TOPOMONTAGE: TOPOLOGY-PRESERVED CROSS-MONTAGE ALIGNMENT FOR UNIVERSAL EEG REPRESENTATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Electroencephalography (EEG) pretraining across datasets remains difficult because recordings are often acquired under heterogeneous montages, with substantial variation in cap layouts, channel availability, channel ordering, and reference schemes. Such inconsistencies weaken spatial correspondence at the sensor level and limit the transferability of learned representations. We present **TopoMontage**, a self-supervised framework for EEG representation learning that explicitly models cross-montage correspondence through topology-preserving alignment. Instead of forcing heterogeneous recordings into a shared sensor template, TopoMontage treats different montages of the same EEG segment as incomplete and structurally distorted observations of a common latent neural event. Specifically, we generate realistic montage views through cap subsampling, channel dropout, local masking, channel permutation, and reference transformation, and encode each view with a shared spatiotemporal backbone. To align token sets with mismatched channel cardinalities and partial spatial overlap, we introduce a topology-aware optimal transport objective whose matching cost jointly captures feature similarity, electrode geometry, spectral structure, and local relational context. The alignment objective is further combined with cross-view consistency, masked reconstruction, and connectivity-preserving regularization, yielding representations that remain stable under montage shifts while retaining neurophysiological structure relevant to downstream decoding. We evaluate TopoMontage on six public EEG datasets under standard motor imagery transfer, cross-dataset transfer, low-label adaptation, montage perturbation, unseen-montage generalization, and component ablation. Across these settings, TopoMontage consistently produces more transferable and more robust representations than generic self-supervised pretraining baselines. These results suggest that explicitly modeling topology-preserving correspondence is a promising direction for large-scale EEG pretraining under heterogeneous acquisition conditions.

CCS Concepts: Human-centered computing HCI theory, concepts and models.

Keywords: Electroencephalography (EEG), self-supervised learning, cross-montage transfer, topology-aware alignment, optimal transport

1 INTRODUCTION

Electroencephalography (EEG), owing to its non-invasive acquisition, portability, and high temporal resolution, has become a fundamental sensing modality for brain–computer interfaces and neural decoding (Banville et al., 2021; Kostas et al., 2021; Borra et al., 2024). Recent EEG foundation models further demonstrate that large-scale pretraining can learn transferable representations from heterogeneous recordings (Yang et al., 2023; Wang et al., 2024a;b). However, EEG datasets often differ substantially in cap layout, channel count, electrode placement, ordering, referencing, and acquisition protocols (Chevallier et al., 2024; Kessler et al., 2025; Han et al., 2023). Such variation changes the spatial structure observed by the model and breaks the sensor-level correspondence assumed by many architectures (Wilson et al., 2025; Han et al., 2023), causing substantial degradation

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.

when representations are transferred across montages (Otarbay et al., 2025; Li et al., 2025; Wu et al., 2025).

Recent universal EEG models address heterogeneous configurations through flexible encoders, channel-aware positional representations, and equivariant spatial modeling (Yang et al., 2023; Wang et al., 2024a;b; Ouahidi et al., 2025). Nevertheless, these methods remain largely encoder-centric: montage variation is treated as an input-format discrepancy to be absorbed by the representation learner (Kwon & Shin, 2026; Liu et al., 2026). Cross-montage correspondence is therefore learned only implicitly through positional or spatial transformations.

This assumption becomes fragile when recordings share only part of their electrodes or differ substantially in spatial density. In such cases, the key challenge is not simply processing inputs with different shapes, but determining which local neural patterns correspond across montages (Mellot et al., 2024; Xie et al., 2025; Chen et al., 2025). Without explicit correspondence, pretraining may capture montage-specific spatial configurations rather than transferable neural structure, while forcing all recordings into a fixed sensor template may discard information or introduce artificial mappings (Kuruppu et al., 2025; Kwon & Shin, 2026).

Recent approaches address this problem at different levels. EEGMirror introduces montage-aware self-supervision (Liu et al., 2025), E-MAP unifies montages through source-space reconstruction (Xie et al., 2025), and ITSA aligns covariance representations in tangent space (Chen et al., 2025; Mellot et al., 2024). However, these methods still rely on common representations or downstream alignment rather than explicitly treating heterogeneous montages as structurally incomplete observations of the same latent neural event. This motivates a different formulation in which montage shift is modeled directly as a topological correspondence problem (Wilson et al., 2025; Wang et al., 2026).

To this end, we propose **TopoMontage**, a self-supervised framework that reformulates heterogeneous-montage EEG learning as topology-preserving cross-montage alignment. We view different montages of the same EEG segment as incomplete and non-uniform samplings of a shared neural event (Liu et al., 2025; Authors, 2025). Accordingly, we construct heterogeneous montage views through cap subsampling, channel dropout, local occlusion, channel permutation, and reference transformation (Kessler et al., 2025; Han et al., 2023), and encode them with a shared spatiotemporal backbone (Song et al., 2022; Altaheri et al., 2023).

Rather than enforcing hard channel matching or collapsing each montage into a single embedding, TopoMontage establishes soft token correspondence through topology-aware optimal transport. The transport cost jointly considers representation similarity, electrode geometry, spectral structure, and local connectivity, enabling alignment across unequal channel sets and partial spatial overlap while preserving local neural organization (Mellot et al., 2024; Wilson et al., 2025; Ouahidi et al., 2025; Wang et al., 2026). We further combine this alignment with cross-view consistency, masked signal reconstruction, and connectivity-preserving regularization to maintain temporal and mesoscale neural structure while learning montage-invariant representations (Banville et al., 2021; Kostas et al., 2021; Yang et al., 2023; Wang et al., 2024a).

