STEP-AWARE RESIDUAL-GUIDED DIFFUSION FOR EEG SPATIAL SUPER-RESOLUTION

Hongjun Liu[1]* **Leyu Zhou**[1]* **Zijianghao Yang**[1] **Chao Yao**[2]†
[1]School of Intelligence Science and Technology, University of Science and Technology Beijing
[2]School of Computer and Communication Engineering, University of Science and Technology Beijing

ABSTRACT

For real-world brain–computer interface (BCI) applications, lightweight Electroencephalography (EEG) systems offer the best cost–deployment balance. However, such spatial sparsity of EEG limits spatial fidelity, hurting learning and introducing bias. EEG spatial super-resolution methods aim to recover high-density EEG signals from sparse measurements, yet is often hindered by distribution shift and signal distortion and thus reducing fidelity and usability for EEG analysis and visualization. To overcome these challenges, we introduce SRGDiff, a step-aware residual-guided diffusion model that formulates EEG spatial super-resolution as dynamic conditional generation. Our key idea is to learn a dynamic residual condition from the low-density input that predicts the step-wise temporal and spatial details to add and uses the evolving cue to steer the denoising process toward high density reconstructions. At each denoising step, the proposed residual condition is additively fused with the previous denoiser feature maps, then a step-dependent affine modulation scales and shifts the activation to produce the current features. This iterative procedure dynamically extracts step-wise temporal rhythms and spatial-topographic cues to steer high-density recovery and maintain a fidelity–consistency balance. We adopt a comprehensive evaluation protocol spanning signal-, feature-, and downstream-level metrics across SEED, SEED-IV, and Localize-MI and multiple upsampling scales. SRGDiff consistently achieves higher SNR than the baseline ESTformer and STAD among Localize-MI, SEED and SEED-IV datasets, with up to roughly 75% relative SNR improvement in the most challenging $8\times$ setting. Moreover, topographic visualizations comparison and substantial EEG-FID gains jointly indicate that our SR EEG mitigates the spatial–spectral shift between low- and high-density recordings. Our code is available at https://github.com/DhrLhj/ICLR2026SRGDiff.

1 INTRODUCTION

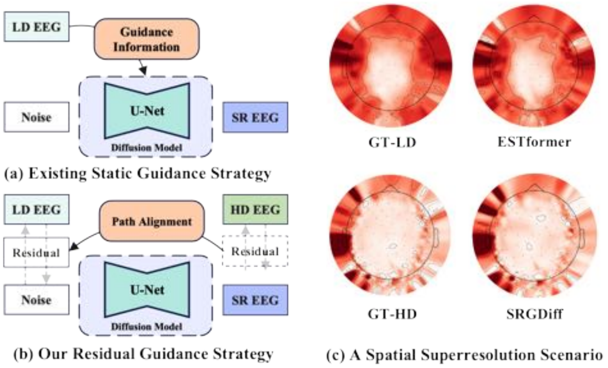
Electroencephalography (EEG) is a noninvasive technique for monitoring the brain's electrical activity, with widespread applications in neuroscience and clinical practice—ranging from brain–computer interfaces and epilepsy diagnosis to emotion recognition (Jiang et al., 2025). However, EEG's spatial resolution is inherently constrained by the number of scalp electrodes and the volume-conduction effect (Li et al., 2025a). High-density (HD) systems with hundreds of channels can mitigate these issues but are costly, cumbersome to deploy, and uncomfortable for extended wear, whereas low-density (LD) setups (e.g., 8 or 16 electrodes) are far more practical yet suffer from severe under-sampling bias (Wang et al., 2025). Indeed, as illustrated in Figure 1(c), the inter-channel activation patterns of 256-channel HD EEG diverge dramatically from those of 16-channel LD EEG, highlighting the strong bias in sparse recordings. EEG spatial super-resolution (SR) has therefore garnered growing attention, with methods that reconstruct high-density EEG from sparse recordings increasingly explored and applied.

Traditionally, EEG spatial super-resolution has relied on direct feature-mapping techniques that learn an end-to-end mapping from low-density to high-density representations. These methods fall

*Equal contribution.

†Corresponding author.

into two main categories: one employs convolutional neural networks or Transformers to upsample LD feature maps into HD ones (Tang et al., 2022), and the other leverages generative adversarial networks-based architectures that synthesize SR EEG signals conditioned on LD inputs (Wang et al., 2024). However, by treating the mapping as a static projection, these approaches often oversimplify the complex, nonlinear inter-electrode dependencies or demand vast amounts of training data and compute, resulting in overly smooth, detail-poor reconstructions that fail to capture true spatial consistency. Figure 1(c) indicates that such feature-mapping methods merely extend LD information, rather than recovering authentic HD channel relationships.



(a) Existing Static Guidance Strategy

(b) Our Residual Guidance Strategy

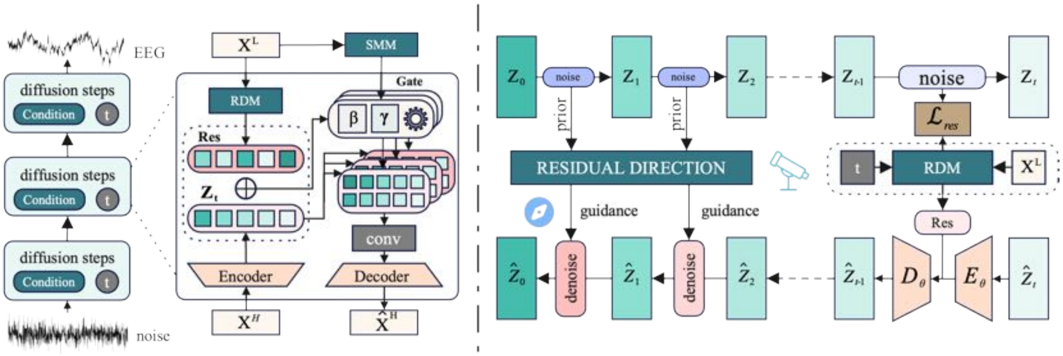
(c) A Spatial Superresolution Scenario

Figure 1: (a) Existing static guidance strategy vs. (b) our residual guidance strategy for EEG super-resolution, and (c) corresponding topographical maps of LD input, ESTformer output, GT HD EEG, and SRGDiff reconstruction.

Recently, diffusion models have been widely applied to time-series generation and missing-data imputation (Huang et al., 2025; Yuan & Qiao, 2024; Li et al., 2025b). In this context, EEG spatial super-resolution can be cast as conditional generation, where LD observations guide the recovery of HD signals. Within this line of work, researchers mainly focused on conditioning strategies through concatenating low-density features with the noise input (Vetter et al., 2024) or using cross-attention between modalities (Wang et al., 2025) as shown in Figure 1(a). While effective in practice, these approaches remain susceptible to a consistency–fidelity trade-off. Interpolation-oriented SR tends to cause distribution shift, making reconstructions adhere too closely to the LD observation and deviate from the HD ground truth. Conversely, generation-oriented SR often introduces distortion, producing HD-like content that fails to remain consistent with the LD input.

To tackle these challenges, we introduce Step-aware Residual-Guided Diffusion (SRGDiff) for EEG Spatial Super-Resolution, which reframes super-resolution as a dynamic conditional generation task. The core idea is to estimate the forward-noising residual from low-density channels, and use it as a per-step corrective direction in the reverse process. Technically, SRGDiff first encodes the low-density EEG with a pre-trained VAE encoder to obtain a compact latent and multi-scale features, and applies forward diffusion to the high-density latent. At each reverse step, a lightweight residual head predicts a path residual from the low-density features and uses it as a directional correction that is additively fused with the previous denoising features to form an incremental feature. The feature is then weighted with a step-dependent affine modulation estimated from the low-density features and the timestep embedding, yielding the current denoised features. This loop repeats over timesteps, coupling the low-density forward-noising and high-density reverse-denoising trajectories and progressively steering samples toward the high-density manifold. Our main contributions in this work can be summarized as follows:

- We recast EEG spatial super-resolution as **dynamic conditional generation**, coupling the LD forward–noising trajectory with the HD reverse–denoising trajectory to balance consistency with the LD observation and fidelity to the HD target.
- We propose a **dynamic residual guidance** paradigm: the path residual estimated from LD inputs serves as a per-step directional correction and is fused additively for incremental updates, yielding a stable, step-aware sampling scheme that remains effective across datasets and a wide range of SR factors.
- We establish a **three-level evaluation protocol** across three datasets, covering signal-level (temporal consistency, spectral fidelity, spatial topology), feature-level (representation quality), and downstream-level (classification accuracy), which provides a comprehensive assessment beyond pointwise error.



(a) Overall Architecture

(b) Illustration of Residual-Guided Diffusion

Figure 2: **SRGDiff overview.** (a) **Overall architecture:** Low-density EEG X^L conditions the latent reverse process. RDM predicts a residual direction from X^L and current decoder features. SMM provides step-aware affine parameters to fuse the residual and modulate activations. (b) **Residual learning:** At each step, the predicted residual guides denoising, and the residual derived from the forward noising process provides supervision via a residual loss.

2 Related Work

Diffusion Models for Missing Data Imputation. Diffusion-based models have emerged as a powerful framework for time-series imputation, leveraging denoising diffusion processes to reconstruct missing values. Diffusion-TS (Yuan & Qiao, 2024) demonstrates interpretable conditional generation across diverse time-series without domain-specific priors. More recently, RDPI (Liu et al., 2025) further enhances precision and efficiency by first generating coarse estimates of missing values through deterministic interpolation, then conditioning a diffusion model on both observed data and these estimates to iteratively refine residual errors. SaSDim (Zhang et al., 2024) introduces self-adaptive noise scaling to preserve spatial dependencies within sensor networks, while SADI (Islam et al., 2025) integrate self-attention mechanisms to handle partial data missing.

EEG Spatial Super-Resolution. Early attempts at EEG spatial super-resolution adapted image-based frameworks to reconstruct dense electrode maps from sparse recordings. EEGSR-GAN (Corley & Huang, 2018) first applied adversarial training to hallucinate missing channels. ESTformer (Li et al., 2025a) then introduced spatiotemporal transformers to model long-range dependencies across electrodes and successfully capture global patterns. More recent diffusion-based and attention-driven approaches have sought to address these limitations. DDPM-EEG (Vetter et al., 2024) leverages denoising diffusion probabilistic models to iteratively refine spatial patterns. STAD (Wang et al., 2025) tackles this by decomposing spatial–temporal interactions into spatial-temporal attention streams. This diffusion-based generative paradigm improves diversity and spectral fidelity compared to GANs, yet their static condition can still lead to distribution drifts and distortion.

Residual Diffusion in Related Domains. Several recent works incorporate residual signals into diffusion models. Ou et al. (2024) synthesize PET from MRI by learning a modality residual under prior information, i.e., a static cross-modal gap that conditions generation. Zhu et al. (2024) reconstruct event-driven video by predicting temporal residuals with inter-frame differences as the generation target to recover dynamics. Mao et al. (2025) address medical segmentation by learning a residual-to-prior that corrects a coarse segmentation, improving calibration and efficiency. These designs either treat the residual as a fixed target/offset or inject it once as a global prior, with weak coupling to the step-by-step reverse dynamics. In contrast, our method targets EEG spatial super-resolution and introduces a dynamic, step-aware residual direction that is re-estimated at every reverse step from the LD observation, and the timestep embedding.

3 Preliminaries and Dynamic Conditional Formulation

Data and Latent Space. Let $X^L \in \mathbb{R}^{C_L \times Length}$ and $X^H \in \mathbb{R}^{C_H \times Length}$ denote low- and high-density EEG with $C_H > C_L$. A pre-trained VAE encoder E maps signals to a latent space $z = E(X^H)$ and a feature extractor F provides LD features as condition $c = F(X^L)$ to condition generation. In practice, we reuse the VAE encoder as the feature extractor to obtain condition c.

