Mapping EEG Signals to Image Representations Using Deep Learning

Sara Alzarooni , Lara Monagel, Rema Alblooshi, Mawaddah Ali, Ruam Saleh Hadi, Meshall Alshalaan
Supervisor: Elias Al Ghazal

Abstract:

Electroencephalography (EEG) is a cost-effective and noninvasive method that tracks cerebral activity with high resolution. Even though EEG has limited spatial clarity, which limits the ability to interpret it visually, it is still considered more convenient for real-time users than other multifunctional approaches. Researchers developed classification-based methods to recognize object categories and reconstruction-based methods, which are enhanced with deep learning to generate an approximate visualization of what the person is seeing. The classification methods achieved accurate results on a variety of categories, but were not as accurate when capturing fine-grained distinctions. On the other hand, reconstruction methods are still inaccurate and coarse. For this reason, we use the EEG-ImageNet dataset, which is the largest and most comprehensive benchmark for EEG-based visual decoding. It consists of 63,850 EEG–image pairs from 16 participants, spanning 80 categories with both coarse- and fine-grained labels. This dataset enables robust evaluation of classification and reconstruction approaches. The objective of this study is to develop a deep learning-based method to predict and reconstruct the images viewed by subjects with higher accuracy.

Keywords: Electroencephalography (EEG), deep learning, image reconstruction, brain signal processing.

1. Introduction:

Decoding human visual experiences directly from brain activity is a classic issue in cognitive neuroscience and brain–computer interface research. Electroencephalography (EEG) is a non-invasive, inexpensive, and temporally precise way to measure neural activity, making it extremely attractive for practical uses. However, image reconstruction from EEG remains largely difficult due to the low spatial resolution, high noise, and intricacy of visual processing intrinsic to the modality. Success in this domain has important implications for neuroscience, clinical communication devices, and future human AI interface systems. Recent advances have demonstrated the feasibility of EEG-based visual decoding. Convolutional and recurrent networks have enabled object categorization from EEG signals, and more recent approaches entail coupling pre-trained vision–language models such as CLIP with robust diffusion backbones to generate class-consistent images [1][2].

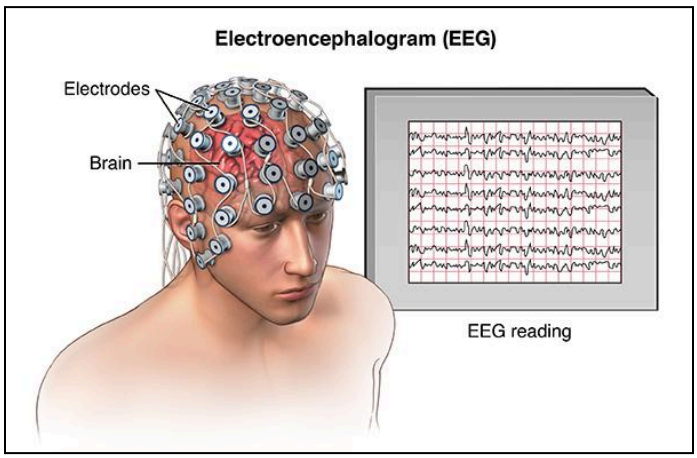


Fig 1: Illustration of EEG Reading [9]

Additional work has shown that spatio-temporal graph neural networks (GNNs) are better suited to model the relational structure of EEG channels [4][6], and that dual-stage pipelines well established in fMRI decoding can disentangle low-level perceptual and high-level semantic information [3].

Despite these advances, most EEG-based methods squish the entire signal into a single holistic vector. This combines the "what" (semantic identity) and the "how" (perceptual features), causing a loss of fine-grained detail and limiting the interpretability and fidelity of reconstructions [5].

