fMRI vision reconstruction methods robustly generalize to mental imagery

Reese Kneeland (rek@umn.edu)
Department of Computer Science, University of Minnesota
Minneapolis, MN 55455 USA

Ghislain St-Yves (gstyves@umn.edu)
Department of Neuroscience, University of Minnesota
Minneapolis, MN 55455 USA

Jesse Breedlove (jbreedlo@umn.edu)
Department of Psychology, University of Minnesota
Minneapolis, MN 55455 USA

Kendrick Kay (kay@umn.edu)
Department of Radiology, University of Minnesota
Minneapolis, MN 55455 USA

Thomas Naselaris (nase0005@umn.edu)
Department of Neuroscience, University of Minnesota
Minneapolis, MN 55455 USA

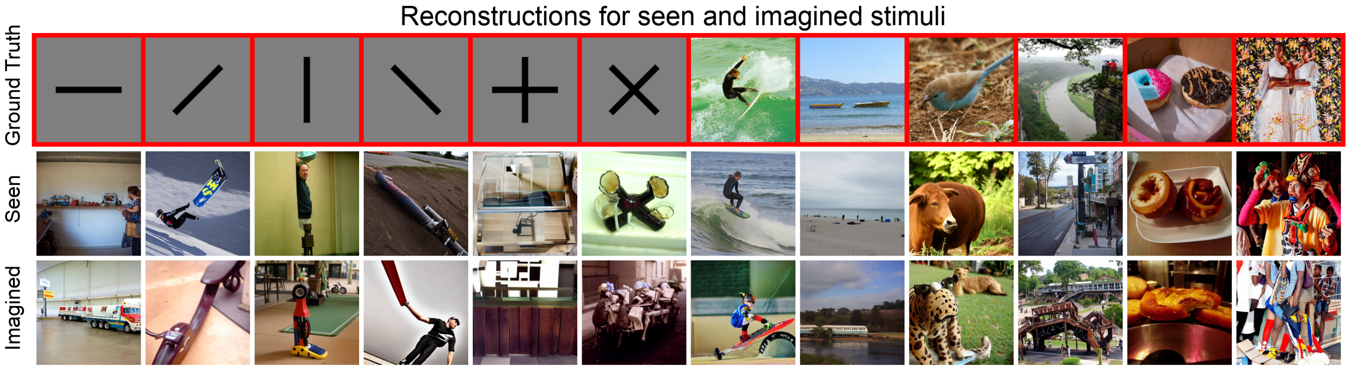

Figure 1: Qualitative comparison of reconstruction methods on simple (left) and complex (right) stimuli across both vision and mental imagery experiments. Reconstructions selected for the figure are the best samples for each stimulus as assessed by quantitative performance on a set of image feature metrics.

Abstract

Fueled by recent leaps in generative AI and the release of the Natural Scenes Dataset (NSD), researchers have been able to reconstruct seen images from human brain activity with unprecedented accuracy. If it were possible to extend these visual decoding methods to mental imagery, they could potentially be useful in clinical settings, e.g., by introducing new diagnostic tools for patients with traumatic brain injuries and mental health disorders, or by providing a communication channel to patients with locked-in syndrome. We tested the application of several recent vision decoding methods to brain activity collected during a special NSD scanning session in which subjects imagined a small set of memorized target stimuli. We show that most of these methods generalize robustly to mental imagery, yielding reconstructions of mental images that human raters consistently identify as corresponding to the target stimuli in a forced-choice task. Interestingly, we find that—by this identification accuracy measure—reconstructions of imagined natural scenes are of slightly better quality than reconstructions of much simpler seen stimuli of bars and crosses. Finally, we observe a strong correlation between stimuli that reconstruct better across vision and imagery trials, suggesting that further improvements in vision decoding methods will afford improvements to mental imagery decoding.

Keywords: mental imagery; decoding; vision; generative models; diffusion models; fMRI; NeuroAI

Introduction

A persistent challenge in modern medicine is diagnosing and treating disorders whose only symptoms are changes in private conscious experiences, including chronic pain, mental health disorders, and unresponsive patients with traumatic brain injuries. There are broad areas of research that would benefit tremendously from brain decoding methods that could transform a patient's internally generated visual experience into a picture that is viewable by external observers.

The open releases of deep learning models such as CLIP (Radford et al., 2021) and Stable Diffusion (Rombach, Blattmann, Lorenz, Esser, & Ommer, 2021), as well as large-scale functional magnetic resonance imaging (fMRI) datasets like the Natural Scenes Dataset (NSD) (Allen et al., 2022), where tens of thousands of images were shown to human subjects, has sparked the rapid developments of vision decoding methods that can reconstruct seen images from brain activity with impressive accuracy (Takagi & Nishimoto, 2023; Ozcelik & VanRullen, 2023; Scotti et al., 2023; Kneeland, Ojeda, St-Yves, & Naselaris, 2023b, 2023c, 2023a). However, to realize the utility of such approaches in clinical settings, decoding methods must be able to handle cases of mental imagery.

