GEOMETRY- AND RELATION-AWARE DIFFUSION FOR EEG SPATIAL SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent electroencephalography (EEG) spatial super-resolution (SR) methods, while showing improved quality by either directly predicting missing signals from visible channels or adapting latent diffusion-based generative modeling, often lack awareness of physiological spatial structure, thereby constraining spatial generation performance. To address this, we reformulate EEG spatial SR as a structured multimodal conditional generation problem and introduce TopoDiff, a geometry- and relation-aware diffusion framework. Inspired by how human experts interpret spatial EEG patterns, TopoDiff conditions generation on sparse EEG signals together with two complementary structured views: topology-aware image embeddings derived from EEG topographic representations, which provide global geometric context, and dynamic channel-relation graphs, which encode inter-electrode relationships that evolve with temporal dynamics. This design yields a spatially grounded EEG spatial super-resolution framework with consistent performance improvements. Across multiple EEG datasets spanning diverse applications, including SEED/SEED-IV for emotion recognition, PhysioNet motor imagery (MI/MM), and TUSZ for seizure detection, our method achieves substantial gains in generation fidelity and leads to notable improvements in downstream EEG task performance.

1 INTRODUCTION

Electroencephalography (EEG) provides a non-invasive window into distributed cortical dynamics of brain and is widely used across affective computing (Zheng & Lu, 2015; Zheng et al., 2018; Chen et al., 2023), brain–computer interfaces (BCIs) (Goldberger et al., 2000; Schalk et al., 2004; Cho et al., 2017), and clinical neurophysiology such as epilepsy diagnosis (Herman et al., 2015; Waak et al., 2023). Despite its broad applicability, standard EEG acquisition typically requires dense electrode placement over the scalp, which introduces hardware constraints and user discomfort, thereby limiting its use in out-of-lab scenarios. In contrast, low-density recording systems, such as wearable EEG devices, enable practical and longitudinal monitoring (Biondi et al., 2024; Cho et al., 2025; Vincenzo et al., 2025; Markov et al., 2024). However, these systems rely on sparse electrode layouts that discard fine-grained spatial information, leading to reduced performance in downstream analysis.

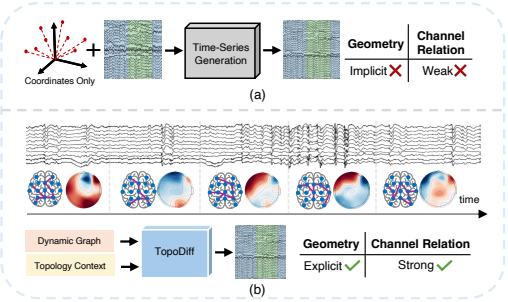

Figure 1: Comparison between previous pipelines and the proposed framework. (a) Prior methods only implicitly encode spatial geometry through coordinates and positional encoding. (b) **TopoDiff** enables geometry- and relation-aware generation by incorporating spatial and relational modeling.

This gap has driven increasing interest in EEG spatial super-resolution (EEG SR) (Courellis et al., 2016; Corley & Huang, 2018; Svantesson et al., 2021), which aims to generate high-density EEG signals from sparse recordings. More recently, STAD (Wang et al., 2024a) employs a latent diffusion model to generate missing channels, while ESTFormer (Li et al., 2025) directly predicts missing signals using a masked autoencoder (He et al., 2022). However, as shown in Figure 1(a), these methods encode spatial information only implicitly through positional embeddings. Although such

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

designs provide channel identity and location information, they cannot explicitly capture the global spatial organization of scalp activity or dynamic inter-electrode dependencies that evolve over time.

These observations motivate explicitly modeling electrode geometry and structured inter-channel relationships, rather than treating spatial positions as passive metadata. In clinical practice, EEG channels are rarely interpreted in isolation; instead, EEG signals are often summarized using *topographic maps* (topoplots), which visualize spatial distribution of signal amplitudes over the scalp (Koles & Paranjape, 1988). Although topographic maps are coarse summaries rather than exact signal representations, they provide a global, geometry-aware view that highlights spatial organization and patterns, and naturally generalize across different electrode montages. This perspective suggests a complementary inductive bias for EEG spatial super-resolution: leveraging topographic representations to provide global spatial context. Moreover, electrode distributions can be naturally modeled using graph structures, which enable more structured and relational representations than treating channels as independent signals (Kotoge et al., 2025).

Building on these insights, as shown in Fig. 1(b), we reformulate EEG spatial super-resolution as a structured multi-modal conditional generation problem. We propose **TopoDiff**, a spatially grounded EEG SR framework that conditions diffusion on sparse EEG signals together with two complementary structured views: (i) **topology-aware image representations** derived from EEG topographic visualizations, which provide global scalp geometry context; and (ii) **dynamic channel-relation graphs** that model time-varying cross-channel dependencies. This formulation enables the generator to jointly leverage signal-level observations, global geometric structure, and local relational dynamics for spatially coherent EEG reconstruction.

Our core contributions are summarized as follows: (1) We propose TopoDiff, a structured multi-modal conditional diffusion framework for EEG spatial super-resolution. (2) We introduce a dual structured-conditioning mechanism that combines **topology-aware modeling** via EEG topographic image embeddings for global scalp geometry with **relation-aware modeling** via channel-relationship graphs for structured and time-varying inter-channel dependencies, enabling generation of fine-grained spatiotemporal EEG patterns. (3) Extensive experiments on SEED, SEED-IV, TUSZ, and PhysioNet MI/MM demonstrate improved reconstruction fidelity and stronger downstream performance compared to competitive baselines.

2 Related Work

Conditional Diffusion Methods. Conditional generation enables data synthesis guided by rich contextual information, ranging from class labels (Peebles & Xie, 2023), input signals (Wang et al., 2024a) to auxiliary modalities (Esser et al., 2024; Wan et al., 2025; Wei et al., 2025). Recently, diffusion models have emerged as a dominant paradigm for conditional generation due to their strong modeling flexibility and high sample quality (Dhariwal & Nichol, 2021; Ho & Salimans, 2022; Saharia et al., 2022; Rombach et al., 2022; Esser et al., 2024; Wan et al., 2025). Motivated by this success, we adopt diffusion models for EEG SR and incorporate structured, domain-specific priors as conditioning guidance, rather than relying solely on signal-level inputs.

Generative Models for Time Series and EEG Super-Resolution. Generative modeling for time series has been widely explored for data augmentation (Shu et al., 2025), privacy-preserving synthesis (Tian et al., 2024), and probabilistic forecasting or imputation (Tashiro et al., 2021), with prominent model families including VAEs, GANs (Yoon et al., 2019), and, more recently, diffusion models for uncertainty-aware conditional generation (Rasul et al., 2021) (e.g., **RDPI** improves spatiotemporal imputation via a two-stage deterministic–diffusion framework (Liu et al., 2025b)). When applied to EEG, generative models have primarily been used for signal synthesis and data augmentation to mitigate data scarcity, but they typically operate on discrete channel tensors and encode electrode layout only implicitly, without explicitly modeling montage geometry (Shu et al., 2025; Williams et al., 2025). Recent EEG SR baselines adapt sequence models or diffusion frameworks specifically for spatial upsampling. **ESTFormer** (Li et al., 2025) employs Transformer attention with positional encodings and masking to reconstruct missing channels, while **STAD** (Wang et al., 2024a) and **SRGDiff** (Liu et al., 2025a) adopt diffusion-based generation with signal-derived conditioning and step-wise guidance. However, these methods largely rely on positional encodings and learn inter-channel relations implicitly within the model, which can overlook global scalp geometry and structured inter-electrode organization.

Downstream EEG Tasks. EEG has been used for tasks such as emotion recognition (Zheng & Lu, 2015), motor imagery (MI) classification (Schalk et al., 2004), and seizure detection (Herman et al., 2015), which are commonly adopted to evaluate EEG signal representations. EEG-based emotion recognition (Zheng & Lu, 2015; Craik et al., 2019) depends on spatial characteristics such as frontal asymmetry and cross-regional connectivity, potentially impacting multiple regions of the brain, indicating that any degradation in spatial fidelity can affect performance. MI classification (Schirrmeister et al., 2017; Lawhern et al., 2018) decodes limb movement from EEG signals and is highly sensitive to the spatial characterization of sensorimotor rhythms, which low-density montage often fails to capture. Finally, seizure detection provides a clinically grounded downstream task that requires preservation of both localized and distributed pathological patterns. Epileptic activity may manifest as focal or generalized abnormalities (Shah et al., 2018), making it extremely important to have an accurate spatial and temporal characterization of EEG signals (Shoeb, 2009; Roy et al., 2019). Collectively, these tasks span affective computing, BCI, and clinical neurophysiology–motivating work in spatially-grounded EEG SR methods.

3 PRELIMINARY

3.1 EEG SPATIAL SUPER-RESOLUTION

Let $X^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T}$ denote a low-resolution (sparse-montage) EEG segment with C_{LR} channels and T time steps, and let $X^{\mathrm{HR}} \in \mathbb{R}^{C_{\mathrm{HR}} \times T}$ denote the corresponding high-resolution EEG with C_{HR} channels recorded over the same time window. The EEG spatial super-resolution (EEG SR) task is defined as estimating $X^{\mathrm{SR}} \in \mathbb{R}^{C_{\mathrm{HR}} \times T}$ that approximates X^{HR}, given X^{LR} as input. Since the low-resolution channels correspond to a known subset of the high-resolution montage, we define the target as reconstructing the unobserved signals $X^{\mathrm{Unseen}} \in \mathbb{R}^{C_{\mathrm{Unseen}} \times T}$.

3.2 DENOISING DIFFUSION MODELS

We adopt a conditional diffusion model in this work for its ability to incorporate diverse conditioning information into generation (Tashiro et al., 2021; Saharia et al., 2022), allowing sparse EEG recordings and structured spatial information to jointly guide missing-channel reconstruction. This choice is supported by our empirical results in Appendix M, where TopoDiff outperforms direct regression using the same architecture and topological and graph conditioning.

Diffusion models generate data by learning to reverse a gradual noising process, which can be interpreted through ordinary differential equations (ODEs) (Song et al., 2022; Chen et al., 2018; Albergo et al., 2025). During training, a noisy sample z is constructed by interpolating between a clean signal $x \sim p(\mathbf{X})$ and Gaussian noise $\epsilon \sim \mathcal{N}(0, \mathbb{I})$. Following (Li & He, 2025), we adopt a linear interpolation parameterization: $z = tx + (1-t)\epsilon$, where $t \in (0,1)$ is sampled from a logit-normal distribution with $t = \mathrm{sigmoid}(\tau)$ and $\tau \sim \mathcal{N}(0, 1.2^2)$.

In our training, noise is applied only to the unobserved channels, while the observed signals X^{LR} are provided as conditioning inputs. Similar to (Li & He, 2025), we employ an *x-prediction* objective, where a denoising network f_θ directly predicts the clean signal $X^{\mathrm{Unseen}}_{\mathrm{pred}} = f_\theta\left(z, X^{\mathrm{LR}}, t\right)$, and is trained using an ℓ_2 reconstruction loss $\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{x,\epsilon,t}\left[\|X^{\mathrm{Unseen}}_{\mathrm{pred}} - x^{\mathrm{Unseen}}\|_2^2\right]$.

At inference time, generation is performed by solving the reverse-time ODE conditioned on X^{LR}. Under the x-prediction formulation, the corresponding velocity field is $v_\theta(z, X^{\mathrm{LR}}, t) = \frac{X^{\mathrm{Unseen}}_{\mathrm{pred}} - z}{1-t}$, which defines the ODE $\frac{\mathrm{d}z}{\mathrm{d}t} = v_\theta(z, X^{\mathrm{LR}}, t)$. Starting from Gaussian noise, we integrate the ODE via $z' = z + \Delta t \cdot v_\theta(z, X^{\mathrm{LR}}, t)$ from $t = 0$ to $t = 1$ to obtain the reconstructed unseen channels $X^{\mathrm{Unseen}}_{\mathrm{gen}}$. In practice, we use a single Euler step for efficient inference, which provides a strong trade-off between computational cost and reconstruction quality.

