ID3A: A Multi-Source Domain Adaptation Framework with Identity Disentanglement for Cross-Subject Emotion Recognition

1st Yufan Yi
School of Electronic Information and Communications
Huazhong University of Science and Technology
Wuhan, China
yufanyi@hust.edu.cn

2nd Yan Tian
School of Electronic Information and Communications
Huazhong University of Science and Technology
Wuhan, China
tianyan@hust.edu.cn

3rd Yiping Xu*
School of Electronic Information and Communications
Huazhong University of Science and Technology
Wuhan, China
xuyiping@hust.edu.cn

4th Bo Yang
School of Emergency Technology and Management
North China Institute of Science and Technology
Hebei, China
boyang@ncist.edu.cn

***Abstract*—Cross-subject EEG-based emotion recognition remains a long-standing and challenging task due to the significant domain shifts in EEG signals across individuals, which severely hinders model generalization. To address this issue, we propose ID3A, a novel identity-disentangled multi-source domain adaptation framework. First, the ID3A model performs explicit disentanglement of identity-dependent and identity-independent features via an identity disentanglement (ID) module. Then, a subject domain adaptation (SDA) module maps identity-dependent features to subject-specific representations and aligns them across domains. Meanwhile, an emotion domain adaptation (EDA) module projects identity-independent features into the emotion space and performs cross-domain alignment of emotion-related features. Finally, the fused emotion features and subject-level features are combined to perform the final emotion recognition task. ID3A achieves 73.73% / 71.81% accuracy/F1 on SEED IV and 39.93% / 39.24% on FACED, outperforming state-of-the-art methods by 0.59% / 4.40% and 1.08% / 0.89%, respectively, and demonstrating strong generalization and efficiency under complex cross-subject scenarios.**

***Index Terms*—Electroencephalogram(EEG), emotion recognition, cross-subject, domain adaptation**

I. Introduction

In emotion recognition research, signals are typically categorized into non-physiological signals, such as facial expressions, body gestures, and speech, and physiological signals, including electroencephalogram (EEG), electrocardiogram (ECG), electromyography (EMG), and heart rate. While approaches based on facial expressions, speech, and text provide intuitive cues for emotion recognition, they rely primarily on external manifestations and fail to capture individuals' internal, genuine emotional states. In contrast, EEG signals, governed by the autonomic nervous system and largely beyond voluntary control, offer a more objective and reliable modality for emotion recognition [1]–[3].

EEG-based emotion recognition has become a significant focus in brain-computer interface (BCI) research [4], [5]. However, a persistent challenge remains: models trained on EEG data from a subset of users often experience substantial performance degradation when applied to unseen subjects [6]–[8]. This performance drop is primarily due to significant inter-subject variability in EEG responses to the same emotional stimuli, resulting in domain shifts across individuals [9]–[11]. To address this issue, researchers have turned to transfer learning techniques, particularly domain adaptation methods, to improve the cross-subject generalizability of EEG-based emotion recognition models.

Domain Adaptation (DA) has been widely used to alleviate domain shift by aligning source and target distributions [12], [13]. Methods such as DANN adopt adversarial training to learn domain-invariant features. At the same time, Deep CORAL aligns second-order statistics between domains. Other approaches like ADA, JDA, and DDA further explore feature association, joint distribution alignment, and local-global discrepancy minimization to enhance cross-domain generalization [14]–[17].

Previous studies typically treated all known subjects as a single source domain. In contrast, the unseen subject was regarded as the target domain. Domain adaptation techniques have been applied in this setting, and specific performance improvements have been achieved. Unlike this unified-source approach, multi-source domain adaptation (MSDA) methods treat each subject as an independent source domain. By explicitly modeling the distributional discrepancies among

This study was supported by the National Key R&D Program of China under Grant 2020YFC0833102, the Central Government Guides Local Funds for Science and Technology Development (No. 236Z0307G) and the Fundamental Research Funds for the Central Universities (No.3142021002, 3142023021). (*Corresponding author)

multiple source domains, MSDA methods further enhance cross-subject emotion recognition performance [18]. Representative approaches achieve this by employing multiple encoders/decoders [19] or self-supervised strategies [20] to learn domain-invariant and domain-specific representations, or by dynamically reducing the discrepancy between the source and target domains through global and subdomain-level alignment [17].

Compared to earlier EEG emotion datasets with only a limited number of subjects [1], recent large-scale fine-grained EEG datasets (e.g., those with up to 123 subjects) offer greater diversity [4]. However, it is difficult to fully cover the entire spectrum of human emotional states, even on such a large scale. As the number of source domains increases (e.g., to 123), applying existing MSDA methods substantially increases both time and space complexity, making scalability a critical concern.

This paper proposes a subject-disentangled multi-source domain adaptation framework to enhance the effectiveness of multi-source domain adaptation without incurring significant increases in model complexity. Experimental results on the SEED IV [1] and FACED [4] datasets demonstrate that the proposed method outperforms current state-of-the-art domain adaptation approaches. The main contributions of this work are summarized as follows:

- The proposed framework decomposes the overall multi-source domain adaptation problem into two sub-problems: subject domain adaptation for identity features and emotion domain adaptation for emotion features, via an identity-disentanglement module.
- These sub-problems are addressed by designing a subject domain adaptation module and an emotion domain adaptation module, which jointly optimize the disentangled adaptation process.
- The proposed method improves cross-subject recognition accuracy by adapting generic identity features into the emotion feature space, demonstrating superior generalization capability across individuals.

