EEG to fMRI Synthesis Benefits from Attentional Graphs of Electrode Relationships

David Calhas
DAVID.CALHAS@TECNICO.ULISBOA.PT
INESC-ID and Instituto Superior Tecnico, Lisboa, Portugal

Rui Henriques
INESC-ID and Instituto Superior Tecnico, Lisboa, Portugal

Abstract

Topographical structures represent connections between entities and provide a comprehensive design of complex systems. Currently these structures are used to discover correlates of neuronal and haemodynamical activity. In this work, we incorporate them with neural processing techniques to perform regression, using electrophysiological activity to retrieve haemodynamics. To this end, we use Fourier features, attention mechanisms, shared space between modalities and incorporation of style in the latent representation. By combining these techniques, we propose several models that significantly outperform current state-of-the-art of this task in resting state and task-based recording settings. Additionally, we show how the developed mapping functions are able to extrapolate to a diagnostic setting. We report which EEG electrodes are the most relevant for the regression task and which relations impacted it the most. Complementary, we observe that haemodynamic activity at the scalp, in contrast with sub-cortical regions, is relevant to the learned shared space. Overall, these results suggest that EEG electrode relationships are pivotal to retain information necessary for haemodynamical activity retrieval.

1. Introduction

Human brain activity upholds cognitive and memory functions. *Neuronal* activity, a set of action potentials localized at the synapses of a neuron Sherrington (1952), can be retrieved through electroencephalography (EEG), while *haemodynamics*, linked to the blood supply Buckner (1998), measured via functional magnetic resonance imaging (fMRI). Electrophysiological and haemodynamical activity have been widely studied, with several key discoveries on their relationships Shibasaki (2008); Yu et al. (2016); He et al. (2018); Rojas et al. (2018); Bréchet et al. (2019); Daly et al. (2019); Cury et al. (2020); Abreu et al. (2021). Though, there is still a research gap in predicting one modality from the other, a problem commonly formulated as a *regression* task also known as cross-mode mapping or synthesis. The work conducted by Liu et al. (2019) places the state-of-the-art in the synthesis of fMRI from EEG. As simultaneous EEG and fMRI recordings become publicly available Deligianni et al. (2014); Walz et al. (2015); Pereira et al. (2020), the neuroscience and machine learning communities can unprecedentedly address these tasks. A *synthesized fMRI modality*, sourced only from EEG, promote ambulatory diagnoses, longitudinal monitoring of individuals, and the understanding of synergistic electrophysiological and haemodynamical activity. Given the EEG reduced recording costs de Beeck et al. (2019), the target synthesis task can further ensure resource availability to communities in need Ogbole et al.

(2018), providing a proxy to fMRI screens with significant impact on the quality of life van Beek et al. (2019).

Here, we push the ability of neural processing techniques to perform EEG to fMRI synthesis, and assess what links these two modalities with explainability methods. The gathered results show statistically significant improvements on the target synthesis task against state-of-the-art. The observed breakthroughs are driven by four major and novel contributions:

- conditioning of the fMRI synthesis task with an attention graph that explicitly captures relationships between electrodes;
- convolution layering and Fourier feature projection from EEG-based spectrograms and blood oxygen level dependent (BOLD) signals;
- shared latent space between modalities under dedicated losses to aid the training;
- incorporation of latent styling principles Gu et al. (2021) of the attention scores via a *style* posterior on the latent space features;
- extrapolation of the synthesized fMRI to a classification setting, achieved by the shift invariant trait of Fourier features and performing separation of the data at the sinusoid.

The main findings of this work are:

- occipital and parietal relationships with frontal EEG electrodes are the most relevant for haemodynamical retrieval in resting state and temporal electrodes only had relevant relations in task based fMRI synthesis (see Section 6). These long topographical links promote Markovian/locality properties to EEG representations (see Section 3.2) and are in accordance with the phenomena of frequencies of the same source being observed in distant electrodes da Silva (2013);
- *style* aid in latent representations, a widely used technique in computer vision research Gu et al. (2021), is a natural fit when conditioned on these relationships (posterior information), inherently providing robustness for the fMRI synthesis (see Section 5.2);
- results support the claim by Laufs et al. (2003); Deligianni et al. (2014) and Walz et al. (2015) on how neuronal activity frequency correlates with fMRI, and further highlight the relevance of spectral features to retrieve haemodynamics (see Section 5.1);

Altogether, these contributions and findings provide a solid ground EEG to fMRI synthesis framework. To the best of our knowledge, this is the first study to synthesize fMRI using EEG, with real simultaneous EEG and fMRI datasets, whereas the state-of-the-art is applied on synthetic data.

1.1. Generalizable Insights about Machine Learning in the Context of Healthcare

EEG to fMRI mappings represent a notable advance for the neuroscience community to acquire an integrative understanding of the brain's electrophysiology and haemodynamics. Amongst the unlocked healthcare potentials, fMRI synthesis from EEG signals is a pivotal

generative artificial intelligence task that can support ambulatory diagnostics, reduced neuroimaging costs, and augmented vision/interpretability for aided clinical decisions. Despite the relevance of cross-modality mappings, their applicability in healthcare settings is still largely untapped. The main motivation of this work is to bring machine learning techniques, inept at tackling problems of high dimensionality in limited observations settings, to health care systems and provide affordable solutions to all.

2. Problem formulation

Let $\boldsymbol{x} \in \mathbb{R}^{C \times F \times T}$ be an encoding of an EEG recording, where $\boldsymbol{x}_i \in \mathbb{R}^{F \times T}$ defines the spectral features of the i^{th} electrode, and F and T correspond to the frequency and temporal dimensions, respectively. Let $\boldsymbol{y} \in \mathbb{R}^{V_x \times V_y \times V_z}$ be an fMRI volume representation at a given time, where V_x, V_y and V_z correspond to the dimensions across the three dimensional referential axes. Given an arbitrary transformation $f : \mathbb{R}^{C \times F \times T} \rightarrow \mathbb{R}^{V_x \times V_y \times V_z}$, the learning objective becomes learning the multi-output regression model, such that $\text{argmin}_\theta ||f(\boldsymbol{x} \mid \theta) - \boldsymbol{y}||_1^1$.

3. Methods

Mathematical operations, such as addition and subtraction, are performed over the EEG and fMRI feature representations, $\boldsymbol{x}$ and $\boldsymbol{y}$, to map the original spaces onto encoded spaces that are identical in structure, $\boldsymbol{z}_x$ and $\boldsymbol{z}_y$, respectively. This is performed in accordance with the methodology described in Appendix 9. To this end, architecture modules of Resnet-18 blocks are set up using Calhas et al. (2022) framework, which hyperparameterizes kernel and stride sizes, potentially differing from layer to layer. The heterogeneity of this convolutional layering structure is beneficial for the performance of the model according to Riad (2022). Further, the algorithmic nature of this framework benefits our methodology as it distances itself from domain biased assumptions. Following, $\boldsymbol{z}_x$ is processed by a densely connected layer with a linear activation to perform the decoding from the encoded representation onto the fMRI volume. The architecture is described by three components: E_x, E_y and D_y. These map $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}_x$, respectively, to $\boldsymbol{z_x}$, $\boldsymbol{z}_y$ and $\hat{y}$. The gradients for these components are $\nabla_{\theta_{E_x}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) + \Omega(\boldsymbol{z}_x, \boldsymbol{z}_y)$, $\nabla_{\theta_{E_y}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) + \Omega(\boldsymbol{z}_x, \boldsymbol{z}_y)$ and $\nabla_{\theta_{D_y}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})$, respectively. The parameters, θ, minimize both $\mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) = ||\hat{\boldsymbol{y}} - \boldsymbol{y}||_1^1$ and $\Omega(\boldsymbol{z}_x, \boldsymbol{z}_y) = \cos(\boldsymbol{z}_x, \boldsymbol{z}_y)$. Figure 1 illustrates the described neural processing pipeline. For the sake of simplicity, please consider $\theta = \theta_{E_y} \bigcup \theta_{E_x} \bigcup \theta_{D_y}$.

3.1. Latent EEG Fourier features

Fourier feature extraction Rahimi et al. (2007) is applied, in addition to spectral analysis, to address the representational dissimilarity gap between EEG and fMRI features spaces. The underlying contributions in computer vision Rahimi et al. (2007); Tancik et al. (2020); Gu et al. (2021) have shown the ability of Fourier features to capture functions with high resolutions. The high degrees of freedom of such an operation Li et al. (2019) support the synthesis of an fMRI representation with rich spatial resolution, as well as the nature of the sum of sinusoids being ingrained in the representation of natural images.

