1. Masked signal modeling with large-scale noisy EEG data

EEG Signals

Split & Mask

EEG Encoder

Latent

EEG Decoder

Reconstruction

2. Fine-tuning with EEG-image pairs

Generated Image

$\times (T-1)$

D

Z_0

Denoising U-Net

Z_{T-1}

Denoising U-Net

Z_T

Pre-trained Stable Diffusion

3. Align EEG, text and image spaces

Paired Image

Projection Layer

EEG Embed

CLIP Image Encoder

Image Embed

Cosine similarity