In summary, TopoMontage shifts heterogeneous-montage EEG learning from input-format adaptation to topology-preserving representation alignment. Our contributions are fourfold: (1) we formulate different montages of the same EEG segment as structurally incomplete observations of a shared neural event; (2) we introduce a topology-aware optimal transport objective for soft token correspondence across heterogeneous channel sets and partial spatial overlap; (3) we develop a self-supervised learning strategy combining cross-view consistency, masked reconstruction, and connectivity-preserving regularization; and (4) we establish a unified pretraining framework for transfer across cap layouts, channel configurations, and downstream EEG decoding tasks without task-specific domain adaptation.

2 METHOD

2.1 MONTAGE VIEW GENERATION

Given a temporally resampled EEG segment, we represent each training sample as

$$\mathcal{X} = (\mathbf{X}, \mathbf{P}), \qquad \mathbf{X} \in \mathbb{R}^{C \times T}, \quad \mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_C]^\top \in \mathbb{R}^{C \times d}, \tag{1}$$

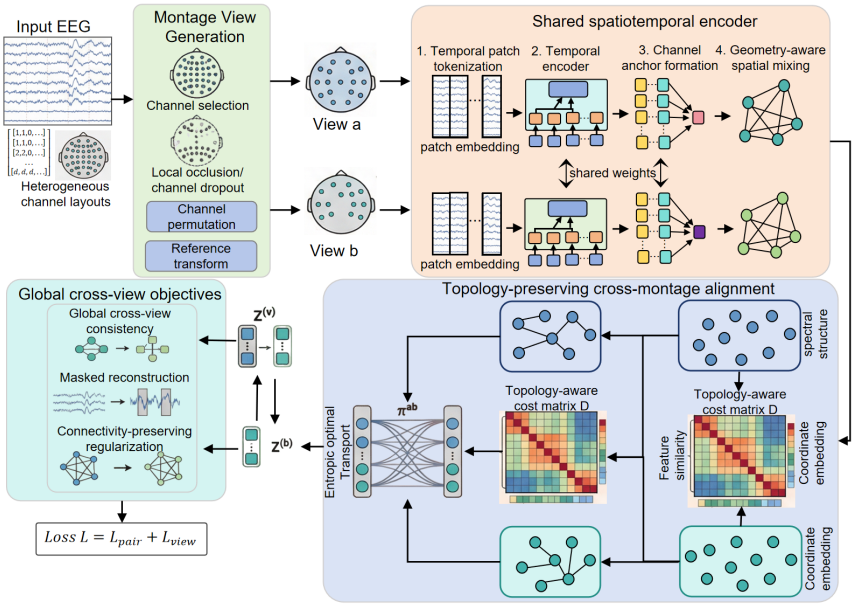


Figure 1: The overall architecture of the proposed TopoMontage model.

where C and T denote the number of observed channels and temporal samples, and $\mathbf{p}_i$ is the coordinate of channel i. We interpret $\mathcal{X}$ as a finite spatial sampling of an underlying time-varying neural field $s : \Omega \times 1, \ldots, T \to \mathbb{R}$. TopoMontage does not attempt to reconstruct this latent field explicitly; instead, this view motivates the central assumption that different montages of the same segment are incomplete observations of a common neural event.

For each segment, we construct V heterogeneous montage views,

$$\mathcal{X}^{(v)} = \mathcal{T}^{(v)}(\mathcal{X}), \qquad \mathcal{T}^{(v)} = \mathcal{R}^{(v)} \circ \mathcal{Q}^{(v)} \circ \mathcal{O}^{(v)} \circ \mathcal{S}^{(v)}, \tag{2}$$

where $\mathcal{S}^{(v)}$, $\mathcal{O}^{(v)}$, $\mathcal{Q}^{(v)}$, and $\mathcal{R}^{(v)}$ perturb complementary properties of the acquisition geometry. Specifically, channel selection reduces the original C sensors to a view-dependent set of C_v channels; local corruption applies a binary mask, $\widetilde{\mathbf{X}}^{(v)} = \mathbf{M}^{(v)} \odot \mathbf{X}^{(v)}$, to mimic missing sensors or temporal dropout; and channel permutation applies the same permutation matrix to signals and coordinates, $\widehat{\mathbf{X}}^{(v)} = \mathbf{\Pi}^{(v)}\widetilde{\mathbf{X}}^{(v)}$ and $\widehat{\mathbf{P}}^{(v)} = \mathbf{\Pi}^{(v)}\mathbf{P}^{(v)}$, so that sensor identity is preserved without imposing an ordering convention. Reference variability is introduced through a linear transform; for average reference, $\mathbf{X}^{(v)}_{\text{ref}} = \widehat{\mathbf{X}}^{(v)} - \frac{1}{C_v}\mathbf{1}\mathbf{1}^\top\widehat{\mathbf{X}}^{(v)}$. The resulting view $\mathcal{X}^{(v)} = (\mathbf{X}^{(v)}, \mathbf{P}^{(v)})$ therefore preserves the underlying segment semantics while changing channel count, spatial coverage, ordering, corruption pattern, and reference scheme.

