FreqSelect : Frequency-Aware fMRI-to-Image Reconstruction

Junliang Ye [1] Lei Wang [2,3] Md Zakir Hossain [1,3,4]

[1]Australian National University [2] Griffith University [3] Data61/CSIRO [4] Curtin University

Motivation

Reconstructing visual experiences from fMRI signals has emerged as a powerful window into the neural code of human vision, yet existing pipelines often struggle to disentangle task-relevant structure from the rich but noisy frequency content inherent in fMRI. While deep generative models like VDVAE(Very Deep VAEs) and diffusion-based decoders capture low-level and high-level features, they treat all frequency bands uniformly, leaving both informative and spurious components to compete in the learning process. Neurophysiological and image-processing studies alike suggest that distinct frequency ranges carry differential semantic and structural cues—some supporting texture and edge detail, others dominated by physiological noise or scanner artifacts. Without a mechanism to adaptively emphasize the bands most predictive of visual content, downstream models must expend capacity filtering noise during training, which can undermine reconstruction fidelity and slow convergence.

To address this, we introduce an **Adaptive Selective Frequency Filter** that learns, which spectral bands to **retain** or **suppress** before latent encoding. By dynamically tuning frequency responses according to reconstruction loss, our module explores the relationship between fMRI and image information of different frequencies,suppresses noise,yields sharper, more semantically faithful reconstructions.

Method

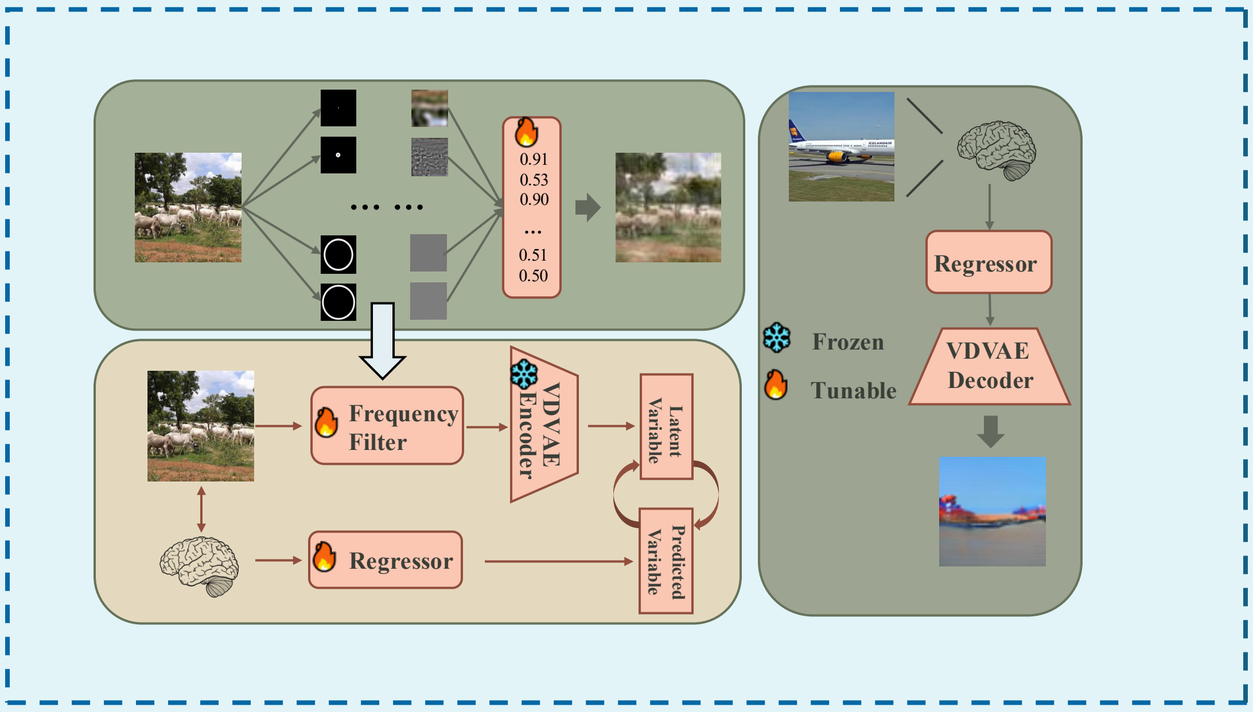


Figure 1： Adaptive Selective Frequency Filter with VDVAE fMRI-to-Image Reconstruction

