MEG2fMRI: Generation of Accurate Synthetic Functional Magnetic Resonance Imaging Slices Based on Magnetoencephalography Data

Bugra Aldal

Student Email: *bugra.aldal[at]bahcesehir.edu.tr*

Student Number: *2101669*

Instructor: *Lavdie Rada Ulgen*

Abstract—Combining hemodynamic and electrophysiological neuroimaging techniques provides a strong window into the spatiotemporal dynamics of brain activity. Direct data fusion is severely hampered by the differences between magnetoencephalography and functional magnetic resonance imaging, particularly with regard to resolution, dimensionality, and acquisition timing. In this study, we introduce a new deep learning approach for creating functional magnetic resonance imaging slices from magnetoencephalography signals using a conditional generative adversarial network. In order to create 32×32 functional magnetic resonance imaging slices, we converted magnetoencephalography epochs into low-dimensional embeddings and trained a Pix2Pix-style generative adverserial network on aligned magnetoencephalography and functional magnetic resonance imaging event pairs. Mean squared error and structural similarity index were used to assess our model. On held-out data, we obtained a mean squared error of 0.046 and structural similarity index of 0.622. Our results show the viability of magnetoencephalography-to-functional-magnetic-resonance-imaging synthesis and highlight the expanding potential of synthetic neuroimaging data, despite obstacles such as data alignment, computing limitations, and training instability. We suggest a number of potential options, such as time-series modeling, 3D or 4D volumetric and timed pipelines, better alignment via unsupervised learning, and postprocessing for higher-quality images. This work serves as a proof-of-concept for deep generative models that are being developed for data augmentation and cross-modal neuroimaging translation.

Keywords—*Magnetoencephalography, Functional Magnetic Resonance Imaging, Cross-modal neuroimaging, Generative adversarial networks, Modularity transfer*

1. Introduction

Multimodal neuroimaging has emerged as an important method in cognitive neuroscience, offering a more comprehensive view of brain function and halping the understanding of diagnostics by combining data from different techniques with complementary strengths [1]. Among these modalitites; Magnetoencephalography (MEG) provides high temporal resolution, capturing neural dynamics on the millisecond scale, whereas functional magnetic resonance imaging (fMRI) offers superior spatial resolution, allowing detailed mapping of brain activity [2], [3]. Integrating these modalities has immense potential for studying cognition, clinical diagnostics, and brain-computer interfacing. However, substantial differences in signal characteristics and dimensionality make aligning of MEG and fMRI data a significant challenge [4].

Recent advances in deep learning have opened new avenues for bridging these modality gaps. Among these, Generative Adversarial Networks (GANs) [5] have gained prominence for their ability to learn complex mappings between high-dimensional data domains. Originally introduced for image generation, GANs have been successfully applied in medical imaging tasks such as cross-modality synthesis [6]. However, the use of GANs to directly learn mappings from time-series electrophysiological data like MEG to spatially resolved fMRI images remains underexplored.

In this work, we present a novel deep generative framework that learns to translate MEG signals into fMRI representations using a Pix2Pix-style conditional Generative Adversarial Network (cGAN) [7]. Our model is trained and evaluated on the SMN4Lang dataset, a large-scale multimodal corpus containing synchronized MEG and fMRI recordings during naturalistic language comprehension [8]. By processing MEG epochs around aligned events and generating corresponding fMRI image slices, we demonstrate that GANs can learn meaningful mappings between these distinct modalities. We report quantitative and qualitative results from our trained model, showcasing its ability to generate spatially plausible synthetic fMRI images from raw MEG inputs.

Additionally, our work contributes to the growing interest in synthetic medical data generation. Synthetic data offers a promising solution to key challenges in neuroscience, including limited data availability, cross-subject variability, and privacy concerns [9]. By generating realistic fMRI images from MEG data, we take a step toward developing flexible and privacy-preserving tools that may support data augmentation, anomaly detection, and multimodal brain decoding in future applications.

2. Literature Review

Recent advances in multimodal neuroimaging have highlighted the potential of integrating neuroimaging techniques to construct a comprehensive, high-resolution understanding of brain function, and numerous studies have tackled the complementary nature and fusion strategies of these modalities to address challenges in spatial-temporal mapping, image synthesis, and diagnostic inference.

