Generalization Across Subjects and Sessions for EEG-based Emotion Recognition Using Multi-source Attention-based Dynamic Residual Transfer

Wanqing Jiang
University of Chinese Academy of Sciences, Beijing, China
jiangwanqing2020@ia.ac.cn

Gaofeng Meng
Institute of Automation, Chinese Academy of Sciences, Beijing, China
gfmeng@nlpr.ia.ac.cn

Tianzi Jiang
Brainnetome Center & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
jiangtz@nlpr.ia.ac.cn

Nianming Zuo
Brainnetome Center & NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
nmzuo@nlpr.ia.ac.cn

***Abstract*—As an important element of emotional brain-computer interfaces, electroencephalography (EEG) signals have made significant progress in emotion recognition due to their high temporal resolution and reliability. However, EEG signals vary widely among individuals and do not satisfy temporal non-stationarity. Furthermore, trained models cannot maintain good classification accuracy for new individuals or new sessions during the inference stage. Although domain adaptation has been employed to address these issues, most approaches that consider different subjects or sessions as a single source domain ignore the large discrepancies between source domains, while methods that consider multi-source domains need to construct a domain adaptation branch for each source domain. Here, we propose a novel emotion recognition method, i.e., multi-source attention-based dynamic residual transfer (MS-ADRT). We introduce a dynamic feature extractor, in which the model uses an attention module to induce parameters to vary with the sample, implicitly enabling multi-source domain adaptation by adapting to the sample, thus reducing multi-source domain adaptation to single-source domain adaptation. Maximum mean discrepancy (MMD) and maximum classifier discrepancy (MCD)-based adversarial training are also used to narrow distances between source and target domains and facilitate the feature extractor to mine domain-invariant and sentiment-distinguishable features. We compared our algorithm with representative methods using the SEED and SEED-IV datasets, and experimentally verified that our method outperforms other state-of-the-art approaches. The proposed method provides a more effective transfer learning pathway for EEG-based sentiment analysis under multi-source scenarios.**

***Index Terms*—Electroencephalogram (EEG), emotion recognition, multi-source domain adaptation, subject-independent**

I. INTRODUCTION

Emotions influence interpersonal interactions and decision-making in humans and play an important role in many neurological and cognitive sciences, especially in the diagnosis of psychiatric disorders. Many studies have confirmed the relationship between psychiatric disorders and emotional state [1]. Emotional brain-computer interfaces [2] detect the user's emotional state through spontaneous electroencephalographic (EEG) signals, thereby enriching the user's experience during interaction. Compared to behavioral signals, such as vocalizations, facial expressions, gestures, and body postures, which are easily masked in emotion recognition, physiological EEG signals are difficult to disguise and exhibit high temporal resolution. EEG signals can reliably identify human emotions and direct recognition can be achieved by analyzing the immediate brain activity induced by emotional stimuli [1]. Fluctuations in EEG signals can directly reflect changes in emotional state in humans. Extensive studies have already been conducted using EEG signals to identify emotional states. Furthermore, EEG-based emotion recognition has made significant progress due to its high accuracy and reliability [1, 3].

Due to the non-stationarity and large individual discrepancies of EEG signals [4], emotion classification can be poorly extrapolated to new individuals or time points. In the early days of EEG emotion recognition, most studies used support vector machines (SVM) [5] and linear discriminant analysis (LDA) [6]. The premise of traditional machine learning is that training and test data are independent and identically distributed. However, this is not the case for EEG data, leading to the wide use of domain adaptation in subsequent research. There are two types of transfer tasks in EEG emotion recognition, i.e., cross-subject (transfer from training subjects to target subjects) and cross-session (transfer of the same subject in a different session). From the perspective of feature transfer methods [7, 8], we divided domain adaptation methods into three categories. The first is marginal distribution adaptation (MDA), which aims to reduce the distance between the marginal probability distributions of the source and target domains. Transfer component analysis is a representative MDA method [9], which utilizes maximum mean discrepancy (MMD) to project the source and target domains into reproducing kernel Hilbert space [10]. The deep adaptation network (DAN) [11] replaces MMD with multi-

