



جامعة أبو بكر بلقايد - تلمسان

Université Abou Bakr Belkaïd de Tlemcen

Faculté de Technologie

Département de Génie Biomédical

MEMOIRE DE PROJET DE FIN D'ETUDES

pour l'obtention du Diplôme de

MASTER en GENIE BIOMEDICAL

Spécialité : Imagerie Médicale

présenté par : BERRAHOU Zoulikha

Segmentation des polypes colorectaux par les réseaux de neurones entièrement convolutifs

Soutenu le 30 juin 2022 devant le Jury

Mme ILES Amel

MCA Université de Tlemcen Présidente

Mme SETTOUTI Nesma

MCA Université de Tlemcen Encadreur

Mme KHEMIS Kamila

MCA Université de Tlemcen Examinatrice

Année universitaire 2021-2022

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABOU BEKR BELKAID
FACULTÉ DE TECHNOLOGIE
DÉPARTEMENT DE GÉNIE BIOMÉDICAL

MÉMOIRE DE FIN D'ÉTUDES

pour obtenir le grade de

MASTER EN GÉNIE BIOMÉDICAL

Spécialité : **Imagerie Médicale**

présenté et soutenu publiquement
par

Zoulikha BERRAHOU

le 30 Juin 2022

Titre:

Segmentation des polypes colorectaux par les réseaux de neurones entièrement convolutifs

Jury

Présidente du jury. Dr. ILES Amel,

MCA UABB Tlemcen

Examinatrice. Dr. KHEMIS Kamila,

MCA UABB Tlemcen

Directrice de mémoire. Dr. SETTOUTI Nesma,

MCA UABB Tlemcen

Je dédie ce travail à :

Ma chère mère,

Mon cher père,

Qui m'ont comblé avec leur amour, sacrifices et précieux conseils. Qui m'ont soutenu moralement et financièrement jusqu'à ce jour, qui m'ont encouragé tout au long de mon parcours d'études. Que ce travail soit pour eux un modeste témoignage de ma Profonde gratitude. Qu'ALLAH les préserve, et leurs accorde santé, bonheur et longue vie.

Mes frères, mes sœurs pour leurs soutien moral et leurs conseils précieux lors de la réalisation de ce travail que Dieu les protège et leur offre la chance et le bonheur et surtout la réussite.

Ma famille, mes proches et mes ami(e)s pour leurs aides et supports.

Ma collègue et sœur Malek MEDJADJI la meilleure amie qu'une personne puisse demander.

A tous ceux que j'aime et ceux qui m'aime.

Remerciements

Je remercie Dieu le Tout-puissant, le Très-Généreux, et le Très-Miséricordieux de m'avoir donné la santé, la patience et la volonté d'entamer et de terminer ce modeste travail qui présente le fruit de plusieurs années de sacrifices.

Au terme de ce travail, je tiens tout d'abord à exprimer ma profonde gratitude à mon encadreur Mme. **SETTOUTI NESMA** qui a tant déployé d'efforts pour la réussite de ce travail et mes sincères remerciements pour sa sympathie, sa patience, sa disponibilité, sa rigueur et sa précieuse aide.

Je tiens à remercier Melle. **ELAOUABER ZINEB AZIZA** et Mr. **MERAD LOTFI** pour leur aide pratique et leur soutien moral et leur encouragement.

Je souhaite remercier également tous mes professeurs pour la qualité de l'enseignement qu'ils m'ont prodigué au cours de ces années passées au département de Génie biomédical, Université de Tlemcen.

Mes remerciements s'adressent aux membres du jury pour l'honneur qui m'est fait d'examiner ce travail.

Enfin, sans oublier de remercier toutes les personnes qui ont participé de près ou de loin à la réussite de ce travail.

Résumé

La détection et la segmentation des polypes en coloscopie est une tâche difficile en raison des variations de taille, de forme, de texture des polypes ainsi que les différents types de cellules. De plus, cette procédure dépend fortement des compétences et de l'expérience de la personne qui manipule l'endoscope, ce qui se traduit par une forte variation de l'opérateur et des performances. Actuellement, les avancées des modèles à base d'apprentissage profond (DL Deep Learning) ont bouleversé le monde de la santé par la proposition de nouvelles méthodes de segmentation d'images aboutissant à des taux de précision les plus élevés sur les bases d'images médicales.

L'objectif de ce travail est d'acquérir une compréhension approfondie du domaine de la segmentation sémantique d'images ainsi qu'une connaissance du domaine de DL, en proposant une étude comparative de deux réseaux entièrement convolutifs (FCN Fully Convolutional Networks) pour la segmentation des polypes colorectaux, à savoir un modèle hybride U-Net avec post-traitement par les morphologies mathématiques et le modèle SegNet.

Mots clés

Polypes colorectaux, segmentation sémantique, apprentissage profond, FCN, U-Net, SegNet.

Abstract

Polyp detection and segmentation in colonoscopy is a difficult task due to variations in polyp size, shape, texture, and cell types. In addition, this procedure is highly dependent on the skill and experience of the person handling the endoscope, resulting a high variation in operator and performance. Currently, advances in deep learning (DL) models have revolutionized the world of healthcare by proposing new image segmentation methods leading to the highest accuracy rates on medical image databases.

The objective of this work is to acquire a deep understanding of the field of semantic image segmentation as well as a knowledge of the DL domain, by proposing a comparative study of two fully convolutional networks (FCN) for the segmentation of colorectal polyps, namely a hybrid U-Net model with post-processing by mathematical morphologies and the SegNet model.

Keywords

Colorectal polyps, semantic segmentation, deep learning, FCN, U-Net, SegNet.

ملخص

يعد اكتشاف الاورام الحميدة وتجزئتها في تنظير القولون مهمة صعبة بسبب الاختلافات في حجم وشكل وملمس الاورام الحميدة وكذلك أنواع الخلايا المختلفة. بالإضافة إلى ذلك، يعتمد هذا الإجراء بشكل كبير على مهارة وخبرة الشخص الذي يتعامل مع المنظار الداخلي، مما يؤدي إلى تباين عالٍ في التشغيل والأداء. في الوقت الحالي، التطورات في النماذج القائمة على التعلم العميق (DL Deep Learning) أحدثت تغييرات كبيرة في مجال الصحة من خلال اقتراح طرق جديدة لتجزئة الصور مما أدى إلى رفع معدلات الدقة في قواعد بيانات الصور الطبية.

الهدف من هذا العمل هو اكتساب فهم معمق لمجال تجزئة الصور الدلالية وكذلك معرفة مجال التعلم العميق DL، من خلال اقتراح مقارنة بين شبكتين ملتفتين بالكامل (FCN Fully Convolutional Networks) لتجزئة الاورام الحميدة للقولون ، وهي نموذج U-Net ملحقة بمعالجة عن طريق الأشكال الرياضية ونموذج SegNet.

كلمات البحث

الاورام الحميدة للقولون ، التجزئة الدلالية ، التعلم العميق ، FCN ، U-Net ، SegNet.

Table des matières

Remerciements	i
Résumé	ii
Abstract	iii
Table des matières	iv
Table des figures	vi
Liste des tableaux	viii
Glossaire	ix
Introduction générale	1
1 Les méthodes de détection du cancer colorectal assistées par ordinateur	3
1 Introduction	3
2 État de l’art du domaine	4
2.1 Les méthodes Ad-Hoc	4
2.2 Les méthodes basées sur de l’apprentissage profond	7
2.3 Les méthodes Hybrides	15
3 Synthèse des travaux	19
4 Conclusion	19
2 Approches de segmentation par réseaux convolutionnels	21
1 Introduction	21
2 La segmentation par les méthodes d’apprentissage profond	21
2.1 La segmentation sémantique	21
2.2 La segmentation instance	22
3 Étude comparative	23
4 Réseau entièrement convolutionnel (U-Net)	23
4.1 Le chemin de contraction (Encodeur)	24
4.2 Le chemin d’expansion symétrique (décodeur)	27
4.3 Bloc de connexion résiduelle	28
4.4 Backbones et ses utilisations	29
4.5 Avantages et inconvénients	29
5 Réseau entièrement convolutionnel (SegNet)	29
5.1 Encodeur	30
5.2 Décodeur	30
5.3 Softmax	31
6 Différences entre SegNet et U-Net	31
7 Conclusion	31

3	Étude comparative : expérimentations et résultats	32
1	Introduction	32
2	Base d'images d'étude	32
3	Modélisation et paramétrages	33
3.1	Environnement de développement et outils utilisés	33
3.2	Redimensionnement des images	33
3.3	Hyperparamétrage	34
4	Les architectures proposées	35
4.1	U-Net	35
4.2	SegNet	36
5	Apprentissage et expérimentations	37
5.1	Expérimentation 1	37
5.2	Expérimentation 2	39
6	Résultats et discussion	40
6.1	Résultats	41
7	Synthèse comparative	45
8	Interface de segmentation	45
9	Conclusion	46
	Conclusion générale	47
	Bibliographie	48

Table des figures

1.1	Schéma de la méthode de Bernal et al. [1].	4
1.2	Comparaison des résultats de la segmentation [1] : (a) Images originales (b) Masques de polypes (c) Résultat de méthode (d) Résultat des coupes normalisées.	5
1.3	Schéma de la méthode de Bernal et al. [2].	5
1.4	Résultats de l'approche avec : (a) image originale, (b) image pré-traitée, (c) carte d'énergie WM-DOVA, (d) la localisation du polype.	5
1.5	Schéma de la méthode de tajbakhsh et al. [3].	6
1.6	Résultats de la méthode contexte-forme.	6
1.7	Schéma de la méthode de Du et al. [4].	7
1.8	Résultat d'approche de propagation de l'attraction (AP).	7
1.9	Schéma du système d'apprentissage profond Offline/Online 3D [5].	8
1.10	Résultats du modèle SegNet modifié [6].	9
1.11	schéma d'architecture de U-Net-MobileNetV2 [7].	10
1.12	Résultats de modèle U-Net-MobileNetV2 [7].	10
1.13	schéma d'architecture de Tomar et al. [8].	11
1.14	Architecture proposée par Brandao et al. [9].	11
1.15	Comparaison entre le modèle proposé (BN FCN) et le FCN classic.	12
1.16	Schéma fonctionnelle du masque R-CNN [10].	12
1.17	Résultats de segmentation de la méthode.	13
1.18	Résultats de segmentation du modèle.	13
1.19	Schéma d'architecture de modèle Park et al. [11].	14
1.20	Résultats du modèle avec (a) Image d'entrée. (b) Carte de probabilité en NG (c) Carte de probabilité filtrée après post-traitement (d) Résultat final.	14
1.21	Schéma d'architecture d'Iwahori et al . [12].	15
1.22	Résultats de la méthode avec (a) image originale, (b) image améliorée, (c) image segmentée, (d) régions candidates.	15
1.23	Schéma d'architecture de méthode de Bae et al. [13].	16
1.24	Résultats de segmentation.	16
1.25	Schéma de méthode Zhang et al. [14].	17
1.26	Schéma d'architecture de méthode de Polyp Localize and Spot Team [15].	17
1.27	Résultats de segmentation avec PLS.	17
1.28	Schéma de méthode FCN +Otsu [16].	18
1.29	Résultat de la méthode FCN +Otsu.	18
2.1	La segmentation sémantique par U-Net [17].	22

2.2	La segmentation instance d'image coloscopique par mask-RCNN [10].	22
2.3	Illustration du processus de comparaison.	23
2.4	L'architecture de Réseau entièrement convolutionnel U-Net pour une image 32×32 [17].	24
2.5	Opération de convolution [18].	25
2.6	les types de padding SAME et VALID [19].	26
2.7	Illustration de l'opération de pooling [18].	26
2.8	Les fonctions d'activation les plus populaires [18].	27
2.9	Illustration de la convolution transposée [20].	28
2.10	Illustration de l'architecture de SegNet [21].	30
2.11	Le sur-échantillonnage par les indices de max pooling [21].	30
2.12	La fonction Softmax.	31
3.1	Exemple d'un polype (Image originale) avec le masque.	33
3.2	Représentation des cas de sur et sous apprentissage par rapport à l'optimal.	35
3.3	Modélisation de l'architecture U-Net utilisée.	36
3.4	Modélisation de l'architecture SegNet utilisée.	36
3.5	Progression de l'apprentissage d'U-Net expérimentation 1.	37
3.6	Tableau de performances d'apprentissage d'U-Net expérimentation 1.	38
3.7	Progression de l'apprentissage de SegNet expérimentation 1.	38
3.8	Tableau de performances d'apprentissage de SegNet expérimentation 1.	38
3.9	Évolution des performances d'apprentissage d'U-Net expérimentation 2.	39
3.10	Évolution des performances d'apprentissage de SegNet expérimentation 2.	39
3.11	Le coefficient Dice.	40
3.12	Intersection over Union.	41
3.13	a) image originale, b) image segmentée avec U-Net 10 epochs, c) image segmentée avec U-Net 20 epochs.	41
3.14	Image test segmentée avant et après post-traitement d'U-NET.	42
3.15	Image test segmentée avant et après post-traitement d'U-NET.	43
3.16	Image test segmentée avant et après post-traitement d'U-NET.	43
3.17	Exemples de segmentation avec SegNet.	44
3.18	Les métriques IoU et la matrice de confusion de test SegNet.	45
3.19	Capture de l'interface	46

Liste des tableaux

1.1	Synthèse des méthodes de détection du cancer colorectal assistée par ordinateur.	20
3.1	Résultats de performances de l'architecture U-Net.	43
3.2	Résultats de performances de l'architecture SegNet.	44
3.3	Synthèse des résultats de performances.	45

Glossaire

3D : Trois Dimension
AP : Attraction propagation
BDD : Base De Données
BN : Batch normalization
CAD : computer aided detection
CNN : Convolution Neural Network
CPU :Central Processing Unit (unité centrale de traitement)
DL : Deep Learning
FC : Fully Connected
FCN : Fully convolutional network
GI : Gastro-Intestinal
GPU :Graphics Processing Unit (unité de traitement graphique)
LR : Learning Rate
ML : Machine Learning
PLS : polyp localize and Spot
ReLU : Rectified linear Unit
ROI : Region Of Interest
RPN : Réseau de Proposition de Région
RVB : espace couleur : Rouge Vert Bleu
SegNet : a deep convolutional encoder-decoder architecture for Segmentation
SLIC : Simple Linear Iterative Clustering
SVM : Support Vector Machine
U-Net : U-Network
VGG16 :Very Deep Convolutional Networks for Large-Scale Image Recognition

Introduction générale

L'Algérie enregistre près de 6.500 nouveaux cas du cancer colorectal (CCR), selon les registres du cancer de l'Institut national de santé publique (INSP) en 2021 [22].

Cette affection classée la première sur l'ensemble des types de cancer répandus en Algérie [22]. Un cancer du côlon est une maladie des cellules qui tapissent la paroi interne du côlon, dans 60 à 80 % des cas, les cancers colorectaux se développent à partir d'un polype. Les polypes sont les croissances anormales de la muqueuse, Il existe de nombreux types de polypes lesquels peuvent se développer à plusieurs endroits y compris le colon. Ils ont des tailles et des formes différentes. Cependant, il est possible pour certains polypes de devenir cancéreux (malins) au fil du temps s'ils ne sont pas décelés et retirés. S'ils sont détectés à un stade précoce, ces polypes peuvent être facilement éliminés. Il est donc très important de détecter tous les polypes. Cependant, la coloscopie est une tâche longue et répétitive. Parfois, le manipulateur peut avoir les yeux fatigués et souffrir d'un déficit d'attention. Selon l'expérience et les normes d'évaluation des médecins, le taux de la détection actuel montre un taux d'erreur important (certains taux d'échec pouvant atteindre 20 %) qui affecte le diagnostic et le traitement en raison des variations de taille, de forme, de texture des polypes ainsi que les compétences et de l'expérience du manipulateur d'endoscope. Afin de réduire le taux d'erreur (taux de polypes non détectés) plusieurs méthodes de segmentation ont été proposées pour aider les médecins à détecter des lésions grâce à l'analyse automatique des images réalisée en direct pendant les examens de coloscopie.

Au cours des dernières années, les modèles d'apprentissage profond (Deep Learning DL) ont donné naissance à une nouvelle génération de modèles de segmentation d'images qui surpassent les performances des anciennes méthodes, obtenant généralement les taux de précision les plus élevés sur les bancs d'essai standards. Dans le domaine, cela a conduit à un changement de paradigme.

À travers ce projet de fin d'études, nous avons proposé une étude comparative en expliquant de manière très détaillée les étapes et stratégies qui permettent la mise en œuvre de la segmentation d'images coloscopiques par les réseaux de neurones entièrement convolutifs.

Le contenu de ce mémoire peut être décrit comme suit :

- Nous commençons le premier chapitre par l'état de l'art du domaine, en ci-

tant les quelques travaux existants dans la littérature pour la segmentation des polypes colorectaux.

- Dans le second chapitre, nous introduisons plusieurs termes et concepts et les notions théoriques nécessaires dans notre étude proposée pour la segmentation d'images.
- Dans le dernier chapitre, nous présentons les résultats et expérimentations engagés dans notre étude comparative entre les architectures SegNet et U-Net.

Chapitre 1

Les méthodes de détection du cancer colorectal assistées par ordinateur

1 Introduction

Les systèmes d'aide à la détection numérique (CAD : Computer Aided Detection) sont des systèmes qui assistent les médecins pour l'interprétation des images médicales. Ces systèmes analysent donc les images des examens médicaux dans le but de mettre en évidence des zones suspectes. En général, un système de CAD comprend plusieurs procédures : le prétraitement de l'image, la segmentation, la sélection des caractéristiques et la classification. Parmi ces procédures, la segmentation de l'image est une procédure clé dans le système de CAD qui consiste à classer chaque pixel comme appartenant à un label préalablement défini. Cependant, la détection automatique des polypes dans les vidéos de coloscopie est très difficile en raison des fortes variations de l'apparence des polypes (taille, couleur, forme, texture) et de la présence d'autres structures telles que les parois du côlon, les reflets spéculaires et les bulles d'air.

Au cours des dernières décennies, de nombreux algorithmes ont été développés et plusieurs approches ont été proposées pour l'évolution et le développement de la segmentation des polypes dans les images de coloscopie. Dans ce but, plusieurs méthodes traditionnelles utilisées telles que le seuillage, détection des bords, croissance des régions, regroupement et contours actifs, Fuzzy C means, Mean Shift, Spectral Clustering, Quick Shift et Entropy-rate clustering (ERC). Ces méthodes ne nécessitent pas de conseils de la part des utilisateurs pour spécifier les régions à segmenter et elles sélectionnent un ensemble approprié de caractéristiques en fonction de différents critères. Puis Les méthodes interactives/semi-automatiques de segmentation d'images ont attiré beaucoup d'attention et elles sont devenues assez développées, parmi ces méthodes : les Contours Actifs (AC), Lazy Snapping, Level Set (LS), Graph Cuts (GC), GrabCut, Random Walks (RW) et Shortest Paths (SP).

Plus récemment, avec l'adoption de la méthodologie d'apprentissage profond, les méthodes de segmentation plus traditionnelles (basées sur les caractéristiques artisanales) sont progressivement remplacées par des approches basées sur les ré-

seaux neuronaux convolutifs (CNN). La segmentation profonde des polypes unifie l'extraction de caractéristiques et la classification en un seul algorithme combiné, ce qui améliore considérablement la précision de la segmentation. Parmi celles-ci, les CNN, FCN, R-CNN, U-NET et Masque RCNN sont populaires.

Dans ce chapitre, nous allons détailler par la suite les articles les plus récents sur la segmentation des polypes dans les images de coloscopie.

