NEUROX-EEG: SCALP FIELD INPAINTING FOR ZERO-SHOT EEG SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG electrodes sparsely sample a smooth spatial field on the scalp surface. Existing learning-based super-resolution methods typically reconstruct a fixed set of target channels repeatedly supervised during training, leaving their ability to recover unseen electrode locations unclear. We introduce NeuroX-EEG for zero-shot reconstruction at locations excluded from both conditioning and supervision, a protocol that tests whether reconstruction generalizes over scalp position. Unlike dataset-level zero-shot transfer, this protocol holds out electrode sites within the same recordings. NeuroX-EEG performs diffusion-based inpainting on a 2D scalp grid, where convolution exploits the local smoothness implied by volume conduction, and incorporates pretrained EEG foundation-model representations to supply broader temporal, spectral, and cross-channel structure. Across three datasets spanning 62- and 19-channel montages and multiple observation densities, NeuroX-EEG achieves the lowest NMSE among the evaluated methods under our zero-shot protocol in six of seven dataset–scale settings. Region-wise zero-shot evaluation shows the clearest gains for band-shaped temporal and parietal holdouts, and the reconstructed signals yield higher mean downstream brain–computer interface (BCI) decoding performance than low-density input under the evaluated protocol.

1 INTRODUCTION

EEG provides a non-invasive window into brain dynamics, but its utility depends partly on how neural activity is spatially sampled across the scalp. Cortical currents propagate through the skull and scalp to form a smooth scalp-potential field, which EEG electrodes observe only at sparse, discrete locations (Nunez & Srinivasan, 2006; Srinivasan et al., 1996). Reconstructing EEG at unrecorded electrode positions is therefore an instance of a broader problem: estimating a function from measurements at known locations and querying it at new ones, rather than learning a fixed mapping between two channel sets. An EEG montage, i.e., the set of electrode positions used in a recording, can thus be viewed as one sampling of the underlying field rather than as the signal domain itself.

In practice, EEG recordings vary in both sampling density and spatial coverage across acquisition protocols. Standardized systems such as the 10–20, 10–10, and 10–5 montages define progressively denser electrode sets within a shared cranial reference framework (Stoyell et al., 2021). Individual studies and datasets nevertheless select different subsets of these positions, resulting in recordings with only partially overlapping channel sets and spatial coverage. Reduced coverage can obscure spatially localized patterns. It also makes recordings hard to compare with one another and incompatible with analyses or models that assume a denser, fixed channel set. When only partial measurements are available, spatial super-resolution estimates unavailable locations and maps sparse or previously collected recordings to a common spatial support for downstream analysis.

Most learning-based EEG super-resolution methods, however, are optimized for a prescribed low-to-high-density mapping. Their target channels are repeatedly supervised during training, so the model can specialize to those locations without learning a reconstruction rule that extends to new targets. This fixed-target paradigm does not match settings in which the missing electrodes vary across recordings or datasets. Moreover, accuracy on repeatedly supervised targets does not reveal whether a model has learned the spatial structure of the underlying field or has merely memorized a mapping specific to the supervised channels. We therefore address **zero-shot EEG channel reconstruction**:

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.

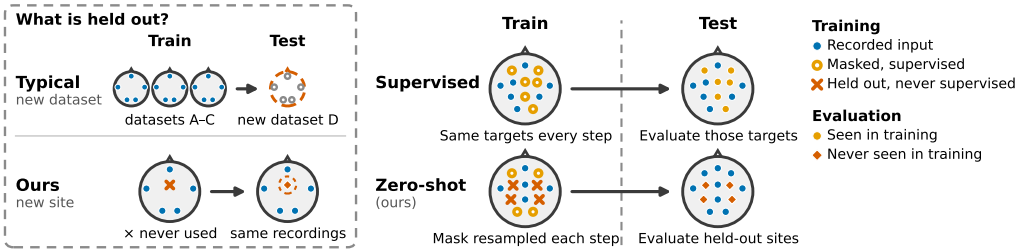


Figure 1: **Supervised vs. zero-shot EEG super-resolution. Left:** typical zero-shot transfer holds out an entire new dataset, whereas ours holds out electrode *sites* within the same recordings. **Right:** supervised methods train and evaluate on the same target channels. In our protocol, an observed subset $\mathcal{O}$ (blue) is resampled from the seen set $\mathcal{S}$ at each step, the remaining seen channels $\mathcal{S} \setminus \mathcal{O}$ (amber rings) are supervised, and a fixed subset $\mathcal{U}$ (×) is withheld from both the conditioning input and the reconstruction loss throughout training and evaluated only at test time (Section 3.1).

a target electrode set is held out from both conditioning and supervision throughout training and reconstructed only at test time (Figure 1). Success here requires a reconstruction rule that transfers across scalp positions, in line with the field view above, rather than a memorized channel mapping.

We introduce **NeuroX-EEG**, a diffusion framework designed for zero-shot EEG super-resolution. NeuroX-EEG represents partial recordings and reconstruction targets on a shared 2D scalp grid and performs spatial inpainting over this domain. The grid exposes local electrode neighborhoods to convolution, while pretrained EEG foundation-model representations provide broader signal context when the available spatial evidence is limited. Together, these components enable the model to reconstruct electrode locations that are excluded from supervision during training. We make three **contributions**:

- Motivated by a scalp-field view of EEG, we formulate and evaluate spatial super-resolution under a strict zero-shot setting, in which target electrodes are excluded from both conditioning and supervision during training.
- We introduce a diffusion inpainting framework on a shared 2D scalp grid, providing an explicit local spatial inductive bias while accommodating different observed channel subsets within the same generative domain.
- We incorporate pretrained EEG foundation-model representations into diffusion training through representation alignment, complementing the grid's local spatial modeling with broader signal context.

2 Related Works

2.1 EEG Spatial Super-Resolution

Reconstructing a partially observed field at unmeasured locations is a recurring machine-learning problem, but methods differ in how they represent the domain. Coordinate-conditioned neural fields query a continuous function at arbitrary locations (Mildenhall et al., 2020), whereas structured grid or graph models make neighborhood relations explicit in the representation (Lam et al., 2023). Across these domains, the central representation choice is between coordinate-queryable flexibility and an explicit spatial inductive bias. EEG spatial super-resolution (SR) poses this choice on the scalp by reconstructing signals at unobserved electrode locations from a sparse recording (Stoyell et al., 2021; Yazıcı et al., 2025).

Classical spherical splines already treat electrode measurements as samples of a smooth scalp field (Perrin et al., 1989), but use a fixed interpolation rule. Most learning-based methods instead bind electrode position to a fixed channel index, including CNNs (Saba-Sadiya et al., 2020), GANs (Corley & Huang, 2018), transformers (Li et al., 2025), and diffusion models (Wang et al., 2025a; Liu et al., 2025). Other methods add topology as conditioning: TopoDiff (Yao et al., 2026) uses topographic image embeddings and channel-relation graphs, and CAFE (Liu et al., 2026a) reconstructs channels in a proximal-to-distal rollout.

Coordinate-conditioned reconstruction provides an alternative representation of the same problem. ZUNA (Warner et al., 2026) conditions on continuous 3D coordinates, Neural Brain Fields (Kedem et al., 2025) fit a coordinate-based space–time field to individual recordings, and ScalpINR (Liu et al., 2026b) amortizes such an implicit field across recordings through support conditioning, studying unseen-electrode generation with a small held-out electrode set under varying observed support. These methods place the field in a coordinate-queryable function space, where spatial regularity must be learned through the coordinate encoding. NeuroX-EEG takes the complementary explicit-domain position: observed and target locations share a regular scalp lattice, so neighborhood structure is imposed architecturally and reused by convolution. In short, coordinate-based models can be queried at any scalp position but must learn spatial smoothness from data, whereas our grid is evaluated only at lattice cells but builds local neighborhood structure into the architecture (Section 3.2; Appendix B).

2.2 Representation Alignment

Pretrained foundation models provide representations that integrate temporal, spectral, and cross-channel regularities learned across large neural-signal corpora. Such representations are particularly valuable for zero-shot channel reconstruction: when observations are sparse or support a target from only one side, the local neighborhood available to a convolutional denoiser may not determine a plausible signal at an unseen location, whereas a foundation model pretrained across many recordings and montages encodes spectral and cross-channel regularities that do not depend on a fixed channel layout. Representation alignment offers a direct way to transfer this information to a generative model: REPA (Yu et al., 2024) aligns intermediate denoiser features with a pretrained encoder without adding inference-time computation, while subsequent work improves the alignment schedule across diffusion timesteps and training stages (Ton et al., 2025; Wang et al., 2025b). NeuroX-EEG brings pretrained EEG foundation-model representations into spatial signal reconstruction, complementing grid-based local modeling with broader temporal, spectral, and cross-channel structure learned at scale.

3 Approach

NeuroX-EEG performs diffusion inpainting on a 2D scalp grid and uses foundation-model alignment only during training (Figure 2). We define the zero-shot channel reconstruction problem, then describe the 2D scalp-grid representation, observation masking, diffusion-based inpainting, and representation alignment objective.

3.1 Problem Setting

Referenced scalp field. Let $\mathcal{M}$ denote the scalp surface and let $f : \mathcal{M} \times \{1, \dots, T\} \to \mathbb{R}$ denote the scalp-potential field under a common referencing and preprocessing convention. An electrode c at position $\mathbf{p}_c \in \mathcal{M}$ records

$$X_{c,t} = f(\mathbf{p}_c, t) + \epsilon_{c,t}, \tag{1}$$

where $\epsilon_{c,t}$ includes measurement noise and modeling error. The field f is defined under the acquisition reference and preprocessing convention shared by the input and target recordings. This formulation makes electrode identity secondary to spatial position and provides the domain on which zero-shot reconstruction is learned.