4 TOPODIFF FRAMEWORK

The core idea of TopoDiff is to integrate geometric and relational inductive biases into the diffusion process in order to ensure spatially and temporally coherent EEG generation. We first introduce

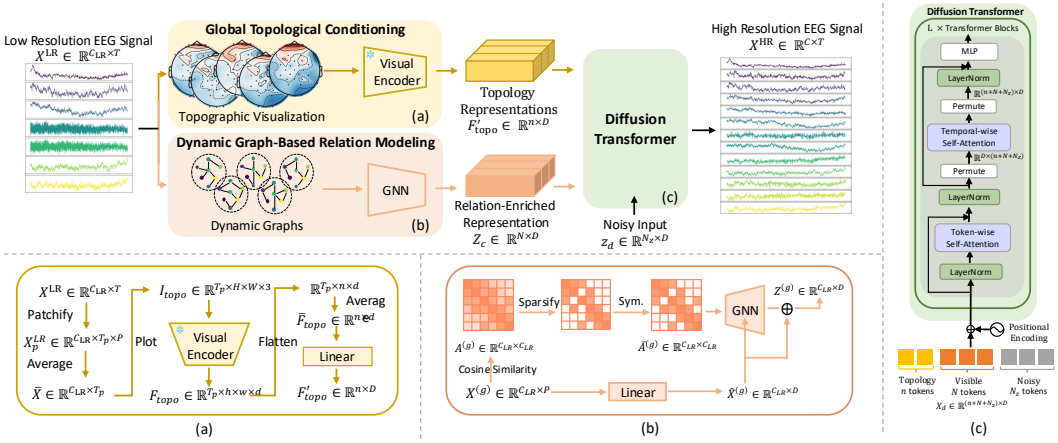

Figure 2: Overview of the **TopoDiff** framework. (a) Global topological conditioning, where EEG topographic representations are used to provide global scalp geometry and spatial context. (b) Dynamic graph-based inter-channel relation modeling, in which graphs are constructed in a time-varying manner and a GNN is used to enrich channel representations with explicit relational information. (c) The diffusion Transformer architecture employed for EEG spatial super-resolution.

the global topological conditioning derived from EEG topographic representations (Section 4.1), followed by the dynamic graph-based inter-channel relation modeling (Section 4.2). The diffusion Transformer architecture used for EEG signal generation is described in Section 4.3

4.1 Global Topological Conditioning

To provide the diffusion model with explicit global scalp geometry information, we incorporate high-level features extracted from EEG topographic representations as conditioning signals. These representations capture coarse-grained spatial organization across the entire scalp, providing global structural context that complements local channel-wise relational modeling. This design can be viewed as a form of cross-modal conditioning, in which image-based representations deterministically derived from EEG signals serve as auxiliary geometric priors to guide time-series generation.

Topographic Representation Construction. We leverage EEG topographic visualization images (Gramfort et al., 2013), constructed by interpolating EEG signals onto a two-dimensional projection of the scalp (Perrin et al., 1989; Koles & Paranjape, 1988). Such spatial visualizations are widely used by human experts to analyze global spatial patterns in EEG data. Given low-resolution EEG input $\mathbf{X}^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T}$, we partition it along the temporal dimension into non-overlapping patches, yielding $\mathbf{X}_p^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p \times P}$. Within each temporal patch, we apply a channel-wise temporal averaging operation to obtain $\bar{\mathbf{X}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p}$, summarizing the dominant signal pattern within each time slot. Each averaged slice is then visualized as a color-coded two-dimensional EEG topographic image $\mathbf{I}_{\mathrm{topo}} \in \mathbb{R}^{H \times W \times 3}$.

Global Feature Extraction and Conditioning. For each plotted image, we apply a frozen pretrained image feature extractor (*e.g.*, DINOv3 (Siméoni et al., 2025)) to obtain a spatial feature map of shape $\mathbb{R}^{h \times w \times d}$, where h and w denote spatial resolution and d is the feature dimension. These features capture high-level spatial and semantic context from the EEG topographic representations. By concatenating the feature maps across all T_p temporal slices, we obtain the topographic feature tensor $\mathbf{F}_{\mathrm{topo}} \in \mathbb{R}^{T_p \times h \times w \times d}$. To enable efficient conditioning in generative diffusion models (Wan et al., 2025; Yang et al., 2024), we flatten the spatial dimensions and compute the average over temporal slices, yielding $\bar{\mathbf{F}}_{\mathrm{topo}} \in \mathbb{R}^{n \times d}$, where $n = h \times w$.

4.2 Dynamic Graph-Based Inter-Channel Relation Modeling

In addition to global topological conditioning, we use a dynamic graph neural network (GNN) over observed EEG channels to capture time-varying inter-channel dependencies for coherent multi-channel generation.

Group-wise Relation Estimation. Given the patchified input $X_p^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p \times P}$, we estimate time-varying inter-channel relations for each temporal group $X_p^{\mathrm{LR}}(:, g, :)$, where $g \in \{0, \ldots, T_p - 1\}$. Specifically, a dense relation matrix $A^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times C_{\mathrm{LR}}}$ is computed using cosine similarity, with $A_{ij}^{(g)} = \frac{\langle v_i^{(g)}, v_j^{(g)} \rangle}{(\|v_i^{(g)}\| + \epsilon)(\|v_j^{(g)}\| + \epsilon)}$, where $v_i^{(g)}$ denotes the input of channel i in group g, and ϵ is a small constant to ensure numerical stability for low-energy signals. This group-wise estimation captures cross-channel dependencies that evolve over time.

Graph Construction. Given $A^{(g)}$, we construct a sparse, relation-aware graph for GNN processing. For each temporal group, nodes correspond to visible EEG electrodes, and spatial locality is enforced using the international 10–20 system (Klem et al., 1999). Each node is first connected to its n_s nearest neighbors based on electrode coordinates, defining a candidate neighborhood. Edge weights are then assigned using the absolute value of the relation matrix, $\tilde{A}^{(g)} = |A^{(g)}|$, allowing both positive and negative correlations to contribute to relational strength. However, the dense relation matrix may contain noisy or weak correlations. To promote robust relational structure, we retain only the top-k strongest edges per node within the candidate neighborhood, preserving dominant local dependencies that are more consistent with the spatial organization of EEG signals. Since row-wise pruning may break symmetry, we symmetrize the pruned adjacency matrix as $\tilde{A}^{(g)} \leftarrow \max\Big(\tilde{A}^{(g)}, (\tilde{A}^{(g)})^\top\Big)$. Finally, self-loops are added and the adjacency is normalized before GNN propagation.

Relation-Aware Channel Encoding with GNN. For each temporal group g, we construct the normalized adjacency matrix using the GCN formulation (Kipf & Welling, 2016), where $\bar{A}^{(g)} = \big(D^{(g)}\big)^{-\frac{1}{2}} \hat{A}^{(g)} \big(D^{(g)}\big)^{-\frac{1}{2}}$, $\hat{A}^{(g)} = \tilde{A}^{(g)} + \mathbb{I}$, and $D_{ii}^{(g)} = \sum_j \hat{A}_{ij}^{(g)}$. Given the group-wise input $X^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times P}$, we project it to a D-dimensional embedding $\hat{X}^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times D}$. We then perform graph-based relation propagation with $X_1^{(g)} = \sigma\Big(\bar{A}^{(g)} \hat{X}^{(g)} W_1\Big)$, $X_2^{(g)} = \bar{A}^{(g)} X_1^{(g)} W_2$, and $Z^{(g)} = \hat{X}^{(g)} + X_2^{(g)}$, where $W_1, W_2 \in \mathbb{R}^{D \times D}$ are learnable parameters and $\sigma(\cdot)$ is a nonlinear activation. The residual connection lets the module act as a lightweight relational adapter, enriching channel representations with graph interactions while preserving signal information.

4.3 Diffusion Transformer Architecture

Building on the representations introduced in Sections 4.1 and 4.2, we incorporate both global topological features and relation-aware graph features as conditioning inputs. Specifically, the topographic representation $\bar{\mathbf{F}}_{\mathrm{topo}} \in \mathbb{R}^{n \times d}$ is projected to the latent dimension via a linear layer, yielding $\mathbf{F}'_{\mathrm{topo}} \in \mathbb{R}^{n \times D}$. The relation-enriched features from the GNN, $Z \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p \times D}$, are reshaped into a token sequence $\mathbf{Z}_c \in \mathbb{R}^{N \times D}$, where $N = C_{\mathrm{LR}} \times T_p$.

We apply the same patchification to the noisy input z, yielding $z_c \in \mathbb{R}^{N_z \times P}$, where $N_z = C_{\mathrm{Unseen}} \times T_p$, followed by a linear projection to obtain $z_d \in \mathbb{R}^{N_z \times D}$. The Transformer input is constructed by concatenating all tokens along the sequence dimension as $X_d = [\mathbf{F}'_{\mathrm{topo}}; \mathbf{Z}_c; z_d] \in \mathbb{R}^{(n+N+N_z) \times D}$, enabling joint conditioning on global geometry and local relation-enriched features. We add fixed sinusoidal positional encodings to this input sequence: standard 2D sine–cosine encodings are used for the topographic tokens on the image grid, while EEG tokens use coordinate-aware 3D sine–cosine encodings following ESTFormer (Li et al., 2025), obtained by normalizing each electrode coordinate axis and concatenating axis-wise sinusoidal embeddings.

The Transformer architecture used in our method is illustrated in Fig. 2(c). Each Transformer block consists of two attention modules designed to capture spatiotemporal interactions: a token-wise self-attention layer and a temporal-wise self-attention layer. The token-wise self-attention operates over the full token set of size $n+N+N_z$, enabling global interactions across modalities and channels. Temporal-wise self-attention is inspired by ESTFormer, where the representation is transposed as $X_t \in \mathbb{R}^{D \times (n+N+N_z)}$, and attention is performed along the first dimension. The features are then transposed back to the token-first format for subsequent processing. After each attention operation, *LayerNorm* and residual connections are applied, followed by an MLP at the end of each block.

After all Transformer blocks, we retain the features corresponding to the noisy unseen-channel tokens and project them back to the signal space, yielding $X_{\mathrm{pred}}^{\mathrm{Unseen}} \in \mathbb{R}^{C_{\mathrm{Unseen}} \times T}$. Supervision is applied

Table 1: Results of signal fidelity across different datasets, comparing against recent baseline methods under $2\times$, $4\times$, and $8\times$ SR settings. Results with (mean + std) are shown in Appendix E.

	$2\times$ SR			$4\times$ SR			$8\times$ SR		
Method	NMSE↓	SNR↑	PCC↑	NMSE↓	SNR↑	PCC↑	NMSE↓	SNR↑	PCC↑
SEED									
STAD (Wang et al., 2024a)	0.6780	1.689	0.530	0.7623	1.179	0.450	0.7517	0.697	0.359
RDPI (Liu et al., 2025b)	0.5667	2.470	0.667	0.6345	1.979	0.625	0.7238	1.406	0.544
ESTFormer (Li et al., 2025)	0.5538	2.566	0.634	0.5499	2.597	0.635	0.6585	1.815	0.551
TopoDiff (Ours)	**0.4982**	**3.026**	**0.674**	**0.5447**	**2.639**	**0.639**	**0.6303**	**2.004**	**0.574**
SEED-IV									
STAD (Wang et al., 2024a)	0.4763	3.221	0.720	0.6555	1.834	0.559	0.7075	1.503	0.517
RDPI (Liu et al., 2025b)	0.3371	4.722	0.796	0.4117	3.854	0.730	0.4915	3.085	0.687
ESTFormer (Li et al., 2025)	0.2648	5.770	0.840	0.3111	5.071	0.813	0.3912	4.076	0.765
TopoDiff (Ours)	**0.2628**	**5.805**	**0.843**	**0.3091**	**5.099**	**0.814**	**0.3607**	**4.429**	**0.783**
MI/MM									
STAD (Wang et al., 2024a)	0.5705	2.438	0.871	0.5985	2.229	0.875	0.6677	1.754	0.675
RDPI (Liu et al., 2025b)	0.1686	7.651	0.872	0.2132	6.712	0.851	0.3457	4.600	0.795
ESTFormer (Li et al., 2025)	0.1788	7.477	0.902	0.2496	6.028	0.880	0.3085	5.108	0.845
TopoDiff (Ours)	**0.1591**	**7.983**	**0.912**	**0.1880**	**7.258**	**0.896**	**0.2326**	**6.334**	**0.871**

only to these unobserved channels, and the final super-resolved montage is obtained by combining the generated unseen channels with the observed low-resolution channels.

5 Experiments

5.1 Datasets and Preprocessing

We evaluate on four EEG benchmarks: **SEED** (Zheng & Lu, 2015; Duan et al., 2013), **SEED-IV** (Zheng et al., 2018), **TUSZ** (TUH EEG Seizure Corpus) (Shah et al., 2018), and **PhysioNet MI/MM** (Schalk et al., 2004; Goldberger et al., 2000). SEED contains 62-channel EEG from 15 subjects with three emotion labels (positive, neutral, negative) sampled at 200 Hz, while SEED-IV extends this setup to four emotion categories. MI/MM comprises 64-channel EEG sampled at 160 Hz for specific limb movements, and TUSZ provides large-scale clinical scalp EEG for seizure detection with standard 10–20 montages (at least 19 channels) and sampling rates of at least 250 Hz.

Table 2: Comparisons on TUSZ at $2\times$ super-resolution. Metrics include NMSE, SNR, and PCC for signal fidelity, and AUROC for seizure detection.

Method	NMSE↓	SNR↑	PCC↑	AUROC↑
Ground Truth	–	–	–	0.880
STAD	0.4519	3.450	0.634	0.861
RDPI	0.5263	2.790	0.701	0.560
ESTFormer	0.3600	4.437	0.732	0.871
TopoDiff (Ours)	**0.3285**	**4.835**	**0.760**	**0.880**

Collectively, these datasets span affective, motor, and clinical settings, enabling us to verify that the proposed method generalizes across diverse tasks and data collection conditions. For each dataset, we simulate wearable EEG configurations by subsampling channels from the original montage and treat the full montage as ground truth. Electrode coordinates are obtained from the standard 10–20 system (Klem et al., 1999) and used to construct topology-aware spatial representations. We use subject-based train/test splits to ensure performance reflects cross-subject generalization rather than within-subject memorization.

