TRACE: LEARNING TO ADAPT SPATIAL–SPECTRAL RELATIONS FOR CROSS-DATASET EEG EMOTION RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-dataset EEG emotion recognition is challenging because spatial–spectral relations that are useful for one subject or recording condition may not transfer to another. We show that even within predictive spatial relation groups, frequency-specific relations can have different effects, and useful adjustments vary across targets. We therefore propose **TRACE** (*Target-conditioned Relation Adaptation for Cross-dataset EEG*), which learns rules for adjusting spatial–spectral relations from labeled source domains and applies these rules to each unlabeled target. TRACE represents channel–band interactions with shared relation types and adapts target-specific adjustments to their relative weights from a domain representation aggregated from unlabeled target trials. It learns these rules from source-domain adjustments validated by lower classification loss on separate source trials and constrains target coefficients to a source-supported region defined by source-domain predictions. TRACE projects the difference between graph representations before and after relation adjustments into a residual that refines the base EEG representation learned directly from the input features for emotion classification. Across eight transfer directions among four public EEG emotion datasets, TRACE achieves the highest accuracy among the compared methods in every direction. Target-substitution experiments further show that adjustments inferred for the current target outperform those transferred from other targets, while TRACE remains robust to electrode masking and feature noise. These results support learning how relations adapt rather than enforcing a fixed structure across domains.

1 INTRODUCTION

Electroencephalography (EEG) emotion recognition depends on interactions across scalp locations and frequency bands (Zheng & Lu, 2015). Graph-based methods increasingly model these relations, from neurophysiological spatial topology (Zhong et al., 2020) and input-adaptive connectivity (Song et al., 2020) to spatial–spectral dependencies (Jia et al., 2021) and personalized cross-frequency coupling (Tian et al., 2025). However, relations that are useful for emotion prediction in one domain need not remain equally useful in another domain because EEG connectivity and spectral features vary across individuals and recording conditions (Li et al., 2025; Wang et al., 2024a). Cross-dataset EEG emotion recognition therefore requires not only learning useful relations, but determining *how those relations should adapt to each unlabeled target*.

Existing approaches address complementary parts of this problem. Relation-based methods learn personalized frequency relationships or align spatio-temporal relationship distributions (Tian et al., 2025; Liu et al., 2024a), while cross-domain methods primarily adapt or learn transferable representations (Wang et al., 2024a; Zhou et al., 2025; Liu et al., 2024b; 2025; Zhang et al., 2026a). However, these approaches do not explicitly learn a transferable rule for *how relations should adapt to a new target*. We investigate this distinction through two questions: ***(1) Should frequency-specific relation types within a useful regional group receive the same adjustment?*** and ***(2) Do useful relation adjustments transfer across targets, or should they be target-specific?***

Controlled diagnostics on SEED→SEED-IV answer both questions using target labels only for these diagnostics (Figure 1). First, selectively attenuating frequency-specific relation types within useful

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

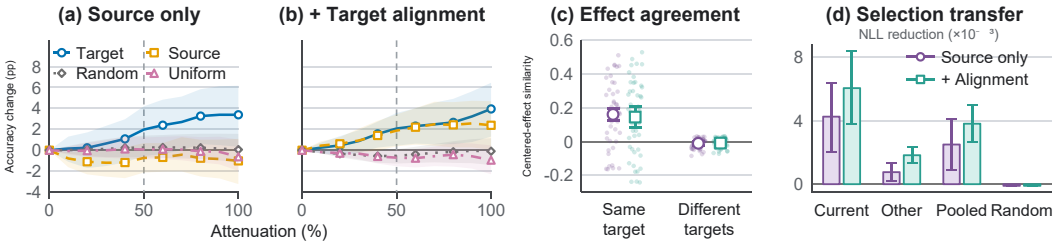

Figure 1: **Relation attenuation on SEED→SEED-IV.** (a–b) Changes in balanced accuracy under progressive attenuation for source-only and target-aligned predictors. Target ranks relations using the current target; Source uses source information with the same target-selected regional groups and selection counts; Random matches the selection count; Uniform distributes the same total attenuation across each group. (c) Cosine similarity of centered relation-effect profiles within the same target and across different targets. (d) Negative-log-likelihood reductions when adjustments are selected using the current target (Current), other targets (Other), their mean profile (Pooled), or random selection (Random). Bands and error bars show 95% bootstrap confidence intervals.

regional groups improves prediction over the unmodified model and matched random or uniform controls (Figure 1(a–b)). Second, relation-effect profiles agree more strongly across disjoint trials from the same target than across targets, and adjustments selected for the current target reduce loss more than those transferred from other targets (Figure 1(c–d)). Together, these results indicate that useful relation adjustments differ across frequency-specific relation types and across targets.

Motivated by these findings, we propose **TRACE** (*Target-conditioned Relation Adaptation for Cross-dataset EEG*), which transfers a rule for *how relations should adapt* rather than a fixed relational structure. TRACE represents channel–band interactions through shared spatial–spectral relation types and learns from labeled source domains how domain context relates to relation adjustments validated by lower classification loss on separate source trials. For an unlabeled target, the rule infers target-specific adjustments from its aggregated context. A source-supported constraint then limits target adjustment to convex hull of source-domain predictions. The induced change in graph representation forms a relational residual that refines EEG representation for emotion prediction.

Across eight transfer directions among SEED, SEED-IV, SEED-V, and MPED, TRACE achieves the highest accuracy and macro-F1, improving accuracy over the strongest baseline in each direction by 2.22–3.94 percentage points, indicating improved generalization. Ablations and target-substitution experiments further support relation adaptation and target conditioning. TRACE also achieves the highest balanced accuracy across all test-time input perturbation settings, demonstrating robustness.

Our contributions are threefold:

- We provide controlled evidence that useful spatial–spectral relation adjustments differ across frequency-specific relation types and across targets.
- We propose **TRACE**, which learns *how relations should adapt* from source adjustments validated on separate source trials. It uses unlabeled target context to infer target-specific adjustments under a source-supported constraint.
- We provide a comprehensive evaluation across eight cross-dataset transfer directions, where TRACE consistently improves over the compared methods; ablations, target substitution, and input perturbations experiments evaluate its key mechanisms and robustness.

2 Related Work

Relation modeling for EEG emotion recognition. EEG relation modeling has progressed from spatial topology and input-adaptive connectivity to spatial–spectral dependencies and personalized frequency relationships (Zhong et al., 2020; Song et al., 2020; Jia et al., 2021; Tian et al., 2025). Recent work (Li et al., 2025; Liu et al., 2024a) further considers variation or alignment of relations across subjects. In particular, VBH-GNN aligns spatial and temporal relationship distributions

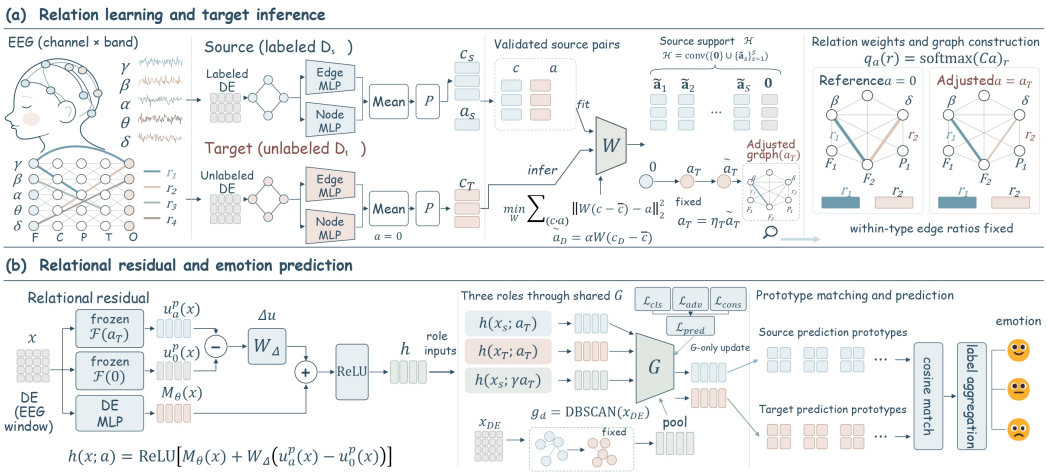

Figure 2: **Overview of TRACE.** (a) TRACE learns from labeled source domains how relation weights should change with domain context and applies the learned rule to unlabeled target context under a source-supported constraint. The inferred target relation coefficients determine the total weight of each relation type. (b) The induced change in graph representation forms a relational residual that refines the EEG representation for target adaptation and emotion prediction.

across subjects (Liu et al., 2024a). These studies make EEG relation modeling increasingly adaptive, but mainly focus on learning, personalizing, or aligning the relations themselves.

Cross-domain EEG emotion recognition. Cross-domain EEG emotion recognition mainly uses domain adaptation and transferable representation learning. Existing methods use predictive balance or source selection (Jiang et al., 2023; Wang et al., 2024a; 2026). Others use graph adaptation or style transfer (Wang et al., 2024a; Zhou et al., 2025). More broadly, hypernetworks and adaptive risk minimization provide mechanisms for conditional parameter generation and learning to adapt under domain shift (Ha et al., 2017; Zhang et al., 2021). TRACE brings this learn-to-adapt perspective to spatial–spectral relations in cross-domain EEG: it learns a transferable rule for how relations should adapt from source adjustments validated on separate trials and applies the rule to each unlabeled target under a source-supported constraint.

3 METHOD

Problem setup. We consider cross-dataset EEG emotion recognition with labeled source domains $\{D_s\}_{s=1}^{S}$ and an unlabeled target domain D_T. Each domain corresponds to one subject recording, with repeated sessions treated as separate domains. All source domains come from one dataset and D_T comes from a different dataset. Each EEG window is represented by five-band EEG features $\mathbf{x} \in \mathbb{R}^{N_c \times N_b}$, where N_c is the number of channels shared across datasets and $N_b = 5$ denotes the δ, θ, α, β, and γ bands. All datasets are mapped to $K = 3$ emotion classes. We use a transductive setting: trials from D_T are available without labels during adaptation, while target labels are used only for evaluation.

Overview. We introduce **TRACE** (*Target-conditioned Relation Adaptation for Cross-dataset EEG*), which learns from source domains *how* spatial–spectral relations should adapt rather than transferring a fixed relational structure. TRACE groups channel–band graph edges into shared relation types and learns from source domains how their weights should change with domain context, using adjustments validated by lower classification loss on separate source trials. It applies the learned rule to unlabeled target context under a source-supported constraint. The induced change in graph representation forms a relational residual that refines the EEG representation for emotion prediction. Sections 3.1–3.4 detail these components and their training.

3.1 Encoding Spatial–Spectral Relations

As EEG patterns related to emotion depend on scalp location and frequency band (Zheng & Lu, 2015), we represent each EEG window as an undirected graph with $N_c N_b$ nodes. Each node carries the feature value of one channel in one frequency band, and every pair of different nodes is connected. A node's type combines its frequency band with one of five scalp regions: frontal, central, parietal, occipital, or temporal. An edge's relation type r is determined by the unordered pair of its endpoint types. The 25 node types give $R = 25 \times 26/2 = 325$ relation types, including edges between different nodes with the same type. We use these shared relation types as the unit of target adjustment. We denote the set of edges of type r by $\mathcal{E}_r$. To construct base edge weights, we use a symmetric channel matrix $\mathbf{A}_{\text{channel}} \in \mathbb{R}^{N_c \times N_c}$ shared across EEG windows. Its diagonal is fixed to one to set the reference scale. Its off-diagonal entries are positive and learned. For an edge $e = ((i, b), (j, b'))$, we define

$$A_e^0 = A_{\text{channel}}[i, j]. \tag{1}$$

The relation type already encodes the endpoint frequency bands. Thus A_e^0 determines relative edge weights within each relation type. We then adjust each relation's total weight using a coefficient vector $\mathbf{a} \in \mathbb{R}^d$. A fixed relation basis matrix $\mathbf{C} \in \mathbb{R}^{R \times d}$ maps these coefficients to the total relation weights $q_{\mathbf{a}}(r)$ through softmax. Each relation's total weight is distributed among its edges by normalizing their base weights, giving the final edge weights $w_{\mathbf{a}}(e)$:

$$\begin{aligned} q_{\mathbf{a}}(r) &= \big[\text{softmax}(\mathbf{C}\mathbf{a})\big]_r, \\ w_{\mathbf{a}}(e) &= q_{\mathbf{a}}(r)\, \frac{A_e^0}{\sum_{e' \in \mathcal{E}_r} A_{e'}^0}, \qquad e \in \mathcal{E}_r. \end{aligned} \tag{2}$$

Updating $\mathbf{a}$ adjusts each relation's total weight while keeping the graph topology and within-type edge ratios fixed. These edge weights control how EEG features are combined into a graph representation. A shared graph encoder uses $w_{\mathbf{a}}(e)$ to aggregate edge features derived from the connected nodes' feature values and node types, and mean-pools the resulting node representations to obtain $\mathbf{u}_{\mathbf{a}}(\mathbf{x})$. A projection network P then gives the projected graph representation $\mathbf{u}_{\mathbf{a}}^{\text{p}}(\mathbf{x}) = P(\mathbf{u}_{\mathbf{a}}(\mathbf{x}))$.

