CANDI: CONTINUOUS AMORTIZED IMPLICIT NEURAL REPRESENTATIONS FOR NEUROIMAGING DATA

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural activity unfolds as a continuous spatiotemporal field, yet most neuroimaging models operate on discrete sampling geometries imposed by acquisition hardware, such as voxel grids and fixed electrode montages. Implicit neural representations (INRs) provide a natural framework for modeling continuous signals, but conventional INRs are fitted independently to each instance, requiring costly test-time optimization and producing representations that do not transfer across subjects. We introduce **CANDI** (**C**ontinuous **A**mortized **N**euroimaging **D**ecoding via **I**NR), a common continuous-field representation framework that can incorporate modality-specific observation geometry and inductive biases to amortize continuous INR inference across subjects. CANDI uses a selective state-space backbone to encode long spatiotemporal contexts in linear time and introduces spherical harmonic positional encodings for scalp-based imaging motivated by the differential operators underlying spherical spline interpolation. These encodings capture the intrinsic geometry and smoothness of continuous fields observed on two-dimensional scalp manifolds. CANDI outperforms current amortized INR baselines on 4D fMRI spatiotemporal reconstruction and outperforms classical spherical splines on EEG unseen-electrode spatial querying on hold-out sets of up to 87.5% unseen electrodes, while producing representations and reconstructions that retain strong downstream-task utility.

1 INTRODUCTION

Neural activity is a continuous field in space and time, but every neuroimaging modality observes it through a discrete sampling geometry fixed by its instrument. The two dominant modalities sample this field in complementary ways; Electroencephalography (EEG) records scalp potentials with millisecond resolution, but only at a few dozen electrodes on a curved 2D surface, in layouts that vary across montages (Jasper, 1958; Nunez & Srinivasan, 2006). Functional MRI (fMRI) measures the blood-oxygen-level-dependent signal, a hemodynamic proxy for neural activity (Logothetis et al., 2001; Boynton et al., 1996), on a dense 3D voxel grid with high spatial but low temporal fidelity, and often resorts to Region-of-Interest (ROI) based generalizations for analysis. These measurements are necessarily discrete, but a further limitation arises when models bind their representations to the acquisition geometry rather than integrating the continuous structure of the underlying signal.

Recent work has loosened this framing; fMRI foundation models have moved from atlas-level ROI inputs (Ortega Caro et al., 2024) to atlas-free voxel and patch inputs that avoid within-parcel information loss (Kim et al., 2023b; Wang et al., 2026), yet their decoders still emit values on the training grid. EEG spatial super-resolution methods (Tang et al., 2023b; Li et al., 2025; Liu et al., 2026b; Yao et al., 2026) interpolate within a predefined electrode set, so they cannot transfer across montages or predict at arbitrary positions absent from training. The classical alternative, spherical spline interpolation (Perrin et al., 1989), models a continuous scalp field but fits each recording independently and does not have the ability to learn across subjects or model latent representations. Existing learned models remain tied to discrete sampling geometries, while classical continuous interpolants do not learn across subjects.

Generalizable implicit neural representations (GINRs) sit at this intersection. An INR maps coordinates to signal values and so represents a field independently of how it was sampled (Sitzmann et al., 2020; Tancik et al., 2020), but it must be fit to each instance. GINRs amortize this fit by conditioning

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.

a shared coordinate decoder on latents that an encoder infers from each observation (Chen & Wang, 2022a; Doyup et al., 2023). Existing GINRs use Transformer encoders, whose quadratic cost forces 4D fMRI into coarse patches, and their standard coordinate encodings ignore the spherical geometry of the scalp.

Our model, CANDI, is a GINR that pairs a bidirectional Mamba encoder (Gu & Dao, 2024a) with a cross-attention coordinate decoder across either modality. Linear encoder scaling lets CANDI tokenize fMRI at patch sizes that exhaust Transformer memory. For EEG, we carry the smoothness bias of spherical splines into the learned model through spherical harmonic positional encodings and a bending energy penalty. Within-montage evaluation cannot separate a learned field from a learned montage, so we hold out $\sim 40\%$ to 87.5% of electrodes from training and score interpolation only on these. Unlike sparse EEG, fMRI is spatially dense and its temporal frequency information is physically constrained by the slow BOLD response; therefore our primary objective is showing that scalable continuous reconstruction provides useful representation pretraining for dense 4D fMRI. Across EEG interpolation, fMRI reconstruction, and downstream tasks on both modalities, we demonstrate the utility of the learned continuous representations. Figure 1 summarizes the two complementary settings studied in this work on both modalities. Our contributions are threefold:

- To our knowledge, **CANDI** is the first **GINR representation framework demonstrated across both 4D fMRI and EEG**, providing a shared amortized, coordinate-queryable formulation while accommodating modality-specific observation geometry.
- We introduce a Mamba-based GINR with interleaved learnable tokens and propose **Sequential Locality Bias (SLB)**, which uses the Mamba's serialized sequence structure as a retrieval prior for coordinate-to-latent cross-attention, empirically enhancing fine-grained reconstruction of long 4D fMRI sequences.
- For EEG, we translate the geometry underlying spherical-spline interpolation into learned field reconstruction by proposing **Spherical Harmonic Positional Encoding (SHPE)** paired with field bending-energy regularization, and introduce an **unseen-electrode evaluation paradigm** spanning $\sim$ 40–87.5% held-out electrode locations.

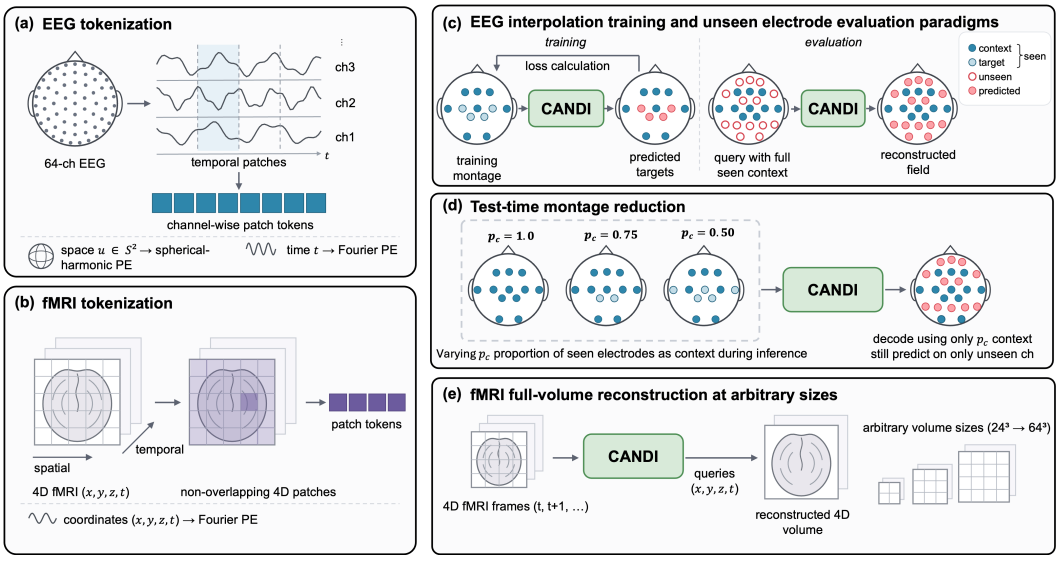

Figure 1: **CANDI problem settings.** (a)-(b) illustrate specifics of tokenization for each data modality. (c)-(d) describe the sparse training and unseen electrode evaluation paradigms, where sparsity can vary in the number of seen electrodes and context provided at test time. (e) describes the fMRI 4D volume reconstruction setting.

2 Related Work

Generalizable INRs and their backbones. GINRs fall into three families: hypernetworks that emit decoder weights (Wu et al., 2023; Szatkowski et al., 2023; Fons et al., 2022), meta-learned initializations that adapt to each instance in a few gradient steps (Dupont et al., 2022), and conditioned decoders that read an instance latent through attention (Chen & Wang, 2022a; Kim et al., 2023a; Doyup et al., 2023; Zhang et al., 2024). Conditioned decoders avoid inner-loop optimization and weight generation, so we adopt them. Current conditioned decoders encode with Transformers, whose cost grows quadratically with input length, and their latents carry no explicit link to input position, so locality must emerge from learned cross-attention (Doyup et al., 2023). Efficient Transformers approximate attention with low-rank projections or kernels (Wang et al., 2020; Choromanski et al., 2021; Jaegle et al., 2022) and trade fidelity for speed (Tay et al., 2022). Structured state space models instead reach linear cost with an exact recurrence (Gu et al., 2022), and Mamba makes that recurrence input-selective (Gu & Dao, 2024b). ALiBi shows that a bias on sequence distance improves attention without learned position embeddings (Press et al., 2022). CANDI brings both ideas into INR conditioning, pairing a Mamba encoder with a cross-attention bias on sequence distance.

fMRI representation learning. BrainLM models activity in atlas parcels and discards within-parcel signal (Ortega Caro et al., 2024). SwiFT and OmniFMRI operate on voxel patches but learn representations for prediction and provide no means to query the signal at a coordinate (Kim et al., 2023b; Wang et al., 2026). Single-instance INRs reconstruct MR images at high fidelity but fit each scan separately and return no latent that transfers across subjects (Wu et al., 2021; Shen et al., 2024). Bayesian nonparametric approaches, including Gaussian-process and kriging models, have also represented neuroimaging signals as continuous spatiotemporal fields through explicit covariance priors (Gössl et al., 2001; Woolrich et al., 2004; Harrison et al., 2008). CANDI instead focuses on amortized neural-field inference with learned latent representations and coordinate decoding across subjects.