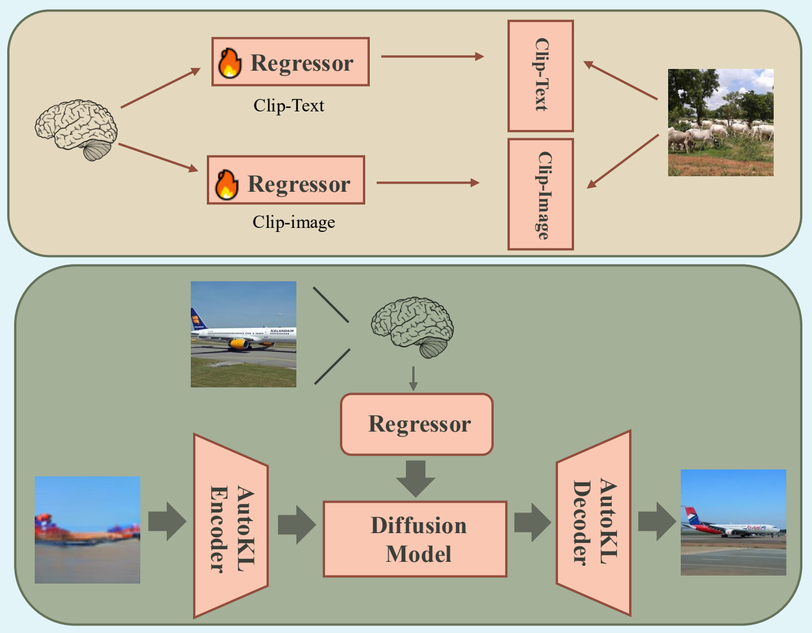


Figure 2:Diffusion Denoising Reconstruction of fMRI

Training Phase

Image Branch

The original scene image first passes through the **Adaptive Selective Frequency Filter**🔥. This module divides the image into ***N*** frequency-band channels in the Fourier domain, each corresponding to a specific frequency range, and assigns a learnable weight to each channel; a weight of 1 retains that band completely, while 0 fully filters it out.

The weighted frequency-domain image is then fed into the **frozen** VDVAE encoder ❄, yielding the latent variable.

fMRI Branch

The fMRI signals recorded while the subject views the same image are input to a **tunable** regressor 🔥, which predicts the corresponding latent variable.

Loss & Optimization

The discrepancy between the predicted and latent variable is used as the **MSEloss** to jointly update the **Adaptive Selective Frequency Filter** and the regressor, while the VDVAE encoder remains frozen.

Testing Phase

New fMRI signals are mapped to latent representations using only the trained regressor.

The frozen VDVAE decoder ❄ decodes these latent back into images, completing the reconstruction.

Training Phase

For each training sample, the ground-truth caption and original scene image are fed into the frozen CLIP encoders.

