Visual Decoding and Reconstruction via EEG Embeddings with Guided Diffusion

Dongyang Li*, Chen Wei*, Shiying Li, Jiachen Zou, Quanying Liu‡

Presented by Dongyang Li (李东洋)

Advisor: Prof. Quanying Liu（刘泉影）

2024.12.01

The core of BCI is neural decoding

NEURAL INFORMATION PROCESSING SYSTEMS

External World

A model to encode (simulate) neural signals

10^{12} neurons

A model to decode neural signals

Senses
Motion
Emotion
Attention
Cognition
…

Neural encoding

- Given an external stimulus, predict how the nerve will react

Neural decoding

- Given a neural signal and interpret the intention contained in it

There are a variety of implementations of neural decoding depending on the task paradigm, the brain area involved, and the computational model.

The key to neural decoding lies in **neural representations**

NEURAL INFORMATION PROCESSING SYSTEMS

The existing experimental paradigm of brain computer interface has many drawbacks: scalability, interpretability, performance

SSVEP

- 7Hz/10Hz

Motor imagery

- Up/down/left/right
- Body part: Hands/feet

Speech imagination

- Apple/watermelon
- Sentence： ......

Emotional arousal

- Happy/sad

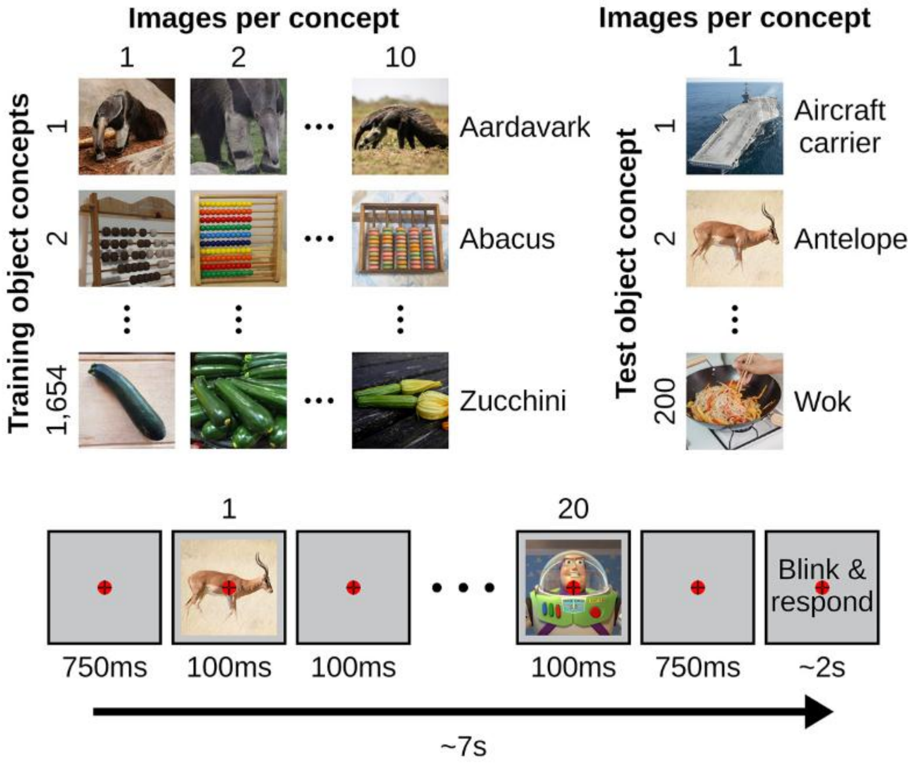
Attention recognition

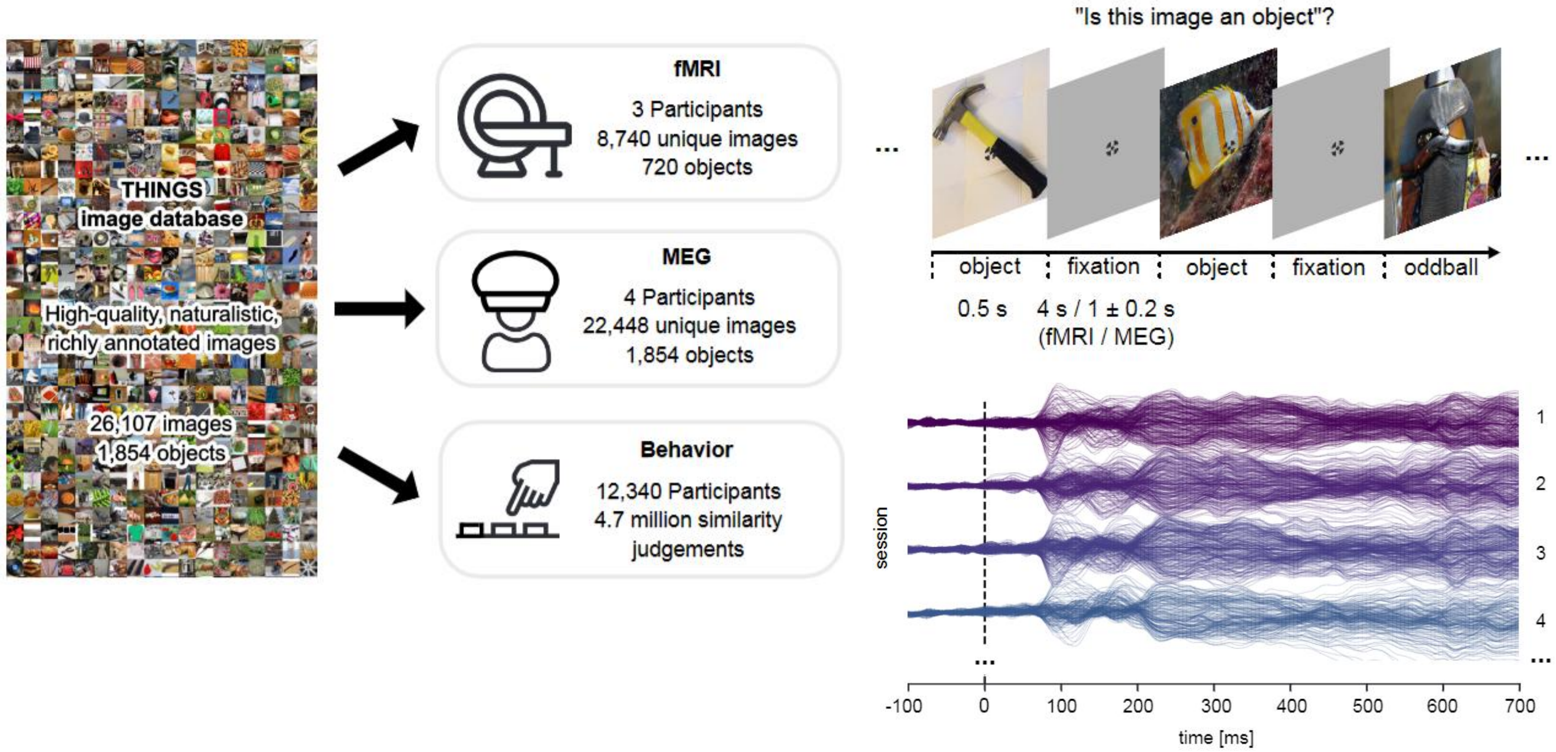
- Zone out/concentrate

- P300

- Normal/abnormal

EEG/MEG/fMRI datasets under visual stimulation





Dataset: THINGS-EEG

Dataset: THINGS-MEG

fMRI dataset: BOLD5000, NSD, THINGS-fMRI...