This mismatch motivates the development of architectures that formally separate perceptual and semantic streams in EEG decoding. We present the Neuro-Perceptual Diffusion Model (NPDM), a novel framework that connects neurobiological discoveries with recent advances in generative modeling. NPDM introduces three novelties: (1) a Spatio-Temporal Graph Encoder to capture

fine-grained neural dynamics; (2) a Dual-Stream Hierarchical Conditioning module that disentangles "what" and "how" and injects each into different phases of the generative process; and (3) a pre-trained diffusion backbone to integrate these complementary signals to reconstruct images that are semantically faithful and perceptually detailed.

2. Data:

2.1. Dataset:

The EEG-ImageNet dataset [7] consists of EEG recordings from 16 individuals, each exposed to 4,000 images sourced from the ImageNet-21k database, yielding a total of 63,850 EEG-image pairs. The photographs are organized into 80 categories, each containing 50 photos, and are annotated with multi-granularity labels comprising 40 coarse-grained and 40 fine-grained classes, for detailed analysis at various semantic levels [7]. Electroencephalogram (EEG) signals were acquired via standard 62-channel EEG caps sampled at a frequency of 1000 Hz. The dataset, accessible via GitHub, provides recommended preprocessing steps and predefined train, test, and evaluation splits as described in the authors' paper. Compared to previous benchmarks, EEG-ImageNet offers five times more EEG–image pairs and establishes strong baselines for both object classification and image reconstruction tasks, making it a significant resource for advancing computational neuroscience and brain–computer interface research.

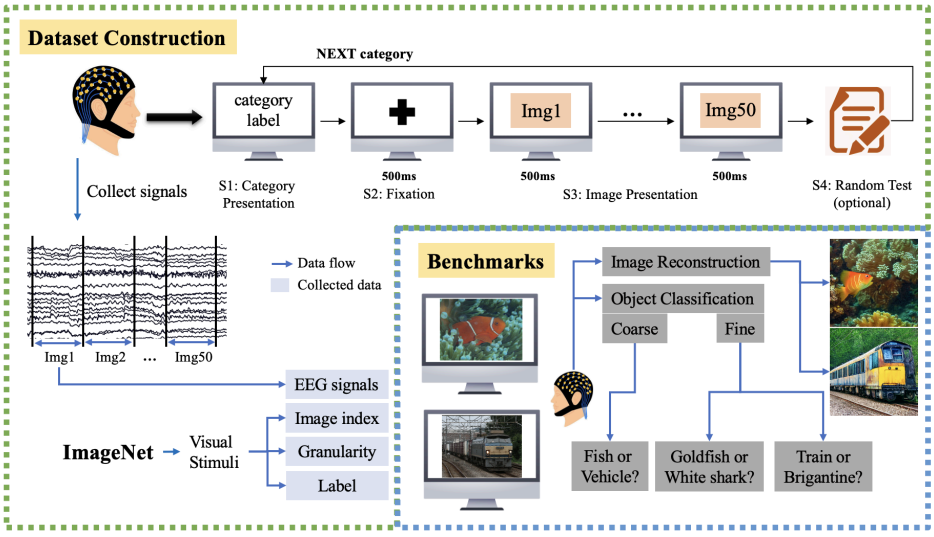


Fig 2: Dataset Construction [8]

2.2. Data Preprocessing:

The EEG-ImageNet dataset used in this study [7] was provided in serialized .pth format and contains synchronized pairs of multi-channel EEG recordings and visual stimuli drawn from ImageNet categories. Each trial includes an EEG signal matrix, a WordNet identifier (WNID)

specifying the object class, and auxiliary metadata such as subject information, stimulus filename, and trial index. To manage heterogeneous data blobs across subjects and sessions, we implemented a PyTorch-based lazy-loading class (EEGImageNetLazyDataset). This class dynamically traverses serialized files, extracts EEG signals, parses WNIDs, and assembles each trial into the standardized structure (X, wnid, image_id, meta), where $X \in R^{(C \times T)}$ denotes the EEG trial with C channels and T timepoints. Dataset reliability was improved using a filtering class (FilteredEEGDataset), which excluded WNIDs with insufficient examples or inconsistent labeling across subjects. A coverage audit was conducted to verify alignment between EEG trials and corresponding ImageNet images, reporting statistics on exact filename matches, token-based matches, class-only fallbacks, and missing categories. Following these steps, the final dataset consisted of thousands of high-quality EEG–image pairs spanning diverse object categories, each suitable for both decoding and generative modeling experiments.

