EEG to fMRI Synthesis: Is Deep Learning a candidate?

David Calhas
INESC-ID Instituto Superior Técnico
Lisbon, Portugal *david.calhas@tecnico.ulisboa.pt*

Rui Henriques
INESC-ID Instituto Superior Técnico
Lisbon, Portugal *rmch@tecnico.ulisboa.pt*

Abstract

Advances on signal, image and video generation underly major breakthroughs on generative medical imaging tasks, including Brain Image Synthesis. Still, the extent to which functional Magnetic Ressonance Imaging (fMRI) can be mapped from the brain electrophysiology remains largely unexplored. This work provides the first comprehensive view on how to use state-of-the-art principles from Neural Processing to synthesize fMRI data from electroencephalographic (EEG) data. Given the distinct spatiotemporal nature of haemodynamic and electrophysiological signals, this problem is formulated as the task of learning a mapping function between multivariate time series with highly dissimilar structures. A comparison of state-of-the-art synthesis approaches, including Autoencoders, Generative Adversarial Networks and Pairwise Learning, is undertaken. Results highlight the feasibility of EEG to fMRI brain image mappings, pinpointing the role of current advances in Machine Learning and showing the relevance of upcoming contributions to further improve performance. EEG to fMRI synthesis offers a way to enhance and augment brain image data, and guarantee access to more affordable, portable and long-lasting protocols of brain activity monitoring. The code used in this manuscript is available in Github and the datasets are open source.

1. Introduction

Signal Generation approaches explore how structural transformations can be learned from a given signal collection to encode and/or produce new signals. When the goal is to learn a mapping function between signals from heterogeneous sources, the focus is placed on translations between those sources, transferring the task from Generation to Synthesis. Recent breakthroughs on brain image generation [2, 10], reconstruction [4], enhancement [22] and synthesis [8] are driven by the simultaneous analysis of multiple imaging modalities, mostly Computed Tomography (CT), functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET) [6]. Yi et al. [6] survey recent works that establish mappings between CT, fMRI and PET modalities. In spite of the increasing number of contributions, a noticeable lack in the existing research is the absence of mappings between electroencephalography (EEG) and fMRI data. A contributor factor is the inherent difficulty of mapping electrophysiological and haemodynamic signals, given their contrasting spatial and temporal resolution. Nevertheless, the importance of this task has been largely evidenced:

1. MRI units are still largely scarce in countries worldwide [37]. [43] estimate the presence of 0.24 units per million people in West African countries;

2. in contrast with other brain imaging modalities, electroencephalography is non-invasive, safe, inexpensive, and yields almost no restriction on the extent of recordings [19]. Research and medical-wise, EEG to fMRI synthesis opens up the possibility to perform more

affordable, portable and long-lasting protocols of brain activity monitoring;

3. simultaneous EEG and fMRI monitoring provides a way of complementing strengths and addressing the limitations of both signals: EEG offers a fine temporal and spectral resolution of the brain electrophysiology, while fMRI offers a precise spatial resolution of blood flow changes (associated with brain activity) [41, 55, 31];

4. EEG-based generation of new fMRI images can be further used as a way of augmenting data or guaranteeing its proper privacy [56];

Understanding the extent to which these modalities can be mapped is critical to answer key research problems:

- unravel the complex neurophysiological relationships between the brain's cortical electrophysiology and its haemodynamic;

- access the components of EEG signals that are decodable and non-decodable into fMRI signals, unraveling their role;

- identify the brain regions from each modality that support the synthesis process. This knowledge can be, for instance, used to reveal the semantics of brain activity and its underlying connectivity.

This manuscript proposes an approach to synthesize fMRI from EEG signals based on the composition of convolution (encoding) and transposed convolution (decoding) layers. Given the rich spatiotemporal nature of both modalities, this problem is formulated as learning a mapping function between multivariate time series with highly dissimilar structures (regarding both spatial and temporal resolution). To this end, we provide a comprehensive comparison of state-of-the-art principles from Neural Processing towards multivariate time series data analysis, together with well-established principles for the integrative analysis of EEG and fMRI data:

1. Autoencoders (AE) [48] as a baseline;

2. Generative Adversarial Networks (GAN) [20] with the entropy and Wasserstein (WGAN) losses;

3. linear combination of top-k most correlated fMRI-EEG training data instances;

4. novel class of neural networks combining contrastive loss [13] at the encoder level with reconstruction loss at the decoder level (details in Section 4).

Results show, on one hand, the feasibility of EEG-based synthesis of fMRI signals, pinpointing the role of current advances on the field to tackle this challenging task. On the other hand, they highlight a significant space to improve performance, stressing the relevance of upcoming contributions to the targeted task.

This article is organized as follows. Section 2 provides essential background on deep generative models and relevant work on brain imaging synthesis. Section 3 describes the datasets and preprocessing protocols considered in the context of our study. Section 4 introduces the proposed approaches, covering principles on how to synthesize BOLD signals from EEG signals. Section 5 discusses the gathered empirical results. Finally, concluding remarks and future directions are presented.

