



HAL
open science

Élagage de réseaux de neurones convolutifs sur graphes pour la sélection de fréquences significatives pour le décodage d'IRMf

Yassine El Ouahidi, Hugo Tessier, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup, Vincent Gripon

► **To cite this version:**

Yassine El Ouahidi, Hugo Tessier, Giulia Lioi, Nicolas Farrugia, Bastien Padeloup, et al.. Élagage de réseaux de neurones convolutifs sur graphes pour la sélection de fréquences significatives pour le décodage d'IRMf. GRETSI 2022: 28ème colloque du Groupement de Recherche en Traitement du Signal et des Images, Sep 2022, Nancy, France. hal-03696640

HAL Id: hal-03696640

<https://hal.science/hal-03696640>

Submitted on 16 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Élagage de réseaux de neurones convolutifs sur graphes pour la sélection de fréquences significatives pour le décodage d'IRMf

Yassine EL OUAHIDI, Hugo TESSIER, Giulia LIOI, Nicolas FARRUGIA, Bastien PASDELOUP, Vincent GRIPON

IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest F-29238, France

prenom.nom@imt-atlantique.fr

Résumé – Le traitement du signal sur graphe permet de manipuler des signaux évoluant sur des structures irrégulières, comme par exemple les signaux cérébraux, en exploitant les dépendances spatiales entre les régions d'intérêt dans le cerveau. Dans ce contexte, nous nous intéressons à mieux comprendre quelles sont les fréquences du graphe les plus utiles pour décoder des signaux d'Imagerie par Résonance Magnétique Fonctionnelle (IRMf). À cette fin, nous introduisons une architecture d'apprentissage profond et adaptons une méthode d'élagage pour identifier automatiquement ces fréquences. Nos expériences montrent que les basses fréquences des graphes sont systématiquement identifiées comme les plus importantes, avec une contribution plus forte pour le graphe fonctionnel que pour le graphe structurel.

Abstract – Graph signal processing defines tools to manipulate signals evolving on irregular domains, such as brain signals, by encompassing the spatial dependencies between regions of interest in the brain. In this work, we are interested in better understanding what are the graph frequencies that are the most useful to decode Functional Magnetic Resonance Imaging (fMRI) signals. For that, we introduce a deep learning architecture and adapt a pruning methodology to automatically identify such frequencies. Our experiments show that low graph frequencies are consistently identified as the most important for fMRI decoding, with a stronger contribution for the functional graph over the structural one.

1 Introduction

Le développement de l'IRMf a permis d'observer le cerveau *in vivo* et d'aborder l'une des questions centrales des neurosciences cognitives : comprendre la relation entre l'activation cérébrale et les fonctions cognitives. Les approches classiques reposent sur l'*inférence directe* (identification d'effets localisés dans le cerveau correspondant à un changement de condition expérimentale) ou l'*inférence inverse* (inférence de la fonction cognitive à partir des schémas d'activation cérébrale). Plus récemment, il a été suggéré que le décodage cérébral (*i.e.*, *identifier* une condition expérimentale à partir des schémas d'activation cérébrale) est une manière plus formelle d'effectuer une inférence inverse [1], car il permet d'identifier les structures cérébrales qui contribuent à la réalisation d'une tâche cognitive spécifique. De nombreux travaux ont étudié le décodage cérébral via des paradigmes classiques d'apprentissage automatique appliqués à des données IRMf provenant de différents sujets et tâches (pour un exemple récent, voir [2]). Certaines de ces études ont considéré des réseaux convolutifs sur graphes pour décoder des signaux IRMf en prenant en compte la connectivité structurelle [3] ou fonctionnelle [4] du cerveau, grâce aux outils du Traitement du Signal sur Graphe (TSG) [5].

Le TSG vise à étendre l'analyse de Fourier à des signaux évoluant sur des domaines irréguliers représentés par des graphes. Les notions de fréquences et de modes de Fourier, ainsi que les opérateurs associés tels que les convolutions ou le filtrage sont

définis par analogie avec le traitement du signal classique.

