Interpretable EEG-to-Image Generation with Semantic Prompts

Arshak Rezvani, Ali Akbari, Kosar Sanjar Arani, Maryam Mirian, Emad Arasteh, Martin J. McKeown

Paper

I. Introduction & Core Problem

- **Objective:**

To reconstruct visual experiences from EEG signals in order to advance both machine learning and cognitive neuroscience.

- **Challenge:**

EEG signals suffer from a low signal-to-noise ratio and limited spatial resolution, which restricts the generation of coherent, high-quality images.

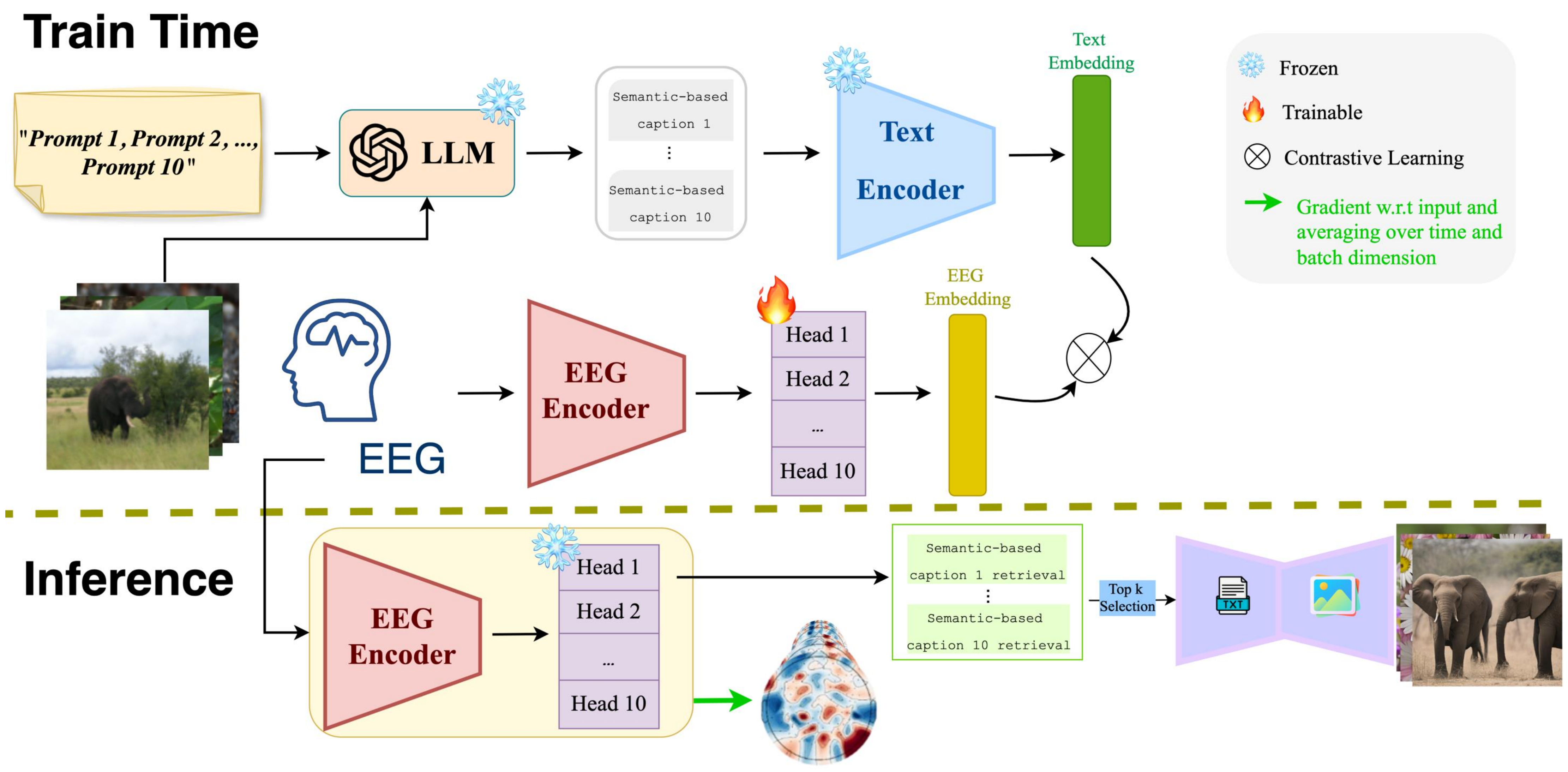
As a result, outputs are often ambiguous, biased, or visually incoherent.

