Connecting Minds and Machines: Decoding Brain Activity during vision perception through the Lens of AI

Matteo Ferrante
Tommaso Boccato
matteo.ferrante@uniroma2.it
University or Rome, Tor Vergata
Rome, Italy

Nicola Toschi
University of Rome, Tor Vergata
Rome, Italy

ABSTRACT

The human brain processes an immense volume of visual information daily, relying on intricate neural mechanisms to perceive and interpret these stimuli. Recent advancements in functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG) have enabled scientists to decode this visual information from brain activity patterns. This research aims to integrate multiple neural data modalities to enhance the decoding of brain activity into meaningful images and text. Utilizing cutting-edge image captioning models, contrastive learning, and latent diffusion models, our approach shows significant progress in brain decoding. We employ datasets like the Natural Scenes Dataset (NSD) and THINGS-MEG, and present methods that outperform existing techniques in brain captioning and image reconstruction. Our findings highlight the potential for future research and applications in brain-computer interfaces and clinical diagnostics.

ACM Reference Format:
Matteo Ferrante, Tommaso Boccato, and Nicola Toschi. 2018. Connecting Minds and Machines: Decoding Brain Activity during vision perception through the Lens of AI. In *Proceedings of (Conference acronym KDD).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In our everyday lives, we primarily interact with the world through our sense of vision. We are constantly able to recognize our surroundings, categorize objects, infer relationships, and extract semantic meanings from what we see. Our brain is perpetually processing this sensory information, generating useful representations that facilitate further inferences and thoughts. Our research addresses two main questions: how can we understand and decode these representations, and how well can we measure these representations using non-invasive methods? To address these questions, we developed AI pipelines to map stimuli representations between the latent spaces of large pretrained models and the brain. Our work explores various tasks, including decoding (reconstructing the actual seen image from neural activity), encoding (predicting brain activity from the stimulus using computational models), and neural modality conversion (predicting, for example, fMRI activity from EEG). Advancing our understanding of the human brain through artificial intelligence is a primary goal of our research, while the ultimate objective is to decipher brain activity and develop powerful brain-computer interfaces (BCIs). In this work, we focused on image perception measured using fMRI because it is non-invasive and provides high spatial resolution compared to other modalities [2]. The cornerstone of our research is the idea that the brain extracts semantic and detailed information from sensory outputs, and these representations are nonlinear with respect to the stimuli [12]. Furthermore, these representations can be captured by non-invasive neural recordings to some extent. Based on this hypothesis, we demonstrate how deep datasets (few subjects, extensive data per subject) coupled with large pretrained models can be used to map between these two spaces (brain and latent space of these models) using simple regularized methods. Our research journey began by showing that decoding semantics is possible and progressively became more ambitious. We have shown that it is possible to decode text and images from the visual cortex [6] and to overcome individual differences [5], achieving cross-subject decoding [7]. Recently,

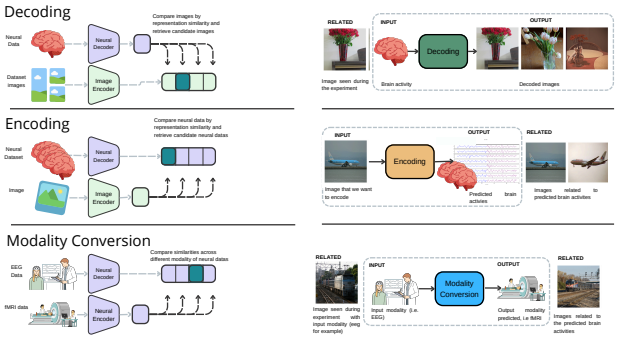

Figure 1: Schematic examples of the tasks involved in this work include decoding, encoding, and modality conversion. The decoding pipeline involves training a neural decoder to predict the latent representation of images, starting from brain activity to retrieve or generate plausible images. The encoding pipeline has the reverse objective: training an image-to-brain encoder to predict brain activity based on image content. Finally, the modality conversion experiment focuses on reconstructing or retrieving one type of neural data from another, such as converting EEG data to fMRI data.

Conference acronym KDD, 2024, Barcellona, SP

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

our research has expanded to align different datasets and modalities to encode, decode, and convert modalities between EEG, MEG, and fMRI [4, 8]. Fig 1 shows some of the task involved in this research.