Forward Diffusion on the HD Latent. We corrupt the HD latent with a standard Gaussian forward process

$$q(z_t \mid z_{z-1}) = \mathcal{N}\Big(\sqrt{\alpha_t}\, z_{t-1},\, (1-\alpha_t)I\Big). \qquad t = 1, \dots, T \tag{1}$$

Dynamic Conditional Reverse Process. In the reverse generation stage, sampling begins from an isotropic Gaussian noise initialization $\hat{z}_T \sim \mathcal{N}(0, I)$, and the model iteratively predicts $\hat{z}_{t-1}$ from $\hat{z}_t$ until recovering the final latent representation $\hat{z}_0$. Unlike conventional diffusion models that rely solely on a base denoiser, SRGDiff explicitly conditions each reverse step on low-density EEG observations, thereby coupling the LD forward-noising trajectory with the HD reverse-denoising trajectory. The reverse denoiser is defined as:

$$p_\theta(\hat{z}_{t-1} \mid \hat{z}_t, c) = \mathcal{N}\Big(\underbrace{\mu_\theta(\hat{z}_t, c)}_{\text{base denoiser}} + \underbrace{(\gamma_t, \beta_t, r_\phi(c,t))}_{\text{dynamic conditional update}}\ , \beta_t I\Big). \tag{2}$$

Here, the base denoiser $\mu_\theta(\hat{z}_t, c)$ is implemented as a U-Net that estimates the noise component. To further enhance temporal fidelity and spatial coherence, we augment the base denoiser with two lightweight modules that inject step-wise conditional guidance from LD features:

- *Residual Direction Module (RDM).* At each timestep, RDM predicts a path residual $r_\phi(c, t)$ from the LD features and applies it as a directional correction:

$$\hat{z}_{t-1}^{RDM} = \hat{z}_t + r_\phi(c, t). \tag{3}$$

- *Step-Aware Modulation Module (SMM).* SMM calibrates the residual update with timestep-aware affine modulation. Specifically, it predicts a scale γ_t and bias β_t from LD features and the timestep embedding, and applies them to the residual-corrected state:

$$\hat{z}_{t-1}^{SMM} = \gamma_t \odot \hat{z}_{t-1}^{RDM} + \beta_t. \tag{4}$$

Together, the pair $(r_\phi, \gamma_t, \beta_t)$ realizes dynamic conditional generation: at every denoising step, LD features provide both a directional residual and a step-dependent modulation strength, yielding a stable and temporally consistent correction of the reverse diffusion process.

4 Proposed Method

This section presents the step-aware residual-guided diffusion framework, and outlines its architecture and core components. SRGDiff reframes SR as dynamic conditional generation that couples the LD forward-noising trajectory with the HD reverse-denoising trajectory, using an LD-estimated residual as a per-step corrective direction and a step-aware calibration to modulate its strength. The framework comprises four parts: the latent diffusion model backbone, Residual Direction Module (RDM) for additive directional updates, Step-Aware Modulation Module (SMM) for step-dependent modulation, and the overall training strategy. An overview is provided in Figure 2, and the following subsections describe each component in detail.

4.1 Latent Diffusion Model Backbone

Our backbone consists of a VAE that builds the latent space and a denoising U-Net that performs diffusion in that space. The VAE follows the EEG autoencoding setup of Aristimunha et al. (2023).

We train an encoder–decoder (E, D) on HD EEG X^H to obtain $z = E(X^H)$ and $\widehat{X}^H = D(\widehat{z})$, and optimize a reconstruction–regularization objective

$$\mathcal{L}_{\text{VAE}} = \|\widehat{X}^H - X^H\|_2^2 + \lambda_{\text{spec}}\big\|\text{STFT}(\widehat{X}^H) - \text{STFT}(X^H)\big\|_1 + \lambda_{KL}\,\text{KL}\big(q_E(z\,|\,X^H)\,\|\,\mathcal{N}(0, I)\big)\,, \tag{5}$$

where $\text{STFT}(\cdot)$ denotes the short-time Fourier transform applied along the temporal dimension of each EEG channel to encourage spectral fidelity.

Empirically, we set the spectral weight to 0.1 and the KL weight to 10^{-4}. After convergence, (E, D) are frozen. On top of this latent space, we adopt a latent-diffusion model in the style of Rombach et al. (2022).

4.2 Residual Direction Module

In the context of EEG spatial SR, most diffusion approaches condition the U-Net via feature concatenation or cross-attention. We instead turn EEG spatial SR into finer and step-aware conditioning by learning residual direction from low-density recordings. We use the VAE encoder to extract multi-scale condition c and an RDM head R_ϕ takes $(c,\, \tau(t))$ to predict a residual Res_t in the encoder feature space, which acts as a per-step directional correction to the reverse process. Concretely, we first sample a timestep t, encode the HD EEG to obtain the latent $z_0 = E(X^H)$, and draw

$$z_t \;=\; \sqrt{\bar{\alpha}_t}\, z_0 \;+\; \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{6}$$

In the forward process, we obtain at each step the noise-corrupted latent of the HD EEG and use this sequence of step-dependent features as supervision targets for the residual. The residual labels span $t = 0, \ldots, T$ and are defined as $\delta z_t \;:=\; z_0 \;-\; z_t,$.

In the reverse process, we estimate δz_t from the low-density EEG to supply the step-wise temporal and spatial details required for denoising. We introduce a lightweight convolutional predictor R_ϕ that takes the timestep embedding $\tau(t)$ and LD features $c = F(X^L)$ as input and outputs the residual feature Res_t, and trained by

$$Res_t \;=\; R_\phi\big(\tau(t),\, c\big), \quad \mathcal{L}_{\text{res}} \;=\; \sum_{t=0}^{T} \big\|\, Res_t - \delta z_t \big\|_2^2. \tag{7}$$

Finally, the predicted residual feature is then added to $\hat{z}_t$ as an incremental update:

$$\hat{z}_t^{RDM} \;=\; LayerNorm(\hat{z}_t) \;+\; Res_t. \tag{8}$$

4.3 Step-Aware Modulation Module

After obtaining $\hat{z}_t^{\text{RDM}}$, we further modulate the current step to control the extent to which the residual condition influences denoising. To enforce temporal fidelity, SMM explicitly weighted the current diffusion timestep with a step-dependent affine modulation estimated from the low-density features and the current timestep embedding.

Specifically, SMM first encodes the low-density EEG through a lightweight 1D convolutional network E_{SMM} to produce a feature map h_t. To enable the conditioning to recognize the current diffusion step, SMM maps each sampled timestep t into a sinusoidal time embedding e_t. A learnable weight σ_t that decays linearly with t balances these two streams, yielding a fused feature:

$$\widetilde{h}_t = \sigma_t h_t + (1 - \sigma_t) e_t = \sigma_t E_{SMM}(c) + (1 - \sigma_t) e_t. \tag{9}$$

For spatial coherence, we adopt an affine calibration mechanism. The fused feature $\widetilde{h}_t$ is passed through two MLPs MLP_γ and MLP_β to predict channel-wise scale γ_t^c and bias β_t^c:

$$\hat{z}_t^{SMM} = \gamma_t \,\odot\, \hat{z}_t^{RDM} \;+\; \beta_t^c = MLP_\gamma(\widetilde{h}_t, t) \,\odot\, \hat{z}_t^{RDM} \;+\; MLP_\beta(\widetilde{h}_t, t). \tag{10}$$

Finally, SRGDiff feeds the updated latent $\hat{z}_t$ into the U-Net decoder to obtain the next denoised state $\hat{z}_{t-1}$.

4.4 Training Strategy

To stabilize optimization and decouple latent representation learning from conditional diffusion modeling, we adopt a two-stage training strategy.

Stage 1: VAE Pre-training. We first train the VAE encoder-decoder pair on high-density EEG data to obtain a stable and structured latent space. After convergence, the VAE parameters are frozen to provide fixed latent representations for subsequent diffusion modeling.

Stage 2: Residual-Guided Latent Diffusion. On the frozen latent space, the final training objective is a weighted combination of the three terms:

$$\mathcal{L}_{\text{Stage 2}} = \mathbb{E}_{z_0,\epsilon,t}\big[\|\epsilon-\epsilon_\theta(z_t,t,c)\|_2^2\big]+\lambda_{res}\sum_{t=1}^{T}\|R_\varphi(c,t)-(z_0-z_t)\|_2^2+\lambda_{SMM}(\|\gamma_t-1\|_2^2+\|\beta_t\|_2^2). \tag{11}$$

Empirically, we set the residual weight to 1 and the SMM weight to 10^{-2}. The term $\lambda_{\text{SMM}}\big(\|\gamma_t - 1\|_2^2 + \|\beta_t\|_2^2\big)$ serves as a regularization component to prevent excessively large values of γ_t and β_t, thereby stabilizing the training dynamics.

5 Experiments

5.1 Downstream Datasets

In this study, we employ three publicly available EEG datasets. The **SEED dataset** (Zheng & Lu, 2015) uses 15 film clips of approximately four minutes each as emotional stimuli to induce stable and continuous emotional responses in three categories: positive, neutral, and negative. Data were acquired via 62 channels at 1000 Hz, downsampled to 200 Hz, band-pass filtered (0–75 Hz) and segments with faulty sensors removed. **SEED-IV** (Zheng et al., 2018) extends SEED by using the same 15 subjects and 62-channel setup (1000 Hz) but adds music and image stimuli to evoke happiness, sadness, fear and neutral states; preprocessing mirrors that of SEED. **Localize-MI** (Mikulan et al., 2020) contains 61 presurgical sessions from seven drug-resistant epilepsy patients, where 256-channel scalp EEG was recorded at 8000 Hz during 0.1–5 mA intracerebral single-pulse stimulation; preprocessing includes a 0.1 Hz high-pass filter, notch filter, bad-channel/trial removal and alignment of trials to the -300 ms to +50 ms stimulus-artifact window.

5.2 Experiment Setup

Data Preprocessing. The experimental setup for the EEG super-resolution task follows the ESTformer and STAD frameworks. The preprocessed EEG signals were segmented into fixed-length windows: continuous, non-overlapping 4-second windows for SEED and SEED-IV datasets, and 260 ms windows (from 250 ms before stimulation to 10 ms after) for Localize-MI. In SEED and SEED-IV, we designed different super-resolution scale factors ($2\times$, $4\times$ and $8\times$) to evaluate reconstruction performance. The selection of visible channels and the super-resolution scaling factors follow the configurations used in ESTformer. For Localize-MI, due to the high channel density, we applied more extensive scale factors ($2\times$, $4\times$, $8\times$, $16\times$).

Training & Environment Settings. For each dataset, we split the data into train/test with an $80\%/20\%$ ratio and reserve 10% of the training portion as a validation set. Stage I is trained only on the HD signals from the training split. Stage II is trained on paired LD/HD samples constructed from the same training split by masking HD channels according to the target SR scale; the validation set is used for early stopping and hyperparameter selection. The held-out test split is used once for final reporting, with no fine-tuning.

