The Promise and Limitations of Sinusoidal Representation Networks for EEG-fMRI Translation

Ziyuan Xu
Department of Computer Science
Vanderbilt University
Nashville, TN 37235
ziyuan.xu@vanderbilt.edu

Jiliang Li
Department of Computer Science
Vanderbilt University
Nashville, TN 37235
jiliang.li@vanderbilt.edu

Abstract

In modern neuroscience, functional magnetic resonance imaging (fMRI) has been a crucial and irreplaceable tool that provides a non-invasive window into the dynamics of whole-brain activity. Nevertheless, fMRI is limited by hemodynamic blurring as well as high cost, immobility, and incompatibility with metal implants. Electroencephalography (EEG) is complementary to fMRI and can directly record the cortical electrical activity at high temporal resolution, but has more limited spatial resolution and is unable to recover information about deep subcortical brain structures. The ability to obtain fMRI information from EEG would enable cost-effective, naturalistic imaging across a wider set of brain regions. Further, beyond augmenting the capabilities of EEG, cross-modality models would facilitate the interpretation of fMRI signals. However, as both EEG and fMRI are high-dimensional and prone to noise and artifacts, it is currently challenging to model fMRI from EEG. Indeed, although correlations between these two modalities have been widely investigated, few studies have successfully used EEG to directly reconstruct fMRI time series. To address this challenge, we propose a novel architecture that can predict fMRI signals directly from multi-channel EEG without explicit feature engineering. Our model achieves this by implementing a Sinusoidal Representation Network (SIREN) to learn frequency information in brain dynamics from EEG, which serves as the input to a subsequent encoder-decoder to effectively reconstruct the fMRI signal from a specific brain region. We demonstrate that our proposed SIREN-based model achieves state-of-the-art performance on the Eyes Open - Eyes Closed dataset. However, the success of the model does not generalize well to a different dataset, the Resing-State dataset. The work highlights the potential of leveraging periodic activation functions in deep neural networks for certain neuroimaging tasks, as well as the approach's limitations with respect to generalizability to different, more noisy data.

1 Introduction

As the two most frequently used non-invasive neuroimaging modalities, functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) play essential roles in advancing our understanding of the human brain and its complexities. The simultaneous recording and analysis of these two modalities have also gained substantial attention, owing to their complementary strengths and physiological views of brain functioning [Huster et al., 2012]. Though fMRI helps to more precisely locate the source of neural activities, it suffers from hemodynamic blurring that introduces uncertainty in the timing of neural activity, which can be more precisely achieved by EEG. Moreover, the high cost and incompatibility with metal implants also hinder MRI from some application scenarios. In particular, MRI units are scare globally, with an estimate of 0.24 units per

Preprint. Under review.

million people in West African countries [Ogbole et al., 2018]. In contrast, electroencephalography is in-expensive, non-invasive, safe, and rarely restricted by the extend of recordings [Fowle and Binnie, 2000]. Consequently, the promise of EEG to fMRI synthesis can open up new avenues to perform more affordable, portable, and long-lasting brain activity proctoring [Calhas and Henriques, 2020]. Thus, in this paper, we take another step in this direction and propose a novel approach to improve existing EEG to fMRI translation, leveraging the Sinusoidal Representation Network (SIREN) to conduct end-to-end translation without explicit feature engineering.

A number of existing studies have tried to build bridges between the two modalities, such as through correlational or machine learning frameworks [Chang and Chen, 2021, Meir-Hasson et al., 2014]. However, factors such as the high dimensionality, complexity, and low signal-to-noise ratio of both EEG and fMRI data create challenges for accurately modeling the translation between EEG and fMRI and for directly reconstructing fMRI time series from EEG.

The rapid development of deep learning in the past decades has facilitated multimodal learning and cross-modal prediction. Recently, Kovalev et al. [2022] proposed a deep learning framework to predict subcortical fMRI signals from 30-channel EEG signals. While this work contributed a major step toward the feasibility of EEG-fMRI translation, the median performances were correlations on the order of 0.3-0.5, leaving room for performance improvement. One potential limitation of this model is that it is not explicitly designed to represent frequency features from EEG; yet, prior work has noted that resolving spectral information in the EEG may yield better prediction accuracy than that obtained without considering the frequency-band distribution of EEG [Meir-Hasson et al., 2014, Kovalev et al., 2022].