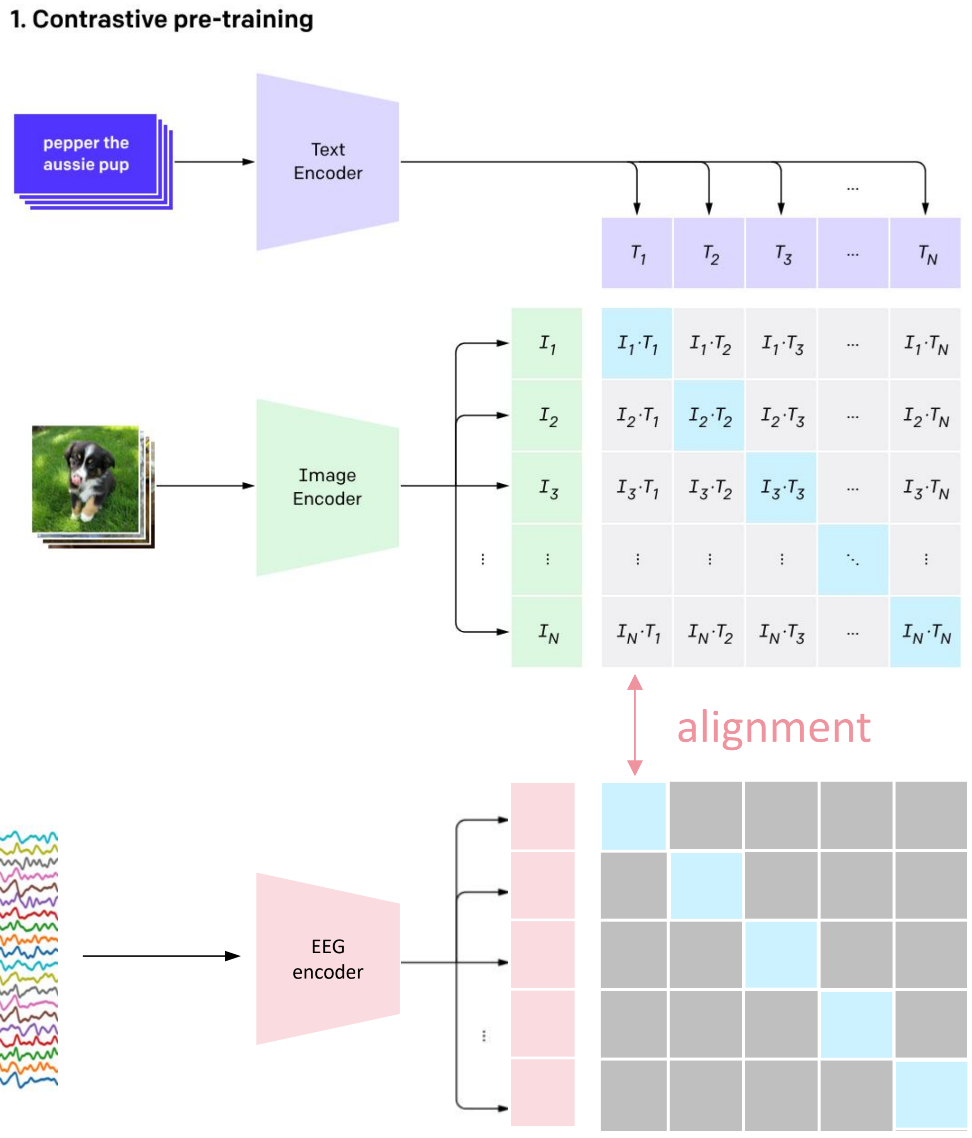
Alessandro et al., NeuroImage, 2022

Martin et al., eLife, 2023

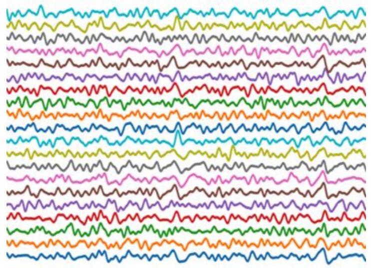
Brain-AI alignment: benefits

CLIP

一只小猫咪



Brain-AI alignment



Benefits for AI

1) better abstraction
2) better generalizability
3) Interpretability
4) AI safety

...

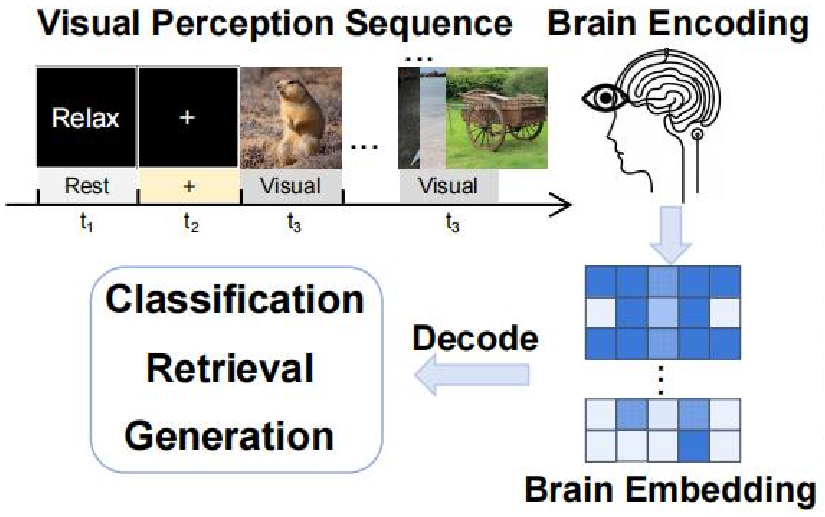
Benefits for neuroscience:

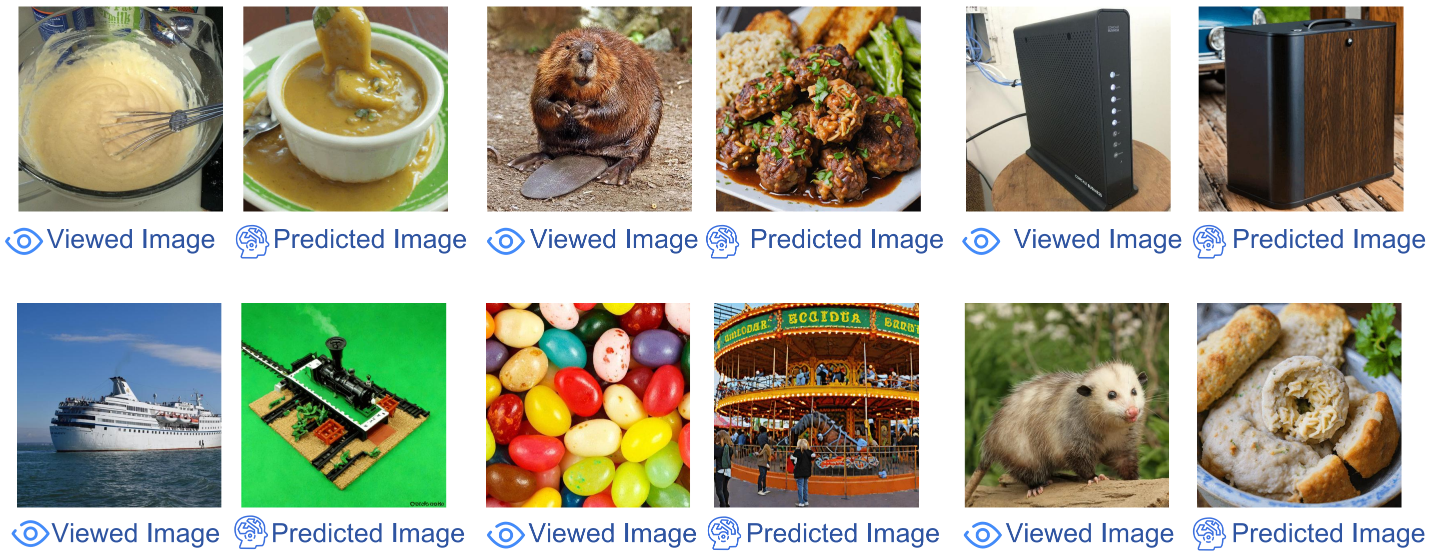
1) less neural data
2) multiple downstream tasks
3) zero-shot / few-shot capability
4) Virtual experimental platform
5) New science discovery

...