Dans ce travail, nous nous intéressons à mieux comprendre quelles sont les fréquences du graphe qui sont les plus utiles pour décoder les signaux IRMf. À cette fin, nous introduisons une architecture d'apprentissage profond, exploitant un graphe, appelée *ResNet spectral*. Les fréquences du graphe y sont exploitées pour définir des convolutions. En utilisant une technique d'élagage, nous identifions dynamiquement les fréquences les plus importantes pour décoder les signaux IRMf. Nos résultats indiquent que les basses fréquences sont les plus significatives, et sont robustes aux changements de jeux de données, de graphes (structurel ou fonctionnel) et d'architectures considérées¹.

2 Contexte scientifique

Le TSG généralise l'approche de Fourier aux signaux évoluant sur des structures irrégulières, en fournissant un espace spectral adapté pour les décomposer en fréquences [5]. Soit un graphe pondéré et non dirigé $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, W \rangle$ avec des sommets $\mathcal{V} = \{v_1, \dots, v_N\}$ de cardinal $|\mathcal{V}| = N$, des arêtes $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, et une fonction de pondération $W : \mathcal{E} \mapsto \mathbb{R}$. Un tel graphe peut être représenté de manière équivalente par sa matrice de poids $\mathbf{W} \in \mathbb{R}^{N \times N}$ telle que $\mathbf{W}[i, j] = W(\{v_i, v_j\})$ si $\{v_i, v_j\} \in \mathcal{E}$ et 0 sinon. De plus, on note $\mathbf{D} \in \mathbb{R}^{N \times N}$ la matrice des degrés de \mathcal{G} , telle que $\mathbf{D}[i, j] = \sum_{k=1}^N \mathbf{W}[i, k]$ si $i = j$ et 0 sinon. À partir de ces deux matrices, nous pouvons calculer la matrice laplacienne normalisée $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}$ de \mathcal{G} , où

¹ Nous tenons à remercier Yu Zhang pour avoir fourni le graphe fonctionnel. Nous remercions également la région Bretagne pour son soutien.

1. <https://github.com/elouayas/gspconv>

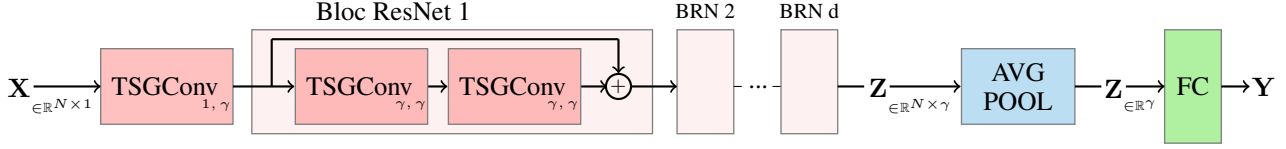


FIGURE 1 – Schéma du ResNet spectral. L’entrée est vue comme un signal sur graphe à 1 canal, transformé en un signal à γ canaux par une couche TSGConv. Il passe ensuite par d blocs ResNet, dont la sortie est la somme entre son entrée et une succession de 2 couches TSGConv. Une couche dense est appliquée à une moyenne selon l’axe des fréquences pour prédire la classe des signaux.

\mathbf{I}_N est la matrice identité de dimension N . \mathbf{L} peut être diagonalisée sous la forme $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, où \mathbf{U} est une matrice de vecteurs orthonormaux associés aux valeurs propres formant une matrice diagonale $\mathbf{\Lambda}$, triés par ordre croissant. Ces valeurs propres sont analogues aux fréquences de l’analyse de Fourier, et sont appelées *fréquences du graphe*.

