Perceptogram: Interpreting Visual Percepts from EEG

Teng Fei[a], Srinivas Ravishankar[a], Abhinav Uppal[b], Ian Jackson[a], David Wang[c], Virginia R. de Sa[a,d]
Departments of [a]Cognitive Science, [b]Bioengineering, [c]Computer Science and Engineering, [d]Halıcıoğlu Data Science Institute
{tfei, srravishankar, auppal, ijackson, dyw001, desa}@ucsd.edu
University of California San Diego, 9500 Gilman Dr., La Jolla, CA, 92093, USA

Abstract

Recent advances in EEG-based visual decoding utilize diffusion models to generate realistic images from neural activity. Typically, these methods project EEG signals into latent spaces–most commonly, Contrastive Language–Image Pretraining (CLIP)–which define visuosemantic features for subsequent image reconstruction. Prior methods rely on deep and opaque models, overlooking the neural origins of decoded information. Here, we introduce Perceptogram, a unified, interpretable framework that uses paired linear mappings between EEG signals and CLIP latents, leveraging CLIP's inherent structure. Perceptogram achieves state-of-the-art reconstruction quality and generates latent-filtered EEG maps, isolating neural activity relevant to specific visual attributes. These maps reveal clear spatiotemporal organization: ≈ 100 ms post-stimulus, lateral posterior negativity encodes smooth textures and blue hues, while medial negativity captures textured images, red hues, and food semantics; ≈ 180 ms, lateral negativity signals animate objects. By identifying these distinct neural signatures, Perceptogram transparently delineates how visual features—from basic textures and colors to high-level object categories—are temporally and spatially represented in the brain [1].

Keywords: Representational Alignment; BCI; Visual Decoding

Approach

Our approach goes beyond reconstructing images from EEG—it identifies and explains which neural signals drive decoding. We do this in two ways: **1. Test-time Perturbation:** We systematically modify EEG signals (e.g., swapping electrodes or time segments) to see how reconstructions change. **2. Decoding-Encoding Loop:** EEG signals are linearly mapped into CLIP's latent space and then back to EEG. This process filters EEG signals, isolating electrodes and time points carrying visual meaning. The resulting patterns ("latent-filtered EEG") are directly interpretable and similar to common spatial patterns used in BCI Blankertz et al. (2008). We further extend this decoding–encoding approach to low-level features such as color and texture by leveraging other latent spaces, revealing their distinct neural signatures.

Datasets: We used the publicly available THINGS-EEG2 (Gifford et al., 2022) and Natural Scenes Dataset (NSD) (Allen et al., 2022) for EEG and fMRI analyses, respectively, to validate findings from our EEG analysis.

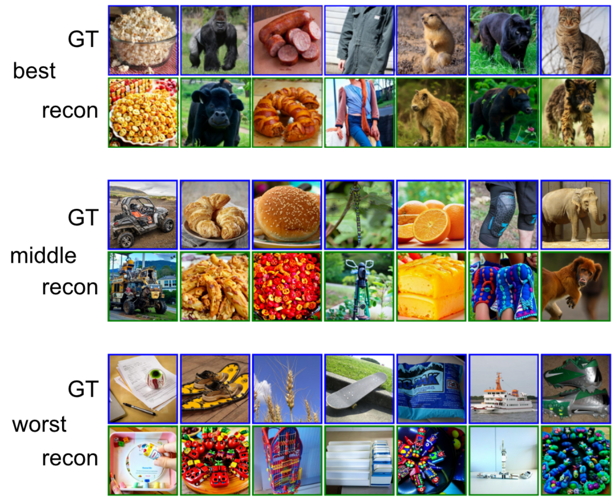


Figure 1: Reconstructed (recon) examples of different qualities. GT is ground truth.

