Journal of HyoJeong Academia

Article

Bridging Multi-Imaging Modalities Using Deep Learning for Comprehensive Insights on Resting State Brain Activities

Grace Dominique Bruno[1], Gyuseok Lee[1], Jörg Stadler[2], André Brechmann[2] and Wonsang You[1,*]

[1] AIIP Lab, Department of Information and Communication Engineering, Sun Moon University, Asan 31460, Korea
[2] Combinatorial Neuroimaging Core Facility, Leibniz Institute for Neurobiology, 3918 Magdeburg, Germany
* Correspondence: wyou@sunmoon.ac.kr

Abstract: In the resting state, EEG and fMRI have functional correlations in the low-frequency band, and integrating the two modalities can provide a more comprehensive understanding of brain activity. However, multimodal imaging faces challenges such as high cost and complexity of data fusion. In this study, we developed a Transformer-CNN to generate fMRI data from EEG signals and introduced spatial normalization to compensate for differences in brain structures between subjects. Our results showed that the brain structures were normalized to the same extent, so that the model could focus only on predicting the signal values of fMRI, and compared with actual fMRI scans, we obtained PSNR of 25.92 and SSIM of 0.56, which were quantitatively and qualitatively evaluated. Although there are some qualitative limitations for medical device utilization, our approach opens new avenues in neuroscience, especially in environments where simultaneous EEG-fMRI acquisition is not possible. This study highlights the potential of deep learning in advancing multimodal imaging and provides enhanced insights into brain function.

Keywords: Multimodal neuroimaging, deep learning, EEG, fMRI

Citation Grace Dominique Bruno, Gyuseok Lee, Jörg Stadler, André Brechmann and Wonsang You. Bridging Multi-Imaging Modalities Using Deep Learning for Comprehensive Insights on Resting State Brain Activities. *JHJA 3*(1), **2025**.

Received: February 10, 2025
Accepted: April 11, 2025
Published: April 20, 2025

Publisher's Note: JHJA stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Understanding the intricate workings of the human brain is one of the most profound challenges in contemporary neuroscience. In this pursuit, integrating Electroencephalography (EEG) and functional Magnetic Resonance Imaging (fMRI) has emerged as a powerful multimodal approach [1,2]. Each technique offers unique advantages: EEG excels in capturing the rapid temporal dynamics of neural activity with millisecond-level precision, while fMRI provides detailed spatial localization of brain functions, offering insights into anatomical specificity [3,4]. Combining these techniques has the potential to overcome their limitations and yield a more comprehensive understanding of neural processes. Despite the promise of EEG-fMRI integration, several technical challenges hinder its widespread adoption. Simultaneous acquisition of EEG and fMRI requires complex equipment setups, and the high cost of this technology further restricts accessibility. Furthermore, the fusion of data from these modalities poses difficulties, as EEG reflects direct electrophysiological activity, while fMRI measures hemodynamic changes related to neural activity [2,5]. Nevertheless, recent advances in deep learning have opened new avenues for overcoming these challenges by transforming EEG data into fMRI-like representations.
Initial research has demonstrated the feasibility of this approach. For instance, models based on Auto-Encoders and Generative Adversarial Networks (GANs) have been employed to synthesize fMRI images from EEG data, revealing promising results [6]. Attention-based models have also been explored, showing fMRI patterns from generated images that align with known brain abnormalities, such as schizophrenia-associated regions in the prefrontal cortex and temporal lobes [1,2]. However, these methods often fail to fully capture the subtle functional correlations between EEG signals and fMRI networks

[7]. A key aspect of EEG-fMRI correlation lies in the relationship between low-frequency brain waves (e.g., alpha and beta bands) and the Resting State Networks (RSNs) observed in fMRI. Prior research has shown that specific oscillatory activity in the occipital region, detected through EEG, corresponds with functional connectivity patterns within RSNs [8]. This highlights the importance of embedding such neuroscientific insights into the generative models to improve the accuracy of fMRI synthesis.

In this study, we propose an advanced framework for EEG-to-fMRI generation that explicitly integrates these neuroscientific correlations. Our model focuses on transforming resting-state EEG data into time-frequency spectrograms after filtering for low-frequency brain waves. We employ a Transformer-CNN architecture as our generative model, building on recent advancements in deep learning. Additionally, we use a U-Net model as a comparative generative architecture [9]. Unlike previous studies that primarily emphasize model performance, our work prioritizes the neuroscientific validity of the generated fMRI data, aiming to bridge the gap between EEG signals and corresponding fMRI networks. This work represents a step toward more accurate and interpretable EEG-to-fMRI synthesis, with potential applications in multimodal neuroimaging research. By capturing both the temporal and spatial aspects of brain activity, this framework could contribute to a deeper understanding of the brain's functional architecture, especially when simultaneous acquisition of EEG and fMRI is not feasible.