Baselines. We compare SRGDiff with strong EEG SR and time-series imputation baselines: **ESTformer** (Li et al., 2025a) and **STAD** (Wang et al., 2025) (transformer-/diffusion-based EEG SR), **DDPMEEG** (Vetter et al., 2024) (diffusion for ECoG SR), **SaSDim** (Zhang et al., 2024) and **SADI** (Islam et al., 2025) (advanced missing data imputation), and the two-stage residual method **RDPI** (Liu et al., 2025). We use authors' official implementations when available and otherwise provide

Model	Ref	Metric	2	4	8	16
SaSDim	IJCAI 2024	NMSE	0.2675±0.003	0.3427±0.001	0.4174±0.004	0.4613±0.003
		PCC	0.8194±0.002	0.7246±0.007	0.6926±0.003	0.6476±0.002
		SNR	5.7443±0.007	4.3796±0.003	3.5549±0.009	2.7678±0.005
SADI	AAAI 2025	NMSE	0.2637±0.003	0.3442±0.001	0.4164±0.004	0.4566±0.003
		PCC	0.8243±0.002	0.7391±0.007	0.6944±0.003	0.6554±0.002
		SNR	5.7511±0.007	4.3724±0.003	3.5498±0.008	2.8942±0.009
RDPI	AAAI 2025	NMSE	0.2561±0.003	0.3562±0.001	0.4076±0.004	0.4531±0.003
		PCC	0.8246±0.002	0.7396±0.007	0.7062±0.003	0.6549±0.002
		SNR	5.7311±0.007	4.3966±0.003	3.5643±0.009	2.7731±0.007
DDPMEEG	Patterns 2024	NMSE	0.2046±0.003	0.3108±0.001	0.3554±0.004	0.4076±0.002
		PCC	0.8516±0.002	0.8163±0.007	0.7306±0.003	0.6739±0.002
		SNR	6.2151±0.008	5.5126±0.003	3.9891±0.009	3.2715±0.005
ESTformer	KBS 2025	NMSE	0.2721±0.003	0.3578±0.001	0.4466±0.004	0.4837±0.002
		PCC	0.8061±0.002	0.7205±0.007	0.6867±0.003	0.6319±0.002
		SNR	5.5403±0.008	3.8671±0.003	3.3023±0.007	2.5671±0.004
STAD	TCE 2025	NMSE	0.1902±0.003	0.3067±0.001	0.3649±0.004	0.4106±0.003
		PCC	0.8635±0.002	0.8194±0.007	0.7216±0.003	0.6694±0.002
		SNR	7.2591±0.008	5.5234±0.003	3.8715±0.009	3.2642±0.005
SRGDiff	OURS	NMSE	**0.1449**±0.003	**0.2384**±0.001	**0.2957**±0.004	**0.3457**±0.002
		PCC	**0.9213**±0.002	**0.8854**±0.007	**0.8323**±0.003	**0.7322**±0.002
		SNR	**8.3755**±0.008	**6.3617**±0.003	**5.2249**±0.009	**4.0197**±0.006

Table 1: Performance of all methods on Localize-MI across different channel settings.

carefully verified reimplementations, applying their recommended hyperparameters and unifying training epochs and sampling steps across methods.

5.3 Evaluation Protocol.

We assess SR quality at three complementary levels to balance faithfulness to the ground truth, preservation of neurophysiological structure, and practical utility.

Signal level (does the waveform match?): We follow **ESTformer** and report normalized mean squared error (NMSE), Pearson correlation coefficient (PCC), and reconstruction signal-to-noise ratio (SNR) with respect to the HD reference, plus topology maps for qualitative inspection (formal definitions in the Appendix).

Feature level (does the representation distribution match?): We adopt **EEG-FID** following Lai et al. (2025), using a frozen EEGNet trained per dataset on its training split; the embedding dimension is 256 for SEED/SEED-IV and 512 for Localize-MI. In addition, we report a **frequency-domain MAE**: we first compute channel-wise STFTs of the reconstructed and reference HD EEG, form their power spectra, and then compute the normalized mean squared error between these spectra averaged over channels and frequency bins, so as to capture spectral distortions that are not reflected by time-domain NMSE alone.

Downstream level (is it useful?): We evaluate SEED/SEED-IV subject-dependent emotion recognition without cross-validation and binary epileptic classification on Localize-MI, both reporting accuracy. All results are summarized as mean±std over subjects; implementation details and metric formulas are provided in the Appendix.

5.4 Main Results

We report signal-level reconstruction quality in Tables 1 and 2. For the most demanding 16× upsampling on Localize-MI, SRGDiff attains an NMSE of 0.3457, over 15% lower than DDPMEEG's 0.4076, indicating that dynamic conditioning effectively guides the diffusion model to generate

Model	Metric	SEED (62)			SEED-IV (62)		
		2	4	8	2	4	8
SaSDim	NMSE	0.4399±0.004	0.6234±0.002	0.7767±0.007	0.3633±0.004	0.5543±0.002	0.7122±0.005
	PCC	0.7341±0.002	0.5649±0.001	0.4349±0.004	0.7249±0.002	0.6211±0.009	0.5009±0.003
	SNR	4.1154±0.096	2.2940±0.046	1.1349±0.127	4.5940±0.009	2.6004±0.004	1.6211±0.111
SADI	NMSE	0.4439±0.004	0.6049±0.002	0.8106±0.007	0.3557±0.004	0.5349±0.002	0.6844±0.005
	PCC	0.7234±0.002	0.5819±0.001	0.4064±0.004	0.7624±0.002	0.6293±0.009	0.5243±0.003
	SNR	4.2419±0.097	2.5160±0.046	1.0137±0.127	4.7093±0.009	2.6044±0.004	1.6610±0.112
RDPI	NMSE	0.4064±0.004	0.6134±0.002	0.7916±0.007	0.3491±0.004	0.5416±0.002	0.6915±0.005
	PCC	0.7416±0.002	0.5716±0.001	0.4216±0.004	0.7861±0.002	0.6316±0.009	0.5164±0.003
	SNR	4.2619±0.097	2.3160±0.046	1.0316±0.127	4.7190±0.009	2.6194±0.004	1.6492±0.115
DDPMEEG	NMSE	0.4916±0.004	0.7319±0.002	0.8634±0.007	0.5136±0.004	0.6513±0.001	0.7916±0.005
	PCC	0.6941±0.002	0.5134±0.001	0.3419±0.004	0.7346±0.001	0.5316±0.009	0.4305±0.003
	SNR	4.1943±0.095	1.5391±0.042	0.9431±0.125	4.4165±0.008	2.1064±0.003	1.6105±0.110
ESTformer	NMSE	0.3288±0.004	0.3483±0.002	0.4149±0.007	0.3448±0.004	0.3911±0.0015	0.5125±0.005
	PCC	0.8368±0.002	0.8012±0.001	0.7670±0.004	0.8106±0.002	0.7822±0.009	0.7048±0.003
	SNR	5.0560±0.097	4.5838±0.044	3.8871±0.126	4.7535±0.008	4.1933±0.003	2.9821±0.113
STAD	NMSE	0.4319±0.004	0.6913±0.002	0.8671±0.007	0.3819±0.004	0.6713±0.002	0.7193±0.005
	PCC	0.7136±0.002	0.4946±0.001	0.3441±0.004	0.7316±0.002	0.5219±0.009	0.4319±0.003
	SNR	4.1364±0.099	1.4349±0.043	0.9134±0.125	4.4930±0.008	2.0492±0.003	1.6193±0.114
SRGDiff	NMSE	**0.1632**±0.004	**0.2977**±0.002	**0.3494**±0.007	**0.1663**±0.004	**0.2115**±0.002	**0.2603**±0.005
	PCC	**0.9102**±0.002	**0.8445**±0.001	**0.8167**±0.004	**0.9113**±0.002	**0.8846**±0.009	**0.8210**±0.003
	SNR	**7.8413**±0.097	**5.2606**±0.043	**4.5912**±0.127	**7.8660**±0.008	**6.6402**±0.003	**6.0346**±0.120

Table 2: Performance comparison of different models on SEED and SEED-IV datasets across different channel settings.

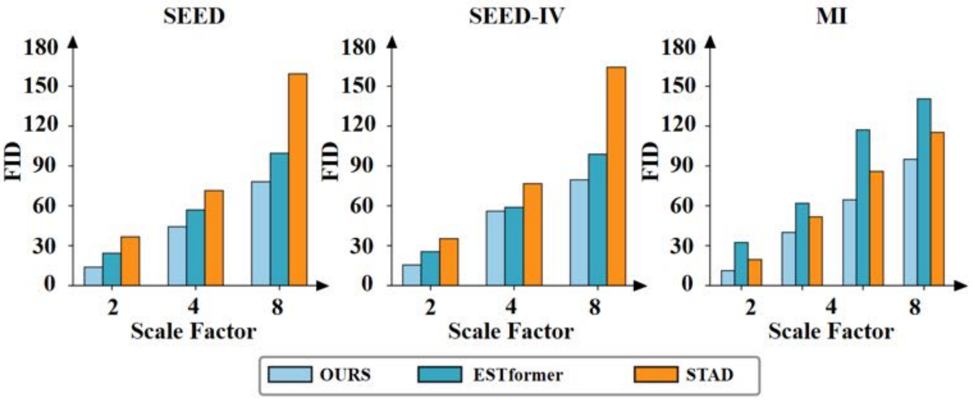
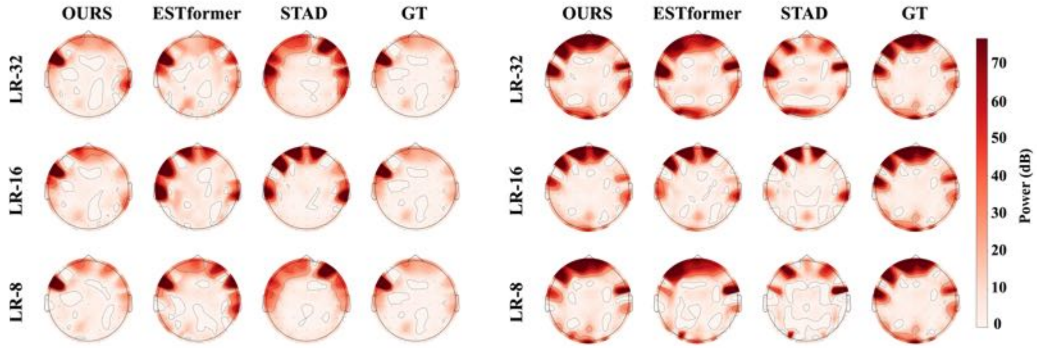


Figure 3: EEG-FID evaluation across three datasets compared with ESTformer and STAD.

super-resolved signals that closely approximate the true high-density data. Its PCC improves from 0.6739 to 0.7322, and its SNR increases from 3.27 dB to 4.02 dB (over 22%), demonstrating that under challenging settings the dynamic conditioning still learns the HD trend while maintaining a favorable signal-to-noise ratio. On SEED with high temporal variability and frequent outliers, SRGDiff reduces the $2\times$ NMSE from ESTformer's 0.3288 to 0.1632 (a reduction of more than 50%) and raises PCC from 0.8368 to 0.9102. A similar pattern is observed on SEED-IV, where NMSE drops to 0.1663 versus 0.3448 and PCC increases to 0.9113 versus 0.8106, indicating that despite lower SNR, the dynamic conditioning exhibits strong generalization.

5.5 Feature-level Reconstruction Evaluation

We report EEG-FID results in Figure 3, and our method consistently achieves the lowest FID scores across SEED, SEED-IV, and Localize-MI datasets under different scale factors. These results indi-



(a) SEED

(b) SEED-IV

Figure 4: Visualization of EEG topographic maps between ground-truth and reconstructed EEG signals by ESTformer, STAD and SRGDiff.

Model	SEED			SEED-IV			Localize-MI			
	2×	4×	8×	2×	4×	8×	2×	4×	8×	16×
ESTformer	6.96	9.31	9.73	7.11	8.30	8.86	7.03	16.67	32.32	35.73
STAD	9.19	11.04	14.40	9.50	10.95	13.12	8.76	13.11	22.53	25.37
SRGDiff	**3.89**	**5.12**	**4.95**	**3.99**	**4.08**	**4.84**	**3.86**	**7.30**	**11.50**	**13.76**

Table 3: Frequency-domain MAE between reconstructed and real HD topomaps on SEED, SEED-IV, and Localize-MI under different SR factors.

cate that our approach generates EEG signals that are statistically closer to the real distribution in the temporal domain.

We further analyze the spectral fidelity of generated signals by visualizing EEG topographic maps under different scale factors. An EEG topographic map projects the power spectral density (PSD) of each channel onto the scalp surface, providing an intuitive representation of the spatial distribution of oscillatory energy. As shown in Figure 4 and Figure 15, although our reconstructed signals still exhibit minor deviations from the original data, they preserve a high degree of overlap in critical regions with strong PSD responses.

Beyond qualitative inspection, we also report a frequency-domain error metric that quantifies the mean absolute error between reconstructed and real HD topomaps. As shown in Table 3, SRGDiff consistently achieves the lowest frequency-domain MAE across datasets and SR scales, indicating better preservation of the spatial distribution of spectral power.

