Learning Domain-Invariant Representations for EEG Emotion Generalization via Representational Similarity Alignment

Yubin Sun & Liying Yang* & JiuLin Fu & Yifan Bai & Ruiqi Gao
Department of Computer Science and Technology, Xidian University

Abstract

Cross-subject electroencephalography (EEG) emotion recognition is challenged by pronounced inter-subject variability, which hinders generalization to unseen subjects in the domain generalization setting. Most existing approaches focus on aligning feature distributions, whereas the invariance of relational structure in latent representations has received far less attention. Inspired by representational similarity analysis (RSA), we revisit cross-subject EEG generalization from a structural perspective and propose an RSA-based structural consistency constraint (RSASC). This regularizer aligns representational dissimilarity matrices (RDMs) computed from different subjects' latent representations, encouraging the model to learn emotion-relevant representations that are more consistent across subjects. Experiments on the SEED and SEED-IV datasets show that the proposed regularizer yields stable improvements in cross-subject generalization: it achieves the best accuracy on SEED, and on SEED-IV it attains competitive performance with the lowest variance. These results suggest that structural invariance offers a novel perspective that complements distribution alignment for learning subject-invariant EEG representations.

Keywords: EEG emotion recognition; cross-subject; domain generalization; representational similarity analysis; structural alignment

Introduction

Emotion is a psychophysiological response elicited by external stimuli in humans and other animals, commonly characterized as a dynamic, multi-component psychological process (Mauss & Robinson, 2009). As a core component of cognition, emotion plays a critical role in adaptive behavior by modulating attention, perception, learning, and memory, and by shaping judgment and decision making (Lerner et al., 2015). Affective computing seeks to equip computational systems with the ability to recognize and respond to emotional states, enabling applications in personalized medicine, digital mental health, and psychiatric intervention (Schlicher et al., 2025).

Electroencephalography (EEG) is a non-invasive technique for measuring brain activity with millisecond-level temporal resolution, enabling near real-time analysis of neural states (Zanetti et al., 2025). Its practicality and cost-effectiveness in engineering deployments (LaRocco et al., 2020), together with its direct reflection of neural processes underlying emotion, allow more direct and reliable assessment of internal affective states (Lim et al., 2024). As a result, EEG has been widely adopted for emotion recognition and affective computing research in recent years.

In prior work, a subset of EEG-based emotion recognition methods adopt a subject-dependent modeling paradigm, where the model is trained and evaluated separately for each individual. Under this setting, the model can fully exploit subject-specific temporal dynamics and spectral patterns in EEG signals, typically achieving high classification accuracy and stable performance (Gu et al., 2023; He et al., 2024; Yang et al., 2018).

Despite strong within-subject performance, EEG emotion recognition models often suffer substantial degradation in cross-subject settings due to pronounced inter-individual variability in EEG signals, arising from anatomical, recording, and cognitive differences (Alghamdi et al., 2025). Such performance drops are widely observed on public datasets such as DEAP (Koelstra et al., 2011) and SEED (Zheng & Lu, 2015), underscoring the challenge of cross-subject distributional shift.

To alleviate cross-subject distributional discrepancies, prior work has conducted extensive exploration from the perspectives of model architecture and representation learning. Typical strategies include designing deeper or more sophisticated network structures (Cheng et al., 2024; Shen et al., 2022), adopting joint time–frequency–spatial modeling schemes (Zhu et al., 2024), and enhancing feature discriminability via attention mechanisms (Gong et al., 2023; Sun et al., 2025), or adaptive normalization techniques (Fdez et al., 2021). Although these methods improve EEG emotion modeling, the learned representations remain largely shaped by the training objective. Consequently, modeling emotion-relevant and cross-subject stable structures at the loss level has become a key direction in cross-subject EEG emotion recognition.

Most existing cross-subject EEG emotion recognition studies address subject shift by introducing distribution-alignment constraints into the loss function to reduce discrepancies between source and target domains and to learn subject-invariant representations. Representative methods include Maximum Mean Discrepancy (MMD) based on statistical distribution matching (Sejdinovic et al., 2013), CORrelation ALignment (CORAL), which aligns second-order statistics (Jayaram & Barachant, 2018), and adversarial domain adaptation approaches such as DANN (Ganin et al., 2016) and DAAN (Yu et al., 2019). Despite their effectiveness, these methods primarily rely on moment matching or adversarial training to en-

*Corresponding author: yangliying1208@163.com

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

force distribution-level consistency, while the effectiveness of emotion-relevant representations at the level of relational structure remains largely unexplored.