2 METHODS

The key pillars of this research are: **adopting an AI-based approach to neural data analysis** and **mapping brain activity to latent representations** of pretrained models. Unlike traditional methods that use shallow data from many patients, our focus is on deep datasets with fewer subjects but a large number of stimuli per subject. We explore advanced methods like convolutional neural networks, multimodal models, knowledge distillation, and contrastive learning to map brain activity to latent representations of stimuli. This enables us to reconstruct perceived images and predict brain activity based on visual stimuli. The datasets used in this research include Generic Object Decoding [11], BOLD5000 [3], and Natural Scenes Dataset for fMRI [1], ImageNet-EEG [14] and THINGS-EEG2 for EEG [10], and THINGS-MEG [9] for MEG, covering complex scenes with 1200 to 10000 stimuli per subject. This provides rich, detailed data for advanced AI models. The study contributions are as follows:

Multimodal Decoding of Human Brain Activity [5]: Using fMRI data from the NSD, this study decodes brain activity into meaningful images and captions with the Generative Image-to-text Transformer (GIT) [15] and latent diffusion models [13], showcasing the potential of AI in understanding brain functions and reconstructing textual and visual descriptions of what was seen during the experiment.

We further extended this framework to work across subjects and generalize beyond subjective differences using a linear functional alignment technique [7].

Finally, in **Towards Neural Foundation Models for Vision** [8]: In this study we use contrastive learning to align neural data with visual stimuli representations across EEG, MEG, and fMRI datasets. It performs decoding, encoding, and modality conversion tasks, creating unified representations that offer deeper insights into brain activity. These experiments assess the models' performance in decoding visual information from neural data and encoding images into neural representations, including predicting fMRI activity from EEG data. The models demonstrate high accuracy in both tasks, confirming their robustness and versatility. This research bridges the gap between neural activity and AI-based representations, paving the way for advancements in brain-computer interfaces and cognitive neuroscience.

3 RESULTS

Our research demonstrated significant advancements in decoding brain activity using advanced AI models. Across multiple experiments involving fMRI, EEG, and MEG datasets, we achieved high accuracy in both decoding and encoding tasks. In [5] starting from fMRI data from the Natural Scenes Dataset (NSD), our model, which combines the Generative Image-to-text Transformer (GIT) and latent diffusion models, accurately decoded brain activity into meaningful images and textual descriptions. This supports the model's capability to capture and reproduce complex semantic features of visual stimuli in their neural representations. We followed up this work with [7] where we demonstraed that using functional alignment is possible to align neural representations decode images even across subjects, saving up to 90% of the scanning time. We also found evidence that is possible to perform decoding across different datasets, acquired with different machines and paradigms. Finally, by aligning neural data with visual stimuli representations through contrastive learning, we tested our framework across EEG, MEG, and fMRI datasets. The results demonstrated the model's ability to capture semantic information and perform tasks such as decoding, encoding, and modality conversion. Our experiments on projecting neural data into a common representation space and retrieving visually related images confirmed the model's robustness. We achieved high accuracy in both decoding visual information from neural data and encoding images into neural representations. The results indicate that it is feasible to predict fMRI activity from EEG data, further showcasing the versatility of our approach. Overall, the results affirm that our computational approach effectively captures and decodes various semantic features from neural data, aligning well with human evaluations and providing a solid foundation for further research in brain-computer interfaces. Details about all the pathway of this research can be found at https://shorturl.at/Ezw5i along with reconstructed images from brain activity, methodological details, full papers and code for reproducibility.

4 DISCUSSION AND CONLUSIONS

This research demonstrates the potential of advanced AI models to decode complex semantic information from neural activity, with significant implications for brain-computer interfaces (BCIs). Our high accuracy in decoding and encoding tasks shows that AI can effectively interpret and reconstruct visual experiences, validating our computational methods against human evaluations. However, the study has limitations, including a limited number of subjects, focus on visual stimuli, and substantial computational resources required. Future research should extend these techniques to other modalities such as videos, language, imagery, and music, and explore cross-modal decoding and encoding. Developing real-time decoding systems, creating adaptive models personalized to individual users, and focusing on underrepresented stimuli like abstract concepts and emotions will further enhance the applicability and accuracy of BCIs. By addressing these areas, future work can overcome current limitations and expand the scope of cognitive neuroscience and AI applications. Furthermore, while we use AI models to simulate brain activity, these models may not fully capture the complexity of our minds. When decoding information from the brain using this pipeline, we must be mindful of these limitations and respect neural privacy. Decoded information may reflect model errors and training data biases rather than the human subject's true brain activity, so it is crucial not to regard these outputs as definitive representations of brain function.

