



HAL
open science

Acquisition et analyse de lames de cytologie digitales multi-plans pour l'aide à la détection précoce du cancer de la vessie

Alexandre Bouyssoux

► **To cite this version:**

Alexandre Bouyssoux. Acquisition et analyse de lames de cytologie digitales multi-plans pour l'aide à la détection précoce du cancer de la vessie. Intelligence artificielle [cs.AI]. Université Paris-Saclay, 2022. Français. NNT : 2022UPAST182 . tel-04001112

HAL Id: tel-04001112

<https://theses.hal.science/tel-04001112>

Submitted on 22 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Acquisition et analyse de lames de
cytologie digitales multi-plans pour l'aide
à la détection précoce du cancer de la
vessie

*Acquisition and analysis of 3D digital cytology slides for
early bladder cancer diagnosis*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 575 : electrical, optical, bio : physics and engineering
(EOBE)

Spécialité de doctorat : Imagerie et physique médicale
Graduate School : Sciences de l'ingénierie et des systèmes
Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche Dynamique cellulaire physiologique
et pathologique (CNRS, Institut Pasteur), sous la direction de Jean-Christophe
OLIVO-MARIN, Professeur des universités et co-encadrée par Riadh FEZZANI,
Docteur en sciences.

Thèse soutenue à Paris-Saclay, le 16 décembre 2022, par

Alexandre BOUYSSOUX

Composition du jury

Laurent Mugnier Directeur de recherche à l'ONERA, Université Paris-Saclay	Président du Jury
Auguste Genovesio Directeur de recherche à l'IBENS, INSERM	Rapporteur et Examineur
Dominique Béréziat Maître de conférence à LIP6, Sorbonne Université	Rapporteur et Examineur
Riadh Fezzani Docteur, VitaDX International	Examineur
Jean-Christophe Olivo-Marin Directeur de recherche à l'Institut Pasteur	Directeur de thèse

Titre : Acquisition et analyse de lames de cytologie digitales multi-plans pour l'aide à la détection précoce du cancer de la vessie

Mots clés : Cytologie urinaire - Cancer de la vessie - Lames digitales - Analyse d'images - Apprentissage machine

Résumé : L'analyse automatique de lames de cytologie digitales a le potentiel d'améliorer la sensibilité de l'examen de cytologie urinaire pour la détection du cancer de la vessie. Parce que la répartition du matériel cellulaire dans les lames de cytologie urinaire est tri-dimensionnelle, il est nécessaire de numériser les lames selon plusieurs plans focaux pour imager nettement l'ensemble des cellules qu'elles contiennent. Se pose alors la question de l'analyse efficace, à l'aide d'algorithmes de traitement et d'analyse d'images, des lames de cytologie digitales multi-plans pour le diagnostic du cancer de la vessie. Les travaux menés au cours de cette thèse portent sur la mise au point de méthodes algorithmiques rapides et robustes, permettant l'utilisation d'information tri-

dimensionnelle pour l'analyse d'objets cellulaires provenant de lames de cytologie digitales. En particulier, des méthodes de traitement et d'analyse d'images sont développées pour la segmentation de régions cellulaires, la détection d'objets cellulaires anormaux ainsi que pour le diagnostic automatique de carcinomes urothéliaux. Cette thèse évalue également l'impact de l'utilisation d'acquisitions multi-plans pour l'analyse d'objets cellulaires d'intérêt et pour l'analyse de lames digitales. Pour cela, les performances de diagnostic de carcinomes urothéliaux obtenues à l'aide d'acquisitions multi-plans sont comparées aux performances obtenues à l'aide d'acquisitions bi-dimensionnelles des mêmes lames de cytologie urinaires.

Title : Acquisition and analysis of 3D digital cytology slides for early bladder cancer diagnosis

Keywords : Urine cytology - Bladder cancer - Digital slides - Image analysis - Machine learning

Abstract : Automatic analysis of digital cytology slides has the potential to improve the sensitivity of urine cytology examination for bladder cancer diagnosis. Because the distribution of cellular material in urine cytology slides is three-dimensional, it is necessary to scan the slides with multiple focal planes to clearly image all the cells they contain. The question then arises of the efficient analysis, using image analysis processing algorithms, of 3D digital cytology slides for bladder cancer diagnosis. The work carried out during this thesis focuses on the development of fast and robust algorithms,

allowing the use of 3D information for cells analysis from digital cytology slides. In particular, image processing and analysis methods are developed for cell regions segmentation, abnormal objects detection and urothelial carcinomas automatic diagnosis. This thesis also evaluates the impact of multi-plane acquisitions for the analysis of cells of interest and digital slides analysis. For this purpose, the diagnostic performances of urothelial carcinomas based on 3D acquisitions are compared to the performances based on simple 2D acquisitions of the same urinary cytology slides.

université
PARIS-SACLAY

ÉCOLE DOCTORALE

Physique et ingénierie:
Electrons, Photons,
Sciences du vivant (EOBE)

anRT

ASSOCIATION NATIONALE
RECHERCHE TECHNOLOGIE



**INSTITUT
PASTEUR**

U○ VitaDX

Remerciements

Je souhaite remercier toutes les personnes qui ont su m'apporter leur aide, leurs conseils, leurs critiques, leurs avis et surtout leur soutien pendant ces années de thèse.

En premier lieu mon directeur de thèse, Professeur Jean-Christophe Olivo-Marin, pour m'avoir accueilli au sein de l'unité Analyse d'Images Biologiques à l'Institut Pasteur, pour avoir assuré l'encadrement de ce projet, pour tes nombreux conseils et ton soutien.

Je tiens également à remercier l'ensemble des membres du jury, en particulier Dr. Dominique Béréziat, maître de conférence à LIP6, Sorbonne Université, et Dr. Auguste Genovesio, directeur de recherche à l'IBENS, INSERM, de l'honneur que vous me faites en acceptant de rapporter cette thèse. Je souhaite également chaleureusement remercier Professeur Laurent Mugnier, directeur de recherche à l'ONERA, d'avoir accepté d'examiner ma thèse et de présider le jury.

Je remercie Thibaut Troude, CTO à VitaDX, et Allan Rodriguez, CEO à VitaDX, pour votre confiance, votre soutien et votre accueil. Je souhaite également remercier l'ensemble des équipes de VitaDX, notamment Laetitia Lallement et Kathleen Jarnouen, pour votre expertise lors des travaux que nous avons menés ensemble, ainsi que l'équipe Recherche et Développement de m'avoir aussi bien intégré dans l'équipe et aidé dans mes recherches. Je souhaite tout particulièrement remercier ici Dr. Riadh Fezzani, d'avoir accepté de co-encadrer cette thèse et pour tous les conseils que tu as pu me donner pendant ces trois dernières années.

Je remercie l'ensemble des membres de l'unité Analyse d'Images Biologiques de l'Institut Pasteur, pour l'ensemble des discussions enrichissantes et pour tous les bons moments que nous avons pu passer ensemble au cours de ces trois années, en particulier Aleix, Daniel, Diana, Robin, Suvadip, Ritu, et plus récemment Samuel, Yekta, Erwan, Antoine, Benjamin, Mounib et Raphaël.

Je remercie également à tous mes proches, amis et famille, pour votre soutien inconditionnel. Un immense merci à Agathe, pour ton soutien quotidien et tes multiples relectures de ce document. Finalement, merci à tous ceux qui ont pu faire le déplacement pour assister à la soutenance de cette thèse.

Merci à vous tous.

Table des matières

Remerciements	i
Table des matières	ii
1 Introduction	1
1.1 Introduction générale	1
1.1.1 Le cancer de la vessie dans le monde	1
1.1.2 Types de cancers de la vessie	1
1.1.3 Sous-types de cancers de la vessie	2
1.1.4 Grades du cancer de la vessie	2
1.1.5 Diagnostic du cancer de la vessie	2
1.2 Pathologie digitale	3
1.2.1 Introduction à la pathologie digitale	3
1.2.2 Applications	3
1.2.3 Contraintes	4
1.3 État de l'art: analyse automatique d'images en pathologie digitale	5
1.3.1 Analyse d'images biologiques	5
1.3.2 Introduction à l'analyse de lames de pathologie digitales	6
1.3.3 Applications à l'histopathologie	7
1.3.4 Applications à la cytopathologie	10
1.4 Acquisition multi-plans en cytopathologie digitale	13
1.4.1 Besoin de l'acquisition multi-plans	13
1.4.2 Fonctionnement sur les scanners de lames	14
1.4.3 Optimisation de l'acquisition de lames digitales multi-plans	15
1.5 Objectifs des travaux de recherche	16
2 Représentation 2D de l'information 3D	17
2.1 Section centrale	17
2.2 Section de meilleur focus	17
2.3 Projection par extension de profondeur de champ	18
2.3.1 Description de la méthode d'extension de profondeur de champ multi-canal	19
2.3.2 L'extension de profondeur de champ pour la cytologie digitale	27
2.3.3 Extension de profondeur de champ pour le diagnostic automatique	29
2.4 Conclusion	31
3 Préparation et acquisition de lames digitales multi-plans	33
3.1 Paramètres considérés	33
3.1.1 Paramètres de préparation de lames physiques	33
3.1.2 Paramètres de numérisation	34
3.2 Optimisation du protocole de coloration	34
3.2.1 Objectifs	35
3.2.2 État de l'art	35
3.2.3 Méthode	37
3.2.4 Distribution des métriques par lame	42
3.2.5 Analyse de l'impact de la solution d'hématoxyline	43
3.2.6 Analyse de l'impact du temps d'immersion dans la solution d'hématoxyline	45
3.2.7 Conclusion	46

3.3	Recherche du nombre de plans focaux optimal	48
3.3.1	Choix de l'espacement inter-plans	48
3.3.2	Recherche du nombre de plans focaux optimal connaissant l'espacement inter-plans	49
3.4	Conclusion	50
4	Segmentation à partir de z-stacks	51
4.1	Segmentation sémantique à partir de profils d'intensité selon l'axe z	52
4.1.1	État de l'art	52
4.1.2	Segmentation à partir de profils-z	53
4.1.3	Métriques d'évaluation	54
4.1.4	Jeux de données	55
4.1.5	Hard mining	56
4.1.6	Méthodes	57
4.1.7	Résultats	59
4.1.8	Discussion	60
4.2	Segmentation sémantique de noyaux à partir de z-stacks	61
4.2.1	État de l'art	61
4.2.2	Méthodes de segmentation sémantique	61
4.2.3	Métriques d'évaluation	64
4.2.4	Résultats	65
4.2.5	Évaluation du temps d'inférence	67
4.2.6	Étude d'ablation et choix de design	69
4.2.7	Discussion: application au diagnostic de cancer du col	70
4.3	Segmentation d'instances cellulaires à partir de z-stacks	71
4.3.1	État de l'art	71
4.3.2	Jeu de données	72
4.3.3	Architecture modifiée pour traiter des z-stacks	73
4.3.4	Métriques d'évaluation	73
4.3.5	Méthodes de segmentation d'instances	74
4.3.6	Résultats	78
4.3.7	Évaluation du temps d'inférence	78
4.3.8	Discussion	79
4.4	Conclusion	81
5	Diagnostic automatique du cancer de la vessie	82
5.1	Introduction	82
5.1.1	La classification du Système de Paris	83
5.2	État de l'art	86
5.2.1	Méthodes avec annotations d'atypies locales	86
5.2.2	Méthodes sans annotations d'atypies locales	86
5.3	Méthode	87
5.3.1	Acquisition des lames de cytologie digitales	88
5.3.2	Identification des objets d'intérêt	88
5.3.3	Calcul des caractéristiques expertes	89
5.3.4	Paradigme de l'apprentissage à partir d'instances multiples	92
5.3.5	Classification de lames par apprentissage à partir d'instances multiples	93
5.4	Résultats	99
5.4.1	Métriques d'évaluation	99
5.4.2	Évaluation par validation croisée	100
5.4.3	Stratégie ensembliste pour l'inférence sur l'ensemble de données de test	101
5.4.4	Performances de diagnostic	101

5.4.5	Comparaison avec les performances des anatomopathologistes	103
5.4.6	Visualisation des prédictions	104
5.4.7	Performances en fonction des examens de cystoscopie et d'histologie	106
5.5	Discussion	108
5.5.1	Impact de l'apprentissage ensembliste sur les performances de classification	108
5.5.2	Nombre minimum de cellules urothéliales basales détectées pour analyser une lame	109
5.5.3	Impact du nombre d'instances sélectionnées pour représenter une lame	110
5.5.4	Choix de design pour les méthodes d'apprentissage à partir d'instances multiples .	111
5.6	Conclusion	114
6	Impact des z-stacks sur le diagnostic automatique du cancer de la vessie	116
6.1	Introduction	116
6.2	État de l'art	116
6.3	Méthode	117
6.3.1	Acquisition des lames de cytologie digitales	117
6.3.2	Détection des objets d'intérêt	118
6.3.3	Calcul des caractéristiques expertes	118
6.3.4	Classification de lames par représentation	120
6.4	Résultats	120
6.4.1	Évaluation statistique	120
6.4.2	Performances de diagnostic par représentation	120
6.5	Discussion	121
6.5.1	Répétabilité des expériences	121
6.5.2	Sensibilité à la segmentation des noyaux	122
6.5.3	Apprentissage ensembliste à partir de différentes représentations 2D des z-stacks . .	124
6.6	Conclusion	125
7	Conclusion générale et perspectives	126
7.1	Contributions et principaux résultats	126
7.1.1	Représentations 2D de l'information 3D	126
7.1.2	Préparation et acquisition multi-plans de lames de cytologie digitales	127
7.1.3	Segmentation à partir de z-stacks	128
7.1.4	Diagnostic automatique de carcinomes urothéliaux	129
7.1.5	Impact des acquisitions multi-plans sur le diagnostic automatique	129
7.2	Limitations des études menées	130
7.3	Perspectives	130
7.3.1	Caractéristiques profondes pour le diagnostic	130
7.3.2	Ajout de l'analyse d'amas cellulaires	132
7.3.3	Filtrage des cellules urothéliales hors-focus dans les acquisitions 2D	133
7.4	Conclusion	133
	Bibliographie	135

1.1 Introduction générale

1.1.1 Le cancer de la vessie dans le monde

Avec 10 626 cas chez l'homme et 2 448 cas chez la femme estimés en France en 2018 ¹, le cancer de la vessie se situe au 9^{ème} rang des tumeurs solides les plus fréquentes en France. Dans le monde, le cancer de la vessie ² représente environ 573 000 nouveaux diagnostics en 2020 (10^{ème} rang), et est responsable d'environ 213 000 décès [1]. L'Agence Internationale de Recherche sur le Cancer (IARC) prévoit une augmentation du nombre de cancers de la vessie diagnostiqués dans le monde jusqu'à 913 000 en 2040 ³.

90% des cancers de la vessie sont des carcinomes à cellules transitionnelles [2] provenant de l'épithélium transitionnel, qui tapisse la paroi interne de la vessie. Ainsi par la suite, le terme général *cancer de la vessie* fait référence aux carcinomes à cellules transitionnelles de la paroi de la vessie. Ces carcinomes sont également appelés carcinomes urothéliaux.

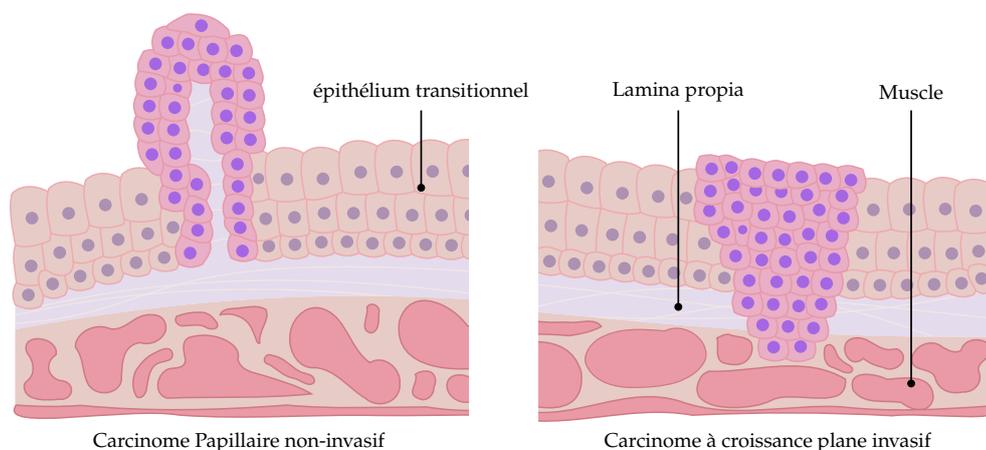
1.1.2 Types de cancers de la vessie

En fonction des couches de cellules atteintes, on distingue deux types de carcinomes urothéliaux : les carcinomes *invasifs* et les carcinomes *non-invasifs*.

Un **carcinome non-invasif** est une tumeur dont la croissance est limitée aux couches de l'épithélium transitionnel, ou urothélium.

Un **carcinome invasif** est une tumeur qui s'est développée dans les couches plus profondes de la vessie, notamment jusqu'à la couche musculaire.

Des illustrations de carcinomes à cellules transitionnelles invasifs et non-invasifs sont données Figure 1.1.



1: Groupe d'Experts de l'Institut National du Cancer, Rapport INCa 2019.

2: Bladder GLOBOCAN 2020 (accès 13/01/2022) : gco.iarc.fr/today/data/factsheets/cancers/30-Bladder-fact-sheet.pdf

[1]: SUNG et al. (2021), « Global Cancer Statistics 2020 »

3: Source (accès 13/01/2022) : gco.iarc.fr/tomorrow/en/dataviz/tables?cancers=30

[2]: ANDREASSEN et al. (2016), « Incidence and Survival of urothelial carcinoma of the urinary bladder in Norway 1981-2014 »

1.1.3 Sous-types de cancers de la vessie

On distingue également deux sous-types de cancers en fonction de la structure et du développement de la tumeur : les *carcinomes papillaires* et les *carcinomes à croissance plane*.

Un **carcinome papillaire** se développe sous forme d'excroissances ⁴ de la surface interne vers la lumière de la vessie.

4: Également appelées polypes.

Un **carcinome à croissance plane** possède une structure plane qui ne dépasse pas ou peu de la couche superficielle de l'épithélium. Un carcinome à croissance plane et non-invasif est aussi appelé carcinome in-situ.

Des illustrations de carcinomes papillaires et de carcinomes à croissance plane sont proposées Figure 1.1.

1.1.4 Grades du cancer de la vessie

Les carcinomes urothéliaux peuvent aussi être décrits par leur grade, défini en fonction du niveau de différenciation des cellules atypiques.

Carcinome urothélial de bas-grade (LGUC) Les carcinomes urothéliaux de bas-grade sont composés de cellules urothéliales atypiques bien différenciées, c'est-à-dire qui ressemblent aux cellules urothéliales saines.

Carcinome urothélial de haut-grade (HGUC) Les carcinomes urothéliaux de haut-grade, au contraire, sont composés de cellules urothéliales atypiques peu différenciées, qui n'ont ni l'aspect ni l'organisation des cellules urothéliales saines.

Les carcinomes urothéliaux de haut-grade sont plus à risque de se propager et de former des métastases que les carcinomes urothéliaux de bas-grade ⁵.

5: American Society of Clinical Oncology, Bladder Cancer. (accès 04/03/2022) : <https://www.cancer.net/cancer-types/bladder-cancer>

1.1.5 Diagnostic du cancer de la vessie

En présence de symptômes d'un cancer de la vessie chez le patient tels que la présence de sang visible dans les urines, des envies fréquentes d'uriner ou des brûlures urinaires en l'absence d'infection, plusieurs examens décrits ci-dessous peuvent permettre d'établir le diagnostic.

Examen cytobactériologique des urines (ECBU)

Le recueil d'un échantillon d'urine auprès du patient permet de réaliser deux examens :

Bactériologie Des bactéries sont recherchées dans l'urine après sa mise en culture afin d'identifier notamment des cas d'infections urinaires responsables des symptômes qui ont amené à la consultation.

Cytologie urinaire Les cellules présentes dans l'échantillon d'urine sont déposées et fixées sur une lame de microscope. Cette lame est ensuite colorée, par exemple à l'aide d'une coloration de Papanicolaou [3], pour ensuite être étudiée au microscope lors d'un examen cytopathologique. Les cellules de la paroi interne de la vessie desquament naturellement, et ces cellules se retrouvent évacuées dans les urines. Le praticien recherche alors si l'échantillon prélevé contient des cellules atypiques, pouvant indiquer la présence d'une tumeur.

Cystoscopie

Un tube fin et souple, équipé d'une caméra, est introduit à l'intérieur de la vessie par l'urètre afin d'examiner visuellement la paroi interne de la vessie pour y détecter la présence d'éventuelles tumeurs. La cystoscopie est un examen intrusif et douloureux, pratiqué sous anesthésie.

Résection trans-urétrale de la vessie et examen histopathologique

En présence d'anomalies lors d'un examen de cystoscopie, le praticien effectue un prélèvement de tissu atypique. Ce tissu prélevé est ensuite fixé sur une lame de microscope et analysé lors d'un examen histopathologique. La biopsie suivie de l'examen histopathologique permet de confirmer le diagnostic et bien qu'il s'agisse de l'examen le plus intrusif, il représente également à ce jour l'examen le plus fiable. En cas de forte suspicion de cancer avant l'examen de cystoscopie grâce à l'examen de cytologie urinaire, la tumeur peut être retirée complètement à l'aide d'une résection trans-urétrale.

1.2 Pathologie digitale

1.2.1 Introduction à la pathologie digitale

La pathologie digitale est l'analyse des causes et conséquences de maladies à partir d'échantillons biologiques numérisés. Des lames physiques de microscopie, contenant des échantillons de tissus ou de cellules individuelles, sont numérisées à l'aide d'un scanner de lames (voir Figure 1.2) et peuvent par la suite être analysées sur ordinateur, partagées et archivées.

1.2.2 Applications

La numérisation de lames digitales permet de nouvelles applications par rapport aux lames de microscopie physique classique.

Télé-médecine De part leur relative facilité de partage, les lames de pathologie digitales peuvent facilement et rapidement être soumises à l'avis de plusieurs médecins, sans que ceux-ci aient besoin d'être dans la même pièce que la lame de microscopie physique. Les lames digitales sont ainsi utiles pour le diagnostic de pathologies rares nécessitant l'avis d'experts ainsi que dans la lutte contre les déserts médicaux.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »



FIGURE 1.2 – Exemple de scanner de lames disponible sur le marché : le Hamamatsu NanoZoomer®S360.

Éducation La pathologie digitale facilite également la formation en pathologie, à l'aide de ressources centralisées pouvant être visualisées simultanément par de nombreux étudiants. De plus, contrairement aux manuels traditionnels de pathologie, les étudiants ont accès aux lames entières, ce qui permet d'apprendre à reconnaître les zones d'importance pour le diagnostic à partir de l'échantillon complet [4].

[4]: HAMILTON et al. (2012), « Virtual microscopy and digital pathology in training and education »

Analyse par ordinateur Il est également possible d'appliquer des méthodes de traitement et d'analyse automatiques d'images aux lames de pathologie digitales afin d'améliorer ou d'accélérer le diagnostic. De telles méthodes peuvent également être appliquées à la recherche de nouveaux marqueurs biologiques (voir section 1.3).

1.2.3 Contraintes

Bien que prometteuse, l'adoption de la pathologie digitale pose certaines contraintes.

Temps d'acquisition Le temps d'acquisition des lames digitales varie grandement en fonction du scanner utilisé, du grossissement choisi, du nombre de plans focaux numérisés, et peut ainsi devenir conséquent. À titre d'exemple, les temps de numérisation d'une lame de microscopie sur une section de 15 × 15 mm au grossissement ×40, pour différents scanners de lames commercialisés par Hamamatsu sont reportés Tableau 1.1.

scanner de lames	temps de numérisation
NanoZoomer ® SQ	275 s
NanoZoomer ® S210	150 s
NanoZoomer ® S360	30 s

TABLE 1.1 – Temps de numérisation d'une lame de microscopie sur une section de 15 × 15 mm au grossissement ×40 pour différents scanners de lames commercialisés par Hamamatsu.

Volume des fichiers numériques Les lames digitales représentent des fichiers numériques larges et volumineux. Pour illustration, la numérisation d'une région de 10 × 10 mm au grossissement ×40 représente une image de l'ordre de deux giga-pixels. De même, une lame de cytologie digitale numérisée selon trois plans focaux au grossissement ×40 produit un fichier numérique de deux à trois gigabits (Gb). La manipulation de grandes quantités de fichiers numériques volumineux nécessite la mise en place d'infrastructures adaptées pour le stockage et le partage des lames digitales, et pose également des difficultés techniques pour l'analyse et le traitement automatique d'images. En effet, l'analyse d'images par ordinateur nécessite de charger en mémoire toute l'information contenue dans une lame digitale, ce qui pose des difficultés en présence de fichiers numériques volumineux et nécessite généralement une étape d'extraction de sous-images d'intérêt.

Formats de numérisation propriétaires La plupart des formats de numérisation utilisés par les scanners de lames commerciaux sont des formats propriétaires et différents pour chaque constructeur : *.ndpi* pour les scanners Hamamatsu ou encore *.mrxs* pour les scanners 3DHISTECH et Zeiss. Les constructeurs proposent des logiciels qui permettent de visualiser les lames digitales sauvegardées en format propriétaire, cependant le traitement de ces images par ordinateur nécessite le développement d'outils adaptés afin de pouvoir les manipuler, comme par exemple OpenSlide [5].

[5]: SATYANARAYANAN et al. (2013), « OpenSlide »

Analyse multi-échelle Lors de l'analyse d'une lame de pathologie, l'information d'intérêt peut être contenue dans les petits détails visibles aux forts grossissements ainsi qu'au niveau du contexte global accessible aux faibles grossissements. Le traitement par ordinateur de lames digitales nécessite alors le développement d'approches multi-échelles afin de considérer ces détails visibles à différents grossissements.

1.3 État de l'art : analyse automatique d'images en pathologie digitale

1.3.1 Analyse d'images biologiques

Avec l'acquisition d'images biologiques est rapidement apparu le besoin de méthodes d'analyses d'images automatiques [6]. Le traitement par ordinateur d'images biologiques permet par exemple d'effectuer des analyses quantitatives de la structure de tissus, de la morphologie de noyaux ou encore du mouvement de cellules [7].

Les méthodes d'analyse d'images biologiques communément utilisées adressent une multitude de problèmes :

Pré-traitement d'images Des algorithmes de pré-traitement ont été développés afin d'améliorer la qualité des images biologiques acquises en compensant les limitations des systèmes imageurs. Des méthodes de déconvolution sont par exemple fréquemment utilisées afin de corriger la contribution d'éléments hors-focus dans une image de microscopie. Des approches de déconvolution classiques reposent sur une méthode d'optimisation permettant d'approcher l'image nette la plus probable à partir de l'image originale et d'une modélisation ou d'une estimation du noyau de convolution du microscope [8]. Afin de limiter la dégradation des échantillons biologiques, certaines acquisitions nécessitent une faible illumination et des temps d'exposition faibles. Les images biologiques résultant de telles acquisitions sont alors généralement bruitées. Des méthodes de dé-bruitage ont alors été appliquées afin d'améliorer le rapport signal sur bruit d'images biologiques. Parmi les méthodes de dé-bruitage courantes, on retrouve par exemple la maximisation de la vraisemblance étant donné une image bruitée et un modèle statistique du bruit. Une image biologique peut également être dé-bruitée grâce à l'atténuation des plus petits coefficients, supposés correspondre au bruit, de sa décomposition par transformée en ondelettes [9].

Détection de particules Afin de pouvoir compter des objets biologiques tels que certaines protéines, l'immunofluorescence peut être utilisée afin de mettre en évidence les objets d'intérêt sous forme de points lumineux. Des méthodes ont été développées afin de détecter automatiquement ces points lumineux d'intérêt par rapport à l'arrière plan des images biologiques, en se reposant notamment sur la décomposition en ondelettes [10].

Segmentation de cellules La segmentation automatique de cellules à partir d'images de microscopie est une tâche importante permettant par la suite de compter ces cellules afin d'étudier l'évolution d'une population, ou de quantifier la morphologie cellulaire [11][12]. De nombreuses méthodes de segmentation cellulaire ont été proposées, parmi lesquelles on retrouve des seuillages automatiques basés sur les valeurs d'intensité [13][14], l'utilisation de modèles déformables [15] plus robustes aux va-

[6]: LEDLEY (1964), « High-Speed Automatic Analysis of Biomedical Pictures »

[7]: MEIJERING et al. (2016), « Imagining the future of bioimage analysis »

[8]: SAGE et al. (2017), « DeconvolutionLab2 »

[9]: MEINIEL et al. (2018), « Denoising of Microscopy Images »

[10]: OLIVO-MARIN (2002), « Extraction of spots in biological images using multiscale products »

[11]: MEIJERING (2012), « Cell Segmentation »

[12]: VICAR et al. (2019), « Cell segmentation methods for label-free contrast microscopy »

[13]: HYE-JIN JEONG et al. (2005), « Comparison of thresholding methods for breast tumor cell segmentation »

[14]: AL-HAFIZ et al. (2018), « Red blood cell segmentation by thresholding and Canny detector »

[15]: ZIMMER et al. (2005), « Coupled parametric active contours »

riations d'illumination, ou plus récemment de réseaux de neurones convolutionnels (CNNs) [16].

Suivi de mouvement Après détection ou segmentation des objets d'intérêt dans des séquences d'images biologiques, des méthodes de suivi de mouvement peuvent être appliquées à des particules [17] ou à des cellules [18].

Classification d'objets biologiques L'analyse d'images biologiques permet également la classification automatique d'objets, par exemple afin d'identifier des cellules d'intérêt parmi une grande variété de cellules, ou directement de classer des cellules comme étant saines, atypiques ou malignes. La classification d'objets biologiques reposent généralement sur des méthodes d'apprentissage, entraînées soit à partir de caractéristiques morphologiques ou de textures extraites au préalable soit directement à partir des images ⁶ [19].

Ainsi la généralisation des acquisitions d'images biologiques a conduit au développement de méthodes d'analyse d'images adaptées. Une quantité croissante de lames digitales est disponible grâce à l'adoption progressive des scanners de lames et, similairement, un nombre croissant d'approches d'analyse d'images pour la pathologie digitale ont été proposées.

1.3.2 Introduction à l'analyse de lames de pathologie digitales

Comme expliqué précédemment sous-section 1.2.2, un des intérêts de la numérisation de lames d'histologie et de cytologie est la possibilité d'y appliquer ensuite des méthodes de traitement et d'analyse d'images. L'utilisation de telles approches en pathologie digitale permet de répondre à plusieurs besoins.

Certaines tâches, comme le décompte de cellules d'intérêt dans des coupes histologiques, sont cruciales pour établir un diagnostic ou quantifier l'évolution d'une pathologie. Effectué manuellement par un pathologiste, un tel décompte est fastidieux et peut souffrir d'un manque d'objectivité. Des approches de détection et de segmentation ont donc été développées avec succès afin d'estimer précisément le nombre de cellules d'intérêt dans des lames d'histologie [20][21].

Lors de l'analyse de lames de pathologie, le praticien établit le diagnostic notamment en évaluant des caractéristiques morphologiques cellulaires, comme le rapport entre l'aire apparente du noyau et l'aire apparente du cytoplasme, ou la quantité de cellules en cours de mitose ⁷. L'évaluation de telles caractéristiques par des observateurs humains est sujette à interprétation. Afin de diminuer la subjectivité du diagnostic, différents systèmes de gradation ont été proposés. En 2008, Rakha et al. ont proposé le score de Nottingham [22] pour définir le grade histologique d'un cancer du sein en se basant sur la structure de la tumeur, le phénotype des cellules atypiques et l'activité mitotique. Epstein et al. ont proposé en 2014 le score de Gleason [23] afin de définir le degré d'agressivité du cancer de la prostate. Barkhan et al. ont définis le Système de Paris [24] pour uniformiser la caractérisation de carcinomes urothéliaux de haut-grade à partir de lames de cytologie. Ces systèmes de gradation permettent une harmonisation du diagnostic entre anatomopathologistes, mais restent basés sur des caractéristiques évaluées à l'œil. Des méthodes de segmentation et classification automatique d'images peuvent alors être appliquées pour fournir des mesures quantitatives et objectives des caractéristiques d'intérêt [25][26], afin de diminuer la variabilité du diagnostic entre anatomopathologistes.

[16]: RONNEBERGER et al. (2015), « U-Net »

[17]: CHENOARD et al. (2014), « Objective comparison of particle tracking methods »

[18]: ULMAN et al. (2017), « An objective comparison of cell-tracking algorithms »

6: Dans le cas de méthodes d'apprentissage profond, souvent à l'aide de réseaux de neurones convolutionnels.

[19]: IRSHAD et al. (2014), « Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology »

[20]: LOUKAS et al. (2003), « An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections »

[21]: MERUVIA-PASTOR et al. (2011), « Estimating Cell Count and Distribution in Labeled Histological Samples Using Incremental Cell Search »

7: Étape de division cellulaire chez les cellules eucaryotes.

[22]: RAKHA et al. (2008), « Prognostic Significance of Nottingham Histologic Grade in Invasive Breast Carcinoma »

[23]: EPSTEIN et al. (2015), « The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma »

[24]: BARKHAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[25]: NAGPAL et al. (2019), « Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer »

[26]: VAICKUS et al. (2019), « Automating the Paris System for urine cytology—A hybrid deep-learning and morphometric approach »

Finalement, dans le but de fournir un second avis pertinent, des approches d'analyse d'images cherchent à prédire directement un diagnostic à partir des lames de pathologie digitales. Ces méthodes, souvent basées sur des techniques d'apprentissage⁸, tirent profit de larges bases de données afin d'en extraire des heuristiques complexes.

8: Méthodes dites de *machine learning*.

1.3.3 Applications à l'histopathologie

Pré-traitement des lames d'histologie digitales Comme décrit dans la sous-section 1.2.3 le traitement des lames d'histologie digitales est compliqué par la très haute résolution des images, typiquement autour de $100\,000 \times 100\,000$ pixels (voir Figure 1.3). La solution couramment utilisée consiste à extraire un ensemble de sous-images⁹, typiquement de dimensions 256×256 pixels, des lames d'histologie digitales. La méthode de sélection de ces sous-images reste un sujet de recherche ouvert, et plusieurs approches ont été proposées. Généralement, uniquement les sous-images contenant au moins une portion de tissu sont conservées pour l'analyse, et les sous-images contenant uniquement de l'arrière-plan sont rejetées. L'analyse d'une lame d'histologie digitale repose ensuite sur l'analyse des sous-images extraites de celle-ci.

9: Ces sous-images sont également appelées *patches*.

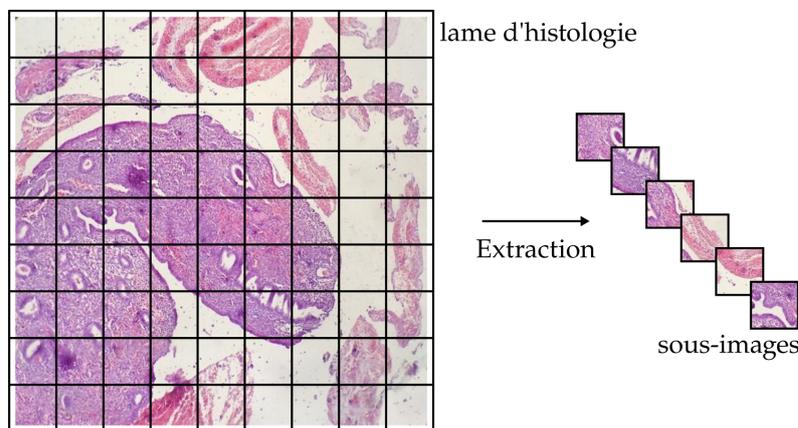


FIGURE 1.3 – Extraction des sous-images d'une lame d'histologie à partir d'une grille régulière sans superposition. Image d'histologie tirée de : www.ihearthisto.com/post/99813629042

Après un éventuel pré-traitement, de nombreuses approches d'analyse d'images histologiques ont été proposées afin de résoudre, entre autres, les problèmes suivants.

La segmentation supervisée de noyaux Plusieurs études ont cherché à obtenir automatiquement les masques de segmentation des noyaux présents dans des sous-images extraites d'une lame d'histologie digitale. Dans [27], Fatakdawala et al. proposent une méthode de segmentation automatique des noyaux de lymphocytes à partir d'un modèle de contours actifs suivi par un post-traitement permettant d'identifier les noyaux superposés. Plus récemment, de nombreuses études ont appliqué des CNN au problème de segmentation automatique de noyaux, avec d'excellents résultats. Xing et al. [28] utilisent un CNN pour effectuer une première détection des noyaux, suivi d'un algorithme itératif de fusion de régions pour obtenir la segmentation finale. La segmentation automatique de noyaux dans des tissus est complexe en présence de régions denses dans lesquelles les nombreux noyaux sont souvent en contact. Pour surmonter cette difficulté, Jung et al. [29] proposent d'appliquer le modèle Mask-RCNN [30] à des sous-images d'histologies provenant de multiples organes pour y détecter puis segmenter chaque noyau cellulaire. Qu et al. [31] utilisent un modèle U-Net [16], afin de prédire directement les masques de segmentation des noyaux et leurs contours à partir de sous-images d'histologie de tissu pulmonaire, évitant ainsi l'étape de détection. Dans [32], Naylor et al. utilisent également un modèle U-Net, cette fois afin de prédire une carte de distance par rapport aux centres des noyaux, plutôt que les contours, permettant de séparer les

[27]: FATAKDAWALA et al. (2010), « Expectation–Maximization-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR) »

[28]: XING et al. (2016), « An Automatic Learning-Based Framework for Robust Nucleus Segmentation »

[29]: JUNG et al. (2019), « An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

[31]: QU et al. (2019), « Joint Segmentation and Fine-Grained Classification of Nuclei in Histopathology Images »

[16]: RONNEBERGER et al. (2015), « U-Net »

[32]: NAYLOR et al. (2019), « Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map »

masques de segmentation des noyaux en contact. Une amélioration de cette idée est proposée par Graham et al. [33], qui utilisent un CNN pour la prédiction des masques de segmentation des noyaux présents dans une sous-image d'une lame d'histologie, ainsi que des cartes de gradients horizontaux et verticaux associées, permettant l'identification de chaque instance nucléaire en post-traitement.

La segmentation de régions d'intérêt Des méthodes d'analyses automatiques d'images ont également été appliquées à la segmentation automatique de régions d'intérêt. Ces régions peuvent être des types de tissus, comme dans [34] où les auteurs utilisent des contours actifs pour la segmentation de glandes sur des sous-images d'histopathologie de la prostate. Dans [35], Linder et al. utilisent une machine à vecteurs de support (SVM) et des caractéristiques extraites de motifs binaires locaux¹⁰ afin de déterminer si une sous-image d'histologie colorectale contient principalement du tissu stroma ou du tissu épithélial. L'inférence sur chaque sous-image de la lame d'histologie digitale permet d'obtenir une segmentation des régions stroma et épithéliale. Pour éviter la construction manuelle de caractéristiques décrivant les tissus, dans [36] les auteurs ont proposé un CNN afin de prédire directement des masques de segmentation de glandes, ainsi que les contours de ces régions, à partir de sous-images d'histologie. Similairement, Bulten et al. [37] appliquent un modèle U-Net à la segmentation de l'épithélium à partir de sous-images d'histologie de la prostate. Les régions d'intérêt à segmenter peuvent également correspondre aux régions atypiques ou tumorales, plutôt qu'aux types de tissus. Des CNNs ont été appliqués à des images d'histologie du sein afin de détecter des sous-images contenant des régions tissulaires de carcinome canalaire invasif [38] ou encore des métastase des ganglions lymphatiques [39]. La prédiction sur chaque sous-image permet ainsi d'obtenir une segmentation des régions tumorales au niveau d'une lame d'histologie.

La détection de mitoses comme marqueurs pronostiques de cancer Une forte activité mitotique peut être marqueur de cancer, ainsi de nombreuses méthodes ont cherché à détecter les cellules en cours de mitoses dans des lames d'histologie. Plusieurs études se sont intéressées à la classification de sous-images pour adresser le problème de détection automatique de mitoses. Dans [40] les auteurs proposent de classer chaque pixel de la lame d'histologie comme faisant partie d'une cellule en mitose ou non. Pour cela un CNN est entraîné à prédire la présence de mitose dans des sous-images centrées sur chaque pixel. Dans [41], Wang et al. combinent caractéristiques expertes et caractéristiques profondes extraites par un CNN sur des sous-images d'histologie afin de détecter les patches contenant des mitoses. Plusieurs études [42][43] appliquent des CNNs profonds, nécessitant de grandes quantités de données annotées, pour la classification automatique de sous-images d'histologie contenant des mitoses. Dans [43], Tellez et al. utilisent des biomarqueurs colorés permettant de mettre en évidence les cellules en cours de mitose, facilitant ainsi la récolte de données annotées nécessaires à l'entraînement du CNN. Afin de limiter la quantité d'images à analyser, des travaux se sont intéressés à des méthodes de détection d'objets, appliquées à de larges champs de vue, pour la détection automatique de cellules en cours de mitose. Dans [44], Chen et al. utilisent un CNN afin de prédire une carte de distance au centre de cellules en cours de mitoses. Une recherche de maximums locaux permet alors de localiser et de compter les cellules en cours de mitoses dans l'image analysée. L'étude [45] propose une architecture de CNN inspirée de Mask-RCNN [46] afin de prédire directement des boîtes englobantes autour des mitoses.

La quantification automatique du statut HER2 de cancers du sein Le statut HER2¹¹ de tissus provenant de biopsies du sein peut être automatiquement quantifié à l'aide de

[33]: GRAHAM et al. (2019), « Hover-Net »

[34]: XU et al. (2011), « A high-throughput active contour scheme for segmentation of histopathological imagery »

[35]: LINDER et al. (2012), « Identification of tumor epithelium and stroma in tissue microarrays using texture analysis »

10: *Local binary patterns*.

[36]: CHEN et al. (2017), « DCAN »

[37]: BULTEN et al. (2019), « Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard »

[38]: CRUZ-ROA et al. (2014), « Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks »

[39]: LIN et al. (2019), « Fast ScanNet »

[40]: CIREŞAN et al. (2013), « Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks »

[41]: WANG et al. (2014), « Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features »

[42]: ALBARQOUNI et al. (2016), « AggNet »

[43]: TELLEZ et al. (2018), « Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks »

[44]: CHEN et al. (2016), « Automated mitosis detection with deep regression networks »

[45]: RAO (2018), *MITOS-RCNN*

[46]: HE et al. (2017), « Mask RCNN »

11: Human Epidermal growth factor Receptor 2, est une protéine qui, lorsque sur-exprimée, est associée à une croissance rapide de la tumeur.

méthodes d'analyse d'images. Le statut HER2 d'une tumeur du sein est généralement quantifié à partir de lames d'immunohistochimie (IHC). Dans [47], les cellules présentes dans des sous-images d'histologie sont détectées puis classées, à l'aide d'un CNN, comme cellules immunitaires, cellules de stroma ou cellules tumorales avec différents niveaux de sur-expression de HER2. Dans [48], Qaiser et al. proposent un réseau de neurones récurrent (RNN) afin de prédire le statut HER2 de lames d'histologie, en analysant des sous-images extraites à différents grossissements. La méthode proposée imite la pratique des pathologistes en traitant les sous-images à fort grossissement uniquement si elle ont été préalablement identifiées comme intéressantes à de plus faibles grossissements. Dans [49], les auteurs cherchent à quantifier le statut HER2 de tumeurs du sein non pas à partir de lames IHC mais en utilisant des lames d'histologies préparées à l'aide des colorants hématoxyline et éosine (H&E). A partir de sous-images extraites de la lame d'histologie H&E, Lu et al. construisent un graphe à l'aide d'une triangulation de Delauney, puis appliquent un réseau de neurones en graphe (GNN) afin de prédire le statut HER2 de la tumeur.

La prédiction du taux de survie Dans [50] les auteurs appliquent des réseaux de neurones convolutionnels à des groupes de sous-images extraites de lames d'histologie pour la prédiction du taux de survie. De même, Courtiol et al. [51] entraînent un CNN à prédire pour chaque sous-image d'une lame d'histologie un score de prédiction de survie uniquement à partir d'annotations globales¹². Les scores de survie des sous-images les plus faibles et les plus élevés sont concaténés et utilisés afin de prédire le taux de survie final pour une lame d'histologie donnée. Dans [52], les auteurs s'intéressent à la prédiction de la probabilité de récurrence d'une tumeur. Pour cela, des CNNs auto-encodeurs sont utilisés pour extraire des caractéristiques profondes de sous-images d'histologie de la prostate à deux grossissements. Ces caractéristiques sont ensuite combinées et passées dans une SVM pour obtenir un score de probabilité de récurrence.

Le diagnostic du cancer Finalement, des méthodes d'analyse d'images ont été appliquées afin d'établir directement un diagnostic à partir d'une lame d'histologie. Dans [53] les auteurs cherchent à prédire automatiquement le score de Gleason¹³ à partir de lames digitales d'histologie de la prostate. Pour cela, des métriques sont calculées à partir d'une décomposition en ondelettes des images d'histologie et de la matrice de co-occurrence des niveaux de gris (GLCM)¹⁴. À partir de ces caractéristiques, un modèle K-NN [54] a été utilisé pour associer, à chaque sous-image d'histologie, le score de Gleason correspondant. Dans [55], les auteurs montrent que l'aire apparente moyenne des noyaux segmentés automatiquement dans une lame d'histologie du sein est fortement corrélée au score de Bloom-Richardson, utilisé pour évaluer la gravité des cancers du sein. Toujours pour évaluer la gravité des cancers du sein, Basavanhally et al. [56] proposent une méthode d'analyse d'image permettant de prédire automatiquement le score de Bloom-Richardson à partir de lames d'histologie du sein. Pour cela, des caractéristiques de la structure du tissu sont extraites de graphes, construits à partir des noyaux segmentés à l'aide de contours actifs. Une forêt d'arbres décisionnels est ensuite entraînée à prédire le score de Bloom-Richardson à partir de ces caractéristiques de graphes ainsi que des caractéristiques de Haralick [57] permettant de décrire la texture des noyaux cellulaires. Plus récemment, de nombreuses études se sont intéressées au diagnostic automatique de lames d'histologie digitales à l'aide de réseaux de neurones convolutionnels. Certaines études reposent sur la collection d'annotations locales.¹⁵ Par exemple, Zhang et al. [58] proposent une méthode afin de diagnostiquer automatiquement les cancers urothéliaux à partir de lames d'histologie de la vessie. Un premier CNN est utilisé pour détecter les régions tumorales à faible grossissement. Puis, un deuxième CNN recherche des anomalies au niveau des noyaux ou de la structure du tissu dans ces régions tumorales à

[47]: VANDENBERGHE et al. (2017), « Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer »

[48]: QAISER et al. (2019), « Learning Where to See »

[49]: LU et al. (2021), « SlideGraph+ »

[50]: ZHU et al. (2017), « WSISA »

[51]: COURTIOL et al. (2019), « Deep learning-based classification of mesothelioma improves prediction of patient outcome »

12: Une seule annotation est disponible pour toutes les sous-images de la lame d'histologie, correspondant au taux de survie du patient.

[52]: YAMAMOTO et al. (2019), « Automated acquisition of explainable knowledge from unannotated histopathology images »

[53]: SOLTANIAN-ZADEH et al. (2002), « Multiwavelet grading of prostate pathological images »

13: Classification histologique basée sur la structure du tissu permettant d'évaluer la gravité des cancers de la prostate.

14: *Grey Level Co-occurrence Matrix*.

[54]: CUNNINGHAM et al. (2022), « k-Nearest Neighbour Classifiers - A Tutorial »

[55]: VETA et al. (2012), « Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer »

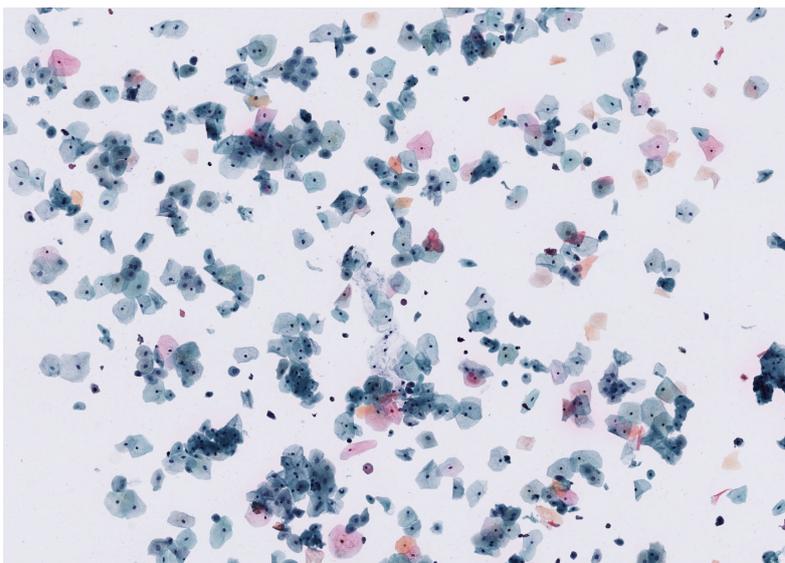
[56]: BASAVANHALLY et al. (2013), « Multi-Field-of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides »

15: Annotations des pixels ou des sous-images par des anatomopathologistes.

[58]: ZHANG et al. (2019), « Pathologist-level interpretable whole-slide cancer diagnosis with deep learning »

un plus fort grossissement. Finalement, un troisième CNN permet d'établir un diagnostic à partir des scores d'anomalies obtenus par sous-images à l'étape précédente. Dans [25], un CNN est entraîné, à partir d'annotations locales, à prédire le score de Gleason associé à chaque sous-image de la lame. Le diagnostic global est ensuite déterminé à partir de la distribution des sous-images dans chaque classe. Wei et al. [59] proposent une méthode afin d'automatiser la classification des profils histologiques dans des lames d'histologie pulmonaires. Un modèle ResNet [30] est entraîné afin de prédire, pour chaque sous-image, le profil histologique correspondant. Finalement, parmi tous les profils histologiques prédits, le profil prédominant est choisi comme prédiction globale pour la lame. Parce que la collection d'annotations locales est coûteuse et parfois difficile à obtenir, des travaux ont été menés afin de prédire le diagnostic de lames d'histologie uniquement à partir d'annotations globales¹⁶. Ces travaux reposent sur une approche faiblement supervisée appelée apprentissage à partir d'instances multiples (MIL), qui cherche à prédire le diagnostic d'une lame à partir de sous-images¹⁷ pour lesquelles aucune annotation locale n'est disponible. Dans [60], Hou et al. entraînent un CNN itérativement à prédire pour chaque sous-image d'une lame d'histologie de gliome le label global correspondant. Parce que toutes les sous-images d'une lame d'histologie positive ne sont pas discriminatives¹⁸, les sous-images utilisées à l'itération i sont les sous-images prédites avec une forte probabilité comme ayant le même label que la lame à l'itération $i - 1$. Une fois le CNN entraîné, un score de diagnostic est inféré pour chaque sous-image de la lame. Une régression logistique est finalement entraînée à prédire le diagnostic global de la lame à partir de l'histogramme des prédictions locales. Dans [61] les auteurs proposent un CNN qui prédit un score de probabilité de tumeur pour chaque sous-image d'une lame d'histologie. Seul le score de probabilité de tumeur le plus haut est utilisé lors de l'entraînement pour calculer la fonction de coût d'entropie croisée avec le label global de la lame. Finalement, Ilse et al. [62] proposent une méthode de MIL pour laquelle la combinaison des instances est effectuée à l'aide d'une moyenne pondérée dont les poids sont appris lors de l'entraînement par un mécanisme dit d'*attention*.

1.3.4 Applications à la cytopathologie



Pré-traitement des lames de cytologie digitales Comme pour les lames d'histologie, le traitement par ordinateur des lames de cytologie digitales¹⁹ est compliqué par la très haute résolution des images. Pour résoudre ce problème, il est possible d'extraire des

[25]: NAGPAL et al. (2019), « Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer »

[59]: WEI et al. (2019), « Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

16: Annotations au niveau de la lame uniquement.

17: Appelées *instances*.

[60]: Hou et al. (2016), « Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification »

18: Toutes les sous-images ne contiennent pas nécessairement une région tumorale.

[61]: CAMPANELLA et al. (2018), *Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology*

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

FIGURE 1.4 – Illustration d'une lame de cytologie digitale, provenant de l'essai clinique VisioCyt® 01.

19: Voir Figure 1.4.

sous-images à partir d'une grille comme pour les lames d'histologie. Il est également possible de détecter à faible grossissement les cellules isolées dans la lame de cytologie, puis d'extraire des sous-images centrées sur chacune de ces cellules isolées.

Après pré-traitement, plusieurs problèmes ont été étudiés à l'aide de méthodes d'analyse d'images de cytologie.

La segmentation de noyaux Dans [63], Angulo et al. proposent un pré-traitement basé sur des opérations morphologiques²⁰ suivi d'une segmentation par ligne de partage des eaux²¹ afin d'obtenir automatiquement les masques de segmentation de leucocytes présents dans des lames de cytologie digitales de frottis sanguins.

La segmentation de cellules superposées Une difficulté majeure dans l'analyse de lames de cytologie par rapport aux lames d'histologie est la présence d'objets d'intérêt fortement superposés. Ainsi de nombreux travaux se sont intéressés à la détection et la segmentation des noyaux et cytoplasmes de cellules superposées. Dans [64] Song et al. proposent une méthode de segmentation d'instances de noyaux et des régions cytoplasmiques dans des lames de cytologie cervicale. Pour cela, un CNN est utilisé afin d'extraire des caractéristiques profondes à partir des super-pixels d'une sous-image de la lame de cytologie. Un graphe est ensuite construit à partir de ces super-pixels et des caractéristiques calculées, dont la partition permet ensuite d'obtenir la segmentation finale. Dans [65], les mêmes auteurs étendent leur travail précédent en cherchant cette fois à segmenter les instances cellulaires, c'est-à-dire à prédire un masque de segmentation pour chaque cellule unique. Ainsi, à partir des masques d'instances des noyaux n_i et du masque de segmentation sémantique du cytoplasme obtenus par la méthode [64], chaque pixel du cytoplasme se voit associé la probabilité d'appartenir à la cellule i . La segmentation d'instances cellulaires est approximée en sélectionnant pour chaque pixel la cellule qui maximise cette probabilité, puis raffinée à l'aide de modèles déformables.

Phoulady et al. [66][67] proposent également une méthode de segmentation d'instances cellulaires dans des lames de cytologie cervicale à partir d'acquisitions multi-plans. Pour cela, une détection de noyaux est réalisée à l'aide de seuillages successifs et d'une recherche des régions dont les caractéristiques²² sont comprises entre des seuils définis au préalable. Un modèle de mélange gaussien (GMM)²³ à deux composantes est ensuite appliqué aux intensités des pixels pour segmenter les pixels des cellules des pixels de l'arrière-plan. Le masque de cytoplasme obtenu est ensuite divisé en blocs suivant une grille régulière. Une mesure de similarité est calculée entre chaque bloc, selon la distance euclidienne entre leurs barycentres, et la distance euclidienne entre des vecteurs de focus calculés sur l'ensemble des plans focaux. Une première estimation des masques de segmentation d'instances est obtenue par groupement des blocs de cytoplasme en fonction de leur similarité. Les masques de segmentation d'instances cellulaires sont finalement raffinés afin de faire correspondre aux bords des masques les régions de plus fort gradient d'intensité. Dans [68] Tareef et al. proposent une série de segmentations par ligne de partage des eaux, appliquée à la segmentation d'instances cellulaires à partir de lames de cytologie cervicale. Après segmentation des noyaux, une deuxième segmentation, initialisée à partir de ces derniers permet d'obtenir des masques de segmentation pour les cellules isolées ainsi que pour les groupes de cellules superposées. Une troisième segmentation permet alors d'obtenir les masques d'instances des cellules appartenant aux groupes de cellules superposées détectés à l'étape précédente. Zhou et al. [69][70] utilisent une version améliorée du modèle Mask-RCNN [46] afin de détecter et segmenter des instances cellulaires à partir de sous-images de lames de cytologie cervicale. Notamment, un premier CNN est appliqué après la branche de détection afin de supprimer les doublons et un deuxième CNN est appliqué après la branche de

[63]: ANGULO et al. (2003), « Microscopic image analysis using mathematical morphology : Application to haematological cytology »

20: Ouverture et fermeture.

21: Également appelée *watershed*.

[64]: SONG et al. (2015), « Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning »

[65]: SONG et al. (2017), « Accurate Cervical Cell Segmentation from Overlapping Clumps in Pap Smear Images »

[66]: PHOULADY et al. (2016), « A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images »

[67]: PHOULADY et al. (2017), « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images »

22: Aire, intensité, solidité et circularité.

23: *Gaussian Mixture Model*.

[68]: TAREEF et al. (2018), « Multi-Pass Fast Watershed for Accurate Segmentation of Overlapping Cervical Cells »

[69]: ZHOU et al. (2019), « IRNet »

[70]: ZHOU et al. (2020), « Deep Semi-supervised Knowledge Distillation for Overlapping Cervical Cell Instance Segmentation »

[46]: HE et al. (2017), « Mask R-CNN »

segmentation afin d'utiliser l'information des instances similaires dans la sous-image pour raffiner la prédiction du masque.

La détection de parasites La cytologie digitale a également permis le développement de méthodes de détection automatique de parasites sanguicoles, dont le plasmodium responsable du paludisme. Dans [71] Tsai et al. utilisent des opérateurs morphologiques de Sobel afin de détecter les érythrocytes et de segmenter automatiquement les parasites à partir d'images de frottis sanguin. Plusieurs études proposent une détection de plasmodium à partir d'une détection automatique des érythrocytes à l'aide d'un seuillage d'Otsu [72], permettant l'extraction de sous-images centrées sur ces cellules d'intérêt. À partir de ces détections, Pan et al. [73] entraînent un CNN LeNet-5 [74] afin de différencier les globules rouges sains des globules rouges infectés. Cette classification automatique a également été adressée par Maqsood et al. [75] à l'aide d'une architecture de CNN personnalisée, ou encore par Li et al. [76], qui ont développé un CNN afin d'extraire des caractéristiques profondes à partir des images d'érythrocytes, puis appliqué une SVM afin de différencier les cellules saines des cellules infectées. Plutôt qu'une détection préalable des globules rouges, Hung et al. [77] appliquent le modèle Faster-RCNN [78] afin de détecter les cellules à partir de sous-images de cytologie, puis de leur attribuer une classe parmi *érythrocyte sain* ou *autre*²⁴. Un second CNN est ensuite appliqué afin de différencier les différentes cellules contenues dans la classe *autre*. Suivant la même idée, dans [79] les auteurs utilisent le modèle Faster-RCNN [78] pour la détection des érythrocytes à partir d'images de cytologie sanguine. Un modèle ResNet [30] est ensuite utilisé afin de séparer les érythrocytes sains des cellules infectées. Finalement, un second modèle ResNet [30] est utilisé afin de déterminer le stade du cycle de vie des parasites pour les cellules infectées.

La classification de cellules anormales À partir d'annotations locales, de nombreuses études ont appliqué des CNNs à la classification de cellules cervicales afin de différencier des cellules saines de cellules anormales. Plusieurs travaux, tous appliqués à l'analyse de lames de cytologie cervicale, reposent sur une détection des noyaux de cellules cervicales suivie d'une extraction de sous-images centrées sur les noyaux et finalement une classification des sous-images. Zhang et al. [80] utilisent un CNN pour la classification binaire de sous-images de cellules cervicales, i.e. différencier les cellules cervicales atypiques des cellules cervicales saines. Pour résoudre le même problème, Nguyen et al. [81] appliquent un ensemble de trois CNNs, dont un modèle ResNet [30] et un modèle Inception [82]. Diniz et al. [83] utilisent une stratégie ensembliste à partir de trois modèles EfficientNet [84], et Yu et al. [85] proposent un CNN contenant notamment des modules Inception [82] dont les noyaux de convolution de différentes tailles permettent d'adresser les dimensions variables des cellules d'intérêt. Dans [86], les auteurs adressent un problème plus compliqué, et utilisent une cascade de CNNs pour différencier les cellules saines et cinq classes de cellules atypiques. Mohammed et al. [87] proposent l'utilisation d'un modèle DenseNet [88] pour le même problème de classification de cellules cervicales. Lin et al. [89] appliquent un modèle GoogLeNet [82] aux images combinées aux masques de segmentation des noyaux et cytoplasmes afin de classer les cellules cervicales en sept classes, en fonction de leur type cellulaire et degré d'atypie. D'autres études ont cherché à identifier les cellules atypiques à partir de plus larges champs de vue d'une lame de cytologie digitale. Dans [90], les noyaux des cellules contenues dans des lames de cytologie de liquide pleural sont segmentés par un groupement non-supervisé de super-pixels à l'aide de la méthode K-Means, suivi par un ensemble d'arbres de décision pour différencier les noyaux atypiques des noyaux sains à partir de caractéristiques extraites manuellement, comprenant des mesures de morphologie, de couleur et de texture. Dans [91], les auteurs utilisent le modèle Mask-RCNN [46] afin de détecter et

[71]: TSAI et al. (2015), « Blood Smear Image Based Malaria Parasite and Infected-Erythrocyte Detection and Segmentation »

[72]: OTSU (1979), « A Threshold Selection Method from Gray-Level Histograms »

[73]: PAN et al. (2018), « Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks »

[74]: LECUN et al. (1998), « Gradient-Based Learning Applied to Document Recognition »

[75]: MAQSOOD et al. (2021), « Deep Malaria Parasite Detection in Thin Blood Smear Microscopic Images »

[76]: LI et al. (2022), « Residual attention learning network and SVM for malaria parasite detection »

[77]: HUNG et al. (2017), « Applying Faster R-CNN for Object Detection on Malaria Images »

[78]: REN et al. (2016), « Faster R-CNN »

24: Contenant les cellules infectées ainsi que les leucocytes.

[79]: DAVIDSON et al. (2021), « Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

[80]: ZHANG et al. (2017), « DeepPap »

[81]: NGUYEN et al. (2019), « Biomedical image classification based on a feature concatenation and ensemble of deep CNNs »

[82]: SZEGEDY et al. (2014), « Going Deeper with Convolutions »

[83]: N. DINIZ et al. (2021), « A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification »

[84]: TAN et al. (2020), « EfficientNet »

[85]: YU et al. (2021), « Automatic Classification of Cervical Cells Using Deep Learning Method »

[86]: GAUTAM et al. (2018), *Considerations for a PAP Smear Image Analysis System with CNN Features*

[87]: MOHAMMED et al. (2021), « Single-cell conventional pap smear image classification using pre-trained deep neural network architectures »

[88]: HUANG et al. (2018), « Densely Connected Convolutional Networks »

[89]: LIN et al. (2019), « Fine-Grained Classification of Cervical Cells Using Morphological and Appearance Based Convolutional Neural Networks »

segmenter automatiquement les noyaux des cellules cervicales puis prédire une classe parmi *normale* ou *atypique*. Dans [92] Ke et al. appliquent une succession de deux CNNs à des sous-images extraites de lames de cytologie cervicales suivant une grille, de façon similaire aux méthodes généralement utilisées en histologie. Le premier modèle est un modèle de segmentation entraîné à prédire des masques des noyaux et cytoplasmes des cellules cervicales, le second modèle est un modèle ResNet [30] utilisé pour classer les sous-images parmi douze classes en fonction du type cellulaire et du degré d'atypie.

Le diagnostic du cancer Comme pour les lames d'histologie, des méthodes d'analyse d'images ont été développées afin d'établir directement un diagnostic à partir d'une lame de cytologie. En se basant sur des annotations d'atypie locales, Pirovano et al. [93] et Cheng et al. [94] appliquent des CNN afin de prédire respectivement une classe et un score d'atypie, à partir de sous-images contenant de multiples cellules endocervicales. Toujours à partir d'annotations locales, Awan et al. [95] utilisent un seuil automatique suivi d'opérations morphologiques pour détecter des cellules urothéliales isolées dans une lame de cytologie urinaire, puis un modèle Xception [96] afin de séparer les cellules urothéliales en trois classes : saines, atypiques et cancéreuses. Parce que les annotations locales sont coûteuses, difficiles à obtenir et parfois bruitées, des études ont cherché à établir un diagnostic uniquement à partir d'une annotation globale, au niveau de la lame. Ainsi, dans [97], les auteurs adoptent une approche faiblement supervisée basée sur des labels bruités. Un CNN est entraîné à prédire le label global de la lame pour toutes les sous-images qui en sont extraites et qui contiennent des cellules d'intérêt. L'apprentissage à partir d'instances multiples est une approche faiblement supervisée populaire pour le diagnostic à partir de lames de cytologie, pour laquelle un diagnostic global cherche à être prédit à partir d'un ensemble d'instances, appelé *sac*. Ce type d'approche a été utilisé avec des vecteurs d'attention permettant l'identification des instances les plus informatives dans [98] pour le diagnostic à partir de sous-images extraites de lames de cytologie pulmonaire, et dans [99] pour le diagnostic à partir de sous-images de cellules urothéliales isolées provenant de lames de cytologie urinaire.

1.4 Acquisition multi-plans en cytopathologie digitale

1.4.1 Besoin de l'acquisition multi-plans

A l'inverse de l'histopathologie digitale, la cytopathologie digitale peine à être largement adoptée [100]. Plusieurs raisons peuvent expliquer ce retard :

- La cytopathologie comporte différents types d'échantillons comme les cytoponctions, les frottis ou les squames qui présentent chacun leurs spécificités et nécessitent éventuellement des traitements adaptés.
- Contrairement aux lames d'histologie qui sont bi-dimensionnelles, les lames de cytologie contiennent des amas cellulaires tri-dimensionnels²⁵, même avec l'utilisation de préparations en couche mince [101].

Pour analyser les objets tri-dimensionnels présents dans les lames de cytologie, les anatomopathologistes et cytotechniciens peuvent ajuster la mise au point du microscope à différentes profondeurs selon l'axe *z*. Afin de reproduire le même comportement, certains scanners de lames peuvent imager la lame de cytologie selon plusieurs plans focaux, aboutissant à l'acquisition d'une série d'images à différentes profondeurs *z*, appelée *z-stack* [102][103]. Il a été montré que les anatomopathologistes analysant des lames de cytologie digitales ont plus confiance en leurs diagnostics lorsqu'ils ont accès à des *z-stacks*, plutôt qu'à de simples acquisitions 2D [104].

[90]: WIN et al. (2018), « Computer Aided Diagnosis System for Detection of Cancer Cells on Cytological Pleural Effusion Images »

[91]: SOMPAWONG et al. (2019), « Automated Pap Smear Cervical Cancer Screening Using Deep Learning »

[46]: HE et al. (2017), « Mask R-CNN »

[92]: KE et al. (2021), « Quantitative analysis of abnormalities in gynecologic cytopathology with deep learning »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

[93]: PIROVANO et al. (2021), « Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier »

[94]: CHENG et al. (2021), « Robust whole slide image analysis for cervical cancer screening using deep learning »

[95]: AWAN et al. (2021), « Deep learning based digital cell profiles for risk stratification of urine cytology images »

[96]: CHOLLET (2017), *Xception*

[97]: DOV et al. (2021), « Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images »

[98]: TERAMOTO et al. (2021), « Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning »

[99]: BUTKE et al. (2021), « End-to-end Multiple Instance Learning for Whole-Slide Cytopathology of Urothelial Carcinoma »

[100]: HANNA et al. (2017), « Why is digital pathology in cytopathology lagging behind surgical pathology? »

25: Voir Figure 1.5.

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

[102]: YAMASHIRO et al. (2009), « Comparison between a traditional single still image and a multiframe video image along the *z*-axis of the same microscopic field of interest in cytology »

[103]: KHALBUSS et al. (2011), « Digital Imaging in Cytopathology »

[104]: DONNELLY et al. (2013), « Optimal *z*-axis scanning parameters for gynecologic cytology specimens »

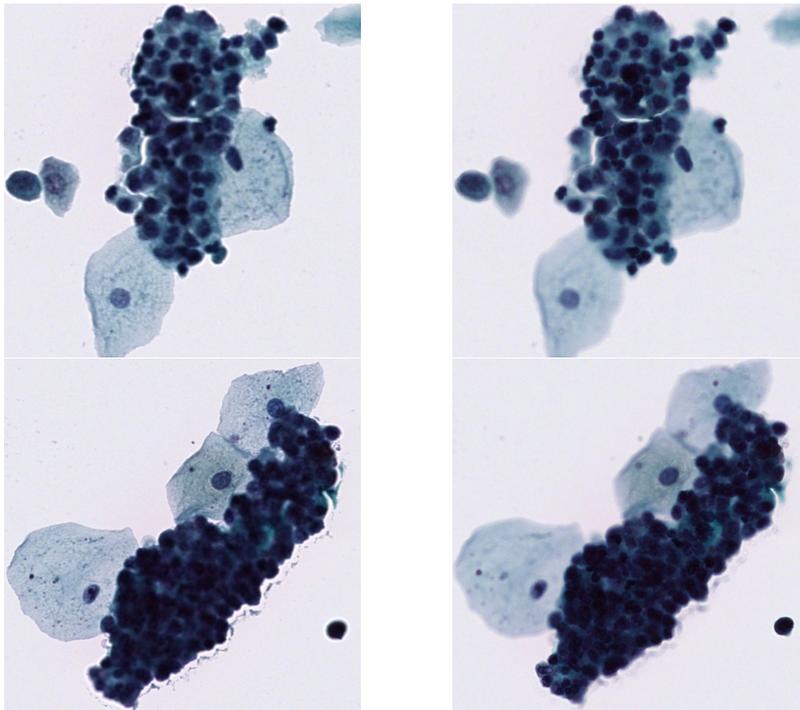


FIGURE 1.5 – Exemple d’amas cellulaire tri-dimensionnels présents dans les lames de cytologie. À gauche, un amas imagé à un plan focal $\mathcal{F}_{z=+1\mu m}$. À droite, un amas imagé à un plan focal $\mathcal{F}_{z=-3\mu m}$. Les objets cellulaires sont visibles nettement à différents plans focaux, une acquisition multi-plans est donc nécessaire pour leur analyse.

Dans le cas de l’analyse automatique de lames de cytologie, il apparaît alors naturel de chercher à imiter l’analyse d’experts et de fournir à l’algorithme des lames multi-plans. Dans le cas où l’acquisition de plusieurs plans focaux n’est pas disponible, il reste nécessaire de prendre en compte la présence d’objets tri-dimensionnels dans l’analyse, par exemple en filtrant les objets hors-focus [105].

[105]: FRIEDRICH et al. (2016), « Removing defocused objects from single focal plane scans of cytological slides »

1.4.2 Fonctionnement sur les scanners de lames

Les scanners de lames sont équipés d’une fonctionnalité d’auto-focus automatique ou semi-automatique, basée sur une carte de focalisation pré-calculée, et la plupart intègrent désormais des fonctions d’acquisitions multi-plans. Une profondeur de mise au point optimale est calculée pour plusieurs points de focalisation²⁶ modifiables par l’opérateur. Pour chacun de ces points, la position selon l’axe z du plan focal le plus net est calculée. A partir de ces mesures il est possible de définir une carte de focalisation, en extrapolant une surface de meilleur focus, ou en calculant un plan de meilleur focus minimisant la distance quadratique moyenne aux points de focalisation, comme illustré Figure 1.6.

26: Dans ce travail, nous avons utilisé 9 points de focalisation par lame de cytologie avec le scanner Hamamatsu NanoZoomer® S360.

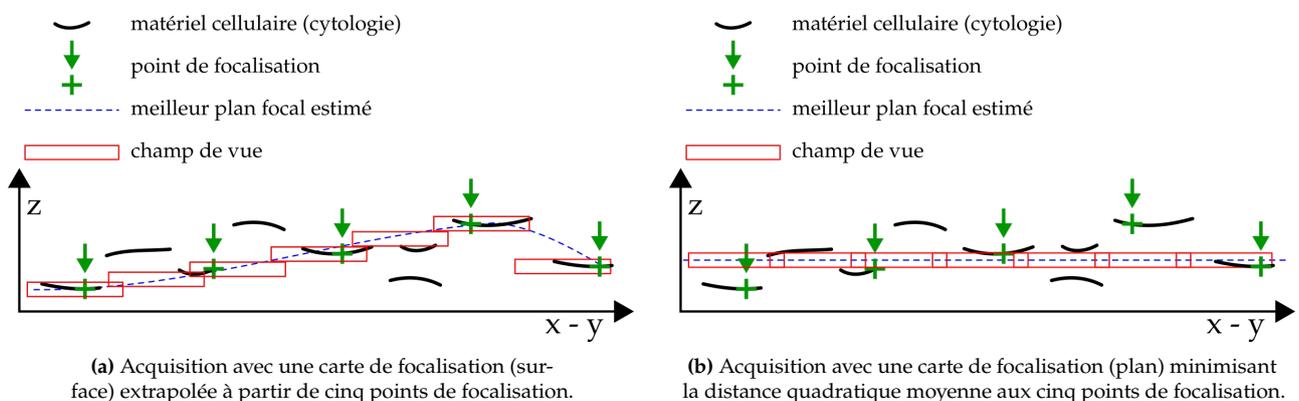


FIGURE 1.6 – Deux méthodes de création de carte de focalisation pour une lame de cytologie à partir de cinq points de focalisation. Adapté de [106].

L'acquisition de z-stacks s'effectue par champ de vue, en prenant pour référence le plan focal le plus net z_0 défini par la carte de focalisation du scanner et en imageant $2n + 1$ plans focaux de profondeur $\mathbf{z} \in \mathbb{R}^{2n+1}$ séparés de $\delta \mu\text{m}$ (voir Figure 1.7) :

$$\mathbf{z} = z_0 \pm k \times \delta \mid k \in [0, n] \quad (1.1)$$

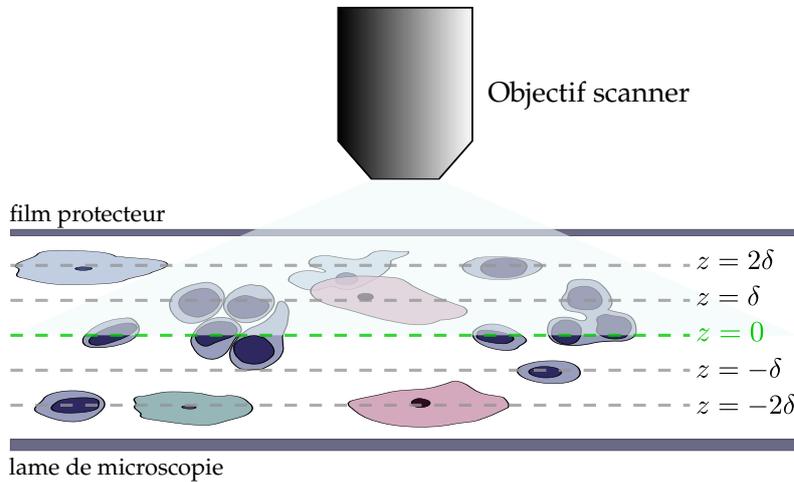


FIGURE 1.7 – Acquisition de multiples plans focaux (z-stack) séparés de $\delta \mu\text{m}$ pour un champ de vue. La répartition des objets cellulaires à différentes profondeurs et la présence d'amas cellulaires tri-dimensionnels encouragent l'acquisition de z-stacks pour l'analyse de lames digitales de cytologie.

1.4.3 Optimisation de l'acquisition de lames digitales multi-plans

L'acquisition multi-plans de lames digitales augmente le temps de numérisation nécessaire proportionnellement au nombre de plans focaux supplémentaires. Le temps moyen d'une analyse manuelle d'une lame de cytologie étant de 5.6 minutes [107], il est nécessaire de limiter le temps d'acquisition et d'analyse à 5.6 minutes pour ne pas perdre en efficacité par rapport à l'analyse manuelle. De plus, augmenter le nombre de plans focaux numérisés augmente également la taille des fichiers numériques, ce qui complique l'analyse, le transfert ou l'archivage de lames digitales. Pour palier à ces difficultés, il est important de trouver un compromis entre la quantité d'information numérisée, le temps d'acquisition nécessaire et la capacité de stockage nécessaire.