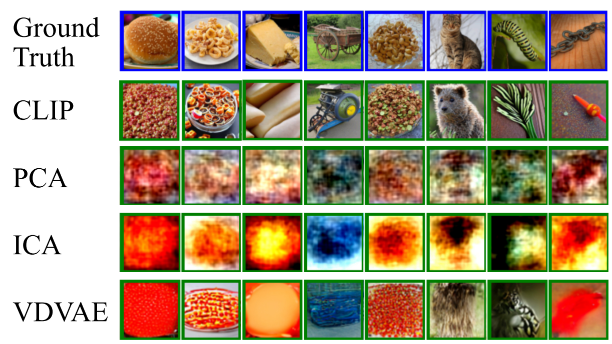


Figure 2: Reconstructions from different latent spaces

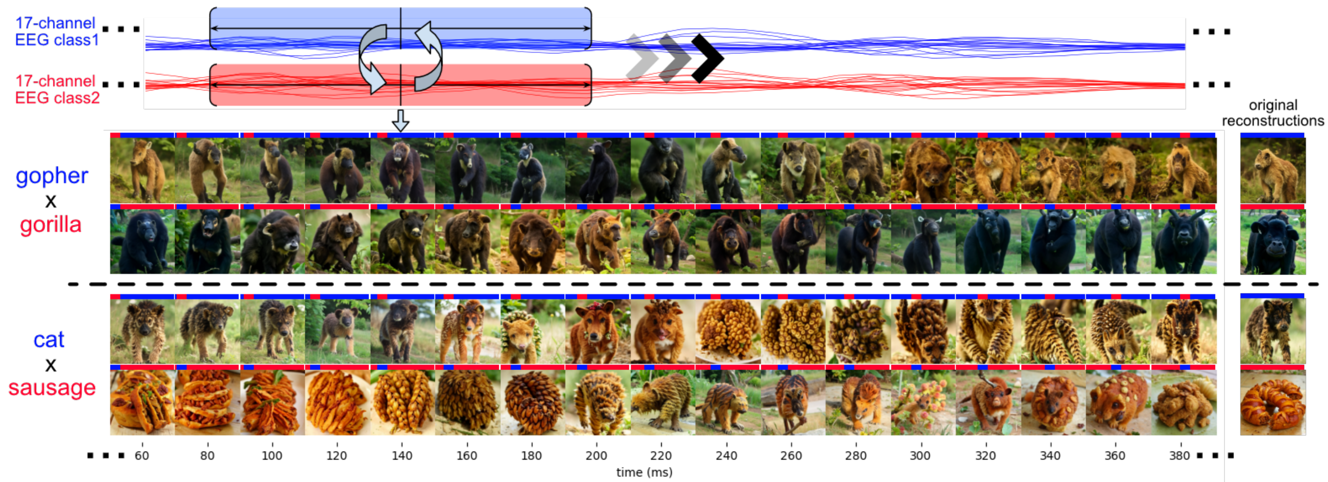
Results & Significance

Table 1 demonstrates state-of-the-art reconstruction performance, despite learning only a simple linear map. Reconstructions of varying quality are shown in Fig 1, and reconstructions from different latent spaces captures various visual features, as shown in Fig 2. Temporal perturbation experiments shown in Fig 3a illustrate later representation of semantic information relative to lower-level visual features. Electrode mirroring results in Fig 3b reveal interesting semantic changes. EEG's temporal resolution enables generation of novel spatiotemporal maps (Fig 3c) representing various visual features that spatially agree with those obtained from fMRI (Fig 3d). Overall our results challenge the idea that EEG-based visual decoding requires deep, nonlinear and non-interpretable models.

[1]This is a brief version of the Arxiv paper Fei et al. (2024)

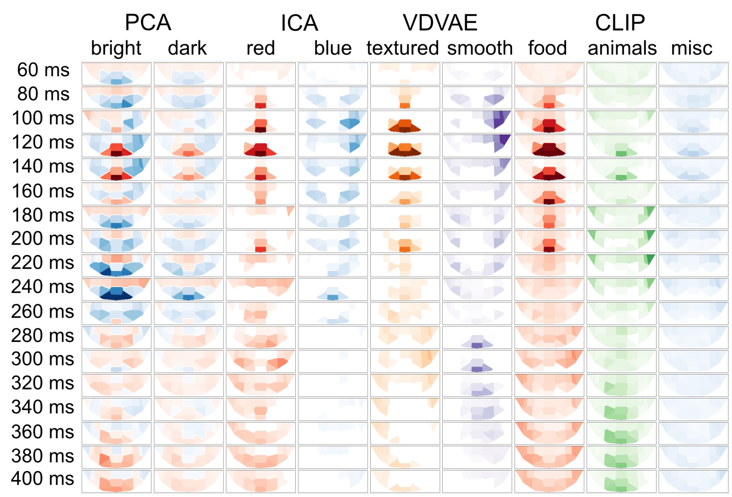
Table 1: Quantitative assessments of the reconstruction quality for EEG and MEG.