Notably, Representational Similarity Analysis (RSA) models the relative relational structure among representations, providing an effective tool for characterizing high-dimensional embeddings. In recent years, RSA has been widely used to analyze and compare the geometric structure of representations learned by artificial models (Du et al., 2025; Dwivedi & Roig, 2019), offering a new perspective for understanding representation alignment from a relational-structure viewpoint.

Motivated by this perspective, we explore whether transferable information in cross-subject EEG emotion recognition is not only characterized by distribution-level alignment objectives, but also encoded in the relative relational structure among representations. We hypothesize that different subjects share an emotion-related, cross-subject invariant latent representational structure that is more stable and identifiable across individuals.

Building on this hypothesis, we introduce RSA as a principled mechanism for constraining representational structure. Departing from its conventional use as a post hoc analytic tool, we incorporate RSA as a structural consistency regularizer that explicitly guides model training toward emotion-relevant representations with strong cross-subject generalization.

The main contributions of this study are summarized as follows:

- We propose a structure-oriented cross-subject modeling hypothesis, positing that transferable, emotion-relevant information can be characterized by the relative relational structure formed among latent EEG representations.
- We introduce an RSA-based structural consistency constraint (RSASC) that aligns representational relational structures across subjects during training, thereby enhancing cross-subject semantic consistency.
- Experiments on public EEG emotion datasets demonstrate that the proposed structural constraint yields superior cross-subject generalization compared with representative distribution-alignment baselines.

Related Work

Representation Similarity Analysis in Neuroscience

Representational Similarity Analysis (RSA) was originally introduced in cognitive neuroscience to compare neural representational patterns elicited by different stimuli, experimental conditions, or brain regions (Cichy et al., 2014; Gauthaman et al., 2025; Kriegeskorte et al., 2008).

With the advancement of deep learning, RSA has increasingly been introduced into the AI community (Hermann & Lampinen, 2020). For example, prior work used RSA to align MEG signals with deep neural network representations, revealing the temporal dynamics of scene representations (Cichy et al., 2017). A recent study applied RSA to compare multimodal large language models with object representations in the human visual cortex, showing that such models can spontaneously develop object-concept geometries highly consistent with those observed in humans (Du et al., 2025).

Beyond post hoc analysis, RSA has further been employed before or during training to quantify relational structure between tasks or model representations. Dwivedi and Roig used RSA to construct a task taxonomy and guide source-model selection in transfer learning (Dwivedi & Roig, 2019), while subsequent work introduced RSA as a training loss for knowledge distillation, improving knowledge transfer and student model performance (McClure & Kriegeskorte, 2016).

Collectively, these studies show that RSA links representational structure to transfer effectiveness. Building on this insight, we incorporate RSA as a training-stage structural constraint to promote cross-subject consistency in relational geometry and improve transferability. To our knowledge, this is the first work to explore RSA as a training-time structural regularization for cross-subject EEG emotion generalization.

Method

Problem Formulation

We consider cross-subject EEG emotion recognition under the domain generalization (DG) setting. Due to substantial inter-individual differences in physiology and cognitive processing, each subject is treated as an independent domain.

Let the source-domain dataset be $\mathcal{D} = \{\mathcal{D}^{(s)}\}_{s=1}^{S}$, where S is the number of source subjects. For subject s, the labeled samples are $\mathcal{D}^{(s)} = \{(x_i^{(s)}, y_i^{(s)})\}_{i=1}^{N_s}$, where N_s denotes the number of EEG samples for subject s, $x_i^{(s)} \in \mathbb{R}^{C\times T}$ denotes an EEG segment with C channels and T time points, and $y_i^{(s)} \in \mathcal{Y}$ is the corresponding emotion label. The training set contains only source domains, i.e., $\mathcal{D}_s = \{\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(S)}\}$, while testing is conducted on an unseen target subject domain $\mathcal{D}_t = \{(x_j^{(t)}, y_j^{(t)})\}_{j=1}^{N_t}$.

Our goal is to learn a function that maps EEG signals into a latent space and generalizes to unseen subjects:

$$f_\theta(\cdot) : \mathcal{X} \rightarrow \mathcal{Z}. \tag{1}$$

The learned representation is further used for emotion classification.

Overview of the Framework

As illustrated in Fig.1, we introduce an RSA-based structural consistency constraint (RSASC), within a standard cross-subject DG EEG emotion recognition framework. Given multi-source subject EEG data as input, a shared feature encoder is used to extract latent representations. In addition to conventional classification supervision, RSASC is imposed to encourage the learning of emotion-related representations with cross-subject generalization ability.

