

École nationale supérieure d'électrotechnique, d'électronique, d'informatique,  
d'hydraulique et des télécommunications



## Projet long

### Recherche Bibliographique

2020-2021

---

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

---



**INSTITUT UNIVERSITAIRE  
DU CANCER DE TOULOUSE**  
**Oncopole**

## Sommaire

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Défis dans le domaine de pathologie digitale</b>	<b>2</b>
<b>3</b>	<b>Transfer learning et Apprentissage Non-Supervisé</b>	<b>5</b>
3.1	Transfer Learning . . . . .	5
3.1.1	Introduction . . . . .	5
3.1.2	Pourquoi utilise-t-on le Transfer Learning? . . . . .	5
3.1.3	Application du Transfer Learning . . . . .	6
3.1.4	Méthodes utilisées en Transfer Learning . . . . .	6
3.2	Unsupervised learning et pathologie digitale . . . . .	7
3.2.1	Approches non-supervisées dans la pathologie digitale . . . . .	7
3.2.2	Deep Cluster . . . . .	7
3.2.3	Combinaison de K-means et apprentissage profond . . . . .	10
3.2.3.a	Stabilité . . . . .	10
3.2.3.b	Solutions . . . . .	11
3.2.4	Tâche de classification . . . . .	11
3.2.5	Utilisation des connaissances de domaines . . . . .	12
3.2.6	Data augmentation . . . . .	14
<b>4</b>	<b>Choix d'implémentations</b>	<b>15</b>
<b>5</b>	<b>Conclusion</b>	<b>15</b>
	<b>References</b>	<b>17</b>

# 1 Introduction

L'anatomopathologie (anapath) est la discipline médicale qui établit les diagnostics à partir de caractéristiques macroscopiques et microscopiques des tissus.

Le besoin de précision dans le diagnostic histopathologique du cancer augmente car le traitement personnalisé du cancer nécessite une évaluation précise des biomarqueurs. Récemment, l'apprentissage automatique, et en particulier l'apprentissage profond, ont permis des progrès rapides en pathologie computationnelle. Ainsi l'intégration de l'apprentissage automatique dans les soins de routine sera une étape importante pour le secteur de la santé au cours de la prochaine décennie, et l'histopathologie est au centre de cette révolution [16].

Plus précisément, dans la pathologie digitale, les approches l'apprentissage automatique -IA- sont conçues pour extraire initialement des représentations d'images appropriées, qui peuvent ensuite être utilisées pour former un classificateur de machine à une tâche de segmentation, de diagnostic ou de pronostic particulière. Plusieurs applications de l'IA en pathologie numérique se sont concentrées sur la nécessité d'automatiser les tâches qui prennent du temps pour le pathologiste, ce qui lui permet ensuite de consacrer plus de temps à des tâches décisionnelles de haut niveau [9].

Dans ce rapport, on va discuter des modèles de pathologie digitale utilisant l'apprentissage profond et des algorithmes de classification. Pour ce faire, on va commencer par présenter les défis dans le domaine de la pathologie digitale. Ensuite, on se plongera dans le monde du Transfer learning et de l'apprentissage non-supervisé afin d'explorer les possibilités de les intégrer dans notre travail. Finalement, on discutera des choix de notre implémentation.

# 2 Défis dans le domaine de pathologie digitale

De nos jours, l'intelligence artificielle a connu un épanouissement dans le domaine de la pathologie. L'usage des outils d'apprentissage automatique et profond peuvent donner de meilleures précisions que les autres moyens classiques. L'introduction des outils d'intelligence artificielle ont permis de simplifier les procédures et les étapes des diagnostics médicaux, comme le montre le schéma au-dessous, et même ils ont permis de traverser certaines limitations que posent les outils de la pathologie traditionnelle [15].

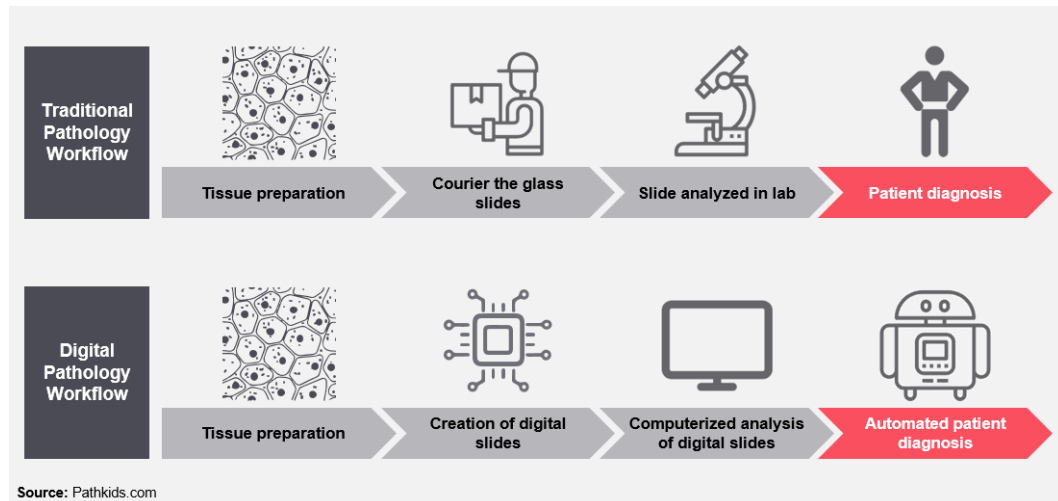


Figure 1 – Comparaison entre les approches manuelles et la digitalisation suivies dans la pathologie [15].

Cependant, la pathologie digitale aura à faire face à plusieurs défis qui menacent son déploiement et son optimisation.