Afin de réduire le temps d'acquisition des lames de cytologie numérique, Fan et al. [108] ont proposé en 2014 un algorithme permettant de délimiter l'échantillon d'intérêt pour réduire la zone à scanner par le scanner de lames au nécessaire. De plus, une identification automatique de points de focalisation permet d'accélérer la recherche du plan de meilleur focus par le scanner, notamment en évitant les mesures de focus sur des régions qui ne contiennent pas de matériel cellulaire.

Plusieurs travaux ont également porté sur le nombre optimal de plans focaux à numériser. En se basant sur le ressenti de cinq experts et la concordance entre leurs diagnostics, Donnelly et al. [104] recommandent l'utilisation de trois plans focaux séparés de $1 \mu\text{m}$. Dans [101], Fan et al. proposent un algorithme afin de quantifier la distribution tri-dimensionnelle des cellules d'intérêt dans les lames de cytologie cervicale, et d'en déduire un nombre de plans focaux optimal à numériser. Les auteurs recommandent ainsi l'acquisition de 21 plans focaux séparés de $1 \mu\text{m}$.

[107]: BOOST (2009), « A comparison of screening times between the Thin-Prep imager and conventional cytology »

[108]: FAN et al. (2014), « An algorithm for microscopic specimen delineation and focus candidate selection »

[104]: DONNELLY et al. (2013), « Optimal z-axis scanning parameters for gynecologic cytology specimens »

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

1.5 Objectifs des travaux de recherche

Cette introduction montre que la cytologie digitale, couplée à des algorithmes d'analyse automatique d'images, a le potentiel d'améliorer la sensibilité de la cytologie urinaire pour le diagnostic précoce du cancer de la vessie.

Comme décrit sous-section 1.3.4, un certain nombre d'études se sont intéressées à l'analyse d'images en cytologie digitale. Cependant plusieurs questions restent en suspens :

Comment analyser efficacement²⁷ des lames de cytologie digitales multi-plans pour le diagnostic du cancer de la vessie ?

Quel gain de sensibilité l'application d'algorithmes d'analyses d'images peut-elle apporter à l'examen de cytologie urinaire et quelle est la contribution de l'analyse multi-plans à ce gain ?

Pour tenter de répondre à ces questions, l'objectif principal de cette thèse est de développer des méthodes d'analyses de lames virtuelles de cytologie urinaire multi-plans pour le diagnostic du cancer de la vessie. Pour y parvenir, plusieurs sous-objectifs sont définis :

- Déterminer des paramètres optimaux de création et de numérisation de lames virtuelles multi-plans.
- Développer des méthodes d'extraction d'information à partir de z-stacks.
- Proposer une méthode de diagnostic du cancer de la vessie à partir de lames de cytologie digitales.
- Évaluer l'impact de l'acquisition multi-plans et de l'information 3D sur les performances de diagnostic.

La suite de ce manuscrit est composée comme suit : Le chapitre 2 introduit plusieurs représentations 2D de l'information 3D contenue dans les z-stacks. Les paramètres optimaux de création et de numérisation de lames digitales de cytologie urinaire sont discutés chapitre 3. Le chapitre 4 présente plusieurs méthodes afin d'extraire efficacement de l'information à partir de z-stacks en utilisant des réseaux de neurones. Une méthode de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales est proposée chapitre 5. Le chapitre 6 évalue l'impact des z-stacks sur les performances de diagnostic. Finalement le chapitre 7 résume les contributions de ce travail, discute des perspectives et conclut ce manuscrit.

²⁷: Par *efficacement* s'entend une analyse tirant au mieux parti de toute l'information contenue dans des lames digitales multi-plans tout en limitant le coût et temps de calcul supplémentaire par rapport à une simple lame digitale 2D.

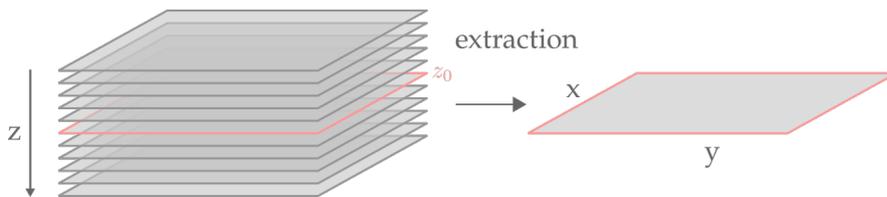
Représentation 2D de l'information 3D

Ce chapitre présente une partie des travaux visant à extraire de l'information à partir de lames digitales multi-plans. Afin de pouvoir tirer partie des nombreuses méthodes développées en analyse d'image et de faciliter l'analyse et le stockage de données multi-plans, ce chapitre étudie plusieurs représentations 2D de l'information 3D contenue dans les z-stacks.

Soit \mathcal{X} une z-stack centrée en (x, y) sur un objet d'intérêt¹. Cette z-stack est composée de $2n+1$ plans focaux², espacés de δ_z et centrés selon l'axe z sur la section centrale z_0 . Ainsi n plans focaux sont numérisés au-dessus du plan z_0 aux positions $z = z_0 + k \times \delta_z$ $k \in [1, n]$. Par symétrie, n plans focaux sont numérisés au-dessous du plan z_0 aux positions $z = z_0 - k \times \delta_z$ $k \in [1, n]$. Pour \mathcal{X} , trois représentations 2D sont définies : la section centrale S_{centre} , la section de meilleur focus S_{sharp} et la projection par extension de profondeur de champ S_{edf} .

2.1 Section centrale

Soit S_{centre} la section centrale de la z-stack. Cette section, illustrée Figure 2.1, correspond à l'image acquise au plan focal z_0 , soit le plan de meilleur focus déterminé par le scanner. S_{centre} est donc l'équivalent d'une simple numérisation 2D de la lame de cytologie. Cette représentation sert de référence pour évaluer l'impact de l'information 3D pour l'analyse locale et globale de lames de cytologie digitales dans la suite de ce travail. Il est à noter qu'en utilisant pour référence la section centrale S_{centre} d'une lame digitale multi-plans plutôt qu'une autre acquisition digitale 2D de la même lame, on s'affranchit des problématiques de recalage et de variations qui peuvent survenir lors de la re-numérisation, comme le bruit numérique, une variation de la balance des blancs ou une variation de la carte de focalisation.



1: Dans le cas de lames de cytologie urinaires, les objets d'intérêt sont les cellules urothéliales et amas de cellules urothéliales.

2: Un scanner de lames équipé d'une fonction d'acquisition multi-plan commence généralement par identifier un plan de meilleur focus z_0 , puis numérise les plans focaux additionnels par rapport à celui-ci. Il est donc naturel de numériser des z-stacks contenant un nombre impair de plans focaux pour respecter une symétrie par rapport à z_0 .

FIGURE 2.1 – Section centrale S_{centre} d'une z-stack. Cette section est l'équivalent d'une numérisation 2D de la lame et sert de référence pour évaluer la pertinence de l'information 3D dans la suite de ce travail.

2.2 Section de meilleur focus

Lors de la numérisation, le plan de meilleur focus, noté z_0 , est calculé par le scanner pour une tuile ou pour la lame entière, contenant donc une multitude d'objets d'intérêt. Bien que le plan z_0 soit globalement le plan contenant le plus d'objets en focus, de part la répartition tri-dimensionnelle de ces objets d'intérêt dans une lame de cytologie, tous les objets ne sont pas nets pour $z = z_0$ et ainsi la section de meilleur focus S_{sharp} de la z-stack pour une cellule donnée n'est pas nécessairement la section centrale S_{centre} . Pour chaque z-stack, la section de meilleur focus S_{sharp} est extraite à la position $z = z_{sharp}$.

La position z_{sharp} est définie par la variance de l'opérateur Laplacien, une mesure locale d'autofocus calculée sur chaque section de la z-stack et définie Équation 2.1.

$$z_{sharp} = \underset{z}{\operatorname{argmax}} \sum_{x,y} \operatorname{var}_{[3 \times 3]} \left(\frac{\delta^2 I_z}{\delta x^2} + \frac{\delta^2 I_z}{\delta y^2} \right) \quad (2.1)$$

Avec :

$\frac{\delta^2 \bullet}{\delta x^2} + \frac{\delta^2 \bullet}{\delta y^2}$: l'opérateur laplacien

$\operatorname{var}_{[3 \times 3]}$: la variance calculée sur un voisinage de 3×3 pixels

Une fois la position z_{sharp} calculée pour une z-stack donnée, la section de meilleur focus S_{sharp} peut être extraite (voir Figure 2.2).

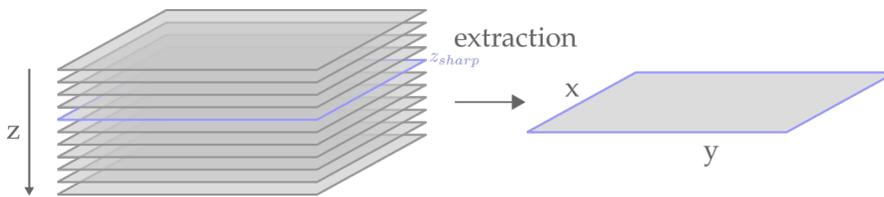


FIGURE 2.2 – Section de meilleur focus S_{sharp} d'une z-stack \mathcal{Z} . Cette section correspond au plan de meilleur focus pour l'objet contenu dans \mathcal{Z} .

2.3 Projection par extension de profondeur de champ

Publication

Les travaux conduits dans cette section ont été décrits dans un article de conférence, accepté à IEEE International Conference on Pattern Recognition (ICPR) 2020 :

A. Bouyssoux, R. Fezzani, and J.-C. Olivo-Marin, "Extended Depth of Field Preserving Color Fidelity For Automated Digital Cytology," in 2020 25th International Conference on Pattern Recognition (ICPR), Milan, 2021.

L'extraction de la section de meilleur focus ne permet pas de tirer parti de toute l'information disponible dans une z-stack. En effet, seule l'information contenue dans la section S_{sharp} est prise en compte, et l'information potentiellement contenue dans les autres plans focaux est alors perdue. De plus, il est fréquent que les objets ne soient pas entièrement nets sur un seul plan focal : dans [101], Fan et al. ont évalué à $10 \mu\text{m}$ l'épaisseur médiane d'amas cellulaires dans des lames de cytologie cervicale, alors que la profondeur typique d'un scanner de lames au grossissement $\times 40$ est d'environ $1 \mu\text{m}$.

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

Pour surmonter ce problème, des algorithmes d'extension de profondeur de champ (EDF) ont été proposés. Ces méthodes permettent de fusionner les sections d'une z-stack en une image 2D composite dont toutes les parties sont apparemment nettes, comme illustré Figure 2.3.

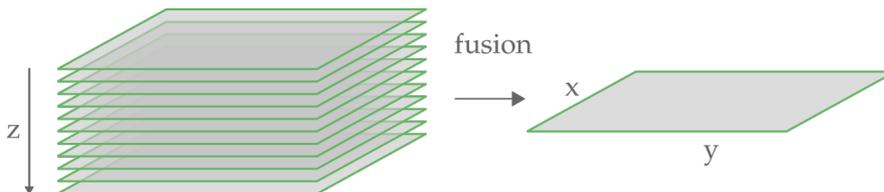


FIGURE 2.3 – Projection par extension de profondeur de champ S_{edf} : création d'une image dont toutes les parties sont apparemment nettes.

Ces méthodes ont été largement étudiées et peuvent être divisées principalement en 3 catégories :

1. Les approches spatiales, directement appliquées sur les valeurs d'intensité de la z-stack, pour lesquelles la règle de sélection des voxels est souvent basée sur des mesures d'autofocus classiques telles que la variance de l'opérateur Tenengrad [109][110].
2. Les méthodes basées sur une transformation du signal, telle que la transformée en ondelettes. Une sélection de coefficients contenant un maximum d'information est effectuée à partir de la décomposition de chaque section de la z-stack dans le domaine de la transformation choisie. La projection EDF est ensuite obtenue par transformation inverse des coefficients sélectionnés [111][112].
3. Les méthodes basées sur l'apprentissage profond, où des réseaux de neurones sont entraînés à prédire une carte de mise au point à partir d'une z-stack [113]. Cette carte de mise au point est ensuite utilisée afin de sélectionner, pour chaque position (x, y) , le voxel de la z-stack contenant l'information nette, produisant ainsi une image EDF.

Les approches des catégories 1. et 3. fusionnent directement les voxels de la z-stack originale afin de produire l'image EDF finale. Ces méthodes ont donc une bonne fidélité couleur, mais peuvent perdre de l'information ou produire des artefacts en présence d'objets superposés et transparents, communs en cytologie comme illustré Figure 2.10 et Figure 2.11.

Les approches de la catégorie 2. sont bien adaptées aux images de cytologie [111], mais sont définies pour des volumes en niveaux de gris, et obtenir une reconstruction avec une haute fidélité des couleurs n'est pas trivial. En cytopathologie, des colorants spécifiques sont utilisés pour mettre en évidence des biomarqueurs pour le diagnostic du cancer. Cela motive le besoin de méthodes d'EDF qui préservent fidèlement les couleurs pour la cytologie digitale.

Cette section présente une méthode d'extension de profondeur de champ adaptée à la cytologie digitale et qui préserve la fidélité des couleurs.

2.3.1 Description de la méthode d'extension de profondeur de champ multi-canal

L'approche proposée appartient aux méthodes basées sur une transformation du signal. Elle s'appuie sur une sélection de coefficients adaptée à la fusion de transformées en ondelettes, suivie d'une étape de reconstruction couleur permettant de préserver la fidélité des couleurs de la projection EDF.

Transformation en ondelettes et sélection de coefficients

Pour calculer l'image EDF d'une z-stack, une transformée en ondelettes est appliquée à toutes les sections qui la composent. Pour chaque position (x, y) , chaque niveau de décomposition en ondelettes et chaque sous-bande³, les coefficients correspondant au maximum d'information en focus sont sélectionnés à partir de la z-stack de transformées en ondelettes. L'image EDF finale est obtenue par la transformée en ondelettes inverse de ces coefficients sélectionnés.

Deux algorithmes de transformée en ondelettes ont été proposés pour la cytologie digitale : la transformée en ondelettes discrètes (DWT)⁴ et la transformée en ondelettes stationnaires (SWT)⁵. Dans [111], les auteurs montrent que la SWT donne de meilleurs résultats pour l'extension de profondeur de champ appliquée à la cytologie digitale. Les expériences que nous avons menées sont en accord avec cette affirmation et ont permis d'identifier une méthode d'EDF basée sur la SWT avec des filtres biorthogonaux 3.3⁶,

[109]: HELMLI et al. (2001), « Adaptive shape from focus with an error estimation in light microscopy »

[110]: SHIHAVUDDIN et al. (2017), « Smooth 2D manifold extraction from 3D image stack »

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

[112]: FORSTER et al. (2004), « Extended depth-of-focus for multi-channel microscopy images »

[113]: LIU et al. (2017), « Multi-focus image fusion with a deep convolutional neural network »

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

3: Horizontale, verticale et diagonale.

4: Discrete Wavelet Transform.

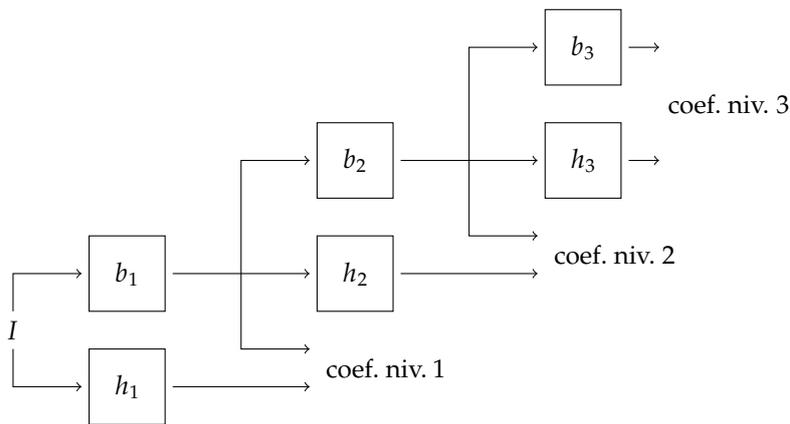
5: Stationary Wavelet Transform.

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

6: <http://wavelets.pybytes.com/wavelet/bior3.3/>

adaptée à nos images de cytologie urinaire.

La transformée en ondelettes stationnaires La transformée en ondelettes stationnaires⁷ fonctionne de façon similaire à la transformée en ondelettes discrètes : à chaque niveau de décomposition k , un filtre passe-bas b_k et une banque de filtre passe-haut h_k sont appliqués au signal. Le signal filtré obtenu pour un niveau de décomposition k correspond alors aux coefficients de niveau k de la SWT. Contrairement à la DWT où le signal filtré est sous-échantillonné par un facteur 2 à chaque niveau de décomposition, la SWT sur-échantillonne les filtres passe-bas et passe-haut utilisés dans la décomposition. Ainsi le filtre h_{k+1} est le filtre h_k sur-échantillonné par un facteur 2. Une illustration de la décomposition par SWT est proposée Figure 2.4.



7: Également connue sous le nom de transformée en ondelettes à trous.

FIGURE 2.4 – Transformée en ondelettes stationnaires (SWT) sur 3 niveaux appliquée à l'image I . À chaque niveau k , un filtre passe-bas b_k et une banque de filtres passe-haut h_k sont appliqués au signal. Le signal filtré par les filtres passe-haut et passe-bas du niveau k correspondent aux coefficients de niveau k de la transformée en ondelettes stationnaires.

En utilisant la SWT, le signal filtré n'est pas sous-échantillonné. Ainsi pour une image de dimensions (h, w) décomposée par SWT, pour chaque niveau de décomposition, les coefficients de détails (passe-haut) et d'approximation (passe-bas), sont également de dimensions (h, w) . Une illustration de la décomposition d'un amas de cellules urothéliales est proposée Figure 2.5.

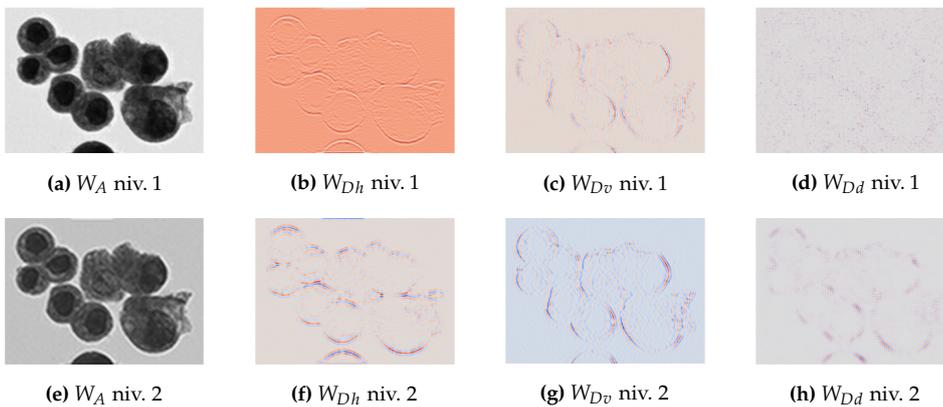


FIGURE 2.5 – Décomposition par transformée en ondelettes stationnaires (SWT) d'une image d'un amas de cellules urothéliales. Pour chaque niveau de décomposition, la SWT produit un coefficient d'approximation W_A et trois coefficients de détails W_{Dh} , W_{Dv} et W_{Dd} .

Règle de sélection des coefficients La règle de sélection des coefficients vise à produire une image EDF contenant un maximum de détails tout en évitant d'inclure trop de bruit, à partir de la transformée en ondelettes d'une z-stack. L'indice z des coefficients d'ondelettes sélectionnés définit une carte de niveau pour la projection EDF. La règle de sélection des coefficients proposée est décrite ci-dessous.

À chaque niveau de décomposition en ondelettes k , la fonction de transformation en ondelettes produit quatre sous-bandes : la sous-bande d'approximation W_A^k ainsi que

trois sous-bandes de détails $W_{D_h}^k$, $W_{D_v}^k$, $W_{D_d}^k$ respectivement pour les détails horizontaux, verticaux et diagonaux, comme décrit Équation 2.2.

$$\begin{aligned} W_A^k(n) &= \sum_{m=-\infty}^{+\infty} W_A^{k-1}(m) b_k(n-m) \\ W_D^k(n) &= \sum_{m=-\infty}^{+\infty} W_A^{k-1}(m) h_k(n-m) \end{aligned} \quad (2.2)$$

Dans la règle de sélection utilisée, pour chaque niveau de décomposition k , deux cartes de niveau LM_A et LM_D sont produites (voir équation 2.3⁸).

$$\begin{aligned} LM_A &= \operatorname{argmax}_z \operatorname{var}_{[3 \times 3]} \left(\frac{\delta^2 W_A}{\delta x^2} + \frac{\delta^2 W_A}{\delta y^2} \right) \\ LM_D &= \operatorname{argmax}_z |W_{D_h}| + |W_{D_v}| + |W_{D_d}| \end{aligned} \quad (2.3)$$

8: Les indices de niveau de décomposition k ont été omis pour simplifier la notation.

Avec :

- W_A : la sous-bande d'approximation de la SWT
- W_{D_h} : la sous-bande de détails horizontaux de la SWT
- W_{D_v} : la sous-bande de détails verticaux de la SWT
- W_{D_d} : la sous-bande de détails diagonaux de la SWT
- $\frac{\delta^2 \bullet}{\delta x^2} + \frac{\delta^2 \bullet}{\delta y^2}$: l'opérateur laplacien
- $\operatorname{var}_{[3 \times 3]}$: la variance calculée sur un voisinage de 3×3 pixels

LM_D tire parti des propriétés des coefficients des sous-bandes de détails de la transformée en ondelettes d'une z -stack, la plus grande valeur absolue le long de l'axe z correspondant au maximum de hautes fréquences et donc notamment aux détails. Pour éviter une reconstruction fortement bruitée⁹ ou altérée par des artefacts, il est important d'assurer la cohérence spatiale entre les sous-bandes de détails dans la règle de sélection des coefficients. Ainsi, en sélectionnant pour LM_D l'indice de la valeur maximale le long de z de la somme des valeurs absolues des trois sous-bandes de détails, la cohérence entre les sous-bandes de détails est assurée.

9: Les hautes fréquences sélectionnées peuvent aussi simplement correspondre à du bruit.

Pour la carte de niveau LM_A , la variance de l'opérateur Laplacien de la sous-bande d'approximation est calculée afin de récupérer les détails de fréquence plus faible que dans les sous-bandes de détails. Cela produit une carte de niveau plus lisse, une partie du bruit ayant été filtrée par le filtre passe-bas de la transformée en ondelettes.

Les deux cartes de niveau LM_A et LM_D calculées précédemment indiquent donc, pour chaque pixel (x, y) , la position selon l'axe z du maximum de détails à différentes fréquences. La carte de niveau LM_A , obtenue à partir de la sous-bande d'approximation de la décomposition en ondelettes d'une z -stack, peut-être assimilée à une version filtrée spatialement de la carte de niveau LM_D .

Pour assurer une certaine cohérence spatiale dans la règle de sélection des coefficients, les cartes de niveau LM_A et LM_D sont filtrées par un filtre médian 3×3 , moyennées et arrondies pour obtenir la carte de niveau finale, notée LM . Le résultat est une carte de coefficients finale utilisée pour sélectionner les coefficients dans la transformées en ondelettes d'une z -stack, comme illustré dans Figure 2.6.

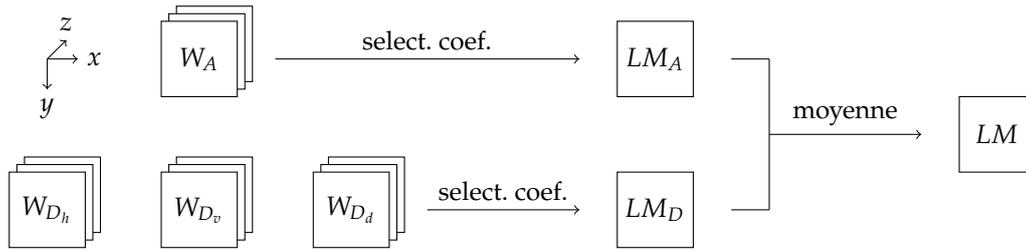


FIGURE 2.6 – Calcul de la carte de niveau LM permettant la sélection des coefficients de la transformée en ondelettes d'une z-stack. À chaque niveau de la transformée en ondelettes correspond trois z-stacks de coefficients, correspondant aux hautes fréquences horizontales (W_{D_h}), verticales (W_{D_v}) et diagonales (W_{D_d}), ainsi qu'une z-stack de coefficients correspondant aux basses fréquences complémentaires (W_A). Une première carte de niveau LM_A est calculée à l'aide du maxima de la variance de l'opérateur Laplacien appliqué aux coefficients basses fréquences W_A . Une seconde carte de niveau LM_D est calculée à partir des coefficients hautes fréquences et assure la cohérence entre les sous-bandes de détails. La carte de niveau finale LM est obtenue en moyennant les cartes de niveau obtenues précédemment.

La règle de sélection de coefficient définie ici peut-être appliquée pour un nombre arbitraire de niveau de décomposition par transformée en ondelettes, en moyennant les cartes de niveau obtenues pour chaque niveau. Une évaluation visuelle des projections EDF des z-stacks de cytologie urinaire suggère des performances similaires pour différents nombre de niveaux de décomposition utilisées¹⁰. Ainsi, afin de minimiser le temps de calcul des projections EDF, seul le premier niveau de décomposition en ondelettes est utilisé.

10: niveaux évalués : {1, 2, 3}

Pré-traitement : pré-sélection des plans focaux

L'acquisition automatisée de z-stacks, lorsque disponibles sur les scanners de lames commerciaux, ne prend généralement pas en compte la position individuelle des objets d'intérêt le long de l'axe z , ni l'épaisseur totale du matériel cellulaire. Ainsi, le nombre de plans focaux numérisés est généralement choisi suffisamment large afin que tout le matériel cellulaire soit imagé. Par conséquent, pour un objet d'intérêt donné, de nombreux plans focaux imagés peuvent ne pas contenir d'information utile, mais seulement un flou de défocus qui peut être à l'origine d'artefacts dans l'image EDF reconstruite. Pour palier à cette éventualité, une étape de pré-traitement est introduite afin d'éliminer les sections hors-focus.

Ainsi, une mesure de focus¹¹ est calculée pour chaque section de la z-stack et est utilisée comme un score global de netteté caractérisant la section. À partir de cette mesure, un algorithme de clustering K-Means à deux classes est appliqué pour classer chaque section d'une z-stack comme *en focus* ou *hors-focus*. En appliquant ce pré-traitement à 1032 z-stacks de cellules urothéliales basales, en moyenne 46 sections sur 201 sont conservées pour la reconstruction EDF.

11: Ici la variance de l'opérateur Laplacien définie Équation 2.1.

Stratégie de reconstruction de la couleur

La règle de sélection des coefficients est définie pour des z-stacks en niveaux de gris et doit être adaptée aux données en couleurs. En cytopathologie urinaire, la coloration est un des marqueurs biologiques important pour la détection du cancer de la vessie, décrit dans le Système de Paris [24]. De plus, pour l'analyse automatique de lames de cytologie digitales, des algorithmes entraînés tels que des modèles de deep learning sont appliqués aux images EDF pour des tâches de détection, segmentation et classification. Le processus de décision complexe de ces algorithmes peut entraîner des résultats imprévisibles lorsqu'ils sont appliqués à des images légèrement différentes des données d'entraînement [114]. Pour ces deux raisons, la méthode d'extension de profondeur de champ proposée vise à préserver une grande fidélité des couleurs lors de la projection.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[114]: MÅRTENSSON et al. (2020), « The reliability of a deep learning model in clinical out-of-distribution MRI data »

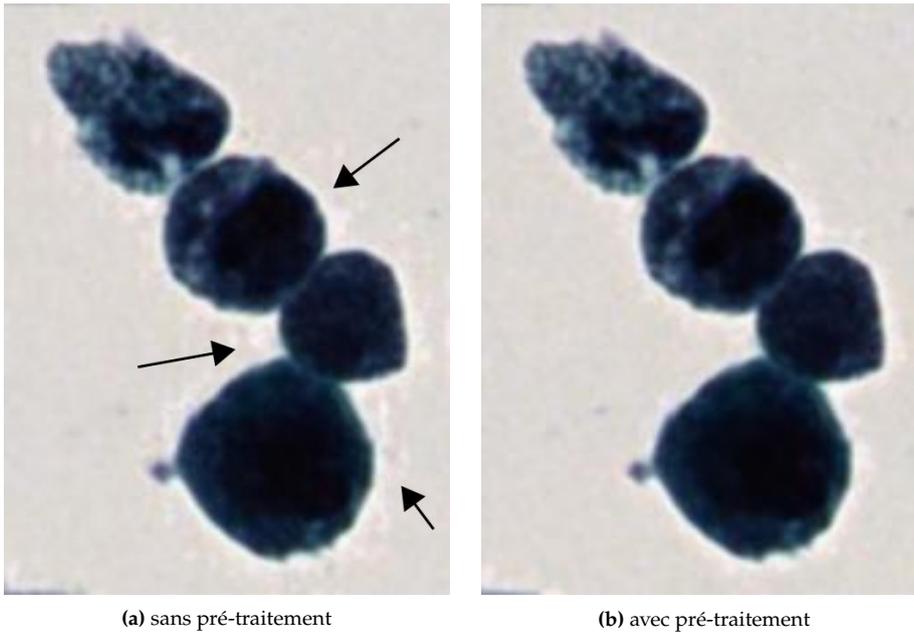


FIGURE 2.7 – Impact de la pré-sélection des plans focaux par clustering non-supervisé. Sans pré-traitement, plus d'artefacts sont visibles à proximité des bords des cellules. Une correction gamma a été appliquée aux images.

Une méthode naïve pour obtenir une image EDF en couleur consiste à appliquer la projection EDF à chaque canal de couleur de la z-stack, puis à concaténer les projections obtenues. Cependant, ce genre d'approche n'apporte aucune certitude quant à la fidélité couleur de la projection et peut produire de fausses couleurs. Forster et al. [112] ont proposé de calculer l'image EDF sur une z-stack en niveaux de gris correspondant à la première composante de la décomposition de la z-stack originale par analyse en composantes principales (PCA)¹². Ensuite, la projection obtenue est reconvertie en une image couleur en utilisant l'inverse de la PCA, puis une réassignation des couleurs est appliquée. Grâce à cette étape de réassignation, les couleurs obtenues dans l'image EDF finale appartiennent assurément à la z-stack originale, cependant cette méthode peut tout de même conduire à des imprécisions de couleur, comme illustré Figure 2.9 (c) et Tableau 2.2. C'est notamment le cas lorsque le nombre de sections considérées dans la z-stack est important. La réassignation peut alors se baser sur des voxels hors-focus ne correspondant pas aux couleurs originales de l'échantillon observé.

La méthode proposée introduit donc une étape supplémentaire de correction des couleurs afin d'améliorer la fidélité couleur de la reconstruction (voir Figure 2.8). La carte de niveau LM est utilisée pour sélectionner les voxels correspondants dans la z-stack originale et calculer une projection 2D en couleur notée C_{rgb} . L'image EDF et la projection C_{rgb} sont ensuite converties dans l'espace de couleur YUV. Dans cet espace couleur, le canal Y contient l'information de luminance¹³ tandis que les canaux U et V contiennent l'information de chrominance. La projection EDF finale est finalement la concaténation du canal Y de l'image EDF, et des canaux U et V de la projection C_{rgb} , combinant les détails fins de l'image EDF et les couleurs réelles du volume original.

Cette stratégie permet d'obtenir une meilleure fidélité des couleurs pour la projection EDF. Des projections EDF obtenues par différentes stratégies de reconstruction des couleurs sont illustrées Figure 2.9 (d).

[112]: FORSTER et al. (2004), « Extended depth-of-focus for multi-channel microscopy images »

12: *Principal Component Analysis*

13: Donc la plupart des informations de détails.

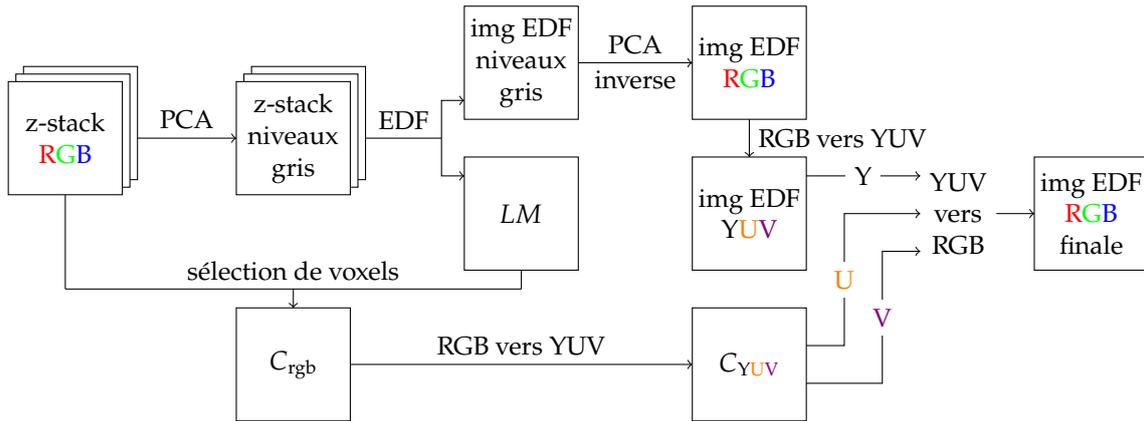


FIGURE 2.8 – Stratégie de reconstruction de la couleur pour la projection EDF. La z-stack originale est convertie en niveaux de gris en utilisant une analyse en composantes principales (PCA) avant de calculer la projection EDF. Cette dernière est projetée à nouveau en couleurs en utilisant la PCA inverse. La correction des couleurs est effectuée dans l'espace de couleurs YUV, par concaténation du canal Y de l'image EDF obtenue et des canaux U et V de l'image composite C_{rgb} calculée à partir de la z-stack originale et de la carte de niveau LM.

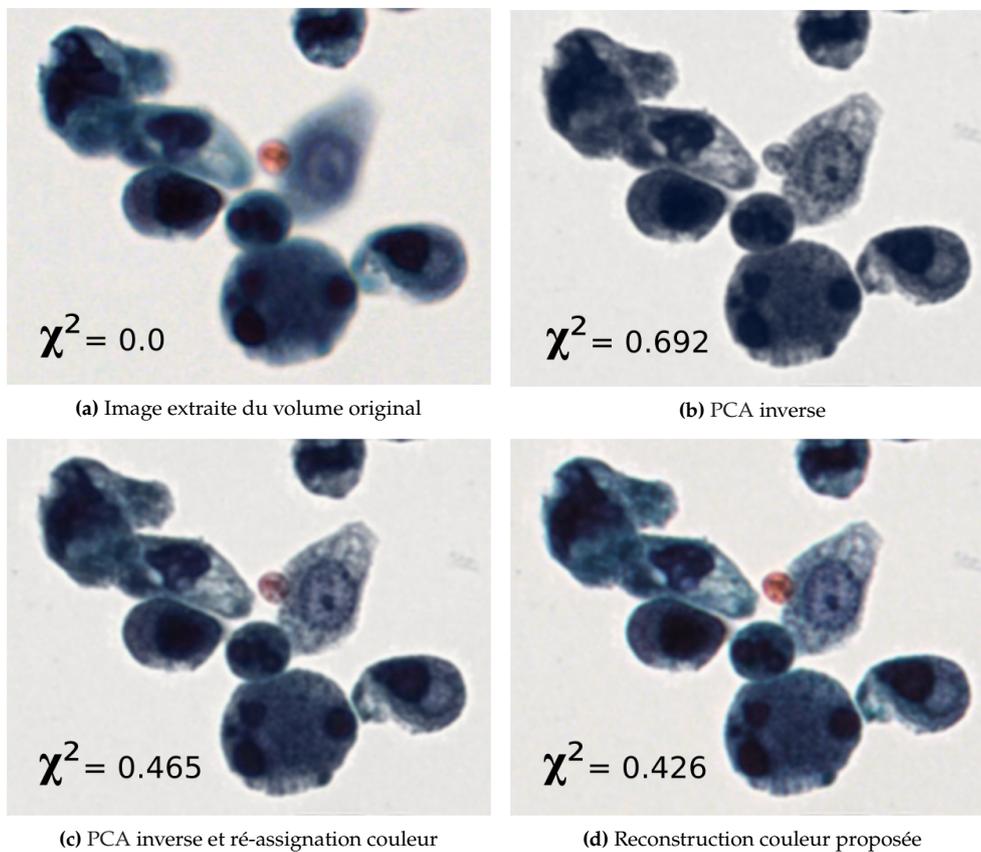


FIGURE 2.9 – Comparaison de la fidélité des couleurs de différentes méthodes de projection EDF par rapport à une section extraite de la z-stack originale. Les couleurs les plus rares sont perdues en utilisant uniquement la PCA inverse. La stratégie proposée, composée d'une PCA inverse suivie d'une fusion dans l'espace de couleurs YUV permet de reconstruire les couleurs originales avec une plus grande fidélité par rapport à la méthode PCA inverse suivie d'une simple réassignation des couleurs, décrite dans [112].

Évaluation de la méthode

Dans cette section, la méthode proposée, notée SWT-EDF, est comparée à trois algorithmes de fusion d'images multifocales : la transformée en ondelettes complexes (CWT-EDF) [112], la Smooth Manifold Extraction (SME) [110] et une méthode de fusion d'images basée sur un réseau de neurones convolutionnel (CNN) [113]. La projection à variance locale maximale [109] produit d'importants artefacts¹⁴, c'est pourquoi elle n'a pas été prise en compte pour la comparaison.

[112]: FORSTER et al. (2004), « Extended depth-of-focus for multi-channel microscopy images »

[110]: SHIHAVUDDIN et al. (2017), « Smooth 2D manifold extraction from 3D image stack »

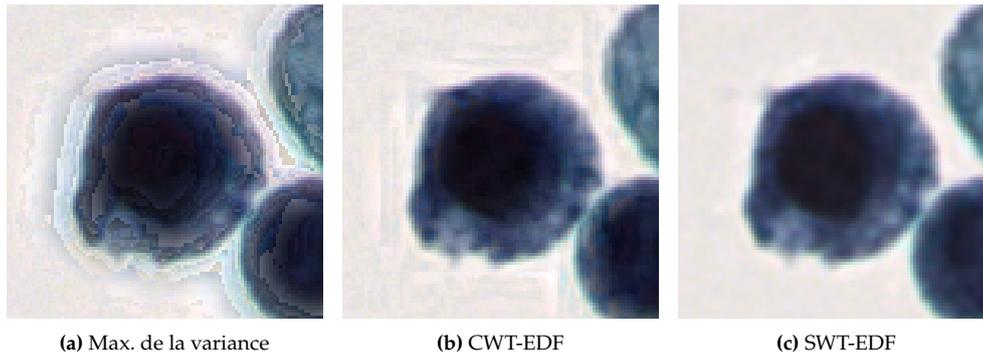
[113]: LIU et al. (2017), « Multi-focus image fusion with a deep convolutional neural network »

[109]: HELMLI et al. (2001), « Adaptive shape from focus with an error estimation in light microscopy »

14: Voir Figure 2.10.

Évaluation de la préservation des détails dans les projections Pour les différentes méthodes de projection EDF considérées, l'évaluation de la préservations des détails est principalement conduite visuellement¹⁵ à l'aide du jeu de données ISBI 2015 et d'images extraites de lames de cytologie urinaires. La compétition ISBI 2015 : *Segmentation of Overlapping Cervical Cells from Multi-layer Cytology Preparation Volumes* [115] fournit un jeu de données de 17 z-stacks de cytologie cervicale et des images EDF associées.

Comme avancé dans [111], la méthode SWT-EDF est visuellement plus performante que la méthode CWT-EDF, qui a tendance à produire plus d'artefacts, comme illustré dans la Figure 2.10.



15: Il n'existe pas de vérité terrain bien définie pour les projections EDF qui sont, par définition, des images composites.

[115]: ZHI LU et al. (2015), « An Improved Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells »

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

FIGURE 2.10 – Comparaison entre (a) la projection selon le maximum de la variance locale, (b) l'EDF par transformée en ondelettes complexes (CWT) et (c) l'EDF par transformée en ondelettes stationnaires (SWT). Les projections selon le maximum de la variance et la CWT-EDF ont tendance à produire plus d'artefacts que la SWT-EDF.

Une évaluation visuelle des projections sur le jeu de données ISBI 2015 montre également que la méthode SWT-EDF surpasse les méthodes SME et CNN-EDF. D'une part, l'algorithme SME suppose que la localisation selon l'axe z des objets d'intérêt peut être approximée par un manifold 2D dans la z-stack observée. D'autre part, la méthode CNN-EDF prédit une carte de fusion afin de sélectionner les voxels d'intérêt et d'effectuer ensuite une projection EDF. En raison de la transparence des images de cytologie, ces stratégies basées sur la prédiction de cartes de niveaux peuvent échouer à récupérer les détails localisés à différentes positions z pour la même position (x, y), comme illustré Figure 2.11.

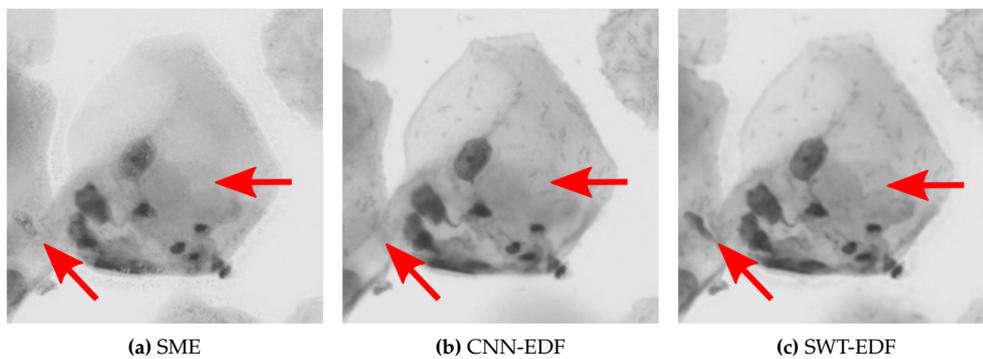


FIGURE 2.11 – Comparaison entre SME, CNN-EDF et SWT-EDF sur deux sous-images du jeu de données ISBI 2015. La superposition des cellules dans ce cas empêche les méthodes SME et CNN-EDF (a) et (b) de récupérer autant de détails (voir flèches) que la méthode SWT-EDF proposée (c).

Afin de quantifier les détails récupérés par les différentes méthodes de projection EDF considérée, pour chaque z-stack du jeu de données ISBI 2015, la similitude structurelle (SSIM) est calculée entre les images EDF fournies et les reconstructions EDF :

$$SSIM_{ab} = \frac{(2\mu_a\mu_b + c_1)(2\sigma_a\sigma_b + c_2)(cov_{ab} + c_3)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)(\sigma_a\sigma_b + c_3)} \quad (2.4)$$

Avec :

μ_a : intensité moyenne sur la fenêtre a.

μ_b : intensité moyenne sur la fenêtre b.

σ_a : écart-type de l'intensité sur la fenêtre a.

σ_b : écart-type de l'intensité sur la fenêtre b.

$c_1 c_2 c_3$: trois constantes pour la stabilité.

La SSIM moyenne et son écart-type, calculés entre les images EDF du jeu de données ISBI 2015 et les projections des méthodes évaluées, sont reportés Tableau 2.1.

SSIM	SME	CNN-EDF	SWT-EDF	CWT-EDF
moyenne	0.955	0.959	0.972	0.946
écart-type	0.009	0.004	0.003	0.004

TABLE 2.1 – Similitude structurelle (SSIM) calculée entre les images EDF du jeu de données ISBI 2015 et les projections des méthodes évaluées. La meilleure performance mesurée est reportée en bleu. Sauf mention contraire, cette convention est conservée dans la suite du document.

La méthode proposée, basée sur la SWT, produit des images EDF plus proches des images EDF de la compétition ISBI 2015¹⁶ que les autres méthodes.

Évaluation de la fidélité des couleurs La fidélité des couleurs de plusieurs méthodes est évaluée sur un ensemble de données de 20 images en couleurs de cellules urothéliales, extraites de lames de cytologie urinaire fournies par VitaDX. Les images EDF en couleur des méthodes évaluées sont comparées aux sections de meilleur focus S_{sharp} de la z-stack originale en utilisant la distance Chi-squared (χ^2), définie Équation 2.5, sur les histogrammes RGB normalisés :

$$\chi^2 = \sqrt{\sum_1^b \frac{(r_i - h_i)^2}{r_i}} \quad (2.5)$$

$$\chi_{rgb}^2 = \chi_r^2 + \chi_g^2 + \chi_b^2$$

Avec :

b : nombre de groupes, ou *bins*, dans l'histogramme.

r_i : valeur de l'histogramme de référence r pour le groupe i .

h_i : valeur de l'histogramme comparé h pour le groupe i .

Les résultats obtenus sont résumés dans le Tableau 2.2. Les méthodes SME et CNN-EDF sélectionnent les voxels focalisés dans le volume original et peuvent être considérées comme ayant une fidélité de couleur parfaite. Les méthodes SWT-EDF et CWT-EDF proposées utilisent une reconstruction des couleurs et, parmi elles, la méthode SWT-EDF préserve mieux la fidélité des couleurs (voir Figure 2.9).

χ^2	SME †	CNN-EDF †	SWT-EDF	CWT-EDF
moyenne	0.173	0.331	0.586	1.106
écart-type	0.07	0.11	0.29	0.39

TABLE 2.2 – Comparaison de la fidélité couleur entre les méthodes évaluées. † : Les méthodes SME et CNN-EDF sélectionnent les voxels focalisés dans le volume et ont une fidélité des couleurs parfaite. Parmi les méthodes basées sur les transformées en ondelettes avec couleurs reconstruites, la méthode proposée obtient la meilleure fidélité des couleurs.

16: Également calculées à partir de la transformée en ondelettes stationnaires.

2.3.2 L'extension de profondeur de champ pour la cytologie digitale

Comparaison entre S_{edf} et S_{sharp} Comme décrit précédemment, l'acquisition de plusieurs plans focaux est nécessaire en cytologie digitale afin d'imager tout le matériel cellulaire d'intérêt, situé à différentes profondeurs dans la lame. Ce paragraphe illustre certains avantages de l'extension de profondeur de champ par rapport à la sélection des sections de meilleur focus S_{sharp} pour certains types d'objets d'intérêt.

Parmi les objets d'intérêt présents dans la lame numérisée, certains sont inclinés et donc l'information focalisée se trouve dans différents plans focaux. Dans ces cas, une acquisition multi-plans suivie par la sélection du plan de meilleur focus ne permet pas de récupérer toute l'information nette, contrairement à une reconstruction EDF (voir Figure 2.12).

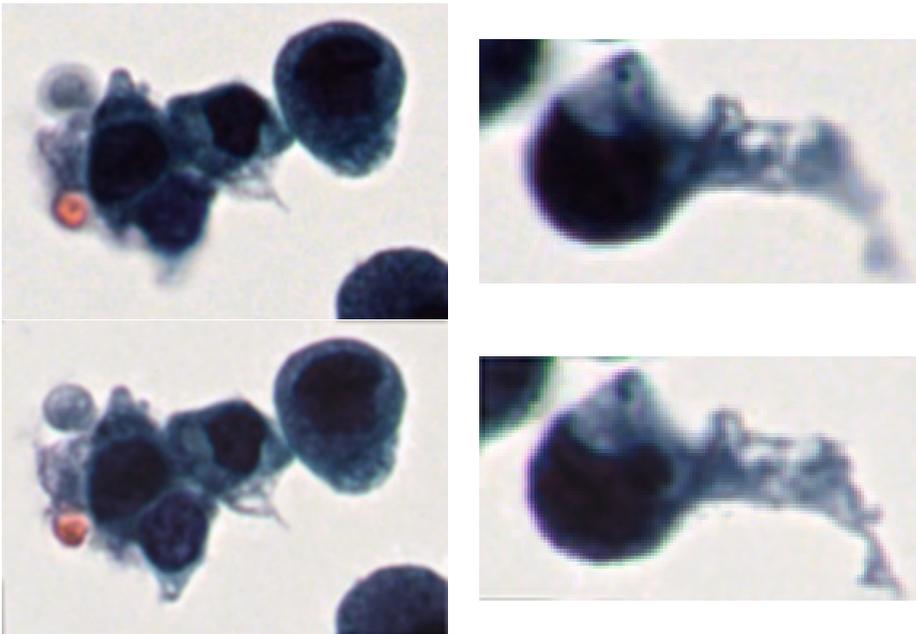


FIGURE 2.12 – Comparaison des sections de meilleur focus S_{sharp} (haut) et des projections EDF S_{edf} (bas) pour des objets cellulaires en focus sur plusieurs plans. Dans ces cas, la méthode d'EDF permet une reconstruction 2D où tout le matériel cellulaire apparaît net.

Les lames de cytologie contiennent également des amas de cellules, avec des cellules superposées. Bien que la sélection du plan de meilleur focus produise une image de bonne qualité générale pour ces objets, certaines cellules de l'amas sont de fait situées au-dessus ou au-dessous du plan de meilleur focus imagé. Pour ces cellules, la projection EDF permet d'obtenir une image contenant plus de détails et de texture que dans la section de meilleur focus, comme illustré Figure 2.13.

Dépendance de la projection EDF aux paramètres d'acquisition Le besoin d'acquisition multi-plans pour la cytologie digitale a été discuté dans la section 1.4. Ces différents paramètres d'acquisitions multi-plans, tels que le nombre de plans focaux à numériser et l'espacement inter-plans, peuvent altérer la qualité des projections EDF. Ce paragraphe étudie la dépendance de la projection EDF aux paramètres d'acquisition.

Un ensemble de 1032 z-stacks sont extraites d'une lame de cytologie scannée à l'aide du scanner de lames Hamamatsu NanoZoomer® S60, avec 201 plans focaux espacés de 0.1 μm . Un échantillonnage de ces z-stacks permet, à posteriori, de simuler plusieurs espacements inter-plans. 201 plans étant le nombre maximal de plans que le scanner utilisé peut numériser pour une lame et 0.1 μm étant le plus petit espacement inter-plans disponible, l'épaisseur des z-stacks sans sous-échantillonnage est de 20 μm [101]. Cette épaisseur, environ deux fois l'épaisseur moyenne des cellules, est maintenue constante dans cette expérience.

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

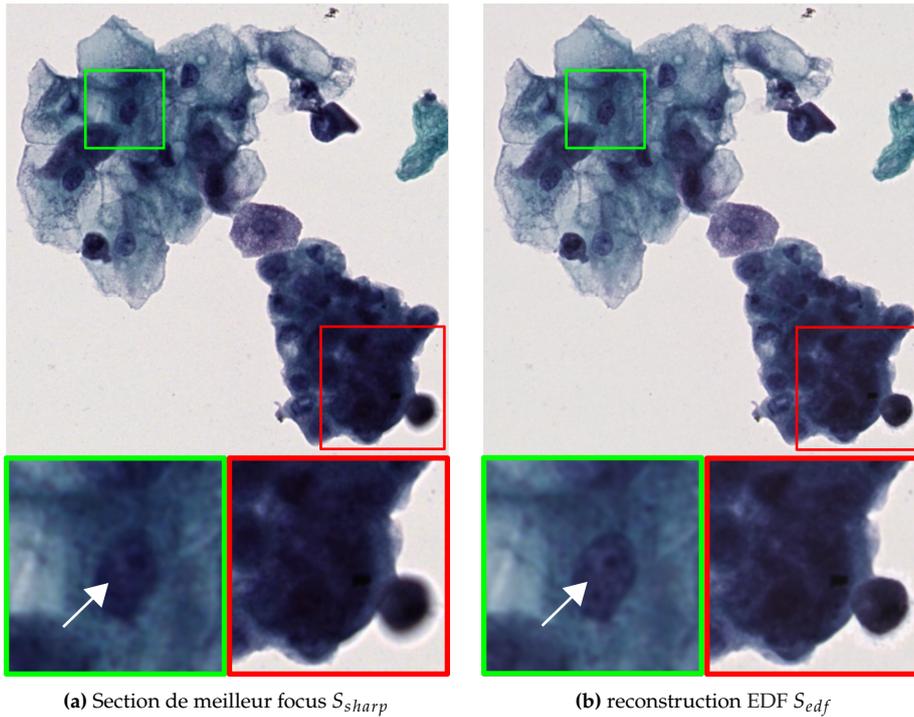


FIGURE 2.13 – Illustration des avantages de l’extension de profondeur de champ pour l’acquisition d’amas cellulaires. (a) est un amas cellulaire imagé au plan de meilleur focus, (b) est une projection EDF de la z-stack. Les sous-images illustrent comment l’image EDF préserve plus de détails (voir les flèches), et élimine le flou hors-focus.

Soit δ_z la distance entre deux plans focaux consécutifs, les z-stacks échantillonnées sont créées avec $\delta_z \in [0.1, 10]$ μm . Pour toutes ces z-stacks échantillonnées, la projection EDF est calculée en utilisant la méthode proposée et est comparée à l’image EDF calculée à partir de la z-stack originale¹⁷. La Figure 2.14 reporte la moyenne, sur tous les volumes de cytologie, de l’erreur quadratique moyenne entre ces comparaisons.

17: C’est-à-dire non échantillonnée.

On peut observer que l’erreur augmente presque linéairement avec δ_z , avant d’atteindre un plateau pour $\delta_z = 6.0$ μm . Après ce point, seule la section centrale du volume échantillonné contient de l’information, et appliquer une projection EDF équivaut à sélectionner la section centrale de la z-stack.

Pour une épaisseur donnée à numériser¹⁸, diminuer δ_z revient à augmenter le nombre n de plans focaux imagés. L’augmentation du nombre de plans focaux augmente également le temps d’acquisition. Le choix de la valeur optimale de δ_z se traduit donc par un compromis entre la qualité de la projection EDF et le nombre de plans focaux à numériser. La Figure 2.15 représente la moyenne, sur l’ensemble des z-stacks, de l’erreur quadratique moyenne entre l’image EDF calculée avec $n = 201$ et les images EDF calculées avec des valeurs décroissantes de n .

18: Ici 20 μm .

On peut remarquer que l’erreur ne diminue pas linéairement avec le nombre de plans acquis et pour plus de 21 plans imagés, correspondant à une distance inter-plans de 1.0 μm , l’erreur quadratique moyenne ne diminue que légèrement. Le scanner utilisé dans cette expérience a une profondeur de champ donnée par :

$$\text{DoF} = \frac{\lambda n}{\text{NA}^2} \quad (2.6)$$

Avec $\text{NA} = 0.75$, $n = 1$ et $\lambda \approx 0.6$ μm , la profondeur de champ du scanner est : $\text{DoF} \approx 1$ μm . On peut noter que la distance inter-plan optimale trouvée pour les données utilisées dans cette expérience correspond à la profondeur de champ du scanner.

Il n’est pas surprenant qu’un espacement inter-plans supérieur à la profondeur de champ du scanner entraîne une perte d’information pour les objets épais, et donc une projection

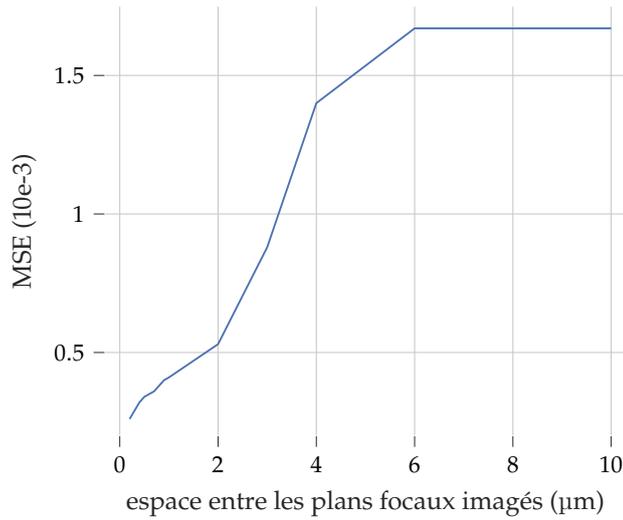


FIGURE 2.14 – Erreur quadratique moyenne entre les images EDF de z-stacks non échantillonnées et les images EDF de z-stacks sous-échantillonnées pour différents espacements inter-plans δ_z . L'erreur augmente linéairement avec le sous-échantillonnage, avant d'atteindre un plateau pour $\delta_z = 6.0 \mu\text{m}$. À ce point et pour un espacement plus élevé, toute l'information est située dans le plan central et l'application de la projection EDF est équivalente à la sélection de la section centrale du volume.

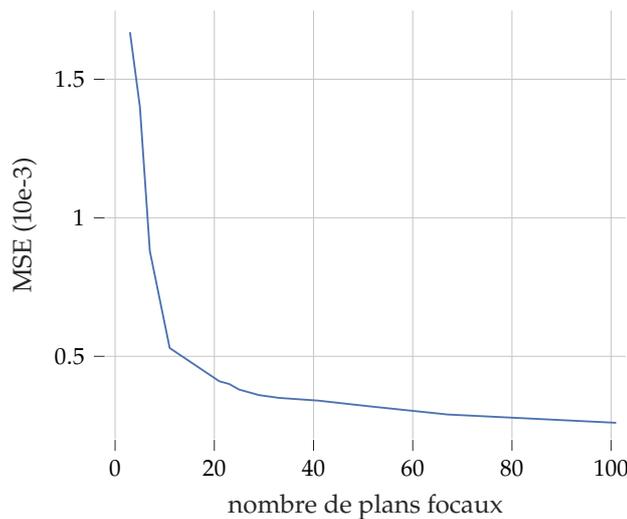


FIGURE 2.15 – Erreur quadratique moyenne entre l'image EDF de volumes non échantillonnés et les images EDF de z-stacks échantillonnées avec différents nombres de plans focaux. L'erreur diminue rapidement avec le nombre de plans focaux imagés, avant d'atteindre un plateau. Suivant cette expérience, un optimum suivant la méthode du coude est trouvé pour 21 plans acquis, correspondant à un espacement inter-plans de $1 \mu\text{m}$.

EDF de moindre qualité. Cependant, on peut observer qu'un espacement inter-plans δ_z plus faible améliore peu la qualité de la projection au prix d'un nombre élevé de plans focaux supplémentaires à numériser.

2.3.3 Extension de profondeur de champ pour le diagnostic automatique

Dans le contexte de la cytologie numérique automatisée, le Ratio Noyau-Cytoplasme (RNC) est une caractéristique importante pour le diagnostic du cancer de la vessie [24]. Le calcul du RNC par cellule repose sur une segmentation précise du cytoplasme et du noyau.

Pour analyser des échantillons cellulaires épais ou composés d'objets multiples situés à différentes profondeurs z , une stratégie courante consiste à utiliser la section de meilleur focus S_{sharp} à partir d'une acquisition multi-plans [101]. Avec des paramètres d'acquisition appropriés, cette méthode est efficace pour analyser des échantillons cellulaires qui se trouvent dans un même plan focal, et dont l'épaisseur ne dépasse pas sensiblement la profondeur de champ du scanner de lames utilisé. Cependant, pour l'analyse d'objets cellulaires épais ou situés sur plusieurs plans focaux, nous allons

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

montrer que l'analyse de l'ensemble du volume avec des méthodes d'extension de profondeur de champ est nécessaire.

Paramètres de l'expérience Une expérience est proposée pour quantifier l'épaisseur d'un échantillon cellulaire à partir de laquelle la sélection de la section de meilleur focus n'est plus suffisante pour assurer une segmentation précise des cytoplasmes et noyaux. La segmentation est réalisée à l'aide d'un modèle U-Net [16]. Son architecture entièrement convolutionnelle permet l'entraînement et l'inférence à partir d'images de différentes tailles. Le modèle est entraîné sur un ensemble de 825 images S_{sharp} de cellules urothéliales, annotées manuellement et réparties en 577 échantillons d'entraînement (Train), 124 échantillons de développement (Dev) ¹⁹ et 124 échantillons de Test. Soit Y_{true} la vérité terrain et Y_{pred} le masque de segmentation prédit, les performances du modèle appris sont évaluées en utilisant la métrique Intersection sur Union (IoU) :

$$IoU = \frac{Y_{true} \cap Y_{pred}}{Y_{true} + Y_{pred} - Y_{true} \cap Y_{pred}} \quad (2.7)$$

Après entraînement ²⁰, le modèle obtient un score IoU moyen de 0.905 pour les 2 classes, cytoplasme et noyau, sur l'ensemble de Test.

Des z-stacks synthétiques contenant un échantillon cellulaire d'épaisseur croissante ont été créés pour réaliser cette expérience :

Soit $S_{sharp} = S(z_{sharp})$ la section de meilleur focus pour une cellule donnée, les autres sections composant la z-stack synthétique sont calculées à partir de S_{sharp} comme suit :

$$S(z) = S(z_{sharp}) * G_{\sigma(z-z_{sharp})} + B \quad (2.8)$$

Avec :

G_{σ} : noyau gaussien d'écart type σ .

B : bruit blanc gaussien additif.

$*$: l'opérateur de convolution.

Les valeurs σ caractérisant le noyau gaussien G_{σ} , dépendent de la distance $(z - z_{sharp})$ et sont définies pour $z \in [0, 200] \setminus \{z_{sharp}\}$ par deux fonctions polynomiales du second ordre. Ces deux polynômes sont définis respectivement pour des sections au-dessus et en-dessous de la section S_{sharp} afin de minimiser la différence entre les gradients d'une section numérisée $S(z)$ et d'une section synthétique correspondante $S(z_{sharp}) * G_{\sigma(z-z_{sharp})}$:

$$\sigma(z - z_{sharp}) = \underset{\sigma}{\operatorname{argmin}} \|\nabla [S(z)] - \nabla [S(z_{sharp}) * G_{\sigma(z-z_{sharp})}]\| \quad (2.9)$$

Avec ∇ l'opérateur de gradient.

Une fois connus les polynômes paramétrisant $\sigma(z - z_{sharp})$ pour $z \in [0, 200] \setminus \{z_{sharp}\}$, il est possible de calculer une z-stack synthétique à partir d'une image unique.

Deux images contenant chacune une cellule urothéliale basale sont utilisées pour créer un ensemble de données de z-stacks synthétiques. Une z-stack synthétique est d'abord créée pour chaque image de cellule, avec $z_{sharp} = i$ et $z_{sharp} = j$ respectivement. La fusion des z-stacks obtenus crée la z-stack synthétique finale, contenant deux cellules dont les sections de meilleur focus sont séparées d'une distance $|i - j|$ le long de l'axe z . Les

[16]: RONNEBERGER et al. (2015), « U-Net »

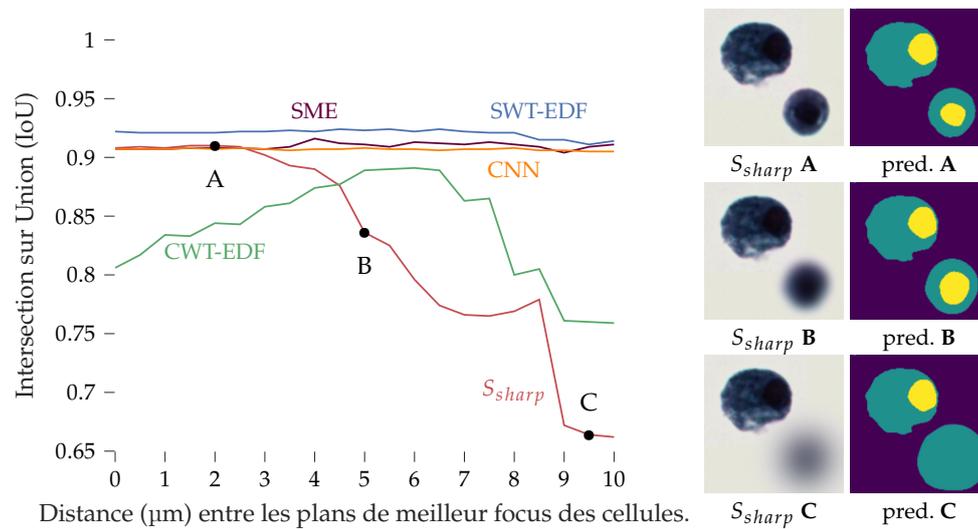
19: Cet ensemble Dev permet de fixer les hyper-paramètres d'entraînement du modèle.

20: Entraînement pendant 150 epochs avec un learning rate de $1e-4$ et une taille de batch de 1, permettant l'apprentissage sur des images de tailles variables.

paramètres i et j permettent de créer une séquence de z-stacks synthétiques contenant deux cellules séparées de distances variables le long de l'axe z , et ainsi d'évaluer les performances d'un modèle U-Net sur un échantillon cellulaire d'épaisseur variable.

Segmentation de matériel cellulaire d'épaisseur simulée croissante Selon l'analyse de Fan et al. [101], l'épaisseur médiane des amas de cellules dans les lames de cytologie cervicales est de $10\ \mu\text{m}$. L'expérience suivante compare donc les performances de segmentation du modèle U-net défini précédemment, appliqué à la section S_{sharp} et aux projections CWT-EDF, SME, CNN-EDF et SWT-EDF d'échantillons cellulaires synthétiques dont l'épaisseur varie de $0\ \mu\text{m}$ à $10\ \mu\text{m}$. Un total de 21 z-stacks synthétiques, chacune composée de 201 sections optiques ²¹, sont créés en augmentant la distance entre i et j de 0 à $10\ \mu\text{m}$ avec un pas de $0.5\ \mu\text{m}$.

Pour chaque z-stack synthétique, la projection par chaque méthode évaluée est ensuite segmentée à l'aide du modèle U-Net entraîné précédemment. La Figure 2.16 illustre les différents scores IoU obtenus, pour différentes épaisseurs simulées de l'échantillon cellulaire.



Cette expérience montre une nette baisse des performances de segmentation du modèle U-Net en utilisant la section S_{sharp} pour un échantillon cellulaire d'épaisseur supérieure à $5\ \mu\text{m}$. Toutes les méthodes EDF évaluées permettent une amélioration de la segmentation par rapport à l'utilisation de la section de meilleur focus S_{sharp} . Il est à noter que parmi les méthodes EDF, la méthode CWT-EDF présente les moins bonnes performances. Cela peut-être dû à la présence d'artefacts ou à des imprécisions de couleur dans la projection, comme on peut l'observer dans Tableau 2.1 et Tableau 2.2, ce qui souligne l'importance d'une projection à haute fidélité de couleur comme celle présentée dans cette section.

2.4 Conclusion

Dans ce chapitre, nous avons définis trois représentations 2D de l'information 3D présente dans les z-stacks. La section centrale de référence S_{centre} qui est équivalente à une simple acquisition 2D, la section correspondant au plan de meilleur focus S_{sharp} ainsi qu'une projection par extension de profondeur de champ S_{edf} . La méthode d'EDF proposée

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

21: Espacées de $100\ \text{nm}$.

FIGURE 2.16 – À gauche : le score IoU des prédictions du modèle U-Net en fonction de l'épaisseur simulée de l'échantillon cellulaire. À droite : les masques prédits par le modèle U-Net à partir de la section S_{sharp} . Pour de faibles épaisseurs, les performances de segmentation à partir des images SWT-EDF, SME, CNN-EDF et de la section S_{sharp} sont similaires (A). Avec l'augmentation de l'épaisseur simulée, deux chutes de performances sont visibles pour les prédictions basées sur la section S_{sharp} , (B) lorsque une des deux cellules devient floues et le modèle ne parvient plus à segmenter précisément le noyau, et (C) lorsqu'un des deux noyaux n'est plus détecté par le modèle. De plus l'IoU sur les projections CWT-EDF est plus faible que sur les images SWT-EDF, SME ou CNN-EDF, probablement à cause d'artefacts et d'un manque de fidélité des couleurs.

dans ce chapitre est adaptée à l'analyse automatique de lames de cytologie digitales colorées. Elle s'appuie sur une approche originale pour obtenir une reconstruction précise de la couleur grâce à la combinaison des informations de détails présentes dans la projection EDF par transformée en ondelettes de la z-stack et des informations de couleur présentes dans les voxels originaux de la z-stack. Les résultats expérimentaux montrent une reconstruction d'image très précise, adaptée aux images de cytologie transparentes avec une haute fidélité couleur. Les avantages du traitement du volume par EDF sur les performances de la segmentation sont également analysés pour différentes épaisseurs de matériel cellulaire. Cette expérience montre l'importance de la fidélité des couleurs sur les tâches de segmentation et d'analyse automatiques.

Préparation et acquisition de lames digitales multi-plans

3

Il a été montré section 1.2 que la pathologie digitale a de nombreuses applications, dont le diagnostic qui peut être posé par un anatomopathologiste observant la lame sur un écran d'ordinateur en remplacement de l'analyse traditionnelle de lames de cytologie ou d'histologie au travers d'un microscope. Dans le cas particulier des lames de cytologies, de par la présence d'amas cellulaires tri-dimensionnels, une numérisation multi-plans est nécessaire (voir section 1.4) pour imager l'ensemble de l'information présente dans la lame.

L'apparence des lames de cytologies digitales multi-plans dépend de différents paramètres de préparation et d'acquisition, et peut avoir des conséquences sur leur analyse [104]. Ce chapitre étudie donc l'impact de ces paramètres de préparation et d'acquisition sur l'analyse automatique de lames de cytologie digitales.

Plus précisément, l'objectif de ce chapitre est de définir un protocole de préparation et de numérisation de lames de cytologie adapté à l'analyse automatique d'images pour la détection du cancer de la vessie. Pour cela, le protocole de préparation de lames physiques doit permettre la visualisation de biomarqueurs caractéristiques de carcinomes urothéliaux, comme l'aspect de la chromatine des cellules urothéliales basales [24], et une bonne reproductibilité des préparations, afin de limiter les variations de colorations auxquels les algorithmes d'analyse d'images ont tendance à être sensible [116]. Le protocole de numérisation de lames doit permettre l'acquisition du maximum d'information pertinente pour le diagnostic à partir de lames digitales, tout en restreignant au nécessaire le nombre de plans focaux imagés pour minimiser le temps d'acquisition et l'espace de stockage nécessaire.

3.1 Paramètres considérés

Les paramètres variables qui influent sur l'apparence finale de la lame digitale comprennent les paramètres de préparation de la lame, qui jouent un rôle dans l'apparence intrinsèque de la lame physique, ainsi que les paramètres d'acquisition qui définissent comment le scanner de lames numérise la lame de microscopie.

3.1.1 Paramètres de préparation de lames physiques

Les lames de cytologie urinaire sont créées par dépôt de cellules présentes dans un échantillon d'urine sur une lame de microscopie, puis sont colorées avant d'être analysées. Cette création de lames physiques est nécessaire pour la cytologie conventionnelle ainsi que pour la cytologie numérique. Comme préconisé par le Système de Paris [24], une coloration de type Papanicolaou [3] est utilisée pour colorer les lames de cytologie urinaire. Afin de normaliser l'étape de création de lames, et ainsi améliorer la répétabilité du protocole de création de lames, des automates sont utilisés pour le dépôt de l'échantillon et la coloration. Dans cette étude, les lames de cytologie ont été systématiquement créées à l'aide du processeur Hologic ThinPrep® 5000 et de la coloreuse Agilent Dako CoverStainer®.

[104]: DONNELLY et al. (2013), « Optimal z-axis scanning parameters for gynecologic cytology specimens »

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[116]: TELLEZ et al. (2019), « Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology »

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

Il existe plusieurs variations de la coloration de Papanicolaou. Une méthode est donc proposée section 3.2 afin d'évaluer plusieurs colorations de Papanicolaou et de sélectionner celle qui facilite la visualisation de biomarqueurs dans les lames de cytologie urinaire.

3.1.2 Paramètres de numérisation

Propres à la cytologie digitale, les paramètres d'acquisition multi-plans définissent comment un scanner numérise une lame de cytologie selon plusieurs plans focaux. En particulier, l'acquisition multi-plans de lames de cytologie est paramétrisée principalement par le nombre de plans à numériser N_z et à l'espacement entre ces plans δ_z .

L'utilisation de scanners de lames commerciaux, marqués CE pour le diagnostic in-vitro, afin de numériser les lames de cytologie est une contrainte fixée de ce projet. Les technologies disponibles sur ces scanners de lames définissent un ensemble de valeur possibles à priori pour N_z et δ_z . En particulier, le scanner Hamamatsu NanoZoomer® S360 utilisé dans ce travail peut imager jusqu'à $N_z = 201$ plans focaux, espacés d'au moins $\delta_z = 0.1 \mu\text{m}$. Ce scanner image par défaut un plan focal $S_{z=z_0}$, globalement de meilleur focus pour la lame, puis numérise $(N_z - 1)/2$ plans au dessus et $(N_z - 1)/2$ plans en dessous du plan $S_{z=z_0}$. Chaque plan focal imagé est espacé d'une distance selon l'axe z paramétrée par une unique valeur δ_z .

Un protocole de numérisation adapté au diagnostic assisté par ordinateur doit permettre de maximiser les performances d'algorithmes de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales. Le développement de tels algorithmes nécessitant l'acquisition au préalable de lames digitales, il est difficile de conditionner la numérisation des lames de cytologie aux performances de ces algorithmes d'analyse. Ainsi, la section 3.3 cherche plutôt à déterminer les paramètres d'acquisition N_z et δ_z permettant de maximiser l'information totale présente dans la lame digitale, tout en gardant N_z suffisamment faible afin de répondre aux contraintes de stockage et de manipulation des lames digitales.

3.2 Optimisation du protocole de coloration

Publication associée

Les travaux conduits dans cette section ont été décrits dans un article de journal, accepté pour publication à Cytometry part A. :

A. Bouyssoux, K. Jarnouen, L. Lallement, R. Fezzani, and J.-C. Olivo-Marin, "Automated Staining Analysis In Digital Cytopathology And Applications," in Cytometry A, 2022.

Les progrès de la pathologie digitale au cours des dernières années ont conduit au développement d'algorithmes d'analyse d'images permettant notamment d'effectuer automatiquement des mesures quantitatives de biomarqueurs. Avec ces nouvelles méthodes vient le besoin d'une haute qualité de coloration et de reproductibilité, car les outils d'analyse d'images sont généralement plus sensibles aux légères variations de coloration que les pathologistes [116]. Cette section présente une méthode d'analyse automatisée de la coloration de lames de cytologie spécifiquement adaptée aux défis rencontrés en cytopathologie digitale. Les difficultés rencontrées sont notamment liées à la présence d'une variété de types cellulaires dans les lames de cytologie,

[116]: TELLEZ et al. (2019), « Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology »

à la distribution tri-dimensionnelle du matériel cellulaire, à la présence de cellules superposées et à la nécessité d'une analyse indépendante des noyaux et des cytoplasmes sont abordées. L'approche proposée est appliquée à la sélection d'un protocole de coloration permettant de maximiser les détails visibles dans le noyau des cellules, afin de faciliter la caractérisation de la chromatine, un biomarqueur important dans le diagnostic du cancer de la vessie [24].

3.2.1 Objectifs

De nombreux facteurs peuvent être à l'origine de variations dans l'aspect des lames de pathologie digitales, notamment au moment du dépôt et de la coloration des échantillons cellulaires. Le scanner de lames utilisé pour la numérisation, ainsi que les paramètres d'acquisition peuvent également être source de variations. L'évaluation de la qualité des lames digitales par un observateur humain est une tâche longue et fastidieuse, qui peut souffrir d'un manque d'objectivité et de mesures quantitatives. Pour résoudre ce problème, cette section présente une méthode d'analyse et de comparaison automatique des lames de cytologie digitales, en termes d'intensité, de coloration et de texture.

