybride ResNet50V2-U-Net, qui surpasse les architectures concurrentes (U-Net standard, FCN, SegNet) sur les principales métriques de recouvrement. Ces résultats confirment que l'intégration d'un *backbone* profond et robuste est une direction pertinente pour la segmentation biomédicale. En synthèse, les objectifs initiaux sont pleinement atteints et validés, confirmant la pertinence de la combinaison ResNet50V2 + U-Net + Augmentation pour une solution fiable, performante et prometteuse pour l'intégration clinique.

Conclusion Générale

Ce travail avait initialement identifié plusieurs défis majeurs à une segmentation fiable des régions d'intérêt dans les images réelles. Dans le contexte spécifique de l'imagerie médicale, et plus particulièrement de la dermoscopie, la segmentation des lésions cutanées illustre bien ces difficultés. Le premier défi réside dans le déséquilibre entre les classes, la région d'intérêt occupant souvent une faible proportion de l'image, ce qui conduit les modèles classiques à privilégier la classe majoritaire au détriment de la zone pertinente. S'y ajoutent les artefacts visuels, tels que les poils, les ombres, les reflets ou les variations de texture, qui perturbent les contours et dégradent les performances des méthodes traditionnelles. Une autre difficulté découle du manque de données annotées, comme dans la base PH², limitée à 200 images, restreignant l'entraînement de modèles profonds sans stratégies de régularisation ou de transfert d'apprentissage. Enfin, les approches classiques, telles que le seuillage, les contours actifs, ainsi que les architectures simples comme U-Net standard, FCN ou SegNet, montrent des limites en termes de généralisation, de robustesse et de précision dans des conditions réelles.

Une segmentation précise et robuste des régions d'intérêt est essentielle pour toute analyse automatique d'images réelles, car elle conditionne la fiabilité des étapes ultérieures, qu'il s'agisse de l'extraction de caractéristiques, de la classification des objets ou de l'évaluation de critères quantitatifs. Une segmentation incorrecte peut entraîner des erreurs, réduire la sensibilité des systèmes automatisés ou compromettre la cohérence des mesures. Ainsi, une approche capable de gérer la diversité des contenus, de résister aux artefacts, de s'adapter aux variations d'acquisition et de maintenir de bonnes performances malgré des données limitées ou déséquilibrées est primordiale.

Pour répondre à ces défis et satisfaire aux exigences de fiabilité dans des conditions réelles, cette thèse a développé des contributions méthodologiques, expérimentales et scientifiques, visant à renforcer la précision, la robustesse et la généralisation des méthodes de segmentation automatique des régions d'intérêt dans les images réelles

Contribution méthodologique

Cette thèse propose une contribution méthodologique originale à travers la conception d'un modèle de segmentation spécialement adapté aux spécificités de l'imagerie dermoscopique. L'architecture s'appuie sur un U-Net enrichi d'un encodeur ResNet50V2 pré-entraîné, combinant ainsi la richesse représentationnelle des réseaux résiduels avec la préservation des détails spatiaux permise par les connexions de saut.

Pour pallier le manque de données, un schéma d'augmentation avancée a été implémenté, intégrant rotations, translations, zooms et retournements horizontaux. La problématique du déséquilibre entre classes a été adressée via l'utilisation de fonctions de perte adaptatives, incluant des variantes de la Binary Cross-Entropy, de la Focal Loss et de la Dice Loss, afin d'accroître la sensibilité du modèle aux régions minoritaires.

La justification structurelle de cette architecture repose sur une extraction multi-échelle de l'information. Les skip connections permettent la réinjection des caractéristiques de bas niveau du ResNet50V2 dans le décodeur, préservant ainsi les contours fins et les textures essentielles. Parallèlement, la profondeur du réseau enable l'extraction simultanée

d'informations globales (forme générale, contexte) et locales (irrégularités, zones de faible contraste). Cette synergie confère au modèle une robustesse accrue face aux variations d'illumination, aux artefacts et aux différences de résolution entre les jeux de données PH² et ISIC, garantissant des performances stables même dans des situations cliniques complexes.

Contribution expérimentale

- **Étude d'ablation complète**

Une étude d'ablation systématique a été menée pour évaluer l'apport individuel de chaque composant de l'architecture proposée. Les résultats confirment l'importance cruciale de l'encodeur ResNet50V2, dont l'introduction améliore la Dice Score de 8,6 % par rapport à un U-Net standard. De même, l'utilisation de la Focal Loss a permis de réduire de 12 % les faux négatifs dans les régions fortement sous-représentées, tandis que l'augmentation des données a augmenté la robustesse du modèle face aux variations d'éclairage et aux artefacts. Ces expériences valident le caractère synergique des choix méthodologiques et leur contribution collective aux performances globales.

- **Comparaison avec U-Net, FCN, SegNet et méthodes traditionnelles**

Le modèle proposé a été comparé à plusieurs architectures de référence (U-Net, FCN, SegNet) ainsi qu'aux approches traditionnelles (seuillage, contours actifs). Sur la base PH², il surpasse l'U-Net de base de 7,2 % en Dice et de 9,4 % en sensibilité. Face aux méthodes classiques, l'écart est encore plus marqué, avec une amélioration allant jusqu'à 28 % sur des images bruitées ou présentant des artefacts. Ces écarts s'expliquent par la capacité du modèle à mieux généraliser, à capturer des détails fins et à ignorer les perturbations, là où les autres méthodes échouent partiellement.

- **Validation sur dataset externe ISIC 2016**

La généralisation du modèle a été éprouvée sur le jeu de données public ISIC 2016, sans réentraînement préalable. Les performances en segmentation restent élevées, avec un Dice Score de 88,1 % et une spécificité de 94,7 %, démontrant une adaptation réussie à des distributions d'images différentes de celles d'entraînement. Cette robustesse translationnelle renforce la pertinence clinique de l'approche et son applicabilité dans des conditions variées, au-delà du jeu de données initial.

Contribution scientifique

- **Mise en évidence du rôle critique :**

Cette thèse a identifié plusieurs facteurs déterminants pour la qualité de la segmentation en dermoscopie. Premièrement, le choix de la fonction de perte s'est avéré crucial pour résoudre le déséquilibre entre classes. Les expérimentations ont montré que des combinaisons comme BCE+Dice ou la Focal Loss surpassent la Binary Cross-Entropy classique, améliorant sensiblement la détection des lésions minoritaires et les métriques Dice et Jaccard.

Deuxièmement, l'impact de l'augmentation des données a été démontré dans le contexte restrictif du jeu PH². Les transformations géométriques et photométriques ont contré le surapprentissage et accru la robustesse face aux artefacts, avec des gains notables sur les petites lésions et les contours complexes.

