PRUNING GRAPH CONVOLUTIONAL NETWORKS TO SELECT MEANINGFUL GRAPH FREQUENCIES FOR FMRI DECODING

Yassine El Ouahidi, Hugo Tessier, Giulia Lioi, Nicolas Farrugia, Bastien Pasdeloup and Vincent Gripon
IMT Atlantique
Lab-STICC, UMR CNRS 6285
F-29238 Brest, France
name.surname@imt-atlantique.fr

***Abstract*—Graph Signal Processing is a promising framework to manipulate brain signals as it allows to encompass the spatial dependencies between the activity in 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 fMRI signals. To this end, we introduce a deep learning architecture and adapt a pruning methodology to automatically identify such frequencies. We experiment with various datasets, architectures and graphs, and 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. We believe that this work provides novel insights on how graph-based methods can be deployed to increase fMRI decoding accuracy and interpretability.**

***Index Terms*—graph signal processing, residual networks, functional magnetic resonance imaging, neural networks pruning**

I. Introduction

The development of Functional Magnetic Resonance Imaging (fMRI) has allowed to observe the brain *in vivo* and to address one of the central questions of cognitive neuroscience: understanding the relation between brain activation and cognitive functions or experimental conditions. Classical approaches have relied on *forward inference* (*e.g.*, identifying localised effects in the brain corresponding to a change in the experimental condition) or *reverse inference* (*e.g.*, inferring the cognitive function from brain activation patterns, which is typically done using single subjects linear statistical models). More recently, it has been suggested that brain decoding (*e.g.*, *predicting* an experimental condition from patterns of brain activation) is a more formal way to perform reverse inference [1], as it allows to identify brain structures that are selectively engaged during a specific cognitive task. A few works have tested large-scale brain decoding using classic machine learning or deep learning paradigms applied to fMRI data from different subjects and tasks [2], [3]. Some of these studies have taken a step forward and tested graph convolutional networks to decode fMRI brain activity by also taking in account brain structural [4] or functional [5] connectivity, exploiting the promising framework of Graph Signal Processing (GSP).

GSP is a mathematical framework that aims at extending classical Fourier analysis to irregular domains represented by graphs. In particular, notions of graph frequencies, modes, and associated operators such as convolutions or filtering can then be defined. In the context of fMRI decoding, the use of GSP-based methods also brings forward interpretability questions.

In this work, we are interested in better understanding what are the graph frequencies that are the most useful to decode fMRI signals. To this end, we introduce a neural network architecture called *spectral ResNet* in which graph frequencies are used to define convolutions. Using a pruning technique, we dynamically identify which frequencies are the most useful to decode fMRI signals. We show that these frequencies are robust to changes in the datasets and the considered architectures. We perform experiments with both structural or functional graphs.

II. Related Work

In this document, tensors and matrices are noted in bold uppercase, and vectors in bold lowercase. We access their entries using Python-style indexing, where : denotes all entries along the corresponding dimension. Constants are noted in uppercase, variables in lowercase, and sets in calligraphic.

GSP generalizes Fourier's approach to signals evolving on irregular structures by providing an adapted spectral space to decompose them in meaningful frequencies [6], [7]. In this framework, we consider a weighted and undirected graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, W \rangle$ with vertices $\mathcal{V} = \{v_1, \dots, v_N\}$ of cardinal $|\mathcal{V}| = N$, edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, and a weighting function $W : \mathcal{E} \mapsto \mathbb{R}$. Such a graph can be equivalently represented by its weights matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ such that $\mathbf{W}[i,j] = W(\{v_i, v_j\})$ if $\{v_i, v_j\} \in \mathcal{E}$ and 0 otherwise. Additionally, we note $\mathbf{D} \in \mathbb{R}^{N \times N}$ the degrees matrix of $\mathcal{G}$, such that $\mathbf{D}[i,j] = \sum_{k=1}^{N} \mathbf{W}[i,k]$ if $i = j$ and 0 otherwise. From these two matrices, we can compute the normalized Laplacian matrix $\mathbf{L} = \mathbf{I}_N - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ of $\mathcal{G}$, where $\mathbf{I}_N$ is the identity matrix of dimension N. Since $\mathbf{L}$ is real and symmetric, it can be diagonalized as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U}$ is a matrix of orthonormal vectors associated with eigenvalues forming the diagonal matrix $\mathbf{\Lambda}$, sorted in increasing order. These

eigenvalues are analogous to frequencies in Fourier analysis, and are called *graph frequencies*.