Scalp field interpolation. Spherical spline interpolation (the field standard for EEG interpolation) fits each recording by minimizing bending energy on the sphere, which damps high-degree spherical harmonics (Perrin et al., 1989; Wahba, 1981). Spherical harmonics also serve as positional encodings for learned fields on the sphere (Cohen et al., 2018; Rußwurm et al., 2024), but no prior work pairs them with a curvature penalty on scalp potentials. Almost all EEG super-resolution methods learn a map from one fixed low-density montage to one fixed high-density montage and cannot query a new position (Tang et al., 2023b; Li et al., 2025; Liu et al., 2026b; Yao et al., 2026). Most EEG foundation models accept heterogeneous montages but target classification rather than field reconstruction (El Ouahidi et al., 2025; Döner et al., 2025). Two recent models do query new positions. Liu et al. (2026a) hold out only 10% of electrodes, and ZUNA encodes positions with a Euclidean rotary encoding after pretraining on dense montages from 208 datasets (Warner et al., 2026). Both learn from dense ground truth. Learning the scalp field from sparse montages alone, and querying positions that no training recording contains, therefore remains open.

3 Methods

3.1 Preliminaries

Implicit Neural Representations. A data instance $x^{(i)} = \{(c_j^{(i)}, y_j^{(i)})\}_{j=0}^{m_n}$ of some dataset $X = \{x^{(i)}\}_{i=0}^{n}$ is defined as having a set of m_n paired coordinate locations $c_j^{(i)} \in \mathbb{R}^{d_{in}}$ and values $y_j^{(i)} \in \mathbb{R}^{d_{out}}$ (e.g., a 2D color image where $d_{in} = 2$ and the RGB value of each pixel is $d_{out} = 3$). An INR f_θ parameterized by parameters θ is a neural network:$f_\theta : \mathbb{R}^{d_{in}} \to \mathbb{R}^{d_{out}}$ that takes coordinates $c \in \mathbb{R}^{d_{in}}$ as inputs and outputs data values $y \in \mathbb{R}^{d_{out}}$. Generalizable INRs (GINRs) extend this concept, parameterizing the implicit representation such that a single network can represent multiple signals from a family and generalize to unseen signals. This is often achieved by conditioning the network on some latent representation or embedding z.

Mamba SSM. Structured state space models (SSMs) Gu et al. (2022) map an input sequence x to an output y through a latent state h governed by the linear ODE $h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t)$,

$y(t) = \mathbf{C}h(t)$. Discretizing with a step size Δ gives the recurrence $h_t = \overline{\mathbf{A}}h_{t-1} + \overline{\mathbf{B}}x_t$, $y_t = \mathbf{C}h_t$, where $\overline{\mathbf{A}} = \exp(\Delta\mathbf{A})$. Earlier SSMs such as S4 share these parameters across all time steps, but Mamba Gu & Dao (2024b) makes $\mathbf{B}$, $\mathbf{C}$, and Δ functions of the current input, so each token thereby controls how much of the state it either overwrites or preserves. Mamba runs in $O(L)$ time and memory linear in sequence length L, compared with the $O(L^2)$ cost of self-attention.

Spherical Spline Interpolation. Spherical spline interpolation (SSI) Perrin et al. (1989) is the state-of-the-art method to reconstruct a continuous scalp potential field from measurements at a finite set of electrodes. Given electrode potentials z_i at locations E_i on the unit sphere, the spherical-spline formulation Wahba (1981) obtains a smooth interpolant by minimizing a spherical smoothness (bending-energy) functional subject to the electrode constraints:

$$J_m[U] = \int_{S^2} \left\| \Delta_{S^2}^{m/2} U \right\|^2 dS, \qquad U = \underset{U(E_i)=z_i}{\arg\min}\, J_m[U], \tag{1}$$

where Δ_{S^2} is the Laplace–Beltrami operator on the sphere. Thus, among fields consistent with the electrode measurements, SSI favors smooth fields with low spherical curvature Wahba (1981). SSI is also naturally expressed in the spherical-harmonic basis. The corresponding spherical-spline kernel has the Legendre expansion

$$g(x) = \frac{1}{4\pi} \sum_{\ell=1}^{\infty} \frac{2\ell+1}{[\ell(\ell+1)]^m} P_\ell(x), \tag{2}$$

where the degree-dependent weighting follows from the spherical-harmonic eigenvalues of the Laplace–Beltrami operator and gives the standard spherical-spline kernel used for scalp interpolation Perrin et al. (1989). By the spherical-harmonic addition theorem, this kernel is equivalently a degree-weighted expansion in spherical harmonics, with $[\ell(\ell+1)]^{-m}$ suppressing higher-degree components. These two properties motivate our design: we use spherical harmonics as the spatial positional basis and impose an analogous bending-energy regularizer on the learned field.

3.2 CANDI

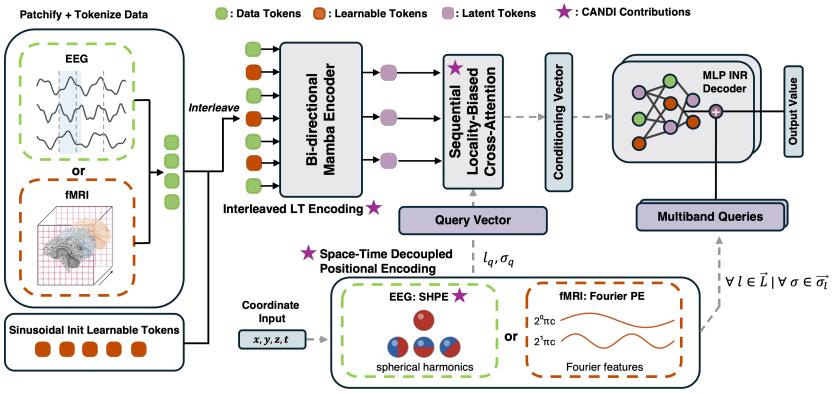

Figure 2: **CANDI architecture.** Stars denote CANDI contributions. Neuroimaging observations are tokenized and encoded with interleaved learnable tokens using a bidirectional Mamba encoder. Coordinate queries use spatiotemporally decoupled positional encodings and retrieve latent conditioning through cross-attention. The conditioning vector is passed through a hierarchical coordinate-conditioned MLP to predict signal value

CANDI is a conditioned-decoder-style GINR method, composed of a Mamba encoder which generates the set of latents $z^{(i)} = \{z_j^{(i)}\}_{j=0}^{n}$ (i representing the data instance and j representing each latent) and the INR Decoder which uses sequence locality-biased cross-attention between a coordinate query and $z^{(i)}$ as the keys to create an aggregate modulation vector that is then used to condition linear layers and output the corresponding values.

Mamba Encoder. The data instance $x^{(i)}$ is first patchified and tokenized as a set of tokens $d^{(i)} = \{d_j^{(i)} \in \mathbb{R}^d\}_{j=0}^{L_d}$, where L_d and d represent the number and dimension of the data tokens, respectively. A set of instance-agnostic learnable tokens $t = \{t_i \in \mathbb{R}^d\}_{i=0}^{L_l}$ is initialized using a fixed sinusoidal function (Valencia et al., 2025). The learnable tokens are then interleaved equidistantly within the data tokens (as shown in Figure 2) as studied by Wang et al. (2025) before being passed into the Mamba encoder to create a full token set with length $L = L_d + L_l$. The Mamba encoder itself is a set of k bi-directional Mamba blocks, where k is the depth of the encoder. Each block takes in the previous sequences and runs it forwards and backwards through the original Mamba block, applying an RMSNorm Zhang & Sennrich (2019) and residual connection. The scan order of the Mamba encoder can be varied depending on the input shape. The encoder output at the learnable token positions is extracted as the latent set $z^{(i)} = \{z_j^{(i)} \in \mathbb{R}^d\}_{i=0}^{L_l}$

Cross-Attention Based INR Decoder. We adopt the hierarchical cross-attention INR decoder of Doyup et al. (2023) (LAGINR) as the base INR architecture, but modify its coordinate-conditioning pathway in three key ways for neuroimaging: (i) spatiotemporally decoupled positional encodings, (ii) a sequential locality-biased (SLB) cross-attention motivated by the ordered Mamba latent sequence, and (iii) spherical harmonic positional encoding (SHPE) for scalp-based coordinates. Given a query coordinate $c = (c_{xyz}, c_t)$, we first construct its query representation using separate spatial and temporal positional-frequency banks:

$$\gamma_\sigma(c) = [\cos(\pi\omega_j c_i), \sin(\pi\omega_j c_i)]\,, \qquad q_c = \mathrm{ReLU}(W[\gamma_{\sigma_s}(c_{xyz}) \| \gamma_{\sigma_t}(c_t)] + b)\,. \tag{3}$$

Unlike a single shared Fourier bank, this allows spatial and temporal coordinates to use different frequency scales. The query q_c then retrieves a coordinate-specific conditioning vector v_c from the latent set $z^{(i)}$ through cross-attention. In LAGINR, "locality-aware" latent relevance is learned entirely through coordinate-based attention. CANDI's encoder provides additional structure: learnable tokens are interleaved with data tokens before the bidirectional Mamba scan, so every latent retains a well-defined position in the serialized sequence. For a query whose coordinate lies in data patch $p_c^{(i)}$, we therefore know its sequence position relative to every latent token. Because information in the SSM is propagated through sequential state updates, we hypothesize that sequence proximity provides a useful retrieval prior, without requiring the latent representations themselves to be local. We encode this prior as **Sequential Locality Bias (SLB)** in the latent-coordinate cross-attention:

$$b_{c,l} = -\alpha\left(\frac{l}{L_l} - \frac{c}{L_d}\right)^2, \qquad v_c = \mathrm{softmax}\left(\frac{q_c {z^{(i)}}^\top}{\sqrt{d}} + b_c\right) z^{(i)}. \tag{4}$$

Here, $b_{c,l}$ biases retrieval toward latents near the corresponding data token in sequence position. SLB biases locality on decoder retrieval, while the Mamba encoder remains free to mix information globally. We use SLB for fixed-grid fMRI but omit it for EEG, since variable context layouts do not define a fixed query-to-token sequence position. Ablations are provided in Table 6.