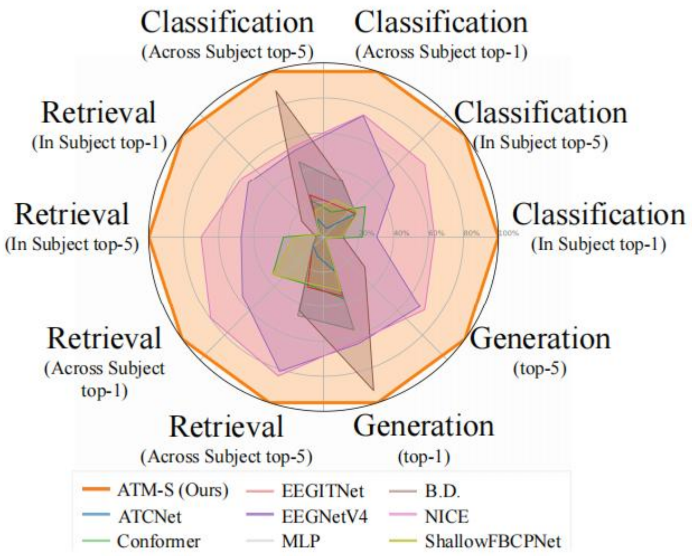
Visual decoding using brain signal

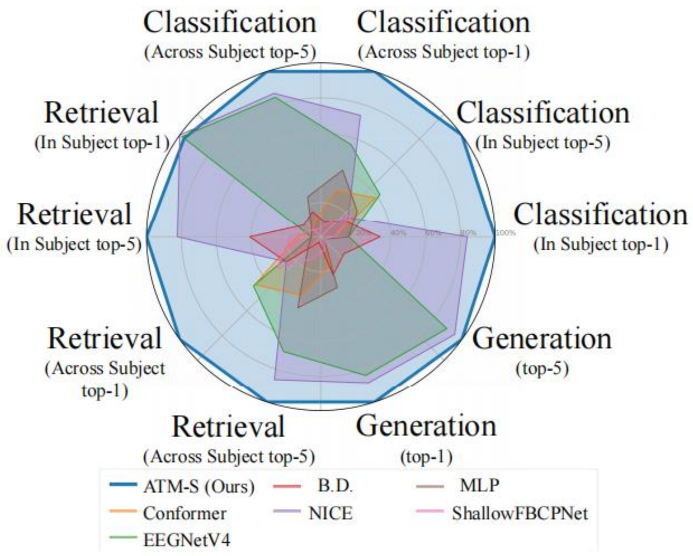
- Brain decoding and reconstruction



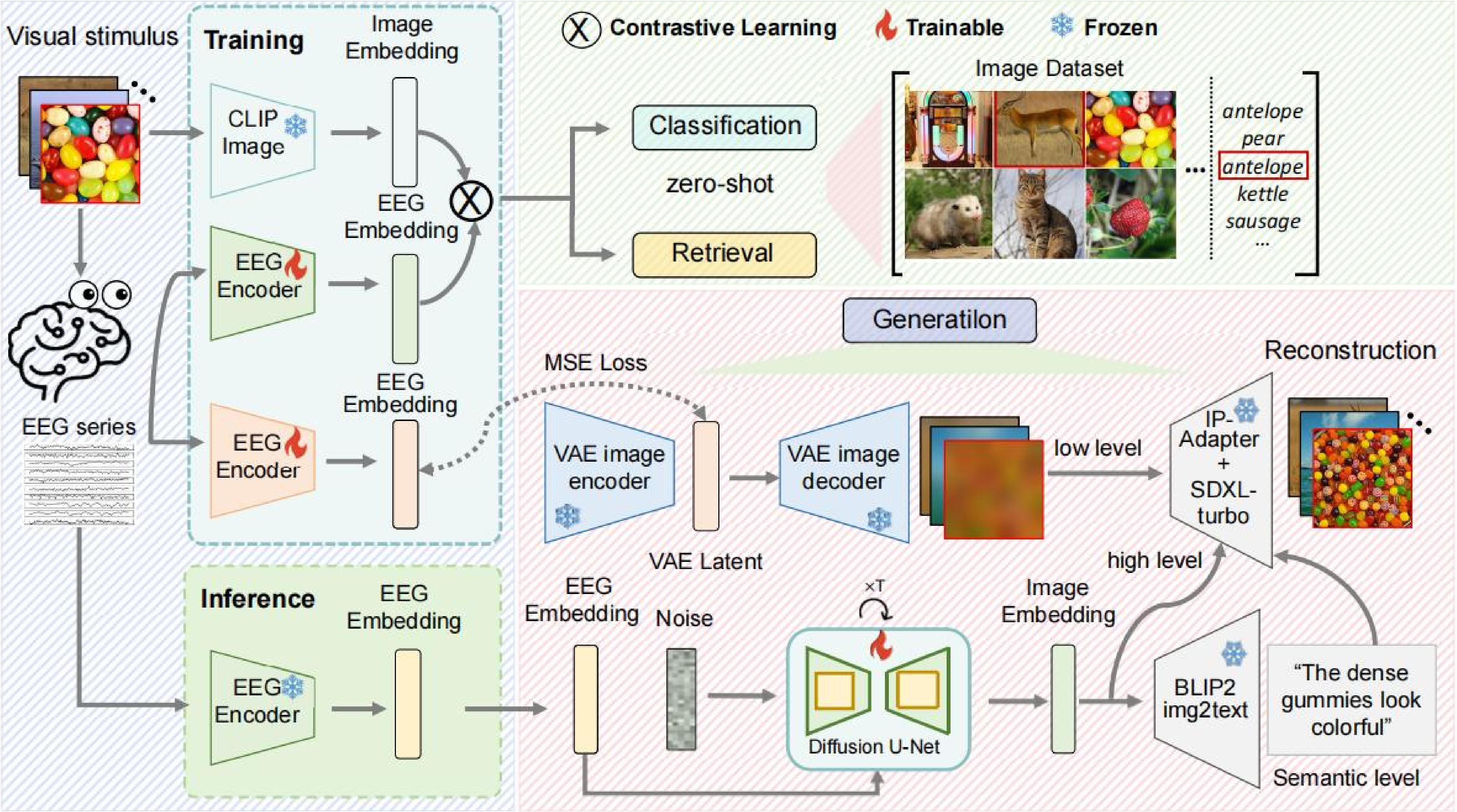