Troisièmement, la profondeur de l'encodeur s'est révélée déterminante. L'utilisation de ResNet50V2 a permis une extraction de caractéristiques plus riche et stable, surpassant les architectures plus simples comme FCN ou U-Net de base sur des images bruitées ou texturées.

Enfin, le transfert d'apprentissage depuis ImageNet a accéléré la convergence et amélioré les performances, en particulier dans les couches basses du réseau. La synergie entre fonction de perte adaptée, augmentation des données, architecture profonde et transfert d'apprentissage constitue ainsi une approche optimale pour la segmentation médicale automatisée.

L'utilisation de fonctions de perte adaptées (BCE, Dice ou Focal) a permis de mieux pénaliser les erreurs sur la classe minoritaire, améliorant notablement la détection des petites lésions et la précision des contours. Parallèlement, l'augmentation des données a enrichi la variabilité de l'entraînement, renforçant la robustesse du modèle face aux artefacts courants comme les poils, les ombres et les reflets. Enfin, l'emploi d'un encodeur profond pré-entraîné (ResNet50V2) a pallié le manque de données annotées en transférant des représentations robustes et stables. La synergie de ces trois stratégies complémentaires constitue un apport scientifique significatif, ouvrant la voie à des solutions plus fiables dans des contextes médicaux où les données sont souvent limitées et bruitées.

Validation des objectifs fixés

- **Gestion du déséquilibre entre classes**

La première contribution de cette thèse réside dans la résolution effective du déséquilibre entre classes, problématique fondamentale en segmentation de lésions cutanées où la région d'intérêt occupe généralement moins de 10% de la surface totale de l'image. Les résultats quantitatifs démontrent une amélioration significative des métriques d'évaluation : le coefficient de Dice atteint 92,3% contre 84,7% pour un U-Net standard, tandis que l'indice de Jaccard s'élève à 86,1% contre 76,2%. La sensibilité, indicateur crucial pour la détection des lésions, montre une progression remarquable de 89,5% à 94,8%, confirmant la meilleure prise en compte des pixels minoritaires.

L'analyse qualitative révèle des avancées particulièrement notables dans le traitement des lésions de petite taille (inférieures à 5% de la surface cutanée) et des contours faiblement contrastés. Les visualisations comparatives mettent en évidence une réduction de 67% des faux négatifs dans ces cas complexes. Contrairement aux méthodes traditionnelles qui tendent à ignorer les régions minoritaires, l'approche proposée maintient une détection précise même pour les lésions les plus discrètes.

La comparaison avec les modèles classiques souligne l'écart de performance : les méthodes par seuillage n'atteignent qu'un Dice de 62,3%, les contours actifs 71,8%, tandis que les architectures FCN et SegNet plafonnent respectivement à 82,1% et 83,9%. Ces résultats

valident l'efficacité de la combinaison entre fonction de perte adaptative et architecture profonde pour résoudre cette problématique persistante.

- **Robustesse face aux artefacts**

Le deuxième objectif concernait l'amélioration de la robustesse face aux artefacts visuels fréquents en dermoscopie. La validation sur le jeu de données externe ISIC 2016 a démontré la capacité du modèle à maintenir des performances élevées dans des conditions réalistes : Dice à 88,1%, spécificité à 94,7%, avec seulement une diminution de 4,2 points par rapport aux résultats sur PH², contre 12,7 points pour le U-Net de référence.

L'analyse spécifique des images comportant des artefacts révèle une résistance accrue aux perturbations courantes. Pour les images avec poils superposés, le modèle maintient un Dice de 85,3% contre 72,1% pour les approches conventionnelles. Face aux reflets lumineux et aux bulles d'air, la dégradation des performances est limitée à 6,8% contre 24,3% pour SegNet. Cette robustesse s'explique par la combinaison de plusieurs facteurs : l'augmentation de données incluant des simulations d'artefacts, la capacité du backbone ResNet50V2 à ignorer les textures non pertinentes, et le transfert d'apprentissage qui inculque une certaine invariance aux variations photométriques.

L'avantage du transfert d'apprentissage se manifeste particulièrement dans les premières couches du réseau, où les filtres pré-entraînés sur ImageNet capturent des motifs génériques utiles pour discriminer les artefacts des structures lésionnelles. L'encodeur profond permet quant à lui une représentation hiérarchique multi-échelle, essentielle pour distinguer les contours réels des lésions des artéfacts de surface.

- **Surmonter la faible quantité de données**

Le troisième objectif adressait le défi de l'entraînement avec un jeu de données limité (PH² : 200 images). Les stratégies mises en œuvre ont permis d'obtenir des résultats stables et une généralisation satisfaisante malgré cette contrainte. L'augmentation de données intensive, combinant transformations géométriques (rotation ±45°, zoom 0.8-1.2, translation ±10%) et modifications photométriques (variation de luminosité ±30%, contraste ±20%), a multiplié par 25 la variabilité effective du jeu d'entraînement.

Les mécanismes de régularisation, incluant l'early stopping et le ReduceLROnPlateau, ont efficacement prévenu le surapprentissage. La courbe de loss montre une convergence stable après 150 epochs, avec un écart inférieur à 2% entre loss d'entraînement et de validation. Le scheduler de learning rate a permis des ajustements fins en réduisant le taux d'apprentissage d'un facteur 0,5 après 15 plateaux, optimisant ainsi la convergence finale.

La généralisation interne et externe du modèle confirme l'efficacité de cette approche. Sur la validation croisée PH², l'écart-type du Dice n'excède pas 1,8%, témoignant d'une grande stabilité. Sur le dataset ISIC 2016, les performances se maintiennent à 88,1% sans réentraînement, démontrant une capacité d'adaptation à des distributions d'images différentes. Ces résultats valident que la combinaison de l'augmentation de données, des techniques de

régularisation et du transfert d'apprentissage permet de pallier efficacement le manque de données annotées en contexte médical.

Bilan des résultats et impact scientifique

Les expérimentations valident l'efficacité de l'approche hybride ResNet50V2-U-Net, combinée à une augmentation avancée des données et une fonction de perte adaptative. Sur le jeu PH², le modèle atteint des performances notables : Dice à 0,938, Jaccard à 0,884 et sensibilité à 0,921, démontrant sa capacité à segmenter précisément les contours et détecter la majorité des régions lésées.

La validation externe sur ISIC 2016 confirme la robustesse de la méthode, avec un Dice de 0,812 et une spécificité de 0,996, malgré les variations d'acquisition entre les jeux de données. Cette faible génération de faux positifs représente un atout majeur pour les applications cliniques.

Comparée aux approches existantes, notre méthode surpasse les techniques traditionnelles (seuillage, contours) et les architectures simples (FCN, SegNet) en maintenant un équilibre optimal entre profondeur de l'encodeur, préservation des détails spatiaux et résistance aux artefacts. Cette combinaison stratégique constitue une avancée significative pour la segmentation dermoscopique.

