

République Algérienne Démocratique et Populaire
Université Abderrahmane MIRA de Béjaia
Faculté des Sciences Exactes

Département de Recherche Opérationnelle



Mémoire Présenté pour L'obtention du Diplôme de Master
en Mathématiques Appliquées

Spécialité : Sciences de Données et Aide à la Décision

**Système d'aide au diagnostic médical basé sur l'apprentissage
profond pour la détection des mitoses dans les images
histopathologiques.**

Présenté par :
MESSAOUI BILLAL
IMCHAL LOUANES

Sous la direction de : Dr L. Asli
Et de : Dr A. BOUDRIOUA

Défendu le 29/06/2025, devant le jury composé de :

M ^r B Brahmi	M.C. classe/ A	Président de jury	UAMB - Bejaia.
M ^r A. Laouar	M.A. classe/ A	Examineur	UAMB - Bejaia.
M ^{elle} I. Tetah	Doctorante	Examinatrice	UAMB - Bejaia.

Année Universitaire 2024 – 2025

Remerciements

Tout d'abord, nous exprimons notre gratitude à Dieu Tout-Puissant pour nous avoir accordé le courage, la force, la patience et la santé nécessaires pour mener à bien ce modeste travail.

Nous tenons à adresser nos sincères remerciements à notre encadrant, M. **Larbi Asli**, pour sa disponibilité, ses orientations pertinentes et ses conseils avisés tout au long de ce projet.

Nous remercions également chaleureusement notre encadrante, Mme **A. Boudrioua**, pour sa rigueur, sa patience et son accompagnement constant, qui ont grandement contribué à la qualité de ce travail.

Nous souhaitons également exprimer notre gratitude envers les membres du jury qui ont accepté de juger notre travail.

Nous aimerions aussi nous remercier mutuellement pour le travail accompli en équipe, pour la collaboration, la persévérance et les efforts fournis tout au long de ce projet.

Enfin, nous tenons à remercier toutes les personnes qui, de près ou de loin, ont contribué à la réalisation de ce travail, par leurs conseils, leurs encouragements ou leur soutien moral.

L'équipe tient à exprimer sa reconnaissance envers tous ceux qui ont soutenu et participé à cette réalisation.

Cordialement,
Billal, Louanes



Dédicace

Je dédie ce travail

À mon père et à ma mère

À mes chères sœurs Meriem, Lamia, Aïda, et à mon frère Alaoua

Et à tous les membres de ma famille Messaoui et Mendi

Et plus encore...

Particulièrement à tous mes amis et collègues

À tous mes professeurs et à tous ceux qui m'ont appris

Tout au long de ma vie scolaire

Billal

Dédicace

Je dédie ce travail

À mon père et à ma mère, pour leur amour, leur soutien et leurs sacrifices.

À mes chères sœurs Lydia et Chanez, et à mon frère Nouari.

À son fils Mohamed, mon neveu adoré, qui apporte lumière et joie à notre famille.

À tous les membres de ma famille IMCHAL et AMIROUCHE.

À tous mes amis et collègues, pour leur présence et leur encouragement.

À mes professeurs, et à tous ceux qui m'ont transmis le savoir tout au long de ma vie scolaire.

Louanes

Table des matières

Remerciements	I
Dédicace	II
Dédicace	III
Liste des figures	VIII
Liste des tables	IX
Liste d'abréviations et notations	X
Introduction générale	1
1 Cadre conceptuel et terminologique	4
Introduction	4
1.1 La mitose et son rôle en pathologie	5
1.1.1 Définition de la mitose	5
1.1.2 Importance clinique des figures mitotiques	6
1.1.3 Problèmes liés à l'identification manuelle	7
1.2 La pathologie numérique	9
1.2.1 Du microscope à l'image numérique	9
1.2.2 La lame virtuelle (WSI – Whole Slide Image)	11
1.3 Introduction à la pathologie computationnelle	12
1.3.1 Motivation pour l'automatisation	12
1.3.2 Définition de la pathologie computationnelle	12
1.3.3 Enjeux spécifiques à la détection des mitoses	13
1.4 L'intelligence artificielle en imagerie médicale	14
1.4.1 Introduction à l'intelligence artificielle (IA)	14
1.4.2 L'apprentissage automatique (Machine Learning)	15
1.4.3 L'apprentissage profond (Deep Learning)	18
Conclusion	20
2 Revue de littérature et choix méthodologiques	21
Introduction	21
2.1 Méthodes de détection existantes	22
2.1.1 Détection des mitoses à l'aide de caractéristiques conçues manuellement	22
2.1.2 Transition vers les méthodes par deep learning	23
2.1.3 Détection des mitoses à l'aide de caractéristiques profondes	24
2.2 Revue de modèles modernes de détection d'objets	25
2.2.1 R-CNN et ses évolutions	26
2.2.2 RetinaNet	27

2.2.3	YOLO (You Only Look Once)	28
2.3	Justification de la démarche retenue : Choix du modèle RetinaNet	29
2.3.1	Une détection robuste dans les contextes de classes déséquilibrées : la Focal Loss	29
2.3.2	Une architecture hiérarchique pour détecter les objets de toutes tailles : le Feature Pyramid Network (FPN)	29
2.3.3	Un compromis pertinent entre précision et efficacité	30
2.3.4	Adaptabilité au contexte de la pathologie numérique	30
	Conclusion	31
3	Méthodologie et développement du modèle	32
	Introduction	32
3.1	Données utilisées	33
3.1.1	Présentation du jeu de données MIDOG 2022	33
3.1.2	Structure du dataset et diversité inter-domaines	34
3.1.3	Format et structure des annotations	34
3.1.4	Spécifications techniques des images	35
3.2	Pipeline général du système	36
3.3	Architecture du modèle – RetinaNet	37
3.3.1	Détails de l’architecture	37
3.3.2	Fonction de perte – Focal Loss	37
3.3.3	Backbone : ResNet-50	38
3.3.4	Feature Pyramid Network (FPN)	40
3.3.5	Génération des boîtes d’ancrage (Anchors)	41
3.3.6	Sous-réseaux de classification et de régression	42
3.3.7	Normalisation des images	42
3.4	Préparation des données	43
3.4.1	Découpage en patches	43
3.4.2	Normalisation des couleurs – Méthode de Macenko	44
3.4.3	Séparation des données : train / validation / test	45
3.5	Configuration de l’apprentissage	47
3.5.1	Hyperparamètres d’apprentissage	47
3.5.2	Stratégie d’optimisation	48
3.5.3	Augmentation dynamique des données	49
	Conclusion	50
4	Expérimentation,résultats et discussions	52
	Introduction	52
4.1	Environnement matériel et logiciel	53
4.1.1	Les logiciels	53
4.1.2	Configuration matérielle	55
4.2	Évaluation du modèle	55
4.2.1	Précision (Precision)	56
4.2.2	Rappel (Recall)	56
4.2.3	F1-score	56
4.2.4	mean Average Precision	57

4.2.5	Motivation du choix des métriques d'évaluation	57
4.3	Résultats métriques du modèle	57
4.4	Évaluation qualitative par visualisation des détections	61
4.5	Interprétabilité et visualisation	64
4.6	Interface utilisateur (prototypage)	67
4.7	Perspectives	70
	Conclusion	70
	Conclusion générale	71
	Bibliographie	77
	Résumé	78

Table des figures

1.1	Différentes phases de la mitose	5
1.2	Figures mitotiques observées au microscope dans une coupe histologique colorée à l'H&E.	7
1.3	Exemple de variation inter-observateurs dans le nombre de mitoses comptées sur la même coupe histologique.	8
1.4	Exemple d'un champ de vision avec une faible densité de mitoses, nécessitant une recherche minutieuse.	9
1.5	Exemples illustrant les défis de la classification des mitoses en histopathologie	9
1.6	Exemple de scanner (Aperio AT2) utilisé pour la numérisation de lames. [9]	10
1.7	Comparaison des scanners utilisés dans la base de données MIDOG 2022, illustrant la diversité des équipements et leur impact potentiel sur la qualité des WSI.[8]	11
1.8	Exemple de lame virtuelle ouverte dans QuPath	12
1.9	Processus de diagnostic assisté par ordinateur	13
1.10	Les relations entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.	15
1.11	Principe de classification par SVM	16
1.12	Exemple d'arbre de décision binaire	16
1.13	Principe d'une forêt aléatoire	17
1.14	Architecture d'un MLP à deux couches cachées[18].	17
1.15	Neurone biologique.	19
1.16	Architecture d'un réseau de neurone convolutif.	19
1.17	Schéma des modèles de détection et classification par CNN	20
2.1	architecture de model Faster R-CNN [15].	27
2.2	: Architecture de RetinaNet [53].	27
2.3	Architecture de model YOLO [56].	28
3.1	Exemples d'images histopathologiques issues du dataset MIDOG 2022.	33
3.3	Exemple d'annotation COCO (extrait JSON affichant une bbox avec category_id).	35
3.4	Diagramme circulaire illustrant la proportion de mitoses (rouge) et de négatifs durs (bleu) dans un échantillon.	35
3.5	Pipeline général du système de détection de mitoses.	36
3.6	Architecture globale de RetinaNet avec backbone ResNet-50, pyramide FPN et sous-réseaux prédictifs[34].	37
3.7	Courbe de la Focal Loss comparée à la Cross-Entropy pour différentes valeurs de γ [34].	38
3.8	Bloc bottleneck utilisé dans ResNet-50.	39

3.9	Correspondance entre niveaux de ResNet et entrées FPN[57].	39
3.10	Schéma détaillé du bloc FPN avec upsampling + addition + conv 3×3 [61]. . .	40
3.11	Visualisation des ancres sur différents niveaux de la pyramide (P3 à P7)[58]. . .	41
3.12	Pipeline de découpage d'une WSI en patches chevauchants avec intégration des annotations mitotiques.	44
3.13	Effet visuel de la normalisation Macenko	46
3.14	Répartition des patches annotés et non annotés dans les ensembles d'entraînement, de validation et de test à partir des lames histopathologiques (WSI)	47
3.15	Illustration de l'effet de l'augmentation dynamique.	50
3.16	Courbe de loss et mAP sur TensorBoard.	50
4.1	Évolution du Train Loss et du mAP au cours des époques.	58
4.2	Graphique des métriques de performance (précision, rappel, F1-score) par classe.	58
4.3	Évaluation des Performances par Classe : Analyse des TP/FP/FN et des Métriques de Précision, Rappel et F1-Score.	59
4.5	Exemples de True Positives (mitoses bien détectées).	61
4.6	Exemples d'erreurs de détection : faux positifs (objets confondus avec des mitoses) et faux négatifs (mitoses non détectées)	62
4.7	Carte annotée des motifs d'erreur fréquents.	63
4.8	Comparaison des prédictions du modèle et des annotations expert	63
4.9	visualisation des prédictions RetinaNet sur une lame complète	64
4.10	Carte de chaleur montrant la densité des prédictions mitotiques sur une lame entière (les régions rouges indiquent une concentration élevée de détections). . .	65
4.11	Exemple de patch avec surimpression d'une heatmap basée sur les scores de détection (plus le rouge est intense, plus la confiance est élevée).	66
4.12	Écran d'accueil de l'application.	67
4.13	Interface principale avec affichage des prédictions, informations image, et navigation par patches détectés.	68
4.14	Interface de visualisation des zones détectées (mitoses et hard negatives). . . .	69
4.15	Interface d'exploration et d'export des résultats de détection.	69

Liste des tableaux

1.1	Système de score histologique selon trois critères	6
1.2	Correspondance entre le score total et le grade histologique final	7
2.1	Comparaison des méthodes de détection de mitoses basées sur des caractéristiques conçues manuellement.	23
2.2	Comparaison des méthodes de détection de mitoses basées sur des caractéristiques profondes.	25
2.3	Comparaison des modèles de détection d'objets.	28
3.1	Informations sur les fichiers selon le type de cancer, espèce, centre médical et scanner utilisé.	34
3.2	Hyperparamètres du mécanisme d'arrêt anticipé (early stopping)	48
3.3	Transformations d'augmentation et de prétraitement appliquées aux images histopathologiques	49
4.1	Configuration matérielle utilisée pour les expérimentations	55
4.2	Scores de détection pour chaque classe : mitoses et faux positifs (hard negatives).	58
4.3	Nombre de vrais positifs (TP), faux positifs (FP) et faux négatifs (FN) par classe.	59
4.4	Principaux motifs d'erreurs rencontrés dans la détection automatique des mitoses.	62

Liste d'abréviations et notations

Abréviations	Signification
AI / IA	Intelligence Artificielle (Artificial Intelligence)
AP	Average Precision
AUC	Area Under the Curve
BBox	Bounding Box (Boîte Englobante)
CNN	Convolutional Neural Network
COCO	Common Objects in Context (format d'annotations)
CSV	Comma-Separated Values
DL	Deep Learning
FC	Fully Connected
FCM	Fuzzy C-Means
F1	F1-score (moyenne harmonique de la précision et du rappel)
FN	False Negative
FP	False Positive
FPN	Feature Pyramid Network
GLCM	Gray-Level Co-occurrence Matrix
Grad-CAM	Gradient-weighted Class Activation Mapping
GPU	Graphics Processing Unit
H&E	Hématoxyline et Éosine
HPF	High Power Field
IoU	Intersection over Union
JSON	JavaScript Object Notation
LBP	Local Binary Pattern
MI	Index Mitotique
ML	Machine Learning
MLP	Multi-Layer Perceptron
mAP	Mean Average Precision
PR	Precision-Recall
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RetinaNet	Réseau de détection d'objets à ancres focalisés
R-CNN	Region-based Convolutional Neural Network
ROI	Region of Interest
SVM	Support Vector Machine
TP	True Positive
VOC	Visual Object Classes

WSI	Whole Slide Image
YOLO	You Only Look Once

Introduction générale

La médecine contemporaine fait face à une explosion des données médicales, rendant nécessaire le recours à des outils diagnostiques plus rapides, plus standardisés et plus fiables. En oncologie, le comptage des figures mitotiques dans les images histopathologiques constitue un critère essentiel dans l'évaluation du grade tumoral. Ce paramètre permet de mesurer l'activité proliférative des cellules tumorales et influence directement les décisions thérapeutiques. Cependant, cette tâche demeure largement manuelle, chronophage et sujette à une forte variabilité inter-observateurs, même parmi des pathologistes expérimentés.

Au fil des années, plusieurs approches ont été proposées pour automatiser cette tâche[23, 25, 26, 29]. Les premières méthodes reposaient sur des techniques classiques de traitement d'images, telles que le seuillage, la détection de contours ou encore l'analyse morphologique. Bien que simples à mettre en œuvre, ces approches se sont révélées insuffisamment robustes face à la variabilité et à la complexité des images histopathologiques.

Dans un second temps, des méthodes hybrides ont tenté de combiner des caractéristiques manuelles avec des descripteurs extraits automatiquement à l'aide de réseaux de neurones convolutifs (CNN)[36, 37, 38, 39, 40]. Ces approches, qui utilisaient les CNN comme simples extracteurs de caractéristiques, ont permis d'améliorer les performances par rapport aux méthodes classiques. Toutefois, elles restaient limitées par des coûts computationnels élevés et une complexité qui freinait leur intégration en milieu clinique.

Plus récemment, l'émergence de l'apprentissage profond (deep learning) a permis de franchir un nouveau cap, grâce à des modèles entraînés de bout en bout. Ces réseaux de neurones convolutifs (CNN), désormais optimisés directement pour la tâche de détection, se sont imposés comme des outils puissants en analyse d'images médicales, capables d'apprendre automatiquement des représentations discriminantes, sans intervention manuelle.

Parmi les architectures de détection d'objets les plus performantes, le modèle **RetinaNet** se distingue par sa capacité à traiter efficacement des objets rares et de petite taille, comme les figures mitotiques, grâce à sa combinaison d'un réseau pyramidal de caractéristiques (*Feature Pyramid Network* – FPN) et d'une fonction de perte adaptée au déséquilibre des classes (*Focal Loss*). Ces propriétés en font une architecture particulièrement bien adaptée au contexte de la pathologie numérique.

Malgré ces avancées, plusieurs défis subsistent : la rareté des mitoses dans les lames histologiques, la diversité morphologique des cellules tumorales, ainsi que les exigences cliniques en matière de fiabilité, de reproductibilité et d'interprétabilité des résultats. Il est donc nécessaire

de concevoir des solutions intelligentes, capables de s'intégrer dans des flux de travail médicaux tout en respectant les contraintes opérationnelles et cliniques.

C'est dans cette perspective que s'inscrit le présent mémoire, qui vise à concevoir et évaluer un système de détection automatique des figures mitotiques basé sur le modèle **RetinaNet**. L'objectif principal est de démontrer l'intérêt de l'apprentissage profond pour améliorer la rapidité, la précision et la reproductibilité du diagnostic histopathologique, tout en analysant les limites et les perspectives d'intégration de ces outils dans la pratique médicale réelle.

Ce travail cherche à répondre à la problématique suivante :

Comment concevoir une méthode efficace, fiable et automatisée pour détecter les mitoses dans les images histopathologiques, malgré la variabilité des tissus et la ressemblance entre mitoses et autres structures cellulaires ?

Cette problématique soulève plusieurs questions spécifiques :

- Comment rendre un modèle robuste face aux variations d'échantillons et à la diversité morphologique des tissus tumoraux ?
- Quelle architecture d'apprentissage profond permet d'optimiser la précision, le rappel et la généralisation du modèle ?
- Comment améliorer l'interprétabilité du modèle pour renforcer la confiance des pathologistes dans ses décisions ?
- Quelles sont les conditions nécessaires pour intégrer efficacement ce type de système dans un environnement clinique contraint ?

Afin d'apporter des éléments de réponse à ces enjeux, ce mémoire propose :

- Une revue critique de la littérature scientifique sur les méthodes de détection des mitoses, classiques et modernes.
- La mise en œuvre d'un pipeline de détection automatique basé sur RetinaNet, appliqué au jeu de données MIDOG 2022.
- Une évaluation quantitative et qualitative des performances du modèle développé.
- Des recommandations concrètes pour une intégration clinique, en tenant compte des exigences de fiabilité et d'interprétabilité.

Organisation du mémoire

Ce mémoire débute par une introduction générale qui présente le contexte, la problématique et les objectifs du travail.

Il est ensuite structuré en quatre chapitres :

- Le premier chapitre introduit le contexte médical et les enjeux liés à la détection des mitoses en pathologie.
- Le deuxième chapitre présente les approches classiques et modernes, et justifie le choix de RetinaNet.
- Le troisième chapitre décrit la méthodologie expérimentale appliquée au dataset MIDOG 2022.

- Le quatrième chapitre évalue les résultats obtenus et propose des perspectives d'amélioration du modèle ainsi qu'une discussion sur ses limites.

Enfin, une conclusion générale résume les apports du mémoire et ouvre sur des perspectives futures.

1

Cadre conceptuel et terminologique

Introduction

Ce chapitre introduit l'importance des mitoses en pathologie, considérées comme un indicateur clé de l'activité tumorale, ainsi que les limites de leur détection manuelle, souvent subjective, complexe et chronophage. Il présente les avancées de la médecine numérique, en particulier l'apport de l'intelligence artificielle et du deep learning, qui permettent d'automatiser et d'optimiser l'analyse des images histopathologiques. Le chapitre est structuré en deux parties : la première aborde les enjeux liés à la détection des mitoses, et la seconde explore les approches computationnelles basées sur l'IA appliquées à la pathologie.

Sommaire

Introduction	4
1.1 La mitose et son rôle en pathologie	5
1.2 La pathologie numérique	9
1.3 Introduction à la pathologie computationnelle	12
1.4 L'intelligence artificielle en imagerie médicale	14
Conclusion	20

1.1 La mitose et son rôle en pathologie

1.1.1 Définition de la mitose

La mitose est un processus fondamental du cycle cellulaire au cours duquel une cellule eucaryote se divise pour produire deux cellules filles génétiquement identiques. Ce mécanisme est indispensable au renouvellement des tissus, à la croissance de l'organisme et à la réparation cellulaire. Toutefois, dans un contexte pathologique, notamment tumoral, une activité mitotique excessive ou anormale peut indiquer une prolifération cellulaire désordonnée, caractéristique des cancers.

Les figures mitotiques évoluent à travers un continuum de quatre à cinq phases distinctes, pendant lesquelles le noyau cellulaire subit des transformations morphologiques importantes. Ces phases sont : prophase, métaphase, anaphase, et télophase (la prométaphase étant parfois distinguée comme une étape intermédiaire entre la prophase et la métaphase). Chacune de ces phases est identifiable par une morphologie nucléaire spécifique, principalement déterminée par la forme et la distribution des chromosomes, souvent colorés en violet sur les images histopathologiques H&E (voir la figure 1.1).

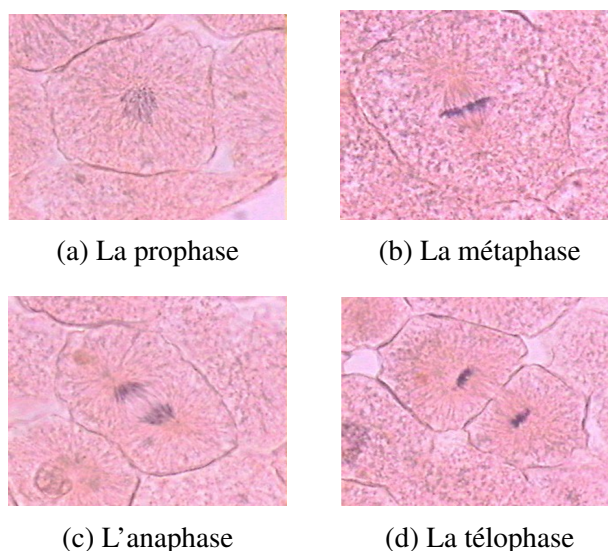


FIGURE 1.1 – Différentes phases de la mitose

- La prophase :** Il s'agit de la première phase de la mitose, au cours de laquelle la chromatine se condense pour former des chromosomes visibles et organisés. Les centrosomes commencent à migrer vers les pôles de la cellule, initiant la formation du fuseau mitotique.
- La métaphase :** Les chromosomes condensés s'alignent au centre de la cellule, sur la plaque équatoriale. Cette disposition symétrique est cruciale pour assurer une répartition équitable du matériel génétique.

- c) **L'anaphase** : Cette phase débute avec la séparation des chromatides sœurs, qui sont ensuite tirées vers les pôles opposés de la cellule par les microtubules du fuseau mitotique. Cela garantit que chaque cellule fille recevra une copie exacte du génome.
- d) **La télophase** : Les noyaux se reforment autour des groupes de chromosomes à chaque pôle. Les chromosomes commencent à se décondenser, les enveloppes nucléaires se reforment et la cellule se prépare à se diviser totalement via la cytokinèse.

Il est important de noter que la détection visuelle des figures mitotiques dans les images histologiques est parfois rendue difficile par des artefacts, des flous de mise au point, ou encore la présence de mitoses atypiques (tripolaires, multipolaires, etc.), qui compliquent leur identification. De plus, des stades tardifs de la mitose, comme l'anaphase ou la télophase, peuvent apparaître sous forme de deux noyaux condensés proches, souvent interprétés comme une seule entité.

1.1.2 Importance clinique des figures mitotiques

Les figures mitotiques jouent un rôle crucial dans l'évaluation du comportement biologique des tumeurs. En pathologie, le comptage des mitoses sur les coupes histologiques constitue une méthode standard pour estimer l'activité proliférative d'une tumeur. Cette évaluation aide à poser un diagnostic, à déterminer le grade tumoral, et à prédire l'agressivité de la maladie.

Dans de nombreux cancers, notamment les carcinomes, le nombre de figures mitotiques observées dans des zones représentatives de la tumeur est utilisé pour évaluer son degré de différenciation. Cette donnée est intégrée dans des systèmes de grading histopronostique comme le système Scarff-Bloom-Richardson (SBR) modifié, aussi appelé classification de Nottingham, particulièrement utilisé dans le cas des cancers du sein. Ce système repose sur trois critères notés chacun de 1 à 3 : la formation tubulaire, l'atypie nucléaire, et le nombre de mitoses (voir la Table 1.1). Le total obtenu permet de classer la tumeur en un grade histologique (voir la Table 1.2) [1].

