Title: Structure-Preserved Image Reconstruction from Brain Recordings

Authors and Affiliations:

Zijiao Chen[1*], Jonathan Xu[4*], Jiaxin Qing[5], Ruilin Li[1], Juan Helen Zhou[1,2,3#]

1. Centre for Sleep and Cognition & Centre for Translational Magnetic Resonance Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
2. Integrative Sciences and Engineering Programme (ISEP), NUS Graduate School, National University of Singapore, Singapore, Singapore
3. Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore
4. Cheriton School of Computer Science, University of Waterloo
5. The Department of Information Engineering, The Chinese University of Hong Kong

Categories:

1. AI & Machine Learning **(primary)**, AI Image Reconstruction **(secondary)**
2. fMRI **(primary)**, fMRI analysis **(secondary)**

Keywords: fMRI; brain decoding; fMRI-Image reconstruction; self-supervised learning; deep learning; image generation

Synopsis:

Driven by the need for more accurate decoding of perceptual experiences from brain recordings, this study addresses the limitations imposed by traditional one-dimensional analysis of fMRI data. Our goal was to maintain the spatial information of fMRI data by employing a two-dimensional cortical surface-based analytical framework, aiming to enhance fMRI representation of neural responses. After converting volumetric fMRI scans into 2D surface data, we use Vision Transformers and Latent Diffusion Models to learn from fMRI and generate precise image. Our model achieved better representation abilities and generated clearer, more accurate high-resolution natural images from 2D fMRI inputs.

Impact:

By preserving spatial information in fMRI analysis and self-supervised learning stage, this study not only establishes innovative methods for fMRI representation learning and perception reconstruction but also paves the way for sophisticated multimodal brain models in the future.

Introduction:

Decoding and reconstructing perceptual experiences from brain recordings has gathered lots of attention these two years, driven by significant advances in neuroimaging data and machine learning techniques[1,2]. Functional Magnetic Resonance Imaging (fMRI), with its ability to capture brain activity patterns with high spatial resolution, presents a rich modality for such explorations. Traditional methods have treated this volumetric data by vectorize it into a one-dimensional format, which leads to losing important spatial information in neural activation. Recognizing the brain's inherent multi-dimensional structure, this study seeks to retain the spatial integrity of the fMRI data by utilizing a two-dimensional cortical surface-based framework during the decoding process. By doing so, we aim to enhance the fidelity of reconstructed perceptual experiences, while also enriching our understanding of the brain's representational mechanisms.

Methods:

In our study, we adopted a novel approach to preserve the spatial structure inherent in fMRI data by transforming 3D volumetric scans into 2D representations via a projection method aligned with the fsaverage brain template, as shown in Figure 2A. This conversion was crucial to maintaining the brain's spatial hierarchies and allowed us to use advanced image processing techniques tailored for 2D data. Following this projection, Figure 2B demonstrates that the 2D fMRI data were systematically divided into discrete patches, which were then tokenized into large embeddings to facilitate the subsequent neural network training.

Capitalizing on the strengths of Vision Transformer (ViT) architectures, we designed a self-supervised learning regimen where a ViT-based autoencoder was tasked with reconstructing randomly masked segments of the fMRI data[3, 4]. This training was pivotal in enabling the model to learn and interpret the latent structures within the fMRI data, which is essential for accurate image reconstruction. Especially when the paired data is limited. With this self-supervised learning scheme, the model can be trained in a large fMRI-only dataset first and then tuned in the downstream target dataset. Later on, we bridged the representation gap between the fMRI embeddings and the needs of a Latent Diffusion Model (LDM)[5] with a latent dimension projector as shown in Figure 2C, ensuring that the fMRI-derived information could be effectively utilized for image generation.

Our methodology involved an extensive pre-training phase using the expansive Human Connectome Project dataset[6] to establish a broad understanding of brain activity patterns. Further refinement was achieved through finetuning on the Natural Scenes Dataset[7, 8], which paired fMRI data with corresponding images to increase the model's reconstruction accuracy. The final stage of our research involved a rigorous evaluation of the model's performance, focusing on its capacity to generate perceptually coherent images from new fMRI data, with success measured against various quantitative metrics.

Results:

The empirical outcomes of this study were twofold. Initially, during the fMRI representation learning phase, our model demonstrated an enhanced ability to capture and reconstruct the structural complexities of fMRI data. Notably, as visualized in our results (Figure 3), the correlation coefficients between the reconstructed and the original fMRI data were significantly higher compared to previous one-dimensional approaches[1], with Pearson Correlation Coefficient (PCC) of 0.92 vs. 0.62. This suggests a superior retention of spatial information, which is integral for accurate brain decoding. Secondly, as shown in Figure 4A, when tasked with generating high-resolution images from the 2D fMRI input, the model yielded outputs with exceptional clarity and detail. The quantitative metrics (shown in Figure 4B) used for evaluating image quality indicated the efficacy of our method in terms of semantic understanding and pixel-level similarity.