2 État de l'art du domaine

Les systèmes d'aide à la détection assistée par ordinateur (CAO) qui concerne la détection et la segmentation des polypes colorectaux peuvent généralement être divisés en trois catégories [15] :

- Les méthodes Ad-hoc ("Hand-crafted")
- Les méthodes basées sur de l'apprentissage profond ("end-to-end learning")
- Les méthodes hybrides

2.1 Les méthodes Ad-Hoc

appelée aussi les méthodes Hand-Crafted , elles sont basées sur l'extraction des caractéristiques intrinsèques (forme , texture , contour , couleur, ...) telles que déterminées par l'utilisateur afin de détecter et localiser les polypes colorectaux [6], parmi ces méthodes on a :

- **Bernal et al.** [1] dans le but de minimiser nombre de régions, ils ont mis au point une méthode (voir Figure1.1) qui consiste à appliquer un algorithme de segmentation de base tel que le partage des eaux, puis une fusion de régions qui prend en compte le modèle d'apparence des polypes qui permet de définir les contours à l'aide des informations des vallées. Cette méthode est plus performante que les coupes normalisées en ce qui concerne la précision de la détection des régions de polype et offre une bonne caractérisation des régions non informatives. La figure 1.2 montre les résultats de cette méthode.

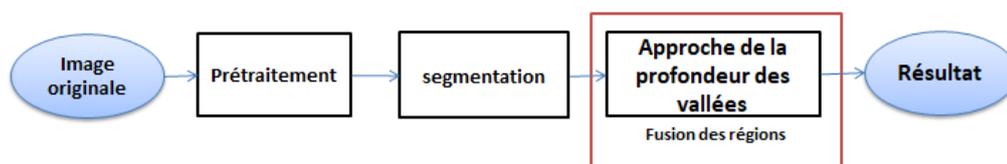


FIGURE 1.1 – Schéma de la méthode de Bernal et al. [1].

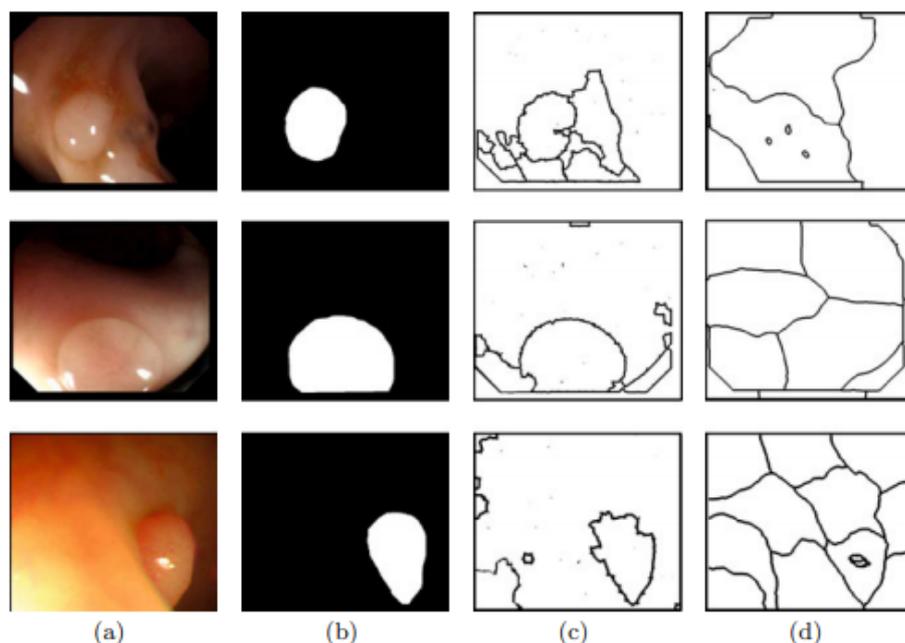


FIGURE 1.2 – Comparaison des résultats de la segmentation [1] : (a) Images originales (b) Masques de polypes (c) Résultat de méthode (d) Résultat des coupes normalisées.

- Par la suite, **Bernal et al.** [2] a proposé une nouvelle méthode de Segmentation des polypes basée sur la forme appelée cartes à fenêtre de profondeur médiane d’accumulation de vallées (WM-DOVA : A Window Median Depth of Valleys Accumulation) (voir Figure1.3) pour intégrer l’information sur les vallées par la diminution de l’influence du sang et des contours. Cette méthode montre de bonnes performances, en particulier pour les petits polypes qui sont signalés comme étant les principales sources de taux d’échec des polypes. La figure 1.4 montre les résultats de cette méthode.

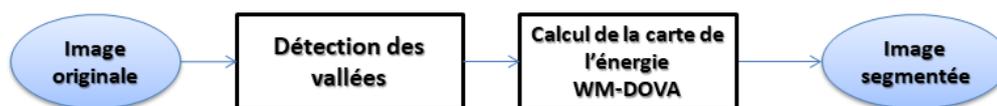


FIGURE 1.3 – Schéma de la méthode de Bernal et al. [2].

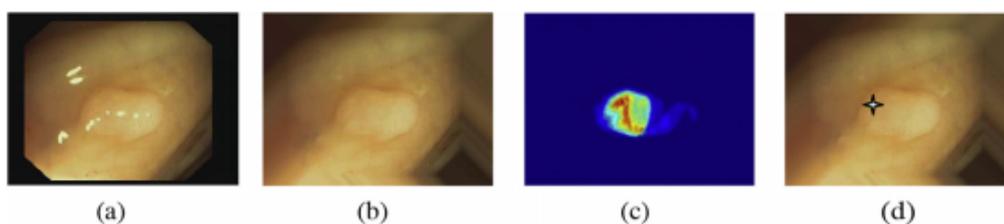


FIGURE 1.4 – Résultats de l’approche avec : (a) image originale, (b) image prétraitée, (c) carte d’énergie WM-DOVA, (d) la localisation du polype.

- **Tajbakhsh et al.** [23] ont proposé plusieurs méthodes de détection automatique des polypes basée sur la forme. La première méthode (voir Figure1.5) fait appelle à la forêt aléatoire pour classer les caractéristiques extraites par un descripteur Haar. Ils ont ensuite amélioré leur système dans [24] en remplaçant les caractéristiques Haar par un nouveau descripteur de patch et en remplaçant la forêt aléatoire par un schéma de classification des contours a deux niveaux en exploitant les caractéristiques contextuelles pour éliminer ces structures trompeuses des images.

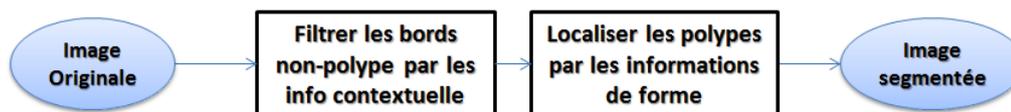


FIGURE 1.5 – Schéma de la méthode de tajbakhsh et al. [3].

- Par la suite dans [3], **Tajbakhsh et al.** ont proposé une autre approche hybride contexte-forme pour la détection des polypes permettant de capturer l'information géométrique discriminante des polypes. Cette approche consiste en 3 étapes : premièrement, la construction d'une carte de contours (edge map) a l'aide du détecteur de Canny dans chacun des trois canaux RVB. Deuxièmement, affiner cette carte en classant chaque pixel de contour en catégories de polypes et de non-polypes en utilisant des informations contextuelles. Enfin, un système de vote est conçu pour localiser les polypes en identifiant leurs têtes courbées. Ceci est réalisé en générant une carte thermique où une température plus élevée est attribuée aux régions qui sont entourées de contours courbés. La figure2.6 illustre les résultats de cette méthode.

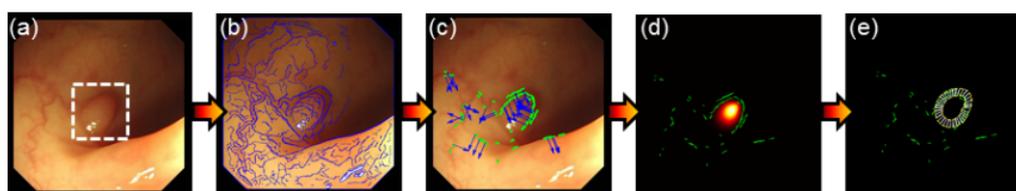


FIGURE 1.6 – Résultats de la méthode contexte-forme.

- **Du et al.** [4] Ont développé une nouvelle structure pour la segmentation interactive d'images basée sur la forme appelé la propagation de l'attraction (AP) (voir Figure1.7) a l'aide d'un système linéaire pour donner une solution d'optimisation. Cette approche propose de déterminer la région d'attraction et d'augmenter les probabilités de transition pour les nœuds dans la région et de réduire les probabilités de transition des pixels dans les régions sans polypes sans aucun pré-traitement des images. L'algorithme proposé permet de commencer avec un seul germe de premier plan et détecter automatiquement les contours des polypes après avoir réglé le paramètre de seuil. La figure1.8 illustre les résultats de cette méthode.

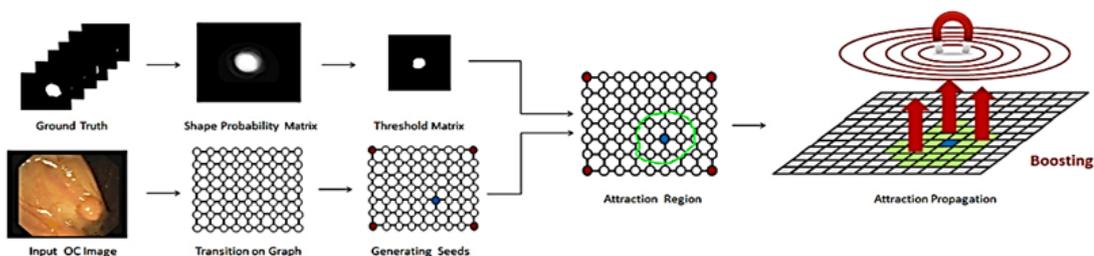


FIGURE 1.7 – Schéma de la méthode de Du et al. [4].

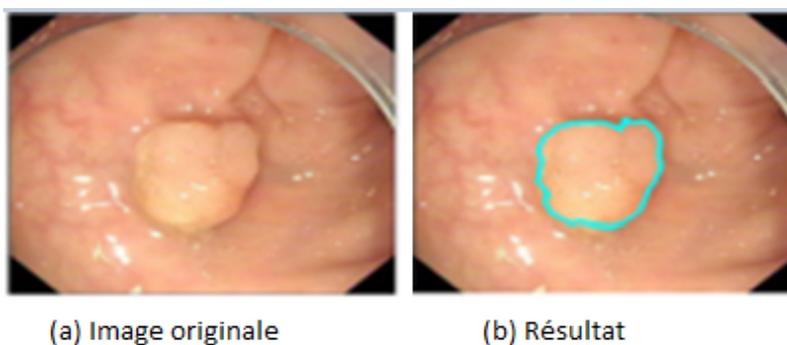


FIGURE 1.8 – Résultat d'approche de propagation de l'attraction (AP).

Avantages de méthodes Ad-Hoc : D'une part, les méthodes Hand-crafted ne nécessitent pas une grande base de données pour l'apprentissage. D'autre part, dans le cas où les caractéristiques sont bien déterminées (par exemple, la couleur ou la forme), l'extraction des polypes est pratiquement facile et efficace en termes de calcul.

Inconvénients : Les méthodes traditionnelles ont eu un succès limité en raison de la nécessité d'une expertise du chercheur et des nombreux facteurs qui affectent l'environnement du côlon; en particulier, il est difficile de faire la différence entre les polypes et les non-polypes, qui présentent des caractéristiques très similaires [25]. De plus, une caractéristique unique ne peut pas résoudre le problème de sorte que de multiples fonctions sont généralement nécessaires. Il est difficile de trouver la bonne combinaison et les bons choix d'un tel ensemble de fonctionnalités.

2.2 Les méthodes basées sur de l'apprentissage profond

Les techniques d'apprentissage profond ont été largement adoptées dans le domaine de segmentation d'images et ont surpassé les méthodes traditionnelles (ad-hoc). Les méthodes d'apprentissage profond permettent à un modèle informatique d'apprendre les caractéristiques et de les utiliser dans la prise de décision

en extrayant les caractéristiques des données automatiquement. Ces méthodes peuvent être divisées en deux catégories :

- La première catégorie utilise une approche basée sur les patches (sliding window). Dans ce cas, le CNN utilise uniquement les informations locales de l'image.
- La deuxième catégorie est développée sur la base du modèle d'apprentissage end-to-end [26].

Dans la littérature, de nombreuses méthodes sont proposées et parmi eux :

A. Modèles d'apprentissage End to end

- **Yu et al.** [5] ont adopté un nouveau cadre d'intégration d'apprentissage profond tridimensionnel (3D) hors ligne et en ligne pour détecter automatiquement les polypes en exploitant les 3D-FCN (Fully Convolutional Network) (voir Figure1.9). Le 3D-FCN a été appliqué pour une détection rapide en adoptant le concept de convolution complète qui permet de distinguer efficacement les polypes des imitations dures telles que les taches spéculaires et les bulles d'air en utilisant l'information spatio-temporelle. Le 3D-FCN hors ligne est d'abord développé et exploité pour apprendre les caractéristiques spatio-temporelles à partir des échantillons d'entraînement extraits des vidéos coloscopiques. Ce processus permet donc de réduire les calculs redondants et d'accélérer la détection par rapport à l'approche traditionnelle par fenêtre glissante.

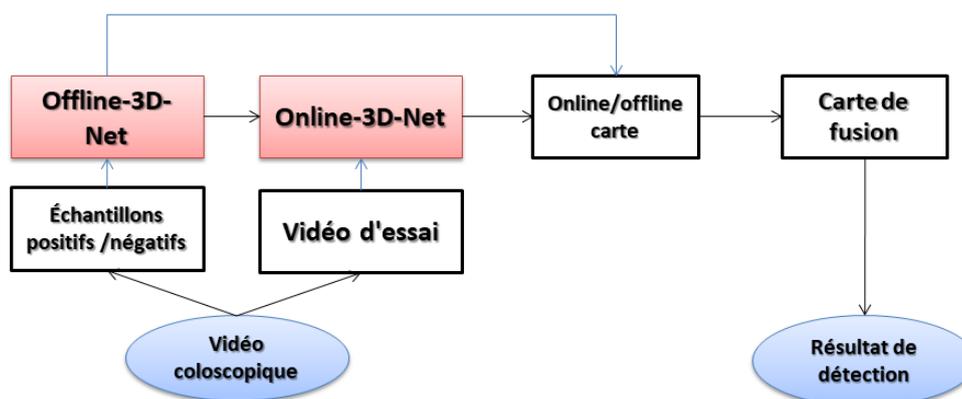


FIGURE 1.9 – Schéma du système d'apprentissage profond Offline/Online 3D [5].

- Dans [6], **Eu et al.** ont proposé un modèle SegNet modifié, avec VGG-19 comme structure, pour la segmentation automatique des polypes. SegNet est une architecture bien connue pour les problèmes de segmentation sémantique en imagerie médicale. Le réseau encodeur d'un SegNet VGG-19 contient une série de blocs de convolution construits à partir de couches de convolution, de normalisation par batch, et de couches d'activation ReLU. L'architecture SegNet modifiée a été obtenue par des améliorations appor-

tées au modèle SegNet VGG-19 original, les auteurs ont changé la structure originale SegNet VGG-19 en ajoutant des connexions de passage et des filtres convolutifs 5×5 , et en concaténant les convolutions dilatées parallèles. Les connexions de saut ont été introduites pour connecter les chemins de sous-échantillonnage aux chemins de sur-échantillonnage, dans le but de préserver l'information sur l'emplacement des polypes. Le modèle SegNet modifié a donné des meilleurs résultats par rapport à l'approche d'origine. La figure 1.10 montre les résultats de cette méthode.

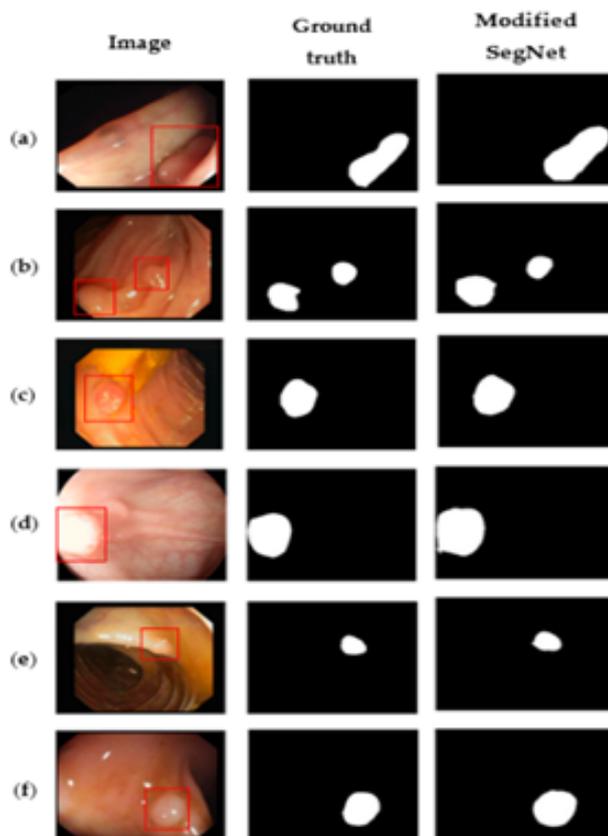


FIGURE 1.10 – Résultats du modèle SegNet modifié [6].

- **Branch et al.** [7] ont présenté un modèle pour la segmentation automatique des polypes basé sur U-Net-MobileNetV2 (voir Figure 1.11) où l'on applique un MobileNetV2 pré-entraîné sur ImageNet en tant qu'encodeur. Le système proposé reçoit les images d'entrée, qui sont insérées dans l'encodeur pré-entraîné, qui est basé sur des blocs résiduels inversés afin de réduire le nombre de paramètres ce qui rend le modèle plus facile et plus rapide à entraîner. Ce modèle inclut une combinaison de convolutions spatiales avec des noyaux 3×3 , une activation ReLu et des couches de normalisation par lots. Dans ce cas l'encodeur utilise convolution compacte de profondeur à profondeur pour filtrer et apprendre les caractéristiques des images qui alimentent le réseau. Finalement, le dernier bloc du réseau est un bloc avec une couche convolutive 1×1 et une activation Sigmoïde, de sorte que

la segmentation effectuée par le réseau puisse être générée. L'encodeur U-Net est capable d'optimiser les performances du modèle dans la tâche de segmentation et de réduire le coût de calcul et le temps de formation. 1.12 montre les résultats de cette méthode.

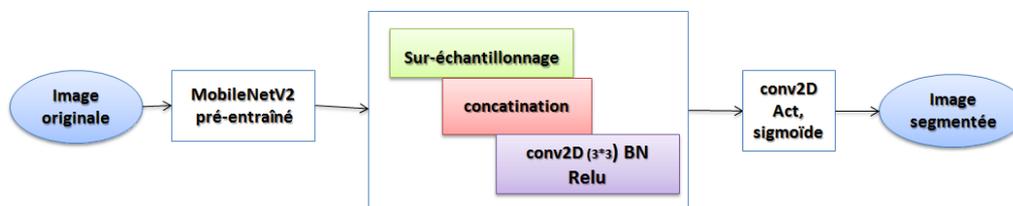


FIGURE 1.11 – schéma d'architecture de U-Net-MobileNetV2 [7].