Preprocessing. EEG data was first segmented into 4s then resampled to 200Hz (except for MI/MM in which we maintained the original 160Hz) to maintain homogeneity across all datasets. Then, a bandpass filter of 0-75Hz was applied to remove high frequency noise (Zheng & Lu, 2015). Finally, the amplitudes of the EEG segments were clamped to $\pm 200 \mu v$ to prevent extreme outliers and noise.

Wearable Subsampling Protocol. We define super-resolution factors of $2\times$, $4\times$, and $8\times$ by selecting sparse subsets of channels from the full montage and reconstructing the full set. Unlike data-driven channel selection techniques that optimize for specific tasks (Tam et al., 2011; Gaur et al., 2021; Pane et al., 2018), the SR subsampling was done to ensure that the subsampled channels had adequate

Table 3: Downstream task accuracy (Acc) across datasets and SR factors. Ground Truth denotes downstream evaluation using the full real recordings.

Method	SEED (Emotion Recognition)			SEED-IV (Emotion Recognition)			PhysioNet MI/MM (Motor Imagery)		
Settings	2× SR	4× SR	8× SR	2× SR	4× SR	8× SR	2× SR	4× SR	8× SR
Ground Truth	0.5817	0.5817	0.5817	0.3689	0.3689	0.3689	0.6910	0.6910	0.6910
STAD	0.5131	0.4988	0.3915	0.3441	0.2746	0.2410	0.6836	0.6340	0.4938
RDPI	0.5193	0.5055	0.4636	0.3117	0.3034	0.2981	0.6221	0.5600	0.5097
ESTFormer	0.5546	0.5241	0.4690	0.3444	0.3094	0.2940	0.6700	0.6360	0.5980
TopoDiff (Ours)	**0.5624**	**0.5560**	**0.4798**	**0.3518**	**0.3185**	**0.3157**	**0.6840**	**0.6409**	**0.6251**

scalp coverage. The sampling followed the symmetrical positions commonly used in the 10-5, 10-10, and 10-20 systems (Michel & Murray, 2012), which ensure that the subsampled channels maintained adequate scalp coverage. With adequate scalp coverage corresponding to the underlying cortical regions, the selected channels can thus be used to capture the large-scale brain dynamics and distributed cortical activity (Jurcak et al., 2007). The channels selected using this paradigm are shown in the Appendix (Table 5).

5.2 Implementation Details

We implement our method with PyTorch, and all experiments are conducted on four NVIDIA RTX PRO 6000 GPUs. For all models, we use the AdamW optimizer with a learning rate of 5×10^{-4} and a weight decay of 0.01, together with a cosine learning rate scheduler. The diffusion Transformer consists of $L = 4$ layers with a hidden dimension of 800. Input signals are patchified with a patch size of $p = 50$ (and $p = 40$ for MI/MM to match the slightly different frequency), and all models are trained for 300 epochs. For global topological feature extraction, we adopt DINOv3 (Siméoni et al., 2025) and apply it to topographic visualizations constructed from temporally averaged EEG signals. For dynamic graph construction, we utilized $n_s = 12, k = 4$ for the $2\times$ factor and $n_s = 12, k = 6$ for $4\times$ and $8\times$ factors (further visualization in the Appendix). Further implementation details can be found in Appendix C.

5.3 Baselines

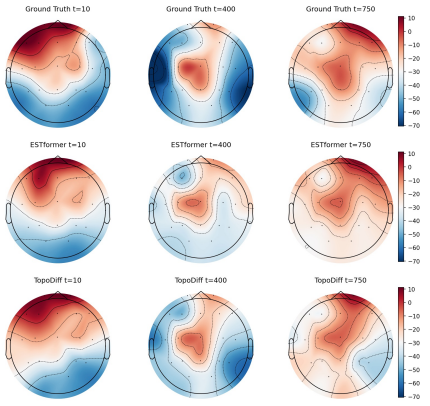

Figure 3: Topographical visualization of Ground Truth, ESTFormer, and TopoDiff super-resolution outputs for the TUSZ 2× setting at 3 different time steps.

For EEG spatial SR, we compare against **ESTFormer** (Li et al., 2025) and **STAD** (Wang et al., 2024a) as strong EEG super-resolution baselines, and include **RDPI** (Liu et al., 2025b) as a representative diffusion-based method for general spatiotemporal imputation. All methods are evaluated under identical super-resolution factors ($2\times$, $4\times$, $8\times$) using the same channel selections and identical subject-based train/test splits. We use the authors' official implementations when available; otherwise, we provide faithful reimplementations and cross-check key architectural and training hyperparameters against the original papers.

For downstream evaluation, we utilize three recent EEG recognition models: **DMMR** (Wang et al., 2024b) for SEED and SEED-IV, **3D-CLMI** (Hao & Cheng, 2025) for MI/MM, and **EvoBrain** (Kotoge et al., 2025) for TUSZ. Using fixed downstream architectures isolates the effect of super-resolution quality, allowing us to quantify how improved reconstruction fidelity translates into gains in cross-subject decoding performance in realistic deployment scenarios with sparse wearable EEG. Each downstream model is trained and tested on the full electrode set at the corresponding SR factor, combining the original recordings with our generated channels. This matches how the method would be used in practice: a low-density system is upsampled to a denser montage.

5.4 Evaluation Metrics

We report signal-level reconstruction fidelity and downstream task performance. For better model comparison, fidelity is evaluated on unseen/generated channels using NMSE (normalized mean squared error; lower is better), SNR (signal-to-noise ratio between reconstructed and ground-truth signals; higher is better), and PCC (Pearson correlation coefficient capturing waveform similarity; higher is better). These decoder-agnostic metrics provide assessments of super-resolution quality across datasets and super-resolution factors.

For downstream evaluation, we quantify how reconstruction quality translates to end-task utility: emotion recognition on SEED/SEED-IV, movement (motor imagery) classification on PhysioNet MI/MM, and seizure detection on TUSZ. We report classification accuracy at each super-resolution factor ($2\times$, $4\times$, $8\times$). For TUSZ, since there are only 19 channels, we report metrics under $2\times$ only and utilize AUROC for downstream task performance following (Kotoge et al., 2025).

5.5 Main Results

Table 1 and Table 2 summarize signal-level fidelity on SEED/SEED-IV, PhysioNet MI/MM, and TUSZ across multiple super-resolution (SR) factors. Across all datasets and SR levels, our TopoDiff framework achieves the best overall reconstruction quality, consistently improving NMSE, SNR, and PCC relative to strong baselines (ESTFormer, RDPI, STAD). The gains are most pronounced in the challenging $8\times$ regime: compared to the best competing method, we improve SNR by **+10.4%** on SEED, **+8.7%** on SEED-IV, and **+24.0%** on MI/MM, alongside substantial NMSE reductions (e.g., MI/MM: 0.2326 vs. 0.3085). Lastly, as seen in Table 2, TopoDiff achieves the lowest NMSE and highest SNR/PCC, with a **+9.0%** increase in SNR overall. We further evaluate stability across five random seeds and observe consistent performance across all datasets and metrics (see Appendix E).

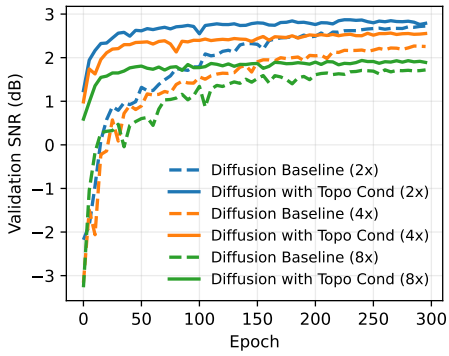

Figure 4: Convergence behavior with topological conditioning on the SEED dataset under $2\times$, $4\times$, and $8\times$ super-resolution settings.

Downstream Performance. Importantly, these signal-level gains translate into consistently improved downstream decoding (Table 3). TopoDiff achieves the highest accuracy among SR methods across all three datasets and all SR factors. On SEED, it improves over the strongest baseline from 0.5546 to 0.5624 at $2\times$ SR, from 0.5241 to 0.5560 at $4\times$ SR (**+6.1%** relative), and from 0.4690 to 0.4798 at $8\times$ SR. A similar pattern holds on SEED-IV, where TopoDiff improves accuracy at $2\times$, $4\times$, and $8\times$ SR, with the largest relative gain at $8\times$ (**+5.9%**). On PhysioNet MI/MM, TopoDiff remains close to full-data performance at $2\times$ SR (0.6840 vs. 0.6910) and increasingly outperforms the strongest baseline as channel sparsity increases, reaching 0.6251 at $8\times$ compared with 0.5980 for ESTFormer (**+4.5%**). Overall, these results indicate that the reconstruction gains preserve class-discriminative information relevant to downstream inference, including under substantial channel sparsity. This trend also extends to clinical seizure detection on TUSZ at $2\times$ SR (Table 2), where TopoDiff improves AUROC from 0.871 for ESTFormer to 0.880, matching the performance obtained from the full recordings.

Visualization. Beyond quantitative gains, Figure 3 provides qualitative evidence that our reconstructions better preserve the spatial structure of the underlying EEG. The topographic maps show that TopoDiff reproduces regionally consistent patterns that more closely align with the Ground Truth, whereas ESTFormer exhibits larger deviations in both the location and magnitude of activity. This closer agreement in spatial distribution suggests that TopoDiff recovers inter-channel relationships more faithfully, yielding signal distribution that are better matched to the original signals. More visualizations are in the Appendix.

5.6 Ablation Studies

Effect of Graph Relation Modeling. Removing graph relation modeling causes a clear drop in reconstruction fidelity. As shown in Table 4, excluding the graph and relying only on topology

conditioning (Diffusion + Topo) reduces SNR by 0.206, and similarly degrades NMSE and PCC. This suggests that the graph module provides essential dynamic inter-channel conditioning, enabling the diffusion model to capture time-varying electrode couplings that are not fully recoverable from global montage context alone.

Effect of Topological Conditioning. As shown in Table 4, under the $2\times$ setting, incorporating topological conditioning improves the SNR by 0.07, demonstrating its effectiveness in facilitating spatial generation. Although topographic representations are coarse and do not perfectly represent the underlying EEG signals, they provide complementary global geometric context that is difficult to obtain from discrete channel signals alone. And pretrained models can provide sufficient spatial conditioning information. To further analyze how this global topology-aware conditioning influences learning dynamics, we plot the validation SNR over the course of training in Fig. 4. With the global geometric prior provided by the topographic representations, the diffusion model converges significantly faster than the vanilla model. Together, the convergence behavior in Fig. 4 and the performance gains in Table 4 suggest that the topographic branch acts as a useful spatial prior, guiding the diffusion model toward more geometry-consistent generation.

Impact of Sparsity Factors. We evaluate several sparsity settings for the two graph-pruning controls in our construction (Table 4). Overall, graph sparsity materially affects reconstruction quality: overly dense graphs introduce noisy, weakly relevant couplings, while overly sparse graphs remove informative spatial interactions, and both regimes degrade signal fidelity. Across configurations, moderate sparsity provides the best trade-off. In particular, using a sufficiently rich candidate pool with controlled pruning (n_s=12, top-k=4; see Sec. 4.2) achieves the strongest performance, outperforming denser or less selective settings. This suggests that preserving adaptivity is critical for stable and spatially consistent super-resolution.

Table 4: Ablation, graph construction, GNN-depth sensitivity and topological conditioning studies on the SEED dataset under the $2\times$ super-resolution setting.

Method / Setting	NMSE↓	SNR↑	PCC↑
Component Ablation			
Diffusion	0.5307	2.752	0.650
Diffusion w/ Topo	0.5224	2.820	0.656
Diffusion w/ Topo + Graph	**0.4982**	**3.026**	**0.674**
Graph Construction (w/o Topo)			
n_s=14, k=8	0.5185	2.852	0.659
n_s=6, k=all	0.5062	2.957	0.669
n_s=20, k=4	0.5248	2.800	0.654
n_s=12, k=all	0.5190	2.848	0.659
n_s=12, k=4 (default)	0.5012	3.000	0.672
GNN Depth (w/ Topo, n_s=12, k=4)			
1 layer	0.5268	2.783	0.653
2 layers (default)	0.4982	3.026	0.674
4 layers	0.5199	2.841	0.658
Topo Conditioning (w/ Graph)			
Learnable CNN	0.5090	2.930	0.667
Pretrained DINOv3	0.4982	3.026	0.674

Sensitivity to GNN Depth. We further analyze the sensitivity of the model to GNN depth under the SEED $2\times$ setting. As shown in Table 4, our chosen 2-layer GNN with n_s=12 and k=4 achieves the best performance. A single GNN layer underperforms due to limited relational modeling capacity, while 4 layers leads to a slight degradation, likely because deeper message passing can over-smooth channel representations. These results indicate that a moderate GNN depth provides the best balance between relational expressiveness and feature discriminability.

6 Conclusion and Discussion

Our results highlight that *spatial representation* is a key bottleneck for EEG super-resolution. Rather than treating electrode coordinates as passive positional encodings, our approach conditions a diffusion model on **global topology-aware topoplot embeddings** and **channel-relationship graphs**, allowing the denoising process to jointly respect scalp geometry and structured inter-electrode dependencies. Additionally, we highlight the importance of dynamic conditioning for EEG time series which reflects the change in channel relationships across time. In the $2\times$, $4\times$, and $8\times$ super-resolution settings, we observe consistent gains in reconstruction quality and downstream decoding accuracy, indicating that the proposed dual conditioning remains effective as the upsampling task becomes more difficult. Limitations and future work can be found in Appendix A.