During training, EEG signals from multiple source subjects are provided to the model as source data. To facilitate subsequent structural modeling, a Sample Pair Construction module is first applied prior to feature extraction to organize and sample EEG instances for relational analysis.

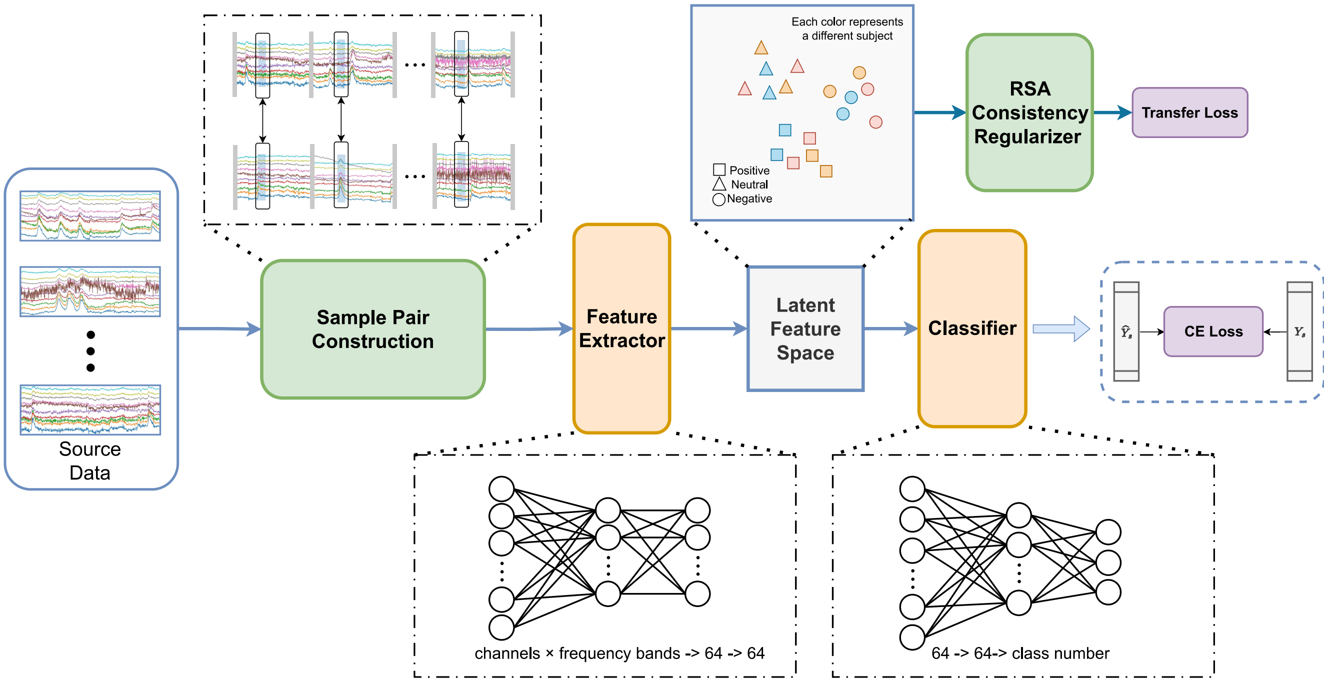


Figure 1: Overview of the proposed cross-subject DG EEG emotion recognition framework with an RSA-based structural consistency constraint.

The constructed EEG samples are then fed into a shared Feature Extractor $f_\theta(\cdot)$, implemented as a multi-layer perceptron (MLP). This module maps raw EEG signals into a latent space to obtain more discriminative representations. The encoder output is given by $z = f_\theta(x)$, where $z \in \mathcal{Z}$ denotes the latent EEG representation. The latent space is designed to capture task-relevant information associated with emotional states.

Based on the learned latent representations, the model is optimized along two parallel objectives. On the one hand, the latent features z are passed to a Classifier, also implemented as an MLP, to predict emotion categories. This branch is trained using standard cross-entropy loss to ensure discriminative performance. On the other hand, as the core contribution of this work, RSASC is directly imposed on the latent representations in the form of a training loss. This constraint encourages structural consistency across subjects, and its formulation and optimization are detailed in the following sections.

Representational Similarity Analysis: Preliminaries

Representational Similarity Analysis (RSA) analyzes and compares representations by characterizing the relational structure among samples in a representation space. Instead of constraining individual feature values, RSA captures the representational structure formed by a set of latent representations, which is typically summarized by a Representational Dissimilarity Matrix (RDM).