Preprocessing ensured that EEG signals were standardized and reproducible across experiments. Each trial was first validated to maintain the canonical orientation ($C \times T$). In cases of erroneously transposed arrays, the matrix was automatically corrected. Channel-wise normalization was then applied using z-scoring, thereby standardizing amplitude distributions and reducing inter-channel variability. To mitigate noise, a predefined set of problematic WNIDs was excluded. For quality control, visualization routines plotted normalized EEG traces from representative subsets of channels alongside their paired ImageNet stimuli, enabling inspection of temporal dynamics relative to object categories. An additional audit procedure quantified dataset integrity by comparing EEG metadata with available ImageNet files, reporting alignment fidelity through exact, partial, and class-level matches. Importantly, preprocessing was implemented directly within the dataset loader. As a result, normalization and integrity checks occurred dynamically at data access time, reducing memory overhead and ensuring reproducibility across different experimental runs.

3. Methodology

3.1 Decoding Pipeline

The decoding pipeline was designed to map EEG trials into latent features that disentangle semantic identity from perceptual attributes. Formally, each trial is represented as $X \in R^{(C \times T)}$. The Spatio-Temporal Graph Encoder (STGE) transforms X into two complementary outputs:

- A semantic embedding $E \in R^{768}$, analogous to a CLIP text embedding.
- A set of perceptual tokens $T \in R^{(12 \times 320)}$, capturing style, texture, and pose.

The STGE architecture consists of: (i) per-channel LayerNorm across time, (ii) a temporal convolutional stem to capture local patterns, (iii) a spatial attention module treating EEG

channels as tokens, (iv) a temporal Transformer encoder modeling long-range dependencies, and (v) a learned pooling mechanism producing both global and token-level outputs. These outputs are processed by the Dual-Stream Conditioner. The semantic stream maps E into a surrogate embedding aligned with the 768-dimensional CLIP space, enabling injection into Stable Diffusion cross-attention layers. In parallel, the perceptual stream transforms T into FiLM parameters for GroupNorm layers and additive key/value biases for cross-attention modules, thereby modulating both categorical semantics and perceptual fidelity. This dual-stream design leverages EEG's temporal resolution to drive object-level decoding while preserving fine-grained perceptual features.

3.2 Reconstruction Pipeline

The reconstruction pipeline integrates EEG-derived control signals into a frozen Stable Diffusion backbone (v1-5). A custom hook manager (NPDMHookManager) injects FiLM parameters into selected GroupNorm layers and key/value biases into cross-attention modules, enabling EEG-derived signals to influence both low-level perceptual statistics and high-level semantic attention distributions. During training, paired EEG–image samples are processed as follows:

- Images are encoded into latent space using the frozen VAE encoder.
- Noise is added at a random timestep according to the diffusion scheduler.
- The UNet denoiser predicts the noise residual conditioned on EEG-derived controls.

The training objective is the mean-squared error between predicted and ground-truth noise. Optimization is restricted to the parameters of the Dual-Stream Conditioner, while the diffusion backbone remains frozen. At inference, unseen EEG trials are passed through the STGE and conditioner to produce semantic and perceptual control signals. These are injected into the diffusion model during iterative denoising, and final images are generated using the frozen VAE decoder. The present study emphasizes the design and implementation of this reconstruction pipeline, with preliminary training runs demonstrating feasibility. Large-scale fine-tuning remains an avenue for future work.