AI USE STATEMENT

In this work, we used generative AI tools to polish and aid in the writing process and discover relevant papers. We have read and checked all edits and content, and all scientific work was done and verified by the authors. We have reviewed all AI-assisted work and take responsibility for the content, claims, text and research artifacts in this paper.

REPRODUCIBILITY STATEMENT

We describe all the datasets, preprocessing, channel selection protocol, model architecture, training settings, baselines, evaluation metrics in our paper in Section 5. The appendix provides more details, such as pseudocode needed to reproduce the experimental results. It also discusses more dataset details and how we conduct preprocessing.

REFERENCES

Michael Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

Andrea Biondi, Sara K. Simblett, Pedro F. Viana, Petroula Laiou, Anna M.G. Fiori, Ewan Nurse, Martijn Schreuder, Deb K. Pal, and Mark P. Richardson. Feasibility and acceptability of an ultra-long-term at-home eeg monitoring system (eeg@home) for people with epilepsy. *Epilepsy & Behavior*, 151:109609, 2024. ISSN 1525-5050. doi: https://doi.org/10.1016/j.yebeh.2023.109609. URL https://www.sciencedirect.com/science/article/pii/S1525505023005280.

Clemens Brunner, Martin Billinger, Martin Seeber, Timothy R. Mullen, and Scott Makeig. Volume conduction influences scalp-based connectivity estimates. *Frontiers in Computational Neuroscience*, 10:121, 2016. doi: 10.3389/fncom.2016.00121.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, Oct 2023. ISSN 2052-4463. doi: 10.1038/s41597-023-02650-w. URL https://doi.org/10.1038/s41597-023-02650-w.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 6572–6583, Red Hook, NY, USA, 2018. Curran Associates Inc.

Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon, and Sung Chan Jun. Eeg datasets for motor imagery brain–computer interface. *GigaScience*, 6(7):gix034, 05 2017. ISSN 2047-217X. doi: 10.1093/gigascience/gix034. URL https://doi.org/10.1093/gigascience/gix034.

Jaeso Cho, Young Jun Ko, Yoon Gi Chung, Anna Cho, and Hunmin Kim. Improving epilepsy monitoring using long-term, in-home-bi-modal seizure monitoring device: clinical utilities and obstacles from a pilot study. *Frontiers in Neurology*, Volume 16 - 2025, 2025. ISSN 1664-2295. doi: 10.3389/fneur.2025.1609838. URL https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2025.1609838.

Andrea Cini, Ivan Marisca, and Cesare Alippi. Filling the g_ap_s: Multivariate time series imputation by graph neural networks. In *International Conference on Learning Representations*, 2022.

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. IEEE, 2018.

Hristos S Courellis, John R Iversen, Howard Poizner, and Gert Cauwenberghs. Eeg channel interpolation using ellipsoid geodesic length. In *2016 IEEE biomedical circuits and systems conference (BioCAS)*, pp. 540–543. IEEE, 2016.

Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, 2021. ISBN 9781713845393.

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for EEG-based emotion classification. In *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 81–84. IEEE, 2013.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, and Robin Rombach. Scaling rectified flow transformers for high-resolution image synthesis. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Pramod Gaur, Karl McCreadie, Ram Bilas Pachori, Hui Wang, and Girijesh Prasad. An automatic subject specific channel selection method for enhancing motor imagery classification in eeg-bci using correlation. *Biomedical Signal Processing and Control*, 68:102574, 2021.

Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. Meg and eeg data analysis with mne-python. *Frontiers in Neuroinformatics*, 7:267, 2013.

Kapil Gururangan, Babak Razavi, and Josef Parvizi. Diagnostic utility of eight-channel eeg for detecting generalized or hemispheric seizures and rhythmic periodic patterns. *Clinical Neurophysiology Practice*, 3:65–73, 2018. doi: 10.1016/j.cnp.2018.03.001.

Yuejiang Hao and Shiwei Cheng. Motor imagery eeg classification method using 3d cnn and lstm for rehabilitation application. *Cognitive Neurodynamics*, 19(1):131, Aug 2025.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Susan T. Herman, Nicholas S. Abend, Thomas P. Bleck, Kevin E. Chapman, Frank W. Drislane, Ronald G. Emerson, Elizabeth E. Gerard, Cecil D. Hahn, Aatif M. Husain, Peter W. Kaplan, Suzette M. LaRoche, Marc R. Nuwer, Mark Quigg, James J. Riviello, Sarah E. Schmitt, Liberty A. Simmons, Tammy N. Tsuchida, and Lawrence J. Hirsch. Consensus statement on continuous eeg in critically ill adults and children, part ii: Personnel, technical specifications, and clinical practice. *Journal of Clinical Neurophysiology*, 32(2), 2015. ISSN 0736-0258. URL https://journals.lww.com/clinicalneurophys/fulltext/2015/04000/consensus_statement_on_continuous_eeg_in.2.aspx.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *ArXiv*, abs/2207.12598, 2022.

Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, 34(4):1600–1611, 2007.

Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1):98–107, 2018. doi: 10.1109/JBHI.2017.2688239.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

George H. Klem, Hans Lüders, H. H. Jasper, and C. Elger. The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, 52:3–6, 1999.

Zoly J. Koles and R. B. Paranjape. Topographic mapping of the eeg: An examination of accuracy and precision. *Brain Topography*, 1(2):87–95, Dec 1988. ISSN 1573-6792. doi: 10.1007/BF01129173. URL https://doi.org/10.1007/BF01129173.

Rikuto Kotoge, Zheng Chen, Tasuku Kimura, Yasuko Matsubara, Takufumi Yanagisawa, Haruhiko Kishima, and Yasushi Sakurai. Evobrain: Dynamic multi-channel EEG graph modeling for time-evolving brain networks. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

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

Tianhong Li and Kaiming He. Back to basics: Let denoising generative models denoise. *arXiv preprint arXiv:2511.13720*, 2025.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Step-aware residual-guided diffusion for eeg spatial super-resolution. *ArXiv*, abs/2510.19166, 2025a.

Zijin Liu, Xiang Zhao, and You Song. Rdpi: a refine diffusion probability generation method for spatiotemporal data imputation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025b. ISBN 978-1-57735-897-8. doi: 10.1609/aaai.v39i12.33335. URL https://doi.org/10.1609/aaai.v39i12.33335.

Karmen Markov, Mohamed Elgendi, and Carlo Menon. Eeg-based headset sleep wearable devices. *npj Biosensing*, 1(1):12, Oct 2024. ISSN 3004-8656. doi: 10.1038/s44328-024-00013-y. URL https://doi.org/10.1038/s44328-024-00013-y.

Christoph M Michel and Micah M Murray. Towards the utilization of eeg as a brain imaging tool. *Neuroimage*, 61(2):371–385, 2012.

Guido Nolte, Ou Bai, Lewis Wheaton, Zoltan Mari, Sherry Vorbach, and Mark Hallett. Identifying true brain interaction from eeg data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307, 2004. doi: 10.1016/j.clinph.2004.04.029.

Paul L. Nunez and Ramesh Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2 edition, 2006. ISBN 9780195050387. doi: 10.1093/acprof:oso/9780195050387.001.0001.

Evi Septiana Pane, Adhi Dharma Wibawa, and Mauridhi Hery Pumomo. Channel selection of eeg emotion recognition using stepwise discriminant analysis. In *2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, pp. 14–19. IEEE, 2018.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4195–4205, 2023.

François Perrin, Jacques Pernier, Olivier Bertrand, and Jean Francois Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, 72(2):184–187, 1989.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In *International Conference on Machine Learning*, 2021. URL https://api.semanticscholar.org/CorpusID:231719657.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, Los Alamitos, CA, USA, June 2022. IEEE Computer Society. doi: 10.1109/CVPR52688.2022.01042. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Raphael Gontijo-Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, 2022. ISBN 9781713871088.

Gerwin Schalk, Dennis Mcfarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Trans. Biomed. Eng.*, 51: 1034–, 07 2004. doi: 10.1109/TBME.2004.827072.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

Vinit Shah, Eva Von Weltin, Silvia Lopez, James Riley McHugh, Lillian Veloso, Meysam Golmohammadi, Iyad Obeid, and Joseph Picone. The temple university hospital seizure detection corpus. *Frontiers in neuroinformatics*, 12:83, 2018.

Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009.

Kai Shu, Le Wu, Yuchang Zhao, Aiping Liu, Ruobing Qian, and Xun Chen. Data augmentation for seizure prediction with generative diffusion model. *IEEE Transactions on Cognitive and Developmental Systems*, 17(3):577–591, 2025. doi: 10.1109/TCDS.2024.3489357.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022. URL https://arxiv.org/abs/2010.02502.

Mats Svantesson, Håkan Olausson, Anders Eklund, and Magnus Thordstein. Virtual eeg-electrodes: Convolutional neural networks as a method for upsampling or restoring channels. *Journal of Neuroscience Methods*, 355:109126, 2021.

Wing-Kin Tam, Zheng Ke, and Kai-Yu Tong. Performance of common spatial pattern under a smaller set of eeg electrodes in brain-computer interface on chronic stroke patients: a multi-session dataset study. In *2011 annual international conference of the IEEE engineering in medicine and biology society*, pp. 6344–6347. IEEE, 2011.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. Csdi: Conditional score-based diffusion models for probabilistic time series imputation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24804–24816. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/cfe8504bda37b575c70ee1a8276f3486-Paper.pdf.

Muhang Tian, Bernie Chen, Allan Guo, Shiyi Jiang, and Anru R Zhang. Reliable generation of privacy-preserving synthetic electronic health record time series via diffusion models. *Journal of the American Medical Informatics Association*, 31(11):2529–2539, 09 2024. doi: 10.1093/jamia/ocae229.

Frederik Van de Steen, Luca Faes, Esin Karahan, Jitkomut Songsiri, Pedro A. Valdes-Sosa, and Daniele Marinazzo. Critical comments on eeg sensor space dynamical connectivity analysis. *Brain Topography*, 32(4):643–654, 2019. doi: 10.1007/s10548-016-0538-7.

Ronca Vincenzo, Cecchetti Marianna, Capotorto Rossella, Di Flumeri Gianluca, Giorgi Andrea, Germano Daniele, Borghini Gianluca, Babiloni Fabio, and Arico Pietro. Beyond the lab: real-world benchmarking of wearable eegs for passive brain-computer interfaces. *Brain Informatics*, 13, 2025. URL https://api.semanticscholar.org/CorpusID:284309128.

Michaela Waak, Joshua Laing, Lakshmi Nagarajan, Nicholas Lawn, and A. Harvey. Continuous electroencephalography in the intensive care unit: A critical review and position statement from an australian and new zealand perspective. *Critical Care and Resuscitation*, 25:9–19, 03 2023. doi: 10.1016/j.ccrj.2023.04.004.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

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*, 71:1034–1045, 2024a.

Yiming Wang, Bin Zhang, and Yujiao Tang. Dmmr: Cross-subject domain generalization for eeg-based emotion recognition via denoising mixed mutual reconstruction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1):628–636, Mar. 2024b. doi: 10.1609/aaai.v38i1.27819. URL https://ojs.aaai.org/index.php/AAAI/article/view/27819.

Haoran Wei, Yaofeng Sun, and Yukun Li. Deepseek-ocr: Contexts optical compression. *arXiv preprint arXiv:2510.18234*, 2025.

Xiaoyan Wei, Tamir Avigdor, Alyssa Ho, Erica Minato, Alfonso Garcia-Asensi, Jessica Royer, Yingqi Laetitia Wang, Vojtech Travnicek, Katharina Schiller, Boris C. Bernhardt, and Birgit Frauscher. Anphy-sleep: an open sleep database from healthy adults using high-density scalp electroencephalogram. *Scientific Data*, 11(1):896, Aug 2024. ISSN 2052-4463. doi: 10.1038/s41597-024-03722-1. URL https://doi.org/10.1038/s41597-024-03722-1.

M. Brandon Westover, Kapil Gururangan, Matthew S. Markert, Benjamin N. Blond, Saien Lai, Shawna Benard, Stephan Bickel, Lawrence J. Hirsch, and Josef Parvizi. Diagnostic value of electroencephalography with ten electrodes in critically ill patients. *Neurocritical Care*, 33(2):479–490, 2020. doi: 10.1007/s12028-019-00911-4.

Chad Williams, Daniel Weinhardt, Joshua Hewson, Martyna Beata Płomecka, Nicolas Langer, and Sebastian Musslick. Eeg-gan: A generative eeg augmentation toolkit for enhancing neural classification. *bioRxiv*, 2025.

Zhuoyi Yang, Jiayan Teng, Wendi Zheng, Ming Ding, Shiyu Huang, Jiazheng Xu, Yuanming Yang, Wenyi Hong, Xiaohan Zhang, Guanyu Feng, et al. Cogvideox: Text-to-video diffusion models with an expert transformer. *arXiv preprint arXiv:2408.06072*, 2024.

Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.

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.

