
Segmentation non-supervisée d'images histopathologiques à l'aide d'auto-encodeurs et d'ensemble clustering

I. Rmouque*, M. Devanne**, J. Weber**, G. Forestier**, C. Wemmert*

*ICube, Université de Strasbourg, France
wemmert@unistra.fr,

**IRIMAS, Université Haute Alsace, Mulhouse, France
prenom.nom@uha.fr

Résumé. L'apprentissage profond non-supervisé à partir d'auto-encodeurs permet d'obtenir d'excellents résultats en matière d'analyse d'images et de vision par ordinateur. Cependant, seules quelques études ont été présentées dans le domaine de la pathologie numérique, où l'étiquetage précis des objets d'intérêt est une tâche particulièrement coûteuse et difficile. Ainsi, le fait de disposer d'une première segmentation entièrement non supervisée pourrait grandement faciliter le processus d'analyse de ces images.

Dans cet article, de nombreuses architectures d'auto-encodeurs convolutifs ont été comparées afin d'étudier l'influence de trois principaux paramètres : 1) le nombre de couches convolutives, 2) le nombre de convolutions dans chacune de ces couches et 3) la taille de l'espace latent. Différents algorithmes de clustering sont également comparés et une nouvelle approche permettant d'obtenir des résultats plus précis en appliquant des techniques de clustering par ensemble est présentée.

1 Introduction

La pathologie est essentielle pour l'évaluation diagnostique et la compréhension de nombreux mécanismes biologiques et physiologiques sous-jacents. Il s'agit généralement d'une évaluation visuelle, par des pathologistes, d'un échantillon de tissu à l'aide d'un microscope pour en identifier les propriétés structurelles. Actuellement, l'évaluation visuelle des échantillons microscopiques est en grande partie un processus non assisté, et la précision du pathologiste est due à une formation approfondie, une analyse comparative, un contrôle de qualité par les pairs et l'expérience personnelle. Cependant, ce domaine a connu plusieurs révolutions technologiques ces dernières années avec l'avènement de la microscopie virtuelle (conversion de lames microscopiques en images à haute résolution), souvent appelée "pathologie numérique". Ainsi, des efforts importants ont été faits pour concevoir des outils d'analyse d'images, par exemple pour identifier les structures biologiques de base (stroma, cellules immunitaires, tumeur, etc.), afin de faciliter l'interprétation (semi-)automatisée des lames par les médecins.

Entre-temps, les algorithmes d'analyse automatique d'images ont à nouveau fait des progrès extraordinaires et ce notamment, grâce aux méthodes d'apprentissage profond introduites

par LeCun et al. (2015). En effet, les performances de ces méthodes ont explosé ces dernières années permettant la détection, la classification et la segmentation d'objets d'intérêt dans les images avec une très grande précision. Mais la plupart de ces approches fonctionnent en mode supervisé, c'est-à-dire qu'elles nécessitent de nombreux exemples pour fournir un modèle efficace. En même temps, les approches non supervisées ont montré leur intérêt pour de nombreuses applications de l'analyse d'images, comme la télédétection (Liang et al., 2018; Mei et al., 2019). Récemment, elles ont également été appliquées à l'analyse d'images histopathologiques sur lames entières scannées (*Whole-Slide Images* - WSI). En particulier, dans Yamamoto et al. (2019), les auteurs décrivent une méthode d'extraction non supervisée d'informations intéressantes à partir d'images complètes de diapositives, qui se révèle plus précise que l'analyse humaine pour le pronostic de la récurrence du cancer de la prostate.

Nous nous intéressons ici à la segmentation automatique afin d'extraire rapidement les régions d'intérêt (les tumeurs par exemple) et d'effectuer une analyse plus précise pour ces seules régions (les WSI sont des images de très grande taille et l'analyse ne pourrait être effectuée que sur certaines régions d'intérêt).