This multimodal interaction is thoroughly described in the foundational work by Hall et al. [10], which clarifies the signals' complimentary nature. fMRI records the slower hemodynamic changes brought on by neural energy demands, whereas MEG records the magnetic fields produced by synchronized post-synaptic neuronal activity. Their analysis explains how direct integration is made more difficult by the different sensitivities, acquisition methods, and processing difficulties of both signals, despite the fact that they are primarily of dendritic origin. However, spatial correlation studies have demonstrated a reasonable co-localization of MEG and fMRI signals employing tasks such as visual stimulation, especially when comparing task-induced oscillations in the beta and gamma bands with evoked responses. These results are consistent with more general patterns found in recent research.

By using scalp field forward modeling and source reconstruction to quantify spatial overlap between electroencephalography (EEG) source models and fMRI activity, Heugel et al. [11] further advance our understanding of the signals. This method circumvents the strict assumptions of joint independent component analysis and reveals significant overlap in temporoparietal regions during auditory oddball tasks.

From a predictive modeling perspective, Engemann et al. [12] present a compelling case for multimodal stacking, showing that predictive models for brain age combining MEG, fMRI, and anatomical MRI outperform unimodal approaches and retain robustness under missing modality scenarios, with MEG features in the 8–30 Hz band contributing significantly to performance.

The SAMBA (Spatiotemporal Alignment of Multimodal Brain Activity) framework is proposed by Afrasiyabi et al. [13]. It establishes a new standard for translating MEG to fMRI and EEG to fMRI. By introducing a shared latent representation that captures brain dynamics independent of modality, SAMBA facilitates translation and downstream cognitive decoding, in contrast to conventional methods that depend on strict mappings or shallow fusion techniques. The model uses a number of innovative elements, graph attention networks (GATs) for learning functional connectivity and spatially upsampling from coarse MEG to fine fMRI resolutions, attention-based wavelet decomposition for selectively filtering electrophysiological signals in the frequency domain, and recurrent neural networks (RNN) for modeling autoregressive patterns and temporal dependencies in brain activity sequences. An important improvement over earlier studies that usually assume a fixed hemodynamic response function (HRF) across brain areas is that SAMBA includes a learnable HRF that permits parcel-specific modeling of neurovascular coupling. The framework outperforms baselines like transformers and convolutional architectures in both short (15s) and long (60s) time windows, achieving state-of-the-art performance. The significance of every architectural element is supported by ablation tests, which show that performance significantly declines with the removal of wavelet, HRF, or LSTM modules. Furthermore, SAMBA's latent embeddings are useful for purposes other than translation, allowing for more accurate stimulus classification than previous models. By modeling the inverse HRF for reconstructing MEG from fMRI and leveraging wavelet reconstruction loss to retain signal fidelity, the model not only supports bidirectional translation but also offers neuroscientific interpretability, such as mapping HRF dispersion across brain regions.

Gao et al. [14] introduce a novel deep learning framework designed to address a shortcoming in multimodal neuroimaging: the challenge of missing or incomplete data in PET scans used for Alzheimer's Disease (AD) studies. By combining two complementary architectures, the Pathwise Transfer Dense Convolutional Network (PT-DCN) for multimodal classification and the Task-Induced Pyramid and Attention GAN (TPA-GAN) for PET image imputation, their work connects image synthesis and clinical inference. The TPA-GAN incorporates a self-attention module that simulates long-range relationships and improves feature relevance, in addition to a pyramid convolution structure to capture multi-scale spatial information. In order to direct the generating process to preserve disease-relevant patterns rather than just photorealistic features, it incorporates a task-induced discriminator trained for Alzheimer's diagnosis.

Significantly, the objectives of models such as SAMBA [13], which seek to develop unified, modality-independent representations that can support both synthesis and inference tasks, are similar to Gao et al. [14] as well as MEG2fMRI's purpose. While SAMBA focuses on dynamic temporal alignment across MEG, EEG, and fMRI using shared latent embeddings and recurrent structures, TPA-GAN-PT-DCN targets anatomical and functional fusion via image-to-image translation and downstream diagnosis, offering a complementary approach optimized for structural imaging datasets. The two frameworks share a commitment to interpretability and clinical applicability specific attention mechanisms and task-aware loss functions, indicating that a clear trend in the field toward systems that are not only accurate but also aligned with neurobiological relevance.