- Overall performance

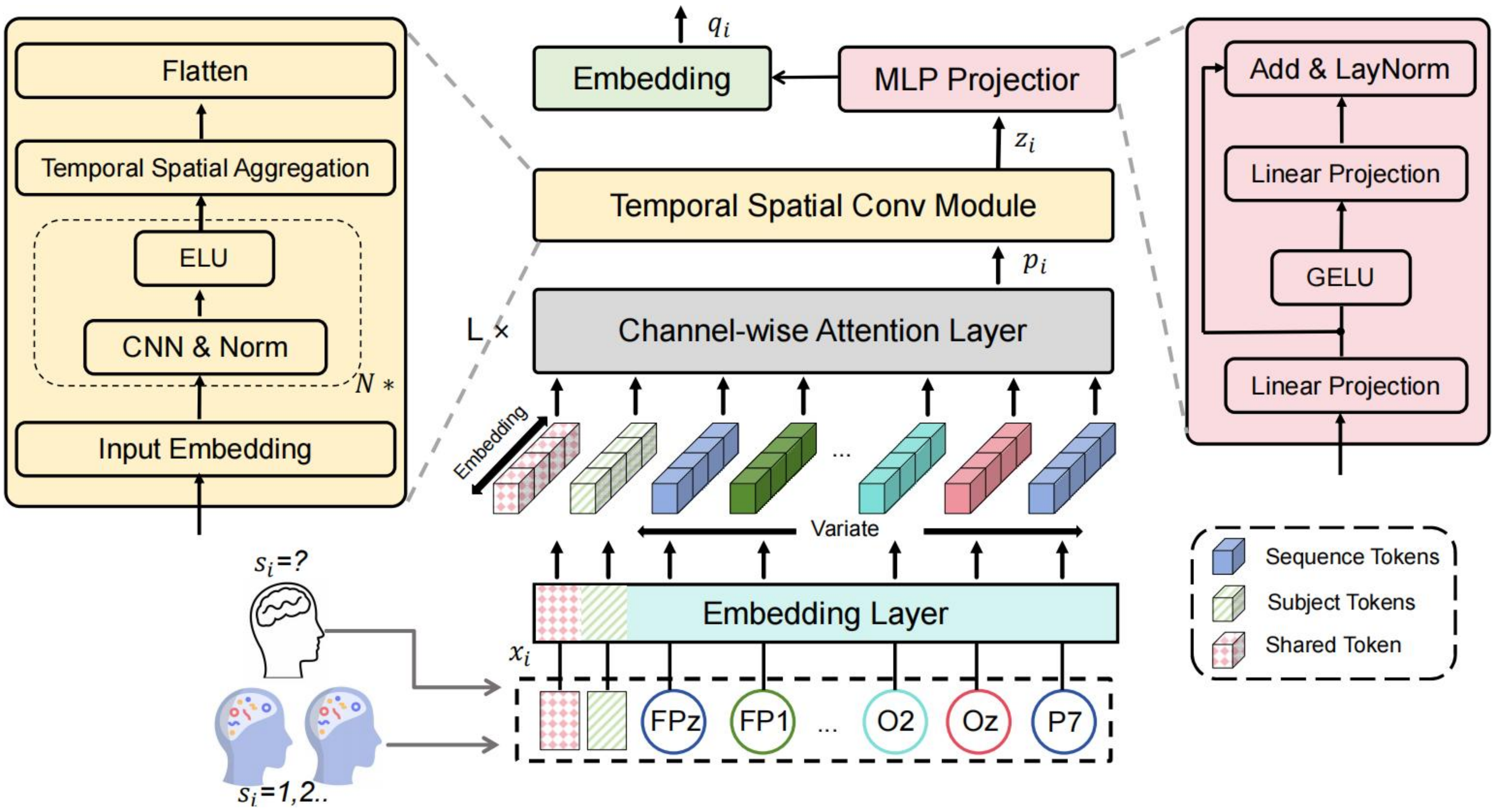




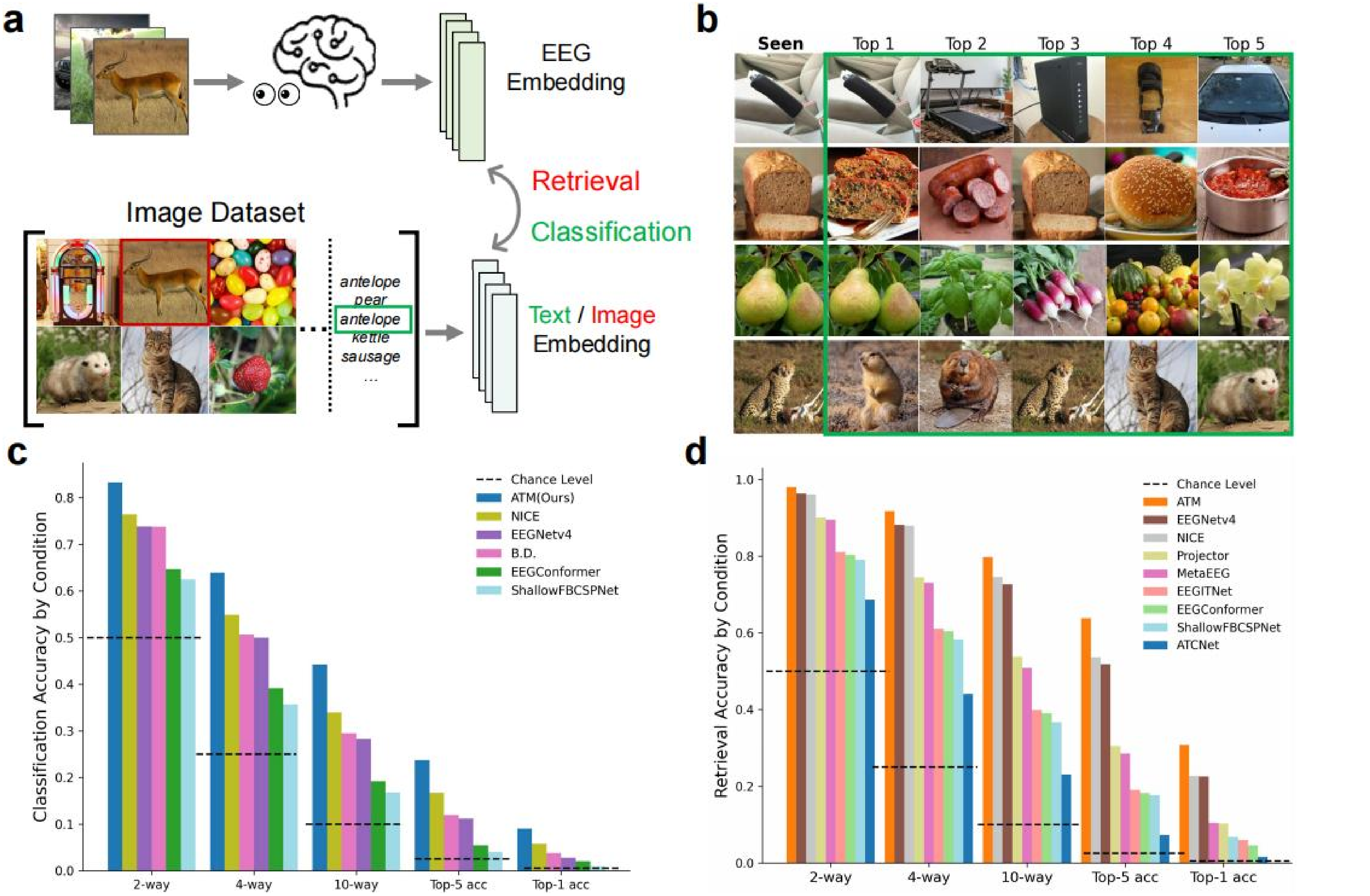
Visual decoding and reconstruction using EEG/MEG