Spherical Harmonic Positional Encoding. As mentioned in Section 1, for scalp-based measurements (such as EEG), the notion of smoothness derived from limiting bandwidth in Fourier positional encoding is not natively accurate; the smoothness should be over the 2-dimensional scalp manifold, not in 3-dimensional space. Inspired by the spherical spline's inherent smoothing limitation using damped spherical harmonic basis, we replace the spatial component of $\gamma_\sigma(c)$ with the same spherical harmonics $Y_\ell^{m'}$ from §3.1, evaluated at the electrode's direction on the unit sphere $u = c_{xyz}/\|c_{xyz}\|$ and truncated at a maximum degree L, analogous to band-limit σ for Fourier PE:

$$\gamma_{\mathrm{SH}}(c_{xyz}) = \left[\, Y_0^0(u),\ Y_1^{-1}(u),\ Y_1^0(u),\ Y_1^1(u),\ \ldots,\ Y_L^L(u) \,\right] \in \mathbb{R}^{(L+1)^2}. \tag{5}$$

We keep the original Fourier encoding for the time coordinate $\gamma_\sigma(c_t)$ and concatenate; the full query vector becomes

$$q_c = \mathrm{ReLU}\big(W\,[\gamma_{\mathrm{SH}}(c_{xyz}) \,\|\, \gamma_\sigma(c_t)] + b\big), \tag{6}$$

Truncating at degree L restricts the angular frequencies available to the coordinate encoding. A small L biases the encoding toward similar smooth, low-degree harmonics as the spline. As with σ in the hierarchical decoder, we schedule ascending degrees so that later layers see a higher L.

Algorithm 1 EEG unseen electrode evaluation

Input: Electrode set $\mathcal{P}$; partition into unseen held-out set $U \subset \mathcal{P}$ (never seen in training) and seen set $\mathcal{S} = \mathcal{P} \setminus U$

Training:
for *each batch* **do**
Sample $n_C \sim \mathrm{Unif}\{2, |S|\}$; partition $\mathcal{S}$ into context C with $|C| = n_C$ and targets $T = \mathcal{S} \setminus C$, $\mathbf{Z} \leftarrow \mathrm{MambaEncoder}(C)$, $\hat{\mathbf{x}}_S \leftarrow \mathrm{Decode}(\mathbf{Z}, S)$, $\mathcal{L} \leftarrow \ell(\hat{\mathbf{x}}_S, \mathbf{x}_S)$, Update parameters by gradient descent on $\mathcal{L}$

Evaluation: $\mathbf{Z}_{\text{eval}} \leftarrow \mathrm{MambaEncoder}(C_{\text{eval}})$ where $C_{\text{eval}} \subseteq S$; decode at U and calculate metrics against ground truth only at unseen electrodes

3.3 EEG Unseen Electrode Interpolation Framework

Within-montage interpolation methods cannot distinguish learning a continuous field from learning a fixed electrode layout. Therefore we propose the training and evaluation paradigm in Algorithm 1, holding out a large unseen set U that receives no supervision during training. We vary how many and which context electrodes C are provided to the model at both train and test time, and then evaluate only on U for performance metrics, vastly extending prior work (Liu et al., 2026a) on 10% holdout evaluation to $\sim$ 40–87.5%. We evaluate varying sparsity on two axes: 1) the training scale $s \in \{2, 4, 8\}$ ($|S| \approx 32, 16, 8$ seen electrodes) and 2) the test-time context proportion $p_c = |C_{\text{eval}}|/|S|$.

3.4 Training Algorithm

For fMRI, *firstly* we use brainmasking for loss calculation. Since fMRI data has lots of empty non-brain background area, our arbitrary querying mechanism allows us to threshold near intensity values of zero to identify a high-resolution per-sample brain mask and avoid querying and training on any background voxels. *Secondly*, for high dimensional data volumes, in each training batch we define a random sampling proportion of the brain masked ground truth which we evaluate and calculate loss on to avoid memory errors associated with decoder querying (this is provided to all baselines, not just CANDI). Our loss function is simply this random-sampled MSE.

For EEG, *firstly* we employ curriculum learning from low to high temporal frequency by low-passnfiltering the data in stages of epochs, with each incremental stage having a higher low-pass cut-off frequency. *Secondly*, we introduce a parameter w_{recon} to weight the contributions of the loss calculation from electrodes in C and T (as defined in 3.3) independently. We find that decaying this parameter within each curriculum stage provides the model with better generalization. *Thirdly*, we introduce a term to penalize the bending energy of the INR function directly to motivate a smooth spatial field. We estimate the same quantity minimized by the spline (Equation 1), but evaluated on the decoded field $u(\theta, \phi, t)$ itself. Writing the surface Laplacian as $\Delta_{S^2} u = u_{\theta\theta} + \cot\theta\, u_\theta + \sin^{-2}\theta\, u_{\phi\phi}$, we penalize mean square over M points sampled uniformly over the scalp at random static times:

$$\mathcal{R} = \frac{1}{M} \sum_{j=1}^{M} \left(\Delta_{S^2} u(\theta_j, \phi_j, t_j)\right)^2. \tag{7}$$

We estimate $u_\theta, u_{\theta\theta}, u_{\phi\phi}$ by central differences, decoding u at each point's four angular neighbors $u(\theta \pm h, \phi)$, $u(\theta, \phi \pm h)$ and substituting into $\Delta_{S^2} u$ above. So overall the EEG loss function is:

$$\mathcal{L} = (1 - w_{\text{recon}})\mathcal{L}_{\text{target}} + w_{\text{recon}}\, \mathcal{L}_{\text{recon}} + w_{\text{smooth}}\, \mathcal{R}, \tag{8}$$

4 Experiments

Our experiments test three consequences of the proposed continuous-field formulation. First, we ask whether CANDI can generalize to EEG coordinates that receive no supervision during training, particularly under severe sensor sparsity. Second, we test whether the Mamba-based encoder scales continuous representations to dense 4D fMRI while preserving reconstruction fidelity. Finally, we evaluate whether the learned representations transfer to downstream tasks and whether the proposed inductive biases are supported by ablation.

Scale	Method	PhysioNet-MM/MI NMSE ↓	PhysioNet-MM/MI SNR ↑	PhysioNet-MM/MI PCC ↑	SEED-V NMSE ↓	SEED-V SNR ↑	SEED-V PCC ↑
2x	SSI	$0.4268_{\pm 0.0056}$	$3.6978_{\pm 0.0505}$	$\mathbf{0.7958_{\pm 0.0017}}$	$0.3393_{\pm 0.0011}$	$4.7871_{\pm 0.0132}$	$0.8269_{\pm 0.0004}$
	ZUNA	$0.5313_{\pm 0.0044}$	$2.7509_{\pm 0.0351}$	$0.7664_{\pm 0.0016}$	$0.5986_{\pm 0.0020}$	$2.3134_{\pm 0.0117}$	$0.7739_{\pm 0.0004}$
	Ours	$\mathbf{0.3699_{\pm 0.0032}}$	$\mathbf{4.3397_{\pm 0.0374}}$	$\mathbf{0.7965_{\pm 0.0015}}$	$\mathbf{0.3096_{\pm 0.0009}}$	$\mathbf{5.1621_{\pm 0.0119}}$	$\mathbf{0.8412_{\pm 0.0003}}$
4x	SSI	$0.6139_{\pm 0.0099}$	$2.1775_{\pm 0.0482}$	$0.7412_{\pm 0.0019}$	$0.5186_{\pm 0.0019}$	$2.9350_{\pm 0.0149}$	$0.7713_{\pm 0.0005}$
	ZUNA	$0.6687_{\pm 0.0101}$	$1.7955_{\pm 0.0401}$	$0.7323_{\pm 0.0018}$	$0.5103_{\pm 0.0014}$	$2.9436_{\pm 0.0112}$	$0.7877_{\pm 0.0004}$
	Ours	$\mathbf{0.4680_{\pm 0.0035}}$	$\mathbf{3.3320_{\pm 0.0346}}$	$\mathbf{0.7554_{\pm 0.0017}}$	$\mathbf{0.3758_{\pm 0.0011}}$	$\mathbf{4.2585_{\pm 0.0134}}$	$\mathbf{0.8158_{\pm 0.0004}}$
8x	SSI	$1.4379_{\pm 0.0485}$	$-0.8771_{\pm 0.0454}$	$0.5858_{\pm 0.0022}$	$1.6821_{\pm 0.0077}$	$-1.1832_{\pm 0.0126}$	$0.6132_{\pm 0.0005}$
	ZUNA	$0.8952_{\pm 0.0144}$	$0.5667_{\pm 0.0339}$	$0.6198_{\pm 0.0019}$	$0.6924_{\pm 0.0020}$	$1.7816_{\pm 0.0096}$	$0.7189_{\pm 0.0004}$
	Ours	$\mathbf{0.5767_{\pm 0.0029}}$	$\mathbf{2.4012_{\pm 0.0233}}$	$\mathbf{0.6324_{\pm 0.0017}}$	$\mathbf{0.4955_{\pm 0.0015}}$	$\mathbf{3.1383_{\pm 0.0113}}$	$\mathbf{0.7653_{\pm 0.0004}}$

Table 1: Reconstruction performance across scales at full context ($p_c = 1.0$) on two datasets.

4.1 Datasets and Tasks

EEG. We evaluate CANDI on two EEG datasets: 1) **SEED-V** collection of 62-channel EEG measurements of 16 subjects provided with emotional stimuli of 5 different classes, and 2) **PhysioNet-MM/MI** collection of 64-channel EEG measurements on 109 subjects during 4 classes of motor movement and mental visualization of motor movement. Recordings are segmented into non-overlapping 4 second windows. For interpolation tasks, we use a train/val/test subject-disjoint split (Appendix A). For evaluation, we use the pre-defined low-resolution subsampling layouts from Tang et al. (2023a), which are the standard for EEG-SR literature (Li et al., 2025; Liu et al., 2026b).

fMRI. Our primary fMRI dataset is the **WU-Minn HCP** dataset of healthy subject resting-state fMRI volumes. The dataset comprises of 1,084 subjects with 4D fMRI volumes (shape H, W, D, T) of size $96 \times 96 \times 96 \times 1200$. We employ an 80/10/10 train/val/test subject-disjoint split, and we chunk these volumes into smaller time segments (exact size varies between tasks) and train on 6 chunks per subject, making the total training set size 5,204 volumes. Full details about datasets and preprocessing are provided in Appendix A.