5.6 Downstream tasks

Table 4 reports results on the three datasets under various super-resolution scales. As the scale grows, accuracy for both raw and super-resolved inputs declines, yet SRGDiff's reconstructions consistently maintain a clear advantage. In particular, SRGDiff greatly outperforms missing-value imputation methods and leads all other spatial imputation approaches. At 2× scale factor, its classification accuracy approaches that obtained from the original full-channel recordings. We further compared the runtime efficiency of different methods as shown in Table 13. Although our proposed SRGDiff is slightly slower than the transformer-based ESTformer, it can still complete EEG super-resolution within 0.1s, which meets the real-time requirement in practical applications. The detailed runtime statistics of all models are provided in the Appendix.

5.7 Ablation Study

To evaluate the contribution of each module in SRGDiff to EEG super-resolution reconstruction, we conducted ablation studies comparing SRGDiff with three variant models. **LDM+LD** keeps only

Method	SEED			SEED-IV			Localize-MI			
	2	4	8	2	4	8	2	4	8	16
GT	0.7152	0.7152	0.7152	0.7027	0.7027	0.7027	0.8368	0.8368	0.8368	0.8368
LR	0.4981	0.4702	0.4424	0.5685	0.5618	0.5011	0.7208	0.6534	0.5219	0.3862
SaSDim	0.5097	0.4793	0.4429	0.5794	0.5692	0.4912	0.7193	0.6519	0.5237	0.3845
SADI	0.5137	0.4834	0.4456	0.5718	0.5644	0.4987	0.7215	0.6634	0.5314	0.3957
RDPI	0.5044	0.4802	0.4531	0.5591	0.5741	0.5071	0.7230	0.6624	0.5210	0.3892
DDPMEEG	0.4738	0.4610	0.4238	0.5548	0.5487	0.4838	0.7015	0.6387	0.5187	0.3767
ESTformer	0.6887	0.6509	0.6057	0.6782	0.6500	0.5084	0.7445	0.6033	0.4739	0.4391
STAD	0.5437	0.5249	0.4610	0.6651	0.6410	0.4977	0.7589	0.6797	0.6344	0.5384
SRGDiff	**0.7019**	**0.6812**	**0.6273**	**0.6821**	**0.6558**	**0.5127**	**0.7641**	**0.7163**	**0.6806**	**0.5887**

Table 4: Classification accuracy comparison across different methods and datasets. GT represents the ground truth performance. Best results are shown in bold, second-best are underlined.

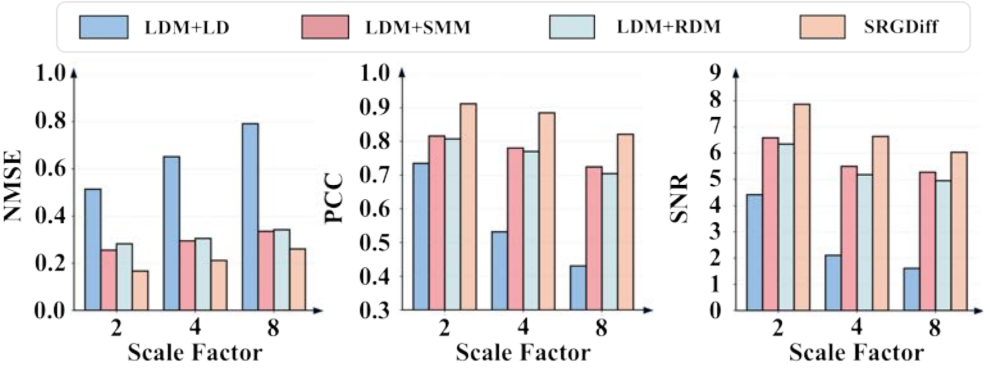


Figure 5: Ablation study performance comparison between SRGDiff and three variant models on the SEED dataset.

the VAE–DDIM backbone and takes the LD EEG as its input condition; **LDM+SMM** preserves the Step-aware modulation module; **LDM+RDM** retains the Residual Direction Module. All models were tested under the same experimental settings.

Figure 5 reports NMSE, PCC and SNR across $2\times$, $4\times$ and $8\times$ upsampling in SEED dataset. At $8\times$, adding SMM to the baseline cuts NMSE from 0.86 to 0.45 with 47% reduction and boosts PCC from 0.34 to 0.69, demonstrating its effectiveness in temporally aligning the denoising trajectory. Incorporating RDM yields a comparable NMSE reduction with 44% and raises PCC to 0.67, highlighting its role in injecting prior information for spatial consistency. When combined in SRGDiff, these modules further decrease NMSE to 0.34 with 60% overall reduction and elevate PCC to 0.81. More ablation results in SEED-IV and Localize-MI datasets are shown in Figure 11 in the Appendix.

6 Conclusion

We introduced SRGDiff, a step-aware residual-guided diffusion model that reframes EEG spatial super-resolution as guided HD generation with a step-aware residual direction and adaptive modulation. Across SEED, SEED-IV, and Localize-MI, SRGDiff consistently improves signal-level metrics, achieves the best EEG-FID across scales, and better preserves spectral and scalp topographies. Downstream evaluations further show higher accuracy on emotion recognition and patient-wise classification, indicating that the reconstructed signals are not only visually and statistically closer to HD EEG but also more useful for analysis. These results validate that explicit, step-wise conditioning on sparse inputs is both necessary and effective for high-fidelity, topology-preserving EEG super-resolution.

7 Acknowledgements

This work was supported by the National Key Research and Development Program of China (Grant No. 2022ZD0118001), the National Natural Science Foundation of China (Grant Nos. U22A2022, 62332017), and was conducted in part at the MOE Key Laboratory of Advanced Materials and Devices for Post-Moore Chips, the Beijing Key Laboratory of Big Data Innovation and Application for Skeletal Health Medical Care, and the Beijing Advanced Innovation Center for Materials Genome Engineering.

8 Ethics Statement

This work adheres to the ICLR Code of Ethics. In this study, no human subjects or animal experimentation was involved. All datasets used, including SEED, SEED-IV and Localize-MI, were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. Details could be found in Appendix. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

9 Reproducibility Statement

We have made every effort to ensure that the results presented in this paper are reproducible. All code and datasets have been made publicly available in an anonymous repository to facilitate replication and verification. The experimental setup, including training steps, model configurations, and hardware details, is described in detail in the paper. We have also provided a full description of SRGDiff, to assist others in reproducing our experiments.

We believe these measures will enable other researchers to reproduce our work and further advance the field.

References

Bruno Aristimunha, Raphael Yokoingawa de Camargo, Sylvain Chevallier, Adam G Thomas, Oeslle Lucena, Jorge Cardoso, Walter Hugo Lopez Pinaya, and Jessica Dafflon. Synthetic sleep eeg signal generation using latent diffusion models. In *DGM4H 2023-1st Workshop on Deep Generative Models for Health at NeurIPS 2023*, 2023.

Isaac A. Corley and Yufei Huang. Deep eeg super-resolution: Upsampling eeg spatial resolution with generative adversarial networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 100–103, 2018. doi: 10.1109/BHI.2018.8333379.

Yu-Hao Huang, Chang Xu, Yueying Wu, Wu-Jun Li, and Jiang Bian. Timedp: Learning to generate multi-domain time series with domain prompts. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence*, 2025.

Mohammad Rafid Ul Islam, Prasad Tadepalli, and Alan Fern. Self-attention-based diffusion model for time-series imputation in partial blackout scenarios. In *AAAI*, volume 39, pp. 17564–17572, 2025. URL https://doi.org/10.1609/aaai.v39i17.33931.

Weibang Jiang, Yansen Wang, Bao-liang Lu, and Dongsheng Li. Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals. In *The Thirteenth International Conference on Learning Representations*, 2025.

Yongfan Lai, Jiabo Chen, Qinghao Zhao, Deyun Zhang, Yue Wang, Shijia Geng, Hongyan Li, and Shenda Hong. Diffusets: 12-lead ecg generation conditioned on clinical text reports and patient-specific information. *Patterns*, 2025.

Dongdong Li, Zhongliang Zeng, Zhe Wang, and Hai Yang. Estformer: Transformer utilising spatiotemporal dependencies for electroencephalogram super-resolution. *Knowledge-Based Systems*, 317:113345, 2025a.

Yang Li, Han Meng, Zhenyu Bi, Ingolv T Urnes, and Haipeng Chen. Population aware diffusion for time series generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 18520–18529, 2025b.

Zijin Liu, Xiang Zhao, and You Song. Rdpi: A refine diffusion probability generation method for spatiotemporal data imputation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12255–12263, 2025.

Fuyou Mao, Beining Wu, Yanfeng Jiang, Han Xue, Yan Tang, and Hao Zhang. Prior-guided residual diffusion: Calibrated and efficient medical image segmentation. *arXiv preprint arXiv:2509.01330*, 2025.

Ezequiel Mikulan, Simone Russo, Sara Parmigiani, Simone Sarasso, Flavia Maria Zauli, Annalisa Rubino, Pietro Avanzini, Anna Cattani, Alberto Sorrentino, Steve Gibbs, et al. Simultaneous human intracerebral stimulation and hd-eeg, ground-truth for source localization methods. *Scientific data*, 7(1):127, 2020.

Zaixin Ou, Caiwen Jiang, Yongsheng Pan, Yuanwang Zhang, Zhiming Cui, and Dinggang Shen. A prior-information-guided residual diffusion model for multi-modal pet synthesis from mri. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 4769–4777, 2024.

Steven M Peterson, Satpreet H Singh, Benjamin Dichter, Michael Scheid, Rajesh PN Rao, and Bingni W Brunton. Ajile12: Long-term naturalistic human intracranial neural recordings and pose. *Scientific data*, 9(1):184, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Yunbo Tang, Dan Chen, Honghai Liu, Chang Cai, and Xiaoli Li. Deep eeg superresolution via correlating brain structural and functional connectivities. *IEEE Transactions on Cybernetics*, 53 (7):4410–4422, 2022.

Julius Vetter, Jakob H Macke, and Richard Gao. Generating realistic neurophysiological time series with denoising diffusion probabilistic models. *Patterns*, 5(9), 2024.

Li Wang, Jungwook Go, and Xiang Chen. An eeg-based emotion recognition model using an interaction design framework and deep learning. *Journal of Mechanics in Medicine and Biology*, 24 (02):2440022, 2024.

Shuqiang Wang, Tong Zhou, Yanyan Shen, Ye Li, Guoheng Huang, and Yong Hu. Generative ai enables eeg super-resolution via spatio-temporal adaptive diffusion learning. *IEEE Transactions on Consumer Electronics*, 71(1):1034–1045, 2025. doi: 10.1109/TCE.2025.3528438.

Xinyu Yuan and Yan Qiao. Diffusion-TS: Interpretable diffusion for general time series generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL `https://openreview.net/forum?id=4h1apFjO99`.

Shunyang Zhang, Senzhang Wang, Xianzhen Tan, Renzhi Wang, Ruochen Liu, Jian Zhang, and Jianxin Wang. Sasdim:self-adaptive noise scaling diffusion model for spatial time series imputation. In Kate Larson (ed.), *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 2561–2569. International Joint Conferences on Artificial Intelligence Organization, 8 2024. doi: 10.24963/ijcai.2024/283. URL `https://doi.org/10.24963/ijcai.2024/283`. Main Track.

W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, pp. 1–13, 2018. ISSN 2168-2267. doi: 10.1109/TCYB.2018.2797176.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development*, 7(3):162–175, 2015. doi: 10.1109/TAMD.2015.2431497.

Lin Zhu, Yunlong Zheng, Yijun Zhang, Xiao Wang, Lizhi Wang, and Hua Huang. Temporal residual guided diffusion framework for event-driven video reconstruction. In *European Conference on Computer Vision*, pp. 411–427. Springer, 2024.

A LLM Usage

Large Language Models (LLMs) were used to aid in the writing and polishing of the manuscript. Specifically, we used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text.

It is important to note that the LLM was not involved in the ideation, research methodology, or experimental design. All research concepts, ideas, and analyses were developed and conducted by the authors. The contributions of the LLM were solely focused on improving the linguistic quality of the paper, with no involvement in the scientific content or data analysis.

