

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université -Ain Temouchent -Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département Math-Informatique



Projet de Fin d'Études

Pour l'obtention du diplôme de Master en : **CYSIA**

Domaine : **Mathématiques et Informatique**

Filière : **Informatique**

Spécialité : **Cyber sécurité et intelligence artificielle**

Thème

Traitement des images médicales par Deep Learning : segmentation des polypes colorectaux pour l'aide au diagnostic

Réalisé par :

- 1) Melle Bentsabet Seloua Amira
- 2) Mr Boulefred Radouane

Devant le jury composé de :

Dr Beddad Fatima	MCB	UAT.B.B (Ain Témouchent)	Présidente
Mme Saidi Samira	MAA	UAT.B.B (Ain Témouchent)	Examinatrice
Dr Belgrana Fatima Zohra	MCA	UAT.B.B (Ain Témouchent)	Encadrante

Année Universitaire 2025/2026

Dédicaces

Avant toute chose, cette dédicace que j'ai écrite à trois heures du matin, après une journée épuisante, ayant senti que c'était le moment opportun pour l'écrire, alors que l'inspiration m'était venue avec des sentiments sincères.

Je tiens à remercier et à louer Dieu, le Tout-Puissant, pour chaque bienfait dont je jouis, pour chaque prière qu'il a exaucée, pour chaque travail que j'ai entrepris et dont j'ai parfois désespéré de venir à bout afin qu'Il m'accorde davantage de persévérance et de soutien pour le mener à terme et pour chaque personne qu'Il a placée sur mon chemin dans cette vie.

*J'ai achevé mon mémoire, qui en **2021** me semblait être un rêve lointain et difficile, mais qui, grâce à Dieu, se réalise aujourd'hui et devient réalité.*

*Je remercie **ma mère** pour son soutien, ses prières et sa présence à mes côtés depuis le début de mes études jusqu'à aujourd'hui. Je la remercie pour sa rigueur, qui a fait de moi l'une des meilleures de ma promotion à chacun de mes trois cycles d'études. Je n'avais pas conscience, à l'époque, que c'était cette rigueur qui me poussait à rester parmi les premières ; je ne l'ai compris que ces dernières années, à l'université : c'est à ma mère que je dois, avant tout, ce que j'ai accompli.*

*Je remercie **mon père** pour son soutien et pour nous avoir offert tous les moyens nécessaires à la vie et aux études. Je le remercie de ne jamais nous avoir laissés manquer de quoi que ce soit : tout ce que je demandais m'était apporté. J'étais, et je suis encore, la princesse de ma famille.*

*Je remercie mes frères **Abdelkader et Sidahmed**, mes sœurs **Mellek et Arzak** pour leur soutien et leur joie tout au long de mon parcours scolaire et universitaire, et je leur souhaite à leur tour beaucoup de réussite.*

*Je remercie également tous **mes enseignants**, à chaque étape de mes études, qui ont eu un rôle et une influence déterminants dans ce que j'ai accompli aujourd'hui.*

*Je remercie aussi une personne qui m'est très chère, qui m'a accompagnée et a éclairé mon chemin alors que j'avais l'impression de me noyer. Je le remercie de m'avoir conduite à bon port : il n'a pas été un simple ami, mais un mentor et un guide, qui m'a fait aimer ma spécialité, aimer mon université et aimer la recherche. À **Abdelillah Belhadj**, mon âme sœur : je te remercie pour chaque nuit où tu es resté éveillé pour me consoler dans mes moments difficiles.*

Je te remercie pour ta confiance en moi, et grâce à Dieu, j'ai été digne de cette confiance.

*Je remercie mes copines : **Fadoua** ,**Sara** et **Hanane** , pour leur soutien. Je remercie aussi l'université de m'avoir offert mes meilleures amies : **Safaa** , **Aya** , **Sara** , **Chaimaa** et **Asma** . Je vous remercie pour votre soutien, je n'oublierai jamais votre bienveillance, et je vous souhaite à toutes la réussite.*

*Je remercie toute personne ayant joué un rôle dans ce mémoire. Je remercie ma tante **Souhila** pour son soutien durant cette période. Je remercie le personnel du service de gastro-entérologie pour leur accueil, **Samra** et **Ilham** , ainsi que les médecins, le professeur **Madani** et le professeur **Boulenouar** . Je remercie également l'équipe du laboratoire d'anapathie pour leurs précieuses informations et leurs efforts pour m'aider, **Soulaf** et **Marwa** , car celui qui ne remercie pas les gens ne remercie pas Dieu.*

*Je tiens à remercier tout particulièrement mon enseignante, Madame **Belgrana** . Elle n'a pas été pour moi une simple encadrante, mais une guide et une véritable enseignante. Je la remercie du fond du cœur pour son soutien et ses encouragements : elle a été l'une des raisons qui m'ont aidée à déterminer ce que je veux pour mon avenir, et elle a été à l'origine de la réalisation de mon rêve et du début d'un nouveau parcours en tant que jeune chercheuse en intelligence artificielle médicale.*

Enfin, merci à vous tous pour ce soutien et cette aide.

Amira

Dédicaces

À mon cher père, pour sa sagesse, sa patience et son soutien inébranlable, qui ont toujours été pour moi une boussole dans les moments d'incertitude.

À ma mère, ma force et ma vie. Ton amour inconditionnel, ta patience silencieuse et tes prières constantes ont façonné chaque pas de mon parcours. Ce travail t'est dédié, car sans toi, rien n'aurait été possible.

*À mon frère **Islem**, à mes sœurs **Esmaa** et **Marwa**, pour leur soutien constant, leur tendresse et leur présence chaleureuse, qui ont toujours été un appui précieux dans mes moments de doute et de joie.*

À mes cousins adorés, sachez que, peu importe la distance ou le temps qui passe, vous aurez toujours une place immense dans mon cœur et mon soutien.

*À mes meilleurs amis **Ilyes** et **Abdssamie**, vous avez toujours été là, fidèles et présents, offrant soutien, écoute et joie à chaque étape de ce parcours. C'est grâce à vous que ce chemin a été plus léger et plus riche.*

Je vous dédie ce travail avec toute ma gratitude. Et à tous les autres membres de ma famille.

Radouane

Remerciements

Au terme de ce travail, nous tenons à exprimer notre profonde gratitude à toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

Nous adressons nos remerciements les plus sincères à notre encadrant, **Dr.Belgrana Fatima Zohra**, pour son soutien constant, ses conseils avisés et son suivi rigoureux tout au long de ce projet.

Nous remercions également les membres du jury **Dr.Beddad Fatima** et **Mme.Saidi Samira** pour l'honneur qu'ils nous font en acceptant d'évaluer ce travail, ainsi que l'ensemble des enseignants du Département de Mathématiques et Informatique pour la qualité de leur formation.

Enfin, nous exprimons toute notre reconnaissance à nos familles et à nos amis pour leurs encouragements et leur soutien indéfectible.

Résumé

Le cancer colorectal représente l'un des problèmes de santé publique les plus préoccupants à l'échelle mondiale, avec près de 1,9 million de nouveaux cas diagnostiqués chaque année. La coloscopie demeure l'examen de référence pour sa prévention, mais des taux de polypes manqués atteignant 22 % à 40 % persistent en raison de la grande variabilité morphologique des lésions et de la fatigue des opérateurs. Le présent mémoire présente CANet+ (Colorectal Attention Network Plus), une nouvelle architecture de Deep Learning dédiée à la segmentation des polypes colorectaux. Cette architecture encodeur-décodeur combinant un encodeur ConvNeXt-Tiny, un décodeur Attention U-Net, des modules d'attention CBAM, un module ASPP pour la modélisation multi-échelle et un mécanisme de supervision profonde. Évalué sur trois benchmarks de référence Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB selon un protocole de validation croisée à quatre plis (four-fold cross-validation), CANet+ atteint des performances comprises entre 0,92 et 0,96 en Dice et entre 0,87 et 0,92 en IoU, surpassant ou rivalisant avec les méthodes de l'état de l'art. Ces résultats démontrent la pertinence de l'architecture proposée et ouvrent des perspectives concrètes d'intégration de systèmes d'aide au diagnostic fondés sur le Deep Learning dans la pratique endoscopique clinique.

Mots-clés : Cancer colorectal, imagerie Endoscopique, segmentation de polypes, Deep Learning, Attention, ConvNeXt, CBAM, ASPP.

Abstract

Colorectal cancer is one of the most pressing global public health challenges, with nearly 1.9 million new cases diagnosed annually. Colonoscopy remains the gold standard for its prevention; however, polyp miss rates ranging from 22% to 40% persist due to the high morphological variability of lesions and operator fatigue. This thesis presents CANet+ (Colorectal Attention Network Plus), an encoder-decoder Deep Learning architecture combining a ConvNeXt-Tiny encoder, an Attention U-Net decoder, CBAM attention modules, an ASPP module for multi-scale context modeling, and a deep supervision mechanism. Evaluated on three standard benchmarks Kvasir-SEG, CVC-ClinicDB, and CVC-ColonDB using a four-fold cross-validation protocol, CANet+ achieves Dice scores between 0.92 and 0.96 and IoU values between 0.87 and 0.92,

demonstrating competitive performance against methods. These results open concrete prospects for integrating Deep Learning based computer aided diagnosis and segmentation systems into real world clinical endoscopic practice.

Keywords: Colorectal Cancer, Endoscopic Imaging, Polyp Segmentation, Deep Learning, Attention, ConvNeXt, CBAM, ASPP.

ملخص

يُعد سرطان القولون والمستقيم أحد أكثر التحديات الصحية العامة إلحاحاً على مستوى العالم، حيث يتم تشخيص ما يقرب من 1.9 مليون حالة جديدة سنوياً. ولا يزال تنظير القولون المعيار الذهبي للوقاية منه؛ ومع ذلك، لا تزال نسبة إغفال الأورام الحميدة تتراوح بين 22% و40% بسبب التباين المورفولوجي الكبير للآفات وإرهاق الطبيب. تقدم هذه الأطروحة CANet+ (شبكة الانتباه لسرطان القولون والمستقيم المحسنة)، وهي بنية تعلم عميق تعتمد على نموذج التشفير-فك التشفير، وتجمع بين مُشفر، ConvNeXt-Tiny وفك تشفير Attention، U-Net، ووحدات انتباه، CBAM، ووحدة ASPP لنمذجة السياق متعدد المقاييس، وآلية إشراف عميق. بعد تقييمها على ثلاثة معايير قياسية هي Kvasir-SEG وCVC-ClinicDB وCVC-ColonDB باستخدام بروتوكول التحقق المتقاطع ذي الأربعة طيات، حققت CANet+ درجات Dice تتراوح بين 0.92 و0.96 وقيم IoU تتراوح بين 0.87 و0.92، مما يدل على أداء تنافسي مقارنةً بالأساليب الأخرى. تفتح هذه النتائج آفاقاً ملهوسة لدمج أنظمة التشخيص والتجزئة بمساعدة الحاسوب القائمة على التعلم العميق في الممارسة السريرية التنظيرية الواقعية.

الكلمات المفتاحية: سرطان القولون والمستقيم، التصوير بالتنظير الداخلي، تجزئة الزوائد اللحمية، التعلم العميق، الانتباه (Attention)، ConvNeXt، CBAM، ASPP.

Table des matières

Introduction Générale	16
1 De l'image numérique à l'imagerie médicale : concepts et contexte clinique	18
1.1 Introduction	18
1.2 Fondamentaux de l'image numérique	18
1.2.1 Codage et types d'images	19
1.2.2 Résolution et échantillonnage	20
1.2.3 Formats de fichiers et compression	20
1.3 Traitement d'images médicales	21
1.3.1 Chaîne de traitement d'image	21
1.3.2 Prétraitement (Bas niveau)	21
1.3.3 Segmentation (Niveau intermédiaire)	22
1.3.4 Analyse et Interprétation (Haut niveau)	23
1.4 L'Imagerie Médicale	23
1.4.1 Panorama des principales modalités d'imagerie médicale	23
1.4.2 Principes de l'imagerie endoscopique	26
1.4.3 La coloscopie : procédure de référence	26
1.4.4 Caractéristiques des images endoscopiques coliques	28
1.5 Contexte clinique : visite au service de gastroentérologie de l'hôpital Ahmed Madaghri (Aïn Témouchent)	29
1.5.1 Stades de développement des polypes colorectaux selon la taille	31
1.6 Conclusion	32
2 État de l'Art en segmentation de Polypes par Deep Learning	33
2.1 Introduction	33
2.2 Deep Learning en imagerie médicale	33
2.2.1 Avènement des CNN en vision par ordinateur	34
2.2.2 Architectures convolutionnelles pour la segmentation	35
2.3 Évolution des approches de segmentation des polypes	37
2.3.1 Limites des approches traditionnelles	37
2.3.2 Le paradigme convolutionnel : de U-Net aux backbones modernes	37

2.3.3	Mécanismes d'attention : vers une modélisation sélective	39
2.3.4	La révolution des Vision Transformers et l'hybridation	39
2.4	Travaux connexes et positionnement	40
2.4.1	PraNet : Attention inverse pour le raffinement des frontières	40
2.4.2	Attention U-Net : Portes d'attention pour le raffinement spatial	41
2.4.3	META-Unet-Attention Transformer multi-échelle	42
2.5	Défis actuels : vers des modèles plus robustes	44
2.6	Conclusion	46
3	Approche proposée, résultats et discussions	47
3.1	Introduction	47
3.2	Contexte clinique et défis de la segmentation des polypes colorectaux	47
3.3	Base de données utilisée et prétraitement	49
3.3.1	Description	49
3.3.2	Prétraitement des Données	50
3.3.3	Augmentation des données	51
3.4	Approche proposée	53
3.4.1	Positionnement et motivations	53
3.4.2	Architecture Globale de notre approche de segmentation	54
3.4.3	Interprétation et analyse du modèle	55
3.5	Conception du modèle proposé CANet+	55
3.5.1	Méthodologie de conception	55
3.5.2	Architecture détaillée de CANet+	56
3.5.3	Fonction de Perte Combinée	61
3.5.4	Protocole d'Entraînement	62
3.6	Expérimentations et Résultats	63
3.6.1	Métriques d'Évaluation	63
3.6.2	Sélection de l'architecture de Deep Learning	64
3.6.3	Analyse de l'impact des différents composants du modèle proposé	65
3.6.4	Résultats Qualitatifs	67
3.6.5	Résultats Quantitatifs et comparaison	68
3.6.6	Règles de détermination du stade du polype	72
3.6.7	Discussion	73
3.7	Implémentation , outils utilisés et présentation de l'application	74
3.7.1	Matériel utilisé	74
3.7.2	Langages de programmation et technologies	74
3.7.3	Environnement de développement	75
3.7.4	Présentation de l'interface de l'application	76
3.8	Conclusion	85

Bibliographie

89

Liste des abréviations

Abréviation	Signification
4K	Ultra Haute Définition (3840×2160 pixels)
AG	Attention Gate (Porte d'attention)
ASPP	Atrous Spatial Pyramid Pooling
BBPS	Boston Bowel Preparation Scale - Échelle de qualité de préparation colique
BCE	Binary Cross-Entropy (Entropie croisée binaire)
CBAM	Convolutional Block Attention Module
CCD	Charge-Coupled Device - Capteur à transfert de charge
CIELAB	CIE 1976 L*a*b* - Espace colorimétrique de la Commission Internationale de l'Éclairage
CLE	Confocal Laser Endomicroscopy - Endomicroscopie confocale par laser
CLAHE	Contrast Limited Adaptive Histogram Equalization
CMOS	Complementary Metal-Oxide Semiconductor
CNN	Convolutional Neural Network - Réseau de neurones convolutifs
CT / TDM	Computed Tomography / Tomodensitométrie - Scanner
CVC-ClinicDB	Computer Vision Center - Clinical Colonoscopy Database
DBN	Deep Belief Network
DL	Deep Learning (Apprentissage profond)
DNN	Deep Neural Network
ETIS	Endoscopic Tissue Image Segmentation
FICE	Flexible Spectral Imaging Color Enhancement
FOV	Field of View - Champ de vision

Abréviation	Signification
GLCM	Gray-Level Co-occurrence Matrix
GELU	Gaussian Error Linear Unit
HD	Haute Définition (1920×1080 pixels)
HSL / TSL	Hue Saturation Luminance / Teinte Saturation Luminosité
IA	Intelligence Artificielle
IoU	Intersection over Union
IRM	Imagerie par Résonance Magnétique
Kvasir-SEG	Base de données endoscopiques pour la segmentation des polypes
MICI	Maladies Inflammatoires Chroniques de l'Intestin
ML	Machine Learning (Apprentissage automatique)
MSA	Multi-head Self-Attention (Auto-attention multi-têtes)
MTA	Multi-scale Transformer Attention
NBI	Narrow Band Imaging - Imagerie en bande étroite
NCI	National Cancer Institute
RA	Reverse Attention (Attention inverse)
ReLU	Rectified Linear Unit
ResNet	Residual Network
RGB / RVB	Red Green Blue / Rouge Vert Bleu - Espace couleur additif
SE	Squeeze-and-Excitation
TDA	Taux de Détection des Adénomes
TIC	Taux d'Intubation Cœcale
TR	Temps de Retrait (endoscope)
U-Net	U-shaped Network - Architecture pour la segmentation
VGG	Visual Geometry Group
ViT	Vision Transformer
WHO / OMS	World Health Organization / Organisation Mondiale de la Santé

Table des figures

1.1	Image binaire (seuil de valeur 100) (Anto59290, 2019)	19
1.2	Image en niveaux de gris (Anto59290, 2019)	19
1.3	Image en couleur (Anto59290, 2019)	20
1.6	Les différents types de segmentations appliquées aux polypes	22
1.7	Exemple d'une image radiographique (HUG, 2018)	24
1.8	Exemple d'une image scanner (CT) (Wikipédia, 2026)	24
1.9	Exemple d'une image IRM (Depositphotos, s.d.)	25
1.10	Exemple d'une image échographique (Biolane, s.d.)	25
1.11	Exemples d'images endoscopiques en lumière blanche (a) et en NBI (b) (Koessler et al., 2015)	26
1.12	Photo prise par nous-mêmes qui présente la salle de diagnostic (a) et son équipement (b, c) de l'hôpital Ahmed Madaghri (Aïn Témouchent).	31
2.1	Architecture générale d'un CNN : couches de convolution, d'activation ReLU, de pooling et de classification. (Singh, 2024)	36
2.2	Architecture U-Net encodeur-décodeur avec connexions de saut (skip connections) pour la segmentation biomédicale. (Elhaddar, s.d.)	37
2.3	Bloc ConvNeXt-Tiny (Vishwkarma, 2025).	39
2.4	Vision Transformers pour la segmentation : ViT, TransUNet et Swin-UNet (Vishwkarma, 2025).	40
2.5	Architecture PraNet : décodeur partiel (PA) et attention inverse (RA) en cascade pour la segmentation précise des contours de polypes (Fan et al., 2020).	41
2.6	Attention U-Net : architecture globale et détail du module Attention Gate (AG) pour le raffinement spatial des skip connections (Oktay et al., 2018).	42
2.7	Architecture META-Unet : module MTA dans chaque étape de l'encodeur pour une modélisation contextuelle globale et locale (Wu et al., 2023).	43
3.1	Anatomie du côlon et vue endoscopique d'un côlon sain. La coloscopie permet d'explorer l'intégralité du cadre colique et d'identifier les lésions suspectes. (Gastroheggp, s.d.)	48

3.2	Progression d'un polype colorectal sur 8 à 10 ans, de la lésion précancéreuse bénigne au carcinome invasif, illustrant l'importance du dépistage précoce. (Lecomte et al., 2026)	48
3.3	Diversité morphologique des polypes colorectaux observés en coloscopie : polypes sessiles, pédiculés, plans et lésions à faible contraste, illustrant la complexité de la segmentation automatique. (Benkalfate et al., 2024)	49
3.4	Distribution Polype/Fond	50
3.5	Pipeline d'augmentation appliqué au jeu de données Kvasir-SEG	51
3.6	Pipeline d'augmentation appliqué au jeu de données CVC-ClinicDB	52
3.7	Pipeline d'augmentation appliqué au jeu de données CVC-ColonDB	52
3.8	Architecture générale de notre approche proposée	54
3.9	Architecture générale d'un réseau encodeur-décodeur convolutif. (Vayssade, 2022)	56
3.10	Architecture générale de CANet+ : encodeur ConvNeXt-Tiny, module ASPP au goulot d'étranglement, décodeur Attention U-Net avec blocs CBAM et supervision profonde.	56
3.11	Architecture de l'encodeur ConvNeXt (Woo et al., 2023)	57
3.12	Module CBAM : attention canal (réseau MLP appliqué aux sorties du pooling moyen et maximum) suivie de l'attention spatiale (convolution 7×7). (Woo et al., 2018)	59
3.13	Module ASPP (Atrous Spatial Pyramid Pooling) : convolutions dilatées parallèles à taux appliqué sur la carte de caractéristiques d'entrée, capturant le contexte à différentes échelles de champ réceptif. (Pan, 2022)	60
3.14	Comparaison qualitative des prédictions des architectures avec les masques de vérité terrain sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB.	67
3.15	Comparaison qualitative des prédictions de CANet+ avec les masques de vérité terrain sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB	68
3.16	Comparaison des performances obtenues sur les 4 folds	69
3.17	Évolution des métriques d'entraînement sur Kvasir-SEG.	70
3.18	Évolution des métriques d'entraînement sur CVC-ClinicDB.	70
3.19	Évolution des métriques d'entraînement sur CVC-ColonDB.	70
3.20	Courbes ROC obtenues sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonD	71
3.21	Chaîne de traitement de l'image en phase d'inférence, du chargement jusqu'à la production des résultats cliniques et techniques.	75
3.22	Architecture logicielle de l'application ColonCare : noyau de traitement commun et double interface.	76
3.23	En-tête de l'application ColonCare Pro affichage du modèle et sélection de l'interface(a :mode clair, b : mode sombre)	77
3.24	Interface Médicale Onglet Gestion des patients : création et liste des patients.	77

3.25	Interface Médicale Onglet Acquisition & Segmentation.	78
3.26	Interface Médicale : Synthèse clinique générée automatiquement après segmentation (application des règles diagnostiques).	79
3.27	Interface Médicale : Génération du rapport clinique en PDF.	79
3.28	Rapport clinique PDF généré par ColonCare Pro conforme au modèle EPH Ahmed Madaghri.	80
3.29	Interface technique.	81
3.30	Interface Technique : (a) Métriques de segmentation et galerie GT/prédiction. (b) Table des métriques à 512×512 (CANet+).	81
3.31	Interface Technique : (a) Métriques de segmentation et courbe ROC. ,(b) Distribution des probabilités et heatmap de prédiction.	82
3.32	Interface Technique : Carte d'incertitude (MC Dropout) et activation Grad-CAM.	82
3.33	Interface Technique : Visualisation des Attention Gates de l'architecture Attention U-Net.	83
3.34	Interface Technique Augmentation Lab : menu de sélection des transformations.	84
3.35	Interface Technique : (a,b)présente Augmentation Lab(comparaison image originale et augmentée.)	85

Liste des tableaux

1.1	Table comparatif de quelques formats d'images	21
1.2	Catégories de taille des polypes et techniques de résection recommandées (Gao et al., 2023)	32
2.1	Comparaison de quelques méthodes récentes de segmentation des polypes colorectaux.	43
3.1	Volumétrie des datasets partition 4-fold CV.	50
3.2	Volumétrie des datasets entraînement et augmentation.	53
3.3	Niveaux de caractéristiques hiérarchiques de l'encodeur ConvNeXt-Tiny.	58
3.4	Hyperparamètres d'entraînement de notre modèle CANet+.	62
3.5	Comparaison progressive des architectures candidates sur Kvasir-SEG.	65
3.6	Étude d'ablation : contribution individuelle des composants de CANet+ (Dice $\pm \sigma$, IoU $\pm \sigma$) sur les trois jeux de données.	66
3.7	Résultats expérimentaux de CANet+ sur les jeux de données (Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB).	69
3.8	Résultats comparatifs de CANet+ avec quelques méthodes de l'état de l'art sur les jeux de données Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB	72

Introduction Générale

Le cancer colorectal représente l'un des défis de santé publique les plus pressants à l'échelle mondiale. Selon les données de l'Organisation Mondiale de la Santé, environ 1,9 million de nouveaux cas sont diagnostiqués chaque année et plus de 900 000 décès lui sont attribués, faisant de cette pathologie le troisième cancer le plus fréquent et le deuxième en termes de mortalité (**World Health Organization, 2022**). Ce fardeau considérable justifie le recours au dépistage précoce comme levier d'intervention prioritaire, d'autant que la carcinogenèse colorectale offre une fenêtre temporelle favorable : la voie adénome–carcinome, qui représente environ 70 % des cas, s'étend sur 10 à 15 ans, tandis que la voie festonnée, portée par les adénomes sessiles festonnés (SSA/P), progresse sur 5 à 10 ans. La coloscopie demeure l'examen de référence pour exploiter cette fenêtre, permettant la détection et la résection des lésions précancéreuses au cours d'une même procédure. Pourtant, malgré les progrès constants des systèmes endoscopiques, des études cliniques révèlent des taux de polypes manqués atteignant 22 à 28% pour les adénomes et jusqu'à 40 % pour les petites lésions de moins de 5 mm (**Levin et al., 2008**) ; (**Rex et al., 2015**). Ces chiffres s'expliquent par la fatigue des opérateurs lors d'examens prolongés, la complexité anatomique du côlon, et surtout la grande variabilité morphologique des polypes : certaines lésions planes et SSA/P présentent un contraste si faible avec la muqueuse environnante qu'elles demeurent difficiles à détecter en lumière blanche standard et représentent une source majeure de lésions manquées lors de la coloscopie.

C'est ce constat, confirmé lors d'une visite au service de gastroentérologie de l'hôpital Ahmed Madaghri d'Aïn Témouchent et d'échanges directs avec les praticiens et le laboratoire d'anatomopathologie, qui est à l'origine de ce travail. L'observation du terrain a mis en évidence un besoin non satisfait : un système capable d'assister l'endoscopiste en temps réel, sans se substituer à son jugement clinique, mais en signalant et délimitant automatiquement les lésions à risque.

Ainsi, l'objectif principal de notre projet de fin d'études est de concevoir une approche capable de segmenter avec précision les polypes colorectaux dans des images endoscopiques, en surmontant la variabilité morphologique et le faible contraste des lésions, tout en maintenant une efficacité computationnelle, avec une possible intégration en contexte clinique.

L'analyse de la littérature révèle que les méthodes de segmentation automatique ont considérablement progressé, depuis les descripteurs manuels classiques jusqu'aux architectures encodeur-décodeur modernes intégrant des mécanismes d'attention et des modules de

modélisation multi-échelle. Toutefois, un compromis persistant entre précision, généralisation et coût computationnelle demeure non résolu, et constitue précisément l'espace dans lequel s'inscrit notre contribution.

Pour répondre à cette problématique, nous avons proposé un modèle de Deep Learning que nous avons nommé CANet+ (Colorectal Attention Network Plus), une architecture hybride encodeur-décodeur dont la conception intègre de manière synergique plusieurs mécanismes complémentaires visant chacun à lever des limitations précises identifiées dans l'état de l'art. L'objectif est de concilier précision de segmentation, robustesse face à l'hétérogénéité des données et efficacité computationnelle dans un cadre unifié. Le modèle est entraîné et évalué selon un protocole rigoureux de validation croisée à quatre plis (four-fold cross-validation) sur trois bases de données publiques de référence suivantes : Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB, et les résultats obtenus, avec des scores Dice compris entre 0,92 et 0,96, confirment la pertinence et la compétitivité de l'approche proposée par rapport aux travaux de la littérature.

Le présent mémoire est structuré en trois chapitres. Le premier parcourt les notions indispensables concernant l'imagerie numérique et endoscopique. Le deuxième dresse un état de l'art de la segmentation des polypes par Deep Learning. Enfin, le troisième présente notre approche proposée y compris l'architecture CANet+, le protocole expérimental adopté ainsi que l'analyse des résultats obtenus, tout en discutant les perspectives d'intégration clinique et les limites identifiées.

Chapitre 1

De l'image numérique à l'imagerie médicale : concepts et contexte clinique

1.1 Introduction

L'imagerie médicale occupe une place essentielle dans le diagnostic moderne en permettant une visualisation non invasive des structures internes. Dans ce cadre, l'endoscopie assistée par l'intelligence artificielle constitue une avancée importante pour améliorer la détection des polypes colorectaux et réduire les lésions manquées (**Misra et al., 2021**).

Le cancer colorectal compte parmi les cancers les plus fréquents et les plus meurtriers à l'échelle mondiale, avec environ 1,9 million de nouveaux cas et plus de 900 000 décès recensés en 2022 (**World Health Organization, 2022**). Le dépistage précoce par coloscopie demeure l'une des stratégies les plus efficaces pour réduire la mortalité, grâce à la détection et à l'ablation des polypes précancéreux avant leur évolution maligne (**National Cancer Institute, 2024**). C'est dans ce contexte que l'analyse automatique des images endoscopiques s'impose comme un axe de recherche porteur, et que s'inscrit notre projet de fin d'études.

Ce chapitre est organisé comme suit : nous présentons d'abord les notions fondamentales de l'image numérique, puis les principales étapes du traitement d'images médicales. Nous décrivons ensuite les modalités d'imagerie médicale, avec une attention particulière portée à l'imagerie endoscopique, avant de conclure sur les enjeux du traitement automatique des images de polypes colorectaux.

1.2 Fondamentaux de l'image numérique

Une image numérique est une représentation discrète d'une scène bidimensionnelle, échantillonnée en un ensemble fini de points appelés pixels. Chaque pixel porte une valeur numérique codant son intensité lumineuse ou ses composantes colorimétriques, formant ainsi une matrice

rectangulaire qui constitue l'image complète.(Gonzalez & Woods, 2018)

1.2.1 Codage et types d'images

Selon le mode de codage des pixels, les images numériques peuvent être classées en plusieurs catégories. Parmi les plus utilisées en traitement d'images, on distingue les images binaires, les images en niveaux de gris et les images couleur.

1.2.1.1 Image binaire (noir et blanc)

Une image binaire est un type d'image dont les pixels ne peuvent prendre que deux valeurs : 0 (noir) et 1 (blanc) (voir Figure 1.1). Un seul bit par pixel suffit à coder ce type d'image, ce qui en fait le format le plus compact.