A Limitations and Future Work

Several directions could further extend this framework. First, while we use topoplot embeddings as a compact global geometric prior, richer subject-specific spatial cues (e.g., individualized electrode coordinates, head models, or montage calibration) may improve anatomical plausibility and cross-subject transfer, allowing for more complete mapping to the brain. Second, our graph conditioning can be expanded to better capture state-dependent connectivity, for example by learning multi-scale or frequency-specific relational graphs and aligning them with known neurophysiological priors. Third, future work can leverage uncertainty for quality control (detecting unreliable reconstructions), active sensing (selecting which additional electrodes would most reduce uncertainty), or downstream-aware training (optimizing SR to maximize task performance). Finally, evaluating robustness under real-world noise, missing channels, and montage shifts enabling unified imputation, denoising, and cross-montage translation would broaden the practical impact of topology- and relation-grounded generative modeling for wearable EEG.

B Diffusion Parameterization and Inference Details

During training, we use the interpolation

$$z = tx + (1-t)\epsilon, \qquad \epsilon \sim \mathcal{N}(0, \mathbb{I}), \tag{1}$$

where $t = \mathrm{sigmoid}(\tau)$ and $\tau \sim \mathcal{N}(0, 1.2^2)$. The denoising network predicts the clean target $x_{\mathrm{pred}} = f_\theta(z, X^{\mathrm{LR}}, t)$. The corresponding velocity field is

$$v_\theta(z,t) = \frac{x_{\mathrm{pred}} - z}{1-t}. \tag{2}$$

At inference, we solve the reverse-time ODE using Euler integration:

$$z_{m+1} = z_m + \Delta t \cdot v_\theta(z_m, t_m). \tag{3}$$

B.1 Detailed Tensor Shapes

Let $T_p = T/P$, $N = C_{\mathrm{LR}} T_p$, and $N_z = C_{\mathrm{Unseen}} T_p$.

Signal patchification. The low-resolution EEG is patchified as

$$X^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T} \to X_p^{\mathrm{LR}} \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p \times P} \to X_c^{\mathrm{LR}} \in \mathbb{R}^{N \times P}. \tag{4}$$

The noisy target is similarly patchified as

$$z \in \mathbb{R}^{C_{\mathrm{Unseen}} \times T} \to z_c \in \mathbb{R}^{N_z \times P}. \tag{5}$$

Topological branch. For each temporal patch, the averaged signal is $\bar{X} \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p}$. Each slice is converted into a topographic image $I_{\mathrm{topo}} \in \mathbb{R}^{H \times W \times 3}$. The image encoder extracts $F_{\mathrm{topo}} \in \mathbb{R}^{h \times w \times d}$, which is flattened and projected to

$$\tilde{F}_{\mathrm{topo}} \in \mathbb{R}^{n \times D}, \qquad n = hw. \tag{6}$$

Graph branch. For each temporal group g, the input is $X^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times P}$. The graph adjacency is $A^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times C_{\mathrm{LR}}}$. After graph propagation, the output is $Z^{(g)} \in \mathbb{R}^{C_{\mathrm{LR}} \times D}$. Concatenating all temporal groups gives

$$Z \in \mathbb{R}^{C_{\mathrm{LR}} \times T_p \times D} \to Z_c \in \mathbb{R}^{N \times D}. \tag{7}$$

Transformer input. The Transformer input is the concatenation of topological tokens, graph-enhanced visible-channel tokens, and noisy unseen-channel tokens:

$$H = \mathrm{Cat}(\tilde{F}_{\mathrm{topo}}, Z_c, z_d) \in \mathbb{R}^{(n+N+N_z) \times D}. \tag{8}$$

Table 5: Channels selected for SR task based on symmetric channel selection.

Dataset	All Channels	2× SR	4× SR	8× SR
PhysioNet MI/MM (Schalk et al., 2004)	Fp1, Fpz, Fp2, AF7, AF3, AFz, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, Oz, O2, Iz	Fp1, Fp2, Fz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, TP7, TP8, CP1, CP2, CP5, CP6, P7, P8, Pz, P3, P4, PO3, PO4, Oz, O1, O2	Fp1, Fp2, F7, F8, T7, T8, F3, F4, C3, C4, P3, P4, P7, P8, O1, O2	F3, F4, C3, C4, P3, P4, O1, O2
SEED/SEED-IV (Zheng & Lu, 2015)	Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, CB1, O1, Oz, O2, CB2	Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, PO4, O1, Oz, O2	Fp1, Fp2, F7, F8, T7, T8, F3, F4, C3, C4, P3, P4, P7, P8, O1, O2	F3, F4, C3, C4, P3, P4, O1, O2
TUSZ (Shah et al., 2018)	Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, T3, T4, T5, T6	Fp1, Fp2, F3, F4, C3, C4, O1, O2, Cz	–	–

C Implementation Details

C.1 Dataset Details

SEED and SEED-IV. The SJTU Emotion EEG Dataset (**SEED**) (Zheng & Lu, 2015; Duan et al., 2013) is a standard benchmark for EEG-based emotion recognition. It contains recordings from 15 subjects watching emotion-eliciting video clips labeled with three emotional states: positive, neutral, and negative. EEG is collected using a 62-channel system following the international 10-20 electrode placement standard with a sampling rate of 200 Hz. **SEED-IV** (Zheng et al., 2018) extends SEED with four emotion categories (happiness, sadness, fear, and neutrality), while using the same acquisition setup as **SEED**.

Physionet MI/MM. The PhysioNet MI/MM dataset (Schalk et al., 2004; Goldberger et al., 2000) is a widely used benchmark for MI and motor execution research in EEG-based brain–computer interfaces. It consists EEG recordings from 109 subjects performing or imagining movements of the left hand, right hand, both hands, or both feet in response to visual cues, thus resulting in a three-class classification problem. EEG signals are recorded with a 64-channel international 10–20 montage at a sampling rate of 160 Hz.

TUSZ. TUSZ (Shah et al., 2018) is a large-scale clinical EEG dataset for seizure detection and classification, consisting of long-term scalp EEG recordings acquired during continuous clinical monitoring. It is a clinically annotated seizure-detection subset of TUEG. Recordings follow standard clinical 10–20 electrode montages (at least 19 channels) with sampling rates of at least 250 Hz.

Dataset versions and access. The PhysioNet MI/MM dataset was downloaded from PhysioNet as version 1.0.0, released on September 9, 2009, under DOI 10.13026/C28G6P. The TUSZ dataset was obtained from the TUH EEG Seizure Corpus release `v2.0.6`. The SEED and SEED-IV datasets were obtained through the BCMI/SJTU dataset application system under the corresponding license agreement. For SEED/SEED-IV, we report the official dataset name because the public application page does not provide explicit numerical release versions.

C.2 Channel Selection

Table 5 lists the original channels and the channels selected for SR tasks. The channels were selected to ensure that the scalp coverage remains similar to the original electrode configuration. As a result the channels maintain a symmetry. For TUSZ, only 2 × SR is considered as it had 19 channels in the full configuration. Furthermore, Figure 5 visualizes the EEG channels in different configurations, showcasing the symmetric coverage of the scalp.

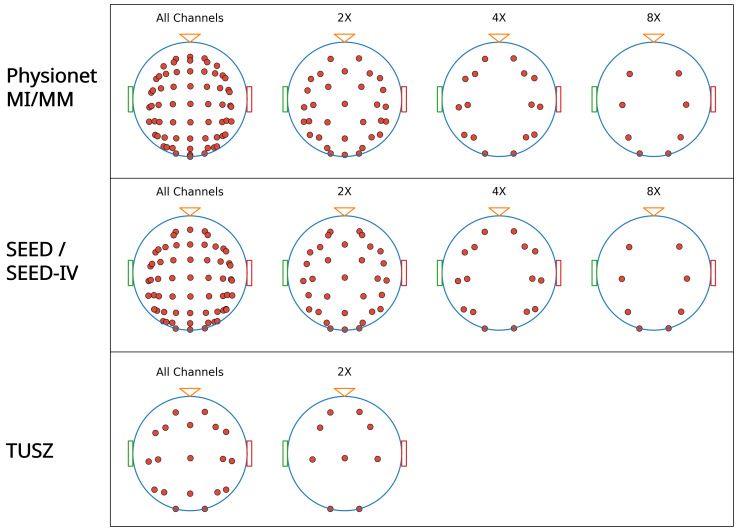

Figure 5: Visualization of EEG channels in full, 2 × SR, 4 × SR, and 8 × SR configurations.

C.2.1 Compute Resources

All experiments in this research are performed on four NVIDIA RTX PRO 6000 GPUs.

C.2.2 Topoplot rendering

To construct the topographic conditioning input, we first average each EEG clip along the temporal dimension, producing a channel-wise vector of size C_{LR}, where C_{LR} denotes the number of observed low-resolution electrodes. We render this vector as a scalp topoplot using the open-source MNE library. Specifically, we use MNE's default cubic interpolation to interpolate the observed electrode values onto a 64×64 internal spatial grid, which is subsequently rasterized as a $3 \times 96 \times 96$ RGB image using the `hot` colormap. Each rendered topoplot includes the standard head outline, markers indicating the observed sensor locations, and six contour lines. Thus, the topographic image provides a spatial representation of the observed EEG values while preserving the geometry of the electrode montage.

C.2.3 DINOv3 input preprocessing

The rendered topoplots are processed using the official DINOv3 image preprocessing pipeline. In particular, RGB values are normalized using the ImageNet channel-wise mean and standard deviation expected by the pretrained encoder. We do not apply random cropping, geometric transformations, color augmentation, or other data augmentation to the topoplots. This ensures that the spatial correspondence between scalp locations and image coordinates remains unchanged before feature extraction.

C.2.4 Electrode coordinate encoding

In addition to the topographic conditioning, the model is provided with the spatial coordinates of the complete electrode montage, including both observed and target electrodes. Importantly, the signal values of the target electrodes remain unavailable to the model; only their known physical

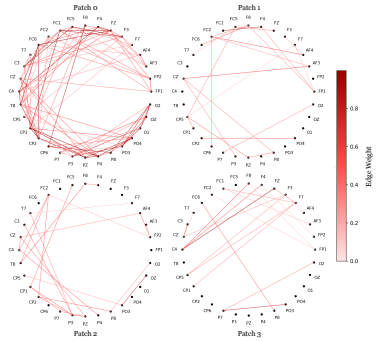

Figure 6: Examples of learned channel-relationship graphs with edge weights visualized by color intensity. The panels show the progression of a sample-specific connectivity pattern over the 10-20 montage at different temporal patches, showing only new edges when compared with the previous patch, highlighting how the relational graph adapts across time while preserving topology-consistent structure.

locations are provided. Following the coordinate-conditioning strategy used in ESTFormer, electrode coordinates are mapped through a cosine positional encoding and added to the corresponding channel representations. Providing the complete montage geometry informs the model of the locations at which missing signals should be reconstructed without exposing their ground-truth EEG values.

C.2.5 Subject-disjoint dataset splits

All experiments use subject-disjoint training and test partitions so that signals from a participant in the test set never appear during training. This protocol prevents subject-specific signal characteristics from leaking across splits and evaluates the ability of the model to generalize to unseen participants. We use 89/20 train/test subjects for MI/MM, 12/3 for SEED, 11/4 for SEED-IV, and 579/43 for TUSZ. For datasets containing multiple sessions from the same participant, all sessions associated with a participant are assigned to the same partition. The splits are randomly constructed at the subject level and then held fixed across the compared methods to ensure a consistent evaluation protocol.

C.2.6 Signal normalization and value scaling

For each dataset, EEG normalization statistics are computed exclusively from the training split to prevent information leakage from the test set. We normalize the signals using the resulting training-set mean and standard deviation. Signal amplitudes are additionally represented consistently in the microvolt (μV) scale across datasets before being passed to the model. The same normalization statistics and scaling procedure derived from the training data are subsequently applied to validation and test samples.

C.3 Graph Construction

Fig. 6 and Fig. 7 show visualizations of how the graphs may change over time patches, adding and removing different edges and shifting the weights of existing edges. Both figures map graph constructions for the SEED dataset at the $2\times$ factor, meaning there are 32 visible channels to potentially map connections between. They both demonstrate our strongest parameter setting for SEED $2\times$ which was $n_s = 12, k = 4$.

C.3.1 Choice of Cosine Similarity

We use absolute cosine similarity to construct the channel graph, as the goal is to capture the strength of dependency between electrodes regardless of whether their relationship is positive or negative. This also ensures a nonnegative adjacency matrix, consistent with the degree-normalized GCN used in TopoDiff. Directly retaining signed similarities would require modifying the current graph formulation to explicitly handle positive and negative edges. We therefore leave dedicated signed-graph modeling as an interesting direction for future work.

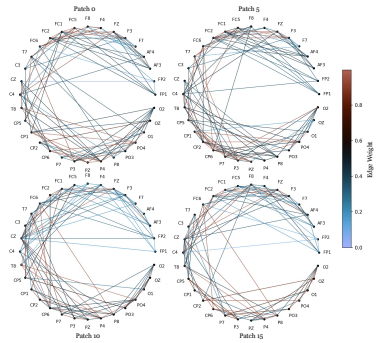

Figure 7: Graph construction over 4 temporal patches with $n_s = 12$ and $k = 4$ on the SEED 2× channel setting.