The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

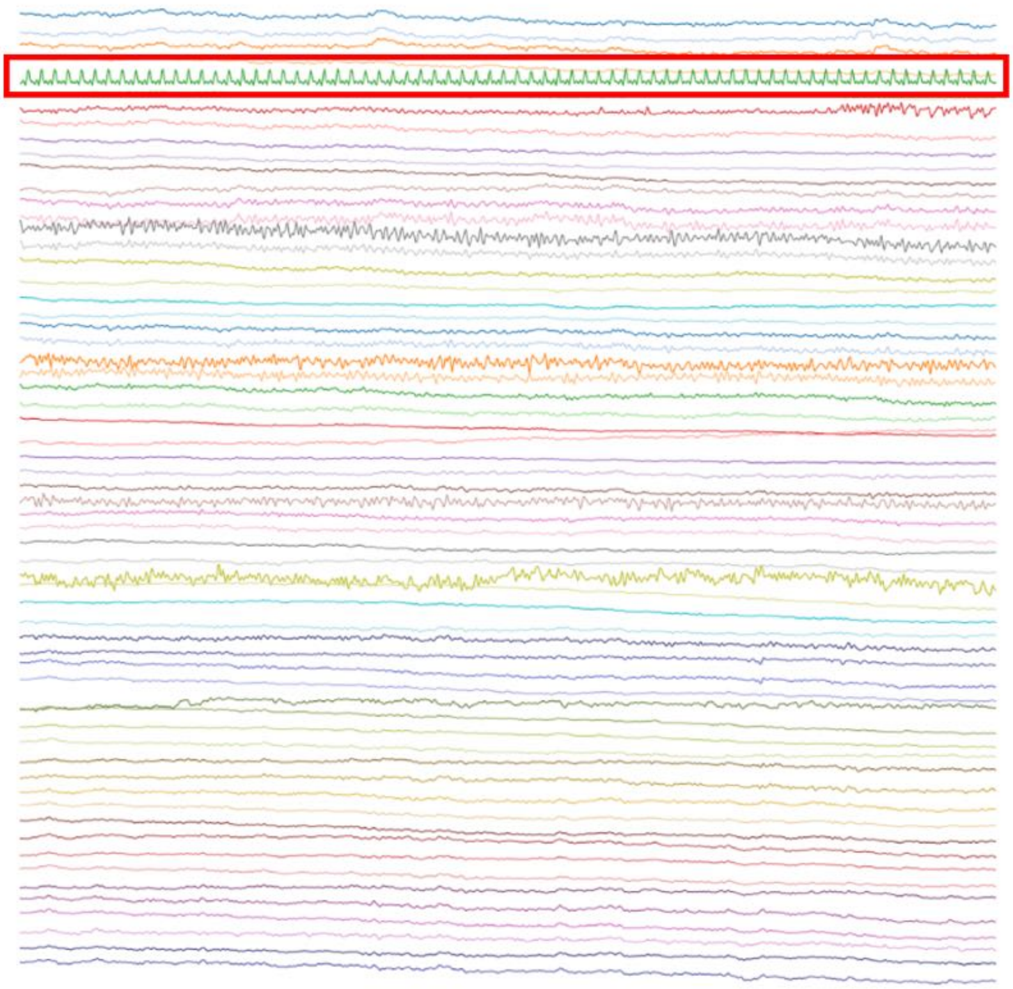


Figure 6: Example of a raw SEED EEG segment with sensor faults highlighted.

B Dataset Preprocess Details

B.1 Dataset Details

We use three publicly available EEG datasets: SEED, SEED-IV, and Localize-MI.

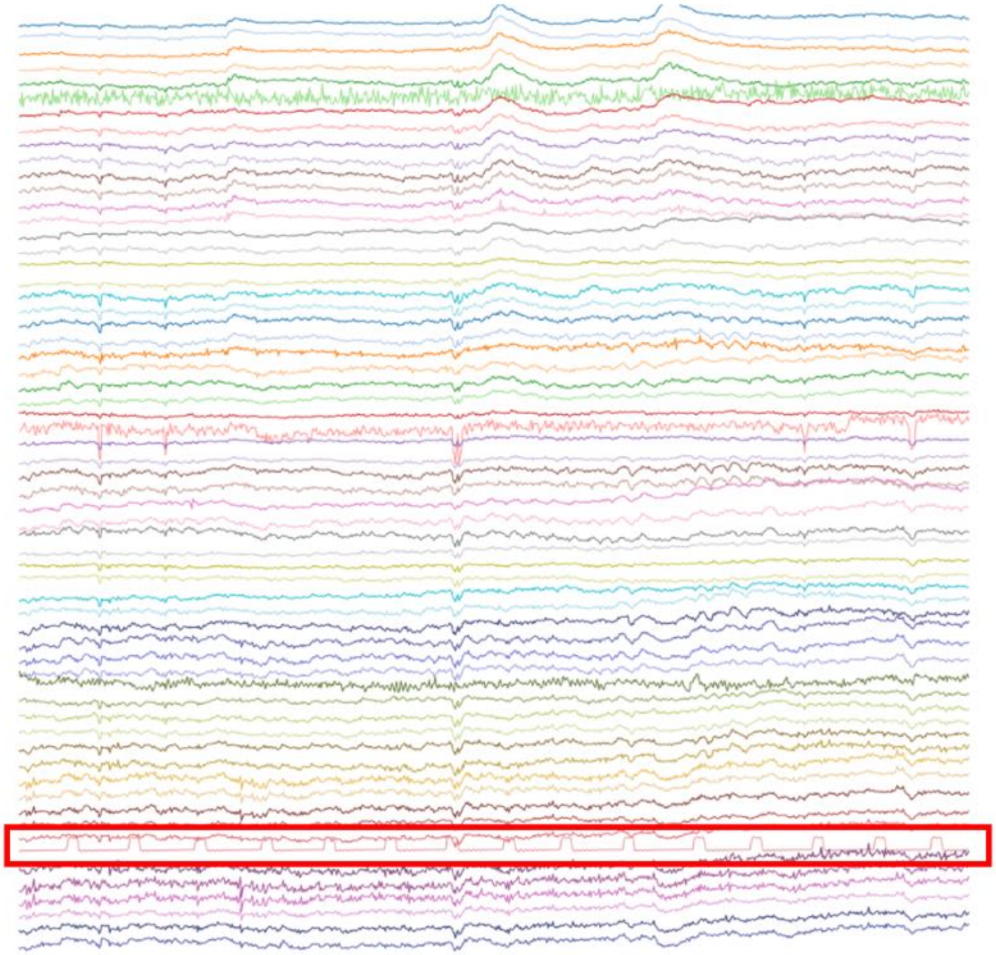


Figure 7: Example of a raw SEED EEG segment with sensor faults highlighted.

SEED (available at `http://bcmi.sjtu.edu.cn/~seed/`) consists of recordings from 15 subjects watching emotion-eliciting film clips (≈ 4 min each) designed to induce positive, neutral, and negative states. Data were acquired with a 62-channel 10–20 montage at 1000 Hz, downsampled to 200 Hz, and band-pass filtered to 0.5–75 Hz. We manually inspected and removed sessions with sensor faults as depicted in Figure 6 and Figure 7.

SEED-IV (available at `http://bcmi.sjtu.edu.cn/~seed/seed-iv.html`) is a publicly available EEG dataset designed for emotion recognition research. It includes four emotional categories: happiness, sadness, neutrality, and fear. Emotional states are elicited using two types of stimuli: music and images. EEG recordings were collected from 15 subjects using a 62-channel system with a sampling rate of 1000 Hz.During preprocessing, the EEG signals were downsampled to 200 Hz and filtered with a bandpass filter ranging from 0.5 to 75 Hz. To ensure data quality, visually corrupted or invalid trials were manually excluded.

Localize-MI (available at `https://doi.org/10.12751/g-node.1cc1ae`) is a high-density intracranial EEG dataset from seven drug-resistant epilepsy patients during 61 presurgical sessions. Stereo–EEG electrodes delivered single-pulse biphasic currents (0.1–5 mA), and 256 channels were recorded at 8000 Hz. Preprocessing included 0.1 Hz high-pass filtering, notch filters at 50/100/150/200 Hz, bad-channel/trial removal, and trial alignment using stimulation artifact peaks (–300 to +50 ms window). In the Localize-MI dataset, we designed a binary classification task (epileptic vs. nonepileptic) to evaluate the effectiveness of synthetic super-resolution EEG (SR

EEG) in detecting epileptic abnormalities. Specifically, EEG signals recorded before electrical stimulation are labeled as nonepileptic, while those recorded during stimulation are labeled as epileptic. The detailed experimental setup follows the description provided in the STAD (Wang et al., 2025) model section.

B.2 More Experimental Details

We follow ESTformer and STAD slicing strategies. Preprocessed signals are windowed into fixed lengths: SEED and SEED-IV use non-overlapping 4 s segments, while Localize-MI retains –250 ms to $+10$ ms around each stimulus (260 ms total). We randomly split 80% for training and 20% for testing, yielding $24265 \times 62 \times 800$ train / $6067 \times 62 \times 800$ test samples for SEED; $29199 \times 62 \times 800$ train / $7300 \times 62 \times 800$ test for SEED-IV; and $1914 \times 256 \times 2081$ train / $479 \times 256 \times 2081$ test for Localize-MI. For SEED and SEED-IV we evaluate $2\times$, $4\times$, and $8\times$ super-resolution; for Localize-MI we additionally include $16\times$. As shown in Figure 8, Localize-MI employs a 256-channel intracranial grid, while Figure 9 shows the 62-channel scalp montage used in SEED-IV and SEED.

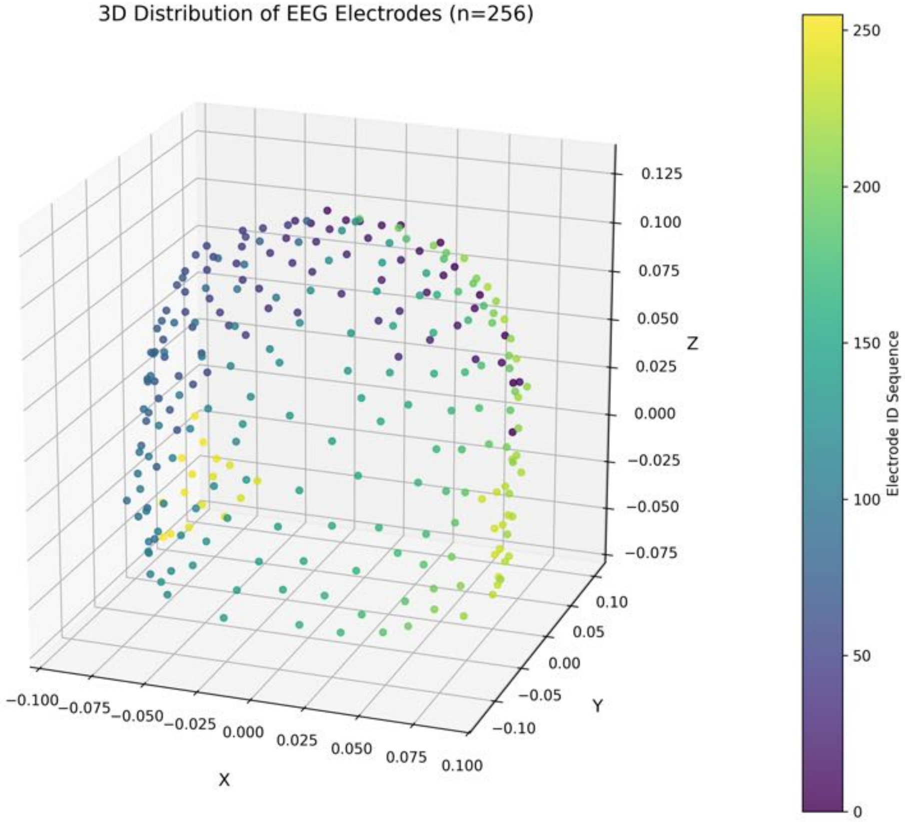


Figure 8: Electrode topology of the Localize-MI dataset (256 intracranial channels).