In recent years, the use of periodic activation functions, specifically the sine function, has been gaining traction because of its excellent performance in implicit neural representations [Sitzmann et al., 2020]. By mapping the periodic patterns and continuous functions underlying a signal, this sinusoidal representation network (SIREN) is proving to be well-suited for representing complex natural signals and their derivatives Sitzmann et al. [2020], Bai et al. [2023]. Therefore, we hypothesize that SIREN would also be effective in learning frequency information in EEG time series, and could thereby significantly improve the prediction of fMRI from EEG. In this paper, we propose a novel framework that embeds SIREN into a deep learning model with an EEG feature encoder and fMRI decoder, aiming to reconstruct fMRI signals directly from EEG raw data without explicit feature engineering.

We tested our SIREN-based network on two indicative datasets: (1) Eyes Open - Eyes Closed and (2) Resting-State. We found that the SIREN-based network outperforms state-of-the-art models on the former dataset. However, on the Resting-State dataset, the SIREN-based model is suboptimal compared to a CNN-Transformer model [Peh et al., 2022]. Moreover, we found a lack of existing models that work well for EEG-fMRI translation when working with subcortical regions or noisy data. Our results provide insights into the feasibility of SIREN for certain well-defined EEG-fMRI translation tasks, as well as its limited generalizable performance across different, more challenging tasks.

We claim the following findings and contributions:

- We propose a novel neural network that leverages sinusoidal representation networks (SIREN) to predict fMRI signal from EEG.
- We conduct thorough evaluations on two indicative datasets to demonstrate SIREN's ability to increase model performance on certain simple tasks. We also unveil SIREN's, as well as ML models' general limitations when conducting EEG-fMRI translation in more challenging tasks (e.g. Resting-State EEG-fMRI translations).

2 Background and Related Work

2.1 EEG, fMRI, with Machine Learning with EEG/fMRI data

Electroencephalography (EEG) and Functional Magnetic Resonance Imaging (fMRI) are pivotal in the exploration of neural mechanisms and brain function. EEG, a non-invasive neurophysiological monitoring method, measures the electrical activity of the brain via electrodes placed on the scalp. It is highly valued for its superior temporal resolution, providing insights into neuronal dynamics on a millisecond scale [Niedermeyer and da Silva, 2005]. EEG data typically consists of multi-channel time series, capturing voltage fluctuations resulting from ionic current flows within neurons, ideally suited for dynamic brain state analysis.

Conversely, fMRI offers a different perspective by detecting changes associated with blood flow, thus reflecting neuronal activation indirectly. This technique stands out for its spatial resolution, enabling detailed anatomical mapping of brain activity. fMRI data is often represented as 3D volumetric images (4D when including the time dimension), which provide a spatially detailed map of brain activity over time. Together, EEG and fMRI are extensively employed in cognitive neuroscience and clinical research, offering complementary data that enhance our understanding of brain functions and pathologies [Logothetis, 2008]. Both aforementioned modalities of EEG and fMRI data are widely used in machine learning applications, where EEG data serves as high-resolution temporal sequences and fMRI as spatially complex volumetric datasets, aiding in diverse computational models for neurological diagnosis and cognitive studies.

2.2 EEG to fMRI Translation

The endeavor to translate EEG to fMRI remains a relatively underexplored area of research. One of the pioneering efforts in this domain was by Meir-Hasson et al. [2014], who employed ridge regression models for EEG to fMRI translation. However, their methodology was relatively straightforward and limited in complexity. More recent efforts by Calhas and Henriques [2020] have demonstrated the feasibility of this translation, albeit using simple models and datasets that resulted in modest performance levels. Subsequent work by Kovalev et al. [2022] also aimed to enhance this translation but reported only moderate success, with median performance correlations ranging between 0.3 and 0.5. We hypothesize that these efforts were partly constrained because they did not focus on representing frequency features from EEG effectively. Research by Rosa et al. [2010] suggests that a more detailed resolution of spectral information and consideration of the frequency-band distribution in EEG could significantly improve prediction accuracy. Building on this insight, our work introduces the use of the Sinusoidal Representation Network (SIREN), which is designed to more effectively capture and leverage frequency information from EEG data.

With respect to experimental settings, EEG to fMRI translation generally relies on simultaneous EEG and fMRI studies [Leite et al., 2013]. In such studies, EEG and fMRI data is collected simultaneously from the same human subjects during some designated task [He et al., 2018, Chang et al., 2013]. As such, each EEG data sample is paired with another fMRI sample.