Figure 1.1 Image binaire (seuil de valeur 100) (Anto59290, 2019)

1.2.1.2 Image en niveaux de gris

Dans une image en niveaux de gris, chaque pixel est codé sur 8 bits (valeurs de 0 à 255), où 0 correspond au noir absolu et 255 au blanc. Ce format simplifie l'analyse quantitative tout en conservant l'information morphologique; il est largement utilisé en radiologie (voir Figure 1.2) (Dougherty, 2009).



Figure 1.2 Image en niveaux de gris (Anto59290, 2019)

1.2.1.3 Image couleur

L'image en couleur repose sur la combinaison de trois canaux chromatiques primaires : rouge, vert et bleu (modèle RVB ou RGB) (voir Figure 1.3). Chaque canal est codé sur 8 bits, offrant 256 niveaux d'intensité par composante, soit un espace de couleur de 16 777 216 teintes distincts (256^3).

Malgré ses avantages, le modèle RVB présente certaines limites : les composantes trichromatiques peuvent prendre des valeurs négatives, certaines couleurs perçues par l'œil humain ne sont pas reproductibles par addition des trois spectres, et deux stimuli visuellement identiques peuvent posséder des triplets (R, G, B) différents selon les conditions d'illumination. (**Gonzalez & Woods, 2018**); (**Wyszecki & Stiles, 2000**)



Figure 1.3 Image en couleur (**Anto59290, 2019**)

D'autres espaces colorimétriques sont utilisés selon les applications. Le CIELAB (L^* , a^* , b^*), défini par la Commission Internationale de l'Éclairage, est conçu pour correspondre à la perception visuelle humaine et calculer des distances colorimétriques perceptuellement uniformes. Le modèle HSL (Teinte, Saturation, Luminosité) est quant à lui plus intuitif pour la manipulation et la segmentation des couleurs en traitement d'images (**Wyszecki & Stiles, 2000**).

1.2.2 Résolution et échantillonnage

La résolution spatiale désigne la densité de pixels par unité de surface (ex. : pixels/pouce). Une résolution élevée permet de distinguer des détails fins, ce qui est essentiel en imagerie médicale pour identifier des lésions de petite taille (**Pizer et al., 1987**).

1.2.3 Formats de fichiers et compression

Un format d'image désigne la structure informatique utilisée pour stocker et encoder les données d'une image numérique. Il détermine notamment le type de compression appliqué (avec ou sans perte), la profondeur de couleur supportée et la compatibilité avec les différents logiciels de traitement. Le choix du format est crucial en imagerie médicale, où la fidélité de l'information doit être préservée. Le tableau ci-dessous présente une comparaison de quelques formats courants (**Hadjira, 2020**).

Table 1.1 Table comparatif de quelques formats d'images

Format	Type	Compression des données	Nombre de couleurs	Animation	Transparence
JPEG	matriciel	Oui, réglable (avec perte)	16 millions	Non	Non
JPEG2000	matriciel	Oui, avec ou sans perte	4 milliards	Oui	Oui
GIF	matriciel	Oui, sans perte	256 max (palette)	Oui	Oui
PNG	matriciel	Oui, sans perte	Palettisé (256 couleurs ou moins) ou 16 millions	Non	Oui (couche Alpha)
TIFF	matriciel	Compression ou pas avec ou sans pertes	de monochrome à 16 millions	Non	Oui (couche Alpha)
SVG	vectériel	compression possible	16 millions	Oui	Oui (par nature)

1.3 Traitement d'images médicales

Le traitement d'images médicales vise à transformer une image brute en information cliniquement exploitable : amélioration de la qualité visuelle, extraction de régions d'intérêt (polypes) et quantification de caractéristiques diagnostiques. (**Gonzalez & Woods, 2018**)

1.3.1 Chaîne de traitement d'image

Ces opérations s'organisent en trois niveaux d'abstraction croissante. Le bas niveau traite directement les pixels bruts (débruitage, correction d'éclairage). Le niveau intermédiaire effectue la transition du pixel vers la structure, par la segmentation des régions d'intérêt. Le haut niveau, enfin, s'oriente vers l'interprétation diagnostique et la prise de décision (**Gonzalez & Woods, 2018 ; Dougherty, 2009**).

1.3.2 Prétraitement (Bas niveau)

La qualité de l'image brute étant conditionnée par les paramètres d'illumination et les dispositifs optiques (**Mesejo et al., 2016**), plusieurs opérations de prétraitement sont appliquées avant toute analyse :

- **Débruitage** : Application de filtres (ex. : non-local means) pour réduire le bruit sans altérer les contours (**Buades et al., 2005**).
- **Correction d'éclairage** : Normalisation de l'illumination par égalisation d'histogramme adaptative (CLAHE) (**Pizer et al., 1987**).
- **Suppression des reflets** : Détection et interpolation des zones spéculaires via analyse de texture (**Chouchane, 2016**).

1.3.3 Segmentation (Niveau intermédiaire)

La segmentation est l'opération qui consiste à partitionner une image en régions homogènes selon des critères définis (couleur, texture, forme). Elle constitue une étape fondamentale du traitement d'images médicales, permettant de délimiter automatiquement les structures anatomiques ou les lésions d'intérêt clinique. Dans le contexte des polypes colorectaux, trois approches principales sont utilisées :

- **Segmentation sémantique** La segmentation sémantique est une tâche permettant de classer une image pixel par pixel. Dans notre contexte, elle vise à attribuer une étiquette binaire (polype ou fond) à l'aide d'architectures dédiées telles que le U-Net (**Ronneberger et al., 2015**) (voir figure 1.6(b)).
- **Segmentation par instance** : La segmentation par instance fournit une analyse plus granulaire que la simple détection d'objets. Elle permet de détecter et de délimiter individuellement chaque objet (exemple : polype) au sein d'une image (voir figure 1.6 (c)), en s'appuyant généralement sur des modèles de référence comme Mask R-CNN (**He et al., 2016**).
- **Segmentation panoptique** : La segmentation panoptique unifie les concepts de segmentation sémantique et d'instances pour générer une compréhension globale de la scène visuelle (**Kirillov et al., 2019**). Ainsi, le modèle ne se limite plus à la simple identification des classes de pixels, mais parvient à extraire les caractéristiques morphologiques exactes (forme et dimension) de chaque entité présente à l'image, comme l'illustre la figure 1.6(d) (**Díaz-Rodríguez, 2025**).

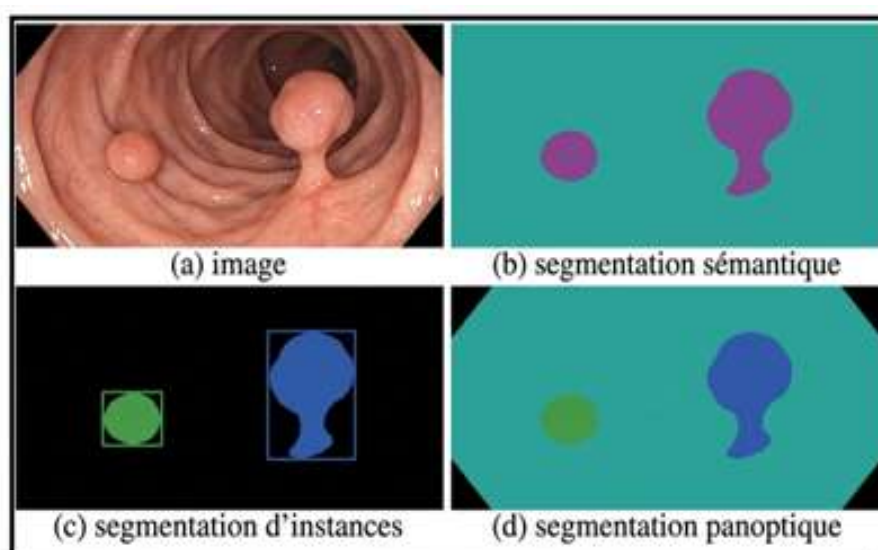


Figure 1.6 Les différents types de segmentations appliquées aux polypes

1.3.4 Analyse et Interprétation (Haut niveau)

Le haut niveau constitue l'étape finale de la chaîne de traitement, où l'on ne manipule plus des pixels mais des concepts. Il regroupe deux phases complémentaires : le post-traitement et l'interprétation.

Le post-traitement vise à affiner les résultats issus de la segmentation, à travers des opérations morphologiques (fermeture, remplissage de trous) pour régulariser les contours, et l'extraction de descripteurs morphologiques (surface, circularité) et texturaux (GLCM) (Moen et al., 2019).

La classification permet ensuite de spécifier le type de pathologie détectée : tumeurs, lésions inflammatoires ou excroissances tissulaires. Dans le cadre de notre étude, elle permet d'identifier le type de polype colorectal (adénomateux, hyperplasique ou inflammatoire).

L'interprétation transforme ces résultats en information cliniquement utile : score de risque histologique, recommandation thérapeutique, contribuant ainsi à l'aide à la décision médicale (Gonzalez & Woods, 2018);(Moen et al., 2019)).

1.4 L'Imagerie Médicale

L'imagerie médicale permet le dépistage précoce, le diagnostic différentiel et le suivi thérapeutique de nombreuses pathologies, en minimisant l'invasivité des procédures diagnostiques (Rumack et al., 2011). Elle repose sur différentes modalités d'acquisition, chacune exploitant un principe physique distinct, que nous présentons dans ce qui suit.

1.4.1 Panorama des principales modalités d'imagerie médicale

1.4.1.1 Radiographie

La radiographie est la technique d'imagerie médicale la plus ancienne et la plus répandue. Elle repose sur l'utilisation de rayons X capables de traverser les tissus mous et d'être partiellement absorbés par les structures denses comme les os. L'image obtenue résulte du contraste entre les différents tissus selon leur densité. Elle est particulièrement utilisée pour l'exploration du squelette, la détection de fractures et le diagnostic de pathologies pulmonaires (Lippincott Williams & Wilkins, 2012).



Figure 1.7 Exemple d'une image radiographique (HUG, 2018)

1.4.1.2 Scanner (CT : Computed Tomography)

Le scanner, ou tomodensitométrie (TDM), utilise des rayons X couplés à un système informatique pour acquérir des coupes axiales du corps et reconstruire des images en trois dimensions. Il permet d'obtenir des images volumétriques avec une excellente résolution spatiale, facilitant la détection et la caractérisation des lésions organiques (Lippincott Williams & Wilkins, 2012).



Figure 1.8 Exemple d'une image scanner (CT) (Wikipédia, 2026)

1.4.1.3 Imagerie par Résonance Magnétique (IRM)

L'IRM repose sur un champ magnétique intense et des ondes radio pour générer des images anatomiques en haute résolution, sans rayonnement ionisant. Elle constitue l'examen de référence pour l'étude des organes complexes, notamment le cerveau, la moelle épinière et les muscles (MSD Manuals, 2023).

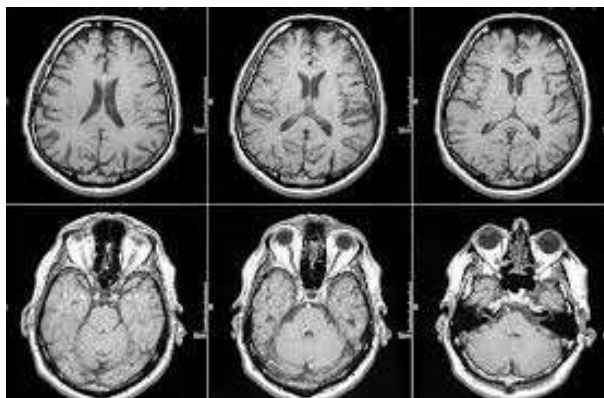


Figure 1.9 Exemple d'une image IRM (Depositphotos, s.d.)

1.4.1.4 Échographie

L'échographie utilise des ondes ultrasonores émises par un transducteur et réfléchies par les interfaces tissulaires pour produire des images dynamiques en temps réel. N'émettant aucun rayonnement ionisant, elle est adaptée à un large éventail de patients (Mersaoui & Fouiri, 2018).



Figure 1.10 Exemple d'une image échographique (Biolane, s.d.)

1.4.1.5 Imagerie endoscopique

L'endoscopie est une technique d'imagerie médicale permettant la visualisation directe des cavités internes du corps humain à l'aide d'un endoscope flexible muni d'une caméra et d'une source lumineuse. Elle est utilisée pour explorer le tube digestif, les voies respiratoires, et d'autres cavités. La coloscopie permet l'exploration directe du côlon avec possibilité de biopsie ou résection. La technique NBI (Narrow Band Imaging) utilise des longueurs d'onde spécifiques (415 nm, 540 nm) pour accentuer le contraste vasculaire et améliorer la détection des lésions précoces (Sano et al., 2009).

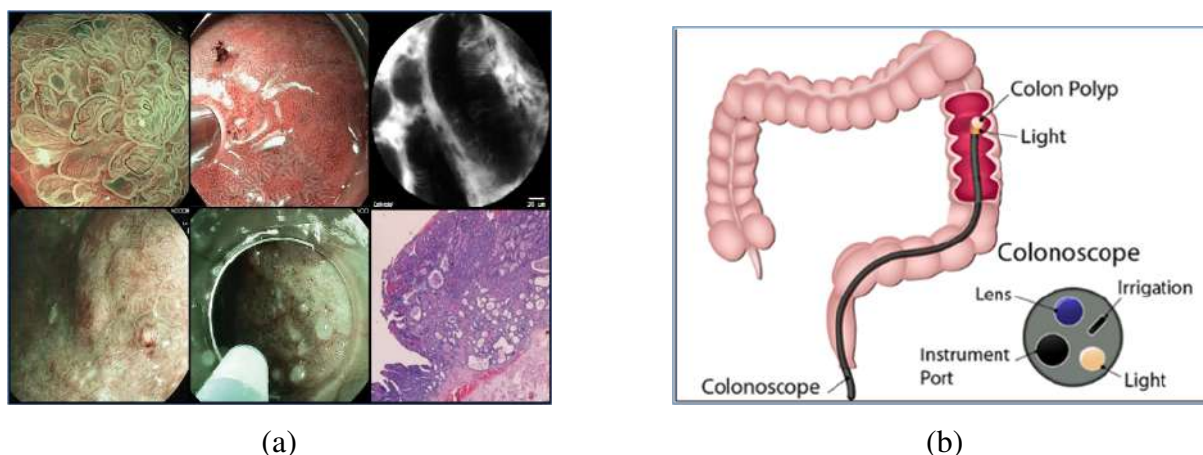


Figure 1.11 Exemples d'images endoscopiques en lumière blanche (a) et en NBI (b) (Koessler et al., 2015)

1.4.2 Principes de l'imagerie endoscopique

1.4.2.1 Composants et formation de l'image endoscopique

Un endoscope est un instrument optique flexible ou rigide, introduit dans une cavité naturelle pour l'observation directe des structures internes (Cotton et al., 2004). Il comprend un système optique grand-angle couplé à un capteur CCD ou CMOS haute définition, une source lumineuse LED/xénon transmise par fibres optiques, un canal opérateur pour les instruments chirurgicaux, et un système de contrôle directionnel (Krizhevsky et al., 2012). L'image est formée par la lumière réfléchiée par la muqueuse, numérisée par le capteur et traitée en temps réel par le processeur vidéo de la tour endoscopique () (Repici et al., 2019).

1.4.2.2 Caractéristiques techniques des endoscopes modernes

Les endoscopes actuels embarquent des capteurs haute définition (HD 1920×1080) ou ultra-haute définition (4K 3840×2160), permettant la visualisation de micro-structures muqueuses (cryptes, capillaires) essentielles au diagnostic (Byrne et al., 2019). À titre de comparaison, les appareils des années 2000 ne dépassaient pas une résolution de 850×480 pixels (Tajiri et al., 2012).

Le champ de vision (FOV) varie entre 140° et 170° , et des systèmes de magnification optique ou électronique ($\times 80$ à $\times 150$) permettent l'analyse fine des patterns muqueux in vivo (Rex et al., 2015); (Spada et al., 2012)

1.4.3 La coloscopie : procédure de référence

1.4.3.1 Déroulement et indications

La coloscopie est un examen endoscopique permettant l'exploration directe de l'ensemble du côlon, depuis la jonction ano-rectale jusqu'au cæcum, en passant par les différents segments co-

liques (sigmoïde, côlon descendant, transverse, côlon ascendant) (**Rex et al., 2015**). Elle constitue l'examen de référence pour :

- Le dépistage du cancer colorectal chez les sujets à risque moyen (dès 50 ans) ou élevé (antécédents familiaux, polypes antérieurs, maladies inflammatoires chroniques) (**Levin et al., 2008**).
- L'exploration diagnostique en cas de symptômes évocateurs (rectorragies, troubles du transit, syndrome anémique inexpliqué) (**Rex et al., 2015**).
- La résection thérapeutique des polypes découverts lors de l'examen, par polypectomie ou mucosectomie (**Zauber et al., 2012**).
- La surveillance post-polypectomie selon des intervalles définis par les recommandations en vigueur (**Spada et al., 2012**).

La préparation colique est une étape indispensable, consistant en l'ingestion de solutions laxatives (polyéthylène glycol, sodium picosulfate) la veille et le matin de l'examen, afin d'obtenir une vacuité colique satisfaisante permettant une visualisation optimale de la muqueuse (Repici et al., 2019).

1.4.3.2 Critères de qualité de la coloscopie

La qualité d'une coloscopie est évaluée par plusieurs indicateurs standardisés, dont les principaux sont :

- Le taux d'intubation cœcale (TIC) : pourcentage d'examens avec atteinte du cæcum, devant être supérieur à 95 % pour les examens de dépistage (**Rex et al., 2015**).
- Le temps de retrait (TR) : durée de l'examen lors du retrait de l'endoscope, recommandé à au moins 6 minutes pour assurer une inspection minutieuse de la muqueuse (**Barclay et al., 2006**).
- Le taux de détection des adénomes (TDA) : pourcentage de coloscopies révélant au moins un adénome, dont le seuil minimal recommandé est de 25 % pour les hommes et 15 % pour les femmes (**Rex et al., 2015**).
- La qualité de la préparation colique, évaluée par l'échelle de Boston (BBPS), dont un score ≥ 2 par segment est requis pour une coloscopie de qualité adéquate (Calderwood et al., 2014).

1.4.3.3 Technologies d'imagerie avancées

Afin d'améliorer la détection et la caractérisation des lésions, plusieurs technologies complémentaires ont été développées et intégrées aux systèmes endoscopiques modernes (Repici et al., 2019) :

a. NBI (Narrow Band Imaging) : filtre la lumière à 415 nm et 540 nm, longueurs d'onde correspondant aux pics d'absorption de l'hémoglobine, pour rehausser les capillaires superficiels et les structures cryptales. Elle permet la différenciation in vivo des lésions bénignes et malignes (Sano et al., 2009) ; (Hewett et al., 2012).

b. Chromoendoscopie virtuelle (i-Scan, FICE) : algorithmes de post-traitement numérique simulant l'effet d'une coloration, pour améliorer la détection des lésions planes et des adénomes festonnés sessiles (Bulois, 2011).

c. Chromoendoscopie au colorant : instillation d'indigo carmin ou de cristal violet pour révéler l'architecture muqueuse. Recommandée dans la surveillance des maladies inflammatoires chroniques de l'intestin (Kiesslich et al., 2003).

d. Endomicroscopie confocale par laser (CLE) : analyse histologique in vivo à l'échelle cellulaire ($\times 1000$), permettant une biopsie optique virtuelle. Réservée aux centres experts (Kiesslich et al., 2003).

1.4.4 Caractéristiques des images endoscopiques coliques

1.4.4.1 Particularités visuelles et artefacts

Les images endoscopiques coliques présentent des caractéristiques visuelles spécifiques qui les distinguent des autres types d'images médicales et qui représentent des défis particuliers pour le traitement automatique (Tajbakhsh et al., 2015) :

a. Reflets spéculaires (glare) : La présence de mucus et de liquide résiduel à la surface de la muqueuse génère des zones de sur-illumination (specular highlights), apparaissant comme des taches blanches brillantes pouvant masquer partiellement les structures sous-jacentes (Samaha et al., 2019). Ces artefacts sont d'autant plus fréquents que l'angle d'incidence entre la source lumineuse et la surface tissulaire est faible.

b. Flou de mouvement : Les mouvements péristaltiques du côlon et les mouvements involontaires de l'endoscope engendrent un flou cinétique affectant la netteté des images (Daubresse-Chasle, 2024). Ce phénomène est plus prononcé dans les segments coliques présentant une activité motrice importante (côlon sigmoïde, côlon transverse).

c. Ombres portées et zones non éclairées : La géométrie tubulaire du côlon et la présence de haustrations (bosselures) créent des zones d'ombre dans les angles morts du champ de vision (Bernal et al., 2012). Ces régions mal éclairées peuvent dissimuler des lésions, contribuant au taux de polypes manqués.

d. Variabilité de la distance focale : La distance entre l'objectif de l'endoscope et la muqueuse varie continuellement lors de l'examen, modifiant l'apparence et la taille apparente des lésions, ainsi que la profondeur de champ (Mesejo et al., 2016).

e. Résidus fécaux et bulles : Malgré la préparation colique, des résidus fécaux ou des bulles d'air persistent dans certains cas et peuvent être confondus avec des lésions polypeuses, ou masquer des structures pertinentes (Sikkandar et al., 2024).

f. Variabilité chromatique : La couleur de la muqueuse colique varie selon les individus, les segments anatomiques et les conditions d'éclairage. L'inflammation, la présence de vaisseaux dilatés ou les résidus hémorragiques modifient significativement l'aspect colorimétrique des images (Bernal et al., 2012).

1.4.4.2 Enjeux pour le traitement automatique des images

Ces particularités visuelles font de l'analyse automatique des images endoscopiques un problème particulièrement délicat. Les systèmes de traitement doivent gérer une forte variabilité intra- et inter-patient, de multiples sources d'artefacts, et la grande diversité morphologique et chromatique des polypes (Daubresse-Chasle, 2024).

De plus, la disponibilité de données annotées de qualité constitue un obstacle majeur au développement de systèmes d'apprentissage supervisé. L'annotation manuelle des polypes par des endoscopistes experts est une tâche chronophage et soumise à une variabilité inter-observateur non négligeable (Bernal et al., 2015). Des bases de données publics ont été constituées pour pallier cette difficulté, tels que ETIS-LaribPolypDB, CVC-ClinicDB (Bernal et al., 2015) et Kvasir-SEG (Jha et al., 2020).

1.5 Contexte clinique : visite au service de gastroentérologie de l'hôpital Ahmed Madaghri (Aïn Témouchent)

Lors de notre déplacement à l'hôpital Ahmed Madaghri d'Aïn Témouchent, nous avons eu l'opportunité d'intégrer l'environnement clinique du service de gastroentérologie et d'assister aux différentes étapes du processus diagnostique liées aux polypes colorectaux. En raison de contraintes techniques temporaires affectant le système vidéo de la salle d'endoscopie au moment de la visite, l'observation détaillée des images endoscopiques en temps réel n'a pas pu être réalisée. Néanmoins, les médecins du service ont fourni des explications approfondies sur le déroulement d'une coloscopie, les méthodes de détection utilisées en pratique ainsi que les critères d'identification des différentes lésions colorectales.

Les praticiens nous ont notamment présenté les principaux types de polypes rencontrés en routine clinique, leurs caractéristiques morphologiques, leur potentiel de transformation maligne

et les éléments permettant d'orienter le diagnostic. Une attention particulière a été portée aux polypes adénomateux, hyperplasiques et festonnés sessiles (SSA/P), qui représentent un enjeu majeur dans la prévention du cancer colorectal.

Notre visite a également inclus le laboratoire d'anatomopathologie, où nous avons pu comprendre le parcours des prélèvements après résection ou biopsie. Les spécialistes ont expliqué les différentes étapes d'analyse histopathologique permettant de déterminer la nature de la lésion (bénigne ou maligne), son type histologique, son degré de dysplasie ainsi que son éventuel stade d'évolution tumorale. Cette étape constitue la référence diagnostique permettant de confirmer les observations réalisées lors de l'examen endoscopique.

Ces échanges avec les professionnels de santé ont permis de mieux cerner les défis de la détection précoce, notamment pour les polypes de petite taille ou à morphologie plane, et ont confirmé l'intérêt clinique d'un système d'intelligence artificielle capable d'assister les praticiens dans l'identification automatique des polypes.



(a)



(b)



(c)

Figure 1.12 Photo prise par nous-mêmes qui présente la salle de diagnostic (a) et son équipement (b, c) de l'hôpital Ahmed Madaghri (Aïn Témouchent).

1.5.1 Stades de développement des polypes colorectaux selon la taille

Les polypes colorectaux sont des lésions tissulaires anormales faisant saillie dans la lumière du côlon ou du rectum. Ils constituent les précurseurs de la majorité des cancers colorectaux (CRC), troisième cancer mondial et deuxième cause de décès par cancer avec environ 900 000 décès annuels. L'évaluation d'un polype repose sur sa localisation, sa taille, sa morphologie, la

profondeur d'invasion pariétale. La table 1.2 présente les catégories de taille et les techniques recommandées (**Gao et al., 2023**) que nous avons considéré dans notre projet de fin d'étude.

Table 1.2 Catégories de taille des polypes et techniques de résection recommandées (**Gao et al., 2023**)

Catégorie	Taille	Risque malin	Technique recommandée
Diminutif	< 5 mm	Très faible	CFP (pince froide)
Petit	6–9 mm	Faible–modéré	CSP (anse froide)
Intermédiaire	10–19 mm	Modéré	CSP/HSP $\pm$ injection
Large	≥ 20 mm	Élevé	EMR / ESD

1.6 Conclusion

Nous avons présenté dans ce chapitre les bases conceptuelles nécessaires à la réalisation de notre travail, à savoir les fondements de l'image numérique, la chaîne de traitement des images médicales, les modalités d'imagerie ainsi que les spécificités de l'endoscopie colique. La segmentation des images, et plus particulièrement celle des polypes colorectaux, constitue une étape cruciale dans l'établissement du diagnostic et représente le cœur de notre projet de fin d'étude. Pour ces raisons, elle sera approfondie dans le chapitre suivant à travers un état de l'art des différentes approches utilisées pour accomplir cette tâche, avec un intérêt particulier porté à l'apprentissage profond.

Chapitre 2

État de l'Art en segmentation de Polypes par Deep Learning

2.1 Introduction

La segmentation précise des polypes constitue une étape essentielle dans les systèmes d'aide au diagnostic assistés par intelligence artificielle. Contrairement à la simple détection, elle permet de délimiter précisément les contours de la lésion, facilitant l'évaluation de sa taille, de sa forme et de son évolution. Cette tâche demeure particulièrement complexe en raison de la variabilité morphologique des polypes et des artefacts présents dans les images endoscopiques.

Historiquement, les premières approches de segmentation des images reposaient sur des méthodes classiques basées sur des caractéristiques conçues manuellement. Cependant, ces méthodes ont rapidement montré leurs limites face à la complexité des données médicales. L'avènement du Deep Learning a marqué une rupture majeure, permettant de passer d'une ingénierie de caractéristiques explicite à un apprentissage automatique de représentations hiérarchiques.

Ainsi, l'évolution des méthodes de segmentation peut être structurée en trois grandes étapes interdépendantes : (i) les approches traditionnelles basées sur des caractéristiques manuelles, (ii) les architectures convolutives profondes, et (iii) les modèles récents intégrant l'attention et les Transformers. Cette progression reflète un passage progressif d'une modélisation locale vers une compréhension globale du contexte visuel. Ce chapitre présente un état de l'art de la segmentation des images médicales, en portant une attention particulière aux polypes colorectaux.

2.2 Deep Learning en imagerie médicale

Le Deep Learning constitue une extension du Machine Learning reposant sur des réseaux de neurones artificiels composés de multiples couches, permettant d'apprendre des représentations de plus en plus abstraites à partir des données (**Hope, 2017**). Contrairement aux approches classiques où les caractéristiques sont définies manuellement, ces modèles apprennent directement

à partir des données brutes.

D'un point de vue fonctionnel, un modèle de Deep Learning peut être interprété comme une fonction non linéaire paramétrée :

$$f(x; \theta) = y \quad (2.1)$$

où :

- x représente l'entrée (image), θ les paramètres du modèle, et y la sortie (masque de segmentation).

Cette capacité d'approximation universelle permet aux réseaux profonds de modéliser des relations complexes, particulièrement adaptées aux tâches de segmentation dense. Toutefois, cette puissance s'accompagne de contraintes importantes, notamment la nécessité de grandes quantités de données annotées et un coût computationnel élevé.

Les différentes architectures de Deep Learning ne sont pas indépendantes : leur évolution répond directement aux limitations des modèles précédents, ce qui justifie leur analyse progressive dans ce chapitre.

2.2.1 Avènement des CNN en vision par ordinateur

Les premières architectures de Deep Learning, telles que les Deep Neural Networks (DNN) (**Hinton & Salakhutdinov, 2006**) et les Deep Belief Networks (DBN) (**Hinton et al., 2006**), ont posé les bases de l'apprentissage profond en montrant qu'il est possible d'entraîner efficacement des réseaux à plusieurs couches. Cependant, ces modèles ne prennent pas en compte explicitement la structure spatiale des images, ce qui limite leur efficacité en vision par ordinateur (**Bengio, 2009**).

L'introduction des réseaux de neurones convolutionnels (CNN) par (**LeCun et al., 1998**) constitue une avancée majeure, en intégrant un biais inductif spatial adapté aux données visuelles. Les opérations de convolution permettent de capturer des motifs locaux (bords, textures), tandis que l'empilement des couches permet une représentation hiérarchique des caractéristiques (**Goodfellow et al., 2016**). La popularisation à grande échelle des CNN doit beaucoup au modèle AlexNet (**Krizhevsky et al., 2012**), qui a démontré la supériorité de ces architectures sur le benchmark ImageNet, ouvrant la voie aux architectures profondes modernes comme VGGNet (**Simonyan & Zisserman, 2015**) et ResNet (**He et al., 2016**).

Ces propriétés sont particulièrement pertinentes pour la segmentation des polypes, où la détection des contours et des textures joue un rôle central. Toutefois, les CNN restent intrinsèquement limités par leur champ réceptif local, qui dépend de la profondeur du réseau et de la taille des noyaux : ils peinent ainsi à modéliser des dépendances contextuelles à longue portée, pourtant essentielles pour distinguer une lésion des structures anatomiques voisines (**Shen et al., 2017**).

