

Proposition de stage 2021 - 2022

Apprentissage auto-supervisé pour la détection d'anomalies en neuroimagerie.
Application au diagnostic précoce de la maladie de Parkinson.

Laboratoire d'accueil : Laboratoire CREATIS, 69 Villeurbanne- MYRIAD Team

Supervision : Carole Lartizien -Nicolas Pinon

Mots clés : Neuroimagerie, Apprentissage profond, Apprentissage auto-supervisé, Apprentissage de représentation

Durée : 6 mois

Date de début : Fin Hiver - début Printemps 2021-22 (ajustable)

Indemnité : ~ 560 euros/mois

Contexte scientifique

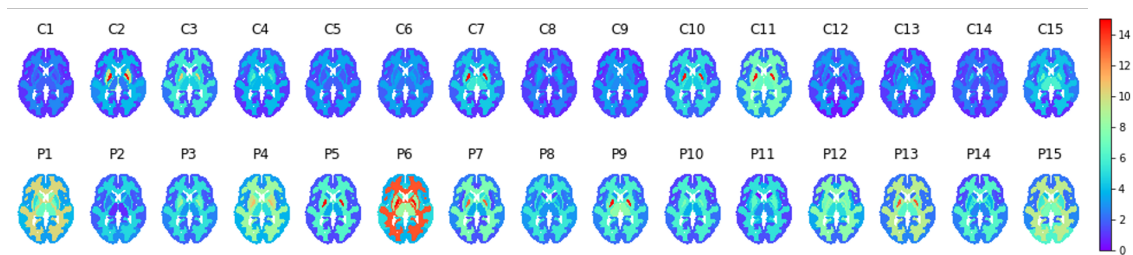


Figure 1: Exemples de cartes paramétriques représentant le pourcentage de voxels anormaux dans des régions cérébrales. haut : 15 témoins (sujets sains) ; en bas : 15 patients atteints de la maladie de Parkinson.[1]

Les progrès récents en apprentissage automatique ont conduit à des résultats très prometteurs en imagerie médicale, allant de la segmentation, recalage, reconstruction au diagnostic et au pronostic.

La grande majorité des méthodes proposées sont supervisées ; elles nécessitent de collecter de nombreux exemples annotés et voient leur performance décroître lorsque les modèles sont appris avec peu de données, avec des données hétérogènes ou encore lorsque les annotations sont fortement déséquilibrées. Ces cas critiques se rencontrent particulièrement dans certaines tâches de détection de modification anormale de forme ou de lésions de petite taille, parfois difficilement décelables à l'œil nu et donc coûteuses voire impossibles à annoter manuellement. C'est le cas pour différentes pathologies cérébrales, comme les micro-hémorragies, les lésions épileptogènes ou de sclérose en plaques, ainsi que des variations fines de volume de certaines petites structures cérébrales profondes, par exemple, dans la maladie de *Parkinson* [2].

Un cadre méthodologique alternatif qui semble prometteur et adapté à l'analyse de ces données est celui de la détection d'anomalies dans un contexte *non supervisé* (dit aussi *auto-supervisé*). Il s'agit d'apprendre un modèle de la normalité, généralement dans un espace de représentation latent, à partir des données saines uniquement, puis de considérer comme des anomalies les échantillons test s'écartant trop de ce modèle. Cette dernière étape est généralement réalisée en calculant l'erreur entre les données originales et les données reconstruites à partir de leur projection dans l'espace de représentation latent.

CREATIS a développé une expertise dans le domaine des méthodes de détection d'anomalies pour l'analyse d'images cérébrales multi-modalités. Notre approche repose sur l'apprentissage de représentation par des *réseaux auto-encodeurs siamois* (SAE) comme illustré sur le bas de la figure 2. Ce type de modèle donne de bonnes performances pour la détection de lésions épileptogènes en IRM multiparamétrique [3]. En collaboration avec le Centre de Neurosciences (GIN) et l'INRIA de Grenoble, nous avons récemment appliqué ce modèle à la détection des formes précoces de la maladie de Parkinson en IRM multiparamétrique, et comparé notre modèle à un auto-encodeur simple (AE) représenté