Because ground-truth field values are available only at electrode positions, we operationalize field recovery as **zero-shot EEG channel reconstruction**: given signals from a fixed subset of *seen* electrodes, reconstruct the signals at *unseen* electrode locations. Throughout this paper, *reconstruction* refers to this zero-shot task of recovering unsupervised target channels from spatial context. Because zero-shot targets are never fit during training, accuracy at these locations reflects how well the learned reconstruction generalizes over scalp position, while remaining an electrode-level measure rather than a validation of the field between electrodes (Appendix A).

Let $\mathbf{X} \in \mathbb{R}^{C \times T}$ denote a multi-channel EEG segment with C electrodes and T time points. We partition the electrode channels into a seen set $\mathcal{S}$ and a zero-shot unseen set $\mathcal{U}$, with $\mathcal{S} \cap \mathcal{U} = \emptyset$ and $\mathcal{S} \cup \mathcal{U} \subseteq \{1, \dots, C\}$. The unseen set $\mathcal{U}$ is fixed before training and is excluded from both the conditioning input and the supervised reconstruction loss throughout training. Within the seen set

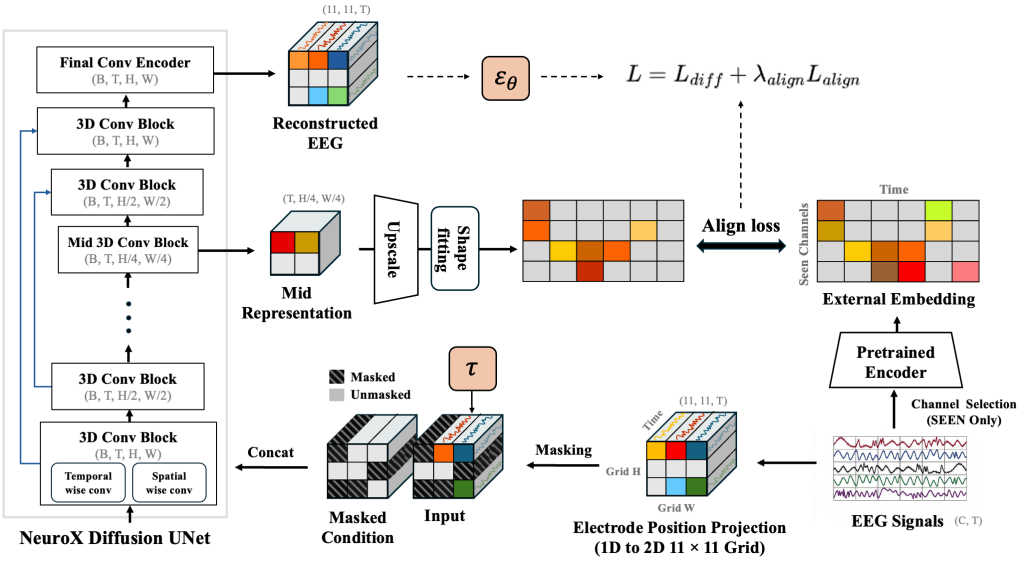


Figure 2: **NeuroX-EEG architecture overview.** Raw EEG signals (bottom right) are projected from 1D channel index to a 2D grid by electrode position. Masked grid cells form the conditioning input to a 3D UNet (left, Section 3.5) which performs conditional denoising over the grid. The UNet bottleneck is upscaled and aligned with per-channel embeddings from a pretrained EEG foundation model (Han et al., 2025) (right), computed only on seen channels. The output reconstructs the full grid, supervised by a combined denoising ($\mathcal{L}_{\text{diff}}$) and alignment ($\mathcal{L}_{\text{align}}$) loss.

$\mathcal{S}$, we further sample a training-time observation subset $\mathcal{O} \subset \mathcal{S}$ containing about half of the seen channels ($\rho = 0.5$; Table 5) at each training step. Channels in $\mathcal{O}$ are provided to the model as observed conditioning signals, while the remaining seen channels $\mathcal{S} \setminus \mathcal{O}$ are masked from the input and used as reconstruction targets. Thus, $\mathcal{S}$ denotes channels available during training, whereas $\mathcal{O}$ denotes the subset actually observed in a given training instance. Detailed algorithms are provided in Appendix C.2.

3.2 Spatial Grid Representation

To exploit the spatial topology of EEG electrodes, we project the EEG channels onto a structured 2D spatial grid. Let $\pi : \mathcal{M} \to \{1, \dots, H\} \times \{1, \dots, W\}$ denote the discretizing projection. Each electrode at $\mathbf{p}_c$ is assigned the fixed grid position $\pi(\mathbf{p}_c) = (r_c, c_c)$ on an $H \times W$ map derived from a standard topographic layout used across EEG analysis tooling (Zhang et al., 2024). Full grid layout details are provided in Appendix C.4. The resulting spatio-temporal volume is $\mathbf{G} \in \mathbb{R}^{T \times H \times W}$, where $\mathbf{G}[t, r, c]$ contains the signal value at time t for the electrode assigned to grid position (r, c). Grid cells with no assigned electrode are set to zero. We define a binary valid-electrode mask $M^{\text{valid}} \in \{0, 1\}^{H \times W}$, where $M^{\text{valid}}_{r,c} = 1$ indicates that grid position (r, c) corresponds to a valid EEG electrode, and $M^{\text{valid}}_{r,c} = 0$ indicates an empty grid cell. The seen and unseen channel sets, $\mathcal{S}$ and $\mathcal{U}$, are therefore defined over valid electrode positions satisfying $M^{\text{valid}}_{r,c} = 1$.

Spatial inductive bias. Topographic projections and discrete electrode lattices are established representations for preserving approximate local electrode relationships in CNN-based EEG analysis (Delorme & Makeig, 2004; Gramfort et al., 2013; Bashivan et al., 2016; Thodoroff et al., 2016; Wang et al., 2020; Su et al., 2022; Zhang et al., 2024). NeuroX-EEG uses this lattice as a generative field domain: measured signals occupy only their electrode cells, unassigned cells remain empty, and a shared convolutional denoiser reconstructs the missing field without a preceding interpolation step. The grid therefore turns a channel-indexed problem into spatial inpainting and enables local filters to be reused across scalp neighborhoods (LeCun et al., 1998).

This design extends the physical premise behind spherical-spline interpolation (Perrin et al., 1989): nearby scalp locations tend to carry correlated potentials because of volume conduction. Rather than

fixing the interpolation kernel, NeuroX-EEG learns the spatiotemporal reconstruction rule from data while retaining neighborhood structure explicitly.

3.3 Observation Masking

Given the seen set $\mathcal{S}$ and the training-time observed subset $\mathcal{O} \subset \mathcal{S}$ defined in Section 3.1, we construct the observed condition $\mathbf{C}$ by retaining only the channels in $\mathcal{O}$: $\mathbf{C}_{:,r,c} = \mathbf{G}_{:,r,c}$ for $(r, c) \in \mathcal{O}$ and zero elsewhere, with $M^{\text{obs}} \in \{0, 1\}^{H \times W}$ the corresponding indicator of $\mathcal{O}$. During training, $\mathcal{O}$ is resampled at each step. At evaluation time, the observation set is fixed to the full seen set, i.e., $\mathcal{O} = \mathcal{S}$. The zero-shot channels $\mathcal{U}$ are never included in $\mathcal{O}$ and therefore receive no signal in $\mathbf{C}$ throughout training or evaluation.

3.4 Diffusion-based Spatial Inpainting

Sparse field completion is generally underdetermined: multiple dense signals can be compatible with the same observed support. We therefore formulate reconstruction as **conditional inpainting** on the grid and solve it with a denoising diffusion probabilistic model (DDPM) (Ho et al., 2020), which learns a conditional distribution over grid completions rather than imposing a fixed interpolation rule.

Forward diffusion. We use a linear noise schedule $\{\beta_t\}_{t=1}^{K}$ with $K = 1000$ steps and $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$. Given a clean grid target $\mathbf{x}_0 \in \mathbb{R}^{1\times T\times H\times W}$ in which all zero-shot target locations $\mathcal{U}$ are set to zero before noising, so that their ground-truth signals are absent from both the conditioning input and the noisy model state throughout training, the noisy sample at step t is

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{2}$$

Conditional noise prediction. The denoiser receives the channel-wise concatenation $[\mathbf{x}_t;\, \mathbf{C}] \in \mathbb{R}^{2\times T\times H\times W}$ and the diffusion timestep t, and predicts the injected noise:

$$\hat{\boldsymbol{\epsilon}} = f_\theta([\mathbf{x}_t;\, \mathbf{C}],\, t)\,. \tag{3}$$

Inference. We run the DDPM reverse process from $\mathbf{x}_K \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, conditioned on $\mathbf{C}$ at each denoising step. For accelerated sampling, we adopt DDIM (Song et al., 2020) with $\eta = 0$, yielding a deterministic trajectory with 50 function evaluations.

3.5 Architecture: 3D Spatio-Temporal UNet

The backbone f_θ is a 3D UNet that processes the spatio-temporal volume of shape $(2, T, H, W)$. The core design principle is **decoupled spatial and temporal convolutions** within each residual block: spatial dependencies are captured by $\mathrm{Conv3d}(1\times3\times3)$ kernels operating on the $H\times W$ plane, and temporal dependencies by $\mathrm{Conv3d}(3\times1\times1)$ kernels operating along the T axis. This factorization keeps the parameter count tractable while modeling electrode topology and temporal structure separately.

The encoder and decoder each have three resolution stages with channel multipliers $(1, 2, 4)$ of base width d and spatial-only downsampling, and the bottleneck feature map $\mathbf{h}_{\text{mid}} \in \mathbb{R}^{4d\times T\times \frac{H}{4}\times \frac{W}{4}}$ serves as the alignment representation (Section 3.6); component details are in Appendix C.3.