3.2.2 État de l'art

Contrôle qualité : détection automatique des régions floues Plusieurs études se sont intéressées à la détection automatisée ou semi-automatisée de lames digitales hors-focus ou partiellement hors-focus. Dans [117], les auteurs présentent une méthode permettant de générer une carte de netteté pour des lames d'histologie digitales, en calculant une mesure de gradient sur des tuiles extraites de la lame. La carte de netteté obtenue est ensuite retournée à l'utilisateur pour une inspection visuelle. Plus récemment, une étude menée par Kohlberger et al. [118] présente un algorithme basé sur un réseau de neurones convolutionnel (CNN) permettant de prédire une carte des régions hors-focus pour des lames d'histologie à partir de données synthétiques. L'étude illustre l'importance du contrôle de la qualité en montrant une diminution des performances d'un algorithme de détection des métastases du cancer du sein lorsque le flou de défocus augmente. Dans [119], les auteurs proposent une méthode permettant d'identifier les lames de cytologie digitales floues à partir de 200 cellules de la lame, tirées au hasard. Une approche similaire est proposée dans [105], où les auteurs proposent de détecter les cellules hors-focus à partir de noyaux segmentés. Les cellules hors-focus ainsi détectées sont ensuite retirées de l'analyse de la lame.

Évaluation des variations causées par le choix du scanner de lames Comme indiqué précédemment, le choix du scanner de lames peut influencer sur l'aspect des lames digitales. Des études ont comparé l'impact des dispositifs de numérisation sur l'analyse de lames d'histologie et de cytologie digitales par des anatomopathologistes. Dans [120], les auteurs présentent une méthode permettant de comparer des lames d'histologie digitales numérisées avec différents scanners. L'utilisateur définit une région d'intérêt pour laquelle des métriques¹ sont calculées et permettent une comparaison en terme de netteté et de quantité de détails. Dans [121], les auteurs évaluent le choix du scanner de lames en demandant à des anatomopathologistes d'évaluer leur niveau confiance en leurs diagnostic pour des lames de cytologie digitales numérisées par différents scanners de lames.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[117]: ZERBE et al. (2011), « Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images »

[118]: KOHLBERGER et al. (2019), « Whole-slide image focus quality »

[119]: LAHRMANN et al. (2013), « Semantic Focusing Allows Fully Automated Single-Layer Slide Scanning of Cervical Cytology Slides »

[105]: FRIEDRICH et al. (2016), « Removing defocused objects from single focal plane scans of cytological slides »

[120]: WALKOWSKI et al. (2011), « Quality evaluation of virtual slides using methods based on comparing common image areas »

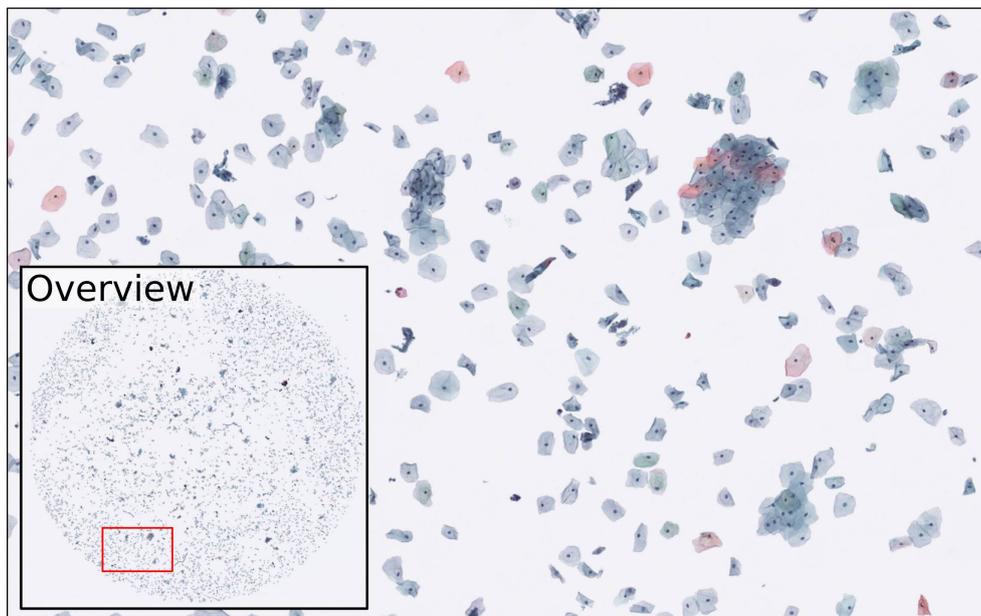
1: Les auteurs proposent les métriques d'Haralick [57] basées sur la matrice de co-occurrence des niveaux de gris.

[121]: HANNA et al. (2017), « Comparison of glass slides and various digital-slide modalities for cytology screening and interpretation »

Évaluation de la coloration pour les lames d’histologie digitales Des études récentes ont proposé des méthodes quantitatives pour l’évaluation de la coloration et de la préparation des lames d’histologie numériques. Dans [122], Eycke et al. présentent une méthode pour évaluer les variations de colorations immunohistochimique (IHC) en se basant sur une mesure dérivée de la Densité Optique du tissu. Cette méthode est ensuite utilisée afin d’évaluer la variabilité de la coloration causée par différents lots de réactifs. Dans [123], les auteurs étudient l’impact de variations du protocole de coloration et de l’utilisation de lots de réactifs périmés sur la Densité Optique moyenne de lames d’histologie digitales. L’étude conclut que l’analyse d’images de lames d’histologie basée sur la Densité Optique est une méthode rapide et efficace pour améliorer le contrôle qualité dans les laboratoires. Enfin, dans [124] les auteurs présentent une méthode automatique de contrôle qualité permettant d’accepter ou de rejeter les lames d’histologie digitales au moment de leur acquisition. Pour différentes sous-images extraites d’une lame digitale, des mesures de netteté, de contraste, de luminosité et de couleur sont calculées et comparées à des seuils prédéfinis.

La méthode proposée et décrite ci-après diffère des travaux précédents à plusieurs égards :

- Elle est spécifiquement adaptée aux lames de cytologie digitales (voir Figure 3.1). Ainsi, l’analyse d’image n’est pas effectuée sur de larges sous-images extraites des lames digitales, comme c’est généralement le cas pour les lames d’histologie, mais plutôt sur des régions cellulaires d’intérêt préalablement segmentées : le cytoplasme et le noyau.
- Bien que l’évaluation du flou de défocus soit une problématique importante en pathologie digitale, elle a été largement étudiée dans des travaux précédents [117][118]. Dans ce travail, l’analyse de cellules hors-focus est évitée à l’aide d’une acquisition multi-plans² associée à un algorithme d’extension de profondeur de champ (EDF). Cela permet d’éviter des variations dans les mesures de couleurs, d’intensité et de texture qui ne seraient pas liées à la préparation de l’échantillon mais plutôt à la présence de matériel cellulaire hors-focus.



[122]: EYCKE (2017), « Image processing in digital pathology : an opportunity to solve inter-batch variability of immunohistochemical staining »

[123]: CHLIPALA et al. (2020), « Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision »

[124]: AMEISEN et al. (2013), « Stack or trash ? »

[117]: ZERBE et al. (2011), « Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images »

[118]: KOHLBERGER et al. (2019), « Whole-slide image focus quality »

2: Aussi appelée z-stack.

FIGURE 3.1 – Extrait d’une lame digitale de cytologie urinaire colorée par un coloration de Papanicolaou. La lame contient différents types d’objets cellulaires : des cellules urothéliales basales, intermédiaires et superficielles, des globules rouges, des cellules neutrophiles, des amas de cellules et des débris.

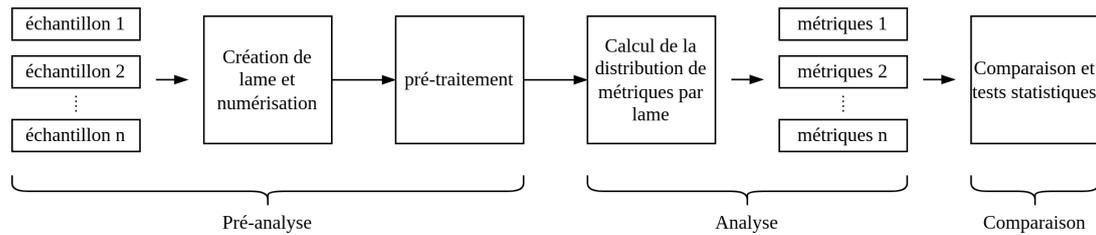


FIGURE 3.2 – Processus de création, numérisation et analyse de lames.

3.2.3 Méthode

Contrairement aux lames d’histologie qui sont analysées à partir de sous-images extraites des régions contenant du tissu, l’analyse de lames de cytologie se concentre sur des cellules isolées. Par conséquent, une étape de détection et de classification des cellules est nécessaire avant l’analyse des lames de cytologie. Comme les régions du noyau et du cytoplasme sont ciblées par différents réactifs de coloration [3], une segmentation de ces régions cellulaires est effectuée pour les cellules pertinentes, permettant le calcul de métriques par région.

De plus, alors que l’épaisseur du tissu dans les lames d’histologie est généralement comprise entre 4 μm et 6 μm , les lames de cytologie peuvent contenir des cellules d’intérêt dans un volume allant jusqu’à 30 μm d’épaisseur [125]. Pour quantifier précisément la coloration des cellules, il est essentiel que toutes les cellules soient en focus. Ceci est assuré par une numérisation multi-plans de la lame et l’utilisation d’un prétraitement par extension de profondeur de champ ³.

Le processus de création, numérisation et analyse de lames considéré dans cette étude est le suivant : les échantillons d’urine sont fixés sur une lame de microscopie, colorés et numérisés. Ensuite, une pré-analyse est effectuée afin de détecter les cellules d’intérêt, ainsi que pour prédire les masques de segmentation des noyaux et cytoplasmes. Puis, les distributions de métriques par région (noyau et cytoplasme) sont calculées pour chaque lame. Enfin, à partir des distributions des métriques extraites, les lames digitales peuvent être comparées quantitativement entre elles. Un aperçu du processus d’analyse est proposé dans Figure 3.2, et chaque étape est décrite en détail dans la suite.

Création de lames et numérisation

Les échantillons d’urine sont collectés auprès de donateurs volontaires en bonne santé. Pour les expériences de coloration, toutes les lames utilisées pour un test sont préparées avec le même échantillon d’urine d’un seul donneur volontaire : ainsi aucune variation indésirable n’est introduite à cette étape. Les lames sont systématiquement créées à l’aide d’un processeur Hologic ThinPrep® 5000 et colorées à l’aide du Agilent Dako CoverStainer® avec une coloration de Papanicolaou [3].

Les lames sont numérisées à l’aide du scanner de lames Hamamatsu Nanozoomer® S360, selon 21 plans focaux espacés de 1 μm et centrés sur le plan de meilleur focus déterminé par l’autofocus du scanner. Les lames digitales sont enregistrées au format Hamamatsu NDPI, en s’appuyant sur la compression JPEG avec une qualité de 80 %. Les deux étapes de préparation et de numérisation des lames, illustrées dans Figure 3.3, peuvent entraîner des variations dans l’aspect final des lames digitales.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

[125]: LEE et al. (2011), « Evaluation and optimization for liquid-based preparation cytology in whole slide imaging »

3: Voir section 2.3 pour le détail de la méthode EDF utilisée.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

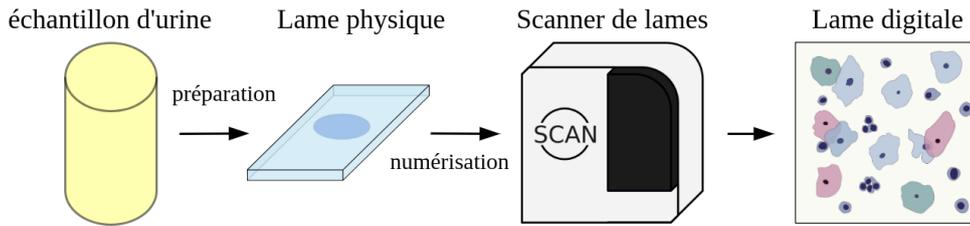


FIGURE 3.3 – Création de lames et numérisation

Pré-traitement

Le pré-traitement appliqué est le suivant : les objets de premier plan sont d'abord extraits de la lame, puis classés en fonction de leur type cellulaire. Un algorithme d'extension de profondeur de champ est appliqué aux cellules d'intérêt pour prévenir un éventuel flou de dé-focus. Les cellules d'intérêt sont ensuite segmentées en trois régions cellulaires comprenant le noyau, le cytoplasme et l'arrière-plan. Enfin, les cellules sont regroupées par défaut en deux groupes en fonction de leur couleur dominante. Chaque étape est décrite en détail dans la suite et résumée dans Figure 3.4.

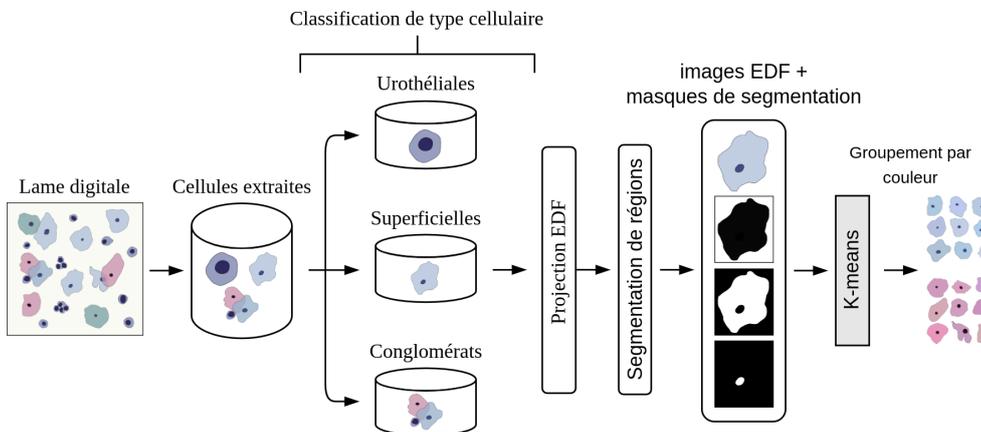


FIGURE 3.4 – Étapes de pré-traitement, comprenant l'extraction et la classification de cellules d'intérêt, la projection par extension de profondeur de champ, la segmentation de régions cellulaires d'intérêt et le groupement non-supervisé de cellules en fonction de leur couleur.

Détection d'objets et classification de type cellulaire Tout d'abord, toutes les composantes connexes de premier plan de la diapositive sont extraites grossièrement comme étant l'ensemble des pixels connectés dont la distance à l'arrière-plan le long d'au moins un canal de couleur est supérieure à un seuil. Ce seuil n'est pas très sensible et peut être déterminé automatiquement⁴ ou fixé à une valeur prédéfinie. Dans ce travail, un pixel de 8 bits est considéré comme un pixel de premier plan si sa distance à l'arrière-plan est supérieure à 14 (valeur déterminée empiriquement) pour au moins un canal de couleur.

La classification des objets extraits d'une lame, en fonction de leur type cellulaire, est ensuite effectuée à l'aide d'un modèle ResNet18 [30]. Cinq types cellulaires sont alors considérés : les cellules urothéliales basales, les cellules urothéliales superficielles, les conglomérats⁵, les cellules neutrophiles et les artefacts⁶. Cette étape de classification permet de comparer des groupes de cellules appartenant à la même classe, et d'éviter les variations potentielles entre les lames causées par une distribution différente des types cellulaires. Les expériences présentées ci-après reposent sur l'analyse des cellules urothéliales superficielles, car elles sont présentes en grand nombre contrairement aux cellules urothéliales basales ou aux neutrophiles. Les conglomérats doivent également être étudiés avec soin, car les cellules superposées introduisent des variations d'intensité et de couleur qui ne sont pas liées à la coloration.

4: Par exemple en utilisant la méthode Otsu [72].

[30]: He et al. (2015), « Deep Residual Learning for Image Recognition »

5: Définis comme un ensemble d'au moins deux objets cellulaires superposés ou en contact.

6: Définis comme l'ensemble des objets cellulaires qui n'appartiennent pas aux quatre catégories citées précédemment.

Traitement du matériel cellulaire hors-focus Une projection par extension de profondeur de champ est calculée pour toutes les z-stacks centrées sur les cellules sélectionnées, produisant une image composite où la cellule apparaît complètement en focus, à l'aide de la méthode décrite section 2.3 qui préserve la fidélité des couleurs. Ainsi, l'analyse de la coloration et de la texture des lames n'est pas affectée par la présence de cellules complètement ou partiellement hors-focus.

Segmentation des noyaux et cytoplasmes Un modèle U-Net [16] est entraîné sur des images EDF pour extraire les masques de segmentation de trois régions cellulaires : noyau, cytoplasme et arrière-plan⁷. Les masques du cytoplasme et du noyau permettent ainsi le calcul de métriques sur différentes régions cellulaires, généralement colorées avec différents réactifs comme dans la coloration de Papanicolaou [3].

Le modèle U-Net, développé avec Pytorch [126], est entraîné pendant 150 à 190 epochs avec un learning rate de 1^{-4} , sur un ensemble de données d'entraînement comprenant 5398 images uniques. L'optimisation est stoppée lorsque aucune amélioration n'est observée pendant 20 epochs sur un ensemble de données de développement composé de 1349 images uniques.

Cette étape de segmentation est cruciale dans la méthode d'analyse proposée, car une mauvaise segmentation conduira à des métriques imprécises. Il est donc primordial que le modèle U-Net produise une segmentation fiable pour différentes colorations. La robustesse du modèle face aux différentes colorations est obtenue en utilisant également de fortes augmentations de couleurs au moment de l'entraînement. Plus précisément, les échantillons d'entraînement sont modifiés de manière aléatoire en faisant varier la luminosité, le contraste, la teinte et la saturation des couleurs afin de simuler une grande variété de colorations. Plusieurs augmentations de données classiques sont également appliquées aléatoirement aux données d'entraînement lors de l'apprentissage, notamment des rotations de 90 degrés, des inversions horizontales et verticales, un recadrage aléatoire associée à une complétion par des zéros si nécessaire afin de produire des images de 256 par 256 pixels.

L'ensemble des paramètres utilisés pour l'augmentation des données d'entraînement est décrit Tableau 3.1 et des illustrations d'augmentations sont proposées Figure 3.5.

Transformation	Paramètres
rotation	[0, 90, 180, 270] degrés
inversions	[horizontale, verticale]
recadrage aléatoire et complétion	dim=(256,256)
variation de la luminosité	[0.75, 1.25]
variation du contraste	[0.75, 1.25]
variation de la saturation	[0.75, 1.25]
variation de la teinte	[0, 1]

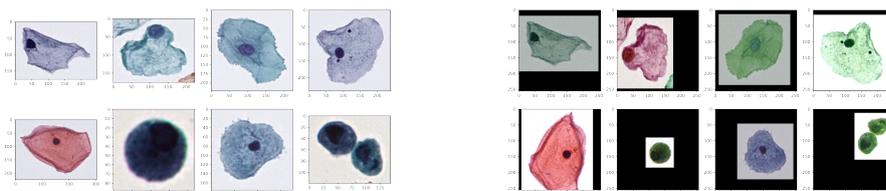
[16]: RONNEBERGER et al. (2015), « U-Net »

7: Le masque de segmentation de l'arrière-plan est utile pour le calcul de métriques normalisées localement.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

[126]: PASZKE et al. (2019), « PyTorch : An Imperative Style, High-Performance Deep Learning Library »

TABLE 3.1 – Paramètres des augmentations de données utilisées pendant l'entraînement du modèle de segmentation.



(a) Exemple d'images d'entraînement avant augmentation de données.

(b) Exemple d'images d'entraînement après augmentation de données.

FIGURE 3.5 – Illustration des augmentations de données utilisées pour l'entraînement du modèle de segmentation.

Groupement non-supervisé de cellules En cytopathologie, les cellules sont colorées pour mettre en évidence les biomarqueurs, et les différents types de cellules ont tendance à être colorés différemment. Par exemple, dans le cas de la coloration de Papanicolaou [3], les cellules sont souvent colorées par les colorants Hématoxyline, Orange G et Éosine. Ces réactifs colorent les cellules urothéliales intermédiaires et superficielles en nuances de bleu ou de rouge en fonction de leur concentration intracellulaire en kératine.

Afin de comparer la coloration de type Papanicolaou des lames de cytologie, sans biais lié à la distribution des types de cellules, les cellules d'intérêt sont classées à priori en deux groupes : les cellules majoritairement bleues et majoritairement rouges. Afin de s'assurer que la séparation des cellules est correctement effectuée pour différentes colorations, la classification est considérée comme un problème de clustering non-supervisé, utilisant le modèle K-Means pour regrouper les cellules sur la base de leurs valeurs médianes de chroma A et B dans l'espace couleur CIE-La*b*. Des exemples de groupes de cellules obtenus par cette méthode sont présentés dans Figure 3.6.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

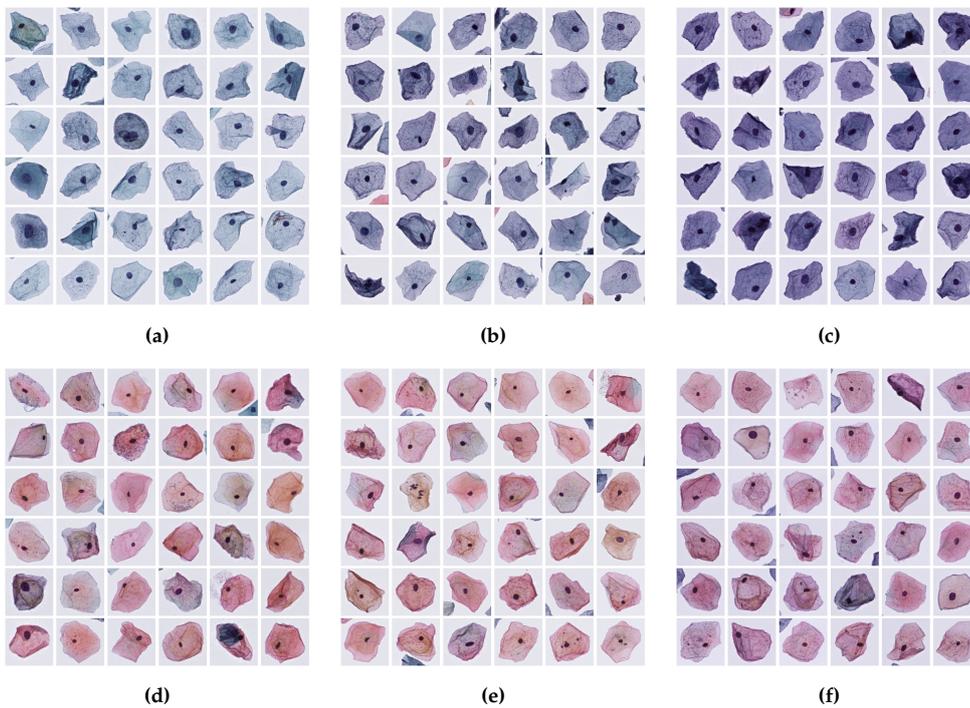


FIGURE 3.6 – Groupes de cellules majoritairement bleues (a), (b), (c) et rouges (d), (e), (f) automatiquement groupées à l'aide du modèle K-Means pour trois colorations de Papanicolaou différentes.

Métriques

Les métriques F_r décrites ci-après sont définies pour différentes régions cellulaires r ⁸. Dans la suite, la métrique F est annotée F_b lorsque calculée pour l'arrière-plan, F_c lorsque calculée pour le cytoplasme et F_n lorsque calculée pour le noyau. Soit M_r le masque booléen de segmentation d'une région r , et \bar{F}_r la valeur médiane de la métrique F_r sur ce masque M_r comme définie Équation 3.1.

$$\bar{F}_r = \text{médiane} (F(x, y) \mid (x, y) \in M_r) \quad (3.1)$$

Métriques de luminance Soit $I(x, y)$ l'intensité de l'image à la position (x, y) . Pour des images en couleur, on considère la valeur de canal L dans l'espace de couleur CIE-La*b*. Les métriques d'intensité utilisées sont définies Tableau 3.2 :

8: Les trois régions cellulaires considérées sont le noyau, le cytoplasme et l'arrière-plan.

Métrique	Formule
Luminance	$L_r = \bar{I}_r$
Transmission	$T_r = L_{r r \in [c,n]} / L_b$
Densité Optique	$OD_r = -\log_{10} T_r$

TABLE 3.2 – Métriques d'intensité.

La transmission définie Tableau 3.2 est une normalisation locale de la luminance de la région cellulaire par rapport à l'arrière-plan, permettant de corriger les éventuelles non-uniformités de ce dernier. La Densité Optique est une mesure de l'absorbance directement dérivée de la transmission, utile pour les méthodes de déconvolution des couleurs qui a été utilisée dans des études précédentes [127][123] pour la comparaison de l'intensité entre les lames d'histologie digitales.

Métriques de couleur Les métriques de couleur sont définies dans l'espace couleur CIE-La*b*. Soit $a^*(x, y)$ et $b^*(x, y)$ les coordonnées chromatiques a^* et b^* de l'image à la position (x, y) dans l'espace CIE-La*b*. Les métriques de couleurs calculées sont reportées Tableau 3.3.

Métrique	Formule
Chroma A	$A_r = \overline{a^*_{r r \in [c,n]}} - \overline{a^*_b}$
Chroma B	$B_r = \overline{b^*_{r r \in [c,n]}} - \overline{b^*_b}$

[127]: ROY et al. (2018), « A study about color normalization methods for histopathology images »

[123]: CHLIPALA et al. (2020), « Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision »

TABLE 3.3 – Métriques de couleur. Les métriques de couleur utilisées pour les régions du cytoplasme et du noyau (A & B) sont normalisées par rapport à l'arrière-plan local pour tenir compte des variations inter-lames et intra-lames de la couleur de l'arrière-plan.

Comme les métriques d'intensité définies précédemment, les métriques de couleur A et B sont normalisées par rapport à l'arrière-plan. L'arrière-plan de la lame ne contient, par définition, pas de matériel cellulaire. Il est donc attendu que la chrominance de cette région soit un niveau de gris. Cependant, des variations de couleurs sont observées au niveau de l'arrière-plan et peuvent être dû à un rinçage insuffisant des réactifs lors de la coloration, ou à une imprécision de la balance des blancs du scanner de lames. Pour tenir compte de ces variations, les valeurs médianes de chrominance des cellules sont normalisées en soustrayant les valeurs médianes de la chrominance de l'arrière-plan local. Des illustrations de cellules avant et après normalisation des couleurs sont visibles Figure 3.7.

Métriques de texture Les métriques de texture suivantes sont calculées sur des images extraites des lames digitales et centrées sur les cellules isolées d'intérêt.

Métriques basées sur les bords : la texture est quantifiée à l'aide de la métrique de Tenengrad (voir Tableau 3.4), basée sur ∇I_v et ∇I_h respectivement les filtres de Sobel verticaux et horizontaux appliqués sur le canal de luminance de l'image.

Caractéristiques de Haralick : la texture est également évaluée à l'aide des caractéristiques de Haralick [57] basées sur la matrice de co-occurrence des niveaux de gris (GLCM) P du canal de luminance d'une image. Le calcul de cette GLCM est défini dans la suite. Soit I une image avec N niveaux de gris. La GLCM non-normalisée, pour cette image I , est une matrice carrée de taille $N \times N$ où le (i, j) ^{ième} élément correspond au nombre d'occurrences où un pixel d'intensité i est adjacent⁹ à un pixel d'intensité j . La GLCM peut être construite de manière symétrique par l'incrémentación simultanée des éléments (i, j) et (j, i) pour deux pixels adjacents d'intensité i et j dans un ordre quelconque. Deux pixels sont adjacents s'ils sont situés à une distance d et un angle a , d et a étant les

[57]: HARALICK et al. (1973), « Textural Features for Image Classification »

9: Selon une distance d et un angle a .

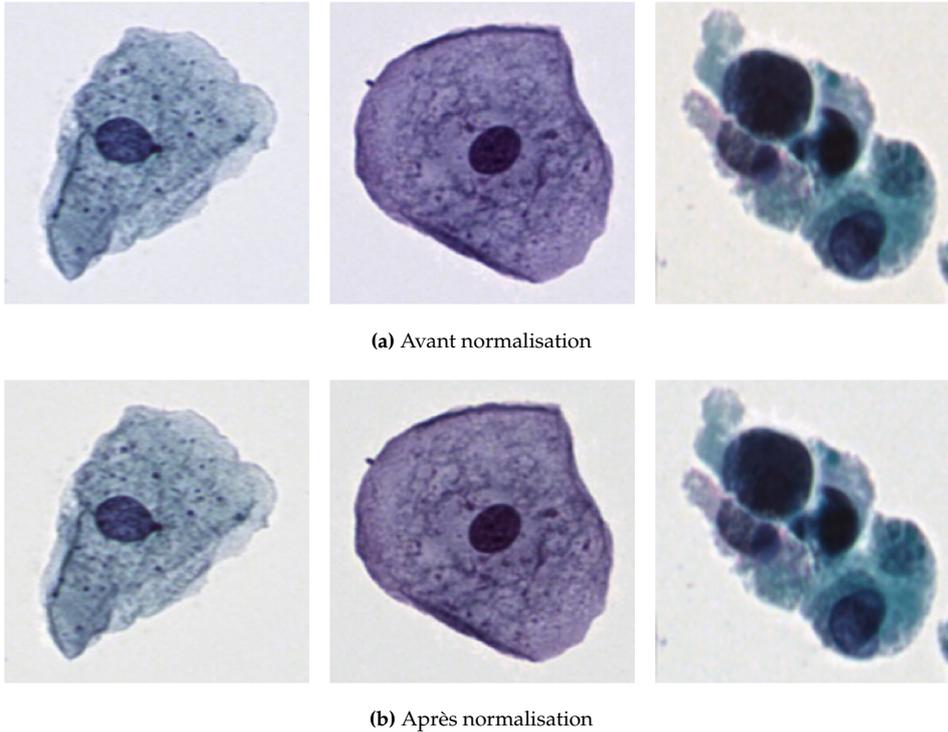


FIGURE 3.7 – Images de cellules avant et après normalisation basée sur la chrominance moyenne de l’arrière-plan local.

paramètres définissant la GLCM. Dans l’analyse proposée, la GLCM doit être calculée sur une région cellulaire d’intérêt uniquement, définie par un masque de segmentation M_r . L’image I considérée a jusqu’à $N = 256$ valeurs correspondant à 8 bits. Ces valeurs $v \in [0, 255]$ sont arbitrairement fixées entre 1 et 256, et la valeur de tous les pixels en dehors de la région d’intérêt sont fixés à 0, produisant une nouvelle image avec $N = 257$ niveaux de gris. Quatre GLCM symétriques de dimensions 257×257 , avec $d = 5$ aux angles $a \in [0, \pi/4, \pi/2, 3\pi/4]$ sont calculées et moyennées pour introduire une certaine invariance par rotation. Par construction, toute co-occurrence avec des pixels situés en dehors de la région d’intérêt contribuera aux entrées de la première ligne et de la première colonne de la GLCM calculée¹⁰. La première ligne et la première colonne de la GLCM sont donc supprimées pour obtenir la matrice P qui est ensuite normalisée de façon à ce que la somme de tous ses éléments soit égale à 1. Ainsi, lorsque de la construction de la GLCM, seul les pixels appartenant à M_r ont été pris en compte. A partir de la GLCM obtenue $P_{i,j}$, quatre mesures de texture sont calculées : l’Entropie, le Second Moment Angulaire (ASM), l’Énergie et l’Homogénéité (voir Tableau 3.4).

*Motifs Binaires Locaux (LBP)*¹¹ : la texture est également caractérisée à l’aide des histogrammes de motifs binaires locaux, définis Tableau 3.4.

3.2.4 Distribution des métriques par lame

Le calcul des distributions de métriques par lame de cytologie digitale est décrit dans cette section et résumé dans la Figure 3.8. La distribution des métriques d’intensité, de texture et de couleurs est calculée à partir des cellules de chaque lame digitale. Ces distributions de métriques permettent ensuite de comparer les colorations entre différentes lames. Une variation du protocole de coloration peut influencer sur la texture du noyau, qui est un des biomarqueurs importants en cytopathologie. En effet, une chromatine nucléaire irrégulière est signe d’atypie pour certains cancers [24], et peut être difficile à détecter lorsque le noyau est sur-coloré ou sous-coloré. La comparaison

10: Correspondant alors aux co-occurrences avec les pixels de valeurs 0 qui appartiennent par construction à l’arrière-plan.

11: En anglais : *Local Binary Patterns*.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

Métrique	Formule
Tenengrad	$Tg_r = \frac{\sum_{(x,y) \in M_r} \nabla I_v(x,y)^2 + \nabla I_h(x,y)^2}{\sum_{(x,y) \in M_r} 1}$
	$\text{normed} Tg_r = \frac{Tg_r}{Tg_b}$
Entropie	$S = - \sum_i \sum_j P(i, j) \log(P(i, j))$
ASM	$ASM = \sum_i \sum_j P(i, j)^2$
Energie	$E = \sqrt{ASM}$
Homogénéité	$H = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i, j)$
LBP	$LBP(P, R)_r = \text{hist}_{(x_c, y_c) \in M_r} \sum_{p=0}^{P-1} \mathbb{1}_{\mathbb{R}^+}(I_c - I_p)$

TABLE 3.4 – Métriques de texture.

de texture entre lames digitales se concentre principalement sur la comparaison des distributions de texture des noyaux des cellules d'intérêt.

Une fois les distributions des métriques calculées pour chaque lame de cytologie digitale, une comparaison inter-lames peut être effectuée. Pour une métrique F , une lame est décrite par la moyenne de F pour toutes les cellules contenues dans celle-ci. Les expériences menées dans la suite ont été conduites avec trois lames¹² pour chaque variable testée. Pour les expériences à plusieurs variables, les différences observées entre groupes de lames sont jugées significatives ou non à l'aide du test statistique ANOVA, corrigé pour les comparaisons multiples à l'aide de l'analyse statistique de Tukey¹³ [128]. Dans le cas d'expériences à deux variables, le t -test de Student [129] est utilisé. Le seuil de confiance α est fixé à 0.05, ce qui signifie que les valeurs- p ajustées inférieures à 0.05 permettent de rejeter l'hypothèse nulle, et de conclure que les différences observées sont significatives.

12: Correspondant à trois répliques techniques.

13: Aussi appelée Tukey HSD
[128]: TUKEY (1949), « Comparing Individual Means in the Analysis of Variance »
[129]: STUDENT (1908), « The Probable Error of a Mean »

3.2.5 Analyse de l'impact de la solution d'hématoxyline

L'impact de la solution d'hématoxyline sur la coloration des lames de cytologie urinaire est étudié en utilisant trois solutions d'hématoxyline différentes : *Hématoxyline Agilent Dako (CS700)*, *Hématoxyline Sigma-Aldrich Gill1 (GHS132-1L)* et *Hématoxyline Agilent Harris (CS709)*. Les solutions d'hématoxyline peuvent être utilisées de façon progressive ou régressive. La technique progressive, généralement adaptée aux solutions peu concentrées, consiste simplement à immerger l'échantillon dans la solution d'hématoxyline jusqu'à atteindre la coloration désirée. La technique régressive, au contraire, sur-colore l'échantillon dans un premier temps, puis une solution de différenciation est utilisée afin de décolorer l'échantillon et ainsi atteindre la coloration désirée. Parmi les solutions d'hématoxyline considérées dans ce travail, les solutions *Dako* et *Gill1* sont recommandées pour une utilisation progressive. La solution d'Hématoxyline de *Harris CS709* peut être utilisée, au choix, de façon progressive ou régressive. Dans le cas d'une utilisation régressive, deux immersions supplémentaires de l'échantillon sont effectuées : dans une solution de différenciation¹⁴ puis dans un bac contenant de l'eau du robinet. Ainsi, quatre groupes de lames sont considérés au total. Des exemples de cellules colorées avec les différentes solutions d'hématoxyline sont proposées Figure 3.9.

14: Ici la solution de différenciation utilisée est la solution *Agilent CS707*.

Analyse de la transmission Les transmissions des noyaux et cytoplasmes des cellules urothéliales superficielles sont comparées pour des lames de cytologie digitales colorées à l'aide des différentes solutions d'hématoxyline considérées. Les résultats obtenus sont

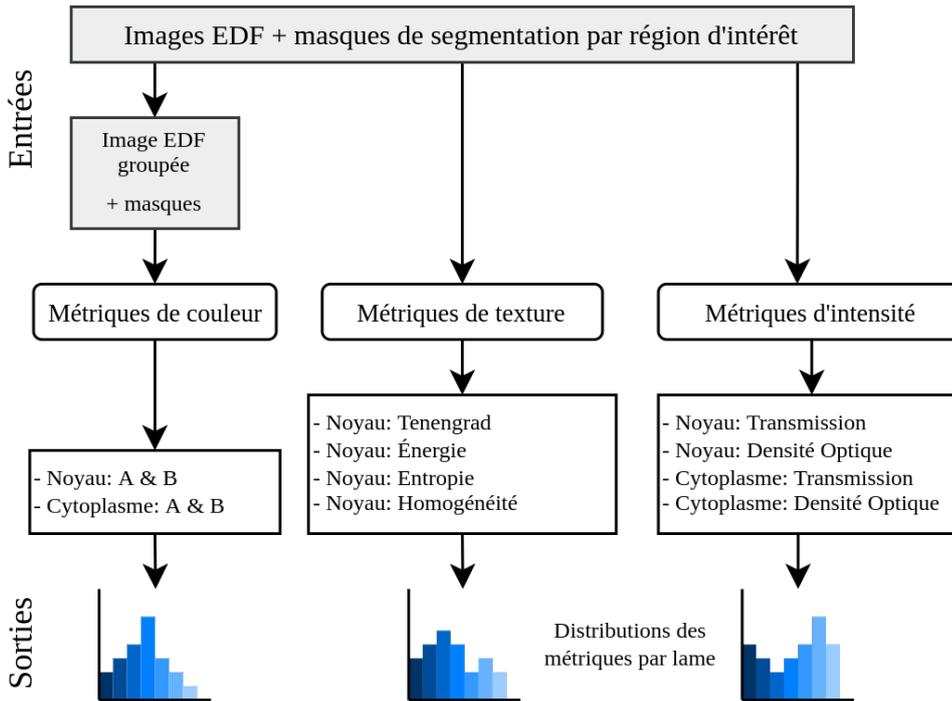


FIGURE 3.8 – Calcul des distributions des métriques par lame de cytologie digitale.

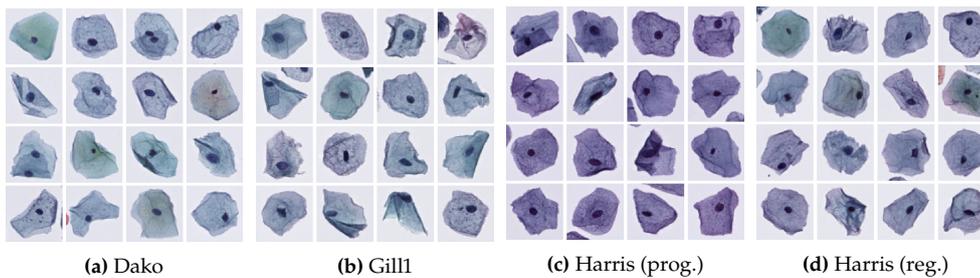


FIGURE 3.9 – Exemples de cellules colorées à l'aide de différentes solutions d'Hémathoxyline : (a) Agilent Dako CS700, (b) Sigma-Aldrich Gill1 GHS132-1L, (c) et (d) Agilent Harris CS709 respectivement utilisée de façon progressive et régressive en ajoutant une solution de différenciation Agilent CS707.

résumés Tableau 3.5. La transmission mesurée au niveau du cytoplasme et du noyau est plus élevée lorsque les cellules sont colorées avec les solutions d'hémathoxyline *Dako* et *Gill1*, un résultat attendu puisque ces solutions sont moins concentrées en hémathoxyline que les solutions de *Harris*.

Analyse de la texture La quantité de détails intra-nucléaires visibles pour différentes solutions d'hémathoxyline est quantifiée à l'aide de métriques de texture calculées sur les noyaux des cellules urothéliales superficielles. La texture des noyaux est ainsi comparée pour toutes les solutions d'hémathoxyline testées et les résultats obtenus sont résumés Tableau 3.6.

Plus une région d'intérêt contient des détails de hautes fréquences, plus la mesure d'Entropie S de cette région est élevée et la mesure d'Énergie E faible. L'expérience est menée en utilisant le même scanner de lames ainsi qu'un unique échantillon d'urine, afin de minimiser la présence de hautes fréquences correspondant au bruit lié à la préparation ou à l'acquisition des lames de cytologie digitales. Ainsi, une différence dans la quantité des détails de hautes fréquences observés au niveau des noyaux correspond à une différence dans la quantité de détails nucléaires visibles. L'expérience montre que la coloration basée sur la solution d'hémathoxyline *Gill1* tend à produire un peu plus de détails nucléaires que d'autres solutions d'hémathoxyline testées. Notamment, l'utilisation de la solution d'hémathoxyline de *Harris* (*progressive*) produit un coloration

Solution	valeur moyenne T_n	valeurs-p			
		<i>Dako</i>	<i>Gill1</i>	<i>Harris (prog.)</i>	<i>Harris (reg.)</i>
<i>Dako</i>	0.239	1	2.3e-1	<1.0e-7*	1.0e-7*
<i>Gill1</i>	0.231		1	<1.0e-7*	3.0e-7*
<i>Harris (prog.)</i>	0.126			1	8.3e-4*
<i>Harris (reg.)</i>	0.155				1

Solution	valeur moyenne T_c	valeurs-p			
		<i>Dako</i>	<i>Gill1</i>	<i>Harris (prog.)</i>	<i>Harris (reg.)</i>
<i>Dako</i>	0.716	1	7.7e-4*	<1.0e-7*	5.0e-7*
<i>Gill1</i>	0.681		1	<1.0e-7*	1.6e-5*
<i>Harris (prog.)</i>	0.540			1	2.2e-6*
<i>Harris (reg.)</i>	0.615				1

Solution	valeur moyenne S_n	valeurs-p			
		<i>Dako</i>	<i>Gill1</i>	<i>Harris (prog.)</i>	<i>Harris (reg.)</i>
<i>Dako</i>	1.683	1	8.7e-3*	2.6e-5*	9.9e-1
<i>Gill1</i>	1.738		1	1.8e-6*	7.8e-3*
<i>Harris (prog.)</i>	1.559			1	2.8e-5*
<i>Harris (reg.)</i>	1.687				1

Solution	valeur moyenne E_n	valeurs-p			
		<i>Dako</i>	<i>Gill1</i>	<i>Harris (prog.)</i>	<i>Harris (reg.)</i>
<i>Dako</i>	0.244	1	5.9e-3*	5.8e-6*	7.1e-1
<i>Gill1</i>	0.219		1	5.0e-7*	1.7e-3*
<i>Harris (prog.)</i>	0.308			1	1.1e-5*
<i>Harris (reg.)</i>	0.249				1

TABLE 3.5 – Comparaison de la transmission T des noyaux et cytoplasmes de cellules urothéliales superficielles colorées à l'aide de différentes solutions d'Hématoxyline. Les différences significatives sont indiquées par *.

TABLE 3.6 – Comparaison de l'Entropie S et de l'Énergie E mesurées au niveau des noyaux de cellules urothéliales superficielles pour des lames de cytologie digitales colorées à l'aide de différentes solutions d'hématoxyline. Les différences significatives sont indiquées par *.

plus sombre, ce qui conduit la plus faible quantité de détails visibles à l'intérieur des noyaux.

3.2.6 Analyse de l'impact du temps d'immersion dans la solution d'hématoxyline

L'impact du temps d'immersion dans l'hématoxyline sur la coloration des lames de cytologie urinaire est évalué en utilisant sept groupes de lames, colorées avec l'hématoxyline *Gill1* pendant 10 min, 3 min, 2 min, 1 min, 30 sec, 5 sec et 0 sec respectivement; les autres paramètres de coloration étant fixes dans cette expérience. Des exemples de cellules colorées avec différents temps d'immersion sont donnés Figure 3.10.

Analyse de la Densité Optique Les Densités Optiques des noyaux et cytoplasmes sont comparées pour les lames colorées avec différents temps d'immersion dans la solution d'hématoxyline. Les résultats obtenus sont résumés dans la Figure 3.11 et dans le Tableau 3.7.

Comme attendu, en utilisant une solution d'hématoxyline *Gill1*, un temps d'immersion plus long implique des Densités Optiques plus élevées. La plupart des valeurs-p ajustées sont inférieures au seuil prédéterminé $\alpha = 0.05$: les différences de Densité Optique induites par les différents temps d'immersion dans la solution d'hématoxyline sont donc jugées statistiquement significatives. On constate que le temps d'immersion dans

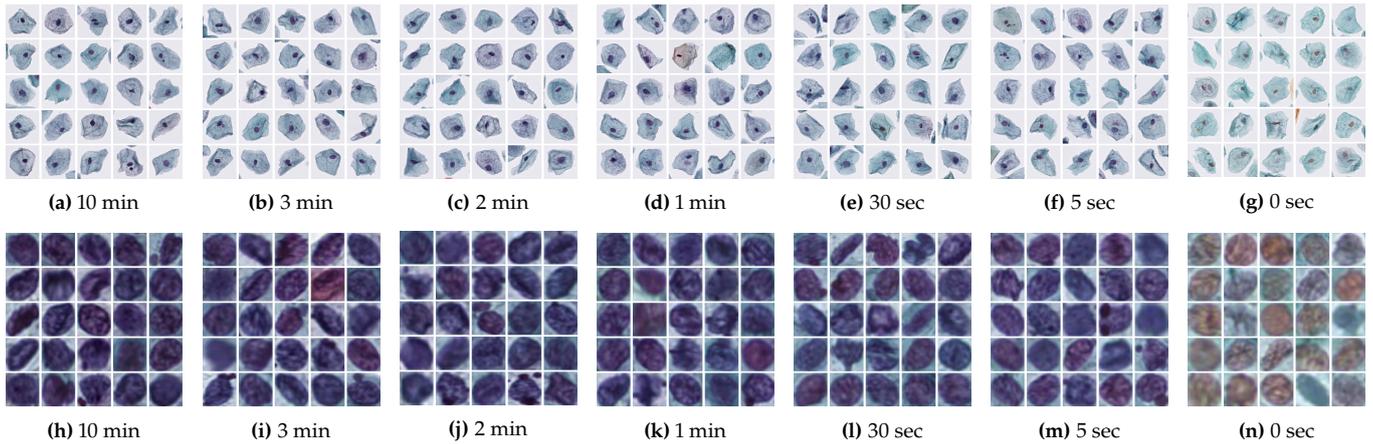


FIGURE 3.10 – Exemples de cellules colorées, avec plusieurs temps d’immersion dans une solution d’hématoxyline Gill1.

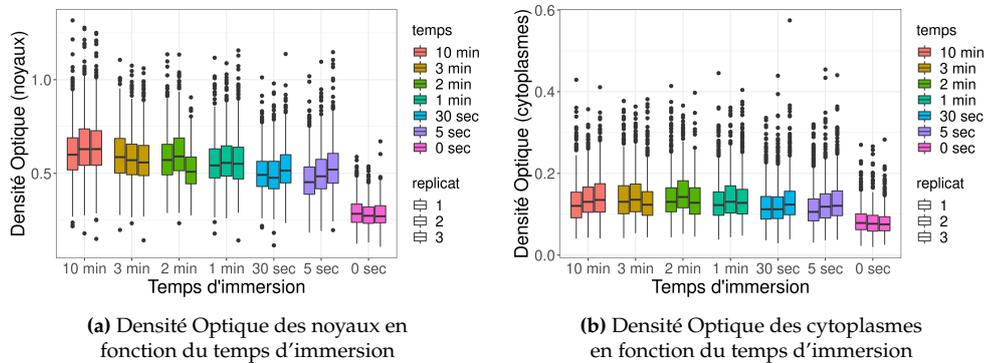


FIGURE 3.11 – Densité Optiques des noyaux et cytoplasmes pour des lames colorées avec différents temps d’immersion dans la solution d’hématoxyline.

l’hématoxyline a moins d’influence sur la Densité Optique des cytoplasmes¹⁵ que pour les noyaux. L’absence d’hématoxyline correspondant au temps d’immersion de 0 seconde a un impact important sur la coloration, même par rapport aux temps d’immersion courts.

15: Voir Figure 3.11.

Analyse de la texture nucléaire La texture des noyaux est comparée pour différents temps d’immersion dans une solution d’hématoxyline Gill1. Les résultats sont résumés Tableau 3.8.

Pour les différents temps d’immersion dans la solution d’hématoxyline Gill1 testés, aucune différence statistiquement significative dans la texture des noyaux n’est observée.

3.2.7 Conclusion

Cette section a introduit une méthode automatique permettant de détecter et de quantifier les différences de colorations entre des lames de cytologie digitales créés par différents protocoles Papanicolaou [3]. Cette méthode a permis de mettre en évidence des différences parfois difficilement visibles par un observateur humain, et encore plus difficilement quantifiables.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

Parmi les différences observées, la choix de la solution d’hématoxyline utilisée pour colorer les lames de cytologie urinaire a un impact important. La solution d’hématoxyline Gill1, la moins concentrée en hématoxyline parmi les solutions testées, est la solution qui conduit à une coloration plus claire et qui permet de maximiser les détails intra-nucléaires. Les expériences menées amènent donc à recommander l’utilisation de

Temps	moyenne OD_n	valeurs-p						
		0"	5"	30"	1'	2'	3'	10'
0"	0.285	1	0.0*	0.0*	0.0*	0.0*	0.0*	0.0*
5"	0.500		1	0.99	4.3e-2*	1.8e-2*	1.8e-2*	2.7e-5*
30"	0.510			1	5.5e-2	2.4e-2*	2.3e-2*	3.4e-5*
1'	0.565				1	0.99	0.99	9.0e-3*
2'	0.580					1	1.0	2.1e-2*
3'	0.585						1	2.1e-2*
10'	0.632							1

Temps	moyenne OD_c	valeurs-p						
		0"	5"	30"	1'	2'	3'	10'
0"	0.082	1	1.0e-7*	8.5e-6	0.0*	1.0e-7*	0.0*	3.0e-7*
5"	0.124		1	3.3e-2*	0.12	0.99	0.17	0.95
30"	0.125			1	2.1e-4*	2.3e-2*	2.9e-4*	0.18
1'	0.135				1	0.17	0.99	2.1e-2*
2'	0.143					1	0.24	0.89
3'	0.139						1	3.1e-2*
10'	0.137							1

TABLE 3.7 – Comparaison de la Densité Optique sur les noyaux (OD_n) et cytoplasmes (OD_c) pour différents temps d'immersion dans une solution d'hématoxyline *Gill1*. Les différences significatives sont indiquées par *.

solutions d'hématoxyline de concentration similaire à la solution *Sigma-Aldrich Gill1 (GHS132-1L)*¹⁶ pour la création de lames de cytologie digitales à partir d'échantillons d'urine.

Le temps d'immersion dans la solution d'hématoxyline *Gill1* a un impact sur la Densité Optique des cellules, néanmoins sans influence significative sur la quantité de détails intra-nucléaires visibles. Les expériences menées ne permettent pas de recommander un temps d'immersion particulier dans la solution d'hématoxyline, en revanche il est préférable de fixer un temps d'immersion pour la création de toutes les lames de cytologie urinaire afin de limiter les variations de coloration et d'éviter l'introduction de biais dans le jeu de données.

16: Concentration reportée par le fournisseur : 2g/L.

Temps	moyenne	valeurs-p						
		0"	5"	30"	1'	2'	3'	10'
0"	1.700	1	0.92	0.09	0.99	0.08	1.00	0.15
5"	1.744		1	0.47	0.79	0.42	0.94	0.63
30"	1.739			1	0.05	1.00	0.10	1.00
1'	1.739				1	0.05	1.00	0.09
2'	1.729					1	0.09	1.00
3'	1.724						1	0.16
10'	1.738							1

Temps	moyenne	valeurs-p						
		0"	5"	30"	1'	2'	3'	10'
0"	0.229	1	0.28	2.6e-3*	1.00	1.4e-3*	0.74	5.4e-3*
5"	0.207		1	0.17	0.53	0.10	0.97	0.32
30"	0.208			1	6.3e-3*	0.99	3.9e-2*	1.00
1'	0.210				1	3.5e-3*	0.95	1.3e-2*
2'	0.211					1	2.1e-2*	0.99
3'	0.210						1	0.08
10'	0.209							1

TABLE 3.8 – Comparaison des mesures d'Entropie S et d'Énergie E au niveau des noyaux et cytoplasmes pour différents temps d'immersion dans une solution d'hématoxyline *Gill1*. Les différences significatives sont indiquées par *.

3.3 Recherche du nombre de plans focaux optimal

Une multitude de combinaison de N_z et δ_z sont possibles pour imaginer une lame sur une profondeur Δ_z . On décide de fixer δ_z à priori de façon à ce que la profondeur de champ du scanner soit proche de la distance inter-plans δ_z , ainsi tout le matériel cellulaire situé entre le premier et le dernier plan focal est imagé net.

3.3.1 Choix de l'espacement inter-plans

Les scanners utilisés en cytopathologie digitale fonctionnent classiquement au grossissement $\times 40$, avec une ouverture numérique $NA = 0.75$ et dans l'air dont l'indice de réfraction est $n = 1$. La profondeur de champ (P_{champ}) du scanner peut être estimée à l'aide de la formule suivante :

$$P_{champ} = \frac{\lambda n}{NA^2} \quad (3.2)$$

Avec :

λ : longueur d'onde

n : indice de réfraction du milieu entre l'objectif et l'échantillon

NA : ouverture numérique de l'objectif

Ainsi, pour des longueurs d'onde dans le visible, de 400 à 800 nm, correspondant à l'éclairage en lumière blanche, la profondeur de champ du scanner varie entre 0.7 μm et 1.4 μm . En fixant $\delta_z = 1 \mu\text{m}$, l'espacement inter-plans est environ égal à la profondeur de champ du scanner. Ainsi, en fixant $\delta_z = 1 \mu\text{m}$ et à condition de numériser un nombre de plans focaux suffisants, tous les objets d'intérêt peuvent être imagés en-focus.

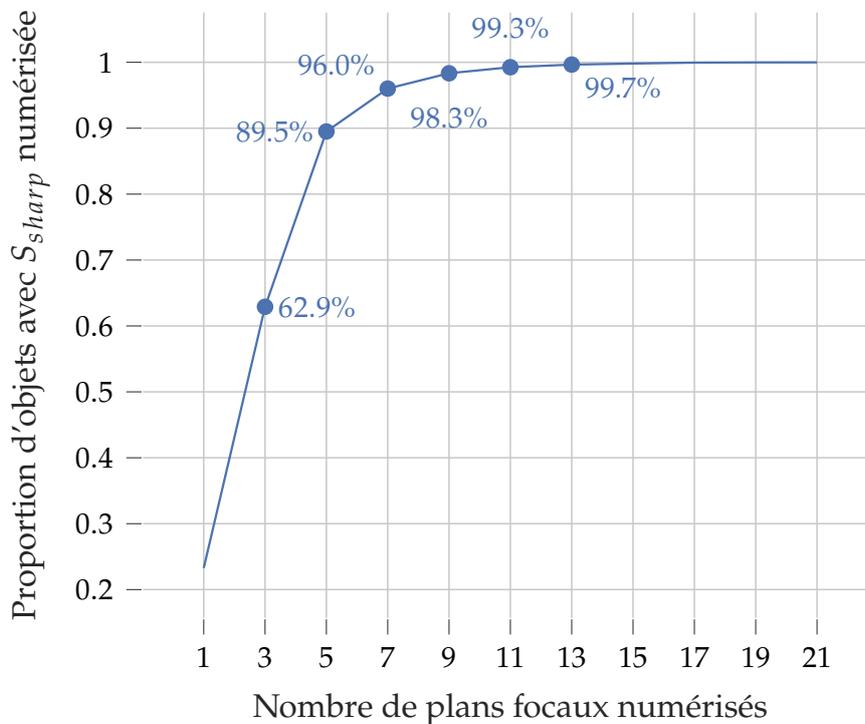
3.3.2 Recherche du nombre de plans focaux optimal connaissant l'espacement inter-plans

Une fois l'espacement inter-plans δ_z fixé à $1\ \mu\text{m}$ en se basant sur la profondeur de champ du scanner, il reste à déterminer un nombre optimal de plans focaux N_z à numériser permettant de maximiser les performances de diagnostic du cancer de la vessie à l'aide d'algorithmes d'analyse d'images, tout en gardant N_z le plus faible possible afin de minimiser le temps d'acquisition et l'espace de stockage nécessaire pour les lames digitales.

Comme expliqué dans la section 3.1, il est difficile de conditionner la numérisation des lames de cytologie aux performances d'algorithmes d'analyse d'images, dont le développement dépend de la numérisation de lames de cytologie.

Ainsi, on cherche dans un premier temps N_z de telle sorte qu'un maximum d'information soit présent dans les lames digitales. Pour cela, pour chaque cellule présente dans la lame, la position z_{sharp} du plan de meilleur focus S_{sharp} , définie dans la section 2.2, est calculée. Ensuite, en fonction de N_z , la proportion des cellules pour lesquelles la section S_{sharp} est incluse dans les plans numérisés est calculée. Pour $N_z = 1$, le plan numérisé est le plan focal déterminé comme optimal par l'autofocus du scanner de lames. Dans cette expérience N_z prend pour valeurs possibles : $N_z \in [1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21]$.

À chaque incrémentation de la valeur de N_z , on considère donc deux plans focaux supplémentaires, espacés de $\delta_z = 1\ \mu\text{m}$ au-dessus et en-dessous les plans focaux extrêmes de l'étape précédente. Pour éviter toute variation causée par une inconsistance de l'autofocus du scanner de lames, les lames digitales ont été scannées chacune une fois selon 21 plans, et les valeurs de N_z inférieures sont obtenues par sous-échantillonnage des z-stacks. Les résultats obtenus, après analyse de l'ensemble des cellules urothéliales basales extraites de 60 lames de cytologie digitales¹⁷, sont illustrés Figure 3.12.



17: Cela représente plus de 170 000 cellules urothéliales basales uniques.

FIGURE 3.12 – Proportion de cellules urothéliales dont le plan de meilleur focus S_{sharp} est numérisé en fonction du nombre de plans focaux total numérisé. Les cellules sont extraites de 60 lames numérisées selon 21 plans focaux espacés de $\delta_z = 1\ \mu\text{m}$.

Le plan z_0 sélectionné par l'autofocus du scanner de lames correspond à la section S_{sharp} de seulement 30% des cellules d'intérêt. Il est nécessaire de numériser cinq plans

focaux espacés de $\delta_z = 1 \mu\text{m}$ pour imager presque 90% des sections S_{sharp} des cellules urothéliales basales, et 11 plans pour voir cette proportion atteindre 99%.

3.4 Conclusion

Dans ce chapitre, plusieurs expériences ont été proposées afin de fixer certains paramètres de création et de numérisation de lames de cytologie urinaire, adaptés à l'analyse automatique de lames digitales.

A l'aide de la méthode décrite section 3.2, il a été trouvé que la solution d'hématoxyline *Gill1*, faiblement concentrée, est bien adaptée à la création de lames de cytologie urinaire, et permet de maximiser les détails intra-nucléaires visibles pour les cellules urothéliales superficielles. Le temps d'immersion dans la solution d'hématoxyline semble ne pas avoir d'impact sur la quantité de détails intra-nucléaires visibles, mais influe légèrement sur la Densité Optique des noyaux et donc sur l'apparence globale des cellules.

La section 3.3 propose une expérience afin de déterminer le nombre de plans focaux N_z optimal, connaissant l'espacement inter-plans δ_z , afin d'imager un maximum de cellules urothéliales basales en focus tout en limitant N_z afin de réduire le temps d'acquisition et l'espace de stockage nécessaire.

L'espace inter-plans δ_z est fixé à $1 \mu\text{m}$ pour être similaire à la profondeur de champ du scanner de lames utilisé au grossissement $\times 40$. Des espaces inter-plans δ_z similaires ont été utilisés dans les études [101][104][130].

Une fois δ_z fixé à $1 \mu\text{m}$, il est trouvé nécessaire de numériser au moins 11 plans focaux pour imager le plan de meilleur focus S_{sharp} de 99% des cellules urothéliales basales d'un ensemble de 60 lames de cytologie digitales. Le choix de $N_z = 11$ plans focaux est délibérément conservateur puisqu'il sera, dans la suite de ce travail, comparé aux performances d'algorithmes de diagnostic automatique du cancer de la vessie. Pour des espacements inter-plans similaires, [104] et [130] recommandent de numériser $N_z = 3$ plans respectivement pour le diagnostic de cytologies cervicales et pour le diagnostic de cytologies de la thyroïde. Dans une autre étude [101], Fan et al. recommandent la numérisation de 18.6 plans focaux en moyenne pour imager l'intégralité du matériel cellulaire contenu dans des lames de cytologie cervicale.

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

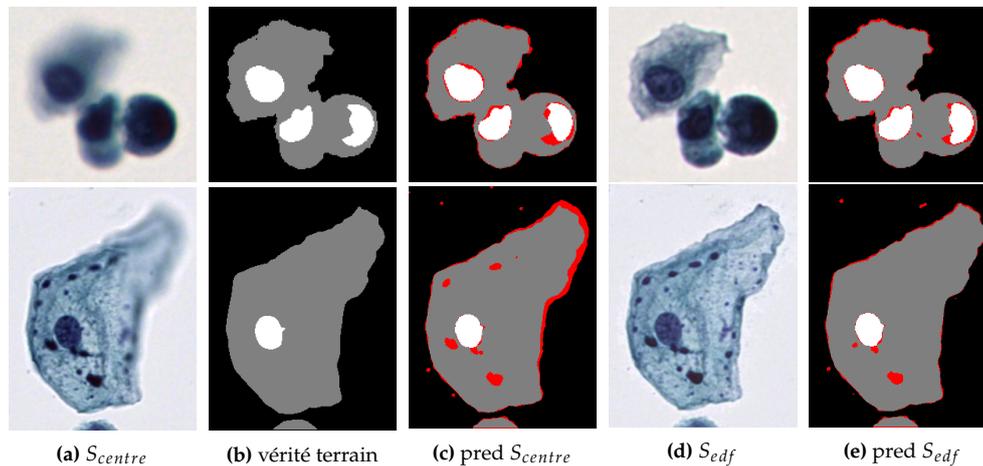
[104]: DONNELLY et al. (2013), « Optimal z-axis scanning parameters for gynecologic cytology specimens »

[130]: MUKHERJEE et al. (2015), « Investigation of scanning parameters for thyroid fine needle aspiration cytology specimens »

[101]: FAN et al. (2016), « A method for quantitative analysis of clump thickness in cervical cytology slides »

Segmentation à partir de z-stacks

En cytopathologie, le diagnostic du cancer de la vessie ou du col de l'utérus repose en grande partie sur l'analyse de caractéristiques morphologiques, telles que la texture de la chromatine du noyau ou encore le Ratio Noyau-Cytoplasme (RNC), correspondant au rapport entre l'aire apparente du noyau et celle du cytoplasme [24]. En cytopathologie conventionnelle, ces caractéristiques sont évaluées visuellement par des anatomopathologistes et des cytotechniciens qualifiés. Avec l'adoption grandissante de la cytologie digitale, des méthodes automatisées et fiables de segmentation du noyau, du cytoplasme et de l'arrière-plan peuvent être développées pour calculer précisément ces caractéristiques d'intérêt. Le matériel cellulaire présent dans les lames de cytologie étant par essence tri-dimensionnel, toute l'information disponible est rarement contenue dans un seul plan focal. Par conséquent, il est parfois difficile d'effectuer précisément une telle segmentation des régions cellulaires à l'aide d'un unique plan focal [131]. Une illustration de la difficulté d'obtenir une segmentation précise à l'aide d'un modèle U-Net [16] en présence de cellules partiellement en focus est proposée Figure 4.1.



[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[131]: AGARD (1984), « Optical Sectioning Microscopy : Cellular Architecture in Three Dimensions »

[16]: RONNEBERGER et al. (2015), « U-Net »

FIGURE 4.1 – Segmentation des régions cellulaires avec un modèle U-Net entraîné sur les sections centrales S_{centre} de chaque z-stack (c) et avec un modèle U-Net entraîné sur les projections extension de profondeur de champ (EDF) S_{edf} des z-stacks (e). Les erreurs de segmentation par rapport à la vérité terrain sont représentées en rouge. En présence de cellules hors-focus ou partiellement hors-focus, le modèle U-Net entraîné sur les sections centrales peut produire une segmentation inexacte.

Pour surmonter ce problème, une approche communément utilisée [111] [112] [132] consiste à projeter l'information contenue dans une z-stack sur une image bi-dimensionnelle, par exemple à l'aide d'une méthode d'EDF, comme décrite dans la section 2.3, suivie d'une méthode de segmentation 2D classique.

Cependant, ces pré-traitements par extension de profondeur de champ n'apportent pas d'information supplémentaire par rapport aux z-stacks brutes, augmentent le temps de traitement nécessaire et peuvent introduire des artefacts dans les images projetées. Pire, dans les cas de cellules superposées, une partie de l'information fusionnée dans un seul plan peut être perdue.

Ainsi dans ce chapitre, différentes approches sont explorées pour extraire de l'information directement à partir de z-stacks, sans pré-traitement. Plusieurs problèmes de segmentation sémantique et de segmentation d'instances sont considérés pour juger de la capacité de différents modèles à extraire de l'information d'intérêt directement à partir de z-stacks.

La segmentation sémantique associe un label de classe à chaque pixel de l'image qui n'appartient pas à l'arrière-plan (voir Figure 4.2 (b)).

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

[112]: FORSTER et al. (2004), « Extended depth-of-focus for multi-channel microscopy images »

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

La **segmentation d'instances** détecte et trouve l'ensemble des pixels appartenant à chaque objet d'intérêt dans l'image. Dans un problème de segmentation d'instance, un même pixel peut appartenir à plus d'un masque de segmentation, contrairement à la segmentation sémantique (voir Figure 4.2 (c)).

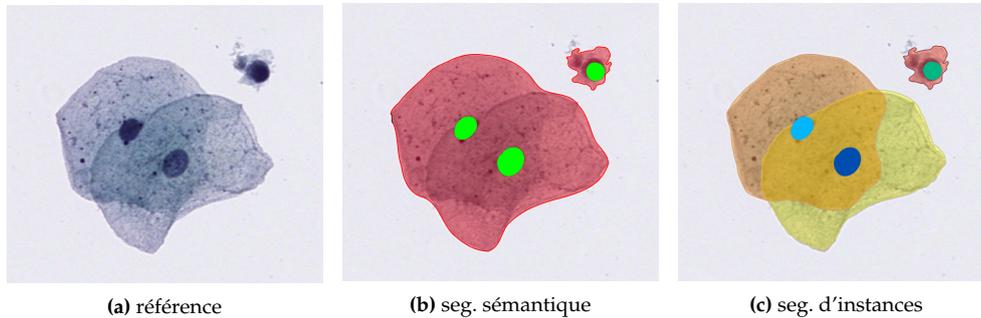


FIGURE 4.2 – Illustration de deux problématiques classiques de segmentation d'une image de référence (a). (b) représente une segmentation sémantique des pixels de cytoplasme (rouge) et des pixels de noyau (vert). (c) représente une segmentation d'instances : chaque organe, cytoplasme et noyau, est détecté et segmenté individuellement.

Puisque les lames de cytologie digitales sont numérisées selon plusieurs plans focaux par un scanner de lames, une z-stack n'est pas considérée comme un volume classique¹, mais plutôt comme une succession de plans qui partagent entre eux une partie de l'information. Les problèmes de segmentation considérés dans ce chapitre cherchent donc à prédire une carte de segmentation bi-dimensionnelle à partir de z-stacks. Les masques de vérité terrain associés sont annotés manuellement à partir des sections S_{edf} qui assurent que tout les objets présents sont en focus.

La suite de ce chapitre est organisée de la façon suivante : la section 4.1 adresse la segmentation sémantique de régions cellulaires comme un problème de classification de profils-z, permettant de considérer l'information contenue dans les z-stacks. Dans la section 4.2, des méthodes de segmentation sémantique de noyaux à partir de z-stacks sont proposées. Ces méthodes améliorent les performances obtenues section 4.1 en tirant parti à la fois de l'information spatiale contenue dans les sections optiques, et de l'information multi-plans propre aux z-stacks. Finalement, dans la section 4.3, les méthodes présentées section 4.2 sont adaptées à la segmentation d'instances de cellules urothéliales superposées : un problème plus complexe que la segmentation sémantique où un unique pixel peut appartenir à plusieurs instances cellulaires.

4.1 Segmentation sémantique à partir de profils d'intensité selon l'axe z

Inspirée par plusieurs études récentes, cette section explore des méthodes afin d'extraire l'information présente dans les z-stacks à partir des variations d'une métrique le long de l'axe z pour une position (x, y) donnée. Dans la suite, ce vecteur de variation d'une métrique selon z sera appelé *profil-z*.

4.1.1 État de l'art

Dans [133] et [134], les auteurs ont calculé des projections de profils-z d'intensité pour augmenter le contraste des cellules non-colorées à partir de lames de microscopie en lumière blanche. Ces projections ont été réalisées en calculant les variations d'intensité : écart-type, écart interquartile, écart absolu moyen, ou en calculant la différence entre deux plans focaux complémentaires par rapport au meilleur plan focal.

1: À la différence d'autres modalités d'imageries, comme l'IRM, permettant d'imager des volumes dont chaque voxel contient une information propre, l'acquisition multi-plans par microscopie photonique produit des images dont les pixels contiennent, en plus de leur information propre, une partie de l'information des plans adjacents modulée par la fonction d'étalement de point du système d'imagerie.

[133]: SELINUMMI et al. (2009), « Bright Field Microscopy as an Alternative to Whole Cell Fluorescence in Automated Analysis of Macrophage Images »

[134]: DEHLINGER et al. (2013), « Dye free automated cell counting and analysis »

Plusieurs études ont appliqué avec succès une classification non-supervisée à des vecteurs de profils-z afin d'identifier différentes régions cellulaires. Dans [110], Shihavuddin et al. ont notamment utilisé un modèle de clustering K-Means afin de regrouper les profils-z d'une métrique de focus en trois classes : *cellules*, *arrière-plan* et *incertains*. Dans l'étude [135], Wang et al. ont appliqué un clustering Fuzzy C-Means sur les profils-z d'intensité pour identifier les régions cellulaires par rapport l'arrière-plan dans des lames de cytologie cervicale, puis pour obtenir des régions candidates à la segmentation des noyaux.

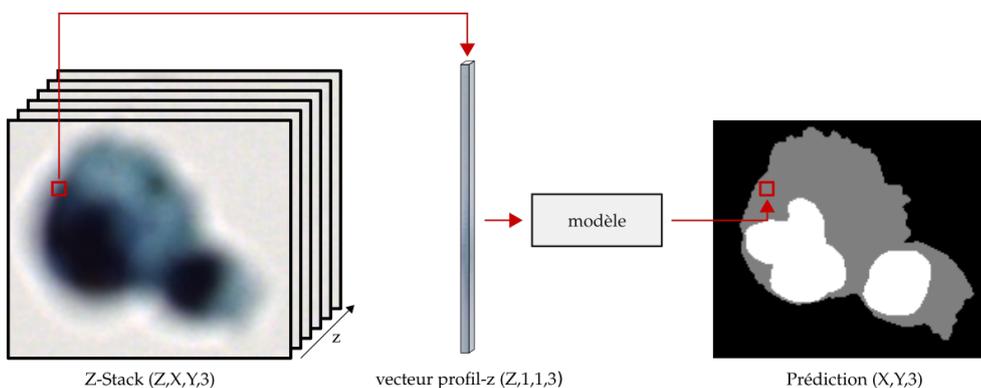
D'autres travaux ont utilisé les profils-z d'intensité directement comme vecteurs de caractéristiques pour la classification supervisée des pixels. Dans [136], Liimatainen et al. ont appliqué une régression logistique avec régularisation L1 pour effectuer une classification binaire des pixels et ainsi détecter les cellules non colorées dans des lames de cytologie. Dans [137], Lugagne et al. ont effectué une classification multi-classes des pixels afin de segmenter les composantes morphologiques des cellules non colorées² en utilisant une machine à vecteurs de support (SVM).

Les lames de cytopathologie digitales sont généralement colorées, notamment à l'aide de coloration de Papanicolaou ou hématoxyline et éosine (H&E), pour mettre en valeur les biomarqueurs. Le problème abordé dans cette section diffère de la plupart des travaux précédemment cités qui se sont concentrés sur le matériel cellulaire non coloré. La segmentation de cellules d'intérêt colorées par rapport à l'arrière-plan est une tâche beaucoup plus facile qu'en présence de cellules non colorées qui manquent souvent de contraste. Cependant, la segmentation automatique de régions cellulaires, comme le noyau et le cytoplasme, reste une tâche difficile. Les travaux menés dans cette section se sont concentrés sur les méthodes d'apprentissage supervisé car des annotations de bonne qualité pour la segmentation des régions cellulaires étaient disponibles.

4.1.2 Segmentation à partir de profils-z

Soit V_{zxyC} une z-stack de dimension (Z, H, W, C) , Z étant le nombre de plans focaux, H, W les dimensions de chaque section dans la z-stack et C le nombre de canaux dans l'espace couleur considéré³. À chaque position $(x, y) \mid x \in [1, H], y \in [1, W]$ sont associés C vecteurs de dimension Z notés $v_{x,y}^c \in \mathbb{R}^Z$. Ces vecteurs sont appelés ci-après vecteurs de profils-z d'intensité.

La segmentation des régions cellulaires à l'aide des caractéristiques du profil-z est un problème de segmentation sémantique, donc de classification de pixels. Pour chaque pixel de l'ensemble de données d'entraînement, les C vecteurs de profil-z associés sont utilisés comme caractéristiques pour entraîner un modèle de classification. Une fois le modèle entraîné, une classe parmi *noyau*, *cytoplasme* et *arrière-plan* est prédite pour chaque position (x, y) , comme illustré dans la Figure 4.3.



[110]: SHIHAVUDDIN et al. (2017), « Smooth 2D manifold extraction from 3D image stack »

[135]: WANG et al. (2020), « Nucleus Segmentation of Cervical Cytology Images Based on Depth Information »

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

[137]: LUGAGNE et al. (2018), « Identification of individual cells from z-stacks of bright-field microscopy images »

2: Les différentes classes considérées sont l'intérieur de la cellule, le contour de la cellule, le halo et l'arrière-plan.

3: $C = 1$ pour une z-stack en niveau de gris et $C = 3$ pour une z-stack en couleur.

FIGURE 4.3 – Segmentation sémantique en utilisant, pour chaque pixel (x_i, y_i) , le vecteur profil-z d'intensité comme vecteur de caractéristiques. Plusieurs types de modèles peuvent être utilisés pour effectuer cette tâche. À noter : la prédiction est de dimension $(X, Y, 3)$ puisqu'il s'agit d'une classification à 3 classes (arrière-plan, cytoplasme et noyau).

L'utilisation de l'information des profils-z pour la segmentation des régions de cellules colorées est motivée par la présence d'une grande variété de profils-z dans les z-stacks, comme illustré dans la Figure 4.4.