Given latent representations, the RDM entry at (i, j) measures the dissimilarity between the i-th and j-th samples. Since an RDM is symmetric with a zero diagonal, RSA commonly keeps only the upper (or lower) triangular part and vectorizes it for subsequent correlation analysis or structural consistency evaluation.

When RSA is used to compare two entities (i.e., two RDMs), the comparison is meaningful only if they share consistent index semantics: the i-th row/column in both RDMs must correspond to the same stimulus condition or experimental instance (e.g., the same stimulus within the same temporal window).

Based on this property, to impose a structural consistency constraint on latent representations across subjects, it is essential to first ensure that the sample sets used to construct each subject's RDM are strictly aligned at the level of stimulus conditions. The next section introduces an RSA-oriented sample organization strategy.

Sample Alignment and Data Organization for RSASC

To ensure interpretability of cross-subject structural comparison, we first retrieve and match EEG segments collected under the same stimulus condition across subjects, so that the RDMs constructed for different subjects share consistent row/column semantics. To this end, we adopt a trial–time-level correspondence scheme: EEG segments elicited by the same set of stimulus items are matched across subjects, forming cross-subject aligned sample sequences.

As illustrated in Fig.2, consider two source subjects subA and subB. For the same stimulus item, both subA and subB

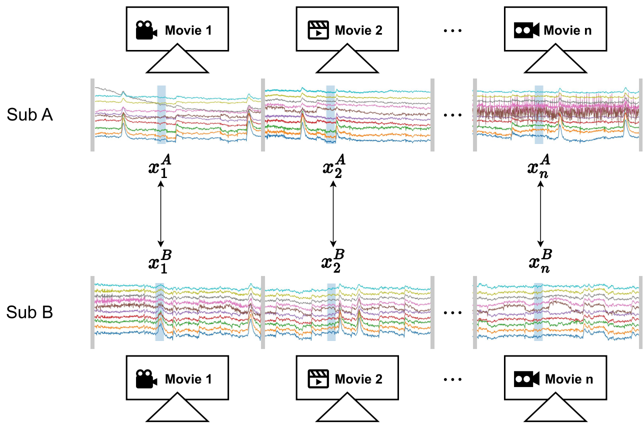


Figure 2: Trial–time-level sample alignment across subjects for RSA.

contain the corresponding EEG observation segment, enabling alignment of samples under an identical stimulus condition.

Specifically, at each iteration we first generate a shared set of trial–time indices:

$$\mathcal{I}_b = \{(k, w)\}_b, \quad k \in \{1, \ldots, m\}, \tag{2}$$

where k denotes the trial index, m is the total number of trials, and w denotes the within-trial time (window) index. Then, for each source subject s participating in the current iteration, we use the same index set $\mathcal{I}_b$ to retrieve the corresponding EEG segments from its dataset, forming the subject-specific input batch:

$$\mathcal{U}_b^{(s)} = \left\{ x_{k,w}^{(s)} \mid (k, w) \in \mathcal{I}_b \right\}. \tag{3}$$

With this shared-index-based sampling scheme, RDMs constructed from different subjects have consistent row/column semantics, enabling clear and interpretable cross-subject structural comparison.

RSA-based structural consistency constraint

Based on the batch input constructed in the previous section, $\mathcal{U}_b = \{\mathcal{U}_b^{(s)}\}_{s \in \mathcal{S}}$, EEG samples from different source subjects are first fed into the shared feature encoder $f_\theta(\cdot)$ to obtain their corresponding latent representations:

$$z_{k,w}^{(s)} = f_\theta\left(x_{k,w}^{(s)}\right), \qquad x_{k,w}^{(s)} \in \mathcal{U}_b^{(s)}. \tag{4}$$

We further define the set of latent representations for subject s within the current batch as $\mathcal{Z}_b^{(s)} = \{z_{k,w}^{(s)} \mid (k, w) \in \mathcal{I}_b\}$. In the latent representation space, a representational dissimilarity matrix (RDM) $\mathbf{R}^{(s)} \in \mathbb{R}^{N \times N}$ is constructed for each subject s, where $N = |\mathcal{I}_b|$. In this work, the squared Euclidean distance is adopted as the dissimilarity measure, and the (i, j)-th entry of the RDM is defined as

$$R_{ij}^{(s)} = \left\| z_i^{(s)} - z_j^{(s)} \right\|_2^2, \qquad i, j = 1, \ldots, N. \tag{5}$$