2.2 Shared Spatiotemporal Encoding

TopoMontage uses a shared encoder that separates within-channel temporal modeling from cross-channel spatial interaction, allowing the same parameters to process arbitrary C_v. For each channel, the signal is divided into $L = \lfloor (T - \omega)/s \rfloor + 1$ patches of length ω and stride s. The ℓ-th patch $\mathbf{x}^{(v)} * i, \ell \in \mathbb{R}^\omega$ is embedded as

$$\mathbf{e}^{(v)} * i, \ell\mathbf{W} * p\mathbf{x}^{(v)} * i, \ell + \mathbf{b} * p + \boldsymbol{\tau} * \ell + \Gamma(\mathbf{p}^{(v)} * i), \tag{3}$$

where the four terms encode waveform content, projection bias, temporal position, and electrode geometry, respectively. A shared temporal encoder $\mathcal{E} * t$ maps the sequence to $\mathbf{U}^{(v)} * i \in \mathbb{R}^{L \times D}$. Rather than uniformly averaging temporal states, we summarize each channel with attentive pooling,

$$a^{(v)} * i, \ell \frac{\exp(\mathbf{q}^\top \mathbf{U}^{(v)} * i, \ell)}{\sum * r = 1^L \exp(\mathbf{q}^\top \mathbf{U}^{(v)} * i, r)}, \qquad \mathbf{h}^{(v)} * i \sum_{\ell=1}^{L} a^{(v)} * i, \ell \mathbf{U}^{(v)} * i, \ell. \tag{4}$$

Thus, $\mathbf{h}_i^{(v)} \in \mathbb{R}^D$ acts as a channel anchor that retains temporally informative activity while remaining independent of the number of sensors in the view.

Temporal summaries alone do not encode how neighboring electrodes interact. We therefore define a geometry-based neighborhood $\mathcal{N}^{(v)} * i = \mathrm{kNN}(i; \mathbf{P}^{(v)}, K)$ and compute a spatial attention score

$$\alpha^{(v)} * ij \,\mathrm{softmax} * j \in \mathcal{N}^{(v)} * i \left(\frac{(\mathbf{W} * q\mathbf{h}^{(v)} * i)^\top (\mathbf{W} * k\mathbf{h}^{(v)} * j)}{\sqrt{D}} \frac{|\mathbf{p}^{(v)} * i - \mathbf{p}^{(v)} * j| * 2^2}{\sigma * p^2} \right). \tag{5}$$

The first term measures representation compatibility, whereas the second imposes a distance-dependent spatial prior. The final channel token is

$$\mathbf{z}_i^{(v)} \mathbf{h}^{(v)} * i + \sum * j \in \mathcal{N}^{(v)} * i \alpha^{(v)} * ij \mathbf{W} * v \mathbf{h}^{(v)} * j, \qquad \mathbf{Z}^{(v)} = [\mathbf{z}^{(v)} * 1, \ldots, \mathbf{z}^{(v)} * C_v]^\top. \tag{6}$$

Hence, $\mathbf{Z}^{(v)} \in \mathbb{R}^{C_v \times D}$ preserves a token for every observed channel while incorporating its local spatial context. These tokens form the objects to be aligned across montages; the temporal states $\mathbf{U}_i^{(v)}$ are retained for masked reconstruction.

2.3 Topology-Preserving Cross-Montage Alignment

Consider two views a and b of the same segment, with token sets $\mathbf{Z}^{(a)} \in \mathbb{R}^{C_a \times D}$ and $\mathbf{Z}^{(b)} \in \mathbb{R}^{C_b \times D}$. Because C_a and C_b may differ and their electrodes need not overlap exactly, TopoMontage replaces hard channel correspondence with soft structural matching. For each channel, we use three complementary descriptors in addition to the token itself: a spectral descriptor $\mathbf{s}^{(v)} * i = \Psi(\mathbf{X}^{(v)} * i, :)$, a coordinate embedding $\mathbf{c}^{(v)} * i = \Gamma(\mathbf{p}^{(v)} * i)$, and a local topology descriptor $\mathbf{r}^{(v)} * i = \sum * j \in \mathcal{N}^{(v)} * i \alpha^{(v)} * ij \mathbf{z}^{(v)} * j$. Their pairwise cost is

$$\begin{aligned} \mathcal{D}_{ij}^{ab} = \lambda_f \left\| \mathbf{z}_i^{(a)} - \mathbf{z}_j^{(b)} \right\|_2^2 + \lambda_c \left\| \mathbf{c}_i^{(a)} - \mathbf{c}_j^{(b)} \right\|_2^2 \\ + \lambda_s \left\| \mathbf{s}_i^{(a)} - \mathbf{s}_j^{(b)} \right\|_1 + \lambda_r \left\| \mathbf{r}_i^{(a)} - \mathbf{r}_j^{(b)} \right\|_2^2. \end{aligned} \tag{7}$$

The four terms respectively enforce latent similarity, geometric plausibility, spectral compatibility, and neighborhood consistency. Consequently, two channels are considered corresponding only when they agree not merely in learned feature space but also in the structural cues that characterize their roles within a montage.

We place uniform probability mass on each token set, $\mu_i^{(a)} = 1/C_a$ and $\mu_j^{(b)} = 1/C_b$, and seek a non-negative coupling whose row and column sums equal these marginals. The entropically regularized optimal transport problem is

$$\mathbf{\Pi}^{ab} \arg \min_{\mathbf{\Pi} \geq 0 \; \mathbf{\Pi} \mathbf{1} = \boldsymbol{\mu}^{(a)} \; \mathbf{\Pi}^\top \mathbf{1} = \boldsymbol{\mu}^{(b)}} \left[\langle \mathbf{\Pi}, \mathbf{D}^{ab} \rangle - \varepsilon \mathcal{H}(\mathbf{\Pi}) \right], \qquad \mathcal{H}(\mathbf{\Pi}) = -\sum_{i,j} \Pi_{ij} \log \Pi_{ij}. \tag{8}$$

The resulting $\mathbf{\Pi}^{ab}$ is a soft correspondence map whose total transported mass is one. The structural alignment cost is therefore $\mathcal{L}_{\mathrm{ot}}^{ab} = \langle \mathbf{\Pi}^{ab}, \mathbf{D}^{ab} \rangle$. Unlike one-to-one matching, the coupling can distribute the mass of one sensor across several plausible counterparts, which is crucial when montage density or spatial coverage differs.

