UNIVERSITÉ DE RENNES 1

Direction de la Recherche et de l'Innovation

Appel à projets - Année 2021

Contrats doctoraux «Ordinaires»

ECOLE DOCTORALE : MathSTIC **POLE DE RECHERCHE : Informatique**

Unité de recherche d'accueil

Unité de Recherche : IRISA UMR 6074

Equipe : LACODAM & Empenn

Encadrement

Directeur de thèse : Elisa Fromont **Unité de recherche** : IRISA UMR 6074

Employeur/Établissement d'origine : Université de Rennes 1

HDR : X oui • non **Nb de thèses en cours** : 8 (deux se terminent en 2021, toutes sont codirigées)

Mail : elisa.fromont@irisa.fr **Tél**. : 02 99 84 73 14

Co-directeur de thèse : Camille Maumet **Unité de recherche** : IRISA UMR 6074

Employeur/Établissement d'origine : Inria

HDR : • oui X non **Nb de thèses en cours** : 1

Mail : camille.maumet@inria.fr **Tél.** : 02 99 84 72 30

Sujet

Titre : Mapping the fMRI pipeline-space towards more robust pipelines

Mots clés : Unsupervised Deep learning, Metric Learning, Brain imaging, Statistics, fMRI, Pipelines

Cotutelle internationale (oui/non) : non **Si oui, préciser l'établissement et le pays de rattachement** :

Sollicitation d'un cofinancement EDUC (oui/non) : non

Descriptif (15-30 lignes) :

Contexte socioéconomique et scientifique ; Les hypothèses et questions posées ; Les grandes étapes de la thèse et démarche ; Approches méthodologiques et techniques envisagées

In a recent study, a collaboration of 200 researchers demonstrated how the choice of brain imaging pipeline can have a non-negligible impact on the findings of a study (Botvinik-Nezer et al., 2020). In this landmark paper, researchers - divided in 70 teams - used the *same* dataset to answer the *same* predefined research questions with the only varying parameter being the analysis pipeline that was selected by each team based on its expertise. In total, the 70 teams chose 70 different pipelines and in multiple instances, the results obtained across pipelines were contradictory.

The pipeline-space in neuroimaging is especially large. Many steps are needed to prepare and analyze neuroimaging data including: motion correction, registration to a standardized space, spatial smoothing, etc. And for each step, different algorithms are available, implemented in various neuroimaging software. While some good practices exist to guide some of those choices, researchers are left with many choices to be made. This thesis will focus on providing a better understanding of the relationships that exist between different pipelines.

The goal of this thesis is to build a "map" of the neuroimaging pipeline-space that will be used to investigate one or more of the following challenges: 1) Pipeline debugging; 2) Identify when pipelines suffer from violations of their underlying assumptions; 3) Eliminate unsuitable pipelines given an input dataset.

Debugging. Some pipelines have been shown to be unstable in the presence of small perturbation in the input data, questioning the robustness of their implementation (Kiar et al., 2020). The goal of pipeline debugging is to identify those pipelines that do not behave as expected due to the presence of a bug.

Assumption violation. Each step of an analytic pipeline is the implementation of a method that comes with some assumptions. If those assumptions are not met, this can lead to an incorrect behaviour of the pipeline.

Dataset-specific unsuitability. In other instances, a pipeline will fail for a given dataset due to the particular properties of these data (e.g. specific noise properties, etc.).

In all of the cases above, a systematic investigation of the pipeline space is impractical. In addition, there is usually no ground truth that can be used to benchmark pipelines results in neuroimaging. In this thesis, we propose instead to provide a representation of fMRI statistics in a lower dimensional space and to use this space to estimate how close or distant different pipelines are. Using this map of distances across pipelines in different contexts, we will derive a solution for the problems exposed above. This work will take advantage of the largest database of fMRI statistic map: NeuroVault (https://neurovault.org/) (Gorgolewski et al., 2015) to build an encoder (GAN, autoencoder, etc.), building on previous work available at: https://github.com/gkiar/analytical-stability.