2. Related Work

Image Reconstruction and Image Synthesis and their *unconditional*, *cross-modal* and *constrained* variants are important tasks towards our end. On Image Reconstruction, learning methods typically establish transformations to reduce noise, remove artifacts and produce descriptors [56, 50]. These approaches can be extended for medical Image Synthesis, where the goal is to learn transformations for auto-encoding signals of medical nature (MRI, fMRI, EEG, CT, PET). Cross-Modality Image Synthesis maps raw signals from one modality, such as magnetic resonance images, onto signals from another modality, such as CT-like images. This cross-modal translation can be achieved in the presence of both *paired* images (multimodal recordings) and *unpaired* images (non-simultaneous recordings). Understandably, unpaired instances do not respect the data fidelity loss term, being unable to preserve small abnormality regions during the translation process. Unconditional Image Synthesis, also known as Image Generation, learns transformations on samples taken from a certain distribution to generate images resembling the ones considered in the learning phase [12]. Constrained Image Synthesis has its base on applying transformations respecting constraints on the modality being synthesized, such as segmentation maps (e.g. suppressing bones, learning mappings between regions of interest).

The interest and necessity to perform synthesis between brain imaging modalities has been largely motivated [6]. Nie et al. [42] inferred CT images from their corresponding MR images using adversarial training from a fully convolutional network. Similar networks are used by Wolterink et al. [54] to map 2D brain MR image slices into a 2D brain CT image. andothers [6] provide an extensive survey on multi-modal brain image synthesis from a stance of adversarial training.

Despite the existing advances on this field, to our knowledge there are not comprehensive attempts to answer EEG to fMRI synthesis. Two observations may explain this observation. First, the still scarce access to simultaneous EEG and fMRI scans. Second, the difficulty of establishing mappings between these two modalities given their highly distinct spatiotemporal dynamics. In particular, unlike the promising role that adversarial training has in the context of aforementioned studies on multi-modality image synthesis, the principles brought forth by adversarial training are indeed insufficient to deal with the complexity of associations between EEG and fMRI signals (results in Section 5).

In this work, we use contributions from Cross-Modality Image Synthesis tackle the task of synthesizing fMRI data from EEG data using paired recordings (simultaneous EEG and fMRI signals). Section 2.1 introduces essential neural processing principles for Image Synthesis, while section 2.2 surveys state-of-the-ark work on simultaneous EEG and fMRI studies.

2.1. Neural Processing for Image Synthesis

Our work builds upon recent deep learning techniques to synthesize multivariate time series, being inspired on: AE [25], Variational AEs (VAE) [30], β-VAE [24], GAN [20], WGAN [5] and Conditional GANs (CGAN) [38]; with some tweaks to each version in order to adapt them to the task at hand (EEG to fMRI synthesis). Since the goal is to synthesize and not to perform signal generation, samples from distributions are not taken (as it happens with VAE, β-VAE, GAN and WGAN). Instead, similar to CGANs, a decoder synthesizes fMRI based on a hidden representation of the EEG signal (with no concatenation of a random sample). This technique is also known as style transfer in GANs [29, 46]. In fact, cross-modality image synthesis is one of the most important application of GANs [6]. Magnetic reasonance (MR) is ranked as the top medical imaging modality explored in GAN-related literature [6], given the current costs and constraints on MR acquisition. GANs hold the potential to reduce MR acquisition time by faithfully generating sequences from already acquired ones. However, the bounds on the available data and convergence difficulties limit their success.

In cross-modality image synthesis, there are not yet reference loss functions and final metrics

for assessing the generative accuracy of the models [6]. Most works opt to use traditional distance metrics such as Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), or Structural Similarity (SSIM) for quantitative evaluation [26]. These measures, however, do not always correspond to the visual quality of the image and disregard time dependencies along image frames. Therefore, additional metrics are proposed in Section 4.3 and, in addition to quantitative results, qualitative results are complementarily presented in Section 5.

2.2. Simultaneous EEG and fMRI studies

He et al. [23] performed a simultaneous EEG and fMRI study to take advantage of both temporal and spatial precision of the EEG and fMRI, respectively. Their work explores the integration between gesture and speech under a thorough analysis between alpha and beta power and BOLD. Results suggest positive correlation between BOLD and alpha power and show that the temporal resolution for spectral content affect the strength of associations. This work leaves open questions, as it reduces electrophysiology to alpha and beta bands.

Chang et al. [9] collected simultaneous EEG-fMRI data under a resting state condition from 10 healthy adults, and examine whether temporal variations in pairwise coupling of functional connectivity networks (based on fMRI) are associated with temporal variations in the amplitude of EEG power, specifically on alpha and theta frequency bands. Functional connectivity networks were defined using an atlas of functional regions of interest that had been defined from a group-level independent component analysis of resting state fMRI. Decreases in alpha and increases in theta over time were associated with relative increases in functional connectivity. Positive correlations with alpha power were also observed in the thalamus and dorsal anterior cingulate cortex. Although these results motivate the possibility to establish EEG to fMRI mappings, they are constrained to specific spectral bands and connectivity maps, neglecting the rich nature of the EEG and haemodynamic signals.

Leite et al. [32] explore different EEG-fMRI transfer functions. For this purpose, metrics extracted from the EEG spectrum (under Morlet wavelet spectral analysis) were associated with haemodynamics for a single epileptic subject. Significant correlations were reported. Yet, the lack of observations and the peculiar electrophysiology of epileptic subjects hamper the target learning of EEG-fMRI transfer functions. Similarly, Rosa et al. [47] estimated EEG-fMRI transfer functions finding changes in BOLD associated with changes in the EEG spectrum. According to them, these changes do not arise from one specific band, but from the relative power of high and low frequencies. This shows how previous studies [9, 23] would possibly improve results by exploring more frequency bands.