A signal $\mathbf{x} \in \mathbb{R}^N$ on $\mathcal{G}$ is an observation on each of its vertices. Its Graph Fourier Transform $\hat{\mathbf{x}} = \mathrm{GFT}(\mathbf{x}) = \mathbf{U}^\top \mathbf{x}$ shows the various contributions of eigenvalues of $\mathbf{L}$ in $\mathbf{x}$. Its inverse $\mathbf{x} = \mathrm{GFT}^{-1}(\hat{\mathbf{x}}) = \mathbf{U}\hat{\mathbf{x}}$ transforms a graph spectrum into a graph signal. Following the usual setting of Fourier Analysis, convolution can be defined as entrywise multiplication $\odot$ in the spectral space. In more details, denote $\mathbf{h} \in \mathbb{R}^N$ a filter, then convolution is obtained as:

$$\mathbf{x} * \mathbf{h} = \mathrm{GFT}^{-1}\left(\mathrm{GFT}(\mathbf{x}) \odot \mathrm{GFT}(\mathbf{h})\right) . \quad (1)$$

There are several standard filters in the literature to perform useful frequency selection (low/high/band-pass and band-stop) as well as several methods to select them [7]. These methods are based on the selection of one or more frequency bands at specific locations in the spectrum. In contrast to these methods, we propose here a method where selection of (non-necessarily contiguous) frequencies is performed automatically while optimizing for a supervised task.

In previous works applying GSP to decode brain activity [4], [5], the focus has been mainly on increasing generalization accuracy. Here we aim at investigating which graph frequencies are the most relevant for brain decoding, and compare results for functional and structural graphs. We also assess the reliability of our results by applying the same analysis to two large scale datasets and testing different deep learning architectures. More specifically we address the following questions:

1) Which graph frequencies are the most relevant to predict the experimental task from corresponding fMRI activation maps?
2) Are these frequencies consistent for different underlying graphs (functional vs. structural) and deep learning architectures?
3) Does selecting specific graph frequency bands increase brain decoding performance?

III. Methodology

In this work, we aim at exploiting the graph on which input signals are defined to improve the classification of these signals and its interpretability in terms of relevant brain spatial patterns. The core idea of the proposed approach is to select by a learning process the appropriate graph frequencies within the signals. Beyond the performance aspect, this also provides an understandable representation of the signal features that are used for training a model. To reach this goal, we build a spectral model around a GSPConv layer, defined below.

A. GSPConv layer

A GSPConv layer is defined using a graph $\mathcal{G}$ with N vertices. It takes as input a matrix $\mathbf{X} \in \mathbb{R}^{N \times c_{\mathrm{in}}}$ seen as a graph signal with c_{in} channels and outputs a matrix $\mathbf{Z} \in \mathbb{R}^{N \times c_{\mathrm{out}}}$ seen as a graph signal with c_{out} channels. It contains a weight tensor $\mathbf{\Theta} \in \mathbb{R}^{N \times c_{\mathrm{in}} \times c_{\mathrm{out}}}$ and its mathematical function is:

$$\mathrm{GSPConv}_{c_{\mathrm{in}}, c_{\mathrm{out}}}(\mathbf{X}) = \sigma\left(\mathrm{GFT}^{-1}\left(\mathrm{GFT}\left(\mathbf{X}\right) \otimes \mathbf{\Theta}\right)\right) , \quad (2)$$

where GFT and GFT^{-1} are applied to each channel independently, σ is a nonlinear activation function, and $\otimes$ is such that $(\hat{\mathbf{x}} \otimes \mathbf{\Theta})[l, c] = \sum_{c'=1}^{c_{\mathrm{in}}} \hat{\mathbf{x}}[l, c'] \mathbf{\Theta}[l, c', c]$.

In other words, a GSPConv is a natural extension of usual convolutional layers used in regular 1d or 2d domains.