Cependant, seuls quelques travaux sur la segmentation entièrement non supervisée des WSI ont été présentés. La première tentative de segmentation des régions d'intérêt des WSI sans aucune information ou exemple préalable a été réalisée dans Khan et al. (2013). Les auteurs ont étudié la morphologie des tissus dans les images histologiques du cancer du sein en calculant un ensemble de filtres de Gabor pour distinguer différentes régions. Dans Fouad et al. (2017), les auteurs utilisent la morphologie mathématique pour extraire des "cellules virtuelles" (par exemple des superpixels), pour lesquelles des caractéristiques morphologiques et de couleur sont calculées pour ensuite appliquer un algorithme d'ensemble clustering pour identifier les différents types de tissu dans l'image. Plus récemment, une approche similaire a été présentée par Landini et al. (2019), ajoutant aux techniques précédentes un classifieur auto-apprenant semi-supervisé qui améliore les résultats au prix de l'abandon de la non-supervision.

Toutes ces approches proposent d'effectuer un clustering sur les image en fonction de caractéristiques prédéfinies. Cependant, des approches d'apprentissage profond, notamment via des architectures d'auto-encodage, permettent d'éviter de définir manuellement les caractéristiques en calculant la représentation condensée de l'image dans un espace latent par l'application de filtres convolutifs. Malheureusement, comme indiqué dans Raza et Singh (2018), la plupart des applications des auto-encodeurs en pathologie numérique ont été développées pour effectuer la segmentation cellulaire ou la détection de noyaux (Xu et al., 2015; Hou et al., 2019), ou la normalisation des taches (Janowczyk et al., 2017).

Nous proposons donc ici d'étudier le potentiel de ces approches pour la segmentation des WSI. Dans cet article, nous présentons une étude sur les performances des auto-encodeurs convolutifs sur la segmentation des WSI en comparant différentes approches. Dans un premier temps, différentes architectures d'auto-encodeurs sont comparées afin de quantifier l'importance des paramètres d'intérêt (nombre de couches, taille de l'espace latent, nombre de convolutions par couche), puis, constatant que les méthodes de clustering conventionnelles donnent un résultat partiel, nous proposons une méthode multi-résolution utilisant des approches d'ensemble clustering.

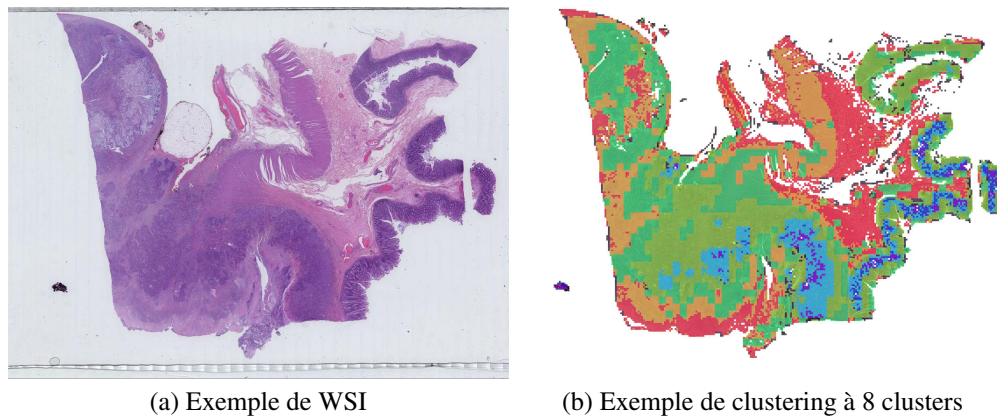


FIG. 1 – Un exemple de WSI et d'un clustering possible

2 Jeu de données

Notre étude a été réalisée sur 8 WSI de tissus colorés à l'hématoxyline et à l'éosine (H&E) extraits d'une cohorte de patients construite dans le cadre du projet AiCOLO (INSERM/Plan Cancer) étudiant le cancer du côlon. Les images ont été fournies par le Centre Georges François Leclerc de Dijon (France). La coloration H&E permet de distinguer les noyaux des cellules en violet, de la matrice extracellulaire et du cytoplasme en rose.