4.2 EEG Experiments

Unseen Electrode Interpolation Using the unseen-electrode framework in Section 3.3, we evaluate CANDI only at electrode locations never observed during training. SOTA Fixed-montage EEG-SR methods (Liu et al., 2026b; Li et al., 2025) cannot support this setting, so we exclude them from baselines and instead compare against Spherical Spline Interpolation (SSI) (Perrin et al., 1989) and ZUNA (Warner et al., 2026), which support arbitrary electrode inputs and coordinate queries. We report normalized mean-squared error (NMSE), signal-to-noise ratio (SNR) and Pearson correlation

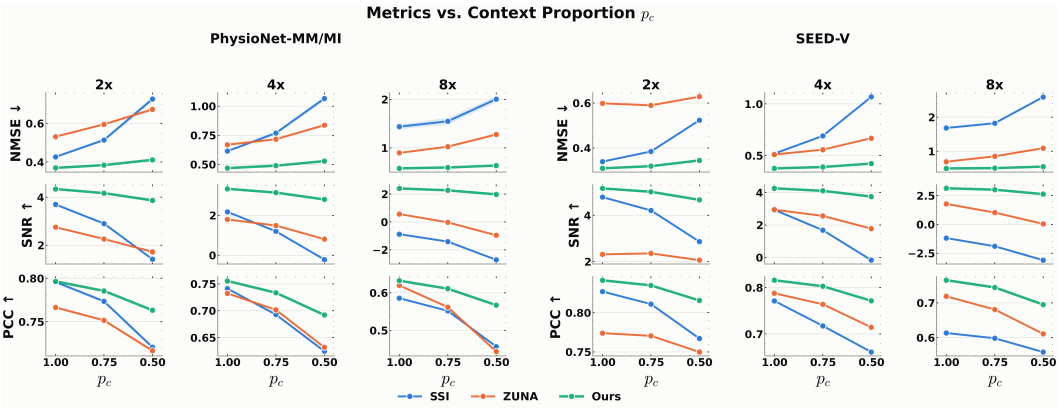

Figure 3: Metrics for interpolation quality as p_c is reduced at test time on both datasets

Model	Sex		Age	
	Acc ↑	AUC ↑	MSE ↓	MAE ↓
SwiFT (scratch)	0.8808	0.9546	9.9225	2.5301
Omni-fMRI (scratch)	0.7957	0.8866	12.2068	2.8800
Omni-fMRI (pre-train + finetune)	0.8727	0.9452	12.7114	3.0222
Ours (scratch)	0.8567	0.9090	15.6120	3.3521
Ours (linear probe)	0.8803	**0.9892**	**8.8443**	**2.4973**
Ours (pre-train + fine-tuned)	**0.9436**	0.9831	10.3182	2.7012

Table 2: Downstream performance on HCP. Sex is binary classification (accuracy, AUC); age is regression (MSE, MAE, in years)

	2×	4×	8×
SSI	**0.5209**	0.4679	0.3791
ZUNA	0.5108	0.4596	0.3709
Ours	0.5154	**0.4763**	**0.3891**
GT (LR)	0.4457	0.3613	0.2912
GT (HR)	0.5819	0.5819	0.5819

Table 3: Downstream classification accuracy on PhysioNet MM/MI using EEGNet classifier

Model	**Patch Size**†	**MSE / HFEN for 4D Volume Size**†							
		$24^3 \times 2$		$36^3 \times 2$		$48^3 \times 2$		$64^3 \times 2$	
		MSE ↓	HFEN ↓	MSE ↓	HFEN ↓	MSE ↓	HFEN ↓	MSE ↓	HFEN ↓
Ours	$2^3 \times 2$	$\mathbf{6.90 \times 10^{-5}}$	**0.0289**	$\mathbf{8.60 \times 10^{-5}}$	**0.0508**	$\mathbf{1.76 \times 10^{-4}}$	**0.0686**	OOM	OOM
Ours	$4^3 \times 2$	1.48×10^{-4}	0.0634	1.74×10^{-4}	0.1003	3.30×10^{-4}	0.1344	$\mathbf{1.84 \times 10^{-4}}$	**0.1197**
LAGINR	$2^3 \times 2$	1.37×10^{-4}	0.0623	OOM	OOM	OOM	OOM	OOM	OOM
LAGINR	$4^3 \times 2$	1.46×10^{-4}	0.0654	1.80×10^{-4}	0.1095	5.18×10^{-4}	0.2163	OOM	OOM
Linformer	$2^3 \times 2$	1.72×10^{-4}	0.0781	2.88×10^{-4}	0.1642	5.31×10^{-4}	0.2214	OOM	OOM
Linformer	$4^3 \times 2$	1.86×10^{-4}	0.0824	2.51×10^{-4}	0.1455	5.20×10^{-4}	0.2147	5.25×10^{-4}	0.3078
TransINR	$2^3 \times 2$	1.93×10^{-4}	0.0856	OOM	OOM	OOM	OOM	OOM	OOM
TransINR	$4^3 \times 2$	1.59×10^{-4}	0.0717	2.08×10^{-4}	0.1239	5.20×10^{-4}	0.2147	OOM	OOM
SCENT	$1^3 \times 1$	2.44×10^{-4}	0.1018	2.88×10^{-4}	0.1525	6.03×10^{-4}	0.2379	6.18×10^{-4}	0.3343

Table 4: 4D fMRI volume reconstruction performance comparisons at different patch sizes, reports MSE / HFEN (High-Frequency Error Norm) pairs. Lower values are better for both. L_l is chosen such that the ratio $L_d : L_l = 9 : 1$ to maintain fairness between models. †: values are represented as spatial$^3\times$ temporal sizes; OOM: Out-of-Memory Error

coefficient (PCC) over two forms of sparsity (described in 3.3). **Results:** CANDI consistently outperforms SSI and ZUNA across both datasets and all seen-electrode scales (Table 1), demonstrating generalization to electrode coordinates that receive no supervision during training. The advantage over SSI is most pronounced under severe sparsity; at 8× with 87.5% unseen electrodes and only 8 context electrodes, SSI yields negative SNR on both datasets, whereas CANDI maintains positive-SNR reconstruction with substantially lower NMSE and higher PCC. CANDI outperforms ZUNA across the board, but ZUNA also demonstrates robustness over SSI as the scale changes. Figure 3 further shows that CANDI degrades more gracefully than SSI and ZUNA as the proportion of observed context electrodes is reduced.

Downstream Motor Task Classification We evaluate CANDI on downstream classification PhysioNet-MM/MI's 4-class motor movement classification task. We use Lawhern et al. (2018), a standard EEG classification model, train it on ground truth EEG signals, and then run inference on interpolation results from CANDI and baselines to demonstrate its ability to learn features that are important for downstream classification. Ground truth low-resolution accuracy is calculated by passing in signals with unseen electrode signals zeroed out to the classifier. **Results:** In Table 3, we show that using the (Lawhern et al., 2018) classifier trained on ground truth data, CANDI-interpolated channels preserve class-relevant structure and its advantage over SSI still grows with sparsity, demonstrated by modest advantages in accuracy at the 8× and 4× scales.

4.3 fMRI Experiments

4D Volume Reconstruction We measure the ability of CANDI to reconstruct a given fMRI volume by tokenizing and encoding it, and querying for the full volume coordinates. The main difficulty in this task comes from overcoming spectral bias in INR methods (Tancik et al., 2020). We compare CANDI to two state-of-the-art Vanilla Transformer-based GINR methods (LAGINR (Doyup et al., 2023) and Trans-INR (Chen & Wang, 2022b)) as well as two efficient backbone GINR methods

Table 5: Ablation of EEG-specific components on SEED-V unseen-electrode reconstruction. DNT denotes that it does not train, with PCC approximately 0.00.

SHPE	Bending Loss	Curriculum	Recon Loss	NMSE↓	SNR↑	PCC↑
✓	✓	✓	✓	**0.3683**	**4.34**	**0.8190**
✗	✓	✓	✓	DNT	DNT	DNT
✗	✗	✓	✓	1.2141	−0.84	0.6565
✓	✗	✓	✓	0.4263	3.70	0.7549
✓	✓	✗	✓	DNT	DNT	DNT
✓	✓	✓	✗	0.4236	3.73	0.7938

Table 6: Ablation of sequence locality bias (SLB) on HCP fMRI reconstruction across latent-token budgets, reporting MSE

	# Learnable Tokens		
	64	**128**	**256**
no SLB	2.07×10^{-4}	1.97×10^{-4}	1.96×10^{-4}
SLB	$\mathbf{2.02 \times 10^{-4}}$	$\mathbf{1.91 \times 10^{-4}}$	$\mathbf{1.71 \times 10^{-4}}$
% change	(↓ 2.42%)	(↓ 3.05%)	(↓ 12.76%)

(LAGINR adapted with a Linformer encoder, and SCENT (Park et al., 2025), a Perceiver-IO (Jaegle et al., 2022) style model) to demonstrate both its ability to scale to large fMRI volumes and also superior performance at smaller patch size. We report mean-squared error (MSE) for reconstruction quality and high-frequency error norm (HFEN) to compare frequency information over various resized volumes and patch sizes. **Results:** CANDI achieves the lowest MSE and HFEN at 24^3, 36^3, and 48^3, with finer $2^3 \times 2$ tokenization consistently improving reconstruction (Table 4). At 64^3, Transformer GINRs become intractable, while CANDI remains feasible and substantially outperforms the efficient baselines. This supports linear-scaling encoding as a means to retain fine tokenization on large 4D volumes. For more analysis on scaling metrics, refer to Appendix C.1.

HCP Downstream Age/Sex Classification Despite reconstruction (rather than continuous-field interpolation) being the fMRI objective, the resulting latents exceed specialized fMRI architectures on several downstream HCP metrics. We pretrain CANDI on reconstructing $72^3 \times 8$ HCP resting-fMRI volumes and evaluate scratch, frozen linear-probe, and end-to-end fine-tuning using a gated-attention classification head. For reference, SOTA models SwiFT and Omni-fMRI are retrained on the identical HCP data and splits; these are controlled architecture comparisons rather than comparisons to their published foundation-model settings. **Results:** Reconstruction pretraining substantially improves CANDI over training from scratch (Table 2). A frozen linear probe achieves the best sex AUC and age-prediction error, while fine-tuning yields the highest sex accuracy. Under our controlled HCP-only setting, CANDI outperforms the retrained SwiFT and Omni-fMRI baselines, supporting the transferability of its learned representations.