D Training Pipeline

We show the full TopoDiff training pipeline pseudocode in Algorithm 1.

Algorithm 1 TopoDiff Training Pipeline

Require: Low-resolution EEG X^{LR}, high-resolution EEG X^{HR}, patch size P, graph parameters n_s, k

Initialize model parameters θ
for each training epoch **do**
  Split X^{HR} into visible channels X^{LR} and target unseen channels x
  Sample $t = \mathrm{sigmoid}(\tau)$, $\tau \sim \mathcal{N}(0, 1.2^2)$
  Sample $\epsilon \sim \mathcal{N}(0, \mathbb{I})$
  Construct noisy target $z = tx + (1-t)\epsilon$
  Patchify X^{LR} and z
  for each temporal group g **do**
    Construct topographic image from averaged EEG patch
    Extract topological embedding $\tilde{F}_{\mathrm{topo}}$
    Compute cosine similarity adjacency $A^{(g)}$
    Apply spatial mask, top-k pruning, symmetrization, self-loops, and normalization
    Compute graph-enhanced representation $Z^{(g)}$
  end for
  Concatenate topological tokens, graph-enhanced EEG tokens, and noisy target tokens
  Predict clean unseen channels $x_{\mathrm{pred}} = f_\theta(z, X^{\mathrm{LR}}, t)$
  Update θ using $\|x_{\mathrm{pred}} - x\|_2^2$
end for

E Stability Across Random Seeds

We evaluate the stability of TopoDiff by training each model with five different random seeds and reporting the mean and standard deviation across all runs. We assess both signal-level reconstruction fidelity (NMSE, SNR, PCC) and downstream task performance (Acc) across multiple datasets and super-resolution (SR) factors.

Overall stability. As shown in Table 6, Table 7, and Table 8, TopoDiff exhibits highly consistent reconstruction performance across all datasets and SR settings. In particular, the standard deviations of NMSE, SNR, and PCC remain extremely small across five random seeds, indicating that the diffusion-based reconstruction is robust to initialization and stochastic training dynamics. Downstream accuracy shows somewhat greater variability, as expected for task-level evaluation, but remains generally stable across seeds while consistently favoring TopoDiff in mean performance.

Table 6: Performance across 5 random seeds (mean ± std) under $2\times$ SR.

Dataset	Method	NMSE↓	SNR↑	PCC↑	Acc↑
SEED	Ours	0.4981±0.0002	3.0266±0.0018	0.6742±0.0004	0.5667±0.0218
MI/MM	Ours	0.1590±0.0001	7.9854±0.0018	0.9126±0.0005	0.6836±0.0060
SEED-IV	Ours	0.2628±0.0000	5.8036±0.0011	0.843±0.0000	0.3525±0.0029
SEED	RDPI	0.6023±0.0199	2.2042±0.1486	0.6423±0.0138	0.5123±0.0024
MI/MM	RDPI	0.1696±0.0030	7.6900±0.0764	0.8824±0.0057	0.6298±0.1286
SEED-IV	RDPI	0.3371±0.0001	4.7259±0.0043	0.7954±0.0026	0.3169±0.0155

Table 7: Performance across 5 random seeds (mean ± std) under $4\times$ SR.

Dataset	Method	NMSE↓	SNR↑	PCC↑	Acc↑
SEED	Ours	0.5447±0.0002	2.639±0.0012	0.639±0.0000	0.5473±0.0077
MI/MM	Ours	0.188±0.0001	7.2592±0.0013	0.8962±0.0004	0.6524±0.0046
SEED-IV	Ours	0.3094±0.0002	5.0952±0.0031	0.814±0.0000	0.3346±0.0210
SEED	RDPI	0.6371±0.0015	1.9582±0.0116	0.6247±0.0002	0.5052±0.0006
MI/MM	RDPI	0.2142±0.0006	6.6911±0.0117	0.8196±0.0096	0.5597±0.0639
SEED-IV	RDPI	0.4092±0.0050	3.8674±0.0251	0.7360±0.0089	0.3034±0.0011

We also evaluated pointwise averaging of reconstructions across the five runs. This reduced SNR by approximately 0.3 dB, likely because small phase or temporal differences among plausible waveform components introduce distortion upon averaging. We therefore retain output diversity, which reflects the inherent ambiguity of EEG super-resolution and the model's ability to represent multiple plausible reconstructions.

Comparison with RDPI. Compared with TopoDiff, RDPI exhibits substantially greater variability in several reconstruction settings, including SEED and MI/MM, with particularly pronounced variation in NMSE and SNR. RDPI also shows large accuracy variance on MI/MM, especially under $2\times$ and $4\times$ SR. Overall, these results indicate that TopoDiff provides not only stronger average reconstruction and downstream performance, but also highly reproducible signal-level results across random seeds.

Clinical robustness (TUSZ). On the TUSZ dataset (Table 9), TopoDiff maintains both strong performance and low variance in AUROC, indicating consistent clinical utility for seizure detection. The significantly lower variance compared to RDPI further highlights the robustness of the proposed approach in real-world settings.

F Empirical Results

Fig. 8 and Fig. 9 further show some results from our super-resolution model when compared at specific time steps to the ground truth and other baseline model generated signals.

G Sensitivity to Electrode Configurations

We evaluate different sparse electrode configurations on SEED under the $2\times$ SR setting (Table 10). The electrode configurations are depicted in Fig. 10. Balanced configurations with broad scalp coverage perform better than imbalanced layouts that concentrate electrodes in a limited scalp region, which substantially degrade performance.

Table 8: Performance across 5 random seeds (mean ± std) under $8\times$ SR.

Dataset	Method	NMSE↓	SNR↑	PCC↑	Acc↑
SEED	Ours	0.6302±0.0001	2.005±0.0007	0.574±0.0000	0.4758±0.0139
MI/MM	Ours	0.2325±0.0001	6.3368±0.0018	0.871±0.0000	0.6226±0.0117
SEED-IV	Ours	0.3606±0.0001	4.4298±0.0013	0.783±0.0000	0.3162±0.0009
SEED	RDPI	0.7065±0.0533	1.3734±0.0182	0.5421±0.0011	0.4643±0.0011
MI/MM	RDPI	0.3405±0.0029	4.6761±0.0425	0.8005±0.0030	0.4938±0.0329
SEED-IV	RDPI	0.4716±0.0118	3.0972±0.0142	0.6878±0.0156	0.2983±0.0008

Table 9: Performance across 5 random seeds on TUSZ ($2\times$ SR).

Method	NMSE↓	SNR↑	PCC↑	AUROC↑
Ours	**0.3279±0.0020**	**4.8332±0.0091**	**0.7602±0.0004**	**0.8782±0.0033**
RDPI	0.5378±0.0100	2.6947±0.0826	0.6855±0.0134	0.5824±0.0731

H Robustness to Input Noise

We evaluate robustness by perturbing the low-resolution input channels with Gaussian noise at inference time. The noise standard deviation is defined relative to the per-channel signal scale. Results on SEED (Table 11) under $2\times$ SR show that the proposed sparsity setting degrades gracefully under input noise.

I Computational Efficiency

We analyze the computational efficiency of TopoDiff and compare it with representative baselines. All methods are evaluated under identical hardware and software settings using a single NVIDIA RTX PRO 6000 GPU.

Inference efficiency. A key advantage of TopoDiff is that it achieves high reconstruction quality using *single-step* diffusion at inference. Unlike conventional diffusion models that require many iterative denoising steps, our formulation solves the reverse-time ODE with a single Euler step, substantially reducing inference cost. TopoDiff requires approximately **0.00237s per sample** on a single NVIDIA RTX 6000 Pro GPU, slightly faster than the per-step latency of STAD (0.00261s) and substantially faster than RDPI (0.01380s per sample).

The topographic plots are precomputed, encoded once with DINOv3, and cached during training, so they introduce no repeated training-time overhead. At inference, plot generation is integrated into the data loader and is not a computational bottleneck: with a batch size of 64, DINOv3 encodes each $96 \times 96 \times 3$ image in approximately 0.0003s. Dynamic graphs are constructed online, but both graph construction and message passing are lightweight.

Importantly, the reported latency for STAD corresponds to a *single denoising step*; its full reverse-diffusion process requires many iterative steps, leading to substantially higher end-to-end inference time. In contrast, TopoDiff achieves strong performance without iterative refinement. TopoDiff also requires approximately 22 GB of GPU memory during inference, substantially less than RDPI (~ 52 GB), owing to the absence of iterative refinement and fewer intermediate activations.

Model size and training efficiency. TopoDiff contains approximately 68M trainable parameters, excluding the frozen DINOv3 encoder. With a batch size of 64, the complete model trains at approximately 1.1 iterations per second on a single NVIDIA RTX PRO 6000 GPU. To quantify the computational overhead introduced by topographic conditioning, we additionally measure training throughput after excluding the DINOv3 encoder and the associated topoplot-conditioning module. Under the same batch size and hardware configuration, throughput increases to approximately 1.35 iterations per second. Thus, incorporating the pretrained topographic representation introduces

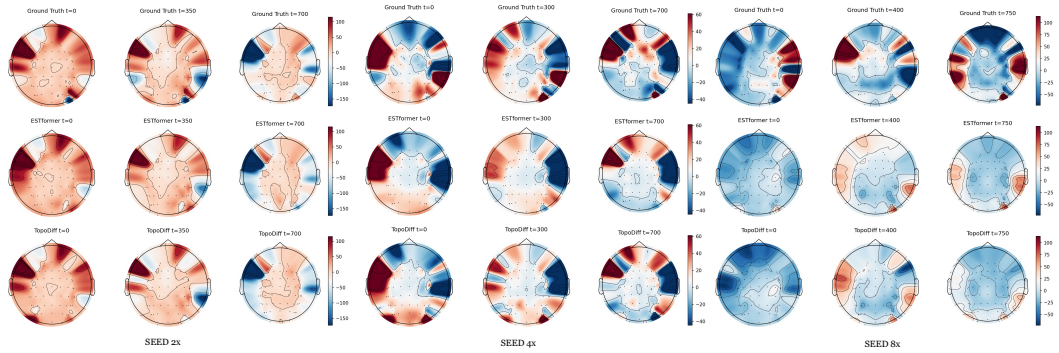

Figure 8: Topographical maps of Ground Truth, ESTFormer, and TopoDiff HR outputs for different SEED settings at 3 different time steps.

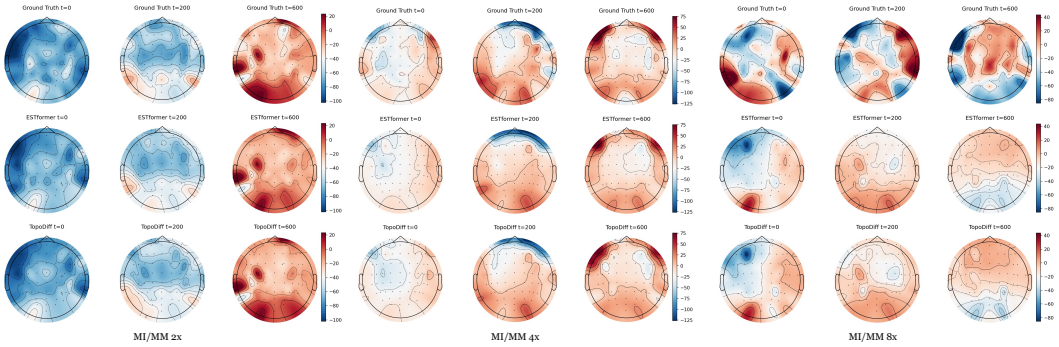

Figure 9: Topographical maps of Ground Truth, ESTFormer, and TopoDiff HR outputs for different Physionet MI/MM settings at 3 different time steps.

a moderate training-time overhead while leaving the remainder of the architecture comparatively lightweight.

Effect of sampling steps. We further evaluate the impact of increasing the number of diffusion steps in Table 12. Increasing the number of steps does not improve performance and instead introduces additional computational overhead. This empirically supports our design choice of single-step inference as an effective trade-off between efficiency and reconstruction quality.

Overall, TopoDiff achieves a favorable efficiency–performance trade-off by combining structured conditioning with single-step diffusion. By eliminating iterative denoising, it provides faster inference, lower memory usage, and competitive or superior reconstruction quality compared to both deterministic and diffusion-based baselines.

J Effect of Dynamic Graph-Based Relation Modeling

We analyze the contribution of the proposed dynamic graph module by isolating its effect on reconstruction performance. Specifically, we compare three variants: (i) a parameter-matched Transformer baseline without graph modeling, (ii) a static graph constructed using fixed electrode connectivity, and (iii) the proposed dynamic graph, which adapts inter-channel relations over time.

Results. As shown in Table 13, the dynamic graph consistently outperforms both the static graph and the no-graph baseline across all metrics. While the static graph captures coarse spatial proximity, it fails to adapt to time-varying dependencies between channels. The parameter-matched Transformer baseline performs the worst, indicating that the improvements do not arise from increased model capacity alone.