Toutes les images ont été acquises à un grossissement de $40\times$ ce qui correspond environ à $0.22\mu\text{m}/\text{pixel}$ mais elles sont stockées à plusieurs résolutions dans un format pyramidal. Les différentes résolutions sont désignées comme niveau de détail (LOD - *Level Of Details*), LOD0 correspondant à un grossissement de $40\times$, LOD1 à $20\times$, etc. La taille de chaque image est d'environ 90000×50000 pixels.

Pour entraîner l'auto-encodeur, 10000 zones de 128×128 pixels ont été extraites aléatoirement au LOD1 de toutes les images, car cela semble être la quantité minimale d'informations requise par un expert humain pour classer le tissu. Entre-temps, des pathologistes ont créé sur les images des annotations manuelles éparées pour les cinq classes de tissus, tumeur, stroma, muqueuse de la couche externe (cryptes de Lieberkuhn et tissu conjonctif), cellules immunitaires et nécrose, et deux classes pour le fond et les artefacts (marques de feutre, etc.), afin de pouvoir évaluer la pertinence du regroupement.

3 Auto-encodeurs convolutifs pour le clustering de WSI

Dans cette section, nous étudions le potentiel de l'utilisation d'auto-encodeurs convolutifs pour le clustering des images histopathologiques. Comme le montre la Figure 2, un *auto-encodeur convolutif* (AEC) est un réseau de neurones convolutif profond composé de deux parties : un encodeur et un décodeur. L'objectif principal de l'AEC est de minimiser une fonction de perte, c'est-à-dire une fonction évaluant la différence entre l'entrée et la sortie de l'AEC (généralement l'erreur quadratique moyenne). Une fois cette fonction minimisée, on peut sup-

poser que la partie encodeur établit un résumé approprié des données d'entrée dans l'espace latent, car la partie décodeur est capable d'en reconstruire une copie fortement similaire à partir de cette représentation encodée. L'encodeur est constitué d'une couche d'entrée (ayant la taille de l'image en entrée) qui est connectée à N couches de convolution de taille décroissante, jusqu'à un goulot d'information de taille Z , appelé espace latent. L'espace latent est connecté à une série de couches de N convolutions de taille croissante, jusqu'à atteindre la taille de l'image d'entrée. Cette seconde partie est appelée le décodeur. Chaque couche de convolution est composée de C convolutions et est suivie de trois autres couches : une normalisation, une fonction d'activation (ReLU) et un max pooling de taille (2,2).

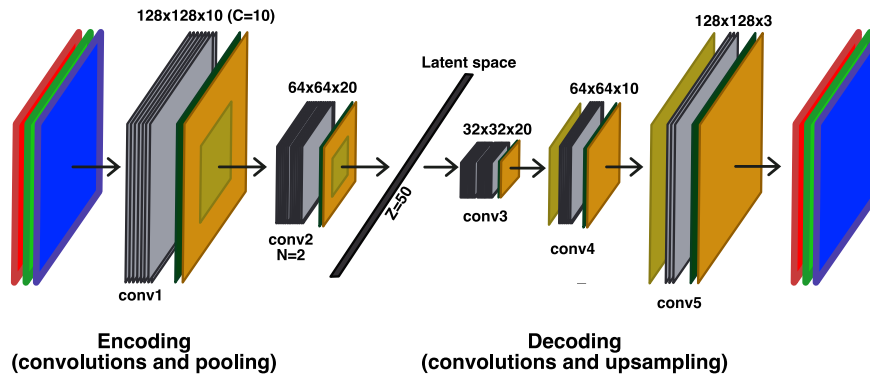


FIG. 2 – Architecture de l'AEC pour $N = 2$, $C = 10$ et $Z = 50$

Pour effectuer un clustering, un AEC entraîné est utilisé pour encoder chaque partie de l'image entière. Ensuite, la représentation encodée dans l'espace latent est donnée en entrée d'un algorithme de clustering qui lui attribue un cluster. Un exemple de clustering d'une WSI est montré sur la Figure 1(b). Nous avons décidé d'évaluer l'influence des trois principaux paramètres N , Z et C . Pour chacun d'eux, des valeurs différentes ont été testées tout en gelant les deux autres. Afin d'évaluer la qualité des résultats, l'indice de Rand ajusté (ARI) est calculé comme la normalisation de l'indice de Rand (RI) défini dans Eq.1.