B. Considered architectures

Using the GSPConv layer, we build a spectral ResNet, inspired by regular ResNets [8]. Such a model consists in assembling a first GSPConv layer, then spectral ResNet blocks (c.f. Figure 1), and finally a fully connected layer applied on globally averaged values over the graph. This architecture is summarized in Figure 1. The main originality and interest of this architecture is its ability to give importance to all the frequencies without filtering them explicitly while doing convolutions, which is the case with standard Graph Convolutional architectures such as Simple Graph Convolution Neural Networks [9].

It is parametrized by a depth d, and a number of channels γ. It contains a total number of trainable parameters $\mathcal{O}(\gamma N + 2d\gamma^2 N + \gamma C)$ where C is the number of classes in the considered problem.

For comparison purposes, we also propose an architecture based on a standard Multi-Layer Perceptron (MLP) which input domain is the graph frequency domain. Such an architecture exploits the structure of graphs in a very different way compared to the previous one and is used for stressing the genericity of our results. For simplicity, we use a fixed number h of hidden neurons in each layer and a depth d. Such a model contains a total of $\mathcal{O}(hN + h^2 d + hC)$ trainable parameters.

C. Using pruning to identify important graph frequencies

In the field of classification, pruning [10] is a very active topic of research that aims at identifying unimportant portions of deep learning architectures and remove them while maintaining a steady level of accuracy. The motivation is usually to reduce the number of parameters, hence the memory usage and the number of computations, resulting in lightweight models that can better fit constrained environments (*e.g.*, edge computing). In this work, our aim is to adapt pruning in order to automatically identify the least important graph frequencies in our models, then prune them.

A pruning method usually combines multiple ingredients: 1- a pruning criterion which associates a subset of parameters in the considered architecture with an importance score, 2- a pruning technique which is typically deployed onto the parameters with the lowest importance score and 3- a pruning schedule which defines how the pruning ingredients interact with the global learning schedule of the considered architecture. We choose to implement Selective Weight Decay (SWD) [11] because it has the asset of allowing any subset of frequencies to be removed at any time during the training procedure, avoiding harsh sudden effects of other methods, while being state-of-the-art in competitive vision benchmarks.

In short, SWD consists of the following: at each step of the learning process, we aggregate all weights associated with a

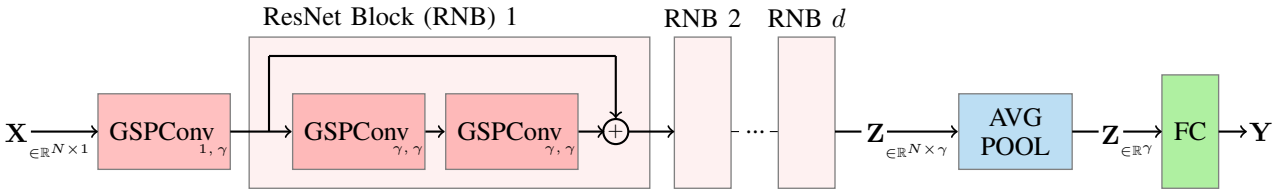

Fig. 1. Illustration of the Spectral ResNet model introduced in this paper. The input is seen as a graph signal with 1 channel, it is first embedded as a graph signal with γ channels using a GSPConv layer. Then it goes through d ResNet blocks consisting of the sum between a shortcut path and a sequential path containing 2 GSPConv layers. The output is seen as a graph signal with γ channels. The values are averaged over the graph before being fed to a classical logistic regression.

particular graph frequency l into a linearized vector. For MLP, only the first layer of the architecture is concerned, whereas for Spectral ResNets, we combine all layers weight tensors $\mathbf{\Theta}[l, \cdot, \cdot]$. We then compute the magnitude of obtained vectors, and sort them in descending order. The K first vectors of weights are left untouched, and the remaining ones are applied an additional weight decay. This additional weight decay grows exponentially during the learning procedure, starting at a very low value $\alpha_{\min}$ so that the effect of SWD is almost unnoticeable, and finishing at a value $\alpha_{\max}$ so large that it boils down to nullifying the considered values. Once the procedure is finished, we identify the least important graph frequencies and remove them from the model, then we retrain the remaining parameters using the same scheduler as for baseline architectures, what is referred to as LR-rewinding [12] in the literature.