We first train the channel matrix together with the graph encoder and projection network on the labeled source domains $\{D_s\}_{s=1}^S$. We use equal total relation weights with $\mathbf{a} = \mathbf{0}$ and $q_{\mathbf{0}}(r) = 1/R$. This defines the reference relation distribution, while the channel matrix and graph encoder remain learned from the source domains. Thus, $\mathbf{a}$ represents relation adjustments around this reference. After this training, the endpoint node-type embeddings are combined using their sum, absolute difference, and product. The resulting symmetric type information is mapped by the FiLM projection to an eight-dimensional relation vector for each relation type (Perez et al., 2018). Centering and orthonormalizing the stacked vectors gives the fixed relation basis $\mathbf{C} \in \mathbb{R}^{R \times 8}$. We then freeze $\mathbf{C}$ together with the channel matrix and the graph encoder and projection network. Subsequent changes in $\mathbf{a}$ therefore affect the representation only through the edge weights.

3.2 Inferring Relation Coefficients

TRACE uses all unlabeled trials in D_T to infer the target relation coefficients $\mathbf{a}_T$, which determine the total weight of each relation type. We first construct a class prototype $\mathbf{b}_{s,k}^0$ for each source domain D_s and emotion class $k \in \{1, \ldots, K\}$ as a reference for estimating emotion probabilities in both source and target trials. Using the frozen encoder and projection network from Section 3.1, we compute projected graph representations $\mathbf{u}_{\mathbf{a}=\mathbf{0}}^{\text{p}}(\mathbf{x})$. Each prototype is the mean projected graph representation for its class, computed by averaging first over windows within each trial in D_s and then over the trials labeled k. We then compare projected graph representations with these prototypes to estimate emotion probabilities for trials in a source or target domain D. For such a domain D, let $\mathcal{T}_D$ denote its set of trials and $\mathcal{W}_{D,t}$ the set of EEG windows in trial $t \in \mathcal{T}_D$. Let $\mathcal{I}_D \subseteq \{1, \ldots, S\}$ contain the indices of the source domains providing the prototypes. For D_T, we use all source domains. For D_s, we exclude all source domains belonging to the subject recorded in D_s. Using the fixed similarity scale σ learned during initial source training, we compute the class logits

$$z_{D,k}(\mathbf{x}) = \log \left[\frac{1}{|\mathcal{I}_D|} \sum_{s \in \mathcal{I}_D} \exp\left\{ \sigma \cos\left(\mathbf{u}_{\mathbf{a}=\mathbf{0}}^{\text{p}}(\mathbf{x}), \mathbf{b}_{s,k}^0 \right) \right\} \right]. \tag{3}$$

Let $\mathbf{z}_D(\mathbf{x}) = (z_{D,1}(\mathbf{x}), \ldots, z_{D,K}(\mathbf{x}))$ collect these logits. For each trial $t \in \mathcal{T}_D$, we average its window logits and apply softmax across classes to obtain the emotion probability vector

$$\mathbf{p}_{D,t} = \mathrm{softmax}\left(\frac{1}{|\mathcal{W}_{D,t}|}\sum_{\mathbf{x}\in\mathcal{W}_{D,t}} \mathbf{z}_D(\mathbf{x})\right). \tag{4}$$

For each trial $t \in \mathcal{T}_D$, we first compute the trial mean $\bar{\mathbf{u}}_{D,t}$ of the graph representations $\mathbf{u}_{\mathbf{a}=\mathbf{0}}(\mathbf{x})$ over its windows. For class k, we then compute a weighted mean of these trial means using $p_{D,t,k}$ as the weight for trial t. We concatenate the means for all K emotion classes in the same order across domains to form the domain context vector $\mathbf{c}_D$ used to infer relation coefficients:

$$\mathbf{c}_D = \mathrm{concat}_{k=1}^{K}\left(\frac{\sum_{t\in\mathcal{T}_D} p_{D,t,k}\bar{\mathbf{u}}_{D,t}}{\sum_{t\in\mathcal{T}_D} p_{D,t,k}}\right). \tag{5}$$

From $\mathbf{c}_D$, we predict the relation coefficients $\widetilde{\mathbf{a}}_D$ using a learned linear map $\mathbf{W}$. Let $\bar{\mathbf{c}}$ be the mean of the context vectors used to train $\mathbf{W}$. Let α be a scalar learned from source labels. We center $\mathbf{c}_D$ by subtracting $\bar{\mathbf{c}}$ and use α to scale the predicted coefficients:

$$\widetilde{\mathbf{a}}_D = \alpha\mathbf{W}(\mathbf{c}_D - \bar{\mathbf{c}}), \qquad \alpha > 0. \tag{6}$$

This centering maps the mean source context to $\widetilde{\mathbf{a}}_D = \mathbf{0}$ and therefore to the reference relation distribution. Thus, $\mathbf{W}$ predicts changes around the reference rather than one shared adjustment. We denote the predicted coefficients for D_s and D_T by $\widetilde{\mathbf{a}}_s$ and $\widetilde{\mathbf{a}}_T$ respectively.

To avoid extrapolation beyond the source predictions in coefficient space, we constrain the final target relation coefficients $\mathbf{a}_T$ to the convex hull of $\{\widetilde{\mathbf{a}}_s\}_{s=1}^{S}$ and $\mathbf{0}$:

$$\mathcal{H} = \mathrm{conv}\left(\{\mathbf{0}\} \cup \{\widetilde{\mathbf{a}}_s\}_{s=1}^{S}\right).$$

We preserve the inferred adjustment direction and scale $\widetilde{\mathbf{a}}_T$ by the largest factor $\eta_T \in [0, 1]$ that places it in $\mathcal{H}$:

$$\eta_T = \max\left(\lambda \in [0,1] : \lambda\widetilde{\mathbf{a}}_T \in \mathcal{H}\right), \qquad \mathbf{a}_T = \eta_T\widetilde{\mathbf{a}}_T. \tag{7}$$

The allowed magnitude therefore depends on the source predictions along the inferred direction. This leaves $\widetilde{\mathbf{a}}_T$ unchanged inside $\mathcal{H}$ and otherwise reduces its magnitude only as needed. The definition also permits $\eta_T = 0$. In this case $\mathbf{a}_T = \mathbf{0}$ and the reference relation distribution is recovered. We compute $\mathbf{a}_T$ once from all unlabeled trials in D_T and freeze it during target adaptation.

3.3 Emotion Prediction with Relation Adjustment

To incorporate the target relation coefficients $\mathbf{a}_T$, we use the difference between the same window's projected graph representations under $q_{\mathbf{a}}(r)$ and $q_{\mathbf{0}}(r) = 1/R$. Using the frozen graph encoder and projection network P from Section 3.1, we define $\Delta(\mathbf{x}; \mathbf{a}) = \mathbf{u}^{\mathrm{p}}_{\mathbf{a}}(\mathbf{x}) - \mathbf{u}^{\mathrm{p}}_{\mathbf{a}=\mathbf{0}}(\mathbf{x})$. Let M_θ be a trainable MLP that maps the EEG features $\mathbf{x}$ to the EEG representation $M_\theta(\mathbf{x})$. We use a trainable linear map $\mathbf{W}_\Delta$ to project Δ to the same dimension as $M_\theta(\mathbf{x})$. We then add the projected difference to $M_\theta(\mathbf{x})$ and apply ReLU:

$$\mathbf{h}(\mathbf{x}; \mathbf{a}) = \mathrm{ReLU}(M_\theta(\mathbf{x}) + \mathbf{W}_\Delta\Delta(\mathbf{x}; \mathbf{a})). \tag{8}$$

We initialize $\mathbf{W}_\Delta$ to zero without bias. Thus $\mathbf{h}(\mathbf{x}; \mathbf{a})$ initially reduces to $\mathrm{ReLU}(M_\theta(\mathbf{x}))$. The same equality holds when $\mathbf{a} = \mathbf{0}$.

During target adaptation (Section 3.4), we train the prediction modules using labeled source windows as references and unlabeled target windows as queries. For a query window $\mathbf{x}$ and a labeled reference window $\mathbf{x}_s$ from source domain D_s, we apply $\mathbf{h}$ as follows:

$$\mathbf{h}_{\mathrm{Query}}(\mathbf{x}) = \mathbf{h}(\mathbf{x}; \mathbf{a}_T), \qquad \mathbf{h}_{\mathrm{Reference}}(\mathbf{x}_s) = \mathbf{h}(\mathbf{x}_s; \gamma\mathbf{a}_T). \tag{9}$$

Here $0 < \gamma \leq 1$. We select γ using source labels before target inference (Appendix B.2). For emotion prediction, a shared network G_ω maps the query and reference representations into a common space. We construct source and target prototypes in this space. These prediction prototypes are distinct from the class prototypes $\mathbf{b}^0_{s,k}$ used to construct domain context in Section 3.2. Each target window is associated with a target prototype. We match it to labeled source prototypes by cosine similarity. The matched source labels are aggregated across source domains to obtain the final emotion prediction. Details of clustering and prototype construction are provided in Appendix C.

3.4 SOURCE TRAINING AND TARGET ADAPTATION

Source training. During source training, all parameters and relation adjustments described below are learned using labeled source data only. We first learn graph representations for emotion classification using the labeled source domains $\{D_s\}_{s=1}^{S}$ with equal total relation weights ($\mathbf{a} = \mathbf{0}$). Each projected graph representation is classified by cosine similarity to K trainable class vectors using a learnable similarity scale σ. We optimize the model by minimizing cross entropy.