Fourier features are applied, as a densely connected layer, to the latent representation $\boldsymbol{z}_x \in \mathbb{R}^L$ according to

$$\cos(\omega \cdot \boldsymbol{z}_x + b), \tag{1}$$

where $\omega \sim \mathcal{N}(0,1)$ and $b \sim \mathcal{U}(0, 2\pi)$. This layer is initialized with L independently drawn samples from $\mathcal{N}(0,1)$ and $\mathcal{U}(0,2\pi)$, such that $[\omega_1, \ldots, \omega_L]$ and $[b_1, \ldots, b_L]$ form the projections for

$$\boldsymbol{z}_x^* = \sqrt{\frac{2}{L}} \begin{bmatrix} \cos(\omega_1 \cdot \boldsymbol{z}_x + b_1) & \ldots & \cos(\omega_L \cdot \boldsymbol{z}_x + b_L) \end{bmatrix}, \tag{2}$$

defining the random Fourier features.

3.2. Topographical attention

The EEG electrodes are placed on the human scalp in accordance to a certain system, e.g. 10-20 system Jasper (1958). It is known that frequencies from the same neuronal source can be present in distant electrodes da Silva (2013) producing an EEG representation without locality, which makes them distant from natural images, a.k.a. Markovian images. This type of schema/relationship between electrodes is not able to be encoded in an Euclidean space, with topographical structures, such as graphs, usually being the go to approach. Since the layers of the encoder, E_x, are convolutional layers, relying on the Markovian property of its inputs others (2012), one needs to promote locality with a reordering operation. To that end, we propose the use of an attention mechanism at the level of the EEG electrodes. Banville et al. (2022) goes further and claims such a mechanism is robust against ill defined EEG electrodes. Let $A \in \mathbb{R}^{C \times F \times T}$ be the attention weights, such that $\forall i \in \{1, \ldots, C\} : A_i \in \mathbb{R}^{F \times T}$, and $E \in \mathbb{R}^{C \times C}$ the context matrix where each column $i \in \{1, \ldots, C\} : e_i^\top = \begin{bmatrix} e_{i1} & \ldots & e_{iC} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_i^\top \cdot A_1 & \ldots & \boldsymbol{x}_i^\top \cdot A_C \end{bmatrix}$. The attention scores,

$$W = \begin{bmatrix} w_{11} & \ldots & w_{1C} \\ \vdots & \ddots & \vdots \\ w_{C1} & \ldots & w_{CC} \end{bmatrix}, \tag{3}$$

where $w_i^\top = \begin{bmatrix} \frac{\exp(e_1)}{\sum_j \exp(e_j)} & \ldots & \frac{\exp(e_C)}{\sum_j \exp(e_j)} \end{bmatrix}$ are used to produce the output $T \in \mathbb{R}^{C \times F \times T}$, such that

$$T_i = \sum^{C} \boldsymbol{x} \odot w_i. \tag{4}$$

The proposed mechanism reorders electrode features to allow locality and performs a element wise (Hadamard) product, denoted by $\odot$, on the electrode dimension, C. The topologically corrected representation, T, enables locality properties in the channel dimension through gradient propagation from convolutional blocks.

Adding style conditioned on attention scores. According to latent style principles, as done by Gu et al. (2021), the attention scores, W, are used to add *style* features to the latent representation, z_x^*. This is done by placing a *style* posterior of the form $z_w | W, \boldsymbol{x} : z_w = B^\top \cdot W$, as

$$\boldsymbol{z}_x^* \odot z_w, \tag{5}$$

with $B \in \mathbb{R}^{C \times C \times L}$.

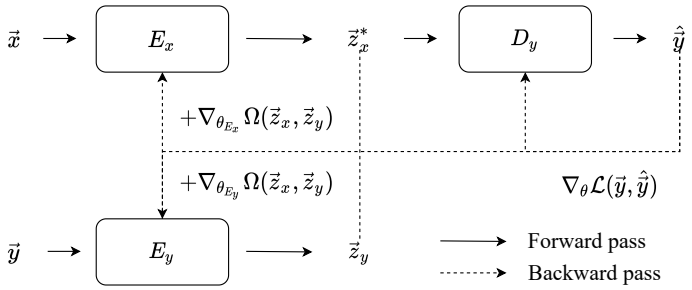


Figure 1: Architectural basis of the EEG to fMRI synthesis with the disclosure of the major forward and backward passes. The fMRI representation, $\boldsymbol{y}$, is processed by the encoder, E_y, such that $\boldsymbol{z}_y = E_y(\boldsymbol{y}, \theta_{E_y})$. The EEG representation, $\boldsymbol{x}$, in its turn is processed by the encoder, E_x, such that $\boldsymbol{z}_x = E_x(\boldsymbol{x}, \theta_{E_x})$. The latent EEG representation, $\boldsymbol{z}_x$, is processed by the decoder, D_y, that performs the mapping to the fMRI estimated instance, $\hat{\boldsymbol{y}} = D_y(\boldsymbol{x}, \theta_{D_y})$. The gradients for E_x, E_y and D_y are $\nabla_{\theta_{E_x}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) + \Omega(\boldsymbol{z}_x, \boldsymbol{z}_y)$, $\nabla_{\theta_{E_y}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}}) + \Omega(\boldsymbol{z}_x, \boldsymbol{z}_y)$ and $\nabla_{\theta_{D_y}} \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})$, respectively.

4. Experimental setting

The mean absolute error (MAE) is used as the loss function, $\mathcal{L}$, and the cosine distance between the latent representations of the EEG and fMRI was added as a regularization term, Ω. MAE is known to be robust under limited data settings and the regularization term on latent representations forces the network to approximate the target representation at the earlier layers Tran et al. (2018). The regularization constant was set to 1 and not optimized for the sake of proof of concept. Please recall the gradient computation illustrated in Figure 1.

4.1. Bayesian optimization

The hyperparameters are optimized in two phases:

1. using the fMRI autoencoder, $D_y(E_y(.; \theta_{E_y}); \theta_{D_y})$, to discover the optimal dimension of the latent space, $L : E_y(.; \theta_{E_y}) \in \mathbb{R}^{K \times K \times K}$;[1]
2. using the complete neural flow, illustrated in Figure 1, that maps EEG to fMRI.

In both phases, Bayesian Optimization Snoek et al. (2012) is applied with a total of 100 iterations. The hyperparameters subject to optimization, with their respective domains, are:

- learning rate $\in [1e-10, 1e-2]$;
- weight decay $\in [1e-10, 1e-1]$;
- filter size $\in \{2, 4\}$;
- max pooling layers (after each convolutional layer) $\in \{0, 1\}$;

1. Recall that the latent dimension is defined as both $\mathbb{R}^{K \times K \times K} = \mathbb{R}^L$, since $L = K \times K \times K$.

- batch normalization layers (after each max pooling/convoutional layer) $\in \{0, 1\}$;
- skip layers (in Resnet block) $\in \{0, 1\}$;
- dropout of convolutional weights $\in \{0, 1\}$.

In phase 1, the latent dimension is $K \in \{4, 5, 6, 7, 8\}$, the max pooling, batch normalization and skip connection layers have a search space of $\{0, 1\}$, 0 means they do not participate in the architecture and 1 means they do. The batch size was fixed at 64 to decrease the time spent on the hyperparameter optimization. The optimization was performed on the Deligianni et al. (2014) dataset, where a total of 8 individuals and 297 volumes per individual were considered. A split of 75/25 was applied to define the training and validation sets.

In addition to the Bayesian optimization, a neural architecture search procedure is also set, as well as an automation of neural architecture generation Calhas et al. (2022), explained in Appendix 9. Altogether, these steps make the methodology as unbiased as possible from prior and domain based assumptions, removing the human bias and strengthening the generalization.

4.2. Datasets

In total, three simultaneous EEG and fMRI datasets were considered to gather experimental results. In this section, a description of each is provided.

NODDI by Deligianni et al. (2014). A dataset with 10 individuals under resting state with eyes open recordings. The EEG recording setup has a total of 64 channels placed according to the 10-20 system, sampled at 250Hz. The fMRI recording was performed with a 2.160 second Time Response (TR) and 30 milliseconds echo time (TE). Each voxel is $3 \times 3 \times 3$mm and the resolution of a volume is $64 \times 64 \times 30$. For this dataset 8 individuals were considered for training and 2 individuals used to form the testing set.

Oddball by Walz et al. (2015). A dataset with 10 individuals subjected to a visual object detection task setting. The EEG recording setup has a total of 43 channels, sampled at 1000Hz. The fMRI recording was performed with a 2 second TR. Each voxel is $3 \times 3 \times 3$mm and the resolution of a volume is $64 \times 64 \times 32$. Similarly, for this dataset 8 individuals were considered for training and 2 individuals for testing.

CN-EPFL by Pereira et al. (2020). A dataset with 20 individuals performing an activity during the recording session. The EEG recording setup has a total of 64 channels placed according to the 10-20 system and sampled at 5000Hz. The fMRI recording was performed with a 1.280 second TR and 31 milliseconds TE. Each voxel is $2 \times 2 \times 2$mm and the resolution of a volume is $108 \times 108 \times 54$. Due to the high spatial resolution of this dataset it can be quite memory consuming to run a neural network that performs an affine transformation to a total of $108 \times 108 \times 54$ voxels. To alleviate memory consumption, the discrete cosine transform (DCT II) Ahmed et al. (1974) was performed and the fMRI volumes were downsampled to $64 \times 64 \times 30$ voxels, by cutting the frequency coefficients and doing the inverse discrete cosine transform (DCT III). For this dataset 16 individuals were considered for training and 4 individuals composed the testing set.