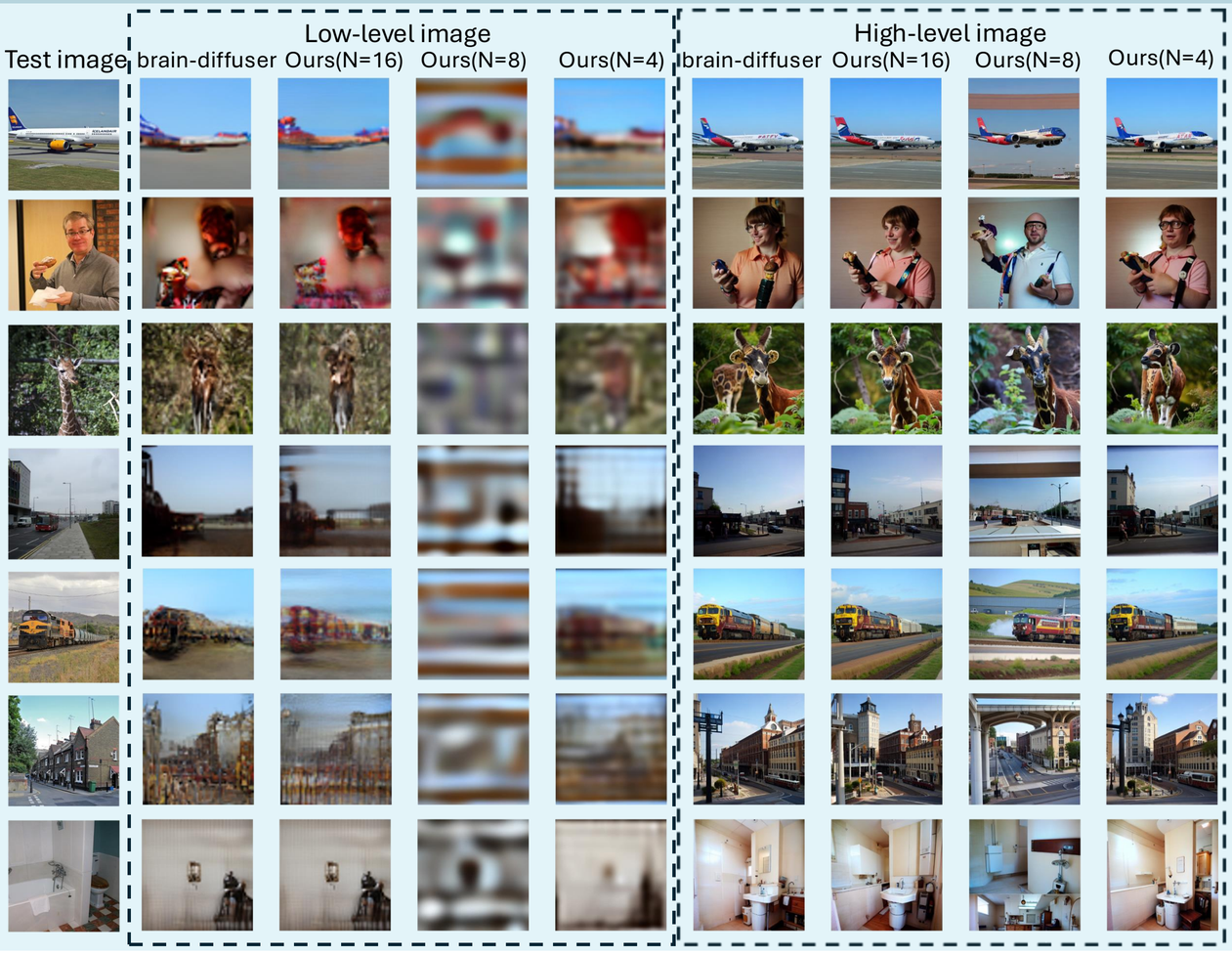
Loss & Optimization

Compute the MSE loss between predicted and true CLIP-Text embeddings, and separately between predicted and true CLIP-Image embeddings.

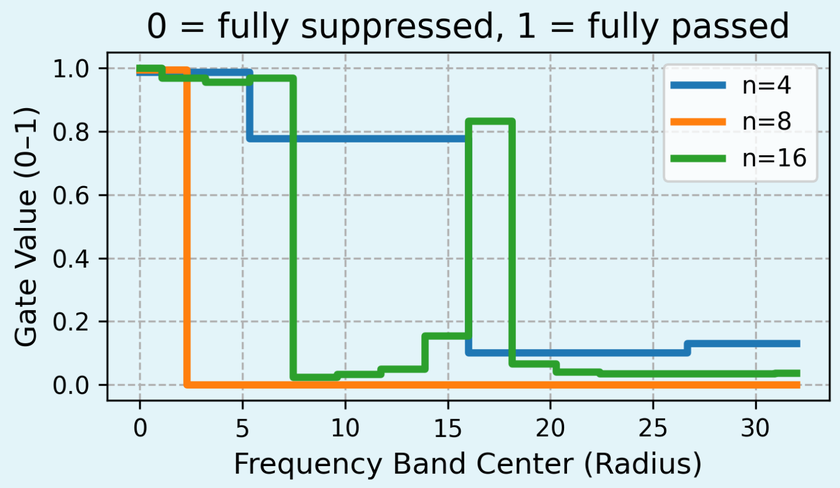
Testing Phase

New fMRI signals are mapped to text and image embeddings via the trained Regressor. The pictures in Figure 1 is denoised by the diffusion model, **conditioned** on these two predicted CLIP embeddings.

Visual result



Channel Pass-Through Rate



Strong low-frequency pass-through

n = 4: broad, coarse band selection

n = 8: retains only low-frequency bands

n = 16: fine-grained, channel-specific weighting

Pass-through rate: 0 = fully suppressed, 1 = fully passed

Adaptive emphasis on informative bands, suppression of noise bands

Evaluation

Low Level Evaluation

Method	PixCorr ↑	SSIM ↑	AlexNet(2) ↑	AlexNet(5) ↑
Brain-Diffuser*	**0.304**	0.293	**96.84%**	**97.48%**
Ours (N=4)	0.2902	0.2914	94.94%	96.80%
Ours (N=8)	0.0738	0.2633	86.82%	93.13%
Ours (N=16)	0.2734	**0.2961**	96.25%	97.46%

High Level Evaluation

Method	Inception ↑	CLIP ↑	EffNet-B ↓	SwAV ↓
Brain-Diffuser*	**88.6%**	92.5%	**0.761**	**0.410**
Ours (N=4)	88.11%	91.85%	0.7725	0.4176
Ours (N=8)	86.00%	91.81%	0.7986	0.4520
Ours (N=16)	88.36%	**92.65%**	0.7672	0.4141

Conclusion

We propose an **Adaptive Selective Frequency Filter** that plugs into fMRI-to-image workflows, learning to keep or discard spectral bands to maintain—and even boost—reconstruction quality on both low- and high-level metrics. The resulting band-pass curves also shed light on which frequencies carry key visual information. Future work will fuse this spectral filtering with spatial attention and extend it to modalities like EEG and MEG.