Ainsi, bien que les CNN constituent une base solide pour l'analyse d'images médicales (**LeCun et al., 2015**), leur évolution vers des architectures encodeur-décodeur, puis vers des modèles intégrant des mécanismes d'attention, s'est imposée pour répondre aux exigences spécifiques de la segmentation médicale.

2.2.2 Architectures convolutionnelles pour la segmentation

Les réseaux de neurones convolutionnels (CNN) constituent la brique fondamentale de la segmentation médicale moderne. Introduits par (**LeCun et al., 1989**) et popularisés par Alex-Net (**Krizhevsky et al., 2012**), les CNN apprennent automatiquement des représentations hiérarchiques sans ingénierie manuelle des caractéristiques.

2.2.2.1 Composants fondamentaux

Un CNN est composé de trois types de couches. La couche de convolution applique un filtre $k \times k$ glissant sur la carte d'activation pour capturer des motifs locaux (bords, textures) invariants à la translation. La couche d'activation ReLU ($\max(0, x)$) introduit la non-linéarité indispensable. La couche de pooling (MaxPool 2×2) réduit la résolution spatiale en renforçant l'invariance aux petites translations.

Mathématiquement, cette opération de convolution discrète (techniquement une corrélation croisée en deep learning) s'exprime par l'équation suivante :

$$Y_{i,j} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X_{i+m,j+n} \cdot W_{m,n} + b \quad (2.2)$$

où :

- X est la carte d'entrée (les valeurs des pixels ou caractéristiques),
- W représente les poids du filtre (le noyau de convolution),
- b est le terme de biais,
- $Y_{i,j}$ est la valeur calculée pour la carte de caractéristiques en sortie à la position i, j .

La couche d'activation ReLU ($\max(0, x)$) introduit ensuite la non-linéarité indispensable. Enfin, la couche de pooling (MaxPool 2×2) réduit la résolution spatiale en renforçant l'invariance aux petites translations.

2.2.2.2 Fonctionnement et représentation hiérarchique

L'empilement des couches crée une hiérarchie de représentations : les premières couches capturent des primitives visuelles (bords, gradients), les couches intermédiaires assemblent des

motifs complexes (formes, textures) et les couches profondes construisent des représentations sémantiques de haut niveau. La Figure 2.1 illustre cette architecture.

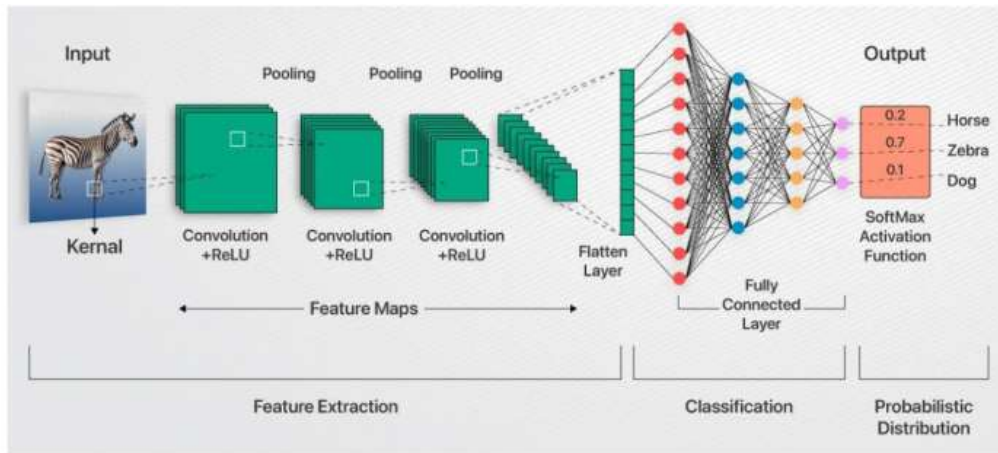


Figure 2.1 Architecture générale d'un CNN : couches de convolution, d'activation ReLU, de pooling et de classification. (Singh, 2024)

2.2.2.3 Limites des CNN pour la segmentation médicale

Malgré leurs succès, les CNN présentent deux limites inhérentes : (i) un champ réceptif local limité qui restreint la capture de dépendances contextuelles à longue portée, essentielles pour distinguer un polype des structures voisines, et (ii) une perte de résolution spatiale progressive par le pooling qui nuit à la localisation précise des frontières. Ces limites ont motivé le développement de l'architecture U-Net (Ronneberger et al., 2015), dont l'innovation clé réside dans les skip connections qui court-circuitent l'information spatiale de l'encodeur vers le décodeur, comme illustré en Figure 2.2.

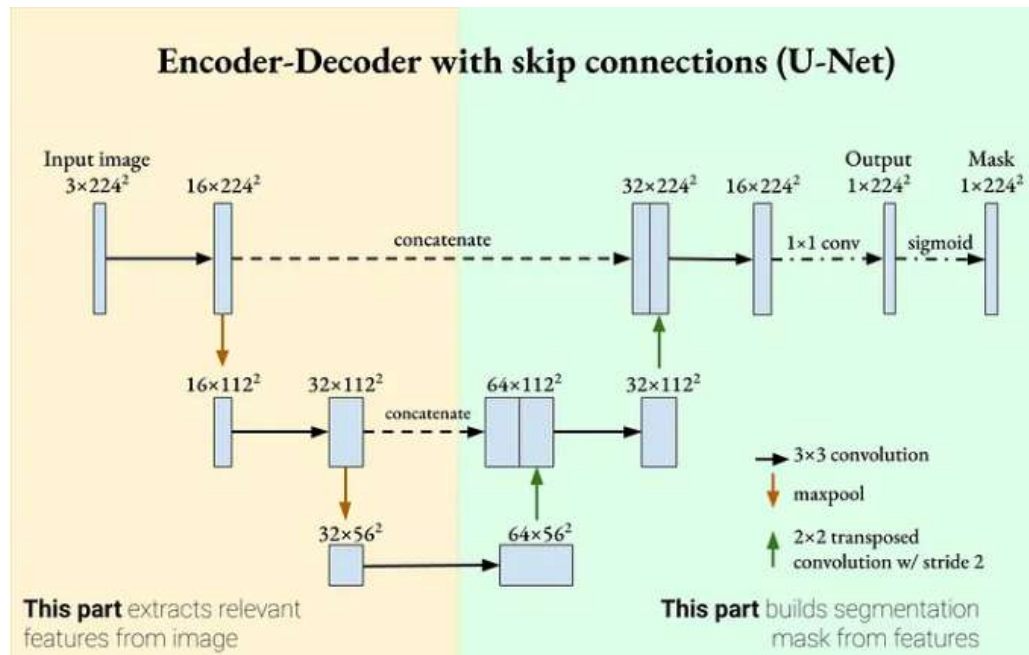


Figure 2.2 Architecture U-Net encodeur-décodeur avec connexions de saut (skip connections) pour la segmentation biomédicale. (Elhaddar, s.d.)

2.3 Évolution des approches de segmentation des polypes

2.3.1 Limites des approches traditionnelles

Les méthodes classiques reposaient principalement sur des descripteurs de bas niveau tels que la couleur, la texture ou la forme (Bernal et al., 2012). Bien que ces approches soient intuitives, elles souffrent d'un manque de robustesse face à la variabilité des données.

Parmi les approches classiques de segmentation, les méthodes basées sur les régions, notamment la croissance de région, ont été largement étudiées en imagerie médicale. Ces méthodes consistent à regrouper des pixels voisins ayant des caractéristiques similaires à partir de points germes. Cependant, elles restent sensibles au bruit et au choix des paramètres (Mubarak et al., 2012).

Les méthodes basées sur le clustering ou les descripteurs de texture (Hwang & Celebi, 2010) présentent des performances limitées, notamment en raison de la similarité visuelle entre les polypes et les tissus environnants.

Ces limitations ont motivé la transition vers des approches basées sur l'apprentissage profond.

2.3.2 Le paradigme convolutionnel : de U-Net aux backbones modernes

La transition vers l'apprentissage profond de bout en bout (end-to-end learning) a permis de s'affranchir de l'extraction manuelle des caractéristiques, laissant le réseau optimiser ses propres

filtres de manière hiérarchique.

2.3.2.1 U-Net : architecture de base et variantes

L'architecture U-Net (**Ronneberger et al., 2015**) constitue le socle de la segmentation biomédicale. Son innovation majeure réside dans l'utilisation de connexions de saut (skip connections) qui concatènent les cartes d'activation de l'encodeur (riches en détails spatiaux de bas niveau) avec celles du décodeur (riches en sémantique globale).

Pour pallier le « fossé sémantique » (semantic gap) entre ces deux flux d'information, des architectures dérivées ont été proposées :

- **UNet++ (Zhou et al., 2020)** : cette architecture introduit des connexions de saut imbriquées et denses (*nested dense skip connections*). Ces modules intermédiaires sont constitués de blocs convolutifs denses qui fusionnent progressivement les cartes de caractéristiques de l'encodeur (riches en détails spatiaux mais pauvres en sémantique) avec celles du décodeur (riches en sémantique mais de basse résolution). Cette agrégation hiérarchique permet de combler le « fossé sémantique » (*semantic gap*) entre les deux flux d'information avant leur concaténation. De plus, l'architecture intègre un mécanisme de supervision profonde (*deep supervision*), qui calcule la fonction de perte à plusieurs niveaux du décodeur. Cela stabilise la rétropropagation du gradient et permet, si nécessaire, d'élaguer (*prune*) le modèle lors de l'inférence pour optimiser le compromis entre vitesse de calcul et précision.
- **ResUNet++ (Jha et al., 2019)** : cette architecture dérivée intègre trois mécanismes clés pour pallier les limites du UNet classique. Premièrement, elle utilise l'apprentissage résiduel (**He et al., 2016**) via des connexions de saut additives pour atténuer la disparition du gradient dans les réseaux profonds. Deuxièmement, elle incorpore des blocs Squeeze-and-Excitation (**Hu et al., 2018**), un mécanisme d'attention qui recalibre dynamiquement l'importance des canaux de caractéristiques pour se concentrer sur les régions sémantiquement pertinentes. Enfin, elle déploie un module ASPP (**Chen et al., 2017**) en bas de l'encodeur, utilisant des convolutions dilatées à taux multiples pour élargir le champ réceptif et capturer un contexte multi-échelle sans perte de résolution spatiale.

2.3.2.2 L'émergence des encodeurs légers

Récemment, la littérature montre une transition des encodeurs traditionnels (VGG, ResNet) vers des architectures de nouvelle génération. L'intégration de ConvNeXt-Tiny dans des modèles comme PolypNextLSTM (**Bhattacharya et al., 2024**) démontre qu'il est possible d'obtenir des performances de pointe grâce à des convolutions en profondeur (depthwise convolutions) et des noyaux larges, offrant un bon compromis entre champ réceptif large et efficacité computationnelle, exigence centrale pour l'inférence clinique en temps réel.

La Figure 2.3 illustre la structure interne d'un bloc ConvNeXt-Tiny, caractérisé par une convolution depthwise à grand noyau (7×7), une normalisation de couche et une connexion résiduelle.

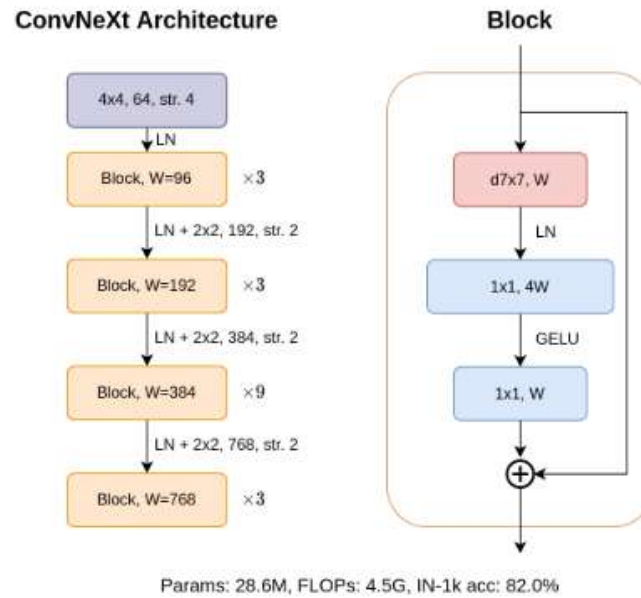


Figure 2.3 Bloc ConvNeXt-Tiny (Vishwkarma, 2025).

2.3.3 Mécanismes d'attention : vers une modélisation sélective

Malgré leur robustesse, les CNN purs traitent toutes les régions de l'image avec une importance égale, ce qui conduit souvent à des faux positifs sur les replis intestinaux. L'intégration de mécanismes d'attention permet au réseau de pondérer dynamiquement les caractéristiques, filtrant le bruit de fond pour se focaliser sur les régions lésionnelles.

Attention canal et spatiale : le module CBAM (Woo et al., 2018) infère des cartes d'attention séquentielles (canal puis spatial) pour souligner les descripteurs pertinents. Son efficacité a été récemment réaffirmée dans des frameworks d'imagerie médicale avancés comme UKAN-CBAM (Ahamed et al., 2026).

Attention inverse pour le raffinement des bords : le modèle PraNet (Fan et al., 2020) a introduit un paradigme novateur : au lieu de segmenter le polype de manière monolithique, le réseau utilise une « attention inverse » pour effacer itérativement les prédictions internes et forcer l'apprentissage sur les frontières, résolvant ainsi le problème des polypes aux limites ambiguës.

2.3.4 La révolution des Vision Transformers et l'hybridation

L'introduction des Vision Transformers (ViT) par (Dosovitskiy et al., 2021) a constitué une rupture paradigmatique. En découpant l'image en patches de 16×16 pixels traités comme tokens

séquentiels, le mécanisme d'auto-attention globale (MSA) capture des dépendances à longue portée dès les premières couches, surmontant le biais inductif local des CNN.

TransUNet (**Chen et al., 2021**) combine un encodeur CNN-ResNet pour l'extraction locale avec un encodeur Transformer pour la modélisation contextuelle globale, suivi d'un décodeur CNN progressif. Swin-UNet (**Cao et al., 2022**) adopte une approche entièrement Transformer en introduisant des fenêtres glissantes décalées (Shifted Window Attention) limitant le coût quadratique, avec Patch Merging et Patch Expanding pour le sous-échantillonnage et l'upsampling hiérarchique.

Dans le domaine des polypes, FCB-SwinV2 (**Fitzgerald & Matuszewski, 2024**) utilise un backbone SwinV2 amélioré pour atteindre des performances état-de-l'art. Ces architectures illustrent la tendance vers une hybridation CNN-Transformer, présentée en Figure 2.4.

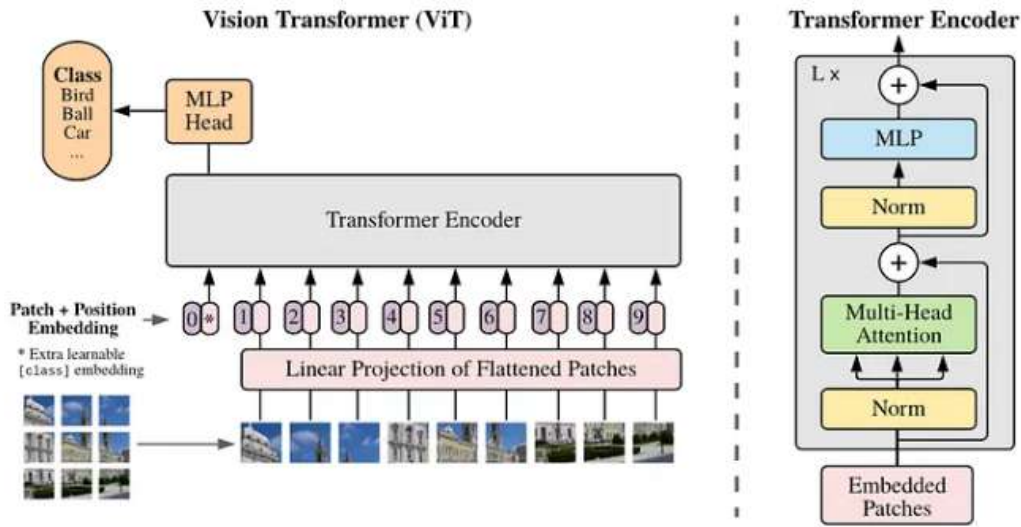


Figure 2.4 Vision Transformers pour la segmentation : ViT, TransUNet et Swin-UNet (**Vishwakarma, 2025**).

Après avoir présenté l'évolution générale des architectures de segmentation, nous analysons dans la section suivante quelques travaux représentatifs de l'état de l'art afin de positionner notre approche par rapport aux méthodes récentes.

2.4 Travaux connexes et positionnement

2.4.1 PraNet : Attention inverse pour le raffinement des frontières

PraNet (**Fan et al., 2020**) propose le paradigme de l'attention inverse (Reverse Attention). Un Partial Decoder (PD) génère une carte de segmentation globale approximative à partir des caractéristiques de haut niveau extraites par un backbone Res2Net. Cette carte est ensuite inversée pour rediriger l'apprentissage vers les régions non segmentées, notamment les contours ambigus. Le mécanisme est appliqué en cascade sur trois niveaux (f_3 , f_4 , f_5), produisant des cartes

S_3, S_4, S_5 supervisées par une perte BCE + IoU. Sur Kvasir-SEG, PraNet atteint Dice = 89,8 % et IoU = 84,0 %, établissant un nouveau standard pour les polypes aux frontières floues.

Il a une excellente précision sur les contours complexes ; très robuste aux variations de forme. Mais avec un coût de calcul élevé ; des performances qui chutent sur les très petites cibles.

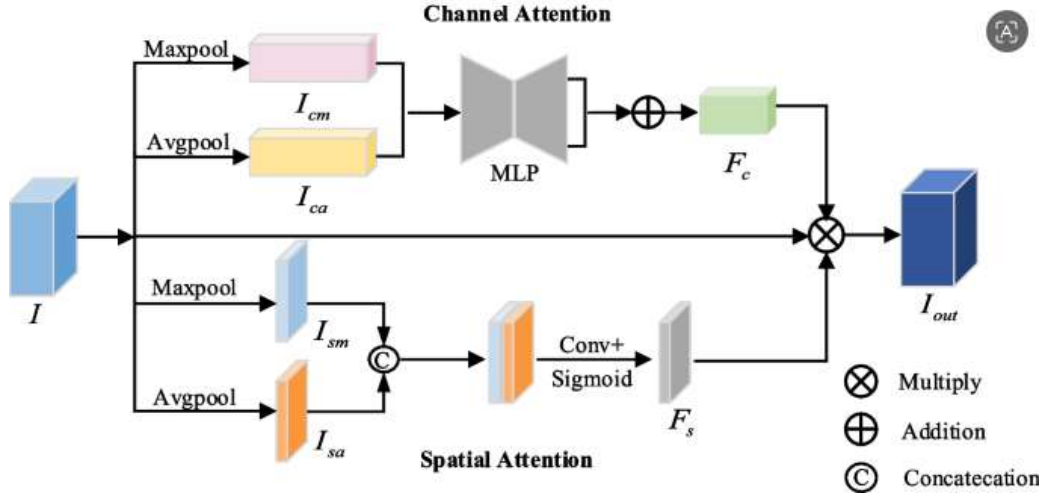


Figure 2.5 Architecture PraNet : décodeur partiel (PA) et attention inverse (RA) en cascade pour la segmentation précise des contours de polypes (Fan et al., 2020).

2.4.2 Attention U-Net : Portes d'attention pour le raffinement spatial

Attention U-Net (Oktay et al., 2018) étend U-Net en intégrant des Attention Gates (AG) dans les *skip connections*. Chaque AG combine les caractéristiques de l'encodeur x et le signal de *gating* g du décodeur pour produire une carte d'attention spatiale :

$$\alpha = \sigma_2(\psi^T \cdot \sigma_1(W_g^T \cdot g + W_x^T \cdot x + b_g)) , \quad (2.3)$$

pondérant les features par $\hat{x} = \alpha \odot x$. Ce mécanisme supprime les faux positifs liés aux plis intestinaux et aux reflets spéculaires, en guidant le décodeur à chaque échelle vers les régions lésionnelles pertinentes. Notre approche proposée s'inspire directement de ce travail en couplant les Attention Gates à un encodeur ConvNeXt-Tiny et un module ASPP. Il supprime efficacement le bruit de fond ; force le réseau à se concentrer sur les zones d'intérêt, mais cela reste un risque de perdre les détails très fins ; sensible au réglage des hyperparamètres.

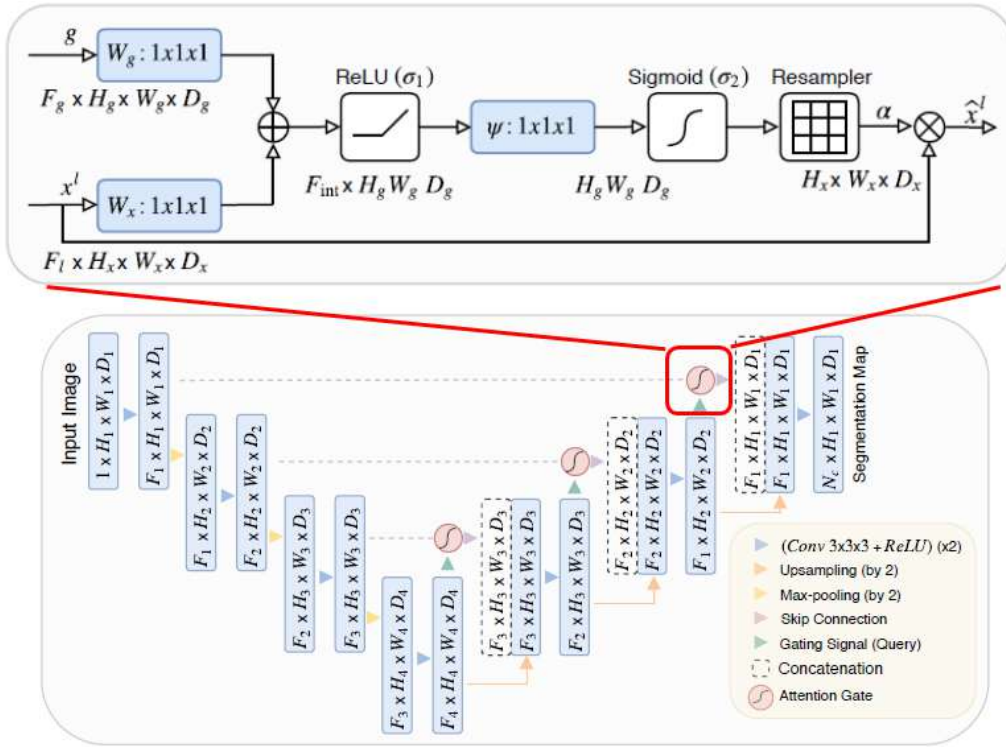


Figure 2.6 Attention U-Net : architecture globale et détail du module Attention Gate (AG) pour le raffinement spatial des skip connections (Oktay et al., 2018).

2.4.3 META-Unet-Attention Transformer multi-échelle

META-Unet (Wu et al., 2023) intègre le module MTA (Multi-scale Transformer Attention) dans chaque étape de l'encodeur U-Net. Le MTA combine une attention multi-têtes globale (MSA) pour les corrélations à longue portée et une attention par fenêtres décalées pour le contexte local, couplées par une normalisation de couche et un réseau Feed-Forward GELU. Cette double attention capture simultanément la position globale du polype et les détails locaux des transitions muqueuse-lésion. Sur Kvasir-SEG : Dice = 91,86 % et IoU = 86,42 %, démontrant qu'une hybridation modulaire CNN-Transformer surpasse les architectures purement convolutionnelles. Il combine détails locaux (CNN) et contexte global (Transformer) ; très haute précision globale.

Et pour ses limites, c'est l'extrêmement gourmand en mémoire et en temps de calcul ; nécessite de très grands jeux de données.

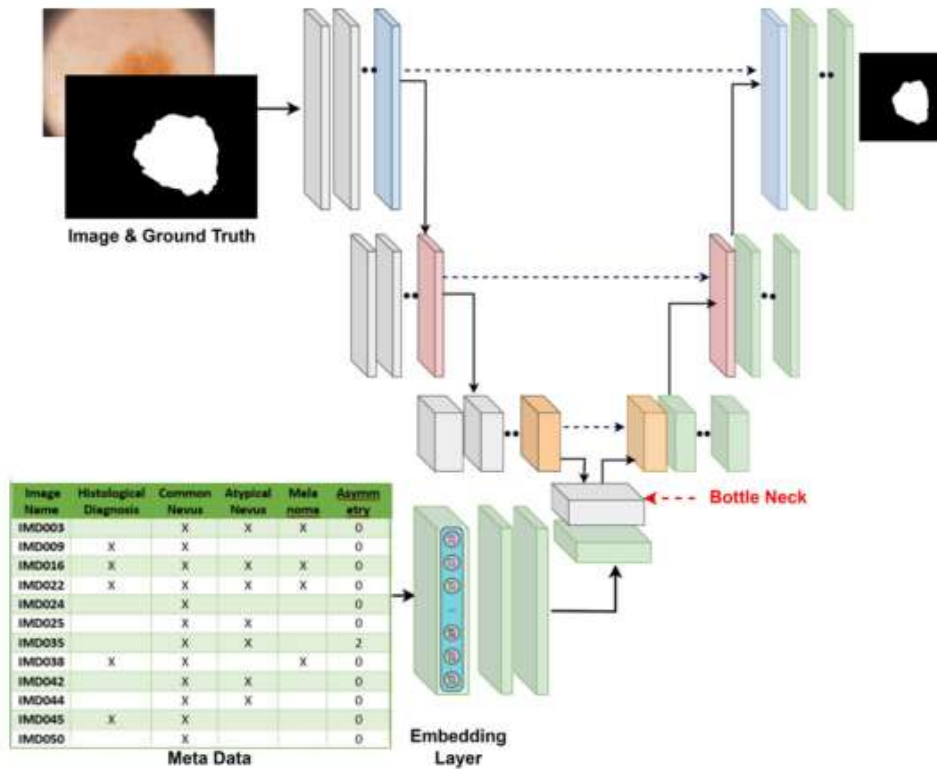


Figure 2.7 Architecture META-Unet : module MTA dans chaque étape de l'encodeur pour une modélisation contextuelle globale et locale (Wu et al., 2023).

Table 2.1 Comparaison de quelques méthodes récentes de segmentation des polypes colorectaux.

Méthode	Année	Architecture	Dataset	Dice (%)	IoU (%)	Référence
PraNet	2020	Res2Net + Reverse Attention	Kvasir-SEG, CVC-ClinicDB	89.8	84.0	Fan et al., MICCAI 2020
TransUNet	2021	CNN + Transformer Encoder	Kvasir-SEG	91.7	86.5	Chen et al., 2021
Focus U-Net	2021	U-Net avec mécanismes d'attention	CVC-ClinicDB	94.1	88.7	Zhang et al., 2021
Polyp-PVT	2023	Pyramid Vision Transformer	Kvasir-SEG	91.8	86.0	Wang et al., 2023
DPE-Net	2024	Dual-Encoder CNN	Kvasir-SEG, CVC-ClinicDB	93.1	89.1	Xu et al., 2024
Dilated U-Net Seg	2024	U-Net avec convolutions dilatées	Kvasir-SEG	92.0	87.0	Liu et al., 2024
CL-Polyp	2025	Apprentissage contrastif + CNN	Kvasir-SEG, CVC-ClinicDB	92.0	88.0	Zhang et al., 2025
VMDU-Net	2025	Fusion Mamba + Transformer	Kvasir-SEG	93.8	87.1	Chen et al., 2025
PDCA-Net	2025	Réseau à double attention	Kvasir-SEG, CVC-ClinicDB	93.0	88.0	Li et al., 2025

L'analyse de la littérature récente met en évidence un défi majeur : trouver un compromis entre précision de segmentation et efficacité computationnelle. Les architectures basées sur les

Transformers offrent des performances élevées, mais leur coût computationnel reste important pour une utilisation clinique en temps réel. À l'inverse, les modèles trop légers présentent souvent des difficultés à segmenter correctement les petites lésions et les polypes plats.

Notre approche s'inscrit dans la continuité de ces travaux. Elle associe un encodeur ConvNeXt-Tiny à un module ASPP pour l'extraction de caractéristiques multi-échelles, enrichi de portes d'attention (Attention Gates) pour le raffinement spatial des prédictions. Cette combinaison vise à concilier précision de segmentation, robustesse aux artefacts endoscopiques et efficacité computationnelle, en se positionnant entre les CNN classiques et les Transformers plus lourds.

2.5 Défis actuels : vers des modèles plus robustes

Malgré les progrès remarquables accomplis dans la segmentation automatique des polypes, plusieurs défis importants subsistent, limitant le déploiement clinique à grande échelle de ces technologies.

- **Variabilité d'apparence des polypes.** Les polypes présentent une grande hétérogénéité en termes de taille (de quelques millimètres à plusieurs centimètres), de forme (pédiculés, sessiles, plats), de couleur (du rose pâle au rouge foncé), et de texture (lisse, granulaire, vil- leux). Cette variabilité rend difficile l'apprentissage de représentations robustes capables de généraliser à tous les types de polypes. Les polypes de petite taille et de morphologie plane sont particulièrement difficiles à segmenter avec précision (**Huang et al., 2023**).
- **Conditions d'acquisition hétérogènes.** Les images endoscopiques sont acquises dans des conditions variables : éclairage non uniforme, reflets spéculaires, présence de bulles d'air ou de résidus fécaux, flou de mouvement, variations de distance et d'angle de vue. Ces artefacts peuvent perturber les algorithmes de segmentation, conduisant à des faux positifs ou des faux négatifs. La robustesse face à ces variations est essentielle pour une utilisation clinique fiable (**He et al., 2024**).
- **Frontières ambiguës.** Les contours des polypes ne sont pas toujours clairement définis, particulièrement pour les lésions planes ou les polypes présentant une transition graduelle avec la muqueuse normale. Cette ambiguïté se reflète dans la variabilité inter-annotateur des masques de vérité terrain, qui peut atteindre 10-15 % en termes de Dice. Les algo- rithmes doivent être capables de gérer cette incertitude intrinsèque (**Huang et al., 2023**).
- **Données d'entraînement limitées.** Contrairement aux applications de vision par ordina- teur sur images naturelles, où des millions d'images annotées sont disponibles, les data- sets médicaux annotés restent relativement petits (quelques centaines à quelques milliers d'images). Cette limitation est due au coût élevé de l'annotation par des experts médicaux et aux contraintes de confidentialité des données patients. Le risque de surapprentissage

est donc important, nécessitant des stratégies de régularisation et d'augmentation de données sophistiquées (Wu et al., 2023) ;(Sikkandar et al., 2024).