Au début, les données doivent être annotées et libellées durant la phase de collecte et la préparation, ce qui peut être onéreux, en termes de temps et de coût, au vu que l’annotation de ces données massives nécessite une retouche humaine par des spécialistes. Ce processus peut présenter des obstacles non seulement pour le temps, la quantité et le coût; mais aussi lors de l’usage et le travail avec des images floues ou à faible résolution, des réseaux lents et des images de microscopie où certaines caractéristiques sont ambiguës [16].

De surcroît, les images de microscopie utilisées pour les diagnostics sont considérées avoir de très grandes dimensions, de l’ordre de Gigapixels, pourtant les réseaux de neurones ne peuvent contenir que des données moyennes, de l’ordre 350 x 350 pixels en général. Pour résoudre ce problème, on recourt à la technique du patching [13], qui permet de diviser une image de très grande taille en petits fragments et carreaux. Il est primordial de sous-échantillonner ces nouvelles données pour pouvoir les alimenter dans un réseau profond. Cependant, ces nouvelles régions peuvent ne pas convenir à de nombreux diagnostics, ainsi que, le sous-échantillonnage de ces correctifs peut entraîner la perte d’informations cruciales. D’autre part, les réseaux profonds avec des tailles d’entrée plus grandes auraient besoin d’une topologie beaucoup plus profonde et d’un nombre beaucoup plus grand de neurones, ce qui les rend encore plus difficiles et peut-être impossibles à entraîner.

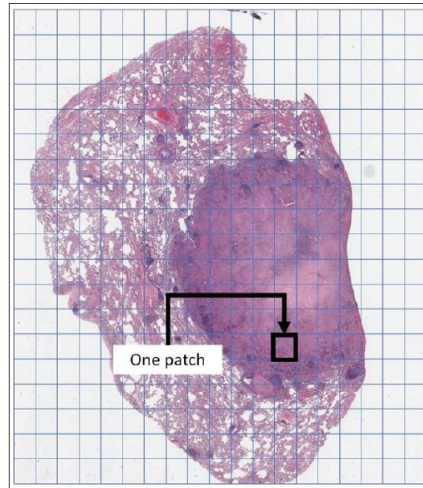


Figure 2 – Exemple d’un patch pris d’une image dédiée pour les diagnostics [16].

Ainsi, La pathologie digitale rencontre un problème qui concerne la variabilité continue et omniprésente des données. En effet, il existe plusieurs types de tissus de base. Cependant, le nombre réel de modèles dérivés de ces tissus d’un point de vue informatique est presque infini si les images histopathologiques doivent être comprises par des algorithmes informatiques. Plusieurs types de tissus construisent un organe qui se reflète également dans une nouvelle variation de texture des types de tissus de base. Ce polymorphisme extrême rend la reconnaissance des tissus par des algorithmes d’image exceptionnellement difficiles. Ainsi, l’architecture inhérente à l’IA profonde nécessite de nombreux cas de formation pour chaque variation. Cependant, cela peut ne pas être facilement disponible, en particulier sous forme de données annotées.

Ensuite, la pathologie digitale doit faire face à plusieurs défis, surtout, ceux qui sont associés à l’interprétation et à la compréhension de la manière dont les technologies et modèles d’intelligence artificielle complexes parviennent à ses décisions, cette contrainte se nomme parfois par le problème de la « boîte noire ». Ce qui est considéré parmi les principales préoccupations qui font obstacle à l’adoption clinique des algorithmes d’apprentissage en profondeur. L’interprétation du modèle pourrait également révéler de nouvelles caractéristiques d’une maladie, comme la présence histologique d’un gliome d’œdème qui n’a pas été précédemment reconnu par les outils classiques [1].

Enfin, la communauté met toujours en question l’efficacité et la performance des outils d’intelligence artificielle. En fait, les rapports de défaillances d’IA dans les soins de santé ne sont pas nécessairement liés à une technologie défaillante, mais plutôt à des difficultés à confier et déployer de tels outils dans la pratique. De plus, la performance instable peut être un inconvénient fatal pour ces technologies. En effet, des résultats récents ont démontré que les modèles d’intelligence artificielle actuels, lorsqu’ils sont entraînés sur des ensembles de données trop petits, même en utilisant des étiquettes méticuleuses, peuvent présenter une baisse de 20% des performances

lorsqu'ils sont testés sur des ensembles de données indépendants, d'où l'importance du travail de stabilité, du rendement et performance de ces technologies.

## 3 Transfer learning et Apprentissage Non-Supervisé

### 3.1 Transfer Learning

#### 3.1.1 Introduction

Les modèles basés sur le Deep learning sont généralement très performants lorsqu'on leur donne un nombre important d'exemples étiquetés. Cependant, quand ils sont confrontés à des conditions pour lesquelles ils n'ont pas été entraînés, les résultats sont moins performants. Ainsi, si on veut appliquer nos modèles dans le monde réel où les possibilités sont infinies et non pas sur une base de données soigneusement construite pour des entraînements, le Transfert learning est l'une des solutions pour transférer les connaissances acquises dans certaines conditions à un état similaire.