3.6 Training Objectives

Masked noise-prediction loss. The primary objective is a weighted MSE on the noise prediction:

$$\mathcal{L}_{\text{diff}} = \frac{\sum_{r,c} w_{r,c} \left\| \boldsymbol{\epsilon}_{:,r,c} - \hat{\boldsymbol{\epsilon}}_{:,r,c} \right\|^2}{\sum_{r,c} w_{r,c}}, \quad w_{r,c} = \begin{cases} \lambda_{\text{obs}}, & (r,c) \in \mathcal{O}, \\ 1, & (r,c) \in \mathcal{S} \setminus \mathcal{O}, \\ 0, & (r,c) \in \mathcal{U} \text{ or } M^{\text{valid}}_{r,c} = 0. \end{cases} \tag{4}$$

We set $\lambda_{\text{obs}} = 0.1$, assigning a small auxiliary weight to the observed conditioning channels while emphasizing the masked seen channels. The zero-shot channels $\mathcal{U}$ and empty grid cells receive zero weight throughout training; ablation details are provided in Appendix D.2.

Representation alignment (REPA). To provide the broader signal context motivated in Section 2.2, we align the denoiser with representations from DIVER-1, a pretrained EEG foundation model learned from a large multi-dataset EEG/iEEG corpus using masking objectives over the frequency and spatial axes (Han et al., 2025). Let $\mathbf{h}_{\mathrm{mid}}$ denote the spatiotemporal bottleneck activation produced by the middle block of the 3D UNet before the decoder. We align $\mathbf{h}_{\mathrm{mid}}$ with per-channel foundation-model embeddings $\mathbf{E}$ computed from the seen EEG channels. Because $\mathbf{h}_{\mathrm{mid}}$ is a spatiotemporal bottleneck feature map whereas $\mathbf{E}$ is defined at the EEG-channel level, we first average both representations over their temporal dimensions. We then flatten the spatial dimensions of $\mathbf{h}_{\mathrm{mid}}$ into a token sequence and linearly interpolate it to match the number of EEG channels in $\mathbf{E}$. A lightweight projection head maps the resulting channel-aligned bottleneck features to the teacher embedding dimension, and the alignment loss maximizes the mean cosine similarity:

$$\mathcal{L}_{\mathrm{align}} = -\frac{1}{N}\sum_{n=1}^{N}\frac{\langle \mathbf{p}_n,\, \mathbf{e}_n\rangle}{\|\mathbf{p}_n\|\,\|\mathbf{e}_n\|}, \tag{5}$$

where N denotes the number of EEG channels, $\mathbf{p}_n$ is the projected bottleneck feature assigned to the n-th channel after interpolation, and $\mathbf{e}_n$ is the teacher embedding of the n-th channel.

We weight alignment down at high-noise timesteps and disable it after the first 20% of training, following scheduling ideas from prior work (Ton et al., 2025; Wang et al., 2025b). This schedule achieves the best NMSE, PCC, and SNR among the evaluated alignment strategies (e.g., NMSE 0.833 versus 0.846 without alignment; Table 4), a small improvement within the evaluated ablation that adds no inference-time computation. The total training loss combines both objectives:

$$\mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \lambda_{\mathrm{align}}\,\mathcal{L}_{\mathrm{align}}. \tag{6}$$

4 Results

Datasets and Baselines. We evaluate on SEED (Zheng & Lu, 2015) and SEED-V (Liu et al., 2021), two 62-channel emotion-recognition datasets, and Mumtaz2016 (Mumtaz, 2016), a 19-channel clinical EEG dataset for which we use eyes-open and eyes-closed resting-state recordings. We compare against spherical spline interpolation (SSI) (Perrin et al., 1989), DeepConv (Saba-Sadiya et al., 2020), SRGDiff (Liu et al., 2025), and ZUNA (Warner et al., 2026). ZUNA is evaluated with its released weights because fine-tuning code is not available, and SRGDiff is paired with SSI because its encoder is fitted to a fixed channel set (Appendix C.5). On SEED and SEED-V, $2\times$, $4\times$, and $8\times$ follow the emotion-related low-density channel subsets of Tang et al. (2022), corresponding to observing approximately 50%, 25%, and 12.5% of the montage. For Mumtaz2016, we adopt the established 10–20 protocol of TopoDiff (Yao et al., 2026), which defines a single split of 9 seen and 10 target channels; we use this split unchanged as the zero-shot configuration, yielding the approximately $2\times$ Mumtaz2016 setting (Appendix C.4). All metrics are computed on zero-shot target channels in raw μV space. Preprocessing, splits, baseline adaptations, and complete evaluation details are provided in Appendix C. SEED and SEED-V scale-based results are averaged over four zero-shot channel configurations at $4\times$ and $8\times$ and three at $2\times$, where two of the four configurations in Tang et al. (2022) coincide. Region-wise results use one predefined configuration per region, and downstream results average the same configurations across three leave-one-subject-out test folds; standard deviations for repeated evaluations are reported in Appendix D.1.

4.1 Zero-Shot Channel Reconstruction

Table 1 summarizes zero-shot reconstruction across seven dataset–scale settings among the evaluated methods under our zero-shot protocol. NeuroX-EEG achieves the best NMSE in six settings and ranks in the top two in all seven, while obtaining the best SNR in five. Because a zero prediction yields NMSE ≈ 1 and SNR ≈ 0 dB (Appendix C.6), SEED at $8\times$ marks the limit of point-wise recovery for all evaluated methods, and comparisons in that setting are best read as relative. PCC is led by other methods on SEED-V at all scales, on SEED at $8\times$, and on Mumtaz2016. On Mumtaz2016 (a single predefined split), the NMSE and SNR ranking extends to 19-channel clinical EEG.

As observations become sparser from $2\times$ to $8\times$, NeuroX-EEG keeps the lowest NMSE in every setting except SEED-V $4\times$, where it ranks second behind SRGDiff+SSI (Figure 3). This NMSE ranking across datasets and densities, though not matched in PCC, is consistent with the value of

Table 1: Comparison of zero-shot channel reconstruction performance across datasets. NMSE indicates lower is better, while PCC and SNR indicate higher is better. Best results are in bold and second-best results are underlined.

Method	Metric	SEED (62)			SEED-V (62)			Mumtaz2016 (19)[¶]
		2×	4×	8×	2×	4×	8×	2×
SSI[†]	NMSE ↓	1.417	1.508	2.182	0.610	0.872	1.836	1.275
	PCC ↑	0.196	0.149	0.139	**0.691**	0.623	0.469	**0.542**
	SNR ↑	-1.510	-1.622	-2.973	2.152	0.650	-2.180	-1.056
ZUNA[‡]	NMSE ↓	1.039	1.061	1.062	1.000	0.999	1.000	0.934
	PCC ↑	0.285	0.246	**0.221**	0.342	0.276	0.235	0.257
	SNR ↑	-0.164	-0.256	-0.255	1.998	1.432	**1.143**	0.297
DeepConv[*]	NMSE ↓	0.956	0.944	1.037	0.836	0.885	0.907	0.999
	PCC ↑	0.314	0.252	0.124	0.423	0.353	0.320	0.039
	SNR ↑	0.223	0.251	-0.151	0.778	0.530	0.425	0.006
SRGDiff (+ SSI)[§]	NMSE ↓	1.130	1.420	1.809	0.668	**0.662**	0.977	1.036
	PCC ↑	0.291	0.166	0.123	0.657	**0.670**	**0.516**	0.530
	SNR ↑	-0.512	-1.423	-2.396	1.764	**1.803**	0.255	-0.155
NeuroX-EEG[*]	NMSE ↓	**0.833**	**0.912**	**1.033**	**0.604**	0.677	**0.868**	**0.912**
	PCC ↑	**0.421**	**0.321**	0.216	0.653	0.584	0.387	0.372
	SNR ↑	**0.803**	**0.401**	**-0.139**	**2.238**	1.732	0.634	**0.398**

[†] Mathematical interpolation method without learning. [‡] Pretrained model evaluated without fine-tuning. [*] End-to-end learned model. [§] Learned reconstruction followed by interpolation (Appendix C.5). [¶] Single predefined split following Yao et al. (2026); no across-configuration standard deviation.

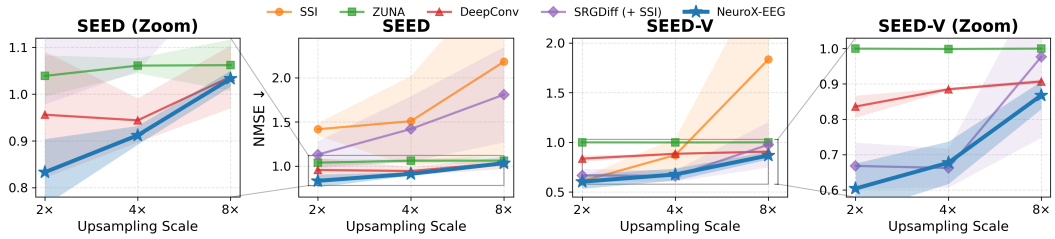


Figure 3: NMSE across upsampling scales under zero-shot EEG super-resolution on SEED and SEED-V. Lower is better. NeuroX-EEG achieves the lowest or second-lowest NMSE at every scale; at 8× on SEED, all methods have NMSE above 1, i.e., no lower error than predicting zero. Shaded regions indicate ±1 standard deviation across zero-shot channel configurations.

explicitly modeling scalp neighborhoods; the ablation in Section 4.4 compares flattened and 2D-grid inputs within the same backbone. Qualitative examples are in Appendix D.3.