Amongst the previous simultaneous EEG and fMRI studies, this work has been motivated byChang et al. [2013]. In particular, the study collected simultaneous EEG and fMRI data from ten healthy adults under resting conditions to investigate if changes in EEG power, particularly in the alpha and theta frequency bands, align with temporal variations in fMRI-based functional connectivity networks. These networks were delineated using a functional atlas developed from a group-level independent component analysis of resting-state fMRI data. The study observed that decreases in alpha power and increases in theta power were linked with increases in functional connectivity, notably showing positive correlations with alpha power in the thalamus and dorsal anterior cingulate cortex. Despite focusing on specific spectral bands and connectivity maps, these results encourage further exploration into translating EEG findings into fMRI analyses, demonstrating the potential for establishing meaningful correlations between these two modalities.

2.3 Sinusoidal Representation Network (SIREN)

The Sinusoidal Representation Network (SIREN) is an innovative neural network architecture that utilizes sinusoidal activation functions as opposed to traditional rectified linear units (ReLU). Introduced by Sitzmann et al. [2020], SIREN has shown remarkable efficacy in modeling and reconstructing complex natural signals and their derivatives, such as audio signals and images. The key strength of SIREN lies in its ability to capture the inherent periodicity of the input data, making it highly suitable for applications involving waveform data.

In the context of EEG-fMRI translation, the ability of SIREN to handle high-frequency details in data makes it exceptionally relevant. EEG signals are inherently oscillatory with significant information contained in their spectral distributions across various frequency bands such as delta, theta, alpha, and beta. Traditional neural network architectures often struggle with accurately modeling such spectral

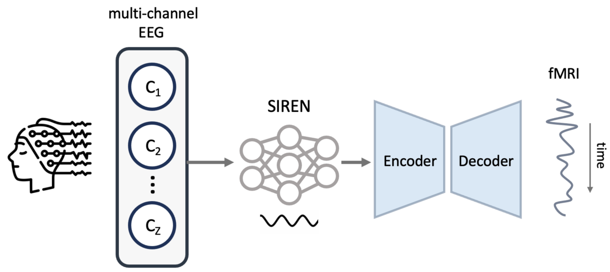

Figure 1: An overview of the model architecture of our proposed SIREN-based network.

information because they typically rely on non-periodic activation functions which are less sensitive to the oscillatory nature of EEG signals.

SIREN, by contrast, leverages its sinusoidal activations to directly encode and process the frequencies present in EEG data. This capability allows for a detailed resolution of the spectral information and ensures that the frequency-band distribution within the EEG is effectively represented and utilized during the translation process to fMRI. This makes SIREN a promising approach for EEG to fMRI translation. By adopting SIREN, we aim to enhance the precision and reliability of neuroimaging translations, capturing nuanced brain activities with greater fidelity.

3 Approach

In this section, we present the architectures of two different models used in this study: (1) our proposed SIREN-based neural network and (2) a CNN-based neural network with Transformer Encoder. The CNN-Transformer model is used as a baseline to assess our proposed SIREN-based neural network.

3.1 SIREN-based Neural Network

As illustrated in Figure 3.1, our proposed model architecture comprises a SIREN layer, followed by an encoder-decoder structure to perform the end-to-end EEG to fMRI translation. We will now present each of these components in detail.

3.1.1 Sinusoidal Representation Network (SIREN)

Inspired by the works of Sitzmann et al. [2020] and Kazemi et al. [2019], we propose a framework that leverages the periodic activation function in each layer of a multilayer perceptron (MLP), i.e., the Sine layer, to extract EEG features, thereby learning common representations between EEG and fMRI without explicit feature engineering:

$$\Phi(x) = W_n(\phi_{n-1} \circ \phi_{n-2} \circ \cdots \circ \phi_0)(x) + b_n, x_i \to \phi_i(x_i) = sin(W_i x_i + b_i) \tag{1}$$

where $\phi_i : \mathbb{R}^{M_i} \to \mathbb{R}^{N_i}$ is the i^{th} layer of the network. It comprises the affine transformation defined by the weight matrix $W_i \in \mathbb{R}^{N_i \times M_i}$ and the bias $b_i \in \mathbf{R}^{N_i}$ applied on the input $x_i \in \mathbb{R}^{M_i}$, followed by the sine non-linearity applied to each component of the resulting vector. This section consists of an input layer and K hidden layers ($K = 1$ in this study), followed by a linear projection layer.

