 DOI: 10.3217/978-3-99161-050-2-158

Perceptogram: Reconstructing Visual Percepts from EEG

T. Fei[1*], I. Jackson[1], A. Uppal[1], V.R. de Sa[1]
[1] *University of California San Diego, CA, USA;*
*E-mail: tfei@ucsd.edu

Introduction: Visual perception is an important aspect of human cognition and a gateway to understanding more complex cognitive processes such as visual mental imagery and dream visualization. EEG has spatial resolution limited by volume conduction which has resulted in its underuse in decoding visual representation. In this work, we reconstruct viewed images from EEG recordings with state-of-the-art quantitative reconstruction performance using a linear decoder that maps the EEG to image latents.

Material, Methods and Results: We used the preprocessed version of THINGS-EEG2 dataset [1]. In the experiment, each image is presented for 100ms followed by a blank screen for 100ms before the next image. The image presentation order is pseudo-randomized across the entire image set. All 10 subjects view the same 16740 images, of which the same 200 images are test images. Each training image is shown 4 times, and each test image is shown 80 times. The image reconstruction is a 2-stage process: the first stage maps the brain signal onto the latent space of a variational auto-encoder (VAE), which provides a rough visual representation. The second stage maps the same brain signal onto each token of the CLIP-Vision. The image generator "unCLIP" [2] combines the CLIP and the encoded images from the VAE and produces the reconstructed images. The reconstructions often capture the semantic meanings as well as the visual features of the ground-truth images (see Fig. 1). The performance across subjects is relatively consistent with a small amount of variation.

Conclusion: We have demonstrated the surprising power of learned linear mappings to different latent spaces. When mapped to the CLIP and VAE latent spaces and reconstructed with unCLIP, realistic reconstructions are obtained that quantitatively and qualitatively outperform previous EEG reconstruction attempts, including those with much more sophisticated transformer-based decoding algorithms.

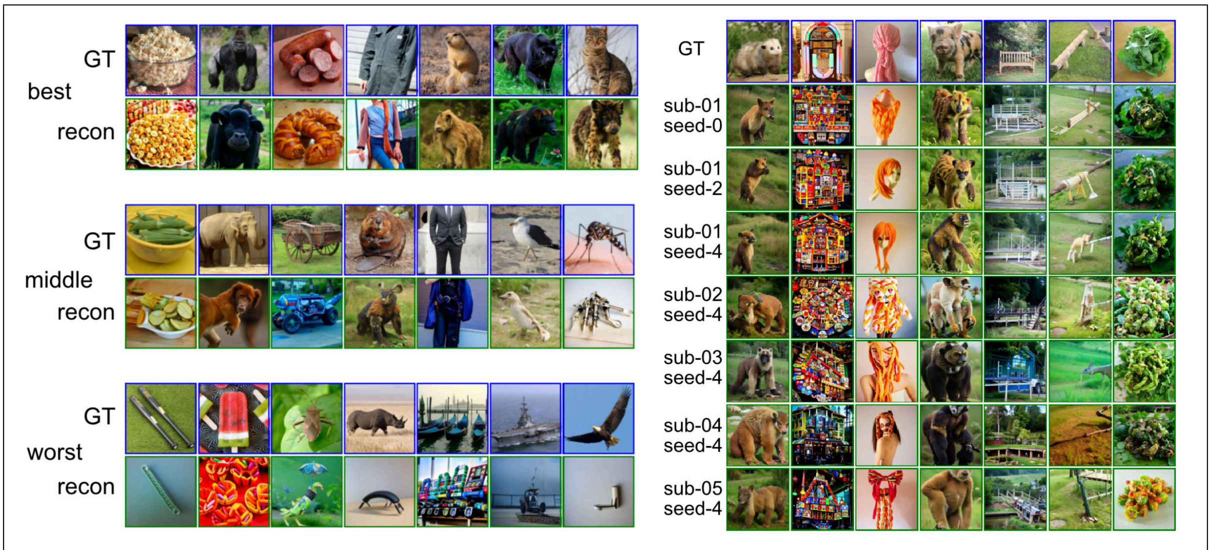

Figure 1: *Left: example reconstructions of viewed images using EEG recorded from the viewing participant. Examples of the best, middle, and worst reconstructions were selected by visual inspection from Subject 1. The rows labeled GT (ground truth) show the image that was shown to the participant. The rows labeled reconstructed (recon) show the image that was created by our system Examples are sampled from among the best (top) middle (middle) and worst (bottom) reconstructions observed. Right: Robust reconstructions across subjects and diffuser random seeds. Top Row: the ground-truth stimulus images; subsequent 3 rows: different random seeds for the same subject; Last 4 rows: robust reconstructions across subjects.* *This work was supported in part by UCSD Social Sciences grant, NSF IIS 1817226 and IIS 2208362.*

References:

[1] Gifford, Alessandro T. and Dwivedi, Kshitij and Roig, Gemma and Cichy, Radoslaw M. A large and rich EEG dataset for modeling human visual object recognition. In *NeuroImage*, 119754, 2022.

[2] Ramesh, Aditya and Dhariwal, Prafulla and Nichol, Alex and Chu, Casey and Chen, Mark. Hierarchical Text-Conditional Image Generation with CLIP Latents. In *arXiv*. 2204.06125, 2022.

Published by Verlag der Technischen Universität Graz