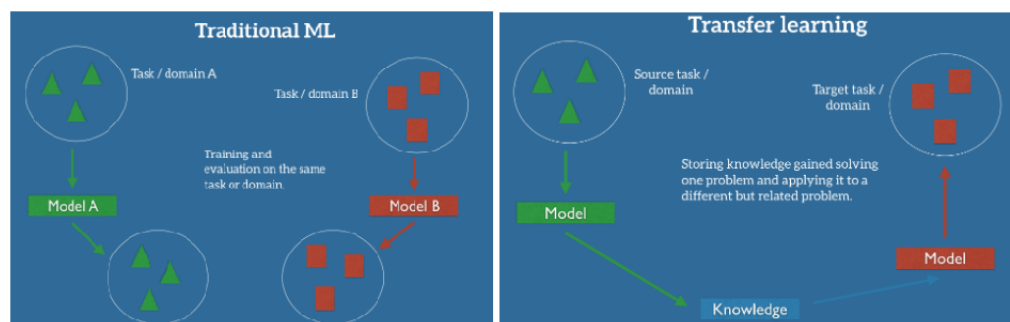


Figure 3 – La configuration traditionnelle d'apprentissage supervisé en ML vs la configuraion du Transfer learning [13]

#### 3.1.2 Pourquoi utilise-t-on le Transfer Learning?

Le succès du Deep learning dans le monde industriel repose principalement sur des modèles très exigeants en quantité de données étiquetées. Ces besoins en matière de données sont atteints dans certains cas où elles ont été accumulées au fil des années [13]. Ce n'est pas toujours le cas, en général elles sont privées, coûteuses ou bien insuffisantes.

De plus, chaque client utilise le modèle avec des données différentes de celles avec lesquelles il a été entraîné et lui demande de réaliser des tâches qui varient un peu de celles pour lesquelles il a été entraîné.

On a donc besoin de se pencher sur le transfert de connaissances acquises précédemment pour résoudre de nouveaux problèmes plus rapidement [12] ainsi on pourra faire évoluer le Machine learning dans le monde industriel surtout dans les domaines où les données se font plus rares.

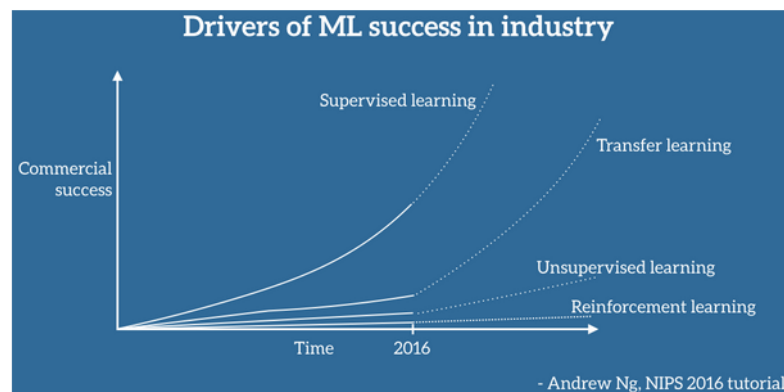


Figure 4 – Les moteurs du succès industriel du ML selon Andrew Ng. [13]

Le Transfer learning est un domaine de l'intelligence artificielle qui s'intéresse à la capacité d'un algorithme de machine learning à améliorer ses capacités d'apprentissage sur une base de données à l'aide d'une exposition préalable à une autre base de données [13]. En pratique, on cherche à transférer le maximum de connaissances obtenus d'un domaine à un autre.

### 3.1.3 Application du Transfer Learning

Plusieurs systèmes de machines learning utilisent des simulations pour les modèles d'entraînements et la collecte de données. Comparé au fonctionnement dans le monde réel, cela permet des gains considérables en termes de temps et d'argent. On utilise alors le Transfer learning pour faire le lien entre la simulation et la réalité en termes de connaissance [13], car même si la simulation est proche de la réalité, la distribution de probabilités marginales est différente.

Les simulations permettent un apprentissage beaucoup plus rapide notamment grâce à un entraînement en parallèle sur plusieurs instances. On peut déjà voir le résultat de ces méthodes sur des technologies comme les voitures autonomes et les robots intelligents.

Le Transfer learning nous aide aussi à résoudre les problèmes d'adaptation de domaine où un modèle qui a été entraîné sur une certaine distribution se retrouve à être utilisé sur une autre qui lui est proche [12]. Ce type de modèle permet donc de s'adapter à chaque utilisateur.

### 3.1.4 Méthodes utilisées en Transfer Learning

La méthode la plus utilisée dans le TL est l'utilisation de caractéristiques de CNN pré-entraîné. Les couches convolutives inférieures nous donnent des détails de bas niveau sur les images tandis que les couches supérieures nous offrent des caractéristiques plus complexes. Ainsi, les couches de fully-connected layers permettent d'identifier des caractéristiques pertinentes dans la classification d'une image

parmi un nombre important d'objets [13]. Une méthode simple mais très efficace est d'entraîner un nouveau modèle à partir de caractéristiques d'un CNN pré-entraîné.

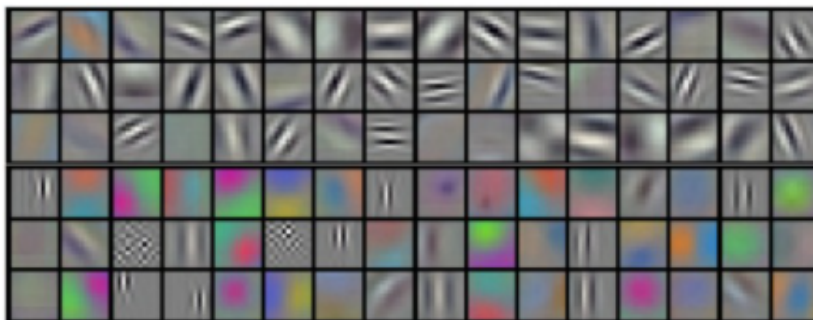


Figure 5 – Exemples de filtre appris par AlexNet. [13]

## 3.2 Unsupervised learning et pathologie digitale

### 3.2.1 Approches non-supervisées dans la pathologie digitale

Le self supervised learning est un sous-groupe des méthodes d'apprentissage non supervisé qui lui aussi se base sur des réseaux convolutifs pré-entraînés pour extraire des caractéristiques. C'est également une forme de Unsupervised learning où les réseaux neuronaux convolutifs sont auto-entraînés à l'aide de labels générés automatiquement. Il nécessite également une excellente connaissance du domaine afin d'élaborer une tâche prétexte qui nous permettrait d'obtenir des caractéristiques transférables [11].