4.4 Ablations

EEG components. The full configuration of components from 3.4 performs best across all metrics(Table 5), supporting our geometry-aware encoding and training choices. Removing bending regularization or reconstruction weighting degrades performance, while removing SHPE and bending loss or the curriculum causes training collapse under our training setup.

Sequential Locality Bias. SLB improves fMRI reconstruction across all latent budgets (Table 6), with the relative MSE gain increasing from 2.42% at 64 tokens to 12.76% at 256, consistent with our hypothesis that sequence-local inductive bias becomes more useful as the latent set grows.

5 Discussion and Conclusion

We propose CANDI, a framework for learning amortized continuous representations of neuroimaging signals while accommodating modality-specific geometry and inductive biases. On 4D fMRI, CANDI's Mamba-based encoder enables finer-grained tokenization at sequence lengths where Transformer GINRs become memory-limited, improving reconstruction while producing representations that transfer to downstream classification. For EEG, spherical harmonic positional encoding and spherical smoothness regularization enable continuous scalp-field reconstruction, with CANDI generalizing to electrode locations withheld entirely during training and preserving information useful for downstream motor-task classification. Ablations further support the contributions of sequential locality-biased cross-attention and EEG training methodologies. Together, these results demonstrate that amortized continuous-field representations provide a common framework for reconstruction, spatial querying, and representation learning across distinct neuroimaging geometries. Future work should evaluate robustness over other axes, such as cross-dataset and cross-montage transfer towards neuroimaging foundation modeling, and possibilities for multimodal fusion analysis.

AI USE STATEMENT

Generative AI tools, including ChatGPT (OpenAI), Claude, and Claude Code (Anthropic), were used during this work for software implementation and debugging, auxiliary analysis and plotting scripts, feedback on experimental design, literature summarization, and editing of manuscript text, figures, and formatting supplemental tables. Specifically, agents were used for the following tasks which require disclosure: **formulate mathematical claims** by formalizing conceptual ideas from the authors, **design or provide feedback on research methodology or experiments** primarily in providing tools to analyze results via coding agents, **implement methods** via coding agents to realize conceptual ideas from the authors or to configure baseline repositories to run with our experimental frameworks, and **interpret results** for evaluating potential explanations for results and creating follow-up experiments.

Generative AI tools were also used for the following tasks for which disclosure is recommended: **create or modify scientific figures or images** via python modules like matplotlib or to make large systematic edits on PowerPoint, **create or edit software code** for experimentation and evaluation scripts, **draft parts of a research paper**, **identify, summarize and/or analyze relevant literature** to situate related works (paired with individual author-led literature review), **edit a research paper to improve readability**

All AI-assisted code, analyses, and written material were reviewed and verified by the authors. All experiments and reported numerical results were produced from the authors' own experimental runs. The authors take full responsibility for the methods, results, claims, and final contents of the paper.

ETHICS STATEMENT

We use existing de-identified neuroimaging datasets, including WU-Minn HCP resting-state fMRI, SEED-V, and PhysioNet-MM/MI, under the terms and data-use policies provided by the respective dataset maintainers. This study involved no new collection of human-subject data.

REPRODUCIBILITY STATEMENT

We provide implementation and experimental details which are sufficient to reproduce the reported results in the Appendix. This includes dataset construction and preprocessing, train/validation/test splits, model and decoder configurations, optimization hyperparameters, EEG curriculum and regularization settings, the complete set of EEG subsampling layouts, downstream-classification settings, and hardware/scaling details. The unseen-electrode training and evaluation procedure is specified in Algorithm 1, and the primary loss functions and architectural components are well-defined. We also pair this submission with a code supplemental which walks the reader through how to reproduce a few main results from the paper.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

REFERENCES

Geoffrey M. Boynton, Stephen A. Engel, Gary H. Glover, and David J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, 16 (13):4207–4221, 1996.

Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations. In *European Conference on Computer Vision (ECCV)*, 2022a.

Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations, 2022b. URL https://arxiv.org/abs/2208.02801.

Krzysztof Choromanski, Valerii Likhosherstov, David Dohan, Xingyou Song, Andreea Gane, Tamás Sarlós, Peter Hawkins, Jared Davis, Afroz Mohiuddin, Łukasz Kaiser, David Belanger, Lucy Colwell, and Adrian Weller. Rethinking attention with performers. In *International Conference on*

Learning Representations (ICLR), 2021. URL https://openreview.net/forum?id=Ua6zuk0WRH.

Taco S. Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical CNNs. In *International Conference on Learning Representations (ICLR)*, 2018.

Berkay Döner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. LUNA: Efficient and topology-agnostic foundation model for EEG signal analysis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2510.22257.

Lee Doyup, Kim Chiheon, Cho Minsu, and Han Wook-Shin. Locality-aware generalizable implicit neural representation, 2023. URL https://arxiv.org/abs/2310.05624.

Emilien Dupont, Hyunjik Kim, S. M. Ali Eslami, Danilo Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one, 2022. URL https://arxiv.org/abs/2201.12204.

Yassine El Ouahidi, Jonathan Lys, Philipp Thölke, Nicolas Farrugia, Bastien Pasdeloup, Vincent Gripon, Karim Jerbi, and Giulia Lioi. REVE: A foundation model for EEG – adapting to any setup with large-scale pretraining on 25,000 subjects. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. arXiv:2510.21585.

Elizabeth Fons, Alejandro Sztrajman, Yousef El-laham, Alexandros Iosifidis, and Svitlana Vyetrenko. Hypertime: Implicit neural representation for time series, 2022. URL https://arxiv.org/abs/2208.05836.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2024a.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces, 2024b. URL https://arxiv.org/abs/2312.00752.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces, 2022. URL https://arxiv.org/abs/2111.00396.

C Gössl, D P Auer, and L Fahrmeir. Bayesian spatiotemporal inference in functional magnetic resonance imaging. *Biometrics*, 57(2):554–562, 2001. ISSN 0006-341X. doi: 10.1111/j.0006-341x.2001.00554.x.

L M Harrison, W Penny, G Flandin, C C Ruff, N Weiskopf, and K J Friston. Graph-partitioned spatial priors for functional magnetic resonance images. *NeuroImage*, 43(4):694–707, 2008. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2008.08.012.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, et al. Perceiver io: A general architecture for structured inputs & outputs. In *International Conference on Learning Representations (ICLR)*, 2022.

Herbert H. Jasper. The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10:371–375, 1958.

Chiheon Kim, Doyup Lee, Saehoon Kim, Minsu Cho, and Wook-Shin Han. Generalizable implicit neural representations via instance pattern composers, 2023a. URL https://arxiv.org/abs/2211.13223.

Peter Yongho Kim, Junbeom Kwon, Sunghwan Joo, Sangyoon Bae, Donggyu Lee, Yoonho Jung, Shinjae Yoo, Jiook Cha, and Taesup Moon. SwiFT: Swin 4D fMRI transformer. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023b.

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. ISSN 1741-2552. doi: 10.1088/1741-2552/aace8c. URL http://dx.doi.org/10.1088/1741-2552/aace8c.

Doyup Lee, Chiheon Kim, Minsu Cho, and Wook-Shin Han. Locality-aware generalizable implicit neural representation. In *Advances in Neural Information Processing Systems*, volume 36, 2023. NeurIPS 2023.

Dongdong Li, Zhongliang Zeng, Zhe Wang, and Hai Yang. ESTformer: Transformer utilising spatiotemporal dependencies for electroencephalogram super-resolution. *Knowledge-Based Systems*, 2025. doi: 10.1016/j.knosys.2025.113373. arXiv:2312.10052.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Coordinate-queryable neural field reconstruction for eeg spatial super-resolution with unseen-electrode generation, 2026a. URL https://arxiv.org/abs/2606.23707.

Hongjun Liu, Leyu Zhou, Zijianghao Yang, and Chao Yao. Step-aware residual-guided diffusion for EEG spatial super-resolution. In *International Conference on Learning Representations (ICLR)*, 2026b. arXiv:2510.19166.

Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843):150–157, 2001.

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

Josue Ortega Caro, Antonio H. de Oliveira Fonseca, Syed Rizvi, Matteo Rosati, Christopher Averill, James Cross, Prateek Mittal, Emanuele Zappala, Rahul Dhodapkar, Chadi Abdallah, and David van Dijk. BrainLM: A foundation model for brain activity recordings. In *International Conference on Learning Representations (ICLR)*, 2024.

David Keetae Park, Xihaier Luo, Guang Zhao, Seungjun Lee, Miruna Oprescu, and Shinjae Yoo. Scent: Robust spatiotemporal learning for continuous scientific data via scalable conditioned neural fields, 2025. URL https://arxiv.org/abs/2504.12262.

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

Ofir Press, Noah A. Smith, and Mike Levy. Train short, test long: Attention with linear biases enables input length extrapolation. In *International Conference on Learning Representations (ICLR)*, 2022. arXiv:2108.12409.

Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. In *International Conference on Learning Representations (ICLR)*, 2024.

Liyue Shen, John Pauly, and Lei Xing. NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):770–782, 2024.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Filip Szatkowski, Karol J. Piczak, Przemysław Spurek, Jacek Tabor, and Tomasz Trzciński. Hypernetworks build implicit neural representations of sounds, 2023. URL https://arxiv.org/abs/2302.04959.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

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, 2023a. doi: 10.1109/TCYB.2022.3178370.

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, 2023b. doi: 10.1109/TCYB.2022.3178370.

Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. Efficient transformers: A survey. *ACM Computing Surveys*, 55(6):1–28, 2022. doi: 10.1145/3530811.