Cury et al. [15] predict combined EEG and fMRI neurofeedback (NF) scores from EEG NF scores. The main goal was to perform a real time NF session using only EEG recording, instead of the costly and non-portable fMRI sessions. The dataset used consisted on a group of 17 subjects. The EEG recording was performed with 64 channels and sampled at 5kHz, while fMRI recordings were produced from a 3 Tesla scanner. The best approach on the training (testing) set claims a Pearson Correlation with mean 0.82 (0.74), an improvement of 10pp against the baseline EEG NF scores. The model is elegant, and more transformations could be added (go from perceptrons to multi-perceptrons) as chains, which is the same as saying deep learning could improve results. In contrast, our work aims at synthesizing BOLD from EEG signals, instead of predicting extracted features (NF scores) from both modalities.

Wei et al. [53] is another study that complements fMRI signal with EEG information using Bayesian fusion. They compare the single use of fMRI signal against complementing fMRI with EEG using a Bayesian belief updating, measuring the added value of EEG. Mosayebi et al. [39] also perform EEG and fMRI fusion by means of a matrix factorization algorithm called Correlated Coupled Matrix Tensor Factorization forcing EEG and fMRI to share the same feature space. The results reported show that there is not a consistent correlation among the

extracted features (please check the original work for more details). Jiang et al. [27] synthesize a functional transcranial brain atlas. Although, they do not perform an actual modality synthesis, this is another example of the upwards trend of the functional neuroimaging modalities synthesis research.

3. EEG-fMRI data

3.1. Simultaneous EEG-fMRI datasets

The contributions of this work are assessed against two distinct neuroimaging datasets with simultaneous EEG-fMRI recordings: i) NODDI dataset with recordings conducted under resting states; and ii) Oddball dataset with stimuli-based recordings. These contrasting settings offer the possibility to acquire a comprehensive understanding on the ability to synthesize fMRI from EEG under different protocols.

NODDI Dataset. NODDI dataset [16, 17] contains 17 individuals (11 males, 6 females) with average age 32.84 ± 8.13 years. 10 out of the 17 individuals are considered, due to corrupted views. Simultaneous EEG-fMRI recordings of resting state with eyes open (fixating a point) were acquired. Subjects were told to stay still on a vaccum cushion during scanning. The fMRI imaging acquisition was done based on a T2-weighted gradient-echo EPI sequence with: 300 volumes, TR of 2160 milliseconds (ms), TE of 30 ms, 30 slices with 3.0 millimeters (mm) (1 mm gap), voxel size of $3.3 \times 3.3 \times 4.0$ mm and a field of view of $210 \times 210 \times 120$ mm. The EEG imaging was acquired during the MRI scan with a 64-channel-MR-compatible electrode cap at 1000 Hz. The electrodes were setup according to the modified combinatorial nomenclature, referenced to the FCz electrode. An electrocardiogram (ECG) was recorded, and the EEG and MR scanner clocks were synchronised. The dataset is available for download by its original source at `https://osf.io/94c5t/`.

Auditory and visual Oddball Dataset. The Oddball simultaneous EEG-fMRI dataset [51, 52, 14] contains visual and auditory stimuli based recordings, each 340 seconds long. In total, there are 17 individuals out of which only 14 were considered, due to corrupted views. The acquisition was made with a 3T Philips Achieva MR Scanner with: single channel send and receive head coil, EPI sequence, 170 TRs per run with a TR of 2000 ms and 25 ms TE, 170 TRs per run with a $3 \times 3 \times 4$ mm voxel size and 32 slices with no slice gap. For a more detailed description of the dataset please refer to [51]. The dataset is available for download in its original source at `https://legacy.openfmri.org/dataset/ds000116/`.

3.2. Pairing EEG and fMRI: data setup

EEG and BOLD data preprocessing was performed in accordance with Deligianni et al. [16] and Walz et al. [51] for the NODDI and Oddball datasets. In addition to the original preprocessing and Lewis et al. [33] principles based on the observation that the distribution of fMRI values on the data at hand follow a lognormal distribution, we decided to log scale the fMRI values as well. The fMRI signal was downsampled (using the nilearn python library [1]) by a factor of 3 due to its fine resolution (i.e. number of voxels). The Short-Time Fourier Transform (STFT) was computed on the EEG signal, due to the variation of frequency intensities being correlated with the BOLD signal [44]. The STFT was taken with a window of 2 seconds and the frequency resolution placed according to the frequency sampling of the corresponding dataset.

According to Liao et al. [34], it is estimated that the neuronal activity is reflected in the BOLD signal with a delay of $s \approx [5.4, 6]$ seconds. Pairs between EEG and BOLD should have a shift of s seconds, such that at time t_{EEG} then the corresponding BOLD pair starts at time $t_{BOLD} (= t_{EEG} + s)$. In addition, there is the need to specify a time window big enough to adequately decode the lower frequency bands from the EEG signal. The interval can further

impact the number of features at input and output, making the problem much more difficult. This balance is extremely important as the network should not be forced to learn these properties.

The datasets described in this section contain EEG and fMRI recordings lengthy enough to be divided into partitions of 25.2 seconds each. In order for the bold shift $s = 5.4$ to be emulated, both STFT EEG and BOLD signals were resampled to 1.8 seconds using the scipy Python library [28].