Dans le cadre de l'étude de pathologies digitales, le problème principal pour appliquer des algorithmes de Deep Learning, est le manque de bases de données suffisamment grandes et bien annotées. De ce fait, l'idée qui vient à l'esprit est d'utiliser des méthodes de Self-supervised learning. Ainsi on arrive à générer des données étiquetées automatiquement pour des datasets larges et diversifiés [5]. C'est pourquoi, l'avancée du SSL permet au Deep Learning de ne plus être limité par le fait d'avoir des dataset non annotés.

Une autre idée serait d'utiliser des réseaux neuronaux pré entraînés d'une manière non supervisée et les adapter avec des algorithmes de clustering pour générer les étiquettes, ce genre de méthode est très prometteuse car elle ne demande pas des connaissances de domaine.

### 3.2.2 Deep Cluster

Dans ce sens, l'emploi des méthodes non supervisées serait plus logique vu que l'étiquetage d'une base donnée est une opération coûteuse et fastidieuse. Ainsi,



Caron et al. [11] proposent une nouvelle approche de clustering nommée Deep-Cluster pour la formation de bout en bout à grande échelle des convnets. Ils montrent qu'il est possible d'obtenir des fonctionnalités visuelles générales utiles avec un framework de clustering. Leur approche, résumée sur la figure 6, consiste à alterner entre le clustering des descripteurs d'images et la mise à jour des poids du convnet en prédisant les affectations de cluster.

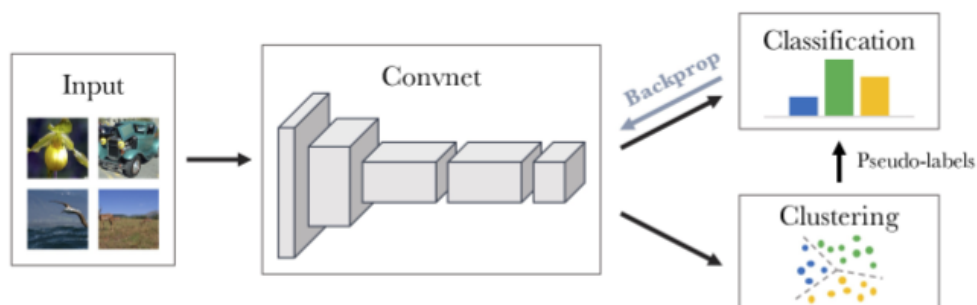


Figure 6 – Illustration de la méthode proposée. nous regroupons de manière itérative les fonctionnalités profondes et utilisons les attributions de cluster comme pseudo-étiquettes pour apprendre les paramètres du convnet, Caron et al. [11]

On commence par utiliser un réseau neuronal pré entraîné sur une base de données, dans son cas ImageNet, cette dernière est une base de données d'images organisée selon la hiérarchie WordNet (actuellement uniquement les noms), dans laquelle chaque nœud de la hiérarchie est représenté par des centaines et des milliers d'images. Cette base de données contient des millions d'images et permet au réseaux de neurones d'atteindre l'état de l'art actuel. Après on élimine la couche de sortie, pour qu'on utilise les descripteurs d'images comme entrées dans l'algorithme de classification comme k-means. Ensuite, on utilise les sorties de l'algorithme de classification comme base pour réentraîner notre réseau neuronal et améliorer les poids. On fait ça itérativement comme montré dans la figure 7:

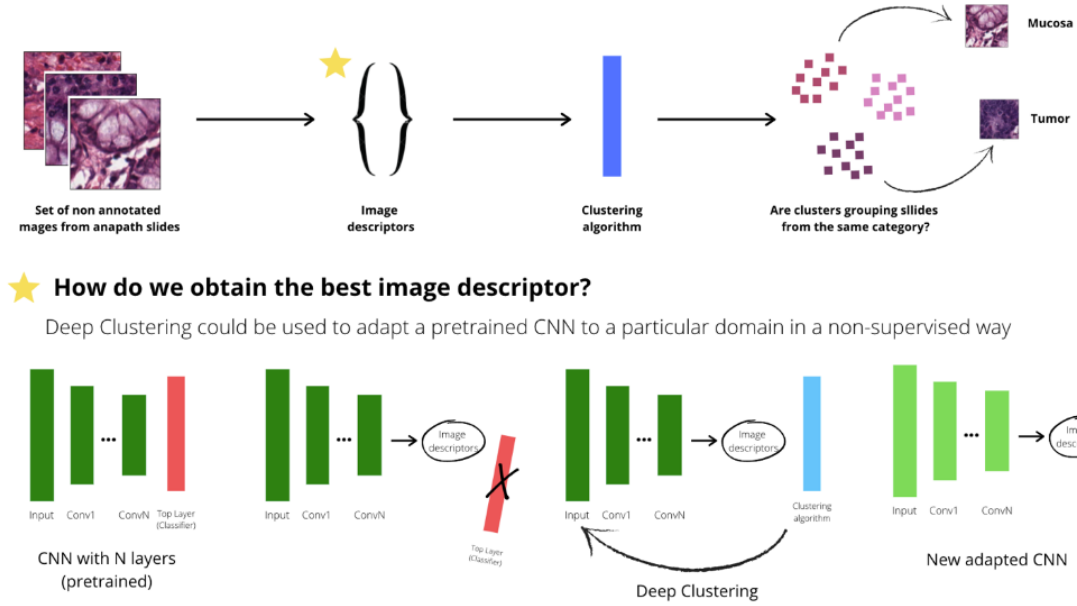


Figure 7 – Description de l’alternance entre des phases d’apprentissage du réseau et des phases de clustering dans l’espace des features des images [3].