Kevin Valencia, Thilina Balasooriya, Xihaier Luo, Shinjae Yoo, and David Keetae Park. Omnifield: Conditioned neural fields for robust multimodal spatiotemporal learning, 2025. URL https://arxiv.org/abs/2511.02205.

Grace Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981. doi: 10.1137/0902002.

Feng Wang, Jiahao Wang, Sucheng Ren, Guoyizhe Wei, Jieru Mei, Wei Shao, Yuyin Zhou, Alan Yuille, and Cihang Xie. Mamba-r: Vision mamba also needs registers, 2025. URL https://arxiv.org/abs/2405.14858.

Mo Wang, Wenhao Ye, Junfeng Xia, Junxiang Zhang, Xuanye Pan, Minghao Xu, Haotian Deng, Hongkai Wen, and Quanying Liu. Omni-fmri: A universal atlas-free fmri foundation model, 2026. URL https://arxiv.org/abs/2601.23090.

Sinong Wang, Belinda Z. Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity. *arXiv preprint arXiv:2006.04768*, 2020. doi: 10.48550/arXiv.2006.04768. URL https://arxiv.org/abs/2006.04768.

Christopher Warner, Jonas Mago, JR Huml, and Beren Millidge. Zuna1.1: A more flexible eeg foundation model for denoising and super-resolution, 2026. URL https://arxiv.org/abs/2607.27308.

Mark W Woolrich, Mark Jenkinson, J Michael Brady, and Stephen M Smith. Fully bayesian spatio-temporal modeling of FMRI data. *IEEE transactions on medical imaging*, 23(2):213–231, 2004. ISSN 0278-0062. doi: 10.1109/TMI.2003.823065.

Qi Wu, David Bauer, Yuyang Chen, and Kwan-Liu Ma. Hyperinr: A fast and predictive hypernetwork for implicit neural representations via knowledge distillation, 2023. URL https://arxiv.org/abs/2304.04188.

Qing Wu, Yuwei Li, Lan Xu, Ruiming Feng, Hongjiang Wei, Qing Yang, Boliang Yu, Xiaozhao Liu, Jingyi Yu, and Yuyao Zhang. IREM: High-resolution magnetic resonance image reconstruction via implicit neural representation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 65–74, 2021.

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.

Biao Zhang and Rico Sennrich. Root mean square layer normalization, 2019. URL https://arxiv.org/abs/1910.07467.

Zizheng Zhang, Peng Liu, Zhen Guo, Yikai Wang, Jiwen Lu, and Jie Zhou. Attention beats linear for fast implicit neural representation generation. In *European Conference on Computer Vision (ECCV)*, 2024.

A Datasets and Preprocessing

A.1 EEG Datasets

A.1.1 Data Overview

PhysioNet-MM/MI. We use the PhysioNet EEG Motor Movement/Imagery database, containing 64-channel EEG recordings from 109 subjects acquired with the international 10–10 system. Recordings are originally sampled at 160 Hz. Each annotated imagery event is segmented into a 4 s epoch, yielding 9,837 epochs in total. The resulting input shape is $(64, 800)$ after resampling to 200 Hz. Labels correspond to four motor-imagery classes: left fist, right fist, both fists, and both feet. For unseen-electrode reconstruction experiments, we use a subject-disjoint split comprising 87/10/12 train/validation/test subjects. For the downstream motor-classification experiment, we instead use a run-level split containing 471/52/131 train/validation/test runs, respectively, and apply the identical partition to both the reconstruction model and EEGNet classifier. This latter split is not subject-disjoint, because state-of-the-art classification models on subject-disjoint classification do not provide reliable results.

SEED-V. SEED-V contains 62-channel EEG recordings from 16 subjects exposed to five classes of emotional stimuli. Data are acquired at 1000 Hz over three sessions per subject. Four recordings that failed preprocessing quality control are excluded, leaving 44 recordings and 660 stimulus trials. Trials are divided into non-overlapping 4 s segments, yielding 26,604 segments with input shape $(62, 800)$ after resampling to 200 Hz. For reconstruction and ablation experiments, we use a subject-disjoint split of 12/1/3 train/validation/test subjects, corresponding to 19,312/1,823/5,469 segments.

For both EEG datasets, electrode coordinates are obtained from the standard 10–05 or 10–10 montage and mapped to 3D Cartesian scalp coordinates. Coordinates are centered and normalized, and unit-sphere directions are used by the spherical-harmonic decoder. Low-resolution electrode layouts for the $2\times$, $4\times$, and $8\times$ unseen-electrode experiments follow the layouts used in prior EEG super-resolution work (Tang et al., 2023a).

A.1.2 Preprocessing

PhysioNet-MM/MI. We apply a 1–75 Hz zero-phase band-pass filter and a 60 Hz notch filter, resample from 160 Hz to 200 Hz, and segment annotation-locked 4 s motor-imagery epochs. Signals are clipped to $\pm 200\ \mu$V and divided by 30 before model input. During training, EEG is additionally passed through the frequency curriculum described in Table 8: a 1 Hz high-pass filter is followed by a stage-dependent low-pass filter whose cutoff is progressively increased during training. No per-channel or per-subject normalization is applied.

SEED-V. Continuous recordings are filtered from 1–75 Hz using a zero-phase filter, notched at 50 Hz, and resampled from 1000 Hz to 200 Hz. Each stimulus trial is extracted using the provided trial boundaries and divided into non-overlapping 4 s windows; incomplete trailing windows are discarded. Signals are divided by 10 prior to model input. Unlike PhysioNet, SEED-V is not clipped and no additional online notch filter is applied. The same stage-wise low-pass curriculum is used during model training.

A.2 fMRI Dataset

A.2.1 Data Overview

HCP S1200. We use resting-state fMRI from the Human Connectome Project Young Adult S1200 release. Each subject contains 1200 BOLD volumes acquired with TR $= 0.72$ s on a native $91 \times 109 \times 91$ voxel grid in MNI 2 mm space. The processed data store contains 1,084 subjects, of which 1,079 have both imaging data and phenotype metadata required by our experiments.

Subjects are partitioned using a fixed subject-level split containing 863/107/109 train/validation/test subjects. Each example consists of consecutive temporal frames, with the number of frames determined by the experiment. The standard reconstruction setting uses two consecutive TRs, while

downstream representation-learning experiments use longer temporal windows as specified in the corresponding experiment.

For the primary reconstruction configuration, each subject contributes 6 non-overlapping two-TR windows, yielding 5,178/642/654 train/validation/test examples. Age and sex labels are used only for downstream phenotype prediction and are not used during reconstruction pretraining.

A.2.2 PREPROCESSING

HCP volumes follow the preprocessing convention used by SwiFT. Foreground voxels are identified by excluding zero-valued background voxels, and each subject's full run is min–max normalized over foreground voxels. The normalized 4D volume is then stored as individual temporal frames.

At loading time, consecutive frames are concatenated along the temporal dimension and padded with background values before spatial resizing. Each frame is resized by trilinear interpolation to the spatial resolution required by the experiment; the primary configuration uses $96 \times 96 \times 96$ volumes, while the reconstruction scaling experiments additionally evaluate 24^3, 36^3, 48^3, and 64^3 resolutions and downstream classification is done at 72^3 spatial resolution. The resulting model input has shape $(1, H, W, D, T)$ before batching.

For reconstruction loss computation, we form a per-sample foreground brain mask and sample query coordinates only from within this mask. A fresh random subset of masked voxel–time coordinates is queried at each optimization step to reduce decoder memory usage; the same query-sampling procedure is applied to all reconstruction baselines.

B TRAINING DETAILS

We report detailed training, implementation and hyperparameter information for each task and experimental setting oto facilitate reproducibility.

B.1 ARCHITECTURE AND HYPERPARAMETERS

B.1.1 RECONSTRUCTION AND INTERPOLATION MODELS

Table 7 provides the general default values for architectural choices, training setting, and hyperparameters used in fMRI reconstruction and EEG spatial interpolation tasks. EEG interpolation specifically has curriculum training and modality-specific parameters which are listed in further detail in Table 8

	TASK	
	fMRI Reconstruction	**EEG Spatial Interpolation**
ARCHITECTURE		
Mamba dimension	512	512
Mamba depth	6	6
Feed-forward dimension	2048	2048
Learnable-token placement	Equidistant	Equidistant
Decoder feature dimension	512	1024
Coordinate encoding	Fourier PE	SHPE (space) + Fourier PE (time)
Query bandwidth / degree	$\sigma_{q,s} = 128, \sigma_{q,t} = 128$	$L_{q,s} = 3, \sigma_{q,t} = 1024$
Hierarchical bandwidth / degree	$\sigma_{\ell,s} = [128, 32], \sigma_{\ell,t} = [128, 32]$	$L_{\ell,s} = [3, 2, 1]$, $\sigma_{\ell,t} = [1024, 256, 64]$
Decoder activation	ReLU	Sine ($\omega_0 = 3$)
OPTIMIZATION		
Training objective	MSE	Eq. 8
Optimizer	AdamW	AdamW
Peak learning rate	1.5×10^{-4}	Stage-dependent; max 5×10^{-4}
Weight decay	0.01	0.01
Training epochs	400	300
Batch size / GPU	2	8
# GPUs	16	16 SEED-V, 4 PhysioNet-MM/MI
Precision	bf16 AMP	bf16 AMP
Minimum LR ratio	0.2	0.1
LR schedule	Warmup + cosine annealing	Per-stage warm-restart cosine
Warmup	10 epochs	Per-stage; see Tab. 8
Gradient clipping	N/A	Dataset-dependent
Randomness	Seed 0	Averaged over 4 fixed layouts
Checkpoint selection	Best validation MSE	Best/last validation MSE

Table 7: Default architecture and optimization hyperparameters for CANDI. Experiment-specific patch sizes, learnable-token counts, EEG context sampling, and curriculum parameters are reported separately.