- **Généralisation inter-centres.** Les modèles entraînés sur des données provenant d'un centre médical spécifique peuvent présenter des performances dégradées lorsqu'ils sont appliqués à des données d'autres centres, en raison de différences dans les équipements endoscopiques, les protocoles d'acquisition, et les caractéristiques démographiques des populations. Cette limitation de généralisation constitue un obstacle majeur au déploiement clinique à grande échelle (Jiang et al., 2024).
- **Efficacité computationnelle.** Pour une utilisation en temps réel pendant la coloscopie, les algorithmes doivent être capables de traiter au moins 25-30 images par seconde (correspondant à la fréquence d'acquisition vidéo standard). Les architectures Transformer, bien que performantes, présentent souvent un coût computationnel élevé qui peut limiter leur applicabilité sur des dispositifs embarqués. Le développement d'architectures efficaces maintenant un bon compromis précision-vitesse reste un défi important (Wu et al., 2023 ; Sikkandar et al., 2024 ; Li et al., 2026).
- **Interprétabilité et confiance clinique.** Les réseaux de neurones profonds sont souvent perçus comme des «boîtes noires» dont les décisions sont difficiles à interpréter. Pour gagner la confiance des cliniciens et faciliter l'adoption clinique, il est important de développer des méthodes d'explicabilité permettant de comprendre quelles caractéristiques visuelles influencent les prédictions du modèle. Les techniques de visualisation d'attention et de cartes de saillance constituent des pistes prometteuses dans cette direction (Ahamed et al., 2026).
- **Validation clinique.** La plupart des études publiées rapportent des performances sur des datasets de recherche, mais peu ont fait l'objet d'essais cliniques prospectifs rigoureux évaluant l'impact réel sur les résultats cliniques (taux de détection des adénomes, taux de cancer d'intervalle). Cette validation clinique est essentielle pour démontrer la valeur ajoutée des systèmes d'IA et justifier leur adoption en routine (Fitzgerald & Matuszewski, 2024 ; He et al., 2024).
- **Aspects réglementaires et éthiques.** Le déploiement clinique de dispositifs médicaux basés sur l'IA nécessite une approbation réglementaire (marquage CE en Europe, approbation FDA aux États-Unis), impliquant des processus de validation rigoureux et coûteux. Les questions de responsabilité en cas d'erreur diagnostique, de protection des données patients, et d'équité d'accès aux technologies avancées soulèvent également des enjeux éthiques importants qui doivent être adressés (Huang et al., 2023).

Ces défis motivent la poursuite des recherches dans plusieurs directions : développement d'architectures plus robustes et efficaces, exploitation de techniques d'apprentissage avec peu de

données (few-shot learning, apprentissage auto-supervisé), amélioration de la généralisation par des approches d'adaptation de domaine, et développement de méthodes d'explicabilité adaptées au contexte médical.

2.6 Conclusion

Ce chapitre a présenté l'évolution des méthodes de segmentation des polypes colorectaux par Deep Learning, des CNN classiques (U-Net, UNet++, ResUNet++) aux architectures intégrant des mécanismes d'attention (CBAM, Attention Gates, attention inverse) et aux modèles hybrides CNN-Transformer. L'analyse des travaux connexes a permis de situer notre approche : un compromis entre la précision des Transformers et l'efficacité des architectures plus légères. Plusieurs défis demeurent, variabilité des polypes, frontières ambiguës, généralisation et efficacité computationnelle et orientent les choix de notre architecture. Le chapitre suivant présente notre approche proposée, CANet+, ainsi que le cadre méthodologique adopté pour son évaluation.

Chapitre 3

Approche proposée, résultats et discussions

3.1 Introduction

Le cancer colorectal (CCR) représente l'un des problèmes de santé publique les plus préoccupants à l'échelle mondiale. En 2022, on estimait à environ 1,9 million le nombre de nouveaux cas de cancer colorectal et à plus de 900 000 le nombre de décès dans le monde. **(OMS, 2026)**.

Dans ce contexte, la segmentation automatique des polypes dans les images endoscopiques est devenue une composante essentielle des systèmes d'aide au diagnostic (CAD). Pour répondre à ces défis, nous présentons dans ce chapitre notre modèle proposé **CANet+**, une architecture hybride de Deep Learning encodeur-décodeur intégrant un encodeur ConvNeXt-Tiny, un décodeur Attention U-Net, des modules d'attention CBAM, un module ASPP dédié à l'extraction multi-échelle, ainsi qu'un mécanisme de supervision profonde visant à améliorer la précision de la segmentation.

3.2 Contexte clinique et défis de la segmentation des polypes colorectaux

La coloscopie constitue l'examen de référence pour la détection précoce des lésions colorectales. La Figure 3.1 illustre l'anatomie du côlon et un exemple d'image endoscopique, tandis que la Figure 3.2 retrace la progression typique d'un polype sur une période de 8 à 10 ans, soulignant ainsi l'urgence clinique du dépistage **(National Cancer Institute, 2024)**

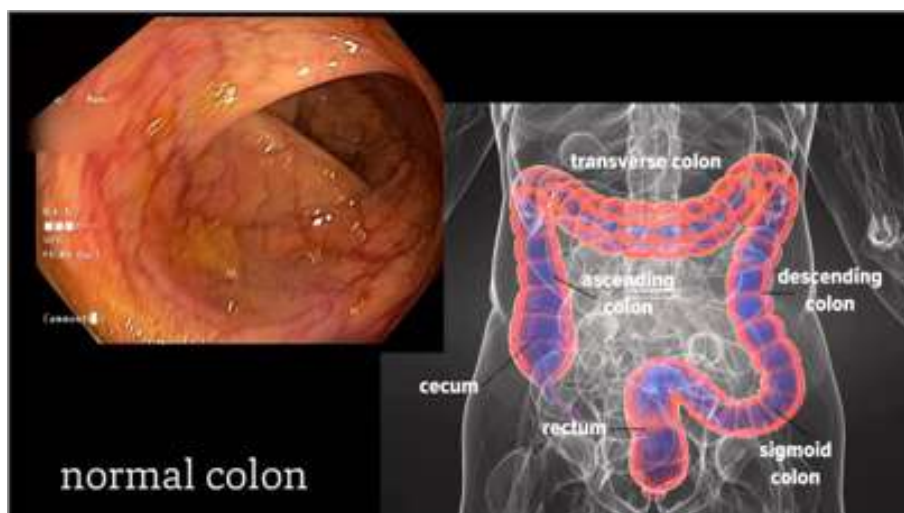


Figure 3.1 Anatomie du côlon et vue endoscopique d'un côlon sain. La coloscopie permet d'explorer l'intégralité du cadre colique et d'identifier les lésions suspectes. (Gastrohegg, s.d.)



Figure 3.2 Progression d'un polype colorectal sur 8 à 10 ans, de la lésion précancéreuse bénigne au carcinome invasif, illustrant l'importance du dépistage précoce.(Lecomte et al., 2026)

Malgré les avancées récentes en vision par ordinateur, la segmentation automatique des polypes demeure une tâche complexe. Les polypes présentent une grande diversité de tailles, de formes, de textures et de couleurs. De plus, leurs contours sont souvent mal définis, le contraste avec les tissus environnants peut être faible et les images endoscopiques contiennent fréquemment des artefacts tels que les reflets spéculaires. La Figure 3.3 illustre cette diversité morphologique et les principaux défis associés à leur segmentation automatique.(Bernal et al., 2012)

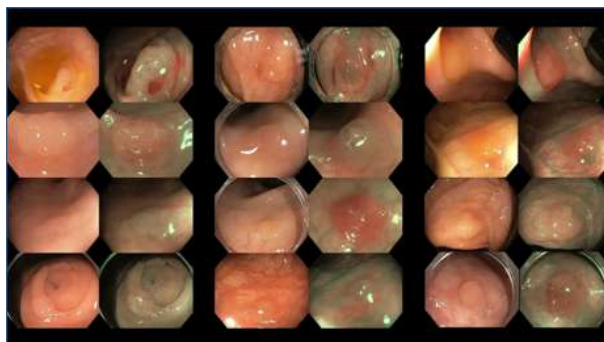


Figure 3.3 Diversité morphologique des polypes colorectaux observés en coloscopie : polypes sessiles, pédiculés, plans et lésions à faible contraste, illustrant la complexité de la segmentation automatique. (Benkalfate et al., 2024)

3.3 Base de données utilisée et prétraitement

Nous avons évalué notre modèle de DL CANet+ sur trois jeux de données publics de référence, couvrant une large gamme de difficultés cliniques : variabilité de taille et de forme des polypes, éclairage variable, reflets spéculaires, présence de mucus, lésions planes et faible contraste. (Adneen et al., 2025); (Bhattacharya et al., 2024). Nous avons également effectué des prétraitements sur ces bases afin de pouvoir les exploiter.

3.3.1 Description

3.3.1.1 Kvasir-SEG

Kvasir-SEG comprend 1 000 images haute résolution (de 332×487 à 1920×1072 pixels) accompagnées de masques binaires annotés manuellement par des gastro-entérologues experts. Ce jeu de données inclut des polypes de toutes tailles et morphologies, avec des reflets, des ombres et des saignements représentatifs des conditions réelles d'une coloscopie (Jha et al., 2020)

3.3.1.2 CVC-ClinicDB

CVC-ClinicDB contient 612 images extraites de 31 vidéos coloscopiques réelles (résolution de 384×288 pixels), chacune associée à un masque de segmentation précis au niveau du pixel. Ce jeu de données reflète des séquences cliniques authentiques, avec apparition et disparition de polypes ainsi que des variations naturelles d'éclairage et d'orientation de la caméra. (Bernal et al., 2015)

3.3.1.3 CVC-ColonDB

CVC-ColonDB comprend 300 images de polypes issues de différentes vidéos coloscopiques, d'une taille moyenne de 574×500 pixels. Il s'agit du jeu de données le plus difficile des trois,

caractérisé par de nombreux polypes plans, des occlusions fréquentes dues aux replis muqueux, un faible contraste et une proportion élevée de reflets spéculaires.(Angermann, 2018)

Table 3.1 Volumétrie des datasets partition 4-fold CV.

Dataset	Total	Train	Val	Test
Kvasir-SEG	1000	600	200	200
CVC-ClinicDB	612	368	122	122
CVC-ColonDB	300	180	60	60
TOTAL	1912	1148	382	382

Les trois jeux de données présentent une diversité importante en termes de nombre d'images, de résolution et de complexité des lésions. Cette complémentarité permet d'évaluer la robustesse du modèle sur différents contextes cliniques.

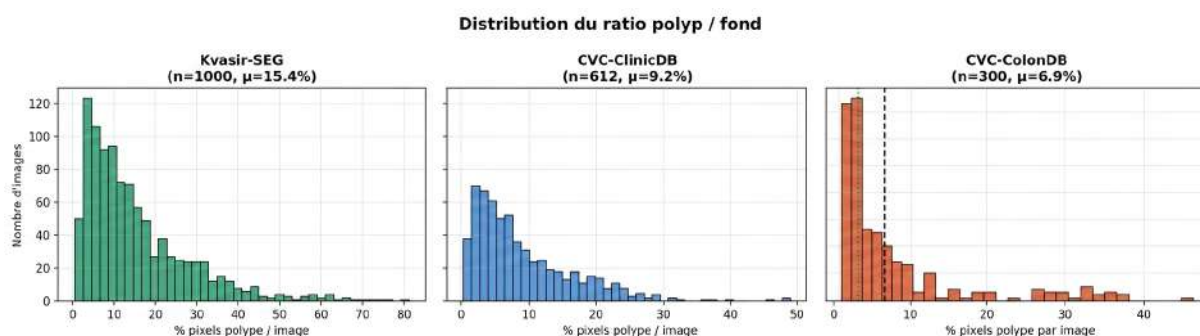


Figure 3.4 Distribution Polype/Fond

La distribution montre un fort déséquilibre entre les pixels appartenant au polype et ceux du fond. Cette observation justifie l'utilisation d'une fonction de perte adaptée ainsi que d'une stratégie d'augmentation des données.

3.3.2 Prétraitement des Données

Toutes les images et leurs masques associés sont redimensionnés à une résolution uniforme de 512×512 pixels, en utilisant une interpolation bilinéaire pour les images et une interpolation par le voisin le plus proche pour les masques, afin de préserver la précision des annotations binaires. Les valeurs des pixels sont normalisées selon les statistiques d'ImageNet (moyenne = [0,485, 0,456, 0,406]; écart-type = [0,229, 0,224, 0,225]), conformément aux prérequis du backbone ConvNeXt-Tiny pré-entraîné.

3.3.3 Augmentation des données

Nous avons adopté une stratégie d'augmentation de données intensive lors de l'entraînement afin de simuler les conditions difficiles d'une coloscopie en situation réelle. Les transformations suivantes sont utilisées :

- **Invariance géométrique** : retournements horizontal et vertical (HorizontalFlip, VerticalFlip), rotation aléatoire de 90° (RandomRotate90) et transformation affine avec changement d'échelle et de perspective (ShiftScaleRotate), pour couvrir les orientations imprévisibles de la caméra endoscopique.
- **Distorsions photométriques** : ajustement aléatoire de la luminosité et du contraste (RandomBrightnessContrast), ajout de bruit gaussien (GaussNoise) et floutage gaussien (GaussianBlur), afin de forcer le modèle à se concentrer sur la morphologie du polype plutôt que sur les artefacts d'éclairage.
- **Occlusion structurelle (CoarseDropout)** : masquage aléatoire de régions rectangulaires de l'image, obligeant le réseau à reconnaître les polypes même lorsqu'ils sont partiellement cachés par des replis tissulaires ou des fluides intestinaux.

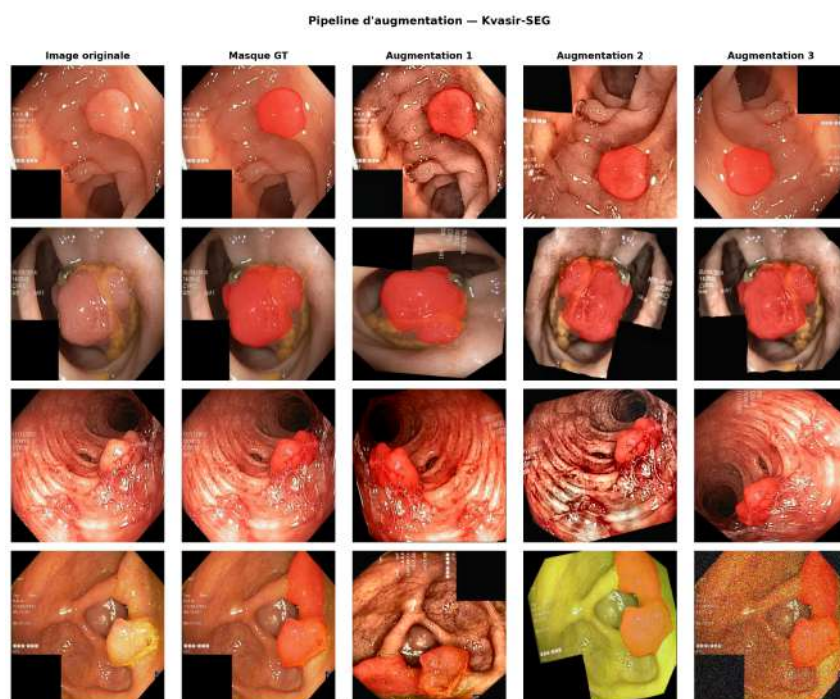


Figure 3.5 Pipeline d'augmentation appliqué au jeu de données Kvasir-SEG

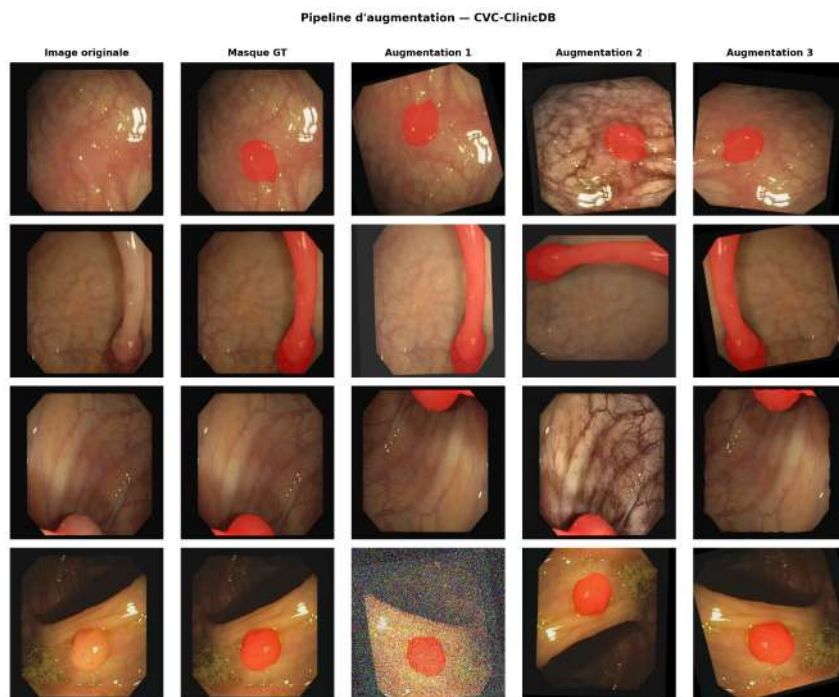


Figure 3.6 Pipeline d'augmentation appliqué au jeu de données CVC-ClinicDB

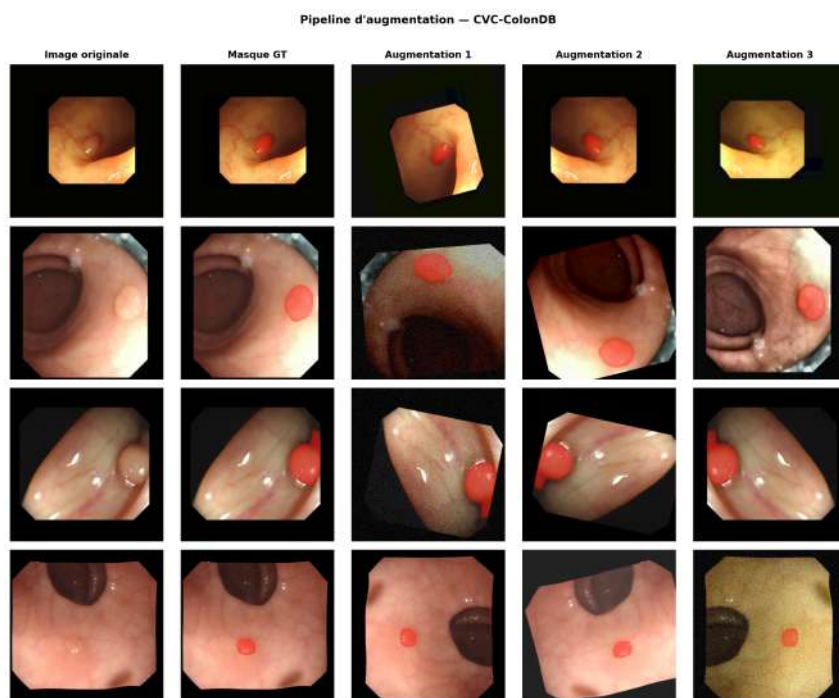


Figure 3.7 Pipeline d'augmentation appliqué au jeu de données CVC-ColonDB

Ces transformations géométriques et photométriques permettent d'améliorer la diversité des données d'entraînement tout en conservant la cohérence entre l'image et son masque.

Table 3.2 Volumétrie des datasets entraînement et augmentation.

Dataset	Train	Train augmenté (×100 époques)
Kvasir-SEG	600	60 000
CVC-ClinicDB	368	36 800
CVC-ColonDB	180	18 000
TOTAL	1148	114 800

Lors des phases de validation et de test, seuls le redimensionnement et la normalisation sont appliqués, garantissant ainsi que les métriques de performance reflètent fidèlement le comportement du modèle sur des images cliniques originales, sans transformation artificielle.

3.4 Approche proposée

3.4.1 Positionnement et motivations

La segmentation automatique des polypes colorectaux a connu des avancées significatives avec l'essor des architectures profondes, notamment les variantes de **U-Net (Ronneberger et al., 2015)** et leurs extensions intégrant des mécanismes d'attention ou de modélisation contextuelle. Toutefois, malgré ces progrès, plusieurs limitations subsistent dans les approches existantes.

Les architectures purement convolutionnelles, telles que U-Net (**Ronneberger et al., 2015**) et ses dérivés, présentent un champ réceptif limité qui restreint leur capacité à modéliser les dépendances spatiales à longue portée, indispensables pour segmenter des polypes de grande taille ou aux contours diffus. À l'inverse, les approches fondées sur des Transformers capturent efficacement le contexte global mais requièrent des volumes de données importants et un coût computationnel élevé (**Liu et al., 2024**), peu compatible avec les contraintes des environnements cliniques.

Par ailleurs, la majorité des architectures classiques traitent uniformément l'ensemble des régions de l'image, sans accorder une importance différenciée aux zones véritablement informatives. Or, en endoscopie, les polypes occupent généralement une faible proportion de l'image et se confondent fréquemment avec la muqueuse environnante, rendant nécessaire un mécanisme d'attention capable de focaliser l'apprentissage sur les régions cliniquement pertinentes.

Par conséquent, nous proposons une architecture de Deep Learning hybride, CANet+, combinant la robustesse hiérarchique d'un encodeur convolutionnel moderne, la précision de reconstruction d'un décodeur à portes d'attention (**Oktay et al., 2018**), l'agrégation multi-échelle par convolutions dilatées (**Chen et al., 2017**), et la stabilisation de l'apprentissage par supervision profonde (**Lee et al., 2015**).

3.4.2 Architecture Globale de notre approche de segmentation

Notre approche de segmentation s'articule autour de cinq étapes successives, de l'image endoscopique brute jusqu'au masque binaire de segmentation.

L'image d'entrée, redimensionnée à une résolution standard, est d'abord traitée par l'encodeur ConvNeXt-Tiny, qui produit une représentation hiérarchique à quatre niveaux de profondeur. Les cartes de caractéristiques extraites alimentent à la fois le décodeur, via les connexions de saut, et le module **ASPP** positionné au goulot d'étranglement.

Le module **ASPP** enrichit la représentation la plus profonde par une agrégation multi-échelle du contexte. Les caractéristiques ainsi enrichies sont ensuite transmises au décodeur Attention **U-Net**, qui reconstruit progressivement le masque de segmentation à pleine résolution. À chaque étape de décodage, les portes d'attention et les blocs **CBAM** affinent les représentations en supprimant les activations non pertinentes.

Enfin, la supervision profonde calcule la perte sur plusieurs sorties intermédiaires en plus de la sortie finale, ce qui guide l'optimisation et améliore la qualité des représentations à toutes les échelles. Le masque binaire produit en sortie indique, pour chaque pixel, la présence ou non d'un polype.

Avant de présenter en détail chacun des composants de **CANet+**, il convient de revenir sur le fondement théorique de l'architecture : le paradigme encodeur-décodeur convolutif, désormais standard en segmentation sémantique, qui structure le flux d'information entre l'extraction des caractéristiques sémantiques et la reconstruction du masque à pleine résolution. Sa formalisation mathématique est décrite dans la sous-section suivante. (Jha et al., 2020) ; (Bernal et al., 2012)

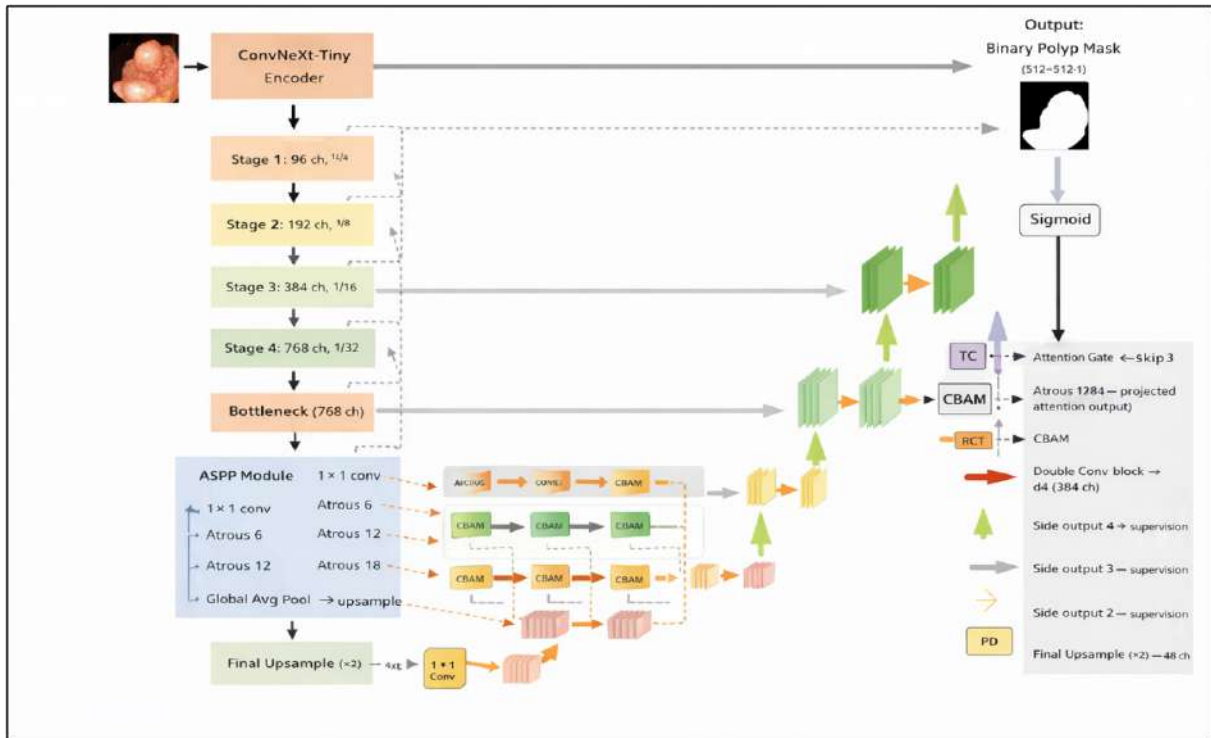


Figure 3.8 Architecture générale de notre approche proposée

3.4.3 Interprétation et analyse du modèle

Afin d'améliorer l'interprétabilité du modèle proposé, **CANet+** intègre un module d'analyse avancée permettant la visualisation des zones d'attention du réseau. Cette analyse repose sur deux composantes complémentaires : l'estimation de l'incertitude via Monte-Carlo Dropout et la visualisation des régions discriminantes à l'aide de **Grad-CAM** appliqué aux caractéristiques du module **ASPP**. Ces mécanismes permettent d'expliquer qualitativement les décisions du modèle en mettant en évidence les régions de l'image les plus influentes dans le processus de segmentation.

La méthode **Grad-CAM** est une technique de localisation discriminante de classe qui génère des explications visuelles pour tout réseau basé sur un CNN sans nécessiter de modifications architecturales ou de réentraînement. (Selvaraju et al., 2020)

3.5 Conception du modèle proposé CANet+

3.5.1 Méthodologie de conception

La conception de notre modèle de DL repose sur quatre principes directeurs, formulés en réponse aux limitations identifiées précédemment. Le paradigme encodeur-décodeur, illustré dans la Figure 3.9, constitue le fondement architectural de **CANet+**. L'encodeur compresse progressivement l'image en représentations sémantiques de haut niveau, tandis que le décodeur reconstruit le masque de segmentation à pleine résolution en s'appuyant sur des connexions de saut (*skip connections*) qui préservent les détails spatiaux fins (Ronneberger et al., 2015). Ces principes se déclinent comme suit :

a. Extraction de caractéristiques hiérarchiques L'encodeur **ConvNeXt-Tiny** combine l'efficacité computationnelle des réseaux convolutionnels modernes avec une capacité de représentation comparable à celle des Transformers. Ce choix offre un compromis avantageux entre performance, vitesse d'inférence et compatibilité avec des jeux de données médicaux de taille modérée.

b. Focalisation attentionnelle Deux mécanismes complémentaires sont mobilisés : les portes d'attention du décodeur Attention U-Net (Oktay et al., 2018), qui filtrent les caractéristiques transmises par les connexions de saut, et les modules CBAM (Woo et al., 2018), qui appliquent une attention conjointe sur les dimensions canal et spatiale au sein du décodeur. Cette double attention supprime le bruit de fond et renforce les régions pertinentes à chaque étape de la reconstruction.

c. Modélisation multi-échelle Un module **ASPP** (Chen et al., 2017) est intégré au goulot d'étranglement et applique en parallèle plusieurs convolutions dilatées à des taux distincts. Cette stratégie élargit considérablement le champ réceptif effectif sans accroître la profondeur du réseau, ce qui est essentiel pour traiter la grande variabilité de taille des polypes.

d. Stabilisation de l'apprentissage Un mécanisme de supervision profonde (Lee et al., 2015) la fonction de perte non seulement sur la sortie finale mais également sur des sorties intermédiaires du décodeur, ce qui atténue le problème de l'évanouissement du gradient et accélère la convergence.