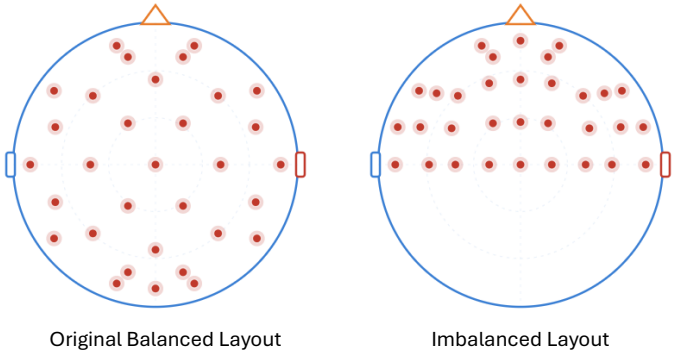

Figure 10: Comparison of electrode configuration in balanced and imbalanced format.

Table 10: Sensitivity to sparse electrode configurations on SEED under 2× SR.

Configuration	NMSE↓	SNR↑	PCC↑
Original balanced layout	0.4982	3.026	0.674
Imbalanced layout	0.7917	1.015	0.444

The results in Table 13 highlight that explicitly modeling *dynamic* inter-channel relationships is critical for EEG spatial super-resolution. By updating graph connectivity based on signal similarity at each temporal group, the proposed approach captures evolving functional relationships between electrodes, leading to more coherent multi-channel reconstruction.

K Generated Channels Complement Sparse Layout

We investigate whether generated channels merely reconstruct missing signals or provide additional utility beyond the original sparse input. To this end, we compare downstream performance using: (i) only the original low-resolution channels, and (ii) an expanded montage that includes both original and generated channels.

Results. On the MI/MM dataset under the $2\times$ setting, using the original 32 channels achieves 67.1% accuracy, while expanding to 64 channels using TopoDiff-generated signals improves performance to 68.4%.

This result suggests that the generated channels are not simply redundant reconstructions, but instead provide complementary spatial information that enhances downstream representations. One possible explanation is that the model learns structured spatial priors during training, allowing it to synthesize channels that capture meaningful spatial patterns beyond those present in the sparse input.

More broadly, this finding indicates that spatial super-resolution can act as a form of *representation augmentation*, improving task performance even when original measurements are available.

L Analysis on Topographic Conditioning

We analyze the role of topographic conditioning by comparing it with standard coordinate-based positional encoding. While positional encodings provide explicit electrode location information, they do not capture global spatial structure or smooth spatial patterns across the scalp.

Setup. We evaluate three variants: (i) diffusion without positional encoding, (ii) diffusion with positional encoding, and (iii) TopoDiff with topographic conditioning.

Table 11: Robustness to Gaussian noise on SEED under $2\times$ SR.

Setting	Noise STD	NMSE↓	SNR↑	PCC↑
Ours ($n_s = 12, k = 4$)	0	0.4980	3.028	0.675
Ours ($n_s = 12, k = 4$)	0.05	0.4984	3.024	0.674
Ours ($n_s = 12, k = 4$)	0.10	0.5000	3.010	0.673
$n_s = 12, k = 12$	0	0.5190	2.848	0.659
$n_s = 12, k = 12$	0.05	0.5196	2.843	0.658
$n_s = 12, k = 12$	0.10	0.5212	2.830	0.657
$n_s = 20, k = 4$	0	0.5248	2.800	0.654
$n_s = 20, k = 4$	0.05	0.5248	2.797	0.654
$n_s = 20, k = 4$	0.10	0.5256	2.785	0.653

Table 12: Effect of sampling steps on SEED ($2\times$ SR).

Sampling Steps	NMSE↓	SNR↑	PCC↑
1 step	**0.4982**	**3.026**	**0.674**
2 steps	0.5030	2.984	0.671
3 steps	0.5153	2.879	0.663

Results. As shown in Table 14, adding positional encoding improves performance over the baseline without spatial information. However, TopoDiff further improves over the positional encoding variant across all metrics.

These results indicate that the benefits of TopoDiff are not simply due to providing electrode coordinates, but rather from modeling *global spatial structure*. The topographic representations encode smooth spatial distributions and large-scale patterns across the scalp, which cannot be captured by independent coordinate embeddings. This highlights the importance of incorporating structured spatial priors for EEG generation.

Effect of Topographic Representation and Smoothing. While topographic maps are a coarse visualization of EEG signals, they are used in our framework to provide complementary global spatial context rather than high-resolution signal reconstruction. This design is consistent with how human experts interpret EEG through spatial distributions over the scalp.

We empirically evaluate the impact of topographic smoothing by replacing the default interpolation with nearest-neighbor interpolation, which produces blocky and less smooth representations. Under this setting, performance degrades (SNR: 2.654 on SEED $2\times$ vs. 3.026 with default interpolation), indicating that smooth spatial structure is beneficial rather than detrimental.

In addition, we observe that incorporating topographic conditioning improves both convergence speed and final reconstruction quality (see main paper), further supporting its role as a meaningful global spatial prior.

Complementarity of Topology and Graph Modeling. Although both topographic conditioning and the graph module incorporate spatial information, they capture complementary aspects of EEG structure. The topographic representation encodes a global, continuous spatial prior reflecting smooth scalp distributions, while the graph module models dynamic, channel-wise relationships that vary over time.

Empirically, both components independently improve performance, and their combination yields the best results. While some redundancy may exist, it is not detrimental; instead, it enables beneficial coupling between global and local spatial cues. Additional analysis shows strong alignment in important regions (high correlation) alongside non-identical focus, indicating that the two components provide consistent yet distinct spatial signals. Together, they enhance spatial modeling and improve reconstruction quality.

Table 13: Ablation of dynamic graph modeling on SEED ($2\times$ SR).

Method	NMSE↓	SNR↑	PCC↑
Transformer (no graph)	0.5274	2.779	0.653
Static graph	0.5208	2.833	0.659
Dynamic graph (Ours)	**0.4982**	**3.026**	**0.674**

Table 14: Effect of spatial conditioning on SEED ($2\times$ SR).

Method	NMSE↓	SNR↑	PCC↑
Diffusion (no positional encoding)	0.5377	2.694	0.645
Diffusion (positional encoding)	0.5307	2.752	0.650
TopoDiff (Ours)	**0.4982**	**3.026**	**0.674**

L.1 Ablation on Visual Backbone for Topographic Conditioning

We analyze the sensitivity of TopoDiff to the choice of visual backbone used for extracting features from EEG topographic representations. Since topographic conditioning provides global spatial context to the diffusion model, it is important to understand whether performance depends strongly on the specific encoder architecture.

Setup. We replace the DINOv3-Base encoder used in the main model with a ConvNeXtV2-Base backbone, while keeping all other components unchanged. Both encoders are used to extract feature maps from topographic images, which are then projected into the Transformer latent space and used as conditioning inputs. We report results on the SEED dataset under $2\times$, $4\times$, and $8\times$ super-resolution (SR) settings.

Results. As shown in Table 15, both visual backbones achieve comparable performance across all SR settings. ConvNeXtV2 yields slightly higher SNR at $2\times$ and $4\times$, while DINOv3 performs marginally better at $8\times$.

These results suggest that TopoDiff is not highly sensitive to the specific choice of visual backbone, as long as the encoder provides sufficiently rich spatial features. This indicates that the performance gains primarily stem from the use of topographic conditioning itself, rather than the particular architecture of the vision encoder. At the same time, the small performance differences suggest that different backbones may capture slightly different spatial characteristics, which could be explored further in future work.

Table 15: Ablation on visual backbone for topographic conditioning (SEED dataset). We report SNR under different super-resolution settings.

Method	$2\times$	$4\times$	$8\times$
w/ DINOv3	3.026	2.639	**2.004**
w/ ConvNeXtV2	**3.052**	**2.655**	1.996

M Comparison with Alternative Approaches

We provide additional comparisons with alternative reconstruction and conditioning strategies to further contextualize the performance of TopoDiff. Unless otherwise noted, all results are reported on SEED under the $2\times$ spatial super-resolution setting.

Comparsion with direct interpolation-based methods. We also compare TopoDiff with conventional interpolation and an alternative topoplot-conditioning design. Compared with bilinear interpolation, TopoDiff achieves a 44.7% reduction in NMSE and more than a $14\times$ improvement in

Table 16: Comparison with direct regression and masked-autoencoding-based reconstruction on SEED at $2\times$ SR.

Method	NMSE ↓	SNR ↑	PCC ↑
Our architecture, direct regression	0.581	2.359	0.608
Our architecture, direct regression + topology and graph	0.578	2.377	0.609
ESTFormer	0.554	2.566	0.634
ESTFormer + topology and graph	0.526	2.791	0.653
TopoDiff	**0.498**	**3.026**	**0.674**

SNR. Compared with spherical-spline interpolation, TopoDiff reduces NMSE by 47.7% and improves SNR by 1347.8%, demonstrating that conventional spatial interpolation cannot recover the complex channel dependencies required for realistic EEG reconstruction.

Imputation-based methods. We further compare against GRIN (Cini et al., 2022), a representative graph-based imputation method. TopoDiff achieves 8.9% lower NMSE and 17.1% higher SNR than GRIN, indicating that conditional generation provides advantages beyond standard imputation-based reconstruction.

Direct regression and masked-autoencoding baselines. To distinguish the diffusion formulation from direct reconstruction, we replace our denoising objective with masked-autoencoding-based regression. This variant achieves only 2.359 dB SNR, and adding topology and graph-relation modeling yields only a marginal improvement as shown in Tab. 16. Although direct regression converges in approximately 20% of the training epochs required by TopoDiff, it generalizes poorly. We also incorporate our topological conditioning and graph-relation modeling into ESTFormer, a masked-autoencoding method with additional architectural and training improvements. These additions substantially improve ESTFormer, but the resulting model still underperforms TopoDiff, indicating that the diffusion-based formulation itself is important for strong generalization.

Although TopoDiff uses a single Euler step by default, this remains distinct from deterministic direct regression: reconstruction begins from a sampled noise state and follows the learned conditional reverse-time ODE. We use one step because it is both efficient and optimal in our experiments; using two steps introduces additional high-frequency components and slightly reduces SNR by approximately 0.04 dB on SEED.

Diffusion-based methods. We additionally compare with SRGDiff (Liu et al., 2025a), a recent diffusion-based reconstruction approach, using its official implementation. TopoDiff achieves 26.1% lower NMSE, 91.6% higher SNR, and 17.7% higher PCC. Thus, TopoDiff's gains are not attributable to diffusion alone, but to the combination of global topological priors and dynamic graph-based relational modeling.

Across interpolation, imputation, direct regression, masked-autoencoding, and diffusion-based baselines, TopoDiff consistently achieves the strongest reconstruction quality. These results support the effectiveness of combining structured topological conditioning with dynamic graph modeling in a conditional diffusion framework for EEG spatial super-resolution.

N Results on Sleep Dataset

We further evaluate TopoDiff on the ANPHY-Sleep dataset (Wei et al., 2024), a high-density EEG benchmark containing 83-channel overnight recordings with detailed sleep-stage annotations. The ANPHY-Sleep dataset was obtained from OSF using DOI `10.17605/OSF.IO/R26FH`. The dataset is distributed under the Creative Commons CC0 1.0 Universal license. Compared to other datasets used in this work, sleep EEG exhibits distinct temporal dynamics and spectral characteristics, making it a challenging and complementary setting for evaluating reconstruction quality and generalization.

Setup. We follow the same super-resolution protocol as in other experiments, training the model to reconstruct high-density EEG signals from a subset of observed channels. We compare TopoDiff against ESTFormer, a recent Transformer-based baseline for EEG modeling.

Results. As shown in Table 17, TopoDiff achieves improved reconstruction performance over ESTFormer across all metrics. In particular, our method reduces NMSE and improves both SNR and PCC, demonstrating consistent gains in signal fidelity.

Table 17: Performance on the ANPHY-Sleep dataset.

Method	NMSE↓	SNR↑	PCC↑
ESTFormer	0.854	0.686	0.916
TopoDiff (Ours)	**0.7962**	**0.99**	**0.925**

Discussion. These results indicate that TopoDiff generalizes effectively to sleep EEG, which differs substantially from emotion and motor imagery datasets in both temporal structure and frequency content. The consistent improvements over ESTFormer suggest that combining global topological conditioning with dynamic graph-based relational modeling provides robust spatial priors that transfer across diverse EEG domains.

O Comparison with SRGDiff

We additionally compare TopoDiff with SRGDiff (Liu et al., 2025a), a recent diffusion-based EEG spatial super-resolution method. We evaluate SRGDiff using the authors' official codebase under the same dataset splits and super-resolution settings used for TopoDiff. Since comparisons with STAD, RDPI, and ESTFormer are already included in the main text, this section focuses only on SRGDiff and TopoDiff. Table 18 reports signal fidelity metrics on SEED, SEED-IV, and PhysioNet MI/MM under $2\times$, $4\times$, and $8\times$ SR settings.

Discussion. TopoDiff consistently outperforms SRGDiff across all datasets, SR factors, and signal fidelity metrics. These results suggest that the improvement is not solely due to using a diffusion-based formulation, but also stems from incorporating global topographic conditioning and dynamic graph-based inter-channel relation modeling.

P Saliency Analysis of Spatial Components

We further analyze the interaction between topographic conditioning and dynamic graph-based relation modeling using saliency-based methods. Although both components incorporate spatial information and operate on temporally patchified inputs, they capture complementary aspects of EEG structure.

