Probing Human Vision via an Image-to-EEG Encoding Model

Zitong Lu (lu.2637@osu.edu)
Department of Psychology, The Ohio State University
Columbus, OH 43210 USA

Julie D. Golomb (golomb.9@osu.edu)
Department of Psychology, The Ohio State University
Columbus, OH 43210 USA

Abstract:

Understanding the complex interplay between visual stimuli and brain activity has been a focal point in cognitive neuroscience. The recent advent of artificial intelligence (AI) provides novel insights for experimental and computational neuroscience research. In this study, we developed a pioneering encoding framework, called "Img2EEG", as an innovative tool for investigating visual mechanisms. Trained on a large-scale EEG dataset of natural images at the individual subject level, Img2EEG effectively learns individualized brain-optimized features and generates highly realistic EEG signals given any image input. Using Img2EEG, we can track the temporal dynamics underlying visual processes, and uncover possible mechanisms of individual differences in visual perception. Moreover, feeding Img2EEG novel sets of images distinctly varied from its original training dataset, the artificially-generated EEG signal reproduced classic face-specific 'N170' ERP and object feature multivariate pattern analysis results. Furthermore, our Img2EEG encoding model can also conduct EEG-to-image zero-shot retrieval task, outperforming current state-of-the-art EEG decoding models. Overall, Img2EEG mapping from visual inputs to high temporal resolution brain signals offers novel and powerful approaches to probe human visual representations.

Keywords: Encoding Model, EEG, Visual Perception

Introduction

Understanding the complex interplay between visual stimuli and human brain activity has been a focal point in cognitive neuroscience. However, traditional techniques still face limitations in deciphering how human brains process visual information: on one hand, analyses of neural data obtained via non-invasive EEG or fMRI techniques often struggle to deeply and directly probe the brain's internal processing of different visual information; on the other hand, obtaining extensive brain recordings from human subjects viewing a large number of image stimuli is challenging due to constraints in experimental time and costs.

Benefiting from the advancements in deep learning (Lecun et al., 2015) and the availability of large neuroimaging datasets (Allen et al., 2022; Hebart et al., 2023), we are now able to better simulate and predict brain activity, overcoming the challenges to gain a deeper understanding of the brain's visual mechanisms. Compared to traditional linear or inverted encoding models (Naselaris et al., 2011, 2012; Samaha et al., 2016; Scotti et al., 2022), deep learning-based image-to-brain encoding models include various visual features and generate brain signals more accurately.

In this study, we introduce a high-performance image-to-EEG encoding model, called "Img2EEG". This model not only generates realistic EEG signals but also captures richer visual features and helps us better understand the dynamic processes involved in the brain's visual processing.

Methods and Results

Img2EEG contains three different modules (low-level vision, high-level semantic, and integration) corresponding to our brain's internal processing (Figure 1A). The low-level vision module is composed of three recurrent convolutional layers from pretrained CORnet-S (Kubilius et al., 2018, 2019) and a nonlinear low-level visual encoder. The high-level semantic module contains three feature extractors corresponding to visual-language (based on CLIP (Radford et al., 2021)), object concept (based on GloVe (Pennington et al., 2014)), and image description information (based on BLIP-2 (J. Li et al., 2023) and MPNet (K. Song et al., 2020)) and a nonlinear high-level semantic encoder. The integration module includes three nonlinear full-connected integration encoders which culminate in generated EEG signals as output.

We trained ten different Img2EEGs on ten human subjects' EEG signals when they viewed 16540 natural images from THINGS EEG2 (Gifford et al., 2022)