Data preprocessing. For all datasets, we applied an upper band filter of 250Hz. All datasets published preprocessed data, which went through a pipeline that removed blink and muscle artifacts. Further, the EEG representations were modified by applying a short-time Fourier transform Allen (1977) with a window of length equal to the length of the TR of the respective dataset, in order to allow a direct temporal synchronization between the EEG and fMRI. For the datasets considered this means one is evaluating frequencies as low as ≈ 0.5Hz for the NODDI and Oddball, and ≈ 0.78Hz for the CN-EPFL. Although lower frequencies are not ranked as the most relevant correlations with haemodynamics Laufs et al. (2003), they are nonetheless informative. Then, the pairing of EEG, $\boldsymbol{x}$, and an fMRI volume, $\boldsymbol{y}$, was done in such a way that 20 seconds of EEG were considered for a single fMRI volume. Only EEG information, from before the previous 6 seconds the fMRI volume was taken. This goes in accordance with the claim that neuronal activity only reflects changes in haemodynamics 5.4 to 6 seconds after Liao et al. (2002). As such, $\boldsymbol{x}$ is taken in an interval $[t-26, t-6]$, being t the referenced time when the fMRI volume was taken. Consequently, formalizing the EEG representation $\boldsymbol{x} \in \mathbb{R}^{C \times F \times 20}$. The frequency dimension is specified by the sampling rate of the dataset, no frequency band mapping was applied.

4.3. Extrapolation for a classification setting

To validate the health care setting fit of the synthesized signal, we tested the representation on the Fribourg EEG-only dataset. The dataset, from Padée et al. (2022), consists of 43 individuals, 24 healthy controls and 19 with diagnosed schizophrenia. The recordings were done while the individual played a game. The EEG had a setup of 128 electrodes, distributed according to the 10-20 system, with a sampling rate of 2048Hz. To extrapolate the proposed model to a classification setting we need: 1) to save a model pretrained on the NODDI dataset; 2) the decoder components need to be fixed, as well as the ω and β Fourier feature parameters; 3) the encoder parameters are initialized and trained with the new EEG representations. Let $\boldsymbol{x}_c \in \mathbb{R}^{128 \times 134 \times 10}$ be the EEG representation and $y_c \in \{0, 1\}$ the target, specifying that the individual is a healthy control, $y_c = 0$, or a schizophrenic, $y_c = 1$. We hypothesize that $\boldsymbol{x}_c$ contains the information necessary to separate the data. The sinusoids, from the Fourier features, allow the neural network to synthesize fMRI style representations, regardless of the distribution shift of the dataset. In addition, the *style* prior encodes the neural network's own spectral coefficients, learned from the NODDI dataset, and help the network to synthesize data identically distributed to the learned fMRI from a different EEG distribution. Therefore the model used is denoted as (iv) w/ a *style prior*.

We perform a leave one individual out cross validation on the Fribourg dataset. At each fold, we train the encoder to minimize a contrastive loss, where positive (negative) pairs refer to instances of the same (opposite) class, $\forall i, j \in \{1, \ldots, S\} : 1[y_{c,i} = y_{c,j}]$, being S the number of individuals.

5. Results

Using the introduced experimental setting, results are produced under the methodology described in Section 3 and assessed against the state-of-the-art approaches by Liu et al.

Model	RMSE			SSIM		
	NODDI	Oddball	CN-EPFL	NODDI	Oddball	CN-EPFL
(i)	0.5261±0.0279	0.7419±0.0290	0.5860±0.0865	0.4304±0.0163	0.1829±0.0332	0.5037±0.0734
(ii) w/ *style* posterior	0.3854±0.0108	0.7728±0.1184	0.5288±0.0355	0.4840±0.0070	0.1580±0.0405	0.5221±0.0707
(iii)	0.4392±0.0103	0.7326±0.0463	0.5282±0.0614	0.4544±0.0070	0.1963±0.0388	0.5074±0.0833
(iv) w/ *style* posterior	0.4141±0.0190	0.7014±0.0855	0.5166±0.0560	0.4698±0.0113	0.2004±0.0172	0.5222±0.0877
Liu et al. (2019)	0.4549±0.0806	0.8591±0.0342	0.5915±0.1083	0.4488±0.0601	0.1885±0.0380	0.5190±0.1062

Table 1: Root mean squared error (RMSE) and structural similarity index measure (SSIM) of the target synthesis task for the proposed and state-of-the-art models across all datasets. (i) refers to the linear projection in the latent space, (ii) refers to topographical attention on the EEG channels dimensions with a linear projection in the latent space, (iii) implements a random Fourier feature projection in the latent space, and (iv) performs topographical attention on the EEG channels dimension with a random Fourier features projection in the latent space.

(2019)[2]. The hyperparameters were optimized in the NODDI dataset and are reported in Table 3. These were used for the experiments of all the other datasets. The baselines subject to comparison with the state-of-the-art are:

- (i) Linear projection on the latent space representation, $\boldsymbol{z}_x$;
- (ii) [with *style* posterior] Topographical attention on the EEG electrode dimension;
- (iii) Random Fourier feature Tancik et al. (2020) projection on the latent space representation, $\boldsymbol{z}_x^*$;
- (iv) [with *style* posterior] Combination of (ii) and (iii), as topographical attention is applied in the EEG electrode dimension, as well as the random Fourier feature Tancik et al. (2020) projection on the latent space representation, $\boldsymbol{z}_x^*$.

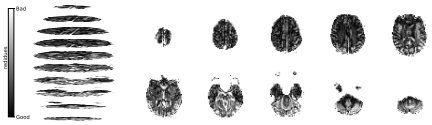
Additionally, experiments of (ii) and (iv), with no style and with a *style* prior learnable vector, are reported in section 5.2.

5.1. fMRI synthesis

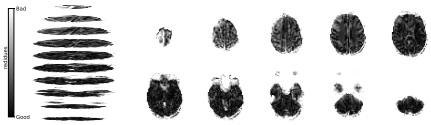
Figure 2 illustrates the distribution of residues (observed vs. estimated differences) on the fMRI volumes for the NODDI dataset. Clearly, by visual inspection, (iv) model has the darker and biggest area of shaded regions, which implies a better coverage across the brain regions and better synthesis quality. Models with topographical attention, (ii) and (iv), corresponding to Figures 2b and 2d, respectively, significantly improve the synthesis, as shown by the darker and bigger areas against (i) and (iii) depicted in Figures 2a and 2c, respectively. Particularly, we notice that models (i) and (iii) report difficulty in the retrieval of haemodynamical activity located in occipital and parietal lobes.

To better address which regions our baselines had more difficulty retrieving, the normalized residues were computed and are illustrated in Figure 3. Baselines – corresponding to

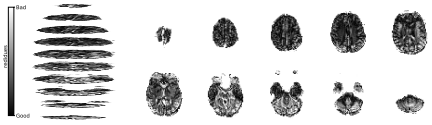
2. Please note that the baseline of Liu et al. (2019) is implemented by the authors as there is no public implementation by the original study. Further, only the EEG to fMRI description was considered and only one volume is synthesized, as opposed to all volumes at once as done in Liu et al. (2018). The model was trained to minimize the MAE loss function.



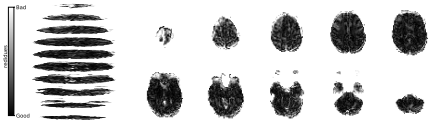
(a) (i) - Linear latent projection. RMSE of 0.5261 ± 0.0279 and SSIM of 0.4304 ± 0.0163 .



(b) (ii) - Topographical attention on the EEG channels dimension, with linear latent projection (i). Attention scores are placed as a *style* posterior on the latent representation, as described in Equation 5. RMSE of 0.3854 ± 0.0108 and SSIM of 0.4840 ± 0.0070.



(c) (iii) - Random Fourier feature latent projection. RMSE of 0.4392 ± 0.0103 and SSIM of 0.4544 ± 0.0070.



(d) (iii) - Topographical attention on the EEG channels dimension, with random Fourier latent projection (i). Attention scores are placed as a *style* posterior on the latent representation, as described in Equation 5. RMSE of 0.4141 ± 0.0190 and SSIM of 0.4698 ± 0.0113.

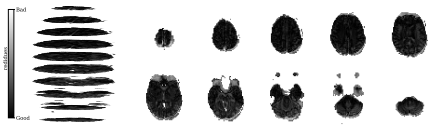
Figure 2: Mean absolute residues for each implemented models. Model (ii), implementing topographical attention with a *style* posterior, and model (iv), additionally transforming the latent features using the random Fourier feature projection (described in Section 3.1), achieve the best performance relative to RMSE and SSIM metrics.

models (i) and (ii), shown in Figures 3a and 3b respectively, which correspondingly implement a linear projection in the latent space and topographical attention –, have difficulty retrieving the prefrontal, occipital and parietal lobes, as the shade tends to a lighter grey in that region. Model (iv), shown in Figure 3d, does not show a noticeable region with a lighter tone of grey, which implies no evident difficulty in retrieving haemodynamical activity across the different brain regions.