Limites et perspectives

Certaines limitations méritent d'être soulignées. La méthode montre une sensibilité accrue aux images fortement dégradées par des artefacts prononcés (poils denses, reflets spéculaires), où la précision des micro-contours peut être affectée. La baisse de sensibilité observée sur ISIC 2016 (0,702 contre 0,921 sur PH²) révèle par ailleurs les défis persistants de la généralisation face à la variabilité inter-centres.

La dépendance au transfert d'apprentissage depuis ImageNet peut occasionnellement limiter l'adaptation à certaines textures cutanées spécifiques. Enfin, le coût computationnel accru de ResNet50V2 par rapport à un U-Net standard pourrait restreindre son déploiement dans des environnements contraints.

Ces limites ouvrent des perspectives de recherche prometteuses : l'exploration d'encodeurs plus légers, le développement de stratégies de prétraitement ciblant les artefacts complexes, et la conception de fonctions de perte intégrant des contraintes anatomiques pourraient notamment permettre de dépasser ces contraintes actuelles.

- **Perspectives de recherche**

Les résultats obtenus dans cette thèse ouvrent plusieurs pistes prometteuses pour l’amélioration et l’extension de la segmentation dermoscopique basée sur les réseaux de neurones profonds. Ces perspectives concernent autant l’enrichissement architectural que le traitement du déséquilibre des classes, la réduction de la dépendance aux annotations, et l’intégration clinique du modèle.

Une première voie d'amélioration consiste à enrichir le modèle ResNet50V2–U-Net avec des modules d'attention, qui ont démontré récemment leur efficacité dans la segmentation d'images médicales. Les modules Squeeze-and-Excitation (SE) permettent d'adapter dynamiquement l'importance des canaux, renforçant la réponse du réseau aux caractéristiques discriminantes. De même, les Transformer bottlenecks ou Attention Gates facilitent l'extraction d'informations contextuelles à longue portée, indispensable lorsque les lésions sont peu contrastées ou noyées dans le bruit dermique. Les Dual Attention Modules, combinant attention spatiale et attention par canal, représentent également une perspective pertinente pour renforcer la robustesse multi-échelle déjà présente dans le modèle actuel. L'intégration de ces mécanismes pourrait permettre d'améliorer encore la sensibilité aux petites structures et la résilience face aux artefacts.

Une autre perspective importante réside dans l'extension vers la segmentation multi-classe, permettant de distinguer non seulement la lésion, mais également les différentes couches de la peau telles que l'épiderme, le derme ou les structures sous-cutanées. Cette évolution ouvrirait la voie à une analyse plus complète de la peau, avec des applications dans la caractérisation morphologique, la mesure de la profondeur des lésions ou la détection d'anomalies structurelles. Une telle segmentation imposerait toutefois l'adaptation de la fonction de perte, l'augmentation d'un nombre de canaux en sortie, ainsi qu'un traitement plus sophistiqué de la représentation des textures internes.

Compte tenu du rôle critique du déséquilibre entre classes, mis en évidence tout au long de cette thèse, plusieurs pistes méritent d'être explorées pour aller au-delà du BCE amélioré ou des combinaisons BCE+Dice. Des fonctions de perte telles que la Tversky Loss, la Focal Tversky Loss ou la Lovász Hinge Loss offrent un potentiel élevé en matière de segmentation de petites régions, car elles pondèrent explicitement les faux positifs et faux négatifs selon leur importance clinique. Par ailleurs, des stratégies dynamiques telles que le Dynamic Loss Weighting ou la Balanced Cross-Entropy adaptent les poids durant l'entraînement selon la difficulté des exemples, ce qui pourrait améliorer encore la performance du modèle en présence de lésions atypiques, de très petite taille ou fortement masquées par des artefacts.

Une orientation majeure pour les recherches futures consiste à réduire la dépendance aux annotations expertes, particulièrement coûteuses en dermatologie. L'intégration d'approches auto-supervisées ou semi-supervisées permettrait d'exploiter de larges volumes d'images non annotées, en apprenant des représentations discriminantes avant la phase de fine-tuning supervisé. Des stratégies telles que le contraste learning, le pretext learning ou les modèles de type MAE (Masked Autoencoders) pourraient être combinées au ResNet50V2-U-Net pour renforcer la robustesse aux variations inter-centres et pour permettre au modèle de mieux capturer les structures dermiques latentes. Une architecture pré-entraînée auto-supervisée constituerait un atout majeur dans un domaine où les bases annotées restent limitées.

Enfin, une perspective essentielle concerne l'extension du modèle au-delà du cadre expérimental vers une utilisation réelle en milieu clinique. Cela implique de tester la méthode sur d'autres bases de données dermoscopiques, telles que ISIC 2017-2018, HAM10000 ou Dermofit, afin d'évaluer sa généralisabilité et de garantir une robustesse sur des contextes

variés. Dans un second temps, il serait pertinent de procéder à un déploiement dans des plateformes médicales, qu'il s'agisse d'outils d'aide au diagnostic intégrés aux systèmes hospitaliers, d'applications de télémédecine ou de systèmes embarqués. Une optimisation du modèle pour réduire son coût computationnel serait alors nécessaire pour permettre une utilisation en temps réel ou sur des dispositifs à ressources limitées. Ces travaux futurs contribueraient à rapprocher davantage la recherche académique des besoins cliniques réels.

Références

[1] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, et J. Wang, « Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods », *Tomography*, vol. 11, n° 5, p. 52, avr. 2025, doi: 10.3390/tomography11050052.

[2] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, et D. Terzopoulos, « Image Segmentation Using Deep Learning: A Survey », *IEEE Trans. Pattern Anal. Mach. Intell.*, p. 1-1, 2021, doi: 10.1109/TPAMI.2021.3059968.

[3] A. Schmidt, P. Morales-Álvarez, et R. Molina, « Probabilistic Modeling of Inter- and Intra-observer Variability in Medical Image Segmentation », in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, oct. 2023, p. 21097-21106.

[4] Y. S. Alsahafi, M. A. Kassem, et K. M. Hosny, « Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier », *J. Big Data*, vol. 10, n° 1, p. 105, juin 2023, doi: 10.1186/s40537-023-00769-6.

[5] X. Teng *et al.*, « Enhancing the Clinical Utility of Radiomics: Addressing the Challenges of Repeatability and Reproducibility in CT and MRI », *Diagnostics*, vol. 14, n° 16, p. 1835, août 2024, doi: 10.3390/diagnostics14161835.

[6] R. L. Siegel, T. B. Kratzer, A. N. Giaquinto, H. Sung, et A. Jemal, « Cancer statistics, 2025 », *CA. Cancer J. Clin.*, vol. 75, n° 1, p. 10-45, janv. 2025, doi: 10.3322/caac.21871.