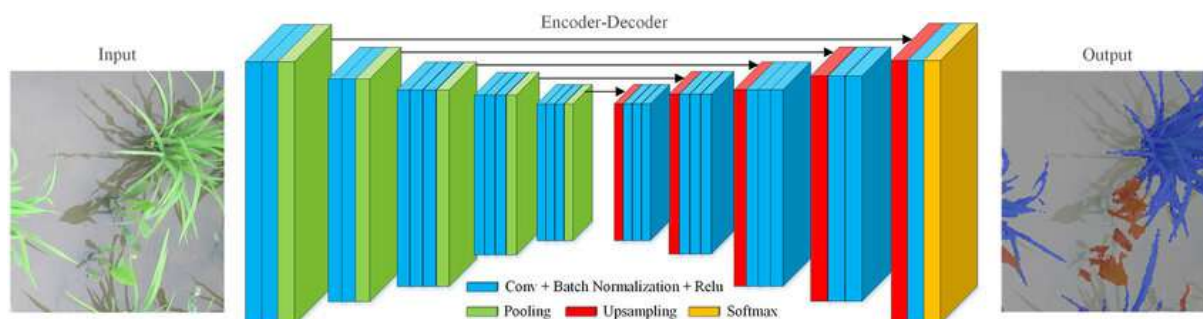


Figure 3.9 Architecture générale d'un réseau encodeur-décodeur convolutif.(Vayssade, 2022)

3.5.2 Architecture détaillée de CANet+

Après avoir présenté l'approche générale et son fondement théorique, cette section décrit en détail l'architecture intégrée de **CANet+**. Comme illustré dans la Figure 3.10, le pipeline de **CANet+** s'articule autour de cinq étapes successives : extraction des caractéristiques par l'encodeur ConvNeXt-Tiny, agrégation multi-échelle par le module **ASPP** au goulot d'étranglement, reconstruction progressive par le décodeur Attention U-Net, raffinement des représentations par les blocs **CBAM**, et stabilisation de l'apprentissage par la supervision profonde, avant la projection finale vers le masque binaire de segmentation.

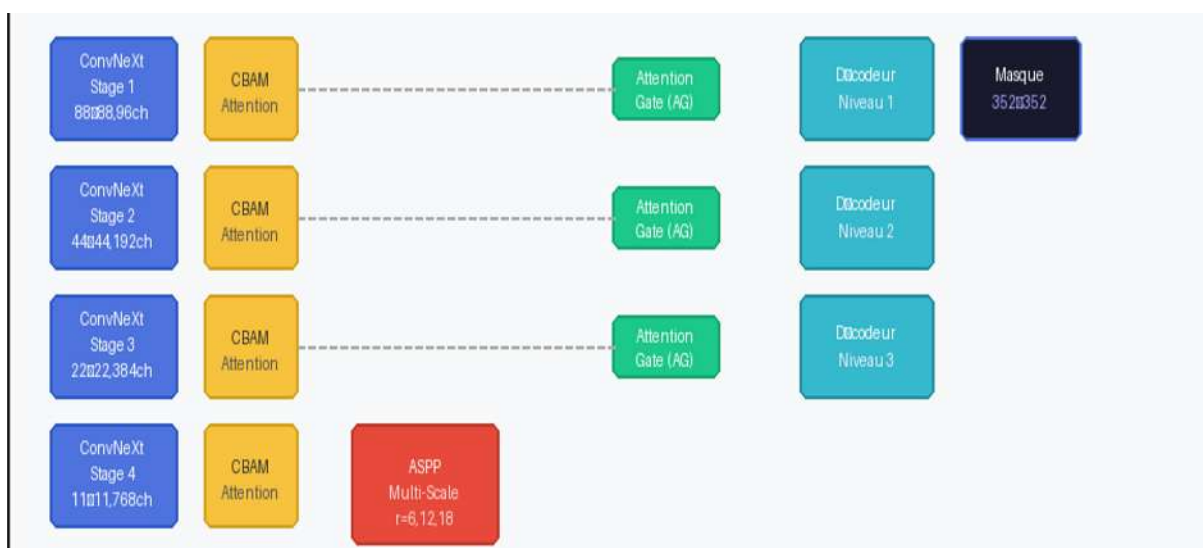


Figure 3.10 Architecture générale de CANet+ : encodeur ConvNeXt-Tiny, module ASPP au goulot d'étranglement, décodeur Attention U-Net avec blocs CBAM et supervision profonde.

3.5.2.1 Encodeur : ConvNeXt-Tiny

L'encodeur de **CANet+** repose sur ConvNeXt-Tiny, un réseau convolutif moderne intégrant de grands noyaux *depthwise* de taille 7×7 , une normalisation par couche (Layer Normalization) en substitution à la normalisation par lots (Batch Normalization), ainsi que des activations **GELU**. Cette architecture, à la fois performante et efficace, a été adoptée avec succès dans plusieurs travaux récents de segmentation de polypes, démontrant sa robustesse face à la variabilité visuelle propre aux images endoscopiques. (Bhattacharya et al., 2024);(Adneen et al., 2025); (Wang et al., 2024).

a) Architecture interne de ConvNeXt La Figure 3.11 illustre l'architecture interne de ConvNeXt, montrant les blocs ConvNeXt répétés à chaque étape avec leurs connexions résiduelles et les couches de sous-échantillonnage. Chaque bloc intègre successivement une convolution *depthwise* 7×7 , une normalisation LayerNorm, une projection **MLP** avec un facteur d'expansion de 4, et une activation GELU. (Adneen et al., 2025)

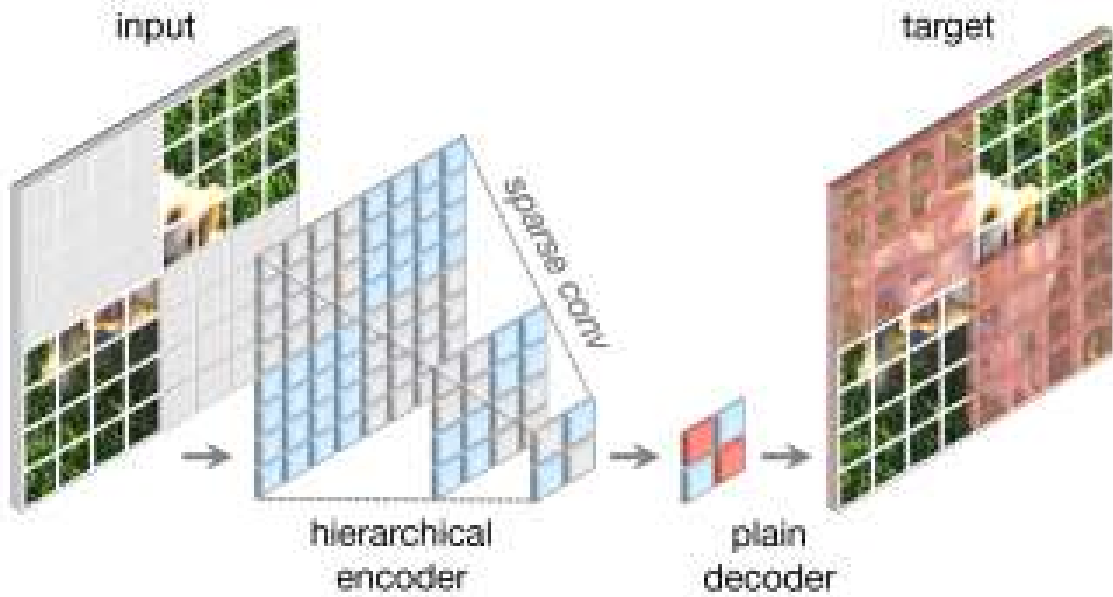


Figure 3.11 Architecture de l'encodeur ConvNeXt (Woo et al., 2023)

b) Niveaux de caractéristiques hiérarchiques L'encodeur produit des cartes de caractéristiques à quatre résolutions spatiales, correspondant à $1/4$, $1/8$, $1/16$ et $1/32$ de la résolution d'entrée (soit 128×128 , 64×64 , 32×32 et 16×16 pixels pour une entrée de 512×512), avec respectivement 96, 192, 384 et 768 canaux. Les couches superficielles capturent les textures fines (contours, reflets, détails locaux), tandis que les couches profondes encodent une compréhension sémantique globale de la scène.

Table 3.3 Niveaux de caractéristiques hiérarchiques de l'encodeur ConvNeXt-Tiny.

Étape	Résolution spatiale	Canaux	Facteur de sous-échantillonnage
Étape 1	128×128	96	$1/4 \times$
Étape 2	64×64	192	$1/8 \times$
Étape 3	32×32	384	$1/16 \times$
Étape 4 (goulot)	16×16	768	$1/32 \times$

L'encodeur est initialisé à partir de poids pré-entraînés sur ImageNet, ce qui lui confère une capacité de représentation initiale particulièrement riche. Cette initialisation est d'autant plus bénéfique dans le contexte médical où les données annotées restent souvent limitées.

3.5.2.2 Décodeur : Attention U-Net

Le décodeur de **CANet+** s'inspire du principe de l'Attention U-Net, adapté pour mieux fusionner les caractéristiques multi-échelles issues de l'encodeur. Des convolutions transposées augmentent progressivement la résolution spatiale, tandis que les connexions de saut réintroduisent les détails haute résolution des étages correspondants de l'encodeur. (Oktay et al., 2018)

Des portes d'attention sont placées sur ces connexions de saut : elles amplifient les régions pertinentes pour la segmentation et atténuent les activations de fond non informatives. Cette stratégie permet de préserver les détails spatiaux fins tout en exploitant pleinement l'information sémantique profonde. (Khan et al., 2025) ; (Ullah et al., 2025)

a) Formalisation des portes d'attention Soient g le signal du décodeur (signal de guidage) et x la représentation de l'encodeur issue de la connexion de saut. Le coefficient d'attention α est calculé selon :

$$\alpha = \sigma(W_{\alpha} \cdot (W_g \cdot g + W_x \cdot x + b_g) + b_{\alpha}) \quad (3.1)$$

Sachant que W_g , W_x et W_{α} désignent des matrices de poids apprenables, b_g et b_{α} des vecteurs de biais, et σ la fonction sigmoïde. Le coefficient α est ensuite multiplié élément par élément avec x pour produire les caractéristiques pondérées transmises au bloc suivant du décodeur. (Oktay, 2018)

3.5.2.3 Module d'Attention CBAM

Afin de raffiner les représentations issues de chaque bloc de fusion, des modules **CBAM** (Convolutional Block Attention Module) (voir Figure 3.12) sont insérés après chaque étape du décodeur. **CBAM** applique successivement une attention canal calculée via un réseau **MLP** appliqué aux statistiques de pooling moyen et maximum puis une attention spatiale réalisée par

une convolution 7×7 . Ce mécanisme permet au réseau de se concentrer dynamiquement sur les canaux et les régions spatiales les plus informatifs. (Woo et al., 2018)

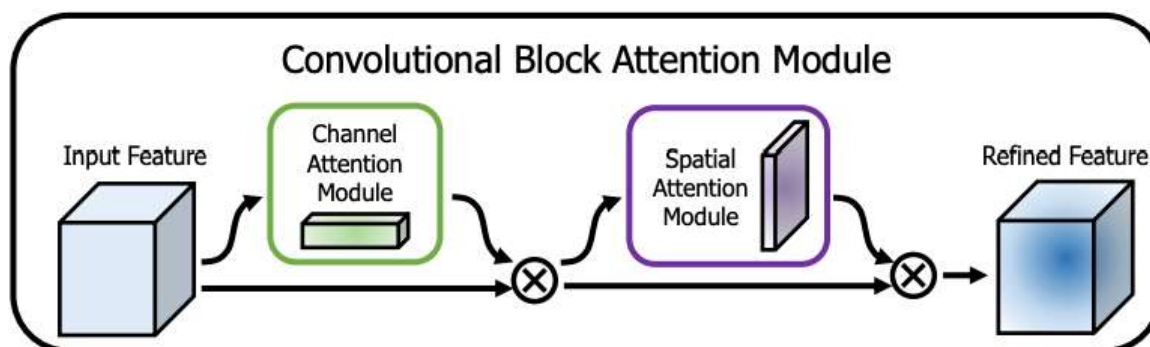


Figure 3.12 Module **CBAM** : attention canal (réseau **MLP** appliqué aux sorties du pooling moyen et maximum) suivie de l'attention spatiale (convolution 7×7). (Woo et al., 2018)

Plusieurs études récentes ont démontré que la combinaison des attentions canal et spatiale améliore significativement la délimitation des contours de polypes dans des images endoscopiques présentant un faible contraste, des reflets spéculaires et du bruit. (Islam et al., 2024); (Ahamed et al., 2026).

Dans **CANet+**, le module CBAM est appliqué après la concaténation des caractéristiques suréchantillonnées avec les connexions de saut guidées par attention. Cette configuration permet d'éliminer le bruit résiduel et de mettre en valeur les structures anatomiques cliniquement significatives.

3.5.2.4 Module ASPP (Atrous Spatial Pyramid Pooling)

Positionné au niveau du goulot d'étranglement, immédiatement après les caractéristiques les plus profondes de l'encodeur, le module ASPP exécute en parallèle des convolutions atrous avec des taux de dilatation de 6, 12 et 18, complétées par une branche de convolution 1×1 et une branche de pooling global moyen (Global Average Pooling). Les sorties de ces cinq branches sont concaténées puis projetées dans un espace de représentation commun. (Chen et al., 2017) (voir Figure 3.13).

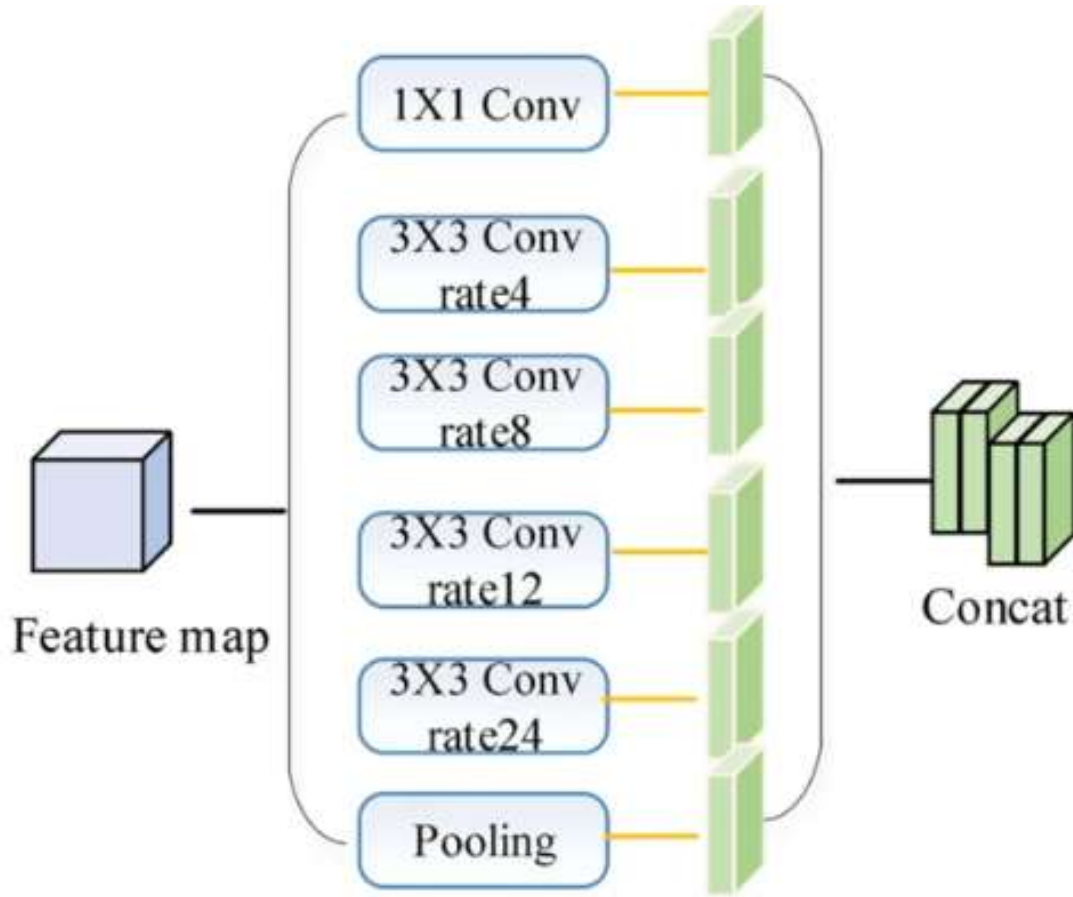


Figure 3.13 Module **ASPP** (Atrous Spatial Pyramid Pooling) : convolutions dilatées parallèles à taux appliqué sur la carte de caractéristiques d'entrée, capturant le contexte à différentes échelles de champ réceptif. (Pan, 2022)

L'intérêt principal de l'**ASPP** réside dans sa capacité à capturer un contexte sémantique à différentes échelles sans sacrifier la résolution spatiale. Cette propriété est essentielle pour la segmentation de polypes, dont la taille peut varier considérablement d'un patient à l'autre. Des travaux récents confirment que les modules de type **ASPP** renforcent la robustesse des modèles de segmentation sur des jeux de données hétérogènes. (Islam et al., 2024) ; (Liu et al., 2024).

3.5.2.5 Supervision Profonde (Deep Supervision)

La supervision profonde est appliquée à plusieurs étages intermédiaires du décodeur, désignés S_2 , S_3 et S_4 . Les sorties auxiliaires correspondantes sont suréchantillonnées à la résolution d'entrée et supervisées par la même fonction de perte que la prédiction finale. (Ullah et al., 2025) ; (Lee et al., 2015)

Cette stratégie assure une propagation efficace du gradient jusqu'aux couches les plus profondes du réseau, évitant ainsi le problème du gradient évanescent. Elle contraint également chaque niveau du décodeur à produire des représentations sémantiquement cohérentes à toutes

les échelles, ce qui se révèle particulièrement bénéfique pour la détection de polypes de tailles très variables. (Hu et al., 2025) ; (Ullah et al., 2025)

La perte totale d'entraînement est définie par la combinaison pondérée suivante :

$$\mathcal{L}_{total} = 0,6 \cdot \mathcal{L}_{final} + 0,13 \cdot \mathcal{L}_{S4} + 0,13 \cdot \mathcal{L}_{S3} + 0,14 \cdot \mathcal{L}_{S2} \quad (3.2)$$

où :

- $\mathcal{L}_{total}$ représente la perte totale utilisée pour la rétropropagation ;
- $\mathcal{L}_{final}$ désigne la perte calculée sur la prédiction finale du réseau ;
- $\mathcal{L}_{S4}$, $\mathcal{L}_{S3}$ et $\mathcal{L}_{S2}$ correspondent aux pertes calculées sur les sorties auxiliaires intermédiaires du décodeur ;
- 0.6, 0.13, 0.13 et 0.14 sont des coefficients de pondération qui déterminent la contribution relative de chaque sortie à la perte totale d'entraînement.

Cette pondération garantit que l'optimisation demeure principalement axée sur le masque final haute résolution utilisé pour le diagnostic clinique, tout en bénéficiant du guidage complémentaire fourni par les sorties auxiliaires intermédiaires.

3.5.3 Fonction de Perte Combinée

La segmentation de polypes est caractérisée par un fort déséquilibre de classes : les pixels de polype ne représentent généralement qu'une infime fraction de l'image totale par rapport aux pixels d'arrière-plan. Pour pallier ce problème, CANet+ adopte une fonction de perte hybride combinant la perte Focal Tversky, la perte de Dice et la Binary Cross-Entropy pondérée :

$$\mathcal{L} = 0,4 \cdot \mathcal{L}_{FT} + 0,4 \cdot \mathcal{L}_{Dice} + 0,2 \cdot \mathcal{L}_{BCE} \quad (3.3)$$

où :

- $\mathcal{L}$ représente la fonction de perte hybride globale ;
- $\mathcal{L}_{FT}$ désigne la perte Focal Tversky, utilisée pour gérer le déséquilibre entre les classes et mettre l'accent sur les exemples difficiles ;
- $\mathcal{L}_{Dice}$ correspond à la perte de Dice, qui optimise le recouvrement entre le masque prédit et la vérité terrain ;
- $\mathcal{L}_{BCE}$ représente la perte Weighted Binary Cross-Entropy, utilisée pour accorder davantage d'importance aux pixels appartenant au polype ;

- 0.4, 0.4 et 0.2 sont des coefficients de pondération contrôlant la contribution de chaque terme dans la fonction de perte globale.

La perte Focal Tversky met l'accent sur les exemples difficiles et régule l'équilibre entre faux positifs et faux négatifs. La perte de Dice optimise directement le recouvrement entre le masque prédit et la vérité terrain. La BCE pondérée pénalise davantage les pixels de polype non détectés. La combinaison de ces trois termes favorise la convergence du modèle et améliore la précision de segmentation, en particulier sur les lésions planes à faible contraste.

3.5.4 Protocole d'Entraînement

Le modèle est implémenté avec le framework PyTorch et entraîné sur un GPU NVIDIA T4 (16 Go de mémoire VRAM). La précision mixte automatique (Automatic Mixed Precision, AMP) via `torch.cuda.amp` est utilisée pour optimiser l'empreinte mémoire et accélérer le calcul, permettant de maintenir une résolution d'entrée élevée de 512×512 pixels sans surcharge du matériel.

Table 3.4 Hyperparamètres d'entraînement de notre modèle CANet+.

Hyperparamètre	Valeur
Optimiseur	AdamW
Taux d'apprentissage initial	8×10^{-5}
Planificateur de taux d'apprentissage	Cosine Annealing
Décroissance des poids	10^{-4}
Taille de batch	4
Résolution d'entrée	512×512 pixels
Nombre d'époques maximum	100
Arrêt précoce (patience)	15 époques
Stratégie de validation	Validation croisée 4 folds
Écrêtage du gradient	<code>max_norm = 1,0</code>

La taille de batch de 4 résulte d'un choix délibéré, imposé par la complexité architecturale du backbone ConvNeXt-Tiny combinée à la haute résolution d'entrée de 512×512 pixels. Associée à un faible taux d'apprentissage et à l'optimiseur AdamW, cette configuration garantit des mises à jour de gradient stables, préservant l'intégrité des cartes de caractéristiques tout au long de la procédure d'entraînement.

3.6 Expérimentations et Résultats

3.6.1 Métriques d'Évaluation

Les performances de CANet+ sont évaluées quantitativement au niveau du pixel, en comparant les masques prédits aux annotations de vérité terrain. Six métriques complémentaires sont utilisées à cet effet.

3.6.1.1 Coefficient de similarité de Dice

Cette métrique mesure le recouvrement entre la prédiction et la vérité terrain. Elle constitue la référence principale en segmentation médicale en raison de sa robustesse face au déséquilibre de classes (Taha et Hanbury, 2015) :

$$\text{Dice} = \frac{2 \times |P \cap G|}{|P| + |G|} \quad (3.4)$$

sachant que P désigne l'ensemble des pixels prédits par le modèle (masque de segmentation prédit) et G l'ensemble des pixels de la vérité terrain (masque de référence annoté par l'expert). Le terme $|P \cap G|$ correspond au nombre de pixels correctement segmentés (vrais positifs), tandis que $|P|$ et $|G|$ représentent respectivement le cardinal du masque prédit et celui du masque de référence. Le coefficient varie entre 0 (aucun recouvrement) et 1 (segmentation identique à la vérité terrain) '(Taha et Hanbury, 2015)'.

3.6.1.2 Intersection sur Union (IoU / Jaccard)

Cette métrique quantifie le rapport entre l'intersection et l'union des masques prédit P et de référence G . Elle est plus exigeante que le coefficient de Dice pour les prédictions imprécises (Taha et Hanbury, 2015) :

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|} \quad (3.5)$$

3.6.1.3 Précision (Precision)

La précision mesure la proportion de pixels correctement identifiés comme polype parmi l'ensemble des pixels prédits positifs par le modèle. Elle évalue le taux de faux positifs (Taha et Hanbury, 2015) :

$$\text{Précision} = \frac{VP}{VP + FP} \quad (3.6)$$

où VP désigne les Vrais Positifs (pixels de polype correctement détectés) et FP les faux positifs (pixels de fond classés à tort comme polype).

3.6.1.4 Rappel (Recall / Sensibilité)

Le rappel (ou sensibilité) mesure la proportion de polype réels correctement détectés. Cette métrique revêt une importance clinique particulière, car minimiser les polypes manqués est prioritaire dans le dépistage du cancer colorectal (**Taha et Hanbury, 2015**) :

$$\text{Rappel} = \frac{VP}{VP + FN} \quad (3.7)$$

où FN désigne les faux négatifs (pixels de polype non détectés par le modèle).

3.6.1.5 F1-Score

Le F1-Score est la moyenne harmonique de la précision et du rappel. Il constitue une mesure équilibrée particulièrement adaptée aux problèmes de segmentation à classes déséquilibrées (**Taha et Hanbury, 2015**) :

$$F_1 - \text{Score} = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (3.8)$$

3.6.1.6 Exactitude globale (Accuracy)

L'exactitude globale représente le taux de classification correcte de l'ensemble des pixels de l'image, incluant les pixels de fond et de polype (**Taha et Hanbury, 2015**) :

$$\text{Accuracy} = \frac{VP + VN}{VP + VN + FP + FN} \quad (3.9)$$

Tous les masques binaires sont obtenus par seuillage à 0,5 de la sortie sigmoid. De plus, une augmentation au moment du test (Test-Time Augmentation, TTA) comprenant des retournements horizontal et vertical ainsi qu'une rotation de 90° est appliquée lors de l'inférence pour améliorer la robustesse et la stabilité des prédictions.

3.6.2 Sélection de l'architecture de Deep Learning

Avant de valider la contribution individuelle de chaque module par l'étude d'ablation, nous avons conduit une phase de sélection systématique de l'architecture de base. Cette étape préliminaire visait à identifier, parmi plusieurs configurations candidates, la combinaison fondatrice de CANet+. Deux variantes ont été évaluées pour chaque architecture : une version sans supervision profonde (« Avant ») et une version avec supervision profonde (+DS, « Après »). La Table 3.5 présente les résultats obtenus sur Kvasir-SEG, sachant que les configurations marquées +DS intègrent la supervision profonde et la dernière ligne correspond à CANet+ final.

Table 3.5 Comparaison progressive des architectures candidates sur Kvasir-SEG.

Phase	Architecture	Dice	IoU
Avant (sans DS)	U-Net	81 %	72 %
	U-Net++	84 %	75 %
	Attention U-Net	85 %	77 %
	Attention U-Net + ResNet-50	87 %	80 %
Après (+DS)	U-Net + DS	86 %	79 %
	U-Net++ + DS	88 %	80 %
	Attention U-Net + DS	89 %	84 %
	Attention U-Net + ResNet-50 + DS	90 %	85 %
	Att. U-Net + ConvNeXt-Tiny + DS	93 %	87 %

Ces résultats mettent en évidence deux enseignements majeurs. D’une part, l’ajout de la supervision profonde améliore systématiquement les scores Dice et IoU de toutes les architectures testées, confirmant son intérêt indépendant de la configuration du backbone. D’autre part, la substitution de l’encodeur ResNet-50 par ConvNeXt-Tiny produit le gain le plus significatif (+0,03 en Dice), justifiant le choix de CANet+ comme architecture finale.

3.6.3 Analyse de l’impact des différents composants du modèle proposé

L’étude d’ablation répond à une question fondamentale en apprentissage profond : chaque composant ajouté à l’architecture contribue-t-il réellement à l’amélioration des performances, ou certains modules sont-ils redondants ?

Plus précisément, cette démarche vise à :

- Valider l’hypothèse de conception : la synergie entre CBAM, ASPP et supervision profonde est nécessaire.
- Quantifier l’apport marginal de chaque module (gain incrémental).
- Éliminer les biais de sur-ajustement : vérifier que les gains ne proviennent pas d’un sur-paramétrage accidentel.

Pour valider la contribution individuelle de chaque composant de CANet+, une étude d’ablation systématique est conduite selon une stratégie d’ajout incrémental, à partir du modèle de référence (baseline) constitué de l’encodeur ConvNeXt-Tiny couplé au décodeur Attention U-Net.

CBAM, ASPP et la supervision profonde sont ensuite intégrés progressivement. Toutes les évaluations sont réalisées par validation croisée à 4 folds sur les trois bases de données. La Table 3.6 rapporte les scores Dice et IoU sous forme de moyenne $\pm$ écart-type.

Table 3.6 Étude d’ablation : contribution individuelle des composants de CANet+ (Dice $\pm \sigma$, IoU $\pm \sigma$) sur les trois jeux de données.

Jeu de données	Configuration	Dice	IoU	Gain Dice
Kvasir-SEG	Baseline (ConvNeXt + Att-UNet)	$0,892 \pm 0,012$	$0,835 \pm 0,015$	—
	+ CBAM	$0,907 \pm 0,010$	$0,852 \pm 0,013$	+0,015
	+ ASPP	$0,918 \pm 0,009$	$0,865 \pm 0,011$	+0,011
	+ Supervision profonde	$0,926 \pm 0,008$	$0,874 \pm 0,010$	+0,008
	CANet+ complet	$0,942 \pm 0,007$	$0,894 \pm 0,009$	+0,050
CVC-ClinicDB	Baseline (ConvNeXt + Att-UNet)	$0,915 \pm 0,010$	$0,862 \pm 0,013$	—
	+ CBAM	$0,929 \pm 0,009$	$0,878 \pm 0,011$	+0,014
	+ ASPP	$0,937 \pm 0,008$	$0,887 \pm 0,010$	+0,008
	+ Supervision profonde	$0,945 \pm 0,007$	$0,896 \pm 0,009$	+0,008
	CANet+ complet	$0,966 \pm 0,006$	$0,918 \pm 0,008$	+0,051
CVC-ColonDB	Baseline (ConvNeXt + Att-UNet)	$0,878 \pm 0,014$	$0,819 \pm 0,016$	—
	+ CBAM	$0,893 \pm 0,012$	$0,836 \pm 0,014$	+0,015
	+ ASPP	$0,904 \pm 0,011$	$0,849 \pm 0,013$	+0,011
	+ Supervision profonde	$0,913 \pm 0,010$	$0,859 \pm 0,012$	+0,009
	CANet+ complet	$0,928 \pm 0,009$	$0,879 \pm 0,011$	+0,050

L’étude d’ablation démontre que l’efficacité de CANet+ résulte de l’intégration progressive et complémentaire de ses composants, plutôt que d’un module isolé. Sur Kvasir-SEG, le score Dice progresse de 0,892 pour le modèle de référence à 0,942 pour CANet+ complet, soit un gain global de +0,050. Une tendance similaire est observée sur CVC-ClinicDB (de 0,915 à 0,966, gain +0,051) et sur CVC-ColonDB (de 0,878 à 0,928, gain +0,050). Ces gains cohérents sur les trois bases de données confirment la complémentarité des choix de conception adoptés.