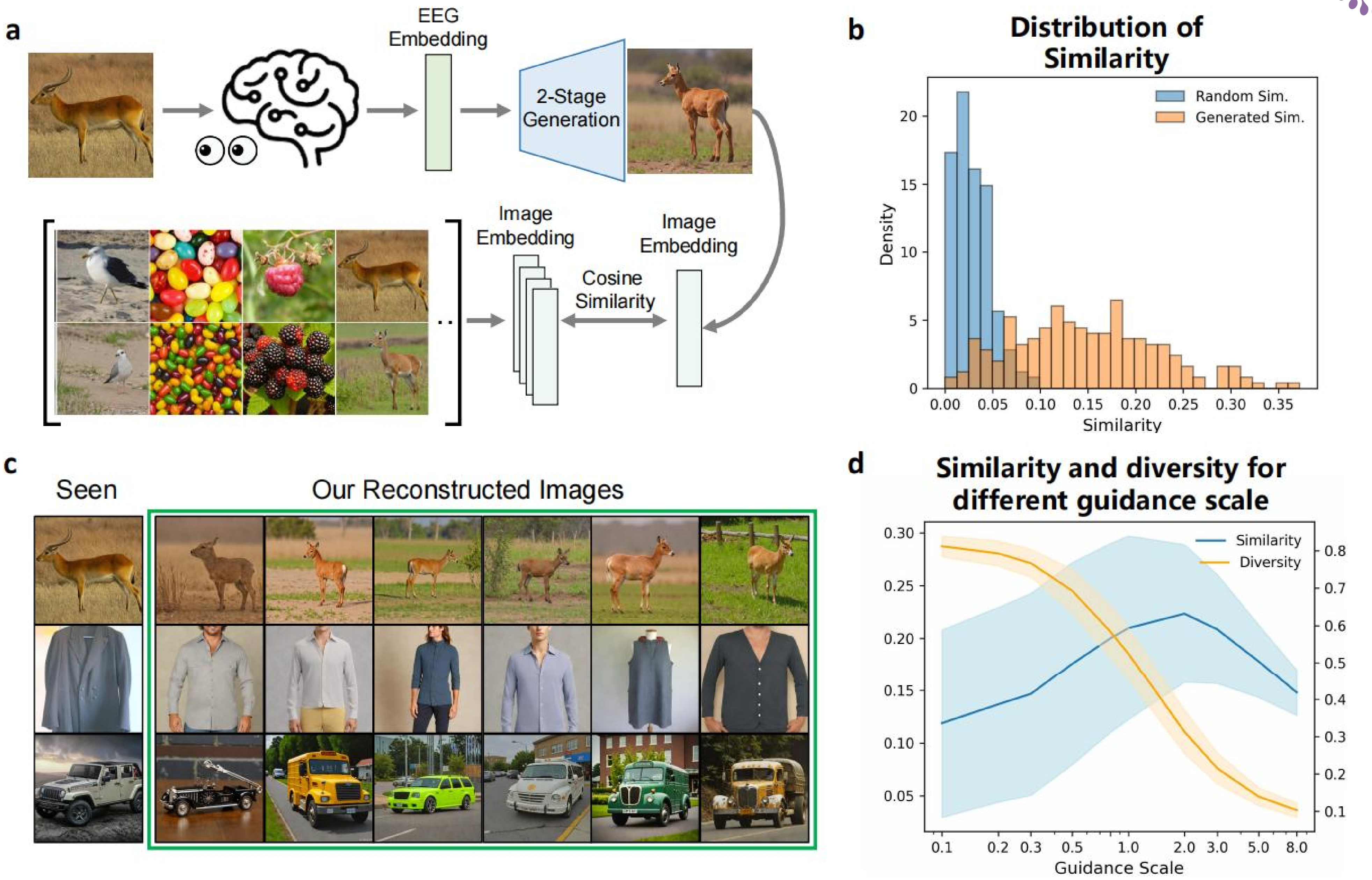
Structure of EEG encoder



EEG-based image retrieval and classification



EEG guidance image generation



Two-stage reconstruction pipeline

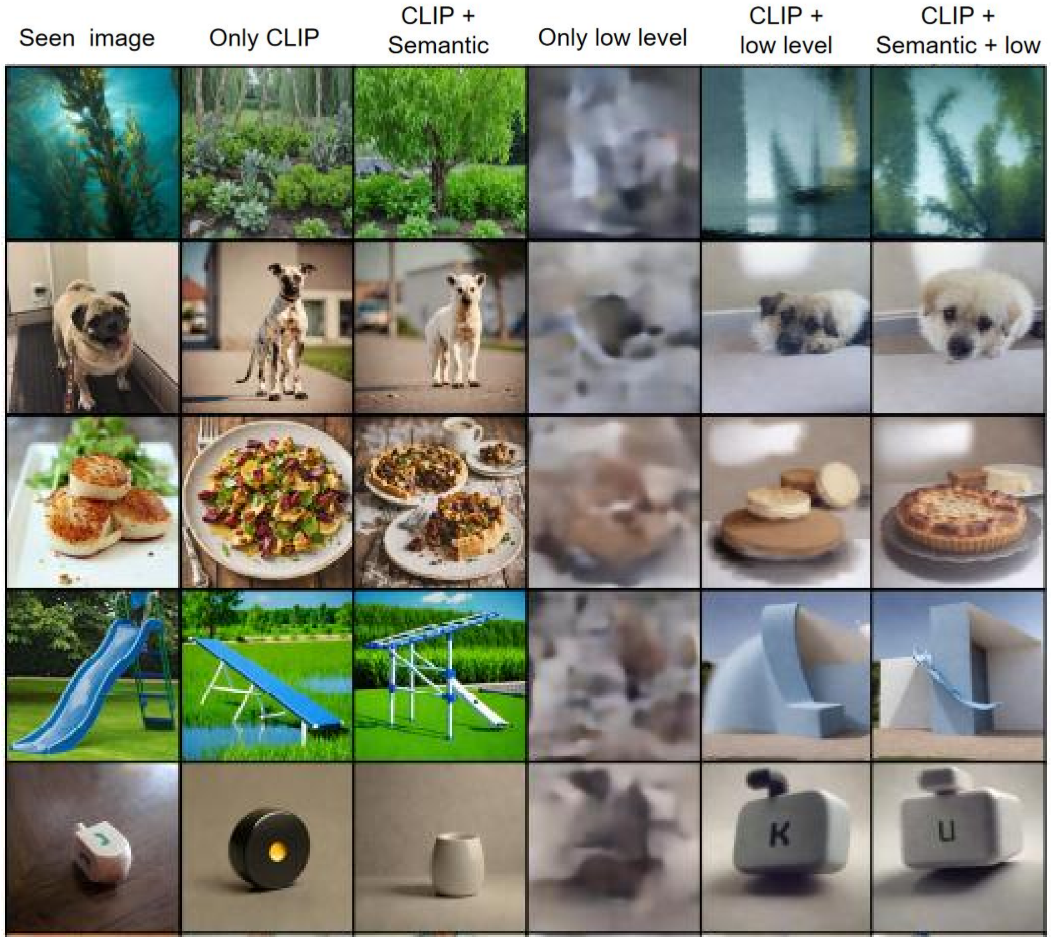
NEURAL INFORMATION PROCESSING SYSTEMS

- Reconstruction performance

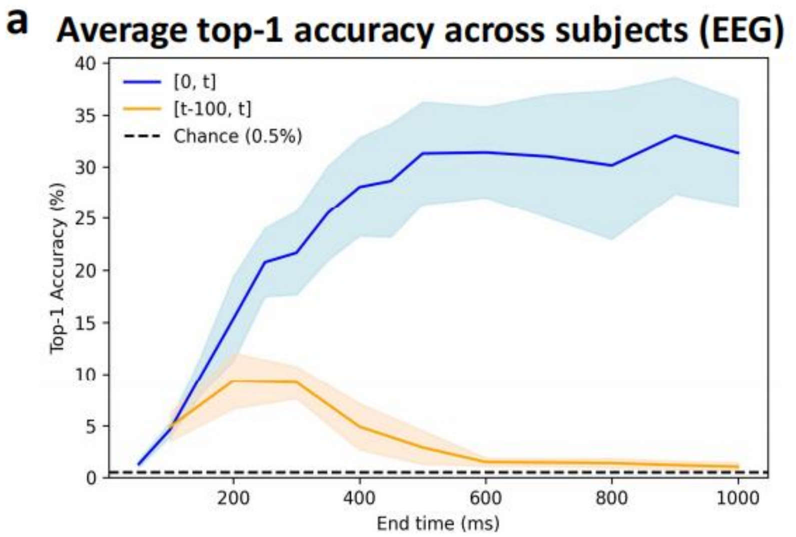
	Low-level		High-level			
Dataset ↑	SSIM ↑	AlexNet(2) ↑	AlexNet(5) ↑	Inception ↑	CLIP ↑	SwAV ↓
NSD-fMRI [4]	0.366	0.962	0.977	0.910	0.917	0.410
NSD-fMRI [33]	0.356	0.942	0.962	0.872	0.915	0.423
NSD-fMRI [41]	0.308	0.917	0.974	0.936	0.942	0.369
THINGS-MEG [4]	0.327	0.695	0.753	0.593	0.700	0.630
THINGS-MEG (averaged) [4]	0.336	0.736	0.826	0.671	0.767	0.584
THINGS-MEG (Ours)	0.340	0.613	0.672	0.619	0.603	0.651
THINGS-EEG (Ours)	**0.345**	**0.776**	**0.866**	**0.734**	**0.786**	**0.582**