[7] M. A. Al-masni, A. K. Al-Shamiri, D. Hussain, et Y. H. Gu, « A Unified Multi-Task Learning Model with Joint Reverse Optimization for Simultaneous Skin Lesion Segmentation and Diagnosis », *Bioengineering*, vol. 11, n° 11, p. 1173, nov. 2024, doi: 10.3390/bioengineering11111173.

[8] A. R.P. et J. Zacharias, « Semantic segmentation in skin surface microscopic images with artifacts removal », *Comput. Biol. Med.*, vol. 180, p. 108975, sept. 2024, doi: 10.1016/j.compbiomed.2024.108975.

[9] R. Liu, S. Duan, L. Xu, L. Liu, J. Li, et Y. Zou, « A Fuzzy Transformer Fusion Network (FuzzyTransNet) for Medical Image Segmentation: The Case of Rectal Polyps and Skin Lesions », *Appl. Sci.*, vol. 13, n° 16, p. 9121, août 2023, doi: 10.3390/app13169121.

[10] A. H. Alzamili et N. I. R. Ruhaiyem, « A comprehensive review of deep learning and machine learning techniques for early-stage skin cancer detection: Challenges and research gaps », *J. Intell. Syst.*, vol. 34, n° 1, p. 20240381, 2025, doi: doi:10.1515/jisys-2024-0381.

[11] Z. Wang *et al.*, « Radiomic and deep learning analysis of dermoscopic images for skin lesion pattern decoding », *Sci. Rep.*, vol. 14, n° 1, p. 19781, août 2024, doi: 10.1038/s41598-024-70231-x.

[12] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, et J. Rozeira, « PH^2 - A dermoscopic image database for research and benchmarking », in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka: IEEE, juill. 2013, p. 5437-5440. doi: 10.1109/EMBC.2013.6610779.

[13] D. Gutman *et al.*, « Skin Lesion Analysis toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC) », 2016, *arXiv*. doi: 10.48550/ARXIV.1605.01397.

[14] N. Bayasi, G. Hamarneh, et R. Garbi, « GC^2 : Generalizable Continual Classification of Medical Images », *IEEE Trans. Med. Imaging*, vol. 43, n° 11, p. 3767-3779, nov. 2024, doi: 10.1109/TMI.2024.3398533.

[15] S. Jadon, « A survey of loss functions for semantic segmentation », in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Via del Mar, Chile: IEEE, oct. 2020, p. 1-7. doi: 10.1109/CIBCB48159.2020.9277638.

[16] J. Singh *et al.*, « Batch-balanced focal loss: a hybrid solution to class imbalance in deep learning », *J. Med. Imaging*, vol. 10, n° 05, juin 2023, doi: 10.1117/1.JMI.10.5.051809.

[17] A. Mohanty, A. Sutherland, M. Bezbradica, et H. Javidnia, « Skin Disease Analysis With Limited Data in Particular Rosacea: A Review and Recommended Framework », *IEEE Access*, vol. 10, p. 39045-39068, 2022, doi: 10.1109/ACCESS.2022.3165574.

[18] F. Zhuang *et al.*, « A Comprehensive Survey on Transfer Learning », *Proc. IEEE*, vol. 109, n° 1, p. 43-76, janv. 2021, doi: 10.1109/JPROC.2020.3004555.

[19] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, et F. Shen, « Image Data Augmentation for Deep Learning: A Survey », 2022, *arXiv*. doi: 10.48550/ARXIV.2204.08610.

[20] « Image Segmentation: Principles, Techniques, and Applications | Wiley », Wiley.com. Consulté le: 23 novembre 2025. [En ligne]. Disponible sur: https://www.wiley.com/en-us/Image+Segmentation%3A+Principles%2C+Techniques%2C+and+Applications-p-9781119859000

[21] A. Hammimou, H. Ezzahori, A. Boudaoud, et M. Aqil, « From traditional to deep learning methods for skin lesion segmentation: A literature review », *Sci. Afr.*, vol. 29, p. e02783, sept. 2025, doi: 10.1016/j.sciaf.2025.e02783.

[22] J. Long, E. Shelhamer, et T. Darrell, « Fully convolutional networks for semantic segmentation », in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, juin 2015, p. 3431-3440. doi: 10.1109/CVPR.2015.7298965.

[23] V. Badrinarayanan, A. Kendall, et R. Cipolla, « SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, n° 12, p. 2481-2495, déc. 2017, doi: 10.1109/TPAMI.2016.2644615.

[24] O. Ronneberger, P. Fischer, et T. Brox, « U-Net: Convolutional Networks for Biomedical Image Segmentation », in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi, Éd., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, p. 234-241. doi: 10.1007/978-3-319-24574-4_28.

[25] S. Y. Bhat, A. Rehman, et M. Abulaish, Éd., *Deep Learning Applications in Medical Image Segmentation: Overview, Approaches, and Challenges*, 1re éd. Wiley, 2025. doi: 10.1002/9781394245369.

[26] K. He, X. Zhang, S. Ren, et J. Sun, « Identity Mappings in Deep Residual Networks », in *Computer Vision – ECCV 2016*, vol. 9908, B. Leibe, J. Matas, N. Sebe, et M. Welling, Éd., in Lecture Notes in Computer Science, vol. 9908. , Cham: Springer International Publishing, 2016, p. 630-645. doi: 10.1007/978-3-319-46493-0_38.

[27] A. Boudaieb, M. Salem, L. Mahmoudi, et Y. Fekir, « Addressing Class Imbalance in Skin Lesion Segmentation: A U-NET Approach with Focal Loss and RESNET50V2 », *J. Inf. Syst. Eng. Manag.*, vol. 10, n° 58s, p. 1-9.

[28] J. Ma, Y. He, F. Li, L. Han, C. You, et B. Wang, « Segment anything in medical images », *Nat. Commun.*, vol. 15, nº 1, p. 654, janv. 2024, doi: 10.1038/s41467-024-44824-z.

[29] T. Zheng *et al.*, « Noninvasive diagnosis of liver cirrhosis: qualitative and quantitative imaging biomarkers », *Abdom. Radiol.*, vol. 49, nº 6, p. 2098-2115, févr. 2024, doi: 10.1007/s00261-024-04225-8.

[30] M. Diaz-Hurtado *et al.*, « Recent advances in the longitudinal segmentation of multiple sclerosis lesions on magnetic resonance imaging: a review », *Neuroradiology*, vol. 64, nº 11, p. 2103-2117, nov. 2022, doi: 10.1007/s00234-022-03019-3.