C More SRGDiff Model Details

C.1 Variational Autoencoder

In this paper, Variational Autoencoder (VAE) follows the AutoencoderKL design (Aristimunha et al., 2023), comprising a convolutional encoder, a latent distribution (mean and variance) with KL regularization toward $\mathcal{N}(0, I)$, and a decoder with deconvolutions and upsampling. We augment both encoder and decoder with attention layers (multi-head and non-local attention), residual connec-

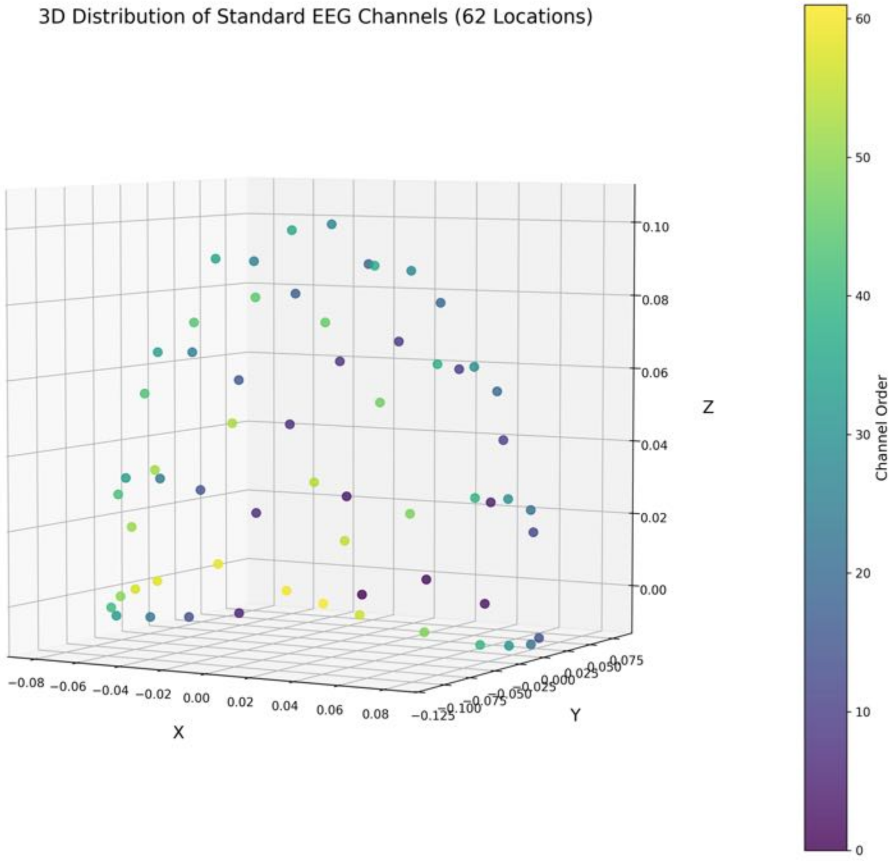


Figure 9: Electrode topology of the SEED-IV and SEED datasets (62 scalp channels).

tions, and GroupNorm to capture global EEG features while ensuring stable training and efficient latent representations.

C.2 DDIM Scheduler

The DDIMScheduler manages noise scheduling and sampling in the forward and reverse diffusion processes. It supports multiple noise prediction types and variance strategies. At each step, it computes the noise coefficient, predicts the denoised sample, clips values for numerical stability, and injects random perturbations to control output diversity.

D Parameter Study

D.1 VAE Latent Shape Selection

We found that the latent shape balances reconstruction precision and generalization. Higher dimensions capture more detail but risk overfitting, while lower dimensions blur outputs. We experimented across the three datasets and selected a latent of 32×400 for SEED/SEED-IV and 64×500 for Localize-MI, which yielded optimal NMSE, PCC, and SNR. Table 5 and Table 6 shows the performance of SRGDiff in different latent shapes.

Shape	**NMSE**	**PCC**	**SNR (dB)**
64 × 400	0.15	0.93	7.75
32×400	**0.12**	**0.95**	**8.72**
16 × 200	0.20	0.89	6.81
8 × 400	0.16	0.92	7.26

SEED

Shape	**NMSE**	**PCC**	**SNR (dB)**
64 × 400	0.16	0.92	7.58
32×400	**0.13**	**0.94**	**8.58**
16 × 200	0.21	0.87	6.86
8 × 400	0.19	0.91	7.23

SEED-IV

Table 5: VAE latent shape selection results on SEED and SEED-IV.

Shape	**NMSE**	**PCC**	**SNR (dB)**
128 × 1000	0.13	0.94	8.60
64×500	**0.09**	**0.96**	**9.01**
32 × 500	0.15	0.92	7.61
32 × 1000	0.13	0.94	8.62

Table 6: Localize-MI: VAE latent shape selection results

D.2 Diffusion Hyperparameters

D.2.1 Diffusion Schedules

We compare linear and cosine noise schedules. Linear adds noise at a constant rate but may cause instability at endpoints; cosine offers smoother transitions and better performance for long diffusion chains. We fixed 1000 timesteps with cosine scheduling and evaluated NMSE, PCC, and SNR on the latent reconstructions to choose this setting as shown in Table 7.

Dataset	**Schedule**	**NMSE**	**PCC**	**SNR (dB)**
SEED	Linear	0.42	0.71	4.18
	Cosine	**0.20**	**0.86**	**7.15**
SEED-IV	Linear	0.51	0.66	4.02
	Cosine	**0.19**	**0.88**	**7.24**
Localize-MI	Linear	0.14	0.93	8.39
	Cosine	**0.11**	**0.95**	**8.88**

Table 7: Comparison of noise schedules on three datasets (NMSE, PCC, SNR).

D.2.2 Training Timestep Lengths

We also tested different training timestep lengths (200, 1000, 2000). Larger values introduce stronger noise but make denoising harder; smaller values lack coverage of high-noise regimes. Using cosine scheduling, the results in Table 8 exhibit that 1000 timesteps to be optimal across datasets.

D.2.3 Cosine Schedule Offset Factor

In the cosine noise schedule for DDIM, the offset factor s adjusts the smoothness and starting point of the noise variance curve to prevent instability from overly small initial noise levels. Concretely, s introduces a phase shift in the cosine function, producing a more gradual noise increase at early timesteps—thereby avoiding abrupt noise jumps—while still covering the full variance range at later steps. Smaller values of s yield gentler initial noise ramp-up, whereas larger s accelerate early noise growth. Table 9 depicts the effect of cosine schedule offset factors on reconstruction quality.

Dataset	**Steps**	**NMSE**	**PCC**	**SNR (dB)**
SEED	200	0.32	0.76	5.95
	1000	**0.20**	**0.86**	**7.15**
	2000	0.29	0.78	6.19
SEED-IV	200	0.36	0.75	5.93
	1000	**0.19**	**0.88**	**7.24**
	2000	0.31	0.76	6.11
Localize-MI	200	0.15	0.92	8.37
	1000	0.11	0.95	8.88
	2000	**0.11**	**0.95**	**8.94**

Table 8: Impact of training timesteps on reconstruction quality

Dataset	s	**NMSE**	**PCC**	**SNR (dB)**
SEED	**0.005**	**0.20**	**0.86**	**7.15**
	0.010	0.24	0.84	7.02
	0.025	0.26	0.83	6.93
SEED-IV	0.005	0.20	0.86	7.17
	0.010	**0.19**	**0.88**	**7.24**
	0.025	0.20	0.85	7.11
Localize-MI	**0.005**	**0.11**	**0.95**	**8.94**
	0.010	0.15	0.93	8.41
	0.025	0.18	0.91	8.23

Table 9: Effect of cosine schedule offset factor s on reconstruction quality

D.3 Effect of λ_{RES} and λ_{SMM}.

To assess the sensitivity of SRGDiff to the weighting coefficients in the loss, we conduct a parameter study on the most challenging SR settings (highest SR factor) for each dataset. We vary the residual-guidance weight λ_{res} and the step-aware modulation weight λ_{SMM} around the default values used in the main paper, while keeping all other hyperparameters fixed.

Concretely, we sweep $\lambda_{\text{res}} \in \{0.1, 0.5, 1.0, 2.0, 5.0\}$ (relative to the default), and $\lambda_{\text{SMM}} \in \{0.001, 0.005, 0.01, 0.02, 0.1\}$. Tables 10 report the NMSE on SEED, SEED-IV, and Localize-MI under these settings.

λ_{res}	SEED	SEED-IV	Localize-MI
0.1	0.3928	0.3286	0.3941
0.5	0.3532	0.2822	0.3578
1.0	**0.3494**	**0.2603**	**0.3457**
2.0	0.3508	0.2810	0.3565
5.0	0.4012	0.3369	0.3827

(a) Effect of λ_{res}

λ_{SMM}	SEED	SEED-IV	Localize-MI
0.001	0.3975	0.3133	0.3904
0.005	0.3539	0.2696	0.3519
0.01	**0.3494**	**0.2603**	**0.3457**
0.02	0.3513	0.2712	0.3489
0.1	0.3990	0.3240	0.3915

(b) Effect of λ_{SMM}

Table 10: Effect of λ_{res} and λ_{SMM} on NMSE (hardest SR setting per dataset).

Overall, the performance is reasonably stable within a broad range around the default values, indicating that SRGDiff is not overly sensitive to these hyperparameters. When λ_{res} becomes too large, the residual guidance term dominates and suppresses learning in the diffusion backbone; when it is too small, the residual guidance has almost no effect. Similarly, if λ_{SMM} is too small, the guidance feature modulation is overly strong, whereas for very large λ_{SMM} the diffusion model effectively

Model	SEED			SEED-IV			Localize-MI			
	2×	4×	8×	2×	4×	8×	2×	4×	8×	16×
ESTformer	6.96	9.31	9.73	7.11	8.30	8.86	7.03	16.67	32.32	35.73
STAD	9.19	11.04	14.40	9.50	10.95	13.12	8.76	13.11	22.53	25.37
SRGDiff	**3.89**	**5.12**	**4.95**	**3.99**	**4.08**	**4.84**	**3.86**	**7.30**	**11.50**	**13.76**

Table 11: Frequency-domain NMSE between reconstructed and real HD topomaps on SEED, SEED-IV, and Localize-MI under different SR factors.

ignores the guidance features. The default configuration achieves the best overall trade-off across datasets.

E Downstream Task

E.1 Classification Feature Extraction

E.1.1 Differential Entropy Feature

Raw EEG at 1000 Hz is downsampled to 200 Hz, band-pass filtered (1–50 Hz) with a 6th-order Butterworth filter, and segmented into non-overlapping 1 s windows (200 samples). Each window is transformed by STFT (Hanning window, 200-point length, 256-point FFT). We compute band-specific power E for δ (1–3 Hz), θ (4–7 Hz), α (8–13 Hz), β (14–30 Hz), and γ (31–50 Hz) , normalize by the number of bins N, and define differential entropy (DE) feature as $\log(E/N)$ with a small constant added for numerical stability.

E.1.2 Power Spectral Density Feature

Power spectral density (PSD) feature features use the same STFT pipeline but report the mean squared magnitude (average power) in each band.

E.2 Random Forest Classifier

For emotion classification on SEED and SEED-IV and epileptic detection on Localize-MI, we employ a random forest with 100 trees. This set of hyperparameters balances nonlinearity modeling with computational efficiency, yielding robust performance on high-dimensional EEG features.

F Supplemental Ablation Results

F.1 Frequency-domain Topomap Error

To complement the qualitative topographic visualizations in Figure 4 and rule out potential visual selection bias, we report a quantitative frequency-domain MAE between reconstructed and real HD topomaps. Concretely, we first transform both reconstructed and reference HD EEG into the frequency domain, aggregate power within standard EEG bands (e.g., θ, α, β), and interpolate the band power of each channel onto a 2D scalp grid using electrode coordinates. We then compute the pixel-wise normalized mean squared error between the reconstructed and real topomaps, averaged over all frequency bands and test samples. A lower value indicates that the model better preserves both the spectral content and its spatial distribution over the scalp.

Table 11 reports frequency-domain MAE on SEED, SEED-IV, and Localize-MI under different SR factors. SRGDiff consistently achieves the lowest error across all datasets and scales, with a larger margin over ESTformer and STAD than in time-domain NMSE/PCC/SNR. This confirms that our residual-guided generative formulation not only improves pointwise reconstruction quality, but also more faithfully recovers the HD spectral–spatial structure.