After this training, we construct relation basis matrix $\mathbf{C}$ as described in Section 3.1 and compute the class prototypes $\mathbf{b}_{s,k}^{0}$ as described in Section 3.2. To obtain training targets for linear map $\mathbf{W}$, we form multiple pairs of disjoint trial subsets from each source domain D_s. We denote them by $\mathcal{T}^{\mathrm{adj}}$ and $\mathcal{T}^{\mathrm{val}}$ for adjustment and validation. Both subsets contain trials from all K emotion classes. We adjust only the relation coefficients $\mathbf{a}$ using $\mathcal{T}^{\mathrm{adj}}$ while keeping all remaining components fixed. We compute class logits $\mathbf{z}_{D_s}(\mathbf{x};\mathbf{a})$ from the projected graph representation $\mathbf{u}_{\mathbf{a}}^{\mathrm{p}}(\mathbf{x})$ using the fixed class prototypes from other subjects. For a trial subset $\mathcal{S}$, we evaluate an adjustment using

$$\mathcal{L}_{\mathrm{rel}}(\mathbf{a};\mathcal{S}) = -\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}_{\substack{t,\mathbf{x}|k \\ t\in\mathcal{S}}}\left[\log[\mathrm{softmax}(\mathbf{z}_{D_s}(\mathbf{x};\mathbf{a}))]_k\right]. \tag{10}$$

We average window losses within each trial, then the trial losses within each class, and finally the losses across classes. Starting from $\mathbf{a} = \mathbf{0}$, we use the loss gradient on $\mathcal{T}^{\mathrm{adj}}$ to determine adjustment direction. The directional generalized Gauss–Newton curvature proposes the initial step size. If needed, we reduce the step size along the same direction and discard the split if no decreasing adjustment is found. We retain $\mathbf{a}$ for training $\mathbf{W}$ only when it also reduces the loss on $\mathcal{T}^{\mathrm{val}}$, which is not used to obtain $\mathbf{a}$:

$$\mathcal{L}_{\mathrm{rel}}(\mathbf{a};\mathcal{T}^{\mathrm{val}}) < \mathcal{L}_{\mathrm{rel}}(\mathbf{0};\mathcal{T}^{\mathrm{val}}). \tag{11}$$

For each retained $\mathbf{a}$, we compute the context vector $\mathbf{c}$ from $\mathcal{T}^{\mathrm{adj}}$ using Equation 5. The trial probabilities are obtained from the fixed class prototypes of other subjects as in Section 3.2. The labels in $\mathcal{T}^{\mathrm{adj}}$ and $\mathcal{T}^{\mathrm{val}}$ are used to obtain and validate $\mathbf{a}$ but not to construct $\mathbf{c}$. The context vector $\mathbf{c}$ and the adjusted relation coefficients $\mathbf{a}$ form a training pair. Using all training pairs, we subtract the mean context vector $\bar{\mathbf{c}}$ and fit $\mathbf{W}$ to predict the relation coefficients:

$$\min_{\mathbf{W}}\sum_{(\mathbf{c},\mathbf{a})}\|\mathbf{W}(\mathbf{c}-\bar{\mathbf{c}})-\mathbf{a}\|_2^2. \tag{12}$$

Details of trial selection, coefficient updates, and the computation of $\mathbf{W}$ are provided in Appendix B. After fitting $\mathbf{W}$, we select γ in Equation 9 and α in Equation 6 using labeled source data. Both are fixed before target inference and through target adaptation (Appendix B.2).

Target adaptation. During target adaptation, labeled source windows provide supervision while target windows remain unlabeled. The target relation coefficients $\mathbf{a}_T$ inferred in Section 3.2 remain fixed. We use $\mathbf{h}_{\mathrm{Query}}$ and $\mathbf{h}_{\mathrm{Reference}}$ from Equation 9 for source selection and alternating training. At each iteration, a trainable classifier f_ξ evaluates each source domain using $\mathbf{h}_{\mathrm{Query}}$. We assign the source with the largest classification loss to the query role and the source with the smallest loss to the reference role:

$$s_{\mathrm{Q}} = \arg\max_{s}\mathcal{L}_{\mathrm{cls}}^{(s)}, \qquad s_{\mathrm{R}} = \arg\min_{s}\mathcal{L}_{\mathrm{cls}}^{(s)}. \tag{13}$$

Inspired by the source reliability idea in HEDN (Wang et al., 2026), we use classification loss only to assign the query and reference source roles. The same TRACE prediction modules are used for both roles with the representations in Equation 9.

Prediction uses classification, domain alignment, and consistency losses. We define $\mathcal{L}_{\mathrm{cls}}$ as cross entropy on the selected source queries, while $\mathcal{L}_{\mathrm{adv}}$ aligns these queries with target queries through gradient reversal (Ganin et al., 2016). Consistency loss $\mathcal{L}_{\mathrm{cons}}$ is cross entropy between f_ξ's predictions on target queries and their pseudo labels. Following Kendall et al. (2018), we learn a scalar ρ_{cons} to weight the consistency loss:

$$\mathcal{L}_{\mathrm{pred}} = \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{adv}} + \exp(-\rho_{\mathrm{cons}})\mathcal{L}_{\mathrm{cons}} + \frac{1}{2}\rho_{\mathrm{cons}}. \tag{14}$$

We alternate between a prediction step that optimizes $\mathcal{L}_{\mathrm{pred}}$ with G_ω fixed and a cluster step that updates G_ω through contrastive learning. After training, we predict target emotion labels through

Table 1: Cross-dataset EEG emotion recognition performance (%).

Metric	MS-MDA Chen et al. (2021)	DMMR Wang et al. (2024b)	UDA-DDA Tang et al. (2025)	DSPB Jiang et al. (2023)	GDDN Chen et al. (2024)	E2STN Zhou et al. (2025)	DBDG Li et al. (2024)	GPTDS+ PC-TTA Imtiaz & Khan (2025)	DBPM Wang et al. (2025)	HEDN Wang et al (2026)	PGDG Zhang et al. (2026b)	PAA-M Li et al. (2026)	TRACE Ours
(a) SEED → SEED-IV													
Acc. ↑	46.21 ±11.78	44.64 ±10.24	50.74 ±14.03	49.13 ±12.88	48.02 ±9.52	47.11 ±12.89	46.92 ±11.93	47.51 ±16.17	51.48 ±13.53	51.68 ±9.16	44.80 ±11.84	47.24 ±6.37	**54.82** ±12.76
Macro-F1 ↑	45.79	45.82	48.37	51.26	44.33	47.96	48.53	45.31	51.25	45.32	47.25	47.10	**51.71**
(b) SEED-IV → SEED													
Acc. ↑	53.11 ±9.32	56.20 ±7.38	54.92 ±10.36	55.81 ±6.47	45.36 ±10.20	53.87 ±9.81	51.25 ±8.70	57.20 ±11.49	53.90 ±13.87	54.01 ±13.18	55.10 ±8.41	58.67 ±8.83	**60.89** ±13.17
Macro-F1 ↑	51.77	53.91	52.84	57.91	43.06	50.98	48.47	54.66	53.64	59.23	52.68	56.72	**60.16**
(c) SEED → SEED-V													
Acc. ↑	52.67 ±6.03	53.80 ±5.21	54.07 ±2.98	54.88 ±9.91	55.61 ±7.17	48.22 ±4.03	51.72 ±7.47	52.29 ±5.11	50.92 ±6.61	53.19 ±7.47	54.63 ±5.90	55.30 ±20.48	**59.55** ±11.21
Macro-F1 ↑	50.30	53.95	52.38	54.44	55.07	40.55	47.07	50.10	54.17	47.29	56.87	46.35	**66.73**
(d) SEED-V → SEED													
Acc. ↑	56.04 ±10.33	55.71 ±6.03	52.55 ±9.54	52.55 ±12.89	47.76 ±3.78	42.63 ±9.24	51.36 ±11.57	50.49 ±11.09	57.11 ±16.93	56.06 ±11.28	47.59 ±7.84	54.73 ±6.02	**60.83** ±9.80
Macro-F1 ↑	55.81	54.66	50.04	50.38	46.76	34.95	47.31	47.76	55.71	54.25	42.90	50.03	**59.75**
(e) SEED-IV → SEED-V													
Acc. ↑	58.23 ±8.36	51.20 ±8.28	60.70 ±9.36	58.45 ±13.39	61.39 ±10.04	56.67 ±11.56	58.47 ±8.98	49.60 ±7.29	55.38 ±10.42	60.50 ±6.25	54.52 ±13.82	49.77 ±3.14	**64.26** ±13.07
Macro-F1 ↑	45.76	45.57	53.37	52.37	50.65	46.02	43.09	40.92	49.14	55.19	56.54	51.60	**59.40**
(f) SEED-V → SEED-IV													
Acc. ↑	57.15 ±7.67	53.04 ±5.53	60.81 ±4.19	53.05 ±11.84	51.04 ±10.46	56.70 ±7.99	56.44 ±9.06	55.68 ±5.24	54.44 ±8.83	57.25 ±5.42	53.06 ±8.05	51.08 ±3.60	**63.32** ±13.19
Macro-F1 ↑	43.47	49.52	54.76	46.70	42.66	42.14	46.33	46.19	47.88	46.22	40.91	49.58	**57.19**
(g) SEED^3 → MPED^3													
Acc. ↑	38.62 ±5.08	40.91 ±2.88	40.30 ±4.14	41.13 ±4.34	37.81 ±3.66	40.78 ±1.91	39.12 ±1.76	38.57 ±3.99	41.29 ±3.70	40.42 ±2.09	40.56 ±3.43	35.37 ±2.30	**44.76** ±3.98
Macro-F1 ↑	36.57	39.26	38.28	39.36	37.19	37.31	35.92	39.04	39.25	37.44	36.12	31.33	**43.52**
(h) MPED^3 → SEED^3													
Acc. ↑	45.29 ±11.61	47.60 ±4.15	48.91 ±12.66	45.34 ±8.56	49.05 ±9.12	48.79 ±10.60	47.38 ±13.80	49.18 ±10.50	47.93 ±8.34	49.61 ±7.04	41.47 ±6.82	45.50 ±3.86	**52.92** ±8.25
Macro-F1 ↑	40.37	42.06	42.25	45.44	45.27	47.71	46.91	49.41	42.79	46.69	40.06	41.18	**49.62**

Acc. is shown as mean ± SD. Superscript 3 denotes the three-class SEED–MPED setting of E2STN (Zhou et al., 2025), using SEED DE and MPED STFT features. The best available value in each row is shown in bold.

prototype matching with all source domains as described in Section 3.3. Details of the cluster objective, prototype updates and optimization schedule are provided in Appendix C.

4 Experiments and Results

4.1 Experimental Setup

Datasets and protocol. We evaluate eight transfer directions across SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2018), SEED-V (Liu et al., 2021), and MPED (Song et al., 2019). Following prior public implementations (Wang et al., 2026; 2025), all transfer directions use the same training and evaluation protocol. Following E2STN (Zhou et al., 2025), we also use its SEED–MPED setting with three emotion classes. SEED, SEED-IV, and SEED-V use released five-band DE features. MPED uses released five-band STFT features. All inputs use 62 channels and five frequency bands. We apply min–max normalization within each domain. All labeled source domains are used for training. TRACE adapts separately to each target domain using its unlabeled windows.

Baselines and evaluation. We compare TRACE with 12 baseline methods in Table 1. All compared methods use the same label mapping, feature inputs, normalization, and target domains. We use a fixed training budget of 1,000 iterations and report accuracy and macro F1 from the final iteration. For each seed, scores are computed over all windows in each recording and averaged equally across recordings. We also report the population standard deviation of accuracy across recordings (Wang et al., 2025). All reported means and SDs are averaged over three random seeds.

Table 2: Ablation study of TRACE (%).

	SEED-IV → SEED-V		SEED-V → SEED-IV		SEED → SEED-IV		SEED-IV → SEED	
Variant	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Full TRACE	**64.26**	**59.40**	**63.32**	**57.19**	**54.82**	**51.71**	**60.89**	**60.16**
w/o Relation Adaptation	59.03	53.29	57.66	50.79	52.09	45.83	56.14	50.43
w/o Target Conditioning	53.67	50.33	56.27	53.81	50.10	46.25	54.59	53.82
w/o Convex-Hull Constraint	63.69	56.21	60.89	55.14	51.18	45.50	53.66	51.60
w/o Relational Residual	60.90	55.83	59.37	54.08	51.92	45.26	53.35	52.79
w/o Graph Relation Adjustment	55.06	52.49	57.92	51.71	50.60	44.72	53.27	52.93

All variants use the same three random seeds as Table 1. We report means across target recordings and random seeds. Bold marks the best mean for each metric and transfer direction.

Table 3: Target specificity of inferred relation adjustments.

Direction	Metric	Own (TRACE)	Other-1	Other-2	Other-3	Other-4	Other-5	Other Avg.	Shared	Zero
SEED-IV →	Acc. ↑	**63.57**	56.53	55.09	53.47	54.56	51.50	54.23	52.96	56.81
SEED-V	F1 ↑	**60.86**	52.52	51.12	50.21	53.85	49.00	51.34	47.11	50.19
SEED-V →	Acc. ↑	**64.01**	56.77	56.05	54.63	53.69	56.81	55.59	55.98	57.81
SEED-IV	F1 ↑	**58.04**	53.34	52.98	51.47	50.40	53.95	52.43	51.06	52.15

This controlled substitution experiment uses an independent set of three random seeds. All conditions within the table use the same seeds. Each cell reports accuracy and macro-F1 (%) from top to bottom.