Lan et al. [15] introduces the Spectrally Normalized Self-Attention Conditional GAN (SC-GAN), a 3D generative adversarial network designed to address the limitations of prior 2D synthesis approaches. SC-GAN integrates several novel mechanisms including spectral normalization, brain region-specific RMSE loss, and feature matching loss to ensure convergence, generalization, and structural fidelity in high-resolution volumetric imaging. Critically, SC-GAN employs 3D self-attention modules in both the generator and discriminator, enabling the modeling of distant dependencies across volumetric neuroimaging data which is relevant for preserving anatomical details in synthesized PET, FA, and MD maps. They note that compared to traditional 2D and 3D cGANs, SC-GAN achieved significantly lower synthesis error and higher structural similarity across multiple tasks.

The utility of synthetic data is emphasized in the study of Vaden et al. [16] on fully synthetic neuroimaging via multiple imputation, which preserves covariance structures while mitigating privacy concerns, providing an ethical framework for developing and distributing datasets.

On the clinical and psychiatric front, Tulay et al. [1] demonstrate the diagnostic power of multimodal integration in psychiatry, leveraging classical methods like principle component analysis and independant component analysis alongside machine learning classifiers to enhance outcome prediction, illustrating how even simple statistical fusion can yield clinically relevant biomarkers.

These papers collectively demonstrate convergent trends in robust multimodal learning frameworks, graph-based and attention-driven modeling, spatiotemporal alignment, and synthetic data creation. By using architectures that can jointly optimize for reconstruction and classification, as well as with provisions for domain generalization through synthetic augmentation and missing modality robustness, they collectively propose that a modularity translation is best approached through latent space learning informed by connectivity patterns, frequency decomposition, and anatomical priors rather than naive image mapping. As such, the MEG2fMRI study has great potential to benefit from incorporating elements such as SAMBA's latent space strategy [13], TPA-GAN's task-guided generative modeling [14], and the late fusion and attention mechanisms that have been repeatedly proven effective in both segmentation and synthesis tasks in the neuroimaging literature.

3. Methodology

3.1. Dataset

This study utilizes the SMN4Lang dataset [8], a publicly available synchronized multimodal neuroimaging dataset designed to support research on naturalistic language processing. It comprises MEG and fMRI data collected from the same 12 healthy, right-handed Mandarin-speaking adults aged between 23 to 30. During several sessions, the participants listened to 60 audio stories for a total of six hours. The same story materials were used for both fMRI and MEG recordings of each participant, allowing for the high spatial and temporal resolution study of language-related cognitive processes. In addition to task-based recordings, the dataset includes high-resolution structural T1 and T2, diffusion-weighted imaging and resting-state fMRI scans, allowing for analyses of structural-functional coupling and connectivity.

MEG recordings were acquired using a 306-channel Elekta Neuromag TRIUX system at a sampling rate of 1000 Hz. The data were preprocessed using advanced artifact removal techniques such as temporal signal space separation and independent component analysis, with bandpass filtering between 0.1 and 40 Hz and additional monitoring of ocular and cardiac signals. A fixed auditory delay of 39.5 ms was accounted for in the alignment. fMRI data were collected using a Siemens Prisma 3T scanner with a BOLD-sensitive protocols that consists of 106×106 image size, 72 2-mm-thick axial slices, and 634 volumes following the Human Connectome Project pipelines to ensure high-quality registration, motion correction, and normalization.

60 spoken stories, each lasting 4 to 7 minutes, covering a variety of subjects like culture and education, made up the auditory stimulus. The total word count of these stories is almost 52,000, with a vocabulary of 9,153. The data is accompanied by linguistic annotations, such as part-of-speech tags, grammatical constituency and dependency trees, character and word frequency counts obtained from extensive Mandarin corpora, and accurate speech-to-text alignments. The dataset also contains a range of textual embeddings.