3.1.2 Feature Encoder and Decover

The SIREN output is then sent into the encoder and decoder blocks to predict the fMRI signals from each ROI. The encoder consists of N encoder blocks following the architecture of the wav2vec 2.0 model [Baevski et al., 2020], followed by the dropout block [Srivastava et al., 2014] (N = 4, dropout rate = 0.3 for all layers in this study). Each encoder block has a down-sampling operation, i.e., max pooling in this analysis, which efficiently increases the receptive field while retaining important information. The decoder comprises the same symmetric building blocks and up-samples the latent space features to produce the fMRI signal of a certain ROI.

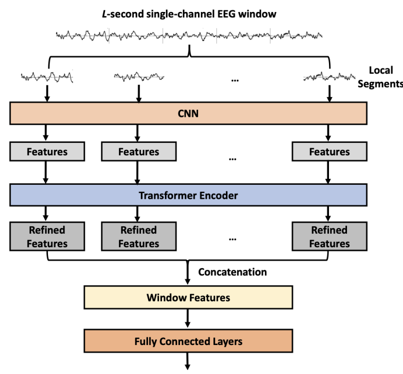

Figure 2: An overview of the CNN-based model with Transformer Encoder

3.1.3 Learning Objective

In this training paradigm, the model optimizes the linear combination of two losses, the mean squared error (MSE) loss and the correlation loss:

$$Loss = L_{mse} + \alpha \cdot L_{corr} \tag{2}$$

Here, α is a hyperparameter we tune, and for the correlation loss, we use the negative Pearson correlation coefficient:

$$L_{corr} = -r(y, \hat{y}) = -\frac{\sum_{i=1}^{n}(y_i - \bar{y_l})(\hat{y}_i - \bar{\hat{y}_l})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y_l})^2(\hat{y}_i - \bar{\hat{y}_l})^2}} \tag{3}$$

3.2 CNN with Transformer Encoder

In addition to the SIREN-based network presented above, we also use another CNN-Transformer model introduced in [Peh et al., 2022] to benchmark our SIREN-based network's performance on our dataset. In particular, we first segment the EEG signals in a window (20.48 seconds) into 32 segments. Then, a Short-time Fourier transform (STFT) is applied to each segment to get a two-dimensional (time x frequency) representation of the signal. 2D convolutional kernel of size 3 is applied to each signal segment and the extracted features are sent to the Transformer Encoder to get refined features. At the end, all the refined features are flattened as a long 1d vector and sent to a MLP to output the 2048 points as the predicted fMRI signal. An overview of the CNN-Transformer architecture can be found in Figure 3.2.

The CNN-Transformer model is trained with the same learning objective as the SIREN-based Neural Network.

4 Experimental Design

As stated in Section 1, we hypothesize that SIREN would also be effective in learning frequency information in EEG time series, and could thereby significantly improve the prediction of fMRI from EEG. In this section, we present the datasets and evaluation design used to test our hypothesis. We also discuss the implementation details for our experiments in Section C.

4.1 Datasets

We employ two different datasets for the training and evaluations of our approach: (1) Eyes Open - Eyes Closed and (2) Resting-State.

Eyes Open - Eyes Closed Simultaneous "Eyes Open – Eyes Closed" EEG-fMRI data was collected from 8 subjects [van der Meer et al., 2016]. The subjects performed a simple eyes open-eyes closed task in 6 consecutive conditions chosen to isolate different kinds of MR-related artifacts. Please see Section A for more information on dataset acquisition and preprocessing.

Resting-State EEG and fMRI data were collected simultaneously from 25 healthy volunteers (age = 34.3 ± 16.0 years, 14 females) in two resting state sessions each lasting 20 minutes. During the rsfMRI scans, subjects rested passively with eyes closed. Please see Section B for more information on dataset acquisition and preprocessing.

4.2 Evaluation Design

To test our hypothesis, we consider the following research questions:

- RQ1: Does the incorporation of SIREN improve existing models' performance on an indicative EEG-fMRI translation task (i.e. on the Eyes Open - Eyes Closed dataset)?

- RQ2: Can the SIREN-based model be used as a universal translator for general-purpose EEG-fMRI tasks?

RQ1: To answer RQ1, we compare our proposed SIREN-based network against the state-of-art EEG-fMRI prediction model BEIRA Kovalev et al. [2022] and ridge regression [Meir-Hasson et al., 2014] on predicting subcortical fMRI signals in the Eyes Open - Eyes Closed dataset.

