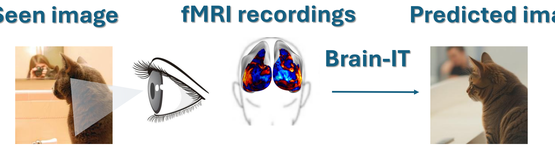
Brain-IT: Image Reconstruction from fMRI via Brain-Interaction Transformer

Roman Beliy, Amit Zalcher*, Jonathan Kogman, Navve Wasserman, Michal Irani*

The Weizmann Institute of Science

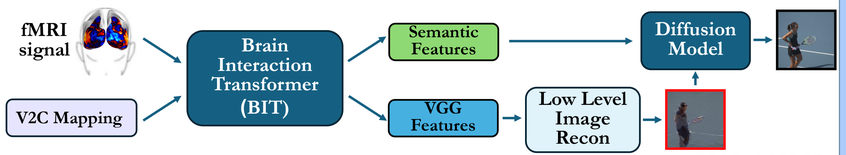
**Equal contributors*

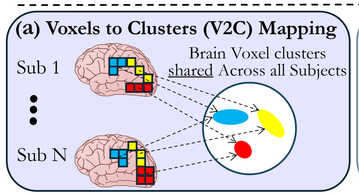
Reconstructing seen images from brain signals (fMRI)

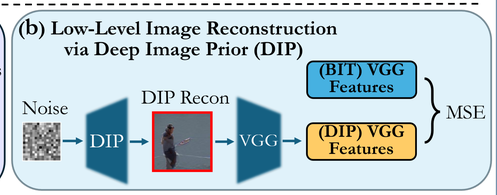


- **A brain-inspired approach for image reconstruction from fMRI.**
- **Cross-Attention between *functional-clusters* of brain-voxels (shared by all subjects) ➔ Integrates information *within* & *across* brains.**
- **Faithful reconstruction of seen images even with only 15 min of fMRI data.**

Overview of the approach



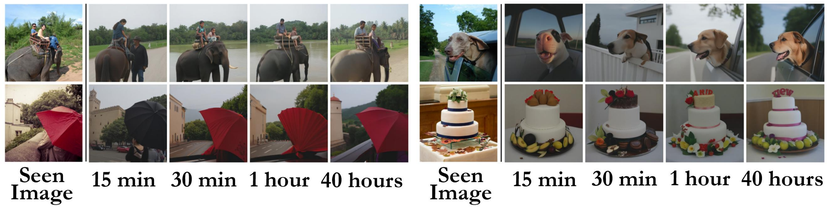




Brain Interaction Transformer (BIT) transforms fMRI signals into Semantic & VGG features: *(i) Low-Level branch:* reconstructs a coarse image from VGG features. *(ii) Semantic branch:* reconstructs semantic CLIP features to guide the diffusion model (initialized by the low-level branch).
(a) Voxel-to-Cluster mapping (V2C): Maps all voxels of all subjects to 128 shared functional clusters. **(b) Low-level Image Reconstruction**: VGG-predicted features are inverted using *Deep Image Prior* to get a coarse image.

Reconstruction of seen images from fMRI using "Brain-IT"

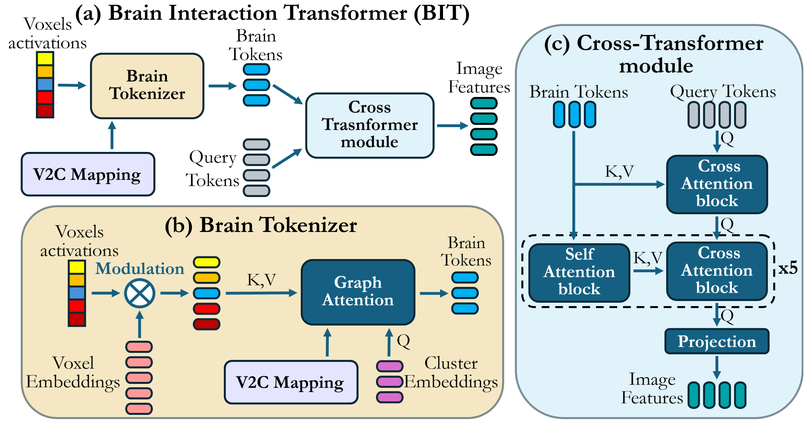
(b) Reconstruction with limited amount of training data:



Quantitative Evaluation

Method	Low-Level				High-Level			
	PixCorr ↑	SSIM ↑	Alex(2) ↑	Alex(5) ↑	Incep ↑	CLIP ↑	Eff ↓	SwAV ↓
NeuroVLA (Shen et al., 2024)	0.265	0.357	93.1%	97.1%	96.8%	**97.5%**	0.633	0.321
MindBridge (Wang et al., 2024a)	0.151	0.263	87.7%	95.5%	92.4%	94.7%	0.712	0.418
NeuroPictor (Huo et al., 2024)	0.229	0.375	96.5%	98.4%	94.5%	93.3%	0.639	0.350
MindEye2 (Scotti et al., 2024)	0.322	0.431	96.1%	98.6%	95.4%	93.0%	0.619	0.344
MindTuner (Gong et al., 2025)	0.322	0.421	95.8%	98.8%	95.6%	93.8%	0.612	0.340
BrainIT (Ours)	**0.386**	**0.486**	**98.4%**	**99.5%**	**97.3%**	96.4%	**0.564**	**0.320**
Results on 1 hour (out of 40)								
MindEye2 (1 hour)	0.195	0.419	84.2%	90.6%	81.2%	79.2%	0.810	0.468
MindTuner (1 hour)	0.224	0.420	87.8%	93.6%	84.8%	83.5%	0.780	0.440
BrainIT (1 hour)	**0.331**	**0.473**	**97.1%**	**98.6%**	**94.4%**	**93.0%**	**0.648**	**0.370**

Brain Interaction Transformer (BIT) Architecture



(a) BIT model: transforms fMRI voxel activations into image features. **(b) Brain Tokenizer**: aggregates voxel activations into cluster-level Brain Tokens (one token per cluster). **(c) Cross-Transformer Module**: integrates and refines Brain Tokens, using query tokens to predict localized image features.

Qualitative comparison against leading methods