2. Materials and Methods

2.1. Dataset

We use the freely available public NODDI dataset for resting multimodal brain studies, which has received ethical approval from the UCL Research Ethics Committee [8]. This dataset consists of EEG-fMRI pairs recorded simultaneously during resting state from a group of 16 patients. The EEG was acquired using a 64-channel MR-compatible electrode cap to ensure that the measurements are compatible with the MRI environment. The fMRI was acquired in 300 volumes with TR/TE=2160/30ms. The scans were acquired in 30 slices of 64x64 size with a voxel size of 3.3x3.3x4.0 mm, providing high spatial resolution for precise localization of neural activity. Figure 1 shows EEG and fMRI samples from this dataset, highlighting the alignment and integration of EEG and fMRI for simultaneous analysis.

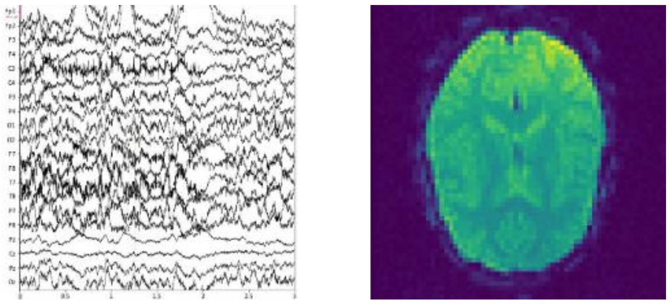


Figure 1. EEG and FMRI samples from dataset.

During the early stages of fMRI acquisition, signal intensity fluctuations occur in the acquired images due to T1 relaxation effects. These fluctuations were eliminated by excluding the first five volumes from the analysis [8]. The dataset can be downloaded from https://osf.io/94c5t/.

2.2. Data Preprocessing

2.2.1.EEG

The aim of the EEG preprocessing is to transform raw EEG signals into spectrograms organized by channels and frequency bands, as outlined in the introduction. The preprocessing pipeline, illustrated in Figure 2, consists of four main steps: data cropping, frequency filtering, sampling adjustment, and spectrogram generation [10, 11].

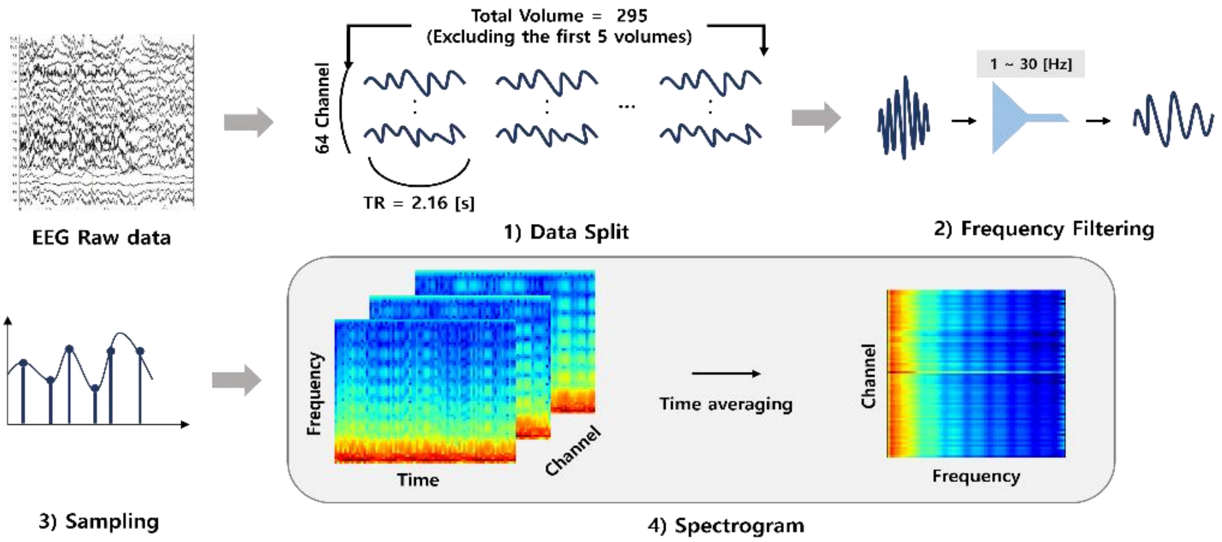


Figure 2. EEG Preprocessing pipeline.

In the first step, the EEG data is segmented in the time domain to align with each fMRI volume. Since the fMRI data is captured at intervals corresponding to a repetition time (TR) of 2.16 seconds, the EEG data is divided into segments that correspond to these intervals. EEG data associated with the initial 5 fMRI volumes, as described in Section 2.1, are excluded from further analysis to minimize any initial noise or instability.

The second step involves isolating the EEG signals within a defined low-frequency range. A band-pass filter is applied to retain frequencies between 1 and 30 Hz, which include brainwave bands commonly used in neuroscience: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz). This ensures that the EEG data reflects the relevant frequency bands associated with neural activity [10].