RQ2: To answer RQ2, we consider the SIREN Neural Network's performance on another dataset, the Resting-State dataset. Specifcally, we compare the model's performance against a CNN-Transformer network in predicting both subcortical and cortical fMRI signals to provide insights into the SIREN Neural Network's generalizability across various EEG-fMRI translation tasks.

5 Results

5.1 RQ1: SIREN on the Eyes Open - Eyes Closed Dataset

Table 1 shows the quantitative performance of our model in all subjects and in one selected subject (the best case). The results show that our model outperforms the current state-of-art deep learning model BERIA. Figure 5.1 depicts the predicted and ground-truth fMRI signals for the single (best-case) subject, and indicates that our model is able to correctly capture most of the phase information of subcortical fMRI signals

5.2 RQ2: SIREN on the Resting-State Dataset

As shown in Figure 5.2 and Figure 5.2, SIREN on Resting-State does not perform well on both cortical and subcortical ROIs. For the 2 chosen cortical ROIs, SIREN does not achieve comparable performance as the CNN-Transformer model. It is also noticeable that neither the SIREN nor the CNN-Transformer models can predict the Resting-State fMRI signals in the subcortical ROIs.

Table 1: Experimental results using the Eyes Open - Eyes Close dataset. Comparison of our proposed SIREN model with existing models for subcortical fMRI time series prediction from EEG. Pallidum, caudate, putamen, accumbens are subcortical regions of the brain.

All subjects	Pallidum	Caudate	Putamen	Accumbens	Average
Ridge regression	0.05 ± 0.15	0.34 ± 0.14	0.38 ± 0.24	0.26 ± 0.10	0.24 ± 0.11
BEIR	0.37 ± 0.04	0.47 ± 0.14	0.48 ± 0.16	0.42 ± 0.05	0.44 ± 0.05
Ours	$\mathbf{0.43 \pm 0.15}$	$\mathbf{0.49 \pm 0.12}$	$\mathbf{0.51 \pm 0.14}$	$\mathbf{0.43 \pm 0.13}$	$\mathbf{0.47 \pm 0.04}$
One subject	Pallidum	Caudate	Putamen	Accumbens	Average
Ridge regression	0.01	0.40	0.48	0.10	0.25 ± 0.20
BEIR	0.39	0.62	0.55	0.37	0.48 ± 0.02
Ours	**0.635**	**0.70**	**0.62**	**0.50**	$\mathbf{0.61 \pm 0.09}$

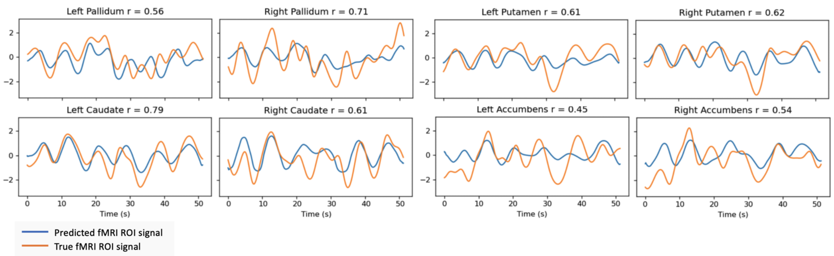

Figure 3: Experimental results using the Eyes Open Eyes Close dataset. Real and predicted fMRI time series of different subcortical ROIs from one subject. Orange line: real signal, blue line: predicted signal.