TABLE 1.1 – Système de score histologique selon trois critères

Critère	Score	Description
3*Formation glandulaire	1	Plus de 75 % de la tumeur forme des glandes
	2	Entre 10 % et 75 % de la tumeur forme des glandes
	3	Moins de 10 % de la tumeur forme des glandes
3*Atypies nucléaires	1	Noyaux petits, réguliers et uniformes
	2	Augmentation modérée de la taille et de la variabilité
	3	Variation marquée de la taille et de la forme
3*Nombre de mitoses	1	Moins de 11 mitoses par 10 champs à fort grossissement (HPF)
	2	Entre 11 et 20 mitoses par 10 HPF
	3	Plus de 20 mitoses par 10 HPF

TABLE 1.2 – Correspondance entre le score total et le grade histologique final

Grade final	Score total
Grade I (bien différencié)	3 à 5
Grade II (modérément différencié)	6 à 7
Grade III (peu différencié)	8 à 9

Un nombre élevé de mitoses conduit donc à un score mitotique plus élevé, ce qui peut faire basculer le grade vers un niveau supérieur. Ce grading a un impact clinique direct, car il influence les décisions thérapeutiques, par exemple le recours à une chimiothérapie adjuvante dans le cancer du sein [2].

L'activité mitotique est également reconnue comme un marqueur pronostic fiable. Elle est généralement corrélée à une plus grande agressivité tumorale, une croissance rapide, un potentiel métastatique accru et un moins bon pronostic [3]. À l'inverse, une faible activité mitotique est souvent associée à un comportement tumoral plus indolent. Il est cependant important de noter que cette évaluation peut être influencée par plusieurs facteurs techniques : la qualité de la coupe histologique, la méthode de coloration (souvent H&E) (voir la figure 1.2), ou encore l'expérience et la subjectivité du pathologiste [4].

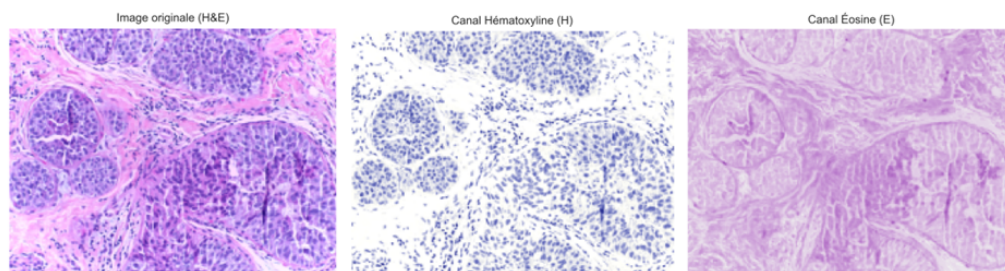


FIGURE 1.2 – Figures mitotiques observées au microscope dans une coupe histologique colorée à l'H&E.

1.1.3 Problèmes liés à l'identification manuelle

Variabilité inter-observateurs

Le comptage manuel des figures mitotiques est une tâche subjective qui dépend fortement de l'expérience, de la formation et de la perception individuelle du pathologiste. En l'absence de critères strictement quantifiables, l'interprétation morphologique des figures mitotiques varie d'un expert à un autre (voir la figure 1.3), ce qui entraîne une hétérogénéité des résultats. Plusieurs études ont démontré une variabilité inter-observateurs importante, avec des écarts parfois significatifs dans le nombre de mitoses identifiées pour une même lame histologique [5]. Ces divergences sont particulièrement marquées dans les cas limites, où les figures mitotiques sont

peu nombreuses, ambiguës ou partiellement visibles.

Cette variabilité a des conséquences cliniques non négligeables. Elle peut conduire à une discordance dans le grading tumoral, modifiant potentiellement la classification pronostique d'un patient, voire les décisions thérapeutiques associées. Par exemple, un grade histologique basé sur un nombre de mitoses sous-estimé pourrait conduire à un traitement moins agressif, alors qu'un surcomptage pourrait entraîner une chimiothérapie inutile.

L'introduction de grilles de lecture standardisées et la formation continue ont permis de réduire partiellement cette variabilité. Toutefois, même avec des recommandations comme celles de l'OMS ou d'organismes spécialisés, le manque de reproductibilité reste une limite majeure du comptage manuel dans la pratique courante.

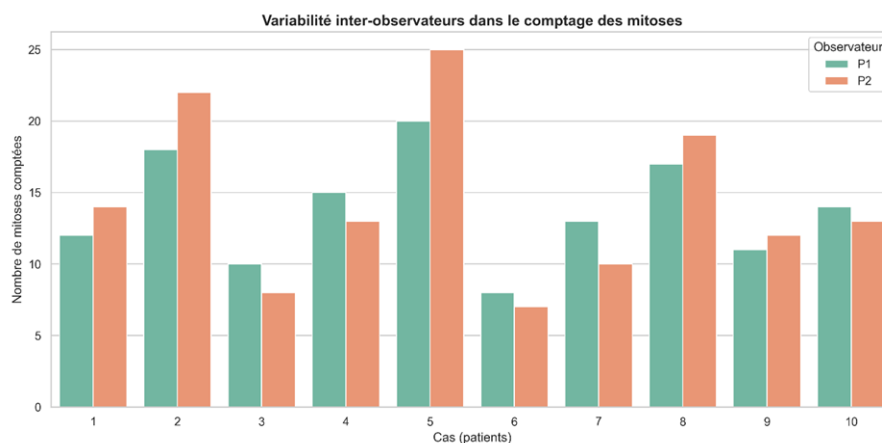


FIGURE 1.3 – Exemple de variation inter-observateurs dans le nombre de mitoses comptées sur la même coupe histologique.

Faible fréquence des mitoses

Les mitoses représentent un phénomène rare et localisé dans les tissus tumoraux. Leur faible fréquence rend leur identification difficile, nécessitant l'examen minutieux de grandes surfaces de tissu sous le microscope (voir la figure 1.4). Cette tâche est non seulement fastidieuse mais également chronophage [6]. En pratique, les pathologistes se concentrent souvent sur des zones dites "hot spots", c'est-à-dire des régions à forte densité mitotique. Cependant, cette approche ciblée peut ne pas être représentative de l'ensemble de la tumeur, ce qui pose un problème d'échantillonnage et de biais de mesure.

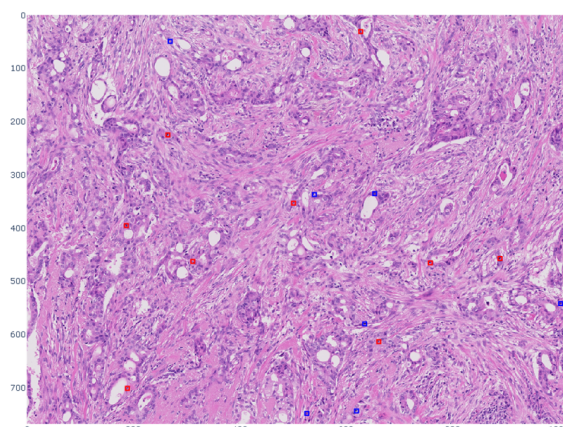


FIGURE 1.4 – Exemple d'un champ de vision avec une faible densité de mitoses, nécessitant une recherche minutieuse.

Difficultés visuelles (bruit, artefacts, apoptose...)

La reconnaissance visuelle des figures mitotiques est également entravée par la présence de nombreux éléments perturbateurs dans les images histologiques (voir la figure 1.5). Les artefacts de préparation (replis tissulaires, bulles, variations de coloration), le bruit de fond, ou encore les cellules en apoptose peuvent imiter la morphologie des mitoses, en particulier dans les phases comme la métaphase ou l'anaphase [7]. Ces similitudes augmentent le risque d'erreurs de classification, aussi bien de faux positifs (confondre une apoptose avec une mitose) que de faux négatifs (ne pas détecter une mitose réelle). Cette complexité visuelle justifie la nécessité de développer des outils d'aide à la décision robustes et reproductibles, notamment basés sur l'intelligence artificielle.

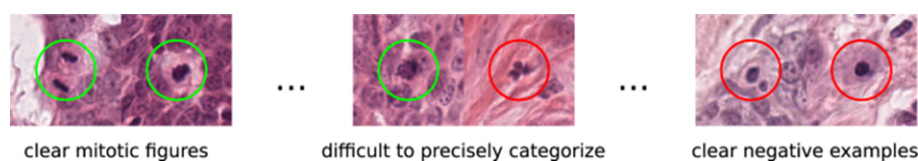


FIGURE 1.5 – Exemples illustrant les défis de la classification des mitoses en histopathologie

1.2 La pathologie numérique

1.2.1 Du microscope à l'image numérique

L'analyse histopathologique traditionnelle repose sur l'observation de coupes tissulaires colorées le plus souvent à l'hématoxyline-éosine (H&E) à l'aide d'un microscope optique. Cette méthode, bien qu'efficace, présente plusieurs limites : subjectivité liée à l'interprétation humaine, difficulté de partage entre pathologistes, impossibilité d'archivage numérique à long

terme et absence d'intégration d'outils d'intelligence artificielle (IA).

Face à ces contraintes, la pathologie connaît depuis plus d'une décennie une transformation majeure avec l'émergence de la pathologie numérique (voir la figure 1.6). Celle-ci repose sur la numérisation des lames histologiques par des scanners optiques haute résolution, capables de générer des WSI (Whole Slide Images) des images numériques intégrales des lames. Ces images peuvent être visualisées avec un zoom interactif, reproduisant l'expérience du microscope traditionnel tout en apportant des bénéfices supplémentaires :

- Archivage numérique sécurisé à long terme.
- Partage instantané entre professionnels (télépathologie).
- Annotation collaborative pour l'enseignement et la recherche.
- Intégration d'algorithmes d'IA pour l'analyse automatisée.
- Réduction de la variabilité inter-observateurs [8].



FIGURE 1.6 – Exemple de scanner (Aperio AT2) utilisé pour la numérisation de lames. [9]

Exemples de scanners de lames histopathologiques

Plusieurs fabricants proposent des scanners utilisés dans les laboratoires à travers le monde (voir la figure 1.7). Parmi les plus courants :

- Hamamatsu XR NanoZoomer 2.0-HT.
- Aperio AT2 (Leica Biosystems).
- Leica GT450.
- 3DHISTECH P1000.

Chaque modèle se distingue par ses spécificités techniques (résolution optique, vitesse de scan, profondeur de champ, colorimétrie...). Cette diversité constitue un obstacle à la standardisation des images numériques. Les algorithmes de traitement d'images doivent ainsi être capables de s'adapter à des entrées hétérogènes [10].

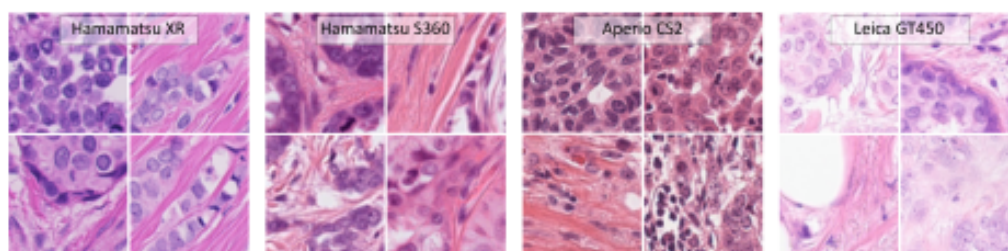


FIGURE 1.7 – Comparaison des scanners utilisés dans la base de données MIDOG 2022, illustrant la diversité des équipements et leur impact potentiel sur la qualité des WSI.[8]

1.2.2 La lame virtuelle (WSI – Whole Slide Image)

La lame virtuelle, ou Whole Slide Image (WSI), désigne une image numérique haute résolution d'une lame histologique complète, obtenue grâce à un scanner spécialisé (voir la figure 1.8). Contrairement aux simples images de microscopie optique prises sur des zones limitées, les WSI permettent une exploration interactive de la totalité de la lame, avec des niveaux de zoom équivalents à ceux d'un microscope traditionnel [11].

Ce type d'image est devenu un élément central de la pathologie numérique, car il permet non seulement l'analyse à distance (télépathologie), mais aussi l'utilisation d'outils d'annotation, d'archivage et surtout d'intelligence artificielle.

Format, taille et résolution

Les WSI sont stockées dans des formats spécifiques tels que .svs, .ndpi, .mrxs ou .tiff, selon le fabricant du scanner. Ces fichiers peuvent atteindre des tailles de plusieurs gigaoctets, car ils contiennent une grande profondeur d'image et plusieurs niveaux de résolution (pyramides multi-échelles). Cela permet de zoomer dynamiquement dans l'image, à la manière des systèmes de cartographie interactifs [11].

La résolution standard d'une WSI en diagnostic est souvent de 40x (soit 0,25 $\mu\text{m}/\text{pixel}$), ce qui permet de visualiser les détails subcellulaires comme les noyaux ou les figures mitotiques.

Problèmes techniques

Malgré leurs avantages, les WSI posent plusieurs défis technologiques :

- Taille des fichiers : Les images peuvent dépasser plusieurs Go, nécessitant un stockage performant.
- Zoom et navigation : Un affichage fluide requiert un logiciel optimisé pour le streaming d'image (ex. : QuPath, SlideViewer, ASAP).
- Formats propriétaires : Chaque fabricant ayant son propre format, cela peut poser des problèmes d'interopérabilité dans les flux multi-institutionnels [10].

- Temps de chargement et latences en cas d'accès réseau ou cloud.

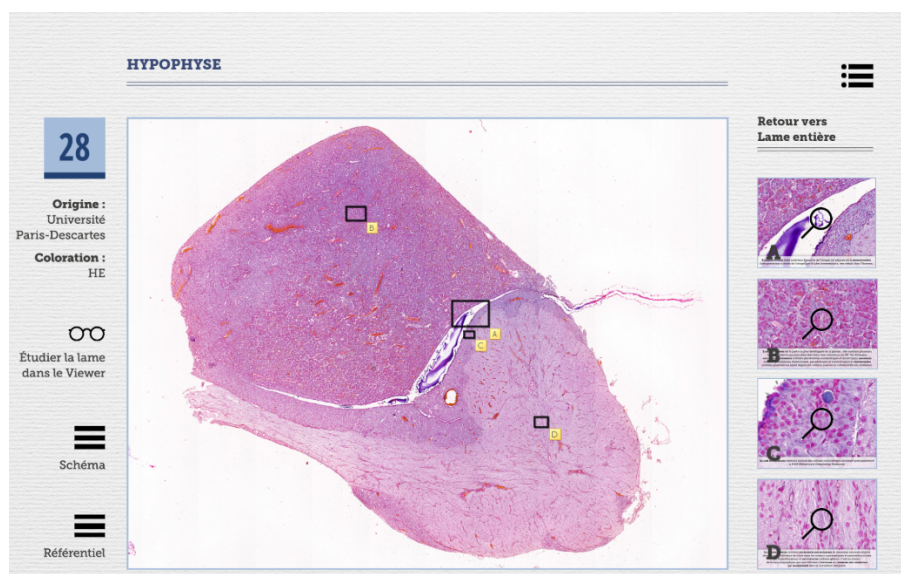


FIGURE 1.8 – Exemple de lame virtuelle ouverte dans QuPath

1.3 Introduction à la pathologie computationnelle

1.3.1 Motivation pour l'automatisation

Le diagnostic histopathologique repose encore aujourd'hui sur une évaluation manuelle des lames par les pathologistes. Bien que cette expertise humaine soit indispensable, elle présente plusieurs limites majeures :

- Une charge de travail croissante due à l'augmentation du nombre de cas à traiter dans les services d'anatomopathologie .
- Une variabilité inter- et intra-observateur dans l'interprétation des images, en particulier pour les tâches complexes comme le comptage mitotique [12] .
- Une nécessité de réduction des délais diagnostiques, notamment en oncologie, où des décisions thérapeutiques rapides sont cruciales.

Dans ce contexte, l'automatisation de certaines tâches diagnostiques à l'aide d'outils computationnels permettrait de soulager les pathologistes, fiabiliser les analyses et standardiser les résultats, tout en conservant l'expertise humaine dans la boucle décisionnelle.

1.3.2 Définition de la pathologie computationnelle

La pathologie computationnelle est un domaine interdisciplinaire qui combine des techniques d'intelligence artificielle (IA), de traitement d'image et de bioinformatique pour analyser automatiquement les images histopathologiques numériques (voir la figure 1.9), en particulier les

WSI (Whole Slide Images).

- Le comptage de mitoses.
- L'évaluation de la densité tumorale.
- La classification des tissus.
- La prédiction de sous-types ou de pronostics.

Ces méthodes s'appuient généralement sur des pipelines d'analyse automatique, composés d'étapes successives : prétraitement de l'image, détection des régions d'intérêt, classification, quantification et interprétation.

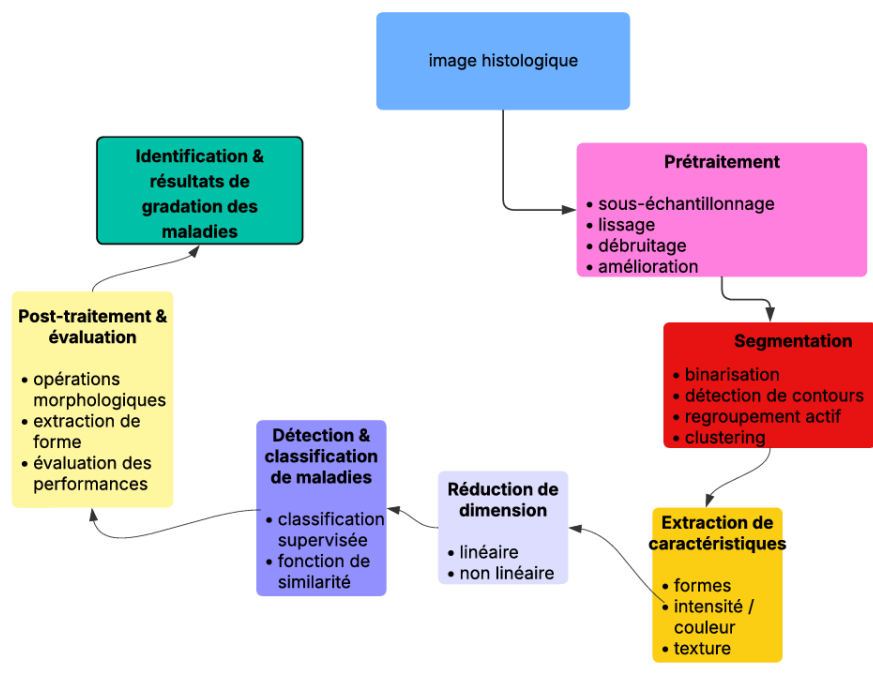


FIGURE 1.9 – Processus de diagnostic assisté par ordinateur

1.3.3 Enjeux spécifiques à la détection des mitoses

L'un des enjeux majeurs en pathologie numérique concerne le comptage mitotique, un indicateur clé de l'agressivité tumorale. L'index mitotique (MI), défini comme le nombre de cellules en division observées sur dix champs de haute puissance (HPF), est un facteur pronostique important.

Cependant, l'estimation manuelle de cet index reste une tâche longue, sujette à erreurs, et marquée par une forte variabilité inter-observateur.

Face à ces limites, l'analyse automatisée d'images médicales représente une opportunité majeure : elle pourrait accélérer les diagnostics, améliorer leur fiabilité, et réduire la subjectivité de l'évaluation.

Les réseaux de neurones convolutifs (CNN) se sont imposés comme des outils de référence

pour détecter automatiquement les figures mitotiques, malgré un environnement visuel souvent complexe, bruité, et hétérogène.

Néanmoins, plusieurs défis subsistent :

- Les mitoses sont des objets rares, difficilement détectables sur des WSI très grandes .
- Leur morphologie varie en fonction du tissu, du scanner, ou de la phase mitotique .
- Le bruit visuel, les artefacts, ou la présence de cellules apoptotiques peuvent induire des confusions, même chez un expert.

Ces contraintes expliquent pourquoi la détection mitotique constitue un benchmark rigoureux pour l'évaluation des performances des modèles d'IA en pathologie. Des bases de données telles que MIDOG 2022 ont été développées précisément pour ce type de tâche, en proposant des images multi-scanners annotées manuellement [8].

L'objectif final est de développer un système robuste et généralisable, capable d'assister efficacement les pathologistes dans l'identification précise des mitoses.

1.4 L'intelligence artificielle en imagerie médicale

1.4.1 Introduction à l'intelligence artificielle (IA)

L'intelligence artificielle (IA) est un domaine de l'informatique qui vise à développer des systèmes capables de simuler certains aspects du comportement humain, comme la prise de décision, le raisonnement logique, la reconnaissance visuelle ou auditive, ou encore l'apprentissage à partir d'expériences. D'après Chollet (2021), l'IA consiste à rendre les machines capables de s'adapter de manière autonome à des situations nouvelles, en optimisant leur comportement en fonction d'un objectif [13].

Historiquement, l'IA est née dans les années 1950, portée par des pionniers comme Alan Turing, John McCarthy, ou encore Marvin Minsky. L'idée initiale était d'imiter l'intelligence humaine à travers des règles formelles : on parlait alors d'IA symbolique. Cette première approche consistait à programmer manuellement des règles logiques permettant de résoudre des problèmes bien définis, comme jouer aux échecs, résoudre des équations, ou diagnostiquer des maladies simples via des systèmes experts [14].

L'IA symbolique a dominé la recherche pendant plusieurs décennies, mais elle montrait ses limites face à des environnements complexes, où les règles sont trop nombreuses ou difficiles à formuler explicitement. C'est dans ce contexte qu'est apparue l'intelligence artificielle statistique, incarnée par le machine learning. Plutôt que de tout programmer, le ML repose sur l'idée de laisser l'algorithme apprendre à partir des données, en construisant ses propres modèles pour généraliser à de nouveaux cas.

Plus récemment, l'avènement de l'apprentissage profond (deep learning) a révolutionné l'IA. Grâce à des avancées en puissance de calcul (GPU), à la disponibilité de grandes bases

de données, et à des architectures neuronales complexes (comme les CNN ou LSTM), le deep learning a permis d'atteindre des performances sans précédent dans des domaines comme la reconnaissance d'images, la traduction automatique, ou la médecine personnalisée [15].

Ainsi, l'IA contemporaine regroupe plusieurs sous-domaines articulés les uns aux autres. La figure suivante illustre la hiérarchie entre IA, machine learning, et deep learning (voir la figure 1.10).

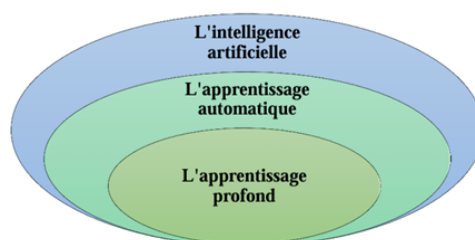


FIGURE 1.10 – Les relations entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

1.4.2 L'apprentissage automatique (Machine Learning)

L'apprentissage automatique, ou machine learning (ML), est un sous-domaine de l'intelligence artificielle qui consiste à permettre à un système informatique d'apprendre automatiquement à partir de données, sans être explicitement programmé pour chaque tâche. Plutôt que de suivre une suite de règles codées à la main, l'algorithme construit son propre modèle à partir d'exemples, qu'il utilise ensuite pour faire des prédictions ou prendre des décisions sur de nouvelles données [13].

Le machine learning repose sur le principe fondamental de la généralisation : un bon modèle doit non seulement s'adapter aux données d'entraînement, mais aussi être capable de traiter des cas inconnus. On distingue généralement quatre grandes familles d'approches selon la nature des données disponibles : supervisé, semi supervisé, non supervisé, et par renforcement.