To ensure the reliability of the data, participants completed comprehension quizzes after each story, achieving mean accuracy rates exceeding 91% for both fMRI and MEG tasks. The fMRI data quality was validated through assessments of temporal signal-to-noise ratio, frame-wise displacement, and inter-subject correlation, all indicating high reliability. Similarly, MEG data were evaluated using inter-subject correlation across multiple frequency bands and neural entrainment analyses, demonstrating strong and consistent brain

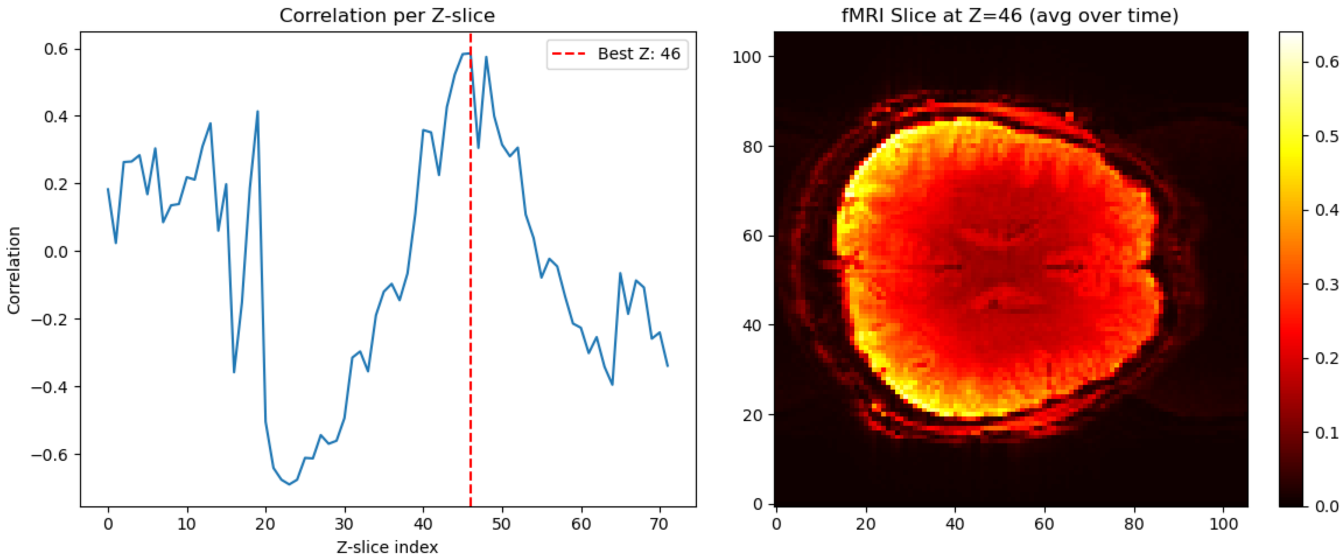

Figure 1. Selection of Best-Matching fMRI Slice Based on Correlation with MEG Time Series. Z-scored MEG-fMRI for each axial slice, with the highest-correlating slice highlighted (left). The corresponding axial fMRI image selected for training at time index 32 (right).

responses particularly in delta and theta bands localized to auditory and language-related regions. The entire dataset is structured following the Brain Imaging Data Structure standard and is publicly available on OpenNeuro, complete with raw and preprocessed data, code, metadata, and annotations, providing a comprehensive resource for multimodal and reproducible neuroimaging research. The dataset's latest version can be accessed via AWS CLI in a S3 bucket with the link s3://openneuro.org/ds004078ds004078-download/.

3.2. Preprocessing and Alignment

A central challenge in multimodal neuroimaging synthesis is establishing correspondence between modalities with fundamentally different spatial, temporal, and signal properties [4]. In our study, we construct a pair of MEG and fMRI data. This is possible since the dataset has both MEG and fMRI data for every subject and run. This dataset is built using a pipeline that emphasizes temporal alignment, normalization, and dimensionality reduction tailored to each modality's properties. The MEG data is preprocessed by first loading raw FIF files using the MNE library. Each raw MEG recording is downsampled to 100 Hz for computational efficiency and to match the timescale relevant to the stimulus presentation paradigm. The continuous MEG signals are then normalized per channel using subtracting the mean and dividing by the standard deviation to standardize signal amplitudes across sensors and recording sessions. Stimulus-related events are extracted from the MEG data using the STI101 trigger channel. We identify the first relevant stimulus event and records its onset time. Around each event onset, a fixed-length 64 timepoints of a MEG epoch is extracted. This results in a temporally localized MEG segment from all 328 sensors. To reduce dimensionality while preserving spatial structure, we collapse the MEG epoch into a single feature vector by averaging across the time axis and then apply z-scoring again to ensure consistent scale across samples.