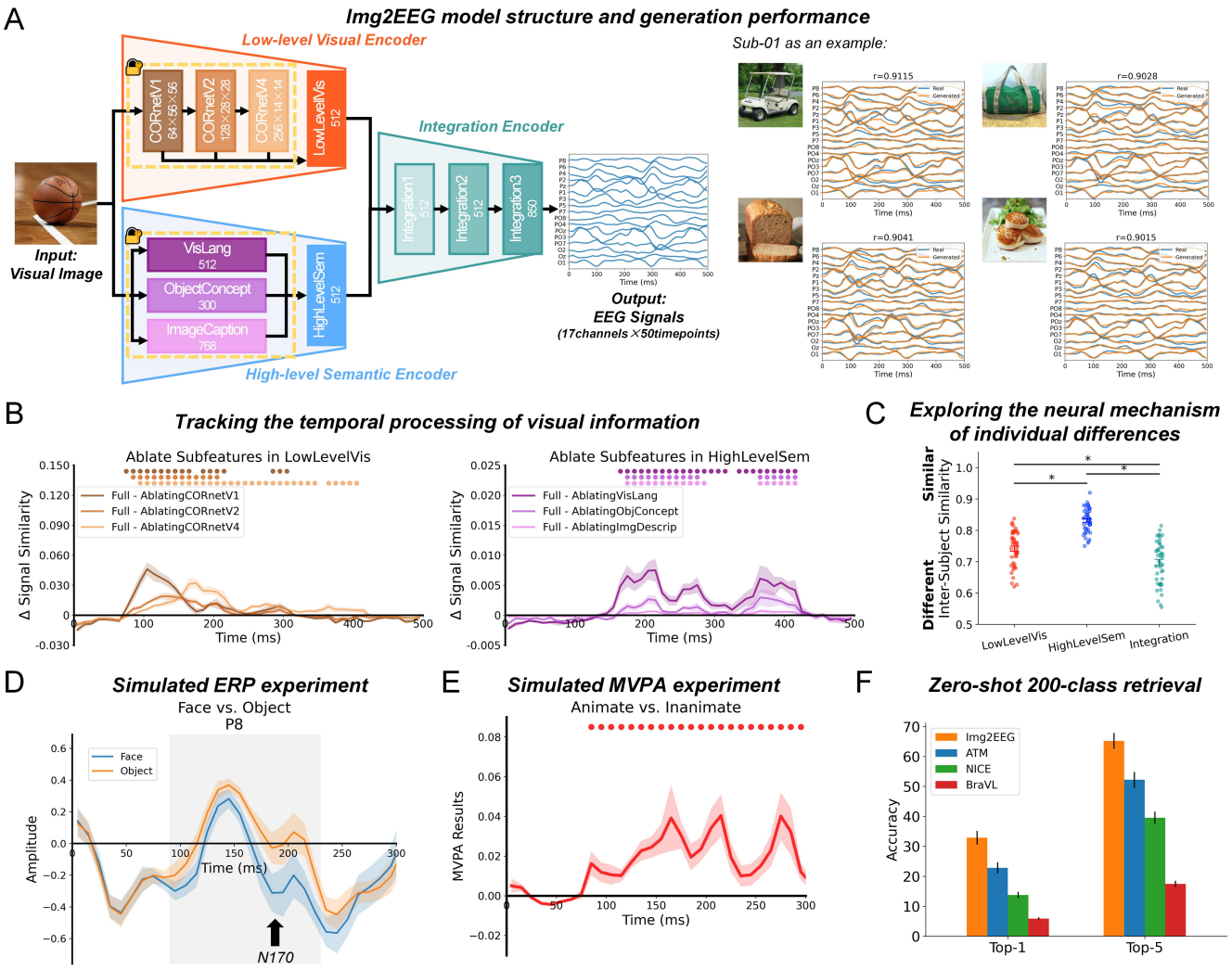

Figure 1: (A) Architecture of Img2EEG and some examples of generated EEG signals from Img2EEG. (B) Temporally-specific signal similarity decreases when ablating different features in Img2EEG. Shaded area reflects ±SEM. Circles indicate timepoints where ablating a given feature results in a significant decrease of generated vs real EEG signal similarity (p<.05). (C) Individual differences in Img2EEG. Each dot indicates a pair of two subjects. The asterisk indicates a significant difference (p<.05). (D) Simulated face versus object ERP results. Orange or blue shaded area reflects ±SEM. Grey shaded area indicates the significant time-window of ERP differences (p<.05). (E) Simulated animate versus inanimate MVPA results. Shaded area reflects ±SEM. Red dots at the top indicate the timepoints where MVPA results were significantly greater than zero (p<.05). (F) Zero-shot 200-class retrieval performance of Img2EEG and other state-of-art EEG-to-image decoding models. Error bar reflects ±SEM.

training set. We then tested the Img2EEGs on a test set of 200 images which had not been presented at all during the training process, coming from entirely novel (untrained) object categories. Img2EEG shows high performance at generating realistic EEG signals (Figure 1A).

Since Img2EEG incorporates multi-level retrievable visual features, we can employ a “feature ablation” approach to sequentially ablate corresponding modules in the model. By comparing the signals generated by these ablated models with those produced by the complete model in terms of similarity to the real signals, we can thereby track the temporal processing of visual information (Figure 1B). Also, we can compare these internal feature representations across ten models to infer the neural mechanism of individual differences (Figure 1C).