4. Proposed EEG to fMRI approach

This section introduces the proposed approaches for EEG to fMRI synthesis (Section 4.1), their hyperparameterization (Section 4.2) and evaluation (Section 4.3).

4.1. Proposed models

Pairwise learning [36] is considered within the proposed approaches, whereby positive and negative pairs of EEG-fMRI recordings are fed as input to guide the learning. The positive pairs correspond to the EEG and BOLD starting at t_{EEG} and t_{BOLD}, respectively. Each positive pair of EEG and BOLD belong to the same individual and the same recording session. In contrast, the negative pairs are all the combinations of EEG and BOLD intances, that verify the following conditions: $t_{EEG} \neq t_{BOLD} + s$ and the individual corresponding to the EEG instance is different from the individual corresponding to the BOLD instance.

The proposed models have traits based on the AE, VAE and β-VAE, where instead of indirectly optimizing the parameters of a distribution (as it is the case for the VAE and β-VAE) the parameters being optimized are only the learnable network parameters (e.g. weights, biases, etc). In particular, the extended version of β-VAE considers reconstruction loss at the output level (similarly to the VAE and AE). This variant is subsequently reintroduced in the form of a linear combination with a distance loss (e.g. Contrastive Loss [13]) at the midlayer level (as done in β-VAE).

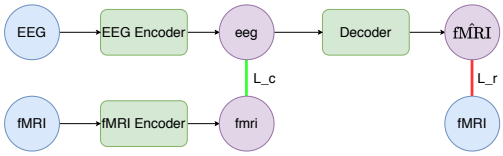


Fig. 1. Pipeline of EEG to fMRI synthesis.

Figure 1 depicts the proposed EEG-to-fMRI synthesis pipeline taking into account positive and negative imaging pairs. In accordance with pairwise learning principles, fMRI Encoder is not consider at testing time. The number of layers, N_L, along the pipeline components is placed by Neural Architecture Search (NAS) [18]. The EEG Encoder and the fMRI Encoder are both trained with the same loss, which varies depending on the training procedure. The Decoder is trained with a loss drawn from its output.

Four classes of networks are proposed: **Linear Combination** (Section 4.1), **AE Baseline** (Section 4.1), **Adversarial** (Section 4.1) and **Top-k Ranking** (Section 4.1). Section 4.1 introduces a technique that is capable of capturing temporal patterns at the encoder level, which is used by all the architectures.

Linear Combination (LCOMB)

The reconstruction loss, L_r, is represented by the **Euclidean Per Volume**, $L_{EPV}(\text{fMRI}, \hat{\text{fMRI}})$,

$$\frac{\sum_{i=0}^{N_{volumes}} \frac{\sqrt{\sum_{v=0}^{N_{voxels}} (\text{fMRI}_{i,v} - \hat{\text{fMRI}}_{i,v})^2}}{N_{voxels}}}{N_{volumes}}. \quad (1)$$

By minimizing this loss function, one converges to an optimal synthesized representation of an fMRI signal from the paired EEG signal. The choice of the Euclidean Per Volume Loss over the Mean Absolute Error Loss was based on the first having lower magnitude values, which may have an impact in the gradients computation.

In addition, to the L_r being introduced at the output level, it is also reintroduced to the EEG Encoder and BOLD Encoder in a linear combination, L_e, with a Contrastive Loss, $L_c(W, Y, \text{EEG}, \text{fMRI})$,

$$Y{D_W}^2 + (1 - Y)\, max(0, m - D_W)^2, \quad (2)$$

$$L_e = \theta L_c + (1 - \theta) L_r. \quad (3)$$

Regarding L_c, (EEG,fMRI) is the input pair, Y=1 if EEG and $fMRI$ are positive pairs and 0 otherwise, D_W the distance between the predicted values of eeg and $fmri$, and m is the margin value of separation.

The Contrastive Loss function forces the neighbors to be pulled together and non-neighbors to be pushed apart. This loss uses a distance metric. The Mean Absolute Error is the chosen metric.

Encoders take into account not only the approximation of the eeg and $fmri$ signals (by mapping the encoder outputs, eeg and $fmri$ signals get closer in space for positive pairs), but also maintain the reconstruction properties of the signal when performing the mapping. Under this premise, Encoders have a loss, L_e, that is a linear combination (set by θ) of L_r and L_c.

AE Baseline (AE)

An AE model incorporating the architecture described in Figure 2 is developed to be used as a baseline. This architecture is the one used in a test phase to perform the transcription from the EEG signal to the synthesized fMRI signal. The AE is treated as a baseline and is included in the results (see Section 5).

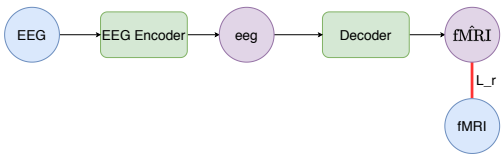


Fig. 2. Auto-Encoder architecture.

Adversarial (GAN and WGAN)

As discussed in Section 2, GANs have shown to be useful in cross-modalilty image synthesis, therefore this work also considers this type of deep learning approach for the synthesis of two functional neuroimaging modalities.