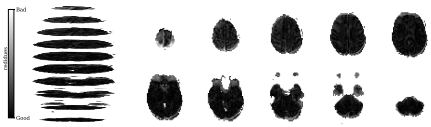
Table 1 contains the results obtained from running the target approaches ((i), (ii), (iii) and (iv)) and the state-of-the-art Liu et al. (2019). For Oddball and CN-EPFL datasets, model (iv) obtained the best RMSE values. In contrast, model (ii) had the best results in the NODDI dataset. Further, our baselines consistently outperform the state-of-the-art, according to the RMSE metric. From analyzing our baselines, we conclude that random Fourier features, described in Section 3.1, benefit models (i) and (ii) in task based recordings and the introduction of topographical attention also benefits both models (i) and (iii) for resting state and task based recordings. The latter, shows the adaptability and robustness of introducing topographical relationships to the synthesis of fMRI. By assessing the experiments from the perspective of the SSIM metric, there is not a concordant superiority across all datasets, as observed with the RMSE. Nonetheless, the state-of-the-art is outperformed by at least one of our baselines on all datasets. Specifically, on the NODDI dataset (resting

	RMSE			SSIM		
	NODDI	Oddball	CN-EPFL	NODDI	Oddball	CN-EPFL
(ii) w/o *style*	0.4820±0.0096	0.9812±0.0847	0.5458±0.0596	0.4408±0.0031	0.1930±0.0543	0.5027±0.0748
(iv) w/o *style*	0.5181±0.0178	0.7221±0.0411	0.5298±0.0636	0.4256±0.0071	0.1991±0.0382	0.5063±0.0830
(ii) w/ *style* prior	0.4834±0.0187	0.9920±0.8901	0.9920±0.8901	0.4426±0.0119	0.1760±0.0402	0.4974±0.1353
(iv) w/ *style* prior	0.5167±0.0063	0.7394±0.0377	0.5568±0.0737	0.4313±0.0007	0.1873±0.0347	0.4960±0.1084

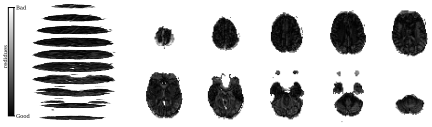
Table 2: RMSE and SSIM scores in the absence and presence of prior styling, all considering the presence of a posterior style vector conditioned on the attention scores. The upper half of this table shows the results of implementing topographical attention, but without using the attention scores to add style to the latent space representation (w/o style). The bottom half, shows the use of a style prior vector, $\in \mathbb{R}^L$, that is not conditioned on any features, and serves to add learnable style features to the latent representation. The latter is widely used in computer vision research, with a recent study applying it to generate images Gu et al. (2021).



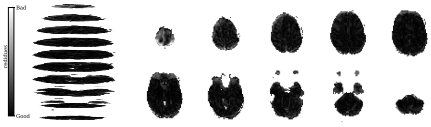
(a) (i) - Linear latent projection.



(b) (ii) - Topographical attention on the EEG channels dimension, with linear latent projection (i).



(c) (iii) - Random Fourier feature latent projection.



(d) (iii) - Topographical attention on the EEG channels dimension, with random Fourier latent projection (i).

Figure 3: Normalized mean absolute residues for the proposed models.

state), we observe that incorporation of topographical attention in model (ii), under a *style* posterior, achieves the best SSIM value.

Figure 4 illustrates the voxel wise comparance, with statistical significance report, between (ii) and (iv).

For the Oddball dataset, the RMSE and SSIM metrics report a worse synthesis ability for all methodologies compared to the other datasets. Our baselines outperform the state-of-the-art, and model (iv) with a *style* posterior is significantly superior to all baselines. Random Fourier projections, (iii), appear to better address the synthesis task than topographical attention alone, (ii). The SSIM is rather poor, with values below 0.2000 being the mean and only model (iv) surpassing this threshold with 0.2004 SSIM.

Models (ii) and (iv), both implementing topographical attention with a *style* posterior, show the best performance in terms of SSIM metric in the CN-EPFL dataset. In spite of the RMSE and SSIM not being in total accordance, the topographical attention superiority is consistent for the metrics considered. This supports our hypothesis that the use of

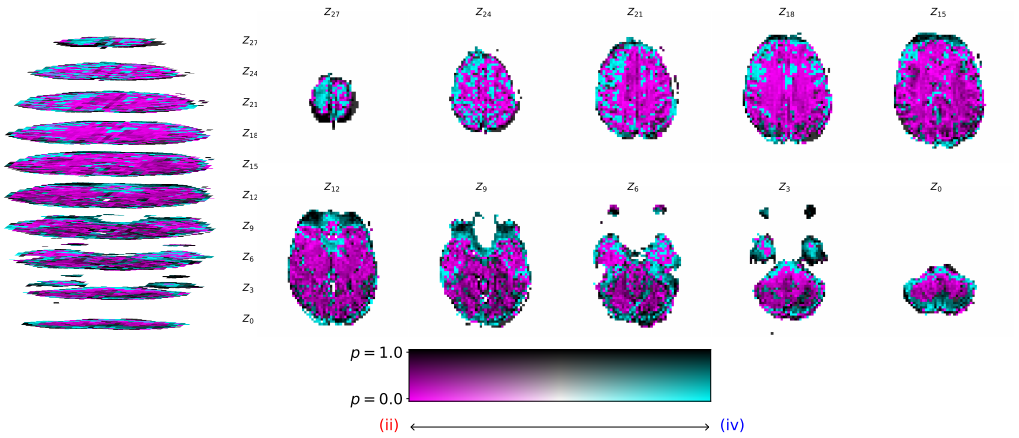


Figure 4: Region-sensitive comparison of models (ii) and (iv), both using *style* posterior, reporting the best model in each voxel according to predictive power (statistical significance under t-test). The report of voxel wise statistical significance further validates the results reported in Table 1, showing the superiority of model (ii) against (iv).

topographical structures plays an important role when studying these two modalities and is hence preferable.

5.2. Role of topographical attention

In Table 1, we reported results of models (ii) and (iv), both implementing a *style* posterior vector that is conditioned on the learned attention graph. This graph is a representation of the relationships between the EEG electrodes, learned during the optimization process, that inherently help the retrieval of haemodynamical activity. To validate this hypothesis, Table 2 shows the RMSE and SSIM metrics obtained from experiments ran on the following models:

- (ii) and (iv) with no *style* induction, but still performing attention in the EEG electrode dimension;
- (ii) and (iv) with *style* prior, reported on the bottom half.

From the previous section, we know that the topographical attention, inducing a *style* posterior on the latent representation (see Section 5), consistently benefits the regression task across all the datasets considered in our experiments. This holds for resting state (NODDI) and task-based (Oddball and CN-EPFL) settings. By comparing the results of models (ii) and (iv) reported in Table 1 with the ones presented in Table 2, the impact of conditioning the *style* posterior vector on the attention scores is quite noticeable. And it goes beyond the simple induction of *style* in the latent space, as Table 2 shows that placing a *style* prior can cause overfitting in some settings.

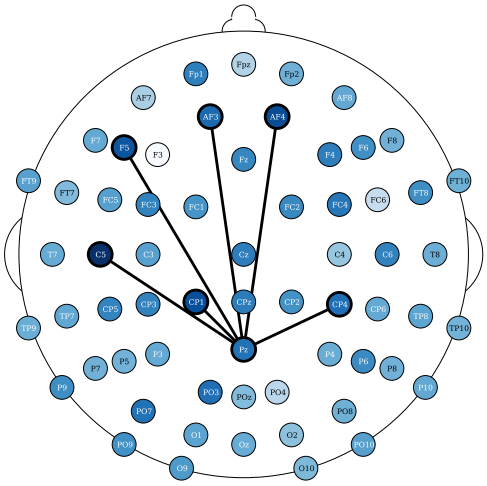
5.3. Discriminative insights

The methodology proposed is able to extrapolate to a diagnostic setting, reporting an area under the receiver operating curve of 0.77. There were two groups in the Fribourg dataset: schizophrenic and healthy controls. The reported score indicates that the synthesized fMRI is able to be applied in a diagnostic setting for schizophrenia.