In the third step, the sampling rate of the EEG data is adjusted to reduce the data volume and improve processing efficiency. The original EEG recordings, sampled at 5000 Hz, are down- sampled to 1000 Hz. This reduction minimizes computational overhead while retaining sufficient temporal resolution for subsequent analysis.

In the final step, a Fourier transform is applied to the channel-specific EEG data, converting it from the time domain to the frequency domain [11]. This process generates spectrograms in the format of (channel x frequency x time). For use in deep learning model training, a time averaging technique is applied to compress the time axis, resulting in a data format of (channel x frequency). This transformation captures detailed frequency dynamics across EEG channels, facilitating more accurate neuroscientific analysis. This preprocessing workflow efficiently transforms raw EEG data into spectrograms, capturing key frequency information that is crucial for the subsequent EEG-to-fMRI generation model.

2.2.2. fMRI

Spatial normalization is a key process in aligning brain images to a standardized space, enabling consistent analysis across different subjects [12]. This step is essential for ensuring that structural patterns are uniformly learned in deep learning model training. The fMRI preprocessing pipeline, summarized in Figure 3, consists of three stages: linear registration, non-linear registration, and masking. The MNI152 template, with dimensions of 91x109x91 and a voxel size of 2 x 2 x 2 mm, is used as the standard reference for spatial normalization [13].

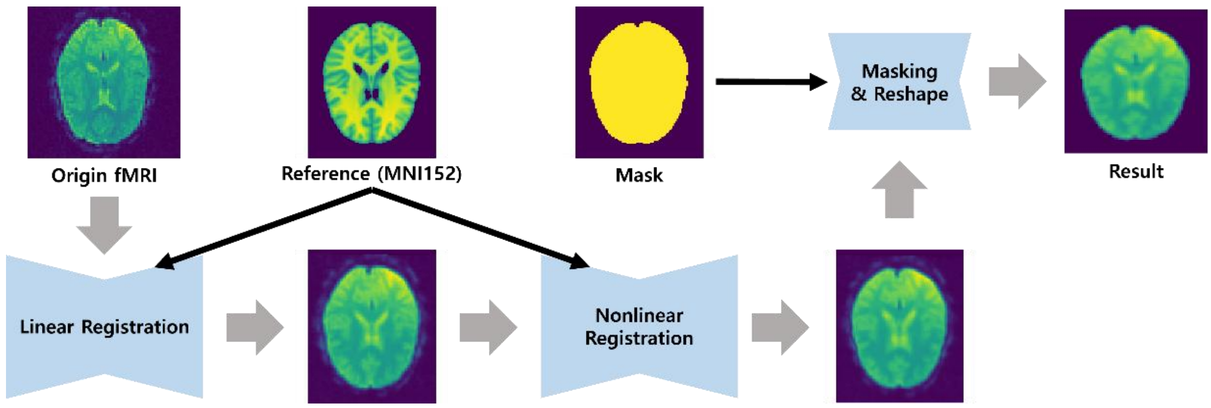


Figure 3. FMRI Preprocessing pipeline.

In the first stage, linear registration is applied to align the fMRI data with the MNI152 template. This step corrects for differences in brain structure, size, and spatial position between the fMRI data and the template. The Affine Transform algorithm is employed to perform this alignment, ensuring that the fMRI data matches the standard template in basic structural terms [14].

After linear registration, non-linear registration is used to address more complex differences in brain structure and shape that cannot be fully corrected through linear alignment. The Symmetric Normalization algorithm is applied to transform the fMRI data to closely match the shape of the standard brain structure in the MNI152 template [15]. This step ensures more precise normalization, accommodating individual anatomical variability.

The final step is masking, where non-brain tissues such as the skull are removed from the fMRI data. Instead of using FSL's conventional single-volume skull stripping method, which can be time-consuming, we utilize a more efficient approach that leverages a brain mask derived from the MNI152 template. This mask is transformed into the same space as the fMRI data, quickly eliminating non-brain areas while preserving the brain structure.

After the masking step, the fMRI data is resized back to its original dimensions of 64 x 64 x 30 to ensure compatibility with subsequent analyses. This step retains the necessary spatial resolution for accurate brain mapping and functional connectivity studies. This comprehensive preprocessing pipeline standardizes the fMRI data, preparing it for deep learning model training and further neuroscientific investigations.

2.3. Transformer-CNN model

2.3.1. Model structure

We propose a method of fMRI synthesis from EEG using Transformer-CNN, as shown in Figure 4, which is a combination of an autoencoder (CAE) composed of convolutional layers and a Transformer [16,17]. The model is trained through two processes, the first is pre-training of the Convolution Auto Encoder (CAE). The CAE learns the process of reconstructing fMRI volumes as input [17]. This aims to utilize the powerful pre-training knowledge of fMRI to reduce the modality difference between EEG and fMRI.