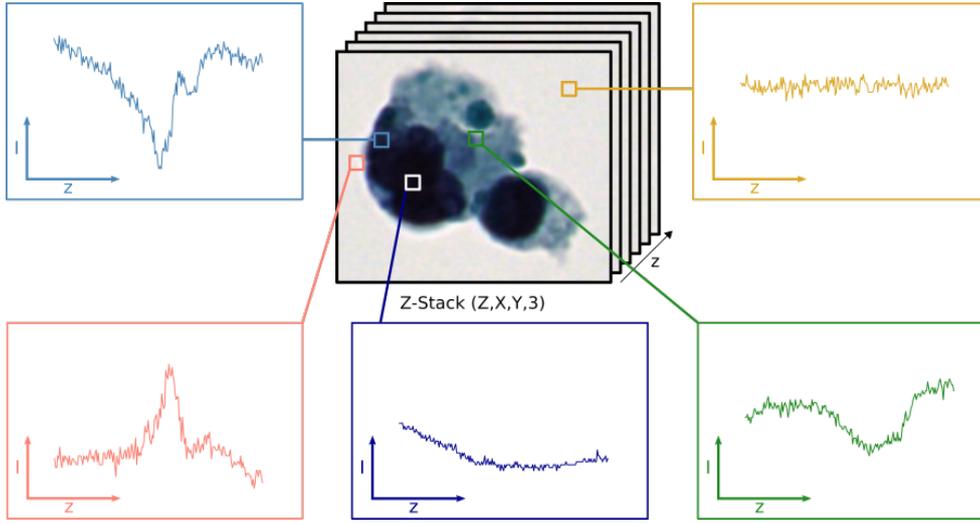


FIGURE 4.4 – Profils-z d'intensité extraits d'une z-stack de 201 plans focaux, pour le canal de couleur vert à différentes positions (x, y) . En particulier, le profil-z en haut à gauche (bleu clair) appartient à un pixel de la membrane cellulaire, et le profil-z en bas à gauche (rouge) appartient à un pixel d'arrière-plan à proximité de la cellule.

Par exemple, les pixels de la cellule à proximité de la membrane cellulaire ont tendance à avoir un profil-z avec un minimum local au niveau du plan de meilleur focus S_{sharp} , tandis que les pixels de l'arrière-plan à proximité de la cellule ont tendance à avoir un profil-z d'intensité avec un maximum local pour $z = z_{sharp}$.

Avec de telles propriétés, les profils-z semblent contenir des informations très utiles pour segmenter précisément le matériel cellulaire de l'arrière-plan.

4.1.3 Métriques d'évaluation

L'entraînement des modèles et l'inférence sont effectués sur les profils-z indépendamment, pour chaque position (x, y) , puis l'évaluation est effectuée au niveau de l'image entière, une fois les prédictions des masques de segmentation sémantique complets obtenus. Soit Y_{true} la vérité terrain pour un problème de segmentation sémantique à K classes, Y_{pred} la prédiction du modèle associée, N_{total} le nombre de pixels dans l'image et $N_{correct}$ le nombre de pixels correctement labellisés. Les métriques de Dice et d'Exactitude sont utilisées pour quantifier la qualité d'une segmentation :

$$\text{score de Dice} = \frac{1}{K} \sum_{i=1}^K \frac{2|Y_{pred}^i \cap Y_{true}^i|}{|Y_{pred}^i| + |Y_{true}^i|} \quad (4.1)$$

$$\text{Exactitude} = \frac{N_{correct}}{N_{total}} \quad (4.2)$$

Pour l'entraînement des modèles, on considère l'entropie croisée catégorielle (CCE)⁴ comme fonction de coût :

$$\mathcal{L}_{CCE} = -\frac{1}{N_{total}} \sum_{x,y} \frac{1}{K} \sum_{k=0}^{K-1} Y_{true}^k(x,y) \log(Y_{pred}^k(x,y)) \quad (4.3)$$

Cette fonction de coût est définie pour N_{total} pixels en faisant la moyenne de l'entropie croisée par classe, pour tous les pixels de l'échantillon. Contrairement au score de Dice,

4: En anglais : Categorical Cross-Entropy.

défini sur un ensemble de pixels, \mathcal{L}_{CCE} peut être calculée pour des pixels individuels et est par conséquent naturellement adaptée à l'entraînement de modèles de classification de pixels.

Le modèle CNN-1D, lorsque entraîné sur le jeu de données complet, peut également être entraîné à l'aide de la fonction de coût de Dice définie comme :

$$\mathcal{L}_{Dice} = 1 - \text{score de Dice} \quad (4.4)$$

4.1.4 Jeux de données

Deux ensembles de données sont utilisés dans cette section.

Un premier jeu de données, nommé ci-après Uro est composé de 837 z-stacks contenant des cellules urothéliales colorées par une coloration de Papanicolaou et numérisées à un zoom $\times 40$ selon 21 plans focaux espacés de $1 \mu\text{m}$, à l'aide d'un scanner de lames Hamamatsu NanoZoomer $\text{\textcircled{R}}$ S60. Cet ensemble de données est divisé en 585 échantillons pour l'entraînement (*Train*), 126 échantillons pour le développement et le choix des hyperparamètres (*Dev*) et 126 échantillons pour l'évaluation finale (*Test*). Afin de capturer une grande variété de morphologie des cellules urothéliales, les z-stacks sont principalement extraites d'une lame d'un patient atteint d'un carcinome urothélial de haut-grade (HGUC). Environ 10% des échantillons testés ont été extraits de deux autres patients, l'un atteint d'un carcinome urothélial de bas-grade (LGUC) et l'autre sain. Les masques des régions d'arrière-plan, de cytoplasme et de noyau ont été annotés manuellement sur les projections EDF des z-stacks.

Le second jeu de données utilisé dans cette section, nommé Cervix67seg, est construit à partir du jeu de données Cervix93 disponible publiquement [132].

Cervix93 Le jeu de données original Cervix93 est composé de 93 z-stacks et de leurs images EDF correspondantes, acquises auprès de patientes saines et de patientes présentant des cas de lésions malpighiennes intra-épithéliales de bas-grade (LSILs) ou des lésions malpighiennes intra-épithéliales de haut-grade (HSILs). Ce jeu de données est composé d'images de cytologie cervicale, relativement semblables aux images de cytologie urinaire qui nous intéressent initialement. Étant disponibles publiquement, ces données permettent ainsi de comparer les performances des méthodes proposées avec l'état de l'art. Pour le jeu de données Cervix93, les annotations fournies sont des points indiquant la position des noyaux des cellules cervicales.

Cervix67seg Le jeu de données Cervix67seg considéré est uniquement basé sur les z-stacks de Cervix93 numérisés avec 20 plans focaux, représentant 67 z-stack sur les 93 disponibles. Les 26 lames rejetées ne contiennent que 10 plans et appartiennent toutes à l'ensemble d'entraînement de la base de données Cervix93. Les masques de segmentation des cellules malpighiennes sont manuellement annotés à partir des projections EDF des 67 z-stacks considérées dans le jeu de données Cervix67seg, et constituent la vérité terrain. Certains masques de noyaux de cellules neutrophiles sont également annotés, et utilisés pour la sélection de sous-images lors de la création du jeu de données Cervix67seg, afin d'inclure des exemples d'arrière-plan non-triviaux dans les données d'entraînement, et ainsi permettre aux modèles de discriminer les noyaux de cellules malpighiennes des noyaux de cellules neutrophiles. Parmi les 67 lames du jeu de données Cervix67seg, 43 lames sont retenues pour l'entraînement ⁵ et 24 lames sont utilisées pour le test ⁶. Les lames d'entraînements sont séparées en un ensemble *Train* et

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

5: 24 lames de bas-grade, 13 lames de haut-grade et 6 lames saines.

6: 12 lames de bas-grade, 8 lames de haut-grade et 4 lames saines.

un ensemble *Dev* de trois façons différentes, comme décrit Tableau 4.1, afin d'évaluer la capacité de généralisation des modèles.

Ensemble	nombre lames LSIL	nombre lames HSIL	nombre lames saines
Train	18	9	4
Dev	6	4	2
Test	12	8	4

TABLE 4.1 – Nombre de lames pour les ensembles Train, Dev et Test dans le jeu de données Cervix67seg.

Comme l'ensemble de données Cervix67seg utilisé dans cette section et l'ensemble de données Cervix93 partagent le même ensemble de Test, une comparaison directe des performances sur les lames de Test des deux jeux de données est donc possible.

Pré-traitement Des z-stacks de dimensions (20, 128, 128) pixels sont extraites des lames composant le jeu de données Cervix67seg, suivant un glissement horizontal et vertical de 64 pixels impliquant un chevauchement. Étant donné que la grande majorité des z-stacks d'une lame ne contiennent que de l'arrière-plan, les z-stacks ne sont ajoutées au jeu de données que si elles contiennent au moins 100 pixels annotés comme noyau de cellule cervicale ou noyau de cellule neutrophile. Bien que les noyaux de neutrophiles soient considérés comme de l'arrière-plan dans nos expériences, il a été nécessaire de les ajouter spécifiquement à l'ensemble de données d'entraînement pour aider le réseau à distinguer les noyaux de cellules cervicales des noyaux de neutrophiles. Ce processus aboutit à un total de 7225 z-stacks extraites, divisées en ensembles Train, Dev et Test en fonction de leur lame d'origine. Tous les z-stacks extraites d'une lame appartiennent au même ensemble, ce qui évite les fuites d'informations entre les ensembles de données. La distribution des sous-volumes est donnée Tableau 4.2.

Split	nb de z-stacks Train	nb de z-stacks Dev	nb de z-stacks Test
1	3352	1365	2508
2	3329	1388	2508
3	3369	1348	2508

TABLE 4.2 – Nombre de z-stacks dans les ensembles de Train, de Dev et de Test, pour chacun des 3 splits.

4.1.5 Hard mining

Chaque vecteur à la position (x, y) dans une z-stack est un échantillon pour les approches de segmentation sémantique décrites dans la sous-section 4.1.6. À titre d'exemple, dans le jeu de données Uro composé de 837 z-stacks, cela correspond à plus de 40 millions d'échantillons uniques. Comme les méthodes SVC et LRC ne s'adaptent pas bien à un grand nombre d'échantillons, un sous-échantillonnage est nécessaire.

Contrairement à [136] qui utilise un sous-échantillonnage aléatoire, ici une approche de hard mining a été appliquée. L'intuition est qu'en sous-échantillonnant l'ensemble de données par une approche de hard mining, les échantillons sélectionnés permettront d'apprendre une frontière de décision plus précise, comme illustré Figure 4.5.

Ce processus de hard mining, détaillé dans la suite, permet la construction d'un jeu de données sous-échantillonné contenant environ un million d'échantillons pertinents (voir illustration Figure 4.6).

Pour sous-échantillonner le jeu de données Cervix67seg, une SVM avec noyau RBF, similaire à l'approche décrite par Lugagne et al. [137], est entraînée tour à tour sur chaque z-stack du jeu de données d'entraînement. Les 512 échantillons situés au plus près des frontières de décision apprises, parmi les échantillons éligibles et suivant la distance

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

[137]: LUGAGNE et al. (2018), « Identification of individual cells from z-stacks of bright-field microscopy images »

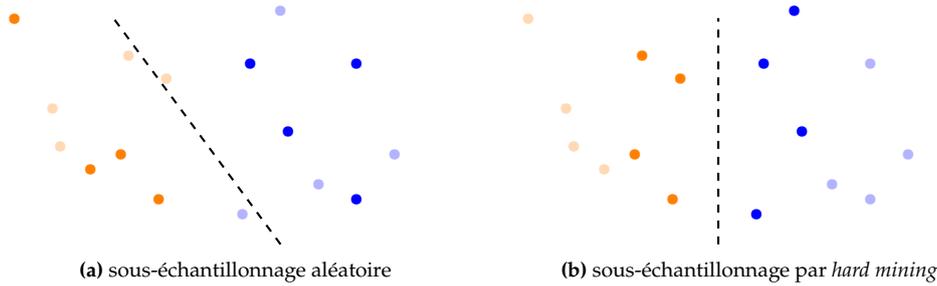


FIGURE 4.5 – Illustration de l’intuition derrière le sous-échantillonnage par hard mining. **(a)** un sous-échantillonnage aléatoire, la frontière de décision alors apprise ne permet pas de séparer correctement les échantillons bleus des échantillons orange. **(b)** un sous-échantillonnage par hard mining permet d’obtenir une meilleure frontière de décision en sélectionnant les échantillons proches de celle-ci. Dans chaque cas, les échantillons non-sélectionnés sont représentés en légère transparence.

euclidienne, sont alors sélectionnés. Les échantillons inéligibles sont l’ensemble des profils-z situés à moins de trois pixels d’une délimitation inter-classes annotée. Cela permet de limiter la probabilité de sélectionner un échantillon mal annotés⁷ comme échantillon pertinent.

Les performances des modèles LRC, SVC et CNN-1D entraînés sur le jeu de données Cervix67seg, sous-échantillonné de manière aléatoire comme proposé dans [136] et sous-échantillonné par extraction d’éléments complexes, sont reportées Tableau 4.3.

Modèle	sous-échantillonnage	moy. Exactitude $\pm \sigma$	moy. Dice $\pm \sigma$
SVC	aléatoire	0.911 \pm 0.082	0.742 \pm 0.131
SVC	hard mining	0.964 \pm 0.036	0.840 \pm 0.099
LRC	aléatoire	0.901 \pm 0.094	0.730 \pm 0.137
LRC	hard mining	0.959 \pm 0.042	0.829 \pm 0.105
CNN-1D	aléatoire	0.912 (0.071)	0.735 (0.121)
CNN-1D	hard mining	0.968 \pm 0.031	0.845 \pm 0.092

La différence de performances induite par l’utilisation d’un jeu de données échantillonné au lieu du jeu de données complet est discutée dans la sous-section 4.1.7 à l’aide du modèle CNN-1D, qui peut être facilement entraîné sur des jeux de données de grande taille grâce à son optimisation par descente de gradient et l’utilisation de batchs de taille limitée.

4.1.6 Méthodes

Trois méthodes de classification de pixels à partir de vecteurs de profils-z sont évaluées : le modèle de régression logistique (LR) proposé dans [136] et nommé ci-après LRC, la machine à vecteurs de support (SVM) proposée dans [137] et nommée SVC dans la suite, ainsi qu’un réseau de neurones convolutionnel (CNN) développé dans cette section et nommé CNN-1D.

Le modèle de régression logistique multi-classes prend en entrée le vecteur $\mathbf{v}_{x,y}$: la concaténation de 3 vecteurs ($\mathbf{v}_{x,y}^r, \mathbf{v}_{x,y}^g, \mathbf{v}_{x,y}^b$) représentant la variation d’intensité le long de l’axe z , associée à une position (x, y) , pour chaque canal couleur. Pour chaque classe k , un ensemble de coefficients de régression β_k est appris à partir des échantillons d’une base de données d’apprentissage. Ensuite, le modèle est utilisé pour calculer la probabilité $P(k|\mathbf{v}_{x,y})$ de chaque position (x, y) d’appartenir à la classe k (voir Équation 4.5). Enfin,

7: Les pixels situés immédiatement à proximité des délimitations inter-classes sont à risque d’être mal annotés.

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

TABLE 4.3 – Performances de segmentation sur les z-stacks du jeu de données Cervix67seg, pour des méthodes entraînées sur le jeu de données sous-échantillonné aléatoirement et par hard mining. Utiliser un sous-échantillonnage par extraction d’éléments complexes permet d’améliorer les performances de façon importante pour toutes les méthodes considérées.

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

[137]: LUGAGNE et al. (2018), « Identification of individual cells from z-stacks of bright-field microscopy images »

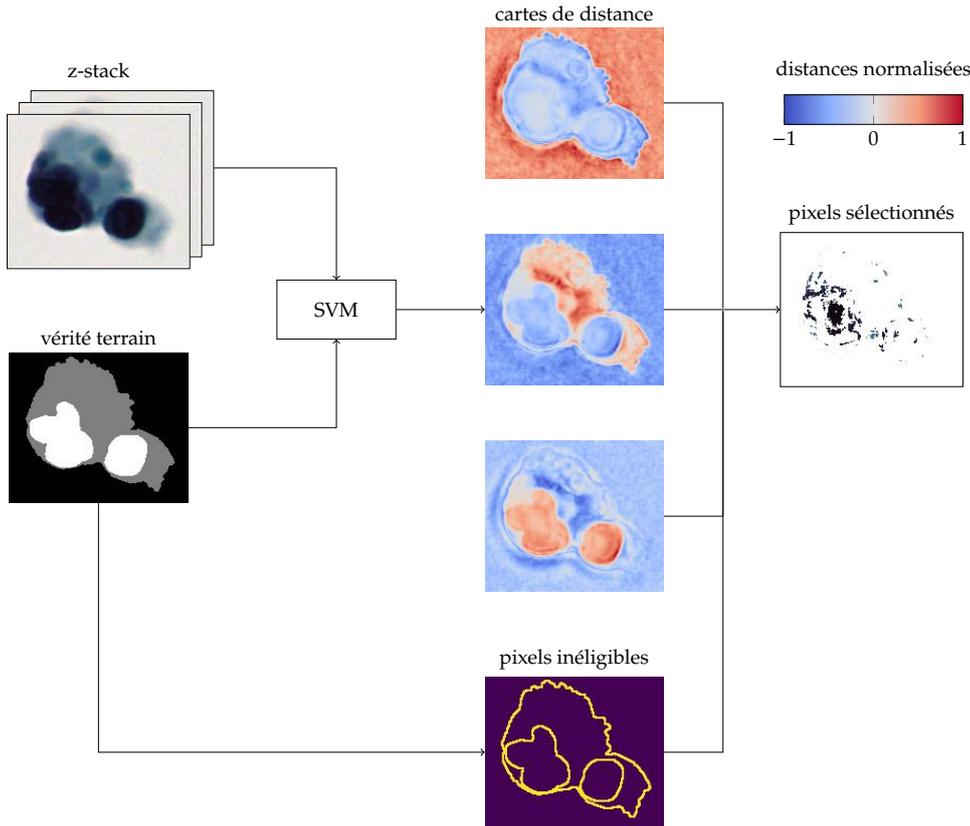


FIGURE 4.6 – Illustration du processus de hard mining utilisé. Pour chaque z-stack, une SVM est entraînée à prédire les classes des pixels parmi *arrière-plan*, *cytoplasme* et *noyau*, à partir des profils-z d'intensité. Les pixels éligibles les plus proches des frontières de décision apprises par la SVM sont finalement sélectionnés.

la prédiction pour la position (x, y) est la classe k qui maximise $P(k|v_{x,y})$.

$$P(k|v_{x,y}) = \frac{1}{1 + \sum_{i=1, i \neq k}^K \exp(\beta_i \cdot v_{x,y})} \quad (4.5)$$

Le modèle SVC prend également en entrée $v_{x,y}$ la concaténation de trois vecteurs $(v_{x,y}^r, v_{x,y}^s, v_{x,y}^b)$ pour chaque position (x, y) . Le modèle SVC utilisé est une SVM linéaire, appliquée sur les vecteurs caractéristiques projetés sur un noyau RBF⁸ approximé à partir des données d'apprentissage. Le noyau RBF est défini pour deux vecteurs $v_{x1,y1}, v_{x2,y2}$ par l'Équation 4.6, où σ est un hyper-paramètre. Pour chaque classe k , un ensemble de coefficients ω_k, b_k est appris, définissant l'hyperplan avec la marge maximale séparant au mieux la classe k des autres classes. Enfin, la prédiction pour la position (x, y) est la classe k avec la plus haute probabilité estimée à partir des scores de classe : $\text{Score}_k(x, y) = \omega_k \cdot v_{x,y} + b_k$. Plus de détails sont disponibles dans [138].

8: Radial Basis Function.

$$\text{RBFK}(v_{x1,y1}, v_{x2,y2}) = \exp\left(-\frac{\|v_{x1,y1} - v_{x2,y2}\|^2}{2\sigma^2}\right) \quad (4.6)$$

[138]: PEDREGOSA et al. (2011), « Scikit-learn : Machine Learning in Python »

Le réseau CNN-1D prend en entrée un tenseur contenant les 3 vecteurs $(v_{x,y}^r, v_{x,y}^s, v_{x,y}^b)$, sans nécessité de concaténation. L'architecture est composée de couches successives de convolutions 1D avec des activations ReLU⁹ et suivies d'opérations de pooling maximum¹⁰ pour la réduction de dimensions. Finalement une couche linéaire dense avec activation *softmax*¹¹ produit un score pour chacune des trois classes possibles : arrière-plan, cytoplasme et noyau. Une illustration du modèle est présentée Figure 4.7.

9: $\text{ReLU}(x) = \max(0, x)$

10: Également noté *maxpool* dans la suite.

11: $\text{softmax}(x)_j = \frac{\exp(x_j)}{\sum_i \exp(x_i)}$

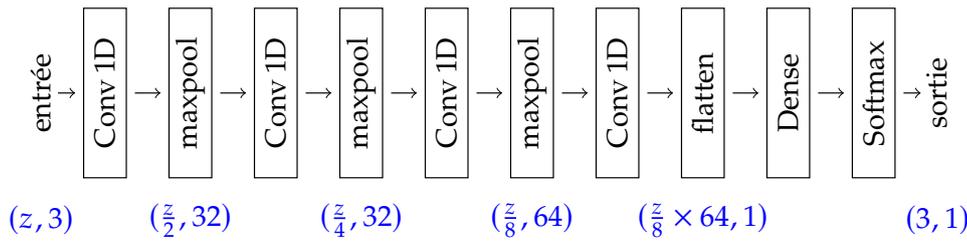


FIGURE 4.7 – Architecture du réseau CNN-1D. Le réseau prend en entrée un ensemble de trois vecteurs de profil-z, un pour chaque canal de couleur. La taille des tenseurs à différentes étapes est rapportée en bleu.

4.1.7 Résultats

Les méthodes CNN-1D, LRC et SVC sont entraînées sur les jeux de données Uro et Cervix67seg sous-échantillonnés en minimisant la fonction de coût \mathcal{L}_{CCE} . Les hyperparamètres des modèles sont choisis en fonction des performances sur le jeu de données Dev, puis les performances sont finalement évaluées sur l'ensemble de données de Test.

Puisque les paramètres du CNN-1D sont optimisés par descente de gradients avec des batchs de taille limitée, le modèle peut être entraîné sans difficulté avec une grande quantité de données. Le modèle peut être conçu pour prendre des z-stacks complètes en entrée et prédire en parallèle une classe pour chaque ensemble de vecteurs $(v_{x,y}^r, v_{x,y}^g, v_{x,y}^b)$. Non seulement ces propriétés permettent d'évaluer les différences de performances induites par l'entraînement sur des ensembles de données plus importants, mais également l'impact de l'utilisation d'une fonction de coût basée sur l'image plutôt que sur les pixels individuellement. Les performances des méthodes évaluées sur les jeux de données Uro et Cervix67seg sont rapportées dans le Tableau 4.4.

TABLE 4.4 – Performances de segmentation sur le jeu de Test pour les modèles entraînés sur le jeu de données Uro et sur le jeu de données Cervix67seg. Pour une comparaison plus approfondie, le CNN-1D est entraîné sur des ensembles de données non échantillonnés en minimisant \mathcal{L}_{CCE} et \mathcal{L}_{Dice} . L'entraînement sur le jeu de données Uro non-échantillonné améliore considérablement les performances du CNN-1D. Probablement parce que le jeu de données Cervix67seg non sous-échantillonné est massivement déséquilibré vis-à-vis des pixels n'appartenant pas à la classe noyaux, l'entraînement du CNN-1D sur celui-ci nuit aux performances.

modèle	Uro		Cervix67seg			
	sous-éch.	Fonction coût	moy. Exactitude $\pm\sigma$	moy. Dice $\pm\sigma$	moy. Exactitude $\pm\sigma$	moy. Dice $\pm\sigma$
SVM	✓	\mathcal{L}_{CCE}	0.934 \pm 0.031	0.885 \pm 0.057	0.964 \pm 0.036	0.840 \pm 0.099
LR	✓	\mathcal{L}_{CCE}	0.930 \pm 0.031	0.875 \pm 0.058	0.959 \pm 0.042	0.829 \pm 0.105
CNN-1D	✓	\mathcal{L}_{CCE}	0.942 \pm 0.030	0.892 \pm 0.053	0.968 \pm 0.031	0.845 \pm 0.092
CNN-1D		\mathcal{L}_{CCE}	0.955 \pm 0.028	0.873 \pm 0.042	0.953 \pm 0.045	0.803 \pm 0.105
CNN-1D		\mathcal{L}_{Dice}	0.954 \pm 0.027	0.913 \pm 0.048	0.951 \pm 0.052	0.805 \pm 0.105

Toutes les méthodes entraînées et évaluées sur le jeu de données Uro sous-échantillonné produisent de bonnes segmentations des cellules par rapport à l'arrière-plan. Cependant, la segmentation des noyaux obtenues par classification des profils-z est peu précise (voir Figure 4.8). Parmi tous les modèles testés, la méthode CNN-1D proposée est la plus performante sur le jeu de données Uro. Pour le jeu de données Uro, la plupart des erreurs sont situées autour des délimitations entre noyaux et cytoplasmes. Ces régions de délimitation inter-classes sont les plus difficiles à segmenter précisément, et les profils-z correspondants n'ont pas été sélectionnés dans le jeu de données sous-échantillonné afin d'éviter l'inclusion de données mal annotées¹².

Pour le jeu de données Cervix67seg, les segmentations obtenues sont beaucoup plus bruitées, comme illustré Figure 4.8. Cela illustre la difficulté de la segmentation de noyaux à l'aide de profils-z en présence d'amas de cellules superposées. Une fois encore, parmi toutes les méthodes testées, la méthode CNN-1D permet d'obtenir les segmentations les plus précises en moyenne.

12: Le détail de la stratégie de hard mining est décrit sous-section 4.1.5.

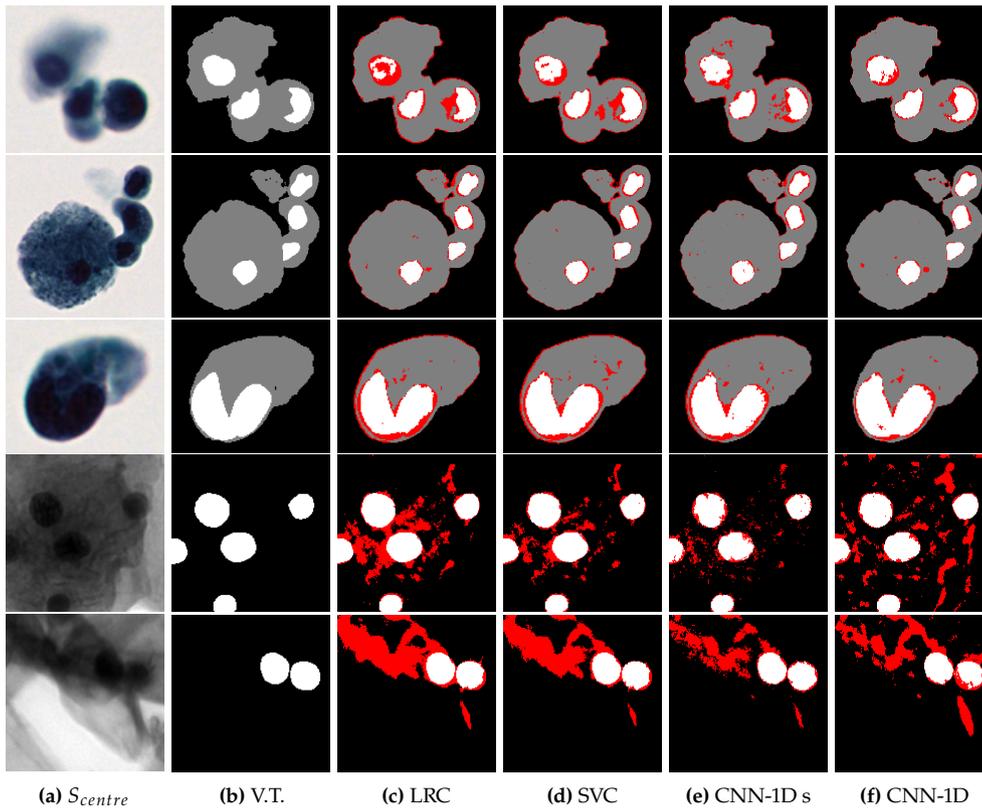


FIGURE 4.8 – Segmentation de régions cellulaires (arrière-plan en noir, cytoplasme en gris et noyaux en blanc dans (b)). Pour les modèles LRC (c), SVC (d) et CNN-1D (e) entraînés sur les jeux de données Uro et Cervix67seg sous-échantillonnés à l’aide de la fonction de coût \mathcal{L}_{CCE} . Pour le modèle CNN-1D (f) entraîné sur les jeux de données Uro et Cervix67seg complets à l’aide de la fonction de coût \mathcal{L}_{Dice} . Les erreurs de segmentation par rapport à la vérité terrain sont représentées en rouge.

Ces expériences permettent également d’étudier l’influence de l’utilisation de jeux de données sous-échantillonnés sur les performances à l’aide du modèle CNN-1D entraîné pour les jeux de données Uro et Cervix67seg respectivement sous-échantillonnés et complets.

Pour le jeu de données Uro, le modèle CNN-1D entraîné en minimisant \mathcal{L}_{Dice} , décrite Équation 4.4, permet d’augmenter à la fois l’Exactitude moyenne et le score de Dice moyen des segmentations obtenues pour les exemples de Test lorsque toutes les données sont utilisées¹³. Au contraire, l’entraînement du modèle CNN-1D sur jeu de données Cervix67seg conduit à une baisse des performances de segmentation. Contrairement au jeu de données Uro, le jeu de données Cervix67seg ne contient que deux classes : *noyau* et *arrière-plan*. Ainsi, le jeu de données complet contient beaucoup plus d’échantillons de la classe arrière-plan et il semble que ce déséquilibre soit à l’origine d’une baisse de performances du modèle.

13: Voir Tableau 4.4.

À l’exception des régions de délimitation inter-classe, la plupart des erreurs dans toutes les méthodes testées et pour les deux jeux de données concernent de petits groupes de pixels sombres classés à tort comme noyau. Bien que l’intensité et la couleur de ces pixels ressemblent à celles de noyaux, une analyse visuelle rapide des pixels voisins suffit à un observateur humain pour classer correctement ces régions comme faisant partie du cytoplasme. Cette observation motive l’utilisation combinée de caractéristiques spatiales en plus des informations extraites de profils-z.

4.1.8 Discussion

Les expériences menées dans cette section montrent qu’il est possible d’utiliser directement des profils-z pour la classification de pixels, et donc pour la segmentation de régions cellulaires.

La méthode CNN-1D proposée dans cette section et appliquée aux profils-z d'intensité, permet d'obtenir de meilleures performances pour la segmentation sémantique de régions cellulaires sur deux jeux de données distincts¹⁴ que les méthodes [136] et [137], qui reposent également sur l'analyse de profils-z.

Cependant, parce que les approches considérées sont des méthodes de classification de pixels individuels, l'information spatiale contenue dans les z-stacks n'est pas utilisée. Ainsi, les masques de segmentation obtenus restent très bruités, surtout en présence d'objets cellulaires fortement superposés comme dans le jeu de données Cervix67seg.

L'utilisation de profils-z semble donc être une approche prometteuse pour l'extraction de l'information contenue dans les plans focaux des z-stacks mais n'est pas suffisante.

4.2 Segmentation sémantique de noyaux à partir de z-stacks

Dans la section précédente, il a été montré que l'information extraite de profils-z peut être pertinente pour segmenter des régions cellulaires. Néanmoins, le manque d'information spatiale conduit à l'obtention de masques de segmentation bruités. Dans cette section, plusieurs méthodes de segmentation sémantique appliquées directement à des z-stacks sont proposées et évaluées sur une tâche de segmentation de noyaux à l'aide du jeu de données Cervix67seg défini section 4.1.

4.2.1 État de l'art

En plus des travaux précédemment cités section 4.1, dans [139] et dans [140], les auteurs ont utilisé des réseaux de neurones convolutionnels afin de prédire les cartes de probabilité des cellules non-colorées à partir de z-stacks. Dans ces deux études, la z-stack a été considérée comme une image avec z canaux couleurs, chaque canal étant un plan focal de la z-stack. Dans [139], les cartes de probabilité prédites sont utilisées pour obtenir des segmentations sémantique 2D des cellules à l'aide de méthodes de "bassins versants". Dans [140], les prédictions du modèle sont utilisées afin de détecter automatiquement les cellules non-colorées.

Comme les cellules utilisées en cytopathologie digitale sont généralement colorées¹⁵ pour mettre en valeur les biomarqueurs, le problème abordé ici diffère de la plupart des travaux précédemment cités qui se sont concentrés sur le matériel cellulaire non coloré. Les travaux [135] portent sur des cellules cervicales colorées, cependant la détection et la segmentation des noyaux sont basées sur l'hypothèse de noyaux elliptiques. Cette hypothèse est raisonnable pour les cellules cervicales, mais peu généralisable et notamment invalide pour les cellules urothéliales basales, dont la forme irrégulière du noyau peut être un critère morphologique d'atypie [24] utilisé pour le diagnostic des HGUCs.

4.2.2 Méthodes de segmentation sémantique

Base Le modèle U-Net [16] permet d'obtenir d'excellentes performances dans les tâches de segmentation sémantique des noyaux. Toutes les méthodes considérées dans cette section sont donc basées sur cette méthode. Différentes stratégies permettant de tirer parti des informations disponibles dans les z-stacks sont décrites ci-après. Une illustration du modèle U-Net utilisé dans ce travail est donnée Figure 4.9.

14: Uro et Cervix67seg.

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

[137]: LUGAGNE et al. (2018), « Identification of individual cells from z-stacks of bright-field microscopy images »

[139]: SADANANDAN et al. (2017), « Automated Training of Deep Convolutional Neural Networks for Cell Segmentation »

[140]: LIIMATAINEN et al. (2019), « Iterative unsupervised domain adaptation for generalized cell detection from brightfield z-stacks »

[139]: SADANANDAN et al. (2017), « Automated Training of Deep Convolutional Neural Networks for Cell Segmentation »

[140]: LIIMATAINEN et al. (2019), « Iterative unsupervised domain adaptation for generalized cell detection from brightfield z-stacks »

15: Avec les colorations Papanicolaou ou H&E.

[135]: WANG et al. (2020), « Nucleus Segmentation of Cervical Cytology Images Based on Depth Information »

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[16]: RONNEBERGER et al. (2015), « U-Net »

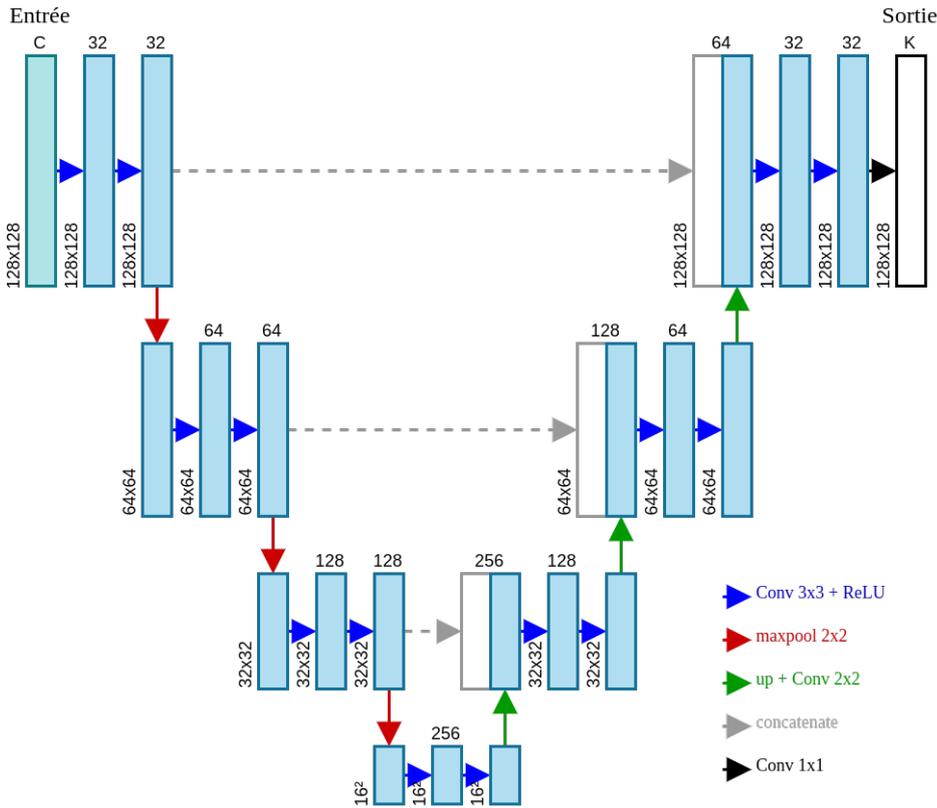


FIGURE 4.9 – Modèle U-Net utilisé comme base pour toutes les méthodes évaluées dans cette section.

Référence 2D La référence 2D utilisée est un modèle U-Net, tel que décrit Figure 4.9 sur la section centrale S_{centre} de chaque z-stack, sans utiliser le reste de l'information disponible. La section centrale S_{centre} est la section la plus nette selon l'autofocus du scanner de lames, et devrait donc contenir en moyenne plus d'informations que les autres sections de la z-stack. Les performances de ce modèle sont utilisées pour quantifier le gain de l'utilisation de toutes les informations présentes dans les z-stacks par rapport à l'utilisation d'une seule section.

Analyse 2D après projection EDF Le prétraitement par extension de profondeur de champ (EDF) est une méthode utilisée pour projeter l'information contenue dans des z-stacks, créant une image apparaissant entièrement en focus. Une fois le prétraitement effectué, le modèle U-Net-2D décrit précédemment peut être appliqué directement sur les images EDF.

Z-stacks considérées comme un volume 3D : U-Net-3D Une première stratégie pour extraire directement l'information des z-stacks, sans prétraitement, consiste à considérer les z-stacks comme des volumes, c'est-à-dire pour un CNN, comme un tenseur \mathcal{T} de dimensions (N, C, Z, H, W) , avec N le nombre de z-stacks dans un batch ¹⁶, C canaux de couleurs, Z plans focaux dans la z-stack, (H, W) la hauteur et la largeur des sections optiques. L'utilisation de tenseurs 5D dans un réseau de neurones convolutionnel se fait facilement à l'aide de convolutions 3D.

Un modèle, nommé ci-après U-Net-3D, est le modèle U-Net défini Figure 4.9, où les opérateurs de convolution 2D de l'encodeur et du décodeur sont remplacés par des convolutions 3D. Afin de produire une carte de segmentation 2D à partir des cartes de

16: Cette dimension supplémentaire, correspondant à la taille de batch choisie permet notamment le traitement en parallèle de plusieurs échantillons dans les réseaux de neurones. Une unique z-stack est de dimensions (C, Z, H, W) .

caractéristiques 3D, une opération de pooling le long de l'axe z précède un opérateur de convolution 2D final. L'opération de pooling utilisée ici est un pooling moyen adaptatif :

$$T_{pool} \in \mathbb{R}^{N \times C \times H \times W} \mid T_{pool}(n, c, h, w) = \frac{1}{Z} \sum_{z=1}^Z T(n, c, z, h, w) \quad (4.7)$$

En utilisant des convolutions 3D, le CNN peut calculer des cartes de caractéristiques basées sur la combinaison simultanée d'information provenant de profils-z et d'information spatiale. Les dimensions des z-stacks sont généralement anisotropes, la hauteur et la largeur étant souvent supérieures au nombre de plans. Pour tenir compte de ces propriétés, les opérations de réduction de dimensions¹⁷ dans l'encodeur, et d'augmentation de dimensions dans le décodeur du modèle U-Net sont effectuées uniquement pour les dimensions h et w des tenseurs. Une illustration est proposée Figure 4.10.

17: Pooling maximum spatial.

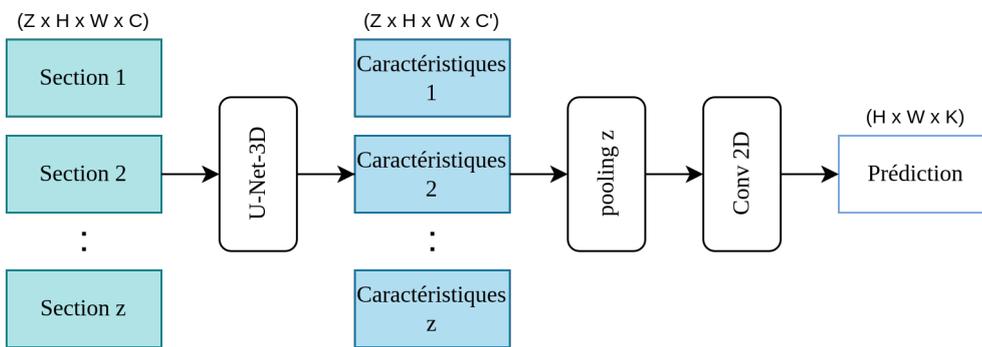


FIGURE 4.10 – Modèle U-Net-3D. L'architecture du U-Net-3D est similaire à celle du U-Net-2D défini précédemment, en remplaçant les convolutions 2D par des convolutions 3D.

Sections de la z-stack traitées en parallèle : U-Net-P Une deuxième stratégie consiste à appliquer le même extracteur de caractéristiques à toutes les sections de la z-stack, puis un opérateur de fusion le long de l'axe z et une couche de convolution 2D pour prédire les masques de segmentation finaux. Le modèle U-Net-2D (Figure 4.9) sans sa dernière couche de convolution, est utilisé comme extracteur de caractéristiques partagé entre les sections optiques.

Deux opérateurs de fusion sont considérés : une simple opération de pooling moyen le long l'axe z, similaire à Équation 4.7 et un opérateur BDCLSTM¹⁸ inspiré de [141] à la différence qu'ici, seuls les deux derniers états des branches LSTM sont conservés et concaténés le long de l'axe c afin d'utiliser le modèle BDLSTM effectivement comme un opérateur de fusion. Une illustration du modèle U-Net-P est proposée Figure 4.11.

18: Bi-Directional Convolutional Long Short-Term Memory.

[141]: CHEN et al. (2016), « Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation »

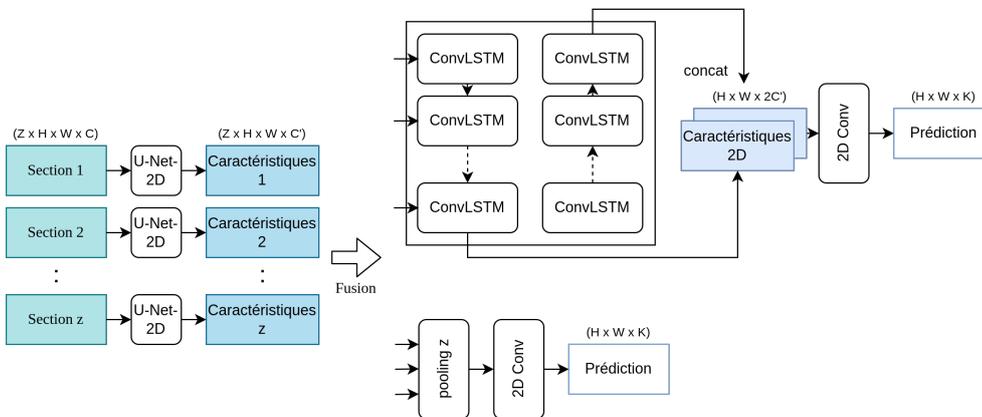


FIGURE 4.11 – Modèle U-Net-P. Deux opérateurs de fusion sont considérés : un modèle BDCLSTM (haut) et un opérateur de pooling moyen (bas).

Sections de la z-stack comme canaux de couleur additionnels : U-Net-CSE Inspirée par [140], une troisième stratégie consiste à considérer les sections d'une z-stack comme des canaux de couleur supplémentaires. Un modèle, nommé ci-après U-Net-CSE, est basé sur le modèle de base U-Net décrit Figure 4.9 auquel est ajoutée une opération de transformation des dimensions des tenseurs avant le premier bloc de convolution. Cette opération de transformation prend en entrée un tenseur T de dimensions (N, C, Z, H, W) ressort un tenseur T' de dimensions $(N, C \times Z, H, W)$. La première couche de convolution du modèle U-Net est modifiée pour prendre en entrée un tenseur avec $N \times C$ canaux au lieu de C canaux.

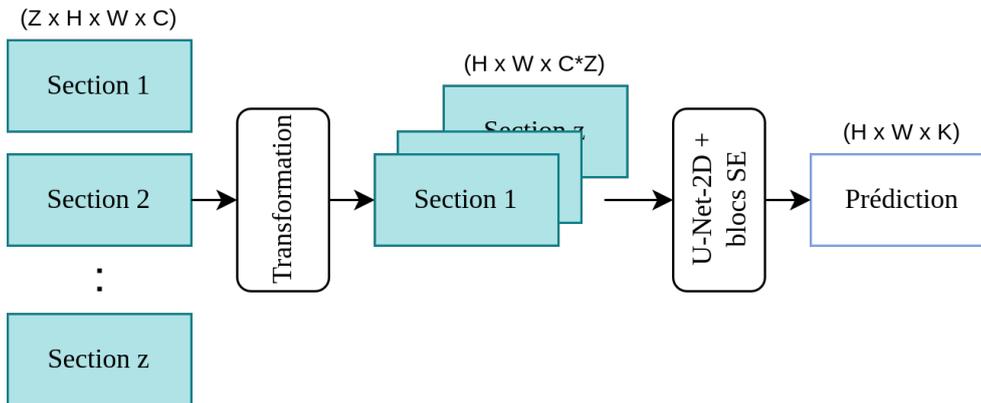


FIGURE 4.12 – Modèle U-Net-CSE.

Comme toutes les sections de la z-stack ne contiennent pas la même quantité d'information, tous les canaux ne devraient pas avoir la même importance dans l'architecture de réseau proposée. Dans [142], les auteurs ont introduit un bloc Squeeze and Excitation (SE) agissant comme un bloc de re-calibration sur l'importance des canaux pour sélectionner les caractéristiques les plus informatives. Ces blocs SE, illustrés dans Figure 4.13, sont ajoutés au modèle U-Net-2D¹⁹ après chaque bloc de convolution : l'intuition est que les blocs SE permettent de sélectionner les caractéristiques les plus utiles à partir des sections de la z-stack contenant le plus d'information.

[142]: HU et al. (2018), « Squeeze-and-Excitation Networks »

19: Voir Figure 4.9.

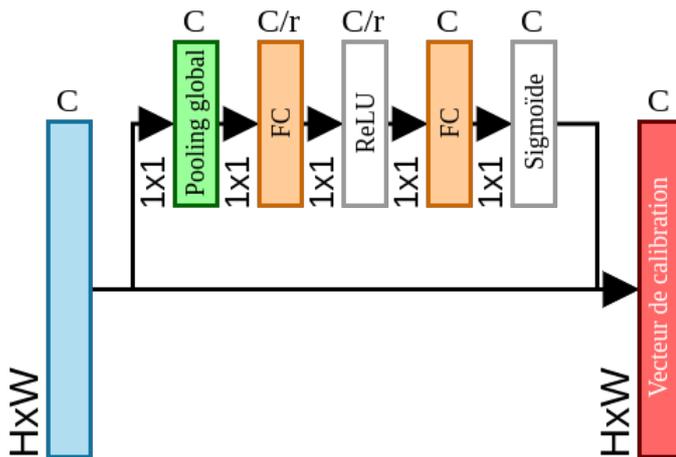


FIGURE 4.13 – Bloc Squeeze and Excitation (SE) permettant de recalibrer l'importance des canaux après les couches de convolutions du modèle U-Net-2D. Dans cette étude, le bloc SE est utilisé avec un facteur de réduction $r = 1$.

4.2.3 Métriques d'évaluation

Métriques pour l'évaluation de la détection des noyaux Le jeu de données Cervix93 a été conçu à l'origine pour la détection des noyaux de cellules cervicales. Tous les résultats des méthodes considérées sont des masques de noyaux. En suivant la méthodologie appliquée dans [132], un noyau est considéré comme correctement détecté²⁰ si le point

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

20: Vrai positif.

annoté se trouve à l'intérieur du masque de segmentation prédit. Un point annoté sans masque de segmentation associé est une détection manquée ²¹. Un masque prédit sans point annoté associé est un faux positif. Un masque prédit contenant deux points annotés ou plus, compte comme un vrai positif, et tous les points annotés restant comptent comme des faux négatifs.

Étant donné le nombre de vrais positifs (TP), de faux positifs (FP) et de faux négatifs (FN) pour une lame digitale, trois mesures sont définies pour évaluer la détection des noyaux : la Précision (AP), le Rappel (AR) et le score F1.

$$AP = \frac{TP}{TP + FP} \quad (4.8)$$

$$AR = \frac{TP}{TP + FN} \quad (4.9)$$

$$F1 = 2 \times \frac{AP \times AR}{AP + AR} \quad (4.10)$$

La Précision donne la proportion de noyaux corrects détectés parmi tous les masques de noyaux prédits, le Rappel indique la proportion de noyaux correctement détectés parmi tous les noyaux présents dans les images et le score F1 agrège les métriques Précision et Rappel pour faciliter la comparaison des méthodes.

Métriques pour l'évaluation de la segmentation des noyaux Le score de Dice, défini Équation 4.1 ²², est à nouveau utilisé ici pour évaluer la segmentation sémantique des noyaux.

Le score de Dice est une mesure de similarité couramment utilisée pour évaluer les problèmes de segmentation. Par définition, le score de Dice prend en compte le déséquilibre des classes et convient bien à la tâche de segmentation sémantique considérée, où 95% des pixels sont étiquetés comme arrière-plan.

4.2.4 Résultats

Tous les modèles ont été développés à l'aide de Pytorch [126], entraînés sur chaque split Train/Dev de l'ensemble de données d'entraînement à l'aide de l'optimiseur Adam avec un learning rate de 1e-4, des batchs de taille 4, et pour un maximum de 200 epochs avec un arrêt précoce de l'optimisation si aucune amélioration sur l'ensemble de données Dev n'est observée pendant 20 epochs. Les modèles ont été entraînés pour minimiser la fonction de coût \mathcal{L}_{Dice} ²³.

Pour effectuer l'inférence sur les lames de cytologie complètes, des z-stacks de dimensions (20, 128, 128) pixels sont extraites de chaque lame avec un chevauchement de 64 pixels selon les axes h et w . L'inférence est effectuée sur chaque z-stack, puis les prédictions obtenues sont assemblées et normalisées. À l'exception des pixels en bords de lame, quatre prédictions sont donc produites par pixel en raison du chevauchement des z-stacks extraites des lames. La prédiction finale au niveau de la lame est obtenue par post-traitement : un seuil de décision t de 0.75 sur la prédiction moyenne des pixels est appliqué.

L'hyper-paramètre du seuil de décision est fixé expérimentalement à 0.75 : pour $t \in [0.25, 0.5, 0.75, 1]$, les performances de détection du modèle U-Net-CSE sont évaluées au niveau de la lame, et reportées Tableau 4.5. Le choix de ce seuil de décision t a un impact important sur les mesures de Précision et de Rappel. En effet, un seuil t faible conduit à

21: Faux négatif.

$$22: \text{Dice} = \frac{2|Y_{\text{pred}} \cap Y_{\text{true}}|}{|Y_{\text{pred}}| + |Y_{\text{true}}|}$$

[126]: PASZKE et al. (2019), « PyTorch : An Imperative Style, High-Performance Deep Learning Library »

23: $\mathcal{L}_{Dice} = 1 - \text{Score de Dice}$

plus de détection de noyaux, conduisant à plus de faux positifs, moins de faux négatifs et donc une mesure de Rappel haute. À l'inverse, un seuil t élevé conduit à moins de détection de noyaux, donc moins de faux positifs et une mesure de Précision haute.

t	split	AP	AR	score F1	score de Dice
0.25	1	0.768 ± 0.099	0.950 ± 0.047	0.846 ± 0.066	0.845 ± 0.063
	2	0.685 ± 0.121	0.959 ± 0.049	0.792 ± 0.085	0.810 ± 0.072
	3	0.774 ± 0.098	0.942 ± 0.053	0.847 ± 0.070	0.846 ± 0.061
0.5	1	0.831 ± 0.099	0.942 ± 0.048	0.880 ± 0.061	0.868 ± 0.051
	2	0.746 ± 0.115	0.955 ± 0.051	0.832 ± 0.077	0.846 ± 0.057
	3	0.824 ± 0.092	0.930 ± 0.065	0.870 ± 0.062	0.867 ± 0.055
0.75	1	0.874 ± 0.075	0.920 ± 0.055	0.894 ± 0.049	0.869 ± 0.047
	2	0.861 ± 0.091	0.935 ± 0.054	0.894 ± 0.061	0.870 ± 0.051
	3	0.872 ± 0.078	0.920 ± 0.066	0.893 ± 0.056	0.875 ± 0.048
1	1	0.863 ± 0.088	0.889 ± 0.073	0.873 ± 0.069	0.844 ± 0.053
	2	0.844 ± 0.111	0.917 ± 0.074	0.876 ± 0.082	0.860 ± 0.051
	3	0.875 ± 0.084	0.905 ± 0.065	0.887 ± 0.061	0.860 ± 0.053

TABLE 4.5 – Performances de segmentation et de détection de noyaux, pour différents seuils t , évalués pour la méthode U-Net-CSE.

Les performances de détection et de segmentation des noyaux obtenues pour toutes les méthodes sont présentées dans la suite.

Évaluation de la segmentation sémantique des noyaux La segmentation des noyaux est évaluée sur l'ensemble Test du jeu de données Cervix67seg, à la fois en considérant uniquement les z-stacks extraites des lames et les lames complètes. L'évaluation effectuée sur les z-stacks de Test permet la comparaison directe entre les méthodes, indépendamment de la méthode de post-traitement choisie, tandis que l'évaluation sur les lames complètes de Test correspond au cas réel d'utilisation des méthodes de segmentation et permet la comparaison avec la méthode de référence introduite par [132]. L'évaluation sur des lames de Test dépend donc des performances de segmentation et de détection des noyaux, mais aussi de la stratégie utilisée pour agréger les prédictions obtenues à partir des z-stacks, comme décrit précédemment. Les performances sont rapportées dans Tableau 4.6.

modèle	moy. Dice ± σ (z-stack)	top Dice (z-stack)	moy. Dice ± σ (lames)
référence [132] †	-	-	0.740 ± 0.094
U-Net-2D (S_{centre})	0.927 ± 0.001	0.929	0.835 ± 0.065
U-Net-2D (S_{edf})	0.940 ± 0.002	0.942	0.884 ± 0.053
U-Net-CSE	0.944 ± 0.001	0.944	0.875 ± 0.048
U-Net-P (pooling moy.)	0.943 ± 0.003	0.947	0.876 ± 0.051
U-Net-P (BDCLSTM)	0.946 ± 0.001	0.947	0.881 ± 0.050
U-Net-3D	0.944 ± 0.003	0.947	0.888 ± 0.045

[132]: PHOULADY et al. (2018), A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection

TABLE 4.6 – Performances de segmentation sémantique sur les z-stacks puis les lames complètes du jeu de données Cervix67seg. L'écart-type sur les sous-volumes est calculé sur les trois splits Train/Dev. Les performances sur les lames complètes sont rapportées pour le modèle le plus performant évalué parmi les trois splits. L'écart-type reporté pour les lames complètes est calculé sur l'ensemble des lames. † : notre implémentation.

Il est observé, à partir des z-stacks de Test, que toutes les approches testées pour la segmentation de noyaux en utilisant directement l'information des z-stacks ont des performances égales ou légèrement supérieures à celles d'un modèle U-Net-2D entraîné sur des images EDF. Le grand écart de performances observé avec le modèle U-Net-2D entraîné sur des sections centrales S_{centre} indique que les méthodes proposées sont effectivement capables d'extraire directement de l'information d'intérêt à partir des z-stacks. C'est également une confirmation que l'utilisation de l'information contenue dans les z-stacks est nécessaire afin d'obtenir une bonne segmentation du noyau dans les amas de cellules cervicales.

Le modèle U-Net-3D obtient les meilleures performances de segmentation pour les lames complètes. Une fois de plus, toutes les méthodes utilisant l'information extraite de z-stacks sont plus performantes que le modèle U-Net entraîné uniquement sur les sections centrales (voir Figure 4.14). De plus, toutes les méthodes CNN explorées dans ce travail surpassent largement les performances de la méthode de référence introduite par [132]. Il est cependant à noter que la méthode de référence, même si elle produit des masques de noyaux, a été optimisée pour la détection des noyaux.

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

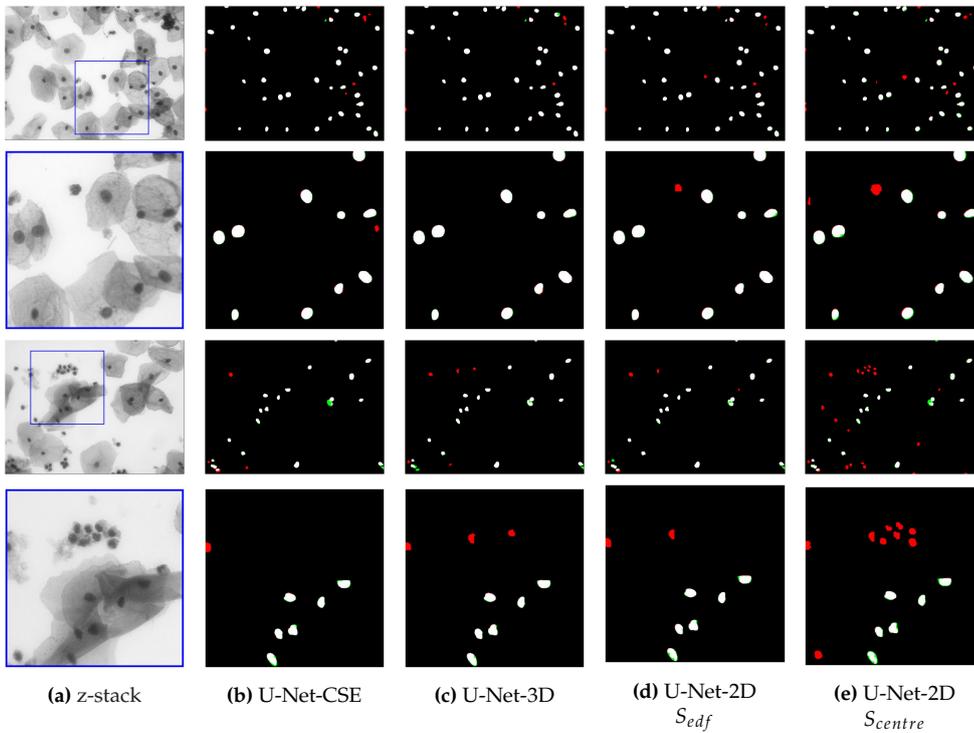


FIGURE 4.14 – Segmentation des noyaux à partir de z-stacks, à l'aide des méthodes U-Net-CSE, U-Net-3D et U-Net-2D. Les pixels blancs sont correctement classés comme noyaux, les pixels rouges sont des faux positifs et les pixels verts sont des faux négatifs. La plupart des erreurs sont des noyaux de cellules neutrophiles mal classés en tant que noyaux de cellules cervicales. L'approche U-Net-2D appliquée aux sections centrales S_{centre} manque de précision et a notamment tendance à segmenter la plupart des zones sombres comme noyaux de cellules cervicales.

Évaluation de la détection des noyaux La détection des noyaux est effectuée sur les lames complètes. Pour chaque méthode, le meilleur modèle parmi les trois splits Train/Dev est sélectionné à partir des performances calculées sur l'ensemble de données Dev. Les performances de ce meilleur modèle sur l'ensemble de données de Test sont reportées dans Tableau 4.7.

La méthode U-Net-3D est la plus performante pour la détection de noyaux, avec une précision moyenne de 0.887, un rappel moyen de 0.925 et un score F1 moyen de 0.903. Il est à noter que toutes les méthodes entraînées pour la segmentation sémantique des noyaux ont des performances égales ou supérieures à celles de la précédente meilleure méthode sur ce jeu de données²⁴ pour le problème de la détection des noyaux.

La précision inférieure obtenue avec le modèle U-Net-2D entraîné sur des sections centrales S_{centre} indique que l'utilisation de données tridimensionnelles, par projection EDF ou directement à partir des z-stacks, permet non seulement d'obtenir des masques de segmentation plus précis, mais permet également d'améliorer la détection des noyaux d'intérêt.

4.2.5 Évaluation du temps d'inférence

L'analyse de z-stacks au lieu d'images bi-dimensionnelles induit un coût de traitement plus élevé, car plus de calculs sont nécessaires. Cependant, il a été démontré sous-

24: Méthode notée CNN, introduite [132].

modèle	moy. Précision $\pm \sigma$	moy. Rappel $\pm \sigma$	moy. score F1 $\pm \sigma$	score F1
U-Net-2D (S_{centre})	0.831 \pm 0.105	0.929 \pm 0.063	0.873 \pm 0.073	0.877
Lu et al. [143] ‡	0.617 \pm 0.258	0.295 \pm 0.274	-	0.399
Ushizima et al. [143] ‡	0.687 \pm 0.217	0.446 \pm 0.230	-	0.541
référence [132] ‡	0.803 \pm 0.137	0.838 \pm 0.128	-	0.820
référence †	0.800 \pm 0.142	0.841 \pm 0.129	0.808 \pm 0.111	0.820
CNN [132] ‡	0.861 \pm 0.095	0.895 \pm 0.086	-	0.878
U-Net-2D (EDF)	0.876 \pm 0.089	0.914 \pm 0.070	0.892 \pm 0.069	0.895
U-Net-CSE	0.872 \pm 0.078	0.920 \pm 0.066	0.893 \pm 0.056	0.895
U-Net-P (pooling moy.)	0.848 \pm 0.108	0.932 \pm 0.058	0.885 \pm 0.078	0.888
U-Net-P (BDCLSTM)	0.889 \pm 0.087	0.918 \pm 0.069	0.902 \pm 0.068	0.903
U-Net-3D	0.887 \pm 0.085	0.925 \pm 0.057	0.903 \pm 0.061	0.906

section 4.2.4 que l'utilisation de données tri-dimensionnelles est nécessaire pour obtenir de bonnes performances de détection et de segmentation. Ce surcoût de calcul doit être comparé à la stratégie alternative : le temps de prétraitement par extension de profondeur de champ (EDF).

Le temps moyen de prétraitement EDF est calculé en évaluant le temps total nécessaire pour effectuer une projection EDF sur les 67 z-stacks²⁵, normalisé par le nombre z-stacks traitées. Le pré-traitement EDF utilisé est la méthode proposée par [111], car c'est le prétraitement utilisé pour produire des images EDF dans le jeu de données original Cervix93 [132]. L'algorithme est exécuté sur un ordinateur équipé d'un processeur Intel Core i7-8750H avec 32 Go de RAM.

Le temps nécessaire à un CNN pour traiter une z-stack de dimensions (20, 128, 128) est calculé en évaluant le temps d'inférence total pour les z-stacks de l'ensemble Train²⁶, normalisé par le nombre total de z-stacks. Le traitement a été effectué sur un ordinateur équipé d'un unique GPU NVIDIA Titan Xp. Le temps de traitement pour une lame complète de dimensions (20, 960, 1280), nécessitant le traitement de 266 z-stacks de dimensions (20, 128, 128) avec chevauchement, est ensuite déduit du temps de traitement nécessaire pour analyser une z-stack.

Pour le prétraitement (EDF) et le traitement (CNN), le temps nécessaire pour charger les z-stacks ou les images en mémoire à partir du disque n'a pas été pris en compte dans l'analyse afin d'éviter les variations de temps non liées aux méthodes évaluées. Cependant, le temps de transfert des données entre le CPU et le GPU pendant le traitement est inclus car utilisé par tous les modèles CNN pour accélérer le temps de calcul.

Les temps d'inférences obtenus sont reportés Tableau 4.8, accompagnés du nombre d'opérations et du nombre de paramètres de chaque modèle.

modèle	pré-traitement temps/z-stack (s)	traitement GPU TitanXP temps/z-stack (s)	total (s)	GMACs	# Params
U-Net-2D (EDF) †	4.47	4.76	9.23	3.1	2.1M
U-Net-CSE	0	5.06	5.06	3.2	2.4M
U-Net-P (pooling moy.)	0	7.77	7.77	62.5	2.1M
U-Net-P (BDCLSTM)	0	13.1	13.1	80.7	2.2M
U-Net-3D	0	14.1	14.1	149.9	5.6M

TABLE 4.7 – Performances de détection de noyaux sur les lames de Test. Les performances sur l'ensemble de données de Test sont reportées ici pour le modèle avec les meilleurs performances parmi les trois splits sur l'ensemble de données Dev. † : notre implémentation. ‡ : reporté directement de [132].

25: Chacune composée de 20 plans focaux.

[111]: BRADLEY et al. (2004), « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform »

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

26: Cela représente 3352 z-stacks.

TABLE 4.8 – Évaluation du temps d'inférence, incluant le temps de pré-traitement, si applicable, et le temps de traitement correspondant à la prédiction des masques de segmentation. Sont également reportés le nombre d'opérations et le nombre de paramètres associés aux modèles évalués. Les temps totaux en bleu correspondent aux temps inférieurs à la méthode †.

La méthode U-Net-CSE, dont les performances sont similaires à celles de l'approche U-Net-2D (EDF), est l'approche la plus rapide. Le léger temps supplémentaire nécessaire au traitement des z-stacks est compensé par l'absence de prétraitement EDF, ce qui permet de réduire le temps de calcul total d'environ 45% par image. Les méthodes U-Net-3D et U-Net-P (BDCLSTM), qui se sont avérées plus performantes que les autres approches considérées, nécessitent toutefois une augmentation du temps de calcul d'environ 50%. La méthode U-Net-P (pooling moy.) nécessite 20 fois plus d'opérations que la méthode U-Net-2D (EDF), mais elle reste plus rapide sur le GPU testé.

4.2.6 Étude d'ablation et choix de design

Les choix de design des différentes méthodes présentées dans ce travail sont discutés dans cette section. Toutes les études d'ablation sont réalisées sur des z-stacks de l'ensemble de données de Test, à l'aide des métriques Exactitude et score Dice.

U-Net-CSE L'influence des blocs Squeeze and Excitation (SE) est évaluée par l'ablation des blocs SE dans le modèle U-Net-CSE. Le modèle obtenu, nommé ci-après U-Net-C, est similaire à la méthode proposée par [140].

modèle	moy. Exactitude $\pm \sigma$	moy. Dice $\pm \sigma$	top Exactitude	top Dice
U-Net-C	0.987 \pm 0.001	0.937 \pm 0.002	0.988	0.942
U-Net-CSE	0.989 \pm 0.000	0.944 \pm 0.001	0.989	0.944

[140]: LIIMATAINEN et al. (2019), « Iterative unsupervised domain adaptation for generalized cell detection from brightfield z-stacks »

TABLE 4.9 – Ablation des blocs SE.

U-Net-P Plusieurs opérateurs de fusion selon l'axe z ont été testés pour les modèles U-Net-P, dont les opérateurs pooling maximum²⁷, pooling moyen²⁸, le modèle LSTM convolutionnel (CLSTM) et le modèle LSTM convolutionnel bi-directionnel (BDCLSTM).

27: Noté pooling max.

28: Noté pooling moy.

modèles	moy. Exactitude $\pm \sigma$	moy. Dice $\pm \sigma$	top Exactitude	top Dice
U-Net-P (pooling max.)	0.988 \pm 0.000	0.940 \pm 0.001	0.988	0.942
U-Net-P (pooling moy.)	0.988 \pm 0.001	0.943 \pm 0.003	0.989	0.947
U-Net-P CLSTM	0.988 \pm 0.001	0.943 \pm 0.002	0.988	0.945
U-Net-P BDCLSTM	0.989 \pm 0.000	0.946 \pm 0.001	0.989	0.947

TABLE 4.10 – Comparaison des opérateurs de fusion selon l'axe z pour les modèles U-Net-P.

U-Net-3D Dans [144], les auteurs proposent d'utiliser des opérateurs de convolution 3D Séparables (S3D) pour tenir compte des dimensions anisotropes des données d'entrée et réduire le coût de calcul des opérateurs de convolution 3D classiques. Une convolution 3D Séparable vise à remplacer une convolution 3D de support $(k \times k \times k)$ en combinant une convolution 2D de support $(k \times k \times 1)$ suivie d'une convolution 1D de support $(1 \times 1 \times k)$. Le modèle U-Net-S3D est l'équivalent du modèle U-Net-3D où toutes les convolutions 3D sont remplacées par des convolutions 3D séparables.

[144]: CHEN et al. (2019), « HSN »

modèles	moy. Exactitude $\pm \sigma$	moy. Dice $\pm \sigma$	top Exactitude	top Dice
U-Net-S3D	0.980 \pm 0.003	0.920 \pm 0.007	0.982	0.927
U-Net-3D	0.989 \pm 0.001	0.944 \pm 0.003	0.990	0.947

TABLE 4.11 – Choix de design pour les opérateurs de convolution 3D.

Pour cette tâche de segmentation sémantique des noyaux de cellules cervicales à partir de z-stacks, les convolutions 3D séparables donnent de mauvais résultats par rapport aux convolutions 3D classiques.

4.2.7 Discussion : application au diagnostic de cancer du col

Certaines approches considérées, comme le modèle U-Net-P (BDCLSTM) et le modèle U-Net-3D, s'avèrent légèrement plus performantes que les autres méthodes en termes de détection et segmentation des noyaux, au prix cependant d'un coût de calcul supplémentaire important. Le compromis entre une méthode plus rapide et une méthode plus précise est établi en fonction de leur impact sur l'extraction de caractéristiques pour le diagnostic du cancer du col de l'utérus.

Le jeu de données Cervix67Seg, composé d'échantillons prélevés sur des patientes saines (N), des patientes présentant des lésions malpighiennes intra-épithéliales de bas-grade (LSILs) et des lésions malpighiennes intra-épithéliales de haut-grade (HSILs), est divisé en 43 lames d'entraînement²⁹ et 24 lames de test³⁰. Le problème de diagnostic considéré est une classification automatique des lames de cytologie cervicale afin de différencier les lames saines des lames cancéreuses. La taille des noyaux étant un critère important pour le diagnostic du cancer du col de l'utérus [145], une caractéristique simple est calculée comme la surface moyenne des cinq plus gros noyaux dans une lame. Cette caractéristique est calculée pour chaque lame, sur la base des masques de segmentation des noyaux obtenus par différentes méthodes.

En utilisant les caractéristiques calculées à partir des masques de segmentation des noyaux obtenus avec différentes approches de segmentation, une machine à vecteurs de support avec noyau RBF est appris sur les images d'entraînement, pour classer chaque lame comme saine ou cancéreuse. Une pondération des échantillons est appliquée pendant l'optimisation pour tenir compte du grand déséquilibre entre les classes. La SVM entraînée est ensuite évaluée sur les 24 images de Test. Les métriques de Précision, Rappel et score F1 caractérisant les performances de diagnostic obtenues sont reportées Tableau 4.12.

modèle	moy. Précision	moy. Rappel	moy. F1	valeur-p score F1 vs U-Net-2D S_{edf}
référence	0.947	0.900	0.923	-
U-Net-2D S_{centre}	0.967 ± 0.024	0.933 ± 0.024	0.949 ± 0.001	0.006*
U-Net-2D S_{edf}	1.000 ± 0.000	0.950 ± 0.000	0.974 ± 0.000	-
U-Net-CSE	0.983 ± 0.024	0.967 ± 0.024	0.975 ± 0.020	1.000
U-Net-P (pooling moy.)	0.984 ± 0.023	0.983 ± 0.024	0.983 ± 0.012	1.000
U-Net-P (BDCLSTM)	0.983 ± 0.024	0.950 ± 0.000	0.966 ± 0.011	1.000
U-Net-3D	1.000 ± 0.000	0.950 ± 0.000	0.974 ± 0.000	1.000

Sur le jeu de données considéré, d'excellentes performances de diagnostic sont rapportées avec au moins 95% de Rappel et 98% de Précision pour toutes les méthodes de segmentation des noyaux utilisant à la fois des informations spatiales et de profondeur. Les méthodes dont les performances de segmentation et de détection sont les plus faibles, comme le modèle U-Net-2D appliqué à S_{centre} ou la méthode de référence proposée dans [132], ont également des performances de diagnostic plus faibles.

Les distributions des scores F1, obtenues pour des modèles entraînés sur trois splits Train/Dev, sont utilisées pour tester si les différences observées sont statistiquement significatives par rapports aux scores F1 obtenus avec l'approche U-Net-2D S_{edf} . Le test statistique utilisé est le *student t-test* [129] corrigé pour les comparaisons multiples par la méthode de Bonferroni [146]. Les valeurs-p obtenues sont reportées Tableau 4.12.

Des différences significatives dans les performances de diagnostic sont observées entre les méthodes U-Net-2D S_{edf} et U-Net-2D S_{centre} , montrant qu'une extraction de caractéristiques plus précise³¹ conduit à de meilleures performances de diagnostic. Pour

29: 24 LSIL, 13 HSIL, 6 N.

30: 12 LSIL, 8 HSIL, 4 N.

[145]: NAYAR et al. (2015), *The Bethesda System for Reporting Cervical Cytology*

TABLE 4.12 – Performances de diagnostic du cancer du col de l'utérus, à partir de caractéristiques extraites de masques de segmentation. Les scores F1, de Précision et de Rappel sont moyennés pour des modèles entraînés sur trois splits Train/Dev.

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

[129]: STUDENT (1908), « The Probable Error of a Mean »

[146]: BONFERRONI (1936), *Teoria statistica delle classi e calcolo delle probabilita*

31: Ici la segmentation des noyaux.

toutes les autres méthodes, le test statistique ne permet pas de conclure à des différences significatives par rapport à l'approche U-Net-2D S_{edf} .

D'après ces observations, une légère amélioration des performances de segmentation ne semble pas améliorer les performances de diagnostic. Il est donc pertinent de privilégier une méthode minimisant le temps de traitement. De ce point de vue, la méthode U-Net-CSE proposée semble intéressante pour extraire de l'information à partir des z-stacks, car elle offre le meilleur compromis entre les performances de segmentation ou de détection des noyaux, les performances de diagnostic et le temps de calcul total.

Bien que toutes les stratégies d'extraction d'information pertinente à partir de plusieurs plans focaux présentées dans ce travail soient basées sur le modèle U-Net, elles sont génériques et peuvent être directement adaptées à de multiples architectures.

4.3 Segmentation d'instances cellulaires à partir de z-stacks

Cette section s'intéresse à la segmentation d'instances cellulaires dans des lames de cytologie digitales, une tâche primordiale pour le diagnostic assisté par ordinateur. La segmentation automatique d'instances cellulaires est un problème plus compliqué que les problèmes de segmentations sémantiques étudiés section 4.1 et section 4.2. En effet, en plus de prédire la classe de chaque pixel, un modèle de segmentation d'instances cherche également à identifier chaque objet individuellement. Les groupes de cellules partiellement superposées sont fréquents dans les lames de cytologie digitales : un unique pixel peut donc appartenir à plusieurs cellules, ce qui complique la tâche de segmentation d'instances.

Publication associée

Les travaux conduits dans cette section ont été décrits dans un article de conférence, accepté pour présentation à IEEE International Symposium on Biomedical Imaging (ISBI), 2022 :

A. Bouyssoux, R. Fezzani and J.-C. Olivo-Marin, "Cell Instance Segmentation Using Z-Stacks in Digital Cytology," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022.

Le code associé à ces travaux est disponible à l'adresse : gitlab.com/vitadx/articles/zstacks_cell_instance_segmentation

4.3.1 État de l'art

La segmentation d'instances de cellules superposées à partir de lames de cytologie digitales a été abordée de nombreuses fois, notamment au travers de deux compétitions proposées lors des conférences *International Symposium on Biomedical Imaging (ISBI)* en 2014 et 2015 [115].

Les jeux de données disponibles pour ces compétitions sont de taille relativement restreinte : 16 images réelles et 945 images synthétiques pour la compétition ISBI 2014 et 17 z-stacks pour la compétition ISBI 2015. En conséquence, les méthodes de segmentation d'instances proposées se sont concentrées sur des approches traditionnelles de traitement et d'analyse d'images [143].