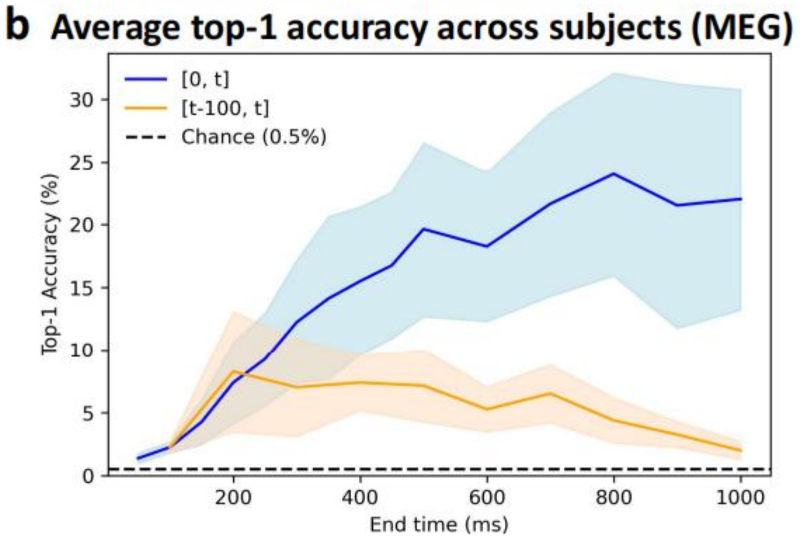
- Two-stage image reconstruction

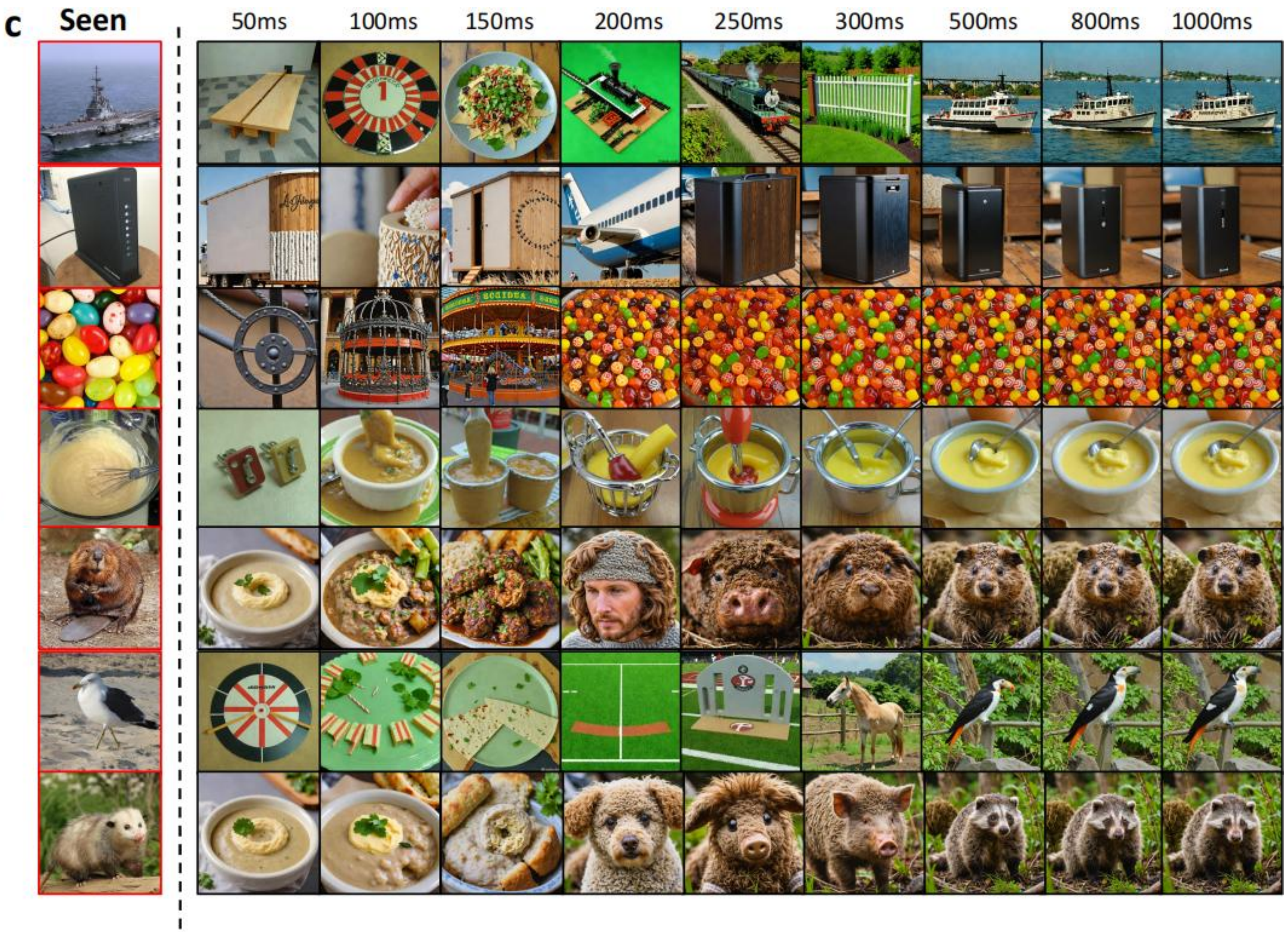
- Two-stage image reconstruction



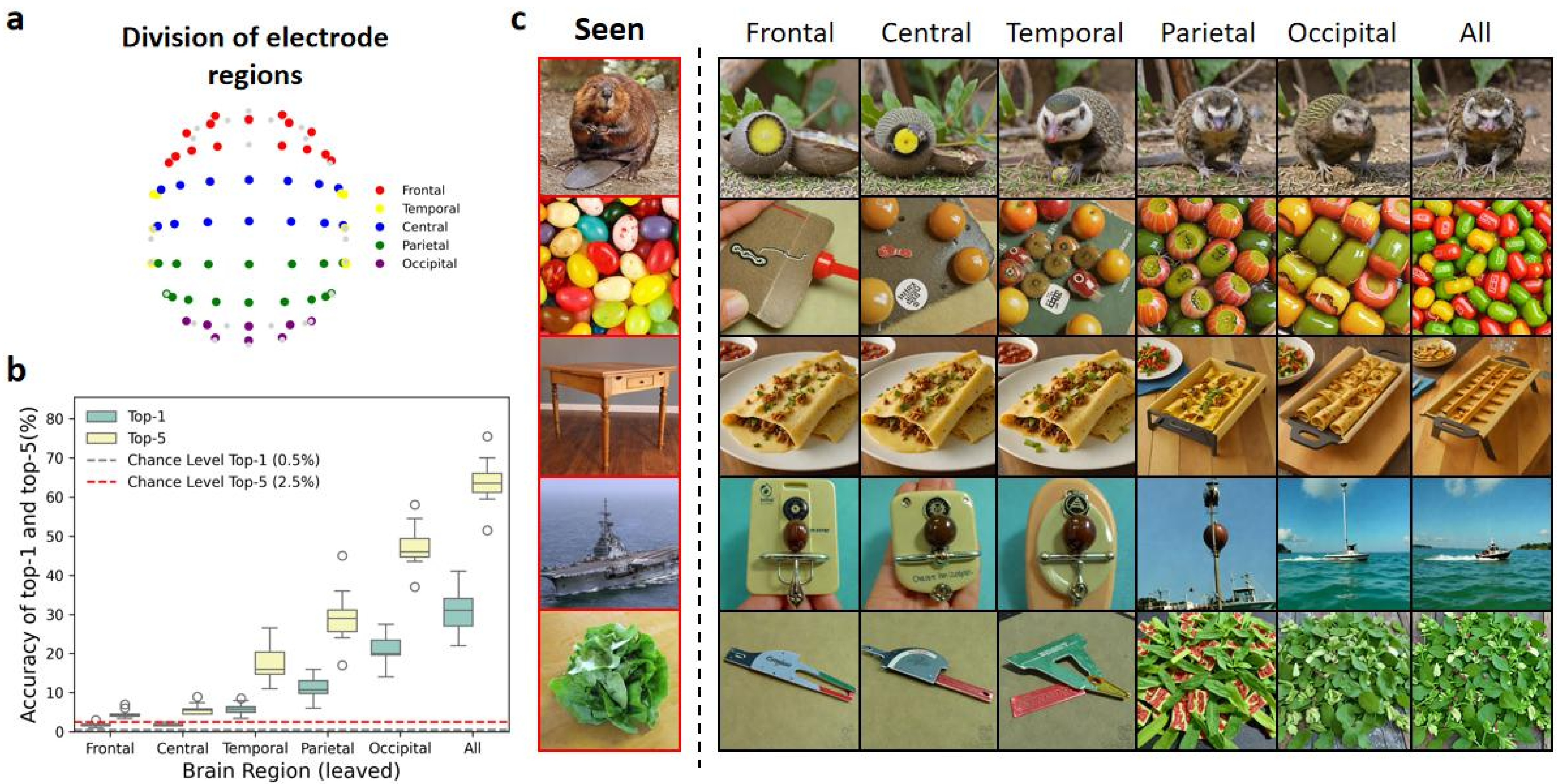
Time windows on visual retrieval and reconstruction