Un signal $\mathbf{x} \in \mathbb{R}^N$ sur \mathcal{G} est une observation sur chacun de ses sommets. Sa transformée de Fourier sur graphe $\hat{\mathbf{x}} = \text{TFG}(\mathbf{x}) = \mathbf{U}^\top \mathbf{x}$ fournit pour le signal un spectre, *i.e.*, une décomposition en une somme de fréquences du graphe. Son inverse $\mathbf{x} = \text{TFG}^{-1}(\hat{\mathbf{x}}) = \mathbf{U}\hat{\mathbf{x}}$ transforme un spectre en un signal sur graphe. Soient $\mathbf{h} \in \mathbb{R}^N$ un filtre, et \odot la multiplication terme à terme. L’opérateur de convolution est défini comme :

$$\mathbf{x} * \mathbf{h} = \text{TFG}^{-1}(\text{TFG}(\mathbf{x}) \odot \text{TFG}(\mathbf{h})) . \quad (1)$$

Contrairement aux méthodes de sélection de fréquences existantes dans la littérature [5], nous proposons ici une méthode où la sélection de fréquences (non nécessairement contiguës) est effectuée automatiquement tout en optimisant pour une tâche supervisée. Dans les travaux précédents utilisant le TSG pour décoder l’activité cérébrale [3, 4], l’accent a été mis sur l’augmentation de la performance de généralisation. Ici, nous cherchons à déterminer quelles fréquences du graphe sont les plus pertinentes pour le décodage de l’activité cérébrale. Nous évaluons la fiabilité de nos résultats en considérant deux jeux de données et en testant différentes architectures d’apprentissage profond. Nous abordons les questions suivantes : 1) Quelles fréquences du graphe sont les plus pertinentes pour identifier la tâche expérimentale à partir des cartes d’activation IRMf correspondantes ? 2) Ces fréquences sont-elles cohérentes pour des graphes (fonctionnel ou structurel) et architectures d’apprentissage profond différents ? 3) La sélection de bandes de fréquences du graphe spécifiques augmente-t-elle les performances de décodage ?

3 Méthodologie

Nous souhaitons exploiter le graphe sur lequel les signaux d’entrée sont définis pour améliorer leur classification et leur interprétation en terme de motifs d’activation spatiale dans le cerveau. L’idée est de sélectionner automatiquement, par un processus d’apprentissage, les fréquences pertinentes du graphe.

3.1 La couche TSGConv

Dans le but de définir une architecture d’apprentissage profond, nous introduisons une couche TSGConv en utilisant

un graphe \mathcal{G} à N sommets. Elle prend en entrée une matrice $\mathbf{X} \in \mathbb{R}^{N \times c_{\text{in}}}$ vue comme un signal sur graphe avec c_{in} canaux et produit une matrice $\mathbf{Z} \in \mathbb{R}^{N \times c_{\text{out}}}$ vue comme un signal sur graphe avec c_{out} canaux. Elle contient un tenseur de paramètres $\Theta \in \mathbb{R}^{N \times c_{\text{in}} \times c_{\text{out}}}$. Sa fonction mathématique est :

$$\text{TSGConv}_{c_{\text{in}}, c_{\text{out}}}(\mathbf{X}) = \sigma(\text{TFG}^{-1}(\text{TFG}(\mathbf{X}) \otimes \Theta)) , \quad (2)$$

où TFG et TFG^{-1} sont appliqués à chaque canal indépendamment, σ est une fonction d’activation non linéaire, et \otimes est tel que $(\hat{\mathbf{x}} \otimes \Theta)[l, c] = \sum_{c'=1}^{c_{\text{in}}} \hat{\mathbf{x}}[l, c'] \Theta[l, c', c]$.

3.2 Architectures considérées

En utilisant la couche TSGConv, nous construisons un ResNet spectral, inspiré des modèles ResNet [6], consistant en l’assemblage d’une première couche TSGConv, puis de blocs ResNet spectraux, et enfin d’une couche dense dont l’entrée est un vecteur moyenné selon l’axe des fréquences (cf. Figure 1). Ce modèle est paramétré par une profondeur d et un nombre de canaux γ , et contient $\mathcal{O}(\gamma N + 2d\gamma^2 N + \gamma C)$ paramètres entraînaibles, avec C le nombre de classes du problème considéré.