While mental imagery exhibits lower signal-to-noise ratios (SNR) (Roy, Breedlove, St-Yves, Kay, & Naselaris, 2023) than vision, the encoding of seen and mental images in visual cortex is similar enough (Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015) for vision to serve as a useful starting point for decoding mental imagery (St-Yves, Breedlove, Kay, & Naselaris, 2023).

Methods

We examine a small unreleased extension of NSD, hereby referred to as NSD-Imagery, in which each NSD subject completed an additional scanning session of mental imagery trials. All of the scanning and experimental procedures for NSD-Imagery remained the same as the main experiment in (Allen et al., 2022). The target stimuli set for this dataset consists of 6 simple stimuli (bars and crosses), and 6 complex stimuli (5 natural scenes and 1 artwork). These stimuli were presented 8 times during an initial set of vision runs, and in separate runs, the subjects were asked to imagine the stimuli 16 times each. We utilized these data to test the generalization capabilities of several contemporary vision decoding methods—which have already been fine-tuned for the NSD subjects—in the domain of mental imagery.

We applied vision decoding methods MindEye1(Scotti et al., 2023), MindEye2 (Scotti et al., 2024), Brain Diffuser (Ozcelik & VanRullen, 2023), and the Brain-Optimized Inference (BOI) reconstruction enhancement algorithm (Kneeland et al., 2023a) to the vision and imagery trials in the NSD-Imagery dataset. We evaluate the BOI algorithm using both the MindEye1 and Brain Diffuser as base methods.

Results

Figure 1 shows selected reconstructions of seen and mental images. Mental image reconstructions are qualitatively further from the target stimuli than seen image reconstructions, but capture many of the essential details of the target stimuli. Interestingly, simple stimuli reconstructions demonstrate strong structural coherence but deviate in semantic detail due to the naturalistic priors baked into the reconstruction methods, which impede generalization to new, out-of-distribution stimuli.

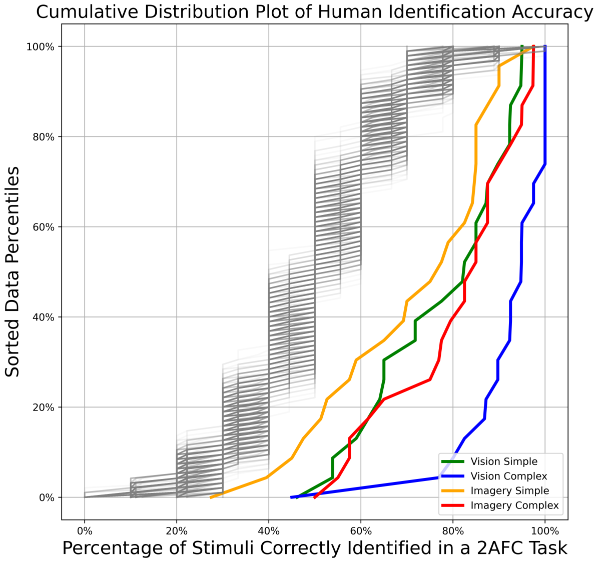

Figure 2: Reconstructed image identification. The percentage of stimuli that are correctly identified in a 2AFC task (x-axis) across percentiles of the sorted data (y-axis) by 266 human raters (averaged over all methods; null distribution in gray) for each stimulus type and modality (colored curves). Less area under the curve is better.

To quantify the quality of seen and mental image reconstructions in a setting relevant to our clinical motivations, we recruited human raters ($n=266$) to perform a 2-alternative forced choice (2AFC) judgment about whether a reconstruction was more similar to a target stimulus image than a randomly selected reconstruction of a different stimulus. In Figure 2, we observe that reconstruction of seen naturalistic images had the highest probability of being correctly identified on these trials, and reconstructions of imagined natural scenes were as roughly as likely to be correctly identified as reconstructions of seen simple stimuli. Given that all decoding methods were optimized for reconstructing seen naturalistic images, our results suggest that decoding accuracy is more reliant on the match between training and test image distributions than on the complexity of the stimuli or whether the image was seen or imagined. Thus, expanding training datasets to include a wider variety of stimuli could enhance reconstruction performance for these simple stimuli.

In a separate task of our online experiment, human raters were simultaneously presented with seen and mental image reconstructions of a target stimulus and used a set of sliders to indicate the similarity of each reconstruction to the target stimulus. Similarity scores of matched seen and mental image reconstructions (Figure 3) were significantly correlated. Averaging all ratings for each ground truth image and pooling data points across the 4 methods that generalize robustly to imagery (excluding MindEye2), the correlation between similarity for seen and mental image reconstructions was 0.56 (p < 0.001, n=48). This indicates that, on average, the similarity of reconstructed mental images to ground truth was more than half of the similarity of the reconstruction of a matched seen image. Collectively, these results suggest that improvements to decoders of seen images will likely translate to improvements in decoding matched mental images.