EEG training parameter	Default value
MODEL	
Temporal patch size	40 samples
# learnable tokens	128
CURRICULUM	
# Curriculum Stages	7
Curriculum cutoffs (Hz)	$[8, 16, 24, 32, 45, 64, 75]$
Epochs per curriculum stage	$[40, 40, 40, 40, 80, 30, 30]$
Peak LR per stage	$[5, 4, 3, 3, 2, 2, 2] \times 10^{-4}$
Max Context w_{recon} per stage	$[0.8, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5]$
LOSS AND REGULARIZATION	
Final context-reconstruction weight	0.2
Smoothness weight w_{smooth}	10^{-2}
Surface-Laplacian estimator	Central finite differences
Finite-difference step h	0.01 rad
Spherical collocation points	512
Time samples per smoothness step	6
Pole exclusion ϵ	0.05

Table 8: Default EEG-specific optimization and regularization parameters.

The schedule in Table 8 is the default used across EEG experiments. A small number of montage configurations required minor stage-wise learning-rate adjustments for optimization stability; these adjustments were selected using training/validation behavior and not test performance.

B.1.2 Downstream Classification Models

We provide training architecture and hyperparameter specifications for downstream classification task models as well.

	TRAINING REGIME		
	Scratch	**Linear Probe**	**Fine-tuned**
MODEL			
Input volume	$72^3 \times 8$	$72^3 \times 8$	$72^3 \times 8$
Encoder	CANDI Mamba	CANDI Mamba	CANDI Mamba
Initialization	Random	Reconstruction pretrained	Reconstruction pretrained
Encoder frozen	No	Yes	No
Pooling	Gated attention	Concat all learnable tokens	Gated attention
Prediction head	MLP	Linear	MLP
Head hidden dimension	16	–	16
Dropout	0.3	0.0	0.3
OBJECTIVES			
Sex objective	Cross-entropy (label smoothing 0.1)	Cross-entropy (label smoothing 0.1)	Cross-entropy (label smoothing 0.1)
Age objective	MSE on standardized age	MSE on standardized age	MSE on standardized age
OPTIMIZATION			
Optimizer	AdamW	AdamW	AdamW
Encoder LR	1.5×10^{-4}	–	2×10^{-5}
Head LR	1.5×10^{-4}	1×10^{-4}	2×10^{-5}
Weight decay	0.01	0.5	0.01
Batch size / GPU	8	8	8
Training epochs	50	20	30
LR schedule	Warmup (5 ep.) + cosine, min ratio 0.2	Warmup (2 ep.) + cosine, min ratio 0.2	Warmup (5 ep.) + cosine, min ratio 0.2
Checkpoint criterion	Best val. accuracy (sex) / MSE (age)	Best val. accuracy (sex) / MSE (age)	Best val. accuracy (sex) / MSE (age)
Seed	0	0	0

Table 9: Training configurations for HCP downstream phenotype prediction. Linear-probe and fine-tuning experiments initialize the encoder from the same CANDI reconstruction-pretrained checkpoint.

	Value
MODEL	
Classifier	EEGNet (Lawhern et al., 2018)
F_1	8
Depth multiplier D	2
F_2	16
Temporal kernel length	64
Separable kernel length	16
Dropout	0.5
OPTIMIZATION	
Classifier objective	Cross-entropy
Optimizer	AdamW
Learning rate	1×10^{-3}
Weight decay	10^{-4}
Batch size	64
Training epochs	150
LR schedule	Warmup (5 ep.) + cosine, min ratio 0.05
Checkpoint criterion	Best validation accuracy
# seeds	1

Table 10: EEGNet configuration for downstream PhysioNet-MM/MI classification. EEGNet is trained only on full-resolution ground-truth signals and is frozen when evaluating signals reconstructed by CANDI and interpolation baselines.

B.2 EEG Training Settings/Procedures

B.2.1 EEG Layouts

Table 11 lists the 24 EEG layouts used in this paper as the seen-electrode set S for interpolation and downstream classification. We use 12 layouts per dataset, with four layouts at each downsampling scale ($2\times$, $4\times$, and $8\times$). These layouts are the same as the subsampling layouts from Tang et al. (2023a) and are listed explicitly here for reproducibility.

Scale	Layout	$\lvert S\rvert$	Seen electrodes S
SEED-V (62-CHANNEL MONTAGE)			
$2\times$	1	34	FPZ, AF3, AF4, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPZ, CP4, TP8, P7, P3, PZ, P4, P8, PO7, PO3, POZ, PO4, PO8, OZ
$2\times$	2	28	FP1, FP2, F5, F1, F2, F6, FC5, FC1, FC2, FC6, C5, C1, C2, C6, CP5, CP1, CP2, CP6, P5, P1, P2, P6, PO5, PO6, O1, O2, I1, I2
$2\times$	3	33	FP1, FPZ, FP2, F7, F5, F3, F1, FZ, F2, F4, F6, F8, T7, C5, C3, C1, CZ, C2, C4, C6, T8, P7, P5, P3, P1, PZ, P2, P4, P6, P8, O1, OZ, O2
$2\times$	4	29	AF3, AF4, FT7, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, PO7, PO5, PO3, POZ, PO4, PO6, PO8, I1, I2
$4\times$	1	16	FPZ, AF3, AF4, FC5, FC1, FC2, FC6, CP5, CP1, CP2, CP6, PO3, PO4, OZ, I1, I2
$4\times$	2	15	FP1, FP2, F5, FZ, F6, T7, C3, CZ, C4, T8, P5, PZ, P6, O1, O2
$4\times$	3	16	FPZ, AF3, AF4, FT7, FC3, FCZ, FC4, FT8, TP7, CP3, CPZ, CP4, TP8, PO7, OZ, PO8
$4\times$	4	16	FP1, FP2, F7, F3, F4, F8, C5, C1, C2, C6, P7, P3, P4, P8, O1, O2
$8\times$	1	8	AF3, AF4, FC5, FC6, CP5, CP6, PO5, PO6
$8\times$	2	8	FP1, FP2, FZ, T7, T8, PZ, O1, O2
$8\times$	3	8	AF3, AF4, FC3, FC4, P3, P4, PO7, PO8
$8\times$	4	8	FPZ, F5, F6, C3, C4, P5, P6, OZ
PHYSIONET-MM/MI (64-CHANNEL MONTAGE)			
$2\times$	1	38	FPZ, AF7, AF3, AFZ, AF4, AF8, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPZ, CP4, TP8, P7, P3, PZ, P4, P8, PO7, PO3, POZ, PO4, PO8, OZ, IZ
$2\times$	2	26	FP1, FP2, F5, F1, F2, F6, FC5, FC1, FC2, FC6, T9, C5, C1, C2, C6, T10, CP5, CP1, CP2, CP6, P5, P1, P2, P6, O1, O2
$2\times$	3	35	FP1, FPZ, FP2, F7, F5, F3, F1, FZ, F2, F4, F6, F8, T9, T7, C5, C3, C1, CZ, C2, C4, C6, T8, T10, P7, P5, P3, P1, PZ, P2, P4, P6, P8, O1, OZ, O2
$2\times$	4	29	AF7, AF3, AFZ, AF4, AF8, FT7, FC5, FC3, FC1, FCZ, FC2, FC4, FC6, FT8, TP7, CP5, CP3, CP1, CPZ, CP2, CP4, CP6, TP8, PO7, PO3, POZ, PO4, PO8, IZ
$4\times$	1	16	FPZ, AF3, AF4, FC5, FC1, FC2, FC6, CP5, CP1, CP2, CP6, PO3, PO4, O1, OZ, O2
$4\times$	2	15	FP1, FP2, F5, FZ, F6, T7, C3, CZ, C4, T8, P5, PZ, P6, O1, O2
$4\times$	3	16	FPZ, AF3, AF4, FT7, FC3, FCZ, FC4, FT8, TP7, CP3, CPZ, CP4, TP8, PO7, OZ, PO8
$4\times$	4	16	FP1, FP2, F7, F3, F4, F8, C5, C1, C2, C6, P7, P3, P4, P8, O1, O2
$8\times$	1	8	AF3, AF4, FC5, FC6, CP5, CP6, PO3, PO4
$8\times$	2	8	FP1, FP2, FZ, T7, T8, PZ, O1, O2
$8\times$	3	8	AF3, AF4, FC3, FC4, P3, P4, PO7, PO8
$8\times$	4	8	FPZ, F5, F6, C3, C4, P5, P6, OZ

Table 11: Seen-electrode layouts S used for EEG interpolation and downstream experiments. Four layouts are used for each downsampling scale on each dataset.

C MAMBA FOR GINR

C.1 ANALYSIS OF SCALING METRICS

We compare the scaling behavior of CANDI with the patch-based GINR baselines used in Section 4.3: LAGINR (Doyup et al., 2023), a Linformer-encoder variant of LAGINR, and TransINR (Chen & Wang, 2022b). All models are evaluated on HCP volumes of size $96^3 \times 10$, while varying the patch size to control the number of input data tokens L_d.

To reduce variation due to decoder cost, we fix the number of learnable tokens to $L_l = 128$ for all models and query the same $96^3 \times 2$ output volume. We report peak GPU memory usage and throughput, with throughput measured as the average number of processed tokens per second over 20 training steps:

$$\text{Throughput} = \frac{20(L_d + L_l)B}{t_{\text{elapsed}}}, \tag{9}$$

where B is the batch size and t_{elapsed} is the total elapsed time over the 20 measured steps.

All measurements are collected on a single NVIDIA A100 40GB GPU with batch size $B = 2$. As shown in Figure 4, the Transformer-based LAGINR and TransINR models exceed the 40GB memory limit at approximately 5,200 input tokens, whereas CANDI and the Linformer variant remain feasible over the evaluated range. CANDI also maintains increasing throughput as sequence length grows, while the Transformer baselines become increasingly memory constrained.

SCENT is omitted from this comparison because its input token count scales with volume size rather than patch size, and therefore does not admit the same patch-size-based sequence-length sweep.