À des fins de comparaison, nous proposons également une architecture basée sur un Perceptron Multi-Couches (PMC) dont le domaine d’entrée est le domaine de fréquences du graphe (*i.e.*, $\text{TSGConv}_{1,1}(\mathbf{X})$). Une telle architecture exploite la structure des graphes d’une manière très différente de la précédente, et est utilisée pour souligner la généralité de nos résultats. Par simplicité, nous utilisons un nombre fixe h de neurones cachés par couche et une profondeur d . Un tel modèle contient un total de $\mathcal{O}(hN + h^2d + hC)$ paramètres entraînaibles.

3.3 Élagage pour la sélection des fréquences

L’élagage [7] vise à identifier les parties non importantes des modèles d’apprentissage profond et à les supprimer tout en maintenant un niveau de performance élevé. Nous exploitons ce paradigme afin de supprimer les fréquences du graphe les moins importantes pour notre application. Nous avons choisi d’implémenter la Décroissance de Poids Sélectif (DPS) [8] qui consiste, à chaque étape du processus d’apprentissage, à réduire de plus en plus drastiquement les poids correspondant aux fréquences les moins utiles, jusqu’à arriver à la conservation d’un nombre K choisi de fréquences, associées aux poids de Θ de plus forte magnitude lors de l’entraînement. Cette étape est suivie d’un ré-entraînement du modèle réduit [9]. Pour le PMC, l’élagage réduit les paramètres à $\mathcal{O}(hK + h^2d + hC)$, contre $\mathcal{O}(\gamma K + 2d\gamma^2 K + \gamma C)$ pour le ResNet spectral.

TABLE 1 – Hyperparamètres optimaux et performances obtenues par modèle. ResNet S (ou ResNet F) fait référence au ResNet spectral utilisant un graphe structurel (ou fonctionnel).

Modèle	Profondeur	h/γ	Params.	IBC	HCP
PMC	3	10	2.5M	61.6 ± 0.5	97.4 ± 0.03
ResNet F	4	7	47M	66.3 ± 0.5	96.9 ± 0.06
ResNet S	4	7	47M	55.1 ± 0.5	97.0 ± 0.08

3.4 Données et graphes de neuro-imagerie

Nous étudions deux jeux de données de signaux IRMF tirés de Neurovault [10] : IBC [11] (collection 6618), composé de 13 sujets ; et HCP, composé de 788 sujets du "Human Connectome Project" [12] (collection 4337). Comme étiquettes pour les signaux, nous avons utilisé les 24 et 7 tâches de IBC et HCP, respectivement. Celles-ci couvrent un large spectre de conditions expérimentales allant de la motricité à la mémoire. Les jeux de données ont été divisés en trois ensembles : entraînement, validation et test (70%, 15% et 15% pour HCP ; 8, 3 et 2 sujets pour IBC) afin d'entraîner nos modèles de manière standard.

Nous avons considéré deux graphes pour modéliser le support des signaux IRMF : 1) un graphe structurel estimé à partir de 56 sujets sains de HCP ; 2) un graphe fonctionnel, estimé à partir des données d'IRMF au repos de 1080 sujets sains de HCP. Les sommets des graphes correspondent aux 360 régions de l'atlas de Glasser [13], et les arêtes sont pondérées par une mesure de la connectivité structurelle [14] pour le graphe structurel, ou par la corrélation de Pearson des signaux IRMF au repos [4] pour le graphe fonctionnel. Enfin, pour chaque graphe, on conserve pour chaque sommet ses 8 plus proches voisins, que l'on binarise.