Although the L_r loss forces the model to learn the spatial properties of the original signal, this may not be enough to make the signal as close as possible to the original. An adversarial learning process introduces penalties given by a Discriminator and Generator components. If the Discriminator recognizes instances synthesized by the Generator, then a penalization is given to the Generator. On the other hand, if the Discriminator does not recognize those synthesized instances, a penalization is given to Discriminator itself. We consider two variations of this type of learning: the Minmax Entropy Loss (also known as Vanilla **GAN**),

$$\mathbb{E}_{x\sim p_{data}(x)}[log(D(X))] + \mathbb{E}_{x\sim p_z(z)}[log(1 - D(G(z)))], \tag{4}$$

and the Earth Mover Distance Loss (also known as **WGAN**),

$$\mathbb{E}_{x\sim p_{data}(x)}[D(X)] + \mathbb{E}_{x\sim p_z(z)}[1 - D(G(z))], \tag{5}$$

where $x \sim p_{data}(x)$ is an instance taken from the real instances and $x \sim p_z(z)$ is a sample taken from a distribution, subsequently decoded by G to $G(z)$.

Top-k Ranking

We found it pertinent to implement another baseline inspired on information retrieval top-k techniques. For that, this next variant concentrates on yet another variation at the encoding level, eeg. Instead of decoding directly the eeg activations, a linear combination of top-k eegs is given to the Decoder. This linear combination is a normalized vector of correlation values from the most correlated eeg instances. The EEG and fMRI Encoders are trained for a fixed number of epochs. Once this training session is over, each instance in the training set is compared to all the others, producing a rank of eeg instances for each eeg instance. Following, the top-k eeg instances are selected and a linear combination of these instances is computed,

$$top\_k\_eeg = \sum_{r=0}^{k} corr(eeg_r, eeg) \times eeg_r. \tag{6}$$

The Decoder then begins its training session with the inputs being the set of instances from the linear combinations and the targets the $fMRI$ associated with each EEG.

Temporal Encoding

Since, both modalities (EEG and fMRI) are functional neuroimaging techniques (i.e. contain temporal properties), there is a need for operations capable of capturing such properties. Most of the variants introduced so far, share a similar architecture to the one shown in Figure 1, only using Convolutional and Convolutional Transposed layers (Bai et al. [7] perform a thorough analysis on the performance of convolutional based networks against recurrent based networks, concluding that convolutions are preferred demonstrating effective memory properties). And although, these convolutions are performed on multiple dimensions (including the time dimension), the combination of these opreations along with recurrent layers is favorable [45, 35, 57, 40]. As such, in order to answer to the main question of this work, a recurrent component that provides time dependent encodings is introduced. Figure 3 shows the setting used to incorporate the rationale explained.

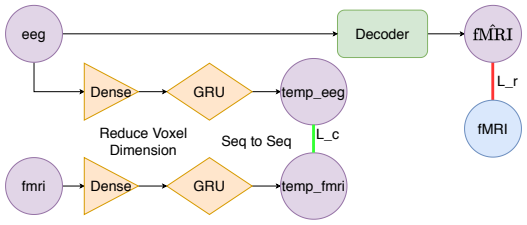


Fig. 3. Temporal Encoding integration in the architecture shown in Figure 1. At the $Decoder$ level, the convolutionally encoded eeg is fed as input and the L_r is computed against the ground truth $fMRI$. At the $Encoder$, the eeg and $fmri$ go through a fully connected layer that squeezes the Voxel dimension to 1, followed by a seq-to-seq Gated Recurrent Unit (GRU) layer that outputs $temp\_eeg$ and $temp\_fmri$, respectively. The pipeline from eeg and $fmri$ are independent, i.e. do not share weights. L_c is computed from the $temp\_eeg$ and $temp\_fmri$ temporal encoded activations.

4.2. Hyperparameter Tuning

Each of the variants described in Section 4.1 has different hyperparameters that are optimized. The tuning is done according to the performance of hyperparameters in a validation set. The hyperparameters that are common to all variants are: learning rate, weight regularization (L1 normalization) and batch size. In addition to those, **Linear Combination** needs the loss coefficient parameter, θ, to be optimized as well. On the other hand, the k value from the **Top-k Ranking** does not impact the performance and was fixed at $k = 5$. With this, as the **Linear Combination** is the procedure that has more hyperparameters to be tuned and containing all the hyperparameters from the other variants, it is chosen to be subjected to a NAS [18]. The Bayesian Optimization (BO) Algorithm is integrated in the search algorithm, therefore the hyperparameters are tuned along with the architecture. The search is done with 100 BO iterations for each depth of the network, stopping when there is no improvement (on the validation set) at a certain depth d against the optimal hyperparameters discovered at $d-1$. The optimal hyperparameters given are discovered at depth $d-1$.

The range of hyperparameters explored by the BO were (p_layer and n_layer represent the shapes of the previous and next layers, respectively): learning rate $\in [1\text{e}{-14}, 1\text{e}{-3}] \in \mathbb{R}$; L1 EEG Encoder regularization $\in [1\text{e}{-5}, 1\text{e}{-1}] \in \mathbb{R}$; L1 BOLD Encoder regularization $\in [1\text{e}{-5}, 1\text{e}{-1}] \in \mathbb{R}$; L1 Decoder regularization $\in [1\text{e}{-5}, 1\text{e}{-1}] \in \mathbb{R}$; loss coefficient, $\theta \in [0, 1] \in \mathbb{R}$; batch size $\in \{2, 4, 8, 16, 32, 64, 128\} \in \mathbb{N}$; EEG Encoder layer shape $\in [p\_layer, n\_layer] \in \mathbb{N}$; BOLD Encoder layer shape $\in [p\_layer, n\_layer] \in \mathbb{N}$; Decoder layer shape $\in [p\_layer, n\_layer] \in \mathbb{N}$. Dropout Layers [49] follow after each added layer with a probability of dropping connections $p = 0.5$.