- **Our Approach:**

We propose a text-mediated framework that bridges EEG signals with semantic captions to guide image synthesis.

This strategy improves not only image quality, but also the interpretability of the decoding process.

II. Methods



Phase 1: Training

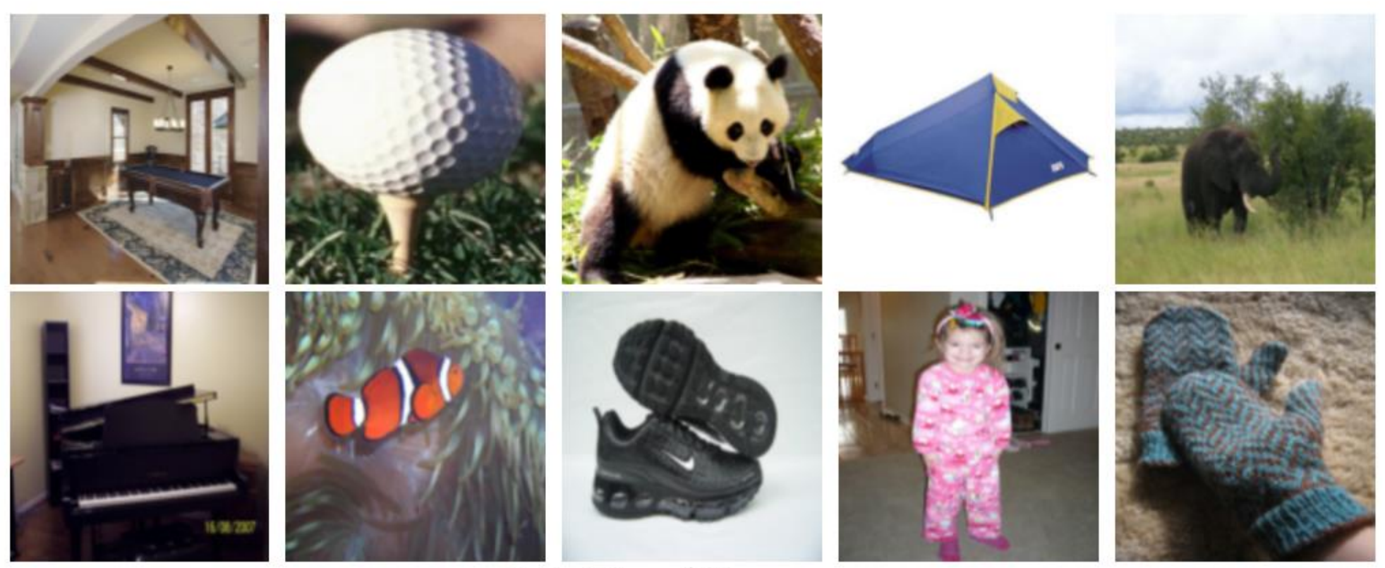
- **Semantic Vocabulary:** Large language model generates multilevel captions (object, spatial, thematic) for each image.
- **EEG-Semantic Alignment:** Transformer encoder aligns EEG signals with captions using contrastive learning [2].

Phase 2: Inference

- **Semantic Retrieval:** EEG input is mapped to the most relevant captions via the trained encoder.
- **Image Generation:** Retrieved captions condition a pretrained latent diffusion model [1] to generate high-quality images.