4 Expérimentations

4.1 Détails expérimentaux

Les hyperparamètres optimaux des modèles considérés, obtenus par recherche exhaustive sur une grille, sont fournis en Table 1. Nous avons appliqué une normalisation de lots à la sortie de chaque couche TSGConv, et des fonctions d'activation ReLU. Les données ont été centrées et normées avant l'apprentissage. Enfin, l'apprentissage utilise du mixup [15]. Nous rapportons les moyennes et les intervalles de confiance à 95% sur plusieurs entraînements.

Nous observons qu'aucun modèle n'est systématiquement meilleur que les autres en terme de performance brute. Toutefois, la version structurelle du ResNet spectral obtient de très mauvaises performances sur IBC par rapport aux autres modèles, ce qui suggère que le graphe structurel est moins adapté aux jeux de données considérés que son homologue fonctionnel.

4.2 Impact des fréquences des graphes

Pour établir une référence pour la méthodologie d'élagage, nous évaluons les performances des modèles proposés lorsque nous sélectionnons une bande de fréquences du graphe. Les Figures 2(a) et 2(b) montrent les parties du spectre qui semblent

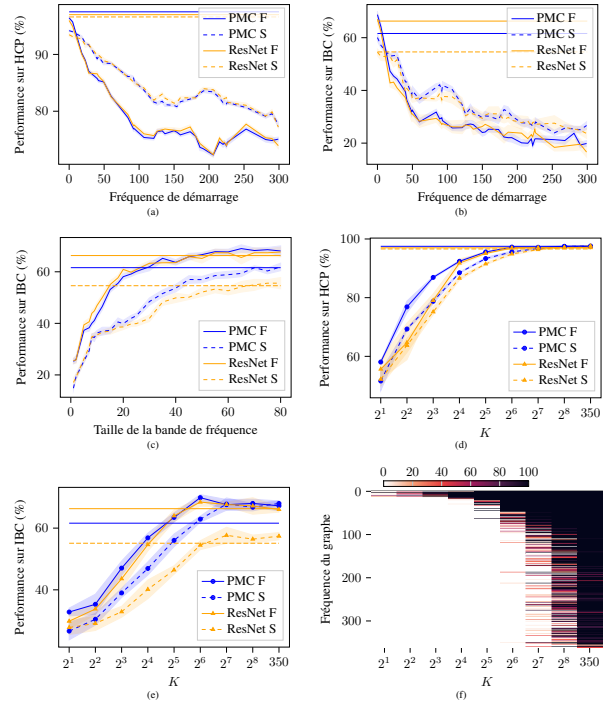


FIGURE 2 – (a)(b) Performance des modèles sur les jeux de données, en fonction du décalage d'une bande de 60 fréquences. (c) Performance des modèles sur IBC, en fonction de la largeur de la bande de fréquence, démarrant à la fréquence 0. (d)(e) Performance des modèles, en fonction du nombre K de fréquences conservées par élagage. Les lignes horizontales correspondent à la performance de référence en Table 1. (f) Nombre de fois où chaque fréquence du graphe fonctionnel est conservée par l'élagage de PMC F entraîné sur IBC, en fonction de K .

les plus utiles pour le décodage des signaux IRMF. Nous observons que, pour tout modèle et graphe, les meilleurs résultats sont obtenus en conservant les basses fréquences du graphe.

En Figure 2(c), nous faisons varier la largeur de bande lorsque le décalage est fixé à 0. Nous observons que le fait de conserver plus de 60 fréquences n'améliore pas la précision de manière significative. À nouveau, nous observons que le graphe fonctionnel donne de meilleurs résultats que le structurel. On note de plus que dans certains cas, conserver une partie des fréquences donne de meilleurs résultats que la référence.

4.3 Élagage des fréquences des graphes

Les figures 2(d) et 2(e) représentent la performance en fonction du nombre de fréquences conservées K . Comme en Figure 2(c), nous observons que la référence peut être surpassée avec un nombre limité de fréquences. À nouveau, le graphe structurel offre de moins bonnes performances, ce qui suggère, en plus d'une inadéquation de ses basses fréquences avec la tâche, une faible capacité de discrimination des données.