As shown in Table 14 and Table 13, each component independently improves reconstruction performance, and their combination yields the best results. The topographic representation encodes a global, continuous spatial prior reflecting smooth scalp structure, while the graph module models dynamic, channel-wise relationships that vary over time.

To better understand their interaction, we perform a projected saliency analysis by computing importance scores over spatial locations for both components. The resulting saliency maps exhibit strong alignment in important regions (correlation ≈ 0.90), indicating that both modules attend to similar spatial areas. At the same time, the overlap among the most salient regions is moderate, suggesting that each component captures distinct but related spatial information.

These results indicate that while some redundancy exists, it is not detrimental. Instead, the two components provide complementary spatial signals that reinforce consistent patterns from different perspectives. Their combination enables more effective spatial modeling, leading to improved reconstruction performance.

Table 18: Comparison between SRGDiff and TopoDiff across datasets and SR factors. SRGDiff results are obtained using the authors' official codebase. We report signal fidelity metrics: NMSE, SNR, and PCC.

Method	2× SR			4× SR			8× SR		
	NMSE↓	SNR↑	PCC↑	NMSE↓	SNR↑	PCC↑	NMSE↓	SNR↑	PCC↑
SEED									
SRGDiff	0.7184	1.436	0.552	0.7938	1.000	0.471	0.8955	0.480	0.328
TopoDiff (Ours)	**0.4982**	**3.026**	**0.674**	**0.5447**	**2.639**	**0.639**	**0.6303**	**2.004**	**0.574**
SEED-IV									
SRGDiff	0.5775	2.380	0.702	0.6883	1.622	0.634	0.8306	0.806	0.451
TopoDiff (Ours)	**0.2628**	**5.805**	**0.843**	**0.3091**	**5.099**	**0.814**	**0.3607**	**4.429**	**0.783**
PhysioNet MI/MM									
SRGDiff	0.2887	5.395	0.855	0.4525	3.440	0.825	0.6064	2.172	0.718
TopoDiff (Ours)	**0.1591**	**7.983**	**0.912**	**0.1880**	**7.258**	**0.896**	**0.2326**	**6.334**	**0.871**

Table 19: **Reconstruction variability at different evaluation granularities.** Each entry reports $\mathrm{mean} \pm \mathrm{std}$ in the order segment/subject/trial. O denotes TopoDiff and E denotes ESTFormer. "—" indicates that trial-level statistics are unavailable for MI/MM. Lower NMSE is better, while higher SNR and PCC are better.

Dataset	SR	M	NMSE (Seg/Sub/Trial)	SNR (Seg/Sub/Trial)	PCC (Seg/Sub/Trial)
SEED	2×	O	$.4981 \pm .2336/.4981 \pm .1418/.4977 \pm .2103$	$3.6044 \pm 2.3786/3.6044 \pm 1.3444/3.6090 \pm 2.0630$	$.6745 \pm .1760/.6745 \pm .1057/.6746 \pm .1587$
		E	$.5855 \pm .2751/.5855 \pm .1735/.5858 \pm .2558$	$2.8931 \pm 1.4618/2.8931 \pm 1.3876/2.8948 \pm 2.0740$	$.5694 \pm .2758/.5694 \pm .1759/.5687 \pm .2644$
	4×	O	$.5447 \pm .2370/.5447 \pm .1545/.5444 \pm .2137$	$3.1245 \pm 2.1678/3.1245 \pm 1.3000/3.1284 \pm 1.8785$	$.6392 \pm .1863/.6392 \pm .1219/.6394 \pm .1696$
		E	$.5890 \pm .2820/.5890 \pm .1912/.5765 \pm .2651$	$2.7271 \pm 2.4285/2.7271 \pm 1.5273/2.7261 \pm 2.1867$	$.5641 \pm .2827/.5641 \pm .1960/.5643 \pm .2718$
	8×	O	$.6300 \pm .2234/.6300 \pm .1390/.6298 \pm .1938$	$2.3593 \pm 1.8909/2.3593 \pm 1.1192/2.3612 \pm 1.6023$	$.5739 \pm .1819/.5739 \pm .1145/.5740 \pm .1593$
		E	$.6702 \pm .2560/.6702 \pm .1654/.6705 \pm .2337$	$2.1486 \pm 2.0348/2.1486 \pm 1.2420/2.1481 \pm 1.7970$	$.4982 \pm .2660/.4982 \pm .1812/.4977 \pm .2509$
SEED-IV	2×	O	$.2754 \pm .1731/.2754 \pm .0884/.2750 \pm .1367$	$6.2818 \pm 2.3724/6.2818 \pm 1.3277/6.2903 \pm 1.8233$	$.8300 \pm .1338/.8300 \pm .0609/.8299 \pm .1066$
		E	$.2804 \pm .1645/.2804 \pm .0820/.2806 \pm .1293$	$6.0990 \pm 2.1626/6.0990 \pm 1.1963/6.0996 \pm 1.6526$	$.8264 \pm .1265/.8264 \pm .0562/.8259 \pm .0988$
	4×	O	$.3402 \pm .1807/.3402 \pm .0897/.3405 \pm .1355$	$5.2314 \pm 2.1734/5.2314 \pm 1.2226/5.2309 \pm 1.6679$	$.7895 \pm .1371/.7895 \pm .0610/.7892 \pm .1041$
		E	$.3663 \pm .1707/.3663 \pm .0816/.3665 \pm .1287$	$5.1804 \pm 2.0989/5.1804 \pm 1.1164/5.1806 \pm 1.5942$	$.7693 \pm .1309/.7693 \pm .0565/.7689 \pm .0998$
	8×	O	$.3710 \pm .1733/.3710 \pm .0861/.3705 \pm .1239$	$4.7309 \pm 1.9055/4.7309 \pm 1.0513/4.7349 \pm 1.3793$	$.7720 \pm .1349/.7720 \pm .0599/.7721 \pm .0979$
		E	$.4059 \pm .1628/.4059 \pm .0521/.4055 \pm .1067$	$4.2272 \pm 1.6330/4.2272 \pm .6190/4.2329 \pm 1.0750$	$.7514 \pm .1282/.7514 \pm .0388/.7513 \pm .0879$
MI/MM	2×	O	$.1590 \pm .1500/.1591 \pm .1075/—$	$9.7884 \pm 4.0992/9.7866 \pm 3.2160/—$	$.9125 \pm .0925/.9124 \pm .0640/—$
		E	$.1788 \pm .1782/.1789 \pm .1380/—$	$9.3979 \pm 4.1825/9.3942 \pm 3.3425/—$	$.9020 \pm .1061/.9019 \pm .0796/—$
	4×	O	$.1880 \pm .1521/.1882 \pm .1124/—$	$8.6734 \pm 3.6616/8.6711 \pm 2.8455/—$	$.8964 \pm .0937/.8963 \pm .0677/—$
		E	$.2496 \pm .3916/.2500 \pm .3488/—$	$8.4650 \pm 4.2392/8.4605 \pm 3.5713/—$	$.8805 \pm .1420/.8803 \pm .1251/—$
	8×	O	$.2327 \pm .1694/.2328 \pm .1250/—$	$7.5361 \pm 3.3852/7.5346 \pm 2.6465/—$	$.8713 \pm .1046/.8712 \pm .0754/—$
		E	$.3085 \pm .3659/.3088 \pm .3177/—$	$6.7738 \pm 3.5991/6.7712 \pm 2.9895/—$	$.8450 \pm .1446/.8449 \pm .1215/—$

Q Reconstruction Variability Across Evaluation Granularities

To further characterize the variability of reconstruction performance, we report the mean and standard deviation of NMSE, SNR, and PCC at three evaluation granularities: individual segments, subjects, and trials. For subject- and trial-level statistics, metrics are first aggregated within each corresponding unit before computing the mean and standard deviation across subjects or trials. MI/MM does not provide trial-level groupings, so trial-level statistics are unavailable for this dataset.

Across datasets and super-resolution ratios, TopoDiff consistently achieves better mean reconstruction performance than ESTFormer, with lower NMSE and higher SNR and PCC at the segment, subject, and, where available, trial levels. The results also show the expected reduction in variability when moving from individual segments to coarser subject-level aggregation. For example, on SEED at $2\times$ SR, the standard deviation of TopoDiff NMSE decreases from 0.2336 across segments to 0.1418 across subjects, while PCC variability decreases from 0.1760 to 0.1057. Similar trends are observed across the remaining datasets and SR settings, indicating that the reconstruction improvements are not driven by a small subset of segments and remain consistent when performance is summarized at the subject and trial levels.

R Channel Subsampling and Low-Density Evaluation

Evaluating EEG spatial super-resolution requires paired low- and high-density recordings of the same underlying neural activity. Metrics such as NMSE, SNR, and PCC require a high-density reference recorded simultaneously with the sparse input, yet, to our knowledge, no public EEG dataset provides simultaneous recordings at multiple channel densities. Collecting such data using separate low- and high-density caps is also nontrivial, since electrodes from the two systems cannot occupy identical scalp locations simultaneously. Consequently, prior EEG super-resolution and channel-imputation methods construct low-density observations by subsampling channels from complete recordings, including Corley & Huang (2018), Svantesson et al. (2021), STAD (Wang et al., 2024a), ESTFormer (Li et al., 2025), SRGDiff (Liu et al., 2025a), CSDI (Tashiro et al., 2021), GRIN (Cini et al., 2022), and RDPI (Liu et al., 2025b). Recent EEG representation-learning frameworks such as HEAR and REVE similarly emphasize standardized channel subsets to accommodate heterogeneous electrode montages across datasets. We therefore follow this field-standard evaluation protocol and construct sparse inputs at three levels of channel reduction: $2\times$, $4\times$, and $8\times$.

The retained electrodes are selected to preserve approximately symmetric whole-scalp coverage under standard 10–5, 10–10, and 10–20 electrode-placement conventions (Jurcak et al., 2007; Klem et al., 1999). This design avoids concentrating observed channels within a particular scalp region and more closely reflects practical low-density EEG configurations. In particular, our $4\times$ setting closely resembles the spatial density of the 14-channel Emotiv EPOC montage, while the $8\times$ setting approximates the channel count of an OpenBCI Cyton configuration. For TUSZ, the $2\times$ subset additionally follows the motivation of abbreviated EEG montages studied for critical-care seizure detection (Gururangan et al., 2018; Westover et al., 2020).

Importantly, our evaluation is not restricted to synthetically subsampled high-density recordings. We additionally evaluate DREAMER (Katsigiannis & Ramzan, 2018), which was acquired directly using a 14-channel Emotiv EPOC headset and therefore provides a complementary test on genuinely low-density EEG. We use TopoDiff to reconstruct the DREAMER recordings onto the 62-channel SEED montage and evaluate whether the reconstructed representation improves downstream emotion recognition. Reconstruction increases classification accuracy from 0.427 using the original low-density signals to 0.434 using the reconstructed montage.

S Biological Interpretation of Learned Channel Relations

The graph edges in TopoDiff should not be interpreted as evidence of functional or anatomical connectivity between brain regions. Scalp EEG measurements are strongly affected by volume conduction: activity from a single neural source can contribute simultaneously to multiple electrodes, producing substantial sensor-level dependence even in the absence of direct interaction between the underlying sources. Consequently, correlations or other dependencies measured in sensor space cannot, by themselves, be interpreted as source-level connectivity (Brunner et al., 2016; Van de Steen et al., 2019). Estimating physiological connectivity generally requires additional procedures designed to mitigate volume-conduction effects, such as imaginary coherency (Nolte et al., 2004), or explicit source-space analysis. TopoDiff does not attempt to estimate such physiological connectivity, and we therefore do not interpret its learned graph edges in this manner.

Instead, the cosine-based similarity used by TopoDiff captures *predictive redundancy* between electrodes: the extent to which the signal observed at one scalp location provides information useful for reconstructing the signal at another. This distinction is important for spatial super-resolution. Each scalp electrode records a mixture of contributions from multiple neural sources propagated through the head volume conductor, and individual sources can contribute to measurements over spatially extended regions of the scalp (Nunez & Srinivasan, 2006). Nearby electrodes therefore tend to share substantial source support by construction. Covariance between distinct underlying sources can introduce additional statistical dependence. For the reconstruction problem considered here, however, these contributions need not be disentangled: either form of shared structure can make an unobserved channel more predictable from the observed channels.

This objective differs fundamentally from that of functional-connectivity analysis. In connectivity estimation, instantaneous or zero-lag dependence caused by volume conduction is typically treated as a confound because it can create apparent interactions that do not reflect genuine coupling between

neural sources (Nolte et al., 2004; Brunner et al., 2016; Van de Steen et al., 2019). In spatial super-resolution, by contrast, this shared sensor-level structure is useful information. The spatial mixing that complicates physiological interpretation is precisely part of what allows the signal at an unobserved scalp location to be inferred from signals measured elsewhere.

Classical EEG interpolation methods exploit the same principle through fixed assumptions about spatial smoothness. For example, spherical-spline interpolation estimates missing scalp potentials from surrounding electrodes using a geometry-defined smoothness prior (Perrin et al., 1989). TopoDiff serves a related reconstruction objective but replaces this fixed interpolation structure with a data-adaptive graph that captures signal-dependent relationships among channels. Thus, the learned edges are best understood as reconstruction-oriented statistical relationships rather than estimates of anatomical or functional neural connectivity.