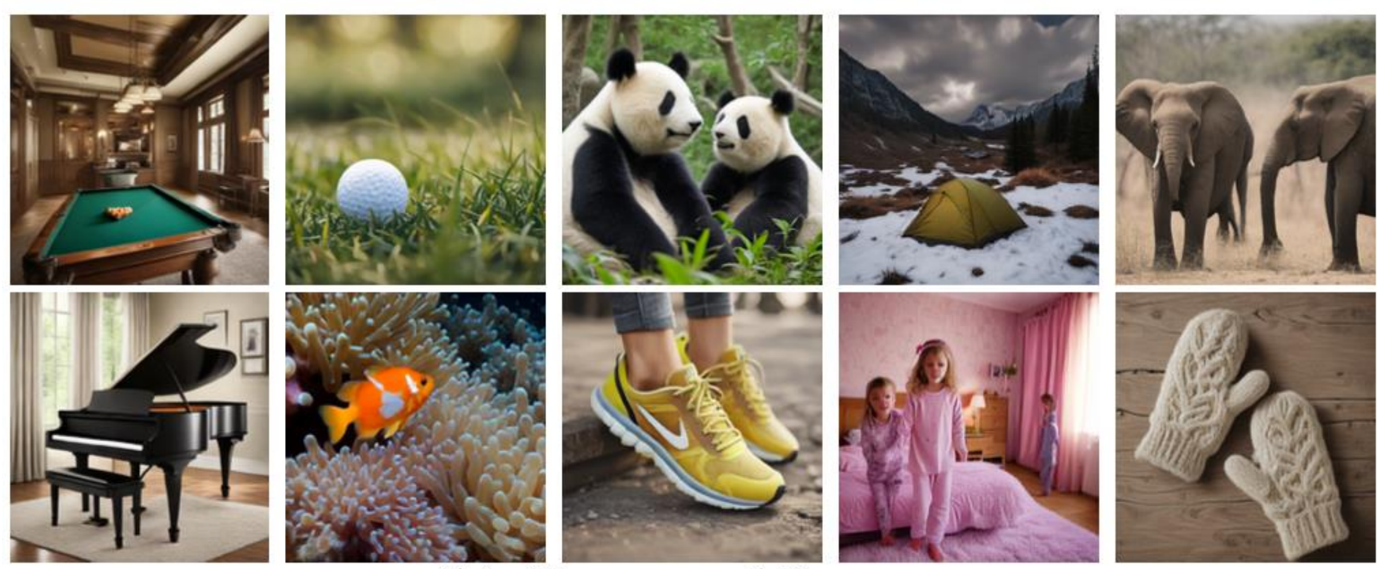
III. Results

Our framework sets a new state-of-the-art EEG-to-image generation schema on the public EEGCVPR dataset [3].



(a) Real Images

Dataset	Model	Type	IS↑	KID↓	PixCorr↑	SSIM↑	Alex2↑	Alex5↑	Inception↑	CS↑	SwAV↓
EEGCVPR (Spampinato et al., 2017)	EEGStyleGAN-ADA (Singh et al., 2024)	GAN	10.82	0.56	-	-	-	-	-	-	-
	EEG-ViT (Akbari et al., 2024)	GAN	12.17	0.05	-	-	-	-	-	-	-
	NeuroVision (Khare et al., 2022)	GAN	5.15	-	-	-	-	-	-	-	-
	Improved-SNGAN (Zheng et al., 2020)	GAN	5.53	-	-	-	-	-	-	-	-
	Brain2Image-VAE (Kavasidis et al., 2017)	VAE	4.49	-	-	-	-	-	-	-	-
	Ours	Diffusion	**37.29 ± 0.32**	**0.009 ± 0.009**	**0.06**	**0.30**	**0.65**	**0.80**	**0.88**	**0.88**	**0.57**