La Table 2 rapporte les meilleures performances sur les jeux de données, en utilisant soit la méthode de la bande de fréquences, soit celle de l'élagage. Nous observons que l'élagage

TABLE 2 – Performances obtenues par jeu de données en utilisant des bandes de fréquences ou un élagage des modèles.

Modèle	IBC		HCP	
	Bande	Élagage	Bande	Élagage
Taille / $K = 60$				
PMC F	68.7 ± 1.1	69.9 ± 1.0	96.4 ± 0.1	97.2 ± 0.2
PMC S	60.0 ± 1.5	62.9 ± 1.0	94.2 ± 0.1	95.6 ± 0.3
ResNet F	67.0 ± 1.7	68.4 ± 1.3	96.0 ± 0.2	96.9 ± 0.2
ResNet S	54.5 ± 1.4	54.7 ± 1.4	93.6 ± 0.2	94.9 ± 0.4

permet d’obtenir des améliorations constantes dans toutes nos expériences, alors que la sélection de bandes dégrade les performances dans la plupart des scénarios.

En Figure 2(f), nous représentons les fréquences conservées par élagage d’un PMC spectral entraîné sur IBC, pour différentes valeurs de K . Nous effectuons 20 entraînements et indiquons le nombre de fois où chaque fréquence a été conservée. Nous observons que les basses fréquences ont globalement été préférées. Il est à noter que ce ne sont toutefois pas strictement des bandes continues de fréquences, ce qui explique la différence de performance avec la méthode basée sur des bandes fixes. Nous avons calculé des mesures d’Intersection sur Union (IsU) pour quantifier la reproductibilité de ces résultats entre les jeux de données et les modèles. Nos scores sont généralement supérieurs à 50%, ce qui suggère une reproductibilité élevée.

En Figure 3, nous visualisons sur le cerveau les 4 vecteurs propres du PMC fonctionnel les plus fréquemment sélectionnés avec $K = 8$: $\mathbf{U}[:, 1]$, $\mathbf{U}[:, 3]$, $\mathbf{U}[:, 4]$, $\mathbf{U}[:, 9]$, associés aux valeurs propres $\Lambda[1, 1]$, $\Lambda[3, 3]$, $\Lambda[4, 4]$, $\Lambda[9, 9]$. Les distributions spatiales des plus grandes valeurs positives et négatives de $\mathbf{U}[:, 1]$ et $\mathbf{U}[:, 3]$ correspondent respectivement aux zones occipitales et temporales supérieures du cerveau, ce qui suggère des contributions des systèmes visuels et auditifs primaires. Les valeurs les plus positives de $\mathbf{U}[:, 4]$ sont confinées au cortex moteur, tandis que $\mathbf{U}[:, 9]$ inclut les contributions du réseau du mode par défaut. Ces résultats montrent que l’élagage permet d’identifier des fréquences significatives et interprétables pour le décodage de l’activité cérébrale. Notons que ces fréquences sont obtenues en entraînant nos modèles sur toutes les tâches disponibles, ce qui explique l’apparition de motifs assez génériques, communs à toutes les tâches [16]. Par la suite, nous souhaitons étudier le potentiel de la méthode introduite dans cet article pour distinguer des tâches individuelles ou des patients.

5 Conclusions

Nous avons introduit un modèle simple d’apprentissage profond basé sur les ResNets pour décoder les signaux IRMF en exploitant un graphe sous-jacent. Par une méthodologie d’élagage, nous avons sélectionné les fréquences des graphes les plus importantes, et observé que conserver un sous-ensemble réduit pouvait conduire à une amélioration des performances. Les fréquences sélectionnées semblent être cohérentes d’un modèle et d’un jeu de données à l’autre. Nous pensons que cette étude pourra aider à concevoir des réseaux profonds sur graphes plus efficaces et interprétables pour ce domaine d’application.