The resulting matrix $\mathbf{R}^{(s)}$ is symmetric with zero-valued diagonal entries. To avoid redundant computations, we introduce an upper-triangular vectorization operator $\mathcal{T}(\cdot)$, which extracts the strictly upper triangular elements (excluding the diagonal) and vectorizes them as

$$\mathbf{r}^{(s)} = \mathcal{T}\left(\mathbf{R}^{(s)}\right) = \mathrm{vec}\left(\mathrm{triu}\left(\mathbf{R}^{(s)}, 1\right)\right) \in \mathbb{R}^{\frac{N(N-1)}{2}}. \tag{6}$$

Here, $\mathrm{triu}(\mathbf{R}, 1)$ retains the strictly upper triangular part of $\mathbf{R}$ with indices satisfying $i < j$, and $\mathrm{vec}(\cdot)$ denotes vectorization according to a predefined and consistent ordering (either row-wise or column-wise).

Based on the above construction, the relational structure vectors $\mathbf{r}^{(s)}$ obtained for different subjects can be directly compared at the structural level. Finally, for any pair of source subjects s and t, the RSA-based structural consistency loss is defined as the mean squared error (MSE) between their corresponding relational structure vectors:

$$\mathcal{L}_{\mathrm{RSA}}(s, t) = \left\| \mathbf{r}^{(s)} - \mathbf{r}^{(t)} \right\|_2^2. \tag{7}$$

The overall computation procedure of $\mathcal{L}_{\mathrm{RSA}}$ is illustrated in Fig.3.

Training Objective and Loss Function

At the t-th iteration, training samples $\{x_{k,w}^{(s)} \mid (k, w) \in \mathcal{I}_b^t\}$ are collected according to the shared index set. Each sample is mapped by the shared encoder $f_\theta(\cdot)$ to a latent representation $z_{k,w}^{(s)}$, which is then fed into the classifier $cls_\phi(\cdot)$ to produce a categorical probability vector:

$$\mathbf{p}_{k,w}^{(s)} = \mathrm{softmax}\left(cls_\phi\left(z_{k,w}^{(s)}\right)\right) \in \mathbb{R}^K,$$

where K denotes the number of emotion categories. Supervised learning is performed using the standard cross-entropy loss:

$$L_{\mathrm{CE}}(s) = -\frac{1}{N_b} \sum_{(k,w) \in \mathcal{I}_b^t} \sum_{c=1}^{K} y_{k,w,c}^{(s)} \log p_{k,w,c}^{(s)},$$

where $y_{k,w}^{(s)} \in \{0, 1\}^K$ is the one-hot ground-truth label, $p_{k,w,c}^{(s)}$ denotes the predicted probability of class c, and $N_b = |\mathcal{I}_b^t|$.

To reduce the $O(|\mathcal{S}|^2)$ complexity induced by fully connected subject pairing, a cyclic pairing strategy is adopted for computing the RSA-based structural consistency loss $\mathcal{L}_{\mathrm{RSA}}$. The subject pair set is defined as

$$\mathcal{P}_{\mathrm{cyc}} = \{(s_i, s_{i+1}) \mid i = 1, \ldots, M, \ s_{M+1} = s_1\},$$

and the RSASC loss is formulated as

$$\mathcal{L}_{\mathrm{RSA}} = \frac{1}{|\mathcal{P}_{\mathrm{cyc}}|} \sum_{(s,t) \in \mathcal{P}_{\mathrm{cyc}}} \left\| \mathbf{r}^{(s)} - \mathbf{r}^{(t)} \right\|_2^2.$$

The overall training objective jointly combines classification supervision and structural consistency:

$$\mathcal{L} = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} L_{\mathrm{CE}}(s) + \lambda \, \mathcal{L}_{\mathrm{RSA}},$$

where λ controls the trade-off between discriminative learning and cross-subject structural alignment.

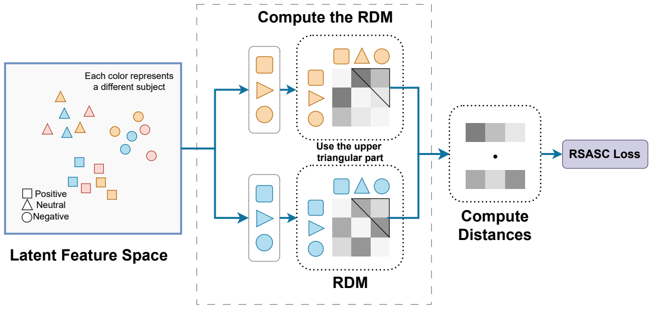


Figure 3: Overview of the computation procedure of the RSA-based structural consistency loss $\mathcal{L}_{\mathrm{RSA}}$.