F.2 Additional Ablation Studies on SEED-IV and Localize-MI datasets

Figure 10 illustrates the **LDM+LD** baseline used in our ablations. In the first stage (top row), we pretrain a VAE on full high-density (HD) EEG: preprocessed HD signals are passed through the encoder E and decoder D to learn a latent space tailored to HD scalp topography. In the second stage (bottom row), low-density EEG is first mapped into this latent space using the pretrained encoder E, yielding a static guidance feature. This guidance feature is then added to the diffusion latent during the denoising process, and the final denoised latent is decoded by D back to HD EEG, without employing RDM or SMM.

Figure 11 presents additional ablation studies on SEED-IV and Localize-MI datasets, reporting NMSE, PCC, and SNR for the baseline LDM+LD, LDM+SMM, LDM+RDM, and the full SRGDiff across various upsampling scales. These plots further illustrate the individual and combined contributions of our two modules to reconstruction quality.

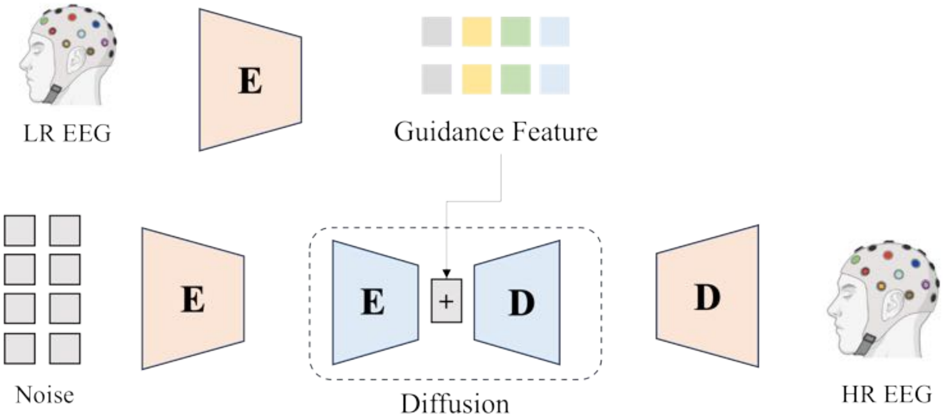


Figure 10: Illustration of the LDM+LD baseline.

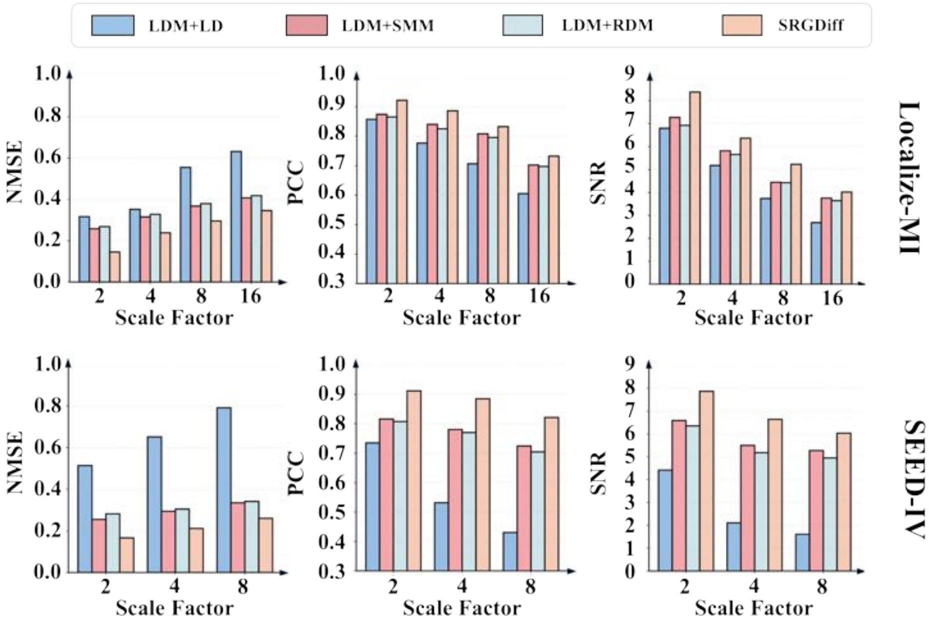


Figure 11: Ablation results on SEED-IV and Localize-MI: comparison of NMSE, PCC, and SNR for LDM+LD, LDM+SMM, LDM+RDM, and SRGDiff at 2×, 4×, and 8× upsampling and 2×, 4×, 8×, and 16× upsampling, respectively.

G Efficiency Analysis

G.1 Computational Cost under Comparable Parameter Budgets

To complement the main-paper results, we further compare the computational cost of SRGDiff with simpler transformer-based SR models under comparable parameter budgets. Table 12 reports, for ESTformer, STAD, and SRGDiff, the total number of trainable parameters and the average computation cost per 4 s EEG window (measured as FLOPs under the same input resolution). Although SRGDiff requires more FLOPs than the single-pass transformer ESTformer due to the iterative denoising process, its parameter count remains in the same order of magnitude as transformer-based baselines, and the additional cost is the price we pay for exploiting a strong latent diffusion prior.

Method	**#Params (M)**	**GFLOPs / window**
ESTformer	12.111	4.302
STAD	13.949	1.650
SRGDiff	2.342	1.38

Table 12: Computational cost of different SR models under comparable parameter budgets. We report the total number of trainable parameters and GFLOPs per 4 s EEG window. SRGDiff uses a diffusion-based denoiser, so its FLOPs are higher than those of ESTformer, but the parameter budget remains comparable.

G.2 Runtime Comparison

In addition to static computational cost, we also measure wall-clock runtime for all methods on the three EEG datasets under a unified implementation and hardware setup (same GPU, batch size, and input length). Table 13 reports the average per-sample latency for a 4 s window. Despite using an iterative denoising process, SRGDiff achieves inference times between 63 ms and 92 ms, which are significantly faster than most diffusion-based baselines and are well below 0.1 s. In contrast, ESTformer attains the smallest latency thanks to its single-pass transformer structure, but, as shown in the main paper, does not match SRGDiff in reconstruction quality. Overall, these results indicate that SRGDiff strikes a favorable balance between accuracy and efficiency, and is suitable for real-time EEG spatial super-resolution.

Method	**SEED**	**SEED-IV**	**Localize-MI**
SasDim	265.4	269.5	374.4
SADI	329.0	324.1	428.7
RDPI	318.1	315.9	421.2
DDPMEEG	549.3	558.2	841.7
ESTformer	3.51	3.55	5.03
STAD	232.6	227.8	385.9
SRGDiff	63.0	62.1	92.5

Table 13: Runtime (ms) of different methods on SEED, SEED-IV, and Localize-MI datasets. Values are average per-sample latency for a 4 s EEG window measured on the same GPU. SRGDiff remains below 0.1 s in all cases while achieving the best reconstruction quality.

H Generalization Analysis

H.1 Extension to Cross-Subject and Cross-Session Settings

EEG signals are known to exhibit strong subject- and session-specific variability. To explicitly examine whether SRGDiff can generalize under this variability, we perform additional experiments on the SEED dataset in a stricter cross-subject and cross-session regime.

For the cross-session setting, we use all subjects in SEED and conduct three experiments: in each experiment, two sessions are used for training and the remaining session is held out for testing.

Model	**Metric**	**Cross-subject (SEED)**			**Cross-session (SEED)**		
		2×	**4×**	**8×**	**2×**	**4×**	**8×**
ESTformer	NMSE	0.4411±0.004	0.4729±0.006	0.5633±0.008	0.3624±0.004	0.4029±0.008	0.5129±0.007
	PCC	0.7393±0.009	0.7189±0.007	0.6515±0.009	0.7924±0.009	0.7742±0.007	0.6954±0.009
	SNR	3.9516±0.031	3.5414±0.004	2.7279±0.039	4.8753±0.034	4.4097±0.047	3.1375±0.043
STAD	NMSE	0.5537±0.004	0.7325±0.002	0.9267±0.008	0.5617±0.003	0.7675±0.002	0.9376±0.007
	PCC	0.6224±0.003	0.4587±0.002	0.2959±0.004	0.6373±0.003	0.4397±0.002	0.2791±0.004
	SNR	3.3872±0.097	1.2449±0.049	0.7276±0.139	3.1794±0.089	1.1297±0.042	0.7168±0.122
SRGDiff	NMSE	**0.2675**±0.003	**0.3829**±0.005	**0.4512**±0.005	**0.2480**±0.003	**0.3529**±0.004	**0.4127**±0.005
	PCC	**0.8023**±0.004	**0.7232**±0.004	**0.6902**±0.004	**0.8508**±0.004	**0.7932**±0.003	**0.7702**±0.007
	SNR	**5.6913**±0.067	**4.5657**±0.055	**4.1189**±0.084	**6.1473**±0.093	**4.1857**±0.052	**3.8189**±0.034

Table 14: Cross-subject and cross-session reconstruction performance on SEED under different SR factors (mean ± std over folds) in terms of NMSE, PCC, and SNR.

We then report averages over all subjects and session splits. For the cross-subject setting, we train SRGDiff on subjects 1-12 and evaluate on held-out subjects 13-15 without any subject-specific fine-tuning.

The reconstruction performance under both cross-session and cross-subject protocols is summarized in Table 14. We observe that SRGDiff degrades gracefully in the cross-session setting, maintaining strong performance at 2× and 4× SR with more noticeable degradation at 8×, while the cross-subject setting is substantially more challenging and leads to larger performance drops across all SR factors than the random division settings. Nevertheless, SRGDiff maintains a clear margin over strong baselines ESTformer, STAD in terms of NMSE, PCC, and SNR, indicating that the proposed partial-observation diffusion formulation is reasonably robust to session- and subject-level variability on SEED.

H.2 Extension to ECoG Channel Super-resolution

To examine whether SRGDiff is specific to scalp EEG or can generalize to other multi-channel neurophysiological signals, we further evaluate it on an invasive electrocorticography (ECoG) dataset. We use the public ECoG benchmark AJILE12 (Peterson et al., 2022) and follow the setting of Vetter et al. (2024). Applying SRGDiff here serves two purposes: (i) it tests whether our partial-observation formulation and dynamic residual guidance are *modality-agnostic* within the family of spatially organized neural recordings, and (ii) it verifies that the proposed method works under a different signal regime (invasive ECoG rather than scalp EEG) without any architecture or hyperparameter changes. As shown in Table 15, SRGDiff consistently improves over transformer-based SR baselines on the ECoG benchmark, supporting our claim that the approach extends beyond EEG to other neurophysiological channel super-resolution tasks.

Model	Metric	2×	4×	8×
ESTformer	NMSE	0.4573	0.7189	0.8517
	PCC	0.7367	0.5299	0.3845
	SNR	3.3991	1.4334	0.6974
STAD	NMSE	0.4932	0.6901	0.7987
	PCC	0.6854	0.5118	0.4312
	SNR	3.1686	1.4449	1.1684
SRGDiff (ours)	NMSE	**0.3575**	**0.6529**	**0.7312**
	PCC	**0.8023**	**0.5332**	**0.4502**
	SNR	**4.8913**	**2.1657**	**1.9089**

Table 15: Channel super-resolution performance (NMSE, PCC, and SNR) on the ECoG dataset from Vetter et al. (2024), following their windowing and data split. SRGDiff consistently improves over transformer-based baselines across all SR factors and metrics.

Train LD chans	Test LD chans	NMSE	PCC	SNR
16	8	0.4542	0.7196	4.0329
16	10	0.4031	0.7650	4.4211
16	12	0.3588	0.8012	4.8705
16	14	0.3245	0.8268	5.1203
16	**16 (base)**	**0.2977**	**0.8445**	**5.2606**
32	8	0.4753	0.6735	3.9518
32	16	0.3585	0.7820	4.4002
32	**32 (base)**	**0.1632**	**0.9102**	**7.8413**

Table 16: SEED dataset: robustness of SRGDiff to variable LD montages. The model is trained with either a 16- or 32-channel LD configuration and evaluated on subsampled LD inputs at test time without retraining.