II. Related work

Domain-Adversarial Neural Network (DANN) introduces an adversarial training mechanism to encourage the feature extractor to learn domain-invariant representations [21], [22]. Deep CORAL achieves deep domain adaptation by minimizing the covariance difference between source and target feature representations, enabling unsupervised alignment without requiring target domain labels [14]. Associative Domain Adaptation (ADA) emphasizes the construction of an association matrix between source and target samples in the feature space to promote distribution alignment and intra-class compactness [15]. To combine the advantages of DANN and ADA, the Joint Domain Adaptation (JDA) method introduces both shallow and deep constraints to reduce discrepancies in marginal and conditional distributions between domains [16]. Going further, DDA addresses global and local distributional differences by minimizing global domain discrepancy and subdomain-level misalignment, effectively tackling the problem of missing emotion labels in the target domain [17].

Although the above methods have achieved specific performance improvements, they overlook inter-subject variability within the source domain. This limitation becomes more pronounced when multiple subjects are involved in the source domain, limiting the effectiveness of these domain adaptation methods. To address this, the MSMDA method exploits the complementarity among multiple source domains by jointly optimizing transfer consistency across source-specific features [18]. Unlike MSMDA, which implicitly models inter-domain differences, DMMR explicitly incorporates subject identity information. By enforcing constraints that encourage the encoder to learn subject-invariant features, DMMR enhances generalization to unseen subjects [19]. While MSMDA and DMMR emphasize the importance of modeling source domain diversity in domain adaptation, they require a separate domain discriminator for each source domain. As the number of source domains increases, these methods suffer from significant growth in computational complexity, which hinders scalability. To overcome these limitations, we propose a more scalable and efficient multi-source adaptation framework based on identity disentanglement.

III. Methods

To enhance the generalization capability of EEG-based emotion recognition models in cross-subject scenarios, we propose an identity-disentangled multi-source domain adaptation framework, as illustrated in Fig. 1. The model consists of three main components: the identity disentanglement (ID) Module, the subject domain adaptation (SDA) Module, and the emotion domain adaptation (EDA) Module.

First, subjects are exposed to emotional stimuli through video clips, during which EEG signals are recorded using acquisition devices. The raw signals are then processed via filtering and denoising and fed into a handcrafted feature extractor. The ID module further processes the extracted features to obtain effective feature representations. In the second stage, the extracted initial deep learning features are disentangled and optimized along two dimensions, subject-invariant representation and emotion-invariant representation, to improve the robustness and transferability of EEG features.

A. Identity-Disentanglement Module

The identity disentanglement module mainly consists of a handcrafted feature extractor and a primary encoder.

1) Handcrafted Feature Extractor: The computation process of the handcrafted feature extractor is as follows: In the data preprocessing stage, this paper adopts commonly used feature extraction methods for the datasets. First, a 0-45 Hz band-pass filter is applied to the raw EEG signals to remove noise and facilitate feature extraction. Then, handcrafted features are computed for five frequency bands. FACED [4] and SEED IV [1] datasets primarily use differential entropy (DE) [23] features. Shannon information entropy formula as entered:

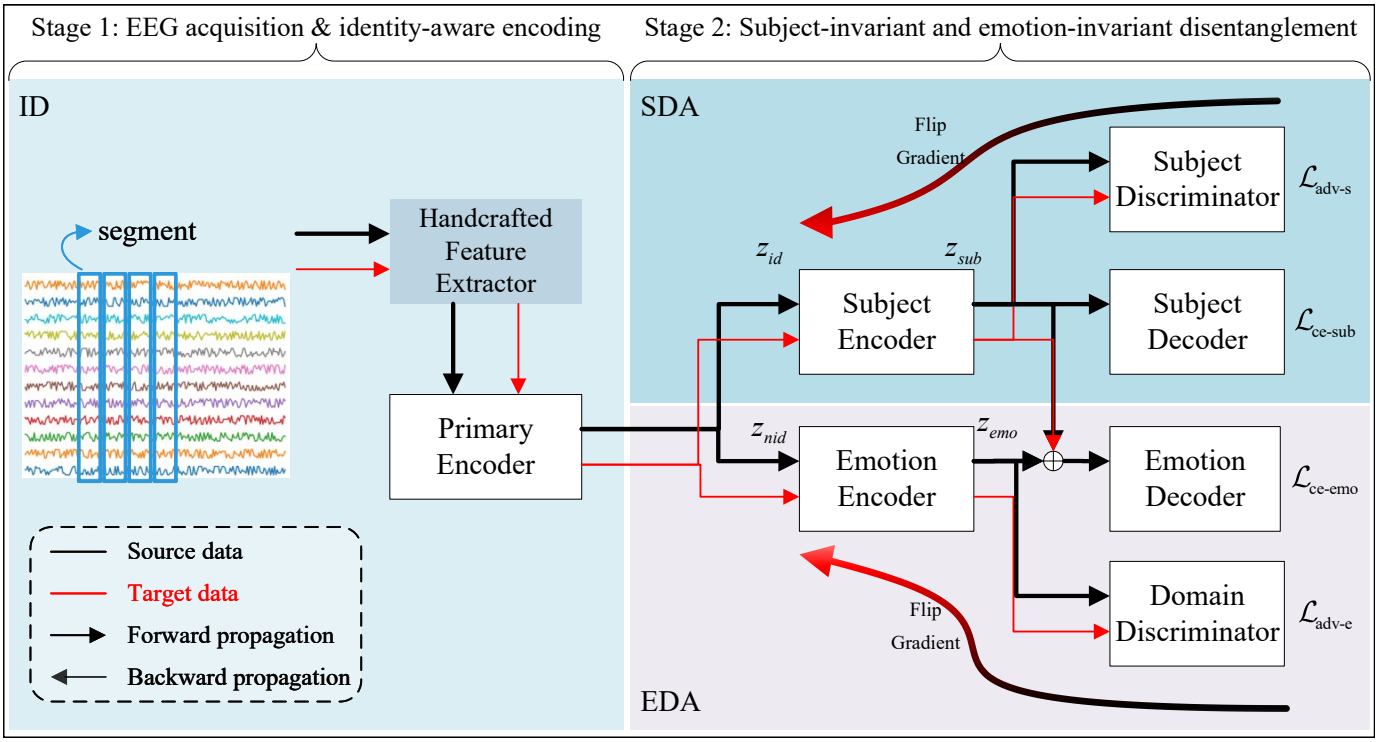

Fig. 1. Multi-Source Domain Adaptation Framework with Identity Disentanglement for Cross-Subject Emotion Recognition

$$DE = -\sum_{x} p(x)\log(p(x)). \tag{1}$$

Differential entropy (DE) [23] is a generalized form of Shannon's information entropy for continuous variables:

$$DE = -\int_{a}^{b} p(x)\log(p(x))dx, \tag{2}$$

where $p(x)$ denotes the probability density function of continuous information, and [a, b] denotes the interval of information values. The differential entropy of an EEG of a given length that approximately follows a Gaussian distribution $N\left(\mu, \sigma_i^2\right)$ is equal to the logarithm of its energy spectrum in a given frequency band:

$$\begin{aligned} DE &= -\int_{a}^{b} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu)^2}{2\sigma_i^2}}\right) dx \\ &= \frac{1}{2}\log\left(2\pi e\sigma_i^2\right). \end{aligned} \tag{3}$$

The resulting feature representation is denoted as $\mathbf{x} \in \mathbb{R}^{C\times F}$, where C represents the number of EEG channels (30 or 62), and F denotes the dimensionality of the handcrafted features, which is set to 5 by default.

2) Primary Encoder: We design a feature encoder using a multilayer perceptron (MLP) to extract discriminative high-level representations from the raw input. The encoder consists of three fully connected (linear) layers, with ReLU activation functions inserted between each layer to enhance the nonlinearity of the network. Let the input feature vector be $\mathbf{x} \in \mathbb{R}^{310}$, which is typically formed by concatenating features from multiple EEG channels and their corresponding frequency bands. The forward propagation of the network is as follows:

$$\begin{aligned} \mathbf{h}_1 &= \mathrm{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1), \quad \mathbf{W}_1 \in \mathbb{R}^{128\times 310},\ \mathbf{b}_1 \in \mathbb{R}^{128}, \\ \mathbf{h}_2 &= \mathrm{ReLU}(\mathbf{W}_2\mathbf{h}_1 + \mathbf{b}_2), \quad \mathbf{W}_2 \in \mathbb{R}^{64\times 128},\ \mathbf{b}_2 \in \mathbb{R}^{64}, \\ \mathbf{h}_3 &= \mathrm{ReLU}(\mathbf{W}_3\mathbf{h}_2 + \mathbf{b}_3), \quad \mathbf{W}_3 \in \mathbb{R}^{32\times 64},\ \mathbf{b}_3 \in \mathbb{R}^{32}. \end{aligned} \tag{4}$$

The final representation $\mathbf{h}_3 \in \mathbb{R}^{32}$ is used as the encoded embedding of the original EEG input. The use of ReLU activation in all hidden layers enhances the non-linearity of the encoder and helps mitigate the vanishing gradient problem.

This hierarchical transformation reduces the input dimensionality from 310 to 32 while preserving critical information relevant to emotion-related neural activity. The resulting latent vector $\mathbf{h}_3$ serves as the input to downstream modules such as classifiers, domain discriminators, or projection heads, depending on the task.

Let the output of the primary EEG feature encoder be $\mathbf{h}_3$, where $\mathbf{h}_3 = \mathbf{z}_{\text{id}} + \mathbf{z}_{\text{nid}}$. We then extract two types of representations using two separate encoders: The subject-related representation: $\mathbf{z}_{\text{sub}} = E_{\text{sub}}(\mathbf{z}_{\text{id}})$, and The subject-independent, emotion-related representation: $\mathbf{z}_{\text{emo}} = E_{\text{emo}}(\mathbf{z}_{\text{nid}})$.

B. Subject Domain Adaptation (SDA)

The SDA module is designed to extract subject-related information from the input features while making the identity

features from the source and target domains indistinguishable through an adversarial process. This module comprises a subject encoder, a subject decoder, and a discriminator.

The Subject Encoder E_{sub} is responsible for extracting subject-specific latent features $\mathbf{z}_{\text{sub}} = E_{\text{sub}}(\mathbf{z}_{\text{id}})$, where $\mathbf{z}_{\text{id}} \in \mathbb{R}^d$ is the input EEG feature vector. To ensure that $\mathbf{z}_{\text{sub}}$ effectively captures individual identity information, a Subject Decoder D_{sub} is employed to perform subject classification.