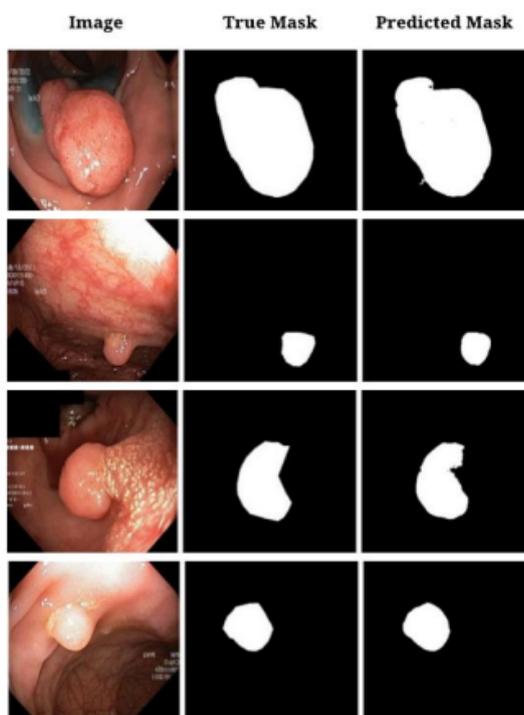


FIGURE 1.12 – Résultats de modèle U-Net-MobileNetV2 [7].

- **Tomar et al.** [8] ont proposé une architecture suivant une approche encodeur-décodeur (voir Figure 1.13). Ce modèle a été entraîné sur la base de données Kvasir-SEG, il combine la force de l'apprentissage résiduel et le mécanisme d'attention du réseau de compression et d'excitation, ce dernier fournit une attention ciblée à ces cartes de caractéristiques, renforçant les caractéristiques importantes et supprimant les caractéristiques non pertinentes. L'approche de compression et d'excitation se fait en deux étapes. Dans un premier temps, les cartes de caractéristiques entrantes sont d'abord comprimées en utilisant une opération de regroupement de la moyenne globale. Dans la deuxième étape, ce vecteur de caractéristiques globales passe par un réseau neuronal à deux couches. Ici, le nombre de caractéristiques

est d'abord réduit et puis étendu à la taille d'origine de l'image. Le modèle proposé constitue un réseau de codage qui se compose de 4 blocs et le réseau de décodage se compose également d'un bloc de 4. Le bloc résiduel est constitué de deux convolutions 3×3 , où la convolution est suivie d'une normalisation par lots (BN Batch Normalization) et d'une fonction d'activation de l'unité linéaire fonction d'activation Rectified Linear Unit (ReLU).

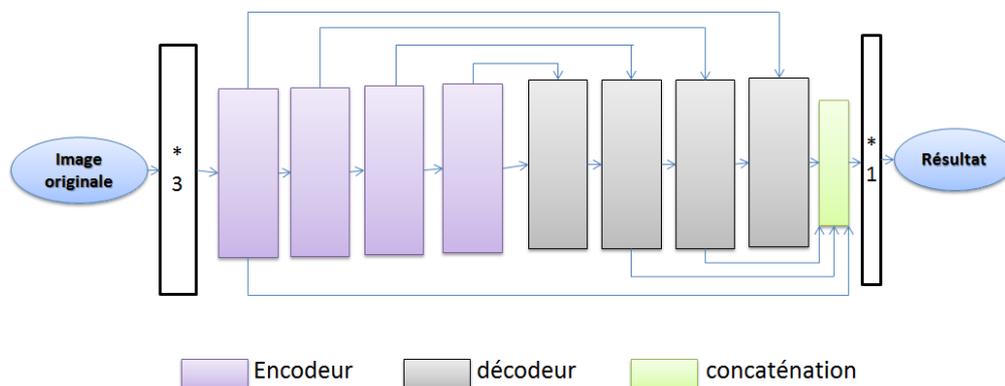


FIGURE 1.13 – schéma d'architecture de Tomar et al. [8].

- **Brandao et al. [9]** ils ont développé des modèles CNN de segmentation des polypes (voir Figure1.14), en utilisant architectures bien connues initialisées avec des poids obtenus lors du pré-entraînement. Ils ont choisi de convertir les CNN pré-entraînés en FCNs en ajoutant des couches de déconvolution afin d'obtenir une carte de classification des pixels en sortie. En outre, ils ont appliqué une technique SfS (shape-from-shading : forme de l'ombrage) qui vise à récupérer la forme 3D d'un objet en analysant la variation de l'illumination dans l'image. Cette technique est efficace pour récupérer la profondeur des images de coloscopie et la fournir comme paramètre supplémentaire, en plus des canaux RVB standard, dans la formulation de modèles de réseaux. Les résultats montrent que l'inclusion de la profondeur peut améliorer la représentation des polypes et conduire à une augmentation des taux de détection et de la précision de la segmentation. La figure1.15 montre les résultats de cette méthode.

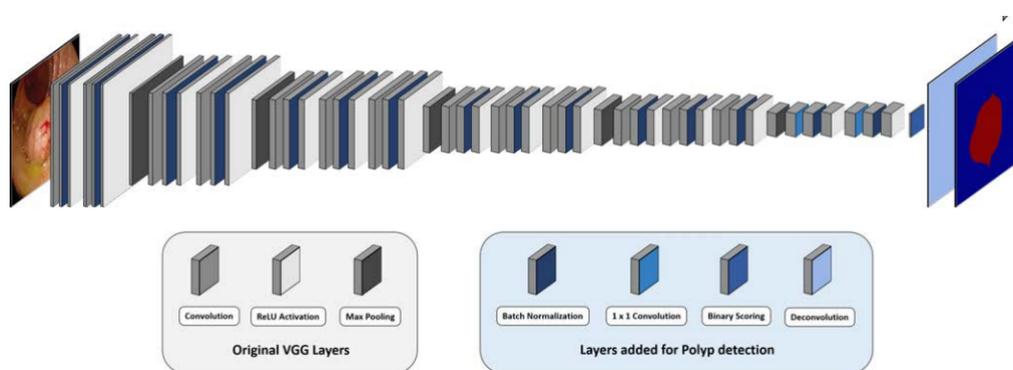


FIGURE 1.14 – Architecture proposée par Brandao et al. [9].

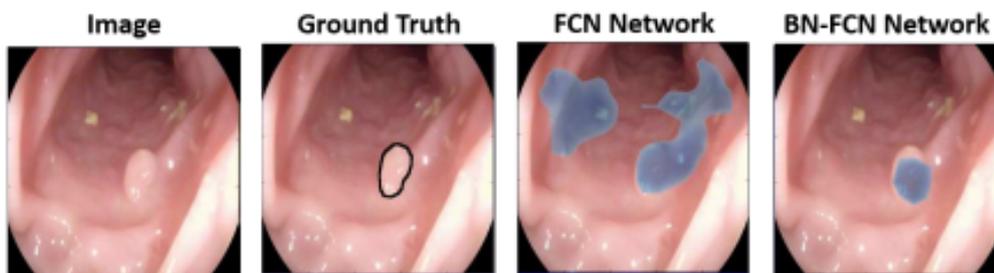


FIGURE 1.15 – Comparaison entre le modèle proposé (BN FCN) et le FCN classic.

Approches basées sur les patches :

- **Meng et al.** [10] ont proposé et exploré un système CAD-CNN pour identifier les polypes colorectaux et extraire les caractéristiques de l'image de coloscopie au lieu de les d'extraire manuellement avec un prétraitement extensif (voir Figure1.16). Ils ont amélioré le modèle mask R-CNN (en particulier la structure de base (backbone)) en construisant des liaisons hiérarchiques de type résiduel au sein d'un seul bloc résiduel afin d'améliorer la capacité de représentation multi-échelle à un niveau plus granulaire. Le système proposé se compose de deux modules : un réseau convolutif profond qui génère des propositions de régions et un système d'analyse des données avec un détecteur Faster R-CNN pour la segmentation et classification. Au préalable du réseau Faster R-CNN, un réseau de proposition de région (RPN) est proposé pour générer la ROI, qui indique la région avec des fractures. Par la suite, le masque R-CNN produit également un masque binaire pour chaque ROI qui est parallèle à la classe prédite et au décalage de la boîte. Ce système diffère de celles de la plupart des systèmes récents où la classification dépend des prédictions de masque. La figure1.17 montre les résultats de détection par cette méthode.

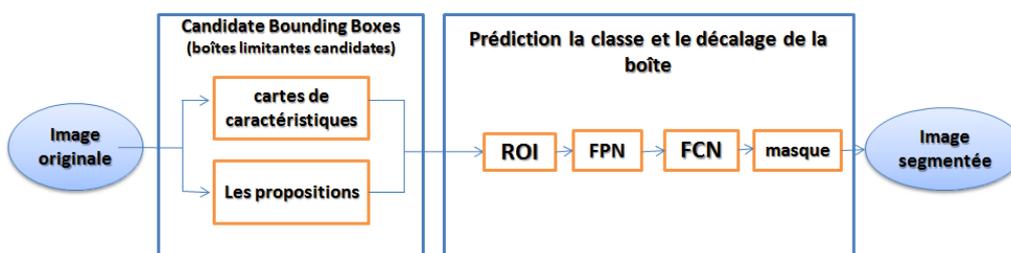


FIGURE 1.16 – Schéma fonctionnelle du masque R-CNN [10].

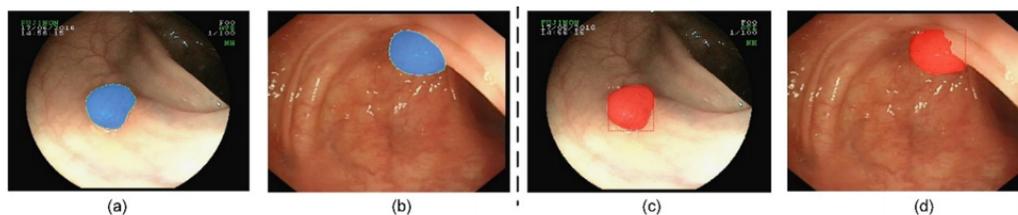


FIGURE 1.17 – Résultats de segmentation de la méthode.

- Dans [27] **Banik et al.** ont proposé un modèle CNN multi-échelle basé sur des patches locaux et globaux sont extraits de chaque pixel de l'image d'entrée (image RVB) et introduits dans deux CNN parallèles et similaires dont l'un est responsable de l'extraction des caractéristiques locales et l'autre de l'extraction des caractéristiques globales, ils ont empilé trois couches convolutives multiplié par une unité linéaire rectifiée (ReLU) afin d'imposer une non-linéarité à la carte des caractéristiques de sortie. Au final, les cartes de caractéristiques des deux CNNs sont concaténées pour l'annotation précise de l'étiquette du pixel de la région du polype. La figure 1.18 illustre les résultats obtenus.

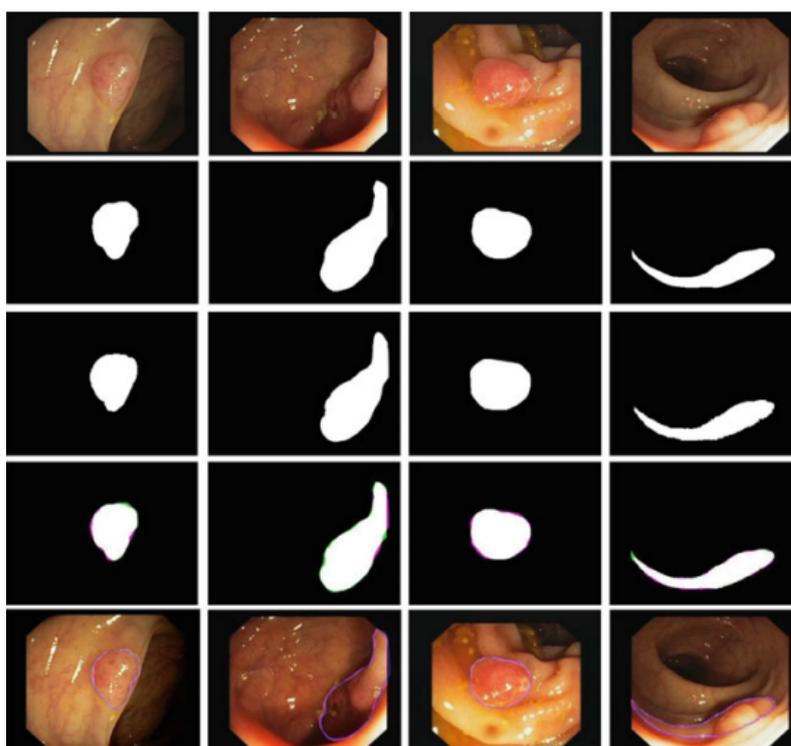


FIGURE 1.18 – Résultats de segmentation du modèle.

- Pour **Park et al.** [11] la méthode proposée (voir Figure 1.19) consiste à apprendre des caractéristiques hiérarchiques à l'aide d'un réseau neuronal convolutif. Les caractéristiques sont apprises à différentes échelles pour fournir des caractéristiques invariantes à l'échelle par le biais du CNN (Le réseau se compose de 3 couches convolutionnelles et de 3 couches de max-

pooling), par la suite, chaque pixel de l'image de coloscopie est classé en deux régions polype ou non polypes par un réseau entièrement connecté (fully connected FC). Pour exploiter la représentation multi-échelle, les caractéristiques du patch du même emplacement avec différentes échelles sont concaténées. Les résultats sont illustrés sur la base de données CVC-ClinicDB et elles montrent que ce modèle donne une précision de 90%. La figure 1.20 montre les résultats de cette méthode.



FIGURE 1.19 – Schéma d'architecture de modèle Park et al. [11].

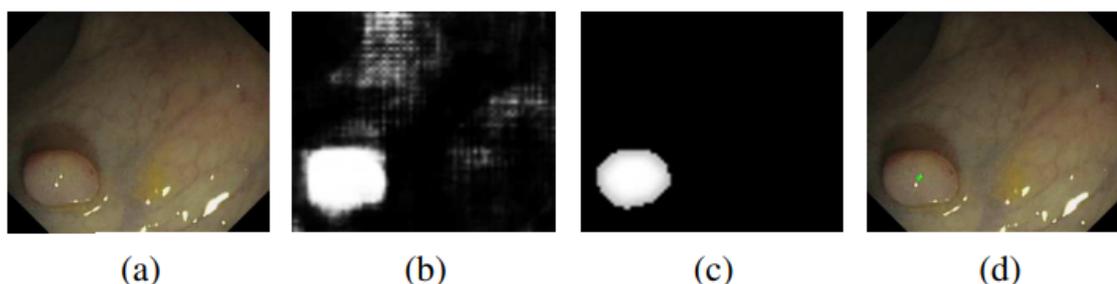


FIGURE 1.20 – Résultats du modèle avec (a) Image d'entrée. (b) Carte de probabilité en NG (c) Carte de probabilité filtrée après post-traitement (d) Résultat final.

Avantages des méthodes basées sur DL : Ces systèmes peuvent fournir une très grande précision à grande vitesse, et avec intervention limitée ou aucune de l'utilisateur, et avec une plus importante qualité des résultats ce qui permet d'économiser beaucoup de temps et d'argent.

Inconvénients : Comme ces méthodes ont des avantages ils ont aussi des limites parmi lesquelles :

- La nécessité de grandes quantités de données de d'apprentissage ;
- Peu de contrôle existe sur la manière dont le système apprend à déduire ;
- L'apprentissage peut être coûteuse en termes de calcul car la puissance de calcul est relative à la complexité et de la masse de données utilisée ;
- Notamment la difficulté à comprendre comment ces systèmes infèrent.

2.3 Les méthodes Hybrides

Ces méthodes combinent principalement des méthodes traditionnelle (Ad-hoc) de différents paramètres (texture, forme, contexte...) pour détecter des régions d'intérêt avec de l'apprentissage machine afin de discriminer ces régions d'intérêt et de détecté les polypes. Voici quelques-unes des méthodes les plus représentatives proposées à ce jour :

- Dans [12] **Iwahori et al.** proposent une technique de détection les régions de polypes candidats (voir Figure1.21) en se basant sur l'hypothèse qu'un polype est convexe (forme de "blob"). en utilisant un filtre Hessien et les techniques d'apprentissage machine. Puis une extraction des paramètres de couleurs et de contours s'en suit pour ces régions. Finalement, SVM (Support Vector Machine) est utilisée sur ces paramètres pour déterminer les possibles régions contenant un polype. Cette méthode permet d'obtenir une grande précision. La figure1.22 illustre les résultats de détection par cette méthode.

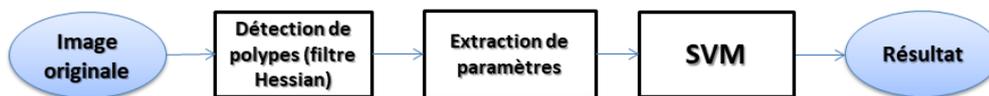


FIGURE 1.21 – Schéma d'architecture d'Iwahori et al . [12].

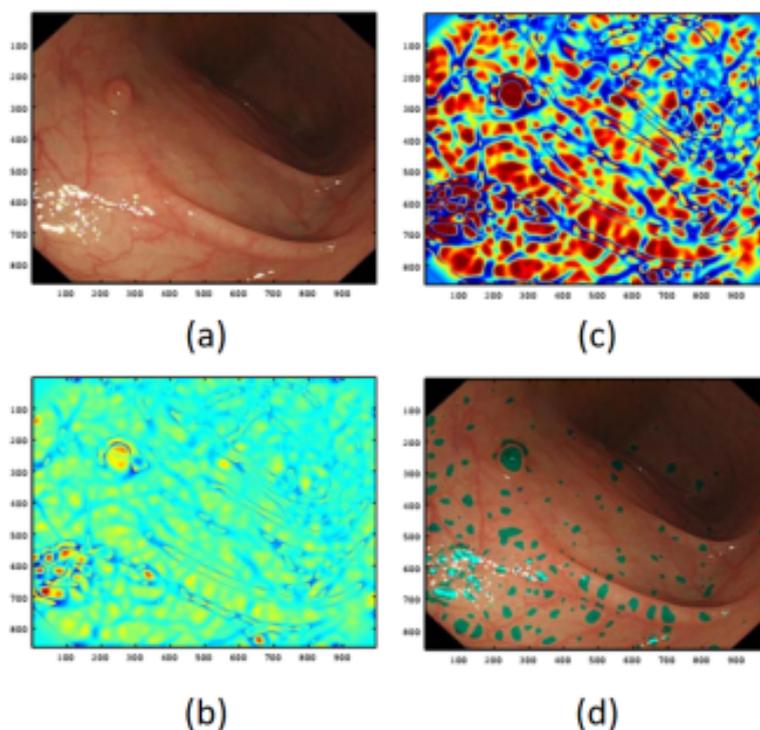


FIGURE 1.22 – Résultats de la méthode avec (a) image originale, (b) image améliorée, (c) image segmentée, (d) régions candidates.

- **Bae et al.** [13] ont mis au point une méthode (voir Figure1.23) permettant de détecter les polypes en utilisant l'apprentissage déséquilibré et l'apprentissage discriminatoire. Ils ont proposé d'utiliser l'apprentissage déséquilibré en raison du déséquilibre qu'il existe dans les bases de données de polypes (entre zones positives et zones négatives). Afin d'y parvenir, dans leur processus d'apprentissage, ils créent des classificateurs faibles avec des jeux de données rééquilibrés par échantillonnage ascendant/descendant, et génère un classificateur fort en combinant les faibles. Puis ils proposent une méthode d'apprentissage d'analyse des moindres carrés partielle pour améliorer la discrimination entre polype et non polypes. Les résultats montrent alors de nettes améliorations par rapport aux détecteurs classiques. La figure1.24 montre les résultats de cette méthode.