4.3. Evaluation Metrics

To address the quality of the synthesized fMRI signals different metrics, exploring both temporal (BOLD) and spatial (fMRI) resolutions of the synthesized signals, are computed in addition to the Loss being minimized (L_{EPV}). The metrics computed are: **Log-Cosine Flattened Voxels (LCFV)**, **Cosine Flattened Voxels (CFV)**, **Euclidean Mean Voxels (EMV)**, **Euclidean Per Volume (EPV)** and **Mean Absolute Error (MAE)**. **LCFV** computes the Log-Cosine of a flattened time series from all the voxels, evaluating the temporal resolution,

$$log(1 - cosine(flatten(\text{BOLD}), flatten(\hat{\text{BOLD}}))). \tag{7}$$

CFV computes the Cosine of a flattened time series from all the voxels, evaluating the temporal resolution,

$$cosine(flatten(\text{BOLD}), flatten(\hat{\text{BOLD}})). \tag{8}$$

EMV computes the Mean of the Euclidean Distance of all the voxels, evaluating the temporal resolution,

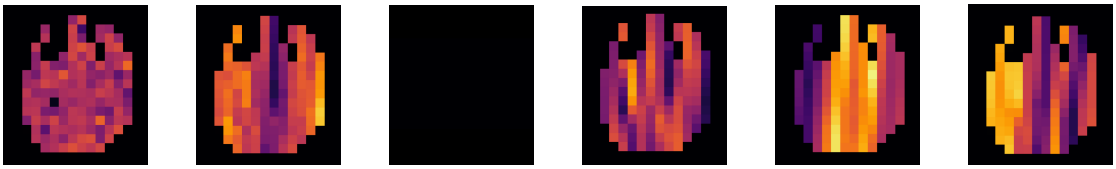
$$\frac{\sum_{i=0}^{N_{voxels}} euclidean(\text{BOLD}_i, \hat{\text{BOLD}}_i)}{N_{voxels}}. \tag{9}$$

EPV computes the Mean of the Euclidean of fMRI Volumes, evaluating the spatial resolution,

$$EPV = \frac{\sum_{i=0}^{N_{volumes}} \frac{\sum_{v=0}^{N_{voxels}} \sqrt{(\text{fMRI}_{i,v} - \hat{\text{fMRI}}_{i,v})^2}}{N_{voxels}}}{N_{volumes}}. \tag{10}$$

MAE computes the Mean Absolute Error of fMRI Volumes, evaluating the spatial resolution,

$$\frac{\sum_{i=0}^{N_{volumes}} \frac{\sum_{v=0}^{N_{voxels}} |\text{fMRI}_{i,v} - \hat{\text{fMRI}}_{i,v}|}{N_{voxels}}}{N_{volumes}}. \tag{11}$$



(a) fMRI **(b)** AE **(c)** GAN **(d)** WGAN **(e)** LCOMB **(f)** TOP-5

Fig. 4. Qualitative results on NODDI data, corresponding to the second individual in the test set (timestep at 26 seconds).

5. Results

The qualitative results are presented in Figure 4.The quantitative results gathered from the models for the test set (2 individuals and 4 individuals for the NODDI Dataset and Auditory and Visual Oddball Dataset, respectively) are presented in Tables 1 and 2. Each row in this table corresponds to the metrics described in Section 4.3, using the following numeration: (i) LCFV, (ii) CFV, (iii) EMV, (iv) EPV and (v) MAE.

As for the results gathered on the NODDI Dataset present in Table 1, AE had the best results in terms of the metrics evaluated (quantitative results), at the naked eye it seems to have good qualitative results (see Figure 4b). GAN and WGAN did not target the loss common to the other models (L_r), which possibly explains their inferiority under spatial metrics (iv) and (v). In regard to pattern based metrics (i) LCFV and (ii) CFV, although there is no model that has a clear superiority, LCOMB had an inferior performance. In contrast, given an Euclidean evaluation along the time axis (pattern based) the LCOMB model performs among the best along with AE. Spatial based metrics showed AE and LCOMB are preferable, this is justifiable with the loss targeted being a spatial loss, described by (iv) EPV.

As for the results gathered on the Auditory and Visual Oddball Dataset present in Table 2, AE, LCOMB and TOP-5 had the best results, in terms of spatial resolution, given by the (iv) and (v) metrics. As for pattern based metrics, there was no clear difference when looking at the (i) and (ii) metrics. On the other hand, GAN and WGAN underperformed according to (iii). Regarding the qualitative results, at the naked eye, the quality of the synthesized signals was very poor for this dataset , with WGAN having a small finer superiority.

Qualitative results of GAN and TOP-5 synthesis were extremely poor with nothing synthesized for most cases, as these methods produced non defined values due to exploding gradients (clipping the loss seemed to have no effect). GAN loss computes a logarithm having a bigger magnitude than the earth mover distance loss of WGAN. As for the TOP-5, it seems a linear combination of eegs at the encoder level does no resemble a good representation at that level. Overall, given that WGAN reaches the other models by the metrics evaluated and most importantly outperforms others in the (vi) metric on both datasets, which is shown in its qualitative results in Figure 4d, it is seen as the best fit candidate for this task, among the ones covered in this manuscript. Further, the need for more data was shown, as a simple model such as AE had good quantitative and qualitative results. This is due to its lower number of learnable parameters. Nonetheless, it can still be claimed that WGAN (while having a high number parameters, for amount of data available) is more suitable for this task.