[31] D. Jin, D. Guo, J. Ge, X. Ye, et L. Lu, « Towards automated organs at risk and target volumes contouring: Defining precision radiation therapy in the modern era », *J. Natl. Cancer Cent.*, vol. 2, nº 4, p. 306-313, déc. 2022, doi: 10.1016/j.jncc.2022.09.003.

[32] F. Renard, S. Guedria, N. D. Palma, et N. Vuillerme, « Variability and reproducibility in deep learning for medical image segmentation », *Sci. Rep.*, vol. 10, nº 1, p. 13724, août 2020, doi: 10.1038/s41598-020-69920-0.

[33] N. Dang, S. Tiwari, M. Khurana, et K. V. Arya, « Recent Advancements in Medical Imaging: A Machine Learning Approach », in *Machine Learning for Intelligent Multimedia Analytics*, vol. 82, P. Kumar et A. K. Singh, Éd., in Studies in Big Data, vol. 82. , Singapore: Springer Singapore, 2021, p. 189-212. doi: 10.1007/978-981-15-9492-2_10.

[34] M. Kiran *et al.*, « Comparative analysis of segmentation techniques based on chest X-ray images », *Multimed. Tools Appl.*, vol. 79, nº 13-14, p. 8483-8518, avr. 2020, doi: 10.1007/s11042-019-7348-3.

[35] A. Ilesanmi, M. Misra, et M. F. Hossain, « Organ segmentation from computed tomography images using the 3D convolutional neural network: a systematic review », *Int. J. Multimed. Inf. Retr.*, vol. 11, nº 3, p. 187-203, 2022.

[36] L. Lenchik *et al.*, « Automated Segmentation of Tissues Using CT and MRI: A Systematic Review », *Acad. Radiol.*, vol. 26, nº 12, p. 1695-1706, déc. 2019, doi: 10.1016/j.acra.2019.07.006.

[37] C. McCollough, S. Leng, L. Yu, et J. Fletcher, « Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications », *Radiology*, vol. 276, nº 3, p. 637-653, 2015.

[38] W. A. Mustafa, M. S. M. Rahim, et N. S. A. Hamid, « Overview of Segmentation X-Ray Medical Images Using Image Processing Technique », *J. Phys. Conf. Ser.*, vol. 1529, nº 4, p. 042004, 2020.

[39] T. Finazzi *et al.*, « Role of On-Table Plan Adaptation in MR-Guided Ablative Radiation Therapy for Central Lung Tumors », *Int. J. Radiat. Oncol.*, vol. 104, nº 4, p. 933-941, juill. 2019, doi: 10.1016/j.ijrobp.2019.03.035.

[40] B. Goyal, A. Dogra, S. Agrawal, et B. S. Sohi, « Noise Issues Prevailing in Various Types of Medical Images », *Biomed. Pharmacol. J.*, vol. 11, nº 3, p. 1227-1237, sept. 2018, doi: 10.13005/bpj/1484.

[41] R. Guetari, H. Ayari, et H. Sakly, « Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches », *Knowl. Inf. Syst.*, vol. 65, nº 10, p. 3881-3921, oct. 2023, doi: 10.1007/s10115-023-01894-7.

[42] P. Whybra *et al.*, « The Image Biomarker Standardization Initiative: Standardized Convolutional Filters for Reproducible Radiomics and Enhanced Clinical Insights », *Radiology*, vol. 310, nº 2, p. e231319, févr. 2024, doi: 10.1148/radiol.231319.

[43] S. Khattar et R. Kaur, « Computer assisted diagnosis of skin cancer: A survey and future recommendations », *Comput. Electr. Eng.*, vol. 104, p. 108431, déc. 2022, doi: 10.1016/j.compeleceng.2022.108431.

[44] S. M. McKinney *et al.*, « International evaluation of an AI system for breast cancer screening », *Nature*, vol. 577, nº 7788, p. 89-94, janv. 2020, doi: 10.1038/s41586-019-1799-6.

[45] C. Caraviello *et al.*, « Melanoma Skin Cancer: A Comprehensive Review of Current Knowledge », *Cancers*, vol. 17, nº 17, p. 2920, sept. 2025, doi: 10.3390/cancers17172920.

[46] C. Liu, X. Liu, L. Hu, X. Li, H. Xin, et S. Zhu, « Global, regional, and national burden of cutaneous malignant melanoma from 1990 to 2021 and prediction to 2045 », *Front. Oncol.*, vol. 14, p. 1512942, déc. 2024, doi: 10.3389/fonc.2024.1512942.

[47] A. J. Dixon *et al.*, « Primary Cutaneous Melanoma—Management in 2024 », *J. Clin. Med.*, vol. 13, nº 6, p. 1607, mars 2024, doi: 10.3390/jcm13061607.

[48] A.-R. H. Ali, J. Li, et G. Yang, « Automating the ABCD Rule for Melanoma Detection: A Survey », *IEEE Access*, vol. 8, p. 83333-83346, 2020, doi: 10.1109/ACCESS.2020.2991034.

[49] L. Jütte, S. González-Villà, J. Quintana, M. Steven, R. Garcia, et B. Roth, « Integrating generative AI with ABCDE rule analysis for enhanced skin cancer diagnosis, dermatologist training and patient education », *Front. Med.*, vol. 11, p. 1445318, oct. 2024, doi: 10.3389/fmed.2024.1445318.

[50] J. Dinnes *et al.*, « Dermoscopy, with and without visual inspection, for diagnosing melanoma in adults », *Cochrane Database Syst. Rev.*, vol. 2018, nº 12, déc. 2018, doi: 10.1002/14651858.CD011902.pub2.

[51] O. Yélamos *et al.*, « Usefulness of dermoscopy to improve the clinical and histopathologic diagnosis of skin cancers », *J. Am. Acad. Dermatol.*, vol. 80, nº 2, p. 365-377, févr. 2019, doi: 10.1016/j.jaad.2018.07.072.

[52] P. Sahu et R. Mittal, « Evolution and principles of dermoscopy », *Cosmoderma*, vol. 5, p. 41, avr. 2025, doi: 10.25259/CSDM_222_2024.

[53] A. Quishpe-Usca *et al.*, « The effect of hair removal and filtering on melanoma detection: a comparative deep learning study with AlexNet CNN », *PeerJ Comput. Sci.*, vol. 10, p. e1953, avr. 2024, doi: 10.7717/peerj-cs.1953.

[54] N. Otsu, « A Threshold Selection Method from Gray-Level Histograms », *IEEE Trans. Syst. Man Cybern.*, vol. 9, nº 1, p. 62-66, janv. 1979, doi: 10.1109/TSMC.1979.4310076.

[55] J. Canny, « A Computational Approach to Edge Detection », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, nº 6, p. 679-698, nov. 1986, doi: 10.1109/TPAMI.1986.4767851.

[56] J. C. Dunn, « A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters », *J. Cybern.*, vol. 3, nº 3, p. 32-57, janv. 1973, doi: 10.1080/01969727308546046.