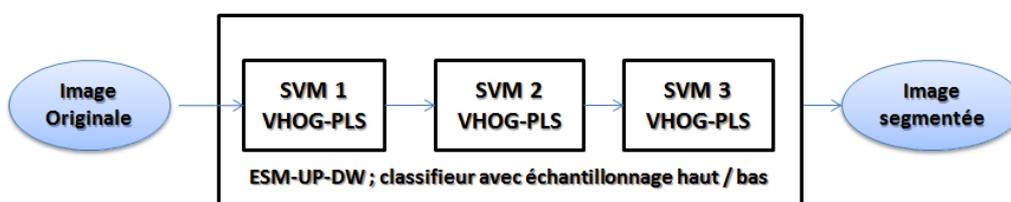


FIGURE 1.23 – Schéma d'architecture de méthode de Bae et al. [13].

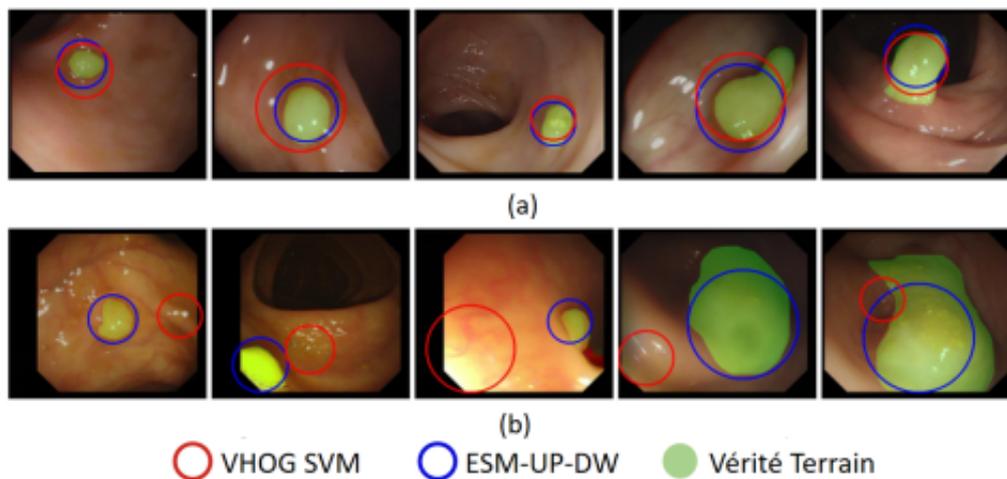


FIGURE 1.24 – Résultats de segmentation.

- **Zhang et al.** [14] s'intéressent à l'architecture FCN-8S pour la segmentation des polypes (voir Figure1.25). Le FCN fournit une prédiction pixel par pixel et des candidats initiaux de région de polype puis, ils ont classé ces régions candidates avec les fonctionnalités Texton et un classifieur de type forêt aléatoire. Les caractéristiques texton sont produites en utilisant l'algorithme de K-means sur la convolution du patch d'entrée et de filtres de Gabor pour différentes orientations.

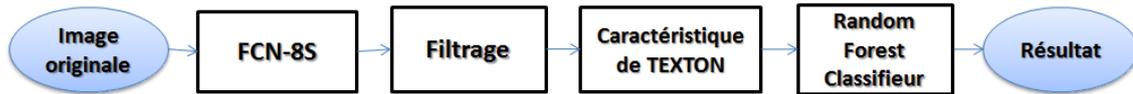


FIGURE 1.25 – Schéma de méthode Zhang et al. [14].

- **Bernal et al.** [15] ont proposé un schéma dénommé PLS (Polyp Localize and Spot Team) (voir Figure1.26) de localisation complète qui se compose de deux parties, la détection et la localisation. En ce qui concerne la détection, deux ensembles d’images, l’un contenant des polypes et l’autre sans polypes, sont utilisés pour l’apprentissage. Les caractéristiques globales de l’image sont utilisées car elles sont faciles et rapides à calculer. Le sous-système de détection décide en temps réel à quelle classe (polype ou pas de polype) appartient l’image d’entrée sur la base des scores de similarité entre l’image d’entrée et celles de l’apprentissage. Concernant le schéma de localisation, il est implémenté comme une séquence de filtres de pré-traitement (conversion de l’espace couleur RVB en YCbCr, suppression des bordures et des sous-images, masquage des effets de lumière et filtrage passe-bas). La décision finale concernant la localisation du polype est prise au travers des valeurs maximales dans la carte d’énergie calculée en utilisant la forme elliptique de l’apparence habituelle du polype. 1.27 montre les résultats de détection par cette méthode.

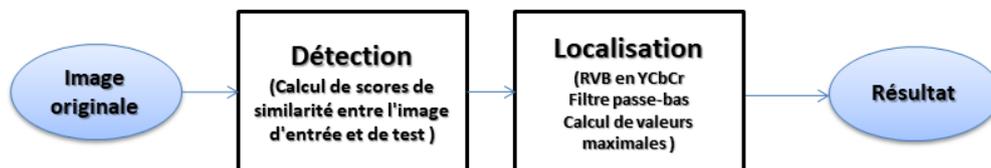
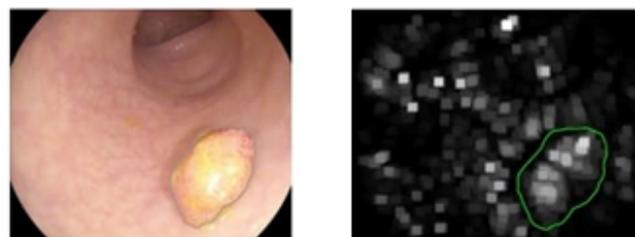


FIGURE 1.26 – Schéma d’architecture de méthode de Polyp Localize and Spot Team [15].



(a) Image originale

(b) Résultat

FIGURE 1.27 – Résultats de segmentation avec PLS.

- **Akbari et al.** [16] ont développé une nouvelle méthode de pour segmen-
tée les polypes colorectaux basés sur le CNN et un seuillage automatique
d'Otsu (voir Figure1.28). Tout d'abord ils ont utilisé FCN-8S pour la seg-
mentation des régions de polypes. Le FCN segmente les régions candidates
en fonction des caractéristiques extraites. Ensuite, dans le post-traitement
pour diminuer le taux de faux positifs de FCN-8S, ils ont appliqué la mé-
thode de seuillage Otsu pour transformer la carte de probabilité résultant
de FCN-8S en une image binaire, puis trouver le plus grand composant et
la considérer comme la plus probable. En outre, ils ont employé des mé-
thodes d'augmentation des données qui contiennent rotation de l'image ou
la sélection de patchs dans l'image. La rotation aide le FCN à apprendre
différentes structures des polypes dans différentes images et pour amélio-
rer la phase de training ils ont également utilisé une méthode de sélection
de patch. La figure1.29 montre les résultats de cette méthode.

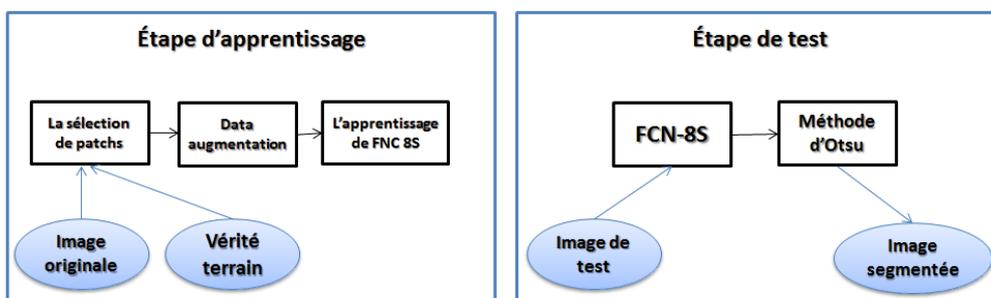


FIGURE 1.28 – Schéma de méthode FCN +Otsu [16].

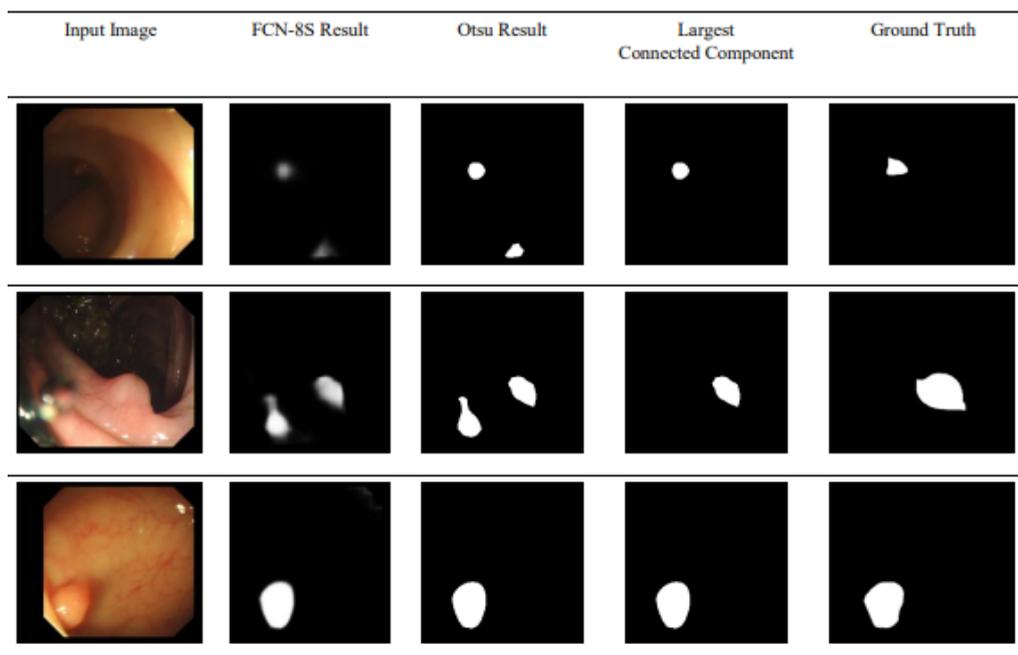


FIGURE 1.29 – Résultat de la méthode FCN +Otsu.

Avantages des méthodes hybrides : Les méthodes hybrides visent à obtenir le meilleur des deux méthodes (traditionnelle Ad-hoc et apprentissage profond DL), ce qui nécessite moins d'efforts d'apprentissage.

Inconvénients :

- La quantité de données requises est importante ;
- Le réglage des paramètres peut être difficile.

3 Synthèse des travaux

Le tableau 1.1 synthétise les différentes méthodes utilisées afin de segmenter les polypes colorectaux qui ont été proposées ces dernières années. Toutes les techniques rapportées peuvent être approximativement regroupées en trois catégories principales, celles basées sur les caractéristiques traditionnelle (Ad-Hoc), celles basées sur les caractéristiques profondes (DL), et celles qui combinent entre les 2 derniers méthodes appelées (Hybride).

Les méthodes Ad-Hoc peuvent être subdivisées en méthodes basées sur la forme et la texture, tandis que les méthodes profondes peuvent être subdivisées en approches basées sur les patches et les approches end-to-end. Bien que toutes ces méthodes aient leurs propres avantages et inconvénients spécifiques.

Pour découvrir plus en détail ce que les méthodes de DL peuvent offrir pour la segmentation sémantique, nous proposons une étude comparative de certaines méthodes et architectures de DL pour produire des résultats réalistes qui nous permettent d'évaluer la technologie et de comparer les méthodes utilisées. Dans ce projet de fin d'études, nous nous focalisons sur deux modèles entièrement convolutifs pour la segmentation des images de coloscopie à savoir : U-Net et SegNet qui ont prouvé leur efficacité dans la segmentation sémantique.

4 Conclusion

Comme nous l'avons vu dans ce chapitre, la segmentation des polypes est un problème très difficile à résoudre dû à la grande variété de formes, de tailles, de couleurs et de textures des polypes, ainsi que la présence d'autres structures de scène endoluminale (reflets spéculaires, aux vaisseaux sanguins et aux plis endoluminaux). Pour pallier ces problèmes, il existe diverses méthodes qui ont été proposées dans la littérature. Parmi elles, les méthodes basées sur l'apprentissage profond (DL) ont été largement adoptées dans la caractérisation, la classification et surtout la segmentation des images de coloscopies. Ces méthodes ont démontré leurs supériorités par rapport aux méthodes traditionnelles car elles offrent actuellement le meilleur rapport entre la précision de la segmentation et l'efficacité du calcul avec intervention limitée ou aucune de l'utilisateur ce qui permet d'économiser beaucoup de temps et d'argent.

Type	Méthode	Auteur	Technique	BDD
Ad-Hoc	Basée sur la forme	Bernal et al.	Ligne de partage des eaux, le modèle d'apparence	COLON-DB
			WM-DOVA	CVC-ClinicDB
		Tajbakhsh et al.	Descripteur Haar	300 images provenant d'une base de données public
			Descripteur de patch	CVC-ColonDB
		Du et al.	La propagation de l'attraction (AP)	800 images collectées
	Basée Contexte-forme	Tajbakhsh et al.	Edge map, un système de vote	CVC-ColonDB
	Basée sur la texture	Alexandre et al.	LBP Local Binary Pattern	4600 images collectées from 35 videos
Ameling et al.		GLCM Grey Level Co-occurrence Matrix	vidéo de 4h	
Deep learning	End-to-end	Yu et al	3D-FCN	ASU-Mayo
		Eu et al.	SegNet modifié	"CVC-ColonDB,CVC-ClinicDB et ETIS-LaribPolypDB"
		Branch et al.	U-Net-MobileNetV2	Kvasir-SEG
		Tomar et al.	FCN (full convolution network)	Kvasir-SEG
		Brandao et al.	FCN , technique SfS	ETIS-Larib, CVC-ColonDB
		Jie Meng et al.	Masque R-CNN	330 images collectées
	Basée sur les patchs	Banik et al.	CNN multi-échelle basé sur des patchs	CVC-Clinic DB
		Park et al.	CNN	CVC-ClinicDB
		Iwahori et al.	filtre Hessian ,SVM	128 images collectées
	Hybride		Bae et al.	ESM-UP-DW (classifieur avec échantillonnage haut /bas)
		Zhang et al.	FCN -8S, les caractéristiques texton	CVC-ColonDB
		Akbari et al.	FCN , Otsu	CVC-ColonDB

TABLE 1.1 – Synthèse des méthodes de détection du cancer colorectal assistée par ordinateur.

Chapitre 2

Approches de segmentation par réseaux convolutionnels

1 Introduction

Dans le premier chapitre, nous avons vu que les algorithmes basés sur l'apprentissage profond sont devenus très populaires ces dernières années, avec des performances en amélioration constante. Cependant, cette popularité ne signifie pas que les méthodes traditionnelles (Ad-Hoc) ont été complètement remplacées. Ce chapitre se consacrera à la présentation de l'approche développée pour la détection des polypes dans les images de coloscopies, nous développerons le principe et l'architecture de l'un des algorithmes d'apprentissage profond le plus populaire dans le domaine de médical à savoir U-Net.

2 La segmentation par les méthodes d'apprentissage profond

Il existe différents niveaux de granularité dans lesquels les ordinateurs peuvent comprendre les images. Pour chacun de ces niveaux, il réside un problème défini dans le domaine de la vision par ordinateur. Au cours des dernières années, les modèles de Deep Learning ont introduit de nouveaux modèles de segmentation d'images avec des améliorations de performance remarquables afin de résoudre ces problèmes. Ces modèles atteignent souvent les meilleurs taux de précision, ce qui a entraîné un changement dans plusieurs domaines. Parmi les types de segmentation les plus populaires on distingue : la segmentation sémantique et celle d'instances.

2.1 La segmentation sémantique

Le but de la segmentation sémantique est d'étiqueter chaque pixel d'une image avec une classe correspondante de ce qui est représenté. Car nous prédisons pour chaque pixel de l'image. Bien que, les résultats attendus de la segmentation sémantique ne sont pas seulement des étiquettes et des paramètres de boîte englobante (bounding box). Le résultat lui-même est une image haute résolution (géné-

ralement de la même taille que l'image d'entrée) dans laquelle chaque pixel est classé dans une classe particulière (voir exemple figure 2.1). Il s'agit donc d'une classification d'image au niveau du pixel, il est à noter que l'approche la plus utilisée dans la segmentation sémantique est : U-Net [17].

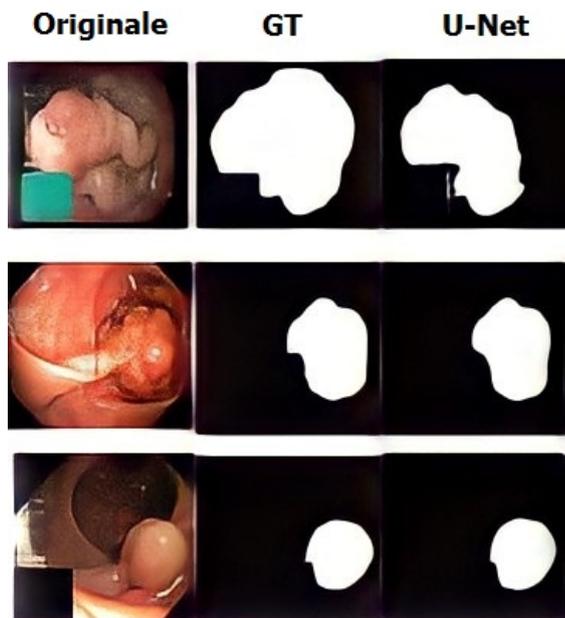


FIGURE 2.1 – La segmentation sémantique par U-Net [17].

2.2 La segmentation instance

La segmentation d'instance va plus loin, où, en plus de la classification au niveau du pixel, La segmentation par instance traite les objets multiples de la même classe comme des instances individuelles distinctes. Par exemple, si une image présente deux polypes, chaque polype est une instance de la classe "polype". Chacun sera classifié séparément, par exemple en utilisant différentes couleurs. Il est à noter que l'une des approches les plus utilisées dans la segmentation instance est : mask-RCNN [10].

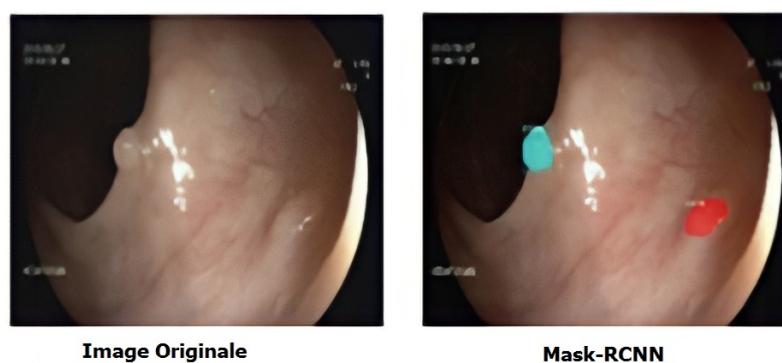


FIGURE 2.2 – La segmentation instance d'image coloscopique par mask-RCNN [10].

3 Étude comparative

Dans le cadre de ce projet de fin d'études, nous proposons une étude comparative avec deux architectures de réseaux de neurones entièrement convolutifs pour la segmentation des polypes colorectaux dans les images de coloscopie. Le premier modèle est hybride à base d'U-Net avec un post traitement à l'aide des méthodes de morphologie mathématique, le second modèle est SegNet. La figure 2.3, illustre le schéma du processus de comparaison qui va être détaillé par la suite.

