PhD Fellowship
Interprétation des réseaux de neurones pour l'imagerie médicale

Mots-clefs : interprétabilité, apprentissage profond, segmentation, recalage, diagnostic

Domaine et contexte scientifiques :

L'apprentissage profond a fait faire un bond en avant fulgurant à l'analyse d'image médicale. Cependant, les réseaux de neurones restent des "boîtes noires": on ne sait ni comment sont prises les décisions ni quelles sont les caractéristiques de l'image d'entrée utilisé pour la décision. Or le domaine médical est un domaine critique dans lequel on shouhaite la décision soit transparente ou au moins vérifiable par clinicien. Par ailleurs, les bases d'apprentissage/validation/test étant souvent issues de la partition d'une même base globale, un biais dans celle ci affectera la généralisation du réseau qui montrerait pourtant de bons résultats.

Les cartes d'attributions [IG,EG], identifiant les pixels d'importance dans le décision d'un réseau, permettent parfois de voir, a posteriori, les zones utilisées par un réseau pour prendre sa décision.Avec ces outils, nous avons montré à CREATIS, que pour un classifieur sujets sains vs pathologiques, on pouvait améliorer l'interprétabilité [Wargnier2021] ou utiliser ces cartes à l'apprentissage pour baser la décision sur les lésions (voir figure ci dessous) [Wargnier 2022].

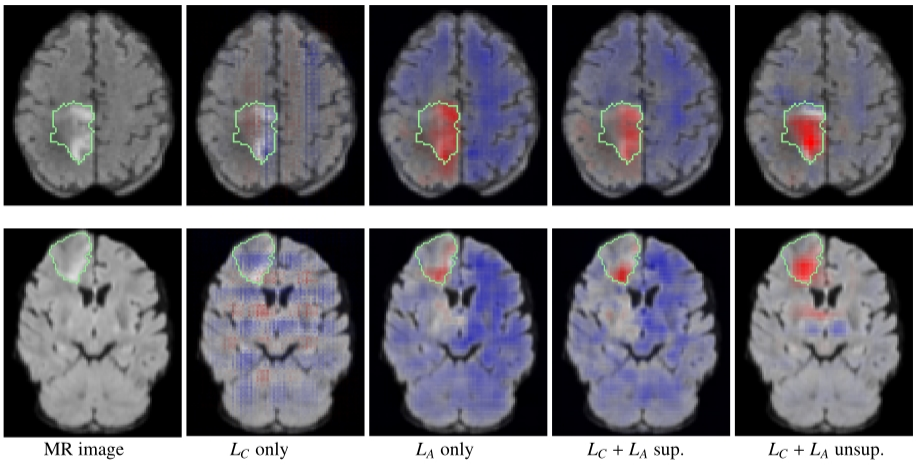

Fig. 1: Brain tumors experiments: Two axial view attribution examples. Tumors manual annotation is drawn in green. Blue represents healthy relevance and red pathological relevance. From left to right: MR image, attribution maps for model trained with classification loss only (L_C only), with attribution loss only (L_A only), with both loss in supervised mode ($L_C + L_A$ sup.) and with both loss in unsupervised mode ($L_C + L_A$ unsup.). Models are trained with Integrated+Expected Gradients and evaluated with Integrated Gradients.

Programme de recherche proposée :

Différents axes pourront être abordés durant la thèse.

Nous avons constaté qu'avec notre terme de régularisation [Wargnier2022], l'apprentissage était plus long et plus difficile. Le premier axe de la thése sera un travail sur la stabilisation de l'apprentissage: comprendre pourquoi l'apprentissage est plus compliqué et proposer des méthodes pour réduire les temps d'apprentissage.

Notre terme de régularisation n'étant adapté qu'à des problèmes de classification, dans un second axe nous pourrons étendre notre régularisation à des tâches de régression. Nous pourrons comme cas d'application des problèmes de recalage affine ou de prédiction du handicap en sclérose en plaque.

Dans un troisième axe, nous étendrons notre approche à la segmentation. Pour cette tâche ce type de régularisation pourrait réduire le nombre de sujets annotées nécessaire. Il faudra prendre en compte deux nouveaux aspects: le changement d'architectures qui sont de type encodeur/decodeur et le fait que la sortie n'est pas qu'une unique valeurs mais une image de même taille que l'image d'entrée.

Profil recherché

Formation : machine/deep learning, mathématiques appliquées, analyse d'images, vision par ordinateur.
Un excellent niveau en développement logiciel, anglais (écrit/oral) et des facilités de rédaction sont aussi attendues.

Merci de fournir les documents suivant avec votre candidature :
- Curriculum Vitae
- Diplômes et notes pour l'ensemble du cursus universitaire
- Lettre(s) de recommandation de l'encadrant de stage
- Lettre de motivation

Superviseurs:
Michaël Sdika CREATIS www.creatis.insa-lyon.fr, michael.sdika [at] creatis.insa-lyon.fr
Christophe Garcia : LIRIS https://liris.cnrs.fr, christophe.garcia[at] insa-lyon.fr

Where and When:

Lieu : CREATIS La Doua Campus, Villeurbanne, France
Démarrage: Octobre 2022

Références bibliographie sur le sujet :

[IG] Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks, in: International Conference on Machine Learning, PMLR. pp.3319–3328.
[EG] Erion, G., Janizek, J.D., Sturmfels, P., Lundberg, S.M., Lee, S.I., 2021. Improving performance of deep learning models with axiomatic attribution priors and expected gradients. Nature Machine Intelligence , 1–12
[Wargnier2021] Wargnier-Dauchelle, V., Grenier, T., Durand-Dubief, F., Cotton, F., Sdika, M., 2021. A more interpretable classifier for multiple sclerosis, in: IEEE ISBI 2021. pp.1062–1066.
[Wargnier2022] Valentine Wargnier-Dauchelle, Thomas Grenier, Françoise Durand-Dubief, François Cotton, Michaël Sdika, Attribution constraints for interpretable and relevant deep classifier, Medical Image Analysis, submitted