Training and testing. We fit and calibrate TRACE's relation model on subject-disjoint source folds, excluding each held-out subject throughout the source-side procedure. We then repeat the same source-only procedure using labeled recordings from all source subjects. Source training and target adaptation use fixed training budgets. We evaluate the final adaptation state. Detailed settings are provided in Appendix D. All code and model weights will be public upon publication.

4.2 Cross-Dataset Emotion Recognition

Table 1 shows that TRACE achieves the highest reported accuracy and macro F1 at the final iteration across all eight transfer directions. Across all eight directions, TRACE improves accuracy over the strongest baseline by 2.22–3.94 percentage points. In particular, on SEED^3 →MPED^3 and MPED^3 →SEED^3, TRACE improves accuracy over the strongest baseline by 3.47 and 3.31 percentage points, respectively. Together, these consistent improvements demonstrate TRACE's strong performance and generalization across cross-dataset EEG emotion recognition settings.

4.3 Ablation Study

We conduct ablation and control experiments on all target recordings in SEED↔SEED-IV and SEED-IV↔SEED-V. **w/o Relation Adaptation** sets the relation coefficient vector to zero and retains the reference relation distribution. **w/o Target Conditioning** uses a single direction-specific coefficient vector optimized on labeled source data and shared across all target recordings. **w/o Hull Constraint** uses the raw target-conditioned proposal without radially contracting it to the convex hull of the source predictions. **w/o Relational Residual** replaces $\Delta(\mathbf{x}; \mathbf{a}_T)$ with $\mathbf{u}_0^{\mathrm{p}}(\mathbf{x})$, retaining graph information without relation adjustment. **w/o Graph Relation Adjustment** uses the same inferred coefficient vector but maps it directly to a feature residual without graph relation adjustment. All variants use the same prediction and target adaptation procedure with the same training budget. Table 2 shows that each of the three component ablations reduces both accuracy and macro F1 in every transfer direction. Accuracy drops by 0.57–10.59 percentage points relative to full TRACE. The largest accuracy degradation occurs without target conditioning on SEED-IV→SEED-V, supporting target-specific relation adaptation. Both additional controls reduce accuracy in all four directions, with a larger accuracy drop after removing graph relation adjustment.

4.4 Target Specificity of Inferred Relation Adjustments

Our label-based diagnostics show that useful relation adjustments vary across targets (Figure 1(c–d)). We test whether our source-learned rule can use unlabeled target context to infer adjustments

that benefit the corresponding target. On SEED-IV→SEED-V and SEED-V→SEED-IV, we evaluate all 48 and 45 target recordings, respectively. We compare coefficients inferred from the target's own context (**Own**) with five fixed coefficient permutations within each session (**Other-1–5**). Each permutation assigns to a target the coefficients inferred from another target recording in the same session. We also compare a source-optimized shared vector (**Shared**) and zero adjustment (**Zero**). The relation encoder and basis remain frozen, and all conditions use the same downstream adaptation procedure. Table 3 shows that Own achieves the highest mean accuracy and macro-F1 in both directions. Its macro-F1 gains over the mean of the five permutations are 9.52 and 5.62 percentage points on SEED-IV→SEED-V and SEED-V→SEED-IV, respectively, while those over the best fixed permutation are 7.01 and 4.09 points.

Table 4: Trial-balanced BAcc (%) under test-time perturbations on SEED → SEED-IV. Clean denotes zero perturbation. Small annotations show changes from each method's clean performance. ↓ and ↑ denote decreases and increases.

Perturbation	Level	**TRACE**	UDA-DDA	DBPM	HEDN	PGDG	DBDG	GDDN	DSPB	E2STN	GPTDS+ PC-TTA	PAA-M
Clean	–	54.81	48.08	49.10	51.89	48.80	48.48	45.66	51.76	48.64	47.90	38.00
Masked electrodes m	3	54.38 ↓ 0.43	47.58 ↓ 0.50	50.09 ↑ 0.99	51.74 ↓ 0.15	49.35 ↑ 0.55	47.80 ↓ 0.68	45.73 ↑ 0.07	51.35 ↓ 0.41	47.46 ↓ 1.18	46.10 ↓ 1.80	37.52 ↓ 0.48
	6	54.47 ↓ 0.34	47.01 ↓ 1.07	49.08 ↓ 0.02	50.42 ↓ 1.47	49.46 ↑ 0.66	46.85 ↓ 1.63	44.36 ↓ 1.30	50.22 ↓ 1.54	48.40 ↓ 0.24	46.17 ↓ 1.73	37.44 ↓ 0.56
	12	54.26 ↓ 0.55	45.42 ↓ 2.66	47.87 ↓ 1.23	48.47 ↓ 3.42	47.70 ↓ 1.10	45.22 ↓ 3.26	43.10 ↓ 2.56	49.81 ↓ 1.95	47.87 ↓ 0.77	45.09 ↓ 2.81	37.08 ↓ 0.92
	24	53.43 ↓ 1.38	43.56 ↓ 4.52	47.63 ↓ 1.47	45.26 ↓ 6.63	43.55 ↓ 5.25	42.58 ↓ 5.90	41.92 ↓ 3.74	45.62 ↓ 6.14	44.99 ↓ 3.65	43.80 ↓ 4.10	35.51 ↓ 2.49
Gaussian noise σ	0.1	54.58 ↓ 0.23	47.95 ↓ 0.13	49.13 ↑ 0.03	51.95 ↑ 0.06	49.29 ↑ 0.49	48.47 ↓ 0.01	45.88 ↑ 0.22	51.70 ↓ 0.06	48.04 ↓ 0.60	47.50 ↓ 0.40	37.81 ↓ 0.19
	0.2	54.15 ↓ 0.66	47.27 ↓ 0.81	49.20 ↑ 0.10	51.82 ↓ 0.07	48.39 ↓ 0.41	46.23 ↓ 2.25	45.15 ↓ 0.51	51.43 ↓ 0.33	48.02 ↓ 0.62	47.59 ↓ 0.31	37.72 ↓ 0.28
	0.4	53.30 ↓ 1.51	46.78 ↓ 1.30	48.76 ↓ 0.34	52.14 ↑ 0.25	46.69 ↓ 2.11	45.68 ↓ 2.80	44.58 ↓ 1.08	50.81 ↓ 0.95	47.42 ↓ 1.22	46.26 ↓ 1.64	37.90 ↓ 0.10
	0.8	52.58 ↓ 2.23	42.77 ↓ 5.31	47.50 ↓ 1.60	49.61 ↓ 2.28	43.36 ↓ 5.44	42.40 ↓ 6.08	42.66 ↓ 3.00	47.96 ↓ 3.80	44.25 ↓ 4.39	42.34 ↓ 5.56	37.86 ↓ 0.14

4.5 Robustness to Test-Time Input Perturbations

On SEED→SEED-IV, we test robustness to channel loss and feature noise using all 45 subject–session target recordings (15 participants; three sessions each). Each method is first adapted to the clean target recording and then frozen; TRACE's inferred relation coefficients and cluster statistics are also frozen. During prediction, electrode masking replaces all five band features of selected electrodes with their clean-target medians. Results are averaged over three random electrode orderings. Each ordering is shared across the four masking levels. For Gaussian noise, results are averaged over 10 independent perturbations per recording and noise level. Noise is scaled by each feature's clean-target standard deviation. All methods receive the same perturbations. We compute trial-balanced BAcc by averaging window-level correctness within each trial and then averaging trials within each class. We then average classes within each recording and weight all 45 recordings equally. Figure 3 and Table 4 show that TRACE achieves the highest BAcc at all eight perturbation levels. At the strongest masking and noise levels, it retains 53.43% and 52.58% BAcc, decreasing by only 1.38 and 2.23 percentage points, respectively. Although DBPM decreases less at $\sigma = 0.8$ (1.60 points), TRACE maintains a higher BAcc.

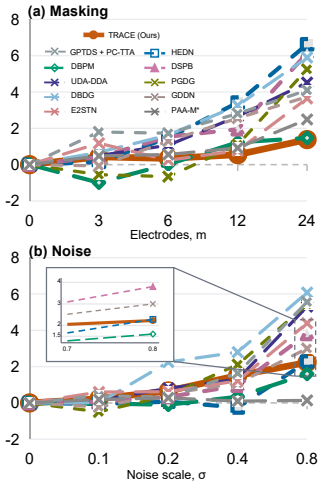

Figure 3: Robustness under input perturbations.

5 Conclusion

We introduced **TRACE** for cross-dataset EEG emotion recognition, which learns from source domains *how* shared spatial–spectral relations should adapt to each unlabeled target rather than transferring a fixed relational structure. Across eight transfer directions among four public EEG emotion datasets, TRACE consistently outperforms the compared methods, while ablations and target-substitution experiments support target-specific relation adaptation and its source-supported constraint. TRACE also remains robust to electrode masking and feature noise. More broadly, our results suggest that when relational structure varies across domains, transferring how relations should adapt can be more effective than enforcing a single transferable structure. Future work will consider heterogeneous electrode configurations and streaming targets.

AI USE STATEMENT

After the manuscript was independently drafted by authors, generative AI tools were used only for language editing and writing suggestions, including grammar correction, academic word choice, and improvements to sentence fluency and readability. Generative AI was not used for research ideation, methodology or experimental design, data analysis, or interpretation of results. All AI-assisted edits were reviewed by the authors, who take full responsibility for final content of the paper.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the model architecture, source training and target adaptation procedures, data processing, evaluation protocol, and implementation settings in Sections 3–4 and Appendices A–F. The appendices include the training hyperparameters, environment, and additional experimental details required to reproduce the reported results. We will release the full source code and trained model weights upon publication.

REFERENCES

Bianna Chen, CL Philip Chen, and Tong Zhang. Gddn: Graph domain disentanglement network for generalizable eeg emotion recognition. *IEEE Transactions on Affective Computing*, 15(3): 1739–1753, 2024.

Hao Chen, Ming Jin, Zhunan Li, Cunhang Fan, Jinpeng Li, and Huiguang He. Ms-mda: Multi-source marginal distribution adaptation for cross-subject and cross-session eeg emotion recognition. *Frontiers in Neuroscience*, 15:778488, 2021.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pp. 226–231, 1996.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

David Ha, Andrew Dai, and Quoc V. Le. Hypernetworks. 2017. URL `https://openreview.net/pdf?id=rkpACe1lx`.

Md Niaz Imtiaz and Naimul Khan. Enhanced cross-dataset electroencephalogram-based emotion recognition using unsupervised domain adaptation. *Computers in Biology and Medicine*, 184: 109394, 2025.

Ziyu Jia, Youfang Lin, Jing Wang, Zhiyang Feng, Xiangheng Xie, and Caijie Chen. Hetemotionnet: two-stream heterogeneous graph recurrent neural network for multi-modal emotion recognition. In *Proceedings of the 29th ACM international conference on multimedia*, pp. 1047–1056, 2021.

Haiting Jiang, Fangyao Shen, Lina Chen, Yong Peng, Hongjie Guo, and Hong Gao. Joint domain symmetry and predictive balance for cross-dataset eeg emotion recognition. *Journal of Neuroscience Methods*, 400:109978, 2023.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7482–7491, 2018.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Guangli Li, Canbiao Wu, Na Tian, Li Zhang, and Zhen Liang. Boundary-aware prototype-driven adversarial alignment for cross-corpus eeg emotion recognition. *arXiv preprint arXiv:2603.26713*, 2026.

Guangqiang Li, Ning Chen, Hongqing Zhu, Jing Li, Zhangyong Xu, and Zhiying Zhu. Uncertainty-aware graph contrastive fusion network for multimodal physiological signal emotion recognition. *Neural Networks*, 187:107363, 2025.

Wei Li, Siyi Wang, Shitong Shao, and Kaizhu Huang. Distillation-based domain generalization for cross-dataset eeg-based emotion recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(3):2474–2490, 2024.

Chenyu Liu, Xinliang Zhou, Zhengri Zhu, Liming Zhai, Ziyu Jia, and Yang Liu. Vbh-gnn: Variational bayesian heterogeneous graph neural networks for cross-subject emotion recognition. In *International Conference on Learning Representations*, volume 2024, pp. 52277–52292, 2024a.