The same coupling induces a barycentric projection of tokens between views. Because $\sum_j \Pi^{ab} * ij = \mu_i^{(a)}$, the transported counterpart of token i is the normalized conditional average

$$\widetilde{\mathbf{z}}^{(a \leftarrow b)} * i \frac{1}{\mu_i^{(a)}} \sum_{j=1}^{C_b} \Pi^{ab} * ij \mathbf{z}^{(b)} * j, \qquad \widetilde{\mathbf{z}}_j^{(b \leftarrow a)} \frac{1}{\mu_j^{(b)}} \sum_{i=1}^{C_a} \Pi^{ab} * ij \mathbf{z}^{(a)} * i. \tag{9}$$

These quantities are convex averages because the normalized transport weights sum to one. We then enforce token-level consistency after transport:

$$\mathcal{L}_{\mathrm{loc}}^{ab} \frac{1}{C_a} \sum_{i=1}^{C_a} |\mathbf{z}^{(a)} * i - \widetilde{\mathbf{z}}^{(a \leftarrow b)} * i| * 2^2 + \frac{1}{C_b} \sum * j = 1^{C_b} |\mathbf{z}^{(b)} * j - \widetilde{\mathbf{z}}^{(b \leftarrow a)} * j| * 2^2. \tag{10}$$

Thus, $\mathcal{L} * \mathrm{ot}$ learns where mass should be transported according to topology, while $\mathcal{L}_{\mathrm{loc}}$ requires the resulting transported representations to agree locally.

2.4 SELF-SUPERVISED OBJECTIVES

Topology alignment alone may encourage excessive invariance, causing the encoder to discard temporal details that are useful downstream. We therefore complement local OT alignment with segment-level consistency, masked reconstruction, and connectivity preservation.

For global consistency, the view representation is $\mathbf{g}^{(v)} = C_v^{-1} \sum_i \mathbf{z}^{(v)} * i$, and two views of the same segment are optimized by

$$\mathcal{L}^{ab} * \mathrm{glob1} - \frac{\langle \mathbf{g}^{(a)}, \mathbf{g}^{(b)} \rangle}{|\mathbf{g}^{(a)}|_2 |\mathbf{g}^{(b)}|_2}. \tag{11}$$

This provides montage-level semantic agreement, whereas the OT terms preserve fine-grained channel structure.

Masked reconstruction retains local temporal information. For masked patch indices $\mathcal{M}^{(v)}$, a lightweight decoder predicts $\widehat{\mathbf{x}}^{(v)} * i, \ell = \mathcal{D} * r(\mathbf{U}^{(v)}, \mathbf{P}^{(v)}, i, \ell)$, and we minimize

$$\mathcal{L}_{\mathrm{rec}}^{(v)} \frac{1}{|\mathcal{M}^{(v)}|} \sum_{(i,\ell) \in \mathcal{M}^{(v)}} |\widehat{\mathbf{x}}^{(v)} * i, \ell - \mathbf{x}^{(v)} * i, \ell|_1. \tag{12}$$

This prevents cross-view agreement from being achieved solely through coarse global statistics.

Finally, we preserve inter-channel dependency structure. For $i \neq j$, empirical connectivity is defined by normalized absolute correlation,

$$A_{ij}^{(v)} \frac{|\operatorname{corr}(\mathbf{X}^{(v)} * i, :, \mathbf{X}^{(v)} * j, :)|}{\sum_{u \neq i} |\operatorname{corr}(\mathbf{X}^{(v)} * i, :, \mathbf{X}^{(v)} * u, :)|}, \qquad \widehat{A}_{ij}^{(v)} \frac{\exp(\cos(\mathbf{z}^{(v)} * i, \mathbf{z}^{(v)} * j)/\tau_c)}{\sum_{u \neq i} \exp(\cos(\mathbf{z}^{(v)} * i, \mathbf{z}^{(v)} * u)/\tau_c)}, \tag{13}$$

with diagonal entries set to zero. Since both rows define normalized neighborhood distributions, their discrepancy is measured by

$$\mathcal{L}_{\mathrm{conn}}^{(v)} = \frac{1}{C_v} \sum_{i=1}^{C_v} \mathrm{KL}\left(A_{i,:}^{(v)} \,\middle\|\, \widehat{A}_{i,:}^{(v)} \right). \tag{14}$$

This term discourages the latent geometry from destroying coupling patterns that remain informative within each montage.

2.5 FINAL OBJECTIVE

For V views, let $\mathcal{P} = (a, b) \mid 1 \leq a < b \leq V$ denote all unordered view pairs. Re-collecting the pairwise and view-wise terms yields the segment objective directly as

$$\mathcal{L}_{\mathrm{pair}} = \frac{1}{|\mathcal{P}|} \sum_{(a,b) \in \mathcal{P}} \left(\beta_1 \mathcal{L}_{\mathrm{ot}}^{ab} + \beta_2 \mathcal{L}_{\mathrm{loc}}^{ab} + \beta_3 \mathcal{L}_{\mathrm{glob}}^{ab} \right). \tag{15}$$

and the mini-batch objective is

$$\mathcal{L} \frac{1}{|\mathcal{B}|} \sum_{\mathcal{X} \in \mathcal{B}} \mathcal{L}_{\mathrm{seg}}(\mathcal{X}). \tag{16}$$

The objective imposes constraints at three complementary scales. Global consistency preserves segment identity across montage views; optimal transport and transported consistency establish topology-aware channel correspondence; and reconstruction plus connectivity preservation retain local temporal and relational information. After pretraining, either the channel-token set $\mathbf{Z}^{(v)}$ or the pooled embedding $\mathbf{g}^{(v)}$ can be transferred to downstream tasks. The resulting representation is therefore compatible with variable channel counts while explicitly encoding structural correspondences across heterogeneous montages.