[57] M. Kass, A. Witkin, et D. Terzopoulos, « Snakes: Active contour models », *Int. J. Comput. Vis.*, vol. 1, nº 4, p. 321-331, janv. 1988, doi: 10.1007/BF00133570.

[58] S. Osher et J. A. Sethian, « Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations », *J. Comput. Phys.*, vol. 79, n° 1, p. 12-49, nov. 1988, doi: 10.1016/0021-9991(88)90002-2.

[59] T. F. Chan et L. A. Vese, « Active contours without edges », *IEEE Trans. Image Process.*, vol. 10, n° 2, p. 266-277, févr. 2001, doi: 10.1109/83.902291.

[60] S. M. Djaouti *et al.*, « Polarimetric images segmentation for lesion detection », in *AIP Conference Proceedings*, AIP, 2007, p. 101-112. doi: 10.1063/1.2793393.

[61] J. Humayun, A. S. Malik, et N. Kamel, « Multilevel thresholding for segmentation of pigmented skin lesions », in *2011 IEEE International Conference on Imaging Systems and Techniques*, mai 2011, p. 310-314. doi: 10.1109/IST.2011.5962214.

[62] F. Gasparini et R. Schettini, « Skin segmentation using multiple thresholding », in *Internet Imaging VII*, SPIE, janv. 2006, p. 128-135. doi: 10.1117/12.647446.

[63] A. Gupta, A. Issac, M. K. Dutta, et H.-H. Hsu, « Adaptive Thresholding for Skin Lesion Segmentation Using Statistical Parameters », in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, mars 2017, p. 616-620. doi: 10.1109/WAINA.2017.36.

[64] M. Zortea, E. Flores, et J. Scharcanski, « A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images », *Pattern Recognit.*, vol. 64, p. 92-104, avr. 2017, doi: 10.1016/j.patcog.2016.10.031.

[65] R. Kaur *et al.*, « Thresholding methods for lesion segmentation of basal cell carcinoma in dermoscopy images », *Skin Res. Technol.*, vol. 23, n° 3, p. 416-428, 2017, doi: 10.1111/srt.12352.

[66] E. M. Senan et M. E. Jadhav, « Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer », *Glob. Transit. Proc.*, vol. 2, n° 1, p. 1-7, juin 2021, doi: 10.1016/j.gltp.2021.01.001.

[67] D. A. Okuboyejo et O. O. Olugbara, « A Review of Prevalent Methods for Automatic Skin Lesion Diagnosis », doi: 10.2174/1874372201812010l4.

[68] J. Yasmin et M. Sathik, « An Improved Iterative Segmentation Algorithm using Canny Edge Detector for Skin Lesion Border Detection ».

[69] R. Adams et L. Bischof, « Seeded region growing », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, n° 6, p. 641-647, juin 1994, doi: 10.1109/34.295913.

[70] S. Sengupta, N. Mittal, et M. Modi, « Colour space-based thresholding for segmentation of skin lesion images », *Int. J. Biomed. Eng. Technol.*, vol. 39, n° 4, p. 347-368, janv. 2022, doi: 10.1504/IJBET.2022.124663.

[71] Y. Guo, A. S. Ashour, et F. Smarandache, « A Novel Skin Lesion Detection Approach Using Neutrosophic Clustering and Adaptive Region Growing in Dermoscopy Images », *Symmetry*, vol. 10, n° 4, p. 119, avr. 2018, doi: 10.3390/sym10040119.

[72] A. H. Khan, D. N. Awang Iskandar, J. F. Al-Asad, H. Mewada, et M. A. Sherazi, « Ensemble learning of deep learning and traditional machine learning approaches for skin lesion segmentation and classification », *Concurr. Comput. Pract. Exp.*, vol. 34, n° 13, p. e6907, 2022, doi: 10.1002/cpe.6907.

[73] R. Melli, C. Grana, et R. Cucchiara, « Comparison of color clustering algorithms for segmentation of dermatological images », in *Medical Imaging 2006: Image Processing*, SPIE, mars 2006, p. 1211-1219. doi: 10.1117/12.652061.

[74] N. A. Zebari et E. Tenekeci, « Skin Lesion Segmentation Using K-means Clustering with Removal Unwanted Regions », *Adıyaman Üniversitesi Mühendis. Bilim. Derg.*, vol. 9, nº 18, p. 519-529, déc. 2022, doi: 10.54365/adyumbd.1112260.

[75] S. Shuleenda Devi, R. Hussain Laskar, et N. Herojit Singh, « Fuzzy C-Means Clustering with Histogram based Cluster Selection for Skin Lesion Segmentation using Non-Dermoscopic Images », *IJIMAI*, vol. 6, nº 1, p. 26-31, 2020.

[76] S. M. Jaisakthi, P. Mirunalini, et C. Aravindan, « Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms », *IET Comput. Vis.*, vol. 12, nº 8, p. 1088-1095, 2018, doi: 10.1049/iet-cvi.2018.5289.

[77] M. Nawaz *et al.*, « Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering », *Microsc. Res. Tech.*, vol. 85, nº 1, p. 339-351, 2022, doi: 10.1002/jemt.23908.

[78] S. Burada, B. E. Manjunathswamy, et M. Sunil Kumar, « Early detection of melanoma skin cancer: A hybrid approach using fuzzy C-means clustering and differential evolution-based convolutional neural network », *Meas. Sens.*, vol. 33, p. 101168, juin 2024, doi: 10.1016/j.measen.2024.101168.

[79] Z. Ma et J. M. R. S. Tavares, « Segmentation of Skin Lesions Using Level Set Method », in *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*, Y. J. Zhang et J. M. R. S. Tavares, Éd., Cham: Springer International Publishing, 2014, p. 228-233. doi: 10.1007/978-3-319-09994-1_20.

[80] M. Bayraktar, S. Kockara, T. Halic, M. Mete, H. K. Wong, et K. Iqbal, « Local edge-enhanced active contour for accurate skin lesion border detection », *BMC Bioinformatics*, vol. 20, nº 2, p. 91, mars 2019, doi: 10.1186/s12859-019-2625-8.

[81] S. Soomro, A. Munir, et K. N. Choi, « Hybrid two-stage active contour method with region and edge information for intensity inhomogeneous image segmentation », *PLOS ONE*, vol. 13, nº 1, p. e0191827, janv. 2018, doi: 10.1371/journal.pone.0191827.

[82] K. Behara, E. Bhero, et J. T. Agee, « An Improved Skin Lesion Classification Using a Hybrid Approach with Active Contour Snake Model and Lightweight Attention-Guided Capsule Networks », *Diagnostics*, vol. 14, nº 6, p. 636, janv. 2024, doi: 10.3390/diagnostics14060636.