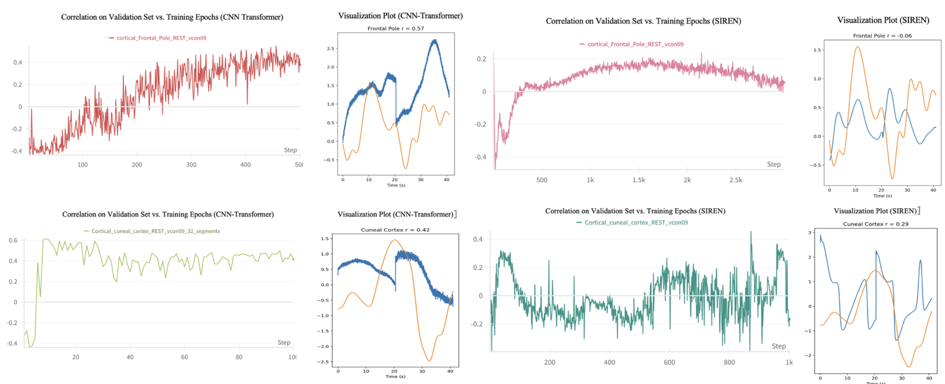

Figure 4: Experimental results using the Resting-State scan of one subject. First Row: Correlation on validation set vs. training epochs for both CNN-Transformer and SIREN on cortical region 'Frontal Pole'. Visualization plot of both real (orange line) and predicted (blue line) fMRI time series. Second Row: Correlation on validation set vs. training epochs for both CNN-Transformer and SIREN on cortical region 'Cuneal Cortex'. Visualization plot of both real (orange line) and predicted (blue line) fMRI time series.

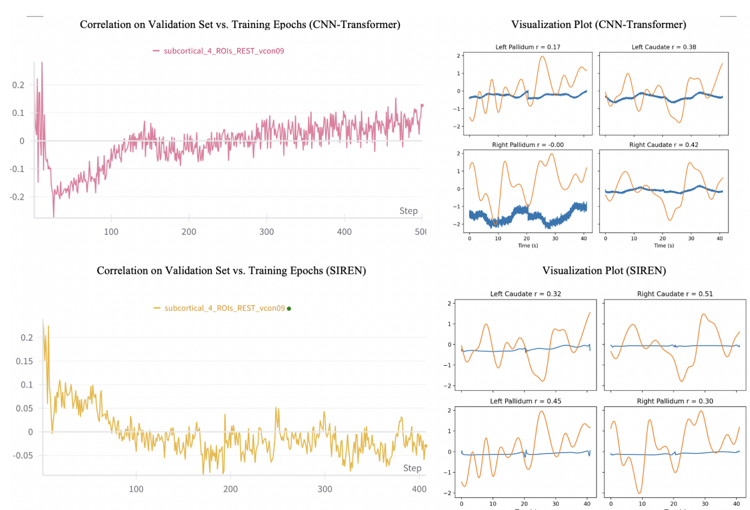

Figure 5: Experimental results using the Resting-State scan of one subject. Left column: Correlation on validation set vs. training epochs for both CNN-Transformer and SIREN. Right column: Real and predicted fMRI time series. Orange line: real signal, blue line: predicted signal. The 4 subcortical ROIs are left and right Pallidum, and left and right Caudate.

6 Discussion

The SIREN model successfully reconstructs several subcortical fMRI signals from EEG time series in the Eyes Open Eyes Closed dataset and significantly improves the prediction accuracy compared with existing models. However, the SIREN model cannot generalize well to the Resting-State dataset in both subcortical and cortical brain ROIs. In fact, a CNN-Transformer model achieves superior performance on the resting-state dataset compared to the SIREN model, indicating that such model architecture may be a more suitable model for learning the EEG-fMRI mapping in more noisy resting-state data.

We hypothesize the following factors contribute to the observed phenomena: (1) the Eyes Open - Eyes Closed dataset has apparent patterns in both EEG and fMRI as participants are asked to open and close their eyes at a regular frequency. (2) Resting-State scan does not require participants to perform tasks, which makes both EEG and fMRI signals consist of more noisy data without apparent patterns. (3) Subcortical ROI is deep inside the brain, making it hard to detect its electrical signals using EEG that is put on participants' scalp - so maybe, machine learning is fundamentally limited by the EEG's inability to capture the nuanced electrical waves in subcoritcal regions.

We contain our work within a novel framework that uses periodic activation function in deep neural networks to learn representations of functional neuroimaging data. Future work can also promisingly examine the CNN-Transformer performance on more brain areas, assess performance on different task conditions and patient populations, and attempted a contrastive learning approach between EEG and fMRI representations, as suggested by Prof. Kolouri.

7 Conclusion

The promise of synthesizing fMRI data from EEG signals provides a new avenue to perform more effective and non-invasive brain activity proctoring. In this paper, we propose to incorporate Sinusoidal Representation Network (SIREN) for the EEG-fMRI signal translation task, aiming to leverage the periodic activation functions to better capture the frequency-band distribution information in EEG. On the Eyes Open - Eyes Closed dataset, we demonstrate that our SIREN-based network surpasses state-of-the-art models at EEG-fMRI translation, demonstrating SIREN's promise in leveraging the frequency-band distribution in EEG. On the other hand, we find that the proposed SIREN-based network is overshadowed by a CNN-Transformer network on the Resting-State dataset. This result opens up new areas of future discussion regarding the limitations of SIREN across tasks and datasets, as well as EEG data's fundamental limitations in capturing subcortical signals.