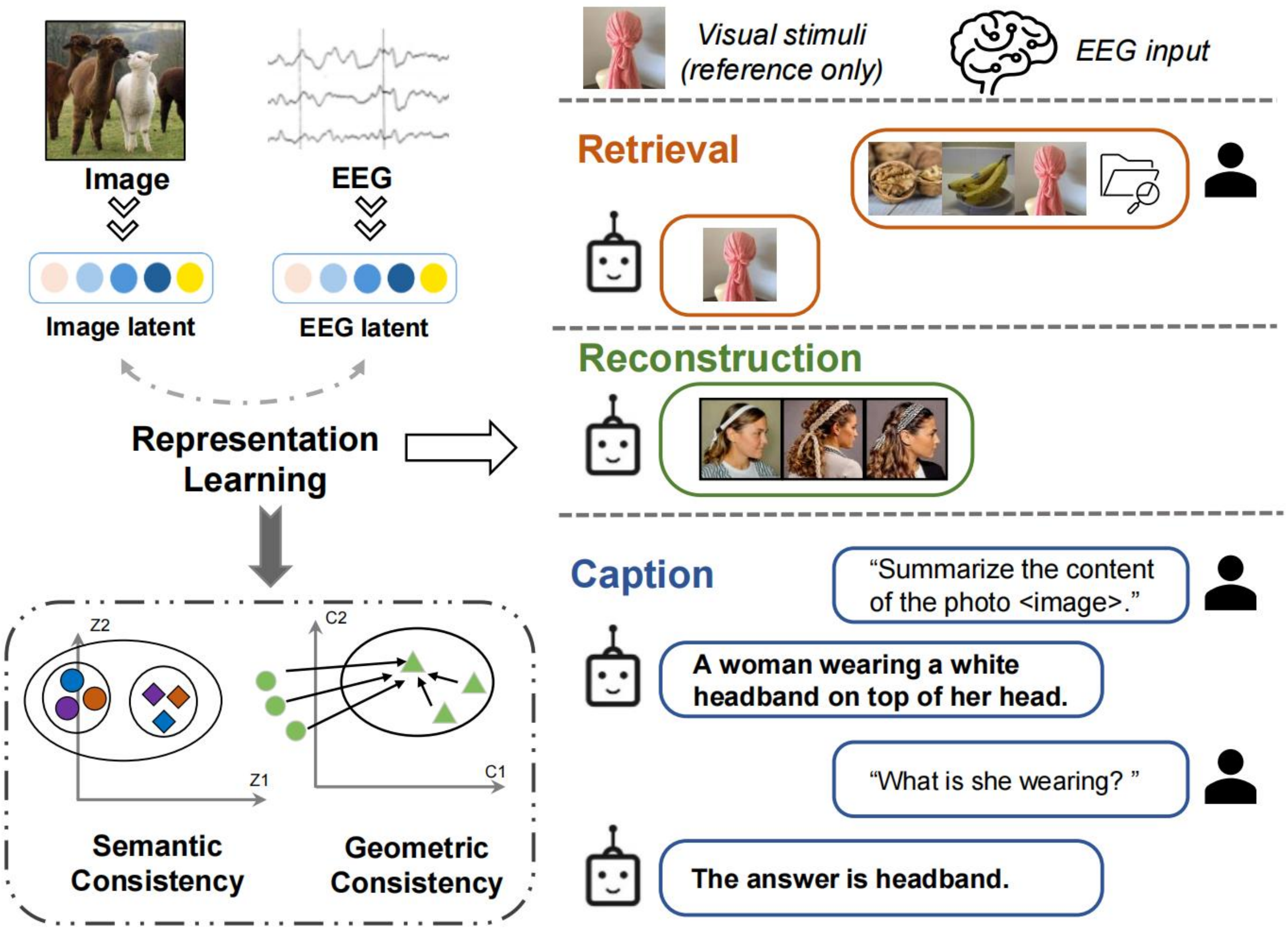




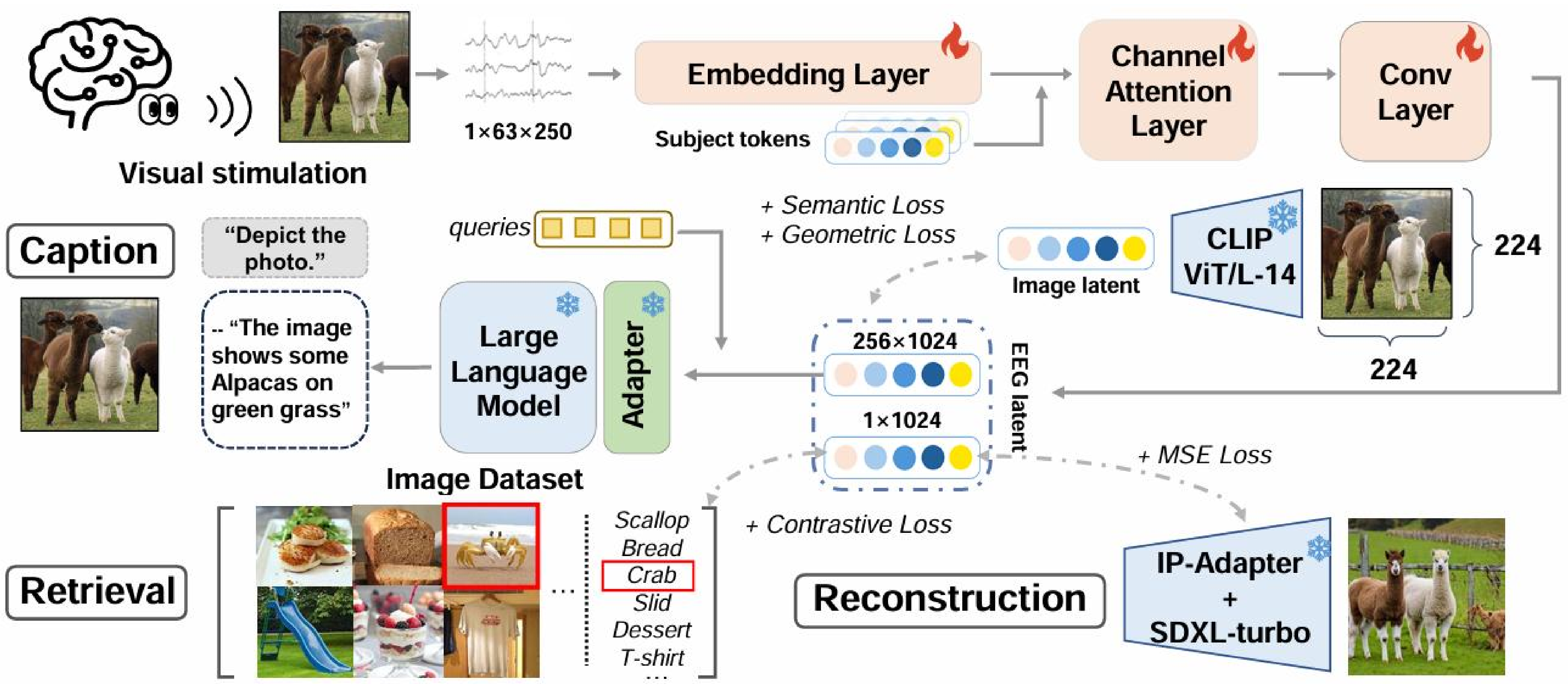
Visual Reconstruction from different brain regions



EEG Multimodal decoding——RealMind framework

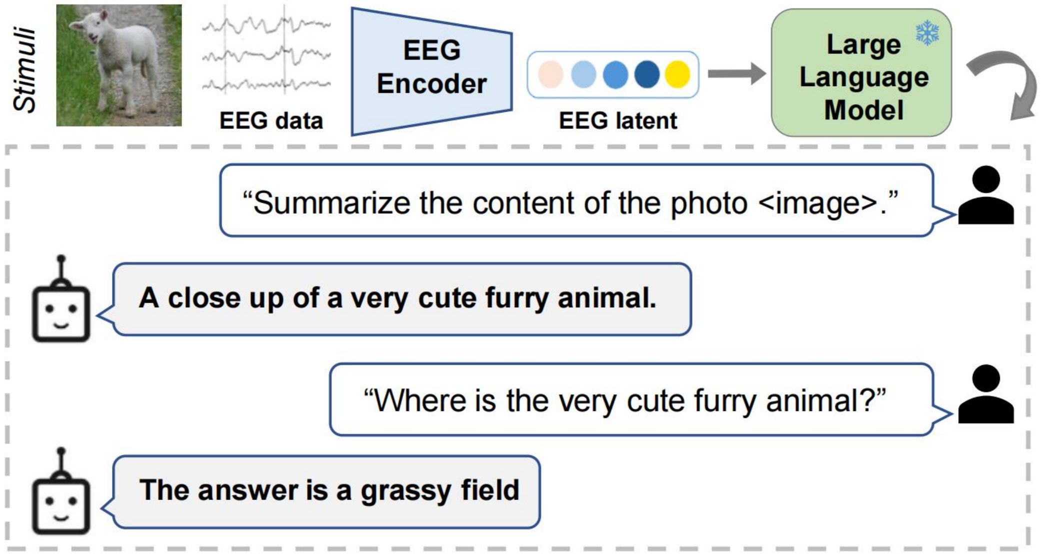


EEG Multimodal decoding——RealMind framework



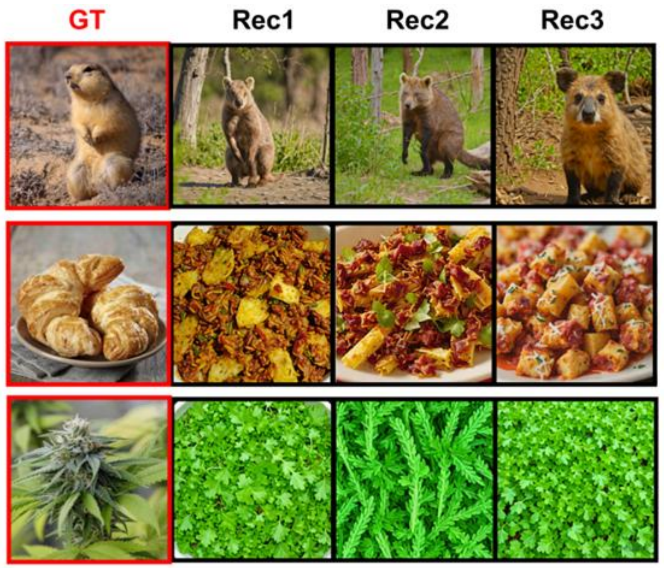
RealMind is able to adapt various down-stream decoding tasks