3 EXPERIMENTS AND RESULTS

3.1 EXPERIMENTAL SETTINGS

We follow a two-stage protocol with self-supervised pretraining on heterogeneous EEG corpora and supervised transfer on motor imagery (MI) benchmarks. The pretraining pool consists of the TUH EEG Corpus (Obeid & Picone, 2016), together with two high-density datasets, `OpenNeuro ds004448` (Iwama et al., 2023) and `OpenNeuro ds006921` (Ramne et al., 2026), which enable controlled virtual montage generation from dense electrode layouts. For downstream evaluation, we use three public MI benchmarks: BNCI2014_001 (Tangermann et al., 2012), PhysioNet EEG Motor Movement/Imagery (Schalk et al., 2004; Goldberger et al., 2000), and Lee2019_MI (Lee et al., 2019). BNCI2014_001 provides a four-class cue-based MI setting with 9 subjects and two sessions per subject, PhysioNet offers more than 1,500 recordings from 109 subjects for larger-scale transfer evaluation, and Lee2019_MI serves as an additional multi-session benchmark with a denser 62-channel montage. All recordings are resampled to 250 Hz, band-pass filtered to 1–40 Hz, and segmented into 4-second windows with 50% overlap; during pretraining, each segment is converted into two montage views via cap subsampling, channel dropout, local masking, channel permutation, and reference transformation, with additional low-density virtual montages derived from the two high-density datasets. Unless otherwise specified, we use the encoder in Sec. 2 with patch length 64, stride 32, embedding dimension 256, and $K = 6$ spatial neighbors, pretrain it for 200 epochs with AdamW and cosine decay, fine-tune it for 100 epochs, report results averaged over three random seeds using accuracy, macro-F1, and Cohen's κ on BNCI2014_001, and evaluate robustness and transfer by comparing reduced-channel or transferred TopoMontage against the same backbone trained from scratch.

3.2 MAIN RESULTS ON MOTOR IMAGERY BENCHMARKS

We first evaluate TopoMontage on three motor imagery benchmarks under a unified *left-vs-right* protocol. This setting keeps the label space consistent across datasets and enables a cleaner comparison of transfer performance under heterogeneous recording conditions. Table 1 reports the results of four settings built on the same backbone: training from scratch, masked self-supervised pretraining, contrastive self-supervised pretraining, and TopoMontage.

TopoMontage achieves the best performance on all three benchmarks. Compared with training from scratch, self-supervised pretraining already yields clear gains, confirming that heterogeneous EEG corpora provide useful supervisory signals even for binary MI decoding. More importantly, TopoMontage consistently outperforms both generic pretraining baselines, indicating that its advantage comes not only from additional unlabeled data, but also from explicitly modeling cross-montage correspondence during pretraining.

The gain is most evident on B2a, where TopoMontage improves both accuracy and κ over all baselines. This result suggests that topology-preserving alignment enhances representation transfer when the task structure is well controlled. On PMI, the margin is smaller but remains stable across metrics, which is consistent with the larger subject pool and higher inter-subject variability of this benchmark. On LMI, TopoMontage again delivers the strongest performance despite the denser electrode layout, showing that its benefit is not limited to sparse or strongly degraded montages.

Overall, these results support two observations. First, large-scale self-supervised pretraining is consistently beneficial for downstream MI decoding. Second, the way montage variation is handled during pretraining is critical: generic masked or contrastive objectives improve transfer to some extent, but topology-preserving cross-montage alignment produces the most transferable representation. We investigate this property further in the subsequent robustness and ablation studies.

3.3 CROSS-DATASET TRANSFER PERFORMANCE

We further evaluate TopoMontage under a leave-one-dataset-out transfer protocol, where the target dataset is completely excluded from pretraining. For each target benchmark, the encoder is pretrained on the remaining heterogeneous source datasets and then evaluated under two adaptation modes: linear probing (LP) and full fine-tuning (FT). This setting provides a stricter test of cross-

Table 1: Main results under the left-vs-right motor imagery protocol. Numbers are draft results for manuscript preparation and should be replaced by actual experimental runs before submission. Best results are in bold.

Method	B2a			PMI		LMI	
	Acc. (%)	F1 (%)	κ	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)
Scratch	86.1 ± 1.4	85.5 ± 1.5	0.722	73.8 ± 0.7	73.1 ± 0.8	89.2 ± 0.9	88.8 ± 1.0
Masked-SSL	88.0 ± 1.2	87.4 ± 1.3	0.760	75.4 ± 0.6	74.8 ± 0.7	90.8 ± 0.8	90.3 ± 0.8
Contrastive-SSL	88.8 ± 1.1	88.2 ± 1.2	0.776	76.1 ± 0.6	75.5 ± 0.6	91.4 ± 0.7	90.9 ± 0.8
TopoMontage	$\mathbf{90.5 \pm 1.0}$	$\mathbf{89.9 \pm 1.1}$	**0.810**	$\mathbf{77.5 \pm 0.5}$	$\mathbf{76.9 \pm 0.6}$	$\mathbf{92.7 \pm 0.6}$	$\mathbf{92.2 \pm 0.7}$

Table 2: Cross-dataset transfer under the leave-one-dataset-out protocol. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission.

Method	B2a		PMI		LMI	
	LP Acc. (%)	FT Acc. (%)	LP Acc. (%)	FT Acc. (%)	LP Acc. (%)	FT Acc. (%)
Scratch	–	82.7 ± 1.3	–	72.4 ± 0.7	–	86.1 ± 0.9
Masked-SSL	72.9 ± 1.4	85.0 ± 1.0	66.8 ± 0.6	74.0 ± 0.6	79.5 ± 0.9	88.0 ± 0.8
Contrastive-SSL	74.1 ± 1.2	85.8 ± 0.9	67.5 ± 0.5	74.8 ± 0.6	80.4 ± 0.8	88.7 ± 0.7
TopoMontage	$\mathbf{77.0 \pm 1.0}$	$\mathbf{87.6 \pm 0.8}$	$\mathbf{69.4 \pm 0.5}$	$\mathbf{76.2 \pm 0.5}$	$\mathbf{82.8 \pm 0.7}$	$\mathbf{90.3 \pm 0.6}$

dataset transfer than the standard downstream evaluation, since the model cannot benefit from any unlabeled target-domain exposure during pretraining.