6. Conclusion

This manuscript provides compelling empirical evidence for the feasibility of relating haemodynamics and electrophysiology in the human brain as given by the study of fMRI data synthesis from EEG data. To this end, we proposed approaches grounded on state-of-the-art principles of neural processing, including pairwise and adversarial learning. Of particular interest, the Contrastive Loss [21] trait for separating neighbours at the encoding level is useful for the targeted

	AE	GAN	WGAN	LCOMB	TOP-5
(i)	$\mathbf{-0.252 \pm 0.324}$	$\mathbf{-0.269 \pm 0.349}$	$\mathbf{-0.267 \pm 0.345}$	-0.071 ± 0.051	$\mathbf{-0.266 \pm 0.343}$
(ii)	$\mathbf{0.193 \pm 0.165}$	$\mathbf{0.202 \pm 0.169}$	$\mathbf{0.201 \pm 0.168}$	0.067 ± 0.047	$\mathbf{0.201 \pm 0.168}$
(iii)	$\mathbf{83.2 \pm 34.4}$	131 ± 40.7	102 ± 38.7	$\mathbf{87.7 \pm 32.8}$	111 ± 39.0
(iv)	$\mathbf{22.0 \pm 8.07}$	34.7 ± 9.99	27.1 ± 9.3	$\mathbf{23.2 \pm 7.57}$	29.4 ± 9.44
(v)	$\mathbf{505 \pm 241}$	888 ± 300	657 ± 279	$\mathbf{538 \pm 225}$	725 ± 282

Table 1. Quantitative results on the NODDI Dataset.

	AE	GAN	WGAN	LCOMB	TOP-5
(i)	-0.030 ± 0.034	-0.030 ± 0.034	$\mathbf{-0.031 \pm 0.034}$	$\mathbf{-0.031 \pm 0.034}$	$\mathbf{-0.031 \pm 0.034}$
(ii)	0.029 ± 0.032	$\mathbf{0.030 \pm 0.032}$	$\mathbf{0.030 \pm 0.032}$	$\mathbf{0.030 \pm 0.032}$	$\mathbf{0.030 \pm 0.032}$
(iii)	$\mathbf{111 \pm 5.11}$	156 ± 5.54	139 ± 5.52	$\mathbf{111 \pm 5.10}$	$\mathbf{112 \pm 5.24}$
(iv)	$\mathbf{29.5 \pm 1.32}$	41.7 ± 1.45	37.0 ± 1.44	$\mathbf{29.5 \pm 1.32}$	$\mathbf{29.9 \pm 1.36}$
(v)	$\mathbf{701 \pm 31.4}$	1024 ± 36.7	892 ± 35.4	$\mathbf{701 \pm 31.3}$	$\mathbf{710 \pm 32.3}$

Table 2. Quantitative results on Auditory and Visual Oddball Dataset.

synthesis task. The gathered results further motivate the relevance of upcoming contributions to the targeted synthesis task, offering solid baselines of performance.

EEG to fMRI synthesis task is expected to have major advances in the following decade, with broad applications in fields such as health, computer vision and neuroscience. Research on these tasks offers new ways of enriching brain imaging modalities, gaining further insights into the brain, and promoting long-lasting and less-expensive monitoring protocols.

Future work. The reverse transformation, fMRI-to-EEG, is also of high relevance to the community. Complementary cohort studies with simultaneous EEG and fMRI monitoring are being undertaken [3], untapping new possibilities. Alternative approaches combining alternative principles from signal processing and time series data analysis are also expected. Finally, the role of emerging state-of-the-art neural processing techniques, such as Neural ODEs [11], to this specific synthesis task is still unexplored.

Acknowledgments

This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) under PhD grant 2020.05762.BD to DC and INESC-ID plurianual (UIDB/50021/2020).

References

1. A Abraham et al. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 2014.
2. D Abramian et al. Generating fmri volumes from t1-weighted volumes using 3d cyclegan. *arXiv*, 2019.
3. R Abreu et al. Eeg-informed fmri: a review of data analysis methods. *Frontiers in human neuroscience*, 2018.
4. P Aggarwal et al. Accelerated fmri reconstruction using matrix completion with sparse recovery via split bregman. *Neurocomputing*, 2016.
5. M Arjovsky and. Wasserstein gan. *ArXiv*, 2017.
6. X Yi andothers. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 2019.
7. S Bai et al. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*, 2018.
8. A Ben-Cohen et al. Cross-modality synthesis from ct to pet using fcn and gan networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence*, 2019.
9. C Chang et al. Eeg correlates of time-varying bold functional connectivity. *Neuroimage*, 2013.
10. G Chao et al. Generating fmri-enriched acoustic vectors using a cross-modality adversarial network for emotion recognition. In *ACM ICMI*, 2018.
11. T Chen et al. Neural ordinary differential equations. In *NIPS*, 2018.
12. X Chen et al. Variational lossy autoencoder. *arXiv*, 2016.
13. S. Chopra et al. Learning a similarity metric discriminatively, with application to face verification. In *IEEE CVPR*, 2005.