1. Apprentissage supervisé :

C'est la forme la plus courante d'apprentissage automatique. Le système apprend à partir de données annotées, c'est-à-dire d'un ensemble d'entrées pour lesquelles la sortie (ou "label") correcte est connue. L'objectif est de trouver une fonction de prédiction $f(x)=y$ à partir d'un grand nombre d'exemples (x_i, y_i) [16].

■ Principaux algorithmes :

● Support Vector Machines (SVM) :

Ces modèles cherchent à séparer les classes par un hyperplan optimal, maximisant la marge entre les exemples des deux classes. Ils sont efficaces sur des données peu bruitées et sont souvent utilisés en imagerie médicale pour des tâches de classification binaire [17].

Ce principe est illustré en (Figure 1.11), où l'on observe la séparation optimale entre deux

classes par un hyperplan, entouré de vecteurs supports.

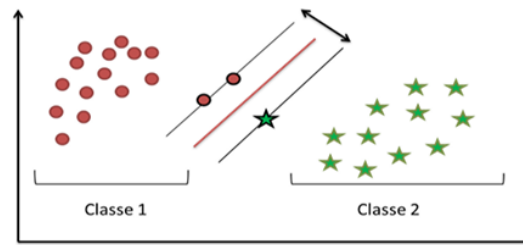


FIGURE 1.11 – Principe de classification par SVM

- **Arbres de décision :**

Méthodes basées sur une structure arborescente où chaque nœud teste une condition sur une variable. Faciles à interpréter mais sensibles au surapprentissage. une représentation graphique typique d'un arbre de décision est donnée en (Figure 1.12).

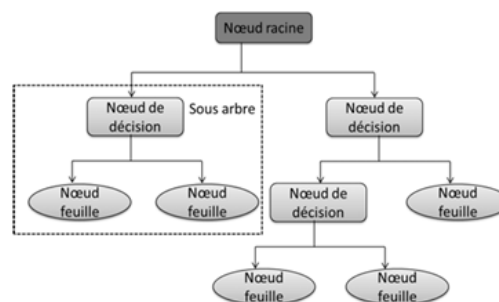


FIGURE 1.12 – Exemple d'arbre de décision binaire

- **Forêts aléatoires (Random Forests) :**

Ensemble d'arbres de décision entraînés sur des sous-échantillons de données. Plus robustes que les arbres seuls.

La Figure 1.13 montre une représentation schématique d'une forêt aléatoire, illustrant l'apprentissage à partir de multiples sous-ensembles de données.

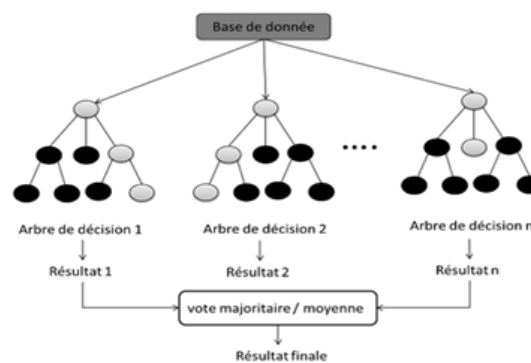


FIGURE 1.13 – Principe d'une forêt aléatoire

- **Régression logistique** : Classificateur probabiliste linéaire. Simple mais souvent efficace dans des contextes bien modélisés.
- **Perceptron multicouche (MLP)** : Réseaux de neurones simples (fully connected) capables d'apprendre des fonctions non linéaires. Considérés comme une transition entre ML classique et deep learning. La figure 1.14 illustre l'architecture du MLP avec deux couches cachées.

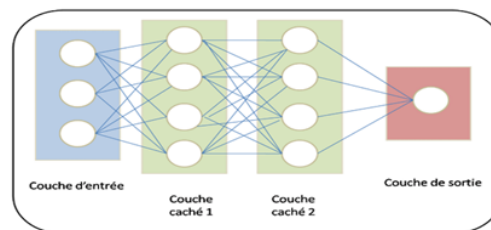


FIGURE 1.14 – Architecture d'un MLP à deux couches cachées[18].

2. Apprentissage semi-supervisé :

L'apprentissage semi-supervisé (ou hybride) combine un petit nombre de données annotées avec un grand volume de données non annotées. Il est particulièrement adapté aux contextes où l'annotation manuelle est coûteuse ou difficile, comme en imagerie médicale. Des techniques telles que l'auto-apprentissage ou la propagation de labels permettent d'exploiter les structures sous-jacentes des données pour améliorer les performances des modèles. Ceux-ci apprennent ainsi à généraliser à partir d'un signal partiel, tout en tirant parti de la richesse des données brutes. Certaines méthodes avancées s'appuient sur des réseaux semi-supervisés ou des modèles génératifs comme les autoencodeurs variationnels (VAE). L'apprentissage semi-supervisé représente ainsi un compromis efficace entre coût, performance et accessibilité des données.

3. Apprentissage non supervisé :

Dans ce cas, aucun label n'est fourni : le système doit découvrir des structures cachées dans les données. Cela inclut par exemple le regroupement (clustering) ou la réduction de

dimension.

■ **Principaux algorithmes :**

- **K-means :** Partitionne les données en k groupes en minimisant la distance intra-cluster. Utilisé en histopathologie pour segmenter des images ou regrouper des profils chromatiques.
- **K-nearest neighbors (KNN) :** Bien que souvent utilisé en classification supervisée, KNN peut aussi être adapté pour des tâches non supervisées, comme la détection d'anomalies ou le pré-clustering.

4. **Apprentissage par renforcement :**

Moins utilisé en imagerie médicale, l'apprentissage par renforcement repose sur un agent autonome qui interagit avec un environnement et apprend par essais-erreurs, via un mécanisme de récompense ou de punition. Il est très populaire en robotique ou en jeu vidéo (ex. : AlphaGo), mais commence à être exploré en médecine pour la planification de traitements ou la navigation dans des images 3D [19].

1.4.3 L'apprentissage profond (Deep Learning)

L'apprentissage profond, ou deep learning, est une sous-branche de l'apprentissage automatique fondée sur l'utilisation de réseaux de neurones artificiels à plusieurs couches. Il repose sur la capacité des algorithmes à apprendre des représentations hiérarchiques complexes à partir de données brutes, sans nécessiter d'extraction manuelle préalable de caractéristiques [20].

Un réseau de neurones artificiels (ANN) est inspiré du fonctionnement du neurone biologique. Dans le cerveau, chaque neurone reçoit des signaux de ses voisins, les combine, et déclenche une réponse en fonction d'un seuil (voir figure 1.15). De manière analogue, un neurone artificiel reçoit des entrées x_i , leur applique des poids w_i , et calcule une sortie via une fonction d'activation :

$$y = f\left(\sum w_i x_i + b\right)$$

où f peut être une fonction sigmoïde, ReLU ou tanh. En empilant ces neurones en couches successives, on construit un réseau profond (deep network).

Les réseaux neuronaux convolutifs (CNN)

Parmi les architectures les plus utilisées en traitement d'image figure le Convolutional Neural Network (CNN) (voir la figure 1.16). Contrairement aux MLP (Multilayer Perceptrons) où chaque neurone est connecté à tous les autres, les CNN exploitent la structure spatiale locale des images en appliquant des filtres (ou noyaux de convolution) sur de petites régions.

Un CNN est généralement constitué de trois types de couches :

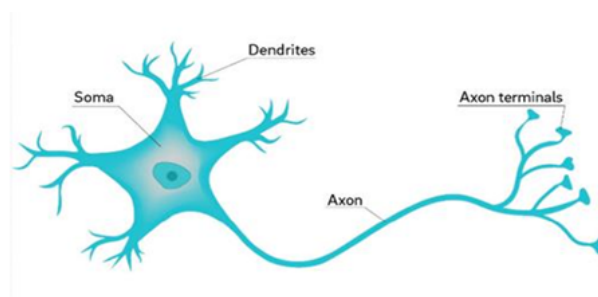


FIGURE 1.15 – Neurone biologique.

1. **Couches de convolution** : appliquent un filtre (matrice 3x3 ou 5x5) pour extraire des motifs locaux (bords, textures, contours). Chaque filtre apprend à détecter un type de motif visuel.
2. **Couches de pooling** : réduisent la taille des représentations (ex. : max pooling), permettant de gagner en invariance aux petites translations et de diminuer le nombre de paramètres.
3. **Couches entièrement connectées (fully connected)** : situées en fin de réseau, elles réalisent la classification proprement dite.

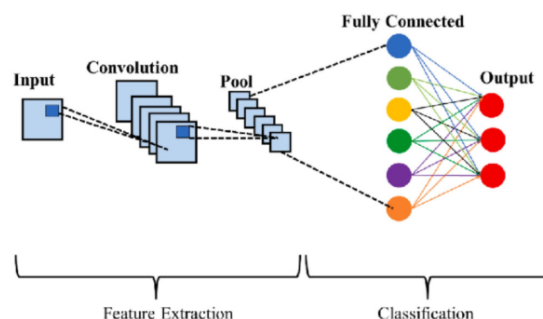


FIGURE 1.16 – Architecture d'un réseau de neurone convolutif.

Il y a plusieurs domaines qui utilisent le CNN pour résoudre les problèmes et parmi ces domaines en a la détection d'objet et classification d'image, dans le schéma suivant nous allons voir les meilleurs modèles dans les deux domaines précédents :

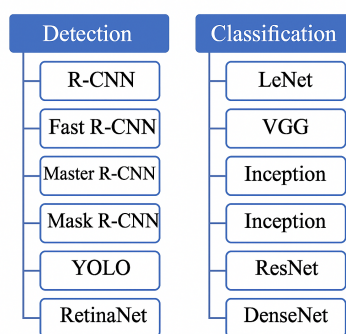


FIGURE 1.17 – Schéma des modèles de détection et classification par CNN

Avantages du deep learning en imagerie médicale

Le deep learning, et en particulier les CNN, a profondément transformé l'analyse d'images médicales. Voici quelques-uns de ses avantages clés dans le contexte de la pathologie numérique :

- Extraction automatique de caractéristiques : le réseau apprend lui-même les représentations pertinentes sans intervention humaine, ce qui évite le biais introduit par les méthodes manuelles.
- Robustesse face au bruit et à la variabilité inter-patients : les CNN généralisent bien sur des images provenant de sources différentes (scanners, colorations, laboratoires), à condition d'avoir suffisamment de données d'entraînement.
- Capacité à traiter des images de très grande dimension, comme les lames histologiques entières, via des approches patch-based.
- Performances supérieures aux méthodes classiques dans des tâches complexes : détection de mitoses, segmentation de noyaux, classification de types tumoraux [21, 22]. Ces propriétés en font un outil puissant pour l'analyse automatisée des lames H&E, notamment dans le cadre de la détection de mitoses, comme exploré dans ce mémoire.

Conclusion

Ce premier chapitre a établi les fondements théoriques nécessaires à la compréhension de notre étude, en mettant en lumière le rôle essentiel des mitoses dans le diagnostic histopathologique. Il a rappelé les difficultés liées à leur détection manuelle, notamment leur rareté, leur variabilité morphologique et la subjectivité des observations. Face à ces défis, nous avons introduit la pathologie numérique et les outils d'intelligence artificielle, en particulier le deep learning, comme solutions prometteuses. Ces approches permettent d'automatiser l'analyse d'images médicales et d'en améliorer la fiabilité. Ce cadre conceptuel servira de base pour le chapitre suivant, dédié à l'état de l'art des méthodes de détection des mitoses.

2

Revue de littérature et choix méthodologiques

Introduction

Ce deuxième chapitre s'inscrit dans la continuité du précédent en explorant l'évolution des méthodes de détection automatique des mitoses dans les images histopathologiques. Il présente d'abord les approches classiques basées sur l'extraction manuelle de caractéristiques et des classificateurs comme les SVM, avant d'aborder les méthodes modernes reposant sur le deep learning. L'apparition des réseaux neuronaux convolutifs (CNN) a permis d'améliorer la précision, notamment grâce à des architectures avancées capables de traiter des objets rares. Ce chapitre propose une revue de littérature des méthodes classiques et récentes en pathologie computationnelle, et justifie le choix méthodologique retenu dans ce mémoire.

Sommaire

Introduction	21
2.1 Méthodes de détection existantes	22
2.2 Revue de modèles modernes de détection d'objets	25
2.3 Justification de la démarche retenue : Choix du modèle RetinaNet	29
Conclusion	31

2.1 Méthodes de détection existantes

La détection des cellules mitotiques dans les images de biopsies colorées à l'hématoxyline et à l'éosine (coloration H&E) a fait l'objet de recherches depuis l'invention des scanners à lames complètes (Whole Slide Imaging). De plus, grâce aux récents développements de l'intelligence artificielle, de nombreuses techniques ont été développées, montrant des performances remarquables et pouvant être utilisées dans des applications en temps réel. Les travaux antérieurs peuvent être divisés en deux grandes catégories : les approches basées sur des caractéristiques conçues manuellement (handcrafted) et celles basées sur des caractéristiques profondes (deep features). Les détails de ces deux catégories sont présentés dans les sections suivantes.

2.1.1 Détection des mitoses à l'aide de caractéristiques conçues manuellement

Des techniques classiques de traitement d'image ont été utilisées pour extraire des caractéristiques conçues manuellement telles que la forme, la texture et la couleur, puis appliquées à des algorithmes d'apprentissage automatique pour détecter les cellules mitotiques. Dans la littérature existante, plusieurs méthodes reposant sur ces caractéristiques ont été proposées.

- Irshad a présenté une méthode [23] où tous les objets attendus étaient d'abord segmentés, puis des caractéristiques statistiques et morphologiques étaient extraites et classifiées à l'aide d'un arbre de décision [24]. Cette technique a obtenu la deuxième place au concours de détection des mitoses de la conférence internationale sur la reconnaissance des formes (ICPR) 2012.
- Tashk et al. ont proposé une technique [25] basée sur les Local Binary Patterns (LBP) et un classifieur SVM [24]. Les caractéristiques LBP sont très discriminantes et invariantes aux changements de niveaux de gris. Cette méthode a obtenu la troisième place au même concours.
- Sommer et al. ont combiné des caractéristiques de forme, d'intensité et de texture dans leur méthode [26] pour distinguer les cellules mitotiques des non-mitotiques, en utilisant un classifieur SVM. Ils ont utilisé deux logiciels open-source d'analyse d'images biomédicales : "ilastik" [27] pour la segmentation et "CellCognition" [28] pour la classification. Cette méthode demande peu de réglages de paramètres et aucun effort de l'utilisateur, grâce à l'utilisation de logiciels libres. Cependant, sa performance de détection reste faible comparée à d'autres méthodes basées sur des caractéristiques manuelles.
- Paul et al. se sont concentrés sur le noyau cellulaire dans leur méthode [29]. Ils ont utilisé un classifieur à forêts aléatoires régénératives qui a montré d'excellentes performances. Toutefois, cette méthode nécessite des ressources de calcul importantes et n'est donc pas adaptée à une application clinique pratique. Dans l'ensemble, la majorité des méthodes basées sur des caractéristiques manuelles affichent une performance de détection inférieure à celle des techniques récentes basées sur des caractéristiques profondes, et elles sont aussi coûteuses en calcul, en raison des opérations de traitement d'image classiques nécessaires à la segmentation des objets (voir la Table 2.1).

Méthode	Vitesse	Précision	Complexité	Avantages	Limites
Irshad (2012)	Moyenne	Moyenne	Moyenne	Simplicité, bien classée au challenge ICPR	Dépend de la qualité de segmentation initiale
Tashk et al. (2013)	Rapide	Moyenne	Faible	LBP efficace, robuste aux variations de gris	Moins performante que les approches profondes
Sommer et al. (2012)	Moyenne	Faible	Moyenne	Utilisation de logiciels open-source, peu de réglages	Faible précision de détection
Paul et al. (2015)	Lente	Bonne	Élevée	Bonnes performances avec forêts aléatoires régénératives	Consommation élevée de ressources, peu adapté en pratique

TABLE 2.1 – Comparaison des méthodes de détection de mitoses basées sur des caractéristiques conçues manuellement.

2.1.2 Transition vers les méthodes par deep learning

Malgré les avancées significatives permises par les approches manuelles de détection d'objets dans les images histopathologiques basées sur la segmentation, l'extraction de caractéristiques (texture, forme) et la classification (SVM, etc.), plusieurs limitations majeures subsistent. Ces méthodes nécessitent souvent une ingénierie manuelle des caractéristiques (feature engineering), ce qui introduit un biais humain, limite leur capacité à généraliser à des jeux de données hétérogènes, et rend leur adaptation à de nouveaux contextes particulièrement laborieuse[30].

Les images histologiques présentent une grande variabilité en termes de coloration, de morphologie cellulaire, de résolution, et de qualité de numérisation. Les descripteurs classiques, bien qu'efficaces dans des contextes contrôlés, peinent à capturer cette richesse visuelle et structurelle. De plus, les mitoses cibles de détection dans de nombreuses applications diagnostiques sont des objets particulièrement difficiles à identifier : elles sont rares, de petite taille, et présentent des variations morphologiques importantes, notamment dans les cas pathologiques complexes[31].

Dans ce contexte, l'émergence du deep learning, et en particulier des réseaux de neurones convolutifs (CNN), a marqué un tournant décisif. À partir de 2012, avec le succès d'AlexNet au concours ImageNet[32], les CNN ont démontré leur capacité à extraire automatiquement des représentations discriminantes à partir des images, sans nécessiter de conception manuelle de descripteurs. Appliqués à l'imagerie médicale, ces modèles ont permis des progrès spectaculaires dans la classification, la segmentation et la détection d'objets[21].

En histopathologie, les CNN permettent :

- d'apprendre directement à partir des pixels d'image les motifs complexes caractéristiques des mitoses (densité chromatinienne, contours irréguliers, texture granuleuse).
- d'absorber la variabilité inter-scanner grâce à l'exposition massive à des données augmen-

tées.

- et d'améliorer significativement la précision de détection, même dans des cas de mitoses atypiques[33].

Ce changement de paradigme ne s'est pas limité à l'architecture des modèles. Il s'est accompagné de l'émergence de nouveaux pipelines d'apprentissage automatique, exploitant des bases de données massives (comme MIDOG), des techniques d'annotation semi-automatisées, et des frameworks de détection modernes (Faster R-CNN, YOLO, RetinaNet)[34, 35].

Ainsi, le deep learning s'est imposé comme la nouvelle norme pour la détection d'objets en pathologie numérique. La section suivante explore en détail les architectures les plus représentatives de cette transition, en mettant en lumière leurs apports spécifiques et leurs limites dans le contexte particulier de la détection de mitoses.

2.1.3 Détection des mitoses à l'aide de caractéristiques profondes

L'intégration des caractéristiques profondes dans les pipelines de détection mitotique a donné lieu à une diversité d'approches, variant selon les choix architecturaux et les stratégies d'inférence. Les méthodes suivantes témoignent de cette dynamique de recherche.

- Ciresan et al. ont présenté une méthode [36] basée sur la fenêtre glissante pour extraire les caractéristiques profondes des images. Cette technique a remporté la première place au concours de détection des mitoses de l'ICPR 2012. Toutefois, l'approche par fenêtre glissante est très coûteuse en calcul, ce qui la rend inadaptée à un usage clinique.
- Malon et al. ont combiné des caractéristiques nucléaires manuelles avec des caractéristiques profondes extraites par un CNN [37]. Ils ont aussi utilisé un second CNN pour réduire la sensibilité lors de l'extraction, mais la méthode souffre de faibles performances et d'une complexité computationnelle élevée.
- Wang et al. ont proposé une méthode en cascade [38] avec deux classifieurs : l'un entraîné avec des caractéristiques manuelles, l'autre avec des caractéristiques CNN. En phase de test, un troisième classifieur est utilisé si les deux premiers donnent des résultats différents. Cette méthode est rapide et peu coûteuse, mais la sélection des régions d'intérêt (ROI) basée sur le traitement d'image classique est moins performante que celle des approches par deep learning.
- Chen et al. ont présenté une méthode en deux étapes [39] : dans la première, les cellules mitotiques sont segmentées par un réseau convolutif entièrement convolutif (FCN) ; dans la deuxième, les objets détectés sont affinés par un CNN supplémentaire.
- Récemment, les R-CNN (Region-based CNN) ont montré d'excellentes performances en vision par ordinateur. Pour la détection des mitoses, Li et al ont proposé une méthode [40] basée sur Faster R-CNN [41] avec un backbone ResNet-50 [42]. Le Faster R-CNN détecte les mitoses, puis celles-ci sont raffinées par ResNet-50. Le Faster R-CNN utilise VGG-16 [43] comme réseau d'extraction de caractéristiques. Cette méthode donne

de bons résultats et un temps d'inférence réduit, mais pourrait être améliorée avec un meilleur backbone, car VGG-16 souffre du problème de disparition de gradient.

- Li et al. ont aussi proposé une autre méthode [44] basée sur des cercles concentriques pour des jeux de données faiblement annotés (ICPR 2014), ce qui est adapté dans ce cas.
- Cai et al. ont utilisé un Faster R-CNN modifié [45] avec ResNet-101 comme extracteur de caractéristiques. Bien que performant, ResNet-101 pourrait être remplacé par un réseau plus léger.
- Li et al. ont aussi proposé une méthode légère [46] développée sur des ordinateurs sans GPU, basée sur une architecture Region-based CNN.
- Enfin, Dodbballapur et al. ont proposé une méthode [47] fondée sur Mask R-CNN [48], intégrant ResNet-50 comme extracteur de caractéristiques et Xception [49] pour réduire les faux positifs. Cette technique est très précise, mais elle nécessite des GPU coûteux et un entraînement intensif, ce qui la rend peu adaptée à un usage clinique courant (voir la Table 2.2).

Méthode	Vitesse	Précision	Complexité	Avantages	Limites
Ciresan et al. (2012)	Très lente	Excellente	Très élevée	1 ^{re} place ICPR 2012	Coûteuse, peu adaptée en clinique
Malon et al. (2013)	Lente	Faible	Élevée	CNN + carac. manuelles	Faible performance globale
Wang et al. (2014)	Rapide	Moyenne	Moyenne	Cascade efficace	Dépend traitement classique
Chen et al. (2015)	Moyenne	Bonne	Moy.-élevée	Segmentation + CNN	Plusieurs étapes nécessaires
Li et al. (2017)	Moyenne	Bonne	Moyenne	Faster R-CNN + ResNet-50	Backbone VGG-16 sous-optimal
Li et al. (2018)	Rapide	Moyenne	Faible	Pour données peu annotées	Moins précis que modèles lourds
Cai et al. (2019)	Moyenne	Bonne	Moyenne	ResNet-101 performant	Peut être allégé
Li et al. (2020)	Très rapide	Moyenne	Faible	Léger, sans GPU	Moins précis que gros modèles
Dodbballapur et al. (2021)	Lente	Excellente	Très élevée	Très précis, peu de faux +	GPU coûteux, peu adapté

TABLE 2.2 – Comparaison des méthodes de détection de mitoses basées sur des caractéristiques profondes.

2.2 Revue de modèles modernes de détection d'objets

La détection d'objets est une tâche fondamentale en vision par ordinateur, qui consiste à localiser et classifier les objets présents dans une image. Dans le domaine de la pathologie numérique, cette tâche permet notamment l'identification d'anomalies cellulaires comme les mitoses. Les réseaux de neurones convolutifs (CNN) ont largement surpassé les méthodes classiques, grâce à leur capacité à extraire automatiquement des représentations visuelles riches à

partir des données.

On distingue deux grandes familles de détecteurs modernes :

- ▶ Les modèles en deux étapes, tels que R-CNN, Fast R-CNN et Faster R-CNN, qui séparent la génération de propositions régionales de la classification.
- ▶ Les modèles en une seule étape, tels que YOLO et RetinaNet, qui réalisent la détection en un seul passage sur l'image.

Les sections suivantes présentent les modèles les plus emblématiques de ces deux familles, en soulignant leurs contributions respectives.