[115]: ZHI LU et al. (2015), « An Improved Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells »

[143]: LU et al. (2017), « Evaluation of Three Algorithms for the Segmentation of Overlapping Cervical Cells »

La segmentation d'instances en général a bénéficié d'une grande attention ces dernières années, notamment grâce à la disponibilité de grands ensembles de données de haute qualité [147]. Les approches couramment utilisées pour la segmentation d'instances peuvent être regroupées en deux catégories : les méthodes qui reposent sur une détection préalable des objets d'intérêt, et les méthodes directes, qui ne nécessitent pas d'étape de détection.

Méthodes avec détection préalable d'objets d'intérêt Ces méthodes considèrent le problème de segmentation d'instances comme une tâche de détection d'instances, souvent à l'aide de boîtes englobantes, suivie d'une segmentation sémantique pour chaque instance détectée. Parmi ces approches on retrouve le très populaire modèle Mask-RCNN [46]. Récemment, la méthode ANCIS [148] a été proposée pour segmenter des instances de cellules neuronales, s'inspirant des idées introduites par la méthode Mask-RCNN et obtenant de meilleures performances de segmentation.

Méthodes sans détection préalable Ces méthodes ne reposent pas sur une détection préalable des objets d'intérêt. Parmi ces approches, on retrouve les modèles SOLO [149] et SOLOv2 [150], utilisés dans la suite. D'autres travaux s'appuient sur la segmentation sémantique des membranes cellulaires pour obtenir une segmentation d'instances à partir d'une segmentation sémantique et de post-traitement [151]. Ces méthodes de segmentation d'instances à partir de masques de segmentation sémantique ont démontré des bonnes performances pour segmentation d'objets en contact, cependant elles ne sont généralement pas adaptées à la segmentation d'instances en présence de cellules superposées.

4.3.2 Jeu de données

Disponibilité des données

Le jeu de données décrit dans cette section est disponible à l'adresse : <https://zenodo.org/record/5938893>

La cytologie digitale est une technique prometteuse permettant notamment le développement de méthodes de diagnostic assisté par ordinateur du cancer de la vessie. Une difficulté connue de la cytologie digitale est la répartition tri-dimensionnelle du matériel cellulaire dans les lames, rendant difficile l'analyse lorsqu'un seul plan focal est numérisé et généralement surmontée à l'aide d'acquisition de z-stacks (voir chapitre 3).

Cette section propose une évaluation de plusieurs architectures récentes de deep learning pour la segmentation d'instances de cellules superposées à partir de lames de cytologie digitales. Basée sur les résultats obtenus section 4.2, une modification simple de ces architectures est proposée afin de pouvoir prédire directement des masques de segmentation d'instances cellulaires à partir de z-stacks.

Pour ce faire, un nouveau jeu de données, nommé ci-après CISC³² est introduit. Il est composé de 3911 échantillons contenant au moins deux cellules urothéliales en contact ou superposées. Les masques d'instances cellulaires 2D ont été annotés manuellement à partir des projections EDF des z-stacks. Tous les échantillons sont extraits de 30 lames cytologiques numériques colorées avec 9 variantes de la coloration de Papanicolaou. Les lames de cytologie sont préparées à partir d'échantillons d'urine de patients sains, à l'aide d'un processeur Hologic ThinPrep® 5000 et colorées systématiquement avec l'automate Agilent Dako CoverStainer®. Les lames sont ensuite numérisées à l'aide du scanner de lames Hamamatsu NanoZoomer® S360 selon 21 plans focaux centrés

[147]: LIN et al. (2015), « Microsoft COCO »

[46]: HE et al. (2017), « Mask R-CNN »

[148]: YI et al. (2019), « Attentive neural cell instance segmentation »

[149]: WANG et al. (2020), « SOLO »

[150]: WANG et al. (2020), « SOLOv2 »

[151]: CHEN et al. (2016), « DCAN »

32: CISC : Cell Instance Segmentation Dataset.

autour du plan S_{centre} déterminé automatiquement par l'autofocus du scanner. A noter que dans le CISD, les petites instances cellulaires ont été automatiquement filtrées car sous-représentées dans les données disponibles. Une petite instance est définie comme un objet dont la largeur ou la hauteur de la boîte englobante est inférieure à 10% de la largeur ou hauteur de la z-stack contenant l'objet. Les instances ainsi filtrées sont principalement des globules rouges, des cellules neutrophiles et des débris.

Chaque échantillon est considéré de trois manières différentes dans le CISD, ce qui permet d'expérimenter différentes méthodes de traitement des z-stacks et de les comparer avec une simple acquisition 2D.

- La section S_{centre} , correspondant à la section de meilleur focus déterminée par l'autofocus du scanner de lames. La section S_{centre} est l'équivalent d'une simple acquisition 2D.
- La projection EDF S_{edf} : les 21 plans focaux sont fusionnés en une image pour laquelle tous les objets apparaissent en focus. La projection EDF est réalisée en utilisant la méthode décrite chapitre 2 section 2.3.
- La z-stack complète, composée des 21 plans focaux numérisés.

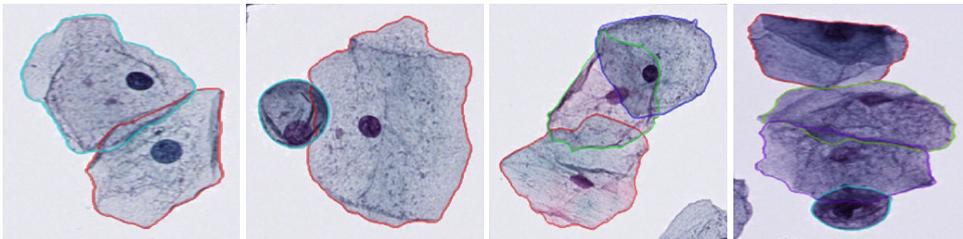


FIGURE 4.15 – Bordure extérieure des masques de segmentation d'instances représentés pour quatre projections EDF extraites du jeux de données CISD.

4.3.3 Architecture modifiée pour traiter des z-stacks

Inspiré par l'étude [66] dans laquelle les auteurs ont tiré parti de toute l'information contenue dans les z-stacks pour segmenter des cellules superposées, une architecture modifiée pour CNN est proposée afin de prédire des masques de segmentation d'instances directement à partir de z-stacks (voir Figure 4.16).

Le modèle ResNet-18 [30], extracteur de caractéristiques populaire dans les approches de vision par ordinateur, est défini pour des tenseurs de dimensions (B, C, H, W) , avec B le nombre d'échantillons dans un batch, C le nombre de canaux couleurs, H, W les hauteur et largeur des images. Pour utiliser des tenseurs de z-stacks, de dimensions (B, C, Z, H, W) , directement en entrée, une modification de l'architecture du modèle ResNet-18 est proposée à partir des résultats obtenus section 4.2.

Avant le premier opérateur de convolution du modèle ResNet-18, le tenseur de dimensions (B, C, Z, H, W) est remanié pour obtenir un tenseur de dimensions $(B, C \times Z, H, W)$, considérant ainsi les plans focaux de la z-stack comme des canaux de couleur supplémentaires. La première couche de convolution du ResNet-18 est suivie par un opérateur Squeeze and Excitation [142] pour pondérer l'importance des canaux de couleur et des plans focaux.

4.3.4 Métriques d'évaluation

Les métriques utilisées dans cette section pour l'évaluation des méthodes de segmentation d'instances sont basées sur le décompte de vrai positifs (TP), faux positifs (FP) et faux négatifs (FN). Un vrai positif correspond à une instance cellulaire correctement segmentée,

[66]: PHOULADY et al. (2016), « A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

[142]: HU et al. (2018), « Squeeze-and-Excitation Networks »

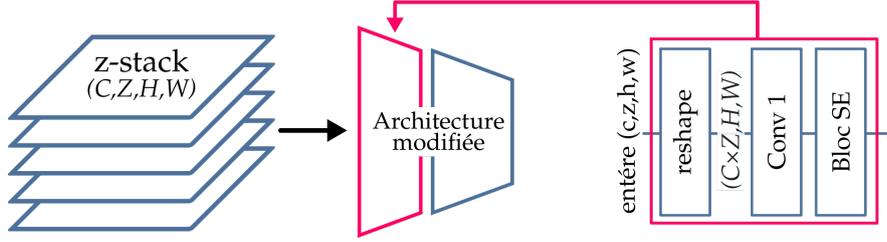


FIGURE 4.16 – Architecture de CNN modifiée pour utiliser directement des z-stacks en entrée : le premier bloc de convolution est précédé par un opérateur de réorganisation des dimensions et suivi par un bloc SE pour pondérer l'importance des différents plans focaux et canaux couleurs.

un faux positif correspond à un objet détecté qui ne correspond pas à une instance cellulaire annotée, et un faux négatif correspond à une instance cellulaire annotée non détectée.

Soit M_{vt}^i le masque de segmentation annoté, associé à l'instance i et M_{pred}^i un masque de segmentation prédit pour cette même instance. L'instance i est considérée comme correctement annotée si l'Intersection sur Union (IoU) entre M_{vt}^i et M_{pred}^i est supérieure à un seuil t :

$$\text{IoU} = \frac{M_{vt}^i \cap M_{pred}^i}{M_{vt}^i \cup M_{pred}^i} \quad (4.11)$$

Le nombre de vrais positifs, faux positifs et faux négatifs dépend donc du seuil t choisi. A partir de ces décomptes les métriques de Précision (AP), Rappel (AR) et score F1 sont définies, en fonction du seuil t :

$$\text{AP}_t = \frac{TP}{TP + FP} \mid \text{IoU} \geq t \quad (4.12)$$

$$\text{AR}_t = \frac{TP}{TP + FN} \mid \text{IoU} \geq t \quad (4.13)$$

$$\text{F1}_t = \frac{\text{AP}_t \times \text{AR}_t}{\text{AP}_t + \text{AR}_t} \quad (4.14)$$

Pour les vrais positifs, la qualité de la segmentation obtenue pour N instances est également évaluée par la métrique IoU :

$$\text{IoU}_t = \frac{1}{N} \sum_{i=0}^N \text{IoU}^i \mid \text{IoU}^i \geq t \quad (4.15)$$

Suivant les pratiques mises en places lors des compétitions ISBI 2014 et ISBI 2015, le seuil $t = 0.7$ est choisi pour définir une bonne segmentation.

4.3.5 Méthodes de segmentation d'instances

Trois méthodes de segmentation d'instances cellulaires sont considérées dans cette section.

Phoulady La méthode proposée par Phoulady et al. [66] est considérée comme référence car elle a obtenu les meilleures performances sur les jeux de données de segmentation de cellules superposées ISBI 2014 et ISBI 2015. Cette méthode de référence est basée sur quatre étapes successives :

1. Une segmentation des noyaux
2. Une segmentation des amas cellulaires
3. Une approximation de masques de segmentation d'instances cellulaires
4. Un affinage des masques de segmentation d'instances cellulaires

L'étape 1 de segmentation de noyaux est définie pour être appliquée sur une image. Les noyaux sont des régions de l'image relativement uniformes, convexes et sombres. L'image est binarisée de manière itérative selon des seuils successifs afin de construire un arbre inversé de composantes connexes. Pour chaque seuil, les composantes connexes trop petites ou trop concaves³³ sont retirées, et les composantes connexes restantes sont considérées comme potentiels noyaux. Au seuil $s + 1$, une composante connexe remplace les composantes connexes qu'elle contient comme potentiel noyau si elle présente une solidité supérieure :

$$\text{solidité} = \frac{\text{Aire région}}{\text{Aire enveloppe convexe}} \quad (4.16)$$

L'étape 2 de segmentation d'amas cellulaires est également appliquée sur une image. En présence de cellules colorées, un simple seuillage suivi par des opérations morphologiques en post-traitement est suffisant pour segmenter les amas cellulaires de l'arrière-plan. Le seuil est appris à l'aide d'un modèle de mélange gaussien à deux composantes : une gaussienne estime la distribution des pixels appartenant aux amas cellulaires tandis que l'autre gaussienne estime la distribution des pixels appartenant à l'arrière-plan.

L'étape 3 d'approximation de masques de segmentation d'instances cellulaires est définie pour être appliquée sur une z-stack. Chaque amas cellulaire segmenté lors de l'étape 2 est divisé en une grille de sous-images. Deux sous-images de cette grille sont assignées en première approximation à la même instance cellulaire si elles sont proches en termes de distance euclidienne et si leurs profils-z³⁴ de mesure de focus sont similaires. La vraisemblance que deux sous-images appartiennent à une même cellule, en fonction de leurs positions et vecteurs de focus, est pondérée par un hyper-paramètre α . De plus, à cette étape, un certain degré de superposition des masques d'instances est autorisé et définit par un hyper-paramètre β .

L'étape 4 d'affinage de la segmentation d'instances cellulaires est appliquée sur une image en deux temps. Un affinage grossier, effectué au niveau des sous-images obtenues à l'étape 3, permet de limiter la prédiction de masques cellulaires concaves. Enfin, un affinage fin est réalisé : la position de la membrane cellulaire est déterminée à l'endroit de plus fort gradient le long de vecteurs radiaux partant du noyau. Une illustration de la méthode est proposée Figure 4.17.

L'étape 3 d'approximation de masques de segmentation d'instances cellulaires est définie pour une z-stack, mais peut également être directement appliquée à une image en considérant le cas particulier d'une z-stack avec $z = 1$ plan focal. Les étapes 1, 2 et 4 sont définies pour des images. Ainsi la méthode *Phoulady* peut-être directement appliquée aux trois représentation des échantillons considérés dans le jeu de données CISD.

Pour les échantillons S_{centre} et S_{edf} , les étapes 1, 2, 3 et 4 sont appliquées respectivement aux images S_{centre} et S_{edf} des z-stacks. Pour les z-stacks complètes, la méthode originale

[66]: PHOULADY et al. (2016), « A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images »

33: Par rapport à des seuils prédéfinis.

34: Variation d'une mesure de focus calculée pour chaque plan focal de la z-stack. Voir section 4.1 pour plus de détails.

[67]: PHOULADY et al. (2017), « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images »

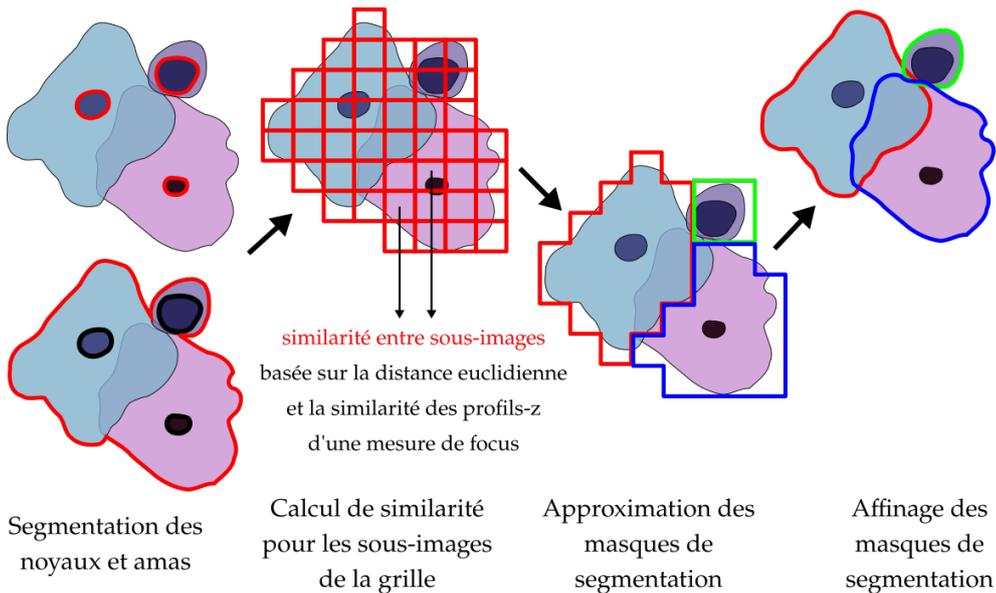


FIGURE 4.17 – Illustration de la méthode de référence proposée par Phoulady et al. [66][67].

proposée dans [67] est utilisée : les étapes 1, 2 et 4 sont appliquées à la projection S_{edf} et l'étape 3 est appliquée à la z-stack complète. Après expérimentation sur les échantillons d'entraînement du jeu de données CISD, les hyper-paramètres de la méthode ³⁵ sont fixés à $\alpha = 2.0$ et $\beta = 20$.

35: Voir [67] pour plus de détails sur la méthode.

Attentive Neural Cell Instance Segmentation (ANCIS) La méthode ANCIS proposée dans [148] est une méthode basée sur une détection préalable des instances cellulaires, suivie par un modèle de segmentation sémantique pour chaque instance détectée. Les auteurs ont montré que cette méthode simple, appliquée avec succès à la segmentation d'instances de cellules neuronales, obtient de meilleures performances que le modèle Mask-RCNN [46] dont elle s'inspire tout en étant plus facile à entraîner. Cette méthode, illustrée Figure 4.18, est basée sur le modèle *Single Shot Multibox Detector* (SSD) [152] avec une architecture ResNet-18 [30] pour la détection d'instances cellulaires. Les boîtes englobantes prédites sont utilisées pour rogner les cartes de caractéristiques du modèle ResNet-18 à différentes profondeurs, lesquelles sont ensuite passées dans un modèle de segmentation sémantique.

[148]: Yi et al. (2019), « Attentive neural cell instance segmentation »

[46]: He et al. (2017), « Mask-RCNN »

[152]: Liu et al. (2016), « SSD »

[30]: He et al. (2015), « Deep Residual Learning for Image Recognition »

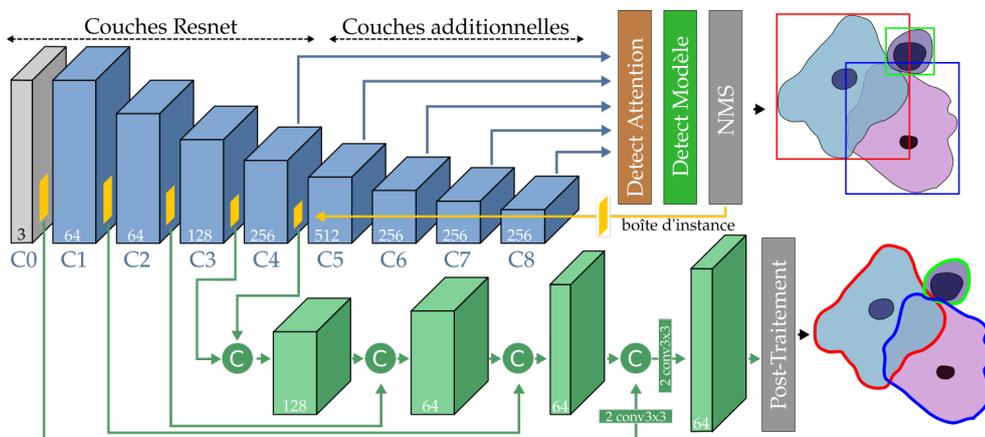


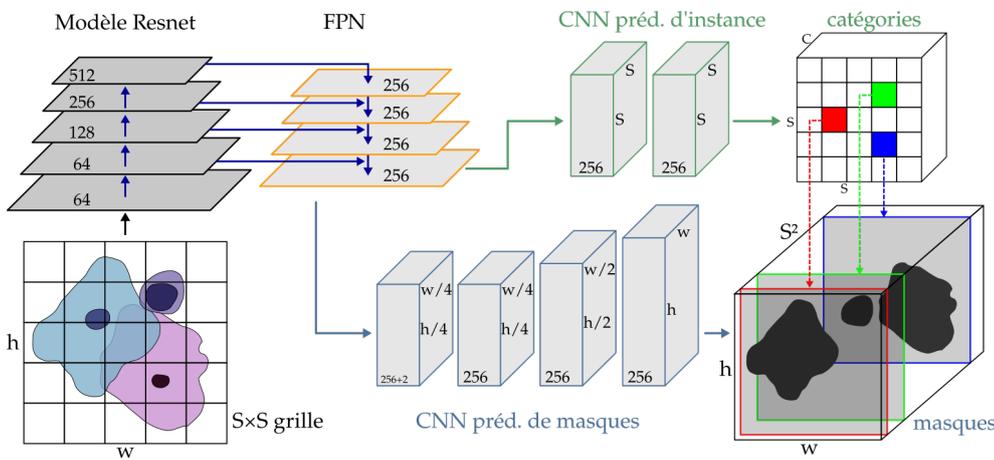
FIGURE 4.18 – Illustration de la méthode ANCIS. La branche supérieure (en bleu) est le modèle SSD utilisé pour la détection d'instances cellulaires. La branche inférieure (en vert) utilise les cartes de caractéristiques rognées du SSD pour prédire des masques de segmentation sémantique pour chaque instance détectée.

Comme recommandé dans la publication originale [148], l'entraînement du modèle ANCIS est effectué en deux étapes. Premièrement le modèle de détection SSD est entraîné seul, puis ses poids sont fixés et la branche de segmentation est entraînée à son tour. L'architecture utilisée dans ce travail est affinée pour être adaptée au CISD,

[148]: Yi et al. (2019), « Attentive neural cell instance segmentation »

comme suggéré par Liu et al. [152] et diffère ainsi légèrement de celle décrite dans [148]. Notamment, pour s'adapter à la taille des objets d'intérêt dans les images, les cartes de caractéristiques C4 à C8 sont utilisées dans le modèle de segmentation sémantique, au lieu des cartes de caractéristiques C3 à C6³⁶.

Segmenting Objects by Location (SOLO) La méthode SOLO [149] et sa version améliorée SOLOv2 [150], désignées ci-après par le terme SOLO sont des méthodes de segmentation d'instances qui ne reposent pas sur une détection au préalable des objets d'intérêt. La méthode est basée sur un paradigme simple : l'image originale est divisée en sous-images à l'aide d'une grille. Chaque case de la grille est alors chargée de prédire une instance et son masque de segmentation si elle contient le centre de masse de cette instance. Cette méthode est basée sur un modèle ResNet [30] combiné à un *Feature Pyramid Network* (FPN) [153] pour l'extraction de caractéristiques à plusieurs échelles. Finalement, deux modèles CNN indépendants sont responsables de la prédiction et de la segmentation des instances pour chaque case de la grille prédéfinie. Une illustration du modèle est proposée Figure 4.19.



[152]: LIU et al. (2016), « SSD »

[148]: YI et al. (2019), « Attentive neural cell instance segmentation »

36: Voir la notation Figure 4.18.

[149]: WANG et al. (2020), « SOLO »

[150]: WANG et al. (2020), « SOLOv2 »

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

[153]: LIN et al. (2017), « Feature Pyramid Networks for Object Detection »

FIGURE 4.19 – Illustration de la méthode SOLO. Les caractéristiques, extraites par un modèle ResNet combiné à un FPN, sont utilisées pour prédire les instances et les masques associés à l'aide de deux modèles CNN dédiés. La branche de segmentation représentée est la branche *vanilla* introduite dans [149].

Plusieurs branches de segmentation ont été proposées dans [149] et [150], notamment la branche de segmentation originale, dite *vanilla*, et une branche de segmentation dite *dynamique*. La branche *vanilla* prédit directement un tenseur de masques de dimensions (H, W, S^2) , alors que la branche de segmentation *dynamique* prédit un tenseur T_{feature} de dimensions (H, W, E) et un tenseur T_{noyau} de dimensions $(S, S, 9E)$. Les valeurs prédites pour le tenseur T_{noyau} sont utilisées comme noyau pour des opérateurs de convolutions, appliqués ensuite au tenseur T_{feature} pour obtenir les masques de segmentation finaux. Dans [150], les auteurs suggèrent d'utiliser les convolutions déformables (DCN) introduite par [154], afin d'augmenter encore les performances du modèle SOLO. Les DCN sont donc testées à la place des convolutions classiques dans le FPN du modèle SOLO. Les performances obtenues avec les différentes variations d'architecture sont reportées Tableau 4.13.

Branche de seg.	AP ₀₇	AR ₀₇	F1 ₀₇	IoU ₀₇
Vanilla	0.924	0.952	0.938	0.947
Dynamique	0.942	0.947	0.944	0.948
Dynamique + DCN	0.938	0.947	0.942	0.949

[149]: WANG et al. (2020), « SOLO »

[150]: WANG et al. (2020), « SOLOv2 »

[150]: WANG et al. (2020), « SOLOv2 »

[154]: DAI et al. (2017), « Deformable Convolutional Networks »

TABLE 4.13 – Impact de la branche de segmentation de SOLO et des convolutions déformables (DCN) sur les performances de segmentation d'instances. Les performances sont reportées pour un sous-ensemble du jeu de données C1SD (Dev) utilisé pour fixer les hyper-paramètres des méthodes évaluées.

Expérimentalement, la branche de segmentation *dynamique*, introduite dans [150] augmente significativement les performances par rapport à la branche *vanilla*, et est donc

[150]: WANG et al. (2020), « SOLOv2 »

utilisée dans la suite. Les DCN en revanche n'améliorent pas les performances de segmentation et les convolutions classiques sont donc utilisées pour des raisons de simplicité.

4.3.6 Résultats

Les méthodes ANCIS et SOLO sont basés sur le modèle ResNet-18 et sont entraînés à partir d'une initialisation aléatoire selon trois splits Train/Dev des données d'entraînement. Les méthodes ont été entraînées sur les échantillons S_{centre} , S_{edf} et les z-stacks complètes. La méthode de référence *Phoulady* ne nécessite pas d'entraînement. Les trois méthodes sont ensuite évaluées sur les 978 échantillons de Test pour les trois types d'échantillons définis précédemment, à l'aide des métriques définies sous-section 4.3.4. Les résultats sont reportés Tableau 4.14.

méthode	AP_{07}	AR_{07}	$F1_{07}$	IoU_{07}
échantillon : S_{centre}				
Phoulady	0.500	0.520	0.510	0.843
ANCIS	0.935 ± 0.004	0.884 ± 0.002	0.909 ± 0.003	0.918 ± 0.002
SOLO	0.942 ± 0.002	0.943 ± 0.002	0.942 ± 0.002	0.944 ± 0.000
échantillon : S_{edf}				
Phoulady	0.587	0.567	0.577	0.853
ANCIS	0.938 ± 0.003	0.888 ± 0.005	0.912 ± 0.004	0.923 ± 0.001
SOLO	0.944 ± 0.002	0.948 ± 0.001	0.945 ± 0.001	0.947 ± 0.000
échantillon : z-stack				
Phoulady	0.589	0.567	0.578	0.854
ANCIS	0.936 ± 0.000	0.886 ± 0.002	0.910 ± 0.001	0.919 ± 0.002
SOLO	0.945 ± 0.001	0.947 ± 0.001	0.946 ± 0.000	0.947 ± 0.000

TABLE 4.14 – AP, AR et score F1 pour un seuil IoU de bonne segmentation défini à $t = 0.7$. Le score IoU pour les vrai positifs parmi les échantillons Test du jeu de données CISD est également reporté. Les méthodes ANCIS et SOLO sont entraînées pour trois splits Train/Dev, les scores ensuite obtenus pour les échantillons de Test sont moyennés et reportés avec leurs écart-type. Pour chaque métrique, les meilleurs performances sont reportées en **bleu**.

Le modèle SOLO s'avère être la méthode la plus performante pour la segmentation des instances cellulaires³⁷ avec une amélioration allant jusqu'à 0.4 AP et 0.4 AR par rapport à la méthode de référence [67]. De plus, les meilleures performances sont obtenues en utilisant les acquisitions multi-plans, que ce soit par l'utilisation des projections S_{edf} EDF ou des z-stacks directement. Ces résultats sont attendus : les sections centrales S_{centre} peuvent contenir des cellules hors-focus, ce qui complique la prédiction de segmentations d'instances cellulaires précises. Des performances de segmentation équivalentes sont obtenues en utilisant des échantillons S_{edf} et les z-stacks.

37: Voir Figure 4.20.

[67]: PHOULADY et al. (2017), « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images »

4.3.7 Évaluation du temps d'inférence

Le temps de calcul total, normalisé par échantillon, comprenant l'éventuel prétraitement par projection EDF et la prédiction des masques de segmentation d'instances, est évalué sur les 978 échantillons de Test et rapporté Tableau 4.15.

méthode	pré-traitement CPU (s)	Inférence CPU (s)	Total CPU (s)
SOLO (S_{edf})	0.726	0.780	1.506
SOLO (z-stack)	0	0.824	0.824
échantillon	pré-traitement CPU (s)	Inférence GPU (s)	Total CPU + GPU (s)
SOLO (S_{edf})	0.726	0.113	0.839
SOLO (z-stack)	0	0.117	0.117

TABLE 4.15 – Temps de calcul pour la méthode SOLO, lorsque appliquée aux projections EDF S_{edf} ou aux z-stacks directement. Les temps totaux les plus faibles sont reportés en **bleu**.

L'inférence des masques d'instances est effectuée à l'aide du modèle SOLO [149], en utilisant les mêmes hyper-paramètres que ceux utilisés dans sous-section 4.3.6, sur une machine équipée d'un CPU Intel Core i7-8750H et d'un GPU Nvidia 1050Ti. Deux configurations sont considérées : tout-CPU et CPU + GPU où le pré-traitement EDF, si nécessaire, est effectué sur CPU et la prédiction des masques de segmentation est effectuée sur GPU. Pour les deux configurations, l'utilisation directe des entrées des z-stacks dans les modèles ResNet-18 à l'aide de l'architecture modifiée présentée sous-section 4.3.3 permet d'économiser de 45% à 86% du temps de traitement total pour des performances de segmentation similaires.

[149]: WANG et al. (2020), « SOLO »

4.3.8 Discussion

Cette section introduit un nouveau jeu de données pour la segmentation d'instances de cellules superposées à partir de z-stacks. Ce jeu de données CISD permet d'entraîner, de comparer et de valider des modèles récents de deep learning pour la segmentation d'instances de cellules extraites de lames de cytologie digitales. Les performances obtenues pour ces méthodes surpassent de loin les performances des approches [67][148] proposées précédemment pour la segmentation d'instances cellulaires.

[67]: PHOULADY et al. (2017), « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images »

[148]: YI et al. (2019), « Attentive neural cell instance segmentation »

Les expériences menées montrent également que la segmentation d'instances appliquée aux sections centrales S_{centre} est moins performante que lorsque appliquée aux projections EDF S_{edf} ou aux z-stacks complètes. Ces résultats étaient attendus, car les sections centrales peuvent contenir des cellules hors-focus, qui rendent difficile la segmentation précise des instances cellulaires. Cela justifie l'utilisation d'acquisitions multi-plans pour la segmentation d'instances cellulaires extraites de lames de cytologie digitales.

La méthode SOLO [149][150], appliquée à la fois sur les projections EDF S_{edf} ou directement sur les z-stacks, obtient d'excellentes performances de segmentation avec un $F1_{07}$ score de 0.946 et 0.947 d'IoU₀₇. L'architecture modifiée proposée sous-section 4.3.3 permet d'utiliser directement les z-stacks en entrée des réseaux de neurones convolutionnels et ainsi d'économiser jusqu'à 86% du temps de traitement total par rapport à l'utilisation des projections EDF, sans compromis sur les performances de segmentation.

[149]: WANG et al. (2020), « SOLO »

[150]: WANG et al. (2020), « SOLOv2 »

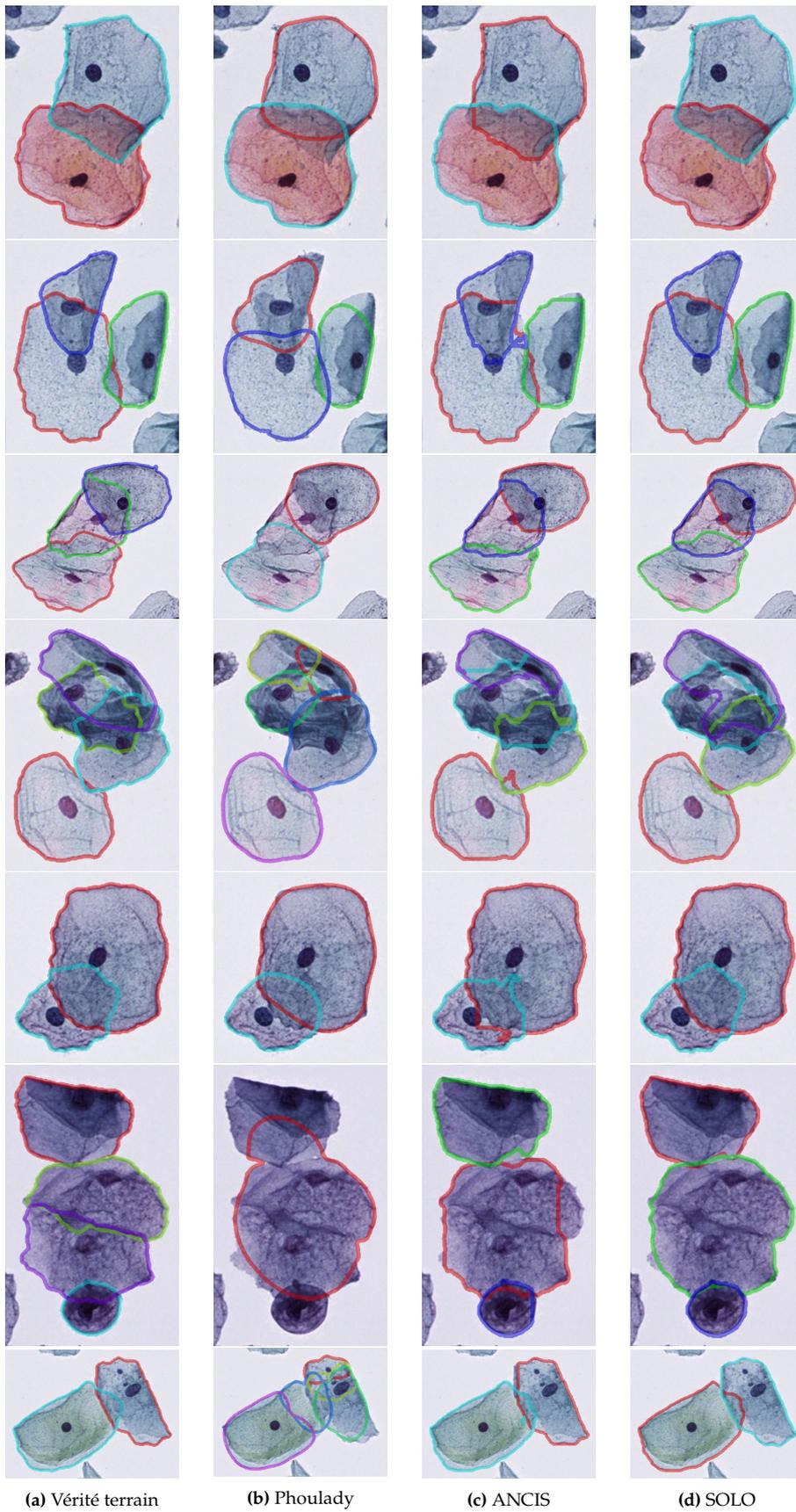


FIGURE 4.20 – Instances cellulaires prédites par les méthodes Phoulady, ANCIS et SOLO pour les échantillons S_{edf} .

4.4 Conclusion

Dans ce chapitre, plusieurs approches ont été développées et évaluées afin d'extraire de l'information directement à partir de z-stacks.

La section 4.1 introduit le problème de segmentation sémantique de régions cellulaires comme un problème de classification de pixels, adressé à l'aide de profils-z d'intensité. Des résultats prometteurs ont été obtenus, mais la méthode proposée ne tient pas compte de l'information spatiale disponible, et les masques de segmentation obtenus restent alors bruités.

La section 4.2 évalue alors plusieurs modifications d'architecture de CNN, appliquées au modèle U-Net, afin de prédire des masques de segmentation sémantique directement à partir de z-stacks. D'excellentes performances de segmentation sémantique sont obtenues lorsque la prédiction est effectuée à partir des z-stacks ou des projections EDF S_{edf} . Les architectures modifiées U-Net-3D et U-Net-P (BDCLSTM) obtiennent les meilleures performances de segmentation d'instances, au prix cependant d'un important surcoût en calculs et donc en temps de traitement. Le modèle U-Net-CSE, qui considère les plans focaux d'une z-stack comme des canaux de couleur additionnels, est un bon compromis qui permet d'obtenir de très bonnes performances de segmentation pour un temps de calcul limité et inférieur à un pré-traitement EDF suivi d'une prédiction 2D classique.

La section 4.3 s'intéresse au problème de la segmentation d'instances cellulaires à partir de cellules superposées extraites de lames de cytologie digitales. À l'aide d'un nouveau jeu de données CISD, la méthode SOLO [149][150] peut être appliquée à ce problème de segmentation d'instances cellulaires, obtenant les meilleures performances parmi les approches évaluées. En utilisant ici encore une architecture modifiée, reprenant les idées du modèle U-Net-CSE, il est possible de prédire des masques d'instances directement à partir des z-stacks, sans perte de performances tout en réduisant jusqu'à 86% le temps de calcul par rapport à un pré-traitement EDF suivi d'une segmentation 2D. Cela confirme les résultats obtenus section 4.2.

Ainsi dans ce chapitre, plusieurs méthodes de segmentation à partir de z-stacks ont été décrites et évaluées. Lorsque ces méthodes sont appliquées aux z-stacks ou aux projections EDF, un gain de performances est systématiquement observé par rapport aux mêmes méthodes appliquées à des acquisitions bi-dimensionnelles³⁸. Cela confirme l'importance de l'acquisition et de l'analyse de lames de cytologie multi-plans pour les tâches de segmentation automatique.

L'utilisation directe des z-stacks dans les méthodes de segmentation ne conduit pas toujours à un gain de performances par rapport à un pré-traitement EDF suivi d'une segmentation 2D. En revanche, il est trouvé systématiquement³⁹ que pour des performances similaires, l'utilisation de z-stacks réduit de façon importante le temps total d'inférence, puisque le pré-traitement par extension de profondeur de champ n'est pas nécessaire.

[149]: WANG et al. (2020), « SOLO »

[150]: WANG et al. (2020), « SOLOv2 »

38: Ici les sections centrales S_{centre} .

39: Voir section 4.2 et section 4.3.

5.1 Introduction

Les chapitres précédents se sont intéressés à l'acquisition et à l'analyse locale¹ de lames de cytologie multi-plans. Le diagnostic automatique du cancer de la vessie nécessite une analyse globale des ces lames, qui doit décider du caractère atypique d'une lame à partir des analyses locales des cellules d'intérêt qu'elle contient. Ce chapitre propose ainsi une méthode de diagnostic automatique de carcinomes urothéliaux à partir de lames de cytologie digitales, créées à l'aide d'échantillons d'urine.

1: C'est-à-dire au niveau d'une cellule isolée.

Dans ce chapitre, la méthode présentée repose sur l'analyse des sections de meilleur focus S_{sharp} de cellules urothéliales basales extraites de lames digitales multi-plans. L'adaptation et les performances de l'approche proposée, pour différentes représentations 2D et méthodes d'analyses des z-stacks, sont discutées chapitre 6 et chapitre 7.

Le diagnostic du cancer de la vessie est considéré comme un problème de classification binaire :

- La classe 0 (négative) est assignée aux lames des patients sains.
- La classe 1 (positive) est assignée aux lames des patients atteints par un carcinome urothélial de bas-grade (LGUC) ou carcinome urothélial de haut-grade (HGUC).

Les données utilisées pour ce problème ont été acquises au cours de l'essai clinique VisioCyt® 01, pour lequel l'utilisation des données récoltées à des fins de recherche a été approuvé par l'Agence Nationale de la Santé et du Médicament et par le Comité de Protection des Personnes.

Les patients éligibles pour inclusion dans l'essai clinique doivent présenter un examen cytobactériologique des urines (ECBU) négatif, afin d'éviter la présence de cellules inflammatoires dans les urines qui pourraient être confondues avec des cellules urothéliales atypiques.

Dans cet essai clinique, un patient est considéré comme sain s'il s'agit d'un patient sans suspicion de cancer de la vessie ou en suivi de carcinome urothélial traité antérieurement pour lequel le résultat de la cytologie conventionnelle, évaluée par un anatomopathologiste, est négatif et le résultat de l'examen de cystoscopie est également négatif.

Un patient est considéré comme malade s'il présente des signes cliniques du cancer de la vessie et que le résultat de l'examen de cystoscopie est positif. Une histologie suivant la cystoscopie positive permet ensuite de connaître le grade du cancer de la vessie. À noter, quelques patients avec une cytologie conventionnelle "très positive"² et une cystoscopie négative ont été inclus comme patients malades dans l'essai clinique, pour tenir compte des carcinomes à croissance plane invasifs, qui sont difficilement visibles par l'examen de cystoscopie.

2: Lame C6 dans la classification du Système de Paris, voir sous-section 5.1.1.

Au total, 1023 patients inclus dans l'essai clinique VisioCyt® 01 forment le jeu de données que l'on nommera simplement par la suite `visiocyt`. Ces patients sont répartis en deux groupes : 731 patients sont disponibles pour l'entraînement de méthodes de diagnostic automatique et 292 patients sont réservés pour l'évaluation de ces méthodes³.

3: Appelé ensemble Test.

Parmi les 731 patients de l'ensemble d'entraînement, 268 sont sains, 233 sont atteints d'un carcinome urothélial de bas-grade et 230 sont atteints d'un carcinome urothélial de

haut-grade. Parmi les 292 patients de l'ensemble de Test, 141 sont sains, 76 sont atteints d'un carcinome urothélial de bas-grade et 75 sont atteints d'un carcinome urothélial de haut-grade.

Le problème considéré dans ce travail diffère donc des études de diagnostic automatique à partir de lames de cytologie digitales menées précédemment. Au lieu d'entraîner un algorithme à prédire le résultat de l'examen de cytologie conventionnel, ici on cherche à prédire le résultat de l'examen de cystoscopie, suivi éventuellement d'un examen histopathologique, à partir de lames de cytologie digitales.

Formulé de cette manière, le problème comporte de nombreuses difficultés. En effet, une lame de cytologie pour un patient malade peut malencontreusement ne pas contenir de cellules cancéreuses, rendant ainsi le diagnostic impossible à partir de la lame de cytologie seule. De plus, la cytologie est peu sensible aux carcinomes urothéliaux de bas-grade, pour lesquels les cellules atypiques sont peu nombreuses et bien différenciées⁴ : c'est pourquoi l'examen cytologique conventionnel se concentre sur la détection des carcinomes urothéliaux de haut-grade. Dans cette étude, les cancers de bas-grades sont également considérés comme positifs.

Néanmoins, le développement d'un algorithme performant pour le diagnostic des cancers de la vessie de haut-grade et de bas-grade permettrait d'éviter le recours systématique à l'examen de cystoscopie. L'intérêt pour le patient serait grand car la cystoscopie est un examen intrusif et douloureux, contrairement à l'examen de cytologie qui nécessite simplement le recueil d'un échantillon d'urine.

4: Les cellules urothéliales basales atypiques des LGUCs ont tendance à être bien différenciées et ressemblent aux cellules urothéliales basales saines, ce qui complique leur détection.

5.1.1 La classification du Système de Paris

Afin de pouvoir apposer un diagnostic fiable sur les carcinomes urothéliaux de haut-grade, une méthode d'analyse normalisée des lames de cytologie a été proposée par le groupe de travail du Système de Paris [24]. Cette méthode, référée dans la suite simplement par Système de Paris, fonctionne comme un arbre de décision basé sur une série de caractéristiques expertes et permet de classer les lames de cytologie en six catégories :

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

- **C1** : La lame ne peut pas être analysée, à cause d'une mauvaise qualité, d'une erreur dans la préparation ou d'un nombre de cellules d'intérêt trop faible.
- **C2** : Lame négative pour le carcinome urothélial de haut-grade.
- **C3** : Lame contenant des cellules urothéliales basales atypiques.
- **C4** : Suspicion de carcinome urothélial de haut-grade.
- **C5** : Lame positive pour un carcinome urothélial de haut-grade.
- **C6** : Lame positive pour un carcinome urothélial de bas-grade.

Les lames de la catégorie **C1** ne sont donc pas traitées. Les lames **C2** et **C3** sont considérées comme négatives et les lames de catégories **C4**, **C5** et **C6** sont positives.

Les caractéristiques expertes suivantes sont utilisées dans le Système de Paris :

Le Ratio Noyau-Cytoplasme (RNC) Le RNC est le ratio entre l'aire apparente du noyau et l'aire apparente du cytoplasme d'une cellule urothéliale basale. Le RNC est le critère d'atypie jugé le plus important par les experts du Système de Paris. Un RNC supérieur à 0.5 est considéré comme un critère d'atypie léger, tandis qu'un RNC supérieur à 0.7 est un critère d'atypie sévère. Des exemples sont donnés Figure 5.1.

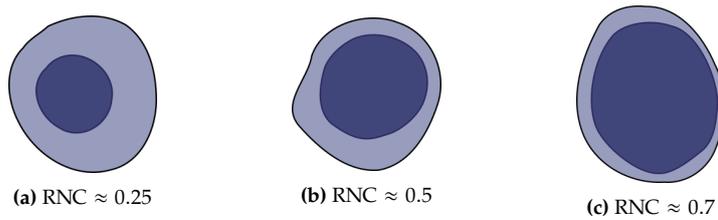


FIGURE 5.1 – Illustration de différents Ratios Noyaux-Cytoplasme correspondants à différents niveaux d'atypie : (a) cellule normale (RNC < 0.5), (b) cellule légèrement atypique (RNC \approx 0.5), (c) cellule sévèrement atypique (RNC \approx 0.7).

Hyper-chromatisme L'hyper-chromatisme est une sur-coloration du noyau, d'apparence plus sombre par rapport au noyau d'une cellule saine. L'hyper-chromatisme est également un critère d'atypie de haute importance, nécessairement observé pour classer une lame de cytologie comme **C4** ou **C5** dans le Système de Paris. Des illustrations sont proposées Figure 5.2.

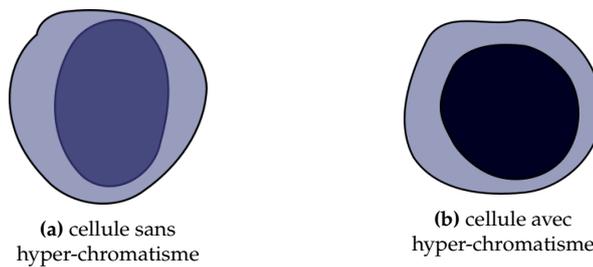


FIGURE 5.2 – Illustration de cellules avec (b) et sans (a) hyper-chromatisme.

Texture de chromatine irrégulière Une texture de chromatine nucléaire irrégulière, et notamment la présence de texture grossière (chromatine agglomérée) est un critère d'atypie secondaire dans la classification du Système de Paris. La chromatine, structure de l'ADN au sein des noyaux cellulaires, peut être plus ou moins compacte. Dans le cas de cellules urothéliales basales atypiques, la chromatine peut former des mottes et ainsi apparaître avec une texture grossière, comme illustré Figure 5.3.

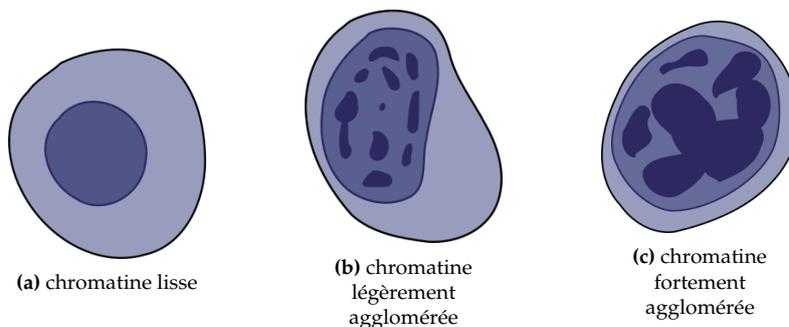


FIGURE 5.3 – Illustration de différentes textures de chromatine.

Bordure de chromatine irrégulière La chromatine peut migrer vers la bordure du noyau, produisant une bordure de chromatine irrégulière et parfois discontinue. Le reste du noyau peut alors apparaître relativement plus pâle (voir Figure 5.4). Une bordure de chromatine irrégulière est un critère d'atypie secondaire dans le Système de Paris.

En se basant sur une évaluation visuelle des caractéristiques expertes définies précédemment, le Système de Paris permet de catégoriser les lames de cytologie en suivant une approche algorithmique de type *arbre de décision*, illustrée Figure 5.5.

Une approche envisageable pour le diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales, est une implémentation de l'arbre de décision du Système de Paris, comme proposé dans les travaux de Vaickus et al. [26] et de Sanghvi et

[26]: VAICKUS et al. (2019), « Automating the Paris System for urine cytology—A hybrid deep-learning and morphometric approach »

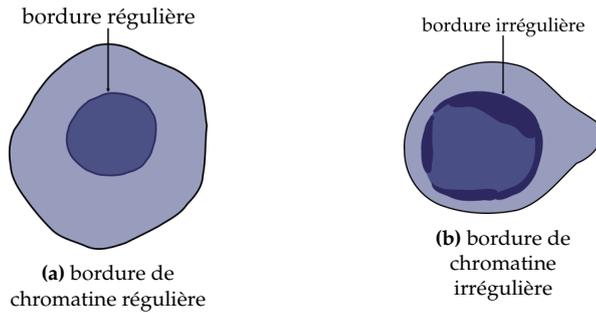


FIGURE 5.4 – Illustration d’une cellule avec une bordure de chromatine “normale” régulière, et d’une cellule avec une bordure de chromatine atypique irrégulière.

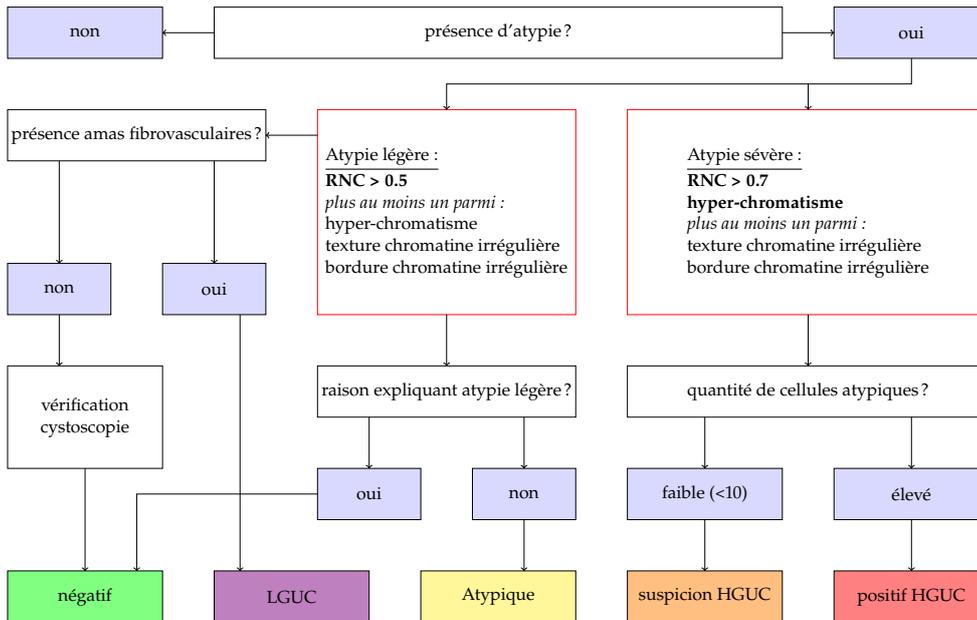


FIGURE 5.5 – Arbre de décision proposé par le Système de Paris [24].

al. [155]. Plusieurs limitations à cette approche apparaissent cependant :

- Il n’y a pas de fort consensus au sein de la littérature scientifique sur la valeur de RNC optimale permettant de définir une cellule urothéliale basale comme atypique. Ainsi, dans [156], Hang et al. préconisent une valeur seuil de RNC de 0.5, tandis que l’étude [157] recommande une valeur seuil de RNC de 0.7 comme critère d’atypie.
- Les seuils critiques de 0.5 et 0.7 du RNC ont été définis pour une évaluation visuelle par des pathologistes. Cependant, des études récentes [158][159] suivant la publication du Système de Paris montrent qu’il est difficile d’évaluer précisément des valeurs correctes de Ratio Noyau-Cytoplasme autour de ces seuils critiques.
- Des seuils ont besoin d’être établis pour définir les caractéristiques expertes d’hyper-chromatisme, de texture de chromatine irrégulières et de bordure de chromatine irrégulières. Ces seuils dépendent de la distribution des autres cellules dans la lame, ainsi que de l’aspect “attendu” pour une cellule saine, et sont en pratique évalués de manière subjective par le pathologiste.
- Les caractéristiques expertes utilisées par le Système de Paris ont été pensées pour une évaluation visuelle par les pathologistes, et se restreindre à celles-ci ainsi qu’aux seuils de décision établis ne permet pas d’utiliser toute l’information quantitative apportée par l’analyse d’images.
- Enfin, le Système de Paris est une méthode d’analyse d’atypie pour chaque instance cellulaire, et le diagnostic global de la lame est basé sur un décompte des cellules atypiques dans une lame de cytologie. Ce décompte est dépendant de la quantité de cellules présentes dans les lames et donc peu robuste si utilisé directement

[155]: SANGHVI et al. (2019), « Performance of an artificial intelligence algorithm for reporting urine cytology »

[156]: HANG et al. (2017), « Digital image analysis supports a NCR cutoff value of 0.5 for atypical urothelial cells »

[157]: McINTIRE et al. (2019), « Digital image analysis supports a NCR cutoff value below 0.7 for positive for HGUC and suspicious for HGUC »

[158]: ZHANG et al. (2016), « Morphologists overestimate the nuclear-to-cytoplasmic ratio »

[159]: LAYFIELD et al. (2017), « Accuracy and Reproducibility of Nuclear/Cytoplasmic Ratio Assessments in Urinary Cytology Specimens »

pour une analyse automatique des lames de cytologie. Pour exemple, les travaux [26] et [155] n'utilisent pas le décompte des cellules atypiques directement pour le diagnostic global, mais plutôt les distributions des caractéristiques d'atypie par lames.

Ainsi, bien que s'en inspirant, les méthodes de diagnostic automatique du cancer de la vessie évaluées dans cette section ne cherchent pas à reproduire fidèlement la classification du Système de Paris, mais plutôt à obtenir les meilleurs performances de diagnostic possible.

5.2 État de l'art

Plusieurs études ont cherché à effectuer un diagnostic automatique, pour différentes pathologies, à partir de lames de cytologie digitales. Ces travaux peuvent être regroupés en deux catégories.

5.2.1 Méthodes avec annotations d'atypies locales

Ces approches reposent sur des annotations d'atypie de cellules ou de sous-images atypiques extraites de lames de cytologie digitales, en plus du label de diagnostic global.

Dans [93], Pironavo et al. cherchent à prédire la classe d'atypie la plus sévère, pour des sous-images contenant de multiples cellules de frottis cervical. La tâche considérée est un problème de classification à cinq classes (normale et quatre niveaux de sévérité d'atypie), effectuée sur un jeu de données synthétiques construit à partir d'images de cellules annotées du jeu de données DTU/Herlev [160].

Dans [94], Cheng et al. cherchent également à diagnostiquer automatiquement le cancer du col de l'utérus à partir de lames de cytologie digitales. Contrairement à l'étude menée par Pironavo et al., les auteurs proposent également une méthode pour extraire les sous-images d'intérêt des lames de cytologie digitales. À partir de 3545 lames de cytologie pour lesquelles 79 911 cellules atypiques ont été annotées, un premier réseau de neurones convolutionnel (CNN) est entraîné à localiser des cellules atypiques à basse résolution ⁵. Un second CNN est ensuite utilisé pour extraire des caractéristiques et prédire un score d'atypie à partir de sous-images haute-résolution ⁶ centrées sur les cellules atypiques détectées à l'étape précédente. Finalement un réseau de neurones récurrent (RNN) est utilisé pour prédire un score global à partir des caractéristiques extraites des dix instances dont le score d'atypie est le plus haut.

Dans [95] Awan et al. proposent une méthode de diagnostic automatique du cancer de la vessie. Un seuillage automatique d'Otsu [72] est utilisé pour détecter les objets cellulaires comme composantes connexes. Un CNN est ensuite entraîné à classer les objets détectés comme normaux, atypiques ou cancéreux, à partir d'annotations locales récoltées auprès d'experts.

5.2.2 Méthodes sans annotations d'atypies locales

Contrairement aux études précédentes, ces approches reposent uniquement sur le label de diagnostic global des lames de cytologie digitales ⁷. Certaines annotations locales non-liées à l'atypie, comme le type cellulaire permettant de détecter les cellules d'intérêt, sont parfois disponibles.

[26]: VAICKUS et al. (2019), « Automating the Paris System for urine cytopathology—A hybrid deep-learning and morphometric approach »

[155]: SANGHVI et al. (2019), « Performance of an artificial intelligence algorithm for reporting urine cytopathology »

[93]: PIROVANO et al. (2021), « Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier »

[160]: JANTZEN et al. (2006), « Analysis of Pap-Smear Image Data »

[94]: CHENG et al. (2021), « Robust whole slide image analysis for cervical cancer screening using deep learning »

5: Environ 0.5µm/pixel.

6: Environ 0.25µm par pixel.

[95]: AWAN et al. (2021), « Deep learning based digital cell profiles for risk stratification of urine cytology images »

[72]: OTSU (1979), « A Threshold Selection Method from Gray-Level Histograms »

7: Ce sont donc des méthodes dites faiblement supervisées.

Dans [97], Dov et al. cherchent à prédire la malignité de tissus thyroïdiens. Pour cela, un CNN est entraîné à reconnaître des sous-images extraites de lames digitales contenant ou non des cellules folliculaires. Un deuxième CNN est entraîné à prédire, pour chaque sous-image contenant des cellules folliculaires, le label global de la lame. Toutes les sous-images d'intérêt détectées dans une lame maligne ne contiennent pas nécessairement des cellules folliculaires atypiques : il s'agit dans ce cas d'un apprentissage faiblement supervisé avec annotations bruitées.

Dans [98], Teramoto et al. cherchent à diagnostiquer automatiquement des cytologies pulmonaires. Pour cela, chaque lame de cytologie digitale est divisée en une grille, et chaque sous-image de cette grille est considérée soit comme *instance* soit comme *arrière-plan*, à partir de la proportion de matériel cellulaire détecté par un seuillage automatique d'Otsu [72]. Chaque sous-image *instance* I_i est ensuite passée dans un CNN pour en extraire un vecteur de caractéristique h_i . Une méthode d'apprentissage à partir d'instances multiples, *Attention-based Multiple Instance Learning (ABMIL)* [62] est ensuite appliquée aux sacs d'instances extraits des lames pour prédire le diagnostic global ⁸.

Dans [99], Butke et al. cherchent à prédire la présence de carcinomes urothéliaux de haut-grade à partir de lames de cytologie urinaire. Pour cela, les auteurs ont à leur disposition des lames digitales de cytologie urinaire et le résultat de l'examen de cytologie associé. Premièrement, un seuillage automatique d'Otsu est appliqué à la lame globale, les composantes connexes sont identifiées, et des sous-images centrées sur ces composantes connexes sont extraites. Similairement à [98], chaque sous-image I_i est passée dans un CNN pour en extraire un vecteur de caractéristiques h_i . Ici encore, la méthode ABMIL [62] est utilisée pour prédire le diagnostic global à partir de tous les vecteurs de caractéristiques h_i contenus dans la lame de cytologie.

5.3 Méthode

La méthode proposée appartient aux approches sans annotations d'atypies locales. En effet, l'annotation de cellules urothéliales basales malignes par des experts est difficile. Les urothéliales basales contiennent divers degrés d'atypie, ce qui rend complexe l'élaboration d'une classification binaire claire et reproductible entre cellules saines et malignes. Contrairement à l'étude menée par Butke et al. [99], les annotations disponibles ici sont le résultats des examens de cystoscopie et d'histologie, les examens de référence pour le diagnostic du cancer de la vessie, et concernent non seulement les carcinomes urothéliaux de haut-grade mais également les carcinomes urothéliaux de bas-grade.

La méthode proposée de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales, illustrée Figure 5.6, est composée de trois étapes :

1. La détection de sous-images d'intérêt à partir de la lame de cytologie digitale.
2. Le calcul de vecteur de caractéristiques pour chaque sous-image d'intérêt.
3. La prédiction du diagnostic global à partir des vecteurs de caractéristiques des objets de la lame de cytologie.

La détection des sous-images d'intérêt est décrite sous-section 5.3.2, le calcul des vecteurs de caractéristiques sous-section 5.3.3 et la prédiction du diagnostic global sous-section 5.3.5.

[97]: DOV et al. (2021), « Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images »

[98]: TERAMOTO et al. (2021), « Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning »

[72]: OTSU (1979), « A Threshold Selection Method from Gray-Level Histograms »

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

8: Les méthodes d'apprentissage à partir d'instances multiples sont décrites plus en détails dans la sous-section 5.3.4.

[99]: BUTKE et al. (2021), « End-to-end Multiple Instance Learning for Whole-Slide Cytopathology of Urothelial Carcinoma »

[98]: TERAMOTO et al. (2021), « Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning »

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

[99]: BUTKE et al. (2021), « End-to-end Multiple Instance Learning for Whole-Slide Cytopathology of Urothelial Carcinoma »

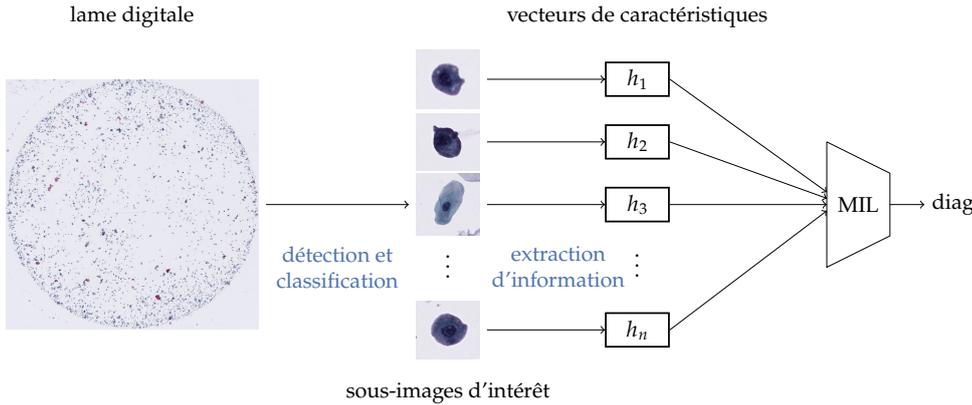


FIGURE 5.6 – Vue d’ensemble de la méthode de diagnostic automatique proposée. Des sous-images d’intérêt sont détectées à partir de la lame de cytologie, des vecteurs de caractéristiques sont calculés pour chaque sous-image et un modèle d’apprentissage à partir d’instances multiples (MIL) est utilisé pour prédire le diagnostic final.

5.3.1 Acquisition des lames de cytologie digitales

Il est possible de créer plusieurs lames de cytologie à partir d’un unique échantillon d’urine. Ainsi, aux 1023 patients composant le jeu de données *visiocyt* correspondent 2323 lames de cytologie digitales. Ces lames ont été numérisées à l’aide d’un scanner de lames Hamamatsu NanoZoomer® S360, selon 11 plans focaux au grossissement $\times 40$ ⁹.

9: Cela correspond à une résolution de 0.23 μm par pixel.

5.3.2 Identification des objets d’intérêt

L’identification des objets d’intérêt à partir d’une lame de cytologie digitale est effectuée en deux étapes. Premièrement, les objets de premier plan sont détectés à partir de la lame entière, puis les cellules d’intérêt¹⁰ sont identifiées parmi tous les objets de premier plan.

10: Dans le cas présent, les cellules d’intérêt sont les cellules urothéliales basales isolées.

Détection de tous les objets de premier plan Les lames de cytologie urinaire utilisées traditionnellement pour le diagnostic du cancer de la vessie sont colorées, à l’aide d’une coloration de Papanicolaou [3]. Les objets de premier plan, colorés, sont donc facilement identifiables par rapport à l’arrière-plan, non-coloré. Ainsi, pour détecter les objets de premier plan, un simple seuillage est suffisant. La valeur de seuil s_{rgb} à utiliser n’est pas très sensible, et peut-être déterminé automatiquement, par exemple à l’aide d’un seuillage automatique de Otsu [72], ou fixée à une valeur prédéfinie.

[3]: PAPANICOLAOU (1942), « A New Procedure for Staining Vaginal Smears »

Dans une première approximation, un pixel $I^p = (I_r^p, I_g^p, I_b^p) \mid I_{\{r,g,b\}}^p \in [0, 255]$ est considéré comme appartenant au premier plan si sa distance euclidienne à l’arrière-plan $I^{bg} = (I_r^{bg}, I_g^{bg}, I_b^{bg}) \mid I_{\{r,g,b\}}^{bg} \in [0, 255]$ dans l’espace de couleur RGB est supérieure à s_{rgb} :

[72]: OTSU (1979), « A Threshold Selection Method from Gray-Level Histograms »

$$I^p = \begin{cases} \text{premier-plan} & \text{ssi } \max_{c \in \{r,g,b\}} |I_c^p - I_c^{bg}| > s_{rgb} \\ \text{arrière-plan} & \text{sinon} \end{cases} \quad (5.1)$$

Les valeurs pour I^{bg} et s_{rgb} sont déterminées expérimentalement à $I^{bg} = (229, 222, 226)$ et $s_{rgb} = 14$.

Les composantes connexes obtenues par le seuillage décrit précédemment sont ensuite traitées par opérations morphologiques pour retirer tous les trous de moins de 128 pixels. Finalement, les objets de premier plan sont les composantes connexes dont l’aire est comprise entre 1200 et 90000 pixels, afin de retirer de l’analyse les objets trop petits ou trop gros pour être des objets d’intérêt¹¹.

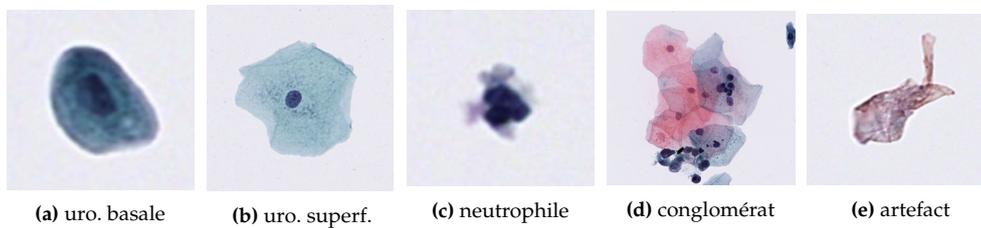
11: Les objets de premier plan ainsi filtrés sont des débris et autre artefacts.

La détection des objets de premier plan est effectuée au grossissement $\times 20$ à partir de la section centrale S_{centre} de la lame de cytologie digitale.

Classification des cellules d'intérêt parmi les objets de premier plan Un réseau de neurones convolutionnel ResNet-18 [30] est entraîné à reconnaître cinq classes d'objets à partir des objets de premier plan obtenues dans l'étape précédente :

- Les cellules urothéliales basales isolées, d'intérêt pour le diagnostic du cancer de la vessie.
- Les cellules urothéliales superficielles.
- Les conglomerats, définis comme amas d'au moins deux cellules en contact.
- Les cellules polynucléaires neutrophiles ¹².
- Les artefacts, pour tout objet n'appartenant à l'une des quatre catégories précédentes.

Des exemples des différents types cellulaires rencontrés dans les lames de cytologie urinaire sont donnés Figure 5.7.



[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

12: Cellules immunitaires.

FIGURE 5.7 – Illustration des différents types cellulaires utilisés pour la classification des objets du premier plan. (a) les cellules urothéliales basales, cellules d'intérêt pour le diagnostic du cancer de la vessie. (b) les cellules urothéliales superficielles, (c) les cellules polynucléaires neutrophiles, (d) les conglomerats et (e) les artefacts.

Le modèle est entraîné à partir d'un jeu de données annoté composé des sections de meilleur focus S_{sharp} ¹³ de 7630 cellules urothéliales basales, 6308 cellules urothéliales superficielles, 1243 cellules polynucléaires neutrophiles, 4100 conglomerats et de 5982 artefacts. Les poids du modèle sont optimisés afin de minimiser la fonction de coût entropie croisée catégorielle (CCE) pendant 100 epochs avec un learning rate de 0.001.

Une fois entraîné, le modèle ResNet-18 permet de conserver uniquement les cellules urothéliales basales pour les étapes suivantes.

5.3.3 Calcul des caractéristiques expertes

Un vecteur de caractéristiques est calculé pour chaque cellule urothéliale basale détectée dans la sous-section 5.3.2. Ces caractéristiques sont dites *expertes* car inspirées des caractéristiques expertes décrites dans le Système de Paris [24]. Le calcul de ces caractéristiques expertes repose au préalable sur une segmentation sémantique précise des cellules urothéliales basales et de leurs noyaux.

13: Les sections de meilleur focus considérées ici sont déterminées par objet, pour des lames digitales numérisées selon trois plans focaux espacés de $\delta_z = 1\mu\text{m}$. Voir section 2.2 pour plus de détails.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

Segmentation précise des cellules par type de représentation La segmentation sémantique précise des cellules par rapport à l'arrière-plan est effectuée pour chaque sous-image à l'aide un seuillage automatique dit *triangle* [161].

[161]: ZACK et al. (1977), « Automatic measurement of sister chromatid exchange frequency. »

[16]: RONNEBERGER et al. (2015), « U-Net »

Segmentation précise des noyaux par type de représentation La segmentation sémantique des noyaux est effectuée, pour chaque sous-image, à l'aide d'un modèle U-Net [16]. Le modèle U-Net est un CNN couramment utilisé pour la segmentation sémantique d'objets, notamment pour des images biologiques.

Le modèle de segmentation de noyau est entraîné, à partir des représentations S_{sharp} des z-stacks, sur un jeu de données composé de 11036 cellules manuellement annotées. Parmi celles-ci, 2921 cellules ont été annotées après un premier entraînement du modèle U-Net afin d'ajouter aux données d'entraînement des exemples pour lesquels les prédictions du modèle étaient incorrectes¹⁴. Le jeu de données est réparti en 8938 cellules utilisées pour l'entraînement du modèle, 994 cellules utilisées pour l'évaluation du modèle lors du développement, notamment pour fixer les hyper-paramètres et 1100 cellules utilisées en test, pour l'évaluation finale des performances du modèle.

Le modèle U-Net a été développé en Python à l'aide de la librairie Pytorch [126]. Le modèle a été entraîné à minimiser la fonction de coût \mathcal{L}_{Dice} ¹⁵, pendant 500 epochs avec une taille de batch de 16 et un learning rate de 1^{-4} .

Pour éviter le sur-apprentissage du modèle, des couches de Dropout avec une probabilité de 0.1 ont été ajoutées après chaque bloc de convolution, à l'exception des premier et dernier blocs. À chaque étape de l'entraînement, une couche de Dropout a une probabilité p de désactiver un neurone d'un réseau de neurones. Cette technique encourage une redondance dans les réseaux de neurones et permet ainsi de limiter le sur-apprentissage.

Après entraînement à partir des sections S_{sharp} , le modèle atteint de très hautes performances avec un score de Dice de 0.929 sur l'ensemble de données de développement et 0.927 sur l'ensemble de données de test.

La segmentation précise des noyaux est une étape importante dans la méthode de diagnostic proposée, car elle conditionne le calcul des caractéristiques expertes. Pour obtenir une segmentation précise même en présence de variations de coloration ou d'intensité dans les lames de cytologie numériques, le modèle U-Net est entraîné avec des fortes augmentations de données, similaires à celles utilisées chapitre 3 sous-section 3.2.3. Les paramètres utilisés pour les augmentations sont reportés Tableau 5.1.

Transformation	Paramètres
rotation aléatoire	$r \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$
inversion verticale	$p = 0.5$
inversion horizontale	$p = 0.5$
recadrage aléatoire et complétion	$\text{dim} = (256, 256)$
variation de la luminosité	$v \in [0.75, 1.25]$
variation du contraste	$v \in [0.75, 1.25]$
variation de la saturation	$v \in [0.75, 1.25]$
variation de la teinte	$v \in [0, 1]$

Une fois les masques de segmentation des noyaux calculés pour chaque cellule urothéliale présente dans une lame, les caractéristiques expertes peuvent être calculées.

Calcul des caractéristiques expertes Un vecteur de caractéristiques expertes est calculé pour chaque cellule d'intérêt¹⁶. Les vecteurs de caractéristiques calculés sont inspirés des caractéristiques morphologiques cellulaires utilisées par les pathologistes et décrits dans le Système de Paris [24] pour le diagnostic du cancer de la vessie à partir de lames de cytologie.

Les caractéristiques expertes utilisées sont décrites ci-dessous :

14: Cette méthode classique dite de hard mining permet d'améliorer les performances d'un modèle en ajoutant au jeu de données des exemples difficiles, pour lequel le modèle a tendance à se tromper.

[126]: PASZKE et al. (2019), « PyTorch : An Imperative Style, High-Performance Deep Learning Library »

15: Rappel : $\mathcal{L}_{Dice} = 1 - \frac{\text{Score de Dice}}{\text{Score de Dice}}$ avec $\text{Score de Dice} = \frac{2 \times |Y_{true} \cap Y_{pred}|}{|Y_{true}| + |Y_{pred}|}$

TABLE 5.1 – Paramètres des augmentations utilisées lors de l'entraînement du modèle U-Net pour la segmentation sémantique des noyaux.

16: Dans le cas présent, une cellule urothéliale.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

Ratio Noyau-Cytoplasme (RNC) : Un fort RNC ¹⁷ est la caractéristique d'atypie la plus importante décrite dans le Système de Paris. À partir du masque de segmentation binaire de la cellule M_{cel} et du masque de segmentation binaire du noyau M_{noy} , le RNC est calculé comme suit :

17: Supérieur à 0.5.

$$\text{RNC} = \frac{\text{Aire}_{M_{noy}}}{\text{Aire}_{M_{cel}}} = \frac{\sum_{x,y} M_{noy}(x,y)}{\sum_{x,y} M_{cel}(x,y)} \quad (5.2)$$

Intensité : l'hyper-chromatisme, également caractéristique d'atypie importante dans le Système de Paris, est définie par une sur-coloration du noyau. Dans un premier temps, une intensité de référence est calculée pour chaque lame. Cette intensité de référence, notée I_{ref} , est l'intensité moyenne des cytoplasmes des cellules urothéliales basales d'une lame et est calculée pour une section $S \in \{S_{centre}, S_{sharp}, S_{edf}\}$ comme suit :

$$I_{ref} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{x,y \in M_{cel} \setminus M_{noy}} S^i(x,y)}{\sum_{x,y \in M_{cel} \setminus M_{noy}} 1} \quad (5.3)$$

Avec :

N : le nombre d'urothéliales basales détectées dans la lame

M_{cel} : le masque de segmentation binaire de la cellule

M_{noy} : le masque de segmentation binaire du noyau

Dans un second temps, pour une image de cellule urothéliale basale S , une métrique d'intensité I est calculée comme suit :

$$I = \frac{1}{I_{ref}} \times \frac{\sum_{x,y \in M_{noy}} S(x,y)}{\sum_{x,y \in M_{noy}} 1} \quad (5.4)$$

Caractéristiques de texture : afin que le modèle puisse reconnaître des textures de chromatine irrégulières et des bordures de chromatine irrégulières, critères d'atypie secondaires du Système de Paris, plusieurs caractéristiques de texture sont calculées :

Écart-type de l'intensité : l'écart-type de l'intensité est calculé pour les pixels appartenant au masque de segmentation du noyau :

$$\sigma_I = \sqrt{\sum_{x,y \in M_{noy}} (S(x,y) - \bar{S}(x,y))^2} \quad (5.5)$$

Trois caractéristiques de texture sont ensuite tirées des descripteurs de Haralick [57], basées sur la matrice de co-occurrence des niveaux de gris (GLCM) $P(i,j)$ calculée sur le masque de segmentation du noyau M_{noy} de façon similaire à la sous-section 3.2.3.

