Neuro-3D: Towards 3D Visual Decoding from EEG Signals

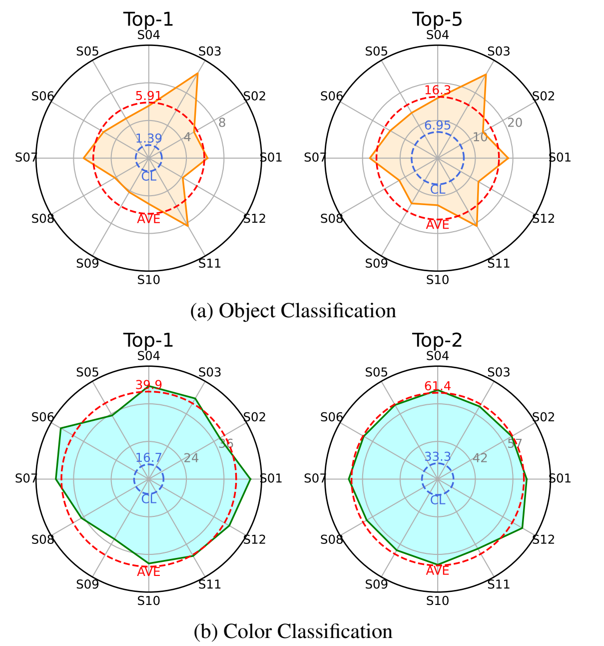
Supplementary Material

1. EEG Data Preprocessing

In this section, we introduce the details of EEG preprocessing pipeline. During data acquisition, static 3D image and dynamic 3D video stimuli were preceded by a marker to streamline subsequent data processing. The continuous EEG recordings were subsequently preprocessed using MNE [5]. The data were segmented into fixed-length epochs (1s for static stimuli and 6s for dynamic stimuli), time-locked to stimulus onset, with baseline correction achieved by subtracting the mean signal amplitude during the pre-stimulus period. The signals were downsampled from 1000 Hz to 250 Hz, and a bandpass filter (0.1–100 Hz) was applied in conjunction with a 50 Hz notch filter to mitigate noise. To normalize signal amplitude variability across channels, multivariate noise normalization was employed [6]. Consistent with established practices [8], two stimulus repetitions were treated as independent samples during training to enhance learning, while testing involved averaging across four repetitions to improve the signal-to-noise ratio, following principles similar to those used in Event-Related Potential (ERP) analysis [11].

2. Evaluation Metrics for Reconstruction Benchmark

To assess the quality of the generated outputs, we adopt the N-way, top-K metric, a standard approach in 2D image decoding [1, 2, 8]. For 2D image evaluation, a pre-trained ImageNet1K classifier is used to classify both the generated images and their corresponding ground truth images. Similarly, we utilize data from Objaverse [3] to pre-train a PointNet++ model [9]. To ensure classifier reliability, the network is trained on all Objaverse data with category labels, excluding the test set used in our study. The point cloud data corresponding to the 3D objects is sourced from [12]. During evaluation, both the generated point clouds and their corresponding ground truth point clouds are classified using the trained network. The results are then analyzed to confirm whether the reconstructed object is correctly identified within the top K categories among N selected. For the efficiency of evaluation, we utilize data from the first five subjects to train and evaluate the reconstruction model. Moreover, a distinct feature of the diffusion model is its dependence on initialization noise, which can influence the generated outputs. We perform five independent inferences for each object and compute the average N-way, top-K metric across these runs. Additionally, to capture the potential best-case performance, we identify the optimal result based on the classifier's predicted scores across the five inferences and compute the N-way, top-K metric.



(a) Object Classification

(b) Color Classification

Figure 1. The results of individual analysis. (a) presents the top-1 and top-5 accuracies for the object classification task across 12 subjects, while (b) depicts the top-1 and top-2 accuracies for the classification task. The blue line in each panel indicates chance-level performance, and the red line represents the average performance across all subjects.