4. Results and Discussion

Although model training and evaluation are currently ongoing, we anticipate that the proposed Neural Perceptual Disentanglement Model (NPDM) will be competitive and possibly outperform existing EEG-to-image reconstruction approaches by effectively separating semantic and perceptual information.

1. **Semantic Accuracy:**
   By disentangling the "what" from the "how," NPDM is expected to improve the accurate identification of high-level semantic categories (e.g., correctly recognizing "dog" vs. "cat") compared to holistic feature-based models.

2. **Perceptual Fidelity:**
   - The model's dual-stream hierarchical conditioning is designed to guide the pre-trained diffusion backbone with fine-grained perceptual details, which should lead to reconstructions that preserve low-level attributes such as texture, orientation, and color.

3. **Quantitative Metrics:**
   - We anticipate higher scores in standard evaluation metrics for image reconstruction, including Structural Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and classification accuracy on reconstructed images using a pre-trained image classifier.

4. **Qualitative Observations:**
   - Generated images are expected to exhibit improved realism and better alignment with EEG stimuli, particularly in capturing subtle variations that prior holistic models tend to blur.

The EEG-ImageNet dataset will serve as the primary benchmark, allowing evaluation across both coarse and fine-grained categories. We expect the NPDM to demonstrate consistent gains across diverse categories, reflecting its ability to capture both semantic identity and perceptual attributes simultaneous

Figure 3 illustrates reconstruction results reported by Zhu et al. [8], which serve as a reference for our anticipated outcomes.



Fig 3: The image reconstruction results from [8]

5. Conclusion

This study underscores the importance of developing accurate EEG-to-image reconstruction models. Existing approaches often struggle to preserve fine-grained perceptual details while maintaining high-level semantic accuracy. By disentangling semantic identity from perceptual attributes, advanced models such as the Neural Perceptual Disentanglement Model (NPDM) have the potential to more effectively bridge this gap. Improved models promise not only higher-quality reconstructions but also clearer visualizations of neural responses and more reliable data for downstream applications.

Ultimately, progress in EEG-based visual decoding will enhance our ability to interpret complex brain signals and open new opportunities in brain computer interfaces, cognitive neuroscience, and neurotechnology, demonstrating the transformative potential of integrating EEG with state-of-the-art deep learning methods.

References

[1] Ferrante, M., Boccato, T., Bargione, S., & Toschi, N. (2024). Decoding visual brain representations from EEG through knowledge distillation and latent diffusion models. *Computers in Biology and Medicine, 178*, 108701.

[2] Guenther, S., Kosmyna, N., & Maes, P. (2024). Image classification and reconstruction from low-density EEG. *Scientific Reports*.

[3] Özçelik, F., & VanRullen, R. (2023). Natural scene reconstruction from fMRI using generative latent diffusion (Brain-Diffuser). *Scientific Reports*.

[4] Shan, X., et al. (2022). Spatial–temporal GCN for EEG classification. *Human Brain Mapping*.

[5] Wilson, H., et al. (2024). Feasibility of decoding visual information from EEG (review). *Network Neuroscience*.

[6] Zeng, Y., et al. (2024). Adaptive node feature extraction in graph-based EEG with task-induced self-supervision. *NeuroImage*.

[7] Tsinghua University. (n.d.). *EEG-ImageNet-Dataset*. Tsinghua University. https://cloud.tsinghua.edu.cn/d/d812f7d1fc474b14bbd0/

[8] Zhu, S., Ye, Z., Ai, Q., & Liu, Y. (2024). *EEG-ImageNet: An Electroencephalogram Dataset and Benchmarks with Image Visual Stimuli of Multi-Granularity Labels*. *arXiv*. https://arxiv.org/abs/2406.07151

[9] Kl, R. (2019, June 6). EEG data analysis. Kaggle. https://www.kaggle.com/code/ruslankl/eeg-data-analysis