L’application de cette méthode sur ImageNet a permis d’atteindre des performances nettement meilleures que l’état de l’art antérieur pour chaque tâche de transfert standard, cependant Il est naturel de se demander si une méthode de clustering telle que K-means, qui est plus basée sur l’approximation des données d’entrée, pourrait être combinée avec l’apprentissage de la représentation, qui utilise un objectif discriminant. En effet, Assano et al. [18] avait montré que combiner l’apprentissage des représentations, qui est une tâche discriminante, avec le clustering n’est pas du tout trivial. En particulier, la combinaison de minimisation d’entropie croisée et de K-means telle qu’adoptée par le DeepCluster ne peut pas être décrite comme l’optimisation d’un objectif d’apprentissage global; au contraire, il existe des solutions dégénérées que l’algorithme évite via des choix d’implémentation particuliers. Afin de remédier à cette lacune technique, Assano et al. proposent une nouvelle formule de principe pour l’apprentissage simultané du regroupement et de la représentation. Le point de départ est de minimiser une fonction de perte unique, la perte d’entropie croisée, pour l’apprentissage du réseau profond et pour l’estimation des étiquettes de données, cela se fait souvent dans le cadre d’un apprentissage semi-supervisé et d’un apprentissage d’instances multiples. Cependant, lorsqu’il est appliqué naïvement au cas non supervisé, il conduit immédiatement à une solution dégénérée où tous les points de données sont mappés sur le même cluster.

Néanmoins, le DeepCluster combine avec succès K-means et apprentissage de la représentation. Le DeepCluster peut être lié à l'approche de Assano et al. comme suit: L'étape 1 de l'algorithme, à savoir l'apprentissage de la représentation via la minimisation d'entropie croisée, est exactement la même. L'étape 2, à savoir l'auto-étiquetage, diffère: lorsque Assano et al. résolvent un problème de transport optimal pour obtenir les pseudo-étiquettes, Caron et al. le font en exécutant K-means sur les vecteurs de caractéristiques extraits par le réseau de neurones. Le DeepCluster a une solution dégénérée évidente: on peut assigner toutes les données pointant vers le même label et apprendre une représentation constante, atteignant simultanément un minimum de perte d'entropie croisée à l'étape 1 et de perte de K-means à l'étape 2. La raison pour laquelle le DeepCluster évite cet écueil est due à l'interaction particulière entre les deux étapes. Tout d'abord, au cours de l'étape 2, les fonctionnalités sont fixes et ne peuvent pas les rassembler. Au lieu de cela, les moyens s'étalent pour couvrir les fonctionnalités telles qu'elles sont, ce qui entraîne un partitionnement équilibré. Deuxièmement, lors de l'étape de classification, les attributions de cluster sont fixées, et l'optimisation des caractéristiques par rapport à la perte d'entropie croisée tend à les séparer. Enfin, la méthode de Caron et al. utilise également d'autres heuristiques telles que l'échantillonnage de la formation à l'inverse de la taille de leurs clusters associés, conduisant à une régularisation plus poussée. Ceci dit, la formulation d'Assano et al. est plus fondée sur des principes, ils montrent que convertir le problème d'affectation de pseudo-étiquettes est équivalent à un problème de transport optimal, cela leur permet de prouver que cette nouvelle approche conduit à des résultats nettement supérieurs au DeepCluster, atteignant le nouvel état de l'art pour les approches d'apprentissage de la représentation. [18]

### 3.2.3 Combinaison de K-means et apprentissage profond

#### 3.2.3.a Stabilité

Un autre problème qui résulte de la combinaison de K-means et l'apprentissage de la représentation est la perte de stabilité : Le Deep Clustering alterne entre le clustering de fonctionnalités approfondi et la mise à jour des paramètres CNN. En particulier, au début de chaque époque, il exécute des algorithmes de clustering hors-ligne sur l'ensemble de données pour obtenir des pseudo-étiquettes en tant que supervision de la prochaine époque. Le clustering hors ligne permute inévitablement les étiquettes attribuées à des époques indifférentes, c'est-à-dire que même si certains des clusters ne changent pas, leurs indices après le clustering seront permutés de manière aléatoire. Par conséquent, les paramètres du classifieur ne peuvent pas être hérités de la dernière époque et ils doivent être initialisés de manière aléatoire avant chaque époque. Le mécanisme introduit une instabilité de formation et expose les représentations à un risque élevé de corruption de représentation.[11, 10]

### 3.2.3.b Solutions

D'une part, Fan et al. [6] contournent ce problème de stabilité en utilisant l'apprentissage progressif non supervisé, sa méthode consiste à affiner le modèle de base à l'aide d'un jeu de données étiqueté non pertinent, qui est stocké en tant que modèle d'origine. Cette étape est également appelée initialisation du modèle d'origine. Deuxièmement, le modèle d'origine est utilisé pour extraire la fonctionnalité des échantillons de l'ensemble de données non étiqueté. Les instances de données sont regroupées et sélectionnées pour générer un ensemble d'apprentissage fiable. Troisièmement, le modèle original est parfaitement réglé par cet ensemble de formation fiable. Enfin, ils utilisent le nouveau modèle pour extraire la fonctionnalité, et les deuxième et troisième étapes sont répétées jusqu'à ce que l'échelle de l'ensemble d'apprentissage fiable devienne stable.