For the fMRI data, we load NIfTI volumes using NiBabel and apply min-max normalization across the entire volume to scale voxel intensities between 0 and 1. The event timings associated with the fMRI scans are extracted from the corresponding event annotation files. We extract a temporal window of 64 consecutive fMRI volumes starting at the stimulus onset time to match time timings with MEG data. This results in a 4D fMRI segment. To match fMRI volumes to the MEG representation more precisely, we search for the most temporally and spatially relevant slice. For each axial slice in the 3D fMRI volume, we compute the mean time series by averaging over the X and Y dimensions. These time series are then z-scored and correlated with the MEG signal's temporal profile to identify the slice with the highest correspondence. The best-matching 2D fMRI slice at time index 32 and best Z-slice is selected for training. An example figure is provided in Figure 1. This slice is resized from its native resolution 106×106 to 32×32 pixels using bilinear interpolation. This resizing balances anatomical fidelity with computational efficiency, making it suitable as a target image for generative modeling.

The result of this pipeline is a dataset of paired MEG–fMRI samples, where each sample consists of a 328-dimensional MEG feature vector and a corresponding 32×32 fMRI image slice. By treating these pairs as input–output mappings, we utilize our Pix2Pix-style cGAN that learns to synthesize spatial brain images from temporal neurophysiological signals. The dataset creation assumes stable cognitive processing across sessions and stimulus trials—an assumption justified by the controlled experimental setup and consistent task structure in the recordings.

3.3. Model Architecture

Our model is highly insipred by the Pix2Pix cGAN which was first introduced for paired image-to-image translation tasks. The purpose of our modification is to convert spatially structured fMRI images from MEG feature vectors. The generator converts a 328-dimensional MEG input vector into a 32x32 fMRI image using a fully connected neural network. The last layer, which produces a flattened image that is then molded into a 1×32×32 tensor, is made up of three linear layers using ReLU activations. A Tanh activation function is applied at the output to constrain pixel values to the range [−1, 1], which is a common choice in GANs to stabilize training. Unlike convolutional architectures typically used in image generation, the use of fully connected layers is more appropriate here due to the non-spatial and vectorized nature of the MEG input.

The discriminator integrates data from both MEG and fMRI modalities using a hybrid design akin to PatchGAN. The MEG vector's dimensions are essentially aligned with the fMRI image by first projecting it into a 32×32 feature map via a linear layer. A two-channel input is then created by concatenating the projected MEG map along the channel axis with the generated or actual fMRI image. To extract hierarchical features and evaluate realism, this unified representation is run through a number of convolutional layers using batch normalization and LeakyReLU activations. Each value in the spatial probability map produced by the final convolution represents the

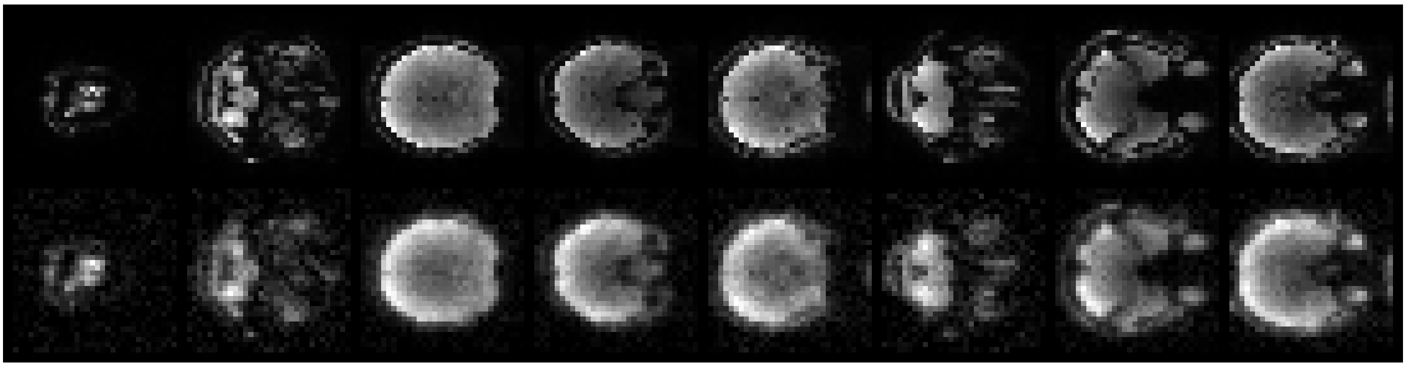

Figure 2. Example images. Ground-truth (top). Generated (bottom).

probability that, given the MEG context, a local patch in the image is real. This patch-level feedback encourages the generator to produce outputs that are both realistic and consistent with the MEG input.