$$RI = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

où TP , TN , FP et FN représentent respectivement les pixels vrais positifs, vrais négatifs, faux positifs et faux négatifs.

L'un des principaux défis du clustering non supervisé est de trouver la correspondance entre un cluster et une classe identifiée par l'expert. Afin de simplifier le problème, nous avons décidé d'évaluer la capacité du clustering à détecter la classe tumeur uniquement (classe d'intérêt dans notre projet). Ainsi, le problème peut être exprimé comme un problème à deux classes : tumeur vs. non tumeur. Pour associer la classe de tumeur à un cluster, nous avons calculé sa densité (nombre de pixels de la tumeur / nombre total de pixels du cluster). Tous les clusters ayant une densité supérieure à 50% sont considérés comme "tumeur", les autres sont étiquetés comme "non tumeur". Chaque AEC a été entraîné sur un ensemble de 10000 extraits de WSI

sélectionnées aléatoirement. Comme les résultats du clustering et de l'entraînement des AEC sont non déterministes, en raison d'une sensibilité élevée aux conditions initiales, 10 auto-encodeurs ont été entraînés, et ont été conservés les résultats moyens pour chaque valeur de chaque paramètre. Nous avons également étudié les performances de plusieurs algorithmes de clustering, Kmeans, Agglomerative clustering (AggCl), Gaussian mixture (GM) et ainsi que la méthode "not too deep clustering" (N2D) de McConville et al. (2020).

Résultats

Comme le montre la Figure 3, il semble que le nombre de convolutions dans chaque couche de convolutions (paramètre C) n'affecte pas beaucoup la qualité de l'auto-encodeur, à part une légère baisse de la variabilité des résultats. Ceci s'explique par le fait que passé un certain seuil, un nombre supplémentaire de convolutions apporte peu d'informations.