D'autre part, Zhang et al. [10] proposent une méthode en ligne du regroupement approfondie (ODC). L'ODC effectue la mise en cluster et la mise à jour du réseau simultanément plutôt qu'en alternance, plus précisément, une itération ODC se compose de propagations avant et de rétropropagation, de réaffectation d'étiquette et de mise à jour des centres de gravité. Pour la mise à jour des étiquettes, ODC utilise les fonctionnalités dans la propagation avant, évitant ainsi l'extraction de fonctionnalités supplémentaires. Pour faciliter la réaffectation des étiquettes en ligne et la mise à jour des centres de gravité, ils conçoivent et entretiennent deux modules de mémoire dynamique, c'est-à-dire une mémoire d'échantillons pour stocker les étiquettes et les fonctionnalités des échantillons, et la mémoire des centres de gravité pour l'évolution des centres de gravité. De cette manière, l'ODC est formé d'une manière ininterrompue similaire à une classification supervisée, sans qu'aucune annotation manuelle ne soit requise. Au cours du processus de formation, les étiquettes et les paramètres du réseau évoluent côte à côte, plutôt qu'en alternance. Étant donné que les étiquettes sont mises à jour à chaque itération en continu et instantanément, le classificateur dans CNN évolue également plus régulièrement et on ne risque plus un problème d'instabilité. Cependant, les méthodes Fan et al. [6], Zhang et al. [10] sont des méthodes qui demandent un ensemble de données à grande échelle, ce qui n'est pas le cas pour la pathologie digitale ou au moins notre base de données.

### 3.2.4 Tâche de classification

On arrive à la partie classification, dans notre implémentation on va utiliser dans un premier temps K-means, cet algorithme a l'avantage de ne pas avoir besoin de connaissance de domaine, si la classification prend beaucoup de temps on peut remédier à k-means++, Caron et al. [11] montre que généralement le Deep Cluster est plus performant en utilisant K-means que Clustering d'itération de puissance (PIC).

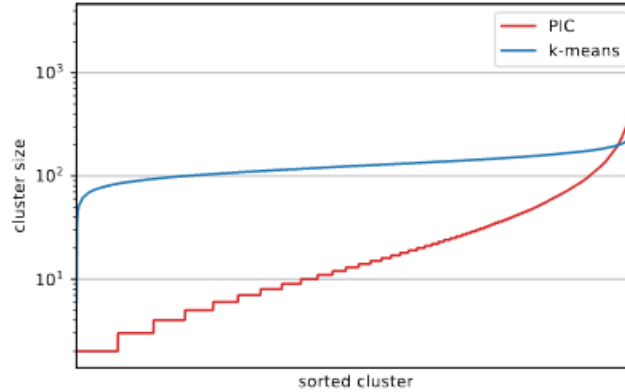


Figure 8 – Tailles des clusters produits par les versions k-means et PIC de Deep-Cluster à la dernière époque de la formation [11].

### 3.2.5 Utilisation des connaissances de domaines

À l’instant, les papiers qu’on a discutés n’utilisent pas des connaissances de domaines or il est logique de penser qu’incorporer une certaine quantité de connaissances spécifiques au domaine dans la méthode d’apprentissage peut s’avérer utile dans ces tâches, dans ce sens là Sari et al. [14] utilisent un nouvel extracteur de fonctionnalités qui utilise les connaissances antérieures du domaine pour l’identification des sous-régions saillantes et conçoit une méthode non supervisée pour leur caractérisation. Pour l’extracteur, il définit les caractéristiques en quantifiant uniquement les caractéristiques des sous-régions saillantes de l’image au lieu de considérer celles de tous les emplacements de l’image. A cet effet, il propose d’abord de définir les sous-régions saillantes autour des composants cytologiques du tissu . Ensuite, pour caractériser les sous-régions / composants de manière non supervisée, il apprend leurs caractéristiques locales par un réseau de croyances profondes constitué de RBM consécutifs et les quantifie en regroupant les caractéristiques locales par l’algorithme k-means. À la fin, il représente et classe l’image avec la distribution de ses sous-régions / composants quantifiées .

La RBM (machine de Boltzmann restreinte) est composée d’une couche de neurones qui reçoit l’entrée, ainsi que d’une couche de neurones cachée, De plus les neurones d’une même couche sont indépendants entre eux. Une vue schématique de la méthode proposée est donnée sur la figure suivante :

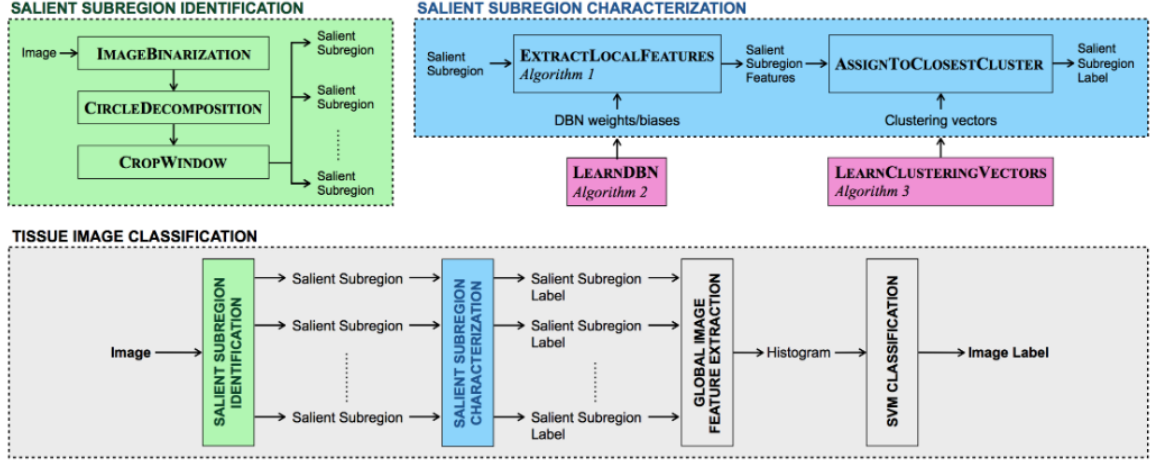


Figure 9 – un aperçu schématique de la méthode proposée [14].