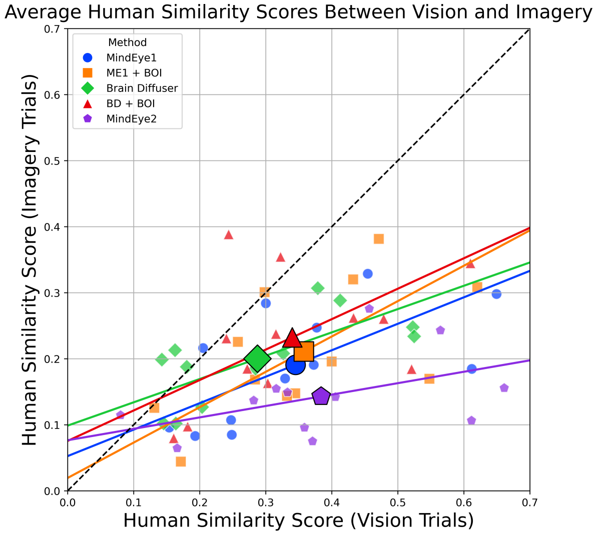

Figure 3: Correlation of seen and mental image reconstruction quality. Average similarity scores of seen image reconstructions (x-axis) and mental image reconstructions (y-axis). Plotted for each reconstruction method (color) and each stimulus (shapes; larger bold shape is the mean similarity across all stimuli), lines are the principal component of the variance for the joint similarity distribution (dashed line at unity indicates equal similarity).

Conclusion

We show that current vision decoding methods can robustly generalize to instances of mental imagery, opening the door to potential clinical applications that include diagnostics and communication for cognitively impaired or unresponsive patients. However, current research has several limitations, including the relatively small size of mental imagery datasets hindering domain-specific training, and the difficulty of generalizing these methods to new subjects with distinct neural codes. The MindEye2 initiative shows promise in multi-subject pretraining (Scotti et al., 2024), but underperforms in mental imagery tasks, indicating the need for further research in this space. As neuroimaging technologies advance toward decoding private visual experiences, ethical considerations demand that brain data be carefully protected and that any companies or organizations collecting such data be transparent with their use.

Acknowledgements

We would like to thank the authors of the seen image reconstruction methods utilized for this paper, including MindEye1 (Scotti et al., 2023), MindEye2 (Scotti et al., 2024), Brain Diffuser (Ozcelik & VanRullen, 2023), and Brain-Optimized Inference (Kneeland et al., 2023a) for providing open-source code for their methods. Thanks to Jasper van den Bosch for assisting with our behavioral experiment. Funding is provided by R01EY023384.

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., . . . Kay, K. (2022, January). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), 116–126.

Kneeland, R., Ojeda, J., St-Yves, G., & Naselaris, T. (2023a, December). *Brain-optimized inference improves reconstructions of fMRI brain activity.* arXiv. (arXiv:2312.07705 [cs, q-bio])

Kneeland, R., Ojeda, J., St-Yves, G., & Naselaris, T. (2023b). Reconstructing seen images from human brain activity via guided stochastic search. Conference on Cognitive Computational Neuroscience.

Kneeland, R., Ojeda, J., St-Yves, G., & Naselaris, T. (2023c, June). *Second Sight: Using brain-optimized encoding models to align image distributions with human brain activity.* arXiv.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215-228.

Ozcelik, F., & VanRullen, R. (2023, March). *Brain-Diffuser: Natural scene reconstruction from fMRI signals using generative latent diffusion.* arXiv.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., . . . Sutskever, I. (2021, 18–24 Jul). Learning transferable visual models from natural language supervision. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (Vol. 139, pp. 8748–8763). PMLR.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2021). High-resolution image synthesis with latent diffusion models. *CoRR*, *abs/2112.10752*. Retrieved from `https://arxiv.org/abs/2112.10752`

Roy, T. S., Breedlove, J., St-Yves, G., Kay, K., & Naselaris, T. (2023). Comparison of signal to noise in vision and imagery for qualitatively different kinds of stimuli. *Journal of Vision*, *23*(9), 5961.

Scotti, P. S., Banerjee, A., Goode, J., Shabalin, S., Nguyen, A., Ethan, C., . . . Abraham, T. M. (2023). Reconstructing the mind's eye: fMRI-to-image with contrastive learning and diffusion priors. In *Thirty-seventh conference on neural information processing systems.*

Scotti, P. S., Tripathy, M., Villanueva, C. K. T., Kneeland, R., Chen, T., Narang, A., . . . Abraham, T. M. (2024). *Mindeye2: Shared-subject models enable fmri-to-image with 1 hour of data.*

St-Yves, G., Breedlove, J., Kay, K., & Naselaris, T. (2023). Do better models of fmri visual response better predict mental imagery responses? Conference on Cognitive Computational Neuroscience.

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).