Table 1: Comparison of classification accuracy (Avg. %) and standard deviation (Std. %) on the SEED and SEED-IV

Method	SEED		SEED-IV	
	Avg. (%)	Std. (%)	Avg. (%)	Std. (%)
Baseline	81.53	5.77	66.46	10.33
DANN	81.86	5.99	68.53	10.18
MMD	83.19	**4.61**	**68.83**	9.88
CORAL	82.02	5.36	68.64	10.63
DAAN	82.16	5.40	68.34	10.05
RSASC (ours)	**83.70**	5.83	68.78	**8.86**

Experiments

Datasets

We conducted experiments on two publicly available EEG emotion recognition datasets: SEED (Zheng & Lu, 2015) and SEED-IV (Zheng et al., 2018).

SEED In the SEED dataset, each subject watches 15 Chinese film clips that elicit three emotional states: *positive*, *negative*, and *neutral*. EEG signals are collected at three different time periods with an interval of approximately one week, corresponding to three sessions. Therefore, SEED contains a total of $15 \times 3 \times 15 = 675$ trials (15 subjects, 3 sessions per subject, and 15 trials per session).

SEED-IV SEED-IV follows the same experimental protocol as SEED, but extends both the number of stimuli and the emotion categories. Each subject watches 24 clips designed to evoke four emotions: *happiness*, *sadness*, *fear*, and *neutral*. Data are also recorded across three sessions, resulting in $15 \times 3 \times 24 = 1080$ trials in total (15 subjects, 3 sessions, and 24 trials per session).

Both datasets are recorded using a 62-channel ESI NeuroScan system at a sampling rate of 1000 Hz, and are subsequently downsampled to 200 Hz with a 1–75 Hz band-pass filter applied. Features are extracted over five canonical frequency bands: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–13 Hz), Beta (13–30 Hz), and Gamma (30–50 Hz).

Implementation Details

We directly use the pre-computed Differential Entropy (DE) features provided by the SEED and SEED-IV datasets and concatenate them into a single feature vector. As a result, each EEG sample is represented as a 310-dimensional feature vector (62×5), i.e., $x \in \mathbb{R}^{310}$.

We evaluate our method in the generalization setting of the cross-subject domain using a Leave-One-Subject-Out protocol (LOSO). In each fold, one subject is held out as an unseen target domain, while the remaining subjects constitute the source domains for training. Importantly, during training we strictly do not access any data or labels from the target subject.

We optimize the network using RMSprop optimizer with a learning rate of 10^{-3}. At each iteration, we sample 12 segments per trial for SEED and 4 segments per trial for SEED-IV under different experimental settings to construct the training data. The maximum number of training iterations is set to 1,000. All experiments are conducted with a fixed random seed of 20 for reproducibility.

All experiments are implemented in PyTorch and run on an NVIDIA GeForce RTX 4060 Ti GPU with CUDA 12.6.

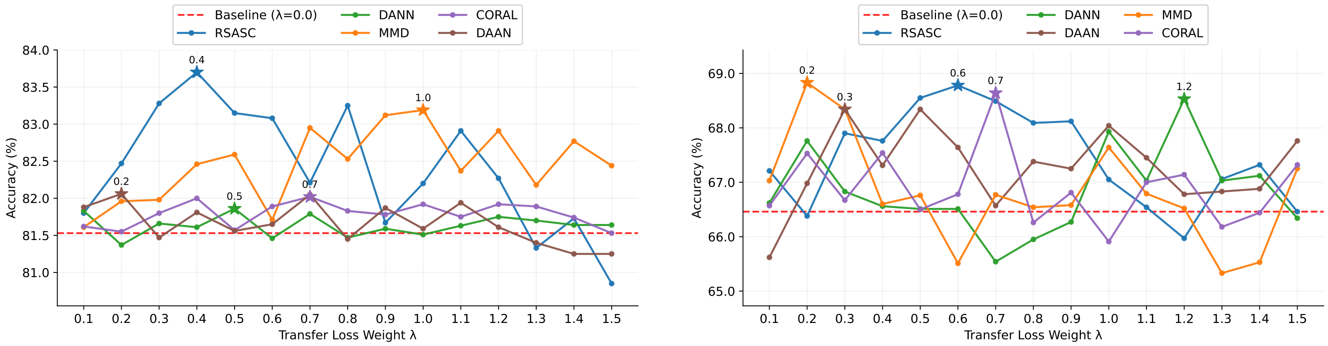
For reproducibility, our code is publicly available at https://github.com/Gekkouka/RSA-Aligned-EEG-Emotion-Generalization.