Une autre approche pour utiliser les connaissances de domaines est abordée par Hedge et al. [7]. Ils proposent des méthodes de recherches de similarités des images histopathologiques en se basant sur des réseaux de neurones convolutifs (deep ranking network) pour apprendre des caractéristiques discriminatives des images d'entrées.

Cette méthode d'analyse d'image est présente également chez Gildenblat et al. [5]. Ils proposent une nouvelle méthode d'apprentissage auto-supervisée qui exploite la continuité spatiale intrinsèque des images histopathologiques des tissus pour former un modèle générant des descripteurs d'images spécifiques au domaine. A cette fin, Gildenblat et al. utilisent l'apprentissage de métriques sur images de diapositives entières -WSI- en entraînant un réseau neuronal siamois (parfois appelé réseau neuronal jumeau) qui est un réseau neuronal artificiel qui utilise les mêmes poids tout en travaillant en tandem sur deux vecteurs d'entrée différents pour calculer des vecteurs de sortie comparables, l'idée est que les WSI ont une continuité spatiale inhérente. Les carreaux spatialement adjacents sont généralement plus similaires les uns aux autres que les carreaux distants de l'image. Afin de générer notre jeu de données d'entraînement, ils définissent une distance maximale entre les paires de tuiles à étiqueter comme similaires, et une distance minimale entre les tuiles à étiqueter comme non similaires en fonction des tailles intrinsèques des régions histo-pathologiques dans l'ensemble de données. Pour chaque tuile de l'ensemble de données, d'autres éléments similaires ou non similaires sont échantillonnés en fonction des seuils de distance prédéfinis, créant un ensemble de données de paires étiquetées automatiquement. Cette stratégie d'échantillonnage permet de créer facilement un ensemble très large et diversifié de paires échantillonnées à partir de WSI d'histopathologie sans aucune annotation manuelle. Le résultat de cette méthode sur la dataset camelyon16 était impressionnant, dans l'expérience mesurant la distance L2 entre les descripteurs des tuiles éloignées et proches, ils comparent l'ADDR global pour ImageNet

pré-entraîné ResNet-50, la méthode NPID et leur approche proposée:

Method	Global ADDR
ResNet-50 pretrained on ImageNet	1.38
Non-Parametric Instance Discrimination	1.28
Ours	<b>1.5</b>

Figure 10 – Rapport de distance L2 entre les descripteurs de tuiles éloignées et proches [5].

Cependant, cette méthode a ses désavantages, dans un premier temps, dans certains cas, les paires d’images ne sont pas étiquetées avec précision. Par exemple, dans les transitions entre différentes zones histologiques fonctionnelles de l’image, il y’a par définitions des tuiles spatialement proximales qui sont visuellement différentes. Dans d’autres cas, deux carreaux distants peuvent être visuellement similaires car ils font partie de zones éloignées de la même fonction histopathologique, de plus d’un point de vue personnel. De plus, dans un cadre plus général on aura pas toujours des bases de données équilibrées [5].

### 3.2.6 Data augmentation

Comme mentionné précédemment, l’un des défis majeurs de la pathologie digitale réside dans le fait que la génération de données médicales annotées telles que des images pathologiques est particulièrement longue et coûteuse. Dans notre travail on va utiliser une bibilothèque qui comporte 5000 exemples, or les modèles d’état de l’art exigent dans la plupart des cas beaucoup plus de données pour atteindre une précision irréprochable. Pour ceci on utilise des stratégies qui permettent aux praticiens d’augmenter considérablement la diversité des données disponibles pour les modèles de formation, sans réellement collecter de nouvelles données. Les techniques d’augmentation des données telles que le recadrage, le remplissage et le retournement horizontal sont couramment utilisées pour entraîner de grands réseaux de neurones. Par exemple, la comparaison de recadrages aléatoires d’une image joue un rôle central en capturant des informations en termes de relations entre des parties d’une scène ou d’un objet. Néanmoins, ces méthodes ne sont pas sans inconvénient, l’augmentation du nombre de “crops” augmente de manière quadratique les besoins en mémoire et en calcul. Dans ce sens, Caron et al. [4] utilise une stratégie nommée “multi-crop”. Cette stratégie consiste à utiliser des images de plus petite taille pour augmenter le nombre de “crops” sans augmenter la mémoire ou les besoins de calcul pendant la formation. En effet, dans ce papier, on montre que la stratégie “multi-crop” semble bénéficier davantage des méthodes basées sur le regroupement que des méthodes contrastives, ce qui s’alignent avec nos objectifs.

## 4 Choix d'implémentations

Dans le cadre de notre projet, on travaille sur la base de données "[colorectal histology](#)": Cette base de données représente une collection de textures dans les images histologiques du cancer colorectal humain. On va commencer dans un premier temps par un réseau neuronal pré-entraîné sur ImageNet suivi par l'algorithme de classification classique k-means. Pour le choix de l'architecture du réseau on compte travailler avec:

- **AlexNet** [2]: AlexNet est le nom d'un réseau de neurones convolutifs (CNN), conçu par Alex Krizhevsky en collaboration avec Ilya Sutskever et Geoffrey Hinton, cette architecture se compose de huit couches: cinq couches convolutifs et trois couches entièrement connectées.
- **VGG16** [8]: VGG16 est un modèle de réseau neuronal convolutif proposé par K.Simonian et A. Zisserman de l'Université d'Oxford dans le papier «Very Deep Convolutional Networks for Large-Scale Image Recognition». Il apporte une amélioration par rapport à AlexNet en remplaçant les grands filtres de la taille d'un noyau (11 et 5 dans la première et la deuxième couche convolutive, respectivement) par plusieurs filtres  $3 \times 3$  de la taille d'un noyau l'un après l'autre.
- **ResNet-50** [17]: est un réseau neuronal convolutif d'une profondeur de 50 couches.