2.2.1 R-CNN et ses évolutions

a) R-CNN (Regions with CNN features)

Introduit par Girshick et al. en 2014 [50], R-CNN est le premier modèle à combiner les sélections régionales classiques (Selective Search) avec des CNN profonds pour extraire des caractéristiques discriminantes. Pour chaque proposition régionale, une classification est effectuée à l'aide d'un SVM. Malgré sa précision, ce modèle est lent à l'inférence en raison de l'extraction individuelle des features pour chaque région.

b) R-CNN (Regions with CNN features)

Proposé par le même auteur en 2015[51], Fast R-CNN améliore l'efficacité en extrayant une seule fois les cartes de caractéristiques de l'image entière, puis en appliquant des RoI Pooling pour chaque proposition. Cela réduit considérablement le temps d'inférence tout en maintenant de bonnes performances.

c) Faster R-CNN

Faster R-CNN, introduit par Ren et al. en 2015 [52], constitue une avancée majeure. Il intègre un Region Proposal Network (RPN) qui remplace les méthodes de sélection régionales classiques. Ce RPN génère directement des régions d'intérêt pertinentes à partir des cartes de caractéristiques (voir la figure 2.1).

L'architecture de Faster R-CNN est composée :

- ★ d'un backbone CNN (comme ResNet ou VGG) pour l'extraction des features.
- ★ d'un RPN pour proposer des régions.
- ★ et d'un classifieur pour prédire la classe et affiner la boîte englobante.

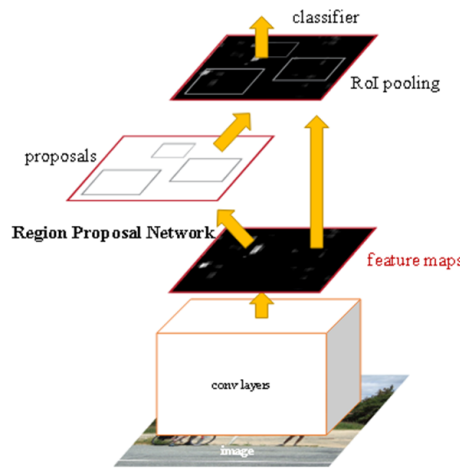


FIGURE 2.1 – architecture de model Faster R-CNN [15].

Faster R-CNN est reconnu pour sa précision, notamment pour les objets petits et rares, ce qui le rend adapté à la détection de mitoses. Cependant, sa vitesse d'inférence reste un point faible pour les applications temps réel.

2.2.2 RetinaNet

RetinaNet, proposé par Lin et al. en 2017 [34], est un modèle à une seule étape qui cherche à combiner vitesse et précision. Il repose sur un backbone ResNet avec un Feature Pyramid Network (FPN), afin de gérer les objets de différentes tailles (voir la figure 2.2).

Son innovation majeure est la Focal Loss, une fonction de perte qui réduit l'impact des exemples faciles (comme le fond), afin de mieux se concentrer sur les objets rares. Cette propriété le rend particulièrement adapté à des tâches médicales, comme la détection de mitoses où le déséquilibre entre classes est fort.

L'architecture de RetinaNet comprend deux têtes :

- ★ une tête de classification.
- ★ une tête de régression pour les boîtes englobantes.

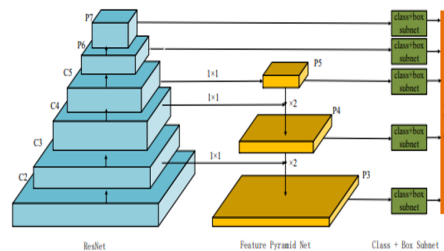


FIGURE 2.2 – : Architecture de RetinaNet [53].

2.2.3 YOLO (You Only Look Once)

YOLO, introduit par Redmon et al. en 2016 [54], est un modèle à une seule étape conçu pour la vitesse. Contrairement aux modèles R-CNN, YOLO divise directement l’image en une grille et effectue la détection en une seule passe(voir la figure 2.3).

La version YOLOv5, développée par Ultralytics en 2020 [55], améliore la précision tout en restant rapide. Elle utilise :

- un backbone CSPDarknet pour l’extraction de caractéristiques.
- une tête PANet pour la fusion des features multi-échelles.
- des techniques modernes comme Mosaic Augmentation, AutoAnchor, et CIOU Loss.

YOLOv5 est adapté aux systèmes temps réel, mais il est parfois moins performant que RetinaNet sur les petits objets comme les mitoses.

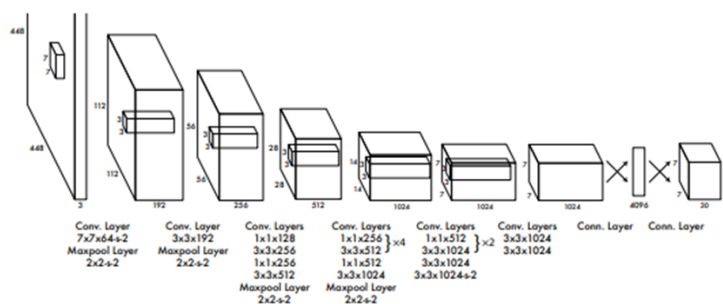


FIGURE 2.3 – Architecture de model YOLO [56].

Modèle	Vitesse	Précision (rares)	Complexité	Avantages	Limites
R-CNN (2014)	Très lente	Bonne	Très élevée	1 ^{re} approche CNN	Temps d’inférence prohibitif
Fast R-CNN (2015)	Moyenne	Bonne	Élevée	Moins de redondance	Dépend de Selective Search
Faster R-CNN (2015)	Lente	Excellente	Moy.-élevée	Bon pour petits objets	Trop lent pour temps réel
RetinaNet (2017)	Moyenne	Excellente	Moyenne	Focal Loss, équilibre	Moins rapide que YOLO
YOLOv5 (2020)	Très rapide	Bonne	Faible	Rapide, temps réel	Faible sur petits objets

TABLE 2.3 – Comparaison des modèles de détection d’objets.

2.3 Justification de la démarche retenue : Choix du modèle RetinaNet

La détection des mitoses dans des images histopathologiques représente un défi majeur en vision par ordinateur. Contrairement aux objets classiques présents dans les images naturelles, les mitoses sont de très petite taille, visuellement subtiles, rares au sein des tissus, et les images histologiques elles-mêmes sont soumises à une forte variabilité inter-scanner (couleurs, résolution, artefacts). Face à ces contraintes, le choix du modèle de détection d'objets doit être mûrement réfléchi.

Dans ce contexte, nous avons retenu RetinaNet, un modèle de détection à une seule étape, comme architecture centrale de notre pipeline de détection. Ce choix repose sur des considérations à la fois théoriques et empiriques, que nous détaillons ci-dessous.

2.3.1 Une détection robuste dans les contextes de classes déséquilibrées : la Focal Loss

L'un des obstacles majeurs à la détection de mitoses est l'extrême déséquilibre entre les classes. Dans une lame entière, les pixels correspondant à une mitose représentent souvent moins de 0,01 % de l'image. Ce déséquilibre engendre un problème de domination de la classe négative (le fond) durant l'apprentissage : un modèle standard tend à minimiser la perte en prédisant constamment l'absence d'objet, ignorant ainsi les régions positives.

RetinaNet répond à ce problème par l'introduction d'une fonction de perte innovante : la Focal Loss [34]. Celle-ci modifie la perte de classification traditionnelle (cross-entropy) en réduisant le poids des exemples faciles (par exemple, les régions de fond correctement classées) et en augmentant celui des exemples difficiles, comme les mitoses.

Concrètement, la Focal Loss introduit un facteur multiplicatif dépendant de la confiance du modèle dans sa prédiction. Les prédictions correctes avec une forte confiance sont ainsi peu pénalisées, tandis que les erreurs ou les prédictions incertaines reçoivent un poids plus élevé. Ce mécanisme permet de focaliser l'apprentissage sur les objets rares et complexes à détecter, rendant RetinaNet particulièrement adapté aux images médicales fortement déséquilibrées[34].

2.3.2 Une architecture hiérarchique pour détecter les objets de toutes tailles : le Feature Pyramid Network (FPN)

Outre leur rareté, les mitoses se distinguent par leur petite taille, parfois à peine perceptible à l'échelle globale de l'image. Les réseaux convolutifs classiques, en particulier dans leurs couches profondes, tendent à perdre la précision spatiale nécessaire à leur localisation.

RetinaNet surmonte cette limite grâce à l'utilisation du Feature Pyramid Network (FPN) [57], associé à un backbone de type ResNet50. Le FPN permet de construire une pyramide

de caractéristiques à plusieurs échelles, en fusionnant les informations extraites à différents niveaux de profondeur du réseau.

Cette structure hiérarchique offre au modèle la capacité de détecter des objets sur une large gamme de tailles. Dans le cas des mitoses, cela lui permet d'exploiter à la fois les résolutions fines des couches peu profondes et la richesse sémantique des couches plus profondes. Cette approche est donc particulièrement efficace pour détecter des objets de petite taille et à faible contraste, comme les noyaux en division[57].

2.3.3 Un compromis pertinent entre précision et efficacité

Les modèles de détection peuvent être classés en deux grandes catégories :

- Les détecteurs à deux étapes, comme Faster R-CNN, qui proposent d'abord des régions d'intérêt, puis les classifient.
- Les détecteurs à une seule étape, comme YOLO ou RetinaNet, qui réalisent directement la détection à partir des cartes de caractéristiques.

RetinaNet, bien qu'appartenant à la seconde catégorie, a été conçu pour atteindre une précision comparable à celle des détecteurs à deux étapes, tout en réduisant considérablement le temps d'inférence. Cette efficacité est rendue possible grâce à une architecture optimisée et à l'abandon de l'étape explicite de génération de propositions régionales, particulièrement coûteuse en calculs.

Dans le domaine médical, où la précision de détection est primordiale, mais où des contraintes de temps et de ressources computationnelles peuvent exister (notamment en vue d'un déploiement clinique), ce compromis entre performance et efficacité est particulièrement attractif.

2.3.4 Adaptabilité au contexte de la pathologie numérique

Enfin, RetinaNet s'intègre de manière fluide dans notre pipeline de détection de mitoses, qui repose sur des prétraitements spécifiques tels que la normalisation de la couleur (Macenko) et le découpage en patches.

Ces étapes visent à réduire la variabilité inter-scanner et à augmenter la probabilité de présence de mitoses dans chaque patch. Le modèle RetinaNet est capable de tirer parti de ces conditions d'entrée homogénéisées pour extraire efficacement des caractéristiques visuelles discriminantes, même dans des régions très limitées de l'image.

Sa structure modulaire (backbone + FPN + têtes de détection) le rend également facile à adapter selon les spécificités du jeu de données, les contraintes matérielles ou encore les objectifs finaux (précision, généralisation, interprétabilité).

Conclusion

Ce chapitre a retracé l'évolution des approches de détection automatique des mitoses, en distinguant les méthodes basées sur des caractéristiques manuelles et celles fondées sur le deep learning. Les premières, simples mais limitées, utilisent des descripteurs visuels avec des classificateurs classiques. Les secondes, plus récentes, apprennent automatiquement des représentations discriminantes grâce à des architectures puissantes comme Faster R-CNN ou Mask R-CNN. Cette revue souligne l'efficacité des approches profondes pour traiter la complexité des images histopathologiques, et justifie le choix d'un modèle adapté aux objets rares : RetinaNet.

3

Méthodologie et développement du modèle

Introduction

Ce troisième chapitre présente le pipeline expérimental développé pour la détection automatique des mitoses à l'aide de RetinaNet. Il détaille la préparation et le traitement des données du dataset MIDOG 2022, en tenant compte de la diversité des scanners et de la rareté des mitoses. Les étapes clés incluent la normalisation des images, le découpage en patches, l'annotation, ainsi que l'adaptation du modèle avec un backbone ResNet-50 et la Focal Loss pour gérer le déséquilibre des classes. Ce chapitre expose l'ensemble du processus, depuis les données jusqu'à la configuration finale du modèle, dans un cadre adapté à la pathologie numérique.

Sommaire

Introduction	32
3.1 Données utilisées	33
3.2 Pipeline général du système	36
3.3 Architecture du modèle – RetinaNet	37
3.4 Préparation des données	43
3.5 Configuration de l'apprentissage	47
Conclusion	50

3.1 Données utilisées

3.1.1 Présentation du jeu de données MIDOG 2022

Le jeu de données utilisé dans ce projet est issu du Mitosis D_Omain Generalization Challenge 2022 (MIDOG 2022). Bien que ce travail ne soit pas directement lié à la compétition, le dataset a été sélectionné en raison de sa richesse et de sa pertinence pour notre problématique, la détection automatique de mitoses dans des images histopathologiques variées [58].

MIDOG 2022 est une initiative internationale visant à évaluer la capacité des algorithmes à généraliser à travers différents domaines incluant des types de cancers, scanners, centres médicaux, et espèces (humaine et canine) [58]. Cette diversité rend le dataset particulièrement adapté à l'évaluation de modèles robustes en contexte biomédical.

Le corpus contient 405 images au format .tiff (non compressé, haute résolution) (voir la figure 3.1), chacune représentant une région d'intérêt (ROI) de 2 mm², équivalente à environ dix champs de haute puissance (HPF) en microscopie optique [59]. Ces images sont extraites de lames complètes (WSI – Whole Slide Images), numérisées à un grossissement de 40×, garantissant une haute fidélité morphologique.

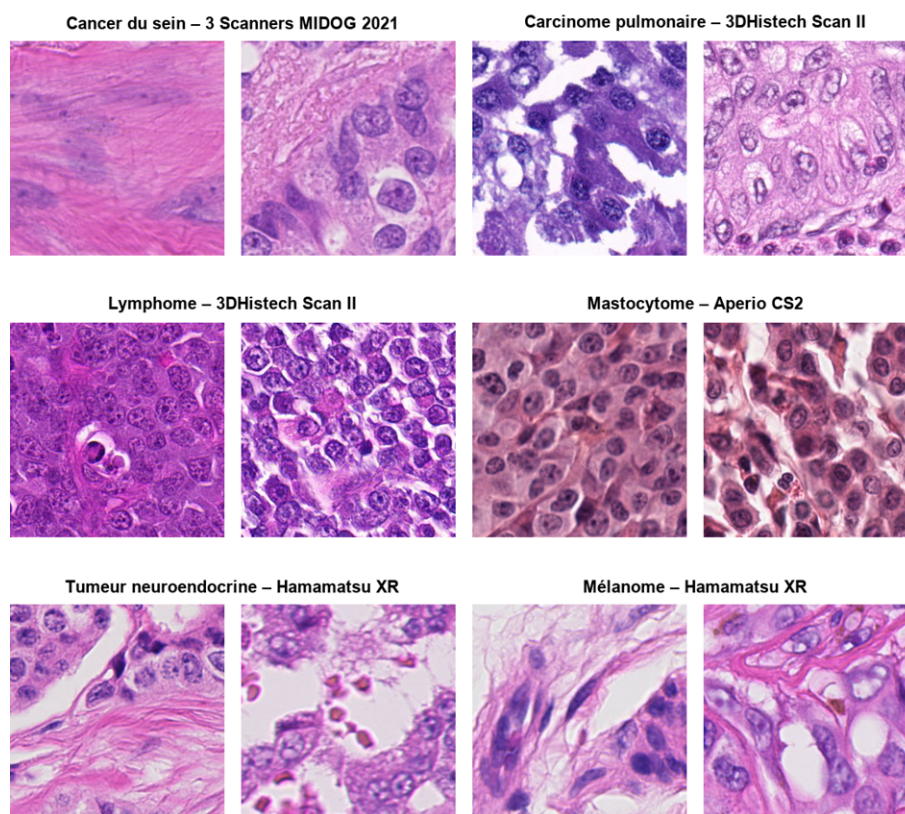


FIGURE 3.1 – Exemples d'images histopathologiques issues du dataset MIDOG 2022.

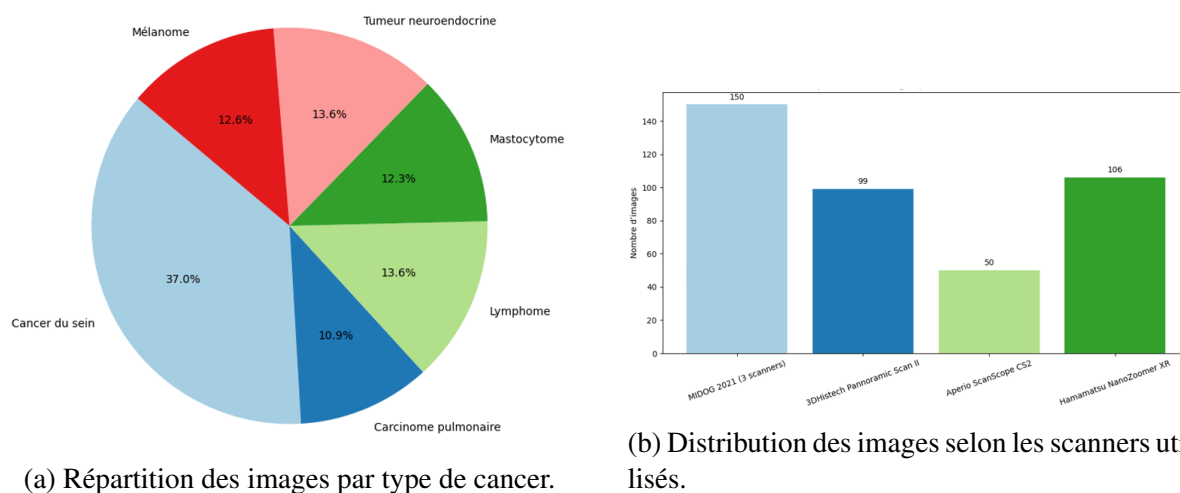
3.1.2 Structure du dataset et diversité inter-domaines

L'un des atouts majeurs du dataset MIDOG 2022 réside dans sa diversité inter-domaines (voir Table 1.1). Les données couvrent plusieurs types de tumeurs, espèces et centres hospitaliers, avec des scanners différents, un critère fondamental pour entraîner des modèles capables de généraliser dans des conditions réelles.

Plage de fichiers	Type de cancer	Espèce	Centre médical	Scanner utilisé
001–150	Cancer du sein	Humain	UMC Utrecht	3 scanners différents (MIDOG 2021)
151–194	Carcinome pulmonaire	Canin	VetMedUni Vienna	3DHitech Pannoramic Scan II
195–249	Lymphome	Canin	VetMedUni Vienna	3DHitech Pannoramic Scan II
250–299	Mastocytome	Canin	FU Berlin	Aperio ScanScope CS2
300–354	Tumeur neuroendocrine	Humain	UMC Utrecht	Hamamatsu NanoZoomer XR
355–405	Mélanome (non annoté)	Humain	UMC Utrecht	Hamamatsu NanoZoomer XR

TABLE 3.1 – Informations sur les fichiers selon le type de cancer, espèce, centre médical et scanner utilisé.

Cette diversité est illustrée par les figures suivantes :



Cette hétérogénéité permet d'examiner la capacité du modèle à généraliser entre domaines, malgré la variabilité des conditions d'acquisition.

3.1.3 Format et structure des annotations

Les annotations sont fournies dans deux formats complémentaires :

- SQLite (compatible avec le visualiseur SlideRunner), utile pour l'inspection manuelle interactive.
- JSON (format COCO), utilisé pour l'entraînement automatique de modèles de détection d'objets(voir la figure 3.3).

Les fichiers COCO contiennent pour chaque image les champs suivants :

- `bbox` : boîte englobante de l'objet détecté au format `[x_min, y_min, x_max, y_max]`.
- `category_id` : 1 pour mitose, 2 pour négatif dur (hard negative).
- `image_id` : identifiant de l'image annotée.

```

{"bbox": [3309, 3268, 3359, 3318], "category_id": 1, "image_id": 94, "id": 2594}, {"bbox": [3171, 3104, 3221, 3154],
{"bbox": [3678, 3217, 3728, 3267], "category_id": 1, "image_id": 94, "id": 2596}, {"bbox": [3611, 3574, 3661, 3624],
{"bbox": [3750, 3834, 3800, 3884], "category_id": 2, "image_id": 94, "id": 2598}, {"bbox": [5036, 3338, 5086, 3388],
{"bbox": [5580, 3326, 5630, 3376], "category_id": 2, "image_id": 94, "id": 2600}, {"bbox": [5768, 3449, 5818, 3499],
{"bbox": [4599, 3837, 4649, 3887], "category_id": 1, "image_id": 94, "id": 2602}, {"bbox": [6048, 3451, 6098, 3501],
{"bbox": [5896, 3622, 5946, 3672], "category_id": 1, "image_id": 94, "id": 2604}, {"bbox": [5945, 3540, 5995, 3590],
{"bbox": [6527, 3394, 6577, 3444], "category_id": 2, "image_id": 94, "id": 2606}, {"bbox": [6969, 3377, 7019, 3427],

```

FIGURE 3.3 – Exemple d'annotation COCO (extrait JSON affichant une `bbox` avec `category_id`).

Le dataset comporte :

- 9501 mitoses annotées (véritables cibles positives).
- 11051 négatifs durs, objets visuellement similaires à des mitoses mais non mitotiques (ex. cellules apoptotiques, artefacts tissulaires)(voir la figure3.4)[59].

Ces négatifs jouent un rôle essentiel dans l'entraînement en réduisant les faux positifs, et en améliorant la précision du classifieur.

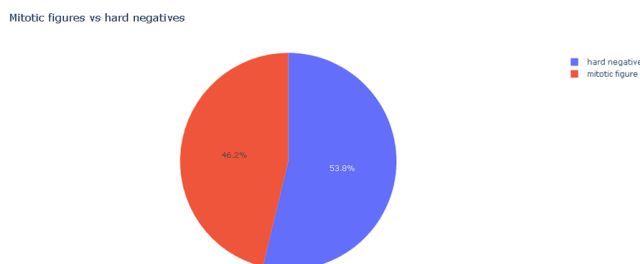


FIGURE 3.4 – Diagramme circulaire illustrant la proportion de mitoses (rouge) et de négatifs durs (bleu) dans un échantillon.

3.1.4 Spécifications techniques des images

Toutes les images du dataset MIDOG 2022 sont dérivées de lames histologiques numérisées à 40×, avec une résolution standardisée de 0,25 $\mu\text{m}/\text{pixel}$, permettant une analyse précise à l'échelle cellulaire.

► **Encadré technique** : Détails des fichiers image

- Format : TIFF (non compressé, RGB 24 bits).
- Taille moyenne : environ 7215×5412 pixels.
- Origine : ROIs extraites automatiquement à partir de WSI standardisées,
- Utilisation : détection de mitoses, évaluation inter-scanner, classification tumorale.

Ce format est couramment utilisé en pathologie numérique pour des tâches nécessitant une forte précision morphologique.

3.2 Pipeline général du système

Cette section présente une vue d'ensemble du pipeline mis en place pour la détection automatique des mitoses à partir des lames histologiques numérisées. Le but est de contextualiser les différentes étapes du traitement, qui seront détaillées plus loin dans ce chapitre.

Le pipeline complet se compose des étapes suivantes :

- **Chargement des WSI** : les lames histologiques entières sont ouvertes via la bibliothèque OpenSlide.
- **Découpage en patches** : les images sont fragmentées en sous-images (patches) de taille fixe pour rendre l'apprentissage tractable.
- **Normalisation des couleurs** : une harmonisation des colorations est appliquée (méthode de Macenko) pour réduire la variabilité liée aux scanners.
- **Augmentation de données** : des transformations (rotation, flip, contraste...) sont utilisées pour enrichir les données d'entraînement.
- **Entraînement du modèle** : le modèle RetinaNet est entraîné à détecter les mitoses à partir des patches annotés.
- **Prédiction et évaluation** : le modèle est testé sur des images non vues pour mesurer sa performance via des métriques comme le mAP.

L'ensemble de ce processus est illustré dans la figure ci-dessous.