Experiment Results

Performance Comparison In this section, we compare different methods under the LOSO cross-subject domain generalization protocol on SEED and SEED-IV, reporting the mean accuracy (Avg.) and standard deviation (Std.) across all folds. The baseline model employs only the classification cross-entropy loss with a fixed weight of 1. We further incorporate several widely used feature alignment losses, including DANN (Ganin et al., 2016), DAAN (Yu et al., 2019), CORAL (Jayaram & Barachant, 2018), and MMD (Sejdinovic et al., 2013), with MMD and DANN being two of the most commonly adopted alignment strategies. For a fair comparison, we report the best performance of each method by selecting its optimal hyperparameter λ.

As shown in Table 1, our method achieves the highest accuracy on SEED, with 83.70%, outperforming all competing approaches and yielding a 2.17% absolute improvement over the baseline (81.53%), which is the largest gain among all methods. On SEED-IV, our method obtains an accuracy of 68.78%, which is comparable to the best result achieved by MMD (68.83%), while attaining the smallest standard deviation (Std. = 8.86%) among all methods, indicating improved stability and robustness in cross-subject evaluation.

Sensitivity Analysis on the Transfer-Loss Weight λ To systematically analyze the effect of the transfer-loss weight on model performance, we conduct sensitivity experiments under the LOSO (leave-one-subject-out) protocol by varying the loss weight λ for different transfer strategies. Specifically, we



(a) SEED

(b) SEED-IV

Figure 4: Sensitivity of accuracy to the transfer loss weight λ under the LOSO protocol.

perform a grid search over $\lambda \in [0.1, 1.5]$ and evaluate the resulting classification performance on both SEED and SEED-IV, as summarized in Fig. 4.

On SEED, different transfer paradigms exhibit clearly different sensitivities to λ. In particular, the proposed RSASC and the MMD baseline show a consistent rise–then–fall pattern as λ increases. When λ is small, the transfer constraint is too weak to effectively mitigate cross-subject feature discrepancies. In contrast, an overly large λ enforces excessively strong alignment, which can suppress emotion-discriminative information and lead to an over-regularization effect, thereby degrading performance; similar observations have also been reported in AC-CNCD (Sun et al., 2025). By comparison, DANN, CORAL, and DAAN exhibit relatively stable behavior over a wider range of λ, although their overall performance gains are limited.

On SEED-IV, performance fluctuations with respect to λ become more pronounced. This is mainly attributed to the increased task difficulty of SEED-IV, including fewer samples, more emotion classes, and larger inter-subject variability, which makes the model more sensitive to hyperparameter choices. Under this setting, several competing methods show sharp peaks or unstable oscillations at specific λ values, whereas the proposed RSASC method yields a smoother performance curve across λ, indicating improved robustness and stability.

Overall, the proposed method achieves stable and competitive performance within a reasonably wide range of λ. On SEED, RSASC performs most stably and achieves the best results in the range $\lambda \in [0.3, 0.6]$, while on SEED-IV the favorable range concentrates around $\lambda \in [0.4, 0.7]$. These results suggest that our method is relatively tolerant to the choice of λ and remains effective under varying dataset scales and task complexities.

Conclusion

In this study, we investigate cross-subject EEG emotion recognition from a representational perspective. Under the DG setting, pronounced inter-subject variability remains a key obstacle to robust generalization. Most existing methods focus on aligning statistical distributions of features, while paying relatively less attention to the role of relational structure in latent representations for cross-domain transfer.

Inspired by cognitive neuroscience, we introduce RSA during training as a structural consistency regularizer. By explicitly aligning the relational structure of latent EEG representations across subjects, the proposed approach encourages the model to learn emotion-related representations that are both discriminative and subject-invariant. Experiments on public EEG emotion datasets show that this strategy consistently improves cross-subject generalization, achieving comparable or even better accuracy and stability than representative distribution-alignment baselines.

More broadly, our findings suggest that beyond matching distributional moments, structural invariance offers a key and complementary view for cross-subject EEG modeling. Integrating RSA into a DG framework provides a theoretically grounded way to impose representational structure constraints in the training objective.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 62374121 and 61974109, and by the Fundamental Research Funds for the Central Universities and the Postgraduate Innovation Fund of Xidian University under Grant YJSJ26014.

References

Alghamdi, A. M., Ashraf, M. U., Bahaddad, A. A., Almarhabi, K. A., Al Shehri, W. A., & Daraz, A. (2025). Cross-subject eeg signals-based emotion recognition using contrastive learning. *Scientific Reports*, *15*(1), 28295.