Specifically, the decoder predicts the subject label $\hat{y}_{\text{sub}} = D_{\text{sub}}(\mathbf{z}_{\text{sub}})$, and the network is optimized using the subject classification loss defined as the standard cross-entropy loss:

$$\mathcal{L}_{\text{ce-sub}} = \text{CE}(\hat{y}_{\text{sub}}, y_{\text{sub}}), \tag{5}$$

where y_{sub} denotes the ground-truth subject identity. This loss encourages the subject encoder to preserve identity-discriminative information in the latent space, which can later be disentangled from emotion-related features.

The Subject Discriminator $D_{\text{dom-s}}$ attempts to predict the subject identity domain y_{d_s} from $\mathbf{z}_{\text{sub}}$, and influences the Subject Encoder E_{sub} via a gradient reversal layer (GRL), forcing it to learn features that are uninformative for subject classification.

$$\mathcal{L}_{\text{adv-s}} = \text{CE}(D_{\text{dom-s}}(\mathbf{z}_{\text{sub}}), y_{d_s}). \tag{6}$$

C. Emotion Domain Adaptation (EDA)

The EDA module extracts emotion-related but domain-invariant features from the input. Like the SDA module, the EDA module consists of an emotion encoder, an emotion decoder, and a domain discriminator. The Emotion Encoder E_{emo} extracts emotion-related features $\mathbf{z}_{\text{emo}}$.

The final feature used for emotion recognition is given by: $\mathbf{z}'_{\text{emo}} = \mathbf{z}_{\text{emo}} \oplus \alpha \cdot \mathbf{z}_{\text{sub}}$, where α is a weighting coefficient for the subject-related representation, and $\oplus$ denotes element-wise combination.

In contrast, the Emotion Decoder D_{emo} imposes constraints on $\mathbf{z}'_{\text{emo}}$ through a cross-entropy loss for emotion classification:

$$\mathcal{L}_{\text{ce-emo}} = \text{CE}(D_{\text{emo}}(\mathbf{z}'_{\text{emo}}), y_e). \tag{7}$$

Domain Discriminator $D_{\text{dom-e}}$: attempts to distinguish the domain or subject domain (e.g., source vs. target) of the emotion-related representation $\mathbf{z}_{\text{emo}}$, and is trained in an adversarial manner. Again, a GRL is placed before $D_{\text{dom-e}}$ to enforce feature invariance, the domain adversarial loss is defined as:

$$\mathcal{L}_{\text{adv-e}} = \text{CE}(D_{\text{dom-e}}(\mathbf{z}_{\text{emo}}), y_{d_e}). \tag{8}$$

D. Final Objective and Joint Training

The full training objective integrates both reconstruction and adversarial losses:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ce-sub}} + \mathcal{L}_{\text{ce-emo}} + \lambda_1 \mathcal{L}_{\text{adv-s}} + \lambda_2 \mathcal{L}_{\text{adv-e}}, \tag{9}$$

where λ_1 and λ_2 are hyperparameters balancing reconstruction and adversarial terms.

IV. Datasets and Evaluation Metrics

A. Datasets

We conduct experiments on the SEED IV [1] dataset, which contains EEG recordings from 15 subjects using 62-channel electrodes. The EEG signals were recorded at a sampling rate of 200 Hz. Each subject participated in three independent sessions, and during each session, they were presented with 24 emotion-eliciting video clips. The target emotional states include neutral, sad, fear, and happy, resulting in four distinct emotion categories.

FACED [4] contains EEG signals recorded from 123 subjects with 30 EEG channels at a sampling rate of 128 Hz. During the experiment, the subjects watched 28 emotion-eliciting video clips covering nine emotion categories: anger, disgust, fear, sadness, neutral, amusement, inspiration, joy, and tenderness.

B. Evaluation Protocols and Metrics

Evaluation Protocols: We employ the leave-one-subject-out (**LOSO**) cross-validation strategy for the SEED IV dataset [1], where each subject is used once as the test set. In contrast, the remaining subjects are used for training. This process is repeated for all subjects. The results of this experiment are shown in Table I.

Unlike the above, FACED [4], the EEG data of the subjects were divided into 10 folds (12 in the first nine folds and 15 in the 10-th fold). Nine folds are used as the training set, and the remaining fold are used as the test set. The results are shown in Table II.

The evaluation metrics include Accuracy, Cohen's Kappa, and F1 score, both reported as the mean and standard deviation over multiple runs.

C. Comparison Methods

We compare our approach with several representative domain adaptation methods, including DDC [12], DANN [21], CORAL [14], ADA [15], JDA [16], MSMDA [18], DDA [17], and DMMR [19]. For fair comparison, the same network architecture as the primary encoder in our method is used as the feature extractor in DDC, DANN, CORAL, ADA, and JDA. All models are retrained on each dataset separately.

D. Implementation Details

All experiments were conducted on a workstation running Ubuntu 20.04, equipped with an Intel Core i7-7700K CPU, 16 GB of RAM, and an NVIDIA RTX 3090 Ti GPU with 24 GB of memory. The CUDA and cuDNN versions used were 11.6 and 8.0.5, respectively. The model was implemented using PyTorch 2.3.1 in a Python 3.8.0 environment. To ensure training stability, all experiments were conducted with fixed random seeds and repeated multiple times, reporting the averaged results. The training employed the Adam optimizer with a learning rate of 1×10^{-3}, a weight decay of 5×10^{-4}, batch sizes ranging from 128 to 512, and total epochs between 100 and 200.