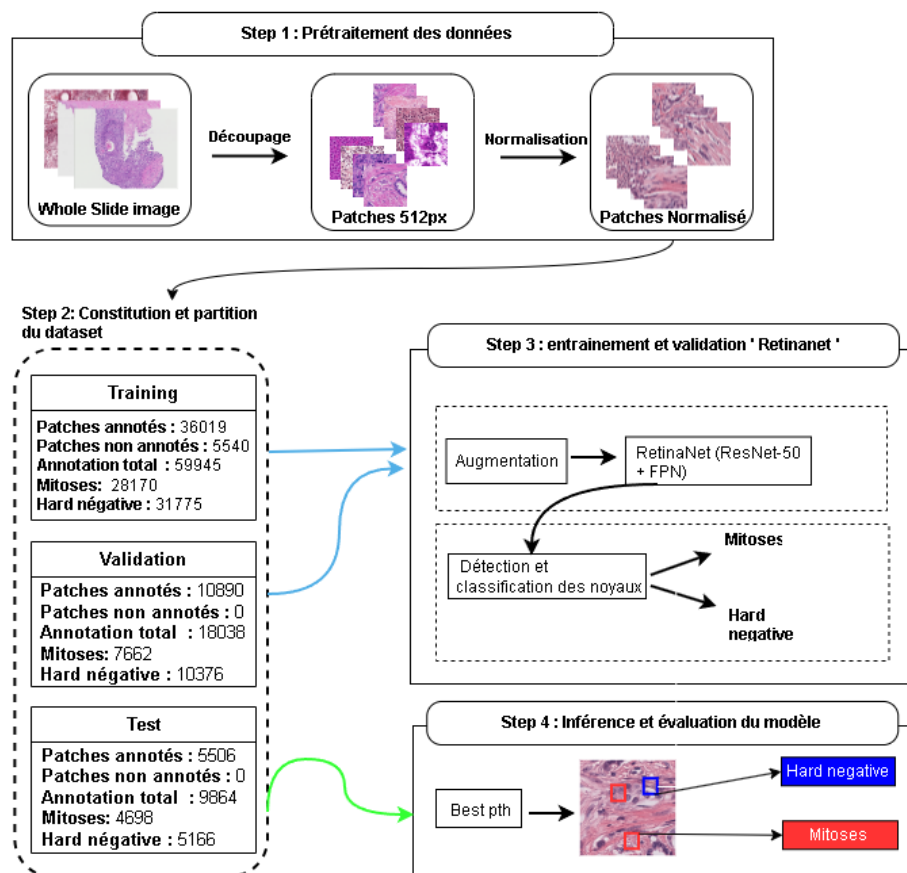


FIGURE 3.5 – Pipeline général du système de détection de mitoses.

3.3 Architecture du modèle – RetinaNet

3.3.1 Détails de l'architecture

Définition

RetinaNet est un modèle one-stage de détection d'objets, introduit par Lin et al. (2017) [34]. Contrairement aux architectures two-stage comme Faster R-CNN, il combine en une seule étape la localisation (régression des boîtes) et la classification des objets (voir la figure 3.6). Cela permet de réduire le temps d'inférence tout en maintenant une précision comparable, notamment grâce à l'utilisation d'une nouvelle fonction de perte appelée Focal Loss .

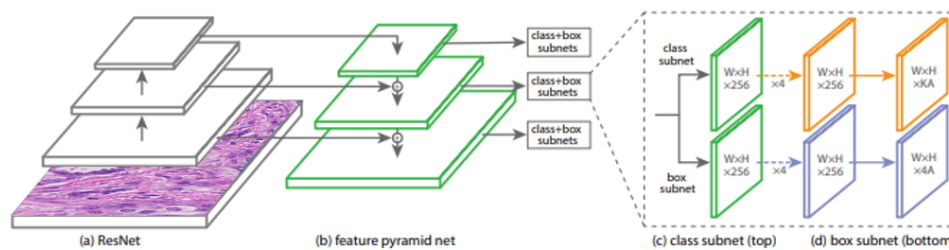


FIGURE 3.6 – Architecture globale de RetinaNet avec backbone ResNet-50, pyramide FPN et sous-réseaux prédictifs[34].

Problème du déséquilibre entre classes

Dans les tâches de détection, les exemples négatifs (fond) sont très largement plus nombreux que les exemples positifs (objets). Dans notre cas, les mitoses sont rares comparées aux pixels non-mitotiques, ce qui cause un déséquilibre important entre classes.

Une fonction de perte classique comme la cross-entropy a tendance à favoriser les exemples faciles, c'est-à-dire les négatifs correctement classés, ce qui pénalise l'apprentissage sur les vrais objets.

3.3.2 Fonction de perte – Focal Loss

Formulation mathématique

La Focal Loss est définie comme suit :

$$\mathcal{L}_{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

avec :

- $p_t = p$ si $y = 1$, et $p_t = 1 - p$ si $y = 0$.
- $\gamma \in [0,5]$: facteur de focalisation.
- α_t : coefficient de pondération des classes.

Intuition et impact des paramètres

- $\gamma = 2$ recommandé par Lin et al[34].
- Plus γ est grand, plus les exemples faciles sont ignorés.
- α compense le déséquilibre global des classes (voir la figure 3.7).

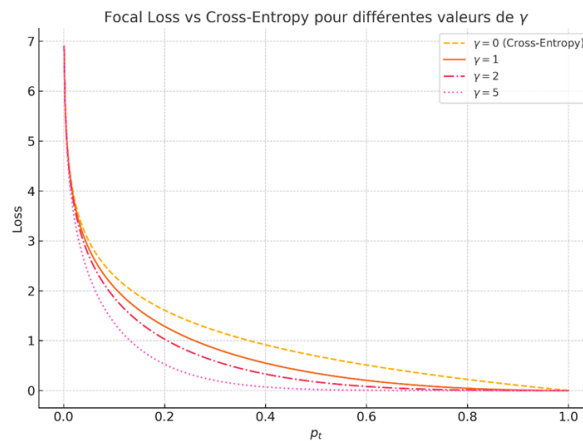


FIGURE 3.7 – Courbe de la Focal Loss comparée à la Cross-Entropy pour différentes valeurs de γ [34].

3.3.3 Backbone : ResNet-50

Architecture de ResNet-50

ResNet-50 est un réseau convolutif profond constitué de blocs résiduels [60]. Il est composé de 5 étages convolutifs notés conv1 à conv5. Chaque étage est lui-même formé de plusieurs blocs de type "bottleneck" (voir la figure 3.8).

Chaque bottleneck est composé de trois couches :

- Une convolution 1×1 pour la réduction de dimension (compression).
- Une convolution 3×3 pour l'extraction locale.
- Une convolution 1×1 pour la rétro-expansion.

Les connexions résiduelles facilitent l'apprentissage profond en contournant le problème de disparition du gradient :

$$\mathbf{y} = \mathcal{F}(\mathbf{x}) + \mathbf{x}$$

- x : l'entrée du bloc résiduel (c'est l'information transmise depuis les couches précédentes).
- $F(x)$: la fonction résiduelle apprise par le bloc (représente la transformation effectuée par les couches convolutives du bloc).
- y : la sortie du bloc résiduel, obtenue en ajoutant l'entrée à la fonction résiduelle, selon la formule : $y = F(x) + x$.

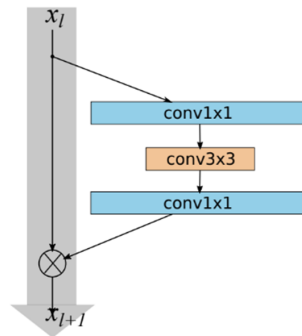


FIGURE 3.8 – Bloc bottleneck utilisé dans ResNet-50.

Pré-entraînement sur ImageNet

ResNet-50 est généralement pré-entraîné sur ImageNet, une base de données contenant plus de 1 million d'images sur 1000 classes. Le pré-entraînement permet :

- de bénéficier de features génériques.
- d'accélérer la convergence de l'apprentissage.
- d'améliorer la généralisation.

Extraction des features pour FPN

Les sorties des blocs conv3, conv4, conv5 (C3, C4, C5) sont utilisées pour construire la pyramide :

- C3 : résolution 1/8 de l'image.
- C4 : 1/16.
- C5 : 1/32.

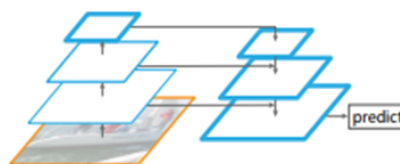


FIGURE 3.9 – Correspondance entre niveaux de ResNet et entrées FPN[57].

Formule du coût computationnel d’une convolution :

$$\text{FLOPs} = 2 \times H \times W \times C_{in} \times K^2 \times C_{out}$$

où :

- H et W : hauteur et largeur de la carte de caractéristiques.
- K : taille du noyau (kernel).
- C_{in} : nombre de canaux en entrée.
- C_{out} : nombre de canaux en sortie.

3.3.4 Feature Pyramid Network (FPN)

Justification de l’utilisation d’une pyramide de caractéristiques

Dans la détection d’objets, la gestion de la variation d’échelle est un enjeu fondamental. Les objets peuvent apparaître à différentes tailles selon leur distance, leur emplacement ou leur grossissement (dans le cas d’images biomédicales).

Les modèles classiques utilisaient des pyramides d’images multi-résolutions, coûteuses en calcul. La Feature Pyramid Network (FPN), proposée par Lin et al. (2017)[57], tire parti de la structure hiérarchique des réseaux convolutifs pour créer une pyramide à partir d’une seule image, avec un coût mémoire et calcul réduit.

Construction de la pyramide de caractéristiques

Le FPN repose sur trois mécanismes :

- Voie ascendante (bottom-up) : C3, C4, C5 issues du ResNet.
- Voie descendante (top-down) : Upsampling ($\times 2$) des cartes plus profondes.
- Connexions latérales : Fusion des features peu profondes et profondes (voir la figure 3.10).

Formule de la construction :

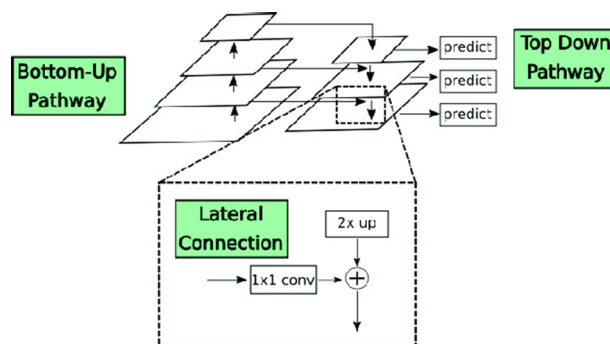


FIGURE 3.10 – Schéma détaillé du bloc FPN avec upsampling + addition + conv 3×3 [61].

Avantages de l'architecture FPN

- meilleure détection des petits objets.
- pas besoin de générer une pyramide d'images.
- apprentissage end-to-end.

3.3.5 Génération des boîtes d'ancrage (Anchors)

Principe des ancrs

Chaque cellule d'une carte P_l génère un ensemble de boîtes d'ancrage (anchors) de tailles et de proportions diverses. Ces boîtes servent de références pour la régression des objets (voir la figure 3.11).

Chaque ancre $a \in A$ est définie par :

- sa position centrale (x_a, y_a) ,
- sa taille (w_a, h_a) ,
- son ratio $r = \frac{h_a}{w_a}$.

Paramètres utilisés dans notre projet

Pour détecter des mitoses de petite taille :

- Tailles : [16, 32, 48, 64, 96] pixels.
- Ratios d'aspect : (0.5, 1.0, 2.0).

Nombre total d'ancres par cellule :

$$\text{Total} = N_{\text{scales}} \times N_{\text{ratios}} = 5 \times 3 = 15$$

Implémentation pratique

```
anchor_sizes = tuple((x, int(x * 2**(1/3)), int(x * 2**(2/3))) for x in [16, 32, 48, 64, 96])
aspect_ratios = ((0.5, 1.0, 2.0,) * len(anchor_sizes))
anchor_generator = AnchorGenerator(sizes=anchor_sizes, aspect_ratios=aspect_ratios)
```

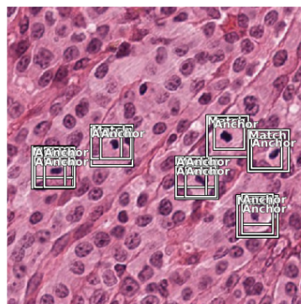


FIGURE 3.11 – Visualisation des ancrs sur différents niveaux de la pyramide (P3 à P7)[58].

3.3.6 Sous-réseaux de classification et de régression

Structure des têtes de prédiction

Chaque niveau P_l de la FPN est connecté à deux sous-réseaux :

- **Classification :**
 - o 4 convolutions 3×3 , 256 filtres,
 - o une couche finale de dimension $A \times C$,
 - o activation : sigmoïde.
- **Régression :**
 - o même structure, sortie de dimension $A \times 4$,
 - o pas d'activation finale.

Fonctions de perte utilisées

- Classification : Focal Loss (voir section 3.2).
- Régression : Smooth L1 Loss :

$$\mathcal{L}_{\text{reg}}(x) = \begin{cases} 0.5x^2 & \text{si } |x| < 1 \\ |x| - 0.5 & \text{sinon} \end{cases}$$

Ajustements spécifiques au projet

- Nombre de classes : 3 (mitose, hard negative, fond).
- Biais initial de la classification :

$$b = -\log\left(\frac{1 - \pi}{\pi}\right), \quad \pi = 0.01$$

3.3.7 Normalisation des images

Objectif de la normalisation

La normalisation vise à centrer et réduire les valeurs des pixels pour :

- accélérer l'apprentissage,
- améliorer la stabilité numérique,
- réduire l'effet de la variabilité de couleur.

Valeurs utilisées

Issues du jeu d'entraînement après normalisation Macenko :

$$\text{Image mean} = [0.7077, 0.5339, 0.6414] \quad \text{Image std} = [0.1758, 0.1892, 0.1399]$$

3.4 Préparation des données

3.4.1 Découpage en patches

Dans le cadre du traitement des images histopathologiques de très haute résolution (WSI – Whole Slide Images), le découpage en sous-images plus petites, appelées patches, constitue une étape préliminaire essentielle. Cette étape permet d'adapter les données au cadre d'entraînement des réseaux de neurones convolutionnels, tout en améliorant l'efficacité de la détection d'objets de petite taille, tels que les mitoses.

Objectif et justification du découpage

Les lames numériques de la base MIDOG 2022 présentent une taille considérable (plusieurs dizaines de milliers de pixels de côté), rendant leur traitement direct par un modèle d'apprentissage profond peu efficace, voire impossible pour des raisons de mémoire. Le découpage en patches permet non seulement de contourner cette contrainte, mais aussi de :

- Localiser finement les mitoses grâce à un traitement localisé de l'image.
- Augmenter le nombre d'échantillons disponibles sans recourir uniquement à des techniques d'augmentation artificielle.
- Favoriser la redondance locale des objets rares par l'utilisation d'un chevauchement (overlap) (voir la figure 3.12), limitant ainsi le risque de perte d'information en bordure de patch.

Paramètres du découpage

Le découpage est réalisé selon une grille glissante sur toute la surface de chaque WSI, avec les paramètres suivants :

- Taille du patch : 512×512 pixels.
- Pas de glissement (stride) : 256 pixels, soit un chevauchement de 50 % entre les patches successifs.
- Niveau de zoom : $40\times$ (niveau 0), correspondant à une résolution de $0,25 \mu\text{m}/\text{pixel}$.

Cette configuration assure une couverture redondante de l'image et une meilleure représentation contextuelle des régions d'intérêt.

Intégration des annotations

Les annotations initiales sont fournies sous forme de boîtes englobantes (bounding boxes), localisant les mitoses (label = 1) et les hard negatives (label = 2). Lors du découpage, chaque patch est associé aux annotations qu'il intersecte. Pour cela :

- Les coordonnées des boîtes sont recalculées relativement au coin supérieur gauche du patch.
- Seules les parties des boîtes situées à l'intérieur du patch sont conservées (rognage si nécessaire).
- Un patch est conservé uniquement s'il contient au moins une annotation.

Les informations relatives aux annotations sont enregistrées dans un fichier CSV structuré, compatible avec le format d'entrée de RetinaNet. Chaque ligne du fichier contient les éléments suivants :

$[chemin_vers_image, x_min, y_min, x_max, y_max, label, annotation_id, image_id_origine]$.

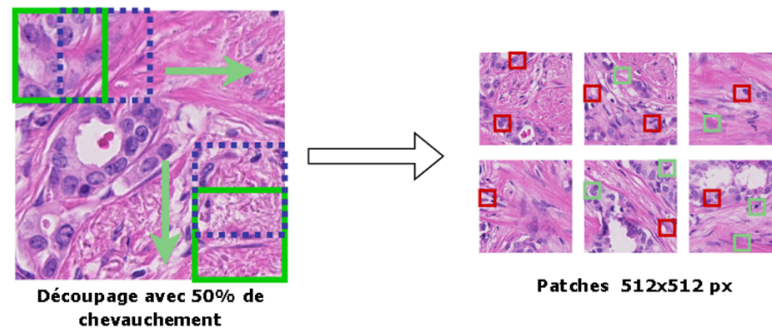


FIGURE 3.12 – Pipeline de découpage d'une WSI en patches chevauchants avec intégration des annotations mitotiques.

3.4.2 Normalisation des couleurs – Méthode de Macenko

L'un des défis majeurs en pathologie numérique est la variabilité de coloration entre les lames histologiques, même lorsqu'elles sont colorées avec les mêmes protocoles H&E (Hématoxyline et Éosine). Cette variabilité est introduite par plusieurs facteurs, notamment :

- la différence de concentration ou de durée d'exposition aux colorants .
- les conditions de fixation des tissus .
- et les variations entre les scanners utilisés (résolution, lumière, traitement optique).

Ces écarts colorimétriques compromettent la performance des algorithmes d'apprentissage profond, qui sont très sensibles aux variations de distribution d'entrée. Afin d'assurer une généralisation robuste du modèle, une étape de normalisation des couleurs est essentielle (voir la figure 3.13).

Méthode de Macenko : principes fondamentaux

La méthode de Macenko, proposée en 2009 [62], est l'une des approches les plus populaires pour standardiser la coloration H&E dans les images histopathologiques. Elle repose sur la décomposition des composantes colorantes (stains) dans l'espace de densité optique (OD), selon les étapes suivantes :

- a) **Transformation en densité optique :**

Les pixels RGB de l'image III sont d'abord convertis en densité optique via :

$$OD = -\log\left(\frac{I+1}{I_0}\right)$$

où :

- $I \in [0, 255]$ est l'intensité du pixel dans chaque canal (R, G, B).
- $I_0 = 240$ est l'intensité de lumière incidente choisie comme constante.
- un décalage de 1 est ajouté pour éviter le logarithme de zéro.

b) Filtrage des pixels faibles

Un seuil $\beta = 0.15$ est appliqué : seuls les pixels pour lesquels toutes les composantes OD sont supérieures à β sont conservés. Cela élimine les zones peu colorées ou blanches.

c) Analyse en composantes principales (SVD) :

Une décomposition en valeurs propres (SVD) est effectuée sur les vecteurs OD pour extraire les deux directions principales représentant les teintes dominantes (Hématoxyline et Éosine). Ces directions forment une base dans laquelle les pixels sont projetés.

d) Extraction des angles extrêmes :

On calcule les angles ϕ des pixels projetés, puis on détermine les angles extrêmes à $\alpha = 1\%$ et 99% , notés ϕ_{\min}, ϕ_{\max} . Ces angles définissent les vecteurs de coloration extrêmes propres à l'image.

e) Rééchelonnement selon un vecteur de référence

Les contributions de chaque stain sont ensuite normalisées par rapport à des valeurs de référence définies empiriquement. Dans notre cas, les vecteurs de référence (HERef) sont :

$$HE_{\text{ref}} = \begin{bmatrix} 0.5626 & 0.2159 \\ 0.7201 & 0.8012 \\ 0.4062 & 0.5581 \end{bmatrix}$$

et les concentrations maximales de référence (maxCRef) sont :

$$\text{maxC}_{\text{ref}} = [1.9705, 1.0308]$$

Chaque image traitée est ainsi alignée sur un espace colorimétrique commun, réduisant considérablement les variations dues à la provenance (scanner, centre médical...).

3.4.3 Séparation des données : train / validation / test

Partition par WSI : principe et justification

La séparation du jeu de données a été réalisée au niveau des lames entières (WSI) et non des patches, afin de prévenir toute fuite d'information entre les ensembles d'entraînement, de validation et de test (voir la figure 3.14).

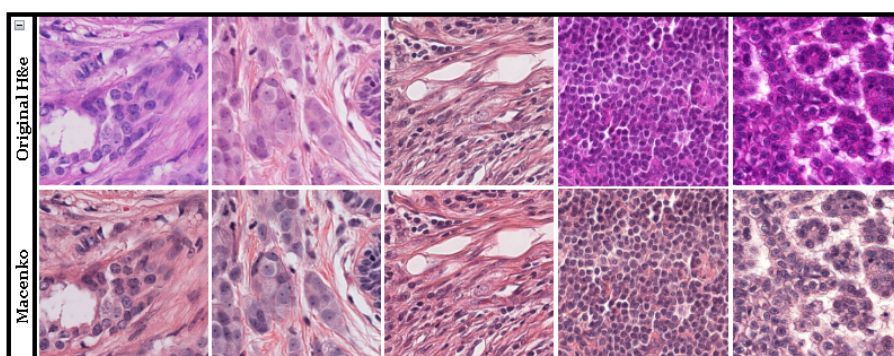


FIGURE 3.13 – Effet visuel de la normalisation Macenko

Ce choix garantit que tous les patches extraits d’une même WSI sont affectés à un seul ensemble, assurant ainsi l’indépendance des splits et la validité des métriques d’évaluation.

La répartition des WSI a été effectuée aléatoirement, mais de manière déterministe (semence fixée), selon les proportions suivantes :

- 70% des WSI pour l’ensemble entraînement (train).
- 20% pour l’ensemble validation (val).
- 10% pour l’ensemble test (test).

Gestion des patches annotés et non annotés

Tous les patches annotés (mitoses et hard negatives) issus des WSI ont été conservés dans les splits correspondants.

En complément, une portion contrôlée de patches non annotés a été ajoutée à l’ensemble d’entraînement, afin de :

- enrichir la diversité du fond (background).
- améliorer la robustesse du modèle en l’exposant à des régions non mitotiques.
- éviter une spécialisation excessive sur les régions annotées.

Concrètement :

- 20% des patches non annotés issus des WSI d’entraînement ont été ajoutés à train de manière aléatoire.
- Les ensembles val et test ont été composés uniquement de patches annotés, afin que l’évaluation se concentre exclusivement sur des cas présentant des objets d’intérêt.

Les patches non annotés sont étiquetés avec un label spécial -1, permettant de les traiter comme négatifs implicites dans le cadre du modèle RetinaNet sans interférer avec la fonction de perte supervisée.

Organisation finale des données

Le processus a permis de générer trois dossiers distincts (train, val, test), chacun contenant les patches correspondant aux WSI allouées.

Un fichier CSV d'annotations au format RetinaNet a été produit pour chaque split, comprenant les informations suivantes :

- Chemin vers le patch.
- Coordonnées des boîtes englobantes (x_min, y_min, x_max, y_max).
- Label de l'objet (1 : mitose, 2 : hard negative, -1 : non annoté).
- ID d'annotation et d'image d'origine.