In general, this design directs the generator to produce fMRI images that are visually realistic and consistent with the MEG signal. The hybrid discriminator structure improves the accuracy of the generated images in capturing brain dynamics by conditioning the image.

3.4. Training Procedure

For training, we used 10 of total 12 subjects. We held-out 2 subjects, one male, one female in order to evaluate our model. The training was performed on a system with NVIDIA GTX 1650 GPU, 9th generation Intel i5 CPU, and 8GB of RAM running NixOS -a Linux distribution. The MEG2fMRI cGAN was trained using a combination of adversarial and pixel-wise reconstruction losses. The generator learns to map MEG feature vectors to corresponding fMRI slices, with supervision from both a binary cross-entropy adversarial loss and a mean squared error reconstruction loss. The adversarial loss encourages the generator to produce realistic fMRI-like images, while the reconstruction loss ensures fidelity to the ground truth by penalizing pixel-wise differences between generated and real fMRI slices. The generator architecture consists of a fully connected network that maps 328-dimensional MEG inputs to 32×32 fMRI outputs, while the discriminator is a hybrid model that combines projected MEG embeddings and fMRI inputs as a two-channel input to a convolutional classifier. Both generator and discriminator were optimized using the Adam optimizer with a learning rate of 0.0001, and momentum terms $\beta_1 = 0.5$ and $\beta_2 = 0.999$. A learning rate scheduler was employed to reduce the learning rate by a factor of 0.5 every 20 epochs. The model was trained for 125 epochs with a batch size of 32.

3.5. Evaluation

For evaluation, we held-out two subjects from the dataset: one female and one male. We used both qualitative visual evaluations and quantitative picture similarity measurements to test our synthesis model's performance. We concentrated on measures that capture both pixel-level accuracy and perceptual fidelity, as our approach aims to produce fMRI images that are accurate to their corresponding MEG signals and structurally realistic. Mean squared error, which calculates the average squared difference between the generated and ground-truth fMRI pictures, was the main metric utilized for pixel-wise accuracy. Although MSE offers a clear indicator of reconstruction error, it does not necessarily correspond with perceptual realism and penalizes high-frequency information. In addition to MSE, we calculated the structural similarity index, a popular metric that compares local patterns of brightness, contrast, and structure to assess an image's perceived quality. Greater resemblance to the reference image is indicated by higher SSIM values, which range from -1 to 1. This measure is especially helpful in neuroimaging applications where maintaining anatomical structure is essential.

Visualizing real and produced fMRI picture pairs across several epochs allowed us to perform qualitative evaluations in addition to numerical ones. The consistency of anatomical features, the evolution of generated outputs throughout training time, and the spatial plausibility of synthesized images were all evaluated using these visual comparisons. These visual evaluations were essential for spotting overfitting patterns, artifacts, or mode collapse that were not immediately visible from numerical scores alone.

4. Results

Our trained model achieved an average MSE of 0.046, indicating a low level of pixel-wise error between the generated and ground-truth fMRI images. The SSIM score averaged 0.622, suggesting that the model preserved much of the anatomical structure present in real images.

We observed that reconstruction accuracy improved consistently across epochs, with a rather steep decrease in MSE during the initial 50 epochs followed by gradual refinement thereafter. SSIM scores also increased steadily. Table 1 summarizes the final metric values on the test set.

Table 1. Quantitative Evaluation Metrics for MEG2fMRI GAN

Metric	Value
Mean Squared Error (MSE)	0.046
Structural Similarity Index (SSIM)	0.622

Early in training, generated images appeared blurry and lacked distinct anatomical boundaries. However, by epoch 30, the outputs began to exhibit sharper contours and finer textural details. By epoch 80, most generated images closely resembled the corresponding ground-truth slices, and consistent spatial patterns could be observed across different subjects. Although the model occasionally produced artifacts in high-contrast regions or around the borders of brain structures, these artifacts were infrequent and tended to diminish with continued training. Overall, the results support the feasibility of using GANs to synthesize spatial fMRI data from MEG recordings. While the current model operates at a reduced resolution of 32×32, it serves as a proof-of-concept for cross-modal brain image synthesis, and lays the groundwork for more complex architectures and clinical-scale applications. Figure 2 shows some example images from the held-out dataset.