Cheng, Z., Bu, X., Wang, Q., Yang, T., & Tu, J. (2024). Eeg-based emotion recognition using multi-scale dynamic cnn and gated transformer. *Scientific Reports*, *14*(1), 31319.

Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *NeuroImage*, *153*, 346–358.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature neuroscience*, *17*(3), 455–462.

Du, C., Fu, K., Wen, B., Sun, Y., Peng, J., Wei, W., Gao, Y., Wang, S., Zhang, C., Li, J., et al. (2025). Human-like object concept representations emerge naturally in multimodal large language models. *Nature Machine Intelligence*, 1–16.

Dwivedi, K., & Roig, G. (2019). Representation similarity analysis for efficient task taxonomy & transfer learning. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12387–12396.

Fdez, J., Guttenberg, N., Witkowski, O., & Pasquali, A. (2021). Cross-subject eeg-based emotion recognition through neural networks with stratified normalization. *Frontiers in neuroscience*, *15*, 626277.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of machine learning research*, *17*(59), 1–35.

Gauthaman, R. M., Ménard, B., & Bonner, M. F. (2025). Universal scale-free representations in human visual cortex. *PLOS Computational Biology*, *21*(11), e1013714.

Gong, L., Li, M., Zhang, T., & Chen, W. (2023). Eeg emotion recognition using attention-based convolutional transformer neural network. *Biomedical Signal Processing and Control*, *84*, 104835.

Gu, Y., Zhong, X., Qu, C., Liu, C., & Chen, B. (2023). A domain generative graph network for eeg-based emotion recognition. *IEEE Journal of Biomedical and Health Informatics*, *27*(5), 2377–2386.

He, H., Yang, L., Du, J., Hai, Q., Tian, Y., & Fu, J. (2024). Mdac: Eeg emotion recognition with multi-scale dual attention capsule network. *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 4164–4171.

Hermann, K., & Lampinen, A. (2020). What shapes feature representations? exploring datasets, architectures, and training. *Advances in Neural Information Processing Systems*, *33*, 9995–10006.

Jayaram, V., & Barachant, A. (2018). Moabb: Trustworthy algorithm benchmarking for bcis. *Journal of neural engineering*, *15*(6), 066011.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18–31.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

LaRocco, J., Le, M. D., & Paeng, D.-G. (2020). A systemic review of available low-cost eeg headsets used for drowsiness detection. *Frontiers in neuroinformatics*, *14*, 553352.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology*, *66*(1), 799–823.

Lim, R. Y., Lew, W.-C. L., & Ang, K. K. (2024). Review of eeg affective recognition with a neuroscience perspective. *Brain Sciences*, *14*(4), 364.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and emotion*, *23*(2), 209–237.

McClure, P., & Kriegeskorte, N. (2016). Representational distance learning for deep neural networks. *Frontiers in computational neuroscience*, *10*, 131.

Schlicher, M., Li, Y., Murthy, S. M. K., Sun, Q., & Schuller, B. W. (2025). Emotionally adaptive support: A narrative review of affective computing for mental health. *Frontiers in Digital Health*, *7*, 1657031.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The annals of statistics*, 2263–2291.

Shen, X., Liu, X., Hu, X., Zhang, D., & Song, S. (2022). Contrastive learning of subject-invariant eeg representations for cross-subject emotion recognition. *IEEE Transactions on Affective Computing*, *14*(3), 2496–2511.

Sun, Y., Yang, L., He, H., & Du, J. (2025). Ac-cdcn: A cross-subject eeg emotion recognition model with anti-collapse domain generalization. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*.

Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2018). Continuous convolutional neural network with 3d input for eeg-based emotion recognition. *International conference on neural information processing*, 433–443.

Yu, C., Wang, J., Chen, Y., & Huang, M. (2019). Transfer learning with dynamic adversarial adaptation network. *2019 IEEE international conference on data mining (ICDM)*, 778–786.

Zanetti, R., Aminifar, A., & Atienza, D. (2025). Eeg glasses for real-time brain electrical activity monitoring. *Scientific Reports*.

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., & Cichocki, A. (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE transactions on cybernetics*, *49*(3), 1110–1122.

Zheng, W.-L., & Lu, B.-L. (2015). 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.

Zhu, X., Liu, C., Zhao, L., & Wang, S. (2024). Eeg emotion recognition network based on attention and spatiotemporal convolution. *Sensors*, *24*(11), 3464.