Excitingly, we can also apply Img2EEG to conduct both ERP and MVPA experiments on simulated EEG data. By simply inputting various stimulus images to Img2EEG and analyzing the generated EEG signals, we observe the classical face-specific “N170” (Bentin et

al., 1996; Rossion & Jacques, 2012) ERP component (Figure 1D), as well as temporally consistent decoding of object animacy (Cichy et al., 2014; Khaligh-Razavi et al., 2018; Wang et al., 2022) (Figure 1E).

Finally, Img2EEG outperforms current state-of-art EEG-to-image decoding models (Du et al., 2023; D. Li et al., 2024; Y. Song et al., 2024) in 200-class zero-shot retrieval based on THINGS EEG2 test set, which can accurately decode which image the human subject sees from EEG signals (Figure 1F).

Conclusion

We propose Img2EEG, a high-performance image-to-EEG encoding model, as a novel framework to probe human vision. Not only can Img2EEG effectively generate highly realistic EEG signals given image input, but also it provides novel approaches to deeply understand human brain internal representations and offer insights to interdisciplinary areas, such as cognitive neuroscience, brain-computer interface, and artificial intelligence.

Acknowledgments

This work was supported by research grants from the National Institutes of Health (R01-EY025648) and from the National Science Foundation (NSF 1848939).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462.

Du, C., Fu, K., Li, J., & He, H. (2023). Decoding Visual Neural Representations by Multimodal Learning of Brain-Visual-Linguistic Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(9), 10760–10777.

Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022). A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, *264*, 119754.

Hebart, M. N., Contier, O., Teichmann, L., Rockter, A. H., Zheng, C. Y., Kidder, A., Corriveau, A., Vaziri-Pashkam, M., & Baker, C. I. (2023). THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *ELife*, *12*, e82580.

Khaligh-Razavi, S.-M., Cichy, R. M., Pantazis, D., & Oliva, A. (2018). Tracking the Spatiotemporal Neural Dynamics of Real-world Object Size and Animacy in the Human Brain. *Journal of Cognitive Neuroscience*, *30*(11), 1559–1576.

Kubilius, J., Schrimpf, M., Kar, K., Rajalingham, R., Hong, H., Majaj, N. J., Issa, E. B., Bashivan, P., Prescott-Roy, J., Schmidt, K., Nayebi, A., Bear, D., Yamins, D. L. K., & Dicarlo, J. J. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. *Advances in Neural Information Processing Systems (NeurIPS)*, *32*.

Kubilius, J., Schrimpf, M., Nayebi, A., Bear, D., Yamins, D. L. K., & DiCarlo, J. J. (2018). CORnet: Modeling the Neural Mechanisms of Core Object Recognition. *BioRxiv*.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Li, D., Wei, C., Li, S., Zou, J., & Liu, Q. (2024). Visual Decoding and Reconstruction via EEG Embeddings with Guided Diffusion. *ArXiv*.

Li, J., Li, D., Savarese, S., & Hoi, S. (2023). BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. *Proceedings of the International Conference on Machine Learning (ICML)*, 19730–19742.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.

Naselaris, T., Stansbury, D. E., & Gallant, J. L. (2012). Cortical representation of animate and inanimate objects in complex natural scenes. *Journal of Physiology Paris*, *106*(5–6), 239–249.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *Proceedings of the International Conference on Machine Learning (ICML)*.

Rossion, B., & Jacques, C. (2012). The N170: Understanding the Time Course of Face Perception in the Human Brain. In *The Oxford handbook of event-related potential components* (Oxford Uni, pp. 115–141). Oxford University Press.

Samaha, J., Sprague, T. C., & Postle, B. R. (2016). Decoding and Reconstructing the Focus of Spatial Attention from the Topography of Alpha-band Oscillations. *Journal of Cognitive Neuroscience*, *28*(8), 1090–1097.

Scotti, P. S., Chen, J., & Golomb, J. D. (2022). An enhanced inverted encoding model for neural reconstructions. *BioRxiv*.

Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). MPNet: Masked and Permuted Pre-training for Language Understanding. *Advances in Neural Information Processing Systems*, *33*, 16857–16867.

Song, Y., Liu, B., Li, X., Shi, N., Wang, Y., & Gao, X. (2024). Decoding Natural Images from EEG for Object Recognition. *International Conference on Learning Representations (ICLR)*.

Wang, R., Janini, D., & Konkle, T. (2022). Mid-level feature differences support early animacy and object size distinctions: Evidence from electroencephalography decoding. *Journal of Cognitive Neuroscience*, *34*(9), 1670–1680.