En résumé, l’étude d’ablation (Table 3.6) démontre que chaque composant de CANet+ contribue de manière additive et complémentaire à l’amélioration des performances. L’intégration du module CBAM apporte le gain marginal le plus important (+0,014 à +0,015 en Dice), validant l’hypothèse que la focalisation attentionnelle canal-spatiale est cruciale pour supprimer le bruit endoscopique. L’ajout subséquent du module ASPP (+0,008 à +0,011) confirme l’utilité de la modélisation contextuelle multi-échelle face à la variabilité morphologique des polypes. Enfin, la supervision profonde, bien qu’apportant un gain numérique plus modeste (+0,008

à $+0,009$), réduit significativement la variance inter-folds (σ divisée par $\sim 1,5$), attestant d'une meilleure stabilité et généralisation. La configuration complète atteint un score Dice de $0,942 \pm 0,007$ sur Kvasir-SEG, surpassant la baseline de $+0,050$, avec une cohérence remarquable sur les trois benchmarks. Une analyse qualitative complémentaire est présentée dans la figure suivante afin de mieux visualiser les différences entre les approches.

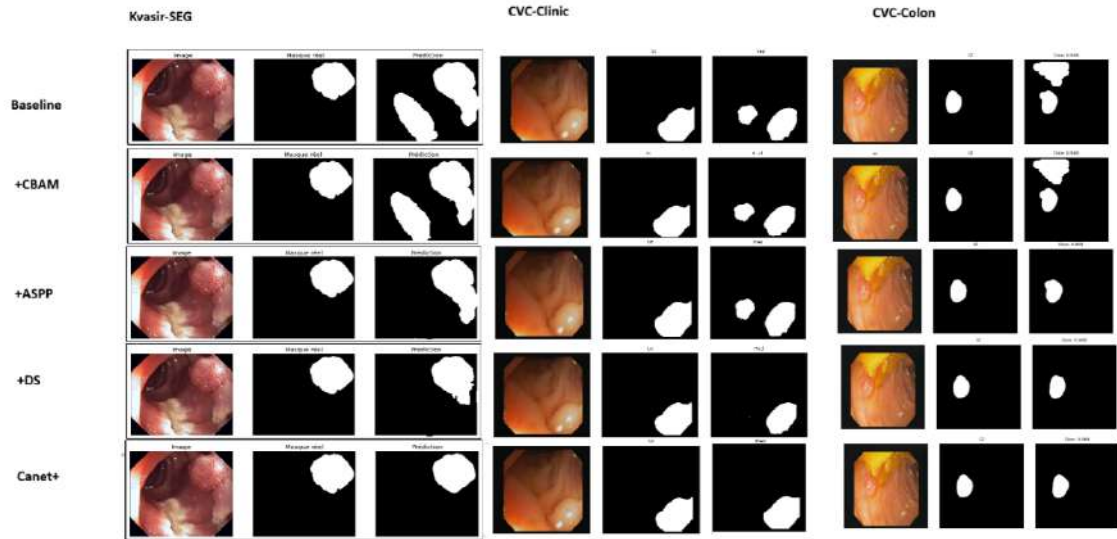


Figure 3.14 Comparaison qualitative des prédictions des architectures avec les masques de vérité terrain sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB.

La Figure 3.14 met en évidence une amélioration progressive des performances de segmentation suite à l'intégration des différents composants proposés dans l'architecture. Par rapport au modèle de base, l'ajout successif des modules CBAM, ASPP et DS permet d'affiner la détection des polypes et d'améliorer la correspondance des masques prédits avec les masques de vérité terrain. Cette tendance est observée sur l'ensemble des bases de données (Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB), où les contours deviennent plus précis et les régions mal segmentées sont progressivement réduites, ce qui confirme l'apport complémentaire de chaque composant dans les performances globales du modèle.

3.6.4 Résultats Qualitatifs

La Figure 3.15 présente une comparaison qualitative des prédictions de CANet+ avec les masques de vérité terrain sur les trois benchmarks. Les images superposées confirment un fort accord spatial entre les régions prédites (en rouge) et les annotations de référence (en blanc), avec un minimum de faux positifs et de polypes manqués, y compris dans les cas difficiles tels que les polypes plans à faible contraste sur CVC-ColonDB et les lésions adjacentes à des reflets spéculaires sur CVC-ClinicDB.

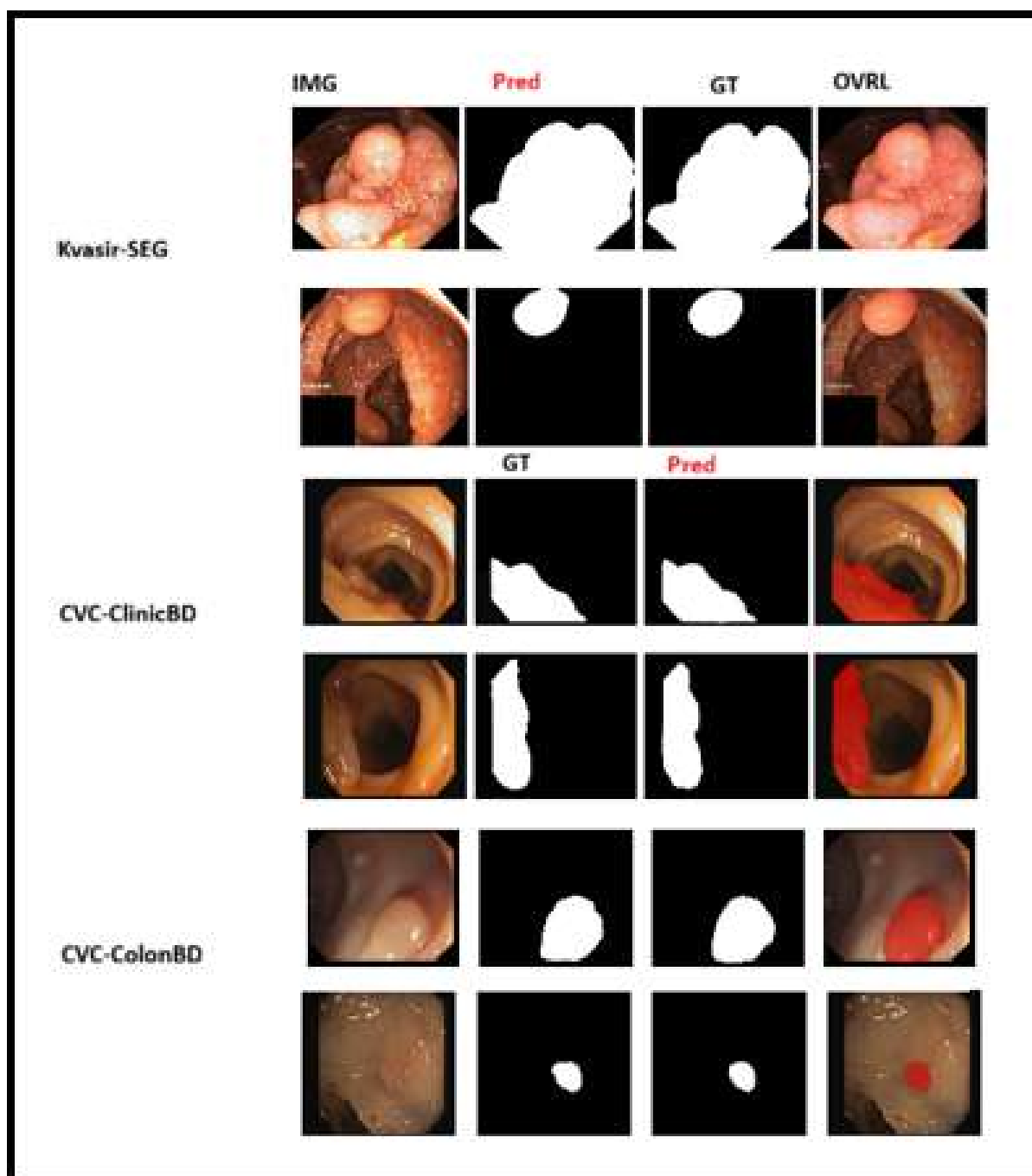


Figure 3.15 Comparaison qualitative des prédictions de CANet+ avec les masques de vérité terrain sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB

3.6.5 Résultats Quantitatifs et comparaison

À l'issue de l'étude d'ablation, l'approche retenue CANet+ a été validée. Les résultats obtenus sont présentés dans la table suivante :

Table 3.7 Résultats expérimentaux de CANet+ sur les jeux de données (Kvasir-SEG, CVC-ClinicDB, CVC-ColonDB).

Métrique	Kvasir-SEG	CVC-ClinicDB	CVC-ColonDB
Dice	0,94	0,96	0,92
IoU	0,89	0,92	0,87
Précision	0,92	0,95	0,91
Rappel	0,96	0,96	0,94
F1-Score	0,94	0,96	0,92
Accuracy	0,97	0,98	0,97

Les résultats montrent que CANet+ obtient de bonnes performances sur les trois jeux de données. Les scores de Dice et d'IoU traduisent une bonne correspondance entre les prédictions et les masques de référence. De même, la précision et le rappel indiquent un équilibre satisfaisant entre la détection correcte des polypes et la limitation des erreurs. Enfin, le F1-score et l'Accuracy confirment la robustesse globale ainsi que la capacité de généralisation de la méthode.

3.6.5.1 Analyse des courbes d'apprentissage

Afin d'analyser la convergence du modèle, les courbes d'entraînement sont présentées ci-dessous.

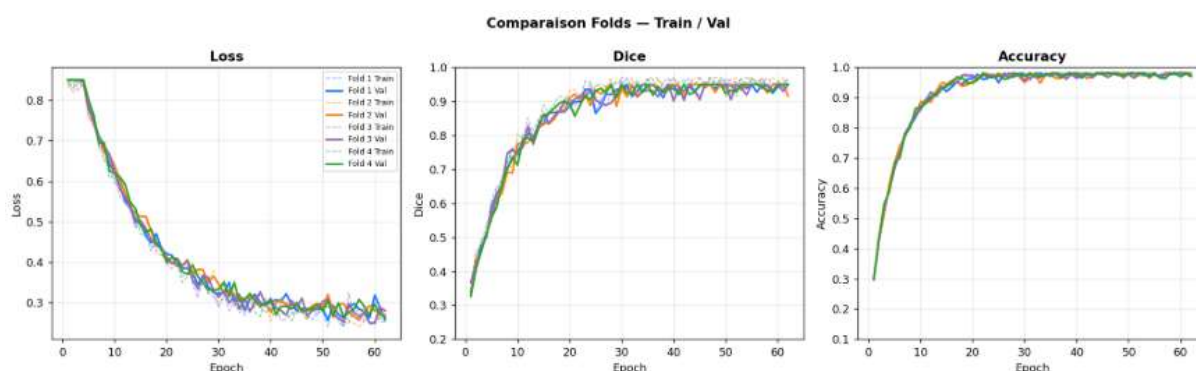


Figure 3.16 Comparaison des performances obtenues sur les 4 folds

a) Convergence Les courbes montrent une convergence stable avec une diminution progressive de la perte et une amélioration continue des métriques Dice et Accuracy.

b) Capacité du généralisation Le faible écart entre les courbes d'entraînement et de validation confirme une bonne capacité de généralisation et un risque limité de surapprentissage.

c) Capacité du modèle Ces résultats démontrent que CANet+ apprend efficacement les caractéristiques des polypes et produit des représentations discriminantes.

3.6.5.2 Analyse des performances sur l'ensemble de test

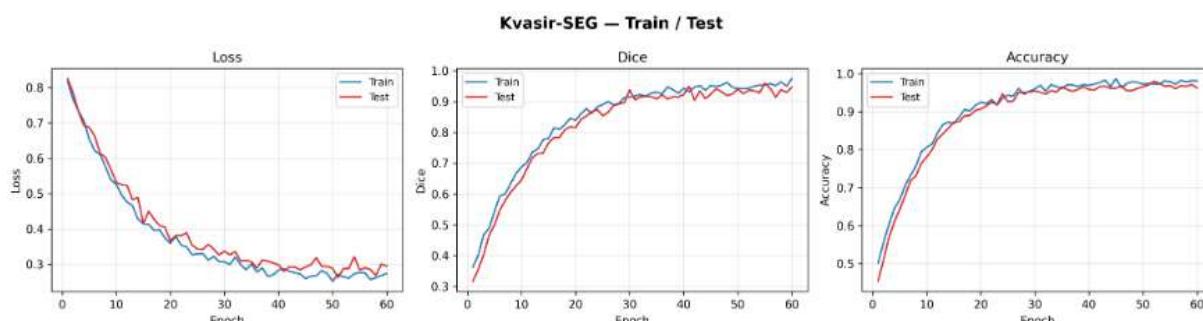


Figure 3.17 Évolution des métriques d'entraînement sur Kvasir-SEG.

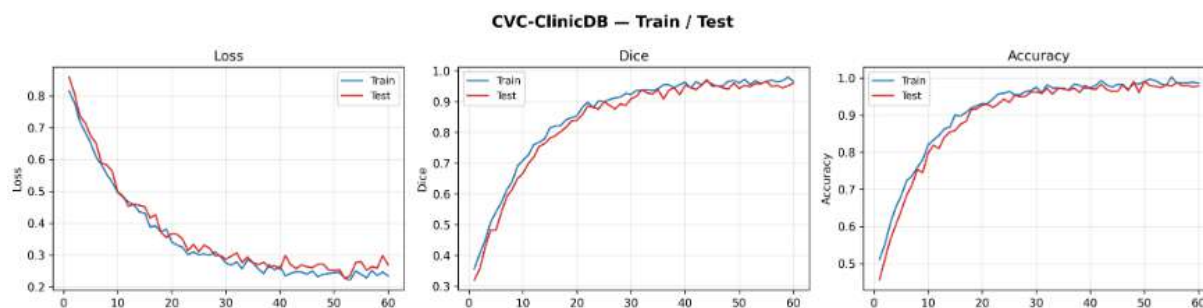


Figure 3.18 Évolution des métriques d'entraînement sur CVC-ClinicDB.

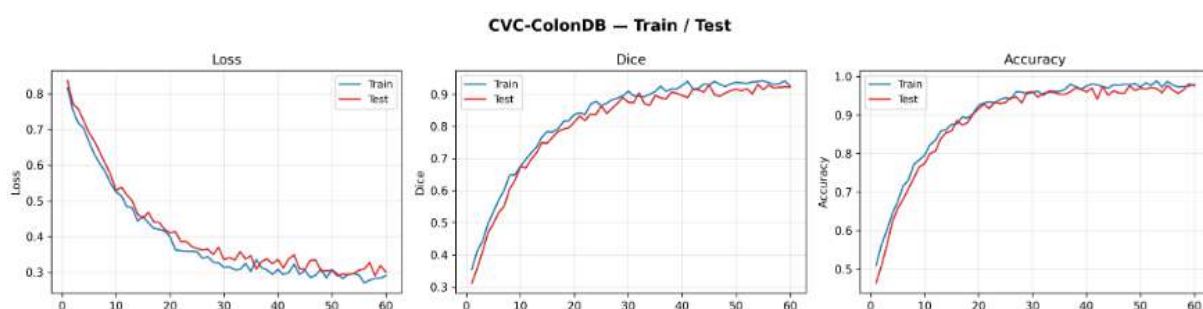


Figure 3.19 Évolution des métriques d'entraînement sur CVC-ColonDB.

a) Robustesse Les performances obtenues sur les données de test confirment la robustesse de CANet+ face aux variations des images endoscopiques.

b) Généralisation Les résultats restent élevés sur les trois bases de données, démontrant une bonne capacité de généralisation sur des données inédites.

3.6.5.3 Analyse des courbes ROC

Les courbes ROC illustrent la capacité discriminante du modèle sur chacun des jeux de données.

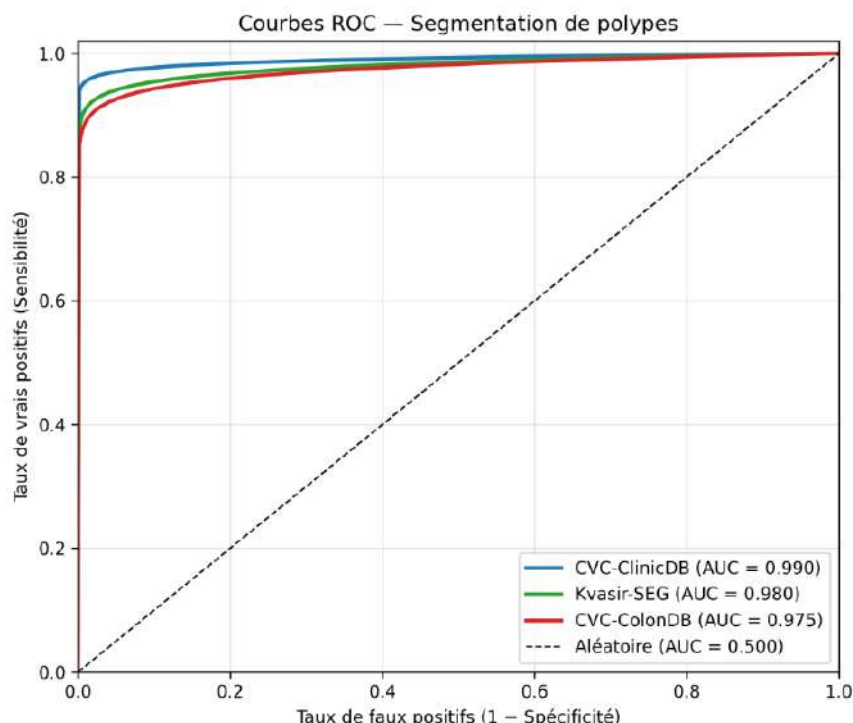


Figure 3.20 Courbes ROC obtenues sur Kvasir-SEG, CVC-ClinicDB et CVC-ColonD

a) Capacité discriminante Les courbes ROC mettent en évidence une excellente capacité de discrimination entre les régions de polype et l'arrière-plan.

b) Performance globale Les valeurs élevées de l'AUC confirment les performances globales et la fiabilité de CANet+ sur les trois jeux de données.

3.6.5.4 Comparaison avec les méthodes de l'état de l'art

Afin de prouver l'efficacité de notre approche, nous avons effectué une comparaison avec quelques travaux de la littérature. Les performances de CANet+ sont comparées à celles de plusieurs méthodes de référence : des approches purement convolutives U-Net (**Ronneberger et al., 2015**), U-Net++ (**Zhou et al., 2020**), ResU-Net++ (**Jha et al., 2019**) et PraNet (**Fan et al., 2020**) ainsi que des architectures hybrides ou basées sur des transformers plus récentes : TransFuse-L (**Zhang et al., 2021**), Polyp-PVT (**Dong et al., 2021**) et ColonFormer-L (**Duc et al., 2022**). La table 3.8 présente les résultats comparatifs sur les trois benchmarks.

Table 3.8 Résultats comparatifs de CANet+ avec quelques méthodes de l'état de l'art sur les jeux de données Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB

Méthode	Kvasir mDice	Kvasir mIoU	ClinicDB mDice	ClinicDB mIoU	ColonDB mDice	ColonDB mIoU
U-Net	0,81	0,74	0,82	0,75	0,51	0,44
U-Net++	0,82	0,74	0,79	0,72	0,48	0,41
ResU-Net++	0,81	0,79	0,79	0,79	—	—
PraNet	0,89	0,84	0,89	0,84	0,70	0,64
TransFuse-L	0,92	0,87	0,94	0,89	0,78	0,70
Polyp-PVT	0,91	0,86	0,93	0,88	0,80	0,72
ColonFormer-L	0,92	0,87	0,93	0,88	0,81	0,73
CANet+ (notre méthode)	0,94	0,89	0,96	0,92	0,92	0,87

CANet+ obtient les meilleures performances parmi les méthodes comparées. Sur Kvasir-SEG, il atteint un score Dice de 0,94 et un IoU de 0,89, avec un rappel de 0,96, attestant une détection fiable des régions polypes. Sur CVC-ClinicDB, il obtient ses meilleures performances avec un Dice de 0,96 et une exactitude globale de 0,98, démontrant sa robustesse sur des séquences vidéo cliniques. Sur CVC-ColonDB la base de données le plus exigeant il maintient un Dice compétitif de 0,92 et un rappel de 0,94, confirmant une excellente capacité de généralisation.

3.6.6 Règles de détermination du stade du polype

Dans une étape d'interprétation des résultats obtenus, et afin d'améliorer la présentation du diagnostic, nous nous sommes appuyés sur l'état de l'art 1.5.1 ainsi que sur notre passage à l'hôpital Ahmed Madaghri pour établir un ensemble de règles permettant de déterminer le stade du polype. Les règles établies sont les suivantes :

- **a) Localisation anatomique :**

$$L = \begin{cases} \text{Côlon,} & \text{si le polype est localisé dans le côlon} \\ \text{Rectum,} & \text{si le polype est localisé dans le rectum} \end{cases} \quad (3.10)$$

où L représente la localisation du polype.

- **b) Classification selon la taille :**

$$P(d) = \begin{cases} \text{Diminutif,} & d < 5 \text{ mm} \\ \text{Petit,} & 5 \leq d < 10 \text{ mm} \\ \text{Intermédiaire,} & 10 \leq d < 20 \text{ mm} \\ \text{Large,} & d \geq 20 \text{ mm} \end{cases} \quad (3.11)$$

où :

- d désigne le diamètre maximal du polype (en mm) ;
- $P(d)$ représente la catégorie de taille associée.

L'évaluation d'un polype colorectal repose principalement sur deux critères fondamentaux : la localisation anatomique et la taille.

3.6.7 Discussion

3.6.7.1 Synergie entre modélisation multi-échelle et focalisation attentionnelle

L'une des forces majeures de CANet+ réside dans la synergie entre le module ASPP et les portes d'attention. En coloscopie, la taille des polypes peut varier considérablement, des lésions planes minuscules aux masses proéminentes volumineuses. Le module ASPP permet au réseau de percevoir simultanément le contexte à plusieurs échelles, garantissant qu'aucun polype ne soit manqué quelle que soit sa taille. Parallèlement, les portes d'attention agissent comme un filtre sélectif, neutralisant les reflets spéculaires sur les parois intestinales humides, souvent confondus avec des polypes par les modèles U-Net classiques.

3.6.7.2 Stabilité de l'entraînement par la supervision profonde

La supervision profonde appliquée aux branches S_2 , S_3 et S_4 s'est révélée indispensable à la stabilité de l'entraînement. En calculant la perte à plusieurs profondeurs du réseau, le signal de gradient demeure fort jusqu'aux couches les plus profondes de l'encodeur, ce qui résout efficacement le problème du gradient évanescent et conduit à des masques de segmentation cohérents, même pour des images à faible contraste. (Ullah et al., 2025) ; (Lee et al., 2015)

3.6.7.3 Pertinence clinique et perspectives

Avec un rappel compris entre 0,94 et 0,96 sur les trois benchmarks, CANet+ présente un fort potentiel d'utilisation comme outil d'aide à la décision clinique, susceptible de réduire le taux de polypes manqués - principale cause des cancers colorectaux d'intervalle. Ce potentiel est d'autant plus pertinent dans des contextes hospitaliers à ressources limitées, tels que Ahmed madaghri d'Ain-Temouchent, où les systèmes CAD peuvent constituer un appui précieux pour les endoscopistes face à des volumes croissants de coloscopies. La complexité architecturale du backbone ConvNeXt-Tiny nécessite toutefois des ressources GPU importantes à la résolution 512×512 . Les travaux futurs porteront sur la quantification et l'élégage du modèle en vue d'un déploiement en temps réel sur le matériel hospitalier standard.

3.7 Implémentation , outils utilisés et présentation de l'application

Après la conception de l'architecture CANet+ et la validation expérimentale du modèle de segmentation présentées dans les sections précédentes, cette section décrit la concrétisation logicielle de l'approche proposée sous la forme d'une application complète, dénommée ColonCare. L'objectif est de transformer le modèle entraîné un fichier de poids issu de la validation croisée en un outil opérationnel et reproductible, exploitable à la fois dans un contexte clinique et dans un contexte de recherche.

3.7.1 Matériel utilisé

Deux phases doivent être distinguées, car elles n'imposent pas les mêmes contraintes matérielles : la phase d'**entraînement** du modèle et la phase de **déploiement** de l'application.

L'entraînement et la validation croisée du réseau CANet+ ont été menés sur la plateforme **Google Colab**, en s'appuyant sur un accélérateur graphique **NVIDIA Tesla T4**. Ce processeur graphique prend en charge le calcul en précision mixte, qui réduit l'empreinte mémoire et accélère l'apprentissage d'une architecture profonde à encodeur ConvNeXt-Tiny sur des images de résolution 512×512 pixels. La mémoire vidéo disponible a conditionné le choix de la taille de lot (*batch size*) ainsi que le recours à des techniques d'optimisation mémoire durant l'apprentissage.

Les besoins matériels de la phase de déploiement ont, au contraire, été maintenus modestes afin de garantir l'accessibilité de l'outil dans un environnement hospitalier ou universitaire ordinaire. ColonCare s'exécute sur une machine locale standard fonctionnant sous Windows. L'application sélectionne automatiquement le périphérique de calcul disponible, privilégiant le processeur graphique compatible CUDA lorsqu'il est présent et basculant de manière transparente sur le processeur central (CPU) dans le cas contraire. Ce mécanisme assure le fonctionnement de l'outil même en l'absence de carte graphique dédiée, au prix d'un temps d'inférence plus élevé.

3.7.2 Langages de programmation et technologies

L'application a été intégralement développée en **Python**, l'un des langages les plus largement adoptés en apprentissage profond et en analyse d'images médicales (**Python Software Foundation, 2026**), en raison de la richesse de son écosystème scientifique. Le choix des bibliothèques a été guidé par la nécessité de couvrir l'ensemble de la chaîne de traitement, depuis le chargement de l'image jusqu'à la production du rapport clinique.

L'apprentissage profond et l'inférence reposent sur **PyTorch** et **torchvision**, ce dernier fournissant notamment l'encodeur ConvNeXt-Tiny pré-entraîné sur ImageNet, qui constitue le tronc de l'architecture. Le traitement des images est assuré par **OpenCV** et **Pillow**, tandis que les transformations et les augmentations de données s'appuient sur **Albumentations**. Le calcul numé-

rique repose sur **NumPy** et **SciPy**, cette dernière étant en particulier mobilisée pour le calcul des distances de surface utilisées dans les métriques de frontière (HD95 et ASSD) via la transformée de distance euclidienne. Les indicateurs d'évaluation et les courbes ROC/AUC sont obtenus à l'aide de **scikit-learn**, et les graphiques d'évaluation sont produits avec **Matplotlib**. La gestion tabulaire des patients et des examens s'appuie sur **pandas**, et la génération des rapports cliniques au format PDF est réalisée avec **fpdf2**.

L'interface utilisateur est construite avec le framework **Gradio**, qui permet de générer une application web interactive directement à partir du code Python, sans développement d'une couche de présentation distincte. Ce choix accélère le prototypage et rend l'application accessible depuis un simple navigateur web, sans installation cliente complexe. L'ensemble des dépendances est consigné dans le fichier de configuration des dépendances, garantissant la reproductibilité de l'environnement logiciel.

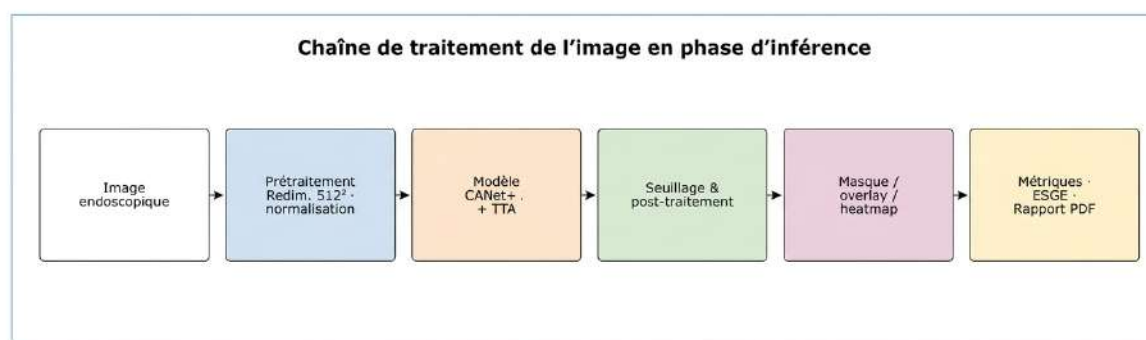


Figure 3.21 Chaîne de traitement de l'image en phase d'inférence, du chargement jusqu'à la production des résultats cliniques et techniques.

3.7.3 Environnement de développement

Le développement a été réalisé dans l'éditeur **Visual Studio Code (Microsoft, 2026)**, retenu pour sa légèreté, son support natif de Python et ses capacités de débogage intégrées. Le projet est organisé selon une structure modulaire regroupant le script principal de l'application, le fichier des dépendances, le dossier hébergeant les poids du modèle entraîné et la configuration de lancement.

L'isolation des dépendances est assurée par un **environnement virtuel Python (Python Software Foundation, 2026)**, dans lequel l'ensemble des bibliothèques requises est installé. Cette pratique garantit la cohérence des versions utilisées et évite toute interférence avec les autres projets de la machine.

L'exécution s'appuie sur une configuration de lancement paramétrable : le chemin du fichier de poids est externalisé au moyen d'une variable d'environnement, ce qui rend l'application portable d'une machine à l'autre sans modification du code source. Au démarrage, l'application vérifie la présence du modèle, charge les poids, instancie le réseau en mode évaluation, puis met

l'interface à disposition de l'utilisateur via le navigateur. Les procédures détaillées d'installation et de lancement sont reportées en annexe.

3.7.4 Présentation de l'interface de l'application

L'application ColonCare s'organise autour de deux interfaces principales (voir Figure 3.22), accessibles par onglets, conformément à la séparation entre usage clinique et usage de recherche évoquée plus haut. Un en-tête commun rappelle le nom du modèle chargé et le périphérique de calcul utilisé, il s'agit de :

a) Interface médicale : pensée pour le praticien et centrée sur le flux de travail diagnostique (gestion des patients, acquisition de l'image, segmentation, synthèse clinique et génération d'un rapport).

b) Une interface technique : destinée au chercheur et exposant les mesures quantitatives, les courbes d'évaluation et les outils d'interprétabilité du réseau. Cette séparation permet à l'interface clinique de masquer délibérément les indicateurs techniques (Dice, IoU, courbes ROC, cartes de chaleur) afin de préserver la lisibilité du flux diagnostique, tandis que le chercheur conserve un accès complet à l'instrumentation du modèle.