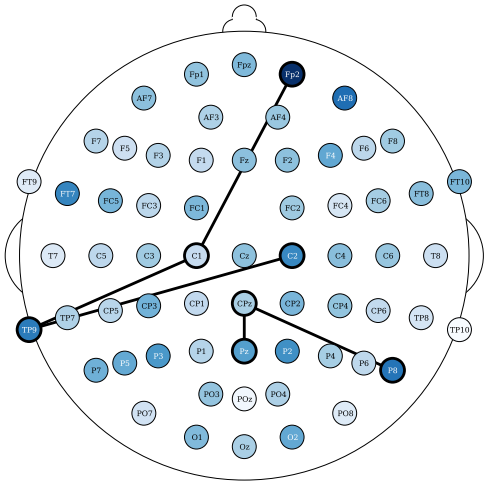
6. Discussion

EEG electrode attentional based relations dependency. The ran experiments with different types of *style*, z_w, in the latent representation (see Equation 5), tell us that conditioning the styling on the attention scores, an EEG electrode topographical representation, is beneficial for the fMRI synthesis task. Further, the fact that, in addition to not conditioning *style*, learning a *style* prior vector is not as informative (no dependency on $\boldsymbol{x}$) for the neural network to better optimize the learning objective. This leads us to believe that a learnable unconditioned *style* acting as a prior, is prone to overfitting the training data, since it is not conditioned on $\boldsymbol{x}$. Our experiments show that the projected random Fourier features (prior), $\boldsymbol{z}_x \rightarrow \boldsymbol{z}_x^*$, if multiplied (conditioned) by data dependent (EEG attention graph scores), Equation 5, not only reduces the empirical risk, but is also preferable to both multiplication of an unconditioned learnable *style* prior and no multiplication at all. Therefore, the placement of a *style* posterior, conditioned on EEG attention scores guides the random Fourier features and removes the inherent assumptions of a prior Tenenbaum et al. (2011). Adding to it, the topographical information retrieved from the attention scores contains information that is highly related to haemodynamical activity, this is in accordance with several neuroscience studies that use topographical structures, such as graphs, to relate EEG and fMRI, used in simultaneous EEG and fMRI studies Yu et al. (2016); Rojas et al. (2018); Bréchet et al. (2019).

Most relevant electrode relations. Consider the relevance of the attention scores, computed from models (ii) and (iv), both having topographical attention at the EEG channel dimension, and model (iv) with projected random Fourier features in the latent space. These relevances were propagated, using the LRP algorithm Bach et al. (2015) described in Appendix 13, through the attention style based posterior. Figure 5 shows the relevances plotted in a white to blue scale, from less relevant to most relevant, respectively. The latter only shows the edges that are above the 99.7 percentile. The presence of an edge between electrodes suggests that either this connection yields a Markovian property for the EEG instance or, otherwise, it is relevant to add fMRI style conditioned on these connections (recall from Section 3.2 that posterior z_w conditions the latent EEG representation z_x^* such that $z_x^* \odot z_w$). For resting state fMRI, both Figures 10a and 5a show connections of parietal and occipital channels (O2 electrode in Figure 10a and Pz electrode in Figure 5a) with frontal and central channels to be the most relevant (above the 99.7 percentile of relevance). Figure 10a reports an additional connection between the Oz and PO9 electrodes, a correspondence between an occipital and a parietal-occipital electrode, which is in accordance with connectivity observations reported by Rojas et al. (2018). There were no reported relevances for the electrodes (T) placed in the temporal regions for resting state settings. In contrast, in task-based fMRI synthesis, relevant relationships between temporal (FT9 and TP9) and



(a) (iv) - Topographical attention on the EEG electrodes dimension, with random Fourier feature projections in the latent space, in NODDI dataset.



(b) (iv) - Topographical attention on the EEG electrodes dimension, with random Fourier feature projections in the latent space, in CN-EPFL dataset.

Figure 5: EEG electrode attention score relevances for resting state NODDI and task based CN-EPFL datasets. Figures 5a and 10b report the attention relevances for the NODDI resting state dataset and the CN-EPFL dataset, respectively.

frontal/central (Fp2 and C1/C2, respectively) electrodes were reported, see Figures 10b and 5b. In both of these figures, connections between central and parietal electrodes were observed. Particularly, there were reported connections between Cz with Pz and CP5 and CP2 electrodes in Figure 10b. And connections between Pz and P8 with CPz electrodes in Figure 5b.

Converging to retrieve near scalp haemodynamical activity. One interesting phenomena that was observed by propagating relevances from the latent representations of the fMRI instance, $\boldsymbol{z}_y$, to the input, $\boldsymbol{y}$, was that the relevances in sub-cortical areas were neither positive nor negative, yielding residual relevance, as seen in Figure 6. This later observation suggests that haemodynamical activity from these areas does not significantly aid the targeted synthesis. Recall that the regularization term, $\Omega(\boldsymbol{z_x}, \boldsymbol{z_y}) = cos(\boldsymbol{z_x}, \boldsymbol{z_y})$, is used with the latent EEG and fMRI representations. This is in accordance with the fact that the retrievable information is in its majority next to the scalp, where the electrodes are placed, and indeed de Beeck et al. (2019) discuss how high frequencies are not able to travel significant distances with obstacles, such as white matter and the scalp, in between. We also report negative relevances on the visual cortex and positive relevances on the occipital and prefrontal lobes. Please note that negative and positive relevances represent relevant features, whereas when one has zero relevance, it means a feature was not relevant for the task. Daly et al. (2019) found that neuronal activity retrieved from EEG can reflect the haemodynamical changes in subcortical areas. Here we claim that haemodynamical activity information in areas next to the scalp are relevant to learn the shared latent space.

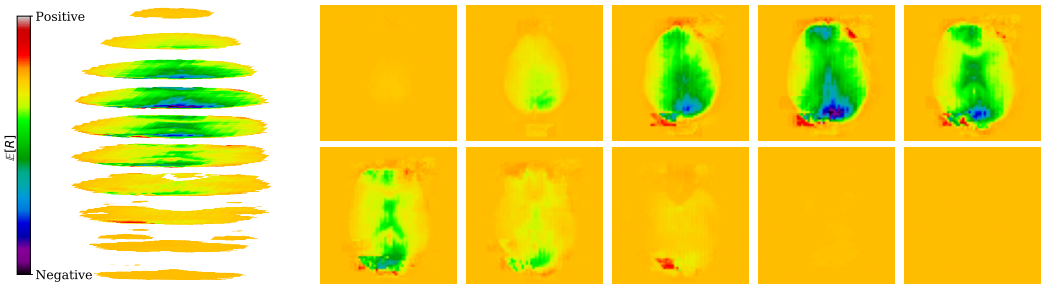


Figure 6: fMRI computed relevances for the NODDI dataset, starting from the latent fMRI representation, $\boldsymbol{z}_y$.

Laboratory setup impacts EEG to fMRI synthesis. The results show that it is more difficult, according to the RMSE metric, to synthesize task-based fMRI than resting state. This observation is in contrast with studies that report that resting state fMRI is inherently more complex than task based fMRI Maknojia et al. (2019). The SSIM metric, in contrast to the RMSE, shows less significant differences for the Oddball recordings in favor of fMRI synthesis in the resting state. However, the CN-EPFL dataset is not in accordance with the latter. This performance heterogeneity across the datasets may not only rise from the characteristics of the recording sessions, but may be also propelled by the different preprocessing techniques employed. Each dataset is publicly available and is supported with published studies, having unique equipment, experimental protocols, and algorithms. CN-EPFL dataset is the most complete one, with a total of 20 individuals and with a resolution of $2 \times 2 \times 2$mm, which makes a total of $108 \times 108 \times 64$ voxels. These differences, caused by working with 3 Tesla (CN-EPFL dataset) versus 1.5 Tesla (NODDI and Oddball datasets) scanners, significantly impact the spatial resolution, which for the datasets NODDI and Oddball produce $64 \times 64 \times 30$ and $64 \times 64 \times 32$ voxels, respectively, with around $3 \times 3 \times 3$mm voxel size. One has to further account for the original recording artifacts and disruptions caused by the applied preprocessing techniques. For instance, Oddball dataset contains intra and inter individual wise misalignments across fMRI volumes. This may be the cause of poorer performance of all methods when compared to the other datasets. In addition, Oddball relies on a different EEG electrode positioning system, having a total of 43 electodes that were not placed in accordance with the 10-20 system Jasper (1958). Although NODDI and CN-EPFL recordings are in accordance with this system, each selected unique electrode locations (see the different electrode placements between Figures 5a and 5b). Finally, the different EEG sampling frequencies, with 250Hz, 1000Hz and 5000Hz considered in NODDI, Oddball and CN-EPFL recordings, respectively, further affect architectural operations and subsequently impact the learning.

The synthesized fMRI extrapolates to a diagnostic setting. We showed the ability of the synthesized fMRI to be applied in an EEG diagnostic setting. Our methodology is robust to shifts in the data distribution, enabling its application in data never before seen. The latter, is due to the nature of the Fourier features incorporated in the (iii) and (iv) model baselines. If one tries to apply the models of (i), (ii) or Liu et al. (2019), we see that the produced fMRI is ill defined since the distribution shift perturbed the learned representation. One of the solutions would be to apply a consistency learning approach, e.g., activations for

all layers, z_l, would be perturbed with gaussian noise, $\sigma \sim \mathcal{N}(0, \epsilon) : z_l = z_l + \sigma$. The latter however, can perturb the data in such a way that one can not extrapolate to a classification setting afterwards.