Qile Liu, Zhihao Zhou, Jiyuan Wang, and Zhen Liang. Joint contrastive learning with feature alignment for cross-corpus eeg-based emotion recognition. In *Proceedings of the 1st International Workshop on Brain-Computer Interfaces (BCI) for Multimedia Understanding*, pp. 9–17, 2024b.

Qile Liu, Weishan Ye, Lingli Zhang, and Zhen Liang. Eeg-scmm: Soft contrastive masked modeling for cross-corpus eeg-based emotion recognition. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 5834–5842, 2025.

Wei Liu, Jie-Lin Qiu, Wei-Long Zheng, and Bao-Liang Lu. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Transactions on Cognitive and Developmental Systems*, (2):715–729, 2021.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Tengfei Song, Wenming Zheng, Cheng Lu, Yuan Zong, Xilei Zhang, and Zhen Cui. Mped: A multi-modal physiological emotion database for discrete emotion recognition. *IEEE Access*, 7: 12177–12191, 2019.

Tengfei Song, Suyuan Liu, Wenming Zheng, Yuan Zong, and Zhen Cui. Instance-adaptive graph for eeg emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 2701–2708, 2020.

Jiahao Tang, Youjun Li, Chun-Wang Su, Xiangting Fan, Yangxuan Zheng, Haoyu Wang, Hadia Naeem, Peng Fang, Jue Wang, Nan Yao, et al. Uda-dda: Unsupervised domain adaptation with dynamic distribution alignment network for emotion recognition using eeg signals. *Neurocomputing*, pp. 131715, 2025.

Dongyuan Tian, Yucheng Wang, Peiliang Gong, Zhewen Xu, Zhenghua Chen, Xiaohui Wei, and Min Wu. Accnet: Adaptive cross-frequency coupling graph attention for eeg emotion recognition. *Neural Networks*, 191:107853, 2025.

Jing Wang, Xiaojun Ning, Wei Xu, Yunze Li, Ziyu Jia, and Youfang Lin. Multi-source selective graph domain adaptation network for cross-subject eeg emotion recognition. *Neural Networks*, 180:106742, 2024a.

Qiang Wang, Liying Yang, Qian Zhang, Jingtao Du, and Yumeng Ye. Fine-grained label propagation via density-based prototype matching for cross-subject eeg emotion recognition. *Knowledge-Based Systems*, pp. 114650, 2025.

Qiang Wang, Liying Yang, Jiayun Song, Yifan Bai, and Jingtao Du. Hedn: A hard-easy dual network with source reliability assessment for cross-subject eeg emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, 2026.

Yiming Wang, Bin Zhang, and Yujiao Tang. Dmmr: Cross-subject domain generalization for eeg-based emotion recognition via denoising mixed mutual reconstruction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 628–636, 2024b.

Marvin Zhang, Henrik Marklund, Nikita Dhawan, Abhishek Gupta, Sergey Levine, and Chelsea Finn. Adaptive risk minimization: Learning to adapt to domain shift. *Advances in neural information processing systems*, 34:23664–23678, 2021.

Qingzhu Zhang, Jiani Zhong, Zongsheng Li, Xinke Shen, and Quanying Liu. Multi-dataset joint pre-training of emotional eeg enables generalizable affective computing. *Advances in Neural Information Processing Systems*, 38:165434–165461, 2026a.

Xuan Zhang, Wang Zheng, Hongxin Cai, Zhigang Li, Yi Yang, Weijia Liu, Junru Zhu, Jingyu Liu, Bin Hu, and Qunxi Dong. Prompt-guided domain generalization for eeg emotion recognition. *IEEE Transactions on Affective Computing*, 2026b.

Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, 49 (3):1110–1122, 2018.

Peixiang Zhong, Di Wang, and Chunyan Miao. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 13(3):1290–1301, 2020.

Yijin Zhou, Fu Li, Yang Li, Youshuo Ji, Lijian Zhang, Yuanfang Chen, and Huaning Wang. Enhancing cross-dataset eeg emotion recognition: A novel approach with emotional eeg style transfer network. *IEEE Transactions on Affective Computing*, 2025.

APPENDIX CONTENTS

A Additional Details of Spatial–Spectral Relation Encoding 14
A.1 Graph Construction and Base Edge Weights 14
A.2 Graph Encoder, Projection, and Relation Basis 15
B Source Learning and Target Coefficient Inference 15
B.1 Initial Training and Source Coefficient Learning 15
B.2 Fitting $\mathbf{W}$ and Source-Only Calibration 16
B.3 Unlabeled Target Inference 19
C Target Adaptation and Emotion Prediction 19
C.1 TRACE-Conditioned Adaptation 19
C.2 Fixed Clusters and Label Aggregation 20
D Reproducibility and Computational Details 21
D.1 Datasets, Label Mapping, and Evaluation Protocol 22
D.2 TRACE Implementation Settings 22
D.3 Baseline Implementation and Evaluation Settings 22
D.4 Computational Complexity 22
E Experimental Protocols 23
E.1 Motivating Relation Diagnostics 23
E.2 Ablation, Target Substitution, and Perturbation Protocols 25
F Additional Analyses 26
F.1 Visualization of Target-Specific Relation Adjustments 26

Table 5: Scalp-region assignment of the 62 common EEG channels used in our experiments.

Region	EEG channels
Frontal	FP1, FPZ, FP2, AF3, AF4, F7, F5, F3, F1, FZ, F2, F4, F6, F8
Central	FC5, FC3, FC1, FCZ, FC2, FC4, FC6, C5, C3, C1, CZ, C2, C4, C6, CP5, CP3, CP1, CPZ, CP2, CP4, CP6
Parietal	P7, P5, P3, P1, PZ, P2, P4, P6, P8
Occipital	PO7, PO5, PO3, POZ, PO4, PO6, PO8, CB1, O1, OZ, O2, CB2
Temporal	FT7, FT8, T7, T8, TP7, TP8

A Additional Details of Spatial–Spectral Relation Encoding

This section provides additional implementation details for the spatial–spectral relation encoder introduced in Section 3.1. We follow the notation of the main text and expand only the operations omitted there for brevity.

A.1 Graph Construction and Base Edge Weights

For a given source–target dataset pair, N_c denotes the number of EEG channels shared across domains, and $N_b = 5$ corresponds to the δ, θ, α, β, and γ bands. The graph therefore contains $N_c N_b$ nodes over the channels shared across the source and target datasets. Each shared EEG channel is assigned to one of five scalp regions: frontal, central, parietal, occipital, or temporal. Combining the five regions with the five frequency bands gives 25 node types. As in Section 3.1, the relation type of an undirected edge is determined by the unordered pair of its endpoint node types. The number of relation types is therefore

$$R = \frac{25 \times 26}{2} = 325,$$

independently of the number of channels assigned to each region.

In our experiments, SEED, SEED-IV, SEED-V, and MPED use the same 62 EEG channels. Thus $N_c = 62$ and each EEG window contains $62 \times 5 = 310$ channel–band nodes. Table 5 gives the scalp-region assignment used for these channels.

For implementation, nodes are stored in channel-major order, with the five frequency bands of one channel kept together. This ordering affects only indexing and not the relation definition. Self-edges are excluded, whereas an edge between two different nodes with the same node type is retained.

The channel matrix $\mathbf{A}_{\text{channel}}$ is learned during initial source training. Its off-diagonal entries are constructed from unconstrained trainable parameters by symmetrizing the parameters and applying a positive transformation. Its diagonal is fixed to one. For $e \in \mathcal{E}_r$, the normalized factor

$$\frac{A_e^0}{\sum_{e' \in \mathcal{E}_r} A_{e'}^0}$$

determines the share of the total weight of relation type r assigned to that concrete edge. Thus, A_e^0 controls relative edge weights within one relation type, whereas $q_{\mathbf{a}}(r)$ controls the total weight assigned to the relation type.

TRACE uses these shared relation types as the unit of target adjustment. The target coefficients change the total weight of each relation type. The graph topology and edge ratios within each relation type remain fixed. Concrete edges are therefore grouped by scalp region and frequency band. TRACE does not assign a separate target adjustment to every edge.

At $\mathbf{a} = \mathbf{0}$, $q_{\mathbf{0}}(r) = 1/R$ assigns the same total weight to every relation type. Concrete edge weights can still differ because each relation total is distributed using A_e^0 within $\mathcal{E}_r$. The channel matrix and graph encoder remain learned from the source domains. Thus, $q_{\mathbf{0}}$ defines the reference relation distribution rather than the complete graph structure.

A.2 Graph Encoder, Projection, and Relation Basis

For an edge connecting nodes with feature values x_i and x_j, the shared edge network uses the symmetric signal statistics

$$x_i + x_j, \qquad |x_i - x_j|, \qquad x_i x_j, \qquad \frac{x_i^2 + x_j^2}{2}.$$

The endpoint node types are represented by learned eight-dimensional type embeddings. For endpoint types p and q, let $\mathbf{e}_p, \mathbf{e}_q \in \mathbb{R}^8$ denote their embeddings. Their symmetric type information is

$$\operatorname{concat}\left(\mathbf{e}_p + \mathbf{e}_q,\ |\mathbf{e}_p - \mathbf{e}_q|,\ \mathbf{e}_p \odot \mathbf{e}_q\right).$$

This type information conditions the hidden edge representation through FiLM (Perez et al., 2018). The rank 8 FiLM projection also gives the relation vector used to construct $\mathbf{C}$. Therefore, exchanging the two endpoint types does not change the symmetric type information.

The edge network uses a 64-dimensional hidden representation and produces a 64-dimensional edge feature, with FiLM rank 8. The same edge network is shared by all concrete edges. Each edge feature is multiplied by $w_{\mathbf{a}}(e)$ before being accumulated at both endpoint nodes. For each node, the weighted incident edge features are aggregated and normalized by their total incoming weight. The shared node network combines the normalized aggregate with the incoming weight and node-type information, using a 64-dimensional hidden layer and producing a 64-dimensional node representation. Averaging over all nodes gives the 64-dimensional graph representation $\mathbf{u}_{\mathbf{a}}(\mathbf{x})$.

The projection network P contains two fully connected layers, $64 \to 64 \to 64$, with ReLU after each layer. Its dropout probability is zero in all reported experiments. The output is the projected graph representation $\mathbf{u}^{\mathrm{p}}_{\mathbf{a}}(\mathbf{x})$ used throughout the remaining stages of TRACE.

After initial source training, the relation vector above is computed for each of the 325 relation types. We use $d = 8$. The relation vectors are stacked into $\mathbf{V} \in \mathbb{R}^{R\times 8}$ and centered column-wise. Reduced QR decomposition of the centered matrix gives the fixed relation basis $\mathbf{C} \in \mathbb{R}^{R\times 8}$. A fixed sign convention removes the sign ambiguity of the QR decomposition. Subsequent relation adjustments are restricted to this eight-dimensional span. The channel matrix, graph encoder, P, and $\mathbf{C}$ are then frozen. Later changes in $\mathbf{a}$ affect the graph representation only through the relation and edge weights.

B Source Learning and Target Coefficient Inference

This section expands the source learning and target coefficient inference procedures summarized in Sections 3.2 and 3.4. The source adjustments used to fit $\mathbf{W}$, the linear map $\mathbf{W}$, and the calibration quantities are determined using labeled source data only. Target coefficient inference then uses unlabeled target trials. Target labels are not used in either procedure.

We first fit and calibrate TRACE's relation model on subject-disjoint source folds. For each held-out subject, all of its recordings are excluded from initial source training and all subsequent source-side steps, including construction of $\mathbf{C}$ and source prototypes, source coefficient learning, fitting $\mathbf{W}$, and calibration of α and γ. We then repeat the same source-only procedure using labeled recordings from all source subjects. The resulting model is used for target inference.

The remaining procedure is organized into four stages. Algorithm 1 constructs validated source adjustment pairs. Algorithm 2 calibrates the source-only scales after fitting $\mathbf{W}$. Algorithm 3 infers one fixed coefficient vector for an unlabeled target. Algorithm 4 then performs target adaptation and final emotion prediction.