4.2 Region-wise Zero-Shot Reconstruction

Random zero-shot channel sets mix the effects of target count and spatial arrangement. We therefore exclude structured Frontal, Temporal, Parietal, and Occipital scalp-region masks, creating structured missingness with substantially different support geometry. Unlike random or spatially distributed holdouts, a scalp-region holdout removes nearby observations over an extended area, so reconstruction depends on how much observed support surrounds the held-out region. Holdouts that remove a polar cap of the scalp (Frontal, Occipital) leave targets near the pole supported from one direction, whereas band-shaped holdouts (Temporal, Parietal) retain observed electrodes on multiple sides of the held-out region. Region definitions are provided in Appendix C.4. Region masks also differ in the number of targets (16–28 channels; Table 8), so support geometry and target count are not fully separated in this comparison.

Table 2: Region-wise zero-shot reconstruction on SEED and SEED-V. Each broad scalp region is excluded during training and evaluated at test time. NMSE indicates lower is better, while PCC and SNR indicate higher is better. Best results are in bold and second-best results are underlined. Macro mean is the unweighted average over the four regional holdouts.

Unseen Region (# Seen Channels)	Method	SEED			SEED-V		
		NMSE ↓	PCC ↑	SNR ↑	NMSE ↓	PCC ↑	SNR ↑
Frontal (39)	SSI	1.782	0.074	-2.509	3.016	0.252	-4.795
	ZUNA	1.014	0.248	-0.061	1.000	**0.359**	**2.114**
	DeepConv	**0.915**	**0.293**	**0.385**	**0.921**	0.281	0.356
	SRGDiff(+SSI)	0.999	0.034	0.004	1.111	0.313	-0.458
	NeuroX-EEG	0.995	0.113	0.020	0.990	0.260	0.044
Temporal (46)	SSI	3.230	0.086	-5.092	1.364	0.518	-1.349
	ZUNA	1.090	0.196	-0.376	0.999	0.368	**2.934**
	DeepConv	1.371	0.012	-1.372	0.866	0.369	0.624
	SRGDiff(+SSI)	0.999	0.038	0.005	0.922	0.459	0.351
	NeuroX-EEG	**0.922**	**0.301**	**0.353**	**0.625**	**0.627**	2.072
Parietal (34)	SSI	2.262	0.092	-3.545	2.356	0.361	-3.721
	ZUNA	1.191	0.233	-0.760	1.000	0.233	**1.660**
	DeepConv	1.122	0.075	-0.500	0.911	0.327	0.403
	SRGDiff(+SSI)	1.040	0.007	-0.171	0.971	0.345	0.128
	NeuroX-EEG	**0.927**	**0.455**	**0.329**	**0.848**	**0.595**	0.779
Occipital (41)	SSI	1.830	0.110	-2.624	5.013	0.396	-7.011
	ZUNA	1.027	**0.300**	-0.114	1.000	0.346	**2.104**
	DeepConv	1.148	0.025	-0.601	0.898	0.320	0.465
	SRGDiff(+SSI)	**0.997**	0.055	**0.012**	**0.808**	**0.536**	0.926
	NeuroX-EEG	1.156	0.142	-0.630	0.958	0.253	0.192
Macro mean	SSI	2.276	0.091	-3.443	2.937	0.382	-4.219
	ZUNA	1.081	0.244	-0.328	1.000	0.327	**2.203**
	DeepConv	1.139	0.101	-0.522	0.899	0.324	0.462
	SRGDiff(+SSI)	1.009	0.034	-0.038	0.953	0.413	0.237
	NeuroX-EEG	**1.000**	**0.253**	**0.018**	**0.855**	**0.434**	0.772

NeuroX-EEG achieves the best macro-average NMSE and PCC across the four regional holdouts on both datasets and the best macro-average SNR on SEED; ZUNA leads macro-average SNR on SEED-V. The clearest gains occur for the band-shaped Temporal and Parietal holdouts. Frontal and Occipital metrics are divided among methods, consistent with the greater difficulty of one-sided spatial support. Overall, NeuroX-EEG is most accurate for the band-shaped holdouts, whereas for the polar Frontal and Occipital holdouts its NMSE stays near 1 (0.958 to 1.156), close to a zero prediction.

4.3 Downstream BCI decoding

We next test whether super-resolved EEG is useful beyond signal reconstruction by evaluating downstream BCI decoding. We compare low-density, super-resolved, and high-density inputs using the same emotion recognition decoder. Table 3 reports downstream decoding performance on the SEED dataset for 3-class emotion recognition. Under a leave-one-subject-out protocol over test subjects 13–15, we report the average across configurations for each scale. With three held-out test subjects, these decoding results indicate the usefulness of reconstructed signals under this protocol rather than subject-general performance.

Overall, all super-resolution methods outperform the low-density baseline in BACC across all scales, showing that zero-shot super-resolution benefits downstream decoding. NeuroX-EEG achieves the best performance at both $2\times$ and $4\times$ across all evaluation metrics. At $8\times$, DeepConv and SSI achieve higher BACC than NeuroX-EEG (0.4918 and 0.4847 vs. 0.4640) despite higher NMSE (Table 1); at this sparsity, lower reconstruction error alone does not guarantee better decoding.

Table 3: Decoding performance across different super-resolution scales on EEG classification. Models are trained on low-density (LD), super-resolved (SR), and high-density ground truth (HD) signals. Each SR method is evaluated by feeding reconstructed signals into the same decoder (EEGNet). HD serves as a high-density reference. Metrics are reported as mean values.

	2×			4×			8×		
Input	BACC	Kappa	F1	BACC	Kappa	F1	BACC	Kappa	F1
LD	0.4118	0.1203	0.3491	0.4498	0.1777	0.3687	0.4485	0.1756	0.3758
SSI	0.4616	0.1972	0.3971	0.4668	0.2037	0.3991	0.4847	0.2326	0.4209
ZUNA	0.4373	0.1589	0.3774	0.4513	0.1729	0.3763	0.4697	0.2073	0.4067
DeepConv	0.4413	0.1650	0.3762	0.4658	0.2050	0.3796	**0.4918**	**0.2409**	**0.4308**
SRGDiff (+ SSI)	0.4323	0.1497	0.3659	0.4510	0.1771	0.4000	0.4494	0.1670	0.3828
NeuroX-EEG	**0.4644**	**0.2003**	**0.4082**	**0.4842**	**0.2138**	**0.4126**	0.4640	0.1971	0.4012
HD (GT)	0.4972	0.2497	0.4565	0.4972	0.2497	0.4565	0.4972	0.2497	0.4565

Table 4: Ablation studies on zero-shot channels (scale 2×, SEED). **Left**: REPA training strategy variants with the same backbone. **Right**: Spatial representation comparison between flattened and 2D grid formulations. NMSE indicates lower is better, while PCC and SNR indicate higher is better.

Method	NMSE ↓	PCC ↑	SNR ↑	Method	NMSE ↓	PCC ↑	SNR ↑
REPA Training Strategy				***Spatial Representation***			
no REPA	0.846	0.411	0.735	Flattened	6.269	0.003	-7.890
vanilla REPA	0.840	0.413	0.765	2D Grid (Ours)	**0.833**	**0.421**	**0.803**
Noise-weight	0.844	0.404	0.744				
Noise-weight + Earlystop 20%	**0.833**	**0.421**	**0.803**				

4.4 Ablation study

Table 4 supports complementary roles for the spatial grid and foundation-model representation. The 2D lattice yields better reconstruction in the evaluated comparison: relative to the flattened sequence, it improves NMSE from 6.269 to 0.833, PCC from 0.003 to 0.421, and SNR from -7.890 to 0.803. Representation alignment provides a small further gain on this backbone: the scheduled strategy improves all three metrics over no alignment (NMSE 0.846 to 0.833, PCC 0.411 to 0.421, and SNR 0.735 to 0.803) and performs best among the evaluated alignment variants.

5 Conclusion

We proposed NeuroX-EEG for zero-shot EEG channel reconstruction, where target electrodes receive neither conditioning nor supervision during training. By representing partial recordings on a shared 2D scalp grid, NeuroX-EEG turns channel reconstruction into field inpainting with explicit local topology. Pretrained EEG foundation-model representations, added through training-time alignment, supply broader temporal, spectral, and cross-channel context and give a small further gain in the evaluated ablation. Across SEED, SEED-V, and Mumtaz2016, NeuroX-EEG achieves the lowest NMSE among the evaluated methods under our zero-shot protocol in six of seven dataset–scale settings, reconstructs signals at electrode locations never used during training, and yields higher mean downstream decoding performance than low-density input. Region-wise experiments show the clearest gains for band-shaped holdouts and reveal how performance varies across holdout geometries. These results support 2D field inpainting with foundation-model guidance as an effective approach to EEG reconstruction beyond fixed supervised channel mappings, within the evaluated datasets and scalp lattice.

AI use statement

In this work, we used generative AI tools (Claude, Anthropic, and ChatGPT, OpenAI) for the following required-disclosure tasks: **implementing methods**, including code for the core model, training, and evaluation; **designing and giving feedback on research methodology**, including discussing how findings from our literature search should inform our method and evaluation design; and **in-**

terpreting results, including evaluating candidate explanations for experimental findings. We have not used generative AI tools for the other required-disclosure tasks (generating synthetic datasets; developing theoretical models or conceptual frameworks; formulating mathematical claims or assisting in proofs; proposing or refining hypotheses; translation; cleaning or reformatting datasets; or qualitative or thematic data analysis).