7. Related work

EEG and fMRI synthesis research state. The learning of mapping functions between structural neuroimaging modalities is increasingly prevalent, e.g. transfer functions between MR and CT scans achieved arguable success using convolutional networks Nie et al. (2017); Wolterink et al. (2017). For a comprehensive description of multi modal brain structural image synthesis, please refer to Yi et al. (2019). In contrast, regression between functional neuroimaging modalities has not received the same amount of attention. The pivotal work by Liu et al. (2019) relies on diverse convolutional operations to pursue the fMRI synthesis from EEG. The work in its entirety goes further and performs bi-directional synthesis. Nonetheless, and despite the efforts made by Liu et al. (2019), advances in this field remain to be explored. Related studies relate haemodynamics with a second (non-neurological) modality using optimized transformations Cury et al. (2020); Raposo et al. (2022), while others work directly in the regression of localized haemodynamical activity in the context of natural language Jain and otehrs (2021) or music retrieval tasks Hoefle et al. (2018). All of these works show the feasibility of haemodynamical retrieval and its usefulness to enrich information. The lack of exploration of cross modal functional neuroimaging arises from various factors inherently present in the structural and representational dissimilarities between modalities. In contrast with structural neuroimaging techniques, the alignment of functional neuroimaging modalities in time has to be further ensured. Simultaneous EEG and fMRI recordings Deligianni et al. (2014); Walz et al. (2015); Pereira et al. (2020) have been increasing conducted, representing a notable effort by the research community that can pave the way to further breakthroughs on the fMRI synthesis from EEG signals.

Simultaneous EEG and fMRI studies. Chang et al. (2013) claim decreases in alpha band and increases in theta band in the time dimension are correlated with relative increases in functional connectivity. Cury et al. (2020) combined EEG and fMRI to build an EEG informed fMRI modality, that is cheaper than fMRI and contains informed features. He et al. (2018) report positive correlation between the haemodynamical activity with alpha band, showing that the temporal resolution of spectral information is important to address the combination of these modalities. Leite et al. (2013) associated EEG spectral features with haemodynamicall activity for a single epileptic subject. Similarly, Rosa et al. (2010) observe that changes in haemodynamical activity were also in accordance with changes in spectral features of neuronal activity. Results gathered in the context of our work further confirm the majority of the aforementioned findings.

8. Conclusion

We found that topographical relationships between EEG channels are highly relevant and beneficial for the targeted fMRI synthesis task. Our experiments conclude that attention-based scores, trained to give Markovian properties to the EEG representation and simultaneously add style features by usage of a posterior, significantly aid the task. Relationships

learned between occipital, parietal and frontal electrodes were observed to be of particular relevance to retrieve haemodynamical activity. We further noticed that haemodynamical information in areas next to the scalp is predominantly considered to learn the shared latent space during the training, aiding fMRI synthesis.

We hope to have motivated researchers to work in this emerging field. Please refer to this github repository (https://github.com/eeg-to-fmri/eeg-to-fmri) to access the code that was used to develop this research. Neuroimaging synthesis and augmentation from more accessible modalities yields unique opportunities for reducing costs and improving diagnostics in health care settings, while offering ambulatory and longitudinal proxy views of haemodynamical activity. Future work is expected to validate the feasibility of the fMRI synthesis task for the aforementioned ends, comprehensively assessing the predictive limits of electrophysiological activity measured at the cortex. The successful extrapolation of the synthesized fMRI motivates further analysis of a new perspective of EEG. Ultimately, explainability and uncertainty quantification analyses are needed for the application of EEG to fMRI synthesis in an health care setting.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e Tecnologia under the PhD Grant SFRH/BD/5762/2020 to David Calhas and INESC-ID pluriannual UIDB/50021/2020. We thank Alexandre Francisco and Sérgio Pereira for their very useful input to the work, in the context of the PhD advising committee. We want to give a special thanks to João Rico, Daniel Gonçalves, Pedro Orvalho and Leonardo Alexandre for giving pivotal feedback on the visual support used in this study. We also want to thank António Gusmão for the enriching discussions had on the role of attention mechanisms used.

References

Charles Sherrington. *The integrative action of the nervous system*. LWW, 1952.

Randy L Buckner. Event-related fmri and the hemodynamic response. *Human brain mapping*, 1998.

Hiroshi Shibasaki. Human brain mapping: hemodynamic response and electrophysiology. *Clinical Neurophysiology*, 2008.

Qingbao Yu et al. Building an eeg-fmri multi-modal brain graph: a concurrent eeg-fmri study. *Frontiers in human neuroscience*, 2016.

Yifei He et al. Spatial–temporal dynamics of gesture–speech integration: a simultaneous eeg-fmri study. *Brain Structure and Function*, 2018.

Gonzalo M Rojas et al. Study of resting-state functional connectivity networks using eeg electrodes position as seed. *Frontiers in neuroscience*, 2018.

Lucie Bréchet et al. Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with eeg and fmri. *Neuroimage*, 2019.

Ian Daly et al. Electroencephalography reflects the activity of sub-cortical brain regions during approach-withdrawal behaviour while listening to music. *Scientific reports*, 2019.

Claire Cury et al. A sparse eeg-informed fmri model for hybrid eeg-fmri neurofeedback prediction. *Frontiers in neuroscience*, 2020.

Rodolfo Abreu et al. Eeg microstates predict concurrent fmri dynamic functional connectivity states. *Brain topography*, 2021.

Xueqing Liu et al. A convolutional neural network for transcoding simultaneously acquired eeg-fmri data. In *NER*. IEEE, 2019.

Fani Deligianni et al. Relating resting-state fmri and eeg whole-brain connectomes across frequency bands. *Frontiers in Neuroscience*, 2014.

Jennifer M Walz et al. Prestimulus eeg alpha oscillations modulate task-related fmri bold responses to auditory stimuli. *NeuroImage*, 2015.

Michael Pereira et al. Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, 2020.

Hans Op de Beeck et al. *Introduction to human neuroimaging*. Cambridge University Press, 2019.

Godwin Inalegwu Ogbole et al. Survey of magnetic resonance imaging availability in west africa. *Pan African Medical Journal*, 2018.

Edwin JR van Beek et al. Value of mri in medicine: More than just another test?, 2019.

Jiatao Gu et al. Stylenerf: A style-based 3d-aware generator for high-resolution image synthesis. *arXiv*, 2021.

Fernando Lopes da Silva. Eeg and meg: relevance to neuroscience. *Neuron*, 2013.

Helmut Laufs et al. Eeg-correlated fmri of human alpha activity. *Neuroimage*, 2003.

David Calhas et al. Automatic generation of neural architecture search spaces. In *AAAI Workshop*, 2022.

Rachid and Riad. Learning strides in convolutional neural networks. *arXiv*, 2022.

Ali Rahimi et al. Random features for large-scale kernel machines. In *Neurips*, 2007.

Matthew Tancik et al. Fourier features let networks learn high frequency functions in low dimensional domains. *arXiv*, 2020.

Zhu Li et al. Towards a unified analysis of random fourier features. In *Icml*, 2019.

Herbert H Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, 1958.

others. Imagenet classification with deep convolutional neural networks. *Neurips*, 2012.

Hubert Banville et al. Robust learning from corrupted eeg with dynamic spatial filtering. *NeuroImage*, 2022.

Ngoc-Trung Tran et al. Dist-gan: An improved gan using distance constraints. In *ECCV*, 2018.

Jasper Snoek et al. Practical bayesian optimization of machine learning algorithms. *Neurips*, 2012.

Nasir Ahmed et al. Discrete cosine transform. *IEEE transactions on Computers*, 1974.

Jonathan Allen. Short term spectral analysis, synthesis, and modification by discrete fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1977.

C.H. Liao et al. Estimating the delay of the fmri response. *NeuroImage*, 2002.

Anna Padée et al. "fribourg ultimatum game in schizophrenia study", 2022.

Hanxiao Liu et al. Darts: Differentiable architecture search. *arXiv*, 2018.

Joshua B Tenenbaum et al. How to grow a mind: Statistics, structure, and abstraction. *science*, 2011.

Sebastian Bach et al. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, 2015.

Sanam Maknojia et al. Resting state fmri: Going through the motions. *Frontiers in neuroscience*, 2019.

Dong Nie et al. Medical image synthesis with context-aware generative adversarial networks. In *MICCAI*, 2017.

Jelmer M. Wolterink et al. Deep mr to ct synthesis using unpaired data. In *Simulation and Synthesis in Medical Imaging*, 2017.

Xin Yi et al. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 2019.

Francisco Afonso Raposo et al. Learning low-dimensional semantics for music and language via multi-subject fmri. *Neuroinformatics*, 2022.

Shailee Jain and otehrs. Interpretable multi-timescale models for predicting fmri responses to continuous natural speech. *bioRxiv*, 2021.