Discussion:

The use of a two-dimensional approach in processing and interpreting fMRI data has proven to be more than a mere technical preference; it aligns closely with the inherent structure and functional organization of the brain. This study extends its inquiry into higher-dimensional spaces, enriching the generalizability and interpretability of the derived neural representations. The marked improvements in our results underscore the importance of respecting the brain's spatial hierarchies in computational modelling. This not only paves the way for more accurate reconstructions of visual stimuli but also suggests potential for a deeper, more nuanced understanding of the neural substrates of perception. Consequently, our findings show the importance of retaining structural information in the design of neural decoding algorithms, which could be pivotal in advancing both theoretical and applied neuroscience.

Conclusion:

We propose to preserve spatial information of neural responses by co-operating higher dimensional analysis. Our methodology not only sets a new standard for image reconstruction from brain activity but also pave the way for future applications in video and audio reconstruction, disease prediction, and brain-behavior analyses. It lays the groundwork for more sophisticated brain foundation models with significant clinical and research implications.

Reference:

1. Chen, Z., Qing, J., Xiang, T., Yue, W. L., & Zhou, J. H. (2023). Seeing beyond the brain: Conditional diffusion model with sparse masked modeling for vision decoding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 22710-22720).
2. Takagi, Y., & Nishimoto, S. (2023). High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14453-14463).
3. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10012-10022).
4. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009).
5. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).
6. Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Wu-Minn HCP Consortium. (2013). The WU-Minn human connectome project: an overview. *Neuroimage*, *80*, 62-79.
7. Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., ... & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, *25*(1), 116-126.
8. Lane, C. (2023). NSD-Flat: Pre-processed brain activity flat maps from the Natural Scenes Dataset. Retrieved from https://huggingface.co/datasets/clane9/NSD-Flat.

Figure 1. **Study Overview**.

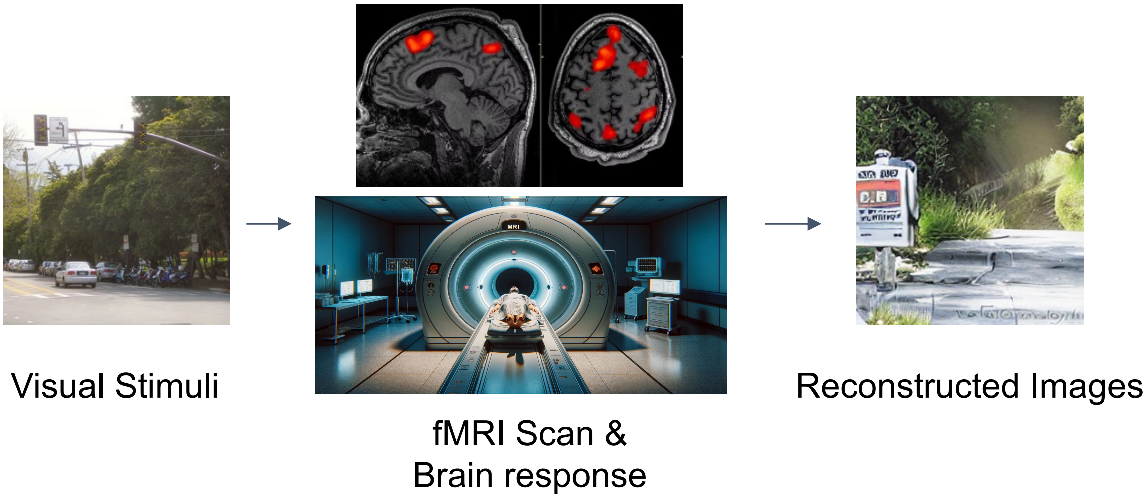

Figure 2. **Model Pipeline for Visual Reconstruction from Brain Activity.**
Panel A illustrates the transformation of brain volume data into a 2D cortical representation, maintaining the neural structure and spatial dynamics. Panel B describes our masked brain modelling approach, using a self-taught Vision Transformer autoencoder to interpret the concealed segments of 2D fMRI data. In Panel C, a Latent Diffusion Model translates neural patterns into detailed visual stimuli.