B.1 Initial Training and Source Coefficient Learning

Initial source training uses equal relation weights, $\mathbf{a} = \mathbf{0}$. The channel matrix, graph encoder, projection network, trainable class vectors, and similarity scale σ are optimized using the cosine classifier described in Section 3.4. This stage uses RMSprop with learning rate 10^{-3}, weight decay 10^{-5}, batch size 96, and a fixed budget of 1,000 optimization steps.

After initial source training, $\mathbf{C}$ is constructed and the graph-related components are frozen. For each source domain D_s and emotion class k, the prototype $\mathbf{b}_{s,k}^0$ is computed by first averaging $\mathbf{u}_{\mathbf{a}=\mathbf{0}}^{\mathrm{p}}(\mathbf{x})$ over the windows of each physical trial and then averaging the resulting trial representations over trials labeled as class k. When prototypes are used to construct a context for a source domain, all source domains belonging to the same physical subject are excluded.

We construct supervision for $\mathbf{W}$ using multiple pairs of disjoint trial subsets from each source domain D_s. We denote the two subsets by $\mathcal{T}^{\mathrm{adj}}$ and $\mathcal{T}^{\mathrm{val}}$. Both subsets contain trials from all K emotion classes. All model components remain fixed and only $\mathbf{a}$ is changed using $\mathcal{T}^{\mathrm{adj}}$.

Let

$$\mathbf{g} = \nabla_{\mathbf{a}} \mathcal{L}_{\mathrm{rel}}(\mathbf{a}; \mathcal{T}^{\mathrm{adj}})\big|_{\mathbf{a}=\mathbf{0}}.$$

If $\mathbf{g}$ is non-finite or $\|\mathbf{g}\|_2 = 0$ in the working precision, the split is discarded. Otherwise, we define

$$\mathbf{d} = -\frac{\mathbf{g}}{\|\mathbf{g}\|_2}.$$

Let κ_{GGN} denote the directional generalized Gauss–Newton curvature along $\mathbf{d}$. If κ_{GGN} is non-finite or non-positive, the split is discarded. Otherwise, the first candidate is

$$\tau_1 = \frac{\|\mathbf{g}\|_2}{\kappa_{\mathrm{GGN}}}, \qquad \mathbf{a}^{(1)} = \tau_1 \mathbf{d}.$$

The gradient and curvature only propose the finite step. We evaluate each candidate using the exact loss on $\mathcal{T}^{\mathrm{adj}}$. If the loss does not decrease, we halve the step size along the same direction and evaluate the next candidate. The search stops at the first strict decrease or when the step becomes zero in the working precision. In the latter case, the split is discarded. No additional numerical tolerance is used for the decrease test.

The accepted coefficient vector is then evaluated on $\mathcal{T}^{\mathrm{val}}$. We retain it for fitting $\mathbf{W}$ only when it also satisfies Equation 11.

For each retained $\mathbf{a}$, we construct the context vector $\mathbf{c}$ from $\mathcal{T}^{\mathrm{adj}}$ using Equation 5. The trial probabilities are obtained from the fixed class prototypes of other subjects as described in Section 3.2. The labels in $\mathcal{T}^{\mathrm{adj}}$ and $\mathcal{T}^{\mathrm{val}}$ are used to obtain and validate $\mathbf{a}$. They are not used to construct $\mathbf{c}$. The pair $(\mathbf{c}, \mathbf{a})$ is retained for fitting $\mathbf{W}$.

B.2 Fitting $\mathbf{W}$ and Source-Only Calibration

Let $\bar{\mathbf{c}}$ denote the mean of the retained source context vectors. If $\mathbf{X}$ stacks the centered contexts $(\mathbf{c} - \bar{\mathbf{c}})^\top$ row-wise and $\mathbf{A}$ stacks the corresponding accepted coefficient vectors, then Equation 12 is solved by

$$\mathbf{W} = \left(\mathbf{X}^\dagger \mathbf{A}\right)^\top,$$

where $\mathbf{X}^\dagger$ is the Moore–Penrose pseudoinverse. We compute it with singular value decomposition, with numerical rank determined by machine precision and the realized singular values. No manually selected regularization coefficient or target data is used.

After fitting $\mathbf{W}$, we calibrate γ and then α using labeled source data only. The scalar γ controls the reference role in Equation 9. For each physical source subject, we evaluate the matched coefficient, the reference coefficient $\mathbf{a} = \mathbf{0}$, and coefficients transferred from the other source subjects. All conditions use the same frozen cosine-prototype scoring rule. The class prototypes and σ remain fixed. The normalized cosine scores are aggregated across source domains by log-mean-exp as in Equation 3. Cross entropy gives the NLL. For subject i, we form the NLL differences

$$\delta_i^0(\gamma) = \ell_i^{\mathrm{match}}(\gamma) - \ell_i^0,$$

and

$$\delta_i^{\mathrm{swap}}(\gamma) = \ell_i^{\mathrm{match}}(\gamma) - \frac{1}{M-1} \sum_{j \neq i} \ell_{i \leftarrow j}^{\mathrm{swap}}(\gamma),$$

Algorithm 1 Source relation adjustment and validation

Require: Source domain D_s, disjoint trial subsets $\mathcal{T}^{\mathrm{adj}}$ and $\mathcal{T}^{\mathrm{val}}$, frozen graph model and source class prototypes
Ensure: A retained pair $(\mathbf{c}, \mathbf{a})$ or DISCARD

Set $\mathbf{a} \leftarrow \mathbf{0}$
Compute $\mathbf{g} \leftarrow \nabla_{\mathbf{a}} \mathcal{L}_{\mathrm{rel}}(\mathbf{a}; \mathcal{T}^{\mathrm{adj}})|_{\mathbf{a}=\mathbf{0}}$
if $\mathbf{g}$ is non-finite or $\|\mathbf{g}\|_2 = 0$ **then**
    return DISCARD
end if
$\mathbf{d} \leftarrow -\mathbf{g}/\|\mathbf{g}\|_2$
Compute directional GGN curvature κ_{GGN} along $\mathbf{d}$
if κ_{GGN} is non-finite or $\kappa_{\mathrm{GGN}} \leq 0$ **then**
    return DISCARD
end if
$\tau \leftarrow \|\mathbf{g}\|_2/\kappa_{\mathrm{GGN}}$
$\mathcal{L}_0 \leftarrow \mathcal{L}_{\mathrm{rel}}(\mathbf{0}; \mathcal{T}^{\mathrm{adj}})$
while $\tau > 0$ **do**
    $\mathbf{a} \leftarrow \tau \mathbf{d}$
    $\mathcal{L}_\tau \leftarrow \mathcal{L}_{\mathrm{rel}}(\mathbf{a}; \mathcal{T}^{\mathrm{adj}})$
    if $\mathcal{L}_\tau < \mathcal{L}_0$ **then**
        break
    end if
    $\tau \leftarrow \tau/2$
end while
if $\tau = 0$ **then**
    return DISCARD
end if
if $\mathcal{L}_{\mathrm{rel}}(\mathbf{a}; \mathcal{T}^{\mathrm{val}}) \geq \mathcal{L}_{\mathrm{rel}}(\mathbf{0}; \mathcal{T}^{\mathrm{val}})$ **then**
    return DISCARD
end if
Construct $\mathbf{c}$ from $\mathcal{T}^{\mathrm{adj}}$ using Equation 5
return $(\mathbf{c}, \mathbf{a})$

where M is the number of physical source subjects. For each difference, both its full subject-average and every leave-one-subject-out average are required to be strictly negative:

$$\frac{1}{M}\sum_{i=1}^{M}\delta_i(\gamma) < 0, \qquad \frac{1}{M-1}\sum_{i\neq m}\delta_i(\gamma) < 0 \quad \text{for every } m.$$

This gives $2(M+1)$ strict source-only conditions: 32 conditions for a 15-subject source dataset and 34 for a 16-subject source dataset such as SEED-V.

We determine $\gamma \in (0, 1]$ using a source-only scalar search. Linearized source constraints first propose an initial candidate. Every candidate is evaluated with the exact frozen source model. If a candidate γ_0 fails, we construct a quadratic surrogate for each failed constraint. The surrogate matches its value and derivative at $\gamma = 1$ and its exact value at γ_0. If no valid smaller candidate is obtained, we use a cubic surrogate that also matches the value at $\gamma = 0$. The next candidate minimizes the largest surrogate constraint over the current interval. A candidate is accepted only after exact source evaluation.

After γ is fixed, α is calibrated along the fixed coefficient directions predicted by $\mathbf{W}$. Within one source domain, $\mathcal{L}_{\mathrm{rel}}$ averages window losses within a trial, trials within an emotion class, and finally the three emotion classes with equal weight. For global scalar calibration, these domain-level risks are then averaged with equal weight over the source domains included in the current source fold. When a physical subject is held out, all of its source domains are excluded before this average is computed.

Algorithm 2 Source-only calibration of γ and α

Require: Frozen source model, fitted $\mathbf{W}$, retained source contexts and coefficients
Ensure: Calibrated γ and α, or calibration failure

Calibration of γ

Evaluate the matched, reference, and transferred source NLLs with the frozen cosine-prototype predictor
Form all full-subject and leave-one-subject-out constraints from $\delta_i^0(\gamma)$ and $\delta_i^{\mathrm{swap}}(\gamma)$
Evaluate their values and directional derivatives at $\gamma = 1$
Use the linearized constraints to propose an initial $\gamma \in (0, 1]$
loop
    Evaluate all constraints with the exact frozen predictor
    if all exact constraints are strictly negative **then**
        Accept γ
        break
    end if
    Construct the quadratic constraint surrogates defined above
    Propose $\gamma_{\mathrm{new}} \in (0, \gamma)$ by minimizing the largest surrogate constraint
    if no valid smaller candidate is obtained **then**
        Construct the cubic constraint surrogates defined above
        Propose $\gamma_{\mathrm{new}} \in (0, \gamma)$ by minimizing the largest surrogate constraint
    end if
    if no valid $0 < \gamma_{\mathrm{new}} < \gamma$ exists **then**
        return CALIBRATION FAILURE
    end if
    $\gamma \leftarrow \gamma_{\mathrm{new}}$
end loop

Calibration of α

Fix the coefficient directions predicted by $\mathbf{W}$
Compute $\mathcal{L}_{\mathrm{src}}(0)$, $\mathcal{L}'_{\mathrm{src}}(0)$, and the directional GGN curvature κ
if any quantity is non-finite, $\mathcal{L}'_{\mathrm{src}}(0) \geq 0$, or $\kappa \leq 0$ **then**
    return CALIBRATION FAILURE
end if
loop
    $\alpha \leftarrow -\mathcal{L}'_{\mathrm{src}}(0)/\kappa$
    if α is non-finite or $\alpha \leq 0$ **then**
        return CALIBRATION FAILURE
    end if
    Evaluate the exact source risk $\mathcal{L}_{\mathrm{src}}(\alpha)$
    if $\mathcal{L}_{\mathrm{src}}(\alpha) < \mathcal{L}_{\mathrm{src}}(0)$ **then**
        return γ, α
    end if
    Update κ using the exact secant curvature implied by the finite evaluation
    if the updated curvature is non-finite or does not increase **then**
        return CALIBRATION FAILURE
    end if
end loop

Let $\mathcal{L}_{\mathrm{src}}(\alpha)$ denote this equally weighted source-domain risk. Its derivative and directional GGN curvature at $\alpha = 0$ propose the first positive candidate. The candidate is accepted only if

$$\mathcal{L}_{\mathrm{src}}(\alpha) < \mathcal{L}_{\mathrm{src}}(0).$$

Otherwise, the observed finite-step source risk is used in a secant curvature update to produce a smaller positive candidate. The first value giving a strict decrease in the exact source risk is retained. Both γ and α are fixed before target inference. A calibration failure terminates the current source-only fit. No target data are used to select a fallback value of γ or α.