[57]: HARALICK et al. (1973), « Textural Features for Image Classification »

$$\text{Énergie} = \sqrt{\sum_i \sum_j P(i,j)^2} \quad (5.6)$$

$$\text{Entropie} = - \sum_i \sum_j P(i,j) \log(P(i,j)) \quad (5.7)$$

$$\text{Homogénéité} = \sum_i \sum_j \frac{1}{1 + (i-j)^2} P(i,j) \quad (5.8)$$

Caractéristiques morphologiques : Deux caractéristiques sont calculées pour tenir compte de la morphologie du noyau dans les cellules urothéliales basales. L'intuition est que les caractéristiques morphologiques du noyau peuvent aider à différencier les cellules atypiques des cellules dégénérées ou abîmées. Bien que ces métriques de morphologie ne soient pas directement utilisées dans le Système de Paris, il a été montré que la morphologie des noyaux peut être une caractéristique d'atypie cellulaire [162]. Ces caractéristiques morphologiques sont décrites ci-dessous :

$$\text{Ratio à l'enveloppe convexe} = \frac{\text{Aire}_{M_{\text{noy}}}}{\text{Aire}_{\text{Conv}(M_{\text{noy}})}} \quad (5.9)$$

$$\text{Circularité} = \frac{4\pi \times \text{Aire}_{M_{\text{noy}}}}{\text{Périmètre}_{M_{\text{noy}}}^2} \quad (5.10)$$

En résumé, pour chaque cellule urothéliale basale détectée dans une lame de cytologie digitale, un vecteur de caractéristique contenant les 8 métriques : *Ratio Noyau Cytoplasme*, *Intensité*, *Écart-type de l'intensité*, *Énergie*, *Entropie*, *Homogénéité*, *Ratio à l'enveloppe convexe* et *Circularité* est calculé.

5.3.4 Paradigme de l'apprentissage à partir d'instances multiples

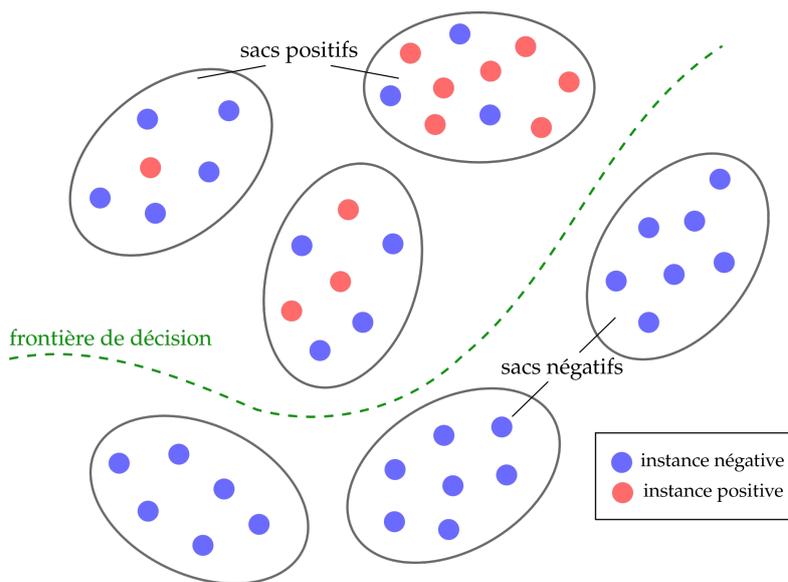


FIGURE 5.8 – Apprentissage à partir d'instances multiples (MIL) : un sac d'instances est positif si il contient au moins une instance positive, sinon il est négatif.

L'apprentissage à partir d'instances multiples (MIL) est une méthode d'apprentissage faiblement supervisée, pour laquelle une unique classe est associée à une collection d'instances appelée "sac d'instances" [163].

Un modèle de MIL binaire (sain / malade) est donc un modèle qui cherche à prédire une unique classe $y \in \{0, 1\}$ à partir d'un sac d'instances $X = \{x_1, x_2, \dots, x_N\}$ avec N le nombre d'instances dans le sac¹⁸ et $x_i \in \mathbb{R}^D$ le vecteur de caractéristiques décrivant l'instance i .

Il est également supposé que pour chaque instance $i \in [1, N]$ il existe un label $y_i \in \{0, 1\}$, qui n'est cependant pas accessible pour l'entraînement du modèle de MIL.

[163]: DIETTERICH et al. (1997), « Solving the multiple instance problem with axis-parallel rectangles »

18: N pouvant varier pour différents sacs dans le cas général.

Un sac d'instances est considéré positif s'il contient au moins une instance positive. Le problème d'apprentissage à partir d'instances multiples dans le cas binaire peut alors s'écrire :

$$y = \begin{cases} 1 & \text{ssi } \max_{i \in \{1, \dots, N\}} y_i = 1 \\ 0 & \text{sinon} \end{cases} \quad (5.11)$$

Le paradigme de l'apprentissage à partir d'instances multiples est illustré Figure 5.8.

Le problème de MIL classique suppose également que la collection de N instances composant un sac est non-ordonnée. Ainsi un modèle de MIL doit être invariant par permutation des instances dans le sac. En reprenant les notations introduites par Ilse et al. [62], le modèle de MIL peut être écrit de la manière suivante :

$$S(X) = g \circ \sigma_{x \in X} \circ f(x) \quad (5.12)$$

Avec :

- $S(X) \in \mathbb{R}$: le score du modèle de MIL pour le sac X
- $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$: une transformation appliquée à chaque instance x_i
- $g : \mathbb{R}^{d'} \rightarrow \mathbb{R}$: une transformation appliquée à un vecteur décrivant le sac
- $\sigma_{x \in X}$: une fonction d'agrégation invariante par permutation des instances

Comme expliqué dans l'étude [62], il existe principalement deux approches d'apprentissage à partir d'instances multiples :

Approches basées sur la classification d'instances Pour ces approches, le modèle cherche d'abord à identifier la classe de chaque instance du sac (i.e. prédire y_i pour chaque x_i). Ainsi en reprenant la notation introduite Équation 5.12, f est une fonction de classification d'instance, σ une fonction d'agrégation de scores invariante par permutation et g est la fonction identité.

Approches basées sur la classification d'une représentation Pour ces approches, le modèle cherche à construire une représentation du sac¹⁹ à partir des vecteurs de caractéristiques associés aux instances x_i . La prédiction globale est ensuite effectuée à partir de cette représentation apprise. En reprenant à nouveau la notation de l'Équation 5.12, f est une fonction qui projette chaque x_i vers une représentation h_i . σ est une fonction d'agrégation de représentations, invariante par permutation, et g est une fonction de classification de la représentation du sac.

5.3.5 Classification de lames par apprentissage à partir d'instances multiples

Les méthodes d'apprentissage à partir d'instances multiples permettent la classification de sacs d'instances, connaissant le label de chaque sac d'entraînement et en ayant un vecteur de caractéristiques pour chaque instance contenue dans chaque sac. Pour le problème de classification de lames de cytologie digitales afin de diagnostiquer le cancer de la vessie, le sac correspond à la lame de cytologie et les instances qui composent ce sac sont les cellules urothéliales basales détectées dans la lame. Ainsi, à partir de la collection de vecteurs de caractéristiques expertes extraite à partir des cellules urothéliales d'une lame de cytologie digitale, l'objectif est de prédire le diagnostic global de la lame : 0 pour

[62]: Ilse et al. (2018), « Attention-based Deep Multiple Instance Learning »

[62]: Ilse et al. (2018), « Attention-based Deep Multiple Instance Learning »

19: Également appelée *embedding*.

un patient sain et 1 pour un patient atteint par un carcinome urothélial de bas-grade ou de haut-grade.

Création des sacs d’instances Les méthodes de MIL acceptent généralement des sacs avec un nombre N d’instances variable [62][164][165]. Pour pouvoir entraîner un modèle de MIL à partir de sacs contenant un nombre d’instances variables, ces méthodes utilisent des batchs de taille 1 lors de l’optimisation. Cependant, l’utilisation de batchs de taille 1 peut perturber l’apprentissage de réseaux de neurones par descente de gradients. En effet, les gradients sont calculés sur un unique exemple à chaque étape avant de mettre à jour les poids du réseau de neurones, et ont donc tendance à être plus bruités que des gradients calculés sur un plus grand nombre d’exemples.

Une autre approche, illustrée Figure 5.9, est proposée pour fixer le nombre d’instances²⁰ N sélectionnées par lame de cytologie à partir de connaissances à priori. Soit une lame de cytologie pour laquelle les cellules urothéliales basales d’intérêt ont été préalablement détectées et un vecteur de caractéristiques expertes a été calculé pour chaque cellule urothéliale basale détectée. Les cellules urothéliales basales sont classées en fonction de leur Ratio Noyau-Cytoplasme, et les N cellules avec le plus haut RNC sont sélectionnées pour former le sac qui représente la lame de cytologie, N étant donc un hyper-paramètre fixé de la méthode. Ainsi, chaque lame l de cytologie est représenté en pratique par un tenseur \mathcal{T}_l de dimensions (N, F) , avec $N = 300$ le nombre d’instances et $F = 8$ la taille du vecteur de caractéristique experte associé à chaque instance. Puisque un haut RNC est le critère d’atypie le plus important dans le Système de Paris, cette approche revient approximativement²¹ à sélectionner, pour chaque lame, les N cellules urothéliales basales les plus atypiques. Outre l’intérêt computationnel, une double intuition est à l’origine de cette approche :

- Pour les lames de patients atteints par un carcinome urothélial : sélectionner les N instances les plus atypiques permet de retirer de l’analyse les cellules saines également présentes, ce qui donne plus de poids aux cellules d’intérêt pour le diagnostic.
- Pour les lames de patients sains : sélectionner les N instances les plus atypiques permet de donner de l’importance aux cellules urothéliales basales parfois légèrement atypiques que l’on retrouve naturellement chez les patients sains. Donner ainsi de l’importance à ces cellules est une recherche d’exemple négatifs non triviaux, qui peut aider le modèle de MIL à produire moins de faux positifs.

Avec un nombre d’instances N par sac fixé, il peut arriver que certaines lames aient un nombre de cellules urothéliales basales détectées inférieur à N . Pour ces cas, le sac est complété par de fausses instances, dont le vecteur de caractéristiques est le vecteur nul. Une attention particulière est portée au design des méthodes de MIL appliquées par la suite, afin que ces fausses instances ne puissent pas influencer la prédiction. Les fausses instances ajoutées servent uniquement à assurer une taille constante des tenseurs \mathcal{T}_l représentant les lames.

Méthodes d’apprentissage à partir d’instances multiples Plusieurs méthodes dites d’apprentissage à partir d’instances multiples sont testées pour résoudre le problème de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales. Tous les réseaux de neurones considérés ont deux blocs communs, permettant des comparaisons justes.

Extracteur de caractéristiques : à partir des vecteurs de caractéristiques expertes, un perceptron multi-couches, nommé par la suite Extracteur, est utilisé pour calculer des caractéristiques plus complexes. Son architecture est décrite Figure 5.10.

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

[164]: LI et al. (2021), « Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning »

[165]: SHAO et al. (2021), « TransMIL »

20: Une instance correspond ici à une cellule urothéliale basale.

21: Approximativement car selon le Système de Paris, un haut RNC (supérieur à 0.5) est une condition nécessaire mais pas suffisante pour caractériser l’atypie cellulaire.

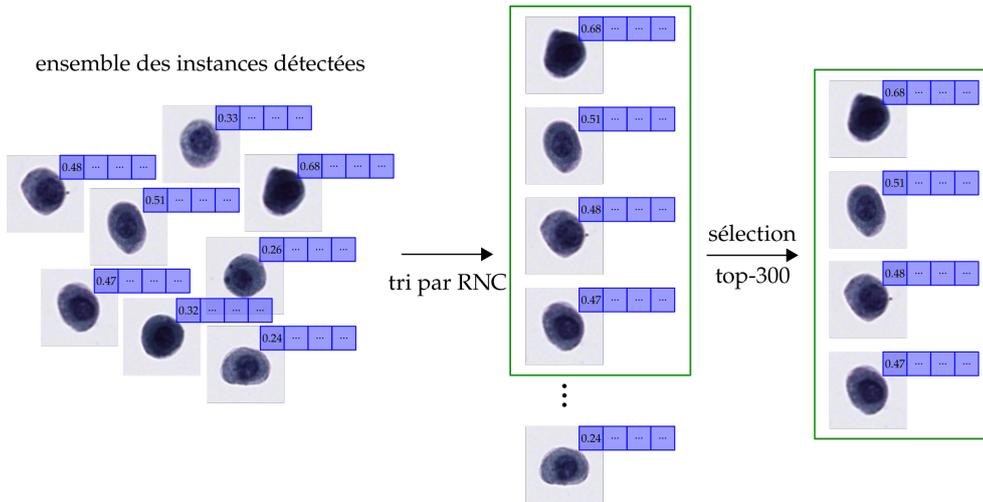


FIGURE 5.9 – Illustration de l’approche utilisée pour la création des sacs d’instances. Au lieu d’utiliser toutes les cellules urothéliales basales détectées dans une lame comme instances, les cellules sont triées en fonction de leur RNC, et seulement les cellules avec les 300 plus haut RNC sont sélectionnées comme instances pour le sac. Dans le cas où moins de 300 cellules ont été détectées dans la lame, toutes les cellules sont sélectionnées, et complétées avec des instances factices dont le vecteur de caractéristiques est le vecteur nul.

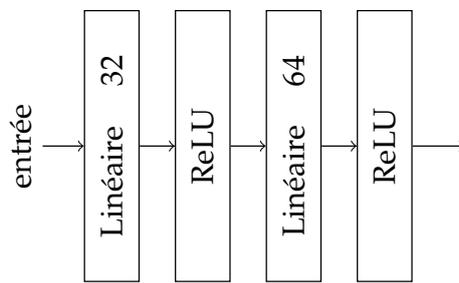


FIGURE 5.10 – Architecture de l’extracteur de caractéristique utilisé pour toutes les méthodes de MIL.

Module de classification : à partir d’un vecteur de caractéristiques, une couche linéaire suivie d’une couche d’activation sigmoïde est utilisée pour prédire un score de malignité noté y_{pred} ou \hat{y} , avec $y_{pred} \in [0, 1]$. Son architecture est décrite Figure 5.11.

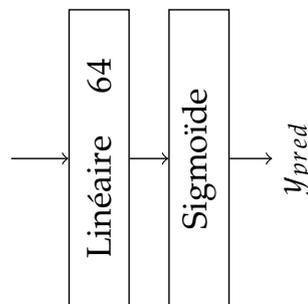


FIGURE 5.11 – Architecture du module de classification utilisé pour toutes les méthodes de MIL.

Méthodes de référence : Quatre réseaux de neurones, qui peuvent être considérés comme quatre variations d’une même approche, sont utilisés comme méthodes de MIL de référence.

MIL par agrégation moyenne des scores de classification d’instances : Cette approche, nommée ci-après *meanpool(inst.)* appartient aux approches de MIL basées sur la classification d’instances. En reprenant les notations introduites Équation 5.12²², f est donc l’extracteur de caractéristiques suivi du module de classification (voir Équation 5.13), σ l’opérateur moyenne défini Équation 5.14 et g la fonction identité.

22: Rappel : $S(X) = g \circ \sigma \circ f$

$$f(x_i) = \hat{y}_i \in [0, 1] \quad \text{avec} \quad x_i \in X \quad (5.13)$$

$$\sigma_{x_i \in X}(Y) = \frac{1}{N} \sum_{i=1}^N \hat{y}_i \quad \text{avec} \quad Y = \{\hat{y}_1, \dots, \hat{y}_N\} \quad (5.14)$$

MIL par agrégation maximale des scores de classification d'instances : Cette approche, nommée ci-après *maxpool(inst.)* appartient également aux approches de MIL basées sur la classification d'instances. Ici, f est l'extracteur de caractéristiques suivi du module de classification, σ l'opérateur maximum et g la fonction identité ²³.

$$\sigma(Y) = \max_{i \in [1, N]} \hat{y}_i \quad \text{avec} \quad Y = \{\hat{y}_1, \dots, \hat{y}_N\} \quad (5.15)$$

MIL par classification de représentation par agrégation moyenne : Cette approche, nommée ci-après *meanpool(repr.)* appartient aux approches de MIL basées sur la classification d'une représentation du sac. f est l'extracteur de caractéristiques, σ l'opérateur moyenne et g le module de classification ²⁴.

$$f(x_i) = \mathbf{h}_i \in \mathbb{R}^d \quad \text{avec} \quad x_i \in X \quad (5.16)$$

$$\mathbf{z} = \sigma(H) = \frac{1}{N} \sum_{i=1}^N \mathbf{h}_i \quad \text{avec} \quad H = \{\mathbf{h}_1, \dots, \mathbf{h}_N\} \quad (5.17)$$

$$g(\mathbf{z}) = y_{pred} \in [0, 1] \quad (5.18)$$

MIL par classification de représentation maximale : Cette approche, nommée ci-après *maxpool(repr.)* appartient également aux approches de MIL basées sur la classification d'une représentation. f est l'extracteur de caractéristiques, σ l'opérateur maximum et g le module de classification ²⁵.

$$\mathbf{z} = \sigma(H) = \max_{i \in [1, N]} \mathbf{h}_i \quad \text{avec} \quad H = \{\mathbf{h}_1, \dots, \mathbf{h}_N\} \quad (5.19)$$

Attention-Based MIL (ABMIL) : Cette méthode a été introduite par Ilse et al. [62]. Dans cette approche, σ est un opérateur de moyenne pondérée :

$$\mathbf{z} = \sigma(H) = \sum_{i=1}^N a_i \mathbf{h}_i \quad \text{avec} \quad H = \{\mathbf{h}_1, \dots, \mathbf{h}_N\} \quad (5.20)$$

Les poids a_i sont calculés à partir des vecteurs de représentation d'instances \mathbf{h}_i , comme détaillé Équation 5.21 en reprenant les notations introduites dans [62].

$$a_i = \frac{\exp \left\{ \mathbf{w}^T \left(\tanh(\mathbf{V} \mathbf{h}_i^T) \odot \text{sigm}(\mathbf{U} \mathbf{h}_i^T) \right) \right\}}{\sum_{j=1}^N \exp \left\{ \mathbf{w}^T \left(\tanh(\mathbf{V} \mathbf{h}_j^T) \odot \text{sigm}(\mathbf{U} \mathbf{h}_j^T) \right) \right\}} \quad (5.21)$$

Avec \mathbf{U} , \mathbf{V} , \mathbf{w} des paramètres entraînaibles du modèle à optimiser.

Dual-Stream MIL (DSMIL) : Cette méthode a été introduite par Li et al. [164]. Contrairement aux approches présentées précédemment, qui sont basées soit sur l'agrégation des classifications d'instances, soit sur la classification d'une agrégation des vecteurs de caractéristiques, la méthode DSMIL cherche à combiner les deux approches simultanément.

À partir des représentations d'instances $H = \{\mathbf{h}_1, \dots, \mathbf{h}_N\}$, une première branche prédit un score par instance, dont est déduit un score global c_m en suivant la même approche que *maxpool(inst.)*. L'instance m ayant le plus haut score d'atypie parmi toutes les instances

23: Ici le module de classification utilisé est le même que Équation 5.13 et l'opérateur maximum est défini Équation 5.15.

24: L'extracteur de caractéristiques f , l'opérateur moyenne σ et le module de classification g utilisés ici sont définis respectivement Équation 5.16, Équation 5.17 et Équation 5.18.

25: L'extracteur de caractéristiques utilisé ici est le même que Équation 5.16, l'opérateur maximum est défini Équation 5.19 et le module de classification est défini Équation 5.18.

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

[164]: LI et al. (2021), « Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning »

du sac est appelée instance critique. Dans une seconde branche, pour chaque vecteur \mathbf{h}_i sont calculés deux vecteurs *requête* \mathbf{q}_i et *valeur* \mathbf{v}_i calculés comme suit :

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i, \quad \mathbf{v}_i = \mathbf{W}_v \mathbf{h}_i, \quad \text{avec } i \in [1, N] \quad (5.22)$$

Avec \mathbf{W}_q et \mathbf{W}_v des matrices de poids, apprises par le modèle DSMIL. À partir des vecteurs \mathbf{q}_i calculés pour chaque instance, une mesure de distance $U(\mathbf{h}_i, \mathbf{h}_m)$ à l'instance critique m est définie :

$$U(\mathbf{h}_i, \mathbf{h}_m) = \frac{\exp(\langle \mathbf{q}_i, \mathbf{q}_m \rangle)}{\sum_{k=0}^N \exp(\langle \mathbf{q}_k, \mathbf{q}_m \rangle)} \quad (5.23)$$

Avec $\langle \cdot, \cdot \rangle$ l'opérateur produit scalaire entre deux vecteurs. Un vecteur de représentation du sac \mathbf{b} , défini Équation 5.24, passe ensuite dans le module de classification pour obtenir un score c_b .

$$\mathbf{b} = \sum_{i=1}^N U(\mathbf{h}_i, \mathbf{h}_m) \mathbf{v}_i \quad (5.24)$$

Le score de classification global de la méthode DSMIL est finalement :

$$y_{pred} = \frac{c_m + c_b}{2} \quad (5.25)$$

Transformer-based MIL (TransMIL) : Cette méthode a été introduite par [165]. Dans cette approche, Shao et al. utilisent le concept de *self-attention* des modèles Transformers [166] afin de modéliser l'importance relative des instances pour une lame.

[165]: SHAO et al. (2021), « TransMIL »

La méthode ABMIL, décrite précédemment, détermine les scores d'importance²⁶ par instance (avant normalisation) uniquement en fonction de \mathbf{h}_i le vecteur de caractéristiques décrivant l'instance i . Au contraire, la méthode TransMIL, par le mécanisme de *self-attention*, identifie les instances importantes après des comparaisons deux à deux entre toutes les instances du sac. L'intuition derrière cette approche, est que l'importance d'une instance pour la classification du sac n'est pas intrinsèque, mais dépend des autres instances présentes dans le sac. En termes biologiques, une cellule légèrement atypique peut ainsi être considérée comme peu importante s'il s'agit de la seule cellule atypique de la lame, ou importante en présence de nombreuses autres cellules également atypiques.

[166]: VASWANI et al. (2017), « Attention Is All You Need »

26: Appelés *attention scores* dans la publication originale.

Pour chaque vecteur de caractéristiques \mathbf{h}_i , trois vecteurs *requête* \mathbf{q}_i , *clé* \mathbf{k}_i et *valeur* \mathbf{v}_i sont calculés :

$$\mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i, \quad \mathbf{k}_i = \mathbf{W}_k \mathbf{h}_i, \quad \mathbf{v}_i = \mathbf{W}_v \mathbf{h}_i, \quad \text{avec } i \in [1, N] \quad (5.26)$$

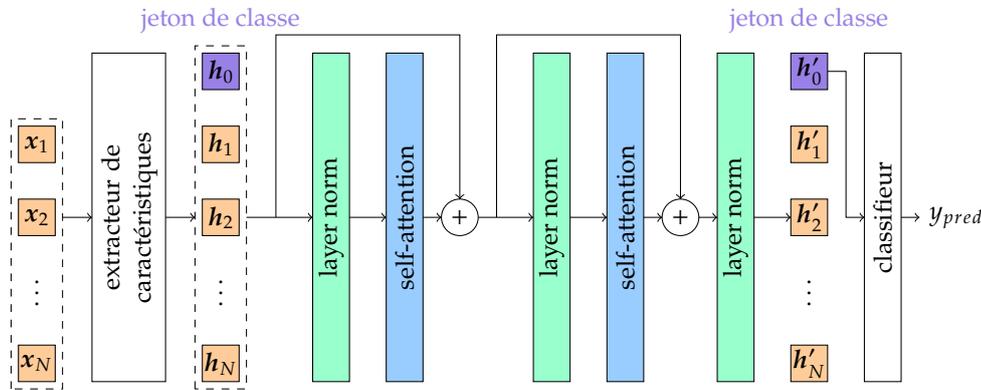
Avec \mathbf{W}_q , \mathbf{W}_k et \mathbf{W}_v des poids appris par le modèle TransMIL.

Une couche de *self-attention* transforme ensuite l'ensemble de vecteurs $H = \{\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_N\}$ en $H' = \{\mathbf{h}'_0, \mathbf{h}'_1, \dots, \mathbf{h}'_N\}$ ²⁷ de la façon suivante :

$$\mathbf{h}'_i = \mathbf{h}_i + \sum_{j=0}^N (\mathbf{q}_i \mathbf{k}_j^T) \mathbf{v}_j \quad (5.27)$$

27: Avec \mathbf{h}_0 le jeton de classe, un vecteur de paramètres propre au module de *self-attention* et de même dimensions que les \mathbf{h}_i .

Dans le modèle de TransMIL utilisé ici, l'opérateur σ est composé de deux couches successives de *self-attention* utilisées avec des couches de normalisation de type *layer norm* [167]. Le jeton de classe h'_0 obtenu après la seconde couche de *self-attention* et normalisation est utilisé comme vecteur de représentation du sac, et passe ainsi dans le module de classification²⁸ pour obtenir la prédiction globale y_{pred} . Une illustration de la méthode utilisée est proposée Figure 5.12.



Une première différence entre la méthode utilisée ici et la méthode TransMIL originale décrite dans [165] porte sur l'approximation de la couche de *self-attention*. La couche de *self-attention* utilisée nécessite le calcul de $q_i k_j^T$ pour $i, j \in [0, N]$, soit le calcul de la matrice QK^T de dimensions $(N + 1, N + 1)$ ²⁹. Dans le cas où le nombre d'instances par sac est important, le calcul de la matrice QK^T est trop coûteux en pratique. Pour y remédier, Shao et al. proposent de calculer une approximation de QK^T à l'aide de la méthode proposée par [168]. Dans ce travail, la création des sacs d'instances est effectuée en sélectionnant au maximum 300 cellules par lame, ainsi la matrice QK^T est calculée directement plutôt qu'approximée.

De plus, la méthode originale TransMIL ajoute, pour l'analyse de lames d'histologie digitales, un module PPEG³⁰ qui n'est pas réutilisé ici. Le module PPEG est utilisé pour encoder l'information de position entre instances : cette information de position est trouvée utile pour analyser les images de tissus [165], mais n'est pas transposable à l'analyse d'images de cytologie dans lesquelles les objets sont disposés aléatoirement et n'ont pas de relation, à priori, avec leurs voisins.

Entraînement des méthodes Toutes les méthodes sont développées en Python à l'aide de la bibliothèque Pytorch [126]. Tous les modèles ont été entraînés à l'aide de l'optimiseur Adam, avec des batchs de taille 128 et un learning rate initial fixé à $1e^{-3}$. Les modèles de référence, ABMIL et DSMIL ont été entraînés pendant 1000 epochs, avec une division par 10 du learning rate aux epochs 900 et 950. Parce que le modèle TransMIL converge beaucoup plus rapidement, comme observé par Shao et al. [165], ce modèle est entraîné pendant 200 epochs, avec une division par 10 du learning rate aux epochs 100 et 150.

Pour les problèmes de classification, une fonction de coût classiquement utilisée est l'entropie croisée :

$$\mathcal{L}_{CE} = y_{true} \cdot \log(y_{pred}) + (1 - y_{true}) \cdot \log(1 - y_{pred}) \quad (5.28)$$

Dans ce travail, une fonction de coût légèrement différente est utilisée : l'entropie croisée

[167]: BA et al. (2016), *Layer Normalization*

28: g dans la notation Équation 5.12.

FIGURE 5.12 – Illustration de la méthode TransMIL utilisée. Le vecteur de caractéristiques expertes X passe par le module extracteur de caractéristiques pour obtenir $H = \{h_1, \dots, h_N\}$ auquel est ajouté un jeton de classe h_0 . Le tenseur H passe ensuite par deux couches de *self-attention* associées à de couches de normalisation. Le jeton de classe h'_0 obtenu en sortie de la deuxième couche de *self-attention* après normalisation est utilisé comme représentation du sac, et passe par le module de classification pour obtenir la prédiction globale y_{pred} .

[165]: SHAO et al. (2021), « TransMIL »

29: Avec N le nombre d'instances dans le sac, plus le jeton de classe h_0 .

[168]: XIONG et al. (2021), « Nyström-former »

30: Pour *Pyramid Position Encoding Generator*.

[165]: SHAO et al. (2021), « TransMIL »

[126]: PASZKE et al. (2019), « PyTorch : An Imperative Style, High-Performance Deep Learning Library »

[165]: SHAO et al. (2021), « TransMIL »

généralisée (GCE), définie ci-dessous.

$$\mathcal{L}_{GCE} = \frac{1 - [y_{true} \cdot y_{pred} + (1 - y_{true})(1 - y_{pred})]^q}{q} \quad (5.29)$$

Avec q un hyper-paramètre de la fonction de coût. La fonction GCE a été introduite par Zhang et al. [169] afin d'entraîner des réseaux de neurones en présence d'annotations bruitées. Pour le problème considéré du diagnostic du cancer de la vessie à partir de lames de cytologie digitales, les annotations sont obtenues par les résultats des examens de cystoscopie et d'histologie : les méthodes de référence à ce jour. Il est donc peu probable que les annotations obtenues soient bruitées. En revanche, comme expliqué section 5.1, les lames de cytologie de patients atteints par un carcinome urothélial peuvent ne pas contenir de cellules atypiques. Ainsi, lors de l'entraînement, un réseau de neurones peut être encouragé à détecter un carcinome urothélial pour une lame d'apparence saine, ce qui est, en pratique, équivalent à des annotations bruitées.

[169]: ZHANG et al. (2018), « Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels »

5.4 Résultats

5.4.1 Métriques d'évaluation

Soit y_{true} la vérité terrain pour le diagnostic d'une lame de cytologie ³¹ et y_{pred} la prédiction d'un modèle pour cette lame. La prédiction y_{pred} obtenue en sortie des modèles peut prendre toutes les valeurs entre 0 et 1. Le diagnostic prédit par le modèle, noté \hat{y} , est alors défini de la façon suivante :

$$\hat{y} = \begin{cases} 1 & \text{si } y_{pred} \geq 0.5 \\ 0 & \text{sinon} \end{cases} \quad (5.30)$$

31: Rappel : 1 si la lame provient d'un patient atteint d'un carcinome urothélial, déterminé par cystoscopie et histologie, 0 si la lame provient d'un patient sain.

A partir de y_{true}^l et de \hat{y}^l , il est possible de définir, pour un ensemble de lames $l \in [1, L]$, le nombre de vrais positifs (TP), de vrais négatifs (TN), de faux positifs (FP) et de faux négatifs (FN) :

$$TP = \sum_{l=1}^L y_{true}^l \cdot \hat{y}^l \quad (5.31)$$

$$TN = \sum_{l=1}^L (1 - y_{true}^l) \cdot (1 - \hat{y}^l) \quad (5.32)$$

$$FP = \sum_{l=1}^L (1 - y_{true}^l) \cdot \hat{y}^l \quad (5.33)$$

$$FN = \sum_{l=1}^L y_{true}^l \cdot (1 - \hat{y}^l) \quad (5.34)$$

Finalement, les métriques d'évaluation suivantes sont utilisées pour quantifier les performances de diagnostic :

$$\text{Exactitude} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.35)$$

$$\text{Sensibilité} = \frac{TP}{TP + FN} \quad (5.36)$$

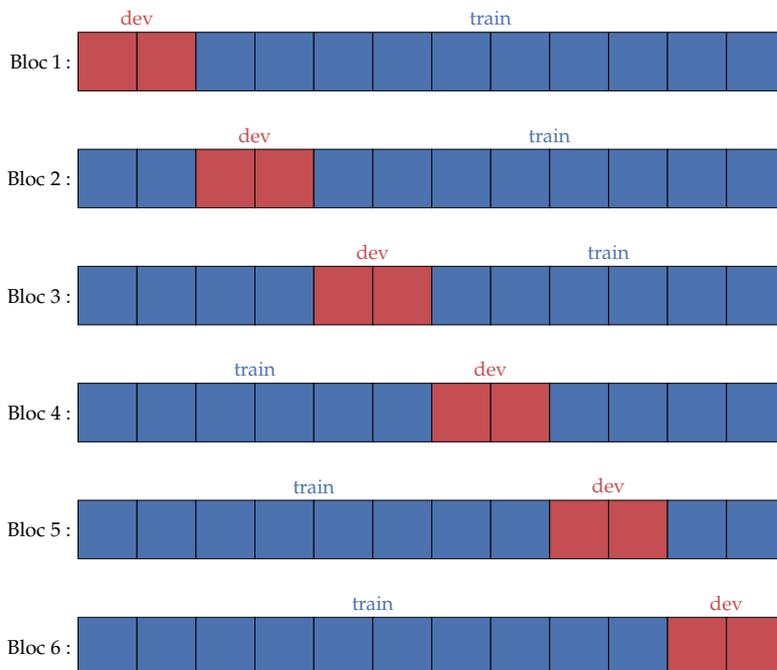
$$\text{Spécificité} = \frac{TN}{TN + FP} \quad (5.37)$$

$$\text{AUC} = \int_0^1 \text{sensibilité} d(1 - \text{spécificité}) \quad (5.38)$$

5.4.2 Évaluation par validation croisée

Pour déterminer certains hyper-paramètres d'architecture ou d'apprentissage, comme le learning rate ou le nombre d'époques à utiliser pour l'optimisation, il est nécessaire d'évaluer de nombreuses fois les performances des modèles. Il convient alors d'éviter d'utiliser l'ensemble de données de Test pour conduire de telles évaluations, afin de ne pas sur-apprendre un ensemble d'hyper-paramètres qui permettrait d'atteindre de bonnes performances de diagnostic sur les données de Test sans être représentatives des performances réelles du modèle.

Pour éviter ce sur-apprentissage, lors de leur développement les modèles sont évalués par validation croisée à six blocs³². La validation croisée consiste à diviser l'ensemble des données d'entraînement en deux sous-ensembles : un sous-ensemble *Train* utilisé pour entraîner une version du modèle et un sous-ensemble *Dev* utilisé pour évaluer les performances de ce modèle. Dans une validation croisée à k blocs, k splits sont effectués et pour chaque split, $\frac{1}{k}$ des échantillons sont sélectionnés dans le sous-ensemble *Dev*. Pour chacun des k splits, une version du modèle, notée Θ^j avec $j \in [1, \dots, k]$, est donc entraînée à l'aide de $\frac{k-1}{k}$ des données et évaluée sur $\frac{1}{k}$ des données. Une illustration du processus de validation croisée est proposée Figure 5.13.



Afin que les sous-ensembles *Dev* soient représentatifs de la base de données d'entraînement, les échantillons *Dev* ne sont pas sélectionnés complètement au hasard, mais

32: Le nombre de blocs utilisé dans la validation, ici six, est choisi arbitrairement. Augmenter le nombre de blocs permet d'augmenter également, pour chaque bloc, la quantité de données disponibles pour l'entraînement de modèles en diminuant d'autant la quantité de données utilisées pour l'évaluation.

FIGURE 5.13 – Illustration du processus d'évaluation par validation croisée à six blocs. Pour chaque bloc, $\frac{5}{6}$ des données est utilisé pour l'entraînement d'une version d'un modèle et $\frac{1}{6}$ des données est utilisé pour l'évaluation des performances.

de façon à ce que chaque sous-ensemble Dev contienne $\frac{1}{k}$ des échantillons sains, $\frac{1}{k}$ des échantillons bas-grades et $\frac{1}{k}$ des échantillons haut-grades ³³.

Chaque lame d'entraînement se retrouve une seule fois dans le sous-ensemble Dev au cours de la validation croisée. Les prédictions obtenues pour les lames Dev lors de la validation croisée sont ensuite utilisées pour évaluer les performances de diagnostic des modèles.

33: Cette stratégie est communément appelée validation croisée stratifiée à k blocs.

5.4.3 Stratégie ensembliste pour l'inférence sur l'ensemble de données de test

Comme décrit précédemment sous-section 5.4.2, à une validation croisée à six blocs correspond six versions $\Theta^{j \in [1, \dots, 6]}$ des poids d'un même modèle, optimisé à chaque fois à l'aide de $\frac{5}{6}$ des données d'entraînement. Pour obtenir une prédiction unique à partir des six versions des poids d'un modèle, une stratégie ensembliste est utilisée. Étant donné X un sac d'instances caractérisant une lame de cytologie, pour chaque version des poids du modèle Θ^j , une prédiction du diagnostic $\hat{y}^j = \Theta^j(X)$ est effectuée. La prédiction finale \hat{y} est alors donnée par :

$$\hat{y} = \frac{1}{6} \sum_{j=1}^6 \hat{y}^j \quad (5.39)$$

Pour étudier la variabilité des performances obtenues par les modèles en fonction de leur initialisation aléatoire et du caractère stochastique de l'apprentissage par descente de gradients, l'entraînement d'un modèle est répété M fois. On note alors Θ_m^j le modèle obtenu lors du m -ème entraînement pour le j -ème bloc de validation croisée ³⁴, et \hat{y}_m^j la prédiction associée. Pour simplification, on note :

34: avec $m \in [1, \dots, M]$

$$\hat{y}_m = \frac{1}{6} \sum_{j=1}^6 \hat{y}_m^j = \frac{1}{6} \sum_{j=1}^6 \Theta_m^j(X) = \Theta_m(X) \quad (5.40)$$

5.4.4 Performances de diagnostic

Les performances de diagnostic sont évaluées pour les différentes méthodes d'apprentissage à partir d'instances multiples, présentées sous-section 5.3.5, à la fois en validation croisée ³⁵ et sur un ensemble de données de test indépendant. Pour toutes les méthodes, les vecteurs de caractéristiques ont été calculés à partir des sections S_{sharp} des z-stacks numérisées.

35: Voir le détail de la validation croisée utilisée sous-section 5.4.2.

Premièrement, les méthodes de MIL sont évaluées en validation croisée à 6 blocs. Chaque modèle est entraîné cinq fois afin d'étudier la stabilité des performances de diagnostic. Les résultats obtenus sont reportés Tableau 5.2.

Une fois entraînés, les modèles Θ_i sont également évalués sur un jeu de données de Test indépendant. Comme décrit sous-section 5.4.3, la prédiction sur le jeu de données de Test est effectuée à l'aide d'une méthode ensembliste, combinant les prédictions \hat{y}^j pour $j \in [1, \dots, 6]$ correspondant aux six blocs de la validation croisée.

Pour le jeu de données de Test, 3.4% des lames de cytologie digitales contiennent moins de dix cellules urothéliales basales et n'ont donc pas été analysées. Les performances obtenues pour différents modèles Θ_m de MIL sont moyennées pour $m \in [1, \dots, 5]$ ³⁶ et reportées Tableau 5.3.

36: L'entraînement de chaque modèle est répété ici cinq fois pour étudier la variabilité des performances de diagnostic obtenues.

modèle	évaluation en validation croisée			
	exactitude	sensi.	spécif.	AUC
<i>maxpool(inst.)</i>	0.686 ± 0.006	0.673 ± 0.010	0.700 ± 0.012	0.748 ± 0.007
<i>meanpool(inst.)</i>	0.716 ± 0.004	0.638 ± 0.003	0.795 ± 0.006	0.792 ± 0.001
<i>meanpool(repr.)</i>	0.724 ± 0.001	0.697 ± 0.002	0.752 ± 0.002	0.792 ± 0.001
<i>maxpool(repr.)</i>	0.728 ± 0.005	0.719 ± 0.008	0.737 ± 0.004	0.793 ± 0.002
ABMIL	0.726 ± 0.009	0.739 ± 0.006	0.714 ± 0.014	0.786 ± 0.004
ABMIL (rés. conv.)	0.733 ± 0.010	0.727 ± 0.011	0.739 ± 0.012	0.789 ± 0.007
DSMIL	0.710 ± 0.002	0.697 ± 0.004	0.723 ± 0.004	0.777 ± 0.002
TransMIL	0.720 ± 0.008	0.733 ± 0.016	0.706 ± 0.030	0.782 ± 0.006
TransMIL (rés. conv.)	0.729 ± 0.007	0.742 ± 0.007	0.716 ± 0.017	0.790 ± 0.006

TABLE 5.2 – Performances de différents modèles de MIL en validation croisée, en utilisant des caractéristiques expertes calculées à partir S_{sharp} . Pour chaque évaluation, cinq modèles sont entraînés selon une validation croisée à six blocs.

modèle	évaluation sur les données de Test			
	exactitude	sensi.	spécif.	AUC
<i>maxpool(inst.)</i>	0.702 ± 0.008	0.666 ± 0.025	0.737 ± 0.022	0.770 ± 0.011
<i>meanpool(inst.)</i>	0.744 ± 0.000	0.655 ± 0.000	0.832 ± 0.000	0.796 ± 0.000
<i>meanpool(repr.)</i>	0.745 ± 0.003	0.695 ± 0.005	0.796 ± 0.005	0.800 ± 0.000
<i>maxpool(repr.)</i>	0.738 ± 0.002	0.712 ± 0.003	0.765 ± 0.007	0.811 ± 0.002
ABMIL	0.771 ± 0.004	0.756 ± 0.007	0.785 ± 0.010	0.824 ± 0.002
ABMIL (rés. conv.)	0.755 ± 0.012	0.723 ± 0.013	0.787 ± 0.014	0.816 ± 0.003
DSMIL	0.699 ± 0.009	0.676 ± 0.008	0.721 ± 0.014	0.771 ± 0.006
TransMIL	0.744 ± 0.006	0.734 ± 0.009	0.753 ± 0.018	0.811 ± 0.002
TransMIL (rés. conv.)	0.754 ± 0.006	0.742 ± 0.013	0.766 ± 0.017	0.812 ± 0.002

TABLE 5.3 – Performances de différents modèles de MIL sur un jeu de données de Test indépendant, en utilisant des caractéristiques expertes calculées à partir S_{sharp} . Pour chaque modèle, cinq entraînements ont été réalisés afin d'évaluer la stabilité des performances obtenues.

Les meilleures performances de diagnostic sur le jeu de données d'entraînement, évaluées par validation croisée, sont obtenues pour les modèles ABMIL et TransMIL avec couche de convolution résiduelle, ainsi que par le modèle *maxpool(repr.)*. Les meilleures performances de diagnostic sur le jeu de données de Test sont obtenues pour le modèle ABMIL, sans couche de convolution résiduelle, devant le modèle TransMIL. Ainsi les modèles ABMIL et TransMIL permettent d'obtenir les meilleures performances de diagnostic. Les métriques évaluées sur le jeu de données de Test sont systématiquement supérieures par rapport aux métriques calculées en validation croisée sur le jeu de données d'entraînement. Cela peut s'expliquer par l'utilisation d'une stratégie ensembliste des prédictions lors de l'évaluation des modèles sur les données de Test, permettant d'améliorer les performances de diagnostic. L'impact de la stratégie ensembliste sur les performances de diagnostic est discutée sous-section 5.5.1.

Intervalle d'incertitude pour le diagnostic Comme décrit précédemment ³⁷, le diagnostic \hat{y} d'un modèle dépend de la comparaison de la sortie brute du modèle $y_{pred} \in [0, 1]$ avec le seuil de décision fixé à 0.5. Dans le cas où la sortie brute du modèle y_{pred} est proche du seuil de décision 0.5, la prédiction peut être considérée comme incertaine. Pour ces cas, il est préférable d'indiquer que le modèle est incertain plutôt que de donner un diagnostic $\hat{y} \in \{0, 1\}$. Pour cela, il est nécessaire de définir un intervalle d'incertitude sur les valeurs y_{pred} , pour lesquelles la prédiction du modèle est considérée incertaine.

$$37: \hat{y} = \begin{cases} 1 & \text{si } y_{pred} \geq 0.5 \\ 0 & \text{sinon} \end{cases}$$

Plusieurs intervalles d'incertitude sont considérés. Si la sortie brute y_{pred} d'un modèle est contenue dans un intervalle d'incertitude, la lame n'est pas comptabilisée dans l'évaluation des performances et est reportée comme lame incertaine. Les performances de diagnostic ainsi que la proportion de prédictions incertaines en fonction de différents intervalles d'incertitude sont reportées Tableau 5.4.

Plus l'intervalle d'incertitude choisi est large, plus les performances de diagnostic

modèle	intervalle	évaluation sur les données de Test				
	d'incertitude	exactitude	sensi.	spécif.	AUC	lames incertaines
ABMIL	[0.5, 0.5]	0.771 ± 0.004	0.756 ± 0.007	0.785 ± 0.010	0.824 ± 0.002	0.000 ± 0.007
ABMIL	[0.45, 0.55]	0.775 ± 0.003	0.758 ± 0.004	0.791 ± 0.005	0.829 ± 0.002	0.043 ± 0.007
ABMIL	[0.4, 0.6]	0.775 ± 0.004	0.761 ± 0.002	0.789 ± 0.007	0.834 ± 0.003	0.082 ± 0.006
ABMIL	[0.35, 0.65]	0.784 ± 0.004	0.770 ± 0.002	0.798 ± 0.008	0.840 ± 0.002	0.130 ± 0.007

TABLE 5.4 – Étude des intervalles d'incertitude sur les performances de diagnostic, évaluées sur l'ensemble de données de Test. En augmentant l'intervalle d'incertitude, les performances de diagnostic augmentent avec la proportion de lames reportées incertaines.

augmentent, ainsi que la proportion de lames reportées comme incertaines. Le choix de l'intervalle d'incertitude est donc un compromis entre les performances de diagnostic et la proportion de lames pour lesquelles le modèle ne se prononce pas. Ce compromis peut-être arbitré en cherchant à obtenir, en sortie du modèle, une proportion de lames incertaines similaire à une analyse manuelle par des anatomopathologistes. L'intervalle d'incertitude est ainsi fixé à [0.45, 0.55], afin de correspondre au taux d'incertitude des anatomopathologistes évalué sur l'essai clinique Visiocyt® 01 (voir sous-section 5.4.5 ci-dessous).

5.4.5 Comparaison avec les performances des anatomopathologistes

Dans la classification du Système de Paris, la classe **C2** correspond à des lames négatives pour le carcinome urothélial de haut-grade (HGUC), tandis que les classes **C4**, **C5** et **C6** correspondent respectivement à des suspicions de HGUC aux lames positives pour un carcinome urothélial de haut-grade et aux lames positives pour un carcinome urothélial de bas-grade. La classe **C3** correspond aux lames contenant quelques cellules urothéliales basales atypiques, en faible nombre ou d'atypie légère. Pour limiter le nombre de faux positifs, le Système de Paris recommande de ne pas considérer les lames **C3** comme positives. Afin d'établir un lien avec les prédictions des modèles de diagnostic décrits précédemment, les lames classées **C3** par les anatomopathologistes sont considérées comme l'équivalent des lames pour laquelle la prédiction du modèle est incertaine.

Pour chaque échantillon urinaire utilisé dans la base de données VisioCyt® 01, une lame de cytologie a été produite et analysée par des anatomopathologistes. Ces derniers ont ensuite diagnostiqué chaque lame selon la classification du Système de Paris. Les performances des anatomopathologistes pour diagnostiquer le cancer de la vessie à partir de lames de cytologie sont évaluées, en considérant comme vérité terrain les résultats des examens de cystoscopie et d'histologie, le cas échéant³⁸.

38: Voir Tableau 5.5.

Analyse	incertitude	exactitude	sensi.	spéci.	lames incertaines	lames C1
Experts	classe C3	0.758	0.515	1.0	0.064	0.048
ABMIL	$y_{pred} \in [0.45, 0.55]$	0.775 ± 0.003	0.758 ± 0.004	0.791 ± 0.005	0.043 ± 0.007	0.034
TransMIL	$y_{pred} \in [0.45, 0.55]$	0.761 ± 0.006	0.746 ± 0.011	0.776 ± 0.018	0.032 ± 0.005	0.034

TABLE 5.5 – Comparaison entre les performances de diagnostic des anatomopathologistes à partir de lames de cytologie et deux méthodes de diagnostic automatique décrites précédemment. La comparaison est effectuée pour la base de données de Test. La vérité terrain est définie à partir des résultats des examens de cystoscopie et d'histologie suivants la cytologie urinaire.

À partir de la proportion de lames classées comme **C3** par les anatomopathologistes³⁹ et des prédictions incertaines des modèles, reportées Tableau 5.4, l'intervalle d'incertitude [0.45, 0.55] est retenu. En effet avec un tel intervalle, la proportion de lames pour lesquelles la prédiction du modèle est incertaine est de 4.3%, soit un taux légèrement inférieur à la proportion de lames classées comme **C3** par les anatomopathologistes.

39: 6.4% de lames **C3** parmi l'ensemble des lames de Test.

Cette comparaison des performances de diagnostic des anatomopathologistes est donnée à titre indicatif, mais comporte deux biais :

- Dans l'essai clinique VisioCyt® 01, dont la base de données utilisée dans ce chapitre est extraite, un patient est inclus dans la catégorie *négative* si à la fois les examens de cytologie et de cystoscopie sont négatifs. La base de données ne contient donc pas de patient dont l'examen de cytologie positif est ensuite déjugé par l'examen de cystoscopie. Par construction, il n'y a donc pas de faux positifs dans les analyses des anatomopathologistes, résultant en une spécificité sur-estimée et fixée à 1.0.
- Le problème considéré est le diagnostic de carcinomes urothéliaux de haut-grade et carcinomes urothéliaux de bas-grade. Les anatomopathologistes qui ont analysé les lames de cytologie se sont basés sur la classification du Système de Paris, créé pour uniformiser le diagnostic des HGUCs. En suivant ces recommandations, les experts ont donc logiquement une sensibilité plus faible pour la détection des LGUCs.

Malgré ces différents biais, les performances obtenues à l'aide des méthodes de diagnostic automatique ABMIL et TransMIL sont compétitives, avec une exactitude légèrement supérieure à l'exactitude obtenue par les experts, une sensibilité largement supérieure et une spécificité largement inférieure.

Par l'analyse objective de toutes les cellules présentes dans une lame, le calcul sans biais des caractéristiques expertes pour chaque cellule d'intérêt et la mise en relation, à l'aide de méthodes de MIL, des caractéristiques expertes des cellules urothéliales basales, plusieurs méthodes sont trouvées capable de détecter efficacement des carcinomes urothéliaux de haut grade et de bas grade à partir de lames de cytologie digitales.

5.4.6 Visualisation des prédictions

Pour que des modèles de diagnostic automatiques soient utilisés par les praticiens, il est nécessaire qu'à la fois médecins et patients puissent avoir confiance en ces modèles. La justification des décisions prise par un modèle est une condition à cette confiance [170].

Au sein de l'Union Européenne, ce besoin d'explication est devenu obligatoire avec le règlement européen : Règlement Général sur la Protection des Données (RGPD), qui explicite dans sa raison 71⁴⁰ au sujet des traitements automatisés :

"En tout état de cause, un traitement de ce type devrait être assorti de garanties appropriées, qui devraient comprendre une information spécifique de la personne concernée ainsi que le droit d'obtenir une intervention humaine, d'exprimer son point de vue, d'obtenir une explication quant à la décision prise à l'issue de ce type d'évaluation et de contester la décision."

Pour le modèle ABMIL, la fonction d'agrégation d'instances σ utilisée est un opérateur de moyenne pondérée⁴¹, où chaque coefficient a_i , associé à l'instance i , est un score d'importance ou d'*attention* de l'instance pour la classification de la lame. À l'inférence, il est donc possible de visualiser les k instances qui ont le plus d'importance dans la représentation d'une lame. Sans expliquer complètement le processus de décision, une telle visualisation permet néanmoins de mettre en évidence les instances jugées les plus importantes pour le diagnostic.

Les scores d'attention, moyennés par instance pour toutes les lames positives et négatives de l'ensemble de données de Test, sont reportés Figure 5.14.

Les instances étant ordonnées par RNC croissant, on peut remarquer qu'en moyenne les scores d'attention plus hauts sont attribués aux cellules avec les RNC les plus élevés. Un RNC élevé étant le premier critère d'atypie d'après la classification du Système de

[170]: DOSILOVIC et al. (2018), « Explainable artificial intelligence »

40: <https://www.privacy-regulation.eu/fr/r71.htm> (consulté le 27/09/2022)

41: Rappel : $\sigma(H) = \sum_{i=1}^N a_i h_i$

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

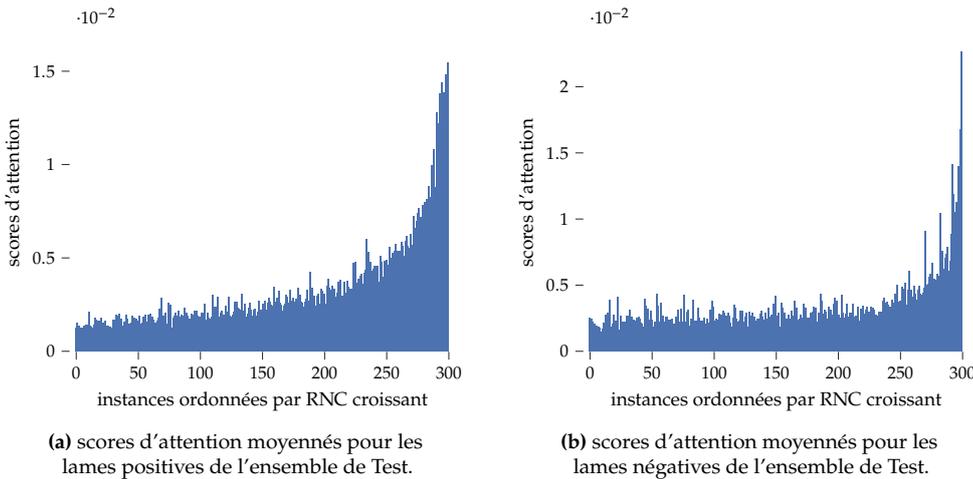


FIGURE 5.14 – Moyenne des scores d'attention pour les instances des lames positives (a) et négatives (b) de l'ensemble de données de Test. Les instances, en abscisse, sont ordonnées par RNC croissant. En moyenne, une plus forte importance est accordée aux cellules avec un RNC élevé.

Paris [24], on peut en déduire que la méthode ABMIL a bien tendance à sélectionner les cellules les plus atypiques pour représenter une lame de cytologie.

Pour une lame de l'ensemble de données de Test, provenant d'un patient sain, les scores d'attention par instance sont reportés Figure 5.15. Pour cette même lame, les dix cellules avec les plus hauts scores d'attention, donc les dix cellules qui ont le plus d'importance dans la représentation de la lame, sont illustrées Figure 5.16.

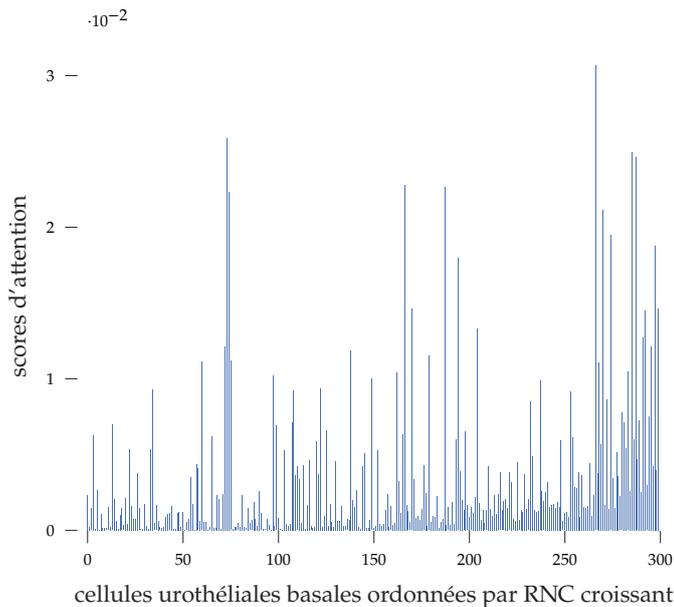


FIGURE 5.15 – Scores d'attention pour les cellules d'une lame de patient sain. Les cellules urothéliales basales, en abscisse, sont ordonnées par RNC croissant. Un plus haut score d'attention signifie que plus d'importance est donnée à la cellule dans la représentation de la lame.

De même, pour une lame de l'ensemble de données de Test, provenant d'un patient atteint par un carcinome urothélial, les scores par instance sont reportés Figure 5.17. Les dix cellules qui ont le plus contribué à la représentation de la lame sont également illustrées Figure 5.18.

Pour les exemples de lames proposés, les instances cellulaires avec les plus hauts scores d'attention ont tendance à être également les cellules avec les plus hauts RNC. Cette observation est valable pour la lame provenant d'un patient sain et pour la lame provenant d'un patient atteint par un carcinome urothélial. Il est remarquable que, dans chaque cas, certaines instances avec des RNC plus faibles ont également des scores d'attention élevés. Ainsi d'autres caractéristiques expertes, comme l'hyper-chromatisme

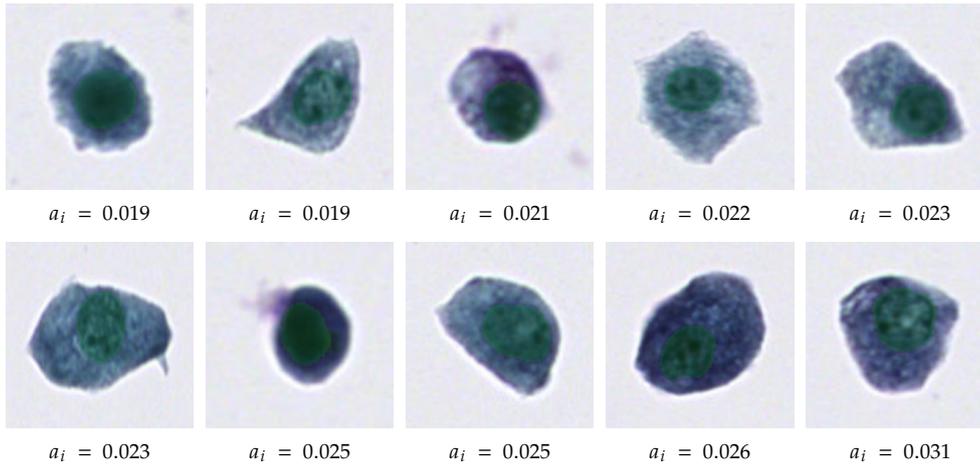


FIGURE 5.16 – Illustration des dix cellules avec les plus hauts scores d’attention dans une lame de patient sain. Chaque cellule est reportée avec son score d’attention a_i . Les masques de noyaux prédits par instance sont représentés en vert pour information.

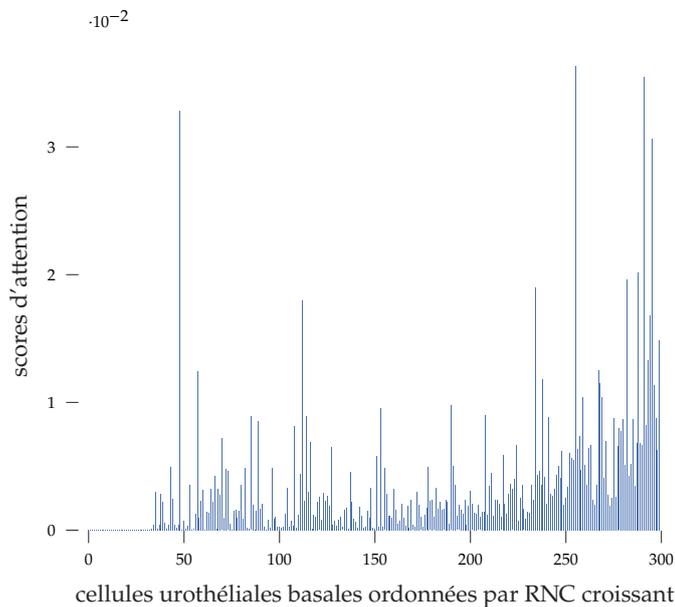


FIGURE 5.17 – Scores d’attention pour les cellules d’une lame de patient atteint d’un carcinome urothélial. Les cellules urothéliales basales sont ordonnées par RNC croissant. Un plus haut score d’attention signifie que plus d’importance est donnée à la cellule dans la représentation de la lame.

ou la texture du noyau, peuvent être suffisantes au modèle ABMIL pour attribuer un score d’attention élevé à une cellule.

La visualisation des cellules avec les plus hauts scores d’attention, en plus de la prédiction du modèle, est ainsi une première justification des décisions prises par ce dernier.

5.4.7 Performances en fonction des examens de cystoscopie et d’histologie

Dans cette sous-section, les performances de diagnostic des méthodes ABMIL et TransMIL sont évaluées en fonction des résultats d’examens de cystoscopie et d’histologie. Les lames de l’ensemble de données de Test sont réparties en trois sous-groupes :

- Négatif : lames pour lesquelles l’examen de cystoscopie est retourné négatif. Pas d’examen d’histologie n’a été effectué dans ce cas.
- Positif LGUC : lames pour lesquelles l’examen de cystoscopie est retourné positif, et l’examen d’histologie suivant est retourné positif pour un carcinome urothélial de bas-grade (LGUC).

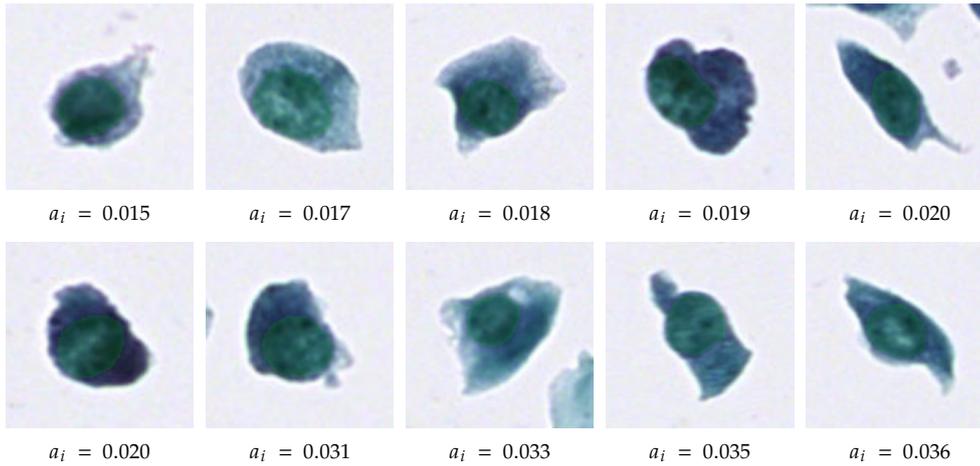


FIGURE 5.18 – Illustration des dix cellules avec les plus hauts scores d’attention dans une lame de patient atteint d’un carcinome urothélial. Chaque cellule est reportée avec son score d’attention a_i . Les masques de noyaux prédits par instance sont représentés en vert pour information.

— Positif HGUC : lames pour lesquelles l’examen de cystoscopie est retourné positif, et l’examen d’histologie suivant est retourné positif pour un carcinome urothélial de haut-grade (HGUC).

Les prédictions de la méthode ABMIL pour chacun de ces sous-groupes, moyennées pour cinq modèles, sont reportées Tableau 5.6. Les lames paucicellulaires ne sont pas analysées par manque d’information, et sont considérées comme des lames C1 dans la classification du Système de Paris. Cela concerne quatre lames du sous-groupe Négatif, deux lames du sous-groupe LGUC et quatre lames du sous-groupe HGUC.

sous-groupe	prédictions de la méthode ABMIL	
	négatif < 0.5	positif \geq 0.5
Négatif	107.6 \pm 1.4 (78.5%)	29.4 \pm 1.4 (21.5%)
LGUC	25.6 \pm 1.0 (34.6%)	48.4 \pm 1.0 (65.4%)
HGUC	9.8 \pm 0.4 (13.8%)	61.2 \pm 0.4 (86.2%)

TABLE 5.6 – Prédications de la méthode ABMIL, sur l’ensemble de données de test, en fonction du sous-groupe de la lame défini par le résultat des examens de cystoscopie et éventuellement d’histologie. Les décomptes des prédictions sont moyennés pour cinq modèles et reportés avec leur écart-type.

Similairement, les prédictions de la méthode TransMIL pour chacun de ces sous-groupes, moyennées pour cinq modèles, sont reportées Tableau 5.7. Ici également, les lames paucicellulaires ne sont pas analysées par manque d’information.

sous-groupe	prédictions de la méthode TransMIL	
	négatif < 0.5	positif \geq 0.5
Négatif	103.2 \pm 2.4 (75.3%)	33.8 \pm 2.4 (24.7%)
LGUC	27.8 \pm 1.2 (37.6%)	46.2 \pm 1.2 (62.4%)
HGUC	10.8 \pm 0.4 (15.2%)	60.2 \pm 0.4 (84.8%)

TABLE 5.7 – Prédications de la méthode TransMIL, sur l’ensemble de données de test, en fonction du sous-groupe de la lame défini par le résultat des examens de cystoscopie et éventuellement d’histologie. Les décomptes des prédictions sont moyennés pour cinq modèles et reportés avec leur écart-type.

Pour les deux méthodes ABMIL et TransMIL, de hautes performances de diagnostic sont obtenues pour le sous-groupe Négatif, avec respectivement 78.5% et 75.3% de lames négatives correctement prédites. De même, de hautes performances sont obtenues pour les deux méthodes sur le sous-groupe HGUC, avec respectivement 86.2% et 84.8% des lames correctement identifiées comme positives pour un carcinome urothélial.

Des performances plus mitigées sont obtenues sur le sous-groupe LGUC avec respectivement 65.4% et 62.4% des lames correctement prédites. Ces résultats ne sont pas surprenants, car les carcinomes urothéliaux de bas-grade sont connus pour être difficilement identifiables à partir de lames de cytologie [24]. De plus, dans la méthode

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

de diagnostic proposée, seules les cellules urothéliales basales isolées sont analysées pour effectuer le diagnostic. Le Système de Paris reporte que la présence d'amas fibrovasculaire dans les lames de cytologie est un marqueur de carcinome urothélial de bas-grade, cependant ces structures ne sont pas prises en compte dans la méthode présentée dans ce chapitre, expliquant certainement les performances mitigées obtenues sur ce sous-groupe de lames.

5.5 Discussion

5.5.1 Impact de l'apprentissage ensembliste sur les performances de classification

Pour évaluer l'impact de l'apprentissage ensembliste, au lieu de moyenniser les prédictions $y_{pred}^j \mid j \in [1, 6]$, les performances de chaque version Θ^j du modèle sont évaluées indépendamment sur l'ensemble de données de Test. Ces métriques de performances sont ensuite moyennées pour les 6 blocs de validation croisée. Pour exemple, le calcul de l'exactitude avec la stratégie ensembliste dans le cadre d'un apprentissage par validation croisée à 6 blocs est donné Équation 5.41 et le calcul de l'exactitude sans stratégie ensembliste est donné Équation 5.42.

$$\text{Exactitude} \left(y_{true}, \frac{1}{6} \sum_{j=1}^6 y_{pred}^j \right) \quad (5.41)$$

$$\frac{1}{6} \sum_{j=1}^6 \text{Exactitude} \left(y_{true}, y_{pred}^j \right) \quad (5.42)$$

Les performances obtenues pour le modèle ABMIL, avec et sans stratégie ensembliste, sont reportées Tableau 5.8 pour l'ensemble de données de Test.

modèle	ensemble	évaluation sur les données de Test			
		exactitude	sensi.	spécif.	AUC
ABMIL	✓	0.771 ± 0.004	0.756 ± 0.007	0.785 ± 0.010	0.824 ± 0.002
ABMIL		0.756 ± 0.004	0.751 ± 0.005	0.762 ± 0.004	0.815 ± 0.002

TABLE 5.8 – Impact de l'apprentissage ensembliste sur les performances du modèle ABMIL pour l'ensemble de données de Test.

Ainsi un gain de performances est systématiquement observé pour toutes les métriques évaluées en utilisant l'apprentissage ensembliste. Pour plus de détails, l'exactitude des modèles sur l'ensemble de données de Test, par bloc de validation croisée j et pour chacune des 5 réplifications de l'expérience sont reportées Figure 5.19.

Il est à noter que l'exactitude avec apprentissage ensembliste est supérieure à l'exactitude moyennée des 6 modèles obtenus par validation croisée pour une expérience ⁴².

Une fois les hyper-paramètres du modèle ABMIL fixés à l'aide de la validation croisée, il est possible de ré-entraîner le modèle sur les ensembles de données Train et Dev combinés ⁴³ puis d'évaluer les performances du modèle obtenu dans l'ensemble de données de Test. Une expérience est donc menée afin de comparer les performances obtenues sur l'ensemble Test en utilisant un entraînement par validation croisée suivi d'une stratégie ensembliste et un entraînement sur l'ensemble de données *Train + Dev*. Les résultats de cette expérience sont reportés Tableau 5.9.

42: Voir comparaison entre les deux dernières colonnes de la Figure 5.19.

43: Noté alors *Train + Dev*.

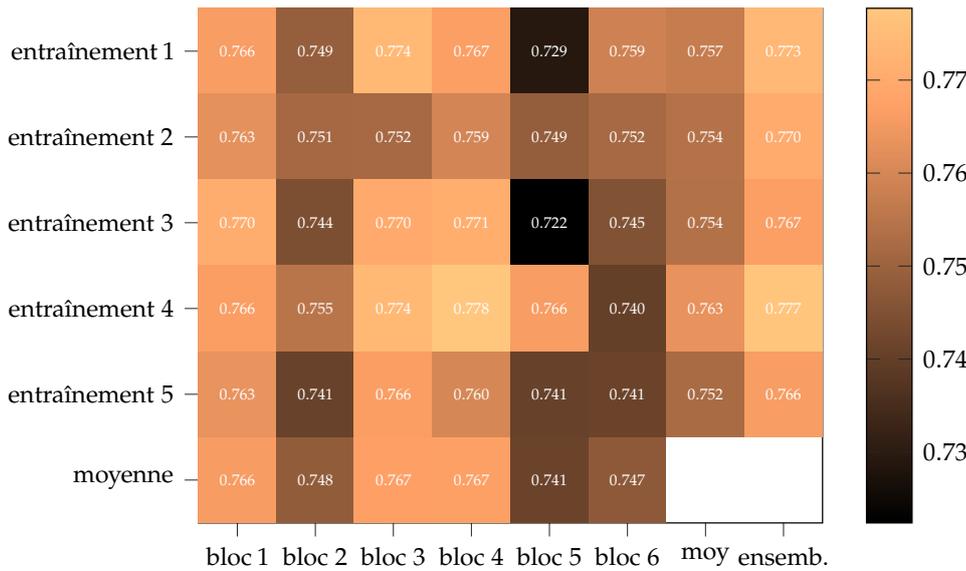


FIGURE 5.19 – Exactitude des modèles sur l’ensemble de données de Test, sans stratégie ensembliste, pour chaque bloc de validation croisée j et chaque réplication de l’expérience. Les moyennes par bloc de validation croisée et par réplication de l’expérience (entraînement m) sont également reportées. Pour comparaison, les performances obtenues par apprentissage ensembliste pour chaque réplication de l’expérience sont également reportées colonne de droite.

modèle	entraînement	ensemble	évaluation sur les données de Test			
			exactitude	sensi.	spécif.	AUC
ABMIL	validation croisée	✓	0.771 ± 0.004	0.756 ± 0.007	0.785 ± 0.010	0.824 ± 0.002
ABMIL	Train + Dev		0.759 ± 0.004	0.759 ± 0.014	0.759 ± 0.015	0.820 ± 0.008
ABMIL	Train + Dev	✓	0.759	0.759	0.759	0.823

TABLE 5.9 – Performances du modèle ABMIL sur l’ensemble de données de Test en fonction de la stratégie d’entraînement. Le modèle est entraîné soit par validation croisée suivie d’une stratégie d’ensemble, soit directement sur l’ensemble de données *Train + Dev*. Chaque modèle est entraîné cinq fois et les performances obtenues sont moyennées.

Les meilleures performances sont obtenues par apprentissage en validation croisée, suivie de la stratégie ensembliste décrite sous-section 5.4.3.

Pour information, les prédictions des cinq répliques du modèle ABMIL entraînées sur *Train + Dev* sont moyennées⁴⁴. Les performances alors obtenues sont également inférieures aux performances du modèle ABMIL entraîné en validation croisée. Il semble donc que le gain apporté par l’entraînement en validation croisée provienne de la combinaison de prédictions de modèles entraînés sur des ensembles de données légèrement différents⁴⁵, plutôt que de la régularisation apportée par la moyenne des prédictions de modèles entraînés sur un même ensemble de données.

44: Voir la dernière ligne Tableau 5.9.

45: Approche similaire à la sélection aléatoire des échantillons utilisée dans les forêts d’arbres de décision aléatoire [171].

5.5.2 Nombre minimum de cellules urothéliales basales détectées pour analyser une lame

Il est possible qu’une lame de cytologie contienne peu ou pas de cellules d’intérêt, dans ce cas elle est dite paucicellulaire. Dans le cas où aucune cellule urothéliale basale n’est présente, il n’est évidemment pas possible d’effectuer un diagnostic : la lame est considérée de catégorie C1⁴⁶ dans la classification du Système de Paris. Il reste à déterminer le nombre minimum de cellules urothéliales basales présentes dans une lame de cytologie pour pouvoir effectuer un diagnostic précis. Pour cela, les métriques exactitude, sensibilité, spécificité et aire sous la courbe ROC (AUC) sont calculées sur l’ensemble des lames de cytologie valides, déterminées par des seuils successifs sur le nombre minimum de cellules d’intérêt détectées. En fonction des seuils, la proportion de lames de cytologie rejetées C1 est également calculée. L’ensemble des résultats obtenus en validation croisée pour la méthode ABMIL est reporté Tableau 5.10

46: Rappel : une lame de catégorie C1 est une lame de cytologie pour laquelle le diagnostic ne peut être posé.

min n instances	évaluation en validation croisée				
	exactitude	sensi.	spécif.	AUC	prop. lames C1
0	0.722 ± 0.004	0.732 ± 0.009	0.711 ± 0.008	0.786 ± 0.004	0.037
10	0.726 ± 0.009	0.739 ± 0.006	0.714 ± 0.014	0.786 ± 0.004	0.045
20	0.726 ± 0.009	0.742 ± 0.016	0.710 ± 0.005	0.787 ± 0.004	0.055
30	0.729 ± 0.009	0.740 ± 0.013	0.717 ± 0.013	0.790 ± 0.003	0.067
40	0.731 ± 0.010	0.745 ± 0.013	0.718 ± 0.015	0.790 ± 0.007	0.083
50	0.714 ± 0.013	0.734 ± 0.013	0.693 ± 0.013	0.784 ± 0.005	0.105
75	0.729 ± 0.007	0.754 ± 0.015	0.704 ± 0.015	0.795 ± 0.003	0.146
100	0.707 ± 0.003	0.735 ± 0.010	0.679 ± 0.012	0.777 ± 0.003	0.187

TABLE 5.10 – Impact du nombre minimum de cellules urothéliales basales nécessaire pour considérer une lame valide sur les performances du modèle ABMIL. Les caractéristiques expertes utilisées sont calculées à partir S_{sharp} . Les performances reportées ont été moyennées pour cinq répétitions de l'expérience.

En augmentant le nombre minimum de cellules urothéliales basales nécessaires pour considérer une lame de cytologie valide, les performances ont tendance à augmenter ainsi que la proportion de lames classées C1. Il est donc nécessaire de trouver un compromis entre les performances et la proportion de lames rejetées. Les échantillons considérés ont été analysés par des anatomopathologistes qui ont classés environ 5% des lames comme C1. Bien que d'autres raisons que le manque de cellules d'intérêt peuvent être à l'origine d'une classification C1, comme un défaut de numérisation, il est donc raisonnable de rejeter environ 5% des lames de cytologie. En sélectionnant un seuil de classification C1 à dix cellules urothéliales basales, la proportion de lames classées C1 est de 4.5% : c'est le seuil choisi pour le reste des expériences.