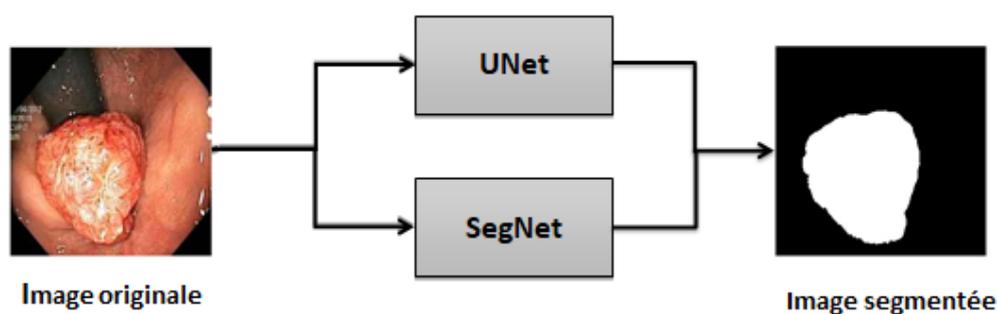


FIGURE 2.3 – Illustration du processus de comparaison.

4 Réseau entièrement convolutionnel (U-Net)

U-Net est un réseau end-to-end de segmentation sémantique d'images, il est construit à l'aide du réseau entièrement convolutionnel FCN. Il s'appuie sur les réseaux convolutifs précédents pour travailler plus efficacement avec moins d'images d'entraînement et pour réaliser une segmentation plus efficace. Dans ce réseau, l'encodeur et le décodeur ont une architecture similaire, à une différence près où la couche de sous-échantillonnage de l'encodeur est remplacée par une couche de sur-échantillonnage dans le décodeur.

Dans la figure 2.4 chaque boîte bleue correspond à une carte de caractéristiques multi-canaux. Le nombre de canaux est indiqué en haut de la boîte, et la taille de la carte est indiquée dans le coin inférieur gauche de la boîte. Les cases blanches représentent des cartes de caractéristiques copiées. Les flèches indiquent les différentes opérations (affichées également dans la légende). La partie gauche du U est l'encodeur, et la partie droite est le décodeur (nous allons détailler dans la section suivante) [28].

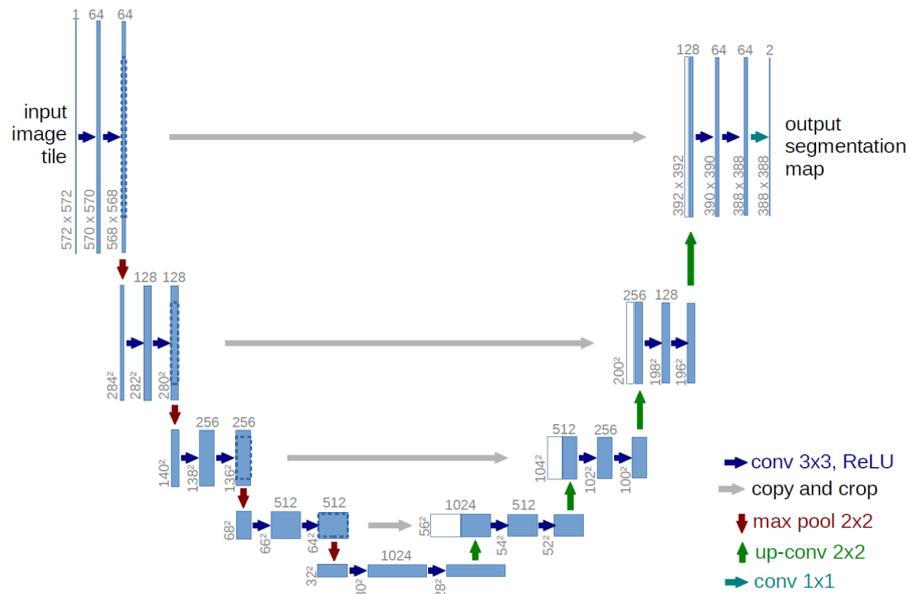


FIGURE 2.4 – L’architecture de Réseau entièrement convolutionnel U-Net pour une image 32×32 [17].

U-Net est l’une des architectures les plus populaires pour la segmentation des images médicales. Il se compose de deux chemins : un chemin de contraction et un chemin d’expansion.

4.1 Le chemin de contraction (Encodeur)

C’est un réseau convolutif typique utilisé pour capturer le contexte d’une image. Il est similaire à un CNN ordinaire, sans les couches entièrement connectées (FC) à la fin. Le rôle de l’encodeur est d’apprendre des représentations hautement abstraites de l’image d’entrée. Il peut être défini comme étant un assemblage de couches de “convolution” et de couches de “max pooling” permettant de créer une carte de caractéristiques d’une image et de réduire sa taille pour diminuer le nombre de paramètres du réseau. C’est ce qu’on appelle le sous-échantillonnage “down-sampling”. L’encodeur se compose de quatre blocs des couches suivantes :

- Deux convolutions 3×3 consécutives cross-channel unpadded avec stride 1.
- Une couche de max pooling 2×2 .
- Activations ReLU.

Chaque étape de sous-échantillonnage double le nombre de cartes de caractéristiques.

Couches de convolutions : Les couches de convolutions sont considérées comme les principaux éléments constitutifs des architectures CNN. Une convolution est définie comme une opération mathématique décrivant une règle pour la fusion de deux ensembles d’informations. Elle prend l’entrée, applique un noyau de

convolution et nous donne une carte de caractéristiques (feature map) comme sortie [18]. La figure 2.5 schématise comment le kernel (noyau) est glissé sur les données d'entrée pour produire les données caractéristiques convoluées (données de sortie).

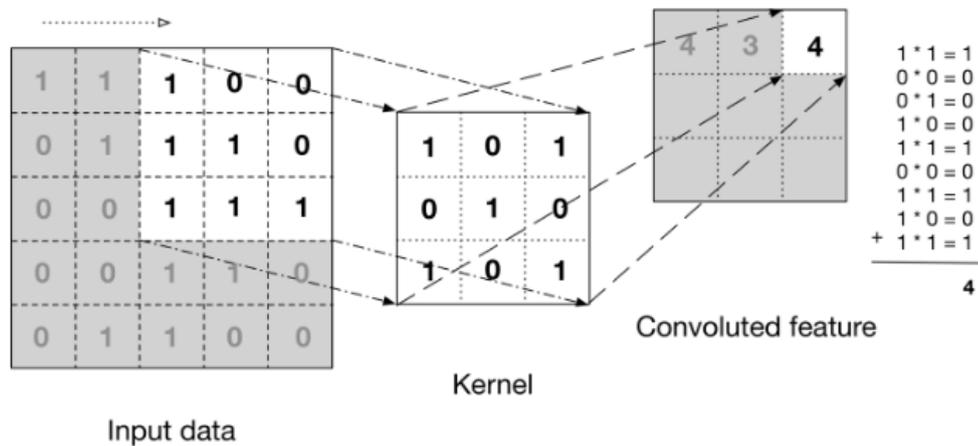


FIGURE 2.5 – Opération de convolution [18].

Voici les hyperparamètres qui déterminent la disposition spatiale et la taille du volume de sortie d'une couche de convolution :

- L'ensemble de filtres « k » (également appelés kernels ou extracteurs de caractéristiques), chacun de taille $(f \times f \times canaux)$, où f est généralement égal à 3 ou 5.
- Le pas (stride) " s " avec lequel nous faisons glisser le filtre. Lorsque le stride est de 1, nous déplaçons les filtres d'un pixel à la fois. Lorsque le pas est de 2, les filtres sautent de 2 pixels à la fois lorsque nous les faisons glisser. Cela produit des volumes de sortie plus petits dans l'espace.
- Le padding " P "; S'il est fixé à " $SAME$ ": la couche de convolution ajoute une marge à zéro et s'il est fixé à " $VALID$ ": la couche de convolution n'utilise pas de marge à zéro. Parfois, il est utile de compléter le volume d'entrée par des zéros sur le contour (voir la figure 2.6). La caractéristique du remplissage zéro est qu'il nous permet de contrôler la taille spatiale des volumes de sortie (le plus souvent, nous l'utiliserons pour préserver exactement la taille spatiale du volume d'entrée, de sorte que la largeur et la hauteur de l'entrée et de la sortie soient identiques) [19].

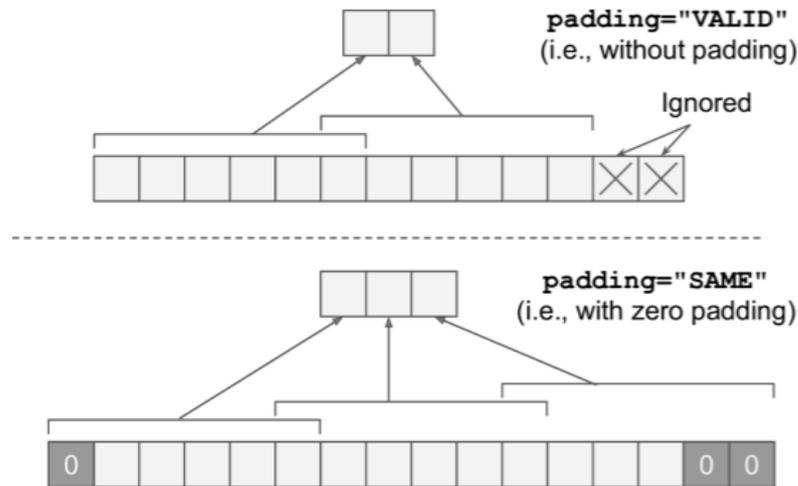


FIGURE 2.6 – les types de padding SAME et VALID [19].

Couches de pooling : Les couches de pooling sont généralement insérées entre des couches de convolutions successives afin de réduire progressivement la taille spatiale (largeur et hauteur) de la carte de caractéristiques et pour avoir moins de paramètres dans le réseau. La couche de pooling fonctionne indépendamment sur chaque tranche de profondeur de l'entrée [18]. L'opération de sous-échantillonnage la plus courante est l'opération max pooling dans laquelle seule la valeur d'entrée maximale peut passer à la couche suivante comme illustrée dans la figure 3.11.

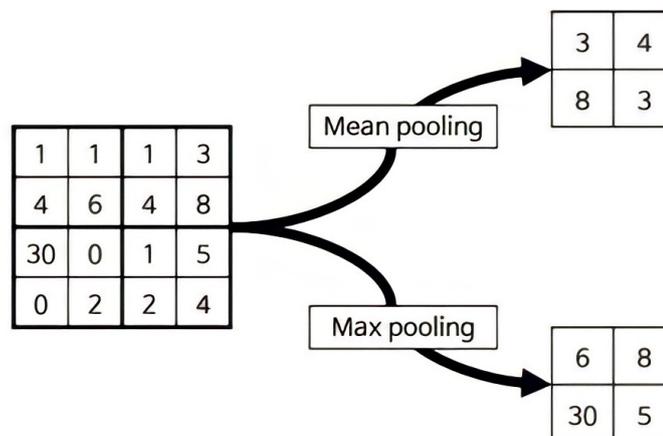


FIGURE 2.7 – Illustration de l'opération de pooling [18].

Fonction d'activation : La sortie des couches précédentes n'est qu'une transformation linéaire de l'entrée. Cette linéarité implique que le réseau ne satisfera pas le théorème d'approximation universelle. Tant que l'on n'élimine pas cette linéarité, le pouvoir de représentation du réseau est limité. Afin de transformer le réseau en un approximateur universel, des non-linéarités doivent être introduites. Dans le domaine de Deep learning ces non-linéarités sont introduites via des couches d'activation, une fonction d'activation est une fonction mathématique utilisée sur un signal en sortie d'un neurone artificiel. Concrètement, elle va

avoir pour rôle de décider si on active ou non une réponse du neurone. Les trois fonctions les plus populaires, *Sigmoïde*, *Tanh* et *ReLU* (représentées sur la figure 3.12) sont brièvement expliquées :

- *Sigmoïde* : cette fonction écrase les résultats dans la plage $[0, 1]$. Bien qu'elle ait été largement utilisée, elle présente des inconvénients qui rendent le processus de formation impossible dans certaines situations : saturation du gradient et sorties non centrées sur le zéro.
- *Tanh* : comme la fonction précédente, *tanh* écrase les résultats dans un intervalle mais, au lieu de $[0, 1]$, *tanh* le fait dans l'intervalle $[-1, 1]$. Comme la fonction précédente, elle sature toujours les gradients mais la sortie est centrée sur zéro, ce qui en fait une meilleure alternative à la fonction sigmoïde.
- *ReLU* : la fonction *ReLU* est la fonction utilisée dans UNet. *ReLU* fonctionne en seillant l'entrée à zéro, ce qui rend l'apprentissage par descente de gradient beaucoup plus rapide, car elle ne sature pas et a un coût de calcul plus faible que les deux fonctions mentionnées ci-dessus. Grâce à ces caractéristiques, elle a amélioré les résultats des fonctions susmentionnées dans plusieurs benchmarks. Cela fait de *ReLU* la fonction d'activation la plus commune pour les CNN [18].

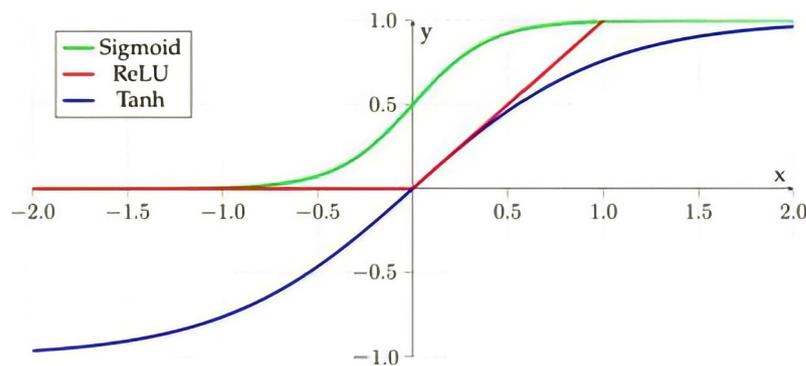


FIGURE 2.8 – Les fonctions d'activation les plus populaires [18].

4.2 Le chemin d'expansion symétrique (décodeur)

Il fait passer les petites cartes de caractéristiques par une série d'étapes de sur-échantillonnage (Up-sampled convolution) pour revenir à la taille de l'image d'origine, Il permet aussi une localisation précise grâce à la convolution transposée. Le sur-échantillonnage fonctionne avec des convolutions transposées 2×2 avec un stride 2 (indiqué par les flèches verticales vertes dans le schéma). La sortie de chaque étape de sur-échantillonnage est concaténée avec les cartes de caractéristique haute résolution recadrée de l'étape de codage correspondante (flèches horizon-tales grises). Chaque convolution transposée est suivie de deux convolutions régulières pour lisser l'image étendue. Les étapes de sur-échantillonnage divisent par deux le nombre de cartes de caractéristiques. La sortie finale utilise une convolution de goulot d'étranglement 1×1 pour mapper le tenseur des cartes de caractéristiques à 64 composants au nombre de classes souhaité.

Up-Sampled convolution Permet de convertir une image à basse résolution en une image à haute résolution pour récupérer l'information "OÙ" (la localisation). Dans la littérature, il existe plusieurs techniques d'échantillonnage, parmi elles ; l'interpolation bi-linéaire, l'interpolation cubique, l'interpolation par le plus proche voisin, le découplage, la convolution transposée, etc. Cependant, la convolution transposée est le choix préféré pour le sur-échantillonnage (Up-Sampling) d'une image.

Convolution transposée : appelée aussi déconvolution ou Fractionally-strided convolution, La convolution transposée est exactement le processus inverse d'une convolution normale (voir figure 2.9), c'est-à-dire que l'entrée est une image à faible résolution et la sortie est une image à haute résolution. La convolution transposée n'est pas une convolution. Mais nous pouvons émuler la convolution transposée en utilisant une convolution. Nous échantillons l'entrée en ajoutant des zéros entre les valeurs de la matrice d'entrée de manière que la convolution directe produise le même effet que la convolution transposée [20].

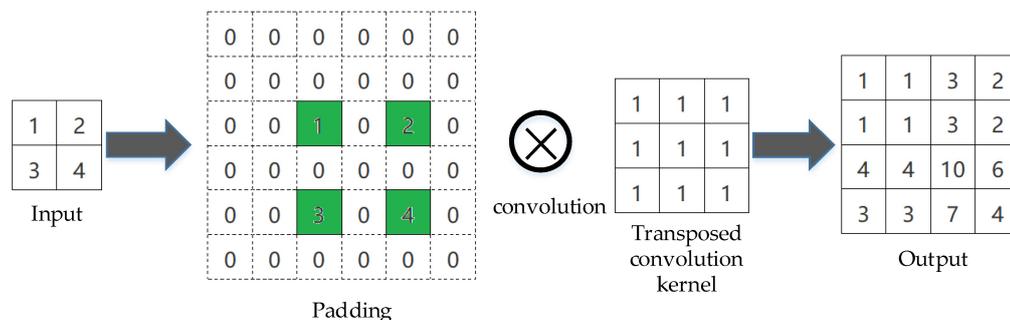


FIGURE 2.9 – Illustration de la convolution transposée [20].

Pour obtenir des localisations plus précises, à chaque étape du décodeur, nous utilisons des connexions de saut en concaténant la sortie des couches de convolution transposées avec les cartes de caractéristiques du codeur au même niveau. Ensuite, après chaque concaténation, nous appliquons à nouveau deux convolutions régulières consécutives afin que le modèle puisse apprendre à assembler une sortie plus précise. C'est ce qui donne à l'architecture une forme symétrique en "U", d'où le nom U-Net.

4.3 Bloc de connexion résiduelle

Pour mieux segmenter les ROIs, le modèle utilise également des connexions de saut (Skip connexion), cette connexion par saut est un concept vital pour préserver les pertes des couches précédentes afin qu'elles se reflètent plus fortement sur les valeurs globales. Il est également prouvé que les connexions de saut produisent de meilleurs résultats et conduisent à une convergence plus rapide du modèle. En concaténant les représentations de sur-échantillonnage à chaque étape avec les cartes de caractéristiques correspondantes à l'étape de contraction.

4.4 Backbones et ses utilisations

Ces backbones sont utilisées comme la première moitié (encodeur) du U-Net ce qui a pour avantage de raccourcir la procédure d'apprentissage, qui est effectuée par les dernières couches du réseau, d'accélérer la convergence et d'obtenir des performances élevées par rapport à un modèle non-pré-entraîné. Les backbones utilisés sont souvent des Vanilla CNNs tels que VGG, ResNet, Inception, EfficientNet, etc. qui effectuent l'encodage et le dé-échantillonnage par eux-mêmes. Ces réseaux sont appliqués et leurs homologues sont construits pour effectuer le décodage et le suréchantillonnage afin de former l'U-Net final [29].

De nombreuses recherches ont été menées pour améliorer ces réseaux en utilisant des backbones capables de mieux extraire les caractéristiques. Il existe plusieurs autres techniques d'amélioration, comme l'utilisation de l'ASPP (Atrous Spatial Pyramid Pooling) pour capturer les informations contextuelles et l'utilisation de blocs de compression et d'excitation pour réduire les informations redondantes.

4.5 Avantages et inconvénients

Les applications d'U-Net sont multiples allant de véhicules autonomes, diagnostic d'images biomédicales, géo-détection, agriculture de précision et problèmes commerciaux.

Une telle architecture est idéale pour préserver la taille initiale. U-Net est symétrique et les connexions de saut entre le chemin de contraction et d'expansion combinent les informations d'emplacement du chemin de sous-échantillonnage avec les informations contextuelles dans le sur-échantillonnage et donc permet d'éviter la perte de caractéristiques.