Additionally, we used generative AI tools for the following recommended-disclosure tasks: **creating or editing software code** for auxiliary scripts (e.g., plotting and preprocessing) separate from the core method implementation above; **identifying and summarizing relevant literature** to inform the methodology decisions noted above; **finalizing the visual design of a scientific figure** (Figure 1), which we conceived, sketched, and specified the content of; and **drafting and editing parts of the manuscript text**. We have reviewed all AI-assisted work: all AI-generated code was checked and verified by the authors before use, and all experimental results reported in this paper were produced by the authors' own experiment runs and confirmed from experiment logs, not generated by AI. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

Ethics statement

This work uses the publicly available, de-identified SEED, SEED-V, and Mumtaz2016 EEG datasets, which are released for research use under their original authors' terms; no new human-subjects data was collected. We do not foresee direct harms from this work beyond those generally associated with EEG-based BCI research.

Reproducibility statement

Preprocessing, training hyperparameters, dataset splits, baseline implementations, and evaluation protocols are described in Appendix C, and the zero-shot masking procedure is given as pseudocode in Algorithm 1. We report standard deviations over repeated zero-shot configurations in Appendix D.1 where applicable.

References

Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from EEG with deep recurrent-convolutional neural networks. In *International Conference on Learning Representations (ICLR)*, 2016.

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.

Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004. doi: 10.1016/j.jneumeth.2003.10.009.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. Meg and eeg data analysis with mne-python. *Frontiers in Neuroscience*, 7:267, 2013. doi: 10.3389/fnins.2013.00267.

Danny Dongyeop Han, Yonghyeon Gwon, Ahhyun Lucy Lee, Taeyang Lee, Seong Jin Lee, Jubin Choi, Sebin Lee, Jihyun Bang, Seungju Lee, David Keetae Park, et al. Diver-1: Deep integration of vast electrophysiological recordings at scale. *arXiv preprint arXiv:2512.19097*, 2025.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Shahar Ain Kedem, Itamar Zimerman, and Eliya Nachmani. Neural brain fields: A nerf-inspired approach for generating nonexistent EEG electrodes. *arXiv preprint arXiv:2601.00012*, 2025.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. GraphCast: Learning skillful medium-range global weather forecasting. *Science*, 2023.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

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

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Step-aware residual-guided diffusion for eeg spatial super-resolution. *arXiv preprint arXiv:2510.19166*, 2025.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, Rujun Han, Shitong Duan, Kuanjian Tang, and Chao Yao. Cafe: Channel-autoregressive factorized encoding for robust biosignal spatial super-resolution. *arXiv preprint arXiv:2602.17011*, 2026a.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Coordinate-queryable neural field reconstruction for EEG spatial super-resolution with unseen-electrode generation. *arXiv preprint arXiv:2606.23707*, 2026b.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 2021.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

Wajid Mumtaz. Mdd patients and healthy controls eeg data (new). *figshare, Dataset*, 2016.

Paul L. Nunez and Ramesh Srinivasan. *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, New York, 2nd edition, 2006.

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.

Sari Saba-Sadiya, Tuka Alhanai, Taosheng Liu, and Mohammad M. Ghassemi. EEG channel interpolation using deep encoder-decoder networks. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2432–2439. IEEE, 2020.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

Ramesh Srinivasan, Paul L. Nunez, Don M. Tucker, Richard B. Silberstein, and Peter J. Cadusch. Spatial sampling and filtering of EEG with spline Laplacians to estimate cortical potentials. *Brain Topography*, 8 (4):355–366, 1996. doi: 10.1007/BF01186911.

Sally M. Stoyell, Janina Wilmskoetter, Mary-Ann Dobrota, Dhinakaran M. Chinappen, Leonardo Bonilha, Mark Mintz, Benjamin H. Brinkmann, Susan T. Herman, Jurriaan M. Peters, Serge Vulliemoz, et al. High-density EEG in current clinical practice and opportunities for the future. *Journal of Clinical Neurophysiology*, 38(2):112–123, 2021.

Yun Su, Zhixuan Zhang, Xuan Li, Bingtao Zhang, and Huifang Ma. The multiscale 3d convolutional network for emotion recognition based on electroencephalogram. *Frontiers in neuroscience*, 16:872311, 2022.

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

Pierre Thodoroff, Joelle Pineau, and Andrew Lim. Learning robust features using deep learning for automatic seizure detection. In *Machine Learning for Healthcare Conference*, volume 56, pp. 178–190. PMLR, 2016.

Tri Ton, Ji Woo Hong, and Chang D Yoo. Taro: Timestep-adaptive representation alignment with onset-aware conditioning for synchronized video-to-audio synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14228–14237, 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(1):1034–1045, 2025a.

Yingdong Wang, Qingfeng Wu, Chen Wang, and Qunsheng Ruan. De-cnn: An improved identity recognition algorithm based on the emotional electroencephalography. *Computational and Mathematical Methods in Medicine*, 2020(1):7574531, 2020.

Ziqiao Wang, Wangbo Zhao, Yuhao Zhou, Zekai Li, Zhiyuan Liang, Mingjia Shi, Xuanlei Zhao, Pengfei Zhou, Kaipeng Zhang, Zhangyang Wang, et al. Repa works until it doesn't: Early-stopped, holistic alignment supercharges diffusion training. *arXiv preprint arXiv:2505.16792*, 2025b.

Christopher Warner, Jonas Mago, JR Huml, Mohamed Osman, and Beren Millidge. Zuna: Flexible eeg super-resolution with position-aware diffusion autoencoders. *arXiv preprint arXiv:2602.18478*, 2026.

Laura Yao, Gengwei Zhang, Moajjem Chowdhury, Yunmei Liu, and Tianlong Chen. Geometry- and relation-aware diffusion for EEG super-resolution. *arXiv preprint arXiv:2602.02238*, 2026.

Mustafa Yazıcı, Mustafa Ulutaş, and Mukadder Okuyan. Effect of eeg electrode numbers on source estimation in motor imagery. *Brain Sciences*, 15(7):685, 2025.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. *arXiv preprint arXiv:2410.06940*, 2024.

Zhi Zhang, Sheng-hua Zhong, and Yan Liu. Torcheegemo: A deep learning toolbox towards eeg-based emotion recognition. *Expert Systems with Applications*, 249:123550, 2024.

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.

A Limitations

Our evaluation measures recovery at zero-shot electrode locations rather than validating a continuous field between electrodes, and because EEG potentials are reference-dependent, every target is defined under the acquisition reference shared by the input and target recordings (Section 3.1). Zero-shot targets are also restricted to cells of the fixed scalp lattice rather than arbitrary off-grid positions. The 2D grid also approximates the curved scalp surface: distortion and incomplete neighborhoods are most pronounced near its boundary, consistent with the greater difficulty of the Frontal and Occipital zero-shot regions. Point-wise recovery also remains limited under the sparsest and most one-sided conditions: at SEED $8\times$ all evaluated methods have NMSE above 1, i.e., no lower error than a zero prediction, and NeuroX-EEG's NMSE stays near 1 for the Frontal and Occipital hold-outs. PCC is led by other methods in five of seven settings, and downstream gains over low-density input are shared by the other reconstruction methods.

Mumtaz2016 broadens the evaluation to 19-channel clinical resting-state EEG, but each dataset-specific model is trained separately. The current results therefore do not establish direct transfer of one trained model to a previously unseen montage. Finally, diffusion inference uses 50 DDIM steps and remains slower than single-pass reconstruction.

B Future Work

Future work will extend the current discrete grid with continuous convolution, spherical or graph-based operators, and hybrid coordinate–grid models that retain local weight sharing while preserving scalp geometry more directly. In particular, continuous convolution over measured electrode coordinates would remove the lattice-cell restriction noted in Appendix A, extending NeuroX-EEG to a lattice-free setting in which reconstruction can be queried and evaluated at arbitrary scalp positions. Training a single model across heterogeneous montages and denser acquisition systems is a further step toward cross-montage field reconstruction. Finally, accelerated samplers, diffusion distillation, and consistency-model objectives could reduce inference cost for real-time BCI use.

C Experimental detail

C.1 Preprocessing Details

Filtering, resampling, and segmentation. SEED and SEED-V are bandpass filtered between 1 and 75 Hz, notch filtered at the applicable line-noise frequency, resampled to 200 Hz, and divided into non-overlapping 4-second segments (800 samples).

For Mumtaz2016, we read the EDF recordings and retain 19 linked-ear reference channels: Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz. The legacy temporal labels T3/T4/T5/T6 are renamed T7/T8/P7/P8 when matched to the standard 10–05 coordinate file. We use the eyes-open and eyes-closed recordings; files labeled TASK are excluded by the preprocessing pipeline. Recordings are resampled from their source rate (256 Hz for these EDFs) to 200 Hz, filtered from 0.3 to 75 Hz with MNE's zero-phase FIR implementation, and notch filtered at 50 Hz. After truncating any incomplete remainder, each recording is divided into non-overlapping 5-second segments, yielding arrays of shape $19 \times 5 \times 200$ before model input reshaping. Electrode coordinates are obtained from the standard 10–05 ELC layout in the processed channel order.

Normalization. Each EEG segment was globally normalized using the channel-wise mean μ and standard deviation σ computed exclusively from the training set:

$$\tilde{x} = \tanh\left(\frac{x - \mu}{\sigma + \epsilon}\right), \tag{7}$$

where x denotes the raw EEG amplitude in μV, $\tilde{x}$ denotes the normalized value, and $\epsilon = 10^{-6}$ is added for numerical stability. This transformation compresses the signal into $[-1, 1]$, consistent with the output range of the diffusion model. The same training-set statistics are applied to the validation and test sets, without estimating new statistics from either split. At evaluation time, reconstructed values are mapped back to raw μV space using the inverse transformation:

$$x = \operatorname{atanh}(\tilde{x}) \cdot (\sigma + \epsilon) + \mu. \tag{8}$$

Artifact Rejection. For SEED and SEED-V, prior to normalization, sessions with poor signal quality were discarded based on three criteria: (i) any channel exhibiting a peak amplitude exceeding 20,000 μV, (ii) the median standard deviation across channels exceeding 100 μV, or (iii) any individual channel whose standard deviation is more than 20 times the median standard deviation across channels. Sessions meeting any of these criteria were removed entirely to prevent corrupted recordings from affecting model training or evaluation.