Note that pruning graph frequencies has a very different impact on the total number of parameters of considered models. For MLP, it reduces the parameters to $\mathcal{O}(hK + h^2d + hC)$ whereas for Spectral ResNets, it reduces to $\mathcal{O}(\gamma K + 2d\gamma^2 K + \gamma C)$, resulting in a way more dramatic effect.

D. Neuroimaging Data and Graphs

Two datasets of spatial brain activation maps (obtained using general linear models of functional MRI signals [13]) were fetched from Neurovault [14]: the second release of IBC [15] (Neurovault collection 6618), consisting of 13 subjects with many experimental tasks, and 788 subjects out of the 900 subjects release of the Human Connectome Project [16] (Neurovault collection 4337). As labels for classification, we used the 24 tasks for IBC, and the 7 tasks for HCP. Here, we consider a large spectrum of experimental conditions ranging from motor to memory and social tasks. The datasets were split in training, validation and test splits. Following the usual supervised learning pipeline, training data was used to tune model parameters, validation data to monitor generalisation accuracy without parameter tuning, and the final accuracy reported is measured on the test set, using the model with the best validation accuracy. For HCP, 70 % of subjects were used for training, 15 % for validation and 15 % for test. For IBC eight subjects were used for training, three for validation and two for test. The obtained splits were all balanced.

In this work, we considered two brain graphs to build spectral filters. The first is an average structural graph estimated from diffusion weighted images of the HCP dataset (56 healthy subjects). Graph nodes correspond to the 360 regions of the Glasser atlas [17] while graph edges are a measure of structural connectivity strength, as described in [18]. We also considered another consensus graph, estimated from resting state fMRI data of 1080 HCP healthy subjects. Also, for the functional graph, the nodes correspond to the 360 regions of the Glasser atlas, while edges are a measure of the Pearson correlation between fMRI time-series during resting state [5]. In order to ensure that the brain graphs had good spectral properties while remaining connected, binarized k-nearest-neighbor graphs were built from the original structural and functional graphs by connecting each node to its $k = 8$ neighbors with strongest connectivity. All the considered models and datasets are available at our Github[1].

TABLE I
HYPERPARAMETERS AND ASSOCIATED PERFORMANCE FOR OUR BASELINE ARCHITECTURES. RESNET S REFERS TO THE STRUCTURAL GRAPH AND RESNET F TO THE FUNCTIONAL GRAPH.

model	depth	h/γ	parameters	IBC	HCP
MLP	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

IV. EXPERIMENTS

We conducted a set of experiments to answer the series of questions established in Section II.

A. Experimental details

As a first step, we ran numerous experiments to find good hyperparameters for our considered architectures. The optimal hyperparameters are displayed in Table I. Note that ResNets contain way more parameters than their MLP counterpart, but as mentioned in Section III, the number of these parameters dramatically reduces as we restrain the number of considered graph frequencies. We used batch normalization at the output of each GSPConv layer and rectified linear units as activation functions. We standardized the datasets (so that data is centered and unit-normed) before training and used mixup [19] during training. We report the averages and 95% confidence intervals over multiple runs.