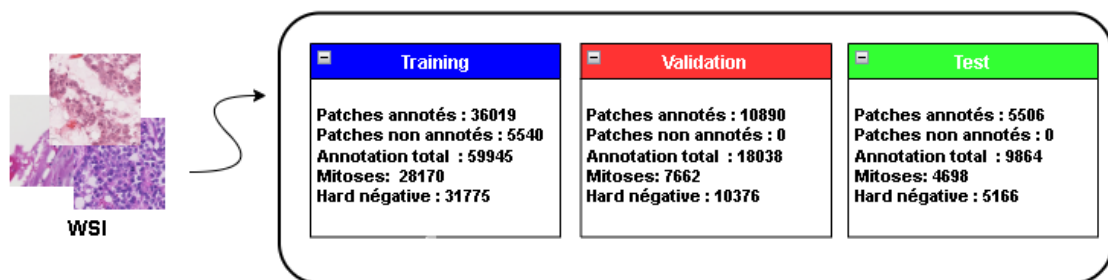


FIGURE 3.14 – Répartition des patches annotés et non annotés dans les ensembles d'entraînement, de validation et de test à partir des lames histopathologiques (WSI)

3.5 Configuration de l'apprentissage

3.5.1 Hyperparamètres d'apprentissage

L'efficacité d'un apprentissage supervisé dépend largement du choix des hyperparamètres, qui doivent être soigneusement adaptés au modèle utilisé, à la nature des données, à la tâche cible, ainsi qu'aux ressources matérielles disponibles. Dans le cadre de ce projet, les hyperparamètres ont été définis à partir d'une analyse empirique préliminaire, tout en s'appuyant sur les recommandations issues de la littérature spécialisée en détection d'objets biomédicaux[21].

► Taille des patches : 512 × 512 pixels

Ce format a été choisi comme compromis optimal entre la conservation du contexte morphologique nécessaire à la reconnaissance des mitoses, et la nécessité de limiter la surface analysée pour éviter une dilution du signal. Des études antérieures ont montré que cette taille permet une bonne représentation des noyaux sans surcharge computationnelle inutile [58].

► Taille du batch : 4

Compte tenu de la taille des images, de la profondeur du modèle RetinaNet (backbone ResNet-50 + FPN), et de la mémoire GPU disponible (8 Go – RTX 3060 Ti), un petit

batch size a été retenu. Cela garantit la stabilité de l'entraînement, tout en permettant l'utilisation de Mixed Precision (AMP) pour optimiser les performances.

► **Taux d'apprentissage initial** : 1×10^{-5}

Un taux d'apprentissage faible est particulièrement recommandé dans le cas des problèmes à classes déséquilibrées, comme celui de la détection de mitoses. Ce choix permet au modèle d'apprendre progressivement sans divergence ni oscillations brutales, en particulier lors des premières itérations [63].

► **Nombre d'époques** : 30

Un plafond de 30 époques a été défini pour encadrer la durée de l'entraînement, tout en offrant au mécanisme d'arrêt précoce (early stopping) la flexibilité nécessaire pour interrompre l'apprentissage en cas de stagnation des performances sur les données de validation.

3.5.2 Stratégie d'optimisation

► **Optimiseur** : Adam

L'optimiseur Adam (Adaptive Moment Estimation) a été adopté dans cette étude pour sa robustesse et sa rapidité de convergence, en particulier dans des contextes à fort gradient de variance comme c'est le cas pour les mitoses rares. Adam combine les avantages de RMSProp et AdaGrad en ajustant dynamiquement les taux d'apprentissage pour chaque paramètre[64].

Ce choix est justifié par :

- la complexité de l'architecture RetinaNet.
- la nature déséquilibrée des classes (mitoses vs. fond).
- le besoin de converger efficacement avec un minimum de réglages manuels.

► **Scheduler** : ReduceLRonPlateau

Pour améliorer la convergence et affiner les derniers stades de l'apprentissage, un scheduler de type ReduceLRonPlateau a été employé. Ce scheduler réduit automatiquement le taux d'apprentissage dès que la métrique mAP de validation stagne pendant un nombre défini d'époques. Les paramètres choisis sont :

Paramètre	Valeur
Mode	max (mAP ↑)
Facteur	0.5
Patience	3 époques

TABLE 3.2 – Hyperparamètres du mécanisme d'arrêt anticipé (early stopping)

Ce mécanisme permet de préserver un apprentissage rapide en début de phase, tout en réduisant la vitesse de descente une fois un palier de performance atteint, pour éviter de sur-apprendre le bruit.

► Arrêt précoce (Early stopping)

Un early stopping a été mis en place avec une patience de 10 époques sans amélioration du mAP. Ce mécanisme permet d'interrompre l'entraînement avant que le modèle ne commence à sur-apprendre les données d'entraînement au détriment de la généralisation.

Ce choix vise à :

- réduire la consommation des ressources (temps, énergie).
- préserver la capacité de généralisation du modèle sur les données non vues.
- favoriser la reproductibilité du système.

3.5.3 Augmentation dynamique des données

En détection d'objets appliquée à la pathologie numérique, les données annotées sont souvent limitées, déséquilibrées, et fortement variables d'un domaine à l'autre (scanner, coloration, coupe, espèce...). Pour faire face à ces défis, une stratégie d'augmentation dynamique a été mise en œuvre afin d'enrichir la diversité des échantillons vus par le modèle à chaque époque [65]. L'augmentation vise à la fois à atténuer le surapprentissage, en exposant le modèle à de multiples variantes d'un même patch, et à améliorer sa robustesse face aux variations inter-domaines (scanners, colorations, tissus...)(voir la figure 3.15).

► Mise en œuvre avec Albumentations

L'augmentation est réalisée dynamiquement à chaque itération d'entraînement, via la bibliothèque Albumentations[66], intégrée au sein du Dataset PyTorch.

Les transformations sont appliquées uniquement sur les données d'entraînement, tandis que les ensembles de validation et de test ne subissent qu'une normalisation simple.

Le pipeline d'augmentation utilisé est résumé dans le tableau suivant :

Transformation	Description	Objectif
D4	Applique l'ensemble des symétries D4 (rotations multiples de 90° et flips)	Simuler différentes orientations de tissus
ColorJitter	Modifie la luminosité, le contraste, la saturation, la teinte	Simuler les variations de coloration H&E
Defocus	Ajoute un flou de défocalisation léger (rayon 1 à 3)	Simuler les imprécisions de mise au point du scanner
Normalize	Centrage des canaux RGB selon les moyennes/écarts-types du dataset	Préparation au traitement CNN
ToTensorV2	Conversion en tenseurs PyTorch + gestion des bounding boxes	Format compatible avec RetinaNet

TABLE 3.3 – Transformations d'augmentation et de prétraitement appliquées aux images histopathologiques

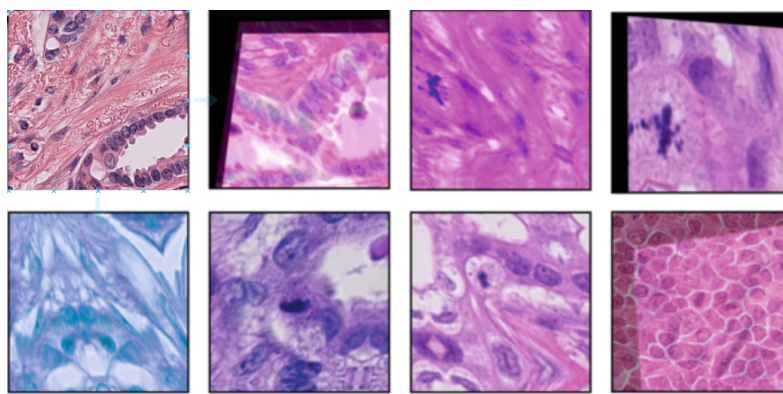


FIGURE 3.15 – Illustration de l'effet de l'augmentation dynamique.

Performance d'entraînement

L'entraînement complet a été effectué sur 30 époques, avec une taille de batch de 4. En raison de la complexité du modèle, de la taille des patches (512×512), et de l'utilisation de calcul mixte (AMP), le temps moyen par époque était d'environ 5389 secondes, soit environ 1h30 par époque.

L'entraînement total a duré environ 45 heures, incluant les phases de validation, de sauvegarde du meilleur modèle (mAP maximal) et les traitements post-détection (voir la figure 3.16).

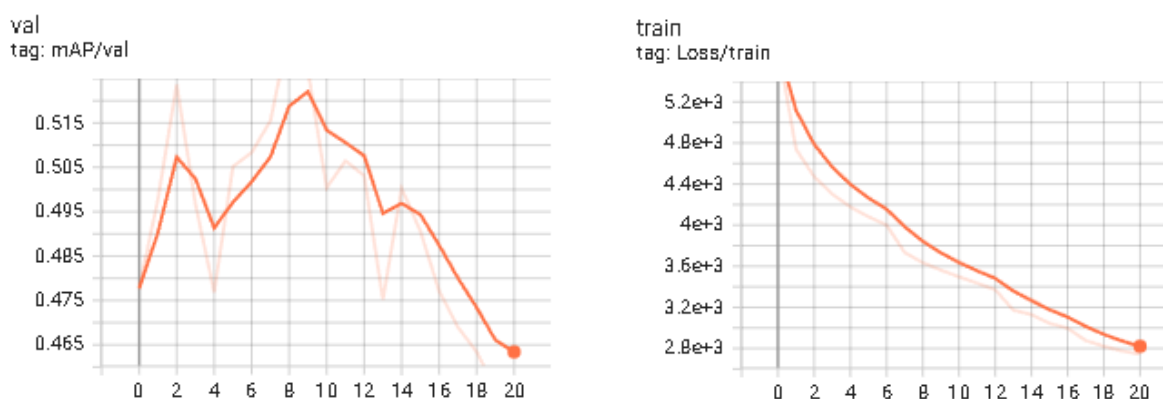


FIGURE 3.16 – Courbe de loss et mAP sur TensorBoard.

Conclusion

Ce chapitre a présenté en détail le pipeline mis en place pour la détection automatisée des mitoses à partir du dataset MIDOG 2022. Il a mis en évidence les défis liés à la diversité des scanners et à la rareté des mitoses, ainsi que les étapes clés de préparation des données : normalisation, découpage en patches et structuration des annotations. L'architecture RetinaNet, avec un backbone ResNet-50 et la Focal Loss, a été choisie pour sa capacité à gérer les objets rares

et déséquilibrés. Ce travail prépare le terrain pour l'implémentation et l'évaluation du système dans le chapitre suivant.

4

Expérimentation, résultats et discussions

Ce quatrième chapitre évalue les performances du modèle RetinaNet appliqué à la détection des mitoses dans les images histopathologiques. Il débute par une description de l'environnement matériel et logiciel utilisé pour l'entraînement et l'évaluation. L'analyse des résultats repose sur des métriques standards telles que la précision, le rappel, le F1-score et le mAP, ainsi que sur une évaluation qualitative à l'aide d'exemples visuels. Une attention particulière est portée à l'interprétabilité du modèle via des cartes de chaleur. Enfin, une section aborde les limites identifiées et propose des perspectives d'amélioration pour une intégration optimale dans une interface destinée aux pathologistes.

Sommaire

Introduction	52
4.1 Environnement matériel et logiciel	53
4.2 Évaluation du modèle	55
4.3 Résultats métriques du modèle	57
4.4 Évaluation qualitative par visualisation des détections	61
4.5 Interprétabilité et visualisation	64
4.6 Interface utilisateur (prototypage)	67
4.7 Perspectives	70
Conclusion	70

4.1 Environnement matériel et logiciel

Le développement, l'entraînement et l'évaluation de notre pipeline de détection des mitoses ont été réalisés dans un environnement matériel et logiciel adapté à des tâches intensives en calcul, typiques de la vision par ordinateur médicale.

4.1.1 Les logiciels

- **Anaconda :**

Anaconda est une distribution libre et open source des langages de programmation Python et R, appliquée au développement d'applications dédiées à la fouille de données et à l'apprentissage automatique (traitement de données à grande échelle, analyse prédictive, calcul scientifique), qui vise à simplifier la gestion des paquets de déploiement.

Les versions de paquetages sont gérées par le système de gestion de paquets conda. La distribution Anaconda est utilisée par plus de 6 millions d'utilisateurs et comprend plus de 250 paquets populaires adaptés à la fouille de données sous Windows, Linux et MacOS [72].



- **Python :**

Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, de l'analyse de données ou encore dans le domaine du développement de logiciels.



- **PyTorch :**

Utilité : Framework principal pour le développement du modèle RetinaNet, la gestion des tenseurs, l'entraînement sur GPU, et les fonctionnalités de deep learning [73].



- **Torchvision :**

Utilité : Fournit des modèles de vision par ordinateur pré-entraînés (dont RetinaNet), des fonctions de transformation d'image, et des utilitaires pour la détection d'objets [74].



- **OpenSlide-Python :**

Utilité : Lecture des lames histologiques numériques (WSI – Whole Slide Images) utilisées dans les bases de données médicales comme MIDOG 2022 [75].



- **NumPy :**

Utilité : Manipulation de tableaux multidimensionnels, opérations vectorielles rapides, traitement des matrices lors de la normalisation des couleurs (Macenko) [76].



- **Pandas :**

Utilité : Traitement des annotations sous forme de tableaux, génération de fichiers CSV pour l'apprentissage, analyse statistique des patches [77].



- **Pillow :**

Utilité : Lecture, modification et sauvegarde des images aux formats PNG/JPG durant les étapes de découpe et de normalisation [78].



- **Scikit-learn :**

Utilité : Division des données en ensembles d'entraînement, validation et test avec la fonction `train_test_split` [79].



- **tqdm :**

Utilité : Affichage de barres de progression conviviales pendant les boucles de traitement, ce qui facilite le suivi des processus longs [80].



- **OpenCV (cv2) :**

Utilité : Traitement d'image avancé, lecture de certains fichiers ou affichage de visuels en complément [81].



- **TensorBoard (via PyTorch) :**

Utilité : Suivi visuel de l'entraînement du modèle à l'aide de courbes (loss, mAP), via la classe `SummaryWriter` de PyTorch [82].



- **customtkinter :**

Utilité : Création d'une interface graphique moderne en Python pour l'affichage des résultats de détection et l'interaction avec l'utilisateur final [83].



- **Autres modules Python :**

- `os`, `shutil`, `pathlib`, `json`, `time` : pour la gestion des fichiers, des chemins, des fichiers JSON d'annotations, et du temps d'exécution.
- `torch.amp` : pour le support du calcul en précision mixte (Mixed Precision Training) à l'aide de `autocast` et `GradScaler`.

4.1.2 Configuration matérielle

Composant	Spécification
GPU	NVIDIA RTX 3060 Ti (8 Go VRAM)
CPU	AMD Ryzen 7 5700X (8 cœurs / 16 threads)
RAM	16 Go DDR4
Stockage	SSD 1 To
Système d'exploitation	Windows 11

TABLE 4.1 – Configuration matérielle utilisée pour les expérimentations

Cette configuration a permis de traiter efficacement les images WSIs et d'entraîner le modèle RetinaNet sans recours à un cluster HPC ou cloud.

4.2 Évaluation du modèle

L'évaluation des performances d'un modèle de détection d'objets constitue une étape clé, notamment en pathologie numérique, où la fiabilité des prédictions peut avoir des conséquences médicales directes. Dans ce travail, plusieurs métriques ont été utilisées pour quantifier les performances du modèle RetinaNet appliqué à la détection des mitoses. Ces indicateurs permettent d'évaluer à la fois la qualité de la classification (mitose vs non-mitose) et la précision de localisation des objets d'intérêt.

4.2.1 Précision (Precision)

La précision mesure la proportion de détections identifiées comme « mitose » qui sont effectivement correctes. Elle se définit par :

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

où :

- TP(True Positives) : nombre de mitoses détectées correctement.
- FP (False Positives) : nombre de faux positifs, c'est-à-dire d'objets détectés comme mitoses mais qui ne le sont pas.

Une précision élevée indique que les prédictions positives sont fiables et que le modèle produit peu de fausses alarmes. Cette métrique est largement utilisée dans les benchmarks de classification et de détection, notamment dans les compétitions VOC et COCO [67].

4.2.2 Rappel (Recall)

Le rappel, ou sensibilité, évalue la capacité du modèle à retrouver toutes les mitoses présentes dans l'image. Il se définit comme suit :

$$\text{Précision} = \frac{TP}{TP + FN}$$

où :

- FN (False Negatives) : nombre de mitoses présentes mais non détectées par le modèle.

Un bon rappel indique que le modèle ne manque que peu d'objets d'intérêt. Cette métrique est primordiale dans des domaines critiques comme la détection de cancer, où l'omission d'un cas peut avoir des conséquences graves [68].

4.2.3 F1-score

Le F1-score combine précision et rappel en une seule valeur via leur moyenne harmonique :

$$F_1 = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}}$$

Le F1-score est particulièrement pertinent dans les cas de déséquilibre de classes, comme ici où les mitoses sont rares par rapport aux tissus sains ou aux fausses alarmes. Il offre un compromis utile pour juger globalement les performances [68].

4.2.4 mean Average Precision

La mAP (Mean Average Precision) est la métrique de référence en détection d'objets. Elle prend en compte à la fois la classification et la localisation à l'aide du critère d'Intersection over Union (IoU) :

$$\text{IoU} = \frac{\text{aire de l'intersection}}{\text{aire de l'union}}$$

Une prédiction est considérée comme correcte si son IoU avec la vérité terrain dépasse un seuil donné (typiquement 0.5). Pour une classe donnée, l'Average Precision (AP) correspond à l'aire sous la courbe précision-rappel. La mAP est la moyenne des AP sur toutes les classes (dans notre cas : mitotic et hard negative), ou sur différents seuils d'IoU (comme mAP@[0.5 :0.95], utilisé dans COCO) [67, 69].

4.2.5 Motivation du choix des métriques d'évaluation

Dans notre étude, les performances du modèle RetinaNet ont été évaluées comme suit :

- Précision, rappel et F1-score calculés à $\text{IoU} \geq 0.5$, pour juger la qualité de classification.
- Le mAP@0.5 a été utilisé comme indicateur principal de performance, particulièrement adapté à la petite taille des objets détectés (mitoses).
- Les métriques ont été extraites à l'aide d'un script personnalisé compatible COCO, suivant les standards établis en vision par ordinateur [67] .

4.3 Résultats métriques du modèle

L'évaluation quantitative du modèle RetinaNet a permis de mesurer ses performances globales et par classe, à partir des prédictions effectuées sur le jeu de validation. Les résultats incluent l'évolution de la fonction de perte au cours de l'entraînement, le score mAP, ainsi que les métriques de classification (précision, rappel, F1-score) pour chaque classe d'objet détecté.

■ Résultats globaux

Le modèle a été entraîné sur 30 époques. La figure ci-dessous présente l'évolution simultanée de la loss d'entraînement et du mAP de validation au fil des époques .

À l'issue de l'entraînement :

- Train Loss final : 0.2501.
- mAP@0.5 (validation) : 0.5722.

Ces résultats montrent une convergence raisonnable du modèle, avec un bon compromis entre sous-apprentissage et sur-apprentissage.

■ Performances par classe

Le tableau suivant présente les métriques de performance pour les deux classes d'intérêt : mitoses et hard negatives. Les mesures ont été effectuées à un seuil $\text{IoU} \geq 0.5$.

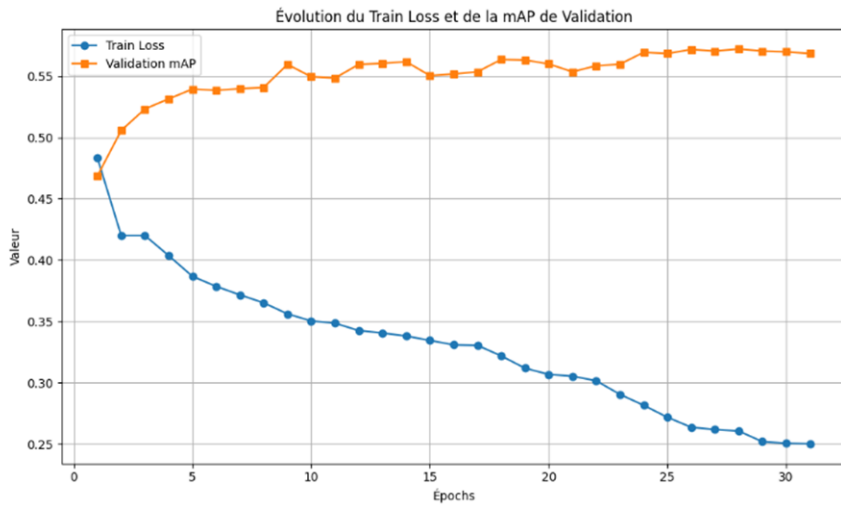


FIGURE 4.1 – Évolution du Train Loss et du mAP au cours des époques.

Classe	Précision	Rappel	F1-score
Mitoses	0.8371	0.7088	0.7676
Hard negative	0.3563	0.4324	0.3907

TABLE 4.2 – Scores de détection pour chaque classe : mitoses et faux positifs (hard negatives).

On observe une meilleure performance sur les mitoses, avec une précision élevée (83.7 %) et un F1-score satisfaisant (76.8 %), tandis que la classe hard negative reste plus difficile à identifier de manière fiable (voir la figure 4.2).

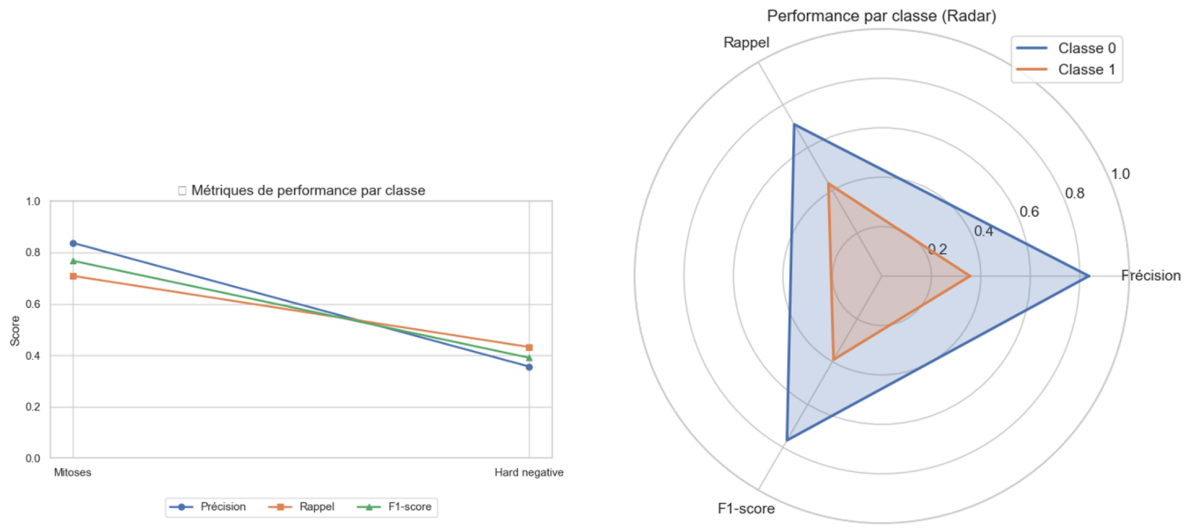


FIGURE 4.2 – Graphique des métriques de performance (précision, rappel, F1-score) par classe.

■ Moyennes globales

En moyennant les résultats sur les deux classes (macro-averaging), on obtient :

- **Précision macro** : 0,5967
- **Rappel macro** : 0,5706
- **F1-score macro** : 0,5791

Ces scores traduisent une performance globalement équilibrée, mais perfectible sur les classes minoritaires ou difficiles.

■ Analyse des TP, FP, FN

Le tableau suivant détaille les vraies détections (TP), faux positifs (FP) et faux négatifs (FN) pour chaque classe :

Classe	TP (vrais positifs)	FP (faux positifs)	FN (faux négatifs)
Mitoses	3330	648	1368
Hard negative	2232	4033	2930

TABLE 4.3 – Nombre de vrais positifs (TP), faux positifs (FP) et faux négatifs (FN) par classe.