References

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in neural information processing systems*, 33:12449–12460, 2020.

Wenjun Bai, Okito Yamashita, and Junichiro Yoshimoto. Learning task-agnostic and interpretable subsequence-based representation of time series and its applications in fmri analysis. *Neural Networks*, 163:327–340, 2023.

David Calhas and Rui Henriques. Eeg to fmri synthesis: Is deep learning a candidate? *arXiv preprint arXiv:2009.14133*, 2020.

Catie Chang and Jingyuan E Chen. Multimodal eeg-fmri: advancing insight into large-scale human brain dynamics. *Current opinion in biomedical engineering*, 18:100279, 2021.

Catie Chang, Zhongming Liu, Michael C Chen, Xiao Liu, and Jeff H Duyn. Eeg correlates of time-varying bold functional connectivity. *Neuroimage*, 72:227–236, 2013.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

Adrian J Fowle and Colin D Binnie. Uses and abuses of the eeg in epilepsy. *Epilepsia*, 41:S10–S18, 2000.

Yifei He, Miriam Steines, Jens Sommer, Helge Gebhardt, Arne Nagels, Gebhard Sammer, Tilo Kircher, and Benjamin Straube. Spatial–temporal dynamics of gesture–speech integration: a simultaneous eeg-fmri study. *Brain Structure and Function*, 223:3073–3089, 2018.

René J Huster, Stefan Debener, Tom Eichele, and Christoph S Herrmann. Methods for simultaneous eeg-fmri: an introductory review. *Journal of Neuroscience*, 32(18):6053–6060, 2012.

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019.

Alexander Kovalev, Ilia Mikheev, and Alexei Ossadtchi. fmri from eeg is only deep learning away: the use of interpretable dl to unravel eeg-fmri relationships. *arXiv preprint arXiv:2211.02024*, 2022.

Marco Leite, Alberto Leal, and Patrícia Figueiredo. Transfer function between eeg and bold signals of epileptic activity. *Frontiers in neurology*, 4:38518, 2013.

Nikos K Logothetis. What we can do and what we cannot do with fmri. *Nature*, 453(7197):869–878, 2008.

Yehudit Meir-Hasson, Sivan Kinreich, Ilana Podlipsky, Talma Hendler, and Nathan Intrator. An eeg finger-print of fmri deep regional activation. *Neuroimage*, 102:128–141, 2014.

Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

Godwin Inalegwu Ogbole, Adekunle Olakunle Adeyomoye, Augustina Badu-Peprah, Yaw Mensah, and Donald Amasike Nzeh. Survey of magnetic resonance imaging availability in west africa. *Pan African Medical Journal*, 30(1), 2018.

Wei Yan Peh, Yuanyuan Yao, and Justin Dauwels. Transformer convolutional neural networks for automated artifact detection in scalp eeg. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 3599–3602. IEEE, 2022.

Maria J Rosa, James Kilner, Felix Blankenburg, Oliver Josephs, and W Penny. Estimating the transfer function from neuronal activity to bold using simultaneous eeg-fmri. *Neuroimage*, 49(2): 1496–1509, 2010.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Johan van der Meer, André Pampel, Eus van Someren, Jennifer Ramautar, Ysbrand van der Werf, German Gomez-Herrero, Jöran Lepsien, Lydia Hellrung, Hermann Hinrichs, Harald Möller, et al. "eyes open–eyes closed" eeg/fmri data set including dedicated "carbon wire loop" motion detection channels. *Data in brief*, 7:990–994, 2016.

A Eyes Open - Eyes Closed Dataset Details

A.1 Data Acquisition

The fMRI acquisition parameters were: TR=1.95s for the first 4 subjects, TR=2.00s for other four subjects; resolution=3mm isotropic; duration=5min per subject. The EEG was recorded using a 30-channel MR-compatible electrode cap with a sampling frequency of 5000 Hz and was then corrected for gradient artifacts and down-sampled to 1000 Hz. Other detailed information about the dataset can be found in van der Meer et al. [2016].