Algorithm 3 Unlabeled target relation inference

Require: Unlabeled target domain D_T, frozen reference graph model, source class prototypes, $\mathbf{W}$, $\bar{\mathbf{c}}$, α, and source predictions $\{\widetilde{\mathbf{a}}_s\}_{s=1}^{S}$
Ensure: Fixed target coefficient vector $\mathbf{a}_T$

```
for each target trial t ∈ T_{D_T} do
    for each window x ∈ W_{D_T,t} do
        Compute u_0^p(x) and z_{D_T}(x) using Equation 3
    end for
    Average the window logits and apply softmax to obtain p_{D_T,t}
    Average u_0(x) over the trial to obtain ū_{D_T,t}
end for
for k = 1, ..., K do
    Compute the probability-weighted trial mean for class k
end for
Concatenate the K class means to obtain c_T using Equation 5
ã_T ← αW(c_T − c̄)
H ← conv({0} ∪ {ã_s}_{s=1}^S)
Find the largest η_T ∈ [0, 1] satisfying η_T ã_T ∈ H
a_T ← η_T ã_T
Freeze a_T for target adaptation
return a_T
```

B.3 Unlabeled Target Inference

For D_T, we first compute the window logits using the frozen reference representation $\mathbf{u}^{\mathrm{p}}_{\mathbf{a}=\mathbf{0}}(\mathbf{x})$ and all source class prototypes. Within each physical target trial, the window logits are averaged before softmax:

$$\mathbf{p}_{D_T,t} = \mathrm{softmax}\left(\frac{1}{|\mathcal{W}_{D_T,t}|}\sum_{\mathbf{x}\in\mathcal{W}_{D_T,t}}\mathbf{z}_{D_T}(\mathbf{x})\right).$$

We separately average $\mathbf{u}_{\mathbf{a}=\mathbf{0}}(\mathbf{x})$ over the same trial. Equation 5 then produces one softly weighted trial mean for each emotion class, and the K means are concatenated in the fixed class order to obtain $\mathbf{c}_T$.

The unconstrained proposal is

$$\widetilde{\mathbf{a}}_T = \alpha\mathbf{W}(\mathbf{c}_T - \bar{\mathbf{c}}).$$

The source predictions, together with $\mathbf{0}$, define $\mathcal{H}$ because both $\{\widetilde{\mathbf{a}}_s\}_{s=1}^{S}$ and $\widetilde{\mathbf{a}}_T$ are produced by the same learned mapping. Thus, the constraint acts on the predictions of the learned rule rather than on the validated source adjustments used to fit it. We then apply the convex-hull constraint in Section 3.2. If the proposal already lies in $\mathcal{H}$, then $\eta_T = 1$; otherwise, only its magnitude is reduced and its direction is retained, using the largest $\eta_T \in [0, 1]$ satisfying $\eta_T\widetilde{\mathbf{a}}_T \in \mathcal{H}$. The resulting $\mathbf{a}_T = \eta_T\widetilde{\mathbf{a}}_T$ is computed once from all unlabeled target trials and remains fixed throughout target adaptation.

C Target Adaptation and Emotion Prediction

This section expands the target-adaptation and prototype-based prediction procedure summarized in Sections 3.3 and 3.4. The target relation coefficients $\mathbf{a}_T$ are computed before adaptation and remain fixed throughout the procedure.

C.1 TRACE-Conditioned Adaptation

For the 62-channel experiments, M_θ maps the flattened 310-dimensional input through two 64-dimensional layers, $310 \to 64 \to 64$. During target adaptation, M_θ, $\mathbf{W}_\Delta$, G_ω, f_ξ, the domain critic, and ρ_{cons} are optimized using the source and target roles described below. The relation map

$\mathbf{W}_\Delta$ is a bias-free $64 \to 64$ linear layer initialized to zero. Together with the frozen relational residual,

$$\Delta(\mathbf{x}; \mathbf{a}) = \mathbf{u}^{\mathrm{p}}_{\mathbf{a}}(\mathbf{x}) - \mathbf{u}^{\mathrm{p}}_{\mathbf{a}=\mathbf{0}}(\mathbf{x}),$$

this gives the representation in Equation 8.

Target windows use $\mathbf{h}_{\mathrm{Query}}(\mathbf{x}) = \mathbf{h}(\mathbf{x}; \mathbf{a}_T)$. A labeled source window used as a reference uses $\mathbf{h}_{\mathrm{Reference}}(\mathbf{x}_s) = \mathbf{h}(\mathbf{x}_s; \gamma\mathbf{a}_T)$. When source domains are evaluated for source selection and classification, their windows are evaluated in the query role.

At each adaptation iteration, f_ξ computes the classification loss for every source domain. The highest-loss source provides labeled queries for $\mathcal{L}_{\mathrm{cls}}$ and for domain alignment with the target through $\mathcal{L}_{\mathrm{adv}}$. The lowest-loss source provides reference representations for target pseudo-label generation. The consistency term $\mathcal{L}_{\mathrm{cons}}$ compares these pseudo labels with the predictions of f_ξ on target queries, and the complete prediction objective is Equation 14.

Target adaptation uses a fixed budget of 1,000 iterations and batch size 96. At every iteration, the main prediction network is updated once using RMSprop with learning rate 10^{-3}, followed by one update of the cluster network using Adam with learning rate 10^{-3}. The two updates therefore follow a 1:1 schedule. The checkpoint after iteration 1,000 is used for the reported final results. Target labels are not used for training, stopping, or checkpoint selection.

C.2 Fixed Clusters and Label Aggregation

Before adaptation, DBSCAN (Ester et al., 1996) is applied once to the domain-normalized 310-dimensional EEG features to determine cluster membership. Its parameters are selected on the first labeled source domain in the fixed source-domain order using cluster purity. We search

$$\mathrm{eps} \in \{1.0, 1.5, 2.0, \ldots, 5.0\}, \qquad \mathrm{min_samples} \in \{3, 4, 5, 6\}.$$

The selected parameters are then applied to each source domain and to the complete unlabeled target domain. Samples assigned the DBSCAN noise label -1 are excluded from the cluster-adaptation samples.

Let K_s be the number of non-noise DBSCAN clusters in source domain D_s. We set

$$K_{\min} = \min_s K_s$$

and retain the $K_{\min}$ largest clusters from each source domain. Let n_s be the number of retained samples in source domain D_s. We then set

$$n_{\min} = \min_s n_s$$

and truncate each retained source set to $n_{\min}$ samples. The DBSCAN memberships remain fixed during adaptation. The representations and cluster statistics computed from these memberships may change. All target windows remain available for final prediction, including windows not included in the cluster-adaptation subset.

The shared mapping network G_ω uses $64 \to 32 \to 64$. For a cluster-adaptation sample set $\mathcal{B}$, let $\widehat{\mathbf{g}}_i$ denote the normalized output of G_ω for sample i, and let c_i denote its fixed cluster assignment. We define

$$\mathcal{P}(i) = \{j \in \mathcal{B} : j \neq i,\ c_j = c_i\}, \qquad \mathcal{V} = \{i \in \mathcal{B} : |\mathcal{P}(i)| > 0\}.$$

Using unit temperature, we use a cluster-based contrastive objective (Khosla et al., 2020):

$$\mathcal{L}_{\mathrm{con}}(\mathcal{B}) = -\frac{1}{|\mathcal{V}|} \sum_{i \in \mathcal{V}} \frac{1}{|\mathcal{P}(i)|} \sum_{p \in \mathcal{P}(i)} \log \frac{\exp(\widehat{\mathbf{g}}_i^\top \widehat{\mathbf{g}}_p)}{\sum_{\substack{j \in \mathcal{B} \\ j \neq i}} \exp(\widehat{\mathbf{g}}_i^\top \widehat{\mathbf{g}}_j)}. \tag{15}$$

When $\mathcal{V}$ is empty, the corresponding loss is zero. For source samples, we compute Equation 15 independently within each source domain and average the resulting domain losses. The target samples form a separate contrastive set. The cluster objective is the sum of the source and target losses. Only G_ω is updated during this step.

Algorithm 4 TRACE-conditioned target adaptation and prediction

Require: Labeled source domains, unlabeled target domain D_T, fixed $\mathbf{a}_T$, calibrated γ, M_θ, $\mathbf{W}_\Delta$, G_ω, f_ξ, and the domain critic
Ensure: Target emotion predictions
Set the prediction modules for target adaptation as specified in Appendix C.1
Select DBSCAN parameters using the designated labeled source domain
Apply DBSCAN to the domain-normalized EEG features of each source domain and D_T
Retain and truncate source cluster samples using the rule in Appendix C.2
Fix the DBSCAN memberships for adaptation
for $j = 1, \ldots, 1000$ **do**
    Evaluate every source domain in the query role
    $s_{\mathrm{Q}} \leftarrow \arg\max_s \mathcal{L}_{\mathrm{cls}}^{(s)}$
    $s_{\mathrm{R}} \leftarrow \arg\min_s \mathcal{L}_{\mathrm{cls}}^{(s)}$
Prediction update
    Compute selected source and target query representations using $\mathbf{a}_T$
    Compute source reference representations using $\gamma\mathbf{a}_T$
    Compute $\mathcal{L}_{\mathrm{cls}}$, $\mathcal{L}_{\mathrm{adv}}$, and target pseudo labels
    Compute $\mathcal{L}_{\mathrm{cons}}$ and $\mathcal{L}_{\mathrm{pred}}$
    Update the prediction parameters with G_ω fixed
Cluster update
    Recompute mapped source and target samples using G_ω
    Compute the source and target cluster contrastive losses
    Update only G_ω
end for
Construct source and target prediction prototypes using the fixed cluster memberships
for each target window **do**
    Associate the window with its closest target prototype
    for each source domain **do**
        Match the target prototype to the closest labeled source prototype by cosine similarity
        Add one vote for the matched source label
    end for
    Predict the class with the largest source-domain vote count
end for
return target predictions

For each fixed cluster, we compute its prediction prototype as the mean of the current G_ω representations of its member samples. Each source prototype is assigned the majority class of its cluster members. Target prototypes remain unlabeled.

For final prediction, each target window is first associated with the target prototype having the highest cosine similarity. For each source domain independently, this target prototype is compared with the labeled source prototypes. The label of the closest source prototype contributes one class count. The predicted emotion is the class with the largest count across source domains. If two or more classes have the same maximum count, the class with the smallest class index is returned; under our label mapping, the order is negative, neutral, and positive.

The class prototypes $\mathbf{b}_{s,k}^0$ used for coefficient inference and the cluster prototypes used here are different quantities. The former are computed under $\mathbf{a} = \mathbf{0}$ and are used to construct the trial probabilities required for $\mathbf{c}_D$. The latter are computed in the G_ω representation space and are used for pseudo-label generation and final emotion prediction.

D Reproducibility and Computational Details

This section provides the dataset processing and evaluation protocol used in Section 4. It also provides TRACE implementation settings and baseline settings. The final subsection reports computational complexity.

D.1 Datasets, Label Mapping, and Evaluation Protocol

We use SEED (Zheng & Lu, 2015), SEED-IV (Zheng et al., 2018), SEED-V (Liu et al., 2021), and MPED (Song et al., 2019). SEED and SEED-IV contain 15 participants with three sessions each. SEED-V contains 16 participants with three sessions. MPED contains 30 participants. All four datasets use 62 EEG channels and five frequency bands.

The six directions among SEED, SEED-IV, and SEED-V use the released five-band DE features. Following E2STN (Zhou et al., 2025), we use its three-class SEED–MPED setting. SEED^3 uses the released five-band DE features. MPED^3 uses the released five-band STFT features. All inputs are arranged as 62×5 channel–band matrices.

The unified class IDs are 0 = negative, 1 = neutral, and 2 = positive. For SEED-IV, the original labels 0 = neutral, 1 = sad, 2 = fear, and 3 = happy are mapped by

$$[1, 0, 0, 2].$$

For SEED-V, the original labels 0 = disgust, 1 = fear, 2 = sad, 3 = neutral, and 4 = happy are mapped by

$$[0, 0, 0, 1, 2].$$

For MPED^3, sadness, neutral, and joy are mapped to classes 0, 1, and 2. Funny, anger, fear, and disgust are excluded. Superscript 3 denotes this three-class setting. Each MPED^3 participant contributes 12 trials.

Each feature dimension is min–max normalized to $[-1, 1]$ using all windows from the same domain. A SEED-family domain is one subject–session recording. An MPED domain is one participant.