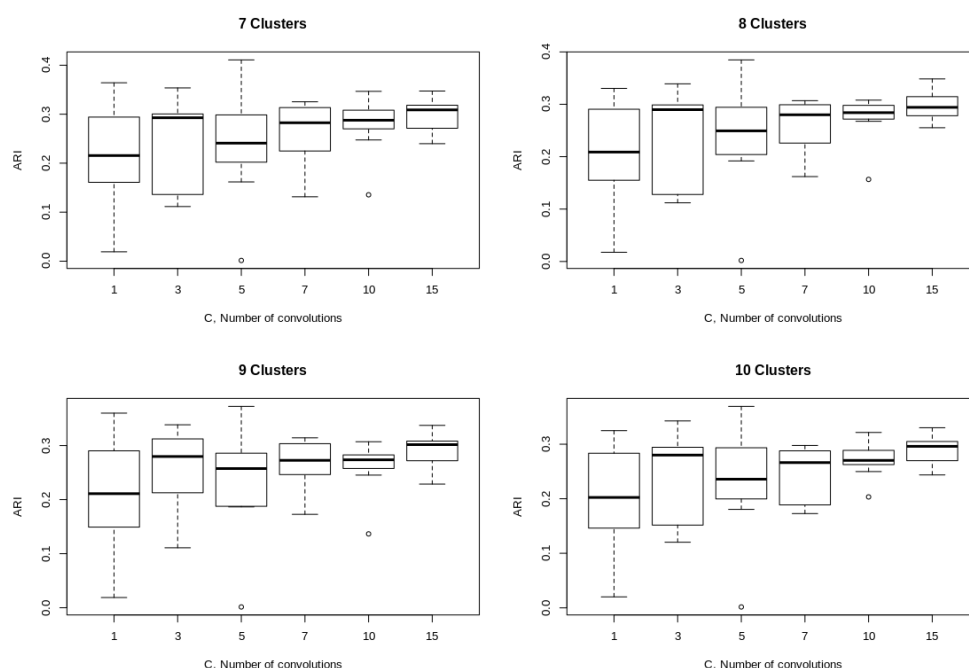


FIG. 3 – Évaluation de l'ARI en fonction du nombre de convolutions C dans chaque couche de convolutions avec l'algorithme de clustering Kmeans pour 7, 8, 9 et 10 clusters.

La Figure 4 montre l'évaluation de l'ARI avec un nombre différent de couches de convolutions dans l'architecture. On remarque une augmentation de l'indice de qualité jusqu'à 4 couches de convolution, puis une chute brutale pour 5, ce qui indique clairement qu'un trop grand nombre de convolutions (et les poolings qui réduisent la taille de l'information) affectent la qualité de l'information qui ne peut plus être traitée correctement.

Néanmoins, comme le montre la Figure 5, la taille de l'espace latent Z , semble avoir une grande influence sur la pertinence du AEC. En effet, l'ARI augmente nettement au fur et

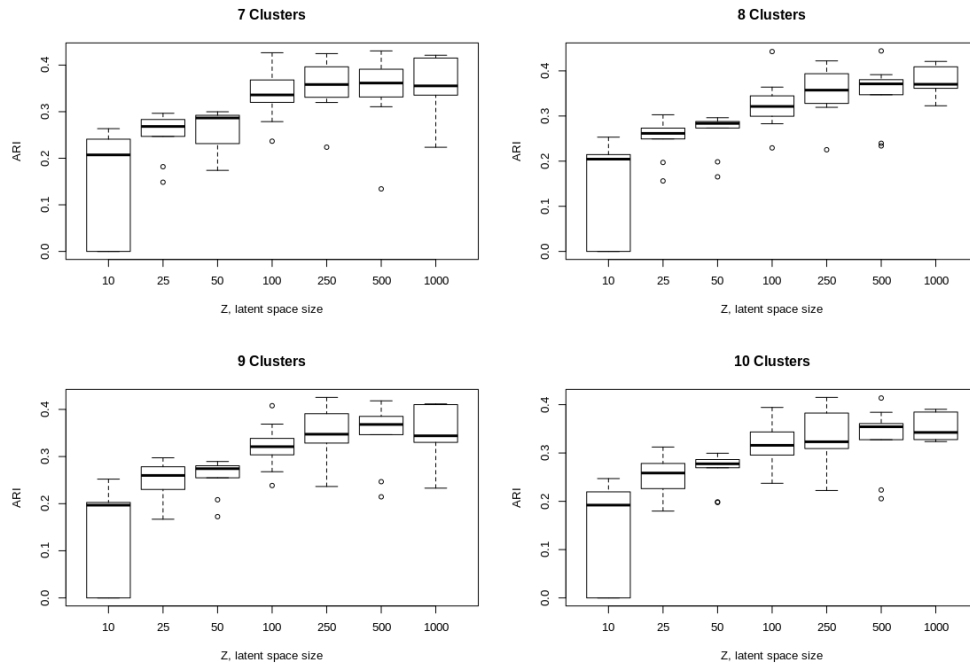


FIG. 4 – Évaluation de l'ARI en fonction de la taille de l'espace latent avec l'algorithme de clustering Kmeans pour 7, 8, 9 et 10 clusters.

à mesure que la taille de la représentation latente augmente. En effet, plus la représentation latente contient d'information, plus il est facile de différencier les classes, mais il est également évident qu'un espace latent trop grand ne pourra pas résumer efficacement les informations et n'aidera donc pas l'algorithme de clustering à distinguer les différents tissus. En outre, plus l'espace latent est grand, plus il faut de mémoire et de temps pour entraîner le réseau.