C.2 Zero-Shot Masking Procedure

Algorithm 1 summarizes the training-time masking procedure used in our zero-shot EEG channel reconstruction framework. At each iteration, a subset of seen channels $\mathcal{O} \subset \mathcal{S}$ is randomly sampled and provided as observed conditioning input, while the remaining seen channels $\mathcal{S} \setminus \mathcal{O}$ are masked and treated as reconstruction targets. In contrast, the zero-shot channel set $\mathcal{U}$ is excluded from both conditioning and reconstruction supervision throughout training.

The algorithm also distinguishes between valid electrode positions and empty grid cells via the mask M^{valid}, and defines the corresponding observation mask M^{obs} and spatially varying loss weights $\mathbf{w}$. This formulation makes explicit the roles of observed, masked, and zero-shot channels, and clarifies how the model is trained to reconstruct missing signals from partial spatial context.

C.3 Training Details and Hyperparameters

Input representation. For SEED and SEED-V, each input segment has length 800 time points (4 seconds at 200 Hz); for Mumtaz2016, each segment has 1,000 time points (5 seconds at 200 Hz), and the model is trained separately on this length. EEG channels are projected onto an 11×11 spatial grid.

Foundation-model teacher. SEED, SEED-V, and Mumtaz2016 were not included in the pretraining corpus of the DIVER-1 model used in our experiments.

Flattened ablation. For the flattened ablation (Table 4), we vectorize the spatial grid along the channel dimension and replace the grid convolutional backbone with its one-dimensional counterpart while retaining the same reconstruction objective.

Algorithm 1 Training-Time Mask Construction for Zero-Shot EEG Channel Reconstruction

Require: Full EEG grid $\mathbf{G} \in \mathbb{R}^{T\times H\times W}$, seen channel set $\mathcal{S}$, zero-shot unseen set $\mathcal{U}$, valid-electrode mask M^{valid}, observation ratio ρ, observed-channel weight λ_{obs}
Ensure: Observed condition $\mathbf{C}$, observation mask M^{obs}, loss weights $\mathbf{w}$

Sample an observed subset $\mathcal{O} \subset \mathcal{S}$ such that $|\mathcal{O}| \approx \rho|\mathcal{S}|$
Define the masked seen target set $\mathcal{M} \leftarrow \mathcal{S} \setminus \mathcal{O}$
Initialize $\mathbf{C} \leftarrow \mathbf{0} \in \mathbb{R}^{T\times H\times W}$
Initialize $M^{\text{obs}} \leftarrow \mathbf{0} \in \{0,1\}^{H\times W}$
Initialize $\mathbf{w} \leftarrow \mathbf{0} \in \mathbb{R}^{H\times W}$
for each grid position (r, c) **do**
  if $M^{\text{valid}}_{r,c} = 0$ **then**
    $\mathbf{C}_{:,r,c} \leftarrow 0$
    $M^{\text{obs}}_{r,c} \leftarrow 0$
    $w_{r,c} \leftarrow 0$
  else if $(r, c) \in \mathcal{O}$ **then**
    $\mathbf{C}_{:,r,c} \leftarrow \mathbf{G}_{:,r,c}$
    $M^{\text{obs}}_{r,c} \leftarrow 1$
    $w_{r,c} \leftarrow \lambda_{\text{obs}}$
  else if $(r, c) \in \mathcal{M}$ **then**
    $\mathbf{C}_{:,r,c} \leftarrow 0$
    $M^{\text{obs}}_{r,c} \leftarrow 0$
    $w_{r,c} \leftarrow 1$
  else if $(r, c) \in \mathcal{U}$ **then**
    $\mathbf{C}_{:,r,c} \leftarrow 0$
    $M^{\text{obs}}_{r,c} \leftarrow 0$
    $w_{r,c} \leftarrow 0$
  end if
end for
return $\mathbf{C}$, M^{obs}, $\mathbf{w}$

Backbone architecture. The UNet backbone has four components. • **Encoder:** Three resolution stages with channel multipliers $(1, 2, 4)$ of base width $d = 128$ (Table 5), each containing two residual blocks, with spatial downsampling by a factor of 2 (stride in H, W only) between stages, preserving the full temporal resolution T throughout. • **Bottleneck:** Two residual blocks at the coarsest spatial scale, producing a feature map $\mathbf{h}_{\text{mid}} \in \mathbb{R}^{4d\times T\times \frac{H}{4}\times \frac{W}{4}}$ that also serves as the intermediate representation for the alignment loss described in Section 3.6. • **Decoder:** A symmetric upsampling path with skip connections from the encoder, followed by a GroupNorm, SiLU, $\mathrm{Conv3d}$ output head that produces $\hat{\boldsymbol{\epsilon}} \in \mathbb{R}^{1\times T\times H\times W}$. • **Timestep conditioning:** The diffusion step t is embedded via a two-layer MLP with SiLU activations and injected into each residual block's feature map as an additive bias.

For diffusion model evaluation, we select the checkpoint with the lowest validation NMSE.

For Mumtaz2016, we use a subject-disjoint split of 43 training, 9 validation, and 11 test participants, stratified by diagnostic group as shown in Table 6. Eyes-open and eyes-closed recordings from the same participant remain in the same partition. The diagnostic label is stored with each segment but is not used by the reconstruction model.

Repeated evaluations. For the SEED and SEED-V scale-based reconstruction results, each scale is evaluated over four distinct zero-shot channel configurations at $4\times$ and $8\times$ and three at $2\times$ (two of the four $2\times$ configurations in Tang et al. (2022) are identical), and we report the corresponding mean and standard deviation. Mumtaz2016 and the region-wise experiments each use one predefined zero-shot configuration and therefore do not have repeated-configuration standard deviations. For EEGNet decoding, each scale is evaluated over its zero-shot configurations and three test subjects (9 evaluations at $2\times$ and 12 at $4\times$ and $8\times$).

Table 5: Diffusion model training hyperparameters.

Hyperparameter	Value
Optimizer	AdamW
Maximum learning rate	4×10^{-5}
Minimum learning rate	8×10^{-7}
Weight decay	10^{-4}
Batch size	4
Learning rate scheduler	Cosine annealing
Warm-up	First 10% of training iterations
Training iterations on SEED	20,000
Training iterations on SEED-V	80,000
Training iterations on Mumtaz2016	80,000
Training diffusion steps	1,000
Inference sampler	DDIM
Inference sampling steps	50
U-Net base dimension	128
Dimension multipliers	$(1, 2, 4)$
Denoising loss	ℓ_2 loss
Alignment loss weight (λ_{align})	0.5
Observation ratio ρ	0.5
Observed-channel weight (λ_{obs})	0.1

Table 6: Dataset splits and checkpoint selection for diffusion model training. Mumtaz2016 uses a subject-disjoint split, with all recordings from a participant assigned to the same partition.

Dataset	Training split	Validation split	Test split
SEED	1–11	12	13–15
SEED-V	1–12	13	14–16
Mumtaz2016	43 subjects (21 H, 22 MDD)	9 subjects (4 H, 5 MDD)	11 subjects (5 H, 6 MDD)

For downstream decoding, we perform leave-one-subject-out evaluation over the three reconstruction test subjects. In SEED, subjects 13–15 serve in turn as the decoder test subject, while the other two are used for training.

Compute. All diffusion model experiments were conducted on a single NVIDIA A100-PCIE-40GB GPU. Training and inference for the SEED dataset required approximately 2 hours, while training and inference for the SEED-V dataset required approximately 7 hours. The diffusion model contains 134.24M parameters.

C.4 Channel Grid Mapping and Zero-Shot Channel Configurations

2D channel grid mapping. To represent EEG signals in a spatially structured format, we mapped each dataset's montage onto a two-dimensional channel grid, using the TorchEEG 2D electrode layout (Zhang et al., 2024) as the specific (r, c) assignment. This representation preserves the approximate topographical arrangement of electrodes and enables the model to process multi-channel signals as spatially organized inputs. SEED and SEED-V use the same 62-channel reference layout for the scale-based and region-wise experiments. Mumtaz2016 uses the corresponding positions of its standard 19-channel 10–20 montage on the same reference grid. The complete 62-channel layout is illustrated in Figure 4.

We use the TorchEEG layout as a reproducible electrode-to-cell assignment consistent with prior 2D EEG representations (Delorme & Makeig, 2004; Bashivan et al., 2016; Thodoroff et al., 2016; Wang et al., 2020; Su et al., 2022). The mapping places each measured channel at one grid cell and leaves unassigned cells empty; it does not spatially interpolate signals before model input.

Table 7: EEGNet decoding hyperparameters.