For each transfer direction, all labeled source domains are used. SEED^3 $\rightarrow$MPED^3 uses 45 SEED source domains and evaluates all 30 MPED target domains. MPED^3 $\rightarrow$SEED^3 uses 30 MPED source domains and evaluates all 45 SEED target domains. Each target domain is adapted independently using all its unlabeled windows. Target labels are used only for final evaluation.

Accuracy and macro F1 are computed separately for each target domain and then averaged equally across target domains. The complete experiment is repeated with three random seeds. For accuracy, we compute the population standard deviation across target domains for each seed and then average over the three seeds.

The SEED–MPED experiments follow E2STN in the three-class mapping and dataset feature types. They use the common training and evaluation settings described below.

For the robustness experiments, trial-balanced BAcc is computed by averaging window correctness within each trial, trials within each emotion class, classes within each target recording, and finally target recordings.

D.2 TRACE Implementation Settings

D.3 Baseline Implementation and Evaluation Settings

We select representative recent methods for cross-domain EEG emotion recognition, covering multi-source adaptation, domain generalization, graph adaptation, and test-time adaptation. All compared methods use the same label mapping, feature inputs, 62-channel arrangement, normalization, target domains, and evaluation protocol. We use the fixed 1,000-iteration budget described in Section 4.1 and evaluate the final iteration. Target labels are not used for training, hyperparameter selection, early stopping, or checkpoint selection.

We use public implementations when available. Otherwise, we implement the method from its published description. Table 7 summarizes the implementation source and any method-specific change required by our unified protocol.

D.4 Computational Complexity

Let $N = N_c N_b$ denote the number of channel–band nodes. The complete undirected graph contains

$$E = \frac{N(N-1)}{2}$$

Table 6: Implementation settings of TRACE.

Setting	Value
Emotion classes K	3
Frequency bands N_b	5
Common channels	62
Node / relation types	25 / 325
Relation coefficient dimension d	8
Type embedding dimension	8
Edge hidden / output dimension	64 / 64
FiLM rank	8
Node hidden / output dimension	64 / 64
Graph representation dimension	64
Projection P	$64 \to 64 \to 64$, ReLU, dropout 0
M_θ	$310 \to 64 \to 64$
$\mathbf{W}_\Delta$	$64 \to 64$, no bias, zero initialized
G_ω	$64 \to 32 \to 64$, ReLU, BatchNorm after the first layer
f_ξ	$64 \to 32 \to 64 \to 3$
Domain critic	$64 \to 64 \to 1$, ReLU, dropout 0.25
Normalization	Within-domain min–max to $[-1, 1]$
Initial source optimizer	RMSprop
Initial source learning rate	10^{-3}
Initial source batch size / steps	96 / 1,000
Initial source weight decay	10^{-5}
Source coefficient update	GGN proposal + exact loss verification
Fitting $\mathbf{W}$	Minimum-norm SVD solution
Target main optimizer	RMSprop, 10^{-3}
Target cluster optimizer	Adam, 10^{-3}
Target batch size / iterations	96 / 1,000
Prediction : cluster updates	$1 : 1$
Random seeds	Three
Hardware	NVIDIA GeForce RTX 5060 Laptop GPU, 8 GB

concrete edges. The shared edge network is applied to every concrete edge, while the node network is applied to the N nodes. For a fixed hidden dimension, the graph representation computation therefore scales quadratically with N. More generally, with hidden dimension h, its dominant computation is $O(Eh^2 + Nh^2)$.

In our experiments, $N = 62 \times 5 = 310$ and $E = 47{,}895$. Counting the multiply–accumulate operations of the edge network, FiLM projection, node encoder, and projection network P gives approximately 0.343 GMAC per EEG window. Concrete edges are processed in chunks of 2,048 to avoid materializing all edge activations simultaneously. The fixed relation basis and data-derived prototypes are not counted as model parameters.

The target coefficient $\mathbf{a}_T$ is inferred once from each unlabeled target recording and remains fixed throughout target adaptation. Thus, target conditioning does not require repeated coefficient inference at every adaptation iteration. All TRACE experiments were run on an NVIDIA GeForce RTX 5060 Laptop GPU with 8 GB memory.

E Experimental Protocols

E.1 Motivating Relation Diagnostics

The diagnostic experiments in Figure 1 use SEED→SEED-IV. They intentionally use target labels to study relation effects. Relation selection uses the discovery trials. Reported confirmation effects

Table 7: Implementation sources and protocol notes for the compared methods.

Method	Implementation source	Protocol note
MS-MDA	Adapted official code (https://github.com/VoiceBeer/MS-MDA)	Retains the released multi-source alignment objective; uses the unified label mapping and fixed 1,000-iteration budget.
DMMR	Adapted official code (https://github.com/CodeBreathing/DMMR)	Retains the public multi-source objective; uses the shared three-class mapping and target evaluation protocol.
UDA-DDA	Adapted official code (https://github.com/XuanSuTrum/UDA-DDA)	Retains the released MMD/CMMD losses with the unified loader, label mapping, and fixed 1,000-iteration budget.
DSPB	Reimplemented from paper	Reproduces the domain-symmetric and predictive-balance objectives; target labels are excluded from adaptation.
GDDN	Reimplemented from paper	Reproduces general/specific graph disentanglement and classifier aggregation under the common target protocol.
E2STN	Reimplemented from paper	Recreates the published method using the shared 62-channel, five-band, three-class inputs and training budget.
DBDG	Reimplemented from paper	Reproduces the published method under the common source–target split and unlabeled-target protocol.
GPTDS+PC-TTA	Adapted official code (https://github.com/RyersonMultimediaLab/EmotionRecognitionUDA)	Adapts the public 32-channel binary-PSD implementation to 62-channel, five-band inputs and three classes while retaining test-time adaptation.
DBPM	Adapted official code (https://github.com/qwangwl/DBPM)	Retains density-based prototype matching; target labels are reserved for post-hoc evaluation.
HEDN	Adapted official code (https://github.com/qwangwl/HEDN)	Uses the public model with the unified loader and label mapping; target labels are excluded from adaptation.
PGDG	Adapted official code (https://github.com/zw-bit101/PGDG)	Adapts the public pipeline to the unified data interface and common target protocol.
PAA-M	Adapted official code (https://github.com/WuCB-BCI/PAA)	Uses the public PAA core and completes the PAA-M-specific path from the published description under the common protocol.

Table 8: Computational characteristics of TRACE.

Quantity	Value
Model parameters	71,762
Graph nodes	310
Concrete graph edges	47,895
Relation types	325
Graph representation computation	0.343 GMAC / window
Edge chunk size	2,048
Hardware	NVIDIA GeForce RTX 5060 Laptop GPU, 8 GB

use the independent confirmation trials. These diagnostics are separate from TRACE deployment. TRACE infers $\mathbf{a}_T$ without target labels.

The split is performed at the physical stimulus level rather than by randomly splitting EEG windows. Within every SEED-IV session and mapped emotion class, 50% of the trials are assigned to an independent confirmation set and the remaining trials form the discovery set. The same stimulus split is shared across subjects within a session. The confirmation trial IDs are listed in Table 9.

For the SEED source data used in these diagnostics, 20% of the trials are additionally held out while locking the same stimulus across all three sessions. The held-out trial IDs are 15 for negative, 11 for neutral, and 9 for positive emotion.

For Figure 1(a–b), the predictor is trained first and then frozen. Within each regional group, frequency-specific relation types are progressively attenuated using

$$\rho \in \{0, 0.1, \ldots, 1.0\}.$$

The Target condition ranks relations using the current target. The Source control retains the target-informed regional groups and the same per-group selection counts but ranks their frequency-specific relations using source information. The Random control uses the same selection counts with randomly chosen relations.

Table 9: SEED-IV confirmation trial IDs used in the motivating diagnostics. The remaining trials of the corresponding class and session form the discovery set.

Session	Negative	Neutral	Positive
1	2, 10, 11, 14, 15, 17	4, 21, 23	3, 19, 20
2	1, 2, 6, 16, 18, 22	5, 14, 21	3, 9, 23
3	2, 4, 8, 11, 13, 21	12, 16, 22	5, 14, 15

The Uniform control uses the same candidate regional group and the same total nominal attenuation as the targeted intervention but distributes that attenuation uniformly over all relations in the group. Specifically, if a group contains m candidate relations and the targeted condition attenuates k selected relations by fraction ρ, the Uniform control attenuates each of the m relations by

$$\rho\frac{k}{m}.$$

It therefore matches the total attenuation within the group rather than uniformly scaling every relation in the complete graph.

For Figure 1(c), each relation-effect profile is centered by subtracting its mean effect before cosine similarity is computed. The within-target comparison uses disjoint trial subsets of the same target, whereas the between-target comparison uses different subjects from the same session. For Figure 1(d), Current uses the current target's own relation selection, Other uses individual same-session donor targets, Pooled uses the mean effect profile of the same-session donors, and Random is the exact expectation under uniform random selection. Five selected relations are evaluated through five separate single-relation interventions, after which their held-out NLL reductions are averaged.

All confidence intervals are based on 2,000 bootstrap resamples at the physical-subject level. All three sessions of a sampled subject remain grouped during resampling.

E.2 Ablation, Target Substitution, and Perturbation Protocols

The ablations in Table 2 change one component while retaining the remaining protocol.

w/o Relation Adaptation sets the relation coefficient vector to zero and retains the reference relation distribution.

w/o Target Conditioning uses one source-optimized coefficient vector shared across all target recordings in a transfer direction.

w/o Hull Constraint uses the raw target-conditioned proposal without radial contraction.

w/o Relational Residual replaces $\Delta(\mathbf{x}; \mathbf{a}_T)$ with $\mathbf{u}_{\mathbf{0}}^{\mathrm{p}}(\mathbf{x})$. It retains graph information without relation adjustment.

w/o Graph Relation Adjustment uses the same inferred coefficient vector. It maps the coefficient vector directly to a feature residual without graph relation adjustment.

These two controls compare the relational residual with reference graph information and with coefficient conditioning without graph relation adjustment, respectively.

All variants use the same prediction and target adaptation procedure. They use the same training budget.

For the target-substitution experiment, Own uses the coefficient inferred from the current target context. The five Other conditions use fixed cross-subject permutations within the same session. Shared uses the direction-specific source-optimized coefficient, and Zero uses $\mathbf{a} = \mathbf{0}$. Each condition independently reruns the complete downstream target-adaptation procedure with the same training budget and settings.

For the robustness experiments, each method is first adapted to the clean target recording and then frozen. For TRACE, $\mathbf{a}_T$ and the clean-target cluster statistics are also frozen. Electrode masking replaces all five band features of $m \in \{3, 6, 12, 24\}$ selected electrodes with their clean-target



Figure 4: Examples of target-specific relation adjustments on SEED-IV→SEED-V. Columns show subjects 1–3 from session 1. The top row displays region-level averages of $q_{\mathbf{a}_T}(r) - q_{\mathbf{0}}(r)$. The bottom row resolves the same temporal–parietal region pair across the 5×5 frequency-band pairs. Color scales are shared across targets within each row.

median values, with results averaged over three nested random electrode orderings. For Gaussian noise, zero-mean noise is added using the clean-target standard deviation of each feature multiplied by $\sigma \in \{0.1, 0.2, 0.4, 0.8\}$. Results are averaged over 10 independent perturbations per target and noise level. All methods receive the same perturbations and are not readapted after perturbation.

F Additional Analyses

F.1 Visualization of Target-Specific Relation Adjustments

We visualize how the shared relation structure changes for individual target recordings using

$$q_{\mathbf{a}_T}(r) - q_{\mathbf{0}}(r).$$

For a compact overview, the relation-type changes are first aggregated over frequency-band pairs to obtain a 5×5 matrix between scalp regions. For the temporal–parietal regional pair, we then retain the frequency-band resolution and display the corresponding 5×5 band-pair matrix. The same regional pair is used for all three targets. This provides both a regional view and the frequency-specific detail used by TRACE.

The three targets exhibit different regional and frequency-specific adjustment patterns despite sharing the same relation definitions. These visualizations describe predictive relation adjustments and are not interpreted as causal brain connectivity.