H.3 Extension to Variable and Irregular LD Electrode Layouts

In many practical deployments, the available low-density electrode layout may differ across subjects, sessions, or hardware configurations. In addition, real-world recordings often contain missing or corrupted channels, leading to irregular montages that deviate from the nominal LD design. A natural question is whether a single model can generalize across such variable and potentially irregular LD layouts.

In SRGDiff, the LD input is treated as a set of spatially localized observations that are first mapped into a common latent representation via the pretrained VAE. Concretely, each LD electrode is embedded into a continuous scalp (or cortical) coordinate space, and its signal is projected onto a fixed latent grid on which the diffusion model operates. The guidance network then consumes these latent features rather than discrete channel indices, so the conditioning is not tied to a specific LD channel configuration. This design makes the model inherently more flexible to changes in the number and spatial arrangement of LD electrodes at inference time.

To verify this empirically, we conducted two sets of experiments on SEED. First, we trained SRGDiff with a 16-electrode LD configuration and evaluated it at test time on 8/10/12/14-electrode inputs obtained by subsampling the 16-channel montage. Second, we trained SRGDiff with a 32-electrode LD configuration and evaluated it on 8- and 16-electrode inputs, again using only subsampling at test time and no retraining. As summarized in Tables 16, SRGDiff degrades gracefully as the LD montage becomes sparser: NMSE increases moderately, while PCC and SNR remain competitive across all tested LD configurations. These results indicate that a single SRGDiff model can handle sparser or irregular LD layouts at inference time without retraining, provided that the new electrodes can be embedded into the same spatial coordinate system and projected onto the latent grid used during training.

I Reconstruction Visualization

Figure 12 through Figure 14 illustrate qualitative reconstructions on the three datasets. For each, we plot a single representative channel over time, comparing the ground-truth high-density EEG (black) against STAD (blue), ESTformer (red), and SRGDiff (green). These overlays demonstrate SRGDiff's closer alignment with the true waveform across diverse temporal patterns.

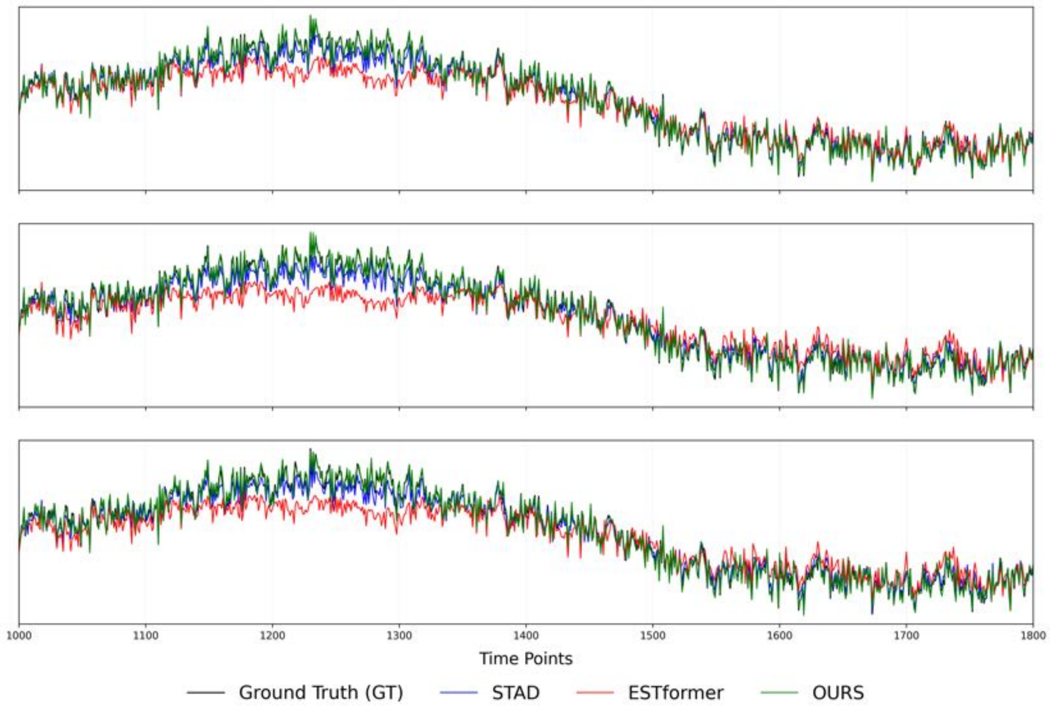


Figure 12: Reconstruction of a representative channel on Localize-MI (motor imagery): ground truth (black), STAD (blue), ESTformer (red), and SRGDiff (green).

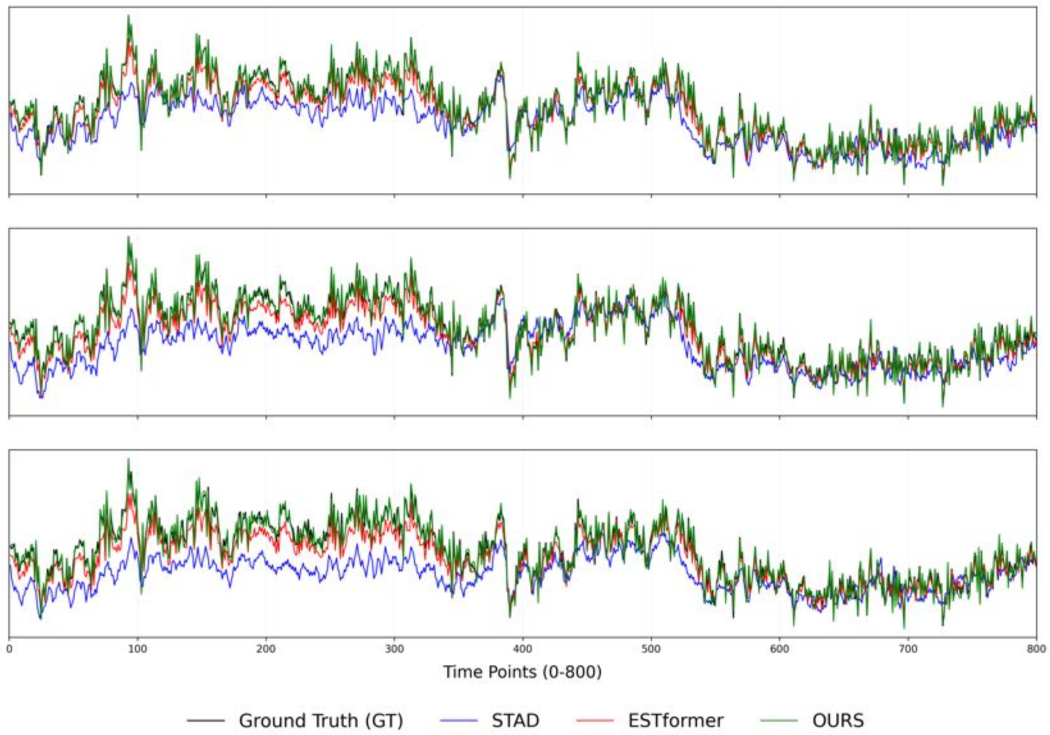


Figure 13: Reconstruction of a representative channel on SEED-IV (emotion recognition): ground truth (black), STAD (blue), ESTformer (red), and SRGDiff (green).

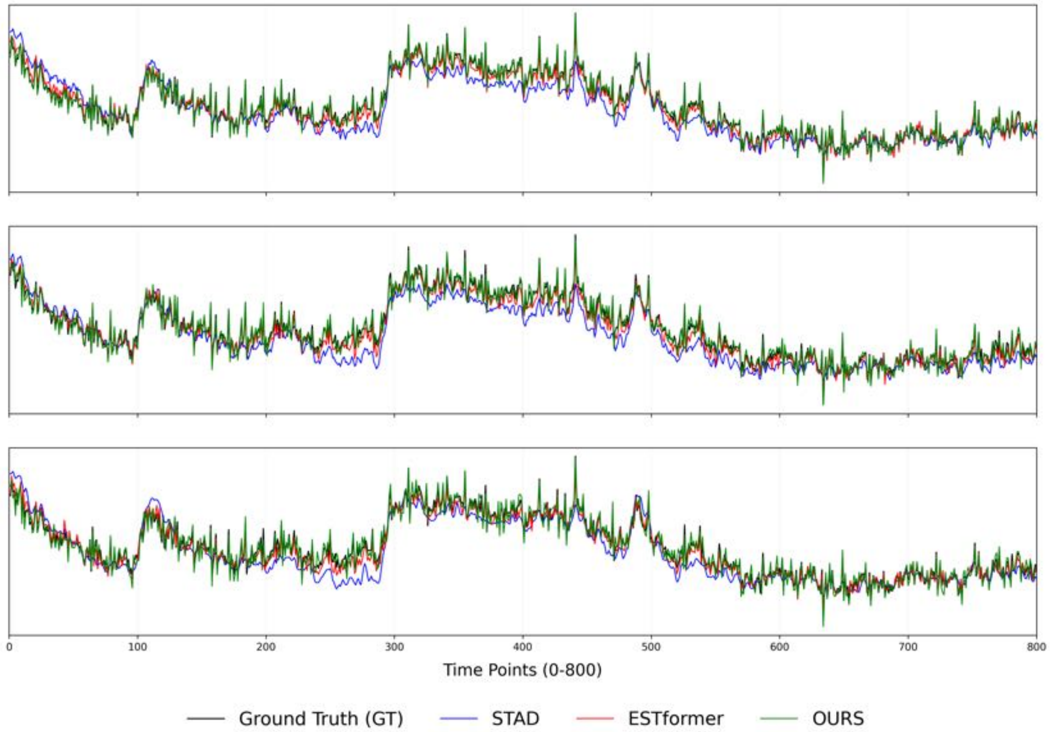


Figure 14: Reconstruction of a representative channel on SEED (emotion recognition): ground truth (black), STAD (blue), ESTformer (red), and SRGDiff (green).

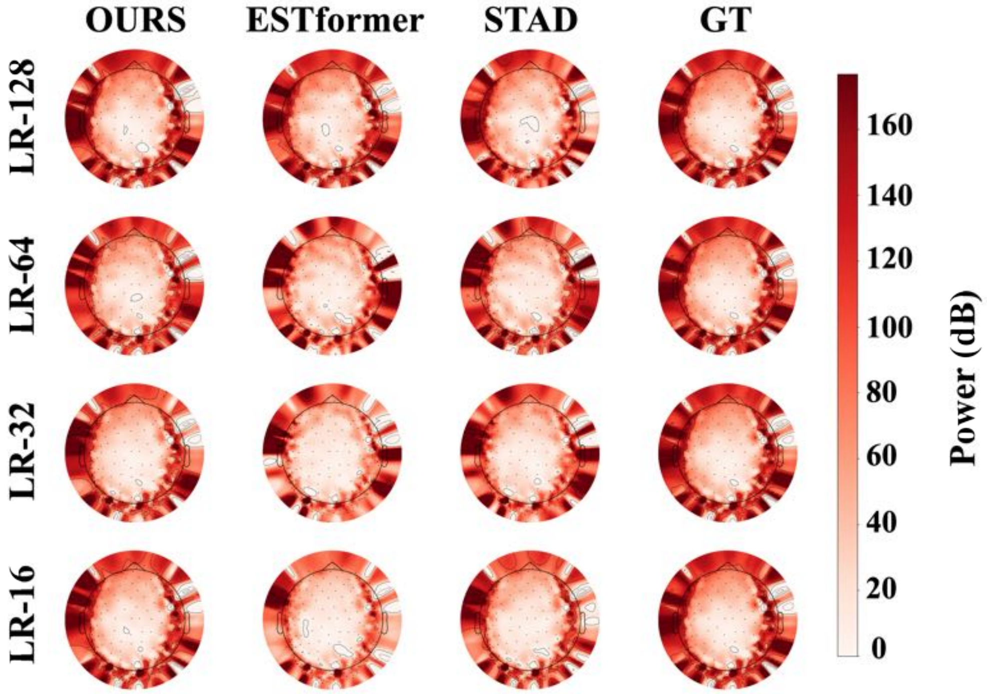


Figure 15: Visualization of EEG topographic maps between ground-truth and reconstructed EEG signals by ESTformer, STAD and SRGDiff on Localize-MI.