The second is to learn the process of fMRI synthesis from EEG by combining the Transformer and CAE. The fMRI data used in the current process and the previous process is the same data to efficiently learn the Transformer based on the pre-learning knowledge of CAE. Transformer utilizes both the Encoder-Decoder structure, and the input of the encoder is an EEG spectrogram, and it consists of a Multi-head attention and Feedforward neural network, while the decoder takes a continuous volume of fMRI as input, and it consists of a Multi-head attention 2-layer and a Feedforward neural network [16].

Transformer encoder embeds the EEG spectrogram and tokenizes it for input and processing. The decoder uses the pre-trained CAE encoder as an embedding network to tokenize the fMRI and input it. At this time, the multi-head attention in the second layer of the decoder is a cross-attention layer that also receives the output of the encoder and reflects important information of the EEG. The processed output of the decoder is input to the CAE decoder to generate a synthetic fMRI. In the training phase, both EEG and fMRI are input to learn the relationship between the two modalities, and in the test phase, without inputting fMRI, previously predicted tokens are accumulated through the autoregressive method of the Transformer and input to the decoder.

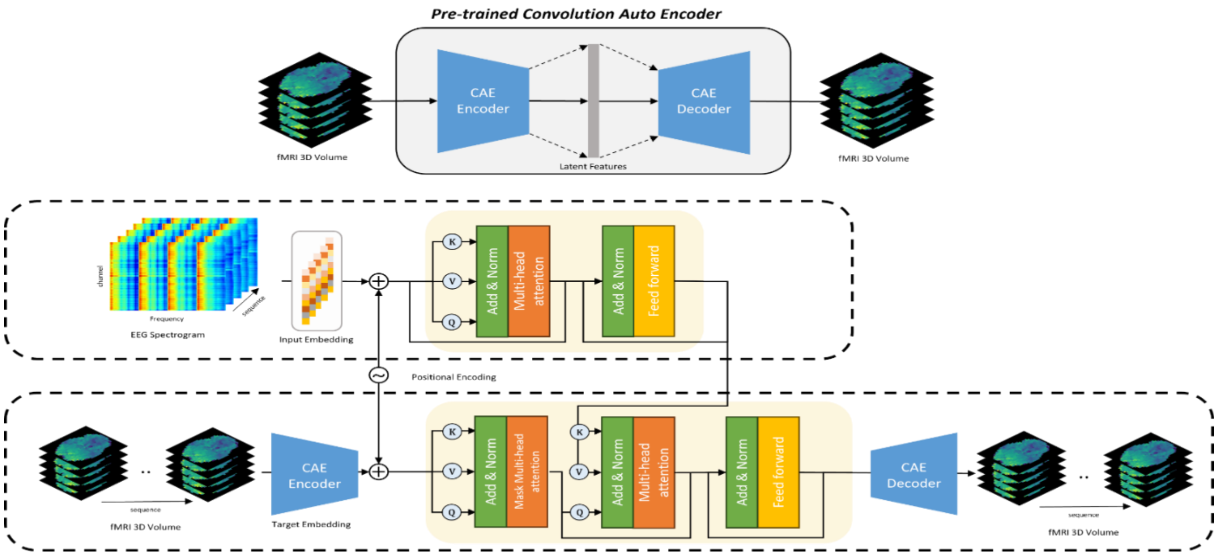


Figure 4. Transformer-CNN architecture.

2.3.2. Model evaluation methods

In this study, the dataset comprises EEG-fMRI pairs obtained from 16 patients, each contributing 295 pairs of data, resulting in a total of 4720 samples. To ensure a robust evaluation of the model, the dataset was split into training, validation, and testing sets using a ratio of 0.5:0.25:0.25. Specifically, 2360 samples were allocated for training, 1180 for validation, and 1180 for testing.

The training process was conducted over 100 epochs, employing Mean Squared Error (MSE) as the loss function to minimize the difference between the predicted and actual fMRI data. The model optimization was performed using the Adam optimizer with a learning rate of 1×10^{-5}, ensuring stable convergence during training. Additionally, two key metrics were used to evaluate the performance of the model: Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) [18,19]. PSNR measures the fidelity of the reconstructed fMRI data, while SSIM assesses the perceptual similarity between the generated fMRI images and the ground truth. This experimental setup, including the dataset split, optimization strategy, and evaluation metrics, ensures that the model is rigorously trained and tested, allowing for accurate assessment of its ability to synthesize fMRI data from EEG inputs.

3. Results

Figure 5 illustrates the progression of Loss, PSNR, and SSIM values over the course of training, providing insight into the model's learning effectiveness. As training progresses, the Loss consistently decreases, while PSNR and SSIM values increase, reflecting the model's improved ability to reconstruct high-quality fMRI images from EEG data. The alignment of trends between the training and validation sets demonstrates that the model is generalizing well to unseen data, with no signs of overfitting. The stability of the PSNR and SSIM values towards the final epochs indicates that the model converges effectively, generating fMRI images that are both high in fidelity and structurally similar to the ground truth. This suggests that the model successfully captures the underlying mapping between EEG and fMRI modalities.