Sebastian Hoefle et al. Identifying musical pieces from fmri data using encoding and decoding models. *Scientific reports*, 2018.

Catie Chang et al. Eeg correlates of time-varying bold functional connectivity. *Neuroimage*, 2013.

Marco Leite et al. Transfer function between eeg and bold signals of epileptic activity. *Frontiers in neurology*, 2013.

Maria J Rosa et al. Estimating the transfer function from neuronal activity to bold using simultaneous eeg-fmri. *Neuroimage*, 2010.

Leonardo Mendonca de Moura et al. Z3: an efficient SMT solver. In *Tools and Algorithms for the Construction and Analysis of Systems*, 2008.

Kaiming He et al. Deep residual learning for image recognition. In *CVPR*, 2016.

Vinod Nair et al. Rectified linear units improve restricted boltzmann machines. In *Icml*, 2010.

Jawad Nagi et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *ICSIPA*. IEEE, 2011.

Sergey Ioffe et al. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Icml*, 2015.

9. Automatic generation of neural architectures

Calhas et al. (2022) proposed a neural architecture generation method that respects the arithmetic of convolutions with a *valid* padding. The relation between the input, I, and output, O, of a downsampling layer (e.g., convolution, pooling) is in accordance with

$$O = \frac{I - k}{s} + 1 \Leftrightarrow I = (O - 1) \times s + k, \tag{6}$$

where k and s are the kernel and stride sizes, respectively. As such, a neural network specification composed of downsampling layers is formalized as

$$\forall n \in \{1, \ldots, N\} : L_n = (k_n, s_n), \tag{7}$$

being L the notation for a layer. In terms of layer processing, the function of a layer is $f_{L_n} : \mathbb{R}^{I_n} \mapsto \mathbb{R}^{O_n}$, where I_n is the input of the nth layer and O_n the output. Similarly, the function that represents the neural network $f : \mathbb{R}^I \mapsto \mathbb{R}^O \wedge f(x) = f_{L_N}(f_{L_{N-1}}(\ldots f_{L_1}(\boldsymbol{x})\ldots))$. Using Satisfiable Modulo Theory and with the encoding

$$O = O_N \wedge \bigwedge_{n=1}^{N} O_n \leq O_{n-1} \wedge k_n > 0 \wedge s_n > 0, \tag{8}$$

one can use a solver de Moura et al. (2008) to get an assignment to the kernel, k, and stride, s, of all layers. The formulation is further extended to multiple dimensions and variable number of layers. For details on the latter please refer to Calhas et al. (2022). This approach is used in order to remove the human bias from the methodology proposed.

Resnet-18 block configuration. The neural architecture specification (Equation 7) is to replace the downsampling blocks of the Resnet-18, illustrated in Figure 7. He et al. (2016) defined kernel and stride sizes set as 1×2 for all blocks. Note that, it is encouraged to use different kernel and stride sizes for the different layers Riad (2022). Therefore, the variable assignments, of Equation 7, are used as the kernel and stride sizes.

10. Hyperparameters and latent dimension

Table 3 reports on the hyperparameters obtained from running the Bayesian optimization algorithm, with the setup described in Section 4.1.

lr	θ decay	batch size	K	filters	max pool	batch norm	skip	dropout
$2.98e-3$	$4.40e-4$	4	7	4	1	1	1	0.5

Table 3: Hyperparameters obtained from the Bayesian optimization algorithm, ran for 100 iterations.

Figure 8 shows the performance of each evaluation made during the hyperparameter search. $K = 7$ achieved the best result, followed by $K = 6$ and $K = 4$. Interestingly, $K = 8$ produced non defined values due to the training being underway, but GPU memory was exceeded. Consequently, these were not counted as evaluations. As for $K = 15$ and $K = 20$, the model was not able to be loaded to the GPU and failed to be evaluated. In contrast with $K = 8$, $K = 15$ and $K = 20$ were evaluated since it did not freeze the GPU. Memory limitation is important, because the target of this framework is to be achievable in a day-to-day laptop, enabling the use of this work in a cheap setup.

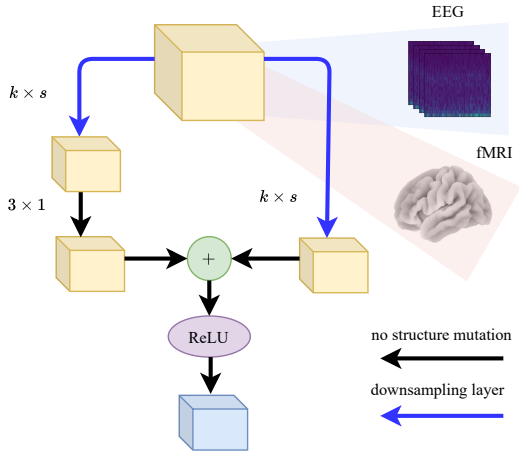


Figure 7: The inspired Resnet-18 block forks the input in two computational flows: (1) the first, represented in the left part of the figure, is processed by a convolutional layer with $k \times s$ as the kernel and stride sizes operate with *valid* padding, following the output goes through a convolutional layer with 3×1 with *same* padding; (2) the second flow, corresponds to the right arrow of the fork, processes the input with a convolutional layer with $k \times s$ with a *valid* padding. The representations of the fork are joined by the *addition* operation, which is followed by a ReLU activation Nair et al. (2010). Please note that max pooling Nagi et al. (2011) and batch normalization Ioffe et al. (2015) layers are optional to follow each downsampling layer. EEG and fMRI feature representations are included in the figure for the reader to understand that this block structure is used to process EEG and fMRI, though differing in the values of $k \times s$ in each network.

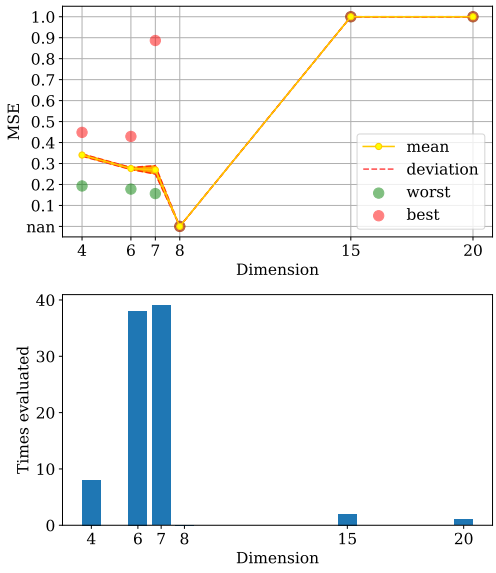


Figure 8: Latent dimension analysis according to the MSE metric.

Candidate	Kernel $\times$ Stride ($\bigwedge_1^N k^{(1)}, k^{(2)}, k^{(3)} \times s^{(1)}, s^{(2)}, s^{(3)}$)	N
1	$11, 86, 2 \times 1, 1, 1 \wedge 17, 20, 2 \times 4, 2, 1 \wedge 2, 7, 2 \times 1, 1, 1$	3
2	$7, 37, 2 \times 3, 5, 1 \wedge 7, 7, 2 \times 2, 2, 1$	2
3	$9, 43, 2 \times 1, 2, 1 \wedge 11, 11, 2 \times 1, 2, 1 \wedge 9, 3, 2 \times 5, 2, 1$	3
4	$28, 15, 2 \times 1, 1, 1 \wedge 30, 77, 2 \times 1, 7, 1$	2
5	$7, 19, 2 \times 1, 1, 1 \wedge 20, 23, 2 \times 1, 4, 1 \wedge 23, 16, 2 \times 2, 1, 1$	3
6	$6, 29, 2 \times 1, 1, 1 \wedge 21, 33, 2 \times 1, 4, 1 \wedge 16, 11, 2 \times 3, 1, 1$	3
7	$32, 47, 2 \times 2, 4, 1 \wedge 4, 15, 2 \times 2, 1, 1$	2
8	$9, 16, 2 \times 3, 1, 1 \wedge 5, 2, 2 \times 1, 1, 1 \wedge 6, 81, 2 \times 1, 5, 1$	3
9	$23, 32, 2 \times 1, 1, 1 \wedge 11, 96, 2 \times 5, 1, 1$	2
10	$16, 31, 2 \times 1, 8, 1 \wedge 24, 6, 2 \times 4, 1, 1$	2

Table 4: From input shape $64 \times 134 \times 10$ to output shape $K \times K \times K$ with $K = 7$. Each layer is followed by a max-pool operation with $2, 2, 1 \times 1, 1, 1$.

11. Neural networks generated

This section describes the generated neural architectures, using the method described in Section 9 and with more detail in Calhas et al. (2022). The neural architecture search was performed with the 01 dataset. The latter, means that the EEG instance $\boldsymbol{x} \in \mathbb{R}^{64\times134\times10}$ and the fMRI instance $\boldsymbol{y} \in \mathbb{R}^{64\times64\times30}$. This search took a total of 3 months, due to the limited GPU resources, as well as the memory of the GPU that was used. In practice, this computational time can be reduced if one does not train all the generated networks at the same time, which means Liu et al. (2018) algorithm would not be used. Nevertheless, the algorithm provides information on the convergence of each network simultaneously, as shown in Figures 9a and 9b.