V. Experiments

To evaluate the effectiveness and robustness of our proposed model, we conduct extensive experiments on two benchmark EEG emotion datasets: **SEED IV** and **FACED**. The experimental section is structured into two major parts:

Comparison with state-of-the-art methods: We benchmark our approach against various classical and recent domain adaptation models under the cross-subject setting.

Ablation Study: To comprehensively assess the effectiveness of our proposed method, we design four groups of ablation studies covering module contribution, feature distribution, model complexity, and fusion weight sensitivity.

A. Comparison with state-of-the-art methods

1) Benchmarking Against State-of-the-Art on SEED IV: To assess the generalization performance of the proposed method under subject-independent conditions, we conduct experiments following the LOSO protocol on the SEED IV dataset. The results are summarized in Table I.

TABLE I
PERFORMANCE COMPARISON UNDER LOSO SETTING ON SEED IV DATASET (MEAN ± STD, %). RESULTS MARKED WITH * ARE CITED FROM ORIGINAL PAPERS. **BOLD** INDICATES BEST, UNDERLINE INDICATES SECOND BEST.

Method	Accuracy	F1 score	Cohen's Kappa
DDC [12]	66.89±9.51	65.64±10.11	55.19±12.80
DANN [21]	66.90±11.08	64.17±13.25	55.32±14.77
CORAL [14]	67.97±11.79	66.28±13.11	56.76±15.88
ADA [15]	66.10±9.79	64.49±10.78	54.06±13.34
JDA [16]	65.40±10.25	63.97±11.36	53.04±13.94
MSMDA [18]	59.34±5.48*	59.70±13.74	46.93±18.39
DDA [17]	73.14±9.43*	67.41±12.44	56.94±16.47
DMMR [19]	72.70±8.01*	66.50±14.05	57.94±13.90
ID3A	**73.73±11.04**	**71.81±12.65**	**64.55±14.63**

As shown in Table I, our method achieves the best performance across all evaluation metrics. Specifically, we obtain an accuracy of 73.73%, an F1 score of 71.81%, and a Cohen's Kappa of 64.55%, outperforming the strongest baseline, DDA, in accuracy and robustness. Compared to the current state-of-the-art domain adaptation method, DDA, our approach improves accuracy, F1 score, and Cohen's Kappa by 0.59%, 4.4%, and 7.61%, respectively.

The improvement can be attributed to our dual-branch adversarial architecture, which explicitly disentangles subject-specific and emotion-relevant components. In contrast, traditional methods such as JDA and CORAL rely on global distribution alignment, which cannot handle individual differences. Adversarial methods like DANN and DDC also struggle due to the absence of explicit feature separation. Our model demonstrates superior cross-subject generalization in EEG emotion recognition tasks. It highlights its practical value in real-world scenarios without subject-specific calibration.

2) Benchmarking Against State-of-the-Art on FACED: To further evaluate the cross-subject transferability of our proposed ID3A model in complex real-world scenarios, we conduct experiments on the FACED dataset following the official cross-subject training protocol. Table II presents a performance comparison between our method and various classic and state-of-the-art domain adaptation models, including DDC, DANN, CORAL, JDA, DDA, ADA, MSMDA, and DMMR. The evaluation metrics include Accuracy, F1 score, and Cohen's Kappa.

TABLE II
PERFORMANCE COMPARISON UNDER CROSS SUBJECT SETTING ON FACED DATASET (MEAN ± STD, %). **BOLD** INDICATES BEST, UNDERLINE INDICATES SECOND BEST.

Method	Accuracy	F1 score	Cohen's Kappa
DDC [12]	36.54±4.10	36.39±3.98	28.51±4.62
DANN [21]	38.53±3.57	38.03±3.43	30.73±4.00
CORAL [14]	38.06±2.86	37.47±2.74	30.22±3.22
ADA [15]	33.55±3.51	32.92±3.07	25.20±3.96
JDA [16]	33.44±2.60	32.71±2.19	24.96±2.87
MSMDA [18]	37.55±3.28	33.79±3.25	26.48±3.79
DDA [17]	38.23±4.00	37.82±3.95	30.43±4.50
DMMR [19]	38.85±3.22	38.35±2.96	31.13±3.60
ID3A	**39.93±2.93**	**39.24±2.93**	**32.30±3.33**

The Table II shows that our proposed ID3A model consistently performs best on the FACED dataset, outperforming all existing mainstream methods across all three evaluation metrics. Compared to the state-of-the-art baseline model, DMMR, our approach achieves improvements of 1.08, 0.89, and 1.17 in Accuracy, F1 score, and Cohen's Kappa, respectively.

Moreover, traditional transfer methods such as JDA and CORAL exhibit inferior performance on this dataset. In contrast, ID3A leverages identity disentanglement and emotion domain alignment to filter out subject-specific variations and structurally transfer emotion-related features, ultimately achieving more stable and reliable cross-domain recognition.

B. Ablation Study

To further validate the effectiveness of the proposed method, we conduct four groups of ablation studies:

1) **Component-Wise Ablation**: To examine the contribution of each individual component within the framework.
2) **Feature Visualization Comparison**: To compare with state-of-the-art domain adaptation methods using feature distribution visualization, providing intuitive insights into how our model components shape the learned representations.
3) **Model Complexity Comparison**: To evaluate the scalability of our approach against recent multi-source domain adaptation methods, especially under scenarios with increased source domain complexity.
4) **Sensitivity Analysis of Fusion Weights**: To investigate the impact of different fusion weight settings on the emotion recognition task, highlighting that incorporating identity-related features can further enhance recognition performance to some extent.