De plus, il n'a pas de couche dense, ce qui signifie que différentes tailles d'image peuvent être utilisées comme entrée puisque les seuls paramètres à apprendre sur les couches de convolution sont le noyau. Le modèle U-Net peut être utilisé sur différents ensembles d'images et les résultats sont tout à fait satisfaisants.

Cependant, ce modèle comprend mieux "QUOI" est représenté dans l'image, mais il perd l'information de "OÙ" il est représenté à cause du down-sampling.

5 Réseau entièrement convolutionnel (SegNet)

SegNet est un modèle de segmentation sémantique Introduit par Badrinarayan et al. [21]. Cette architecture centrale de segmentation entraînable se compose d'un réseau encodeur, d'un réseau décodeur correspondant et d'une couche de classification par pixel.

5.1 Encodeur

L'architecture du réseau encodeur est similaire à celle du modèle VGG-16 [30], elle n'a pas des couches entièrement connectées que des convolutions (l'architecture est montrée dans la figure 2.10).

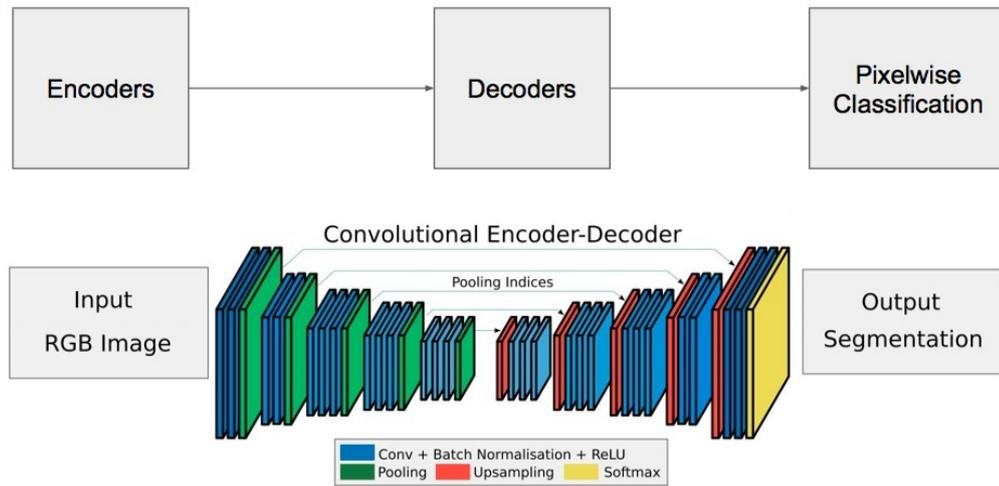


FIGURE 2.10 – Illustration de l'architecture de SegNet [21].

5.2 Décodeur

Le rôle du réseau décodeur est de mettre en correspondance les cartes de caractéristiques de basse résolution du codeur avec les cartes de caractéristiques de pleine résolution de l'entrée pour la classification par pixel. La nouveauté de SegNet réside dans la manière dont le décodeur sur-échantillonne ses cartes de caractéristiques d'entrée à faible résolution (voir figure 2.11). Plus précisément, le décodeur utilise les "indices de pooling" calculés dans l'étape de max-pooling du codeur correspondant pour effectuer un sur-échantillonnage non linéaire cela facilite le processus de formation puisque le réseau n'a pas besoin d'apprendre à nouveau les poids de sur-échantillonnage.

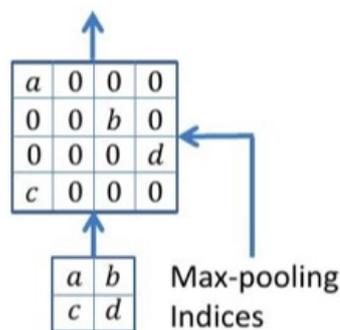


FIGURE 2.11 – Le sur-échantillonnage par les indices de max pooling [21].

5.3 Softmax

Le Softmax est utilisé souvent en sortie des réseaux de neurones pour avoir des "probabilités" pour chaque pixel (voir Figure 2.12). Softmax est une fonction qui normalise une valeur d'entrée dans un vecteur de probabilité dont la somme totale est égale à 1. Par exemple, si on a en entrée 0.1 et 0.2, alors Softmax donnera 0.48 et 0.52.

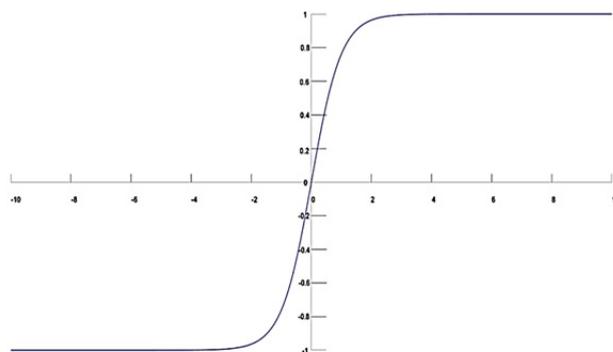


FIGURE 2.12 – La fonction Softmax.

En effet, la fonction softmax est une généralisation de la régression logistique qui peut être utilisée pour la classification multi-classes. Dans la plupart du temps, la fonction Softmax est liée à la fonction Cross Entropy. Sa formule est très similaire à la fonction sigmoïde qui est utilisée pour la régression logistique.

6 Différences entre SegNet et U-Net

SegNet et U-Net sont deux architectures de segmentation sémantique. Dans SegNet, seuls les indices de pooling sont transférés vers le chemin d'expansion à partir du chemin de compression, utilisant moins de mémoire. Alors que dans U-Net, des cartes de caractéristiques entières sont transférées du chemin de compression au chemin d'expansion, ce qui nécessite beaucoup de mémoire.

7 Conclusion

Dans ce chapitre, nous avons présenté les approches utilisées pour réaliser le processus de segmentation des images endoscopiques. Dans un premier temps, nous avons défini les différents type de segmentation, et dans un second temps, nous avons présenté les techniques de segmentation sémantiques basées sur deep learning : U-Net et SegNet, les détails de leurs architectures et la définition de chaque couche étant présentée. Le choix des modèles entièrement convolutionnel est motivé par leurs grandes perspectives dans le domaine de l'analyse des images médicales.

Chapitre 3

Étude comparative : expérimentations et résultats

1 Introduction

Dans ce chapitre, nous allons présenter les expérimentations effectuées afin d'obtenir un modèle permettant de détecter les polypes dans les images coloscopiques afin d'aider les médecins dans le diagnostic.

2 Base d'images d'étude

L'ensemble de données Hyper-Kvasir [31] qui est le plus grand dataset d'images et de vidéos du tractus gastro-intestinal (GI) disponible aujourd'hui dans le contexte de challenge GI annuel tels que (Le challenge Medico 2017, Medico Task 2018 et du BioMedia Challenge 2019).

L'objectif de l'organisation de ces défis est d'évaluer les approches d'apprentissage machine (ML) pour la classification d'images endoscopiques avec un accent spécifique sur la vitesse et la robustesse des méthodes [32]. Le challenge consiste en trois tâches différentes (Tâche de détection, Tâche de détection efficace et la tâche de segmentation) chacune ciblant une exigence différente pour l'utilisation en clinique. Les images et vidéos d'HyperKvasir ont été collectées de manière prospective à partir d'examens cliniques de routine réalisés dans un hôpital norvégien entre 2008 et 2016. Les images ont été récupérés à partir de la base de données de documentation d'images Picsara (CSAM, Norvège), un module d'extension du système de dossiers médicaux électroniques.

La base de données contient un total de 373 vidéos contenant différents résultats et points de repère (landmarks). Cela correspond à environ 11,62 heures de vidéos et 1 059 519 images vidéo qui peuvent être converties en images si nécessaire. Chaque vidéo a été évaluée manuellement par un professionnel de la santé travaillant dans le domaine de la gastro-entérologie et a donné lieu à un total de 171 résultats annotés.

Pour notre sujet d'étude, nous allons travailler avec 1000 images munies du masque de segmentation de la classe des polypes. Dans le masque, les pixels représentant le tissu du polype, la région d'intérêt, sont représentés par le foreground (masque blanc), tandis que le background (en noir) ne contient pas de pixels de polype(Figure 3.1) [31].



FIGURE 3.1 – Exemple d'un polype (Image originale) avec le masque.

3 Modélisation et paramétrages

3.1 Environnement de développement et outils utilisés

Pour accomplir ce travail nous avons besoin de machines de calculs puissantes, car l'apprentissage profond et ses architectures nécessitent de grandes bases de données d'apprentissage. Pour cela nous avons utilisé deux machines de calculs afin d'accélérer les calculs et minimiser le temps.

- Machine 1 : PC HP : Intel(R) Core(R) CPU - 3.40 GHz - RAM 8,00 Go.
- Machine 2 : PC MSI : AMD Ryzen Threadripper 1950X 16-CORE 3.40 Ghz - RAM 16,00 Go.
- Logiciel Matlab 2019b.

3.2 Redimensionnement des images

Malgré toutes les façons dont les réseaux de neurones convolutifs ont révolutionné la vision par ordinateur ces dernières années, un aspect important a reçu étonnamment peu d'attention : l'effet de la taille de l'image sur la précision des tâches pour lesquelles on s'entraîne. En général, pour être efficaces, les images d'entrée sont redimensionnées à une résolution spatiale relativement faible [33].

Lorsqu'une image est de grande taille, elle peut contenir des millions de pixels, cela signifie que stocker toutes les informations d'une telle image nécessite une grande mémoire et prend autant de temps et besoin d'une vitesse importante

pour les calculs. De ce fait, nous avons redimensionné les images, nous avons remarqué que la taille des images disponibles est de tailles différentes (720×576) et (768×576), delà, afin de minimiser le nombre de pixels et de diminuer le temps de calcul, tout en gardant une bonne qualité des images pour améliorer la performance de notre programme, nous avons choisi la taille 256×256 en utilisant la fonction « resize » dans MATLAB.

Ce choix de redimensionnement a été fait selon les recherches effectués à l'université de Lund par Olivier Rukundo sur les effets de la taille de l'image sur les performances de l'apprentissage profond via la segmentation sémantique d'images cardiaques par résonance magnétique avec U-Net [34], la taille optimale de l'image est de 256×256 .

3.3 Hyperparamétrage

Les modèles DL ont plusieurs hyper-paramètres, et déterminer la configuration idéale de ces paramètres n'est pas facile. La détermination des hyper-paramètres nécessite de l'expérience et beaucoup d'observation et d'expérimentation. Configurer les hyper-paramètres, tels que la descente de gradient, le taux d'apprentissage LR (learning rate), le nombre d'epochs, et batchsize, n'est pas simple. Ils peuvent être ajustés tout au long de l'apprentissage du modèle. Nous devons fixer la valeur optimale de ces hyper-paramètres à travers plusieurs expérimentations pour que le modèle génère les meilleurs résultats.

La descente de gradient

La descente de gradient est un algorithme d'optimisation couramment utilisé pour former des modèles d'apprentissage automatique et des réseaux neuronaux. Les données d'apprentissage aident ces modèles à apprendre au fil du temps. Parmi les algorithmes de descente de gradient les plus populaires, citons : Momentum, Adagrad, Adam, AdaMax.

Learning rate

Learning rate (Le taux d'apprentissage) LR est un hyper-paramètre important à optimiser pour un apprentissage effectif des réseaux profonds. Même avec un LR constant comme point de départ, la sélection de ce seuil pour l'apprentissage d'un réseau profond est délicate. Le processus de réglage des hyper-paramètres est un travail d'équilibre précis entre under-fitting et over-fitting. Lorsqu'un modèle est incapable de minimiser l'erreur pour l'ensemble de test ou d'apprentissage, on parle d'under-fitting (sous apprentissage). En revanche, lorsqu'un modèle est si puissant qu'il généralise trop bien que sur l'ensemble d'apprentissage, un over-fitting (sur apprentissage) se produit, et l'erreur de généralisation augmente. Le over-fitting peut également se produire si le LR est trop faible (voir figure3.2).

Nous devons choisir le LR avec précaution, car si la valeur est élevée, la solution optimale est manquée, si valeur basse le réseau met une éternité à converger.

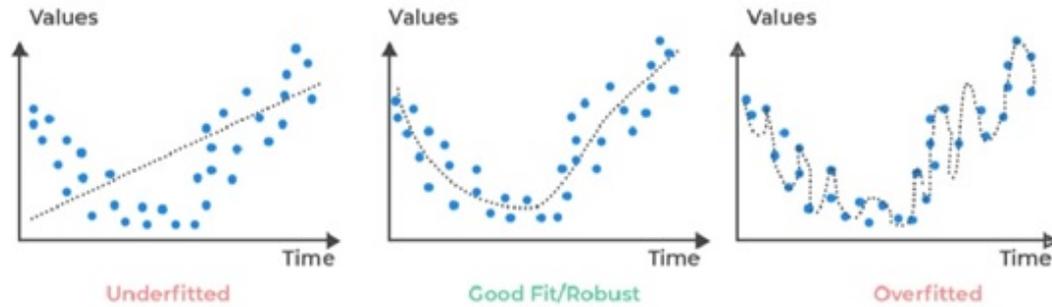


FIGURE 3.2 – Représentation des cas de sur et sous apprentissage par rapport à l’optimal.

Batch size

Batch size est un hyperparamètre qui détermine le nombre d’échantillons qui seront transmis au réseau en une seule fois. Nous pouvons considérer un batch comme une boucle for itérant sur un ou plusieurs échantillons et effectuant des prédictions. Lorsque le batch est terminé, la valeur d’erreur est calculée en comparant les prédictions et la sortie prévue. À partir de cette erreur, l’algorithme mis à jour est utilisé pour évoluer le modèle. Un ensemble de données d’apprentissage peut être divisé en un ou plusieurs batches.

Epochs

Nous parlons d’une epoch lorsqu’un ensemble de données entier n’est traité par le réseau neuronal qu’une seule fois. Nous divisons une epoch en plusieurs batches plus petits. Il n’y a pas de nombre fixe d’époques. Cependant, nous pouvons supposer que le nombre d’epochs est proportionnel à la diversité des données.

4 Les architectures proposées

4.1 U-Net

Dans ce travail nous avons modélisé une architecture d’U-Net qui contient 46 couches et 48 connections, nous avons choisi 3 comme une profondeur (EncoderDepth) avec 2 classes (polype et Background) pour l’optimisation nous avons utilisé l’algorithme d’Adam avec un Batchsize de 32 et 10 epochs.

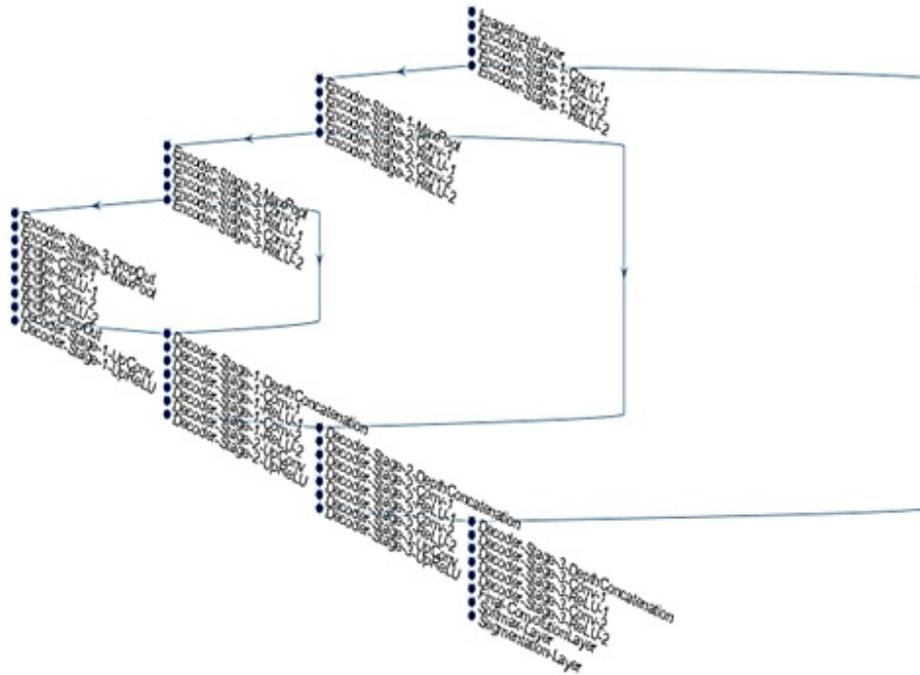


FIGURE 3.3 – Modélisation de l’architecture U-Net utilisée.

4.2 SegNet

A des fins de comparaison, nous avons appliqué le modèle SegNet sur notre base d’images. Dans ce modèle nous avons utilisé 45 couches avec 50 connections (Voir Figure 3.4) et nous avons choisi 3 comme paramètre de profondeur (EncodeurDepth) avec 2 classes avec l’algorithme Adam pour l’optimisation avec un miniBatchsize de 32.

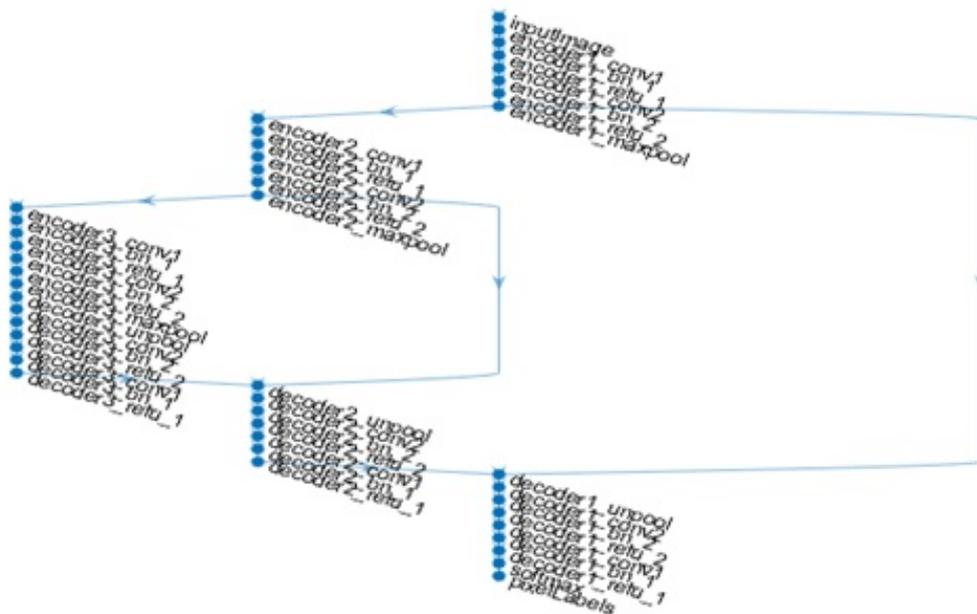


FIGURE 3.4 – Modélisation de l’architecture SegNet utilisée.

5 Apprentissage et expérimentations

Pour entraîner nos modèles, nous utilisons 1000 images de la base de données Hyper-Kvasir. Nous avons lancé l'entraînement plusieurs fois pour qu'on puisse trouver et fixer les performances appropriées pour notre modèle en changeant les hyper-paramètres (Epochs et Minibatch).

Notre modèle a été compilé avec l'optimiseur Adam, et nous utilisons la fonction de perte focale binaire puisqu'il n'y a que deux classes ; le blanc pour les pixels des objets segmentés (polype) et le noir pour le reste des pixels (Background).

5.1 Expérimentation 1

Pour le premier entraînement d'U-Net et SegNet, nous utilisons une taille miniBatch de 32, avec 10 epochs en fixant le taux d'apprentissage (LR) à 0.001.