3. Analysis of Individual Difference

We present the performance variability across individuals on two classification tasks, as illustrated in Fig. 1. On both tasks, individual performance consistently exceeds chance level, demonstrating that EEG signals encode visual perception information and that our method effectively extracts and utilizes this information for decoding. Notably, performance varies across tasks for the same individual. For instance, participant $S12$ performs significantly below average in object classification but achieves above-average results in color classification, suggesting distinct neural mechanisms underlying the processing of different visual attributes and their representation in EEG signals.

Furthermore, it has been widely confirmed that EEG signal has substantial individual variations [4, 7, 10]. As shown in Fig. 1, significant differences are observed between individuals performing the same task, particularly in object classification, where $S03$ and $S11$ exhibit superior perfor-

CVPR
#8343

Figure 2. More results reconstructed by Neuro-3D with different samplings trials, and the corresponding ground truth. The sampling variations arise either from results obtained across different subjects or from inference outputs of the diffusion model for the same subject using distinct noise initializations.

mance, while $S08$, $S09$ and $S12$ fall markedly below average. Similar variability is observed in the color classification task, albeit to a lesser extent. These results underscore the pronounced inter-subject differences in EEG signals and highlight a critical challenge for cross-subject EEG visual decoding tasks, where performance remains suboptimal. Addressing this variability is a key direction for future research.

4. More Reconstructed Samples

Additional reconstructed results alongside their corresponding ground truth point clouds are presented in Fig. 2. The proposed Neuro-3D framework exhibits robust performance, effectively capturing semantic categories, shape details, and the overall color of various objects.

5. Analysis of Failure Cases

Figure 3. Failure cases. (a) highlights reconstructions with significant loss of fine details, while (b) demonstrates several instances of incorrect semantic category prediction.

Fig. 3 illustrates representative failure cases, categorized into two principal types: inaccuracies in detailed shape prediction and semantic reconstruction errors. Despite these limitations, certain features of the stimulus objects, including shape contours and color information, are partially preserved in the displayed reconstructed images. These shortcomings primarily arise from the inherent challenges of the low signal-to-noise ratio and limited spatial resolution of EEG signals, which constrain the performance of 3D object reconstruction. Addressing these issues presents a promising direction for future improvement.

References

[1] Zijiao Chen, Jiaxin Qing, Tiange Xiang, Wan Lin Yue, and Juan Helen Zhou. Seeing beyond the brain: Masked modeling conditioned diffusion model for human vision decoding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 1

[2] Zijiao Chen, Jiaxin Qing, and Juan Helen Zhou. Cinematic mindscapes: High-quality video reconstruction from brain activity. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[3] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3D objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023. 1

[4] Erin Gibson, Nancy J Lobaugh, Steve Joordens, and Anthony R McIntosh. Eeg variability: Task-driven or subject-driven signal of interest? *NeuroImage*, 252:119034, 2022. 1

[5] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroinformatics*, 7:267, 2013. 1

[6] Matthias Guggenmos, Philipp Sterzer, and Radoslaw Martin Cichy. Multivariate pattern analysis for meg: A comparison of dissimilarity measures. *NeuroImage*, 173:434–447, 2018. 1

[7] Gan Huang, Zhiheng Zhao, Shaorong Zhang, Zhenxing Hu, Jiaming Fan, Meisong Fu, Jiale Chen, Yaqiong Xiao, Jun Wang, and Guo Dan. Discrepancy between inter-and intra-subject variability in eeg-based motor imagery brain-computer interface: Evidence from multiple perspectives. *Frontiers in Neuroscience*, 17:1122661, 2023. 1

[8] Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, and Quanying Liu. Visual decoding and reconstruction via EEG embeddings with guided diffusion. In *Advances in Neural Information Processing Systems*, 2024. 1

[9] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 30, 2017. 1

[10] Simanto Saha and Mathias Baumert. Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: a review. *Frontiers in Computational Neuroscience*, 13:87, 2020. 1

[11] Shravani Sur and Vinod Kumar Sinha. Event-related potential: An overview. *Industrial Psychiatry Journal*, 18(1):70–73, 2009. 1

[12] Runsen Xu, Xiaolong Wang, Tai Wang, Yilun Chen, Jiangmiao Pang, and Dahua Lin. Pointllm: Empowering large language models to understand point clouds. *European Conference on Computer Vision*, 2024. 1