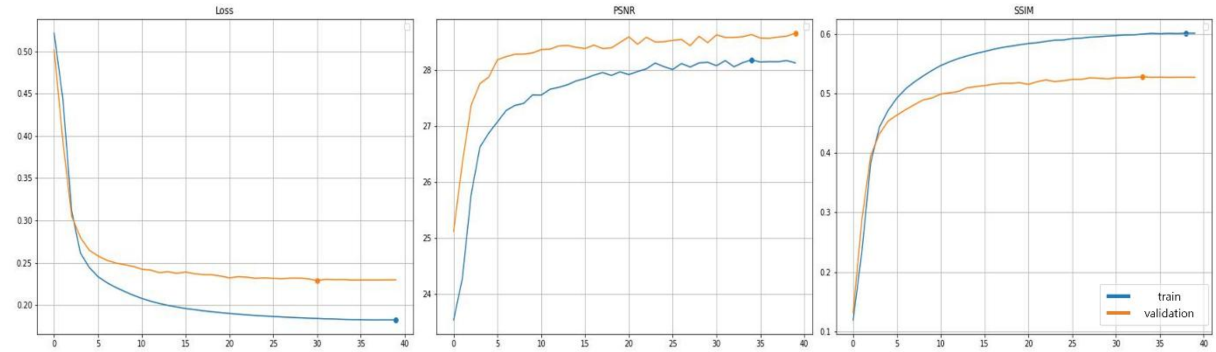


Figure 5. Learning-Validation Loss, PSNR, SSIM Graph.

Table 1 presents the performance metrics for the test results, providing insight into the quality of the fMRI images generated from EEG data. A PSNR value of 25.92 dB suggests a moderate level of similarity between the generated and actual fMRI images. While the generated images are of reasonable quality, they fall short of achieving the high fidelity required for more precise applications. Similarly, the SSIM score of 0.56 indicates a moderate degree of structural similarity. Although the generated fMRI images capture some key structural features, noticeable differences remain when compared to the ground truth. These metrics suggest that while the model is effective at producing general fMRI patterns, further refinement is needed to enhance the accuracy and structural integrity of the generated fMRI data for more nuanced neuroscientific applications.

Table 1. The quantitative results of the proposed model

Loss	PSNR	SSIM
0.20899	25.92	0.56

In a previous study, David Calhas et al. used the same NODDI dataset in both studies. In their work using autoencoders (AE), generative adversarial networks (GANs), and Wasserstein GANs (WGANs), they processed EEG with Short-Time Fourier Transform (STFT) and applied logarithmic scaling and 3x down sampling to fMRI [17]. In his subsequent work, they used an encoder-decoder architecture model combining Resnet and Attention, processed EEG with STFT, omitted any preprocessing process for fMRI, and focused on the lesion classification task in the synthetic fMRI rather than on fMRI synthesis [7]. However, we improved the performance of fMRI synthesis by processing STFT for each frequency band such as delta, theta, and alpha and integrated the brain structures across subjects through spatial normalization of fMRI. Model comparison was excluded because the data preprocessing methods and objectives were not consistent. To assess the quality of the transformed fMRI (predicted) data compared to the actual fMRI, three types of visualizations were conducted. These visual comparisons in this figure highlight the model's ability to approximate key features of the brain.

The generated fMRI images capture the overall shape and main characteristics of the brain regions quite well. However, some predictions exhibit slight blurring or smoothing, a common artifact in generated images. While the predictions retain a reasonable amount of detail, certain finer features may appear less pronounced compared to the ground truth. This suggests that although the model effectively reconstructs the essential structures of the brain, further refinement is needed to enhance the precision of finer details in the predicted fMRI images.