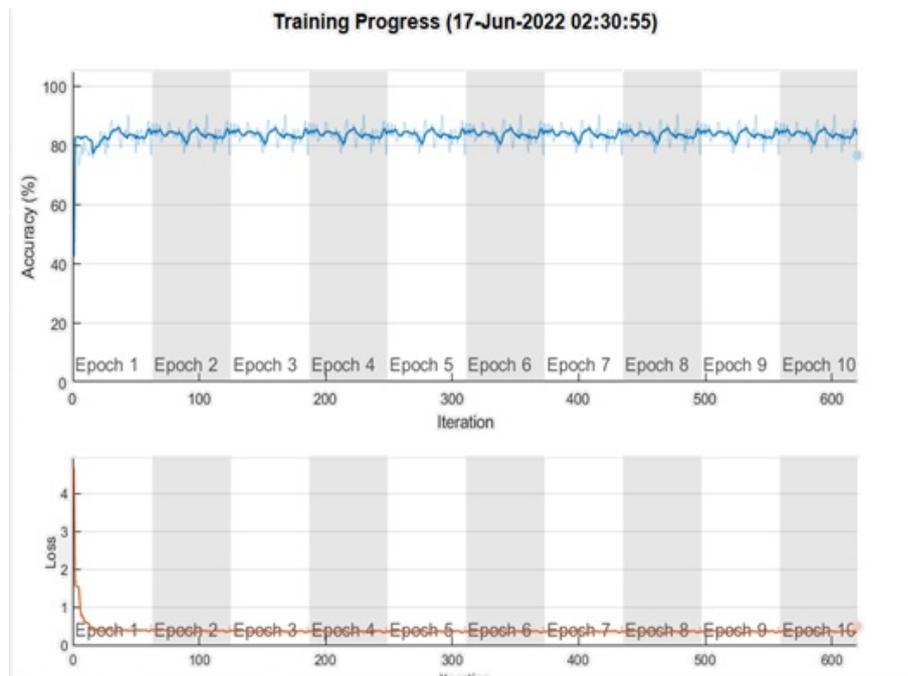


FIGURE 3.5 – Progression de l'apprentissage d'U-Net expérimentation 1.

```
Command Window
Training on single CPU.
Initializing input data normalization.
```

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:40	42.34%	4.7001	0.0010
1	50	00:25:28	77.60%	0.4571	0.0010
2	100	00:50:23	79.05%	0.4143	0.0010
3	150	01:15:20	77.32%	0.4020	0.0010
4	200	01:40:12	87.86%	0.2981	0.0010
5	250	02:05:00	84.89%	0.3756	0.0010
5	300	02:29:54	80.30%	0.4005	0.0010
6	350	02:54:57	83.69%	0.3756	0.0010
7	400	03:19:49	79.25%	0.4038	0.0010
8	450	03:44:04	83.94%	0.3573	0.0010
9	500	04:08:18	87.49%	0.3196	0.0010
9	550	04:32:47	85.52%	0.3334	0.0010
10	600	04:57:10	85.63%	0.3140	0.0010
10	620	05:06:52	76.64%	0.4935	0.0010

FIGURE 3.6 – Tableau de performances d’apprentissage d’U-Net expérimentation 1.

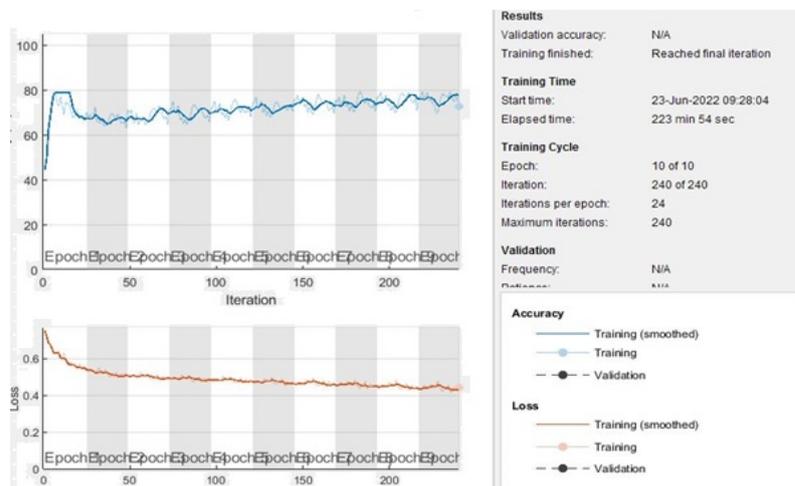


FIGURE 3.7 – Progression de l’apprentissage de SegNet expérimentation 1.

```
Command Window
Training on single CPU.
Initializing input data normalization.
```

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:02:14	44.36%	0.7516	0.0010
3	50	01:08:48	69.92%	0.4988	0.0010
5	100	01:52:24	70.96%	0.4881	0.0010
7	150	02:32:21	77.02%	0.4650	0.0010
9	200	03:12:06	73.68%	0.4705	0.0010
10	240	03:43:54	72.59%	0.4425	0.0010

FIGURE 3.8 – Tableau de performances d’apprentissage de SegNet expérimentation 1.

Les figures 3.5, 3.7 montrent l'évolution des valeurs des paramètres lors de l'apprentissage des deux modèles sous les mêmes conditions fixées ; les figures 3.6 et 3.8 déroulent les performance et les résultats obtenus à chaque epoch du processus d'entraînement de nos modèles. Nous remarquons clairement que la précision d'entraînement d'U-Net augmente nettement mieux que celles de SegNet, en effet, la fonction de perte d'U-Net diminue plus rapidement que SegNet.

5.2 Expérimentation 2

Pour le second entrainement des deux modèles, nous utilisons une taille miniBatch de 32, avec 20 epochs et un taux d'apprentissage (LR) a 0.001.

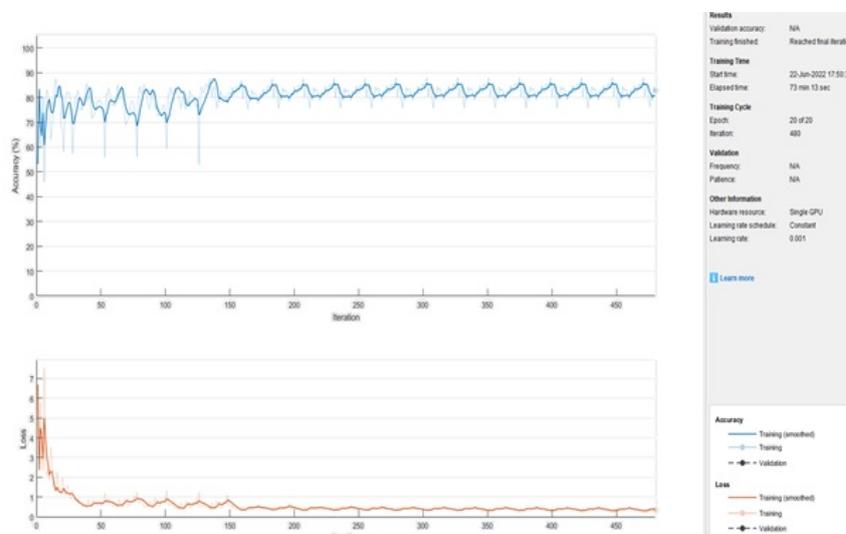


FIGURE 3.9 – Évolution des performances d'apprentissage d'U-Net expérimentation 2.

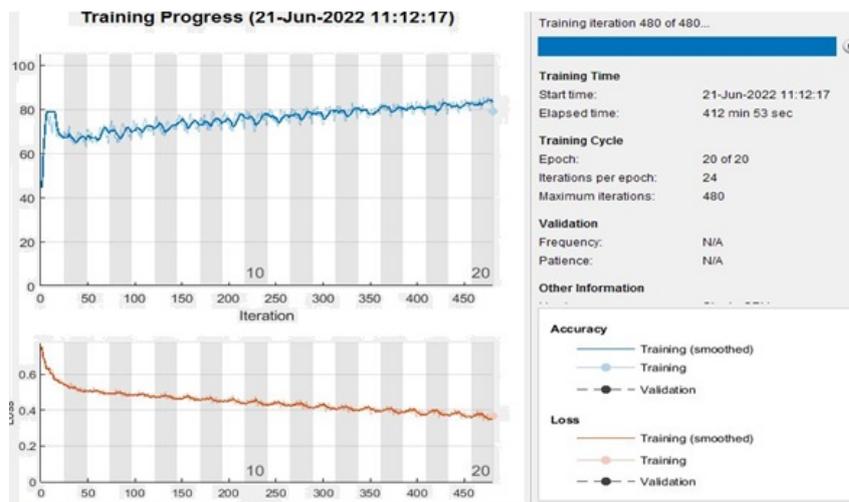


FIGURE 3.10 – Évolution des performances d'apprentissage de SegNet expérimentation 2.

Lorsque nous regardons les courbes de précision et de perte d'apprentissage du modèle U-Net et SegNet représenté dans les figures 3.9 et 3.10, nous pouvons observer que la précision de SegNet augmente jusqu'à la valeur de 82% par ailleurs celle d'U-Net atteint le taux de 85% avec le nombre d'époques 20. Tandis que la perte d'entraînement d'U-Net diminue plus rapidement avec le nombre d'époques. Finalement nous pouvons déduire que le choix des valeurs d'hyperparamètres a affecté les résultats et ceci nécessite une riche connaissance et une grande investigation.

6 Résultats et discussion

Une fois que la phase d'apprentissage a été réalisée, nous devons évaluer la performance du modèle. Pour savoir si notre modèle est performant, nous appliquons les mesures les plus courantes pour la segmentation sémantique, à savoir : la précision des pixels, Coefficient Dice (Score F1) et l'intersection sur l'union (ou IoU).

La précision des pixels (Pixel Accuracy)

La précision des pixels est peut-être le concept le plus facile à comprendre. Il s'agit du pourcentage de pixels de votre image qui sont classés correctement en calculant les Vrai Positif (VP), Vrai Négatif (VN), Faux Positif (FP) et Faux Négatif (FN). Bien qu'elle soit facile à comprendre, ce n'est en aucun cas la meilleure métrique car parfois elle fournit des résultats trompeurs lorsque la représentation de la classe est faible dans l'image,

$$Precision = \frac{(VP + VN)}{(VP + VN + FP + FN)}$$

Le coefficient Dice

Également connu sous le nom de coefficient Sørensen-Dice et de score F1, il est égal à 2 * la zone de chevauchement divisée par le nombre total de pixels dans les deux images.

$$F1(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}$$

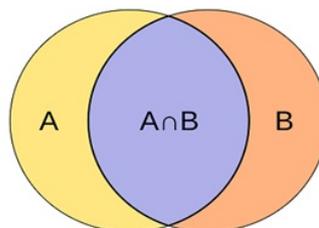


FIGURE 3.11 – Le coefficient Dice.

Intersection over Union (IoU)

Également connu sous le nom d'indice de Jaccard Il s'agit d'une autre méthode pour évaluer les prédictions d'un modèle de segmentation d'image. L'IoU est une mesure très simple et extrêmement efficace. Cette méthode calcule la performance en calculant l'intersection et l'union entre la vérité du terrain et la prédiction.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


FIGURE 3.12 – Intersection over Union.

6.1 Résultats

Dans l'étape de test, nous allons évaluer nos modèles en utilisant les métriques présentées précédemment sur 50 images de test, les résultats obtenus sont présentés comme suit :

Modèle U-Net

La figure 3.13 présente les résultats de segmentation de l'architecture U-Net avec les différents epochs 10 et 20, comme nous pouvons le voir, les résultats sont insatisfaisants, nous n'avons pas pu atteindre une segmentation nette, cela est due à un manque d'exemples d'apprentissage pour mieux généraliser, notons que ces images sont proposées dans un challenge, leur variabilité et complexité étant un facteur important.

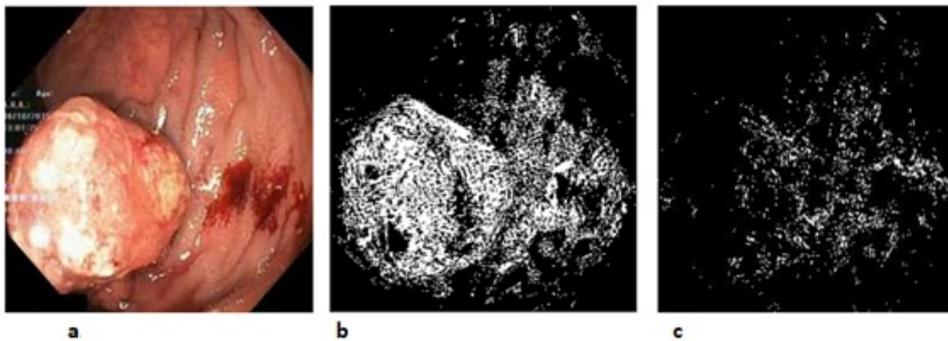


FIGURE 3.13 – a) image originale, b) image segmentée avec U-Net 10 epochs, c) image segmentée avec U-Net 20 epochs.

Afin d'améliorer nos résultats, nous avons décidé d'ajouter un post-traitement à l'architecture U-Net générée avec le nombre d'epochs égal à 10.

Post-traitement appliqué sur les résultats d'U-Net

Afin d'améliorer les résultats obtenus par le modèle U-Net, nous avons appliqué un post traitement avec :

- IMFILL : une fonction permet de remplir les régions et les trous de l'image).
- BWAREAOPEN : supprimer les petits objets d'une image binaire) et des méthodes de morphologies mathématiques.
- DILATATION : une transformation en tout ou rien basée sur l'intersection en utilisant la notion d'élément structurant qui généralise le concept de fenêtre glissante. L'idée est que, pour pouvoir étudier la forme des objets, il faut pouvoir considérer des fenêtres glissantes avec des formes plus complexes. Dans la pratique, un élément structurant sera simplement une image binaire qui contient la forme d'intérêt.
- EROSION : l'inverse de la dilatation et parmi ses propriétés qualitative :
 - La taille des objets décroît,
 - Un objet avec des concavités ou des trous peut être divisés en plusieurs,
 - Les petits objets et les détails disparaissent.

Les figures 3.14, 3.15 et 3.16 illustrent les résultats de segmentation générés après post-traitement. Nous notons une nette amélioration confirmée par les résultats de performances après post-traitement dans le tableau 3.1.

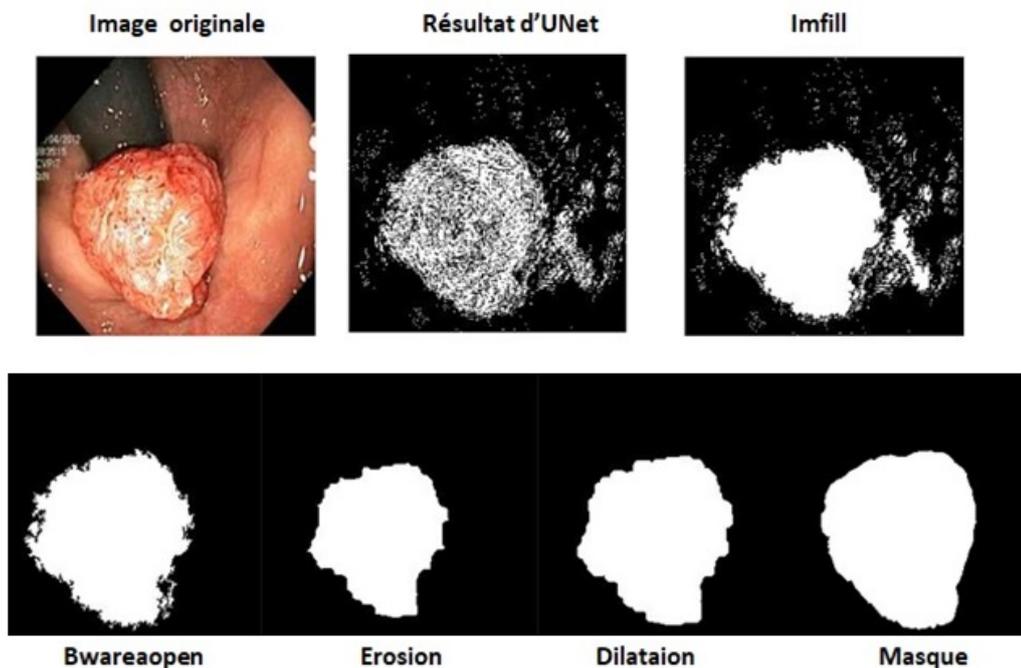


FIGURE 3.14 – Image test segmentée avant et après post-traitement d'U-NET.

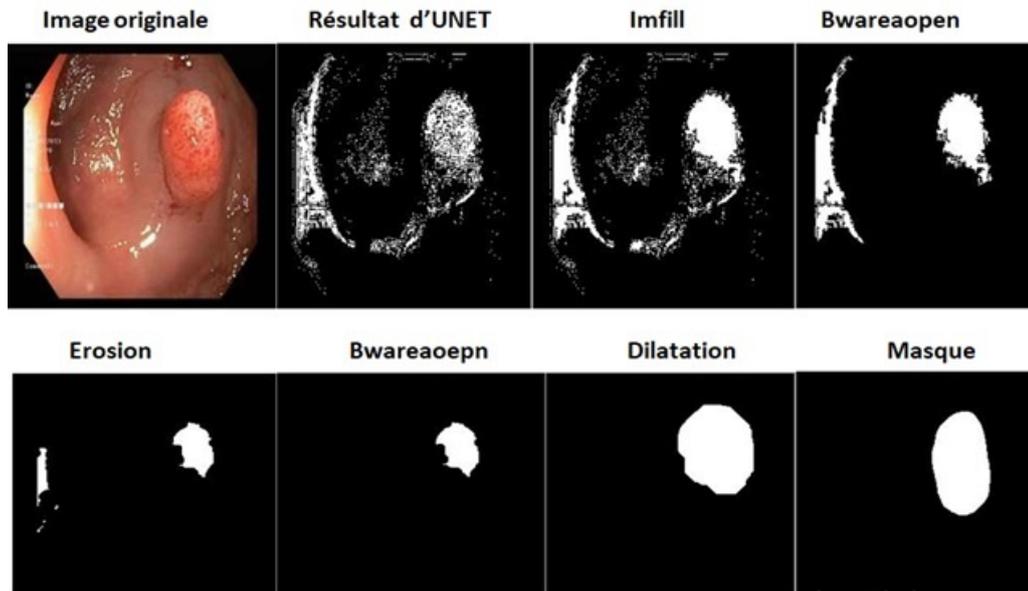


FIGURE 3.15 – Image test segmentée avant et après post-traitement d’U-NET.