However, our research has demonstrated significant results across various datasets and configurations, providing evidence that, to some extent, these models can effectively simulate brain activity. Additionally, vision perception can be decoded from neural activity in a non-invasive manner.

REFERENCES

[1] Emily J. Allen, Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, J. Benjamin Hutchinson, Thomas Naselaris, and Kendrick Kay. 2022. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience* 25, 1 (01 Jan 2022), 116–126. https://doi.org/10.1038/s41593-021-00962-x

[2] Thomas Bazeille, Elizabeth DuPre, Hugo Richard, Jean-Baptiste Poline, and Bertrand Thirion. 2021. An empirical evaluation of functional alignment using inter-subject decoding. *NeuroImage* 245 (2021), 118683. https://doi.org/10.1016/j.neuroimage.2021.118683

[3] Nadine Chang, John A. Pyles, Austin Marcus, Abhinav Gupta, Michael J. Tarr, and Elissa M. Aminoff. 2019. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Scientific Data* 6, 1 (06 May 2019), 49. https://doi.org/10.1038/s41597-019-0052-3

[4] Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. 2023. Decoding visual brain representations from electroencephalography through Knowledge Distillation and latent diffusion models. arXiv:2309.07149 [eess.SP]

[5] Matteo Ferrante, Tommaso Boccato, Furkan Ozcelik, Rufin VanRullen, and Nicola Toschi. 2023. Multimodal decoding of human brain activity into images and text. In *UniReps: the First Workshop on Unifying Representations in Neural Models*. https://openreview.net/forum?id=rGCabZfV3d

[6] Matteo Ferrante, Tommaso Boccato, and Nicola Toschi. 2023. Semantic Brain Decoding: from fMRI to conceptually similar image reconstruction of visual stimuli. arXiv:2212.06726 [cs.CV]

[7] Matteo Ferrante, Tommaso Boccato, and Nicola Toschi. 2023. Through their eyes: multi-subject Brain Decoding with simple alignment techniques. arXiv:2309.00627 [q-bio.NC]

[8] Matteo Ferrante, Tommaso Boccato, and Nicola Toschi. 2024. Towards neural foundation models for vision: Aligning EEG, MEG and fMRI representations to perform decoding, encoding and modality conversion. In *ICLR 2024 Workshop on Representational Alignment*. https://openreview.net/forum?id=nxoKCdmteM

[9] Martin N Hebart, Oliver Contier, Lina Teichmann, Adam H Rockter, Charles Y Zheng, Alexis Kidder, Anna Corriveau, Maryam Vaziri-Pashkam, and Chris I Baker. 2023. THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. *eLife* 12 (feb 2023), e82580. https://doi.org/10.7554/eLife.82580

[10] Martin N Hebart, Adam H Dickter, Alexis Kidder, Wan Y Kwok, Anna Corriveau, Caitlin Van Wicklin, and Chris I Baker. 2019. THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLoS One* 14, 10 (Oct. 2019), e0223792.

[11] Tomoyasu Horikawa and Yukiyasu Kamitani. 2017. Generic decoding of seen and imagined objects using hierarchical visual features. *Nature Communications* 8, 1 (Aug. 2017), 15037. https://doi.org/10.1038/ncomms15037

[12] Subba Reddy Oota, Manish Gupta, Raju S. Bapi, Gael Jobard, Frederic Alexandre, and Xavier Hinaut. 2023. Deep Neural Networks and Brain Alignment: Brain Encoding and Decoding (Survey). http://arxiv.org/abs/2307.10246 arXiv:2307.10246 [cs, q-bio].

[13] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. https://doi.org/10.48550/ARXIV.2112.10752

[14] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah. [n. d.]. Deep Learning Human Mind for Automated Visual Classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Honolulu, HI, 2017-07). IEEE, 4503–4511. https://doi.org/10.1109/CVPR.2017.479

[15] Jianfeng Wang, Zhengyuan Yang, Xiaowei Hu, Linjie Li, Kevin Lin, Zhe Gan, Zicheng Liu, Ce Liu, and Lijuan Wang. [n. d.]. GIT: A Generative Image-to-text Transformer for Vision and Language. arXiv:2205.14100 [cs] http://arxiv.org/abs/2205.14100