EEG candidates. Starting with the EEG encoder, which was generated setting $I = 64\times134\times10$ and $O = 7\times7\times7$, the properties, such as the kernel, stride and number of layers, of the architecture are presented in Table 4. Along with the properties of the architectures generated, the convergence of the Liu et al. (2018) algorithm is reported in Figure 9a. By analyzing the Figure, we conclude that the best architecture obtained was the candidate number 2.

fMRI candidates. Following with the fMRI encoder, which was generated setting $I = 64 \times 64 \times 30$ and $O = 7 \times 7 \times 7$, the properties of the architecture are presented in Table 5. By analyzing Figure 9b, we conclude that the best architecture obtained was the candidate number 2.

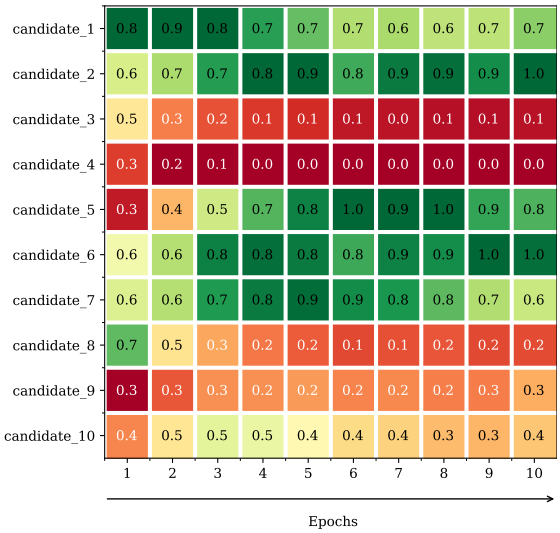
12. Neural architectures generation setup

Table 12 contains the specification of the variables to generate the neural architectures reported in Tables 4 and 5. Following, the DARTS algorithm Liu et al. (2018) was ran on the generated NA search space, for both E_x and E_y, with a total of 10 epochs, learning rate of 0.001. Note that gradients are passed according to zero order, i.e., the weights of each network are trained independently from the weights of the softmax final layer, proposed in Liu et al. (2018).[3] Similar to the hyperparameter optimization, a total of 8 individuals were considered and a split of 75/25 was done to define the training and validation sets.

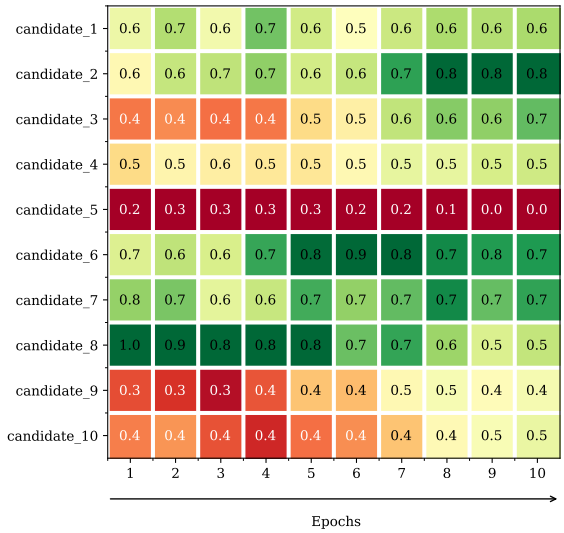
3. The computational resources to compute all the gradients would require large amounts of GPU allocated memory, that surpasses the used NVIDIA GeForce RTX 2080 GPU (8GB) capacity.

Candidate	Kernel × Stride ($\bigwedge_1^N k^{(1)}, k^{(2)}, k^{(3)} \times s^{(1)}, s^{(2)}, s^{(3)}$)	N
1	$16, 8, 8 \times 4, 2, 1 \wedge 2, 16, 9 \times 1, 1, 1 \wedge 3, 5, 6 \times 1, 1, 1$	3
2	$16, 6, 12 \times 2, 1, 1 \wedge 6, 4, 6 \times 1, 1, 1 \wedge 12, 47, 5 \times 1, 1, 1$	3
3	$8, 15, 3 \times 1, 4, 1 \wedge 38, 6, 21 \times 3, 1, 1$	2
4	$8, 7, 15 \times 1, 1, 1 \wedge 20, 5, 2 \times 1, 1, 1 \wedge 15, 10, 6 \times 3, 6, 1$	3
5	$6, 20, 2 \times 5, 1, 1 \wedge 5, 8, 16 \times 1, 6, 2$	2
6	$6, 44, 15 \times 1, 1, 1 \wedge 28, 7, 5 \times 1, 1, 1 \wedge 16, 6, 3 \times 2, 1, 1$	3
7	$14, 13, 5 \times 1, 1, 2 \wedge 18, 16, 2 \times 1, 1, 1 \wedge 11, 21, 3 \times 3, 2, 1$	3
8	$8, 11, 14 \times 1, 1, 1 \wedge 29, 19, 6 \times 1, 1, 1 \wedge 20, 27, 3 \times 1, 1, 1$	3
9	$7, 2, 7 \times 1, 1, 1 \wedge 29, 25, 9 \times 1, 1, 1 \wedge 21, 23, 7 \times 1, 2, 1$	3
10	$17, 28, 5 \times 1, 1, 1 \wedge 19, 16, 7 \times 1, 1, 1 \wedge 7, 6, 4 \times 3, 2, 2$	3

Table 5: From input shape $64 \times 64 \times 30$ to output shape $K \times K \times K$ with $K = 7$. Each layer is followed by a max-pool operation with $2, 2, 2 \times 1, 1, 1$.



(a) DARTS weight deviation for the generated NAs for the EEG encoder.



(b) DARTS weight deviation for the generated NAs for the fMRI encoder.

Figure 9: Liu et al. (2018) algorithm weight convergence for each network, for the EEG and fMRI encoder.

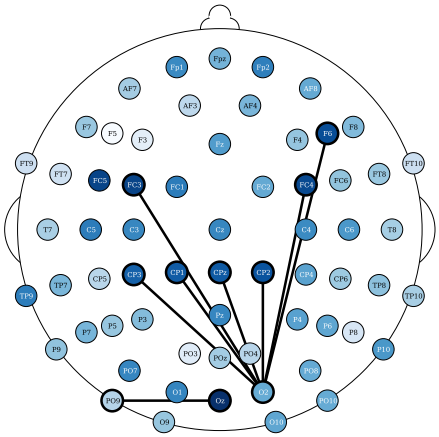
13. Layer-wise relevance propagation

Bach et al. (2015) proposed a method to propagate relevances from the output of a neural network to the input features. This provides relevance features, that have an informative explainability nature, assessing which ones were more relevant (either negatively or positively). Let j be a hidden neuron, following the proposed propagation rule, then its relevance is computed as

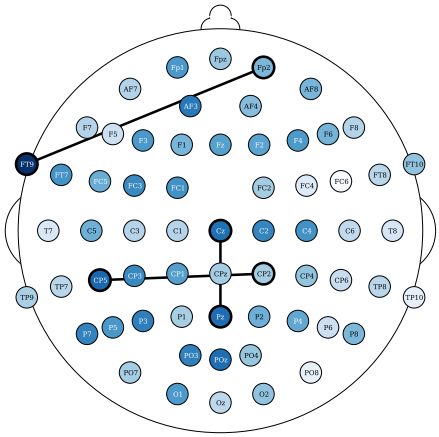
$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_j a_j w_{jk}} R_k, \tag{9}$$

Variable	E_x	E_y
I	$64 \times 134 \times 10$	$64 \times 64 \times 30$
O	$K \times K \times K$	$K \times K \times K$
n	1	1
N	3	5

Table 6: Variable specification for the formula in Equation 8. The I for E_x and E_y correspond to the EEG and fMRI representations, respectively, obtained from the Deligianni et al. (2014) dataset.



(a) (ii) - Topographical attention on the EEG electrodes dimension in NODDI dataset.



(b) (ii) - Topographical attention on the EEG electrodes dimension in CN-EPFL dataset.

Figure 10: EEG electrode attention score relevances for resting state NODDI and task based CN-EPFL datasets. Figures 10a and 10b report the attention relevances for the NODDI resting state dataset and the CN-EPFL dataset, respectively.

where a neuron, k, has a relevance, R_k, associated to it. The relevance of all the neurons of the output layer are by default the output logits and the relevance of all layers are computed by backpropagation of relevances using the rule stated in Equation 9. Note that, this rule does not apply to propagate through sinusoidal activations, which are used in this work (Section 3.1). For EEG features, the relevances are propagated through the proposed *style* posterior (Section 3.2), where standard layers, that enable the use of this rule are used.