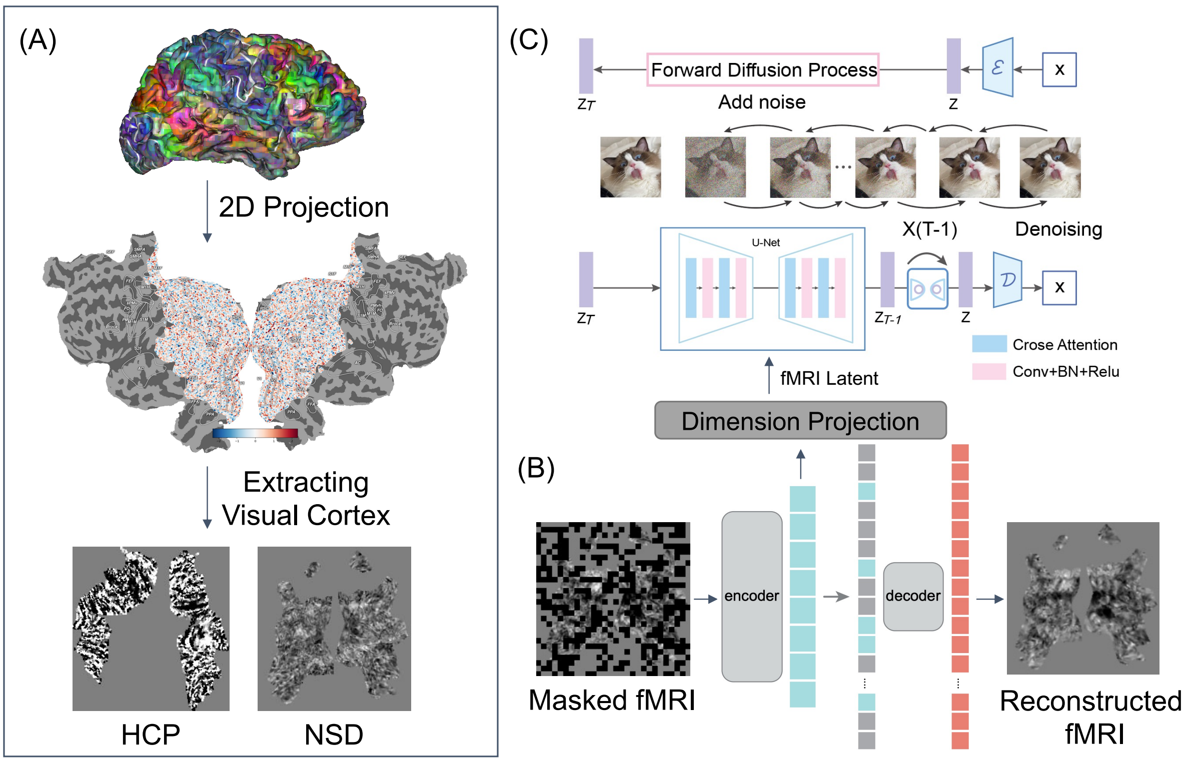

Figure 3: **Visualization of Masked fMRI Reconstruction**. This figure demonstrates our model's capability to accurately reconstruct masked segments of fMRI data. In the pre-training phase utilizing the Human Connectome Project dataset, the model achieved a Pearson Correlation Coefficient (PCC) of 0.82 and 0.87 for Natural Scenes Dataset.

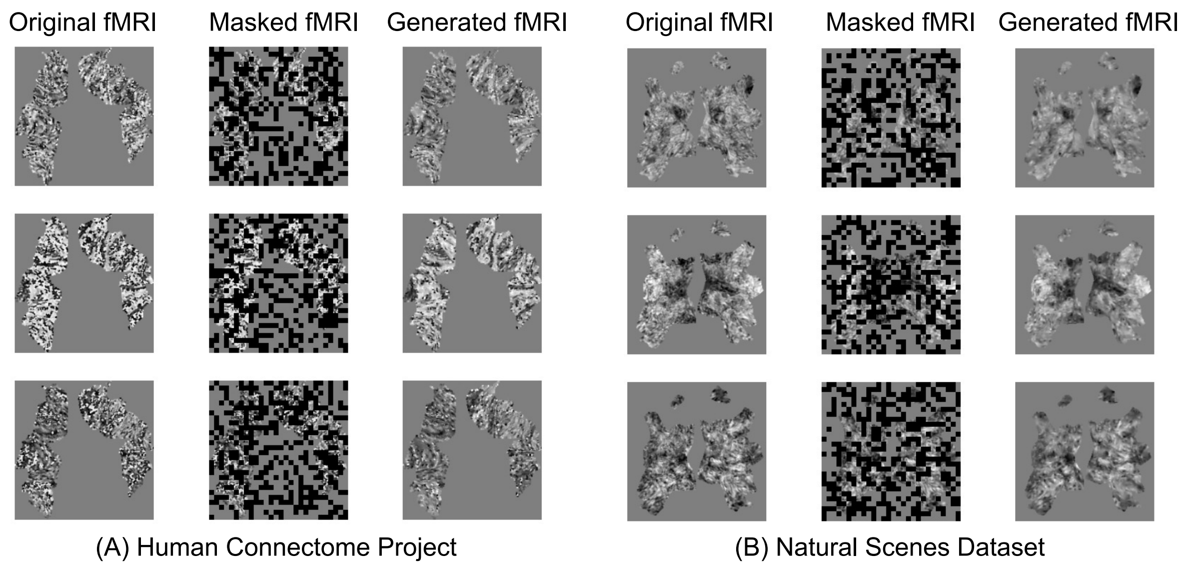

Figure 4. **Efficacy of Latent Diffusion Model in Image Reconstruction from fMRI Data.** The figure presents the ability of our latent diffusion model in generating high-fidelity natural images from fMRI data, showcasing both quantitative accuracy and visual congruence.