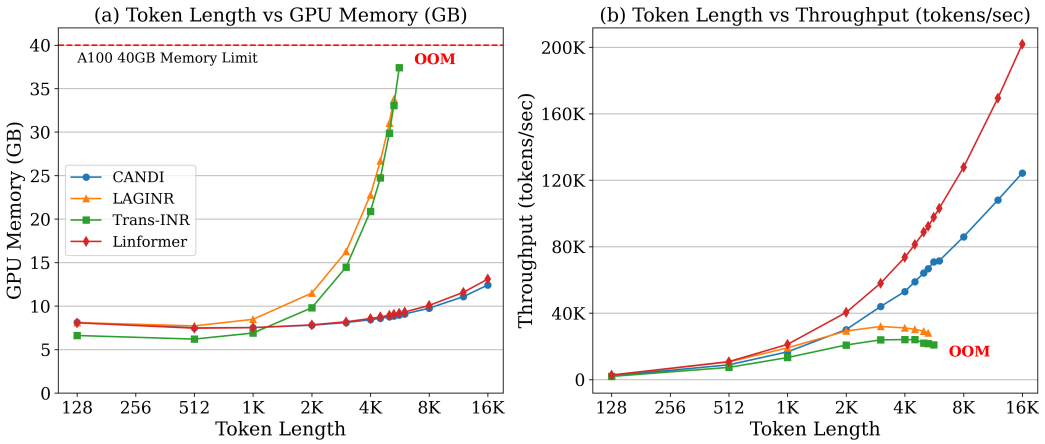

Figure 4: Scaling behavior of CANDI and patch-based GINR baselines as input sequence length increases. a) peak GPU memory usage on a single NVIDIA A100 40GB GPU. b) training throughput measured in processed tokens per second. Transformer-based LAGINR and TransINR exceed the 40GB memory limit at longer sequence lengths, while CANDI and the Linformer-based variant remain feasible over the evaluated range.

C.2 DELTA AND LT T-SNE

We visualize the internal Mamba step-size parameter Δ and encoder token representations using t-SNE on a CIFAR-10 reconstruction model. Figure 5(a) shows separation between data-token and learnable-token positions in the Δ representation, while Figure 5(b) shows a similar separation in CANDI's encoded token representations. In contrast, the corresponding LAGINR representations in Figure 5(c) exhibit substantially greater overlap. These visualizations are exploratory and are included only to illustrate differences in token organization between the two encoders; we do not treat them as quantitative evidence of representation quality. However, this qualitative separation warrants further exploration into the latent representations of these two methods and why they differ.

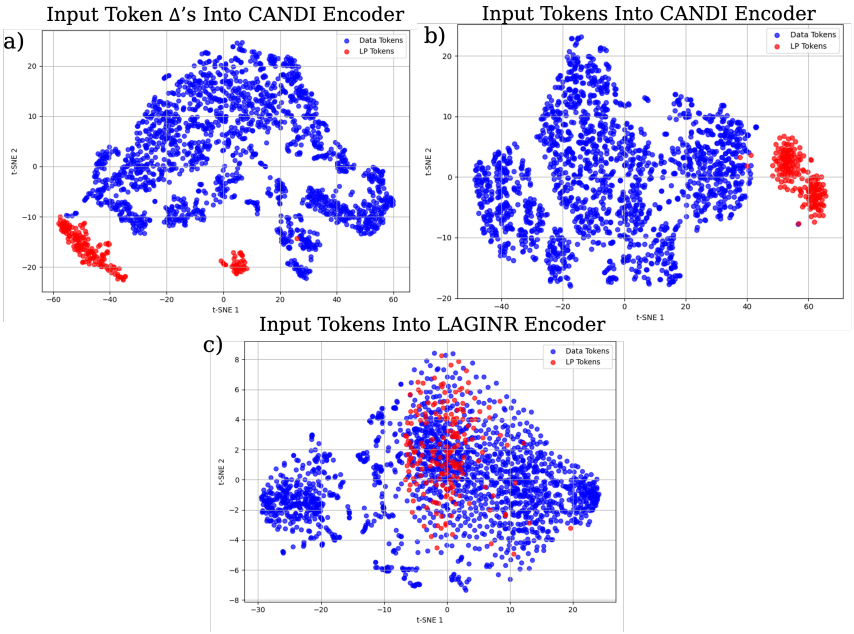

Figure 5: t-SNE visualizations of Mamba and Transformer token representations. (a) Internal Mamba step-size parameter Δ for data-token and learnable-token positions. (b) CANDI encoder token representations. (c) LAGINR encoder token representations.

C.3 Learnable-Token Distribution Studies

Because Mamba processes tokens sequentially, the placement of learnable tokens within the input sequence may affect how information is aggregated. We therefore evaluate both the number and placement of learnable tokens (LTs) using CIFAR-10 reconstruction, similar to the analysis in (Wang et al., 2025).

We compare five placement strategies across LT budgets of $\{64, 128, 192, 256\}$. *Equidistant* inserts LTs uniformly throughout the sequence. *G2*, *G16*, and *G32* partition the data tokens into groups of size 2, 16, and 32, respectively, and place the corresponding LTs contiguously within each group. *Middle* places all LTs near the center of the sequence. All experiments use 2×2 image patches, yielding $L_d = 256$ data tokens. We report reconstruction PSNR.

Table 12 shows that distributing LTs across the full sequence is consistently preferable to concentrating them in a single region. Equidistant and G2 placement perform similarly across LT budgets, while the Middle strategy performs substantially worse. More aggressive grouping also becomes less effective as the LT budget decreases, with G32 showing the largest degradation at 64 LTs.

These results suggest that broad coverage of the input sequence is more important than concentrating learnable tokens locally. We therefore use equidistant LT placement as the default configuration in CANDI.

	LT Distribution Method				
# LTs	Equidistant	G2	G16	G32	Middle
256	<u>54.92</u>	**55.33**	54.72	54.31	51.28
192	**55.04**	<u>54.79</u>	54.15	53.98	49.91
128	**55.11**	<u>54.14</u>	53.77	52.07	48.46
64	**54.52**	<u>53.63</u>	51.85	45.13	48.74

Table 12: Ablation: Learnable Token Distribution. Reports reconstruction PSNR of each configuration. Equidistant refers to a distribution of tokens in every m real data tokens. G2, G16, G32 denote a grouping of 2, 16, and 32, respectively, then every group is placed with equal number of tokens in between.

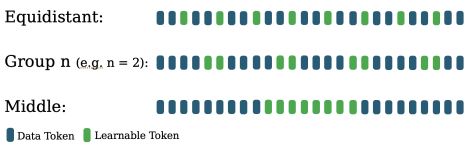

Figure 6: Visual representation of the learnable token groupings described in Table 12 for clarity. Each line represents the sequential layout of all data and learnable tokens.

D fMRI Qualitative Results

Figure 7 shows representative qualitative reconstructions from the HCP fMRI reconstruction task compared to baselines on $48^3 \times 2$ volumes. We visualize matched spatial slices on all three axes from the ground-truth volume and the corresponding reconstructions to complement the quantitative MSE and HFEN results reported in the main text. Across methods, the qualitative differences are consistent with the quantitative reconstruction results: CANDI better preserves fine spatial structure while maintaining the large-scale organization of the target volume.

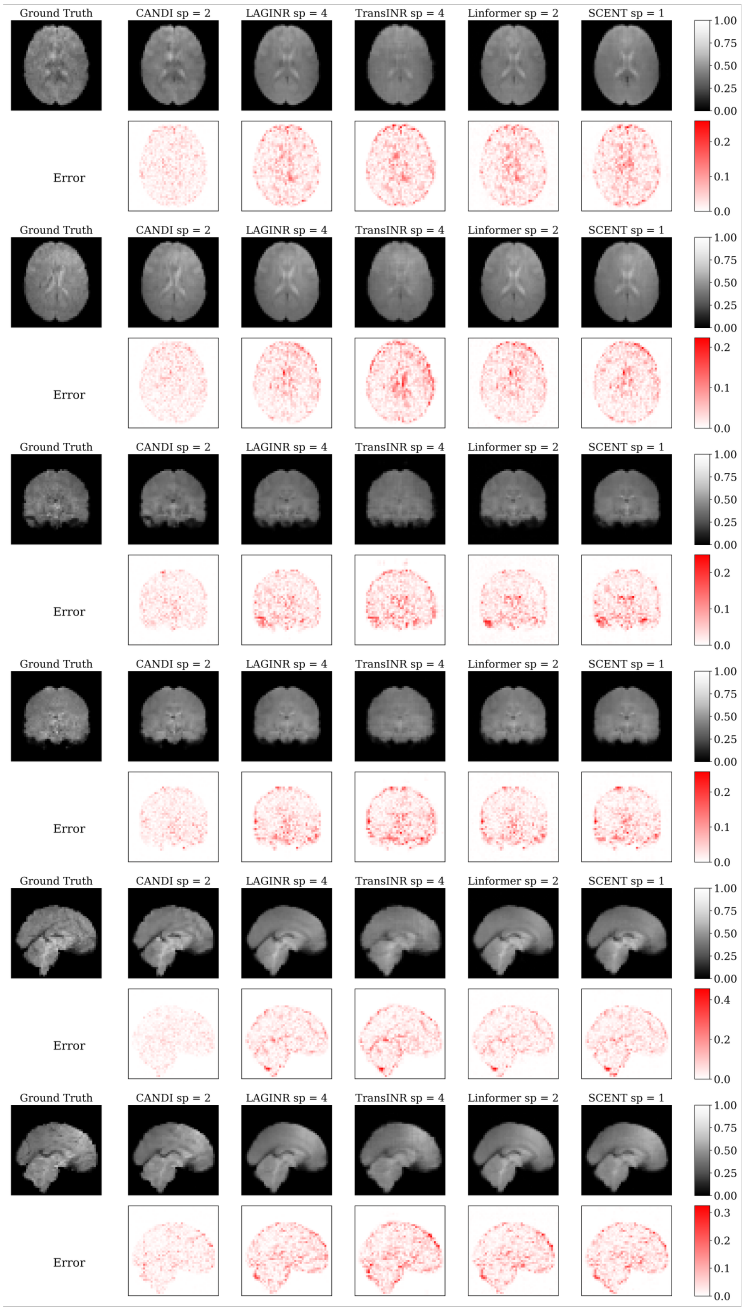

Figure 7: Qualitative fMRI reconstruction examples on HCP at $48^3 \times 2$ resolution. Matched spatial slices from the ground-truth volume are shown alongside reconstructions from CANDI and the compared baselines.