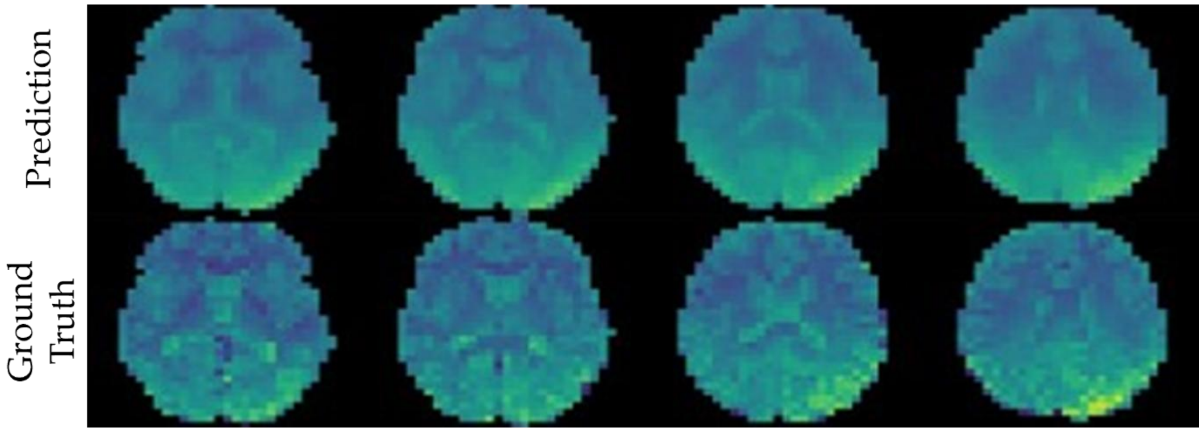


Figure 6. FMRI prediction and target comparison.

The second type of comparison involved analyzing the average BOLD signal across different brain regions, utilizing the Harvard-Oxford cortical and subcortical atlas for brain parcellation [20-22]. Figure 8 illustrates the results of comparing the generated and actual BOLD signals. The generated predictions (represented by the blue line) closely follow the overall trend of the ground truth data (orange line). However, the ground truth contains several high peaks that are less pronounced in the predicted signals. This indicates that the model tends to smooth out some of the natural variability observed in the ground truth, reducing the prominence of extreme peaks. This suggests that while the model effectively captures the general trend of BOLD signal fluctuations, it may underestimate the intensity of certain high-amplitude events or more extreme variations in the actual data. Specifically, for the Frontal Pole region, the predictions align well with the general pattern, but the model appears to minimize the impact of abrupt changes. This smoothing effect points to a potential area for improvement in better capturing the full range of BOLD signal dynamics.

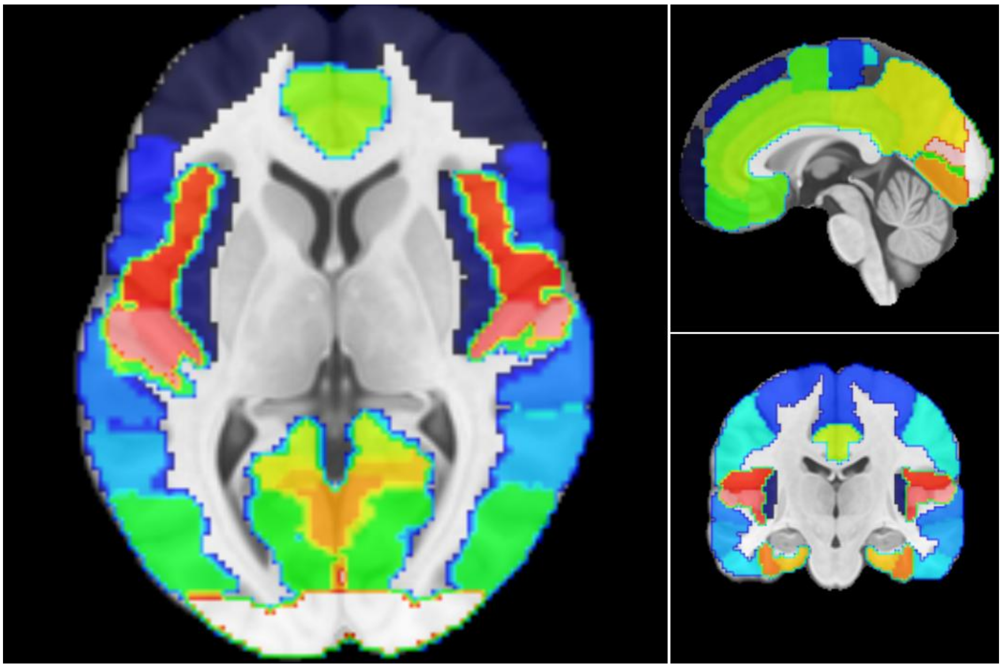


Figure 7. Harvard-Oxford cort maxprob thr25 2mm [20, 23].