5. Discussion

The results of this study demonstrate the feasibility of using cGANs to learn a meaningful cross-modal mapping from MEG data to fMRI representations. By leveraging paired data from the SMN4Lang dataset, we trained a Pix2Pix-style GAN capable of synthesizing low-resolution but structurally plausible fMRI images conditioned on

MEG input. Quantitative metrics and qualitative inspection indicate that the model successfully captures key spatial patterns in brain activity that correspond to temporally localized MEG signals.

Despite these promising results, the development of this framework presented multiple technical and conceptual challenges. One of the primary limitations was computational capacity. The training and evaluation of GANs is inherently resource intensive, and the need to process large neuroimaging datasets increased the demand on available hardware. Combined with the constrained time, this led us limit image size and other capabilities which in turn have limited the final model.

Another significant challenge involved the alignment of MEG and fMRI data, both temporally and semantically. Since the recordings were conducted in separate sessions, precise alignment based on time was not possible. We relied on it and event timings but this strategy is flawed. A more robust approach would incorporate learned alignment between modalities, potentially via unsupervised methods or contrastive embedding learning.

We also encountered I/O bottlenecks during data loading, particularly when streaming files during training. A more robust solution would involve parallelized data pipelines or conversion to faster-access formats such as TFRecords.

A recurring obstacle was the fragility of the MNE-Python library, which we extensively used for MEG data handling. Lack of package maintainance in NixOS and installation issues led to delays and required significant manual configuration. Future iterations of this work may benefit from a more modular architecture where heavy dependencies like MNE are isolated in preprocessing scripts rather than tightly coupled to the model training pipeline.

From a modeling perspective, hyperparameter tuning remained a delicate and often unstable process. GANs are notoriously sensitive to training conditions. This underscores the importance of robust evaluation protocols and multiple training runs in GAN-based neuroimaging studies. A more robust setting might leverage the power of transformer architecures.

6. Conclusion

This study presents a novel deep learning framework, MEG2fMRI, for translating MEG signals into fMRI-like spatial representations using a conditional GAN. Trained on aligned MEG and fMRI pairs from the SMN4Lang dataset, our Pix2Pix-style model generated 32×32 fMRI slices conditioned on MEG activity, achieving promising results in quantitative metrics. These findings validate the idea that learned mappings between neuroimaging modalities are feasible and potentially useful for neuroscientific interpretation and data augmentation.

Overall, this work provides a proof-of-concept for data-driven MEG to fMRI translation, highlighting both the promise and complexity of cross-modal neural synthesis. The framework lays a foundation for future multimodal neuroimaging models that are more flexible, biologically grounded, and capable of handling the inherent challenges of cognitive data.

7. Future Work

Looking forward, several promising directions for future research emerge from this work. One possible extension involves increasing the dimensionality of the output fMRI representation. While this study focused on generating 2D slices for simplicity and efficiency, expanding the generator to produce full 3D fMRI volumes would offer a more faithful spatial mapping and open the door to finer anatomical modeling. Additionally, incorporating time into the setting and generating 4D structures from 2D MEG data could capture richer functional structure and potentially enable real-time decoding or simulation of cognitive states.

Improvements in event alignment represent another compelling avenue. Although our current pipeline relies on explicit event annotations, an unsupervised or weakly supervised model that learns to infer alignment between MEG and fMRI on the basis of latent dynamics could eliminate the need for labeled correspondence and improve generalization to new datasets. Similarly, postprocessing methods could be applied to the GAN output to reduce noise and improve anatomical clarity.

Moreover, the resolution and quality of generated images remain an area for enhancement. Moving to higher-resolution output could increase the utility of generated fMRI for downstream clinical or research applications. Integrating pretrained encoders or perceptual loss functions may also improve fidelity and structural realism.