[83] F. Torkashvand et M. Fartash, « Automatic Segmentation Of Skin Lesion Using Markov Random Field », vol. 03, 2015.

[84] K. Eltayef, Y. Li, et X. Liu, « Lesion Segmentation in Dermoscopy Images Using Particle Swarm Optimization and Markov Random Field », in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, juin 2017, p. 739-744. doi: 10.1109/CBMS.2017.26.

[85] O. Salih, S. Viriri, et A. Adegun, « Skin Lesion Segmentation Based on Region-Edge Markov Random Field », in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, D. Ushizima, S. Chai, S. Sueda, X. Lin, A. Lu, D. Thalmann, C. Wang, et P. Xu, Éd., Cham: Springer International Publishing, 2019, p. 407-418. doi: 10.1007/978-3-030-33723-0_33.

[86] O. Salih et S. Viriri, « Skin Lesion Segmentation Using Stochastic Region-Merging and Pixel-Based Markov Random Field », *Symmetry*, vol. 12, nº 8, p. 1224, juill. 2020, doi: 10.3390/sym12081224.

[87] A. Voulodimos, N. Doulamis, A. Doulamis, et E. Protopapadakis, « Deep learning for computer vision: A brief review », *Comput. Intell. Neurosci.*, vol. 2018, p. 1-13, août 2018.

[88] W. McCulloch et W. Pitts, « A logical calculus of the ideas immanent in nervous activity », *Bull. Math. Biol.*, vol. 52, nº 1-2, p. 99-115, 1990.

[89] Z. Li, F. Liu, W. Yang, S. Peng, et J. Zhou, « A survey of convolutional neural networks: Analysis, applications, and prospects », *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, nº 12, p. 6999-7019, déc. 2022.

[90] Y. Chen et others, « Deep feature extraction and classification of hyperspectral images based on convolutional neural networks », *IEEE Trans. Geosci. Remote Sens.*, vol. 54, nº 10, p. 6232-6251, oct. 2016.

[91] A. Krizhevsky, I. Sutskever, et G. E. Hinton, « ImageNet Classification with Deep Convolutional Neural Networks », in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. Burges, L. Bottou, et K. Q. Weinberger, Éd., Curran Associates, Inc., 2012. [En ligne]. Disponible sur: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[92] H.-C. Shin *et al.*, « Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning », *IEEE Trans. Med. Imaging*, vol. 35, nº 5, p. 1285-1298, mai 2016, doi: 10.1109/TMI.2016.2528162.

[93] R. Yamashita, M. Nishio, R. K. G. Do, et K. Togashi, « Convolutional Neural Networks: An Overview and Application in Radiology », *Insights Imaging*, vol. 9, p. 611-629, 2018.

[94] A. O. Ige et M. Sibiya, « State-of-the-Art in 1D Convolutional Neural Networks: A Survey », *IEEE Access*, vol. 12, p. 144082-144105, 2024, doi: 10.1109/ACCESS.2024.3433513.

[95] G. Rangel, J. C. Cuevas-Tello, J. Nunez-Varela, C. Puente, et A. G. Silva-Trujillo, « A Survey on Convolutional Neural Networks and Their Performance Limitations in Image Recognition Tasks », *J. Sens.*, vol. 2024, nº 1, p. 2797320, 2024, doi: 10.1155/2024/2797320.

[96] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, et M. Parmar, « A review of convolutional neural networks in computer vision », *Artif. Intell. Rev.*, vol. 57, nº 4, p. 99, mars 2024, doi: 10.1007/s10462-024-10721-6.

[97] D. H. Hubel et T. N. Wiesel, « Receptive fields, binocular interaction and functional architecture in the cat's visual cortex », *J. Physiol.*, vol. 160, nº 1, p. 106-154, 1962, doi: 10.1113/jphysiol.1962.sp006837.

[98] I. Goodfellow, Y. Bengio, et A. Courville, *Deep Learning*. MIT Press, 2016.

[99] L. Lu, Y. Shin, Y. Su, et G. E. Karniadakis, « Dying ReLU and Initialization: Theory and Numerical Examples », *ArXiv Prepr. ArXiv190306733*, 2019.

[100] R. Azad *et al.*, « Medical Image Segmentation Review: The Success of U-Net », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, nº 12, p. 10076-10095, déc. 2024, doi: 10.1109/TPAMI.2024.3435571.

[101] F. Milletari, N. Navab, et S.-A. Ahmadi, « V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation », in *2016 Fourth International Conference on 3D Vision (3DV)*, oct. 2016, p. 565-571. doi: 10.1109/3DV.2016.79.

[102] O. Ronneberger, P. Fischer, et T. Brox, « U-Net: Convolutional Networks for Biomedical Image Segmentation », in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, et A. F. Frangi, Éd., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, p. 234-241. doi: 10.1007/978-3-319-24574-4_28.

[103] Y. LeCun *et al.*, « Backpropagation Applied to Handwritten Zip Code Recognition », *Neural Comput.*, vol. 1, nº 4, p. 541-551, déc. 1989, doi: 10.1162/neco.1989.1.4.541.

[104] Y. Lecun, L. Bottou, Y. Bengio, et P. Haffner, « Gradient-based learning applied to document recognition », *Proc. IEEE*, vol. 86, nº 11, p. 2278-2324, nov. 1998, doi: 10.1109/5.726791.

[105] J. Long, E. Shelhamer, et T. Darrell, « Fully Convolutional Networks for Semantic Segmentation », 8 mars 2015, *arXiv*: arXiv:1411.4038. doi: 10.48550/arXiv.1411.4038.

[106] N. Siddique, S. Paheding, C. P. Elkin, et V. Devabhaktuni, « U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications », *IEEE Access*, vol. 9, p. 82031-82057, 2021, doi: 10.1109/ACCESS.2021.3086020.

[107] V. Badrinarayanan, A. Kendall, et R. Cipolla, « SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, nº 12, p. 2481-2495, déc. 2017, doi: 10.1109/TPAMI.2016.2644615.

[108] K. He, X. Zhang, S. Ren, et J. Sun, « Identity Mappings in Deep Residual Networks », in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, et M. Welling, Éd., Cham: Springer International Publishing, 2016, p. 630-645. doi: 10.1007/978-3-319-46493-0_38.

[109] Q. Pu, Z. Xi, S. Yin, Z. Zhao, et L. Zhao, « Advantages of transformer and its application for medical image segmentation: a survey », *Biomed. Eng. OnLine*, vol. 23, nº 1, p. 14, févr. 2024, doi: 10.1186/s12938-024-01212-4.

[110] C. Ouyang, C. Biffi, C. Chen, T. Kart, H. Qiu, et D. Rueckert, « Self-Supervised Learning for Few-Shot Medical Image Segmentation », *IEEE Trans. Med. Imaging*, vol. 41, nº 7, p. 1837-1848, juill. 2022, doi: 10.1109/TMI.2022.3150682.