5.5.3 Impact du nombre d'instances sélectionnées pour représenter une lame

Une lame de cytologie digitale est représentée par un sac contenant $N = 300$ instances⁴⁷. L'impact du choix de N sur les performances de l'approche est discutée dans cette sous-section. La méthode ABMIL est entraînée en utilisant des sacs de $N \in \{100, 200, 300, 400, 500\}$ instances. Les instances sélectionnées sont les N cellules urothéliales basales avec les plus hauts RNC. Dans le cas où une lame de cytologie digitale contient moins de N cellules urothéliales basales, le sac est complété par des fausses instances, dont le vecteur de caractéristiques est le vecteur nul. L'impact du nombre d'instances N sur les performances de la méthode ABMIL, calculées en validation croisée, est reporté Figure 5.20.

47: Dans ce cas, une instance est une cellule urothéliale basale.

Les métriques d'exactitude et d'AUC du diagnostic obtenu avec la méthode ABMIL sont relativement stables en fonction du nombre d'instances sélectionnées N . Un léger maximum local de l'exactitude est atteint pour $N = 300$ et un léger maximum local de l'AUC est obtenu pour $N = 400$. La valeur $N = 300$ est sélectionnée pour le reste des expériences.

La sélection des N cellules urothéliales basales avec les plus haut RNC est comparée avec une sélection aléatoire sans remise de N instances au sein de la lame. Ici encore, dans le cas où la lame de cytologie contient moins de N instances, le sac est complété par autant de vecteurs nuls que nécessaire. L'impact de la méthode de sélection des instances est évalué sur les performances de diagnostic de la méthode ABMIL, pour $N = 300$, calculées en validation croisée. Les résultats obtenus sont résumés Tableau 5.11.

sélection d'instances	évaluation en validation croisée			
	exactitude	sensi.	spécif.	AUC
aléatoire	0.702 ± 0.008	0.709 ± 0.004	0.695 ± 0.014	0.761 ± 0.007
top-k	0.726 ± 0.009	0.739 ± 0.006	0.714 ± 0.014	0.786 ± 0.004

TABLE 5.11 – Performances en *cross-validation* en fonction de la méthode de sélection d'instances. Les caractéristiques expertes utilisées sont calculées à partir S_{sharp} . Pour chaque évaluation, cinq modèles sont entraînés selon une validation croisée à 5 blocs.

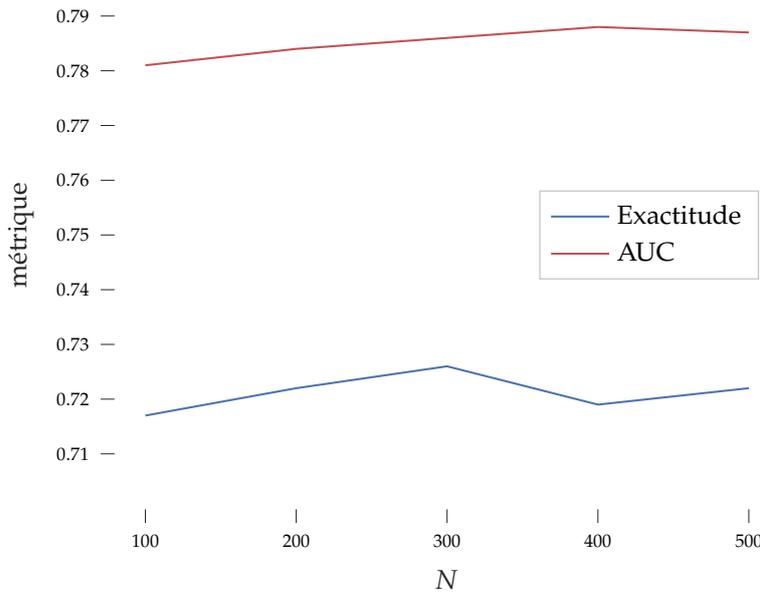


FIGURE 5.20 – Impact du nombre d’instances N sélectionnées pour représenter une lame sur l’exactitude et l’AUC de la méthode ABMIL. Les métriques sont calculées en validation croisée et moyennées pour 5 répétitions de l’expérience.

Ainsi sélectionner les N instances avec le plus haut RNC permet d’obtenir des meilleures performances de classification qu’une sélection aléatoire de N instances dans la lame. Le gain observé en sensibilité et spécificité indique que cette sélection permet d’améliorer la détection des carcinomes urothéliaux tout en diminuant le nombre de faux positifs.

5.5.4 Choix de design pour les méthodes d’apprentissage à partir d’instances multiples

Choix de design pour la méthode TransMIL Dans ce travail, plusieurs modifications ont été appliquées pour adapter la méthode TransMIL proposée par Shao et al. [165]. L’impact de ces modifications, discutées dans la présentation de la méthode TransMIL, sur les performances de diagnostic sont étudiées ici. Notamment l’influence de l’approximation de la matrice QK^T par la méthode [168]⁴⁸ et du module PPEG est évaluée et les résultats obtenus sont reportés Tableau 5.12.

méthode	approx Nyström	PPEG	évaluation en validation croisée			
			exactitude	sensi.	spécif.	AUC
TransMIL	✓		0.731 ± 0.007	0.750 ± 0.013	0.713 ± 0.020	0.790 ± 0.007
TransMIL [165]	✓	✓	0.726 ± 0.005	0.714 ± 0.015	0.739 ± 0.010	0.789 ± 0.004
TransMIL †			0.729 ± 0.007	0.742 ± 0.007	0.716 ± 0.017	0.790 ± 0.006

Contrairement aux résultats obtenus par Shao et al. [165] pour les lames d’histologie, le module PPEG n’apporte pas de gain de performances pour le diagnostic de lames de cytologie : au contraire une baisse de performances est observée. Le module PPEG a été introduit par les auteurs pour apprendre des corrélations spatiales entre des sous-images extraites de lames d’histologie afin de tenir compte de la structure spatiale du tissu. Dans le cas de lames de cytologie, il n’y a pas de relation spatiale entre les instances, ce qui peut expliquer l’absence de gain de performances à l’aide du module PPEG.

Des performances similaires sont obtenues avec et sans approximation de la matrice QK^T par la méthode de Nyström. La méthode d’approximation décrite dans [168] utilise une couche de convolution résiduelle appliquée aux vecteurs v_i pour accélérer la convergence. Pour une comparaison juste, la couche de convolution résiduelle est également appliquée

[165]: SHAO et al. (2021), « TransMIL »

[168]: XIONG et al. (2021), « Nyström-former »

48: Approximation de Nyström.

TABLE 5.12 – Étude d’ablation pour le modèle TransMIL, évaluée en validation croisée. † représente la méthode utilisée dans ce travail, sans module PPEG et sans approximation de la matrice QK^T .

[165]: SHAO et al. (2021), « TransMIL »

[168]: XIONG et al. (2021), « Nyström-former »

sans approximation de la matrice QK^T (méthode † dans le Tableau 5.12). L'impact de cette couche de convolution résiduelle est discutée dans le paragraphe suivant.

Discussions autour de l'ajout d'une couche de convolution résiduelle permettant de mixer l'information entre instances proches Dans la méthode d'approximation de la matrice QK^T proposée par [168] une couche résiduelle est appliquée aux vecteurs v_i . Une couche de *self-attention* classique, avec une couche résiduelle équivalente, transforme alors $H = \{h_0, h_1, \dots, h_N\}$ en $H' = \{h'_0, h'_1, \dots, h'_N\}$:

$$H' = H + \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V + W_\theta V \quad (5.43)$$

Avec $W_\theta V$ le produit de convolution 2D de la matrice V par un noyau paramétré par W_θ . Le noyau de convolution utilisé est de dimensions $(\theta, 1)$ et mixe donc l'information entre les instances proches dans le sac.

Parce que cette couche de convolution résiduelle mixe l'information entre instances proches, cette opération enfreint l'hypothèse d'invariance par permutation d'instances du MIL. En utilisant cette couche de convolution résiduelle, il est donc possible d'obtenir des résultats différents en fonction de l'ordre des instances dans le sac représentant la lame de cytologie.

L'impact de la couche de convolution résiduelle sur les performances de diagnostic est évaluée en validation croisée pour la méthode TransMIL avec *self-attention* sans approximation et $\theta = 33$. Les résultats obtenus sont reportés Tableau 5.13.

méthode	évaluation en validation croisée				AUC
	résiduelle	exactitude	sensi.	spécif.	
TransMIL		0.720 ± 0.008	0.733 ± 0.016	0.706 ± 0.030	0.782 ± 0.006
TransMIL	✓	0.729 ± 0.007	0.742 ± 0.007	0.716 ± 0.017	0.790 ± 0.006

[168]: XIONG et al. (2021), « Nyström-former »

TABLE 5.13 – Étude de l'impact de la couche résiduelle. Les modèles utilisent une couche de *self-attention* sans-approximation.

L'utilisation de la couche de convolution résiduelle permet ainsi d'atteindre de meilleures performances de diagnostic. Deux hypothèses peuvent expliquer ce gain de performances :

Hypothèse 1 : Lors de la sélection des instances pour former le sac représentant la lame de cytologie, les cellules sont ordonnées par RNC croissant. Ainsi ordonnées, les instances dans le sac ont donc tendance à avoir le même degré d'atypie que leurs instances voisines. La convolution résiduelle permet alors d'extraire de l'information pertinente à partir d'une instance et de ses instances voisines.

Hypothèse 2 : Lorsque une lame de cytologie contient moins de $N = 300$ instances, des vecteurs nuls sont ajoutés au sac. L'opération de *self-attention* permet de ne pas tenir compte de ces fausses instances en imposant : $\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)(i, :) = 0$. En revanche, la convolution résiduelle considère ces potentiels vecteurs nuls, et peut en déduire indirectement le nombre de vraies instances dans une lame afin d'utiliser cette information pour le diagnostic.

Si l'hypothèse 1 est vérifiée, alors la couche de convolution résiduelle combinée à la stratégie de sélection d'instances est pertinente pour le diagnostic du cancer de la vessie. Dans le cas où l'hypothèse 2 est vérifiée, le gain de performances peut provenir d'un biais naturel du jeu de données : les lames de patients atteints d'un carcinome urothélial ont tendance à avoir plus de cellules urothéliales basales atypiques que les lames de patients sains [172].

[172]: BRIMO et al. (2015), « Urine cytology »

Pour tester ces hypothèses deux expériences sont proposées. Premièrement, le modèle est appliqué à des sacs dont les instances ne sont plus ordonnées par RNC mais permutées aléatoirement. Dans ce cas, une instance donnée n'a à priori plus de relation avec ses instances voisines ce qui devrait, dans le cas de l'hypothèse 1, conduire à une baisse de performances. Deuxièmement, dans le cas des lames contenant moins de 300 cellules urothéliales basales, les sacs ne sont plus complétés par des vecteurs nuls mais par des instances de la lame re-tirées aléatoirement avec remise. Ainsi, en l'absence de vecteurs nuls dans les sacs, la couche de convolution résiduelle ne peut plus discriminer les lames en fonction de leur nombre de cellules d'intérêt ce qui devrait, dans le cas de l'hypothèse 2, conduire à une baisse de performances. Les résultats obtenus pour ces deux expériences sont reportés Tableau 5.14

méthode	permutation	re-tirage	évaluation en validation croisée			
			exactitude	sensi.	spécif.	AUC
TransMIL			0.729 ± 0.007	0.742 ± 0.007	0.716 ± 0.017	0.790 ± 0.006
TransMIL	✓		0.718 ± 0.006	0.739 ± 0.020	0.697 ± 0.029	0.781 ± 0.005
TransMIL		✓	0.731 ± 0.003	0.754 ± 0.014	0.708 ± 0.020	0.787 ± 0.005

TABLE 5.14 – Discussions autour de l'impact de la couche résiduelle. Pour identifier l'origine du gain de performances permet par la couche de convolution résiduelle.

Les performances de diagnostic baissent lorsque les instances sont permutées, et au contraire restent stable lorsque les vecteurs nuls sont remplacés par des vecteurs de caractéristiques re-tirés aléatoirement avec remise. Ces résultats laissent ainsi penser que l'hypothèse 1 est correcte : en présence d'instances ordonnées, la couche de convolution résiduelle permet un gain de performances pour le diagnostic du cancer de la vessie par extraction d'informations pertinentes à partir d'une instance et des instances voisines dans le sac.

Cette couche de convolution résiduelle peut être également appliquée à d'autres méthodes que la méthode TransMIL, et devrait également permettre une amélioration des performances de diagnostic. Ainsi, une couche de convolution résiduelle est ajoutée au modèle ABMIL après l'extracteur de caractéristiques, transformant le tenseur H de la façon suivante :

$$H_{:,j} = H_{:,j} + H_{:,j} * k_{\theta} \quad | \quad H_{:,j} = \{h_1(j), \dots, h_N(j)\} \quad (5.44)$$

Avec k_{θ} le noyau de convolution dont les poids θ sont appris lors de l'entraînement. Les performances obtenues sont reportées Tableau 5.15.

méthode	résiduelle	n epochs	évaluation en validation croisée			
			exactitude	sensi.	spécif.	AUC
ABMIL		1000	0.726 ± 0.009	0.739 ± 0.006	0.714 ± 0.014	0.786 ± 0.004
ABMIL	✓	1000	0.660 ± 0.016	0.796 ± 0.012	0.523 ± 0.026	0.717 ± 0.020
ABMIL	✓	200	0.733 ± 0.010	0.727 ± 0.011	0.739 ± 0.012	0.789 ± 0.007

TABLE 5.15 – Étude de l'importance d'une couche de convolution résiduelle appliquée au modèle ABMIL.

En ajoutant simplement une couche de convolution résiduelle à la méthode ABMIL comme décrit précédemment, une importante baisse de performances est observée. En effet, l'opération supplémentaire conduit à une convergence de modèle beaucoup plus rapide, et au bout de 1000 epochs, le modèle sur-apprend la base de données d'entraînement et ne généralise plus correctement, comme illustré Figure 5.21.

Lorsque le nombre d'epochs est diminué à 200, comme fixé pour l'entraînement du modèle TransMIL, le sur-apprentissage est évité et les performances obtenues en validation croisée sont supérieure avec utilisation d'une couche de convolution résiduelle.

Dans [165], Shao et al. montrent expérimentalement que le modèle TransMIL converge plus rapidement que d'autres méthodes de MIL, ABMIL compris. Cette expérience

[165]: SHAO et al. (2021), « TransMIL »

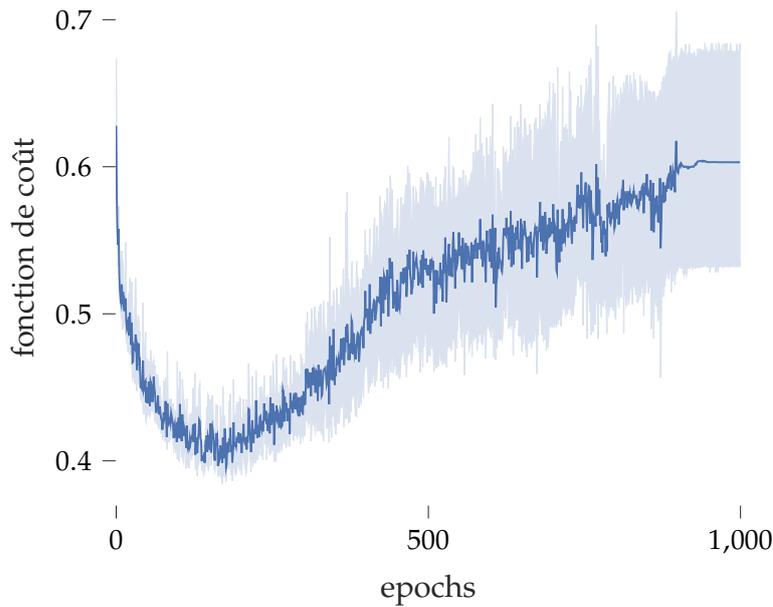


FIGURE 5.21 – Fonction de coût calculée sur l'ensemble de données *Dev* pour la méthode ABMIL avec convolution résiduelle entraînée pendant 1000 epochs.

permet également de mettre en évidence que la couche de convolution résiduelle semble responsable de cette convergence plus rapide.

5.6 Conclusion

Ce chapitre décrit une méthode de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales. La méthode proposée repose sur une détection des cellules d'intérêt dans la lame de cytologie digitale, puis un calcul de caractéristiques expertes inspirées du Système de Paris.

Pour chaque lame, les 300 cellules urothéliales basales avec le plus haut RNC, marqueur d'atypie le plus important, sont sélectionnées. Cette règle simple de sélection d'instances est trouvée beaucoup plus efficace qu'une sélection aléatoire, et permet de limiter le nombre de cellules saines dans les sacs représentant les lames *malades* tout en donnant de l'importance aux cellules légèrement atypiques qui se trouvent naturellement dans les lames *saines*.

À partir des sacs de représentations extraits par lames, plusieurs méthodes d'apprentissage à partir d'instances multiples ont été entraînées en validation croisée, puis évaluées à l'aide d'une stratégie ensembliste sur un ensemble de données de Test. Cette stratégie ensembliste est assimilable aux techniques de sélection aléatoire des échantillons⁴⁹ utilisée dans les forêts d'arbres aléatoires. Cela conduit également à un gain de performances par rapport à un entraînement sur l'ensemble des données d'entraînement.

49: Également appelée *bagging*.

En validation croisée, les méthodes ABMIL et TransMIL, utilisées avec des couches de convolution résiduelles permettant de mixer l'information entre des cellules de RNC équivalents, sont trouvées les plus performantes. Sur l'ensemble de données de Test, avec stratégie ensembliste, le modèle ABMIL permet d'obtenir le meilleur diagnostic. Ainsi ABMIL est l'approche de MIL retenue pour la méthode proposée de diagnostic automatique de carcinome urothéliaux.

Grâce aux scores d'attention, disponibles pour certaines méthodes de MIL dont la méthode sélectionnée ABMIL, il est possible de visualiser les instances qui ont contribué le plus au vecteur de représentation de la lame de cytologie. De cette façon, l'approche

permet à un anatomopathologiste d'examiner les cellules qui ont le plus contribué à la prédiction du modèle.

Il a ainsi été montré qu'il est possible d'obtenir de bonnes performances de diagnostic, en cherchant à prédire le résultat des examens de cystoscopie et d'histologie à partir de lames de cytologie. En détails, sur l'ensemble de données de Test, la méthode de diagnostic proposée atteint 77.5% d'exactitude, 75.8% de sensibilité, 79.1% de spécificité et 82.9% d'AUC. La majorité des erreurs sont commises pour les lames LGUCs, pour lesquelles le diagnostic est notoirement plus difficile [24]. De tels résultats sont encourageants. En effet, en aidant les anatomopathologistes à obtenir, à partir de cytologies urinaires, des performances similaires aux examens de cystoscopie et histologie, ces examens invasifs pour le patient pourraient être en partie évités. De plus, la cytologie urinaire étant un examen non-invasif, obtenir une sensibilité et spécificité élevée pour cet examen pourrait ouvrir la voie à des campagnes de dépistage en population générale.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

Impact des z-stacks sur le diagnostic automatique du cancer de la vessie

6

6.1 Introduction

Il est montré chapitre 2 et chapitre 4 que l'utilisation d'information provenant d'acquisitions multi-plans de lames de cytologie digitales permet d'améliorer les performances de modèles pour certaines tâches locales, comme la segmentation sémantique de noyaux ou la segmentation d'instances cellulaires. Ce chapitre s'intéresse alors à la question suivante : *Quel est l'impact de l'utilisation de z-stacks sur le diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales ?*

La méthode de diagnostic automatique du cancer de la vessie décrite chapitre 5 s'appuie sur des caractéristiques expertes dont le calcul dépend de tâches locales, notamment de segmentation. À partir des résultats obtenus chapitre 2 et chapitre 4, il est attendu que le calcul des caractéristiques expertes soit plus précis lorsqu'il s'appuie sur l'information provenant de z-stacks plutôt qu'en utilisant une simple acquisition 2D. Il est alors nécessaire d'évaluer si le calcul de caractéristiques expertes plus précises, permis par l'acquisition multi-plan, conduit à une amélioration des performances de diagnostic automatique du cancer de la vessie.

L'acquisition multi-plans de lames de cytologie implique une augmentation importante du temps de numérisation, ainsi que de l'espace de stockage numérique nécessaire. Le potentiel gain de performances pour le diagnostic permis par l'utilisation de z-stacks doit donc être suffisant pour compenser ces deux contraintes de l'acquisition multi-plans de lames.

6.2 État de l'art

La cytologie conventionnelle est connue pour contenir des amas cellulaires tri-dimensionnels. Afin de limiter la superposition cellulaire en cytologie, des méthodes semi-automatiques de cytologie en phase liquide (LBC) ¹ ont été développées [173], parmi lesquelles on retrouve les méthodes BD SurePath™ et Hologic ThinPrep®. La LBC consiste à mettre les cellules de l'échantillon prélevé en suspension dans un liquide, avant de les déposer en couche fine sur la lame de microscopie.

Pour les différents types de cytologies, conventionnelle et LBC, quelques études se sont intéressées à l'impact de l'acquisition multi-plans sur les performances de diagnostic en cytopathologie digitale.

Dans [174], les auteurs ont comparé les performances de diagnostic de 79 cytopathologistes et cytotechniciens sur des lames de cytologie cervicale (LBC) numérisées premièrement selon un plan focal (lames 2D) puis selon cinq plans focaux espacés de 3 µm (lames 3D). Les auteurs ont trouvé que l'utilisation de lames 3D n'apporte pas de gain de performances, cependant les cytopathologistes et cytotechniciens ont reporté préférer les lames numérisées avec plusieurs plans focaux afin d'éviter les incertitudes causées par les cellules floues.

Donnelly et al. [104] ont cherché à déterminer le nombre optimal de plans focaux à numériser pour des lames de cytologie cervicale préparées selon la méthode BD SurePath™. Cinq participants ont analysé les lames de cytologie digitales numérisées

1: *Liquid Based Cytology*.

[173]: KARNON et al. (2004), « Liquid-based cytology in cervical screening »

[174]: DEE et al. (2007), « Utility of 2-D and 3-D Virtual Microscopy in Cervical Cytology Education and Testing »

[104]: DONNELLY et al. (2013), « Optimal z-axis scanning parameters for gynecologic cytology specimens »

selon trois, cinq et sept plans focaux espacés de 1 μm , et les meilleures performances de diagnostic ont été obtenues pour les lames numérisées selon trois plans focaux.

Dans [130], Mukherjee et al. ont également cherché à déterminer le nombre optimal de plans focaux à numériser pour les lames de cytologie conventionnelle de biopsies par ponction à aiguille fine de la thyroïde. Comme pour l'étude précédente, la meilleure concordance de diagnostic entre les experts est observée pour les lames de cytologie digitales numérisées selon trois plans focaux, espacés de 1 μm .

Récemment, Kim et al. [175] ont évalué le rôle des z-stacks dans l'analyse visuelle de lames de cytologie digitales par des cytopathologistes pour le diagnostic de carcinomes urothéliaux de haut-grade (HGUCs). Bien que la mise à disposition de z-stacks améliore l'expérience utilisateur et la confiance des cytopathologistes dans l'analyse de lames de cytologie digitales, les auteurs ont reporté peu ou pas de gains de performances pour le diagnostic des HGUCs. En conséquence les auteurs suggèrent de ne pas utiliser de z-stacks pour l'analyse manuelle de lames de cytologie digitales.

Toutes les études précédentes reportent au moins un gain de confort apporté par l'acquisition multi-plans des lames de cytologie digitales, évitant la présence de cellules d'intérêt hors-focus et les incertitudes causées par celles-ci. Cependant pour moitié, ces études montrent qu'aucun gain de performances n'est apporté par l'utilisation des z-stacks, et au vue des temps de numérisation augmentés et de la quantité d'espace de stockage nécessaire, elles recommandent d'utiliser plutôt de simples acquisitions 2D. Il est à noter que l'ensemble des études décrites précédemment reposent sur l'analyse visuelle de lames de cytologie digitales par des experts. À notre connaissance, il n'existe pas encore d'étude évaluant l'impact des acquisitions multi-plans sur les performances de méthodes automatisées basées sur le traitement et l'analyse d'images. Dans [155], Sanghvi et al. proposent une méthode d'analyse automatique de lames de cytologie urinaire en se basant sur de simples acquisitions 2D mais supposent que l'acquisition multi-plans peut améliorer les performances de leur méthode en améliorant la netteté des cellules d'intérêt et par conséquent la quantité d'information disponible pour l'analyse.

Ainsi la suite de ce chapitre étudie l'impact de l'acquisition multi-plans sur les performances d'une méthode d'analyse automatique de lames de cytologie urinaire.

6.3 Méthode

La méthode d'analyse automatique de lames de cytologie digitales pour le diagnostic du cancer de la vessie, utilisée dans ce chapitre, est la méthode décrite précédemment chapitre 5. Cette méthode est appliquée à trois représentations des objets contenus dans les lames de cytologie digitales, comme décrit chapitre 2 :

- S_{centre} : La section centrale de la z-stack, équivalent à un simple acquisition 2D.
- S_{sharp} : La section de meilleur focus de la z-stack, calculée pour chaque objet individuellement.
- S_{edf} : La projection par extension de profondeur de champ de la z-stack, selon la méthode présentée section 2.3.

6.3.1 Acquisition des lames de cytologie digitales

Le même ensemble de lames de cytologie que décrit chapitre 5, provenant de l'essai clinique VisioCyt® 01, est utilisé ici. Pour rappel, les 2323 lames de cytologie ont été numérisées à l'aide d'un scanner Hamamatsu Nanozoomer® S360, au grossissement $\times 40^2$ selon 11 plans focaux espacés de 1 μm .

[130]: MUKHERJEE et al. (2015), « Investigation of scanning parameters for thyroid fine needle aspiration cytology specimens »

[175]: KIM et al. (2022), « Evaluating the role of Z-stack to improve the morphologic evaluation of urine cytology whole slide images for high-grade urothelial carcinoma »

[155]: SANGHVI et al. (2019), « Performance of an artificial intelligence algorithm for reporting urine cytopathology »

2: Grossissement correspondant à une résolution de 0.23 μm par pixel.

6.3.2 Détection des objets d'intérêt

Dans un souci d'équité, l'étape de détection des objets d'intérêt est commune pour les trois représentations considérées S_{centre} , S_{sharp} et S_{edf} .

Les objets de premier plan sont détectés au grossissement équivalent $\times 20$, à partir de la section centrale S_{centre} correspondant au plan de meilleur focus déterminé par le scanner de lames.

L'identification des cellules d'intérêt est ensuite effectuée à l'aide d'un modèle ResNet-18³, en utilisant pour chaque cas la section S_{sharp} .

Cette détection commune d'objets d'intérêt permet d'avoir les mêmes cellules utilisées pour chaque lame de cytologie et pour chaque représentation. En effet, en utilisant les représentations S_{centre} , certains objets peuvent être hors-focus et difficiles à identifier comme cellules d'intérêt. Ainsi, dans le cas où des écarts de performances sont observés entre les différentes représentations, il est possible de conclure que ces variations sont liées à la quantité d'information disponible dans les lames plutôt qu'à une différence dans la détection des cellules d'intérêt.

3: Voir en détail section 5.3.

6.3.3 Calcul des caractéristiques expertes

Les caractéristiques expertes sont calculées pour chaque cellule urothéliale basale détectée dans une lame et pour chaque représentation.

Le calcul de ces caractéristiques expertes nécessite, pour chaque représentation S_{centre} , S_{sharp} et S_{edf} , la segmentation précise des cellules urothéliales basales par rapport à l'arrière-plan et la segmentation précise de leurs noyaux.

Comme décrit dans la sous-section 5.3.3, pour chaque représentation la segmentation précise des cellules est effectuée par seuillage automatique triangle [161]. La précision de la segmentation peut donc varier en présence de flou de défocus, notamment pour les sections S_{centre} (voir Figure 6.1).

[161]: ZACK et al. (1977), « Automatic measurement of sister chromatid exchange frequency. »

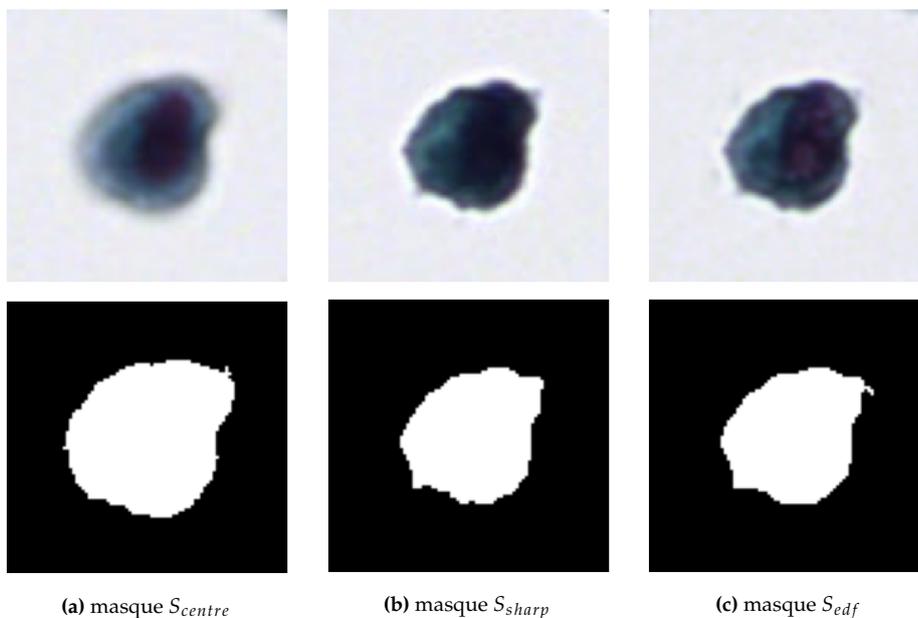


FIGURE 6.1 – Exemple de masques de segmentation sémantique de cellules (bas) pour les trois types de représentation considérées (haut) : (a) S_{centre} , (b) S_{sharp} et (c) S_{edf} .

La segmentation sémantique des noyaux est effectuée, pour chaque représentation, à l'aide du modèle U-Net [16]. Ce modèle U-Net est entraîné pour chaque représentation

[16]: RONNEBERGER et al. (2015), « U-Net »

S_{centre} , S_{sharp} et S_{edf} sur le même jeu de données que celui décrit sous-section 5.3.3. Pour chaque cellule, la même annotation de masque de noyau est utilisée pour les trois représentations S_{centre} , S_{sharp} et S_{edf} .

Les mêmes hyper-paramètres d'entraînement que ceux décrits sous-section 5.3.3 sont utilisés pour chaque représentation, à savoir la minimisation de la fonction de coût \mathcal{L}_{dice} pendant 500 epochs, avec un taille de batch de 16 et un learning rate de 1^{-4} .

De meilleures performances de segmentation sémantique ont été mesurées pour le modèle U-Net entraîné avec les images S_{sharp} et S_{edf} que pour le modèle U-Net entraîné avec les images S_{centre} . Comme attendu, les masques de noyaux prédits à partir des images S_{centre} sont moins précis car les sections S_{centre} peuvent être hors-focus, ce qui rend difficile la segmentation (voir Figure 6.2 et Tableau 6.1).

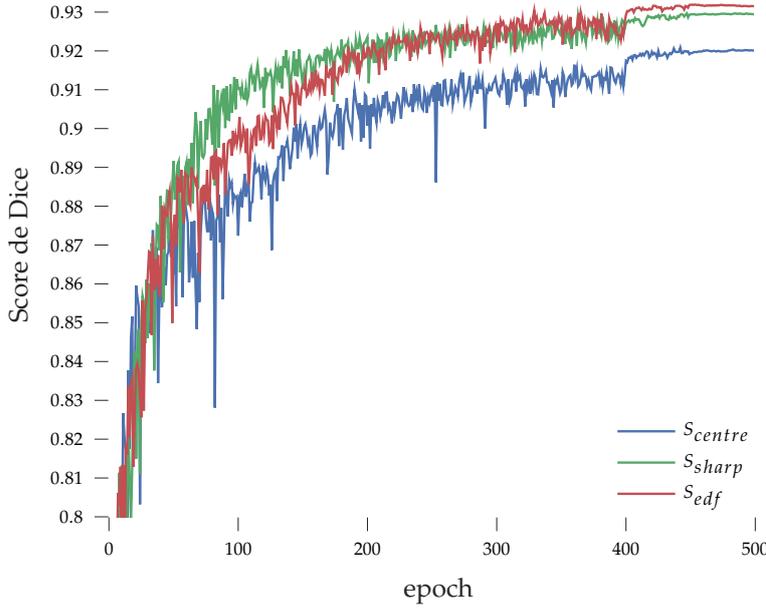


FIGURE 6.2 – Score de Dice au cours de l'entraînement du modèle U-Net pour l'ensemble de données Dev et pour des prédictions effectuées sur les sections S_{centre} , S_{sharp} et S_{edf} . Les performances atteintes sont plus élevées en utilisant les représentations S_{sharp} et S_{edf} .

représentation	score de Dice (<i>Dev</i>)	score de Dice (<i>Test</i>)
S_{centre}	0.920	0.917
S_{sharp}	0.929	0.927
S_{edf}	0.930	0.930

TABLE 6.1 – Performances du modèle U-Net pour la segmentation de noyaux calculées sur l'ensemble de données de développement (*Dev*) et sur l'ensemble de données de (*Test*). Les meilleures performances sont obtenues pour les représentations S_{sharp} et S_{edf} qui ne contiennent pas de cellules hors-focus.

Une fois les masques de segmentation des cellules et des noyaux calculés par représentation, les vecteurs de caractéristiques expertes peuvent être calculés, également pour chaque représentation. Un exemple du calcul de la métrique Ratio Noyau-Cytoplasme (RNC) pour la section S_{centre} est donnée Équation 6.1.

$$\text{RNC}_{centre} = \frac{\sum_{x,y} M_{centre}^{\text{noyau}}}{\sum_{x,y} M_{centre}^{\text{cellule}}} \quad (6.1)$$

Avec :

$M_{centre}^{\text{noyau}}$: Le masque de segmentation du noyau, prédit à partir de S_{centre}
 $M_{centre}^{\text{cellule}}$: Le masque de segmentation de la cellule, prédit à partir de S_{centre}

L'ensemble des caractéristiques expertes décrites dans la sous-section 5.3.3 est calculé de façon similaire, pour chaque représentation et pour chaque cellule urothéliale basale détectée dans une lame.

6.3.4 Classification de lames par représentation

Une fois le vecteur de caractéristiques expertes calculé pour chaque cellule urothéliale basale d'une lame de cytologie, un sac représentant la lame est constitué en sélectionnant les 300 cellules urothéliales basales avec le plus haut Ratio Noyau-Cytoplasme.

À partir de ces sacs, plusieurs modèles de diagnostic automatique de lames de cytologie digitales, décrits en détails chapitre 5, sont entraînés pour les différentes représentations 2D de l'information 3D⁴, à savoir S_{centre} , S_{sharp} et S_{edf} . Les modèles de diagnostic automatique considérés sont les méthodes ayant obtenues les meilleures performances (sous-section 5.4.4), à savoir les méthodes ABMIL et TransMIL.

4: Voir détails chapitre 2.

6.4 Résultats

Les modèles ABMIL et TransMIL sont entraînés à prédire le diagnostic des lames de cytologie, en s'appuyant sur les caractéristiques expertes calculées à partir des représentations S_{centre} , S_{sharp} et S_{edf} . Pour chaque méthode considérée et pour chaque représentation 2D, dix modèles sont entraînés selon une validation croisée à six blocs. Les performances des modèles sont premièrement évaluées en validation croisée, en évaluant la prédiction sur les lames de l'ensemble de développement pour chaque bloc de validation croisée (sous-section 5.4.2). Les performances sont également évaluées sur un ensemble de données de Test indépendant, en adoptant une stratégie ensembliste décrite sous-section 5.4.3.

6.4.1 Évaluation statistique

Pour chaque méthode, afin de déterminer si les différences de performances observées par représentation 2D sont significatives, une analyse statistique de Tukey (ou Tukey HSD) [128] est effectuée. L'analyse de Tukey compare toutes les paires de moyennes possibles pour les différents groupes et cherche à déterminer si ces moyennes sont significativement différentes. Ici, un groupe est l'ensemble des dix valeurs obtenues pour une métrique, correspondant aux dix modèles entraînés pour une représentation 2D. Les trois groupes, correspondant aux trois représentations 2D S_{centre} , S_{sharp} et S_{edf} , sont ainsi comparés à l'aide de l'analyse statistique de Tukey.

[128]: TUKEY (1949), « Comparing Individual Means in the Analysis of Variance »

6.4.2 Performances de diagnostic par représentation

Les performances de ces modèles sont comparées à l'aide des métriques *exactitude*, *sensibilité*, *spécificité* et l'aire sous la courbe ROC (AUC), détaillées sous-section 5.4.1. Les performances moyennées pour les 10 modèles en validation croisée sont reportées Tableau 6.2. Pour chaque méthode et chaque métrique, la valeur la plus haute est reportée en **bleu**, ainsi que toutes les valeurs qui ne sont pas significativement différentes de cette plus haute valeur d'après le test statistique Tukey HSD.

Lors de l'évaluation en validation croisée de la méthode ABMIL, les meilleures performances de diagnostic sont obtenues pour la représentation S_{edf} . En revanche, pour les

représentation	méthode	évaluation en validation croisée			
		exactitude	sensi.	spécif.	AUC
S_{centre}	ABMIL	0.723 ± 0.005	0.735 ± 0.007	0.710 ± 0.007	0.796 ± 0.004
S_{sharp}	ABMIL	0.724 ± 0.009	0.736 ± 0.007	0.712 ± 0.012	0.788 ± 0.005
S_{edf}	ABMIL	0.733 ± 0.008	0.753 ± 0.008	0.712 ± 0.013	0.795 ± 0.004
S_{centre}	ABMIL (rés. conv.)	0.735 ± 0.007	0.721 ± 0.013	0.749 ± 0.013	0.791 ± 0.006
S_{sharp}	ABMIL (rés. conv.)	0.731 ± 0.010	0.722 ± 0.015	0.739 ± 0.013	0.789 ± 0.005
S_{edf}	ABMIL (rés. conv.)	0.734 ± 0.012	0.731 ± 0.014	0.736 ± 0.015	0.791 ± 0.007
S_{centre}	TransMIL (rés. conv.)	0.729 ± 0.008	0.752 ± 0.007	0.706 ± 0.021	0.791 ± 0.004
S_{sharp}	TransMIL (rés. conv.)	0.727 ± 0.006	0.750 ± 0.017	0.704 ± 0.024	0.788 ± 0.005
S_{edf}	TransMIL (rés. conv.)	0.722 ± 0.009	0.752 ± 0.020	0.692 ± 0.028	0.782 ± 0.004

autres méthodes évaluées, aucune représentation 2D ne conduit à une augmentation significative des performances.

Les performances de diagnostic sont ensuite évaluées sur l'ensemble de données de Test. L'inférence sur les données de test est effectuée en utilisant une stratégie ensembliste à partir de six versions des poids d'un modèle obtenus lors de l'apprentissage par validation croisée à six blocs⁵. Les performances obtenues, moyennées pour les dix modèles sur l'ensemble de données de Test, sont reportées Tableau 6.3.

représentation	méthode	métriques sur les données de Test			
		exactitude	sensi.	spécif.	AUC
S_{centre}	ABMIL	0.763 ± 0.004	0.752 ± 0.009	0.774 ± 0.015	0.823 ± 0.002
S_{sharp}	ABMIL	0.766 ± 0.007	0.752 ± 0.007	0.780 ± 0.013	0.823 ± 0.002
S_{edf}	ABMIL	0.742 ± 0.005	0.730 ± 0.011	0.755 ± 0.008	0.825 ± 0.003
S_{centre}	ABMIL (rés. conv.)	0.759 ± 0.008	0.721 ± 0.014	0.797 ± 0.019	0.818 ± 0.005
S_{sharp}	ABMIL (rés. conv.)	0.756 ± 0.009	0.721 ± 0.015	0.790 ± 0.014	0.816 ± 0.004
S_{edf}	ABMIL (rés. conv.)	0.746 ± 0.003	0.719 ± 0.010	0.773 ± 0.011	0.819 ± 0.004
S_{centre}	TransMIL (rés. conv.)	0.754 ± 0.007	0.741 ± 0.012	0.768 ± 0.019	0.819 ± 0.003
S_{sharp}	TransMIL (rés. conv.)	0.752 ± 0.008	0.748 ± 0.013	0.756 ± 0.021	0.812 ± 0.002
S_{edf}	TransMIL (rés. conv.)	0.746 ± 0.012	0.752 ± 0.009	0.740 ± 0.029	0.811 ± 0.004

Sur l'ensemble de données de Test et pour toutes les méthodes évaluées, aucune différence significative de performances de diagnostic n'est observée entre les représentations S_{centre} et S_{sharp} .

Les métriques exactitude, sensibilité et spécificité obtenues à partir de la représentation S_{edf} sont régulièrement significativement inférieures aux métriques obtenues à partir des autres représentations 2D. À noter que pour le modèle ABMIL, il n'y a pas de différence significative observée en fonction des représentations pour la métrique AUC, ce qui laisse penser que la baisse de performances observée pour la représentation S_{edf} peut être dû au seuil de décision sur la prédiction brute du modèle, fixé à 0.5, qui serait alors sous-optimal pour cette représentation.

6.5 Discussion

6.5.1 Répétabilité des expériences

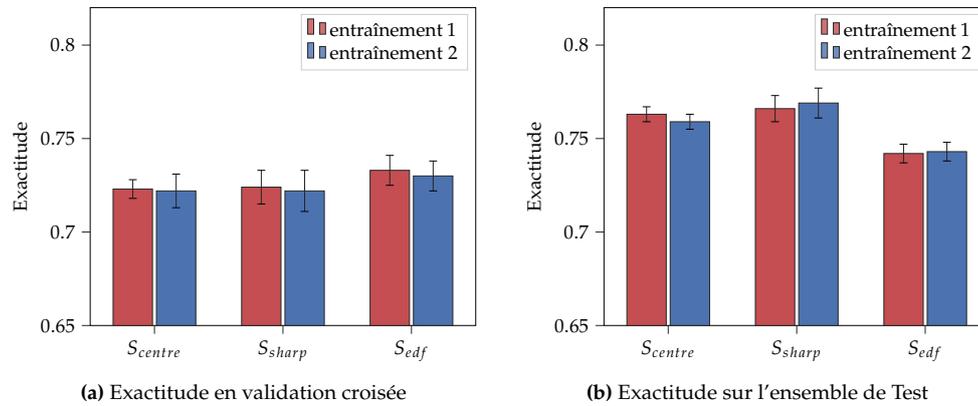
Pour toutes les expériences, chaque modèle est entraîné dix fois selon une validation croisée à six blocs. Cela permet de prendre en compte les variations introduites par l'initialisation aléatoire des poids des modèles et de la stochasticité de l'apprentissage dans l'évaluation des performances.

TABLE 6.2 – Performances des modèles d'apprentissage à partir d'instances multiples (MIL) en validation croisée, en utilisant des caractéristiques expertes calculées à partir de différentes représentations 2D des z-stacks. Pour chaque méthode et chaque métrique, la valeur la plus haute est reportée en **bleu**, ainsi que toutes les valeurs qui ne sont pas significativement différentes de cette plus haute valeur d'après le test statistique Tukey HSD.

5: Voir détail sous-section 5.4.3.

TABLE 6.3 – Performances des modèles de MIL sur l'ensemble de données de Test, en utilisant des caractéristiques expertes calculées à partir de différentes représentations 2D des z-stacks. Pour chaque méthode et chaque métrique, la valeur la plus haute est reportée en **bleu**, ainsi que toutes les valeurs qui ne sont pas significativement différentes de cette plus haute valeur d'après le test statistique Tukey HSD.

Afin d'étudier la stabilité de l'entraînement des modèles et la répétabilité des expériences conduites, deux séries ⁶ de dix entraînements sont réalisées. La précision et l'AUC des modèles ABMIL sont reportées Figure 6.3 et Figure 6.4.



6: Notées *entraînement 1* et *entraînement 2*.

FIGURE 6.3 – Exactitude du modèle ABMIL, calculée par validation croisée (a) et sur l'ensemble de données de test (b), pour deux séries d'entraînements de dix modèles.

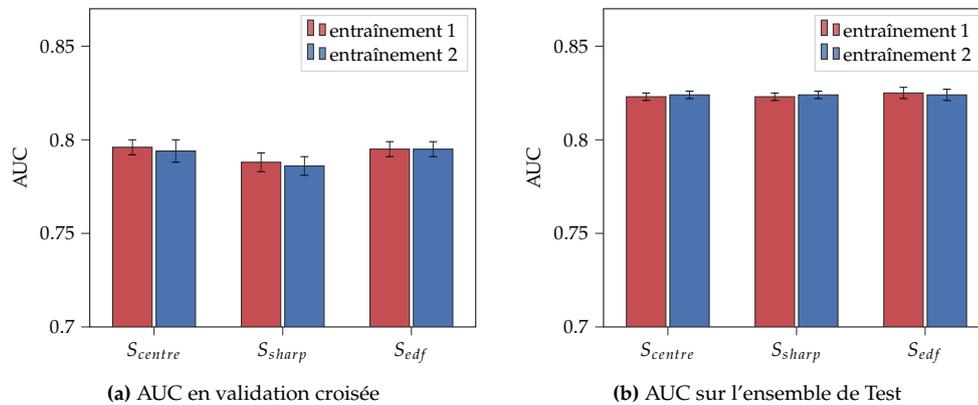


FIGURE 6.4 – AUC du modèle ABMIL, calculée par validation croisée (a) et sur l'ensemble de données de test (b), pour deux séries d'entraînements de dix modèles.

Il en résulte que pour les deux séries *entraînement 1* et *entraînement 2* et pour les représentations S_{centre} , S_{sharp} et S_{edf} , les métriques de précision et d'AUC sont trouvées stables pour une évaluation en validation croisée et sur l'ensemble de données de Test. Ainsi, les performances moyennées sur dix modèles sont stables et permettent de reproduire aisément les expériences menées.

6.5.2 Sensibilité à la segmentation des noyaux

Le calcul des caractéristiques expertes dépend fortement des masques de segmentation des noyaux prédits pour les cellules urothéliales basales. Pour étudier l'impact de petites variations de ces masques de segmentation sémantique des noyaux sur les performances de diagnostic, une deuxième version du même modèle U-Net de segmentation de noyaux est entraînée en suivant exactement la même procédure que décrite sous-section 6.3.3. Cette deuxième version du modèle U-Net présente de légères différences avec la première version, causées par le caractère stochastique de l'apprentissage par descente de gradient ⁷ ainsi que par l'initialisation aléatoire des poids du modèle. Les performances de segmentation obtenues avec les deux versions du modèle U-Net, calculées sur les ensembles Dev et Test du jeu de données de segmentation de régions cellulaires, sont reportés Tableau 6.4.

7: Les exemples composant les batchs servant à l'entraînement du modèle sont choisis aléatoirement.

Les scores de Dice mesurés pour les deux entraînements du modèle Unet ⁸ sont très proches pour les ensembles de données de développement et de test. L'entraînement

8: Notés *version 1* et *version 2*.

poids U-Net	score de Dice (Dev)			score de Dice (Test)		
	S_{centre}	S_{sharp}	S_{edf}	S_{centre}	S_{sharp}	S_{edf}
version 1	0.920	0.929	0.930	0.917	0.927	0.930
version 2	0.920	0.929	0.931	0.914	0.927	0.930

TABLE 6.4 – Performances de deux versions du même modèle U-Net pour la segmentation sémantique de noyaux sur les ensembles de données Dev et Test.

du modèle U-Net pour la segmentation sémantique de noyaux de cellules urothéliales basales est donc stable et facilement reproductible.

À l'aide des caractéristiques expertes, calculés à partir des noyaux de cellules urothéliales basales prédits par les deux versions du modèle U-Net, la méthode ABMIL est entraîné à prédire le diagnostic des lames de cytologie urinaire, suivant la même procédure décrite sous-section 5.3.5. L'expérience est reproduite dix fois pour chaque représentations S_{centre} , S_{sharp} et S_{edf} . La précision des modèles ABMIL entraînés de cette façon est reportée Figure 6.5, et l'AUC est reportée Figure 6.6.

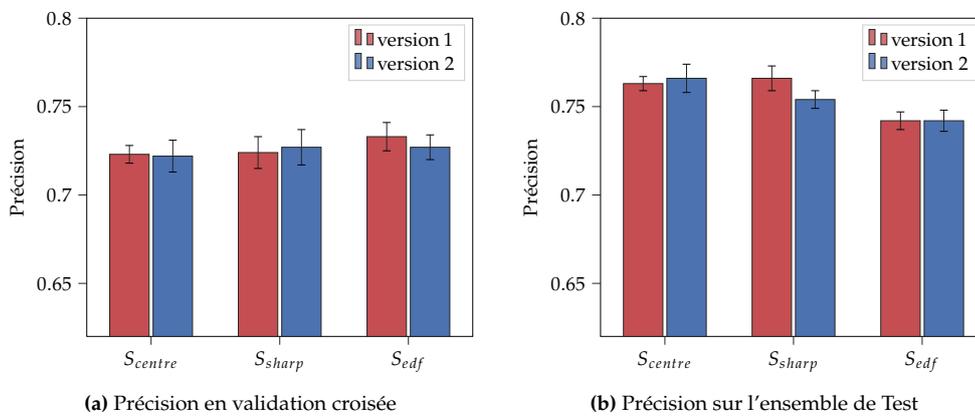


FIGURE 6.5 – Précision du modèle ABMIL, calculée par validation croisée (a) et sur l'ensemble de données de Test (b), lorsque entraîné à partir de caractéristiques expertes calculées à partir de noyaux segmentés par deux versions d'un même modèle U-Net.

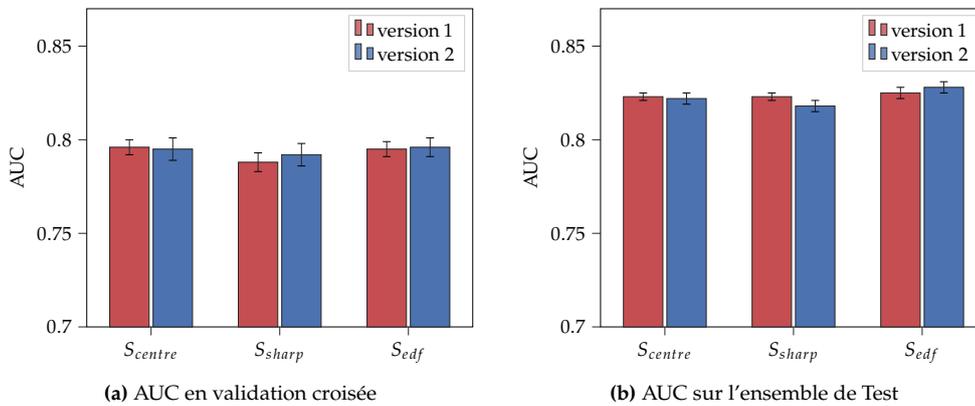


FIGURE 6.6 – AUC du modèle ABMIL, calculée par validation croisée (a) et sur l'ensemble de données de Test (b), lorsque entraîné à partir de caractéristiques expertes calculées à partir de noyaux segmentés par deux versions d'un même modèle U-Net.

La métrique AUC est relativement stable pour les deux versions du modèle U-Net utilisé lors de la prédiction des noyaux de cellules urothéliales basales. La précision du modèle ABMIL est également stable lorsque évaluée en validation croisée, mais plus variable pour les représentations S_{centre} et S_{edf} . Ces variations observées sont plus importantes que les variations liées à la reproduction de l'expérience, évaluées sous-section 6.5.1, ce qui confirme que les performances de diagnostic automatique par la méthode ABMIL dépendent fortement de la précision de segmentation des noyaux des cellules d'intérêt.

6.5.3 Apprentissage ensembliste à partir de différentes représentations 2D des z-stacks

Il a été montré sous-section 5.5.1 que l'apprentissage ensembliste des modèles permet l'amélioration des performances de diagnostic. Ces stratégies ensemblistes ont été trouvées particulièrement utiles lorsque les modèles impliqués sont indépendants [171] et ont ainsi moins tendance à commettre des erreurs de prédictions sur les mêmes échantillons.

[171]: BREIMAN (2001), « Random Forests »

Dans la sous-section 5.5.1, les modèles utilisés pour l'apprentissage ensembliste sont relativement indépendants car chacun est entraîné sur un ensemble de données différent, défini par les blocs de validation croisée.

Parce que les caractéristiques expertes calculées pour les représentations S_{centre} , S_{sharp} et S_{edf} sont différentes, il est alors également possible de combiner les prédictions de modèles indépendants, entraînés sur plusieurs types de représentation.

Pour chaque représentation $S \in \{S_{centre}, S_{sharp}, S_{edf}\}$, dix modèles Θ_i^S avec $i \in [1, 5]$ ont été appris. Soit \mathbf{X} le sac représentant une lame de cytologie digitale, on note $y_{pred}^{S,i} = \Theta_i^S(\mathbf{X})$ la prédiction du modèle Θ_i^S pour cette lame.

On définit alors la stratégie ensembliste suivante pour un nombre N_S de représentations :

$$y_{pred}^i = \frac{1}{N_S} \sum_S \Theta_i^S(\mathbf{X}) \quad \text{avec } i \in [1, 10] \quad (6.2)$$

À noter que la prédiction finale est le résultat de deux combinaisons ensemblistes successives, d'abord entre les 6 modèles obtenus par validation croisée pour obtenir $y_{pred}^{S,i}$ puis entre N_S représentations pour obtenir y_{pred}^i comme décrit Équation 6.2.

Les performances du modèle ABMIL, sur l'ensemble de données de Test, après apprentissage ensembliste entre N_S représentations sont reportées Tableau 6.5.

méthode	représentations utilisées			métriques sur les données de Test			
	S_{centre}	S_{sharp}	S_{edf}	exactitude	sensi.	spécif.	AUC
ABMIL	✓			0.763 ± 0.004	0.752 ± 0.009	0.774 ± 0.015	0.823 ± 0.002
ABMIL		✓		0.766 ± 0.007	0.752 ± 0.007	0.780 ± 0.013	0.823 ± 0.002
ABMIL			✓	0.742 ± 0.005	0.730 ± 0.011	0.755 ± 0.008	0.825 ± 0.003
ABMIL	✓	✓		0.761 ± 0.008	0.743 ± 0.009	0.779 ± 0.015	0.823 ± 0.002
ABMIL	✓		✓	0.758 ± 0.005	0.735 ± 0.012	0.782 ± 0.011	0.828 ± 0.002
ABMIL		✓	✓	0.748 ± 0.006	0.726 ± 0.007	0.769 ± 0.007	0.828 ± 0.001
ABMIL	✓	✓	✓	0.760 ± 0.004	0.741 ± 0.006	0.778 ± 0.007	0.827 ± 0.002

TABLE 6.5 – Performances de la méthode ABMIL avec assemblage des prédictions pour différentes représentations.

Utiliser une stratégie ensembliste entre plusieurs représentations S des z-stacks conduit à une légère augmentation de l'AUC, mais est responsable systématiquement d'une diminution des métriques exactitude, sensibilité et spécificité. Si, après combinaison des prédictions pour différentes représentations S , le modèle semble capable de mieux ordonner les lames de cytologie digitales en fonction de leur label diagnostic (augmentation de la métrique AUC), le seuil de décision sur la prédiction brute fixé à 0.5 se retrouve sous-optimal (diminution de l'exactitude, de la sensibilité et de la spécificité). L'apprentissage ensembliste à partir de différentes représentations 2D des z-stacks ne semble donc pas pertinente.

6.6 Conclusion

Dans ce chapitre, plusieurs méthodes de diagnostic automatique du cancer de la vessie à partir de lames de cytologie en phase liquide⁹ ont été appliquées à des lames digitales numérisées selon 11 plans focaux.

L'information contenue dans les z-stacks ainsi obtenues a été utilisée à l'aide de trois représentations 2D : S_{sharp} la section de meilleur focus de la z-stack, S_{edf} la projection par extension de profondeur de champ de la z-stack, et finalement S_{centre} la section centrale de la z-stack, équivalente à une simple acquisition 2D.

Pour chaque représentation, un détecteur de noyau a été entraîné et évalué. Il a été trouvé que de meilleures performances de segmentation sont obtenues pour les représentations S_{sharp} et S_{edf} qui évitent la présence de cellules floues ce qui rend difficile une segmentation précise du noyau. Ce résultat est similaire aux résultats obtenus chapitre 4 : l'utilisation de z-stacks permet d'améliorer les performances sur les tâches locales, ici de segmentation.

Pour chaque représentation, associée à un détecteur de noyau qui lui est propre, des caractéristiques expertes ont été extraites pour toutes les cellules urothéliales basales détectées par lames de cytologie digitales.

Les méthodes ABMIL et TransMIL, décrites et étudiées chapitre 5, sont entraînées à partir de ces vecteurs de caractéristiques expertes extraites des représentations S_{centre} , S_{sharp} et S_{edf} .

Bien que l'utilisation de z-stacks améliore les performances de segmentation locale, et donc certainement la précision des caractéristiques expertes calculées, il est trouvé que l'utilisation des z-stacks ne conduit pas à une amélioration des performances de diagnostic, pour la méthode de diagnostic automatique de carcinomes urothéliaux proposée.

Ces résultats concordent avec les précédents travaux [174] et [175], qui ont reporté l'absence de gain de performances lors de l'utilisation de z-stacks. Dans ces travaux, les experts interrogés ont cependant préféré l'analyse de z-stacks, évitant la présence de cellules floues pouvant être à l'origine d'incertitudes. Cela concorde également avec les résultats obtenus dans ce chapitre et chapitre 4, reportant de meilleures performances sur des tâches locales lors de l'analyse de lames de cytologie digitales multi-plans.

La numérisation de lames de cytologie digitales selon plusieurs plans focaux augmente le temps d'acquisition et l'espace de stockage nécessaire. Ainsi, parce que l'acquisition multi-plans de lames de cytologie en phase liquide Hologic ThinPrep® ne permet pas d'améliorer le diagnostic automatique du cancer de la vessie, il apparaît préférable d'utiliser de simples acquisitions 2D.

9: Ici la méthode Hologic ThinPrep®.

[174]: DEE et al. (2007), « Utility of 2-D and 3-D Virtual Microscopy in Cervical Cytology Education and Testing »

[175]: KIM et al. (2022), « Evaluating the role of Z-stack to improve the morphologic evaluation of urine cytology whole slide images for high-grade urothelial carcinoma »

Conclusion générale et perspectives

Grâce à la numérisation des lames de pathologie, des algorithmes de traitement et d'analyse automatique d'images peuvent être utilisés pour la quantification objective de certains phénotypes, soit pour la recherche de nouveaux biomarqueurs, soit en routine pour assister le pathologiste lors de la lecture des lames ¹.

D'autres méthodes d'analyse d'images, souvent basées sur des méthodes de machine learning, permettent directement d'extraire de l'information des lames de pathologie digitales et à partir de cette information de prédire automatiquement un diagnostic ou le pronostic d'un patient [51].

Ces méthodes peuvent permettre, par exemple, de compenser en partie le manque de pathologistes dans certaines régions du monde, qui rend difficile le dépistage de certaines pathologies comme le cancer du col de l'utérus [87]. Elles peuvent également soulager la charge de travail des pathologistes en accélérant l'analyse par un pré-traitement automatique des lames de pathologie digitales recommandant les régions les plus atypiques pour relecture par un expert [94]. Finalement, ces méthodes peuvent aider à résoudre des problèmes complexes, comme la prédiction du statut de récepteurs hormonaux pour le cancer du sein, à partir de lames d'histologie colorées en Hématoxyline et Éosine [49].

De telles approches ont été largement étudiées dans le domaine de l'histologie, mais restent peu explorées pour les lames de cytologie [100]. Parmi les difficultés propres à la cytologie, qui peuvent expliquer ce retard, on retrouve la présence d'amas cellulaires 3D et de cellules superposées qui rendent l'analyse difficile et une simple numérisation 2D insuffisante pour imager l'ensemble du matériel cellulaire.

Pour résoudre cette limitation, de nombreux scanners de lames commerciaux proposent maintenant des fonctions dites de *z-stacking* [121] permettant des acquisitions selon plusieurs plans focaux des lames de cytologie digitales.

Cette thèse présente de nouvelles approches permettant d'analyser efficacement des lames de cytologie digitales multi-plans ainsi qu'une méthode de diagnostic automatique du cancer de la vessie à partir de lames de cytologie urinaire. Une étude est également conduite afin d'évaluer l'impact de l'acquisition multi-plans des lames virtuelles sur les performances de diagnostic du cancer de la vessie.

Ce chapitre résume les contributions et principaux résultats des travaux conduits pendant cette thèse, discute les limitations des expériences menées et propose des perspectives de recherches.

7.1 Contributions et principaux résultats

Les contributions et principaux résultats de ce travail sont résumés dans cette section.

7.1.1 Représentations 2D de l'information 3D

Le chapitre 2 introduit plusieurs représentations 2D de l'information 3D contenue dans les *z-stacks*, notamment la section de meilleur focus S_{sharp} et la projection par extension de profondeur de champ (EDF) S_{edf} . Ce chapitre introduit également la représentation S_{centre} correspondant à la section centrale de la *z-stack* et équivalente à une simple

1: Pathologie assistée par ordinateur.

[51]: COURTIOL et al. (2019), « Deep learning-based classification of mesothelioma improves prediction of patient outcome »

[87]: MOHAMMED et al. (2021), « Single-cell conventional pap smear image classification using pre-trained deep neural network architectures »

[94]: CHENG et al. (2021), « Robust whole slide image analysis for cervical cancer screening using deep learning »

[49]: LU et al. (2021), « SlideGraph+ »

[100]: HANNA et al. (2017), « Why is digital pathology in cytopathology lagging behind surgical pathology? »

[121]: HANNA et al. (2017), « Comparison of glass slides and various digital-slide modalities for cytopathology screening and interpretation »

acquisition 2D de la lame de cytologie. Cette représentation sert de référence afin d'estimer l'impact de l'information 3D sur des tâches locales, comme la segmentation de noyaux, ou globales, tel que le diagnostic du cancer de la vessie. Cette étude se concentre sur la définition de la projection par extension de profondeur de champ de la z-stack, notée S_{edf} .

Il est trouvé que pour des z-stacks de cytologie, contenant des cellules superposées et semi-transparentes, les approches basées sur transformée en ondelettes des sections permettent d'obtenir des projections avec visuellement peu d'artefacts et de perte d'information, en comparaison avec d'autres méthodes plus récentes basées sur des réseaux de neurones convolutionnels (CNNs) ou une extraction de *manifold*.

Dans ce chapitre, une méthode permettant de garantir une haute fidélité couleur de la projection EDF est proposée. De plus, un étape de pré-traitement originale est utilisée, permettant de filtrer les sections hors-focus et ainsi de minimiser la présence d'artefacts dans l'image EDF finale.

L'importance de l'utilisation de projection EDF avec peu d'artefacts et une haute fidélité couleur est finalement étudiée au travers d'une segmentation d'amas cellulaires synthétiques d'épaisseur variable. Il est observé qu'à l'aide de la projection EDF, de meilleurs performances de segmentation sont obtenues comparativement aux sections de meilleur focus qui peuvent tout de même contenir des cellules floues. De même, la méthode proposée permet d'obtenir une projection EDF avec une haute fidélité des couleurs, sans laquelle la précision de la segmentation du modèle U-Net utilisé est trouvée affectée.

7.1.2 Préparation et acquisition multi-plans de lames de cytologie digitales

Le chapitre 3 étudie les paramètres de préparation et d'acquisition multi-plans de lames de cytologie digitales.

Dans un premier temps, une méthode d'analyse de la coloration des lames de cytologie est proposée. Cette méthode permet d'évaluer l'impact d'une part de différentes solutions d'hématoxyline commerciales et d'autre part de différents protocoles de coloration, sur l'aspect des lames digitales. Cette approche, appliquée à la cytologie urinaire, repose sur une analyse de la couleur et de la texture des noyaux et cytoplasmes des cellules urothéliales superficielles présentes dans les lames digitales. L'analyse montre que la solution d'hématoxyline *Gill1*, peu concentrée en hématoxyline et adaptée à la cytologie, permet de maximiser la quantité de détails intra-nucléaires visibles, et ainsi d'augmenter la quantité d'information disponible pour le diagnostic.

Dans un second temps, deux paramètres sont considérés pour l'acquisition multi-plans de lames de pathologie digitales : N_z le nombre de plans focaux à numériser et δ_z l'espacement inter-plan. δ_z est fixé à 1 μm comme dans de nombreuses autres études, afin d'être du même ordre de grandeur que la profondeur de champ du scanner. Ainsi entre deux plans focaux numérisés, tous les objets sont imagés nets. Pour la valeur δ_z choisie, le nombre de plans focaux à numériser est fixé à $N_z = 11$, permettant d'avoir plus de 99% des cellules urothéliales basales isolées en focus ², et offrant ainsi un compromis conservateur entre la quantité d'information numérisée, le temps d'acquisition et la taille du fichier digital.

2: Calculé pour 60 lames de cytologie digitales.

7.1.3 Segmentation à partir de z-stacks

Dans le chapitre 4, plusieurs approches sont étudiées afin d'extraire de l'information directement depuis des z-stacks à l'aide de réseaux de neurones convolutionnels. Ces études portent sur différentes tâches de segmentation sémantique et d'instances.

Pour conduire ces travaux, trois jeux de données ont été construits. Le jeu de données Uro, composé de z-stacks dont les masques des régions cellulaires *cytoplasme* et *noyau* ont été annotés. Le jeu de données Cervix67seg, créé à partir du jeu de données publique Cervix93 [132], composé de z-stacks pour lesquelles les masques des noyaux des cellules cervicales ont été annotés. Le jeu de données CISD³, contenant des z-stacks de cellules urothéliales superposés dont les masques d'instances ont été annotés.

Dans un premier temps, pour les problèmes de segmentation sémantique, l'information contenue dans les z-stacks est analysée soit par une projection EDF suivie de l'application d'un CNN classique, soit par une analyse directe des z-stacks à l'aide de CNNs développés à dessein. Inspiré par les travaux [136] et [137], un modèle CNN-1D a été développé pour la segmentation sémantique de régions cellulaires à partir de profils-z. Les mesures de performances montrent que ce modèle CNN-1D surpasse les méthodes [136] et [137], cependant les masques de segmentation sémantique obtenus restent bruités, ne tirant pas partie de l'information spatiale contenue dans les z-stacks. Pour palier à ce problème, un modèle de segmentation U-Net [16] est appliqué premièrement aux projections EDF des z-stacks, puis directement aux z-stacks à l'aide de plusieurs modifications d'architectures proposées. De bonnes performances de segmentation sémantique sont atteintes avec les méthodes proposées, supérieures aux performances des approches décrites dans [132]⁴.

Dans un second temps, pour le problème de segmentation d'instances, trois méthodes de segmentation d'instances : Phoulady [67] ANCIS [148] et SOLO [149] ont été appliquées aux projections EDF de z-stacks pour la segmentation d'instances de cellules urothéliales superposées. La méthode SOLO a été identifiée comme étant la plus performante, puis adaptée afin de pouvoir être appliquée directement à des z-stacks, enlevant ainsi le besoin d'un pré-traitement EDF coûteux en temps de calcul.

Dans chaque cas, l'utilisation de méthodes exploitant l'information tri-dimensionnelle des z-stacks, que ce soit à l'aide d'une projection EDF ou de CNNs adaptés, permet d'atteindre de meilleures performances de segmentation en comparaison avec les méthodes reposant sur de simples acquisitions 2D.

Il est intéressant de noter que, pour des performances de segmentation similaires, l'utilisation directe des z-stacks par des réseaux de neurones convolutionnels adaptés permet de réduire le temps de calcul nécessaire à l'inférence, en évitant le besoin d'une étape de pré-traitement par extension de profondeur de champ. Ainsi, la méthode U-Net-CSE, proposée section 4.2, permet de réduire le temps de calcul nécessaire pour la segmentation sémantique de noyaux de près de 45% par rapport à un pré-traitement EDF suivi d'un modèle U-Net-2D. Similairement, en adaptant la méthode SOLO [149] pour traiter directement les z-stacks, une diminution du temps d'inférence allant jusqu'à 86% est observée, en comparaison avec une projection EDF suivie d'une analyse 2D.

Ainsi, plusieurs méthodes ont été proposées avec succès afin d'extraire de l'information directement à partir des z-stacks. Pour des tâches de segmentation sémantique et de segmentation d'instances, les performances mesurées sont similaires aux performances obtenues sur les représentations S_{edf} et supérieures aux performances obtenues sur les représentations S_{centre} . L'information 3D contenue dans les z-stacks permet donc d'améliorer les performances pour des tâches locales. L'utilisation directe des z-stacks dans les réseaux de neurones permet également de réduire le temps de traitement total

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

3: Ce jeu de données CISD est disponible publiquement.

[136]: LIIMATAINEN et al. (2016), « Supervised method for cell counting from bright field focus stacks »

[137]: LUGAGNE et al. (2018), « Identification of individual cells from z-stacks of bright-field microscopy images »

[16]: RONNEBERGER et al. (2015), « U-Net »

[132]: PHOULADY et al. (2018), *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*

4: État de l'art sur le jeu de données Cervix93.

[67]: PHOULADY et al. (2017), « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images »

[148]: Yi et al. (2019), « Attentive neural cell instance segmentation »

[149]: WANG et al. (2020), « SOLO »

[149]: WANG et al. (2020), « SOLO »

par rapport à l'application d'un pré-traitement par extension de profondeur de champ suivi d'un traitement 2D classique, sans compromis sur les performances.

7.1.4 Diagnostic automatique de carcinomes urothéliaux

Le chapitre 5 décrit une approche de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales. Contrairement aux travaux précédents, cette étude considère pour vérité terrain le résultat des examens de cystoscopie et histologie, examens de référence pour le diagnostic du cancer de la vessie, plutôt que les résultats de l'examen de cytologie. De plus, la méthode proposée cherche à détecter non seulement les carcinomes urothéliaux de haut-grade (HGUCs) mais également les carcinomes urothéliaux de bas-grade (LGUCs).

Plusieurs méthodes d'apprentissage à partir d'instances multiples (MIL) sont évaluées pour résoudre ce problème, où les instances sont alors des vecteurs de caractéristiques expertes inspirés du Système de Paris, calculés à partir des cellules urothéliales basales isolées. L'étude menée montre qu'il est pertinent de sélectionner, pour chaque lame de cytologie digitale, les 300 cellules urothéliales basales avec le plus fort Ratio Noyau-Cytoplasme. Le Ratio Noyau-Cytoplasme (RNC) étant un critère d'atypie majeur dans la classification du Système de Paris, cela revient en première approximation à sélectionner les 300 cellules urothéliales basales les plus atypiques pour chaque lame, afin de créer le sac d'instances qui sera ensuite utilisé dans les méthodes de MIL.

Parmi les méthodes de MIL étudiées, les méthodes ABMIL [62] et TransMIL [165] sont trouvées les plus performantes. Il est également trouvé qu'en entraînant les modèles de MIL avec une validation croisée à six blocs suivie d'une combinaison ensembliste des prédictions pour l'inférence, de meilleures performances de diagnostic sont obtenues comparativement à un entraînement classique sur l'ensemble de données d'entraînement. Sur l'ensemble de données de test, le modèle ABMIL atteint 75.6% de sensibilité, 78.5% de spécificité et 82.4% d'aire sous la courbe ROC (AUC). En détails, en analysant uniquement les cellules urothéliales basales isolées, la méthode ABMIL diagnostique correctement 75.5% des lames négatives, 62.4% des lames LGUC et 84.8% des lames HGUC.

[62]: ILSE et al. (2018), « Attention-based Deep Multiple Instance Learning »

[165]: SHAO et al. (2021), « TransMIL »

7.1.5 Impact des acquisitions multi-plans sur le diagnostic automatique

Finalement, le chapitre 6 étudie l'impact des z-stacks sur les performances de diagnostic automatique du cancer de la vessie.

L'approche décrite chapitre 5 est appliquée aux trois représentations 2D de l'information des z-stacks : S_{centre} , S_{sharp} et S_{edf} . Un détecteur de noyau, nécessaire au calcul de caractéristiques expertes est entraîné pour chaque représentation 2D. Il est trouvé que la segmentation des noyaux est plus précise pour les représentations S_{sharp} et S_{edf} que pour la section centrale S_{centre} qui peut contenir des cellules hors-focus. Des vecteurs de caractéristiques expertes ont été calculés pour ces trois représentations, afin d'être utilisés dans la méthode de MIL. Une fois les vecteurs de caractéristique expertes calculés pour chaque représentation 2D, plusieurs modèles de diagnostic automatique sont entraînés puis évalués. L'évaluation des modèles est effectuée en validation croisée et sur un ensemble de données de test composé de 75 lames positives pour un carcinome urothélial de haut-grade, 76 lames positives pour un carcinome urothélial de bas-grade et 141 lames saines.

Bien que l'utilisation de l'information des z-stacks, au travers des représentations S_{sharp} et S_{edf} permette une amélioration des performances sur des tâches de segmentation locale, il n'a pas été observé de gain de performances pour le diagnostic automatique du cancer de la vessie.

7.2 Limitations des études menées

Les choix arbitraires résultants des contraintes sur les études menées au cours de cette thèse, sont décrits ci-dessous.

L'étude quantitative de l'impact du protocole de coloration de lames sur l'aspect des cellules a été effectué en analysant des cellules urothéliales superficielles provenant uniquement de donneurs volontaires sains. Il n'était pas envisageable d'utiliser des échantillons d'urines de patients avec suspicion de carcinome urothélial pour effectuer de tels tests de coloration. Les lames de donneurs sains ont tendance à contenir peu de cellules urothéliales basales, cellules d'intérêt pour le diagnostic de carcinome urothélial [24], les mesures quantitatives de luminance, de texture et de couleur ont donc été effectuées sur les cellules urothéliales superficielles, présentes en plus grand nombre. Il reste ainsi nécessaire de valider que la solution d'hématoxyline *Gill1*, identifiée comme bien adaptée à l'analyse de lames de cytologie digitales, reste pertinente pour l'analyse de cellules urothéliales basales dans des lames positives à un carcinome urothélial.

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

Dans les études menées chapitre 4, des comparaisons de temps de traitement sont effectuées entre des modèles traitant directement des z-stacks, et des modèles qui reposent sur un pré-traitement par extension de profondeur de champ. Le pré-traitement EDF est lent par rapport à l'inférence des modèles traitant les z-stacks. Il semble possible d'approcher le résultat d'un traitement EDF à l'aide de réseaux de neurones convolutionnels [176]. Une telle approche permet d'accélérer le pré-traitement EDF, grâce aux capacités d'accélération des GPU. Sous réserve que l'approximation de la projection EDF par une telle méthode soit de qualité suffisante et robuste, les modèles reposant sur une projection EDF pourraient devenir compétitifs en temps de calcul avec les modèles traitant directement les z-stacks.

[176]: MANESCU et al. (2022), « Content aware multi-focus image fusion for high-magnification blood film microscopy »

Lors de l'étude de l'impact des z-stacks sur les performances de diagnostic automatique du cancer de la vessie ⁵, dans un souci d'équité, le même modèle de classification de type cellulaire a été utilisé pour toutes les représentations. Ce modèle permet d'identifier les cellules urothéliales basales isolées du reste du matériel cellulaire et a été appliqué dans chaque cas sur la section S_{sharp} . Ainsi, la comparaison des performances de diagnostic pour les représentations S_{centre} , S_{sharp} et S_{edf} repose sur un ensemble de cellules d'intérêt détectées commun, ce qui assure une comparaison juste. Cependant, en procédant ainsi, l'impact de la représentation 2D sur la détection des cellules urothéliales basales isolées n'est pas étudié.

5: Voir chapitre 6.