EEG-to-image captioning results



Caption Performance

Metric	Shikra captions		GIT captions		BLIP2 captions	
	L2Cap	I2Cap	L2Cap	I2Cap	L2Cap	I2Cap
BLEU-1 ↑	**26.59**	23.09	15.43	18.28	18.31	25.97
BLEU-4 ↑	**4.31**	3.88	2.90	3.70	3.25	4.65
METEOR ↓	**17.79**	15.00	15.43	14.20	15.01	18.40
Sentence ↑	**17.76**	19.62	14.26	23.78	15.60	25.99
CLIP-ViT-L ↑	**55.78**	53.91	57.83	61.34	58.77	57.52



Ground Truth: A yellow jackhammer standing on top of a dirt field.
Lat2Rec caption1: A small animal standing in the dirt.
Lat2Rec caption2: A small animal sitting on the ground in the woods.
Lat2Rec caption3: A small animal sitting on the ground.
Latent caption: **A close up of a very cute furry animal.**

Ground Truth: A plate of croissants on a table with a napkin.
Lat2Rec caption1: A white plate topped with a bowl of food.
Lat2Rec caption2: A plate of tofu with cheese and herbs.
Lat2Rec caption3: A plate of food with potatoes and cheese.
Latent caption: **A close-up of some food on a plate.**

Ground Truth: A marijuana plant is growing with lots of leaves.
Lat2Rec caption1: A close up of a bunch of green plants.
Lat2Rec caption2: A close up of a green plant.
Lat2Rec caption3: A close up of a bunch of green plants.
Latent caption: **A pile of green vegetables sitting on top of a table.**

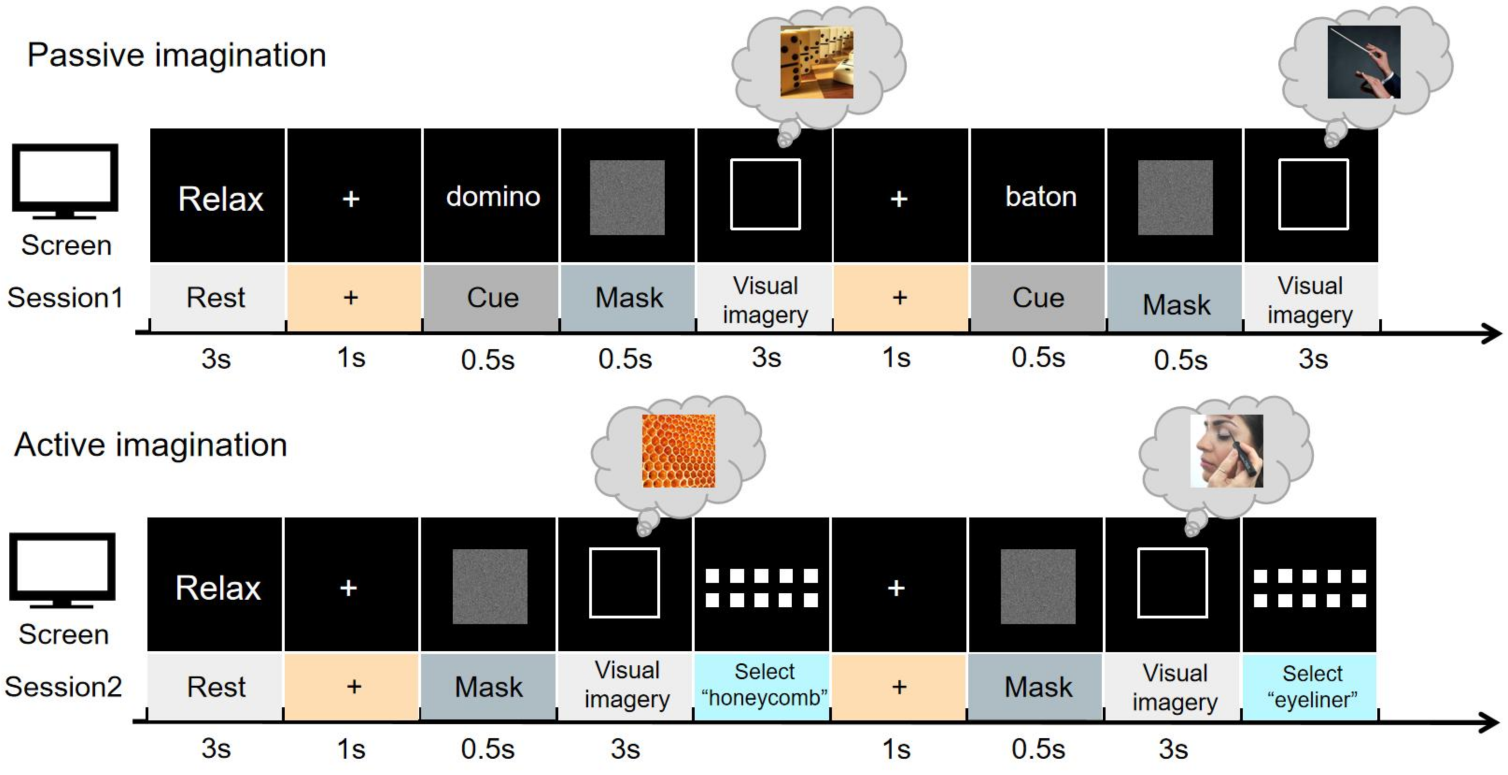
submitted to ICASSP 2025

Visual viewing to visual imagination: from viewing to imagining

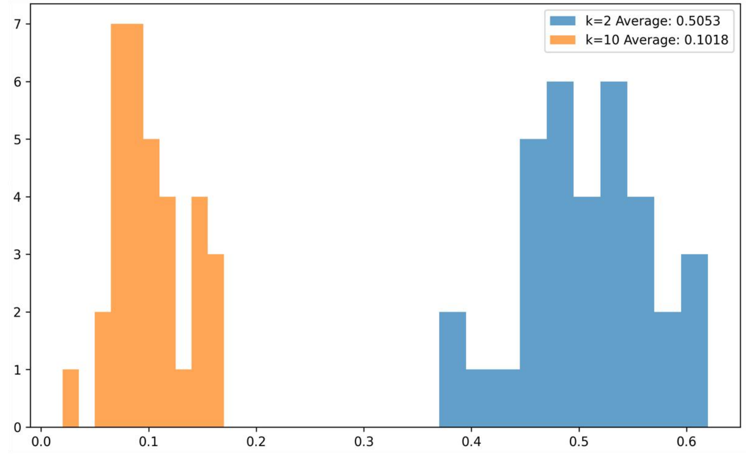
NEURAL INFORMATION PROCESSING SYSTEMS

Question: Can the decoding model for visual viewing (bottom up) be generalized to passive/active visual imagination (top down)?

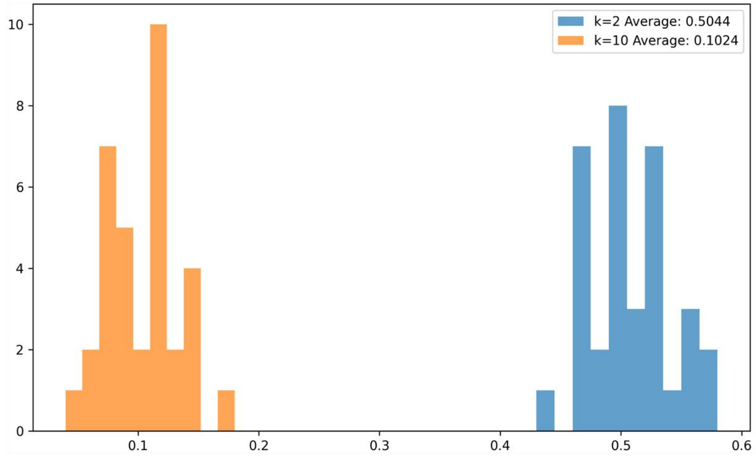
- We recuit college students to conduct visual experiments (n=34, #trial=10).
- The performance drops to **chance-level** without sufficient fine-tuning.



Accuracy distribution (retrieval)



Accuracy distribution (classification)



Summary: EEG-to-Image

NEURAL INFORMATION PROCESSING SYSTEMS

Existing problems

- The existing visual reconstruction methods rely generation model.
- The learned EEG representations in EEG-To-Image models have not been tested with other tasks.

Future directions

- Cross-subjects and multi-task
- More flexible neural network architectures and pretrained brain model
- Tansfer learning and meta-learning
- Unified framework for different downstream tasks

南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

生物医学工程系
Department of Biomedical Engineering

NEURAL INFORMATION PROCESSING SYSTEMS

🔥🔥代码和模型已开源! (已获得120+star⭐): https://github.com/ncclab-sustech/EEG_Image_decode

Thanks!

Project page

Paper link

NEURAL COMPUTING CONTROL
神经计算与控制实验室
Neural Computing & Control Lab