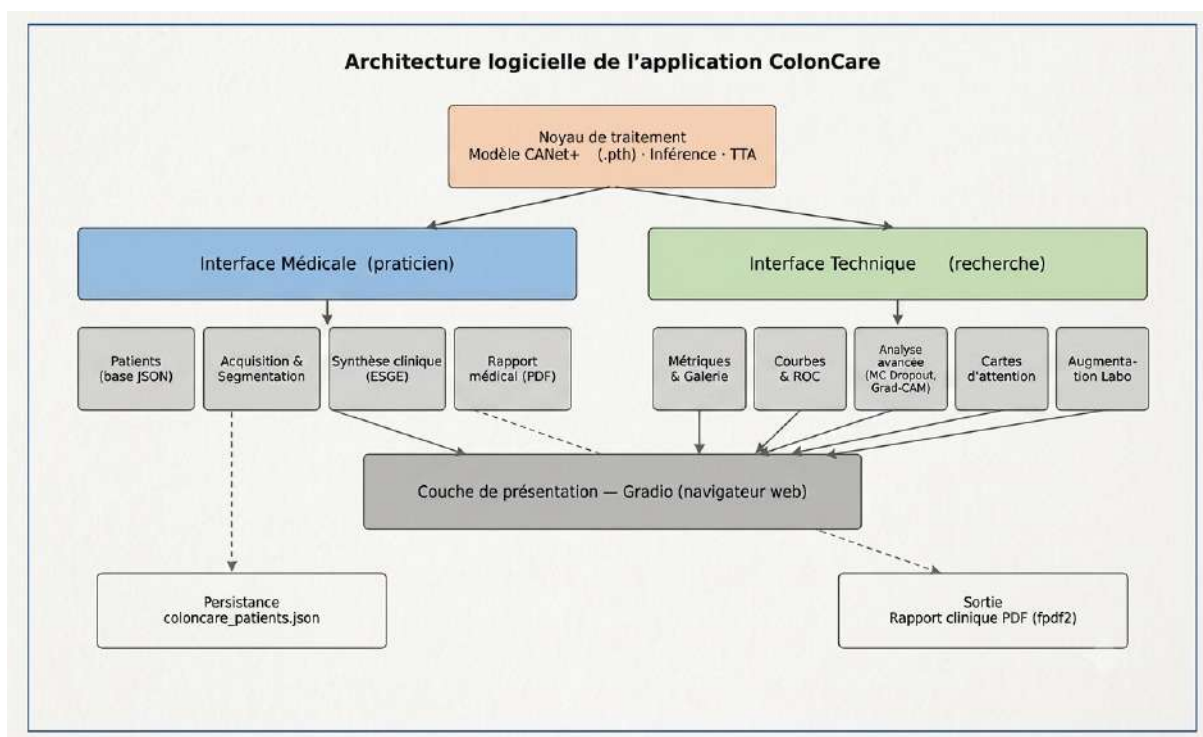
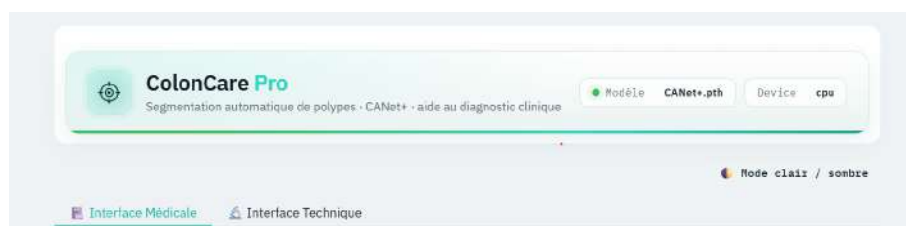
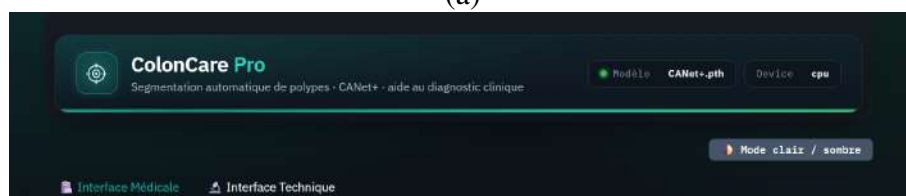


Figure 3.22 Architecture logicielle de l'application ColonCare : noyau de traitement commun et double interface.

Pour accéder à ces deux interfaces, il faut passer par la fenêtre d'accueil de l'application ColonCare Pro, qui affiche le modèle chargé (CANet+.pth) ainsi que le périphérique d'exécution (CPU). Cette fenêtre est présentée par la figure suivante :



(a)



(b)

Figure 3.23 En-tête de l'application ColonCare Pro affichage du modèle et sélection de l'interface(a :mode clair, b : mode sombre)

Dans ce qui suit, nous présentons de manière plus détaillée chacune de ces deux interfaces.

a) Interface médicale : Destinée au médecin, elle est subdivisée en quatre onglets correspondant aux étapes successives du flux diagnostique :

- **Gestion des patients :** module de gestion permettant la création de fiches patients (nom, âge, adresse) et l'affichage de la liste des patients enregistrés. Les données sont conservées de manière persistante, chaque patient se voyant attribuer un identifiant unique et un historique d'examens. La figure 3.24 représente l'onglet patients de l'Interface Médicale, qui comporte un formulaire de création d'un nouveau patient (Nom, Âge, Téléphone) et liste des patients enregistrés avec leur identifiant unique, âge et nombre d'examens.

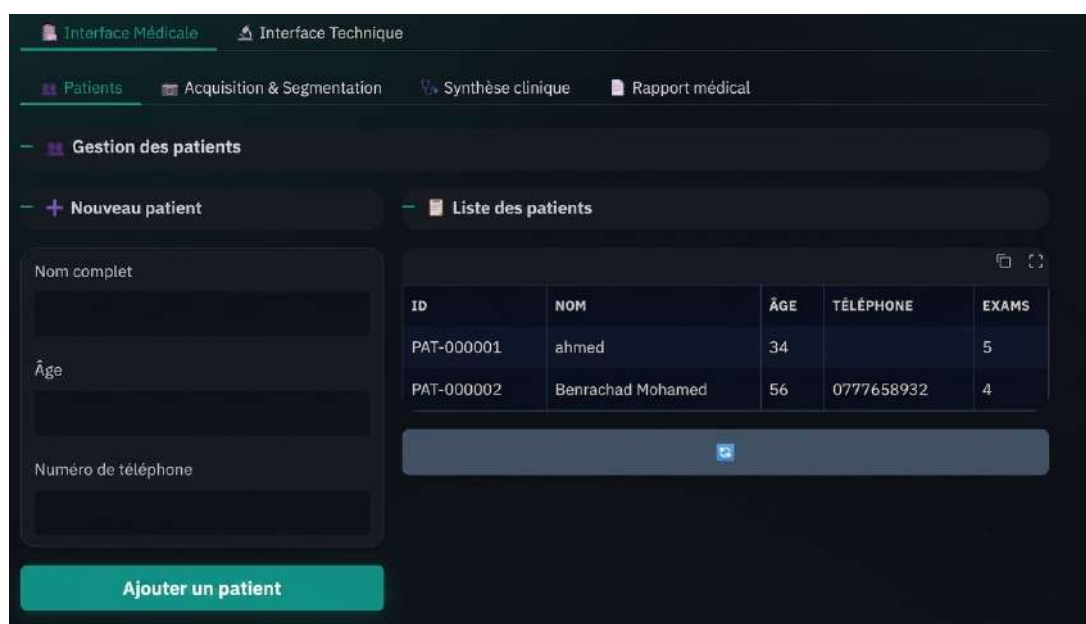


Figure 3.24 Interface Médicale Onglet Gestion des patients : création et liste des patients.

- **Acquisition et Segmentation** : le médecin procède au chargement de l'image endoscopique et exécution de la segmentation. Le résultat est présenté au praticien sous une forme purement visuelle (image originale, masque de la lésion, superposition), sans aucun indicateur technique, afin de préserver la lisibilité clinique. Voir la figure 3.25.

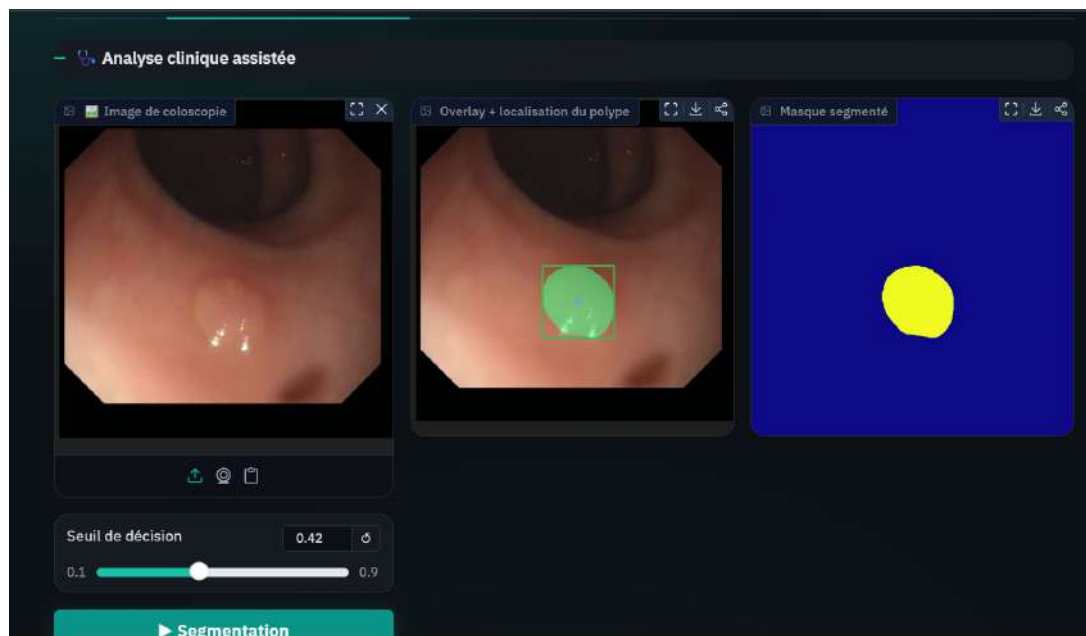


Figure 3.25 Interface Médicale Onglet Acquisition & Segmentation.

- **Synthèse clinique** : module de diagnostic assisté combinant le résultat de la segmentation à une classification inspirée des recommandations ESGE (European Society of Gastrointestinal Endoscopy). À partir de la taille estimée de la lésion et de l'analyse morphologique du masque, le système propose une catégorie de taille, un niveau de risque indicatif et une suggestion de suivi, ainsi qu'une indication morphologique rattachée à la classification de Paris. Toutes ces suggestions sont explicitement présentées comme devant être confirmées par le médecin.

La figure 3.26 représente l'onglet Synthèse clinique affichant les résultats de l'analyse IA, une présentation d'une étude d'un cas où il est question d'un polype détecté, de taille estimée 6.92 mm, avec un niveau de risque intermédiaire et une confiance de 97.7 %, une localisation centre-centre, avec morphologie et conduite à tenir. Notons toutefois que les règles diagnostiques ont été abordées dans les sections 1.5.1 et 3.6.6, permettant de déterminer le stade du polype en fonction de sa taille, de son niveau de risque et de sa catégorisation.



Figure 3.26 Interface Médicale : Synthèse clinique générée automatiquement après segmentation (application des règles diagnostiques).

- **Rapport médical** : Une fois le diagnostic établi, un compte rendu clinique au format PDF est généré automatiquement, intégrant les informations du patient, les résultats de l'analyse, la catégorie clinique ESGE et la conclusion du praticien.

La figure 3.27 présente l'onglet de génération du rapport médical, où l'on procède d'abord à la sélection du patient par son ID, le médecin devrait par la suite saisir son nom ainsi que la conclusion clinique. Le rapport PDF est généré automatiquement en cliquant sur le bouton dédié.

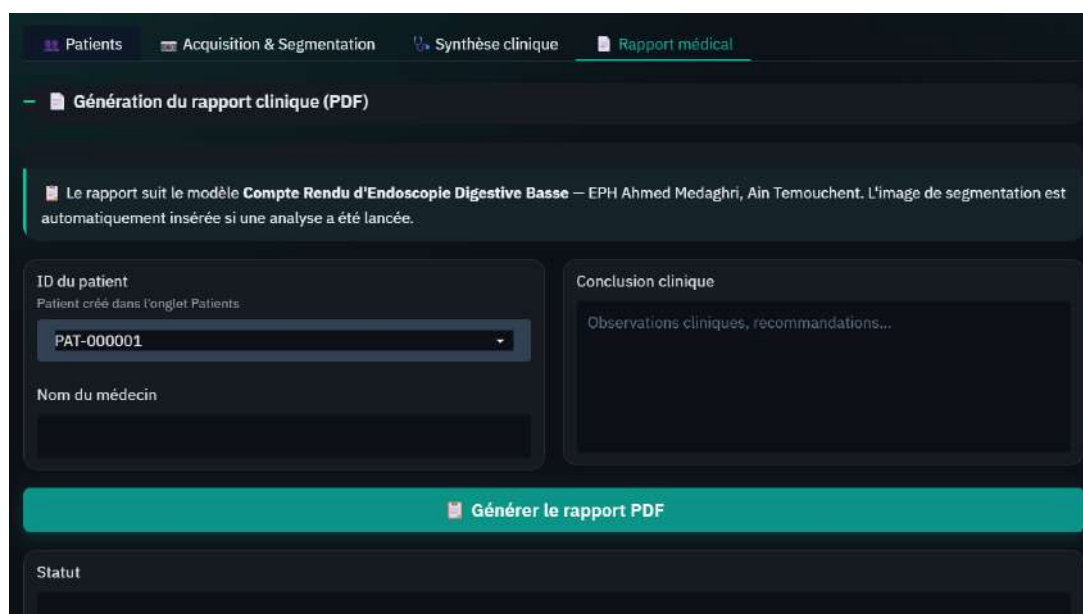


Figure 3.27 Interface Médicale : Génération du rapport clinique en PDF.

La figure présentée ci-dessous représente un exemple de rapport clinique en PDF, généré automatiquement par notre application ColonCare Pro, conforme au modèle du Compte Rendu d'Endoscopie Digestive Basse de l'Établissement Public Hospitalier Ahmed Me-

daghri (Ain Temouchent). Le rapport intègre l'image de segmentation, la synthèse et la conclusion du médecin.

ETABLISSEMENT PUBLIC HOSPITALIER SERVICE D'HEPATO-GASTRO-ENTEROLOGIE	COMPTE RENDU D'ENDOSCOPIE DIGESTIVE BASSE DATE : 19/06/2026
---	---

Type d'examen : Coloscopie Nom de l'opérateur : Dr. Ben Amira	Patient : Nom : Benrachad Mohamed Date de naissance : 56 ans N deg de fiche navette : PAT-000002 Adresse par :
--	--

Indication / Histoire clinique :

EXAMEN PROCTOLOGIE



Polypes detectes : 1 | Taille estimée : 6.92 mm | Confiance : 98%

ANALYSE - SYNTHESE

Polypes detectes :	1
Localisation :	central-centre
Taille estimée :	6.92 mm []
Categorie (ESGE) :	Petite (6-9 mm)
Niveau de risque :	Intermediaire
Conduite a tenir :	Suggestion : coloscopie de suivi a 3 ans - a confirmer par le medecin

IMPORTANT : La taille est ESTIMEE par ratio pixel - calibration endoscopique requise. Cet outil est une aide à la decision. NON un diagnostic autonome.

CONCLUSION

Presence d'une lesion polypoide visualisee au niveau colique et correctement detectee par le systeme de segmentation. Une confirmation histopathologique demeure necessaire pour etabliir le diagnostic definitif.

CACHET ET SIGNATURE



ColonCare Pro - Outil d'aide à la decision - NON destine à un usage clinique autonome

Figure 3.28 Rapport clinique PDF généré par ColonCare Pro conforme au modèle EPH Ahmed Madaghri.

b) Interface technique Destinée au chercheur, elle expose l'instrumentation complète du modèle à travers cinq onglets :

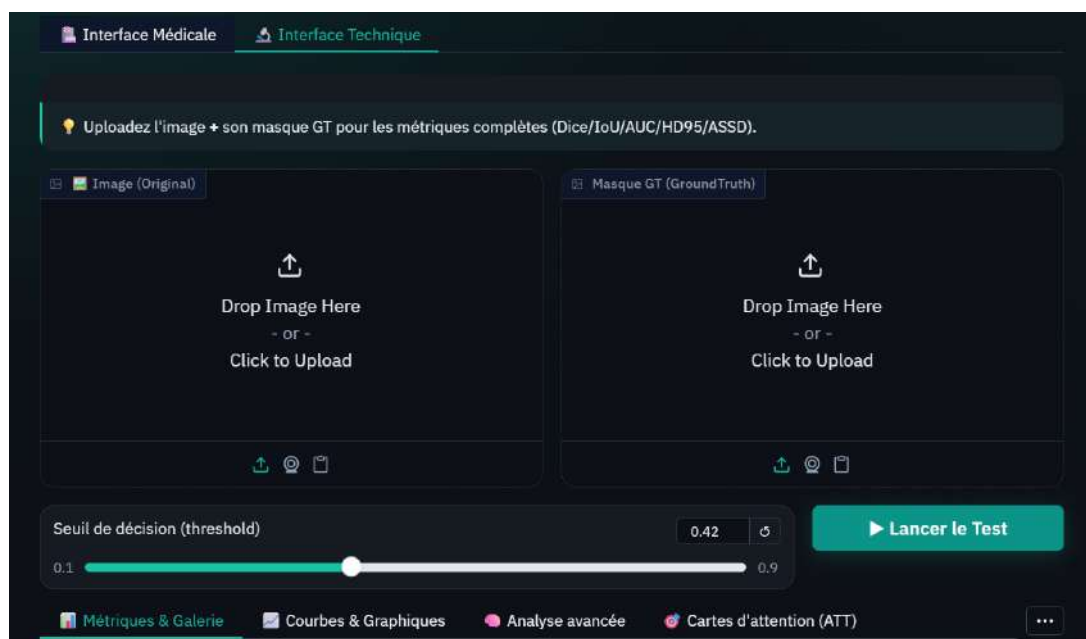


Figure 3.29 Interface technique.

- **Métriques & Galerie** : lorsqu'un masque de vérité terrain accompagne l'image, le système calcule l'ensemble des indicateurs quantitatifs (Dice, IoU, précision, rappel, F1, exactitude) ainsi que les métriques de frontière HD95 et ASSD. Une galerie comparative met en évidence les vrais positifs, faux positifs et faux négatifs, l'ensemble étant calculé à la résolution 512×512 . Onglet Métriques et Galerie de l'Interface Technique (avec masque GT) représente une galerie composée du masque Ground Truth, de l'image originale, du masque prédit, de l'overlay et de la carte TP/FP/FN colorée (voir figure 3.30). Il permet également de visualiser les résultats de segmentation évalués à l'aide des métriques présentées dans 3.6.1.

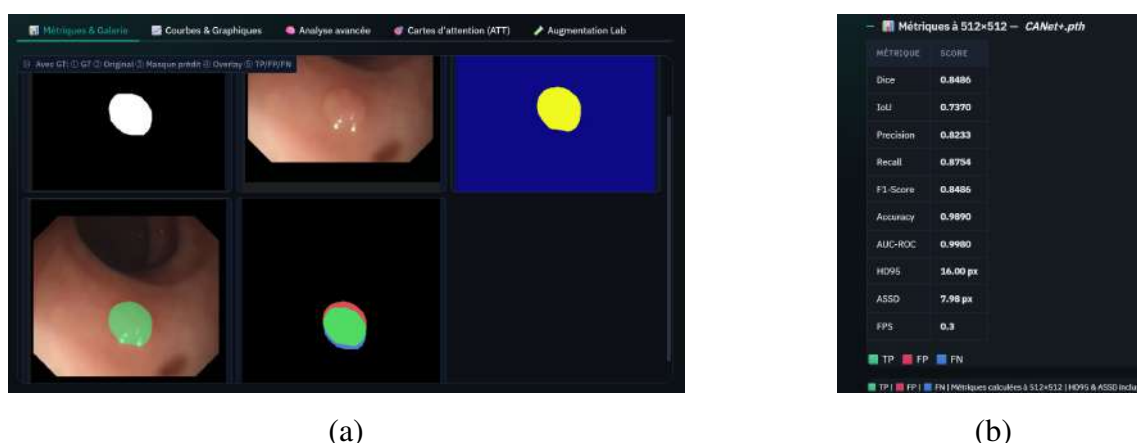


Figure 3.30 Interface Technique : (a) Métriques de segmentation et galerie GT/prédiction. (b) Table des métriques à 512×512 (CANet+).

- **Courbes & Graphiques** : visualisation des courbes d'évaluation, en particulier la courbe ROC et l'aire sous la courbe (AUC), ainsi que les histogrammes de métriques. La figure

3.31 .(a) ci-dessous illustre le Diagramme en barres des métriques de segmentation et courbe ROC (AUC=0.9980) tracée par rapport à la référence aléatoire. Les lignes pointillées indiquent de bon seuils (0.42) et acceptable (0.70). Elle présente également l'Histogramme de distribution des probabilités prédites (Background vs Polyp) avec seuil de décision (ligne verte) et heatmap des probabilités à 512×512 avec colormap plasma(figure 3.31(b)).

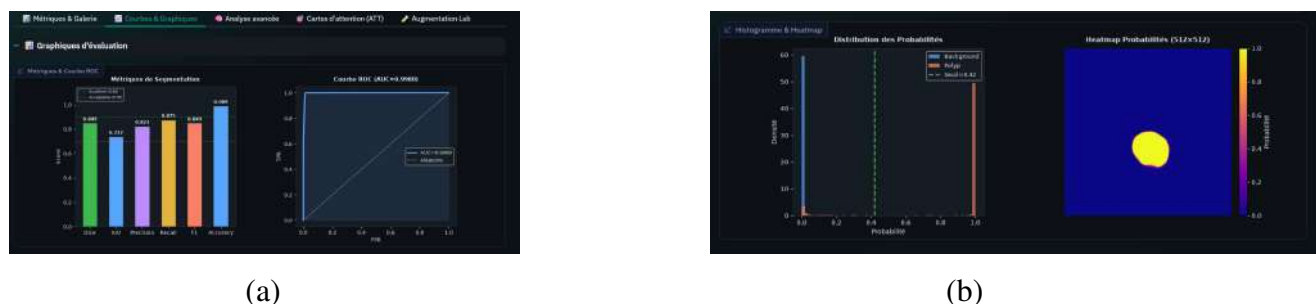


Figure 3.31 Interface Technique : (a) Métriques de segmentation et courbe ROC. , (b) Distribution des probabilités et heatmap de prédiction.

- **Analyse avancée** : outils d'interprétabilité et d'estimation de l'incertitude. L'incertitude prédictive est estimée par Monte-Carlo Dropout (passes successives à probabilité de désactivation paramétrable), tandis que la carte d'activation *Grad-CAM*, calculée sur la sortie du module ASPP, met en évidence les régions ayant le plus contribué à la décision du réseau. La figure 3.32 montre une analyse avancée par MC Dropout (50 passes, $p=0.50$) générant une carte d'incertitude (moyenne=0.042, max=0.417) et par Grad-CAM sur la sortie ASPP visualisant les zones d'activation du modèle.

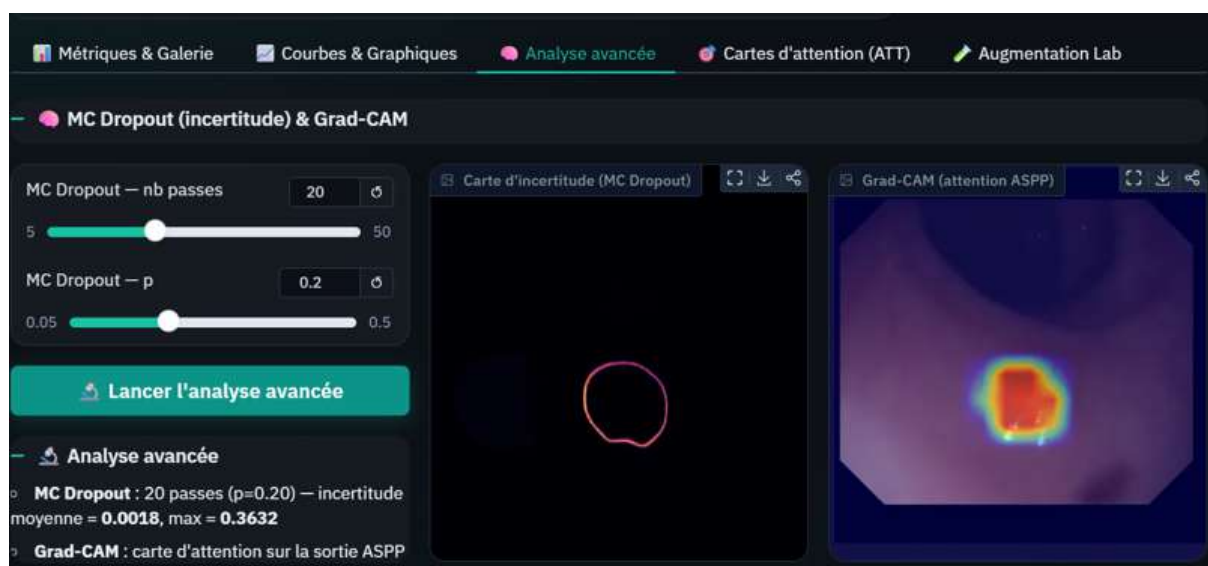


Figure 3.32 Interface Technique : Carte d'incertitude (MC Dropout) et activation Grad-CAM.

- **Cartes d'attention** : visualisation des portes d'attention du décodeur de CANet+, héritées du paradigme Attention U-Net, superposées sur l'image afin d'illustrer la manière dont le

réseau focalise progressivement son attention sur la lésion à travers les différents niveaux. La figure suivante illustre la carte d'attention (Attention Gates a1 → a4) superposées sur l'image endoscopique. Les zones claires (vert/jaune) indiquent où le décodeur concentre son attention spatiale, du niveau fin (a1) au niveau profond (a4).

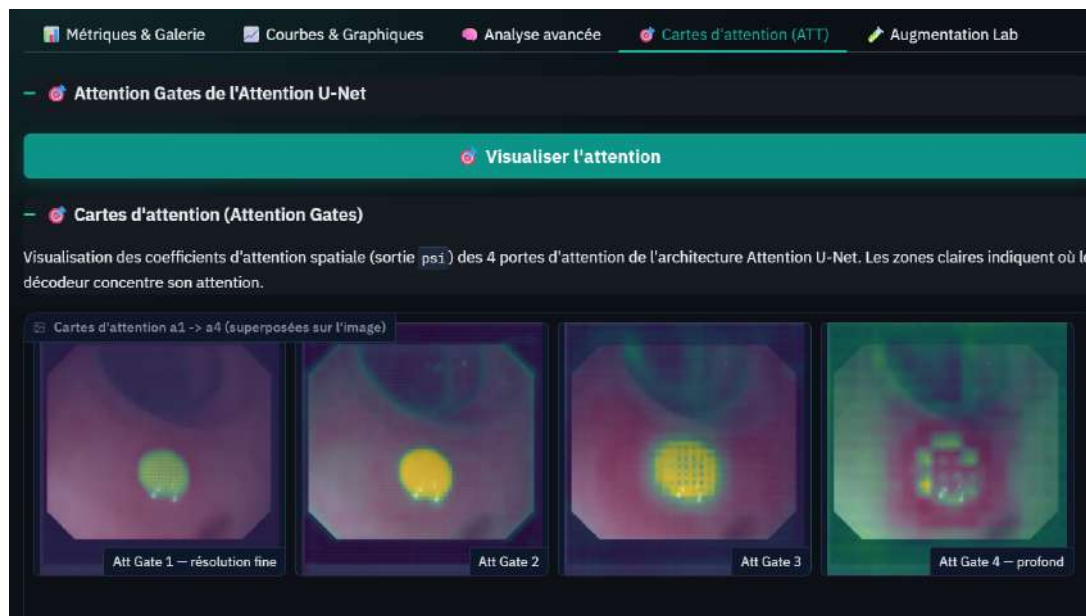


Figure 3.33 Interface Technique : Visualisation des Attention Gates de l'architecture Attention U-Net.

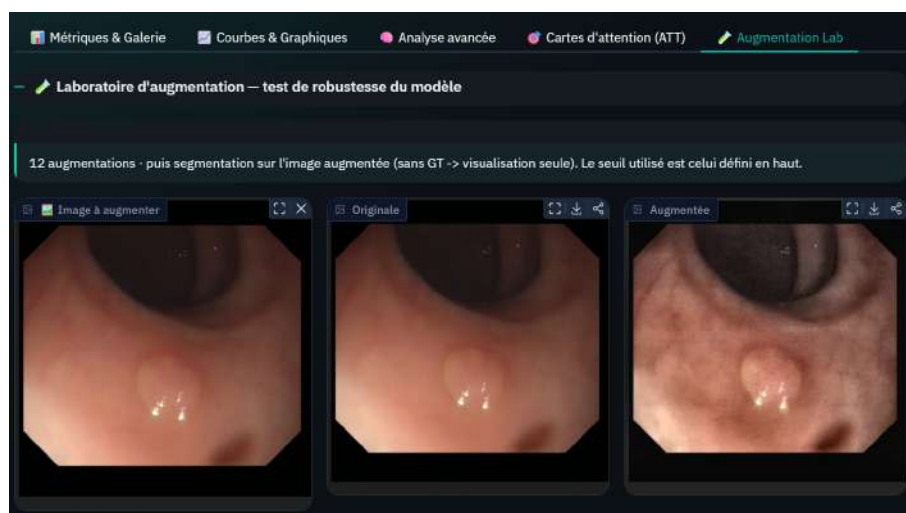
- **Augmentation Lab** :laboratoire de test de robustesse permettant d'appliquer différentes transformations géométriques (retournements, rotations), photométriques (luminosité, contraste, CLAHE) et de dégradation (flou et bruit gaussiens, transformation élastique) puis de relancer la segmentation sur l'image augmentée, dans le but d'évaluer qualitativement la stabilité des prédictions face aux variations d'acquisition. La figure suivante illustre un menu déroulant listant les 12 augmentations disponibles (flip, rotation, luminosité, contraste, flou gaussien, bruit gaussien, CLAHE, transformation élastique). L'augmentation sélectionnée est appliquée instantanément.