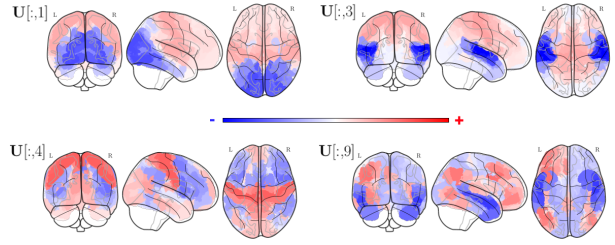


FIGURE 3 – Fréquences du PMC F les plus conservées.

Références

- [1] R. A. Poldrack, “Inferring mental states from neuroimaging data : From reverse inference to large-scale decoding,” *Neuron*, vol. 72, no. 5, pp. 692–697, 2011.
- [2] M. A. Schulz, B. T. Yeo, J. T. Vogelstein, J. Mourao-Miranada, J. N. Kather, K. Kording, B. Richards, and D. Bzdok, “Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets,” *Nature Communications*, vol. 11, no. 1, 2020.
- [3] M. Bontonou, G. Lioi, N. Farrugia, and V. Gripon, “Few-shot decoding of brain activation maps,” in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1326–1330.
- [4] Y. Zhang, L. Tetrel, B. Thirion, and P. Bellec, “Functional annotation of human cognitive states using deep graph convolution,” *NeuroImage*, vol. 231, p. 117847, 2021.
- [5] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, “Graph signal processing : Overview, challenges, and applications,” *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.
- [6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015.
- [7] S. Han, J. Pool, J. Tran, and W. Dally, “Learning both weights and connections for efficient neural network,” *Advances in neural information processing systems*, vol. 28, 2015.
- [8] H. Tessier, V. Gripon, M. Léonardon, M. Arzel, T. Hannagan, and D. Bertrand, “Rethinking weight decay for efficient neural network pruning,” *Journal of Imaging*, vol. 8, no. 3, 2022.
- [9] A. Renda, J. Frankle, and M. Carbin, “Comparing rewinding and fine-tuning in neural network pruning,” in *International Conference on Learning Representations*, 2019.
- [10] K. J. Gorgolewski, G. Varoquaux, G. Rivera, Y. Schwarz, S. S. Ghosh, C. Maumet, V. V. Sochat, T. E. Nichols, R. A. Poldrack, J.-B. Poline *et al.*, “Neurovault. org : a web-based repository for collecting and sharing unthresholded statistical maps of the human brain,” *Frontiers in neuroinformatics*, vol. 9, p. 8, 2015.
- [11] A. L. Pinho, A. Amadon, B. Gauthier, N. Clairis, A. Knops, S. Genon, E. Dohmatov, J. J. Torre, C. Ginisty, S. Becuwe-Desmidt *et al.*, “Individual brain charting dataset extension, second release of high-resolution fmri data for cognitive mapping,” *Scientific Data*, vol. 7, no. 1, pp. 1–16, 2020.
- [12] D. C. Van Essen, K. Ugurbil, E. Auerbach, D. Barch, T. E. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss *et al.*, “The human connectome project : a data acquisition perspective,” *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, 2012.
- [13] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. V. Essen, “A multi-modal parcellation of human cerebral cortex,” *Nature*, vol. 536, no. 7615, pp. 171–178, 2017.
- [14] M. G. Preti and D. Van De Ville, “Decoupling of brain function from structure reveals regional behavioral specialization in humans,” *Nature Communications*, vol. 10, no. 4747, pp. 1–7, 2019.
- [15] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup : Beyond empirical risk minimization,” 2018.
- [16] A. Mensch, J. Mairal, and B. Thirion, “Extracting representations of cognition across neuroimaging studies improves brain decoding,” *PLoS Computational Biology*, pp. 1–20, 2021.