4 Ensemble clustering

Comme indiqué dans Yamamoto et al. (2019), les micro-structures et les macro-structures des objets anatomiques présents dans les WSI fournissent des informations différentes et complémentaires. Les pathologistes s'accordent également sur le fait qu'il est beaucoup plus difficile d'identifier une cellule sans son contexte environnant et ils regardent toujours la WSI à un grossissement plus faible (pour mieux saisir le contexte) avant de zoomer à un grossissement plus élevé.

Nous avons donc essayé d'améliorer nos résultats en utilisant un ensemble de méthodes de clustering, chacune se concentrant sur un niveau de détail différent. L'objectif est de fusionner les informations de bas niveau (contexte) avec les informations de haut niveau (forme des cellules, etc.). Pour ce faire, nous utilisons la méthode proposée dans Wemmert et Gançarski

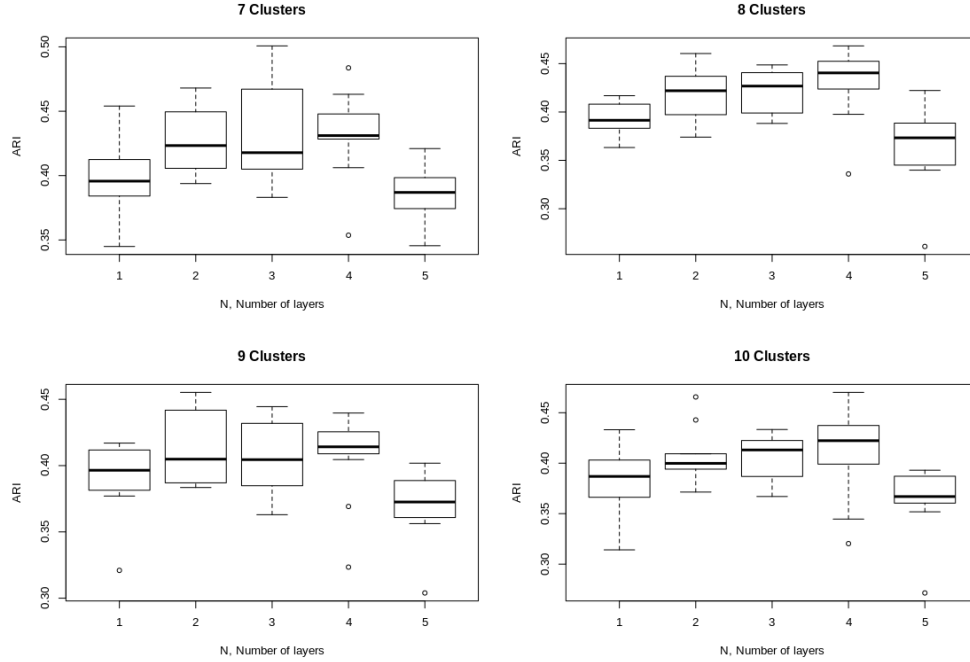


FIG. 5 – Évaluation de l'ARI en fonction du nombre de couches de convolutions N avec l'algorithme Kmeans pour 7, 8, 9 et 10 clusters.

(2002). De nombreuses configurations différentes ont été testées pour cette expérience, mais nous n'en présentons ici que deux, représentatives de l'amélioration de la qualité du résultat qui peut être obtenue en utilisant l'ensemble clustering.

La première, $E_{multires}$, est composée de trois algorithmes de clustering (Kmeans), utilisant la représentation dans l'espace latent construit par différents AEC entraînés à différentes résolutions : LOD2 avec 8 clusters, LOD3 avec 6 clusters et LOD3 avec 8 clusters.