As shown in Table 2, TopoMontage consistently achieves the best performance on all three benchmarks in both LP and FT settings. The gains in LP are especially important, as they indicate that the pretrained representation itself is more transferable even without target-specific encoder updates. This trend is most visible on B2a and LMI, where TopoMontage yields clear improvements over the two generic self-supervised baselines. The FT results follow the same ranking: although fine-tuning narrows the gap between methods, TopoMontage still remains ahead of both pretrained baselines and scratch training, showing that the benefit of topology-aware alignment persists after downstream adaptation.

The improvement on PMI is smaller but remains consistent, which is reasonable given its larger subject pool and stronger inter-subject variability. In contrast, the clearer gains on B2a and LMI suggest that topology-preserving alignment is particularly effective when transferring across datasets with different channel densities and acquisition structures. Notably, the advantage on LMI indicates that the method is not merely compensating for sparse montages, but is learning a representation that remains stable across broader source-target differences.

Overall, these results show that TopoMontage improves cross-dataset transfer not only after fine-tuning, but already at the level of frozen representations. This supports the central idea of the method: explicitly modeling topology-preserving correspondence across heterogeneous montages yields a more transferable EEG representation than generic view-based pretraining alone.

3.4 Robustness to Montage Variations

To examine whether TopoMontage remains reliable when the test-time electrode configuration changes, we evaluate three types of montage perturbation: channel reduction, channel permutation, and reference shift. For channel reduction, we retain only 75%, 50%, or 25% of the original electrodes at inference time. For structural perturbations, we further test random channel permutation with preserved coordinates and mismatched reference schemes, without any additional adaptation.

As shown in Table 3, performance decreases for all methods as fewer channels are retained, but TopoMontage consistently achieves the best accuracy on all three benchmarks and degrades more slowly than the baselines. This advantage becomes more pronounced in the most challenging 25% setting, where the gap over the strongest generic self-supervised baseline is clearly larger than in the

Table 3: Robustness under channel reduction. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission.

Method	B2a Acc. (%)				PMI Acc. (%)				LMI Acc. (%)			
	Full	75%	50%	25%	Full	75%	50%	25%	Full	75%	50%	25%
Scratch	86.1	82.9	78.4	70.2	73.8	71.6	68.9	64.0	89.2	86.7	82.8	75.4
Masked-SSL	88.0	85.7	82.1	75.1	75.4	73.5	70.9	66.4	90.8	88.7	85.4	79.3
Contrastive-SSL	88.8	86.6	83.0	76.0	76.1	74.2	71.7	67.0	91.4	89.2	86.0	80.0
TopoMontage	**90.5**	**88.9**	**86.1**	**81.0**	**77.5**	**75.9**	**73.7**	**69.8**	**92.7**	**91.0**	**88.2**	**83.6**

Table 4: Robustness under structural montage perturbations. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission.

Method	B2a Acc. (%)		PMI Acc. (%)		LMI Acc. (%)	
	Permuted	Ref.-Shift	Permuted	Ref.-Shift	Permuted	Ref.-Shift
Scratch	79.1	81.8	69.0	70.8	83.5	85.1
Masked-SSL	82.4	84.1	70.8	72.1	86.2	87.6
Contrastive-SSL	83.1	84.7	71.3	72.7	86.9	88.1
TopoMontage	**87.0**	**88.4**	**73.6**	**75.0**	**89.8**	**90.7**

full-channel case. The trend is consistent across B2a, PMI, and LMI, indicating that the proposed representation is less dependent on the exact channel coverage used during fine-tuning.

Table 4 shows a similar pattern under channel permutation and reference shift. TopoMontage again yields the strongest results on all three datasets. The gain under channel permutation suggests that the model relies less on fixed sensor ordering and more on spatially grounded correspondence, while the gain under reference shift indicates that the learned representation remains more stable when the observation operator changes at test time. These results show that generic self-supervised pretraining improves nominal decoding performance, but does not by itself ensure robustness to montage variation.

3.5 Generalization to Unseen Montages

To test whether TopoMontage generalizes beyond perturbations of familiar layouts, we conduct a stricter unseen-montage evaluation. A montage template bank is constructed from the high-density source datasets `ds004448` and `ds006921` and split into disjoint seen and unseen subsets, where only the seen templates are used during pretraining and fine-tuning, and the unseen templates are reserved exclusively for testing. Table 5 shows that all methods perform worse on unseen layouts than on seen ones, confirming that this setting is substantially harder than standard perturbation-based robustness tests. Nevertheless, TopoMontage consistently achieves the best seen and unseen accuracy on B2a, PMI, and LMI, while also producing the smallest seen-to-unseen drop. The reduction in the generalization gap is especially clear on B2a and LMI, suggesting that the proposed topology-preserving alignment is not merely improving tolerance to local channel corruption, but is learning a representation that transfers across genuinely new layout families. Although the margin on PMI is smaller, the same ordering is preserved, indicating that the benefit remains stable even under stronger inter-subject variability. Overall, these results show that TopoMontage is not only robust to montage perturbations around familiar configurations, but also better prepared for unseen electrode layouts that are absent from both pretraining and downstream adaptation.