Hyperparameter	Value
Optimizer	Adam
Learning rate	10^{-3}
Loss function	Cross-entropy loss
Epochs	50
Batch size	32
Validation set	None
Checkpoint used	Final checkpoint
Dropout	0.5
Temporal kernel size	64
Separable convolution kernel size	16
F_1	8
F_2	16
Depth multiplier D	2

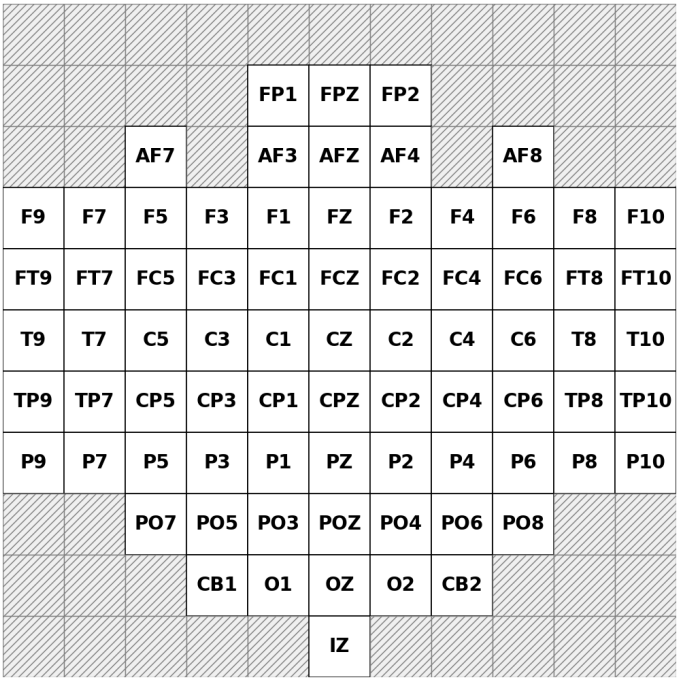


Figure 4: Two-dimensional EEG channel grid mapping used in this study. Gray hatched cells indicate unused grid positions.

Scale-based zero-shot channel configurations. For the scale-based zero-shot settings, we followed the emotion-related channel configurations proposed in previous work (Tang et al., 2022). Specifically, the low-density channel configurations with scale factors of 2, 4, and 8 were used as the seen-channel sets, and the remaining channels in the full montage were treated as zero-shot targets. The full montage consisted of 62 channels distributed over frontal, temporal, central, parietal, and occipital regions.

Mumtaz2016 zero-shot channel configuration. Following the 10–20 protocol of TopoDiff (Yao et al., 2026), the full montage is {Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, O1, O2, T7, T8, P7, P8}, using the standard names of the legacy T3/T4/T5/T6 labels (Appendix C.1). The seen set is {Fp1, Fp2, F3, F4, C3, C4, O1, O2, Cz}; the remaining channels {F7, F8, Fz, P3, P4, Pz, T7, T8, P7, P8} are excluded from both conditioning and supervision during training and evaluated as

zero-shot targets. This protocol defines a single seen/target split; we adopt it unchanged so that the Mumtaz2016 setting remains directly comparable to prior work, and report it as the approximately $2\times$ configuration.

Region-wise zero-shot channel configurations. In addition to the scale-based configurations, we evaluated region-wise zero-shot settings. In these experiments, spatially contiguous electrode groups associated with broad Frontal, Temporal, Parietal, or Occipital scalp locations were designated as zero-shot targets, while all remaining channels were used as seen channels. These names are descriptive sensor-location labels rather than cortical parcellations. Because EEG sensor regions do not form strict non-overlapping anatomical boundaries, groups may partially overlap. The configurations are summarized in Table 8.

Table 8: Region-wise zero-shot channel configurations. For each setting, the listed channels were excluded from conditioning and supervision as zero-shot targets, while all other channels in the full montage were used as seen channels. Region names describe approximate scalp sensor locations rather than cortical parcellations.

Zero-shot region	# Zero-shot	Zero-shot channels
Frontal	23	Fp1, Fpz, Fp2, AF3, AF4, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8
Temporal	16	F7, F8, FT7, FT8, FC5, FC6, C5, C6, T7, T8, CP5, CP6, TP7, TP8, P7, P8
Parietal	28	C1, C2, C3, C4, C5, C6, Cz, CP1, CP2, CP3, CP4, CP5, CP6, CPz, P1, P2, P3, P4, P5, P6, P7, P8, Pz, PO3, PO4, PO5, PO6, POz
Occipital	21	P1, P2, P3, P4, P5, P6, P7, P8, Pz, PO3, PO4, PO5, PO6, PO7, PO8, POz, CB1, CB2, O1, Oz, O2

C.5 BASELINES

Spherical Spline Interpolation (SSI). SSI (Perrin et al., 1989) is applied directly to the observed channels at inference time using standard electrode coordinates from the 10–05 system. We use the default implementation provided by MNE-Python, with all hyperparameters set to their default values. No training is required. SSI serves as a classical geometric baseline that does not leverage any learned spatial representations.

DeepConv. DeepConv (Saba-Sadiya et al., 2020) is reimplemented from the officially released code and adapted to our zero-shot protocol. Specifically, the model is trained using only the seen channels $\mathcal{S}$, with $\mathcal{U}$ excluded from both the input and the reconstruction target throughout training. At test time, the model is evaluated on the zero-shot channels $\mathcal{U}$ only, consistent with our evaluation protocol.

SRGDiff. SRGDiff (Liu et al., 2025) relies on a pretrained VAE whose latent distribution is fitted to a fixed set of training channels. Under the zero-shot setting, zero-shot channels $\mathcal{U}$ are not represented in the encoder's training distribution, which may limit direct reconstruction at those locations. To provide a meaningful diffusion-based reference under this constraint, we adopt a hybrid evaluation strategy: SRGDiff first reconstructs the seen channel set $\mathcal{S}$, and SSI is subsequently applied to extend coverage to the zero-shot channels $\mathcal{U}$. This hybrid approach allows SRGDiff to participate in the zero-shot comparison while respecting the limitations of its VAE encoder.

ZUNA. ZUNA (Warner et al., 2026) is evaluated using the officially released pretrained weights without any fine-tuning. Since ZUNA conditions on continuous 3D electrode coordinates via 4D rotary positional encodings, it can directly perform inference at arbitrary channel locations, including the zero-shot channels $\mathcal{U}$, without any architectural modification.

Note on baseline adaptation. The baselines above are evaluated under different training regimes because no single protocol applies uniformly to all of them under the zero-shot setting. ZUNA's official fine-tuning code is not publicly released, so we evaluate it with its released pretrained weights

as-is rather than fine-tuning it on our splits. SRGDiff is not designed for zero-shot channel reconstruction; however, this limitation is shared by most existing EEG spatial super-resolution methods, and omitting it would leave no recent diffusion-based SR method in the comparison, so we include SRGDiff under the hybrid adaptation described above.

C.6 Evaluation Metrics

We evaluate reconstruction quality using signal-level metrics. Let $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^N$ denote the flattened predicted and ground-truth EEG signals over selected channels and time points.

Normalized Mean Squared Error (NMSE).

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N} y_i^2 + \epsilon} \tag{9}$$

Signal-to-Noise Ratio (SNR).

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_{i=1}^{N} y_i^2}{\sum_{i=1}^{N}(x_i - y_i)^2 + \epsilon} \right) \tag{10}$$

Pearson Correlation Coefficient (PCC).

$$\mathrm{PCC} = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Var}(x)\,\mathrm{Var}(y)}} \tag{11}$$

where

$$\mathrm{Cov}(x, y) = \frac{1}{N} \sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y} \tag{12}$$

D Additional Experimental Results

D.1 Full Quantitative Results with Standard Deviations

This section provides the full quantitative results corresponding to the repeated evaluations reported in the paper. Table 9 reports variability across the zero-shot configurations (three at $2\times$, four at $4\times$ and $8\times$) for each SEED and SEED-V scale. Mumtaz2016 uses one predefined $2\times$ configuration, so no across-configuration standard deviation is defined; its results are therefore reported in the main table (Table 1) but not repeated in this variance table. Table 10 similarly reports decoding performance across the repeated EEGNet evaluations. These results are consistent with the trends in the main text while quantifying variability where repeated configurations are available. For some settings, the margin between the best and second-best mean is within one standard deviation across configurations; we therefore read these rankings as trends across zero-shot channel sets.

D.2 Loss Mask Ablation

Table 11 compares two loss weighting strategies under the zero-shot channel setting at scale $2\times$ on SEED. Supervising only the masked seen channels degrades point-wise reconstruction fidelity across all metrics relative to the weighted variant. Assigning a small auxiliary weight ($\lambda_{\mathrm{obs}} = 0.1$) to the observed channels alongside the primary masked-channel supervision improves all three metrics substantially (NMSE: 0.833 vs. 0.912, PCC: 0.421 vs. 0.386, SNR: 0.803 vs. 0.408). We attribute this to the nature of zero-shot generalization: the model must infer signals at locations that receive no direct supervision, and maintaining a weak reconstruction signal on the observed channels encourages the model to preserve global spatial context throughout training rather than focusing solely on the masked regions.

D.3 Visualization Example

Figure 5 presents an example of the time-wise visualization of EEG signals from the SEED dataset scale 2 setting. Figure 6 illustrates the corresponding topographic brain activity patterns.