A.2 Data Preprocessing

The clean EEG data were filtered into 1-100Hz, and notch filtered at 50, 100, and 150 Hz to remove the power-line signal. The EEG signals were then re-referenced to the average reference. The fMRI signals were slice-timing corrected and spatially registered to standard MNI space, and a spatial Gaussian filter with Full-Width at Half Maximum (FWHM) = 3 mm was applied to increase the signal-to-noise ratio (SNR). Then, confound regression was carried out to remove motionrelated artifacts and slow trends due to scanner drifts. Subsequently, the fMRI signals in several regions of interest (ROIs) were extracted using the Harvard-Oxford structural atlas [Desikan et al., 2006]. In this study, we focused on four bilaterally symmetric basal ganglia regions: pallidum, caudate, putamen, and accumbens. The preprocessed subcortical fMRI signals were interpolated to 100 Hz, and the EEG was downsampled to the same sampling rate. Since the hemodynamic response measured by fMRI is slower and delayed compared to the actual onset of neural activity, we shifted the EEG signals with a time delay that approximates that of the hemodynamic response function (HRF). We set the time delay as 6 seconds, as this value attained the best performance in the previous experiments conducted by Kovalev et al. [2022]. . In our analysis, we trained subject-specific models given the potentially unique response properties of individuals. The preprocessed data for each subject were divided into training and testing sets in a ratio of 4:1, i.e., 4 minutes for training and 1 minute for testing. The data are further divided into windows of length $t_{twin} = 20.48$ seconds (with the time shift already incorporated into the EEG data to accommodate the HRF delay). To form the training sample pair $\{X_j, y_j\}$ for the j^{th} window, we extract a window of EEG data formed at a randomly selected staring time index t_j as $X_j = X(:, t_j : t_j + t_{twin} \times Fs)$ and a corresponding window of fMRI data as $y_j = y(:, t_j : t_j + t_{twin} \times Fs)$, where Fs is the sampling rate of 100 Hz.

B Resting-State Dataset Details

B.1 Data Acquisition

Written informed consent was obtained, and the experimental protocol as well as data acquisition procedure were approved by the Institutional Review Board of the National Institutes of Health and Vanderbilt University.

MRI data were acquired on a 3T Siemens Prisma scanner (Siemens, Erlangen, Germany) with a Siemens 64-channel head/neck coil. BOLD fMRI data were collected using a multi-echo gradient-echo EPI sequence with the following settings: flip angle = 75 degrees, repetition time (TR) = 2100 ms, echo times = 13.0, 29.4, and 45.7 ms, voxel size = 3 × 3 × 3 mm³, slice gap = 1 mm, matrix size = 82 × 50, 30 axial slices, and acceleration factor = 2. Scalp EEG was acquired simultaneously with fMRI using a 32-channel MR-compatible system with FCz as the reference electrode (BrainAmps MR, Brain Products GmbH) at a sampling rate of 5 kHz and was synchronized to the scanner's 10 MHz clock to facilitate reduction of MR gradient artifacts. Respiration and cardiac activity were also simultaneously monitored during the scans using a pneumatic belt and PPG (Biopac, Goleta, CA). MRI scanner (slice) triggers were recorded together with the physiological and EEG signals for data synchronization. Meanwhile, a high-resolution, T1-weighted structural image (TR = 2200 ms, TE = 4.25 ms, flip angle = 9 deg, inversion time = 1000 ms, matrix = 256 × 256, 160 sagittal slices, 1 mm isotropic) was acquired for anatomic reference.

B.2 Data Preprocessing

Slice-timing correction and motion-coregistration were applied to each multi-echo fMRI scan. Then the fMRI data were pre-processed using multi-echo ICA, which was carried out using tedana 0.0.9a. Subsequent preprocessing steps included alignment to an MNI152 standard template using SPM (https://www.fil.ion.ucl.ac.uk/spm/), removal of low-order trend (up to 4th order polynomials), and spatial smoothing (to 3mm FWHM) using AFNI (https://afni.nimh.nih.gov/afni). EEG data were corrected for gradient and ballistocardiogram (BCG) artifacts using BrainVision Analyzer 2. Subsequently, EEG data were downsampled to 250 Hz.

C Implementation Details

The proposed models were implemented using PyTorch. We chose AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.9$, and the correlation loss coefficient $\alpha = 0.1$. The batch size is 32 and initial learning rate is $3e-4$ with weight decay of $3e-4$. The whole analysis was running on a single RTX A5000 GPU.