1) Component-Wise Ablation: To investigate the contribution of each component in our proposed framework, we perform ablation studies by systematically removing the **Subject Domain Adaptation (SDA)** module and the **Emotion Domain Adaptation (EDA)** module, respectively. Table III shows the performance degradation after removing each module under the LOSO protocol on the SEED IV dataset.

TABLE III
ABLATION STUDY ON THE SEED IV DATASET (MEAN ± STD, %). **BOLD** INDICATES BEST.

Method	Accuracy	F1 score	Cohen's Kappa
ID3A	**73.73±11.04**	**71.81±12.65**	**64.55±14.63**
w/o EDA	70.36±10.59	67.79±11.98	59.90±14.08
w/o SDA	72.05±10.86	70.46±12.31	62.12±14.82

The complete ID3A model achieves the best performance across all three evaluation metrics, with an Accuracy of 73.73±11.04, an F1 score of 71.81±12.65, and Cohen's Kappa of 64.55±14.63. When the EDA module is removed, accuracy drops to 70.36±10.59, while F1 score and Kappa decrease by 4.02 and 4.65 points, respectively. These results highlight the critical role of EDA in enhancing the domain invariance of emotion-related features.

By comparison, removing the SDA module leads to a slightly smaller yet still significant performance degradation, with Cohen's Kappa dropping by more than 2 points. This suggests that SDA plays a beneficial role in mitigating subject variability and improving cross-subject consistency.

Overall, the experimental results validate the synergistic effect of the two modules in our proposed identity-disentangled multi-source domain adaptation framework. Specifically, the EDA module performs domain alignment to enhance the ability to generalize emotion-related features. In contrast, the SDA module improves cross-subject robustness by disentangling subject-specific information. These two components contribute to the model's superior performance in multi-subject EEG-based emotion recognition tasks.

2) Feature Visualization of Domain Adaptation Variants: To demonstrate that the proposed multi-domain adaptation model can effectively extract identity-independent emotional features, we perform feature visualization using t-SNE [24]. The original 32-dimensional features are first reduced to two dimensions using PCA. The visualized features include subject-wise and emotion-wise distributions. In the subject distribution visualization, different shapes represent different emotions. In contrast, different colors indicate different subjects, with red denoting the target domain. Different shapes denote emotions in the emotion distribution visualization, with blue representing source domain data and red indicating target domain data.

As shown in Fig. 2, we can find that:

- **Subject-wise Feature Distribution Comparison**: There is a large discrepancy in identity information across multiple source domains in the raw data. The model trained solely on source domains fails to achieve overlap between subject-related features of the source and target domains. After applying the DDA domain adaptation method, the discrepancy between source and target domains is reduced; however, the features become overly concentrated and lose separability. In contrast, the proposed method enables the target domain's subject-related features to be evenly distributed across different source domains.
- **Emotion-wise Feature Distribution Comparison**: The raw emotion features show significant differences between the source and target domains, and the emotional categories are not separable. After training on source domain data, the source emotion features exhibit clear class boundaries. However, the target domain features still deviate significantly from the source distribution. The DDA method somewhat reduces this discrepancy, but some target features remain misaligned. In comparison, the proposed ID3A method disperses the target domain features across the global emotional feature space, achieving the best alignment between source and target features with minimal discrepancy.

  These results fully validate the efficiency of our method in terms of architectural design. Unlike DMMR, which requires separate encoders and decoders for each source domain, our framework adopts a unified and parameter-efficient encoder-decoder structure. This significantly reduces model complexity and inference overhead and maintains competitive performance in cross-subject emotion recognition tasks.

3) Model Complexity and Computational Efficiency Analysis: To further evaluate the computational efficiency of our proposed method, we compare its model complexity against DMMR [19] under both SEED IV and FACED datasets. As shown in Table IV, the comparison includes total parameter count, floating-point operations (FLOPs), parameter memory footprint, and estimated total memory cost.

As shown in Table IV, our method demonstrates a significant advantage in model efficiency across all evaluation metrics. For example:

- On the **SEED IV** dataset, the pretraining stage of DMMR involves over 1.8 million parameters and consumes approximately 31.82M FLOPs. In contrast, our method achieves comparable recognition performance with only 35.9K parameters and 0.03M FLOPs. Notably, our model's total memory footprint on SEED IV is just 0.13 MB, representing a 67× reduction compared to DMMR's 8.69 MB.
- On the **FACED** dataset, the model complexity of DMMR increases significantly. The complete DMMR model (including pretraining and fine-tuning stages) contains over 8 million parameters, requires approximately 146.33M FLOPs, and occupies a total size of 38.33 MB. In contrast, due to the reduced number of EEG electrodes in FACED (32 channels versus 62 in SEED IV), the parameter count of our primary encoder decreases sub-