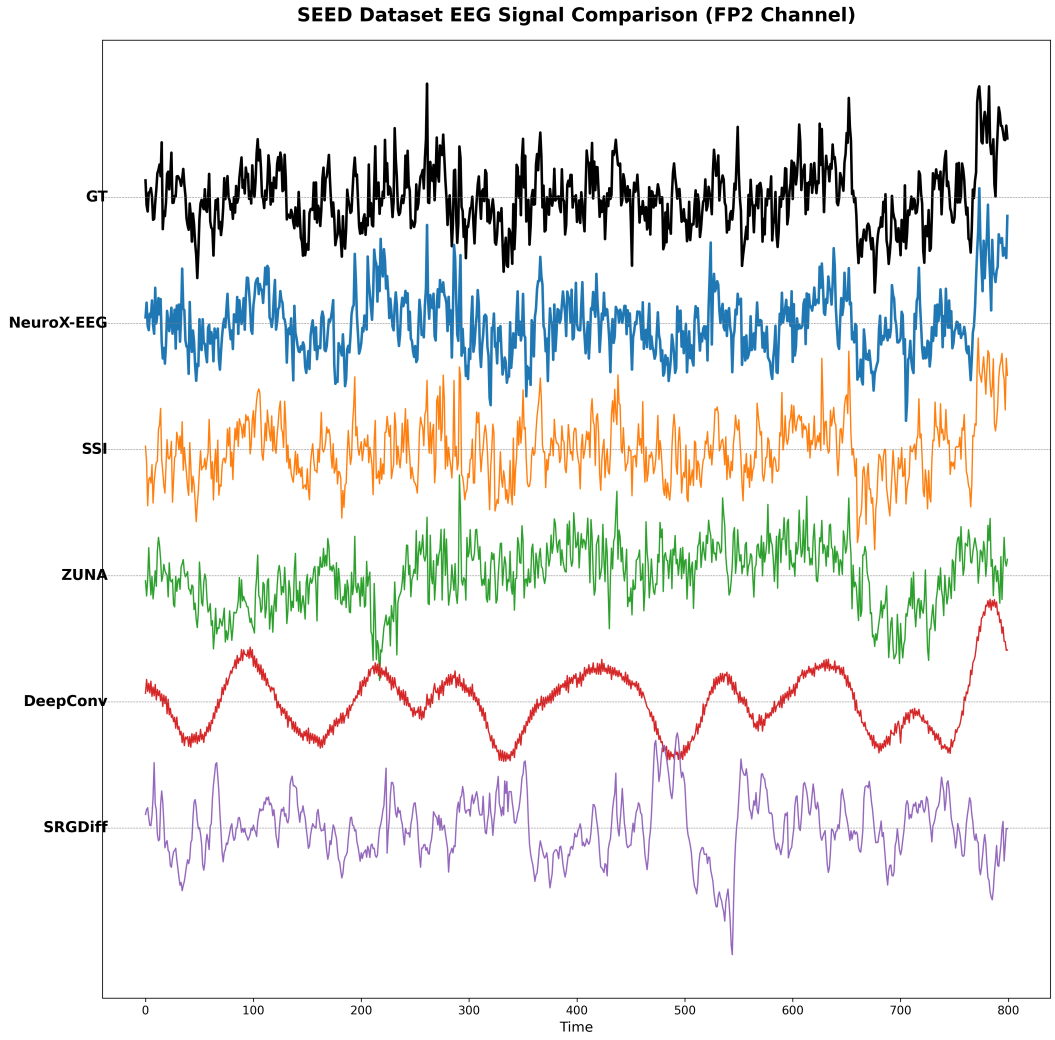


Figure 5: Time-wise visualization of EEG signals from the SEED dataset.

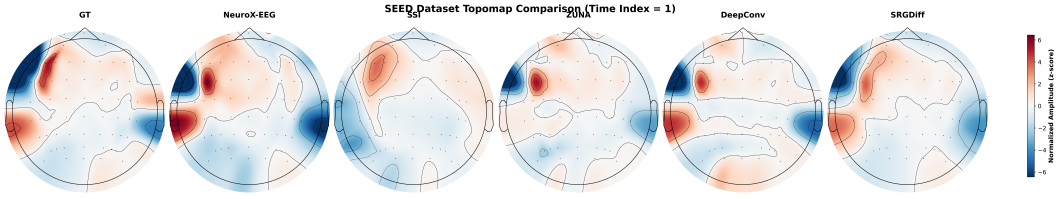


Figure 6: Topographic visualization of the SEED dataset.

Table 9: Zero-shot reconstruction performance with standard deviations across zero-shot channel configurations (three at $2\times$, four at $4\times$ and $8\times$) on SEED and SEED-V. Mumtaz2016 is omitted because it uses a single predefined configuration and therefore has no across-configuration standard deviation. NMSE indicates lower is better, while PCC and SNR indicate higher is better. Best mean results are in bold and second-best mean results are underlined.

Method	Metric	SEED (62)			SEED-V (62)		
		$2\times$	$4\times$	$8\times$	$2\times$	$4\times$	$8\times$
SSI†	NMSE ↓	1.417± 0.071	1.508± 0.512	2.182± 1.042	0.610± 0.040	0.872± 0.170	1.836± 0.960
	PCC ↑	0.196± 0.083	0.149± 0.020	0.139± 0.072	**0.691**± 0.032	0.623± 0.056	0.469± 0.110
	SNR ↑	-1.510± 0.217	-1.622± 1.308	-2.973± 2.254	2.152± 0.285	0.650± 0.786	-2.180± 2.323
ZUNA‡	NMSE ↓	1.039± 0.044	1.061± 0.016	1.062± 0.055	1.000± 0.001	0.999± 0.001	1.000± 0.000
	PCC ↑	0.285± 0.026	0.246± 0.006	**0.221**± 0.019	0.342± 0.031	0.276± 0.013	0.235± 0.011
	SNR ↑	-0.164± 0.181	-0.256± 0.066	-0.255± 0.221	1.998± 0.187	1.432± 0.068	**1.143**± 0.082
DeepConv*	NMSE ↓	0.956± 0.135	0.944± 0.047	1.037± 0.067	0.836± 0.031	0.885± 0.004	0.907± 0.006
	PCC ↑	0.314± 0.094	0.252± 0.083	0.124± 0.018	0.423± 0.043	0.353± 0.005	0.320± 0.012
	SNR ↑	0.223± 0.596	0.251± 0.221	-0.151± 0.270	0.778± 0.160	0.530± 0.020	0.425± 0.028
SRGDiff (+ SSI)§	NMSE ↓	1.130± 0.152	1.420± 0.371	1.809± 0.540	0.668± 0.066	**0.662**± 0.055	0.977± 0.229
	PCC ↑	0.291± 0.054	0.166± 0.016	0.123± 0.042	0.657± 0.039	**0.670**± 0.017	**0.516**± 0.065
	SNR ↑	-0.512± 0.576	-1.423± 1.042	-2.396± 1.508	1.764± 0.428	**1.803**± 0.364	0.255± 1.057
NeuroX-EEG*	NMSE ↓	**0.833**± 0.071	**0.912**± 0.020	**1.033**± 0.017	**0.604**± 0.071	0.677± 0.060	**0.868**± 0.039
	PCC ↑	**0.421**± 0.096	**0.321**± 0.030	0.216± 0.034	0.653± 0.039	0.584± 0.055	0.387± 0.045
	SNR ↑	**0.803**± 0.368	**0.401**± 0.094	**-0.139**± 0.072	**2.238**± 0.530	1.732± 0.398	0.634± 0.194

† Mathematical interpolation method without learning. ‡ Pretrained model evaluated without fine-tuning. * End-to-end learned model. § Learned reconstruction followed by interpolation.

Table 10: Decoding performance across different super-resolution scales on EEG classification with standard deviations. Models are trained on low-density (LD), super-resolved (SR), and high-density ground truth (HD) signals. Each SR method is evaluated by feeding reconstructed signals into the same decoder. HD serves as a high-density reference. Best mean results among SR methods are in bold and second-best mean results are underlined. Standard deviations are shown after ±. Method annotations follow Table 9.

	$2\times$			$4\times$			$8\times$		
Input	BACC	Kappa	F1	BACC	Kappa	F1	BACC	Kappa	F1
LD	0.4118± 0.0842	0.1203± 0.1280	0.3491± 0.1235	0.4498± 0.0949	0.1777± 0.1453	0.3687± 0.1348	0.4485± 0.0692	0.1756± 0.1051	0.3758± 0.1111
SSI	0.4616± 0.0795	0.1972± 0.1213	0.3971± 0.1151	0.4668± 0.0752	0.2037± 0.1152	0.3991± 0.0973	0.4847± 0.0997	0.2326± 0.1532	0.4209± 0.1503
ZUNA	0.4373± 0.0897	0.1589± 0.1363	0.3774± 0.1297	0.4513± 0.0821	0.1729± 0.1241	0.3763± 0.1356	0.4697± 0.1106	0.2073± 0.1682	0.4067± 0.1538
DeepConv	0.4413± 0.0735	0.1650± 0.1123	0.3762± 0.1020	0.4658± 0.0997	0.2050± 0.1588	0.3796± 0.1195	**0.4918**± 0.0927	**0.2409**± 0.1416	**0.4308**± 0.1318
SRGDiff (+ SSI)	0.4323± 0.0982	0.1497± 0.1496	0.3659± 0.1135	0.4510± 0.0548	0.1771± 0.0844	0.4000± 0.0796	0.4494± 0.0855	0.1670± 0.1271	0.3828± 0.1196
NeuroX-EEG	**0.4644**± 0.1000	**0.2003**± 0.1529	**0.4082**± 0.1376	**0.4842**± 0.0969	**0.2138**± 0.1482	**0.4126**± 0.1054	0.4640± 0.0823	0.1971± 0.1280	0.4012± 0.0963
HD (GT)	0.4972± 0.1023	0.2497± 0.1569	0.4565± 0.0765	0.4972± 0.1023	0.2497± 0.1569	0.4565± 0.0765	0.4972± 0.1023	0.2497± 0.1569	0.4565± 0.0765

E Impact Statement

NeuroX-EEG may support broader EEG analysis in sparse or heterogeneous electrode settings by reconstructing unobserved channel signals and improving downstream BCI decoding. We do not identify immediate negative societal impacts specific to the proposed technical method, but reconstructed EEG should not be treated as directly measured neural activity without appropriate validation.

Table 11: Loss mask ablation on zero-shot channels (scale $2\times$, SEED). **Unobs only** supervises only the masked seen channels $\mathcal{S} \setminus \mathcal{O}$ with unit weight, while **1 Unobs + 0.1 Obs** additionally applies a weight of $\lambda_{\text{obs}} = 0.1$ to the observed channels $\mathcal{O}$. Zero-shot channels $\mathcal{U}$ receive no supervision in either setting. NMSE indicates lower is better, while PCC and SNR indicate higher is better.

Method	NMSE ↓	PCC ↑	SNR ↑
Unobs only	0.912	0.386	0.408
1 Unobs + 0.1 Obs	**0.833**	**0.421**	**0.803**