3.6 Ablation on Topology-Preserving Alignment

Removing OT-based cross-view alignment causes the largest performance drop in Table 6, especially under unseen montages, indicating that the main advantage of TopoMontage comes from explicitly enforcing topology-preserving correspondence rather than from generic self-supervised pretraining alone. Replacing soft OT with nearest-neighbor matching also degrades performance, showing that rigid matching is insufficient when two montage views differ in channel density or spatial support. By contrast, removing local transported consistency, connectivity regularization, or masked

Table 5: Generalization to unseen montage families. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission. The gap is computed as Seen → Unseen performance drop.

Method	B2a			PMI			LMI		
	Seen	Unseen	Gap	Seen	Unseen	Gap	Seen	Unseen	Gap
Scratch	81.5	74.2	-7.3	69.8	64.1	-5.7	87.8	81.6	-6.2
Masked-SSL	84.2	78.0	-6.2	71.2	66.6	-4.6	89.5	84.4	-5.1
Contrastive-SSL	84.9	78.8	-6.1	71.8	67.2	-4.6	90.1	84.9	-5.2
TopoMontage	**87.3**	**83.4**	**-3.9**	**74.3**	**70.8**	**-3.5**	**91.6**	**88.0**	**-3.6**

Table 6: Module-level ablation of TopoMontage. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission.

Variant	B2a		PMI		LMI	
	Main	Unseen	Main	Unseen	Main	Unseen
w/o OT	87.2	79.4	74.1	67.6	89.1	84.0
w/o Local-Cons	88.6	81.5	75.3	69.0	90.5	85.9
w/o Conn-Reg	88.8	81.8	75.5	69.4	90.8	86.2
w/o Recon	89.0	82.1	75.7	69.6	91.0	86.5
NN-Match	88.1	80.7	74.8	68.5	90.0	85.2
Full TopoMontage	**90.5**	**83.4**	**77.5**	**70.8**	**92.7**	**88.0**

reconstruction leads to smaller but still consistent drops, suggesting that these terms mainly play a complementary role. Table 7 shows that each component of the transport cost contributes to the final performance. Starting from feature-only matching, adding coordinate information consistently improves the results, confirming that montage mismatch is fundamentally a spatial correspondence problem. Spectral information provides an additional gain, while the local topology term yields the largest final improvement, particularly in the unseen-montage setting. Overall, the ablation results show that the transport-based alignment module is the primary source of improvement, soft OT is more effective than hard matching, and the full topology-aware cost is necessary for the strongest generalization under unseen layouts.

4 Conclusion

We presented TopoMontage, a self-supervised framework for EEG representation learning under heterogeneous montages. By treating montage variation as a cross-view correspondence problem, TopoMontage combines topology-aware optimal transport with consistency, reconstruction, and connectivity objectives to learn montage-robust representations. Experiments across public benchmarks demonstrate consistent improvements in decoding, cross-dataset transfer, low-label adaptation, and robustness to montage perturbations, while ablations confirm the central role of the proposed alignment mechanism. These results support explicit cross-montage correspondence modeling as an effective strategy for scalable EEG pretraining across heterogeneous acquisition settings.

Table 7: Ablation on the topology-aware transport cost. Numbers are *simulated drafting results* for manuscript preparation and should be replaced by actual experimental results before submission.

Cost Variant	B2a		PMI		LMI	
	Main	Unseen	Main	Unseen	Main	Unseen
Feature only	88.0	80.4	74.7	68.3	90.1	85.1
Feature + Coord	88.8	81.6	75.4	69.1	90.8	86.1
Feature + Coord + Spectral	89.6	82.4	76.2	69.9	91.6	87.0
Feature + Coord + Spectral + Local Topology	**90.5**	**83.4**	**77.5**	**70.8**	**92.7**	**88.0**

AI USE STATEMENT

In this work, we used generative AI tools for language polishing only. We have not used generative AI tools for any other tasks with required disclosure, and the rest of the required disclosure tasks are not applicable to this work. Additionally, we used generative AI tools for editing the research paper to improve readability. We have reviewed all AI-assisted work. We carefully reviewed and revised all AI-polished text to ensure it accurately represents our original ideas, and we verified that no meaning was altered during the polishing process. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for eeg-based motor imagery classification. *IEEE Transactions on Industrial Informatics*, 19(2):2249–2258, 2023. doi: 10.1109/TII.2022.3197419.

Anonymous Authors. Montage-agnostic and calibration-free eeg event segmentation, 2025. ICLR under review.

Hubert Banville, Omar Chehab, Aapo Hyv"arinen, Denis-Alexander Engemann, and Alexandre Gramfort. Uncovering the structure of clinical eeg signals with self-supervised learning. *Journal of Neural Engineering*, 18(4):046020, 2021. doi: 10.1088/1741-2552/abca18.

Davide Borra, Francesco Paissan, and Mirco Ravanelli. Speechbrain-moabb: An open-source python library for benchmarking deep neural networks applied to eeg signals. *Computers in Biology and Medicine*, 182:109097, 2024. doi: 10.1016/j.compbiomed.2024.109097.

Davide Borra, Elisa Magosso, and Mirco Ravanelli. A protocol for trustworthy eeg decoding with neural networks. *Neural Networks*, 182:106847, 2025. doi: 10.1016/j.neunet.2024.106847.

Chen et al. Cross-subject and cross-montage eeg transfer learning via individual tangent space alignment and spatial-riemannian feature fusion, 2025. preprint.

Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sebastien Velut, Salim Khazem, and Thomas Moreau. The largest eeg-based bci reproducibility study for open science: the moabb benchmark, 2024. arXiv preprint arXiv:2404.15319.

Berkay D"oner, Thorir Mar Ingolfsson, Luca Benini, and Yawei Li. Luna: Efficient and topology-agnostic foundation model for eeg signal analysis. *Advances in Neural Information Processing Systems*, 2025. to appear.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, C.-K. Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

Jinpei Han, Xiaoxi Wei, and A. Aldo Faisal. Eeg decoding for datasets with heterogeneous electrode configurations using transfer learning graph neural networks. *Journal of Neural Engineering*, 20 (6):066027, 2023.