References

[1] E. E. Tulay, B. Metin, N. Tarhan, and M. K. Arıkan, "Multimodal neuroimaging: Basic concepts and classification of neuropsychiatric diseases", *Clinical EEG and Neuroscience*, vol. 50, no. 1, pp. 20–33, 2019, PMID: 29925268. DOI: 10.1177/1550059418782093. eprint: https://doi.org/10.1177/1550059418782093. [Online]. Available: https://doi.org/10.1177/1550059418782093.

[2] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. Lounasmaa, "Magnetoencephalography: Theory, instrumentation, and applications to noninvasive studies of the working human brain", *Rev. Mod. Phys.*, vol. 65, pp. 413–, Apr. 1993. DOI: 10.1103/RevModPhys.65.413.

[3] K. Kwong, J. Belliveau, D. Chesler, *et al.*, "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 12, pp. 5675–5679, Jun. 1992. DOI: 10.1073/pnas.89.12.5675.

[4] Z. Liu, L. Ding, and B. He, "Integration of eeg/meg with mri and fmri", *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, no. 4, pp. 46–53, Jul. 2006. DOI: 10.1109/MEMB.2006.1657787.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial networks", *Advances in Neural Information Processing Systems*, vol. 3, Jun. 2014. DOI: 10.1145/3422622.

[6] M. Rafiq, H. Ali, G. Mujtaba, Z. Shah, and S. Azmat, *Cross modality medical image synthesis for improving liver segmentation*, 2025. arXiv: 2503.00945 [eess.IV]. [Online]. Available: https://arxiv.org/abs/2503.00945.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, *Image-to-image translation with conditional adversarial networks*, 2018. arXiv: 1611.07004 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1611.07004.

[8] S. Wang, X. Zhang, J. Zhang, and C. Zong, *"a synchronized multimodal neuroimaging dataset to study brain language processing"*, OpenNeuro, 2023. DOI: doi:10.18112/openneuro.ds004078.v1.2.1.

[9] R. Chen, M. Lu, T. Chen, D. Williamson, and F. Mahmood, "Synthetic data in machine learning for medicine and healthcare", *Nature Biomedical Engineering*, vol. 5, pp. 1–5, Jun. 2021. DOI: 10.1038/s41551-021-00751-8.

[10] E. L. Hall, S. E. Robson, P. G. Morris, and M. J. Brookes, "The relationship between meg and fmri", *NeuroImage*, vol. 102, pp. 80–91, 2014, Multimodal Data Fusion, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.11.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811913010975.

[11] N. Heugel, E. Liebenthal, and S. Beardsley, “Method for spatial overlap estimation of electroencephalography and functional magnetic resonance imaging responses”, *Journal of Neuroscience Methods*, vol. 328, p. 108 401, 2019, ISSN: 0165-0270. DOI: https://doi.org/10.1016/j.jneumeth.2019.108401. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027019302584.

[12] D. A. Engemann, O. Kozynets, D. Sabbagh, *et al.*, “Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers”, *eLife*, vol. 9, A. Shackman, F. P. de Lange, K. Tsvetanov, and N. Trujillo-Barreto, Eds., e54055, May 2020, ISSN: 2050-084X. DOI: 10.7554/eLife.54055. [Online]. Available: https://doi.org/10.7554/eLife.54055.

[13] A. Afrasiyabi, D. Bhaskar, E. L. Busch, *et al.*, “Latent representation learning for multimodal brain activity translation”, in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5. DOI: 10.1109/ICASSP49660.2025.10887834.

[14] X. Gao, F. Shi, D. Shen, and M. Liu, “Task-induced pyramid and attention gan for multimodal brain image imputation and classification in alzheimer’s disease”, *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 36–43, 2022. DOI: 10.1109/JBHI.2021.3097721.

[15] H. Lan, the Alzheimer Disease Neuroimaging Initiative, A. W. Toga, and F. Sepehrband, “Three-dimensional self-attention conditional gan with spectral normalization for multimodal neuroimaging synthesis”, *Magnetic Resonance in Medicine*, vol. 86, no. 3, pp. 1718–1733, 2021. DOI: https://doi.org/10.1002/mrm.28819. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28819. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.28819.

[16] K. I. Vaden, M. Gebregziabher, Dyslexia Data Consortium, and M. A. Eckert, “Fully synthetic neuroimaging data for replication and exploration”, *NeuroImage*, vol. 223, p. 117 284, 2020, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2020.117284. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811920307709.