[1]https://github.com/elouayas/gspconv

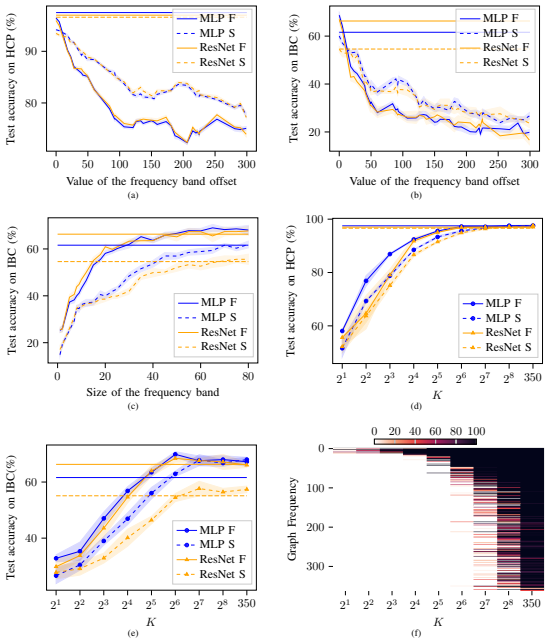

Fig. 2. (a)(b) Evolution of the accuracy of various architectures/datasets as a function of the graph frequency band offset with a bandwith of 60 frequencies. (c) Evolution of the accuracy of various architectures on the IBC dataset as a function of the graph frequency bandwidth, when the offset is 0. There are no improvements measured above 80 graph frequencies inside the band (similar results for the HCP dataset). (d)(e) Evolution of the accuracy of various architectures/datasets as a function of the number K of kept frequencies obtained through our pruning methodology. The horizontal lines correspond to the baseline accuracy. Shaded areas correspond to 95% confidence intervals. (f) Occurrences of each graph frequency for a fixed K with the MLP Functional on the IBC dataset.

We observe that there is no model that is universally better than the others in terms of raw performance. Yet the structural version of ResNet performs very poorly on IBC compared to the alternatives, suggesting the structural graph is less adapted to the considered datasets than its functional counterpart.

B. Impact of graph frequencies

To set up a baseline for the pruning methodology, we first evaluate the performance of the proposed architectures when we apply a band selection of the graph frequencies. To this end, we first identify which parts of the spectrum seem the most useful for decoding fMRI signals. The results are depicted in Figures 2(a) and 2(b). We fixed a bandwidth of 60 frequencies and varied the offset, which corresponds to the x axis. We observe that no matter the architecture nor the graph, best results are obtained by keeping the low graph frequencies. Such similar patterns within both graphs can be explained by previous studies [20].

In Figure 2(c), we vary the bandwidth when the offset is set to 0. We observe that keeping more than 60 frequencies does not benefit accuracy by a large margin. We again observe that the functional graph leads to better results overall than the structural one. Interestingly, in some cases keeping a portion of the frequencies leads to better results than the baselines.

TABLE II
BEST GAINS OBTAINED WHEN USING PREDETERMINED FREQUENCY BANDS OR THE PRUNING STRATEGY ON THE CONSIDERED ARCHITECTURES AND DATASETS.

model	**IBC**		**HCP**	
Size / K = 60	Band	Pruning	Band	Pruning
MLP F	68.7±1.1	**69.9±1.0**	96.4±0.1	**97.2±0.2**
MLP 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**

C. Automated selection of optimal graph frequencies

We now test the proposed pruning methodology. Figures 2(d) and 2(e) depict accuracy as a function of the number of kept frequencies K. Similarly to Figure 2(c), we observe that the baseline cannot be outperformed with a limited number of selected frequencies. Again, the structural graph leads to poorer performance, suggesting that it is not only its lower spectrum that is not aligned with the task but that overall it offers poorer discrimination capabilities on our benchmarks.

Table II reports the best achieved gains with respect to our baselines, when using either the band or pruning method. We observe that pruning can consistently lead to improvements over all our experiments, whereas band selection degrades the performance in most considered scenarios. This table highlights the benefits of using the pruning methodology.

In Figure 2(f), we depict the frequencies selected using the pruning methodology, for various values of K, in the case of the spectral ResNets on the functional graph and IBC dataset. We perform 20 runs, and report the number of times each frequency was selected. Interestingly, we observe that the lower spectrum is overall preferred, even though it is not exactly a band of frequencies that is selected, explaining for the potential benefits over the predetermined frequency bands. We computed Intersection over Union (IoU) measures to quantify how reproducible these results are across datasets and architectures. Our IoU scores are typically above 50%, suggesting high reproducibility.