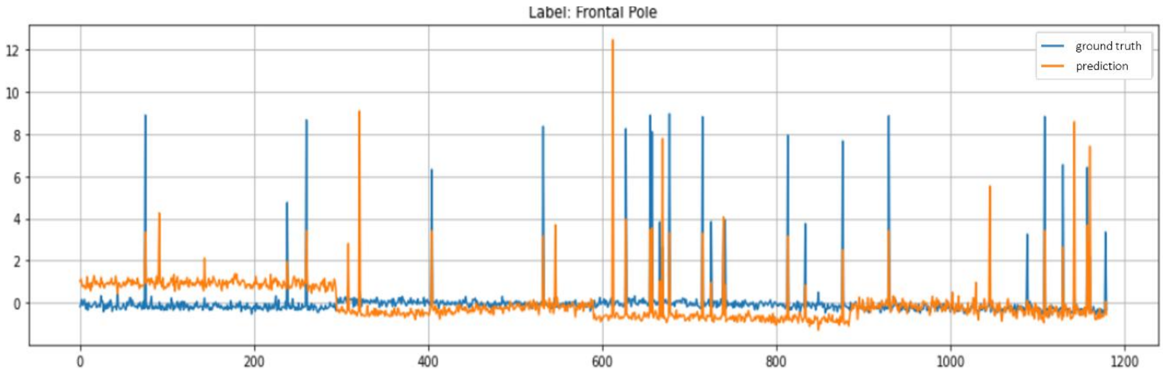
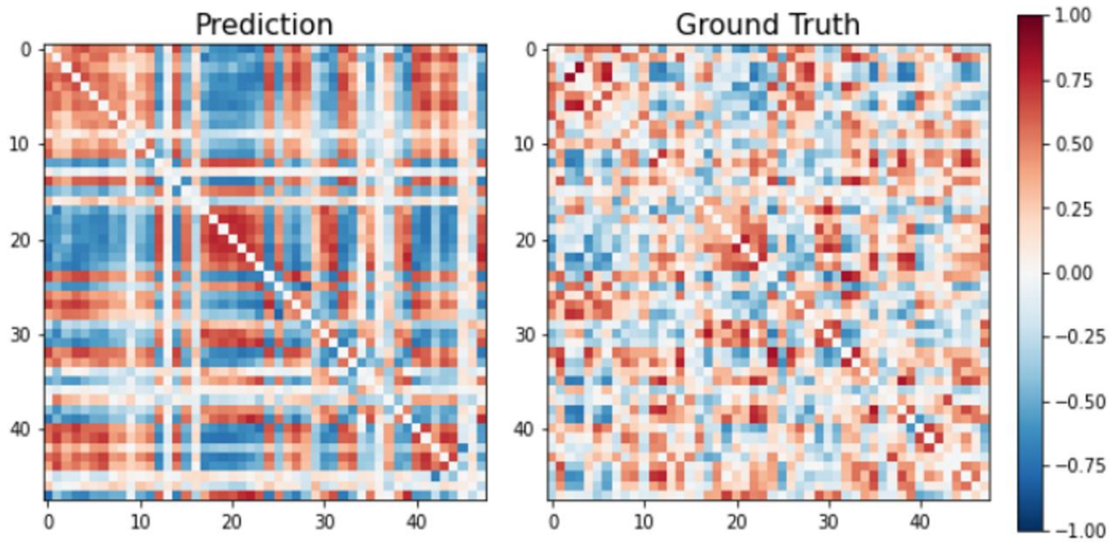


Figure 8. Bold Signals comparison.

The third type of comparison involved visualizing the **correlation matrix** of extracted **BOLD signals** between various brain regions to analyze the inter-regional relationships. **Figure 9** presents the comparison between the predicted and actual correlation matrices. The correlation matrix reflects how BOLD signals from different brain regions relate to each other, and comparing these matrices helps evaluate the model's ability to replicate functional connectivity patterns observed in the actual fMRI data [24]. The predicted correlation matrix generally captures similar patterns to the actual matrix, indicating that the model can approximate the relationships between different brain regions to a reasonable extent. However, there may be subtle differences in the strength and distribution of correlations, suggesting areas where the model could improve its fidelity in capturing the intricate connectivity dynamics across the brain. Overall, this comparison highlights that the model performs well in preserving the broad structure of inter-regional correlations but may smooth or underrepresent some finer connectivity details.



(a) (b)

Figure 9. Correlation matrix comparison : (a) Prediction ; (b) Ground Truth.

4. Discussion

The study proposed a framework to generate fMRI data from EEG recordings, aiming to establish the neuroscientific relationship between these two modalities. The methodology involved filtering EEG data into the low-frequency bands and transforming the signals into spectrograms, which were then processed by a Transformer-CNN model. The model achieved a Peak Signal-to-Noise Ratio (PSNR) of **25.92 dB** and a Structural Similarity Index Measure (SSIM) of **0.56**. While these metrics indicated a moderate level of similarity between the generated and actual fMRI data, there was still a noticeable qualitative disparity. In particular, the analysis of brain functional networks revealed limitations, as the accuracy was reduced in comparisons of average BOLD signals across brain regions and in the correlation matrix visualizations.

Despite these challenges, the study demonstrated the feasibility of generating fMRI from EEG data. There is considerable potential for improvement in the proposed Transformer-CNN model. Enhancing performance could be achieved through advanced training techniques, refinements in model architecture, and more rigorous data preprocessing. One possible improvement is to perform denoising in the postprocessing step to improve the quality of the generated fMRI images. Additionally, leveraging models such as **Super-Resolution GAN (SRGAN) [25] c**ould further improve the resolution and overall quality of the generated fMRI images, bringing them closer to actual fMRI outputs. This paves the way for future work in developing more accurate multimodal neuroimaging models.