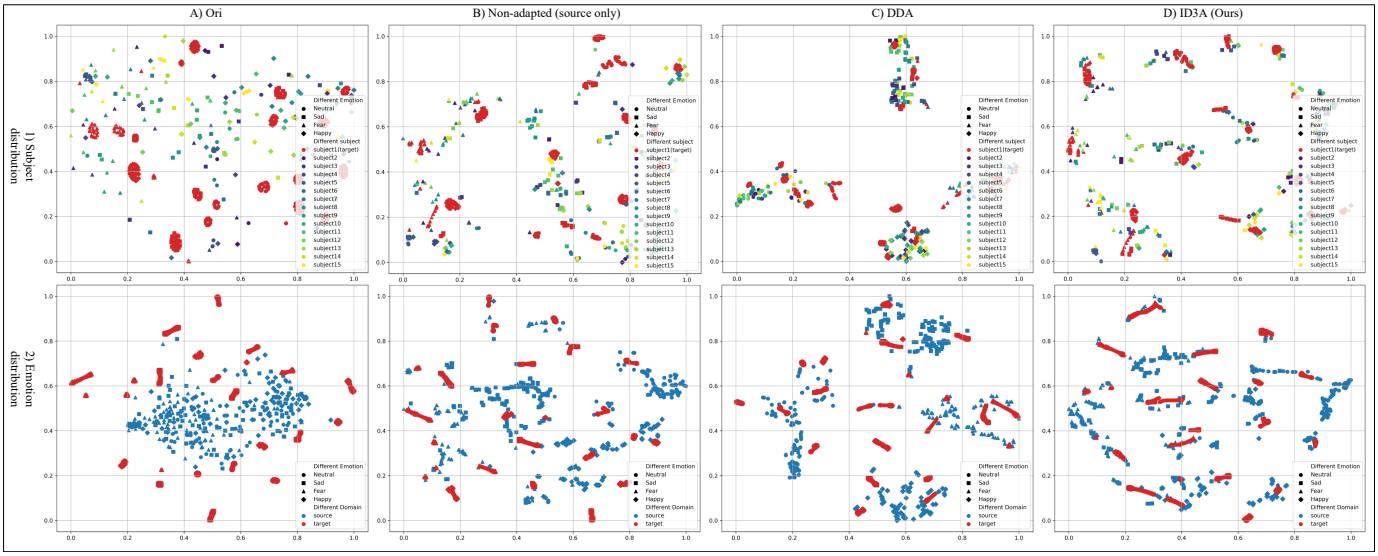

Fig. 2. Visualization of feature distributions under different multi-source domain adaptation methods. Red indicates target domain data, while other colors represent source domain data. The visualizations focus on: 1) subject-wise feature distribution, and 2) emotion-wise feature distribution. The compared methods include: a) Original features, b) Feature representations trained using only source domain data, c) Features of the DDA method, and d) Features of the proposed ID3A method.

TABLE IV
MODEL COMPLEXITY COMPARISON ON SEED IV AND FACED DATASETS.

Dataset	Module	#Params	FLOPs (M)	Params Size (MB)	Total Size (MB)
SEED IV (# Domains =15)	DMMR (PreTrain)	1823260	31.82	7.29	8.69
	DMMR (FineTune)	82755	0.56	0.33	0.36
	Ours	54391	0.05	0.22	0.22
FACED (# Domains =123)	DMMR (PreTrain)	8014468	146.33	32.06	38.33
	DMMR (FineTune)	197150	0.97	0.79	0.83
	Ours	35895	0.04	0.14	0.15

stantially. Moreover, our domain-adversarial module does not introduce additional computational overhead. As a result, our method becomes even more lightweight than in the SEED IV setting. Specifically, it contains only 35.9K parameters, requires 0.04M FLOPs, and has a total memory footprint of just 0.15 MB. Overall, this represents a 223× reduction in parameter count and a 3658× reduction in computational cost compared to DMMR, significantly reducing training and deployment overhead.

4) Sensitivity Analysis of Fusion Weights α: To investigate the balance between emotion-related and identity-related features, we conduct experiments on the SEED IV dataset by varying the fusion weight α between these features, ranging from 0.00 to 1.00. This hyperparameter controls the degree to which identity information complements emotion representations during training, and it plays a critical role in determining the model's generalization ability across individuals in EEG-based emotion recognition.

As shown in Fig. 3, as the fusion weight α increases from 0 to 1.0, the model exhibits a steady improvement across all three evaluation metrics: Accuracy, F1 Score, and Cohen's Kappa, with the best performance observed around $\alpha = 0.07$.

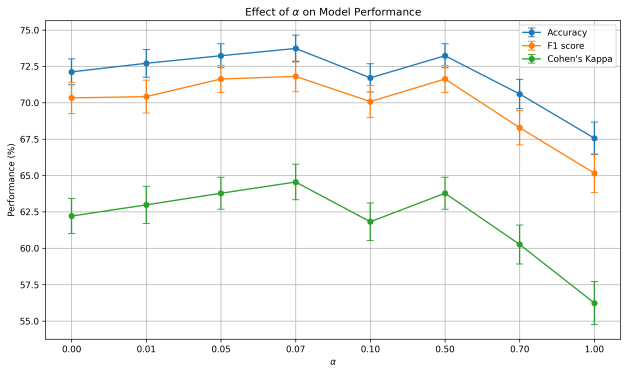

Fig. 3. Performance comparison on the SEED IV dataset under different fusion weight α. A peak is observed around $\alpha = 0.07$, suggesting a proper trade-off between emotion-related and identity-related features.

However, further increasing the weight leads to a performance decline, indicating that excessive identity information may undermine the model's generalization ability. These results suggest that moderate identity-aware supervision can enhance the model's discriminative power while effectively preventing overfitting. Therefore, we set $\alpha = 0.07$ for all subsequent

experiments. Overall, this experiment provides complementary evidence, from identity and emotion representation perspectives, for the model's strong cross-subject generalization capability, surpassing that of mainstream domain adaptation methods.