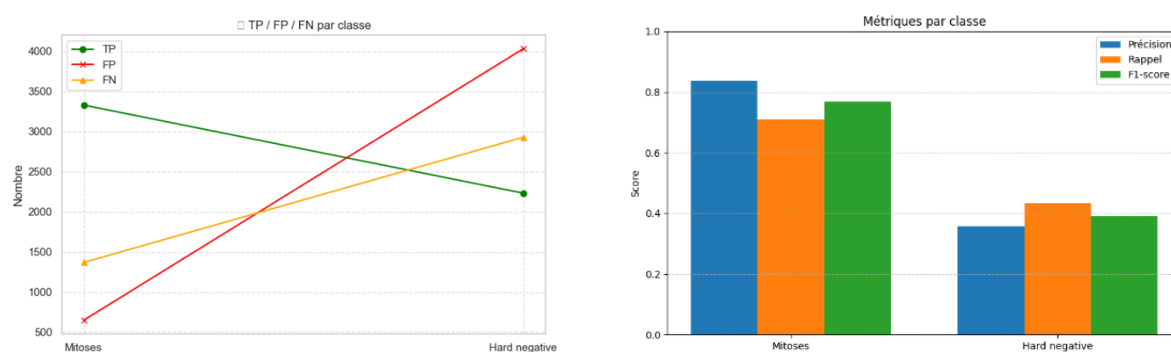


FIGURE 4.3 – Évaluation des Performances par Classe : Analyse des TP/FP/FN et des Métriques de Précision, Rappel et F1-Score.

On constate que :

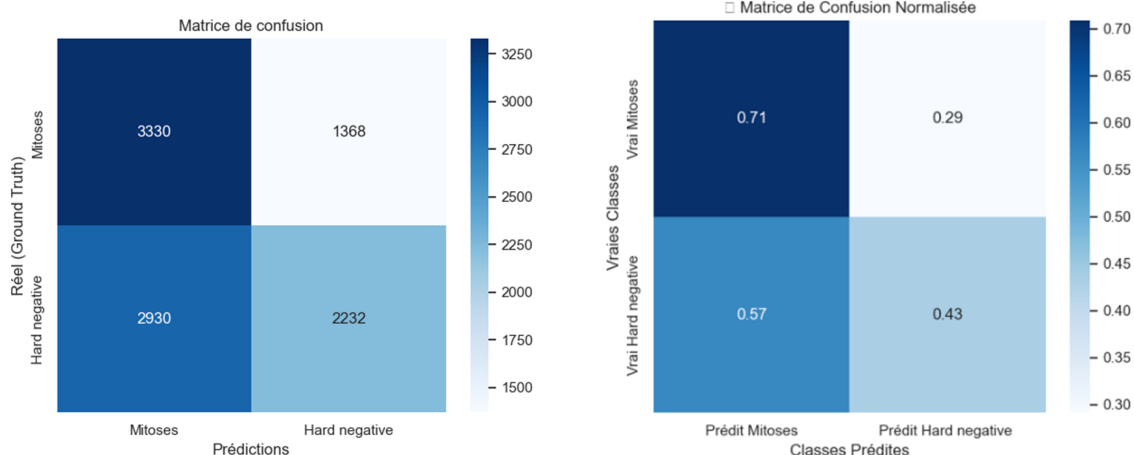
- Le nombre de FP est très élevé pour la classe hard negative (4033), ce qui tire sa précision vers le bas.
- Le modèle manque encore des mitoses (1368 FN), bien que le rappel reste correct.

■ Matrices de confusion

Pour évaluer les confusions entre classes, deux matrices de confusion ont été générées : une en valeurs absolues, l'autre normalisée.

Ces matrices confirment la tendance observée :

- Les mitoses sont généralement bien détectées.
- Les hard negatives sont souvent confondues avec des mitoses, ce qui explique le nombre élevé



(a) Matrice de confusion (valeurs brutes).

(b) Matrice de confusion normalisée.

de FP dans cette classe.

■ Discussion des performances

Les résultats obtenus confirment l'efficacité du modèle RetinaNet dans la détection des mitoses sur les images histopathologiques de la base MIDOG 2022. Avec un F1-score de 0,7676 pour la classe mitose, le modèle atteint des performances satisfaisantes et se rapproche des meilleures approches du challenge MIDOG 2022[84], qui ont obtenu des scores avoisinant 0,78. Malgré ces performances encourageantes, certains points méritent d'être soulignés :

■ Confusion avec les hard negatives :

Le modèle présente des difficultés à différencier les cellules mitotiques des cellules non mitotiques de type hard negatives.

Ces hard negatives correspondent à des cellules qui ne sont pas en division, mais dont l'apparence est visuellement proche de celle des mitoses, ce qui peut induire des erreurs de classification.

Bien que la base de données contienne une proportion plus importante de hard negatives, la stratégie d'apprentissage adoptée dans ce projet s'est principalement concentrée sur la classe mitose. Un meilleur équilibre entre les classes ou une adaptation de la fonction de perte pourrait améliorer la capacité du modèle à mieux distinguer les cellules mitotiques des hard negatives.

■ Défi de la petite taille des objets :

La détection des mitoses reste un défi en raison de leur petite taille et de leur faible contraste dans les images histologiques. Bien que les performances soient globalement satisfaisantes, des optimisations au niveau de la détection des petites structures pourraient encore améliorer les résultats.

■ Importance clinique des faux négatifs :

Dans un contexte médical, les faux négatifs sont particulièrement critiques, car une mitose non détectée peut fausser l'évaluation du grade tumoral. Le rappel obtenu est correct, mais il serait

utile de le renforcer pour assurer une détection plus exhaustive et fiable en situation clinique.

En conclusion, les résultats obtenus sont cohérents et compétitifs par rapport à l'état de l'art. Ils montrent que le modèle développé peut constituer une base solide pour des travaux futurs visant à améliorer encore la précision et la fiabilité de la détection des mitoses.

4.4 Évaluation qualitative par visualisation des détections

L'analyse qualitative des prédictions permet de mieux comprendre le comportement du modèle RetinaNet au-delà des seules métriques numériques. En visualisant des exemples concrets, il est possible d'identifier les points forts du modèle, les types d'erreurs récurrentes, ainsi que les cas limites. Cette démarche est particulièrement importante en pathologie numérique, où la confiance visuelle dans les prédictions est un critère clé d'acceptabilité clinique.

■ Visualisations des prédictions correctes (True Positives)

Les figures suivantes illustrent des cas de mitoses correctement détectées (TP) par le modèle. On observe une bonne concordance entre les prédictions et les annotations de vérité terrain.

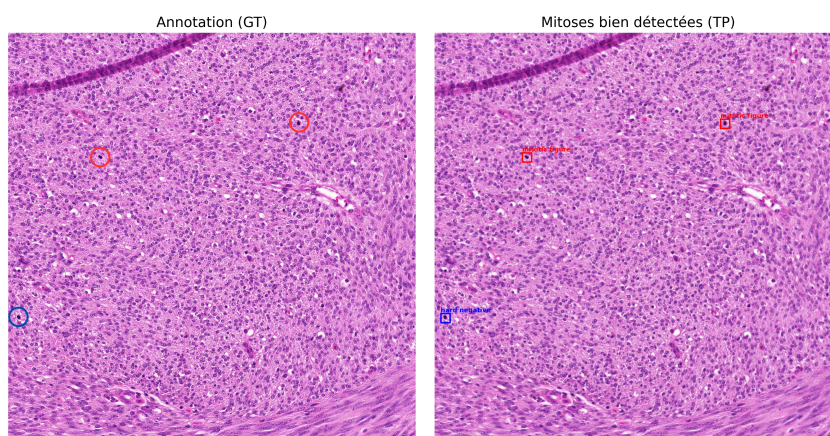


FIGURE 4.5 – Exemples de True Positives (mitoses bien détectées).

Dans la majorité des cas, les mitoses sont détectées avec une précision spatiale adéquate, malgré leur taille réduite et la complexité du contexte tissulaire.

■ Exemples d'erreurs de prédiction

L'identification des erreurs de prédiction permet de cerner les limites du modèle.

- Faux positifs (FP) : le modèle détecte une mitose là où il n'y en a pas.
- Faux négatifs (FN) : le modèle ne détecte pas une mitose annotée par l'expert.

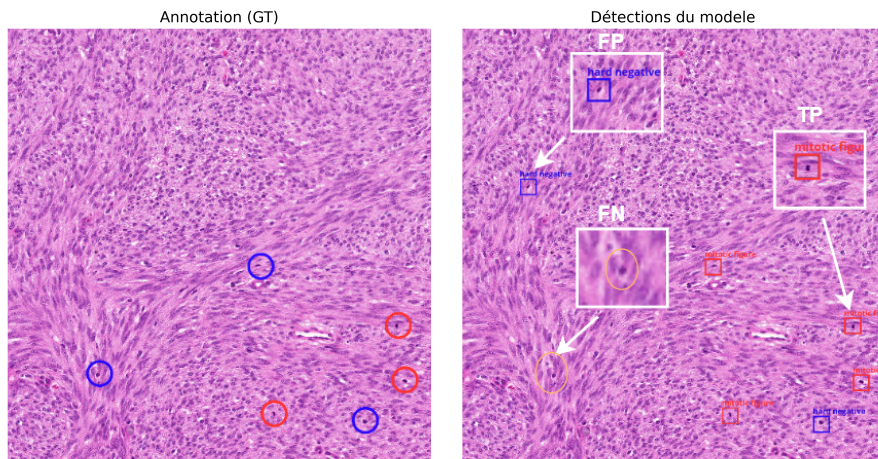


FIGURE 4.6 – Exemples d’erreurs de détection : faux positifs (objets confondus avec des mitoses) et faux négatifs (mitoses non détectées)

On remarque que les FP apparaissent souvent dans des régions denses ou floues, où des structures morphologiquement proches des mitoses (noyaux sombres ou en mitose apoptoïque) peuvent induire des confusions. Les FN sont parfois liées à des mitoses partiellement visibles ou très peu contrastées.

■ **Analyse des motifs d’erreur fréquents**

Les erreurs peuvent être regroupées selon plusieurs motifs récurrents :

Motif d’erreur	Description
Confusion avec artefacts	Certaines zones de pli ou de décoloration sont interprétées comme des mitoses.
Noyaux atypiques	Des noyaux très denses sont confondus avec des mitoses.
Mitoses floues ou partielles	Les mitoses mal sectionnées sont parfois ignorées.
Zones périphériques	Les prédictions en bordure de patch sont souvent imprécises.

TABLE 4.4 – Principaux motifs d’erreurs rencontrés dans la détection automatique des mitoses.

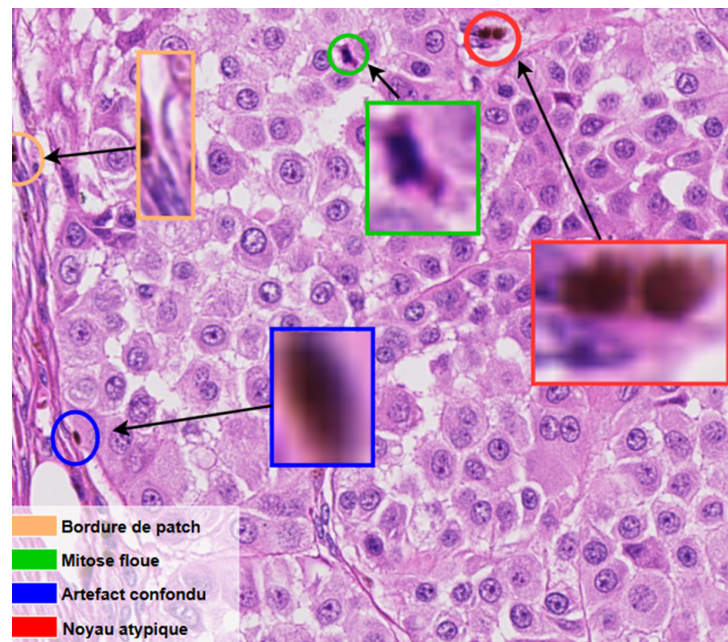


FIGURE 4.7 – Carte annotée des motifs d’erreur fréquents.

■ Comparaison expert vs modèle

Une comparaison directe entre les annotations de l’expert (GT) et les prédictions du modèle est présentée ci-dessous.

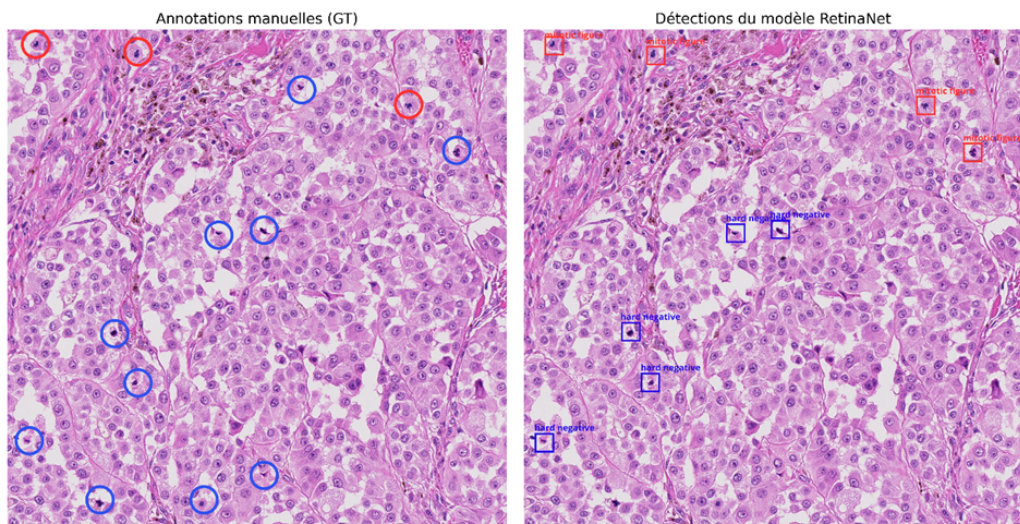


FIGURE 4.8 – Comparaison des prédictions du modèle et des annotations expert .

Ces comparaisons montrent une bonne couverture générale des mitoses par le modèle, avec des écarts localisés dans des zones complexes (amas de cellules, ruptures de colorations).

■ Visualisation intuitive sur lame entière

Pour mieux comprendre le comportement global du modèle sur des images de grande taille, une prédiction a été réalisée sur une lame entière (WSI), en traitant les patches individuellement puis en les reconstruisant(voir la figure 4.9).

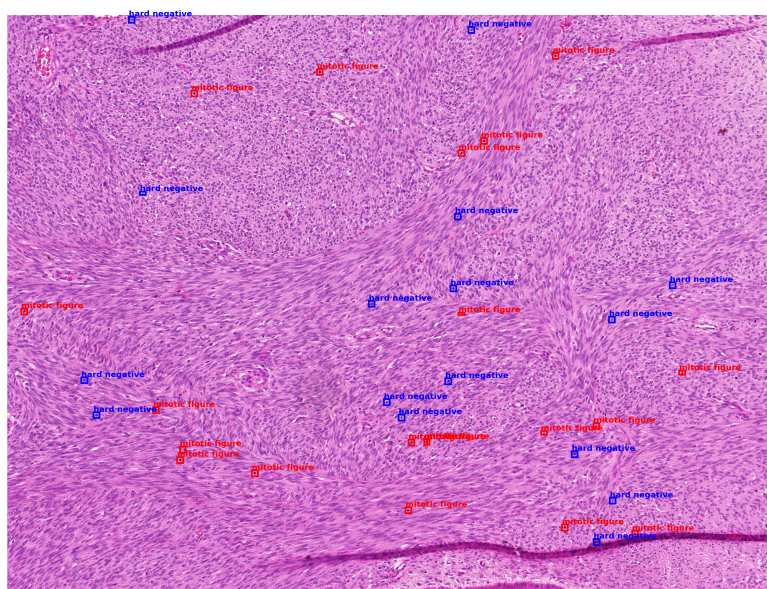


FIGURE 4.9 – visualisation des prédictions RetinaNet sur une lame complète

Cette visualisation intuitive permet :

- De localiser rapidement les zones denses en mitoses.
- De faciliter l'examen humain, en servant de préfiltrage aux experts.
- D'illustrer le potentiel de l'approche pour un déploiement clinique.

4.5 Interprétabilité et visualisation

■ Introduction à l'interprétabilité dans le contexte médical

L'interprétabilité constitue un enjeu central dans le déploiement de modèles d'intelligence artificielle en santé. En pathologie numérique, il est impératif que les prédictions d'un modèle puissent être comprises, vérifiées, et validées par les cliniciens. L'objectif n'est pas uniquement de produire une sortie binaire (mitose ou non-mitose), mais d'offrir un soutien visuel explicite permettant à l'expert de comprendre pourquoi le modèle a pris une décision[70].

Ainsi, l'intégration de méthodes de visualisation telles que les cartes de chaleur (heatmaps) ou les méthodes Grad-CAM (Gradient-weighted Class Activation Mapping) devient un levier d'interprétabilité, particulièrement utile pour les objets de petite taille comme les mitoses [71].

■ Cartes de chaleur et visualisation des régions discriminantes

Dans ce travail, des cartes de chaleur ont été générées afin de visualiser les zones d'attention du modèle sur les images pathologiques. Ces cartes permettent de repérer les régions les plus influentes dans la décision de détection, même sans avoir accès à l'architecture interne du modèle. Deux approches ont été utilisées :

1. Heatmap de densité de prédictions sur lame entière :

En regroupant les prédictions effectuées sur chaque patch (score ≥ 0.5), une carte d'intensité a été construite pour illustrer la distribution spatiale des zones détectées comme mitoses (voir la figure 4.10).

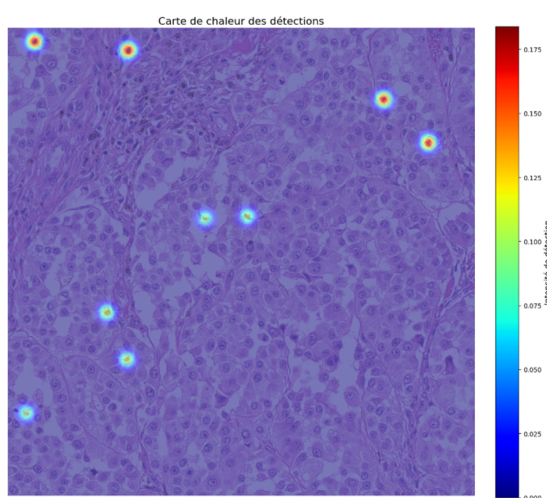


FIGURE 4.10 – Carte de chaleur montrant la densité des prédictions mitotiques sur une lame entière (les régions rouges indiquent une concentration élevée de détections).

2. Visualisation par surimpression des scores de confiance :

Pour chaque patch, une carte peut être créée en superposant les scores de confiance des prédictions à l'image d'origine, en colorant les régions en fonction du niveau de confiance du modèle (voir la figure 4.11).

■ Identification des régions discriminantes

Ces visualisations permettent d'identifier les motifs morphologiques sur lesquels le modèle s'appuie. On observe que :

- Les zones denses en noyaux foncés sont souvent surlignées par les heatmaps.
- Le modèle semble concentrer son attention sur des régions avec des contours bien définis ou des anomalies cytologiques.

Dans certains cas, la heatmap recouvre partiellement une mitose annotée, révélant une attention partielle ou imprécise, ce qui peut expliquer des faux négatifs. À l'inverse, certaines régions non annotées mais marquées par la heatmap sont morphologiquement similaires à des mitoses, révélant des ambiguïtés sémantiques possibles même pour un pathologiste.

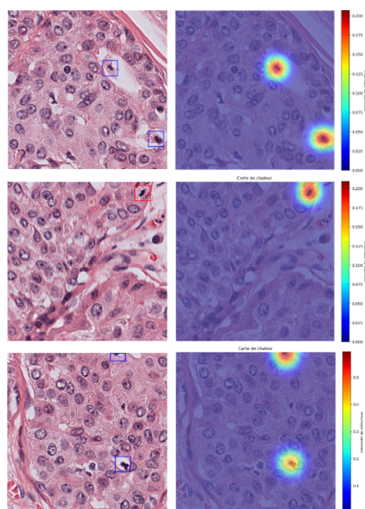


FIGURE 4.11 – Exemple de patch avec surimpression d’une heatmap basée sur les scores de détection (plus le rouge est intense, plus la confiance est élevée).

■ Discussion sur l’acceptabilité clinique

Ces visualisations constituent un outil d’interprétation précieux pour les pathologistes. Elles permettent :

- de comprendre la logique de détection du modèle .
- de valider visuellement certaines décisions du modèle (soutien à l’interprétation humaine) .
- d’identifier les zones douteuses, où le modèle hésite ou surinterprète certains motifs .
- d’évaluer la robustesse du modèle sur des cas hors distribution. L’intégration de ces cartes dans l’interface utilisateur (voir section 5) renforce la transparence du processus décisionnel et facilite une adoption clinique progressive.

4.6 Interface utilisateur (prototypage)

■ Présentation générale

Une interface utilisateur a été développée à l'aide des bibliothèques Tkinter et CustomTkinter pour permettre une interaction intuitive avec le modèle de détection de mitoses. Elle a été pensée pour répondre aux besoins d'utilisateurs techniques ou médicaux, tout en restant simple d'usage.



FIGURE 4.12 – Écran d'accueil de l'application.

■ Fonctionnalités principales

L'interface graphique intègre les fonctionnalités suivantes :

- **Importation d'image** : Chargement d'images histologiques (au format .tiff) via le bouton Importer.
- **Affichage des prédictions** : Après importation, les mitoses détectées sont annotées automatiquement sur l'image.
- **Contrôle du seuil de détection** : Un slider interactif permet de modifier dynamiquement la sensibilité du modèle.
- **Zoom** : Une barre de zoom permet d'ajuster la visualisation de l'image.
- **Statistiques image** : Nombre de mitoses détectées, hard negatives, zoom utilisé, dimensions, etc.
- **Galerie de patches détectés** : Un aperçu visuel de chaque mitose détectée est disponible dans

une galerie.

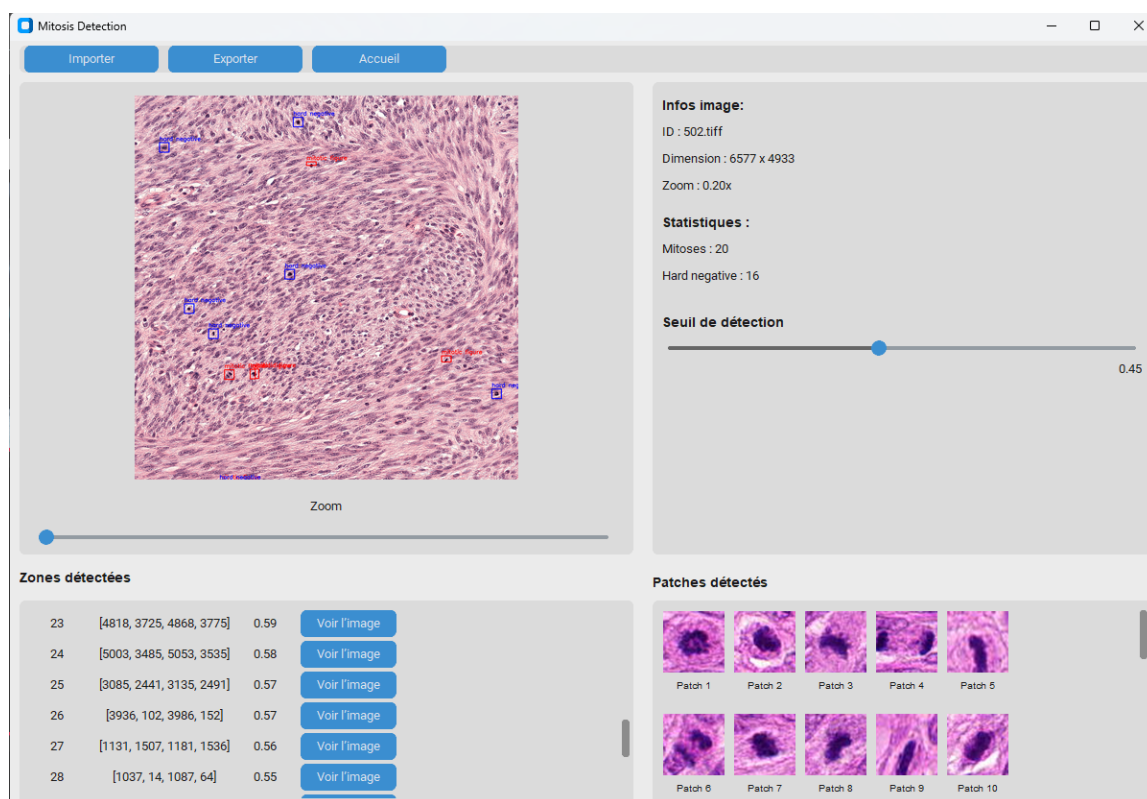


FIGURE 4.13 – Interface principale avec affichage des prédictions, informations image, et navigation par patches détectés.