14. B Conroy et al. Fast bootstrapping and permutation testing for assessing reproducibility and interpretability of multivariate fmri decoding models. *PloS one*, 2013.
15. C Cury et al. A sparse eeg-informed fmri model for hybrid eeg-fmri neurofeedback prediction. *bioRxiv*, 2019.
16. F Deligianni et al. Relating resting-state fmri and eeg whole-brain connectomes across frequency bands. *Frontiers in Neuroscience*, 2014.
17. F Deligianni et al. Noddi and tensor-based microstructural indices as predictors of functional connectivity. *PLOS ONE*, 2016.
18. T Elsken et al. Neural architecture search: A survey. *arXiv*, 2018.
19. A Fowle et al. Uses and abuses of the eeg in epilepsy. *Epilepsia*, 2000.
20. I Goodfellow et al. Generative adversarial nets. In *NIPS*, 2014.
21. R Hadsell et al. Dimensionality reduction by learning an invariant mapping. In *IEEE CVPR*, 2006.
22. X He et al. Ode-inspired network design for single image super-resolution. In *IEEE CVPR*, 2019.
23. Y He et al. Spatial–temporal dynamics of gesture–speech integration: a simultaneous eeg-fmri study. *Brain Structure and Function*, 2018.
24. I Higgins et al. beta-vae: Learning basic visual concepts with a constrained variational framework. In *ICLR*, 2017.
25. G Hinton et al. Reducing the dimensionality of data with neural networks. *science*, 2006.
26. A Hore et al. Image quality metrics: Psnr vs. ssim. In *ICPR*. IEEE, 2010.
27. Y Jiang et al. Targeting brain functions from the scalp: Transcranial brain atlas based on large-scale fmri data synthesis. *NeuroImage*, 2020.
28. E Jones et al. SciPy: Open source scientific tools for Python, 2001.
29. T Karras et al. A style-based generator architecture for generative adversarial networks. In *IEEE CVPR*, 2019.
30. D P. Kingma et al. Auto-encoding variational bayes. *CoRR*, 2013.
31. R Labounek et al. Eeg spatiospectral patterns and their link to fmri bold signal via variable hemodynamic response functions. *Journal of Neuroscience Methods*, 2019.
32. M Leite et al. Transfer function between eeg and bold signals of epileptic activity. 2013.
33. S Lewis et al. Logarithmic transformation for high-field bold fmri data. *Experimental brain research*, 2005.
34. C.H. Liao et al. Estimating the delay of the fmri response. *NeuroImage*, 2002.
35. M Lopez-Martin et al. Network traffic classifier with convolutional and recurrent neural networks for internet of things. *IEEE Access*, 2017.
36. L Mencía et al. Pairwise learning of multilabel classifications with perceptrons. In *IEEE IJCNN*, 2008.
37. M Mikulic. Mri units density by country 2017, 2019.
38. M Mirza et al. Conditional generative adversarial nets. *ArXiv*, 2014.
39. R Mosayebi et al. Correlated coupled matrix tensor factorization method for simultaneous eeg-fmri data fusion. *Biomedical Signal Processing and Control*, 2020.
40. S Mousavi et al. Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific reports*, 2019.
41. T Murta et al. Electrophysiological correlates of the bold signal for eeg-informed fmri. In *Human brain mapping*, 2015.
42. D Nie et al. Medical image synthesis with context-aware generative adversarial networks. In *MICCAI*, 2017.
43. G Ogbole et al. Survey of magnetic resonance imaging availability in west africa. *Pan African Medical Journal*, 2018.
44. G Portnova et al. Correlation of bold signal with linear and nonlinear patterns of eeg in resting state eeg-informed fmri. *Frontiers in human neuroscience*, 2018.
45. D Quang et al. Danq: a hybrid convolutional and recurrent deep neural network for quantifying the function of dna sequences. *Nucleic acids research*, 2016.
46. S E. Reed et al. Generative adversarial text to image synthesis. *ArXiv*, 2016.
47. M Rosa et al. Estimating the transfer function from neuronal activity to bold using simultaneous eeg-fmri. *Neuroimage*, 2010.
48. D Rumelhart et al. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*. 1986.
49. N Srivastava et al. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, 2014.
50. J S Turek et al. A semi-supervised method for multi-subject fmri functional alignment. In *ICASSP*. IEEE, 2017.
51. J Walz et al. Simultaneous eeg-fmri reveals temporal evolution of coupling between supramodal cortical attention networks and the brainstem. *Journal of Neuroscience*, 2013.
52. J Walz et al. Simultaneous eeg–fmri reveals a temporal cascade of task-related and default-mode activations during a simple target detection task. *Neuroimage*, 2014.
53. H Wei et al. Bayesian fusion and multimodal dcm for eeg and fmri. *NeuroImage*, 2020.
54. J Wolterink et al. Deep mr to ct synthesis using unpaired data. In *Simulation and Synthesis in Medical Imaging*, 2017.
55. Z Zhu et al. Semantic unification modulates n400 and bold signal change in the brain: A simultaneous eeg-fmri study. *Journal of Neurolinguistics*, 2019.
56. P Zhuang et al. Fmri data augmentation via synthesis. In *ISBI*. IEEE, 2019.
57. H Zuo et al. Combining convolutional and recurrent neural networks for human skin detection. *IEEE Signal Processing Letters*, 2017.