

TITRE : Deep clustering pour le diagnostic médical des cancers : transfert et adaptation de caractéristiques générales sans supervision dans les images de microscopie.

Mots clé : Transfer learning, domain adaptation, deep clustering, medical imaging, histopathologie

Encadrants:

- Arnaud Abreu – aabreu@unistra.fr
- Pilar Ortega – maria-del-pilar.ortega-arevalo@inserm.fr
- Sandrine Mouysset - sandrine.Mouysset@irit.f

Description :

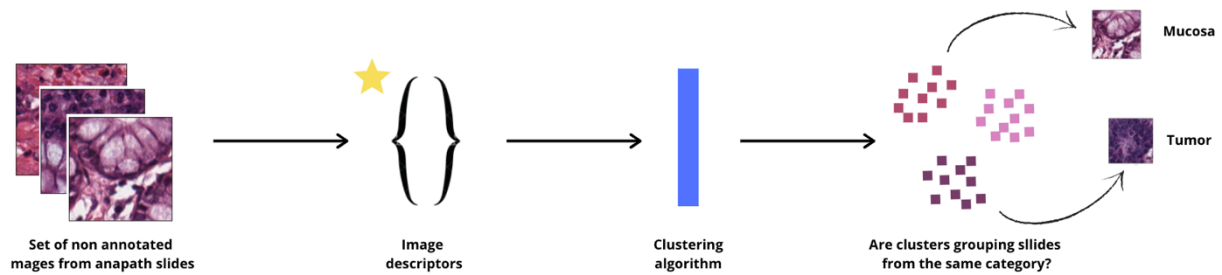
L'anatomopathologie (anapath) est la discipline médicale qui établit les diagnostics à partir de caractéristiques macroscopiques et microscopiques des tissus. La numérisation des images microscopiques (Whole Slide Imaging, WSI) a permis le développement d'algorithmes d'IA destinés à aider les pathologistes dans leur diagnostic [1].

Dans le domaine de la pathologie numérique, le problème principal pour appliquer des algorithmes de Deep Learning, est le manque de bases de données suffisamment grandes et bien annotées. Aussi, une solution développée pour un hôpital ou un type de tissu particulier est rarement généralisable à d'autres environnements avec des conditions différentes. Cela signifie que le développement et l'utilisation du deep learning sont limités aux grands laboratoires capables de générer des quantités d'images suffisantes pour entraîner l'algorithme de façon supervisée.

Le secteur médical, en particulier l'anapath, a besoin d'un moyen pour généraliser et exploiter les connaissances apprises sur différents domaines, car la génération de grandes bases de données est très coûteuse et peu efficace. Le principe du Transfer Learning est de réutiliser ou « transférer » les connaissances apprises sur une tâche pour résoudre un problème différent (problème-cible). Il y a plusieurs exemples d'application du transfer learning, hors du domaine médical, qui sont basés sur l'utilisation de réseaux pré-entraînés (comme par exemple les réseaux VGG16 ou Xception qui sont entraînés à classer IMAGENET) [2], [3].

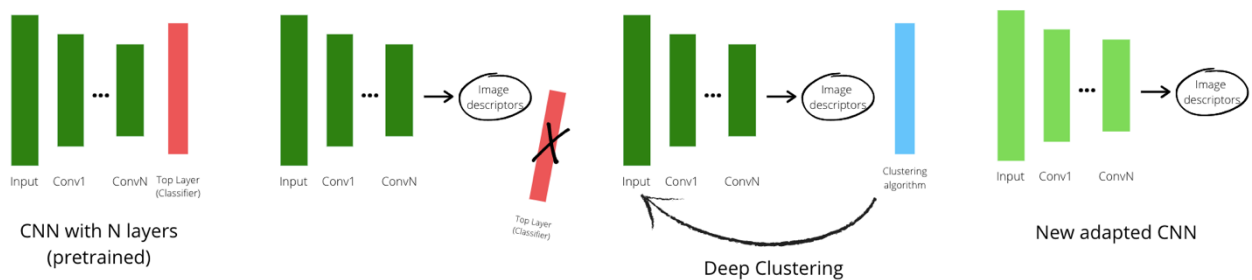
Aussi, l'état de l'art actuel propose notamment de ré-entraîner des réseaux de neurones sans supervision : cette méthode porte le nom de Deep Clustering [4], [5], et est basée sur l'alternance entre des phases d'apprentissage du réseau et des phases de clustering dans l'espace des features des images. L'objectif de ce projet est de proposer une méthode d'adaptation de domaine de réseaux pré-entraînés aux images de la pathologie numérique. Cet algorithme deviendra par la suite un point clef dans le

développement de modèles prédictifs plus généralisables que les réseaux neuronaux classiques entraînés de zéro.



★ How do we obtain the best image descriptor?

Deep Clustering could be used to adapt a pretrained CNN to a particular domain in a non-supervised way



Notre proposition est de faire une adaptation de domaine non supervisée basée sur le principe du Deep Clustering pour obtenir des descripteurs d'images [FIG 2]. Ces descripteurs sont ensuite utilisés pour grouper les images similaires en catégories qui, idéalement, reflètent bien les classes manipulées par le médecin lorsqu'il construit son diagnostic.

Ce projet doit répondre à plusieurs problématiques :

1. Est-il possible de transférer les connaissances générales obtenues pour l'analyse d'un domaine non-médical vers le domaine de l'histopathologie ?
2. Si oui, quelle est la meilleure stratégie pour le faire ?
3. Comment mesurer les performances des différentes méthodes d'adaptation proposées et les comparer aux solutions existantes dans la littérature ?

Au cours du projet, notre équipe pourra proposer des bases de données existantes et en constituer de nouvelles pour tester et évaluer la pertinence des solutions proposées. On donnera les étiquettes des images pour évaluer les différentes solutions obtenues et déterminer la meilleure méthode d'extraction des caractéristiques. La première phase sera bibliographique et reposera sur les articles cités en références ci-dessous. Pour la deuxième phase, on démontrera l'implémentation et la comparaison des architectures issues de la partie bibliographique.

Bibliographie :

- [1] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, "Artificial

- intelligence in digital pathology — new tools for diagnosis and precision oncology,” *Nat. Rev. Clin. Oncol.*, vol. 16, no. 11, pp. 703–715, Nov. 2019, doi: 10.1038/s41571-019-0252-y.
- [2] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” 2009, doi: 10.1109/TKDE.2009.191.
 - [3] S. Ruder, “Transfer Learning - Machine Learning’s Next Frontier,” 2017. [Online]. Available: <https://ruder.io/transfer-learning/>. [Accessed: 12-Oct-2020].
 - [4] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, “Deep clustering for unsupervised learning of visual features,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11218 LNCS, pp. 139–156, 2018, doi: 10.1007/978-3-030-01264-9_9.
 - [5] H. Fan, Y. I. Yang, L. Zheng, C. Yan, and Y. Yang, “Unsupervised Person Re-identification: Clustering and Fine-tuning,” *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 14, no. 4, 2018, doi: 10.1145/3243316.