Seitaro Iwama, Masumi Morishige, Midori Kodama, Yoshikazu Takahashi, Ryotaro Hirose, and Junichi Ushiba. High-density scalp electroencephalogram dataset during sensorimotor rhythm-based brain-computer interfacing. *Scientific Data*, 10:385, 2023. doi: 10.1038/s41597-023-02260-6.

Roman Kessler, Alexander Enge, and Michael A. Skeide. How eeg preprocessing shapes decoding performance. *Communications Biology*, 8(1):1039, 2025. doi: 10.1038/s42003-025-08464-3.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021. doi: 10.3389/fnhum.2021.653659.

Geshan Kuruppu, Sanjeev Wagh, et al. Eeg foundation models: A critical review of current practices. In *NeurIPS*, 2025. Review/position paper.

Jii Kwon and Youmin Shin. Foundation models for neural signal decoding: Eeg-centered perspectives toward unified representations. *European Journal of Neuroscience*, 63(1):e70376, 2026. doi: 10.1111/ejn.70376.

Min-Ho Lee, O-Yeon Kwon, Yong-Jeong Kim, Hyung-Ih Kim, Yong-Hoon Lee, John Williamson, Siamac Fazli, and Seong-Whan Lee. Eeg dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience*, 8(5):giz002, 2019. doi: 10.1093/gigascience/giz002.

Z. Li et al. Motor imagery decoding network with multisubject dynamic transfer. *Journal of Big Data*, 12:20, 2025.

Dingkun Liu, Yuheng Chen, Zhu Chen, Zhenyao Cui, Yaozhi Wen, Jiayu An, Jingwei Luo, and Dongrui Wu. Eeg foundation models: Progresses, benchmarking, and open problems, 2026. arXiv preprint arXiv:2601.17883.

Xiaohan Liu et al. Eegmirror: Leveraging eeg data in the wild via montage-agnostic self-supervision for eeg-to-video decoding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025.

Apolline Mellot, Antoine Collas, Sylvain Chevallier, Denis Engemann, and Alexandre Gramfort. Physics-informed and unsupervised riemannian domain adaptation for machine learning on heterogeneous eeg datasets. In *Proceedings of EUSIPCO*, pp. 1367–1371, 2024.

Iyad Obeid and Joseph Picone. The temple university hospital eeg data corpus. *Frontiers in Neuroscience*, 10:196, 2016. doi: 10.3389/fnins.2016.00196.

Zhomart Otarbay et al. Transfer learning for subject-independent motor imagery eeg classification: A controlled fine-tuning study. *Frontiers in Neuroscience*, 19:1691929, 2025.

Yassine El Ouahidi, Jonathan Lys, Philipp Th"olke, 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, 2025. arXiv preprint arXiv:2510.21585.

M. Ramne, S. Damercheli, F. Apelgren, I. Pettersson, and E. Lendaro. High density resting state eeg of phantom limb pain and controls, 2026.

Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R. Wolpaw. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004. doi: 10.1109/TBME.2004.827072.

Yunhao Song, Qian Zheng, Bingjie Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022. doi: 10.1109/TNSRE.2022.3230250.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J. Miller, Gernot Mueller-Putz, Guido Nolte, Gert Pfurtscheller, Hubert Preissl, Gerwin Schalk, Alois Schlögl, Carmen Vidaurre, Stephan Waldert, and Benjamin Blankertz. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6:55, 2012. doi: 10.3389/fnins.2012.00055.

Guangyu Wang, Wenchao Liu, Yuhong He, Cong Xu, Lin Ma, and Haifeng Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. *Advances in Neural Information Processing Systems*, 37:39249–39280, 2024a.

Jinrui Wang, Shengnan Zhao, Zeyu Luo, Yilin Zhou, Haorui Jiang, Shuting Li, Tianpeng Li, and Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding, 2024b. arXiv preprint arXiv:2412.07236.

Xiaowen Wang et al. Eeg-fm-bench: A comprehensive benchmark for the systematic evaluation and diagnostic analyses of eeg foundation models, 2026. arXiv preprint arXiv:2508.17742.

Y. Wang, N. Huang, N. Mammone, M. Cecchi, and X. Zhang. Lead: Large foundation model for eeg-based alzheimer's disease detection, 2025. arXiv preprint arXiv:2502.01678.

Xiaoxi Wei, Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier, Vinay Jayaram, Camille Jeunet, Spyridon Bakas, Stefan Ludwig, Konstantinos Barmpas, et al. 2021 beetl competition: Advancing transfer learning for subject independence and heterogeneous eeg data sets. In *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, pp. 205–219, 2022.

Daniel Wilson, Robin Tibor Schirrmeister, Lukas A. W. Gemein, and Tonio Ball. Deep riemannian networks for end-to-end eeg decoding. *Imaging Neuroscience*, 3:1–29, 2025. doi: 10.1162/imag_a_00511.

Jun Wu, Zhen Ren, Jiahong Wang, Peng Zhu, Yunhao Song, Mingxuan Liu, Qian Zheng, Le Bai, Wanli Ouyang, and Chenwei Song. Adabrain-bench: Benchmarking brain foundation models for brain-computer interface applications, 2025. arXiv preprint arXiv:2507.09882.

Xie et al. Cross-dataset montage alignment via electroencephalogram source imaging for enhanced motor decoding, 2025. preprint.

Cheng Yang, M. Brandon Westover, and Jimeng Sun. Biot: Biosignal transformer for cross-data learning in the wild. *Advances in Neural Information Processing Systems*, 36:78240–78260, 2023.