Comme l'image reconstruite à partir de l'auto-encodeur semble se concentrer davantage sur l'intensité des couleurs que réellement sur les structures, une seconde configuration a été testée. Afin d'ajouter de la diversité et de forcer le résultat final à se concentrer davantage sur la structure des objets, un clustering travaillant sur une image binaire a été ajouté à l'ensemble. Cette image est construite en seuillant l'intensité de l'image initiale. La seconde configuration, E_{struct} , est donc composée de trois algorithmes de clustering (Kmeans) avec les paramètres suivants : LOD3 sur l'image binaire avec 6 clusters, LOD3 sur l'image binaire avec 8 clusters et LOD2 sur l'image RGB initiale avec 6 clusters.

Deux critères d'évaluation ont été calculés à partir des résultats et sont présentés dans le Tableau 1. L'ARI (voir Eq. 1) est complété par le $Fscore$ défini à partir de la précision et du rappel comme :

$$Fscore = \frac{TP}{TP + 1/2(FP + FN)} \quad (2)$$

	Données brutes		Données encodées											
	Kmeans		Kmeans		AggCl		GM		N2D		$E_{multires}$		E_{struct}	
	ARI	FScore	ARI	FScore	ARI	FScore	ARI	FScore	ARI	FScore	ARI	FScore	ARI	FScore
Image 1	0.39	0.89	0.48	0.96	0.38	0.94	0.27	0.73	0.43	0.97	0.47	0.96	0.42	0.88
Image 2	0.27	0.68	0.33	0.63	0.29	0.62	0.19	0.43	0.29	0.68	0.31	0.66	0.46	0.62
Image 3	0.25	0.76	0.39	0.87	0.35	0.85	0.22	0.76	0.31	0.88	0.37	0.87	0.45	0.91
Image 4	0.08	0.48	0.08	0.50	0.13	0.61	0.05	0.51	0.12	0.55	0.08	0.54	0.0	0.48
Image 5	0.11	0.65	0.11	0.64	0.10	0.60	0.10	0.62	0.11	0.65	0.12	0.65	0.17	0.72
Image 6	0.37	0.68	0.52	0.75	0.51	0.72	0.49	0.67	0.43	0.76	0.51	0.77	0.57	0.75
Image 7	0.28	0.68	0.35	0.73	0.33	0.75	0.14	0.61	0.37	0.74	0.41	0.76	0.36	0.84
Image 8	0.33	0.63	0.44	0.71	0.42	0.70	0.07	0.44	0.37	0.69	0.44	0.75	0.45	0.75

TAB. 1 – Évaluations des résultats de clustering obtenus à l'aide des différentes méthodes

Résultats

L'ensemble des résultats pour chacune des configurations est donné dans le Tableau 1. On y retrouve le résultat pour chacune des méthodes de clustering testées (simples et ensembles), ainsi que le résultat pour un clustering effectué avec l'algorithme Kmeans sur les données brutes (sans aucun traitement de l'AEC), qui a été calculé comme base de référence afin d'évaluer la pertinence de l'encodage des données avec un AEC.

Tout d'abord, on constate que le clustering appliqué directement sur les données brutes a obtenu des résultats de moins bonne qualité que toutes les autres configurations, ce qui confirme que l'encodage des données a un effet positif sur le clustering. Ensuite, les méthodes classiques appliquées sur la représentation de l'espace latent de l'AEC tendent à montrer des résultats acceptables.

Cependant, les deux configurations d'ensemble clustering montrent une efficacité plus grande pour trouver des clusters cohérents correspondant aux classes d'intérêt définies par les pathologistes. En effet, parmi toutes les méthodes proposées, E_{struct} semble donner les meilleurs résultats. Elle tend à confirmer l'importance, à la fois, d'introduire plus de variabilité dans l'ensemble pour améliorer les résultats, mais aussi de l'utilisation d'informations sur la structure des objets. En outre, cela montre que même si les auto-codeurs convolutifs visent à trouver automatiquement les meilleures caractéristiques pour coder l'image, ils peuvent également tirer parti de caractéristiques introduites manuellement par l'expert pour certaines tâches spécifiques.