Author Contributions: Conceptualization, Wonsang You and Jörg Stadler, André Brechmann; Methodology, Gyuseok Lee, Wonsang You; Software, Gyuseok Lee; Validation, Grace Dominique Bruno and Gyuseok Lee; Formal analysis, Grace Dominique Bruno; Investigation, Grace Dominique Bruno and Gyuseok Lee; Materials, Gyuseok Lee; Data curation, Grace Dominique Bruno and Gyuseok Lee; Drafting-manuscript preparation, Grace Dominique Bruno; Review & editing; Gyuseok Lee and Jörg Stadler, André Brechmann, Wonsang You; Visualization, Grace Dominique Bruno; Guidance, Wonsang You; Project administration; Wonsang You; Funding acquisition, Wonsang You. All authors have read and agreed to the published version of the manuscript." Authorship must be limited to those who have contributed substantially to the work reported.

Funding: This work was supported by the Start-up Growth Technology Development Project (S3228660) funded by the Ministry of SMEs and Startups and was also supported in part by the Basic Science Research Program (NRF2022R1F1A1075204), the BK21 FOUR, and the Regional Innovation Strategy Project (2021RIS-004) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, South Korea.

References

1. Menon, V. and Crottaz-Herbette, S. Combined EEG and fMRI studies of human brain function. *International review of neurobiology 66*, 291-321 (2005).
2. Warbrick, T. Simultaneous EEG-fMRI: what have we learned and what does the future hold? *Sensors 22*(6), 2262 (2022).
3. Christoph, M. and Koenig, T. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage 180*, 577-593 (2018).
4. Soares, J. M. et al. A Hitchhiker's guide to functional magnetic resonance imaging. *Frontiers in neuroscience 10*, 515 (2016).
5. Huster, R. J. et al. Methods for simultaneous EEG-fMRI: an introductory review. *Journal of Neuroscience 32*(18), 6053-6060 (2012).
6. Calhas, D. and Henriques, R. EEG to fMRI Synthesis: Is Deep Learning a candidate? *arXiv:2009*, 4133v1 (2020).
7. Calhas, D. and Henriques, R. EEG to fMRI Synthesis for Medical Decision Support: A Case Study on Schizophrenia Diagnosis. *medRxiv 10.1101*, 23293748 (2023).
8. Deligianni, F. and Centeno, M. Relating resting-state fMRI and EEG whole-brain connectomes across frequency bands. *Front. Neurosci. 8*, 258 (2014).
9. Olaf, R., Fischer, P. and Brox. T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of 18th international conference on Medical image computing and computer-assisted intervention, Munich, Germany (5-9 October 2015).
10. Gyorgy, B. and Draguhn. A. Neuronal oscillations in cortical networks. *science 304.5679*, 1926-1929 (2004).
11. Ng, M., Jing, J. and Westover, M. A primer on EEG spectrograms. *Journal of Clinical Neurophysiology 39*(3), 177-183 (2022).
12. Crinion, J. et al. Spatial normalization of lesioned brains: performance evaluation and impact on fMRI analyses. *Neuroimage 37*(3), 866-875 (2007).
13. Fonov, V., Evans, A., Mckinstry, R., Almli, C. and Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood." *NeuroImage 47*(1), S102 (2009).
14. Ganesh, L. et al. Affine based image registration applied to MRI brain. In 2013 International Conference on Information Communication and Embedded Systems (ICICES), Tsukuba, Japan (13-17 May 2013).
15. Noblet, V., Heinrich, C., Heitz, F. and Armspach, J. Symmetric nonrigid image registration: application to average brain templates construction. *Med Image Comput Comput Assist Interv. 11*(2). 897-904 (2008).
16. Vaswani, A. et al. Attention is all you need. In Proceedings of the 31th international Conference on Neural Information Processing systems, Long Beach California, USA (4-9 Dec. 2017).
17. Dor, B., Koenigstein, N. and Giryes, R. Autoencoders. In Machine learning for data science handbook: data mining and knowledge discovery handbook (Springer, 2023).
18. Alain, H. and Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 20th international conference on pattern recognition, Istanbul, Turkey (23-26 Aug. 2010).
19. Nilsson, J. and Akenine-Möller, T. Understanding ssim. *arXiv:2006*, 13846 (2020).
20. Nikos, M. et al. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophrenia research 83*(2), 155-171 (2006).

21. Megan, L., Smyser, C. and Shimony, J. Resting-state fMRI: a review of methods and clinical applications. *American Journal of neuroradiology 34*(10), 1866-1872 (2013).
22. Greicius, M., Krasnow, B., Reiss, A. and Menon, V. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the national academy of sciences 100*(1), 253-258 (2003).
23. https://identifiers.org/neurovault.image:1699
24. Ryan, M., Glonek, G. Tuke, J. and Humphries, M. Capturing functional connectomics using Riemannian partial least squares. *Scientific Reports 13*(17386), (2023).
25. Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA (21-26 July 2017).