Dataset	PixCorr ↑	SSIM ↑	AlexNet(2) ↑	AlexNet(5) ↑	Inception ↑	CLIP ↑	EffNet ↓	SwAV ↓
THINGS-MEG (BrainDecoding) Benchetrit et al. (2024)	0.088	0.333	0.747	0.855	0.712	0.804	-	0.576
THINGS-MEG (Perceptogram with unCLIP)	0.187 ± .004	**0.376** ± 0.007	**0.848** ± 0.036	**0.906** ± 0.031	**0.748** ± 0.032	**0.826** ± 0.027	0.875 ± 0.021	**0.527** ± 0.021
THINGS-EEG2 (ATM-S) Li et al. (2024)	-	0.345	0.776	0.866	0.734	0.786	-	0.582
THINGS-EEG2 (Perceptogram with Versatile Diffusion)	0.267 ± .015	0.347 ± 0.003	**0.910** ± 0.010	**0.927** ± 0.005	**0.752** ± 0.008	**0.807** ± 0.009	0.877 ± 0.004	0.540 ± 0.004
THINGS-EEG2 (Perceptogram with unCLIP)	0.223 ± .029	**0.37** ± 0.005	0.875 ± 0.013	0.915 ± 0.008	0.749 ± 0.024	0.806 ± 0.016	0.87 ± 0.011	**0.530** ± 0.009

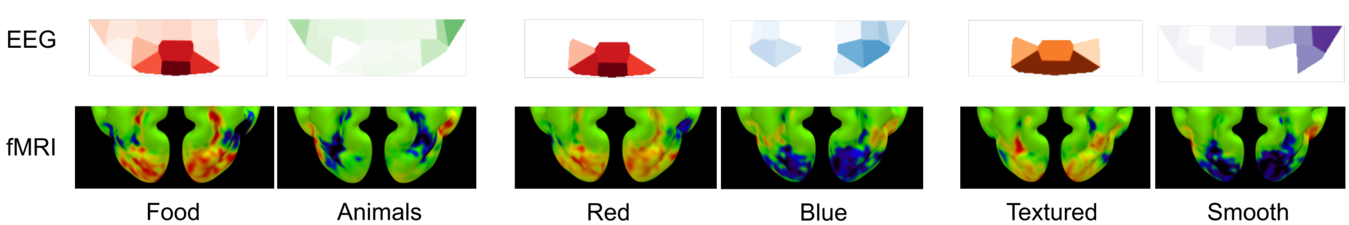


(a) Time-Swapping. Each reconstructed image has a horizontal color bar showing when and how much EEG data was swapped between the two image classes. The gopher reconstruction appears darker when EEG segments from 40–320 ms are swapped with gorilla EEG, and the gorilla appears lighter when the same range is replaced with gopher EEG. In the cat-sausage swap, the cat takes on a food-like appearance when EEG from 240–280 ms is replaced with sausage EEG, and the sausage looks more animal-like when EEG from 200–360 ms is replaced with cat EEG.

(b) Electrode-Mirroring. **Mirror All**: Electrode locations mirrored about the midline during test time **Mirror O1 & O2**: O1 and O2 are swapped during test time. In the "Mirror-all" condition, we found that many images reversed their animacy. For example, `cheetah`, `seagull`, `panther`, and `robot` all produced non-living objects. Conversely, `balance beam` and `sandpaper` produced mirrored reconstructions that look like living creatures.



(c) Average latent-filtered EEG patterns (10 subjects). For PCA, red indicates stronger positive polarity; blue indicates stronger negative polarity. For ICA, VDVAE, and CLIP, stronger color indicates stronger negative polarity.



(d) Cross-subject EEG–fMRI alignment. The increased fMRI signal (red) corresponds to increased EEG signal represented by the darker category-specific color.

Acknowledgments

This work was supported in part by NSF IIS 1817226, CRCNS 2208362, and a seed grant from UC San Diego Social Sciences. The authors are also grateful for the hardware support provided by NVIDIA, Adobe, and Sony.

References

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

Benchetrit, Y., Banville, H., & King, J.-R. (2024). Brain decoding: Toward real-time reconstruction of visual perception. In *The twelfth international conference on learning representations (iclr).*

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., & Müller, K.-R. (2008). Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, *25*(1), 41-56.

Fei, T., Uppal, A., Jackson, I., Ravishankar, S., Wang, D., & de Sa, V. R. (2024). *Perceptogram: Reconstructing visual percepts from eeg.* arXiv. Retrieved from `https://arxiv.org/abs/2404.01250` doi: 10.48550/ARXIV.2404.01250

Gifford, A. T., Dwivedi, K., Roig, G., & Cichy, R. M. (2022, December). A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, *264*, 119754. doi: 10.1016/j.neuroimage.2022.119754

Li, D., Wei, C., Li, S., Zou, J., & Liu, Q. (2024, March). Visual decoding and reconstruction via EEG embeddings with guided diffusion. *arXiv*(arXiv:2403.07721). Retrieved from `http://arxiv.org/abs/2403.07721` (arXiv:2403.07721 [cs, eess, q-bio])