In Figure 3, we visualize on the glass brain the most frequently selected eigenvectors of the functional graph. These 4 eigenvectors – $\mathbf{U}[:,1]$, $\mathbf{U}[:,3]$, $\mathbf{U}[:,4]$, $\mathbf{U}[:,9]$ associated with eigenvalues $\mathbf{\Lambda}[1,1]$, $\mathbf{\Lambda}[3,3]$, $\mathbf{\Lambda}[4,4]$, $\mathbf{\Lambda}[9,9]$ respectively – are among those that appeared the most frequently when using pruning with $K = 8$, under all configurations of models and datasets. Spatial distributions of largest positive and negative values of $\mathbf{U}[:,1]$ and $\mathbf{U}[:,3]$ correspond respectively to occipital and superior temporal brain areas, suggesting contributions of primary visual and auditory systems. $\mathbf{U}[:,4]$ most positive values are confined to the motor cortex, while $\mathbf{U}[:,9]$ includes medial contributions of the default mode network [21]. These results show how the proposed methodology is able to identify meaningful graph frequencies (interpreted here as spatial patterns) for brain activity decoding. As we consider the graph frequencies common to all the tasks combined, general purpose brain patterns were found, consistent with the literature [22]. In future work we will investigate the potential

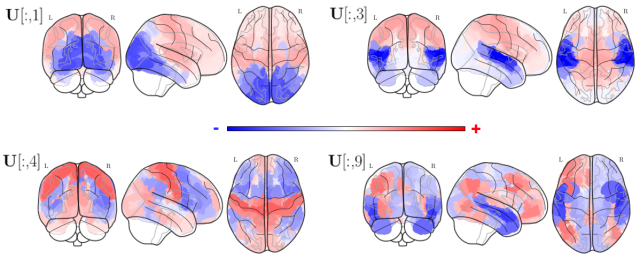

Fig. 3. Visualization of most frequently selected graph eigenvectors of the functional graph on the glass brain.

of this methodology to identify spatial patterns characteristics of a specific decoded task or of a patient profile.

V. Conclusions

In this paper, we have introduced a simple deep learning architecture based on ResNets to process graph-based signals and applied it to fMRI decoding tasks. Using a pruning methodology, we selected the most important graph frequencies and observed that keeping 60 out of the 360 initial frequencies could lead to improved performance. Interestingly, the selected frequencies seem to be reproducible with other architectures and datasets. We believe that this study could help designing more efficient and interpretable graph neural networks for this important domain of application.

VI. Acknowledgment

We would like to thank Yu Zhang for providing the consensus functional graph. We also thank the region of Brittany for its support.

References

[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] G. Varoquaux, Y. Schwartz, R. A. Poldrack, B. Gauthier, D. Bzdok, J.-B. Poline, and B. Thirion, "Atlases of cognition with large-scale human brain mapping," *PLoS computational biology*, vol. 14, no. 11, p. e1006565, 2018.

[3] M.-A. Schulz, B. 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, pp. 1–15, 2020.

[4] M. Bontonou, G. Lioi, N. Farrugia, and V. Gripon, "Few-shot decoding of brain activation maps," in *IEEE conference European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1326–1330.

[5] 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.

[6] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE signal processing magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[7] 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.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[9] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *International conference on machine learning*. PMLR, 2019, pp. 6861–6871.

[10] 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.

[11] 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, p. 64, 2022.

[12] A. Renda, J. Frankle, and M. Carbin, "Comparing rewinding and fine-tuning in neural network pruning," *arXiv preprint arXiv:2003.02389*, 2020.

[13] W. D. Penny, K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[14] 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.

[15] A. L. Pinho, A. Amadon, B. Gauthier, N. Clairis, A. Knops, S. Genon, E. Dohmatob, 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.

[16] 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.

[17] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

[18] 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. 1, pp. 1–7, 2019.

[19] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[20] A. Messé, G. Marrelec, P. Bellec, V. Perlbarg, J. Doyon, M. Pélégrini-Issac, and H. Benali, "Comparing structural and functional graph theory features in the human brain using multimodal mri," *Irbm*, vol. 33, no. 4, pp. 244–253, 2012.

[21] J. Smallwood, B. C. Bernhardt, R. Leech, D. Bzdok, E. Jefferies, and D. S. Margulies, "The default mode network in cognition: a topographical perspective," *Nature reviews neuroscience*, vol. 22, no. 8, pp. 503–513, 2021.

[22] A. Mensch, J. Mairal, B. Thirion, and G. Varoquaux, "Extracting representations of cognition across neuroimaging studies improves brain decoding," *PLoS computational biology*, vol. 17, no. 5, p. e1008795, 2021.