■ Fonctionnalités avancées

● Affichage des BBox (boîtes englobantes) :

Lorsqu'on clique sur Voir l'image pour un patch détecté, une nouvelle fenêtre s'ouvre, affichant précisément la région d'intérêt (qu'il s'agisse d'une mitose ou d'un hard negative), encadrée par une boîte englobante. Cette visualisation permet une vérification fine par l'utilisateur, essentielle pour l'évaluation clinique des résultats, qu'il s'agisse de confirmer une mitose ou d'identifier un faux positif (voir la figure 4.14).

■ Exploration des résultats CSV

En cliquant sur le bouton Exporter, une nouvelle fenêtre s'ouvre avec une table interactive contenant toutes les détections : coordonnées, scores, classes. Cette table est dérivée du fichier .csv généré après l'analyse. Une option est également disponible pour télécharger le fichier CSV localement pour une analyse externe ou archivage (voir la figure 4.15).

■ Objectifs de conception

L'interface vise trois objectifs principaux :

- **Accessibilité** : Elle est utilisable sans connaissances techniques, avec une interface claire et intuitive.
- **Utilité clinique** : Elle facilite la détection et la validation des mitoses par les pathologistes

grâce à des visualisations précises et interactives.

- **Validation humaine** : Grâce à l’affichage des patches, des BBox et à l’exploration des résultats, l’utilisateur peut vérifier manuellement les prédictions du modèle.

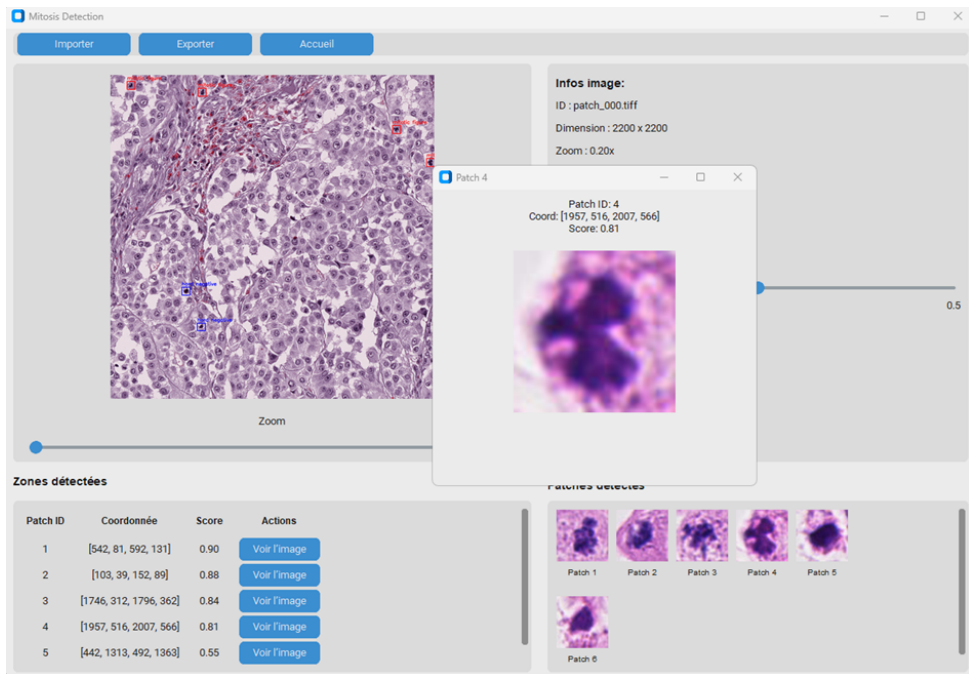


FIGURE 4.14 – Interface de visualisation des zones détectées (mitoses et hard negatives).

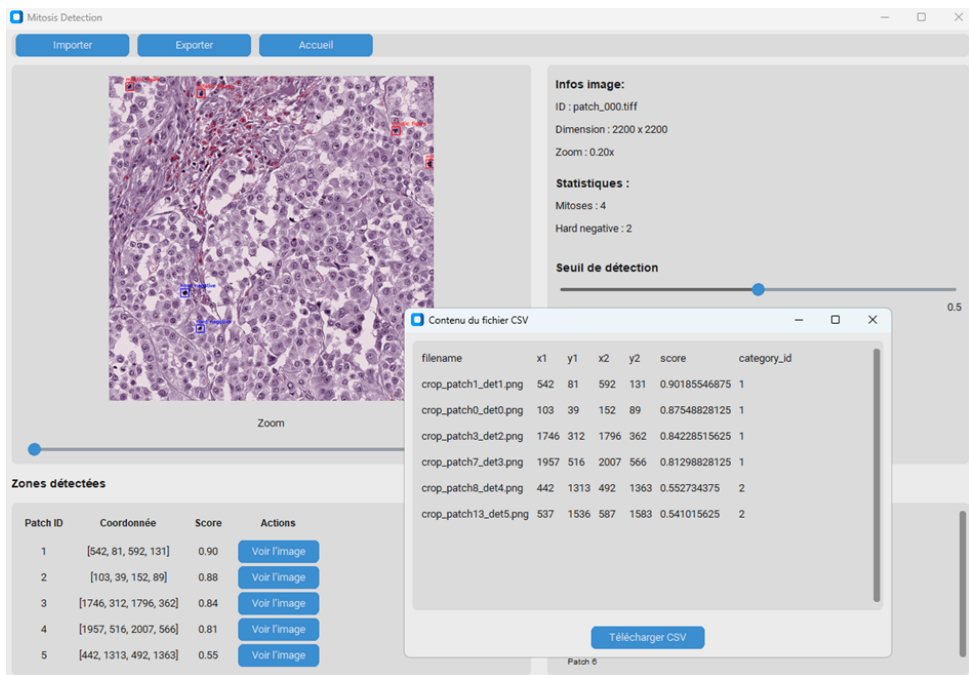


FIGURE 4.15 – Interface d’exploration et d’export des résultats de détection.

4.7 Perspectives

Plusieurs pistes d'amélioration peuvent être envisagées afin de renforcer la performance et la robustesse du modèle développé.

L'enrichissement du jeu de données par génération synthétique ou par diversification des cas d'apprentissage permettrait d'accroître la variabilité des images et d'améliorer la capacité de généralisation du modèle.

Il serait également pertinent d'explorer des architectures plus spécialisées dans la détection d'objets de petite taille, ou de recourir à des approches basées sur des ensembles de modèles afin d'augmenter la fiabilité des prédictions.

Par ailleurs, l'intégration de modèles plus récents tels que les Transformers visuels, reconnus pour leurs performances en vision par ordinateur, pourrait offrir une meilleure prise en compte du contexte spatial et contribuer à une détection plus précise dans des environnements complexes.

Enfin, dans une perspective d'application clinique, une évaluation en conditions réelles en collaboration avec des professionnels de santé serait indispensable afin de valider la fiabilité et l'utilité du système dans un cadre médical concret.

Conclusion

Ce chapitre a présenté une évaluation complète des performances du modèle RetinaNet pour la détection des mitoses. Après avoir décrit l'environnement matériel et logiciel utilisé, nous avons analysé les résultats quantitatifs (précision, rappel, F1-score, mAP) ainsi que des visualisations qualitatives pour interpréter les prédictions. L'ensemble valide l'efficacité du modèle, tout en révélant certaines limites à prendre en compte. Enfin, des perspectives d'amélioration ont été proposées afin d'optimiser la robustesse et de favoriser l'intégration du système dans une interface dédiée aux experts en pathologie.

Conclusion générale et travaux futures

Ce mémoire a porté sur le développement d'un système d'aide au diagnostic basé sur l'apprentissage profond, spécifiquement conçu pour la détection automatique des figures mitotiques dans les images histopathologiques. En s'appuyant sur une méthodologie rigoureuse et sur les approches de l'état de l'art, nous avons exploré les enjeux scientifiques, techniques et cliniques liés à cette problématique.

Notre solution, fondée sur l'architecture RetinaNet, intègre des réseaux de neurones convolutifs, la Focal Loss pour traiter le déséquilibre des classes, et les Feature Pyramid Networks pour gérer la détection multiscalaire. Les résultats obtenus, tant sur le plan quantitatif (mAP, F1-score) que qualitatif (visualisation des prédictions), valident la pertinence de cette approche.

Nous avons également mis en évidence l'intérêt du deep learning pour la détection de structures rares et complexes, comme les mitoses, et évalué les performances du modèle sur des données variées issues du jeu MIDOG 2022. Sur le plan applicatif, le système développé représente un outil prometteur pour assister les pathologistes, en réduisant le temps d'analyse et en améliorant la reproductibilité des diagnostics. L'intégration d'une interface utilisateur adaptée facilite son adoption dans un cadre clinique.

Sur le plan théorique, ce travail propose une méthodologie reproductible, de la préparation des données à l'évaluation, et illustre comment les modèles de détection peuvent être adaptés aux exigences du domaine médical. Cependant, certaines limites subsistent, notamment liées à la variabilité des scanners, à la qualité des lames et à l'absence d'une validation en environnement clinique réel. L'annotation manuelle demeure également une contrainte.

Ces observations ouvrent plusieurs perspectives de recherche : explorer des modèles plus récents comme les Transformers visuels, évaluer la généralisation sur des bases externes, et intégrer des indices morphologiques et contextuels via des approches multimodales. Une validation clinique future, en partenariat avec des pathologistes, serait cruciale pour confirmer la robustesse du système en pratique réelle.

En conclusion, ce mémoire met en lumière le potentiel de l'intelligence artificielle, et en particulier du deep learning, pour répondre à un besoin clinique concret : la détection précise et automatisée des figures mitotiques. Il constitue une contribution concrète et évolutive dans le domaine de la pathologie computationnelle, et ouvre la voie à des outils médicaux plus intelligents, fiables et intégrés.

Bibliographie

- [1] C. W. Elston and I. O. Ellis. Pathological prognostic factors in breast cancer. histological grade. *Histopathology*, 19(5) :403–410, 1991.
- [2] E. A. Rakha et al. Breast cancer histologic grading : past, present and future. *Pathology Research and Practice*, 205(4) :210–218, 2009.
- [3] A. Goldhirsch et al. Personalizing the treatment of women with early breast cancer : highlights of the st gallen international expert consensus. *Annals of Oncology*, 24(9) :2206–2223, 2013.
- [4] S. C. Lester. *Manual of Surgical Pathology*. Elsevier, 3 edition, 2010.
- [5] P. Bándi et al. Comparison of different mitosis detection methods in histopathological images. *Modern Pathology*, 34 :1347–1354, 2021.
- [6] M. Veta et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, 20(1) :237–248, 2015.
- [7] D. Komura and S. Ishikawa. Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*, 16 :34–42, 2018.
- [8] M. Aubreville, C. A. Bertram, R. Klopffleisch, and A. Maier. The midog challenge : Mitotic figure detection across multiple scanners. <https://arxiv.org/abs/2203.12000>, 2022. arXiv preprint arXiv :2203.12000.
- [9] DMI Medical USA. Aperio at2 dx system. <https://dmimedicalusa.com/product/aperio-at2-dx-system>, 2025. Consulté en mai 2025.
- [10] D. N. Louis et al. Computational pathology : From principles to applications. *Archives of Pathology & Laboratory Medicine*, 145(7) :798–809, 2021.
- [11] P. Bankhead, M. B. Loughrey, J. A. Fernández, et al. Qupath : Open source software for digital pathology image analysis. *Scientific Reports*, 7 :16878, 2017.
- [12] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis : A review. *IEEE Reviews in Biomedical Engineering*, 2 :147–171, 2009.
- [13] F. Chollet. *Deep Learning with Python*. Manning Publications, 2 edition, 2021.
- [14] S. Russell and P. Norvig. *Artificial Intelligence : A Modern Approach*. Pearson, 4 edition, 2020.
- [15] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553) :436–444, 2015.
- [16] Y. Zhang and K. Moon. A brief survey on machine learning techniques in medical imaging. *Computers in Biology and Medicine*, 139 :104929, 2021.

- [17] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.
- [18] Kavya Ramkumar, C. Allison Stewart, Kasey R. Cargill, and Della Corte et al. Axl inhibition induces dna damage and replication stress in non-small cell lung cancer cells and promotes sensitivity to atr inhibitors. *Molecular Cancer Research*, 19(3) :485–497, 2021.
- [19] R. S. Sutton and A. G. Barto. *Reinforcement Learning : An Introduction*. MIT Press, 2 edition, 2018.
- [20] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
- [21] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42 :60–88, 2017.
- [22] W. Rawat and Z. Wang. Deep convolutional neural networks for image classification : A comprehensive review. *Neural Computation*, 29(9) :2352–2449, 2017.
- [23] H. Irshad. Automated mitosis detection in histopathology using morphological and multi-channel statistics features. *Journal of Pathology Informatics*, 4 :1–6, 2013.
- [24] G. Madzarov, D. Gjorgjevikj, and I. Chorbev. A multi-class svm classifier utilizing binary decision tree. *Informatica*, 33 :233–241, 2009.
- [25] A. Tashk, M. S. Helfroush, H. Danyali, and M. Akbarzadeh. An automatic mitosis detection method for breast cancer histopathology slide images based on objective and pixel-wise textural features classification. In *Proceedings of the 5th Conference on Information and Knowledge Technology*, pages 406–410, Shiraz, Iran, 2013.
- [26] C. Sommer, L. Fiaschi, F. A. Hamprecht, and D. W. Gerlich. Learning-based mitotic cell detection in histopathological images. In *Proceedings of the 21st International Conference on Pattern Recognition*, pages 2306–2309, Tsukuba, Japan, 2012.
- [27] C. Sommer, C. Straehle, U. Kothe, and F. A. Hamprecht. Ilastik : Interactive learning and segmentation toolkit. In *IEEE International Symposium on Biomedical Imaging : From Nano to Macro*, pages 230–233, Chicago, IL, USA, 2011.
- [28] M. Held, M. H. A. Schmitz, B. Fischer, T. Walter, B. Neumann, M. H. Olma, M. Peter, J. Ellenberg, and D. W. Gerlich. Cellcognition : time-resolved phenotype annotation in high-throughput live cell imaging. *Nature Methods*, 7 :747–754, 2010.
- [29] A. Paul, A. Dey, D. P. Mukherjee, J. Sivaswamy, and V. Tourani. Regenerative random forest with automatic feature selection to detect mitosis in histopathological breast cancer images. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 94–102, Munich, Germany, 2015.
- [30] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.
- [31] O. S. Al-Kadi. Texture measures combination for improved meningioma classification of histopathological images. *Pattern Recognition*, 43(6) :2043–2053, 2010.
- [32] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. 2012.
- [33] M. Veta, J. P. Pluim, P. J. van Diest, and M. A. Viergever. Breast cancer histopathology image analysis : A review. 2014.

- [34] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. pages 2980–2988, 2017.
- [35] G. Litjens, C. I. Sánchez, N. Timofeeva, and et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. 2016.
- [36] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention*, pages 411–418, Nagoya, Japan, 2013.
- [37] C. D. Malon and E. Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. *Journal of Pathology Informatics*, 4 :1–5, 2013.
- [38] H. Wang, A. Cruz-Roa, A. Basavanthally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi. Cascaded ensemble of convolutional neural networks and handcrafted features for mitosis detection. In *Proceedings of the SPIE Medical Imaging*, pages 1–10, San Diego, CA, USA, 2014.
- [39] H. Chen, Q. Dou, X. Wang, J. Qin, and P.-A. Heng. Mitosis detection in breast cancer histology images via deep cascaded networks. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1160–1166, Phoenix, AZ, USA, 2016.
- [40] C. Li, X. Wang, W. Liu, and L. J. Latecki. Deepmitosis : Mitosis detection via deep detection, verification and segmentation networks. *Medical Image Analysis*, 45 :121–133, 2018.
- [41] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39 :1137–1149, 2017.
- [42] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, Las Vegas, NV, USA, 2016.
- [43] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy. Going deeper in spiking neural networks : Vgg and residual architectures. *Frontiers in Neuroscience*, 13 :1–10, 2019.
- [44] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, and J. Huang. Weakly supervised mitosis detection in breast histopathology images using concentric loss. *Medical Image Analysis*, 53 :165–178, 2019.
- [45] D. Cai, X. Sun, N. Zhou, X. Han, and J. Yao. Efficient mitosis detection in breast cancer histology images by rcnn. In *IEEE 16th International Symposium on Biomedical Imaging*, pages 919–922, Venice, Italy, 2019.
- [46] Y. Li, E. Mercan, S. Knezevitch, J. G. Elmore, and L. G. Shapiro. Efficient and accurate mitosis detection—a lightweight rcnn approach. In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, pages 69–77, Funchal, Portugal, 2018.
- [47] V. Dodballapur, Y. Song, H. Huang, M. Chen, W. Chrzanowski, and W. Cai. Mask-driven mitosis detection in histopathology images. In *IEEE 16th International Symposium on Biomedical Imaging*, pages 1855–1859, Venice, Italy, 2019.
- [48] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988, Venice, Italy, 2017.

- [49] F. Chollet. Xception : Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1800–1807, Honolulu, HI, USA, 2017.
- [50] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. pages 580–587, 2014.
- [51] R. Girshick. Fast r-cnn. pages 1440–1448, 2015.
- [52] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. 28, 2015.
- [53] M. Benabdelouahab, H. Hachimi, and M. Bellafkih. Cnn architectures review for medical image segmentation. *IOP Conference Series : Earth and Environmental Science*, 585(1) :012142, 2020.
- [54] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once : Unified, real-time object detection. pages 779–788, 2016.
- [55] Glenn Jocher et al. Yolov5 by ultralytics. 2020. Accessed : 2025-06-02.
- [56] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv :1506.02640*, 2015.
- [57] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. pages 2117–2125, 2017.
- [58] M. Aubreville, C. A. Bertram, M. Veta, C. Marzahl, R. Klopffleisch, A. Maier, and et al. Midog : The mitosis domain generalization challenge. *Scientific Data*, 10 :105, 2023.
- [59] MIDOG Challenge Organizers. Midog 2022 dataset on zenodo. 2022. Accessed : 2025-06-03.
- [60] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. pages 770–778, 2016.
- [61] M. Montiel-Montiel, U. Orozco-Rosas, M. Sanchez-Adame, E. Tlelo-Cuautle, and J. Reyes-Reyes. Evaluation method of deep learning-based embedded systems for traffic sign detection. *IEEE Access*, 9 :101737–101752, 2021.
- [62] Marc Macenko, Marc Niethammer, J.S. Marron, David Borland, J.T. Woosley, Xiaojun Guan, Claus Schmitt, and N.E. Thomas. A method for normalizing histology slides for quantitative analysis. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1107–1110. IEEE, 2009.
- [63] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller, editors, *Neural Networks : Tricks of the Trade*, pages 437–478. Springer, 2012.
- [64] Diederik P. Kingma and Jimmy Ba. Adam : A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015. arXiv :1412.6980.
- [65] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1) :60, 2019.
- [66] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations : Fast and flexible image augmentations. *Information*, 11(2) :125, 2020.

- [67] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco : Common objects in context. In *European Conference on Computer Vision (ECCV)*. Springer, 2014.
- [68] Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, 10(3) :e0118432, 2015.
- [69] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2) :303–338, 2010.
- [70] Andreas Holzinger, Chris Biemann, Constantinos S Pattichis, and Douglas B Kell. What do we need to build explainable ai systems for the medical domain? *arXiv preprint arXiv :1712.09923*, 2017.
- [71] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam : Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.
- [72] Belmehdi Yousra and Messili Kenza. Détection de pathologies pour l’aide au diagnostic en ophtalmologie. Mémoire de fin d’études, Université Saad Dahlab de Blida, 2020.
- [73] Adam Paszke, Sam Gross, Francisco Massa, and et al. Pytorch : An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.
- [74] PyTorch Team. Torchvision. <https://pytorch.org/vision/stable/>, 2025.
- [75] OpenSlide. A c library for reading whole-slide images. <https://openslide.org/>, 2025.
- [76] Charles R Harris, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *Nature*, 585 :357–362, 2020.
- [77] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51–56, 2010.
- [78] Alex Clark. Pillow : Python imaging library. <https://python-pillow.org/>, 2025.
- [79] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn : Machine learning in python. *Journal of Machine Learning Research*, 12 :2825–2830, 2011.
- [80] tqdm Developers. tqdm : A fast, extensible progress bar for python. <https://github.com/tqdm/tqdm>, 2025.
- [81] Gary Bradski. The opencv library. *Dr. Dobb’s Journal of Software Tools*, 2000.
- [82] TensorFlow Authors. Tensorboard : Visualizing learning. <https://www.tensorflow.org/tensorboard>, 2025.

-
- [83] Tom Schimansky. Customtkinter : Modern gui for python. <https://github.com/TomSchimansky/CustomTkinter>, 2025.
- [84] Grand Challenge. Midog 2022 - mitosis domain generalization challenge leaderboard. <https://midog2022.grand-challenge.org/evaluation/challenge/leaderboard/>, 2022. Accessed : 2025-06-24.

Résumé

Dans le contexte croissant de la médecine numérique et face aux limites du diagnostic histopathologique manuel, ce mémoire s'inscrit dans une démarche d'innovation technologique visant à automatiser la détection des mitoses dans les images histopathologiques. L'objectif principal est de concevoir un système d'aide au diagnostic basé sur l'apprentissage profond, en particulier l'architecture RetinaNet, afin d'améliorer la précision et la reproductibilité de l'identification des figures mitotiques, marqueurs essentiels de la prolifération tumorale.

La méthodologie adoptée repose sur l'utilisation du jeu de données MIDOG 2022, composé de lames virtuelles annotées issues de différents centres. Les images ont été prétraitées et découpées en patches, puis utilisées pour entraîner un modèle RetinaNet intégrant la Focal Loss et les Feature Pyramid Networks (FPN). Une interface utilisateur a également été développée pour visualiser les résultats.

Les résultats montrent que le modèle proposé atteint des performances satisfaisantes en termes de précision (mAP, F1-score), tout en conservant une bonne capacité de généralisation. L'étude met en évidence la capacité du deep learning à traiter des données visuelles complexes et à offrir une alternative fiable au comptage manuel.

Mots-clés : pathologie numérique, mitoses, deep learning, RetinaNet, détection d'objets, images histopathologiques, IA médicale.

Abstract

In the growing context of digital medicine and the limitations of manual histopathological diagnosis, this thesis adopts a technological innovation approach aimed at automating the detection of mitotic figures in histopathology images. The main objective is to develop a diagnostic support system based on deep learning, specifically the RetinaNet architecture, to improve the accuracy and reproducibility of mitosis identification, a key marker of tumor proliferation.

The adopted methodology is based on the MIDOG 2022 dataset, which includes annotated whole slide images (WSIs) from multiple medical centers. The images were preprocessed and divided into patches, then used to train a RetinaNet model integrating Focal Loss and Feature Pyramid Networks (FPN). A user interface was also developed to visualize the detection results.

The results demonstrate that the proposed model achieves satisfactory performance in terms of accuracy (mAP, F1-score) while maintaining good generalization capabilities. The study highlights the ability of deep learning to handle complex visual data and offer a reliable alternative to manual mitosis counting.

Keywords : digital pathology, mitosis, deep learning, RetinaNet, object detection, histopathology images, medical AI.