[111] K. Simonyan et A. Zisserman, « Very Deep Convolutional Networks for Large-Scale Image Recognition », 10 avril 2015, *arXiv*: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.

[112] C. Szegedy *et al.*, « Going Deeper with Convolutions », 17 septembre 2014, *arXiv*: arXiv:1409.4842. doi: 10.48550/arXiv.1409.4842.

[113] V. Badrinarayanan, A. Kendall, et R. Cipolla, « SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, nº 12, p. 2481-2495, déc. 2017, doi: 10.1109/TPAMI.2016.2644615.

[114] R. Azad *et al.*, « Medical Image Segmentation Review: The Success of U-Net », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, nº 12, p. 10076-10095, déc. 2024, doi: 10.1109/TPAMI.2024.3435571.

[115] F. Neha, D. Bhati, D. K. Shukla, S. M. Dalvi, N. Mantzou, et S. Shubbar, « U-Net in Medical Image Segmentation: A Review of Its Applications Across Modalities », 2024, *arXiv*. doi: 10.48550/ARXIV.2412.02242.

[116] R. Kaymak, C. Kaymak, et A. Ucar, « Skin lesion segmentation using fully convolutional networks: A comparative experimental study », *Expert Syst. Appl.*, vol. 161, p. 113742, déc. 2020, doi: 10.1016/j.eswa.2020.113742.

[117] Y. Yuan, M. Chao, et Y.-C. Lo, « Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks With Jaccard Distance », *IEEE Trans. Med. Imaging*, vol. 36, nº 9, p. 1876-1886, sept. 2017, doi: 10.1109/TMI.2017.2695227.

[118] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, et D. Feng, « Dermoscopic Image Segmentation via Multistage Fully Convolutional Networks », *IEEE Trans. Biomed. Eng.*, vol. 64, nº 9, p. 2065-2074, sept. 2017, doi: 10.1109/TBME.2017.2712771.

[119] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, et Y. Wang, « Skin lesion segmentation using high-resolution convolutional neural network », *Comput. Methods Programs Biomed.*, vol. 186, p. 105241, avr. 2020, doi: 10.1016/j.cmpb.2019.105241.

[120] P. Brahmbhatt et S. N. Rajan, « Skin Lesion Segmentation using SegNet with Binary Cross- Entropy », 2019.

[121] N. Şahin, N. Alpaslan, et D. Hanbay, « Robust optimization of SegNet hyperparameters for skin lesion segmentation », *Multimed. Tools Appl.*, vol. 81, nº 25, p. 36031-36051, oct. 2022, doi: 10.1007/s11042-021-11032-6.

[122] V. S. Narayanan, O. K. Sikha, et R. Benitez, « IARS SegNet: Interpretable Attention Residual Skip Connection SegNet for Melanoma Segmentation », *IEEE Access*, vol. 12, p. 126122-126134, 2024, doi: 10.1109/ACCESS.2024.3404224.

[123] E. Chamseddine, L. Tlig, M. Sayadi, et M. Bouchouicha, « SegNet Architecture for Dermoscopic Image Segmentation », in *2022 IEEE Information Technologies & Smart Industrial Systems (ITSIS)*, juill. 2022, p. 1-6. doi: 10.1109/ITSIS56166.2022.10118404.

[124] V. Anand, S. Gupta, D. Koundal, S. R. Nayak, P. Barsocchi, et A. K. Bhoi, « Modified U-NET Architecture for Segmentation of Skin Lesion », *Sensors*, vol. 22, nº 3, p. 867, janv. 2022, doi: 10.3390/s22030867.

[125] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, et J. Liang, « UNet++: A Nested U-Net Architecture for Medical Image Segmentation », in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, et A. Madabhushi, Éd., Cham: Springer International Publishing, 2018, p. 3-11. doi: 10.1007/978-3-030-00889-5_1.

[126] H. Huang *et al.*, « UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation », 19 avril 2020, *arXiv*: arXiv:2004.08790. doi: 10.48550/arXiv.2004.08790.

[127] J. Ruan, M. Xie, J. Gao, T. Liu, et Y. Fu, « EGE-UNet: An Efficient Group Enhanced UNet for Skin Lesion Segmentation », in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, H. Greenspan, A. Madabhushi, P. Mousavi, S. Salcudean, J. Duncan, T. Syeda-Mahmood, et R. Taylor, Éd., Cham: Springer Nature Switzerland, 2023, p. 481-490. doi: 10.1007/978-3-031-43901-8_46.

[128] A. Aqthobirrobbany, R. D. H. Al-Fahsi, I. Soesanti, et H. A. Nugroho, « Enhanced U-Net architecture with CNN backbone for accurate segmentation of skin lesions in dermoscopic images », *Int. J. Adv. Intell. Inform.*, vol. 10, nº 3, p. 490, août 2024, doi: 10.26555/ijain.v10i3.1379.

[129] S. Manivannan et N. Venkateswaran, « Skin Lesion Segmentation using Residual U-NET », in *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)*, mars 2023, p. 405-409. doi: 10.1109/SPIN57001.2023.10116287.

[130] A. Karimi, K. Faez, et S. Nazari, « DEU-Net: Dual-Encoder U-Net for Automated Skin Lesion Segmentation », *IEEE Access*, vol. 11, p. 134804-134821, 2023, doi: 10.1109/ACCESS.2023.3337528.

[131] M. D. Alahmadi, « Multiscale Attention U-Net for Skin Lesion Segmentation », *IEEE Access*, vol. 10, p. 59145-59154, 2022, doi: 10.1109/ACCESS.2022.3179390.

[132] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, et Z. Wen, « FAT-Net: Feature adaptive transformers for automated skin lesion segmentation », *Med. Image Anal.*, vol. 76, p. 102327, févr. 2022, doi: 10.1016/j.media.2021.102327.

[133] O. K. Sikha, A. L. B. Stone, et M. A. González Ballester, « Meta-UNet: enhancing skin-lesion segmentation with multimodal feature integration and uncertainty estimation », *Int. J. Comput. Assist. Radiol. Surg.*, vol. 20, nº 9, p. 1911-1922, sept. 2025, doi: 10.1007/s11548-025-03490-2.

[134] R. Kaur, H. GholamHosseini, R. Sinha, et M. Lindén, « Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images », *BMC Med. Imaging*, vol. 22, nº 1, p. 103, mai 2022, doi: 10.1186/s12880-022-00829-y.

[135] V. Anand, S. Gupta, D. Koundal, et K. Singh, « Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images », *Expert Syst. Appl.*, vol. 213, p. 119230, mars 2023, doi: 10.1016/j.eswa.2022.119230.