7.3 Perspectives

7.3.1 Caractéristiques profondes pour le diagnostic

Dans le chapitre 5, des caractéristiques expertes inspirées du Système de Paris ont été utilisées pour résumer l'information contenue dans les cellules urothéliales basales isolées. Ces caractéristiques expertes peuvent être remplacées par des caractéristiques profondes, directement extraites à partir des sous-images de cellules par des réseaux de neurones convolutionnels. Il est envisagé que les caractéristiques profondes, plus

complexes que les caractéristiques expertes, contiennent plus d'information que les caractéristiques expertes et permettent ainsi d'atteindre de meilleures performances de diagnostic. Plusieurs options sont envisageables pour l'apprentissage du CNN permettant l'extraction de caractéristiques profondes :

Apprentissage de bout en bout Le CNN extracteur de caractéristiques est appris en même temps que le modèle d'apprentissage à partir d'instances multiples. Pour chaque image de cellule urothéliale basale dans une lame, l'extracteur de caractéristiques est optimisé afin d'obtenir un vecteur de caractéristiques profondes pertinentes. L'ensemble des vecteurs de caractéristiques ainsi calculés peut alors être directement utilisé dans le modèle de MIL. Ce type d'approche a été appliqué par Butke et al. [99] pour la détection de carcinomes urothéliaux de haut-grade. L'apprentissage de bout en bout a un fort potentiel, permettant à l'extracteur de caractéristiques de calculer des caractéristiques profondes pertinentes pour analyser l'atypie de cellules urothéliales basales. Cependant ce type d'approches doit être utilisé avec précautions, puisque les apprentissages de bout en bout sont connus pour être prompts à apprendre les biais présents dans les données d'entraînement, et ainsi perdre ses capacités de généralisation [177].

Apprentissage à partir d'ImageNet ImageNet [178]⁶ est un jeu de données de grande ampleur qui fait figure de référence pour la reconnaissance d'objets à grande échelle. Avec 1 461 406 images individuelles et 1000 classes d'objets, ce jeu de données est propice à l'entraînement de réseaux de neurones convolutionnels robustes. Ainsi, un CNN peut être entraîné sur le jeu de données ImageNet, et ensuite utilisé comme extracteur de caractéristiques pour d'autres problèmes. Pour les architectures de CNNs classiques comme les modèles ResNet [30], le tenseur en sortie du modèle, juste avant la dernière couche linéaire, peut être utilisé comme vecteur de caractéristiques profondes. Les caractéristiques extraites par un CNN entraîné sur ImageNet sont moins susceptibles d'être biaisées, l'apprentissage n'ayant pas été effectué à partir des lames de cytologie digitales⁷. Cependant, et pour les mêmes raisons, les caractéristiques ainsi obtenues peuvent également être peu pertinentes. Il s'agit de la stratégie adoptée par Shao et al. dans [165].

Apprentissage supervisé d'atypie Un défaut majeur de l'utilisation du jeu de données ImageNet pour entraîner l'extracteur de caractéristique pour des images de cytologie vient de la large différence de domaine entre les images naturelles présentes dans ImageNet et les images médicales. L'information ainsi extraite des images de cytologie peut ne pas être optimale. Afin d'obtenir un extracteur de caractéristiques adapté aux images de cellules urothéliales et pertinent pour le diagnostic du cancer de la vessie, il est possible d'entraîner un réseau de neurones à reconnaître les cellules urothéliales basales atypiques et malignes des cellules urothéliales basales saines. Cette tâche nécessite cependant une base de données annotée conséquente, ce qui est coûteux en ressource. De plus, l'annotation d'atypie par cellule est une tâche particulièrement complexe, parfois subjective et sujette à une forte variabilité entre experts.

Apprentissage auto-supervisé Pour palier aux difficultés d'obtention d'un jeu de données annoté de bonne qualité et de taille suffisante, des méthodes d'apprentissage auto-supervisé⁸ [179][180][181][182][183] peuvent être utilisées. Les méthodes d'apprentissage auto-supervisé peuvent permettre l'apprentissage de caractéristiques adaptées aux images de cytologie sans avoir besoin d'annotations, et ainsi obtenir de meilleures performances qu'avec des caractéristiques apprises sur ImageNet [184].

[99]: BUTKE et al. (2021), « End-to-end Multiple Instance Learning for Whole-Slide Cytopathology of Urothelial Carcinoma »

[177]: VAN GANSBEKE et al. (2020), « SCAN »

[178]: RUSSAKOVSKY et al. (2015), « ImageNet Large Scale Visual Recognition Challenge »

6: ImageNet Large Scale Visual Recognition Challenge (ILSVRC) du nom original de la compétition de vision par ordinateur.

[30]: HE et al. (2015), « Deep Residual Learning for Image Recognition »

7: Contrairement à l'apprentissage de bout en bout.

[165]: SHAO et al. (2021), « TransMIL »

8: *Self-Supervised Learning (SSL)*.

[179]: CHEN et al. (2020), « A Simple Framework for Contrastive Learning of Visual Representations »

[180]: CHEN et al. (2020), *Improved Baselines with Momentum Contrastive Learning*

[181]: CHEN et al. (2020), « Exploring Simple Siamese Representation Learning »

[182]: GRILL et al. (2020), « Bootstrap your own latent »

[183]: KOHPAYEGANI et al. (2021), « Mean Shift for Self-Supervised Learning »

[184]: CIGA et al. (2022), « Self supervised contrastive learning for digital histopathology »

Apprentissage auto-supervisé multimodal Les méthodes d'apprentissage auto-supervisé décrites précédemment permettent ainsi d'entraîner un réseau de neurones à extraire des caractéristiques adaptées aux images de cytologie, sans garantie cependant que ces caractéristiques soient pertinentes pour le diagnostic du cancer de la vessie⁹. Pour adresser ce problème, l'apprentissage auto-supervisé de réseau de neurones peut être appliqué simultanément à plusieurs sources de données¹⁰. Dans [185], Taleb et al. utilisent des images du fond de la rétine et des données génomiques contenant clairement l'information sur la présence de pathologie afin de guider l'apprentissage auto-supervisé. Cette méthode leur permet d'obtenir un réseau de neurones capable d'extraire des caractéristiques profondes pertinentes pour le diagnostic de rétinopathies à partir d'images du fond de la rétine, sans avoir eu besoin d'un jeu de données annoté. Une approche similaire est envisageable pour les images de cellules urothéliales basales, par exemple à l'aide de tests de quantification de la protéine 22 de la matrice cellulaire¹¹ obtenus à partir d'échantillons d'urine [186].

Dans chaque cas décrit ci-dessus, les caractéristiques profondes sont extraites par des réseaux de neurones convolutionnels. Ces caractéristiques peuvent être calculées à partir des représentations 2D définies chapitre 2 : S_{centre} , S_{sharp} et S_{edf} , mais également être directement calculées depuis les z-stacks, en utilisant les approches décrites chapitre 4. Ainsi, l'impact de l'acquisition multi-plans sur les performances de diagnostic nécessite d'être ré-évalué lors de l'utilisation de caractéristiques profondes.

7.3.2 Ajout de l'analyse d'amas cellulaires

Afin d'améliorer les performances de diagnostic automatique du cancer de la vessie à partir de lames de cytologie digitales, il serait intéressant de considérer les amas cellulaires dans l'analyse, en complément des cellules urothéliales basales isolées. En effet, dans le Système de Paris [24], la présence d'amas fibrovasculaires est décrite comme un biomarqueur de LGUC.

De plus, dans [187], Levy et al. ont identifiés que les amas de cellules urothéliales contenant des cellules urothéliales basales atypiques ainsi que le nombre d'amas de cellules urothéliales basales dans une lame sont des marqueurs importants d'atypie de l'échantillon.

Ainsi la combinaison des caractéristiques extraites des amas cellulaires et des caractéristiques extraites des cellules urothéliales basales isolées a le potentiel d'améliorer le diagnostic de carcinomes urothéliaux à partir de lames de cytologie digitales.

Pour extraire des caractéristiques à partir d'amas cellulaires, plusieurs approches sont envisageables. Des caractéristiques expertes, comme le nombre de cellules urothéliales atypiques ou la distribution de la morphologie cellulaire au sein d'un amas, peuvent être calculés après détection et segmentation des instances cellulaires par amas, similairement à la méthode proposée dans [187]. Des caractéristiques profondes peuvent également être extraites par un CNN à partir de sous-images extraites de l'amas cellulaire, comme décrit dans [97].

Les amas de cellules urothéliales étant des structures généralement tri-dimensionnelles, l'impact de l'acquisition multi-plans sur l'analyse de telle structures est nécessaire, suivant la même méthodologie que celle proposée chapitre 6.

9: Contrairement aux caractéristiques profondes appris par apprentissage supervisé d'atypies.

10: Apprentissage multimodal.

[185]: TALEB et al. (2021), « ContIG »

11: Nuclear Matrix Protein 22 - NMP22.

[186]: OTHMAN et al. (2020), « A highly sensitive fluorescent immunosensor for sensitive detection of nuclear matrix protein 22 as biomarker for early stage diagnosis of bladder cancer »

[24]: BARKAN et al. (2016), « The Paris System for Reporting Urinary Cytology »

[187]: LEVY et al. (2022), *Uncovering Additional Predictors of Urothelial Carcinoma from Voided Urothelial Cell Clusters Through a Deep Learning Based Image Preprocessing Technique*

[187]: LEVY et al. (2022), *Uncovering Additional Predictors of Urothelial Carcinoma from Voided Urothelial Cell Clusters Through a Deep Learning Based Image Preprocessing Technique*

[97]: DOV et al. (2021), « Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images »

7.3.3 Filtrage des cellules urothéliales hors-focus dans les acquisitions 2D

Les résultats obtenus chapitre 6 suggèrent que les acquisitions 2D de lames de cytologie digitales contiennent suffisamment d'information pour le diagnostic automatique du cancer de la vessie. Cependant les résultats obtenus chapitre 4 et chapitre 6 montrent également que pour des tâches locales, les acquisitions multi-plans conduisent à de meilleures performances. En effet, en présence de cellules floues, la segmentation de noyaux est peu précise et peut nuire au calcul des caractéristiques expertes.

À partir d'une acquisition 2D d'une lame de cytologie digitale, il est souhaitable de ne pas analyser les cellules urothéliales hors-focus afin de consolider la méthode de diagnostic. Pour cela, une méthode de filtrage automatique peut être utilisée, sous la forme d'un réseau de neurones qui classe les cellules urothéliales basales en deux catégories : *en focus* et *hors-focus*. Seules les cellules urothéliales en focus sont alors prises en compte pour le calcul du vecteur z ¹² de représentation de la lame. Cela permet d'éviter l'analyse erronée de cellules hors-focus.

Plusieurs travaux se sont intéressés au problème de la détection d'objets hors-focus dans les lames de pathologie digitales avec des résultats prometteurs, notamment pour évaluer automatiquement la qualité de la numérisation. Dans [105], les auteurs montrent qu'il est possible d'identifier les noyaux hors-focus dans des lames de cytologie avec une sensibilité de 96% et une spécificité de plus de 99%. Ils en déduisent donc qu'il est raisonnable de numériser les lames de cytologie digitales selon un unique plan focal, et d'utiliser par la suite une étape de détection des objets hors focus. Kohlberger et al. [118] ont développé un réseau de neurones convolutionnel permettant de détecter des sous-images hors-focus provenant de lames d'histologie digitales. Le modèle développé permet également de quantifier le degré de flou des sous-images identifiées hors-focus. Des approches similaires peuvent donc être utilisées pour filtrer les cellules urothéliales basales hors-focus en amont de la méthode de diagnostic proposée.

12: Voir sous-section 5.3.5.

[105]: FRIEDRICH et al. (2016), « Removing defocused objects from single focal plane scans of cytological slides »

[118]: KOHLBERGER et al. (2019), « Whole-slide image focus quality »

7.4 Conclusion

Dans ce document, plusieurs études et méthodes ont été présentées pour le diagnostic du cancer de la vessie à partir de lames de cytopathologie digitales multi-plans.

Une nouvelle méthode de quantification de la coloration des lames de cytologie digitales a été présentée chapitre 3, et a permis le développement d'un protocole de préparation de lames de cytologie urinaire adapté à la cytopathologie digitale. De même, par une analyse de la quantité d'information présente dans les lames digitales, des paramètres conservateurs de numérisation multi-plans des lames de cytologie, à savoir le nombre de plans focaux et l'espacement inter-plans, ont pu être sélectionnés. Cette étude a conduit à la numérisation, selon 11 plans focaux espacés d'1 μm , de 2323 lames de cytologie digitales, provenant de 1023 patients inclus dans l'essai clinique VisioCyt® 01.

À partir de ces données, plusieurs approches ont été proposées pour extraire l'information contenue dans les z-stacks. Ainsi, trois représentations 2D de l'information 3D contenue dans les z-stacks ont été définies chapitre 2, permettant ensuite l'application d'algorithmes de traitement et d'analyse d'images. Parmi ces représentations 2D, une projection par extension de profondeur de champ des z-stacks est proposée avec une nouvelle approche de réassignation des couleurs permettant d'améliorer la fidélité couleur des reconstructions.

Il a également été montré, chapitre 4, que l'information contenue dans les z-stacks peut être directement utilisée par des réseaux de neurones convolutionnels pour la segmentation de régions cellulaires. En particulier, d'excellentes performances ont été obtenues sur des problèmes de segmentation sémantique des noyaux et de segmentation d'instances cellulaires à partir de z-stacks.

À l'aide de l'approche de diagnostic automatique du cancer de la vessie présentée chapitre 5, l'impact de l'utilisation de l'information 3D sur la détection de carcinomes urothéliaux. Bien que l'utilisation d'information 3D conduise à une amélioration des performances pour plusieurs tâches locales de segmentation, paradoxalement il a été montré chapitre 6 que l'utilisation des z-stacks ne permet pas d'améliorer les performances de la méthode proposée de diagnostic automatique du cancer de la vessie. À notre connaissance il s'agit de la première étude de l'impact de l'utilisation de l'information 3D sur les performances d'un algorithme de diagnostic automatique de lames de cytologie digitales.

À cause de l'augmentation du temps d'acquisition et de l'espace de stockage nécessaire à l'acquisition de lames digitales multi-plans, et compte tenu de l'absence de gain de performances mesuré pour la méthode de diagnostic proposée, il apparaît ainsi préférable de numériser les lames de cytologie selon un unique plan focal.

À partir des perspectives évoquées précédemment dans ce chapitre, de futurs travaux pourront s'attacher à confirmer ou discuter les résultats obtenus dans cette thèse. Notamment il sera intéressant d'explorer le diagnostic automatique de lames de cytologie digitales à l'aide de caractéristiques profondes extraites par des CNNs, et d'analyser l'impact de l'information provenant de z-stacks dans ce contexte. L'impact des acquisitions multi-plans sur le diagnostic du cancer de la vessie devra être à nouveau évalué pour ces perspectives.

Bibliographie

Ci-dessous les références utilisées, par ordre de citation.

- [1] Hyuna SUNG et al. « Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries ». en. In : *CA : A Cancer Journal for Clinicians* 71.3 (mai 2021), p. 209-249. DOI : [10.3322/caac.21660](https://doi.org/10.3322/caac.21660). (Visité le 13/01/2022) (cf. p. 1).
- [2] B. K. ANDREASSEN et al. « Incidence and Survival of urothelial carcinoma of the urinary bladder in Norway 1981-2014 ». en. In : *BMC Cancer* 16.1 (déc. 2016), p. 799. DOI : [10.1186/s12885-016-2832-x](https://doi.org/10.1186/s12885-016-2832-x). (Visité le 13/01/2022) (cf. p. 1).
- [3] G. N. PAPANICOLAOU. « A New Procedure for Staining Vaginal Smears ». en. In : *Science* 95.2469 (avr. 1942), p. 438-439. DOI : [10.1126/science.95.2469.438](https://doi.org/10.1126/science.95.2469.438). (Visité le 11/03/2021) (cf. p. 3, 33, 37, 39, 40, 46, 88).
- [4] Peter W. HAMILTON, Yin Hai WANG et Stephen J. McCULLOUGH. « Virtual microscopy and digital pathology in training and education ». en. In : *APMIS* 120.4 (avr. 2012), p. 305-315. DOI : [10.1111/j.1600-0463.2011.02869.x](https://doi.org/10.1111/j.1600-0463.2011.02869.x). (Visité le 04/03/2022) (cf. p. 4).
- [5] Mahadev SATYANARAYANAN et al. « OpenSlide : A vendor-neutral software foundation for digital pathology ». en. In : *Journal of Pathology Informatics* 4.1 (2013), p. 27. DOI : [10.4103/2153-3539.119005](https://doi.org/10.4103/2153-3539.119005). (Visité le 04/03/2022) (cf. p. 4).
- [6] Robert S. LEDLEY. « High-Speed Automatic Analysis of Biomedical Pictures ». en. In : *Science, New Series* 146.3641 (1964), p. 216-223 (cf. p. 5).
- [7] Erik MEIJERING et al. « Imagining the future of bioimage analysis ». en. In : *Nature Biotechnology* 34.12 (déc. 2016), p. 1250-1255. DOI : [10.1038/nbt.3722](https://doi.org/10.1038/nbt.3722). (Visité le 19/07/2022) (cf. p. 5).
- [8] Daniel SAGE et al. « DeconvolutionLab2 : An open-source software for deconvolution microscopy ». en. In : *Methods* 115 (fév. 2017), p. 28-41. DOI : [10.1016/j.ymeth.2016.12.015](https://doi.org/10.1016/j.ymeth.2016.12.015). (Visité le 19/07/2022) (cf. p. 5).
- [9] William MEINIEL, Jean-Christophe OLIVO-MARIN et Elsa D. ANGELINI. « Denoising of Microscopy Images : A Review of the State-of-the-Art, and a New Sparsity-Based Method ». en. In : *IEEE Transactions on Image Processing* 27.8 (août 2018), p. 3842-3856. DOI : [10.1109/TIP.2018.2819821](https://doi.org/10.1109/TIP.2018.2819821). (Visité le 19/07/2022) (cf. p. 5).
- [10] Jean-Christophe OLIVO-MARIN. « Extraction of spots in biological images using multiscale products ». en. In : *Pattern Recognition* 35.9 (sept. 2002), p. 1989-1996. DOI : [10.1016/S0031-3203\(01\)00127-3](https://doi.org/10.1016/S0031-3203(01)00127-3). (Visité le 19/07/2022) (cf. p. 5).
- [11] Erik MEIJERING. « Cell Segmentation : 50 Years Down the Road [Life Sciences] ». en. In : *IEEE Signal Processing Magazine* 29.5 (sept. 2012), p. 140-145. DOI : [10.1109/MSP.2012.2204190](https://doi.org/10.1109/MSP.2012.2204190). (Visité le 19/07/2022) (cf. p. 5).
- [12] Tomas VICAR et al. « Cell segmentation methods for label-free contrast microscopy : review and comprehensive comparison ». en. In : *BMC Bioinformatics* 20.1 (déc. 2019), p. 360. DOI : [10.1186/s12859-019-2880-8](https://doi.org/10.1186/s12859-019-2880-8). (Visité le 19/07/2022) (cf. p. 5).
- [13] HYE-JIN JEONG et al. « Comparison of thresholding methods for breast tumor cell segmentation ». en. In : *Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005*. Busan, South Korea : IEEE, 2005, p. 392-395. DOI : [10.1109/HEALTH.2005.1500489](https://doi.org/10.1109/HEALTH.2005.1500489). (Visité le 19/07/2022) (cf. p. 5).
- [14] Fatimah AL-HAFIZ, Shiroq AL-MEGREN et Heba KURDI. « Red blood cell segmentation by thresholding and Canny detector ». en. In : *Procedia Computer Science* 141 (2018), p. 327-334. DOI : [10.1016/j.procs.2018.10.193](https://doi.org/10.1016/j.procs.2018.10.193). (Visité le 19/07/2022) (cf. p. 5).

- [15] C. ZIMMER et J.-C. OLIVO-MARIN. « Coupled parametric active contours ». en. In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.11 (nov. 2005), p. 1838-1842. doi : [10.1109/TPAMI.2005.214](https://doi.org/10.1109/TPAMI.2005.214). (Visité le 19/07/2022) (cf. p. 5).
- [16] Olaf RONNEBERGER, Philipp FISCHER et Thomas BROX. « U-Net : Convolutional Networks for Biomedical Image Segmentation ». en. In : *International Conference on Medical image computing and computer-assisted intervention*. Mai 2015. (Visité le 10/10/2019) (cf. p. 6, 7, 30, 39, 51, 61, 89, 90, 118, 128).
- [17] Nicolas CHENOUEARD et al. « Objective comparison of particle tracking methods ». en. In : *Nature Methods* 11.3 (mar. 2014), p. 281-289. doi : [10.1038/nmeth.2808](https://doi.org/10.1038/nmeth.2808). (Visité le 19/07/2022) (cf. p. 6).
- [18] Vladimír ULMAN et al. « An objective comparison of cell-tracking algorithms ». en. In : *Nature Methods* 14.12 (déc. 2017), p. 1141-1152. doi : [10.1038/nmeth.4473](https://doi.org/10.1038/nmeth.4473). (Visité le 19/07/2022) (cf. p. 6).
- [19] Humayun IRSHAD et al. « Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology : A Review—Current Status and Future Potential ». en. In : *IEEE Reviews in Biomedical Engineering* 7 (2014), p. 97-114. doi : [10.1109/RBME.2013.2295804](https://doi.org/10.1109/RBME.2013.2295804). (Visité le 19/07/2022) (cf. p. 6).
- [20] Constantinos G. LOUKAS et al. « An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections ». en. In : *Cytometry* 55A.1 (sept. 2003), p. 30-42. doi : [10.1002/cyto.a.10060](https://doi.org/10.1002/cyto.a.10060). (Visité le 20/07/2022) (cf. p. 6).
- [21] Oscar E MERUVIA-PASTOR et al. « Estimating Cell Count and Distribution in Labeled Histological Samples Using Incremental Cell Search ». en. In : *International Journal of Biomedical Imaging* (2011), p. 16 (cf. p. 6).
- [22] Emad A. RAKHA et al. « Prognostic Significance of Nottingham Histologic Grade in Invasive Breast Carcinoma ». en. In : *Journal of Clinical Oncology* 26.19 (juil. 2008), p. 3153-3158. doi : [10.1200/JCO.2007.15.5986](https://doi.org/10.1200/JCO.2007.15.5986). (Visité le 07/03/2022) (cf. p. 6).
- [23] Jonathan I EPSTEIN et al. « The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma ». en. In : *Am J Surg Pathol* 00.00 (2015), p. 10 (cf. p. 6).
- [24] Güliz A. BARKAN et al. « The Paris System for Reporting Urinary Cytology : The Quest to Develop a Standardized Terminology ». en. In : *Acta Cytologica* 60.3 (2016), p. 185-197. doi : [10.1159/000446270](https://doi.org/10.1159/000446270) (cf. p. 6, 22, 29, 33, 35, 42, 51, 61, 83, 85, 89, 90, 104, 105, 107, 115, 130, 132).
- [25] Kunal NAGPAL et al. « Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer ». en. In : *npj Digital Medicine* 2.1 (déc. 2019), p. 48. doi : [10.1038/s41746-019-0112-2](https://doi.org/10.1038/s41746-019-0112-2). (Visité le 08/03/2022) (cf. p. 6, 10).
- [26] Louis J. VAICKUS et al. « Automating the Paris System for urine cytopathology—A hybrid deep-learning and morphometric approach ». en. In : *Cancer Cytopathology* 127.2 (fév. 2019), p. 98-115. doi : [10.1002/cncy.22099](https://doi.org/10.1002/cncy.22099). (Visité le 19/05/2022) (cf. p. 6, 84, 86).
- [27] Hussain FATAKDAWALA et al. « Expectation–Maximization-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR) : Application to Lymphocyte Segmentation on Breast Cancer Histopathology ». en. In : *IEEE Transactions on Biomedical Engineering* 57.7 (juil. 2010), p. 1676-1689. doi : [10.1109/TBME.2010.2041232](https://doi.org/10.1109/TBME.2010.2041232). (Visité le 14/03/2022) (cf. p. 7).
- [28] Fuyong XING, Yuanpu XIE et Lin YANG. « An Automatic Learning-Based Framework for Robust Nucleus Segmentation ». en. In : *IEEE Transactions on Medical Imaging* 35.2 (fév. 2016), p. 550-566. doi : [10.1109/TMI.2015.2481436](https://doi.org/10.1109/TMI.2015.2481436). (Visité le 07/03/2022) (cf. p. 7).
- [29] Hwejin JUNG, Bilal LODHI et Jaewoo KANG. « An automatic nuclei segmentation method based on deep convolutional neural networks for histopathology images ». en. In : *BMC Biomedical Engineering* 1.1 (déc. 2019), p. 24. doi : [10.1186/s42490-019-0026-8](https://doi.org/10.1186/s42490-019-0026-8). (Visité le 07/03/2022) (cf. p. 7).
- [30] Kaiming HE et al. « Deep Residual Learning for Image Recognition ». en. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. Déc. 2015. (Visité le 10/03/2021) (cf. p. 7, 10, 12, 13, 38, 73, 76, 77, 89, 131).
- [31] Hui QU et al. « Joint Segmentation and Fine-Grained Classification of Nuclei in Histopathology Images ». en. In : *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice, Italy : IEEE, avr. 2019, p. 900-904. doi : [10.1109/ISBI.2019.8759457](https://doi.org/10.1109/ISBI.2019.8759457). (Visité le 07/03/2022) (cf. p. 7).

- [32] Peter NAYLOR et al. « Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map ». en. In : *IEEE Transactions on Medical Imaging* 38.2 (fév. 2019), p. 448-459. DOI : [10.1109/TMI.2018.2865709](https://doi.org/10.1109/TMI.2018.2865709). (Visité le 07/03/2022) (cf. p. 7).
- [33] Simon GRAHAM et al. « Hover-Net : Simultaneous segmentation and classification of nuclei in multi-tissue histology images ». en. In : *Medical Image Analysis* 58 (déc. 2019), p. 101563. DOI : [10.1016/j.media.2019.101563](https://doi.org/10.1016/j.media.2019.101563). (Visité le 07/03/2022) (cf. p. 8).
- [34] Jun XU et al. « A high-throughput active contour scheme for segmentation of histopathological imagery ». en. In : *Medical Image Analysis* 15.6 (déc. 2011), p. 851-862. DOI : [10.1016/j.media.2011.04.002](https://doi.org/10.1016/j.media.2011.04.002). (Visité le 14/03/2022) (cf. p. 8).
- [35] Nina LINDER et al. « Identification of tumor epithelium and stroma in tissue microarrays using texture analysis ». en. In : *Diagnostic Pathology* 7.1 (déc. 2012), p. 22. DOI : [10.1186/1746-1596-7-22](https://doi.org/10.1186/1746-1596-7-22). (Visité le 14/03/2022) (cf. p. 8).
- [36] Hao CHEN et al. « DCAN : Deep contour-aware networks for object instance segmentation from histology images ». en. In : *Medical Image Analysis* 36 (fév. 2017), p. 135-146. DOI : [10.1016/j.media.2016.11.004](https://doi.org/10.1016/j.media.2016.11.004). (Visité le 08/03/2022) (cf. p. 8).
- [37] Wouter BULTEN et al. « Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard ». en. In : *Scientific Reports* 9.1 (déc. 2019), p. 864. DOI : [10.1038/s41598-018-37257-4](https://doi.org/10.1038/s41598-018-37257-4). (Visité le 08/03/2022) (cf. p. 8).
- [38] Angel CRUZ-ROA et al. « Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks ». en. In : sous la dir. de Metin N. GURCAN et Anant MADABHUSHI. San Diego, California, USA, mar. 2014, p. 904103. DOI : [10.1117/12.2043872](https://doi.org/10.1117/12.2043872). (Visité le 08/03/2022) (cf. p. 8).
- [39] Huangjing LIN et al. « Fast ScanNet : Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection ». en. In : *IEEE Transactions on Medical Imaging* 38.8 (août 2019), p. 1948-1958. DOI : [10.1109/TMI.2019.2891305](https://doi.org/10.1109/TMI.2019.2891305). (Visité le 08/03/2022) (cf. p. 8).
- [40] Dan C. CIREŞAN et al. « Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks ». en. In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013* 8150 (2013). Sous la dir. de Kensaku MORI et al., p. 411-418. DOI : [10.1007/978-3-642-40763-5_51](https://doi.org/10.1007/978-3-642-40763-5_51). (Visité le 08/03/2022) (cf. p. 8).
- [41] Haibo WANG et al. « Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features ». en. In : *Journal of Medical Imaging* 1.3 (oct. 2014), p. 034003. DOI : [10.1117/1.JMI.1.3.034003](https://doi.org/10.1117/1.JMI.1.3.034003). (Visité le 08/03/2022) (cf. p. 8).
- [42] Shadi ALBARQOUNI et al. « AggNet : Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images ». en. In : *IEEE Transactions on Medical Imaging* 35.5 (mai 2016), p. 1313-1321. DOI : [10.1109/TMI.2016.2528120](https://doi.org/10.1109/TMI.2016.2528120). (Visité le 08/03/2022) (cf. p. 8).
- [43] David TELLEZ et al. « Whole-Slide Mitosis Detection in H&E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks ». en. In : *IEEE Transactions on Medical Imaging* 37.9 (sept. 2018), p. 2126-2136. DOI : [10.1109/TMI.2018.2820199](https://doi.org/10.1109/TMI.2018.2820199). (Visité le 08/03/2022) (cf. p. 8).
- [44] Hao CHEN, Xi WANG et Pheng Ann HENG. « Automated mitosis detection with deep regression networks ». en. In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague, Czech Republic : IEEE, avr. 2016, p. 1204-1207. DOI : [10.1109/ISBI.2016.7493482](https://doi.org/10.1109/ISBI.2016.7493482). (Visité le 08/03/2022) (cf. p. 8).
- [45] Siddhant RAO. *MITOS-RCNN : A Novel Approach to Mitotic Figure Detection in Breast Cancer Histopathology Images using Region Based Convolutional Neural Networks*. en. arXiv : 1807.01788. Juil. 2018. URL : <http://arxiv.org/abs/1807.01788> (visité le 08/03/2022) (cf. p. 8).
- [46] Kaiming HE et al. « Mask R-CNN ». en. In : *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017. DOI : [10.1109/ICCV.2017.322](https://doi.org/10.1109/ICCV.2017.322) (cf. p. 8, 11-13, 72, 76).

- [47] Michel E. VANDENBERGHE et al. « Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer ». en. In : *Scientific Reports* 7.1 (mai 2017), p. 45938. DOI : [10.1038/srep45938](https://doi.org/10.1038/srep45938). (Visité le 08/03/2022) (cf. p. 9).
- [48] Talha QAISER et Nasir M. RAJPOOT. « Learning Where to See : A Novel Attention Model for Automated Immunohistochemical Scoring ». en. In : *IEEE Transactions on Medical Imaging* 38.11 (nov. 2019), p. 2620-2631. DOI : [10.1109/TMI.2019.2907049](https://doi.org/10.1109/TMI.2019.2907049). (Visité le 08/03/2022) (cf. p. 9).
- [49] Wenqi LU et al. « SlideGraph+ : Whole Slide Image Level Graphs to Predict HER2Status in Breast Cancer ». en. In : *Medical Image Analysis* (oct. 2021). (Visité le 08/03/2022) (cf. p. 9, 126).
- [50] Xinliang ZHU et al. « WSISA : Making Survival Prediction from Whole Slide Histopathological Images ». en. In : *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI : IEEE, juil. 2017, p. 6855-6863. DOI : [10.1109/CVPR.2017.725](https://doi.org/10.1109/CVPR.2017.725). (Visité le 08/03/2022) (cf. p. 9).
- [51] Pierre COURTIOL et al. « Deep learning-based classification of mesothelioma improves prediction of patient outcome ». en. In : *Nature Medicine* 25.10 (oct. 2019), p. 1519-1525. DOI : [10.1038/s41591-019-0583-3](https://doi.org/10.1038/s41591-019-0583-3). (Visité le 08/03/2022) (cf. p. 9, 126).
- [52] Yoichiro YAMAMOTO et al. « Automated acquisition of explainable knowledge from unannotated histopathology images ». en. In : *Nature Communications* 10.1 (déc. 2019), p. 5642. DOI : [10.1038/s41467-019-13647-8](https://doi.org/10.1038/s41467-019-13647-8). (Visité le 08/03/2022) (cf. p. 9).
- [53] Hamid SOLTANIAN-ZADEH et Kourosh JAFARI-KHOZANI. « Multiwavelet grading of prostate pathological images ». en. In : sous la dir. de Milan SONKA et J. Michael FITZPATRICK. San Diego, CA, mai 2002, p. 1130. DOI : [10.1117/12.467070](https://doi.org/10.1117/12.467070). (Visité le 14/03/2022) (cf. p. 9).
- [54] Pádraig CUNNINGHAM et Sarah Jane DELANY. « k-Nearest Neighbour Classifiers - A Tutorial ». en. In : *ACM Computing Surveys* 54.6 (juil. 2022), p. 1-25. DOI : [10.1145/3459665](https://doi.org/10.1145/3459665). (Visité le 06/09/2022) (cf. p. 9).
- [55] Mitko VETA et al. « Prognostic value of automatically extracted nuclear morphometric features in whole slide images of male breast cancer ». en. In : *Modern Pathology* 25.12 (déc. 2012), p. 1559-1565. DOI : [10.1038/modpathol.2012.126](https://doi.org/10.1038/modpathol.2012.126). (Visité le 15/03/2022) (cf. p. 9).
- [56] Ajay BASAVANHALLY et al. « Multi-Field-of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides ». en. In : *IEEE Transactions on Biomedical Engineering* 60.8 (août 2013), p. 2089-2099. DOI : [10.1109/TBME.2013.2245129](https://doi.org/10.1109/TBME.2013.2245129). (Visité le 15/03/2022) (cf. p. 9).
- [57] Robert M. HARALICK, K. SHANMUGAM et Its'Hak DINSTEIN. « Textural Features for Image Classification ». In : *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (nov. 1973), p. 610-621. DOI : [10.1109/TSMC.1973.4309314](https://doi.org/10.1109/TSMC.1973.4309314). (Visité le 22/01/2021) (cf. p. 9, 35, 41, 91).
- [58] Zizhao ZHANG et al. « Pathologist-level interpretable whole-slide cancer diagnosis with deep learning ». en. In : *Nature Machine Intelligence* 1.5 (mai 2019), p. 236-245. DOI : [10.1038/s42256-019-0052-1](https://doi.org/10.1038/s42256-019-0052-1). (Visité le 08/03/2022) (cf. p. 9).
- [59] Jason W. WEI et al. « Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks ». en. In : *Scientific Reports* 9.1 (déc. 2019), p. 3358. DOI : [10.1038/s41598-019-40041-7](https://doi.org/10.1038/s41598-019-40041-7). (Visité le 08/03/2022) (cf. p. 10).
- [60] Le Hou et al. « Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification ». en. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. Mar. 2016. (Visité le 08/03/2022) (cf. p. 10).
- [61] Gabriele CAMPANELLA, Vitor Werneck Krauss SILVA et Thomas J. FUCHS. *Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology*. en. arXiv : 1805.06983. Sept. 2018. URL : <http://arxiv.org/abs/1805.06983> (visité le 08/03/2022) (cf. p. 10).
- [62] Maximilian ILSE, Jakub M TOMCZAK et Max WELLING. « Attention-based Deep Multiple Instance Learning ». en. In : *Proceedings of the 35th International Conference on Machine Learning*. Stockholm, Sweden, 2018, p. 10 (cf. p. 10, 87, 93, 94, 96, 129).
- [63] Jesús ANGULO et Georges FLANDRIN. « Microscopic image analysis using mathematical morphology : Application to haematological cytology ». en. In : *Science, Technology and Education of Microscopy : An Overview* (2003), p. 9 (cf. p. 11).

- [64] Youyi SONG et al. « Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning ». en. In : *IEEE Transactions on Biomedical Engineering* 62.10 (oct. 2015), p. 2421-2433. DOI : [10.1109/TBME.2015.2430895](https://doi.org/10.1109/TBME.2015.2430895). (Visité le 09/03/2022) (cf. p. 11).
- [65] Youyi SONG et al. « Accurate Cervical Cell Segmentation from Overlapping Clumps in Pap Smear Images ». en. In : *IEEE Transactions on Medical Imaging* 36.1 (jan. 2017), p. 288-300. DOI : [10.1109/TMI.2016.2606380](https://doi.org/10.1109/TMI.2016.2606380). (Visité le 08/02/2021) (cf. p. 11).
- [66] Hady Ahmady PHOULADY et al. « A new approach to detect and segment overlapping cells in multi-layer cervical cell volume images ». en. In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague, Czech Republic : IEEE, avr. 2016, p. 201-204. DOI : [10.1109/ISBI.2016.7493244](https://doi.org/10.1109/ISBI.2016.7493244). (Visité le 19/05/2021) (cf. p. 11, 73, 75, 76).
- [67] Hady Ahmady PHOULADY et al. « A framework for nucleus and overlapping cytoplasm segmentation in cervical cytology extended depth of field and volume images ». en. In : *Computerized Medical Imaging and Graphics* 59 (juil. 2017), p. 38-49. DOI : [10.1016/j.compmedimag.2017.06.007](https://doi.org/10.1016/j.compmedimag.2017.06.007). (Visité le 19/05/2021) (cf. p. 11, 75, 76, 78, 79, 128).
- [68] Afaf TAREEF et al. « Multi-Pass Fast Watershed for Accurate Segmentation of Overlapping Cervical Cells ». en. In : *IEEE Transactions on Medical Imaging* 37.9 (sept. 2018), p. 2044-2059. DOI : [10.1109/TMI.2018.2815013](https://doi.org/10.1109/TMI.2018.2815013). (Visité le 09/03/2022) (cf. p. 11).
- [69] Yanning ZHOU et al. « IRNet : Instance Relation Network for Overlapping Cervical Cell Segmentation ». en. In : *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Sous la dir. de Dinggang SHEN et al. T. 11764. Cham : Springer International Publishing, 2019, p. 640-648. DOI : [10.1007/978-3-030-32239-7_71](https://doi.org/10.1007/978-3-030-32239-7_71). (Visité le 30/03/2021) (cf. p. 11).
- [70] Yanning ZHOU et al. « Deep Semi-supervised Knowledge Distillation for Overlapping Cervical Cell Instance Segmentation ». en. In : *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Juil. 2020. (Visité le 30/03/2021) (cf. p. 11).
- [71] Meng-Hsiun TSAI et al. « Blood Smear Image Based Malaria Parasite and Infected-Erythrocyte Detection and Segmentation ». en. In : *Journal of Medical Systems* 39.10 (oct. 2015), p. 118. DOI : [10.1007/s10916-015-0280-9](https://doi.org/10.1007/s10916-015-0280-9). (Visité le 10/03/2022) (cf. p. 12).
- [72] Nobuyuki OTSU. « A Threshold Selection Method from Gray-Level Histograms ». en. In : *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (jan. 1979), p. 62-66. DOI : [10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076). (Visité le 23/02/2022) (cf. p. 12, 38, 86-88).
- [73] W. David PAN, Yuhang DONG et Dongsheng WU. « Classification of Malaria-Infected Cells Using Deep Convolutional Neural Networks ». en. In : *Machine Learning - Advanced Techniques and Emerging Applications*. Sous la dir. d'Hamed FARHADI. InTech, sept. 2018. DOI : [10.5772/intechopen.72426](https://doi.org/10.5772/intechopen.72426). (Visité le 15/03/2022) (cf. p. 12).
- [74] Yann LECUN et al. « Gradient-Based Learning Applied to Document Recognition ». en. In : (1998), p. 46 (cf. p. 12).
- [75] Asma MAQSOOD et al. « Deep Malaria Parasite Detection in Thin Blood Smear Microscopic Images ». en. In : *Applied Sciences* 11.5 (mar. 2021), p. 2284. DOI : [10.3390/app11052284](https://doi.org/10.3390/app11052284). (Visité le 10/03/2022) (cf. p. 12).
- [76] Daiyi LI et Zongmin MA. « Residual attention learning network and SVM for malaria parasite detection ». en. In : *Multimedia Tools and Applications* (fév. 2022). DOI : [10.1007/s11042-022-12373-6](https://doi.org/10.1007/s11042-022-12373-6). (Visité le 10/03/2022) (cf. p. 12).
- [77] Jane HUNG et Anne CARPENTER. « Applying Faster R-CNN for Object Detection on Malaria Images ». en. In : *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Honolulu, HI, USA : IEEE, juil. 2017, p. 808-813. DOI : [10.1109/CVPRW.2017.112](https://doi.org/10.1109/CVPRW.2017.112). (Visité le 15/03/2022) (cf. p. 12).
- [78] Shaoqing REN et al. « Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks ». en. In : *Advances in neural information processing systems*. Jan. 2016. (Visité le 03/09/2021) (cf. p. 12).

- [79] Mira S. DAVIDSON et al. « Automated detection and staging of malaria parasites from cytological smears using convolutional neural networks ». en. In : *Biological Imaging* 1 (2021), e2. DOI : [10.1017/S2633903X21000015](https://doi.org/10.1017/S2633903X21000015). (Visité le 10/03/2022) (cf. p. 12).
- [80] Ling ZHANG et al. « DeepPap : Deep Convolutional Networks for Cervical Cell Classification ». en. In : *IEEE Journal of Biomedical and Health Informatics* 21.6 (nov. 2017), p. 1633-1643. DOI : [10.1109/JBHI.2017.2705583](https://doi.org/10.1109/JBHI.2017.2705583). (Visité le 09/03/2022) (cf. p. 12).
- [81] Long D. NGUYEN et al. « Biomedical image classification based on a feature concatenation and ensemble of deep CNNs ». en. In : *Journal of Ambient Intelligence and Humanized Computing* (mar. 2019). DOI : [10.1007/s12652-019-01276-4](https://doi.org/10.1007/s12652-019-01276-4). (Visité le 09/03/2022) (cf. p. 12).
- [82] Christian SZEGEDY et al. « Going Deeper with Convolutions ». en. In : *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014. (Visité le 26/07/2022) (cf. p. 12).
- [83] Débora N. DINIZ et al. « A Deep Learning Ensemble Method to Assist Cytopathologists in Pap Test Image Classification ». en. In : *Journal of Imaging* 7.7 (juil. 2021), p. 111. DOI : [10.3390/jimaging7070111](https://doi.org/10.3390/jimaging7070111). (Visité le 09/03/2022) (cf. p. 12).
- [84] Mingxing TAN et Quoc V. LE. « EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks ». en. In : *International conference on machine learning*. Sept. 2020. (Visité le 26/07/2022) (cf. p. 12).
- [85] Suxiang YU et al. « Automatic Classification of Cervical Cells Using Deep Learning Method ». en. In : *IEEE Access* 9 (2021), p. 32559-32568. DOI : [10.1109/ACCESS.2021.3060447](https://doi.org/10.1109/ACCESS.2021.3060447). (Visité le 30/03/2021) (cf. p. 12).
- [86] Srishti GAUTAM et al. *Considerations for a PAP Smear Image Analysis System with CNN Features*. en. arXiv : 1806.09025. Juin 2018. URL : <http://arxiv.org/abs/1806.09025> (visité le 09/03/2022) (cf. p. 12).
- [87] Mohammed Aliy MOHAMMED, Fetulhak ABDURAHMAN et Yodit Abebe AYALEW. « Single-cell conventional pap smear image classification using pre-trained deep neural network architectures ». en. In : *BMC Biomedical Engineering* 3.1 (déc. 2021), p. 11. DOI : [10.1186/s42490-021-00056-6](https://doi.org/10.1186/s42490-021-00056-6). (Visité le 09/03/2022) (cf. p. 12, 126).
- [88] Gao HUANG et al. « Densely Connected Convolutional Networks ». en. In : *IEEE conference on computer vision and pattern recognition*. Jan. 2018. (Visité le 26/07/2022) (cf. p. 12).
- [89] Haoming LIN et al. « Fine-Grained Classification of Cervical Cells Using Morphological and Appearance Based Convolutional Neural Networks ». en. In : *IEEE Access* 7 (2019), p. 71541-71549. DOI : [10.1109/ACCESS.2019.2919390](https://doi.org/10.1109/ACCESS.2019.2919390). (Visité le 09/03/2022) (cf. p. 12).
- [90] Khin Yadanar WIN et al. « Computer Aided Diagnosis System for Detection of Cancer Cells on Cytological Pleural Effusion Images ». en. In : *BioMed Research International* 2018 (nov. 2018), p. 1-21. DOI : [10.1155/2018/6456724](https://doi.org/10.1155/2018/6456724). (Visité le 09/03/2022) (cf. p. 12, 13).
- [91] N. SOMPAWONG et al. « Automated Pap Smear Cervical Cancer Screening Using Deep Learning ». en. In : *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany : IEEE, juil. 2019, p. 7044-7048. DOI : [10.1109/EMBC.2019.8856369](https://doi.org/10.1109/EMBC.2019.8856369). (Visité le 09/03/2022) (cf. p. 12, 13).
- [92] Jing KE et al. « Quantitative analysis of abnormalities in gynecologic cytopathology with deep learning ». en. In : *Laboratory Investigation* 101.4 (avr. 2021), p. 513-524. DOI : [10.1038/s41374-021-00537-1](https://doi.org/10.1038/s41374-021-00537-1). (Visité le 30/03/2021) (cf. p. 13).
- [93] Antoine PIROVANO et al. « Computer-aided diagnosis tool for cervical cancer screening with weakly supervised localization and detection of abnormalities using adaptable and explainable classifier ». en. In : *Medical Image Analysis* 73 (oct. 2021), p. 102167. DOI : [10.1016/j.media.2021.102167](https://doi.org/10.1016/j.media.2021.102167). (Visité le 10/03/2022) (cf. p. 13, 86).
- [94] Shenghua CHENG et al. « Robust whole slide image analysis for cervical cancer screening using deep learning ». en. In : *Nature Communications* 12.1 (déc. 2021), p. 5639. DOI : [10.1038/s41467-021-25296-x](https://doi.org/10.1038/s41467-021-25296-x). (Visité le 10/03/2022) (cf. p. 13, 86, 126).

- [95] Ruqayya AWAN et al. « Deep learning based digital cell profiles for risk stratification of urine cytology images ». en. In : *Cytometry Part A* 99.7 (juil. 2021), p. 732-742. DOI : [10.1002/cyto.a.24313](https://doi.org/10.1002/cyto.a.24313). (Visité le 10/03/2022) (cf. p. 13, 86).
- [96] François CHOLLET. *Xception : Deep Learning with Depthwise Separable Convolutions*. en. arXiv :1610.02357 [cs]. Avr. 2017. URL : <http://arxiv.org/abs/1610.02357> (visité le 26/07/2022) (cf. p. 13).
- [97] David DOV et al. « Weakly supervised instance learning for thyroid malignancy prediction from whole slide cytopathology images ». en. In : *Medical Image Analysis* 67 (jan. 2021), p. 101814. DOI : [10.1016/j.media.2020.101814](https://doi.org/10.1016/j.media.2020.101814). (Visité le 10/03/2022) (cf. p. 13, 87, 132).
- [98] Atsushi TERAMOTO et al. « Weakly supervised learning for classification of lung cytological images using attention-based multiple instance learning ». en. In : *Scientific Reports* 11.1 (déc. 2021), p. 20317. DOI : [10.1038/s41598-021-99246-4](https://doi.org/10.1038/s41598-021-99246-4). (Visité le 10/03/2022) (cf. p. 13, 87).
- [99] Joshua BUTKE et al. « End-to-end Multiple Instance Learning for Whole-Slide Cytopathology of Urothelial Carcinoma ». en. In : (2021), p. 12 (cf. p. 13, 87, 131).
- [100] Matthew G. HANNA et Liron PANTANOWITZ. « Why is digital pathology in cytopathology lagging behind surgical pathology ? » en. In : *Cancer Cytopathology* (2017), p. 3. DOI : <https://doi.org/10.1002/cncy.21855> (cf. p. 13, 126).
- [101] Yilun FAN et Andrew P. BRADLEY. « A method for quantitative analysis of clump thickness in cervical cytology slides ». en. In : *Micron* 80 (jan. 2016), p. 73-82. DOI : [10.1016/j.micron.2015.09.002](https://doi.org/10.1016/j.micron.2015.09.002). (Visité le 16/10/2018) (cf. p. 13, 15, 18, 27, 29, 31, 50).
- [102] Katsushige YAMASHIRO et al. « Comparison between a traditional single still image and a multiframe video image along the z -axis of the same microscopic field of interest in cytology : Which does contribute to telecytology ? » en. In : *Diagnostic Cytopathology* 37.10 (oct. 2009), p. 727-731. DOI : [10.1002/dc.21078](https://doi.org/10.1002/dc.21078). (Visité le 16/03/2022) (cf. p. 13).
- [103] Walid E. KHALBUSS, Liron PANTANOWITZ et Anil V. PARWANI. « Digital Imaging in Cytopathology ». en. In : *Pathology Research International* 2011 (juil. 2011), p. 1-10. DOI : [10.4061/2011/264683](https://doi.org/10.4061/2011/264683). (Visité le 16/03/2022) (cf. p. 13).
- [104] AmberD DONNELLY et al. « Optimal z-axis scanning parameters for gynecologic cytology specimens ». en. In : *Journal of Pathology Informatics* 4.1 (2013), p. 38. DOI : [10.4103/2153-3539.124015](https://doi.org/10.4103/2153-3539.124015). (Visité le 10/03/2021) (cf. p. 13, 15, 33, 50, 116).
- [105] David FRIEDRICH et al. « Removing defocused objects from single focal plane scans of cytological slides ». en. In : *Journal of Pathology Informatics* 7.1 (2016), p. 21. DOI : [10.4103/2153-3539.181765](https://doi.org/10.4103/2153-3539.181765). (Visité le 10/03/2021) (cf. p. 14, 35, 133).
- [106] MichaelC MONTALTO, RobertJ FILKINS et RichardR MCKAY. « Autofocus methods of whole slide imaging systems and the introduction of a second-generation independent dual sensor scanning method ». en. In : *Journal of Pathology Informatics* 2.1 (2011), p. 44. DOI : [10.4103/2153-3539.86282](https://doi.org/10.4103/2153-3539.86282). (Visité le 23/03/2022) (cf. p. 14).
- [107] Terese BOOST. « A comparison of screening times between the ThinPrep imager and conventional cytology ». en. In : *Diagnostic Cytopathology* (2009), p. 4. DOI : [10.1002/dc.21069](https://doi.org/10.1002/dc.21069) (cf. p. 15).
- [108] Yilun FAN, Yaniv GAL et Andrew P. BRADLEY. « An algorithm for microscopic specimen delineation and focus candidate selection ». en. In : *Micron* 66 (nov. 2014), p. 51-62. DOI : [10.1016/j.micron.2014.05.006](https://doi.org/10.1016/j.micron.2014.05.006). (Visité le 13/11/2018) (cf. p. 15).
- [109] F.S. HELMLI et S. SCHERER. « Adaptive shape from focus with an error estimation in light microscopy ». en. In : *ISPA 2001. Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces (IEEE Cat. No.01EX480)*. Pula, Croatia : Univ. Zagreb, 2001, p. 188-193. DOI : [10.1109/ISPA.2001.938626](https://doi.org/10.1109/ISPA.2001.938626). (Visité le 01/02/2019) (cf. p. 19, 24).
- [110] Asm SHIHAVUDDIN et al. « Smooth 2D manifold extraction from 3D image stack ». en. In : *Nature Communications* 8.1 (août 2017), p. 15554. DOI : [10.1038/ncomms15554](https://doi.org/10.1038/ncomms15554). (Visité le 14/06/2019) (cf. p. 19, 24, 53).

- [111] Andrew P BRADLEY et Pascal C BAMFORD. « A One-pass Extended Depth of Field Algorithm Based on the Over-complete Discrete Wavelet Transform ». en. In : *Image and Vision Computing'04 New Zealand (IVCNZ'04)*. 2004, p. 279-284 (cf. p. 19, 25, 51, 68).
- [112] B. FORSTER et al. « Extended depth-of-focus for multi-channel microscopy images : A complex wavelet approach ». en. In : *2004 2nd IEEE International Symposium on Biomedical Imaging : Macro to Nano (IEEE Cat No. 04EX821)*. T. 2. Arlington, VA, USA : IEEE, 2004, p. 660-663. DOI : [10.1109/ISBI.2004.1398624](https://doi.org/10.1109/ISBI.2004.1398624). (Visité le 30/10/2018) (cf. p. 19, 23, 24, 51).
- [113] Yu LIU et al. « Multi-focus image fusion with a deep convolutional neural network ». en. In : *Information Fusion* 36 (juil. 2017), p. 191-207. DOI : [10.1016/j.inffus.2016.12.001](https://doi.org/10.1016/j.inffus.2016.12.001). (Visité le 18/07/2019) (cf. p. 19, 24).
- [114] Gustav MÅRTENSSON et al. « The reliability of a deep learning model in clinical out-of-distribution MRI data : A multicohort study ». en. In : *Medical Image Analysis* 66 (déc. 2020), p. 101714. DOI : [10.1016/j.media.2020.101714](https://doi.org/10.1016/j.media.2020.101714). (Visité le 16/08/2022) (cf. p. 22).
- [115] ZHI LU, GUSTAVO CARNEIRO et Andrew P. BRADLEY. « An Improved Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells ». en. In : *IEEE Transactions on Image Processing* 24.4 (avr. 2015), p. 1261-1272. DOI : [10.1109/TIP.2015.2389619](https://doi.org/10.1109/TIP.2015.2389619). (Visité le 25/06/2019) (cf. p. 25, 71).
- [116] David TELLEZ et al. « Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology ». en. In : *Medical Image Analysis* 58 (déc. 2019), p. 101544. DOI : [10.1016/j.media.2019.101544](https://doi.org/10.1016/j.media.2019.101544). (Visité le 22/08/2022) (cf. p. 33, 34).
- [117] Norman ZERBE, Peter HUFNAGL et Karsten SCHLÜNS. « Distributed computing in image analysis using open source frameworks and application to image sharpness assessment of histological whole slide images ». en. In : *Diagnostic Pathology* 6.S1 (déc. 2011), S16. DOI : [10.1186/1746-1596-6-S1-S16](https://doi.org/10.1186/1746-1596-6-S1-S16). (Visité le 08/03/2021) (cf. p. 35, 36).
- [118] Timo KOHLBERGER et al. « Whole-slide image focus quality : Automatic assessment and impact on ai cancer detection ». en. In : *Journal of Pathology Informatics* 10.1 (2019), p. 39. DOI : [10.4103/jpi.jpi_11_19](https://doi.org/10.4103/jpi.jpi_11_19). (Visité le 10/03/2021) (cf. p. 35, 36, 133).
- [119] Bernd LAHRMANN et al. « Semantic Focusing Allows Fully Automated Single-Layer Slide Scanning of Cervical Cytology Slides ». en. In : *PLoS ONE* 8.4 (avr. 2013). Sous la dir. de Paul van DIEST, e61441. DOI : [10.1371/journal.pone.0061441](https://doi.org/10.1371/journal.pone.0061441). (Visité le 10/03/2021) (cf. p. 35).
- [120] Slawomir WALKOWSKI et Janusz SZYMAS. « Quality evaluation of virtual slides using methods based on comparing common image areas ». en. In : *Diagnostic Pathology* 6.S1 (déc. 2011), S14. DOI : [10.1186/1746-1596-6-S1-S14](https://doi.org/10.1186/1746-1596-6-S1-S14). (Visité le 08/03/2021) (cf. p. 35).
- [121] Matthew G. HANNA et al. « Comparison of glass slides and various digital-slide modalities for cytopathology screening and interpretation ». en. In : *Cancer Cytopathology* 125.9 (sept. 2017), p. 701-709. DOI : [10.1002/cncy.21880](https://doi.org/10.1002/cncy.21880). (Visité le 10/03/2021) (cf. p. 35, 126).
- [122] Yves-Rémi Van EYCKE. « Image processing in digital pathology : an opportunity to solve inter-batch variability of immunohistochemical staining ». en. In : *Scientific Reports* (2017), p. 15 (cf. p. 36).
- [123] Elizabeth CHLIPALA et al. « Optical density-based image analysis method for the evaluation of hematoxylin and eosin staining precision ». en. In : *Journal of Histotechnology* 43.1 (jan. 2020), p. 29-37. DOI : [10.1080/01478885.2019.1708611](https://doi.org/10.1080/01478885.2019.1708611). (Visité le 19/10/2020) (cf. p. 36, 41).
- [124] David AMEISEN et al. « Stack or trash ? Quality assessment of virtual slides ». en. In : *Diagnostic Pathology* 8.S1 (sept. 2013), S23. DOI : [10.1186/1746-1596-8-S1-S23](https://doi.org/10.1186/1746-1596-8-S1-S23). (Visité le 08/03/2021) (cf. p. 36).
- [125] RoyE LEE et al. « Evaluation and optimization for liquid-based preparation cytology in whole slide imaging ». en. In : *Journal of Pathology Informatics* 2.1 (2011), p. 46. DOI : [10.4103/2153-3539.86285](https://doi.org/10.4103/2153-3539.86285). (Visité le 10/03/2021) (cf. p. 37).
- [126] Adam PASZKE et al. « PyTorch : An Imperative Style, High-Performance Deep Learning Library ». en. In : *Advances in Neural Information Processing Systems* 32 (2019), p. 12 (cf. p. 39, 65, 90, 98).

- [127] Santanu ROY et al. « A study about color normalization methods for histopathology images ». en. In : *Micron* 114 (nov. 2018), p. 42-61. DOI : [10.1016/j.micron.2018.07.005](https://doi.org/10.1016/j.micron.2018.07.005). (Visité le 09/03/2021) (cf. p. 41).
- [128] John W. TUKEY. « Comparing Individual Means in the Analysis of Variance ». en. In : *Biometrics* 5.2 (juin 1949), p. 99. DOI : [10.2307/3001913](https://doi.org/10.2307/3001913). (Visité le 29/03/2021) (cf. p. 43, 120).
- [129] STUDENT. « The Probable Error of a Mean ». In : *Biometrika* 6.1 (mar. 1908). DOI : <https://doi.org/10.2307/2331554> (cf. p. 43, 70).
- [130] MaheswariS MUKHERJEE et al. « Investigation of scanning parameters for thyroid fine needle aspiration cytology specimens : A pilot study ». en. In : *Journal of Pathology Informatics* 6.1 (2015), p. 43. DOI : [10.4103/2153-3539.161610](https://doi.org/10.4103/2153-3539.161610). (Visité le 10/03/2021) (cf. p. 50, 117).
- [131] D A AGARD. « Optical Sectioning Microscopy : Cellular Architecture in Three Dimensions ». en. In : *Annual Review of Biophysics and Bioengineering* 13 (1984), p. 31 (cf. p. 51).
- [132] Hady Ahmady PHOULADY et Peter R. MOUTON. *A New Cervical Cytology Dataset for Nucleus Detection and Image Classification (Cervix93) and Methods for Cervical Nucleus Detection*. en. arXiv : 1811.09651. Nov. 2018. URL : <http://arxiv.org/abs/1811.09651> (visité le 10/02/2021) (cf. p. 51, 55, 64, 66-68, 70, 128).
- [133] Jyrki SELINUMMI et al. « Bright Field Microscopy as an Alternative to Whole Cell Fluorescence in Automated Analysis of Macrophage Images ». en. In : *PLoS ONE* 4.10 (oct. 2009). Sous la dir. de Teresa SERRANO-GOTARREDONA, e7497. DOI : [10.1371/journal.pone.0007497](https://doi.org/10.1371/journal.pone.0007497). (Visité le 20/05/2019) (cf. p. 52).
- [134] Dietrich DEHLINGER et al. « Dye free automated cell counting and analysis ». en. In : *Biotechnology and Bioengineering* 110.3 (mar. 2013), p. 838-847. DOI : [10.1002/bit.24757](https://doi.org/10.1002/bit.24757). (Visité le 12/11/2019) (cf. p. 52).
- [135] Tao WANG et al. « Nucleus Segmentation of Cervical Cytology Images Based on Depth Information ». en. In : *IEEE Access* 8 (2020), p. 75846-75859. DOI : [10.1109/ACCESS.2020.2989369](https://doi.org/10.1109/ACCESS.2020.2989369). (Visité le 20/10/2020) (cf. p. 53, 61).
- [136] Kaisa LIIMATAINEN et al. « Supervised method for cell counting from bright field focus stacks ». en. In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague, Czech Republic : IEEE, avr. 2016, p. 391-394. DOI : [10.1109/ISBI.2016.7493290](https://doi.org/10.1109/ISBI.2016.7493290). (Visité le 22/10/2019) (cf. p. 53, 56, 57, 61, 128).
- [137] Jean-Baptiste LUGAGNE et al. « Identification of individual cells from z-stacks of bright-field microscopy images ». en. In : *Scientific Reports* 8.1 (déc. 2018). DOI : [10.1038/s41598-018-29647-5](https://doi.org/10.1038/s41598-018-29647-5). (Visité le 08/01/2019) (cf. p. 53, 56, 57, 61, 128).
- [138] Fabian PEDREGOSA et al. « Scikit-learn : Machine Learning in Python ». en. In : *Journal of Machine Learning Research* 12 (2011), p. 6 (cf. p. 58).
- [139] Sajith Kecheril SADANANDAN et al. « Automated Training of Deep Convolutional Neural Networks for Cell Segmentation ». en. In : *Scientific Reports* 7.1 (déc. 2017), p. 7860. DOI : [10.1038/s41598-017-07599-6](https://doi.org/10.1038/s41598-017-07599-6). (Visité le 20/10/2020) (cf. p. 61).
- [140] Kaisa LIIMATAINEN et al. « Iterative unsupervised domain adaptation for generalized cell detection from brightfield z-stacks ». en. In : *BMC Bioinformatics* 20.1 (déc. 2019), p. 80. DOI : [10.1186/s12859-019-2605-z](https://doi.org/10.1186/s12859-019-2605-z). (Visité le 22/10/2019) (cf. p. 61, 64, 69).
- [141] Jianxu CHEN et al. « Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation ». en. In : *Advances in neural information processing systems*. Sept. 2016. (Visité le 18/01/2021) (cf. p. 63).
- [142] Jie Hu et al. « Squeeze-and-Excitation Networks ». en. In : *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA : IEEE, 2018. DOI : [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745). (Visité le 06/04/2021) (cf. p. 64, 73).
- [143] Zhi Lu et al. « Evaluation of Three Algorithms for the Segmentation of Overlapping Cervical Cells ». en. In : *IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS* 21.2 (2017), p. 10 (cf. p. 68, 71).

- [144] Wei CHEN et al. « HSN : Hybrid Segmentation Network for Small Cell Lung Cancer Segmentation ». en. In : *IEEE Access* 7 (2019), p. 75591-75603. DOI : [10.1109/ACCESS.2019.2921434](https://doi.org/10.1109/ACCESS.2019.2921434). (Visité le 12/01/2021) (cf. p. 69).
- [145] Ritu NAYAR et David C. WILBUR, éd. *The Bethesda System for Reporting Cervical Cytology : Definitions, Criteria, and Explanatory Notes*. en. Cham : Springer International Publishing, 2015. (Visité le 13/04/2021) (cf. p. 70).
- [146] Carlo Emilio BONFERRONI. *Teoria statistica delle classi e calcolo delle probabilita*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936 (cf. p. 70).
- [147] Tsung-Yi LIN et al. « Microsoft COCO : Common Objects in Context ». en. In : *Proceedings of the IEEE european conference on computer vision*. Springer, fév. 2015. (Visité le 03/09/2021) (cf. p. 72).
- [148] Jingru YI et al. « Attentive neural cell instance segmentation ». en. In : *Medical Image Analysis* 55 (juil. 2019), p. 228-240. DOI : [10.1016/j.media.2019.05.004](https://doi.org/10.1016/j.media.2019.05.004). (Visité le 26/05/2021) (cf. p. 72, 76, 77, 79, 128).
- [149] Xinlong WANG et al. « SOLO : Segmenting Objects by Locations ». en. In : *ECCV 2020*. Glasgow, United Kingdom, 2020 (cf. p. 72, 77, 79, 81, 128).
- [150] Xinlong WANG et al. « SOLOv2 : Dynamic and Fast Instance Segmentation ». en. In : *Advances in neural information processing systems*. Oct. 2020. (Visité le 02/09/2021) (cf. p. 72, 77, 79, 81).
- [151] Hao CHEN et al. « DCAN : Deep Contour-Aware Networks for Accurate Gland Segmentation ». en. In : *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. Avr. 2016. (Visité le 19/05/2021) (cf. p. 72).
- [152] Wei LIU et al. « SSD : Single Shot MultiBox Detector ». en. In : *European conference on computer vision*. T. 9905. 2016, p. 21-37. DOI : [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2). (Visité le 25/05/2021) (cf. p. 76, 77).
- [153] Tsung-Yi LIN et al. « Feature Pyramid Networks for Object Detection ». en. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. Avr. 2017. (Visité le 26/05/2021) (cf. p. 77).
- [154] Jifeng DAI et al. « Deformable Convolutional Networks ». en. In : *Proceedings of the IEEE international conference on computer vision*. Juin 2017. (Visité le 13/09/2021) (cf. p. 77).
- [155] Adit B. SANGHVI et al. « Performance of an artificial intelligence algorithm for reporting urine cytology ». en. In : *Cancer Cytopathology* 127.10 (oct. 2019), p. 658-666. DOI : [10.1002/cncy.22176](https://doi.org/10.1002/cncy.22176). (Visité le 20/01/2020) (cf. p. 85, 86, 117).
- [156] Jen-Fan HANG et al. « Digital image analysis supports a nuclear-to-cytoplasmic ratio cutoff value of 0.5 for atypical urothelial cells ». en. In : *Cancer Cytopathology* 125.9 (sept. 2017), p. 710-716. DOI : [10.1002/cncy.21883](https://doi.org/10.1002/cncy.21883). (Visité le 19/05/2022) (cf. p. 85).
- [157] Patrick J. McINTIRE et al. « Digital image analysis supports a nuclear-to-cytoplasmic ratio cutoff value below 0.7 for positive for high-grade urothelial carcinoma and suspicious for high-grade urothelial carcinoma in urine cytology specimens ». en. In : *Cancer Cytopathology* 127.2 (fév. 2019), p. 120-124. DOI : [10.1002/cncy.22061](https://doi.org/10.1002/cncy.22061). (Visité le 19/05/2022) (cf. p. 85).
- [158] Mingjuan L. ZHANG, Alan X. GUO et Christopher J. VANDENBUSSCHE. « Morphologists overestimate the nuclear-to-cytoplasmic ratio : Overestimation of N :C Ratio ». en. In : *Cancer Cytopathology* 124.9 (sept. 2016), p. 669-677. DOI : [10.1002/cncy.21735](https://doi.org/10.1002/cncy.21735). (Visité le 19/05/2022) (cf. p. 85).
- [159] Lester J. LAYFIELD et al. « Accuracy and Reproducibility of Nuclear/Cytoplasmic Ratio Assessments in Urinary Cytology Specimens : REPRODUCIBILITY OF NUCLEAR/CYTOPLASMIC RATIO ESTIMATES ». en. In : *Diagnostic Cytopathology* 45.2 (fév. 2017), p. 107-112. DOI : [10.1002/dc.23639](https://doi.org/10.1002/dc.23639). (Visité le 19/05/2022) (cf. p. 85).
- [160] Jan JANTZEN et George DOUNIAS. « Analysis of Pap-Smear Image Data ». en. In : *Proceedings of the Nature-Inspired Smart Information Systems 2nd Annual Symposium*. 2006, p. 10 (cf. p. 86).
- [161] G W ZACK, W E ROGERS et S A LATT. « Automatic measurement of sister chromatid exchange frequency. » en. In : *Journal of Histochemistry & Cytochemistry* 25.7 (juil. 1977), p. 741-753. DOI : [10.1177/25.7.70454](https://doi.org/10.1177/25.7.70454). (Visité le 24/05/2022) (cf. p. 89, 118).

- [162] Edgar G FISCHER. « Nuclear Morphology and the Biology of Cancer Cells ». en. In : *Acta Cytologica* (2020), p. 9 (cf. p. 92).
- [163] Thomas G. DIETTERICH, Richard H. LATHROP et Tomás LOZANO-PÉREZ. « Solving the multiple instance problem with axis-parallel rectangles ». en. In : *Artificial Intelligence* 89.1-2 (jan. 1997), p. 31-71. DOI : [10.1016/S0004-3702\(96\)00034-3](https://doi.org/10.1016/S0004-3702(96)00034-3). (Visité le 19/05/2022) (cf. p. 92).
- [164] Bin LI, Yin LI et Kevin W. ELICEIRI. « Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning ». en. In : *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA : IEEE, juin 2021, p. 14313-14323. DOI : [10.1109/CVPR46437.2021.01409](https://doi.org/10.1109/CVPR46437.2021.01409). (Visité le 10/03/2022) (cf. p. 94, 96).
- [165] Zhuchen SHAO et al. « TransMIL : Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification ». en. In : *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*. Sydney, Australia, oct. 2021. (Visité le 31/05/2022) (cf. p. 94, 97, 98, 111, 113, 129, 131).
- [166] Ashish VASWANI et al. « Attention Is All You Need ». en. In : *31st Conference on Neural Information Processing Systems*. Déc. 2017. (Visité le 01/06/2022) (cf. p. 97).
- [167] Jimmy Lei BA, Jamie Ryan KIROS et Geoffrey E. HINTON. *Layer Normalization*. en. Number : arXiv :1607.06450 arXiv :1607.06450 [cs, stat]. Juil. 2016. URL : <http://arxiv.org/abs/1607.06450> (visité le 02/06/2022) (cf. p. 98).
- [168] Yunyang XIONG et al. « Nyströmformer : A Nyström-Based Algorithm for Approximating Self-Attention ». en. In : *The 35th AAAI Conference on Artificial Intelligence*. AAAI, mar. 2021. (Visité le 02/06/2022) (cf. p. 98, 111, 112).
- [169] Zhilu ZHANG et Mert SABUNCU. « Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels ». en. In : *32nd Conference on Neural Information Processing Systems*. Montréal, Canada, 2018, p. 11 (cf. p. 99).
- [170] Filip Karlo DOSILOVIC, Mario BRICIC et Nikica HLUPIC. « Explainable artificial intelligence : A survey ». In : *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. Opatija : IEEE, mai 2018, p. 0210-0215. DOI : [10.23919/MIPRO.2018.8400040](https://doi.org/10.23919/MIPRO.2018.8400040). (Visité le 15/06/2022) (cf. p. 104).
- [171] Leo BREIMAN. « Random Forests ». In : *Machine Learning* 45.1 (2001), p. 5-32. DOI : [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324). (Visité le 29/06/2022) (cf. p. 109, 124).
- [172] Fadi BRIMO et al. « Urine cytology : does the number of atypical urothelial cells matter ? A qualitative and quantitative study of 112 cases ». en. In : *Journal of the American Society of Cytopathology* 4.4 (juil. 2015), p. 232-238. DOI : [10.1016/j.jasc.2015.01.003](https://doi.org/10.1016/j.jasc.2015.01.003). (Visité le 09/06/2022) (cf. p. 112).
- [173] J KARNON et al. « Liquid-based cytology in cervical screening : an updated rapid and systematic review and economic analysis ». en. In : *Health Technology Assessment* 8.20 (mai 2004). DOI : [10.3310/hta8200](https://doi.org/10.3310/hta8200). (Visité le 17/06/2022) (cf. p. 116).
- [174] Fred R. DEE et al. « Utility of 2-D and 3-D Virtual Microscopy in Cervical Cytology Education and Testing ». en. In : *Acta Cytologica* 51.4 (2007), p. 523-529. DOI : [10.1159/000325788](https://doi.org/10.1159/000325788). (Visité le 17/06/2022) (cf. p. 116, 125).
- [175] David KIM et al. « Evaluating the role of Z-stack to improve the morphologic evaluation of urine cytology whole slide images for high-grade urothelial carcinoma : Results and review of a pilot study ». en. In : *Cancer Cytopathology* (mai 2022), cncy.22595. DOI : [10.1002/cncy.22595](https://doi.org/10.1002/cncy.22595). (Visité le 17/06/2022) (cf. p. 117, 125).
- [176] Petru MANESCU et al. « Content aware multi-focus image fusion for high-magnification blood film microscopy ». en. In : *Biomedical Optics Express* 13.2 (fév. 2022), p. 1005. DOI : [10.1364/B0E.448280](https://doi.org/10.1364/B0E.448280). (Visité le 08/07/2022) (cf. p. 130).
- [177] Wouter VAN GANSBEKE et al. « SCAN : Learning to Classify Images Without Labels ». en. In : *Computer Vision – ECCV 2020*. T. 12355. Series Title : Lecture Notes in Computer Science. Springer International Publishing, 2020, p. 268-285. DOI : [10.1007/978-3-030-58607-2_16](https://doi.org/10.1007/978-3-030-58607-2_16). (Visité le 06/07/2022) (cf. p. 131).

- [178] Olga RUSSAKOVSKY et al. « ImageNet Large Scale Visual Recognition Challenge ». en. In : *International journal of computer vision* 115.3 (jan. 2015), p. 211-252. (Visité le 06/07/2022) (cf. p. 131).
- [179] Ting CHEN et al. « A Simple Framework for Contrastive Learning of Visual Representations ». en. In : *International conference on machine learning*. Juin 2020. (Visité le 06/07/2022) (cf. p. 131).
- [180] Xinlei CHEN et al. *Improved Baselines with Momentum Contrastive Learning*. en. arXiv :2003.04297 [cs]. Mar. 2020. URL : <http://arxiv.org/abs/2003.04297> (visité le 06/07/2022) (cf. p. 131).
- [181] Xinlei CHEN et Kaiming HE. « Exploring Simple Siamese Representation Learning ». en. In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nov. 2020. (Visité le 06/07/2022) (cf. p. 131).
- [182] Jean-Bastien GRILL et al. « Bootstrap your own latent : A new approach to self-supervised Learning ». en. In : *Advances in neural information processing systems*. Sept. 2020. (Visité le 06/07/2022) (cf. p. 131).
- [183] Soroush Abbasi KOOHPAYEGANI, Ajinkya TEJANKAR et Hamed PIRSAVASH. « Mean Shift for Self-Supervised Learning ». en. In : *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada : IEEE, oct. 2021, p. 10306-10315. DOI : [10.1109/ICCV48922.2021.01016](https://doi.org/10.1109/ICCV48922.2021.01016). (Visité le 06/07/2022) (cf. p. 131).
- [184] Ozan CIGA, Tony XU et Anne Louise MARTEL. « Self supervised contrastive learning for digital histopathology ». en. In : *Machine Learning with Applications* 7 (mar. 2022), p. 100198. DOI : [10.1016/j.mlwa.2021.100198](https://doi.org/10.1016/j.mlwa.2021.100198). (Visité le 06/07/2022) (cf. p. 131).
- [185] Aiham TALEB et al. « ContIG : Self-supervised Multimodal Contrastive Learning for Medical Imaging with Genetics ». en. In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nov. 2021. (Visité le 06/07/2022) (cf. p. 132).
- [186] Hazha Omar OTHMAN et al. « A highly sensitive fluorescent immunosensor for sensitive detection of nuclear matrix protein 22 as biomarker for early stage diagnosis of bladder cancer ». en. In : *RSC Advances* 10.48 (2020), p. 28865-28871. DOI : [10.1039/D0RA06191C](https://doi.org/10.1039/D0RA06191C). (Visité le 07/07/2022) (cf. p. 132).
- [187] Joshua J. LEVY et al. *Uncovering Additional Predictors of Urothelial Carcinoma from Voided Urothelial Cell Clusters Through a Deep Learning Based Image Preprocessing Technique*. en. preprint. Pathology, mai 2022. DOI : [10.1101/2022.04.30.490136](https://doi.org/10.1101/2022.04.30.490136). (Visité le 05/07/2022) (cf. p. 132).