VI. Conclusions

This paper proposes a multi-source domain adaptation framework for cross-subject EEG-based emotion recognition, which explicitly disentangles identity-related and emotion-related features. By introducing the subject domain adaptation (SDA) and emotion domain adaptation (EDA) modules, the proposed method effectively enhances generalization under subject variability while preserving strong discriminative power. Extensive experiments on two benchmark datasets, SEED IV and FACED, demonstrate that our approach outperforms state-of-the-art domain adaptation baselines and achieves superior efficiency in model complexity and computation. Ablation studies and feature visualizations further confirm the critical roles of each module. Overall, the proposed method provides a novel and efficient solution for cross-subject emotion recognition, especially suitable for resource-constrained or subject-diverse brain–computer interface (BCI) scenarios.

In future work, we plan to explore few-shot and continual learning strategies for more adaptive EEG representation learning.

References

[1] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE Transactions on Cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[2] Y. Yi, Y. Tian, C. He, Y. Fan, X. Hu, and Y. Xu, "Dbt: multimodal emotion recognition based on dual-branch transformer," *The Journal of Supercomputing*, vol. 79, no. 8, pp. 8611–8633, 2023.

[3] Y. Yi, Y. Xu, Z. Ye, L. Li, X. Hu, and Y. Tian, "Stan: spatiotemporal attention network for video-based facial expression recognition," *The Visual Computer*, vol. 39, no. 12, pp. 6205–6220, 2023.

[4] J. Chen, X. Wang, C. Huang, X. Hu, X. Shen, and D. Zhang, "A large finer-grained affective computing eeg dataset," *Scientific Data*, vol. 10, no. 1, p. 740, 2023.

[5] Y. Yi, Y. Xu, B. Yang, and Y. Tian, "A weighted co-training framework for emotion recognition based on eeg data generation using frequency-spatial diffusion transformer," *IEEE Transactions on Affective Computing*, vol. 15, no. 4, pp. 2055–2069, 2024.

[6] T. Zhang, Z. Cui, C. Xu, W. Zheng, and J. Yang, "Variational pathway reasoning for eeg emotion recognition," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 3, pp. 2709–2716, 2020.

[7] R. Zhou, Z. Zhang, H. Fu, L. Zhang, L. Li, G. Huang, F. Li, X. Yang, Y. Dong, Y. Zhang, and Z. Liang, "Pr-pl: A novel prototypical representation based pairwise learning framework for emotion recognition using eeg signals," *IEEE Transactions on Affective Computing*, vol. 15, no. 2, pp. 657–670, 2023.

[8] T. Pan, Y. Ye, H. Cai, S. Huang, Y. Yang, and G. Wang, "Multimodal physiological signals fusion for online emotion recognition," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 5879–5888.

[9] Y. Li, Z. Liu, L. Yao, J. J. Monaghan, and D. McAlpine, "Disentangled and side-aware unsupervised domain adaptation for cross-dataset subjective tinnitus diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 538–549, 2022.

[10] Y. Shen and Y. Lin, "Cross-day data diversity improves inter-individual emotion commonality of spatio-spectral eeg signatures using independent component analysis," *IEEE Transactions on Affective Computing*, 2023.

[11] X. Shen, L. Tao, X. Chen, S. Song, Q. Liu, and D. Zhang, "Contrastive learning of shared spatiotemporal EEG representations across individuals for naturalistic neuroscience," *NeuroImage*, vol. 301, p. 120890, Nov. 2024.

[12] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," 2014.

[13] W. Li, L. Fan, S. Shao, and A. Song, "Generalized contrastive partial label learning for cross-subject eeg-based emotion recognition," *IEEE Transactions on Instrumentation and Measurement*, 2024.

[14] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 443–450.

[15] P. Häusser, T. Frerix, A. Mordvintsev, and D. Cremers, "Associative domain adaptation," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*. IEEE Computer Society, 2017, pp. 2784–2792. [Online]. Available: https://doi.org/10.1109/ICCV.2017.301

[16] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for eeg emotion recognition based on latent representation similarity," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 2, pp. 344–353, 2019.

[17] Z. Li, E. Zhu, M. Jin, C. Fan, H. He, T. Cai, and J. Li, "Dynamic domain adaptation for class-aware cross-subject and cross-session eeg emotion recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 5964–5973, 2022.

[18] H. Chen, M. Jin, Z. Li, C. Fan, J. Li, and H. He, "Ms-mda: Multisource marginal distribution adaptation for cross-subject and cross-session eeg emotion recognition," *Frontiers in Neuroscience*, vol. 15, 2021.

[19] Y. Wang, B. Zhang, and Y. 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*, vol. 38, no. 1, 2024, pp. 628–636.

[20] G. Zhang, V. Davoodnia, and A. Etemad, "Parse: Pairwise alignment of representations in semi-supervised eeg learning for emotion recognition," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2185–2200, 2022.

[21] Y. Ganin and V. S. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, ser. JMLR Workshop and Conference Proceedings, F. R. Bach and D. M. Blei, Eds., vol. 37. JMLR.org, 2015, pp. 1180–1189. [Online]. Available: http://proceedings.mlr.press/v37/ganin15.html

[22] Y. Yi, Y. Xu, X. Hu, and Y. Tian, "Multi-feature fusion graph convolutional network based on domain adaptation for EEG emotion recognition," in *Frontiers in Artificial Intelligence and Applications*. IOS Press, Mar. 2024.

[23] R. Duan, J. Zhu, and B. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[24] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.