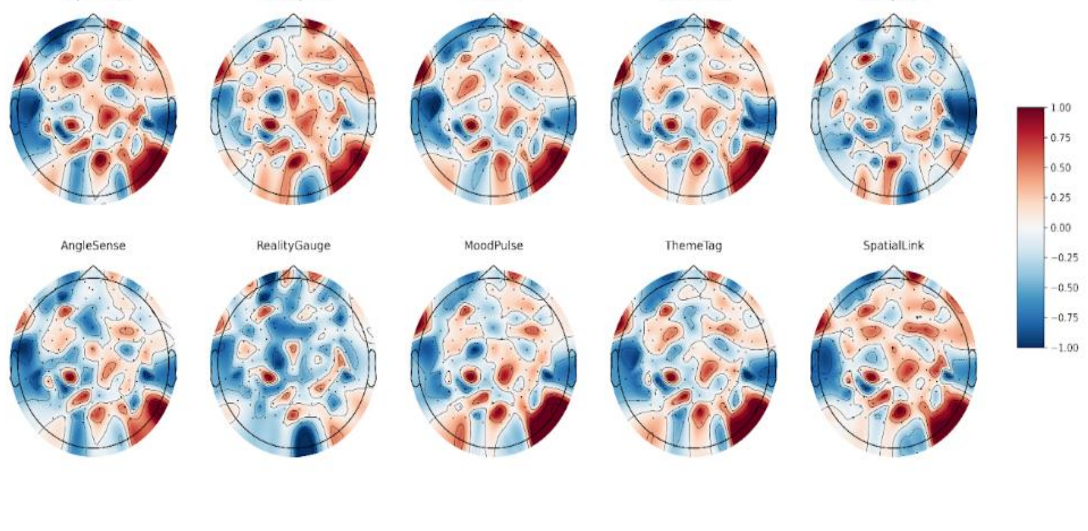


(b) Generated Images

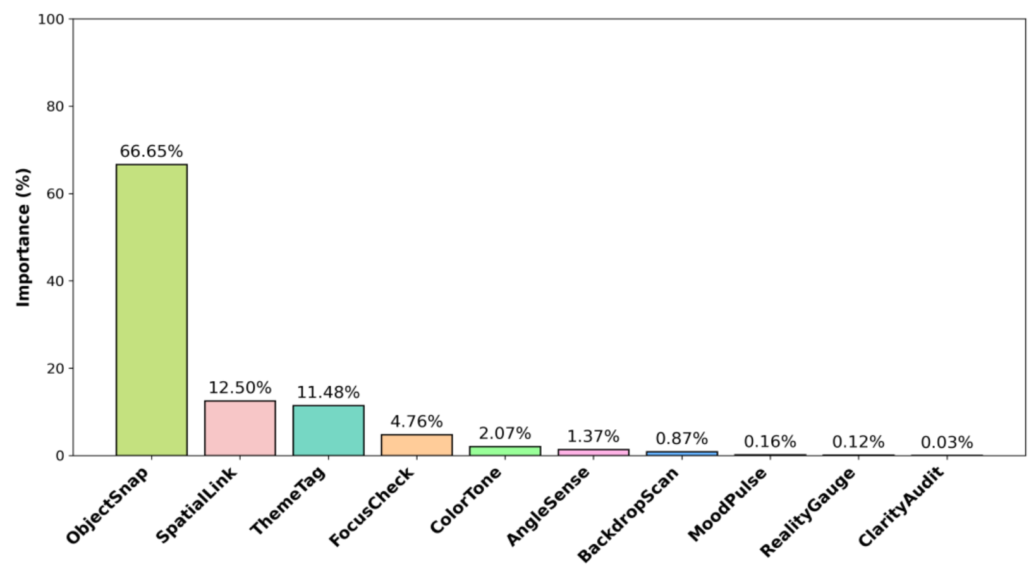
Evidence: Generated images exhibit strong visual fidelity and semantic alignment with ground truth, validated through qualitative and quantitative benchmarks.

