

Classification de Whole Slide Images par Chemins Spatiaux de Patchs sous Supervision Faible

Wassim Chikhi

Master 2 Vision et Machines Intelligentes – 2025/2026

La soumission comprend un notebook Jupyter (code, figures et résultats) ainsi que ce rapport \LaTeX .

Auteur : Wassim Chikhi

1. Introduction

La classification de *Whole Slide Images* (WSI) en histopathologie constitue un défi majeur en analyse d'images médicales. Les WSI sont des images de très haute résolution, rendant toute annotation locale exhaustive extrêmement coûteuse et souvent irréaliste.

Dans de nombreux contextes cliniques, seules des annotations globales au niveau de la lame (WSI-level), indiquant la présence ou l'absence de tissu tumoral, sont disponibles. Ce travail s'inscrit dans ce cadre de **supervision faible**, où l'objectif n'est pas de prédire des labels locaux au niveau des patchs, mais de produire une **décision globale fiable au niveau WSI**.

L'approche proposée exploite la **cohérence spatiale du tissu** en représentant chaque WSI comme un ensemble de chemins spatiaux de patchs, modélisés comme des séquences ordonnées et agrégées selon un schéma de type *Multiple Instance Learning* (MIL).

2. Données et Partitionnement

2.1. Dataset

Les expériences sont menées sur un sous-ensemble du dataset CAMELYON16, composé de **54 WSI** annotées au niveau global.

Split	Total	Normal	Tumoral
Train	39	17	22
Validation	8	3	5
Test	9	4	5

TABLE 1. Répartition des WSI par split et par classe.

2.2. Stratégie de partitionnement

Le partitionnement est réalisé strictement au niveau WSI. Tous les objets dérivés d'une même lame (patchs, embeddings, chemins) appartiennent au même split, garantissant l'absence totale de fuite d'information (*data leakage*).

3. Extraction et Filtrage des Patchs

Chaque WSI est découpée en patchs de taille **256×256 pixels** selon une grille régulière. Un filtrage automatique est appliqué afin de supprimer les patchs non informatifs, notamment ceux dominés par le fond, flous ou uniformes.

Pour chaque patch conservé, les métadonnées spatiales (x, y) et l'identifiant de la WSI sont stockés.

4. Embeddings Patch-Level

Chaque patch est projeté dans un espace latent à l'aide d'un backbone pré-entraîné.

Deux architectures ont été évaluées :

- ViT-B/16 supervisé ImageNet (baseline),

- **DINOv2 ViT-B/14 auto-supervisé** (choix final).

Dans la version finale, chaque patch est représenté par un vecteur $\mathbf{e}_i \in \mathbb{R}^{768}$. Les embeddings sont gelés pour l'ensemble du pipeline. Une réduction de dimension par PCA est utilisée uniquement à des fins exploratoires et de visualisation.

5. Modélisation Spatiale

Les patches sont modélisés comme des noeuds disposés sur une grille spatiale. Une connectivité locale est définie afin de capturer la proximité topologique entre patchs adjacents. Cette modélisation repose exclusivement sur la structure spatiale, sans information sémantique locale.

6. Génération de Chemins Spatiaux

Pour chaque WSI, un ensemble de chemins spatiaux est généré sous la forme de **random walks** sur le graphe spatial.

Les chemins respectent les contraintes suivantes :

- absence de revisite de noeuds,
- arrêt anticipé en cas de blocage local,
- longueur variable, bornée par des seuils minimaux.

Chaque chemin est représenté par une séquence :

$$\mathbf{X} \in \mathbb{R}^{L \times 768}$$

Ces chemins capturent une continuité morphologique locale sans correspondre à une trajectoire physique réelle.

7. Modèle Séquentiel

Chaque chemin est encodé indépendamment à l'aide d'un Transformer de type *encoder*. Les séquences de longueur variable sont gérées via un padding explicite et un masque d'attention.

Une projection linéaire $768 \rightarrow 256$ est appliquée avant l'encodage. Le modèle produit un **score scalaire par chemin**, interprété comme un logit associé à la classe tumorale.

Chaque chemin hérite du label WSI, ce qui introduit un bruit de supervision assumé et intrinsèque au cadre de supervision faible.

8. Agrégation MIL et Entraînement

Une WSI est modélisée comme un sac de chemins. Les scores chemin-level sont agrégés afin de produire une prédiction unique au niveau WSI.

L'agrégation finale retenue est un **Top-K mean pooling** appliqué aux probabilités chemin-level, permettant de mettre l'accent sur les chemins les plus informatifs tout en réduisant la variance par rapport au max-pooling.

À titre de comparaison, une agrégation par **log-sum-exp** sur les logits est évaluée dans une étude d'ablation.

L'entraînement utilise la fonction de perte BCEWithLogitsLoss avec pondération des classes. Un arrêt précoce est appliqué sur la performance WSI-level du jeu de validation (ROC-AUC).

9. Évaluation

Les performances sont évaluées exclusivement au niveau WSI.

Les métriques utilisées sont :

- ROC-AUC (métrique principale, indépendante du seuil),
- Accuracy et balanced accuracy (secondaires),
- Matrice de confusion, avec seuil ajusté pour privilégier la sensibilité.

Le jeu de test contient 9 WSI ; les résultats sont interprétés avec prudence et dans une optique méthodologique.

10. Discussion et Limites

La supervision chemin-level est intrinsèquement bruitée, en particulier lorsque les régions tumorales sont spatialement localisées. De plus, les chemins générés peuvent être partiellement corrélés, ce qui limite l'hypothèse d'indépendance classique en MIL.

L'approche proposée privilégie la validité méthodologique et la cohérence spatiale, au prix d'une performance absolue limitée par la taille du dataset.

11. Conclusion

Ce travail propose une approche de classification de WSI sous supervision faible fondée sur des chemins spatiaux de patchs et une agrégation MIL au niveau WSI. Sans recourir à des annotations locales, la méthode exploite la structure spatiale du tissu pour produire une décision globale cohérente et cliniquement interprétable.

Les résultats obtenus illustrent le potentiel de cette formulation tout en mettant en évidence les limites inhérentes à la supervision faible et aux jeux de données de taille réduite.