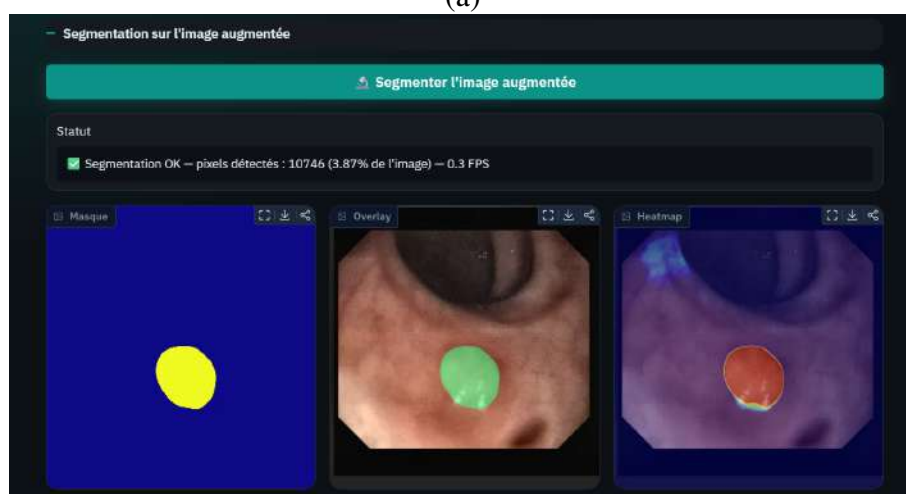
Figure 3.34 Interface Technique Augmentation Lab : menu de sélection des transformations.

La figure suivante montre l'image originale et image augmentée (Clahe) affichées côte à côte. Ce module permet de tester la robustesse du modèle face aux variations d'acquisition endoscopique.

Les résultats de segmentation sur l'image augmentée (CLAHE) sont présentés dans la figure suivante :



(a)



(b)

Figure 3.35 Interface Technique : (a,b)présente Augmentation Lab(comparaison image originale et augmentée.)

L'inférence repose par ailleurs sur une stratégie d'**augmentation au moment du test** (Test-Time Augmentation, TTA), qui moyenne les prédictions du réseau sur l'image originale et sur plusieurs de ses versions transformées (retournements et rotation), dans le but d'améliorer la robustesse des prédictions finales.

3.8 Conclusion

Nous avons présenté dans ce chapitre notre modèle CANet+, une architecture hybride encodeur-décodeur dédiée à la segmentation automatique des polypes colorectaux en imagerie endoscopique. En intégrant de manière complémentaire un encodeur ConvNeXt-Tiny, un décodeur Attention U-Net, des modules d'attention CBAM, un module ASPP et un mécanisme de supervision profonde, le modèle atteint un équilibre supérieur entre le raffinement local des

caractéristiques et la compréhension contextuelle globale.

Les résultats expérimentaux conduits sur trois benchmarks de référence indiquent que le modèle présente une fiabilité satisfaisante, avec un score Dice compris entre 0,92 et 0,96 et un IoU compris entre 0,87 et 0,92. Ces performances surpassent plusieurs méthodes récentes de l'état de l'art, notamment Polyp-PVT (**Dong et al., 2021**) et ColonFormer-L (**Duc et al., 2022**), en particulier sur le benchmark CVC-ColonDB, le plus exigeant des trois. Notons que notre modèle a fait l'objet d'une acceptation et a été présenté dans une conférence internationale sur l'Informatique médicale en environnements dynamiques (**MIDE2026 Program Committee, 2026**) organisée à Tipasa (Algérie), sous le titre «CANet+ : Attention-Guided Colorectal Polyp Segmentation with Multi-Scale Context Modeling».

Les travaux futurs s'orienteront vers l'élagage architectural et la quantification du modèle pour des applications en temps réel sur du matériel embarqué, ainsi que vers l'intégration de données multimodales afin d'améliorer encore les performances sur les lésions planes à faible contraste.

Conclusion Générale

Le présent mémoire s'est inscrit dans une démarche qui articule, de manière cohérente, un besoin clinique réel, une analyse de la littérature scientifique et une contribution architecturale originale. Le point de départ n'était pas une abstraction théorique, mais une observation directe du terrain : les échanges menés au service de gastroentérologie de l'hôpital Ahmed Madaghri d'Aïn Témouchent ont mis en lumière les limites concrètes de la coloscopie standard face à la diversité morphologique des polypes colorectaux, en particulier les lésions planes et les adénomes sessiles festonnés, dont le faible contraste avec la muqueuse environnante en fait une source persistante de lésions manquées. C'est à partir de ce constat que la problématique a été formulée. L'objectif est alors est de concevoir une architecture de Deep Learning capable de segmenter avec précision les polypes colorectaux dans des images endoscopiques, en surmontant la variabilité morphologique et le faible contraste des lésions, tout en maintenant une efficacité computationnelle compatible, avec une possibilité d'intégration en milieu clinique.

L'analyse de l'état de l'art a révélé que, malgré les progrès considérables réalisés depuis U-Net jusqu'aux architectures hybrides les plus récentes intégrant Vision Transformers et mécanismes d'attention, un compromis persistant entre précision de segmentation, robustesse face à l'hétérogénéité des données et efficacité computationnelle demeurait non résolu. C'est précisément dans cet espace que CANet+ a été conçu et évalué.

L'architecture proposée repose sur l'intégration de plusieurs mécanismes complémentaires de représentation, d'attention multi-échelle et de supervision profonde, conçus pour répondre aux principales limites identifiées dans l'état de l'art. Chacun de ces mécanismes a été validé par une étude d'ablation systématique confirmant sa contribution aux performances globales du modèle. Les résultats expérimentaux obtenus sur les trois benchmarks de référence Kvasir-SEG, CVC-ClinicDB et CVC-ColonDB selon un protocole de validation croisée à quatre plis (four-fold cross-validation) confirment la pertinence de l'approche : avec des scores Dice compris entre 0,92 et 0,96 et des valeurs d'IoU comprises entre 0,87 et 0,92, CANet+ démontre des performances compétitives par rapport aux méthodes de l'état de l'art.

Ces résultats doivent cependant être interprétés avec la rigueur qu'impose toute recherche honnête. Les bases de données utilisées, bien qu'elles constituent les benchmarks standards de la communauté, ont été acquises dans des centres spécifiques avec des équipements particuliers, la généralisation à d'autres environnements hospitaliers, notamment en contexte algérien avec des équipements endoscopiques variés, reste à valider. Par ailleurs, la taille relativement modérée

des corpus annotés disponibles constitue une limite structurelle commune à l'ensemble du domaine. Toutefois, l'augmentation de données et la validation croisée permettent d'atténuer sans toutefois l'éliminer. Ces limites, loin de relativiser la contribution, en définissent les contours et tracent naturellement les directions des travaux futurs. Plusieurs perspectives se dégagent en effet de ce travail. À court terme, la quantification et l'élagage du modèle technique de compression permettant de réduire significativement l'empreinte mémoire et la latence d'inférence constitueraient une étape décisive vers un déploiement embarqué sur les systèmes endoscopiques en salle de coloscopie. À moyen terme, l'extension de l'architecture à la segmentation vidéo en temps réel, exploitant la continuité temporelle des séquences endoscopiques, représente une évolution naturelle à fort potentiel clinique. Enfin, à plus long terme, l'intégration de données multimodales combinant l'imagerie en lumière blanche avec la chromoendoscopie virtuelle (NBI, FICE) pourrait permettre de lever les ambiguïtés résiduelles que l'imagerie standard seule ne parvient pas à résoudre.

Ce mémoire constitue ainsi une contribution consacré à la segmentation automatique des polypes colorectaux. Au-delà de l'apport méthodologique proposé, il s'inscrit dans une perspective plus large visant le développement de systèmes d'aide au diagnostic capables de s'intégrer à la pratique endoscopique réelle et de contribuer à la réduction de la mortalité liée au cancer colorectal grâce au dépistage précoce.

Bibliographie

1. Adneen, S. et al. (2025). *PolypSegNet : A hybrid ConvNeXt-Tiny and attention U-Net framework for accurate colorectal polyp segmentation*. Research Square (prépublication). <https://doi.org/10.21203/rs.3.rs-7102819/v1>.
2. Ahamed, M. F., Shafi, F. B., Islam, M. R., Nabi, M. F. & Haider, J. (2026). Rethinking U-Net architecture in medical imaging : Advancing the efficient and interpretable UKAN-CBAM framework for colorectal polyp segmentation. *Artificial Intelligence in Medicine*, 174, 103352. <https://doi.org/10.1016/j.artmed.2026.103352>.
3. Angermann, Q. (2018). *Détection temps réel assistée par ordinateur pour le dépistage précoce du cancer colorectal en vidéocoloscopie* (Thèse de doctorat). Université de Cergy Pontoise.
4. Barclay, R. L., Vicari, J. J., Doughty, A. S., Johanson, J. F. & Greenlaw, R. L. (2006). Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *New England Journal of Medicine*, 355(24), 2533–2541.
5. Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–127. <https://doi.org/10.1561/22000000006>.
6. Benkalfate, L., Ginestet, C., Arnachellum, R., Salin, G., Doucet, L., Sanavio, M., Camus, M., & Quénéhervé, L. (2024). Caractérisation endoscopique des polypes coliques : où en sommes-nous ? *Hépatogastro & Oncologie Digestive*, 31(2), 224-236. <https://doi.org/10.1684/hpg.2023.2734>.
7. Bernal, J., Sánchez, J. & Vilariño, F. (2012). Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 45(9), 3166–3182.
8. Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C. & Vilariño, F. (2015). WM-DOVA maps for accurate polyp highlighting in colonoscopy : Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*.
9. BESBAS, A. & HARB, M. E. (2023). *Reconstruction 3D des images médicales d'IRM cérébrales*.

10. Bhattacharya, D. et al. (2024). PolypNextLSTM : A lightweight and fast polyp video segmentation network using ConvNeXt and ConvLSTM. *International Journal of Computer Assisted Radiology and Surgery*.
11. Breier, M., Gross, S., Behrens, A., Stehle, T. & Aach, T. (2011). Active contours for localizing polyps in colonoscopic NBI image data. *Medical Imaging 2011: Computer-Aided Diagnosis* (Vol. 7963). SPIE.
12. Buades, A., Coll, B. & Morel, J. M. (2005). A review of image denoising algorithms, with a new one. *Multiscale modeling & simulation*, 4(2), 490–530.
13. Bulois, P. (2011). Magnification électronique en endoscopie digestive. *Acta Endoscopica*, 41(4), 206-212.
14. Bushberg, J. T., Seibert, J. A., Leidholdt, E. M. & Boone, J. M. (2012). *The Essential Physics of Medical Imaging* (3^e éd.). Lippincott Williams & Wilkins.
15. Byrne, M. F. et al. (2019). Standardized endoscopy reporting and data systems for gastrointestinal imaging. *Gastrointestinal Endoscopy*.
16. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q. & Wang, M. (2022). Swin-Unet : Unet-like pure transformer for medical image segmentation. European Conference on Computer Vision (ECCV) Workshops, 205–218. https://doi.org/10.1007/978-3-031-25066-8_9.
17. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y. et al. (2021). TransUNet : Transformers make strong encoders for medical image segmentation. arXiv :2102.04306.
18. Chen, L. C., Papandreou, G., Schroff, F. & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. <https://doi.org/10.48550/arXiv.1706.05587>.
19. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision (ECCV).
20. Chouchane, A. (2016). *Analyse d'images d'expressions faciales et orientation de la tête basée sur la profondeur* (Thèse de doctorat). Université Mohamed Khider - Biskra.
21. Cotton, P. B., Durkalski, V. L., Pineau, B. C., Palesch, Y. Y., Mauldin, P. D., Hoffman, B. et al. (2004). Computed tomographic colonography (virtual colonoscopy) : a multicenter comparison with standard colonoscopy for detection of colorectal neoplasia. *Jama*, 291(14), 1713–1719.

22. Zhang, Y., Liu, H., & Hu, Q. (2021). TransFuse : Fusing transformers and CNNs for medical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 14-24.
23. Daubresse-Chasle, M. (2024). Reconstruction de la déformation d'une scène tubulaire en vue de l'utilisation d'une meute de capsules miniaturisées dans l'intestin grêle. *Thèse de doctorat, Université Bourgogne Franche-Comté*.
24. Decker, R. S., Shademan, A., Opfermann, J. D., Leonard, S., Kim, P. C. & Krieger, A. (2017). Biocompatible near-infrared three-dimensional tracking system. *IEEE Transactions on Bio-medical Engineering*, 64(3), 549–556.
25. Depositphotos (s.d.). IRM du cerveau, coupe transversale. *Depositphotos*. Consulté le 18 juin 2026. [Lien en ligne](#).
26. Díaz-Rodríguez, E. (2025). RLHF y RLAIIF, Revolución Silenciosa de la Retroalimentación Educativa. *HETS Online Journal*, 15(2), 33–48.
27. Dong, B. et al. (2021). Polyp-PVT : Polyp segmentation with pyramid vision transformers. *arXiv preprint arXiv :2108.06932*.
28. Dosovitskiy, A. et al. (2021). An image is worth 16x16 words : Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.
29. Dougherty, G. (2009). *Digital Image Processing for Medical Applications*. Cambridge University Press.
30. Duc, N. T., Oanh, N. T., Thuy, N. T., Triet, T. M., & Dinh, V. S. (2022). ColonFormer : An efficient transformer based method for colon polyp segmentation. *IEEE Access*, 10, 80575–80586.
31. Elhaddar, Y. (s.d.). U-Net : Revolutionizing Medical Image Segmentation. *Medium*. Consulté le 18 juin 2026. <https://medium.com/@yasserelhaddar/u-net-revolutionizing-medical-image-segmentation-f080a1b05119>.
32. Fan, D. P., Ji, G. P., Zhou, T., Chen, G., Fu, H., Shen, J. et al. (2020). PraNet : Parallel reverse attention network for polyp segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. https://doi.org/10.1007/978-3-030-59725-2_26.
33. Fitzgerald, K. & Matuszewski, B. (2024). FCB-SwinV2 Transformer for polyp segmentation. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2024.3376228>.
34. Gao, P., Zhou, K., Su, W., Yu, J. & Zhou, P. (2023). Endoscopic management of colorectal polyps. *Gastroenterology Report*, 11, goad027. <https://doi.org/10.1093/gastro/goad027>.

35. Gastrohegp. (s.d.). Endoscopie conventionnelle. <https://www.gastrohegp.fr/?page=endoscopies-conventionnelles>.
36. Gonzalez, R. C. & Woods, R. E. (2018). *Digital Image Processing*. Pearson.
37. Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*. MIT Press.
38. Gu, H., Wang, P., Li, Y., Bao, N., Wang, H., Xie, Y. et al. (2024). MFP-DeepLabv3+ : a multi-scale feature fusion and parallel attention network for enhanced bone metastasis segmentation. *Traitement du Signal*, 41(2), 771.
39. Guo, Y., Du, H., Zhang, Z., Wang, Y., Li, D., & Li, D. (2026). PRANet : Pathological relationship perception and dual attention guided network for diabetic retinopathy grading. *Applied Intelligence*, 56(1), 1. <https://doi.org/10.1007/s10489-025-07036-6>.
40. Hadjira, L. (2020). *L'image numérique*. Cours universitaire.
41. Haggitt, R. C., Glotzbach, R. E., Soffer, E. E. & Wruble, L. D. (1985). Prognostic factors in colorectal carcinomas arising in adenomas. *Gastroenterology*, 89(2), 328–336. [https://doi.org/10.1016/0016-5085\(85\)90333-6](https://doi.org/10.1016/0016-5085(85)90333-6).
42. He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
43. He, L., Lee, K., Lewis, M. & Zettlemoyer, L. (2017). Deep semantic role labeling : What works and what's next. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 473–483.
44. He, et al. (2024). A lighter hybrid feature fusion framework for polyp segmentation. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-72763-8>.
45. Hewett, D. G., Kaltenbach, T., Sano, Y., Tanaka, S., Saunders, B. P., Ponchon, T., Soetikno, R., & Rex, D. K. (2012). Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology*, 143(3), 599-607.
46. Hinton, G. E., Osindero, S. & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7), 1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>.
47. Hinton, G. E. & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507. <https://doi.org/10.1126/science.1127647>.
48. Hôpitaux universitaires de Genève (2018). La radiographie. *Hôpitaux universitaires de Genève*. Consulté le 18 juin 2026. <https://www.hug.ch/pluriprofessionnels-sante/radiographie>.

49. Hu, J., Shen, L. & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7132–7141.
50. Hu, K., Wang, C., Zhu, H., Zhao, L., Fu, C., Yang, W. et al. (2025). Adjacent-differential network with shallow attention for polyp segmentation in colonoscopy images. *Scientific Reports*, 15, 32777. <https://doi.org/10.1038/s41598-025-15470-2>.
51. Huang, et al. (2023). Polyp2Former : Boundary guided network based on Transformer for polyp segmentation. 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). <https://doi.org/10.1109/BIBM58861.2023.10385257>.
52. Hwang, S. & Celebi, M. E. (2010). Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 678–681. <https://doi.org/10.1109/ICASSP.2010.5495103>.
53. Islam, A. M., Bhuiyan, S. S., Mashira, M., Ahmed, M. R., Islam, S. & Shatabda, S. (2025). Polyp segmentation in colonoscopy images using DeepLabV3++. *Computers in Biology and Medicine*. <https://www.sciencedirect.com/science/article/pii/S0010482525013381>.
54. Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., de Lange, T., Halvorsen, P. & Johansen, H. D. (2019). ResUNet++ : An advanced architecture for medical image segmentation. 2019 IEEE International Symposium on Multimedia (ISM), 225–2255. <https://doi.org/10.1109/ISM.2019.00049>.
55. Jha, D. et al. (2020). Kvasir-SEG : A segmented polyp dataset. *Proceedings of the International Conference on Multimedia Modeling (MMM)*.
56. Jiang, et al. (2024). Improving polyp segmentation with boundary-assisted guidance and cross-scale interaction fusion transformer network. *Processes*, 12(5), 1030. <https://doi.org/10.3390/pr12051030>.
57. Khan, R. A. et al. (2025). Bilateral collaborative streams with multi-modal attention network for accurate polyp segmentation. *Scientific Reports*, 15. <https://doi.org/10.1038/s41598-025-15401-1>.
58. Kiesslich, R., Fritsch, J., Holtmann, M., Koehler, H. H., Stolte, M., Kanzler, S. et al. (2003). Methylene blue-aided chromoendoscopy for the detection of intraepithelial neoplasia and colon cancer in ulcerative colitis. *Gastroenterology*, 124(4), 880–888.
59. Kikuchi, R. et al. (1995). Management of early invasive colorectal cancer. *Diseases of the Colon & Rectum*, 38(12), 1286–1295. <https://doi.org/10.1007/BF02049154>.

60. Kirillov, A., Girshick, R., He, K. & Dollár, P. (2019). Panoptic feature pyramid networks. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6399–6408.
61. Kokudo, N., Takemura, N., Hasegawa, K., Takayama, T., Kubo, S., Shimada, M. et al. (2019). Clinical practice guidelines for hepatocellular carcinoma : the Japan Society of Hepatology 2017 (4th JSH-HCC guidelines) 2019 update. *Hepatology Research*, 49(10), 1109–1113.
62. Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems (NeurIPS)*, 25, 1097–1105.
63. Koessler, T., Bichard, P., Puppa, G., Lepilliez, V., Roth, A., & Cacheux, W. (2015). Cancers superficiels de l'œsophage : épidémiologie, diagnostic et traitement. *Revue Médicale Suisse*, 11(475), 1130-1135.
64. Lamy, B. (2020). *Physique-Imagerie médicale-Rayons X, IRM, échographie, scintigraphie, tomographies-Phénomènes, techniques, utilisation*. Editions Ellipses.
65. Latrive, A. (2012). *Tomographie de cohérence optique plein champ pour l'endoscopie : microscopie in situ et in vivo des tissus biologiques* (Thèse de doctorat). Université Pierre et Marie Curie - Paris VI.
66. Lecomte, T., Tougeron, D., Chautard, R., Bressand, D., André, T., Bibeau, F., Blanc, B., Cohen, R., Jacques, J., Lagasse, J., Laurent-Puig, P., Lepage, C., Lucidarme, O., Martin-Babau, J., Panis, Y., Portales, F., Taïeb, J., Aparicio, T., & Bouché, O. (2026). Cancer du côlon non métastatique. *Thésaurus National de Cancérologie Digestive (TNCD)*. <https://www.snfge.org/tncd/cancer-du-colon-non-metastatique>.
67. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>.
68. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>.
69. LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>.
70. Lee, C. Y., Xie, S., Gallagher, P., Zhang, Z. & Tu, Z. (2015). Deeply-supervised nets. *Proceedings of AISTATS*.

71. Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J. et al. (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008. *Gastroenterology*, 134(5), 1570–1595.
72. Li, et al. (2026). HCA-Net : Hierarchical contextual attention network for lightweight and accurate polyp segmentation. *IEEE Journal of Biomedical and Health Informatics*. <https://doi.org/10.1109/JBHI.2026.3657790>.
73. Liu, X. et al. (2024). Attention combined pyramid vision transformer for polyp segmentation. *Biomedical Signal Processing and Control*.
74. Marcq, G., Mach, K., Chang, T., Trivedi, D., Villers, A. & Liao, J. (2018). Biopsie optique des lésions planes de cancer de vessie par endomicroscopie confocale laser et analyse histologique : étude descriptive des critères diagnostiques. *Progrès en Urologie*, 28(13), 717–718.
75. Mersaoui, I. & Fouiri, M. (2018). *Étude sur l'imagerie médicale : prétraitement, segmentation et amélioration des exploitations*.
76. Mesejo, P., Pizarro, D., Abergel, A., Rouquette, O., Beorchia, S., Poincloux, L. et al. (2016). Computer-aided classification of gastrointestinal lesions in regular colonoscopy. *IEEE transactions on medical imaging*, 35(9), 2051–2063.
77. Microsoft (2026). Visual Studio Code – The open source AI code editor. *Microsoft*. Consulté le 18 juin 2026. <https://code.visualstudio.com/>.
78. MIDE2026 Program Committee. (2026). International Conference on Medical Informatics in Dynamic Environments (MIDE 2026). *Tipasa Smart City*. <https://tipasasmartcity.org/mide/>.
79. Misra, S. et al. (2021). Artificial intelligence in colonoscopy : Clinical application and future prospects. *World Journal of Gastroenterology*, 27(35), 5879–5894.
80. Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. & Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature Methods*, 16(12), 1233–1246. <https://doi.org/10.1038/s41592-019-0403-1>.
81. MSD Manuals (2023). *Imagerie par résonance magnétique (IRM)*. MSD Manuals (version grand public). <https://www.msdmanuals.com/fr/accueil/sujets-par>.
82. Mubarak, D. M. N., Sathik, M. M., Beevi, S. Z. & Revathy, K. (2012). A hybrid region growing algorithm for medical image segmentation. *International Journal of Computer Science & Information Technology*, 4(3), 61.
83. National Cancer Institute (2024). *Screening tests to detect colorectal cancer and polyps*. Fact sheet, NCI. <https://www.cancer.gov/types/colorectal/screening-fact-sheet>.

84. Oktay, O., Schlemper, J. et al. (2018). *Attention U-Net : Learning where to look for the pancreas*. arXiv :1804.03999 ; MIDL 2018.
85. Organisation mondiale de la Santé (2026). *Cancer colorectal*. Fiche d'information (mise à jour du 13 février 2026). Consulté le 18 juin 2026. <https://www.who.int/fr/news-room/fact-sheets/detail/colorectal-cancer>.
86. Pan, L. (2022). Improved ASPP module. *A Study of English Learning Vocabulary Detection Based on Image Semantic Segmentation Fusion Network*. ResearchGate. https://www.researchgate.net/figure/Improved-ASPP-module_fig3_361028753.
87. Pizer, S. M. et al. (1987). Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3), 355–368.
88. Python Software Foundation (2026). Welcome to Python.org. *Python Software Foundation*. Consulté le 18 juin 2026. <https://www.python.org/>.
89. Qi, et al. (2024). Colorectal polyp segmentation algorithm with multi-scale global feature enhancement extraction. Proceedings of the 2024 International Conference. <https://doi.org/10.1145/3653781.3653812>.
90. Rex, D. K., Adler, S. N., Aisenberg, J., Burch Jr, W. C., Carretero, C., Chowers, Y. et al. (2015). Accuracy of capsule colonoscopy in detecting colorectal polyps in a screening population. *Gastroenterology*, 148(5), 948–957.
91. Ronneberger, O., Fischer, P. & Brox, T. (2015). U-Net : Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
92. Rumack, C. M., Wilson, S. R., Charboneau, J. W. & Levine, D. (2011). *Diagnostic Ultrasound*. Elsevier.
93. Samaha, E., Colas, C., Dhooge, M., Saurin, J. C., Lecomte, T., Coron, E. et al. (2019). Comparaison de la coloscopie avec chromoendoscopie à l'indigo carmin à la coloscopie avec chromoendoscopie virtuelle au NBI de 3ème génération chez des patients ayant un syndrome de Lynch : étude Lynch-Chromo. *Endoscopy*, 51(3).
94. Sano, Y., Ikematsu, H., Fu, K. I., Emura, F., Katagiri, A., Horimatsu, T. et al. (2009). Meshed capillary vessels by use of narrow-band imaging for differential diagnosis of small colorectal polyps. *Gastrointestinal Endoscopy*, 69(2), 278–283.
95. Sano, Y. et al. (2009). Narrow-band imaging (NBI) classification system for colorectal lesions. *Gastrointestinal Endoscopy*, 70(3), 522–528.

96. Schmidhuber, J. (2015). Deep learning in neural networks : An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>.
97. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2020). Grad-CAM : Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>.
98. Shen, D., Wu, G. & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual Review of Biomedical Engineering*, 19, 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
99. Sikkandar, M. Y. et al. (2024). Utilizing adaptive deformable convolution and position embedding for colon polyp segmentation with a visual transformer. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-57993-0>.
100. Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR).
101. Singh, R. (2024). Decoding CNNs : A Beginner's Guide to Convolutional Neural Networks and their Applications. *Python in Plain English*. Consulté le 18 juin 2026. [Lien en ligne](#).
102. Snover, D. C. (2011). Update on the serrated pathway to colorectal carcinoma. *Human Pathology*, 42(1), 1–10.
103. Spada, C., Hassan, C., Galmiche, J.-P., Neuhaus, H., Dumonceau, J.-M., Adler, S. et al. (2012). Colon capsule endoscopy : European Society of Gastrointestinal Endoscopy (ESGE) guideline. *Endoscopy*, 44(5), 527–536.
104. Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool. *BMC Medical Imaging*, 15, 29.
105. Tajbakhsh, N. et al. (2020). Incomplete polyp resection during colonoscopy — major cause of interval colorectal cancer. *Gastroenterology*.
106. Ullah, I., Alomar, K., Jamel, L., Althobaiti, M. M., Othman, K. M. & Noorwali, A. (2025). Dual pathway attention over attention network with multi-scale supervision for colonoscopy polyp segmentation. *Neurocomputing*. <https://www.sciencedirect.com/science/article/pii/S0925231225031200>.
107. Vayssade, J.-A. (2022). Architecture encodeur-décodeur d'un réseau convolutif. *Approche multi-critère pour la caractérisation des adventices* [Thèse]. ResearchGate. https://www.researchgate.net/figure/Architecture-encodeur-decodeur-dun-reseau-convolutif_fig32_361040063.

108. Vishwkarma, D. (2025). ConvNeXT. *Kaggle*. Consulté le 18 juin 2026. <https://www.kaggle.com/code/divyanshvishwkarma/convnext>.
109. Wang, C., Wang, S., Shao, S. & Zhai, J. (2024). DeepNeXt : A lightweight polyp segmentation algorithm based on multi-scale attention. *Quantitative Imaging in Medicine and Surgery*, 14(12), 8551–8567. <https://doi.org/10.21037/qims-24-985>.
110. Wang, P. et al. (2020). Deep-learning-aided detection of colorectal polyps in colonoscopy. *Gastroenterology*, 159(4), 1229–1238.
111. Wikipédia (2026). Tomodensitométrie. *Wikipédia, l'encyclopédie libre*. Consulté le 18 juin 2026. <https://fr.wikipedia.org/wiki/Tomodensitométrie>.
112. Woo, S., Park, J., Lee, J. Y. & Kweon, I. S. (2018). CBAM : Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19.
113. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://arxiv.org/abs/2301.00808>.
114. Wu, et al. (2023). META-Unet : Multi-scale efficient transformer attention U-Net for fast and high-accuracy polyp segmentation. *IEEE Transactions on Automation Science and Engineering*. <https://doi.org/10.1109/TASE.2023.3292373>.
115. Wyszecki, G. & Stiles, W. S. (2000). *Color Science : Concepts and Methods, Quantitative Data and Formulae*. Wiley.
116. Zauber, A. G. et al. (2012). Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *New England Journal of Medicine*, 366(8), 687–696.
117. Zeste de Savoir (2019). Introduction au traitement d'image. Consulté le 18 juin 2026. <https://zestedesavoir.com/tutoriels/1557/introduction-au-traitement-dimage/>.
118. Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N. & Liang, J. (2020). UNet++ : Redesigning skip connections to exploit multiscale features in medical image segmentation. *IEEE Transactions on Medical Imaging*. <https://doi.org/10.1109/TMI.2019.2959609>.

Annexe

Figures complémentaire

ETABLISSEMENT PUBLIC HOSPITALIER AHMED MEDAGHRI AIN TEMOUCHENT SERVICE D'HEPATO-GASTRO-ENTEROLOGIE Type d'examen: FOGD Nom de l'opérateur: Appareil: FUJINON Patient calme: coopérative Indication/Histoire clinique : EXAMEN PROCTOLOGIE • Inspection: • Touche rectale: Résultats de l'exploration endoscopique: Progression de l'endoscope jusqu'au: Bonne préparation: mauvaise préparation: Tolérance du patient : • Rectum : ▪ Muqueuse : ▪ Valvule de Houston: ▪ Réseau vasculaire: • Sigmoides: • Colon gauche : • Colon transverse: • Colon droit: • Colon droit: • Caecum: <u>CONCLUSION:</u>	COMTE RENDU D'ENDOSCOPIE DIGESTIVE BASSE DATE: Patient Nom: Prénom: Date de naissance: ans N° de fiche navette: Adresse par: CACHET ET SIGNATURE
---	--

Figure 1.1 Représentation du modèle de rapport médical de l'hôpital Ahmed Madaghri