Après avoir comparé les résultats des architectures précédentes, on compte tester d'autres algorithmes de classifications, utiliser des méthodes d'augmentation de données et enfin explorer des algorithmes extracteur de fonctionnalités afin d'améliorer nos modèles.

## 5 Conclusion

Dans cette bibliographie, on explore la pathologie digitale et les défis qu'elle relève. Le problème principal pour appliquer des algorithmes de Deep Learning est le manque de bases de données suffisamment grandes et bien annotées. En effet, la génération de grandes bases de données est très coûteuse et peu efficace et pour relever ce défi, on a discuté des méthodes pour entraîner des réseaux neuronaux d'une manière non-supervisée. Un exemple est Le DeepCluster [11] qui est basée sur l'alternance entre des phases d'apprentissage du réseau et des phases de clustering dans l'espace des features des images. On a vu que cette alternance n'est pas sans inconvénients et qu'il y'a d'autres formulations plus robustes, Fan et al. [6], Zhan



et al. [10], Asano et al. [18]. Jusque là ces méthodes ne nécessitent ni n'utilisent de connaissances préalables spécifiques au domaine. Néanmoins, pour que les humains puissent accomplir certaines tâches, un aperçu qui ne pouvait être acquis que par une formation spécialisée dans le domaine connexe est généralement requis. Par conséquent, incorporer une certaine quantité de connaissances spécifiques au domaine dans la méthode d'apprentissage peut s'avérer utile dans ces tâches [14]. On explore donc des méthodes qui utilisent les propriétés intrinsèques des tissus soit en comparant les images et regroupant les tissus similaires [5, 7], soit en analysant les sous-régions saillantes [14]. Le point commun de tous ces travaux et notre problème actuel est celui de la base de données qui est assez 'petite' et donc non abordable en apprentissage supervisé et pour cela on discute d'une méthode d'augmentation de Data nommée Multi-crop [4]. Dans ce projet on essaie de proposer une méthode d'adaptation de domaine de réseaux pré-entraîné 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.

## References

- [1] Rantalainen M. Hartman J. Acs, B. Artificial intelligence as the next step towards precision pathology. in journal of internal medicine. *Blackwell Publishing Ltd*, 288, Issue 1:62–81, 2020.
- [2] Ilya Sutskever Alex Krizhevsky and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. 2012.
- [3] Pilar Ortega et Sandrine Mouysset Arnaud Abreu. Sujet (23) du projet long – oncopole pour l’année 2020-2021.
- [4] Misra I. Mairal J. Goyal P.-Bojanowski P. & Joulin A.(2020) Caron, M. Unsupervised learning of visual features by contrasting cluster assignments. url: <http://arxiv.org/abs/2006.0988>.
- [5] Klaiman E. (n.d.) Gildenblat, J. Self-supervised similarity learning for digital pathology.
- [6] L. Zheng C. Yan H. Fan, Y. I. Yang and Y. Yang. Unsupervised person re-identification: Clustering and fine-tuning. *ACM Trans. Multimed. Comput. Commun. Appl*, 14(4), 2018.
- [7] Hipp J. D. Liu Y. Emmert-Buck M. Reif E. Smilkov D. Terry M. Cai C. J. Amin M. B. Mermel C. H. Nelson P. Q. Peng L. H. Corrado G. S. Stumpe M. C. (2019). Hegde, N. Similar image search for histopathology: Smily. pages 1–9, May.
- [8] Tian Y. Jing, L. Very deep convolutional networks for large-scale image recognition. 2014.
- [9] D. L. Rimm-V. Velcheti K. Bera, K. A. Schalper and A. Madabhushi. Artificial intelligence in digital pathology —new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.*, 16(11):703–715, nov. 2019.
- [10] Xiaohang Zhan • Jiahao Xie • Ziwei Liu • Yew Soon Ong • Chen Change Loy. *Online Deep Clustering for Unsupervised Representation Learning CVPR 2020*.
- [11] A. Joulin M. Caron, P. Bojanowski and M. Douze. Deep clustering for unsupervised learning of visual features. *Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell*, 11218 LNCS(11):139–156, 2018.
- [12] S. J. Pan and Q. Yang. A survey on transfer learning. 2009.
- [13] S. Ruder. Transfer learning -machine learning’s next frontier. url: <https://ruder.io/transfer-learning/>. [accessed: 12-oct-2020],. 2017.

## REFERENCES

---

- [14] Gunduz-Demir C. Sari, C. T. Unsupervised feature extraction via deep learning for histopathological classification of colon tissue images. *ieee transactions on medical imaging*, 38(5), 1139–1149.
- [15] Digital Pathology – Transforming the Future of Lab Testing. <https://www.futurebridge.com/industry/perspectives-life-sciences/digital-pathology/>.
- [16] Hamid Reza Tizhoosh and Liron Pantanowitz. Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics*, 2018.
- [17] Understanding and Coding a ResNet in Keras. <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>.
- [18] Andrea Vedald (2020): Yuki Markus Asano, Christian Rupprecht. Self-labelling via simultaneous clustering and representation learning.