Performance: Achieves state-of-the-art results, surpassing prior methods [4] in Inception Score (IS), Kernel Inception Distance (KID), and CLIP Score.

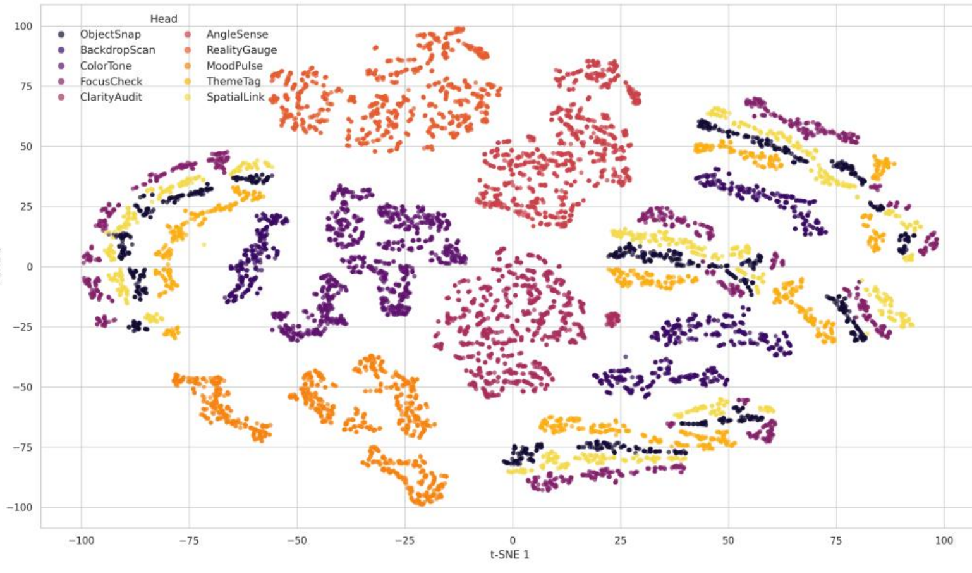
IV. Interpretability



Neural Mapping:
Saliency maps reveal low-level features (e.g., color) in occipital regions and high-level semantics (e.g., theme) in frontal areas, aligning with neurocognitive principles.



Semantic Specialization:
Encoder heads (ObjectSnap, SpatialLink, ThemeTag) specialize in distinct semantic roles, accounting for ~90% of EEG-caption alignment.



Encoder Head Specialization:

- **ObjectSnap**: Captures object-level details (e.g., items, colors); linked to occipital regions.
- **SpatialLink:** Focuses on spatial layouts (e.g., object arrangements, scene structure); tied to parietal regions.
- **ThemeTag:** Encodes themes and emotions (e.g., mood, abstract concepts); engages frontal regions.

Insight: Provides a transparent view into how EEG signals encode visual semantics.

V. Conclusion & References

Summary:
We propose a novel EEG-to-image framework leveraging multilevel semantic prompts to achieve interpretable, high-fidelity visual reconstruction. Our model sets a new benchmark on EEGCVPR and offers insights into the brain's semantic organization.

Contributions:

- Multilevel semantic prompts for EEG-to-image synthesis.
- State-of-the-art performance with interpretable neural mappings.
- Scalable framework integrating EEG with pretrained diffusion models.

- **References**

[1] Rombach, R., et al. (2022). High-resolution image synthesis with latent diffusion models. CVPR.
[2] Radford, A., et al. (2021). Learning transferable visual models from natural language supervision. ICML.
[3] Singh, P., et al. (2024). Learning robust deep visual representations from EEG brain recordings. WACV.
[4] Akbari, A., et al. (2024Joint Learning for Visual Reconstruction from the Brain Activity: Hierarchical Representation of Image Perception with EEG-Vision Transformer. NeurIPS Workshop.