5 Conclusion

Dans cet article, nous avons comparé différentes configurations d'auto-encodeurs convolutifs dans le domaine de l'apprentissage non supervisé pour la segmentation des images histopathologiques WSI. Pour cela, différentes architectures d'AEC ont été comparées pour essayer de trouver la meilleure configuration et d'étudier l'influence de chaque paramètre. Ensuite, nous avons proposé une nouvelle approche qui utilise la technique d'ensemble clustering pour tirer profit des informations multirésolution et des caractéristiques structurelles de l'image. Cela confirme l'importance de la diversité dans un cadre d'apprentissage profond et que le fait

de travailler à différentes résolutions en même temps peut réellement améliorer la qualité des résultats.

Remerciements

Ce travail est issu du projet AiCOLO financé par l'appel *Approches interdisciplinaires des processus oncogéniques et perspectives thérapeutiques : apports à l'oncologie des mathématiques et de l'informatique* Inserm / Plan Cancer (2014-2019).

Références

- Fouad, S., D. Randell, A. Galton, H. Mehanna, et G. Landini (2017). Unsupervised morphological segmentation of tissue compartments in histopathological images. *PloS one* 12(11), e0188717.
- Hou, L., V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, et al. (2019). Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern recognition* 86, 188–200.
- Janowczyk, A., A. Basavanthally, et A. Madabhushi (2017). Stain normalization using sparse autoencoders (stanosa) : application to digital pathology. *Computerized Medical Imaging and Graphics* 57, 50–61.
- Khan, A. M., H. El-Daly, E. Simmons, et N. M. Rajpoot (2013). Hymap : A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images. *Journal of Pathology Informatics* 4.
- Landini, G., S. Fouad, D. Randell, et H. Mehanna (2019). Epithelium and stroma segmentation using multiscale superpixel clustering. *Journal of Pathology Informatics* 10, S45–S48.
- LeCun, Y., Y. Bengio, et G. E. Hinton (2015). Deep learning. *Nature* 521(7553), 436–444.
- Liang, P., W. Shi, et X. Zhang (2018). Remote sensing image classification based on stacked denoising autoencoder. *Remote Sensing* 10(1), 16.
- McConville, R., R. Santos-Rodriguez, R. J. Piechocki, et I. Craddock (2020). N2d : (not too) deep clustering via clustering the local manifold of an autoencoded embedding.
- Mei, S., J. Ji, Y. Geng, Z. Zhang, X. Li, et Q. Du (2019). Unsupervised spatial-spectral feature learning by 3d convolutional autoencoder for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing* 57(9), 6808–6820.
- Raza, K. et N. K. Singh (2018). A tour of unsupervised deep learning for medical image analysis. *arXiv preprint arXiv :1812.07715*.
- Wemmert, C. et P. Gançarski (2002). A multi-view voting method to combine unsupervised classifications. In *Artificial Intelligence and Applications*, pp. 447–452.
- Xu, J., L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, et A. Madabhushi (2015). Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. *IEEE transactions on medical imaging* 35(1), 119–130.

Yamamoto, Y., T. Tsuzuki, J. Akatsuka, M. Ueki, H. Morikawa, Y. Numata, T. Takahara, T. Tsuyuki, K. Tsutsumi, R. Nakazawa, et al. (2019). Automated acquisition of explainable knowledge from unannotated histopathology images. *Nature communications* 10(1), 1–9.

Summary

Unsupervised deep learning with autoencoders have shown excellent results in image analysis and computer vision. However, only few studies have been presented in the field of digital pathology, where proper labelling of the objects of interest is a particularly costly and difficult task. Thus, having a first fully unsupervised segmentation could greatly help in the analysis process of such images. In this paper, many architectures of convolutional autoencoders have been compared to study the influence of three main hyperparameters: number of convolutional layers, number of convolutions in each layer and size of the latent space. Different clustering algorithms are also compared and we propose a new way to obtain more precise results by applying ensemble clustering techniques.