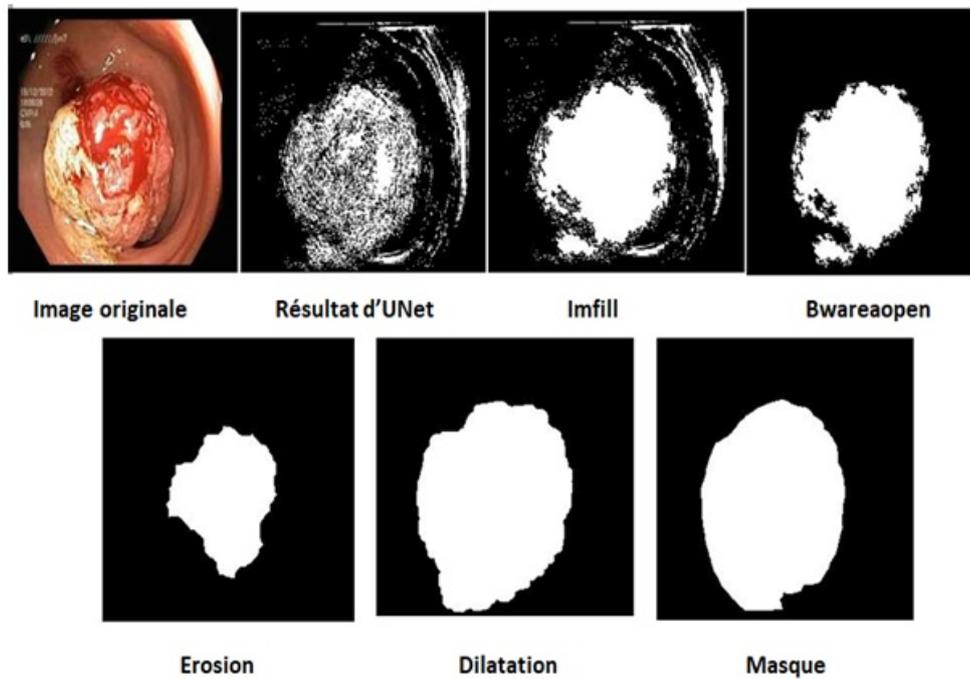


FIGURE 3.16 – Image test segmentée avant et après post-traitement d’U-NET.

Modèle	Précision	IoU	Fscore
U-Net	0.84374	0.53064	0.2951
U-Net + post-traitement	0.85064	0.7735	0.7702

TABLE 3.1 – Résultats de performances de l’architecture U-Net.

Modèle SegNet

La figure 3.17, représente deux exemples de segmentation d'images de test avec SegNet suivant le modèle de l'expérimentation 2 avec 20 epochs. Le tableau 3.2, témoigne des résultats de SegNet avec les différentes métriques moyennées sur la base de test de 50 images. Nous pouvons remarquer que les résultats de performances de SegNet sont relativement bas par rapport à U-Net avec et sans post-traitement.

Modèle	Précision	IoU	Fscore
SegNet	0.77909	0.53748	0.5953

TABLE 3.2 – Résultats de performances de l'architecture SegNet.

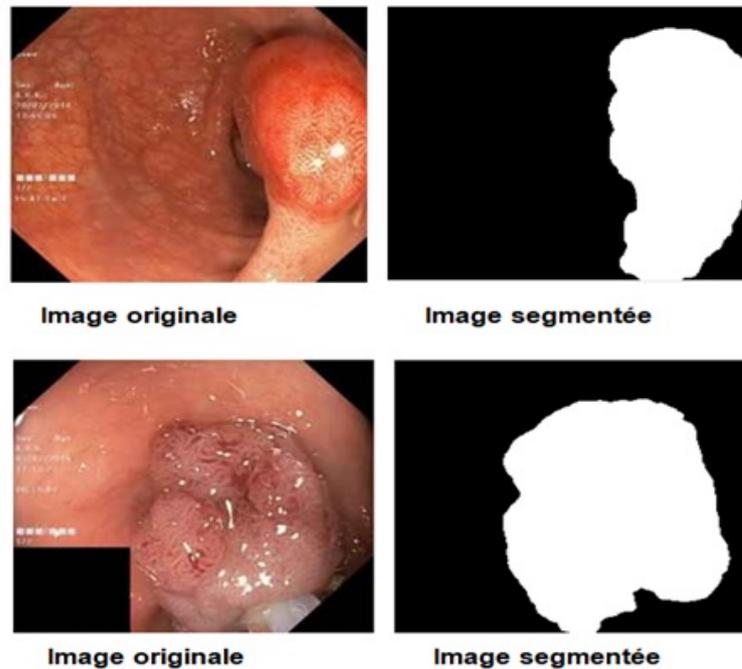


FIGURE 3.17 – Exemples de segmentation avec SegNet.

La figure 3.18 nous indique la matrice de confusion dans laquelle pour 100 images de polypes, le modèle arrive à prédire une bonne réponse pour 83 images, Et pour 100 images de label background, le modèle arrive à prédire 77 images, ceci révèle une bonne reconnaissance de notre modèle mais une moins bonne performance en résultats de segmentation.

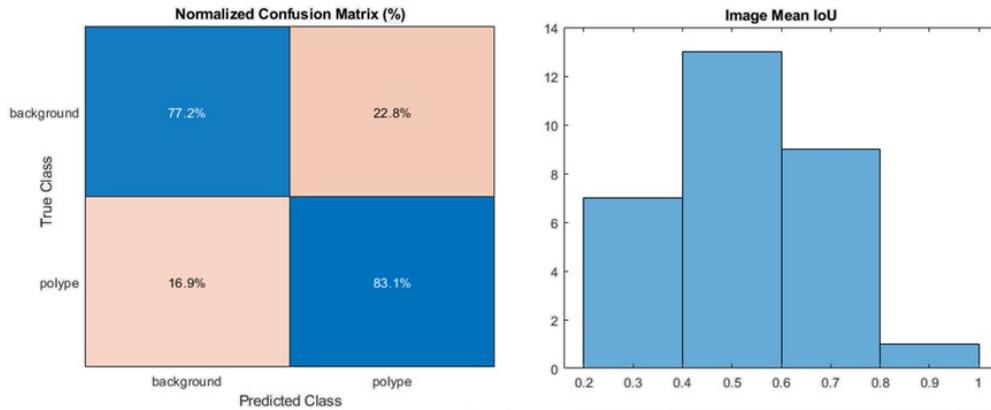


FIGURE 3.18 – Les métriques IoU et la matrice de confusion de test SegNet.

7 Synthèse comparative

Du tableau 3.3, nous pouvons conclure que l’approche hybride d’U-Net avec le post-traitement donne de bons résultats quantitatif et qualitatif. De son côté, l’architecture SegNet assure un résultat qualitatif très satisfaisant n’est pas suivi par un bon résultat quantitatif de plus dans un temps de calcul important.

Modèle	Epochs	Temps	Précision	IoU	Fscore
U-NET	10	5h	0.84374	0.53064	0.2951
U-NET+POST-TRAITEMENT	10	6h	0.85064	0.7735	0.7702
SEGNET	20	6h	0.77909	0.53748	0.5953

TABLE 3.3 – Synthèse des résultats de performances.

8 Interface de segmentation

Nous allons maintenant présenter les résultats obtenus grâce à une interface réalisée de notre application de segmentation des polypes colorectaux des images de coloscopie, le modèle d’interface crée par App Designer sur MATLAB nous permet de choisir un modèle parmi les trois modèles que nous avoir constuit : U-Net, U-Net avec post traitement et SegNet, (voir la figure 3.19).

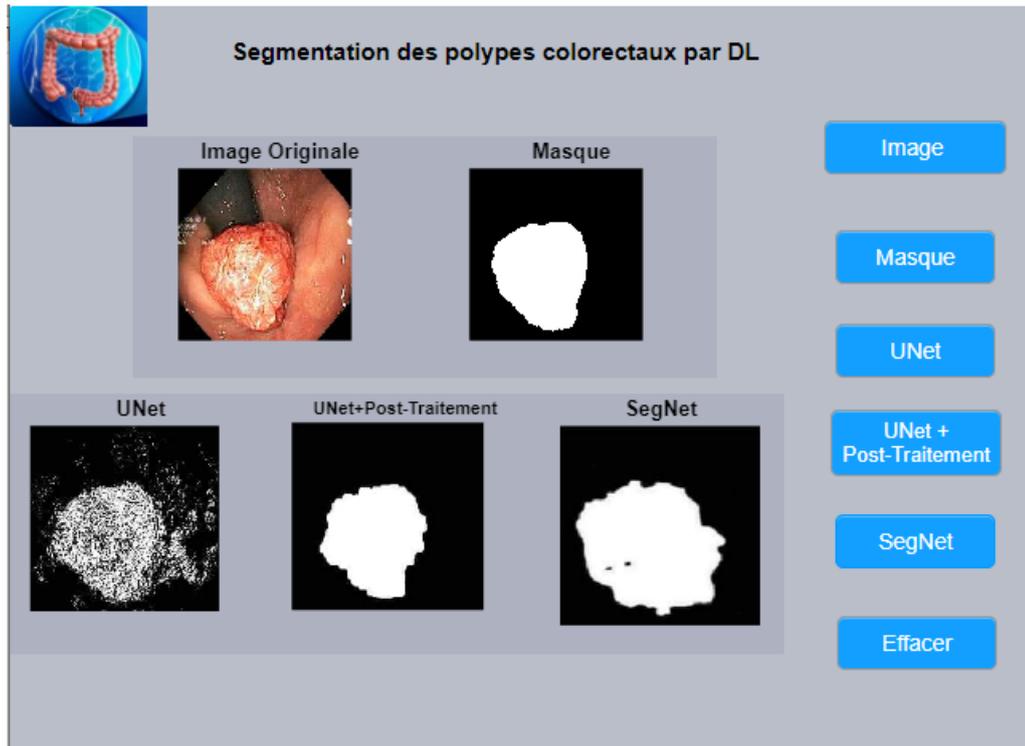


FIGURE 3.19 – Capture de l'interface

9 Conclusion

Dans ce chapitre, nous avons appliqué deux modèles de réseaux entièrement convolutifs U-Net et SegNet en ajoutant une phase de post-traitement pour améliorer la segmentation des polypes dans les images de coloscopie. Nous avons suivi les étapes de construction de ces modèles et fixé les hyper-paramètres appropriés aux images de notre base. Nous avons présenté les expérimentations, les tests et les résultats obtenus. En général après plusieurs heures d'entraînement les résultats ont été très satisfaisants, cependant, nous avons noté que les performances du modèle d'U-Net ont surpassées celles de SegNet les résultats qualitatifs de ce dernier ont eu recours à un post-traitement pour améliorer la précision de segmentation des polypes colorectaux.

Conclusion générale

Ce projet de fin d'études traite la segmentation des polypes colorectaux et les méthodes d'apprentissage profond DL utilisés pour la segmentation sémantique en particulier (U-Net et SegNet).

Notre objectif principal était de créer des réseaux (modèles) qui fonctionneraient au mieux avec un problème de segmentation d'image dans le domaine d'imagerie coloscopique. Des le départ de ce mémoire, nous avons arboré les structures Ad-Hoc, hybride et DL les plus utilisées pour la segmentation des polypes ainsi que les avantages et inconvénients de chacune d'entre elles.

Par la suite, nous avons passé en revue les modèles FCN : U-Net et SegNet utilisés pour le processus d'apprentissage profond.

En finalité, nous avons exposé les expérimentations, les résultats de tests et l'évaluation de comparaison entre les deux architectures. Les résultats ont prouvé que nos modèles étaient très efficaces et adaptables en termes de métriques de performance dans les problèmes de segmentation d'images, mais le modèle U-Net a surpassé le modèle SegNet, nous avons ajouté une phase de post-traitement au modèle U-Net ce qui a amélioré la qualité de segmentation de nos résultats.

Sans aucun doute, l'apprentissage profond DL a considérablement aidé la segmentation d'images, mais il reste encore de nombreuses difficultés à surmonter qui peut être résumé en manque de temps, l'absence d'expérience, la complexité des images avec un nombres qui répond à notre besoin, et les performances de matériels.

Dans des travaux futurs, nous pourrions utiliser des ensembles de données plus variés avec la stadification des types, en appliquant d'autres techniques pour l'amélioration les performances et le temps de réponses avec une approche de sur-segmentation (post-traitement) comme superpixels SLIC [35] afin d'offrir une détection en temps réels. Une autre piste intéressante à développer étant la transparence des résultats avec une technique comme LIME (Local Interpretable Model-Agnostic Explanations) [36] pour l'explicabilité des modèles DL dit boites noires, nous pourrions aussi appliqué le Transfer Learning en utilisant différents Backbones.

Bibliographie

- [1] Jorge Bernal, Javier Sánchez, and Fernando Vilarino, "A region segmentation method for colonoscopy images using a model of polyp appearance," in *Iberian Conference on Pattern Recognition and Image Analysis*. Springer, 2011, pp. 134–142.
- [2] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño, "Wm-dova maps for accurate polyp highlighting in colonoscopy : Validation vs. saliency maps from physicians," *Computerized Medical Imaging and Graphics*, vol. 43, pp. 99–111, 2015.
- [3] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 630–644, 2015.
- [4] Ning Du, Xiaofei Wang, Jianhua Guo, and Meidong Xu, "Attraction propagation : A user-friendly interactive approach for polyp segmentation in colonoscopy images," *Plos one*, vol. 11, no. 5, pp. e0155371, 2016.
- [5] Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng Ann Heng, "Integrating online and offline three-dimensional deep learning for automated polyp detection in colonoscopy videos," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 65–75, 2016.
- [6] Chin Yii Eu, Tong Boon Tang, Cheng-Hung Lin, Lok Hua Lee, and Cheng-Kai Lu, "Automatic polyp segmentation in colonoscopy images using a modified deep convolutional encoder-decoder architecture," *Sensors*, vol. 21, no. 16, pp. 5630, 2021.
- [7] Marcus VL Branch and Adriele S Carvalho, "Polyp segmentation in colonoscopy images using u-net-mobilenetv2," *arXiv preprint arXiv :2103.15715*, 2021.
- [8] Nikhil Kumar Tomar, "Automatic polyp segmentation using fully convolutional neural network," *arXiv preprint arXiv :2101.04001*, 2021.
- [9] Patrick Brandao, Odysseas Zisimopoulos, Evangelos Mazomenos, Gastone Ciuti, Jorge Bernal, Marco Visentini-Scarzanella, Arianna Menciassi, Paolo Dario, Anastasios Koulaouzidis, Alberto Arezzo, et al., "Towards a computed-aided diagnosis system in colonoscopy : automatic polyp segmentation using convolution neural networks," *Journal of Medical Robotics Research*, vol. 3, no. 02, pp. 1840002, 2018.
- [10] Jie Meng, Linyan Xue, Ying Chang, Jianguang Zhang, Shilong Chang, Kun Liu, Shuang Liu, Bangmao Wang, and Kun Yang, "Automatic detection and

- segmentation of adenomatous colorectal polyps during colonoscopy using mask r-cnn," *Open Life Sciences*, vol. 15, no. 1, pp. 588–596, 2020.
- [11] Sungheon Park, Myunggi Lee, and Nojun Kwak, "Polyp detection in colonoscopy videos using deeply-learned hierarchical features," *Seoul National University*, 2015.
- [12] Yuji Iwahori, Takayuki Shinohara, Akira Hattori, Robert J Woodham, Shinji Fukui, Manas Kamal Bhuyan, and Kunio Kasugai, "Automatic polyp detection in endoscope images using a hessian filter.," in *MVA*, 2013, pp. 21–24.
- [13] Seung-Hwan Bae and Kuk-Jin Yoon, "Polyp detection via imbalanced learning and discriminative feature learning," *IEEE transactions on medical imaging*, vol. 34, no. 11, pp. 2379–2393, 2015.
- [14] Ruikai Zhang, Yali Zheng, Tony Wing Chung Mak, Ruoxi Yu, Sunny H Wong, James YW Lau, and Carmen CY Poon, "Automatic detection and classification of colorectal polyps by transferring low-level cnn features from nonmedical domain," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 41–47, 2016.
- [15] Jorge Bernal, Nima Tajkbaksh, Francisco Javier Sanchez, Bogdan J Matuszewski, Hao Chen, Lequan Yu, Quentin Angermann, Olivier Romain, Bjørn Rustad, Ilangko Balasingham, et al., "Comparative validation of polyp detection methods in video colonoscopy : results from the miccai 2015 endoscopic vision challenge," *IEEE transactions on medical imaging*, vol. 36, no. 6, pp. 1231–1249, 2017.
- [16] Mojtaba Akbari, Majid Mohrekesh, Ebrahim Nasr-Esfahani, SM Reza Soroushmehr, Nader Karimi, Shadrokh Samavi, and Kayvan Najarian, "Polyp segmentation in colonoscopy images using fully convolutional network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 69–72.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net : Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [18] Josh Patterson and Adam Gibson, *Deep learning : A practitioner's approach*, " O'Reilly Media, Inc.", 2017.
- [19] A. Géron, *Deep Learning avec Keras et TensorFlow : Mise en oeuvre et cas concrets*, Hors Collection. Dunod, 2020.
- [20] Gang Li, Biao Ma, Shuanhai He, Xueli Ren, and Qiangwei Liu, "Automatic tunnel crack detection based on u-net and a convolutional neural network with alternately updated clique," *Sensors*, vol. 20, no. 3, pp. 717, 2020.
- [21] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla, "Segnet : A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling," *CoRR*, vol. abs/1505.07293, 2015.
- [22] Institut National de Santublique, "Cancer colorectal : pre 6500 nouveaux cas enregistrnnuellement en alge," Minist de la Sante la Population et de la Rrme Hospitali, Mars 29, 2021, Accessed Juin 22,

- 2022., <https://www.insp.dz/images/PDF/ENQUETES/ENQUET-NLE-CANCER-RESULTATS-NATIONAUX.pdf>.
- [23] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang, "A classification-enhanced vote accumulation scheme for detecting colonic polyps," in *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*. Springer, 2013, pp. 53–62.
- [24] Nima Tajbakhsh, Changching Chi, Suryakanth R Gurudu, and Jianming Liang, "Automatic polyp detection from learned boundaries," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014, pp. 97–100.
- [25] Willem Dijkstra, André Sobiecki, Jorge Bernal, and Alexandru C Telea, "Towards a single solution for polyp detection, localization and segmentation in colonoscopy images.," in *VISIGRAPP (4 : VISAPP)*, 2019, pp. 616–625.
- [26] YunBo Guo, *Polyp segmentation in colonoscopy images with convolutional neural networks*, Ph.D. thesis, University of Central Lancashire, 2019.
- [27] Debapriya Banik, Debotosh Bhattacharjee, and Mita Nasipuri, "A multi-scale patch-based deep learning system for polyp segmentation," in *Advanced Computing and Systems for Security*, pp. 109–119. Springer, 2020.
- [28] Ivan Vasilev, *Advanced Deep Learning with Python : Design and implement advanced next-generation AI solutions using TensorFlow and PyTorch*, Packt Publishing Ltd, 2019.
- [29] Mohammed Hakim Bendiabdallah and Nesma Settouti, "A comparison of u-net backbone architectures for the automatic white blood cells segmentation," *WAS Science Nature (WASSN) ISSN : 2766-7715*, vol. 4, no. 1, 2021.
- [30] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv :1409.1556*, 2014.
- [31] Hanna Borgli, Vajira Thambawita, Pia H Smedsrud, Steven Hicks, Debesh Jha, Sigrun L Eskeland, Kristin Ranheim Randel, Konstantin Pogorelov, Matthias Lux, Duc Tien Dang Nguyen, et al., "Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific data*, vol. 7, no. 1, pp. 1–14, 2020.
- [32] Debesh Jha, Sharib Ali, Steven Hicks, Vajira Thambawita, Hanna Borgli, Pia H Smedsrud, Thomas de Lange, Konstantin Pogorelov, Xiaowei Wang, Philipp Harzig, et al., "A comprehensive analysis of classification methods in gastrointestinal endoscopy imaging," *Medical image analysis*, vol. 70, pp. 102007, 2021.
- [33] Hossein Talebi and Peyman Milanfar, "Learning to resize images for computer vision tasks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 497–506.
- [34] Olivier Rukundo, "Effects of image size on deep learning," *arXiv preprint arXiv :2101.11508*, 2021.
- [35] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk, "Slic superpixels compared to state-of-the-art

superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

- [36] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, "“why should i trust you ?” : Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2016, KDD '16, p. 11351144, Association for Computing Machinery.