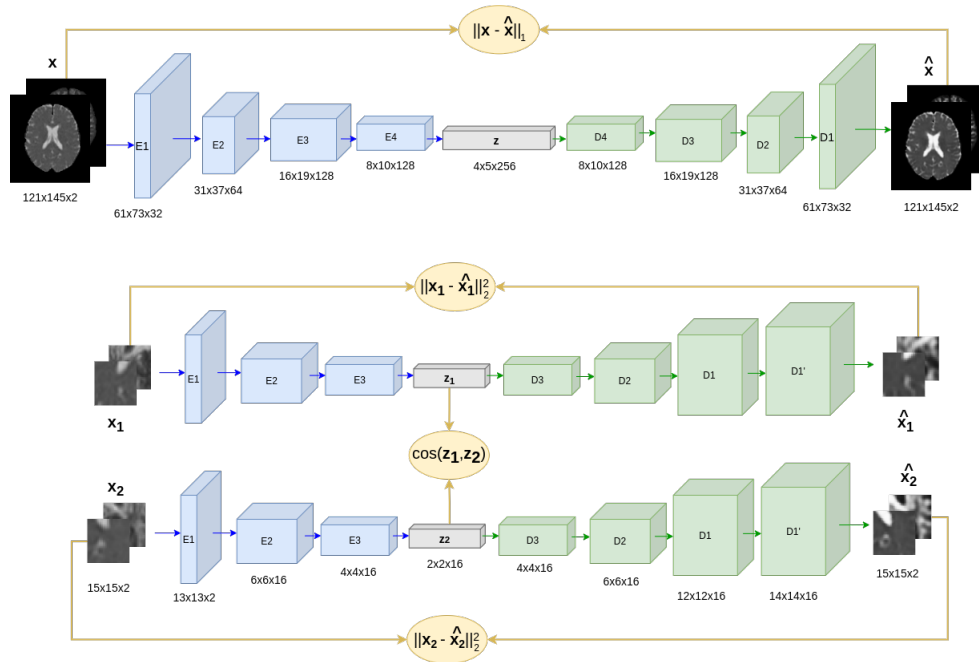


Figure 2: Auto-encodeurs simple (haut) et siamois par patch (bas) pour l'apprentissage de la représentation saine des sujets.

sur le haut de la figure 2. Les résultats de cette étude illustrés sur la figure 1 ont donné lieu à un article de conférence publié au workshop MICCAI MLCN (Machine Learning Clinical Neuroimaging) 2021 [1].

La figure 2 illustre les deux architectures utilisées pour ce projet. Une des particularités du réseau siamois (figure du bas) est de forcer les patches colocalisés dans le cerveau à être proches au sens d'une certaine métrique dans l'espace latent. L'influence du choix de la métrique ou de l'architecture du réseau sur la répartition des patches dans l'espace latent est encore peu étudiée et pourrait être une aide au choix de ces premiers.

Objectif du stage

L'objectif de ce projet est de poursuivre ce projet en explorant les pistes suivantes :

- Comparaison d'approches par patch versus image entière et siamois versus simple :
Les versions actuelles des modèles SAE et AE considèrent en entrée, des patch 2D (image de taille 15X15) extraites des images originales pour le premier, pour le second des coupes transverses 2D du cerveau ou d'un hémisphère. Par rapport au modèle AE, le modèle SAE permet d'augmenter la taille de la base d'apprentissage (extraction de dizaine de milliers de patches par volumes d'images 3D) au détriment d'une perte d'information sur le contexte spatial. De plus, le modèle SAE contraint les patches colocalisés de différents patients à être proches dans l'espace latent, le modèle AE n'impose pas de contraintes sur les tranches colocalisées. Le premier objectif du stage sera d'implémenter les modèles SAE et AE complémentaires (SAE sur image entière / AE sur patch).
- Développement d'outils d'analyse de l'espace latent :
Afin de mieux comprendre les mécanismes de ces modèles, nous souhaitons explorer les caractéristiques de leur espace de représentation. L'objectif sera de prendre en main des méthodes de visualisation de l'espace latent (t-SNE, UMAP, spectral embedding, etc.) et les adapter à notre problématique. Il pourrait aussi être intéressant d'étudier l'influence des fonctions de coût choisies ou de la régularisation variationnelle sur ces représentations.

Le candidat ou la candidate choisi pourra accéder à la base de données PPMI (<https://www.ppmi-info.org/access-data-specimens/download-data>) contenant de multiples images de contrôles et patients parkinsoniens en différentes modalités ainsi qu'à des ressources de calculs (supercalculateur de CREATIS et/ou du CNRS).

Compétences

Le candidat ou la candidate choisi devra avoir de solides connaissances en machine learning et/ou traitement d'image, ainsi que de bonnes compétences en programmation (Python). De l'expérience avec des bibliothèques de machine learning (Tensorflow, Pytorch ou Scikit learn) et/ou les systèmes d'exploitation Linux seront appréciées. Nous sommes à la recherche d'un étudiant ou d'une étudiante motivé et autonome avec un fort intérêt dans la recherche pluridisciplinaire (traitement d'image et machine learning dans un contexte médical). La personne choisie sera en constante collaboration avec un étudiant en thèse, Nicolas Pinon, travaillant sur ce projet.

Candidatures

Les candidats intéressés devront envoyer leur CV, lettre/mail de motivation et relevés de notes à : **carole.lartizien@creatis.insa-lyon.fr** et **nicolas.pinon@creatis.insa-lyon.fr**

References

- [1] Verónica Muñoz-Ramírez, Nicolas Pinon, Florence Forbes, Carole Lartizen, and Michel Dojat. Patch vs. global image-based unsupervised anomaly detection in mr brain scans of early parkinsonian patients. In Ahmed Abdulkadir, Seyed Mostafa Kia, Mohamad Habes, Vinod Kumar, Jane Maryam Rondina, Chantal Tax, and Thomas Wolfers, editors, *Machine Learning in Clinical Neuroimaging*, pages 34–43, Cham, 2021. Springer International Publishing.
- [2] Verónica Muñoz Ramírez, Virgilio Kmetzsch, Florence Forbes, and Michel Dojat. Deep learning models to study the early stages of parkinson's disease. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1534–1537, 2020.
- [3] Zaruhi A., J. Jung, R. Bouet, and C. Lartizien. Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: Application to epilepsy lesion screening. *Medical Image Analysis*, 60:101618, 2020.