kernel MMD (MK-MMD), adapting the last three layers. In addition, domain adversarial neural networks (DANNs) [12] introduce adversarial training strategies to enable shallow networks to learn domain-invariant features to achieve certain improvements. The second is conditional distribution adaptation (CDA), which aims to narrow the distance between the conditional probability distributions of the source and target domains by considering category information. Maximum classifier discrepancy (MCD) leverages class-specific decision boundaries to align source and target domain distributions to obtain domain-invariant and emotion-discriminative features to mitigate domain shift [13]. The third is joint distribution adaptation (JDA), which combines the advantages of the above two approaches to reduce the distance of the joint probability distribution of the source and target domains. Li et al. [14] considered the functional differences between shallow and deep network layers and applied adversarial training to adapt to marginal distribution for task-independent shallow features and associative domain adaptation to adapt to conditional distribution for task-related deep features.

However, most previous studies on EEG-based emotion recognition have not considered the disparities in marginal distributions between different source data, instead opting for a simple concatenation approach. Classification performance of each source domain is degraded due to the distributional discrepancies between multi-source domains that are not considered when using single source domain adaptation [15]. This problem can be solved by multi-source domain adaptation [16], where sources contain multiple domains. Current multi-source domain adaptation algorithms assume that N source domains are independent and use domain-specific feature extractors and classifiers for each source domain. Chen et al. [17] proposed utilizing a unique feature extractor-classifier pair for each source-target domain pair, then averaging the classifier predictions as the target result. Alternative strategies use a small number of labeled target samples to select the closest source domains [14]. While multiple domain-specific networks can address the multi-source domain issue, it can lead to a linear increase in network parameters and training time as source domains increase. Hence, more flexible multi-source domain adaptation methods are still required.

Domain-specific feature extractors require only one set of networks with parameters that change with the domain, without repeated network construction. Inspired by the dynamic residual transfer (DRT) approach [15], we incorporated parallel dynamic convolutional residual blocks into the feature extractor, requiring only a single set of networks whose parameters vary dynamically with the samples. Adjusting the model according to the domain can be achieved by tuning the model for each sample, as each domain is considered a distribution of samples. Squeeze-and-excitation techniques have a wide range of applications in computer vision and need only a small amount of computation to achieve effective improvements [18]. Here, we used an attention module based on squeeze-and-excitation to make the feature extractor vary with each sample. We then used MMD to align marginal distribution and MCD to align conditional distribution. We named the proposed method Multi-Source Attention-Based Dynamic Residual Transfer (MS-ADRT). We conducted extensive experiments and demonstrated that the

proposed multi-source emotion recognition architecture can significantly improve EEG-based emotion recognition.

This paper is organized as follows: A detailed description of the proposed MS-ADRT, including the design of the dynamic extractor and choice of domain alignment algorithms, is presented in section 2. Descriptions of the EEG datasets and experimental details are provided in section 3. Multi-view algorithm validation results are presented and discussed in section 4. The main conclusions of our research are provided in section 5.

II. METHODS

A. General Structure of Proposed Network

As outlined in Fig. 1, the proposed MS-ADRT model is composed of three parts, i.e., attention-based dynamic residual feature extractor, emotion classifiers, and domain adaptation. The feature extractor introduces dynamic perception to focus on domain transfer among source domains. In order to transfer the knowledge learned from the labeled source domain to the unlabeled target domain, domain adaptation methods are necessary to align the feature representation between the source and target domains.

B. Attention-based Dynamic Residual Feature Extractor

Model f_θ with parameter θ is a static model denoted as f_{θ_c} when the parameters are shared and fixed. Therefore, using static networks to map input samples with large distribution gaps to latent feature space using the same transformation function is suboptimal. To overcome this issue, dynamic networks are introduced to alleviate multi-source domain alignment conflicts. The dynamic part of the model learns parameters specifically for input sample x, i.e., $\theta=\theta(x)$. Considering the overlap of source domain distributions and large number of network parameters, not all parameters need to adapt dynamically with the sample. To simplify the model and improve convergence, static and dynamic networks are combined as feature extractors [15], where some network parameters vary dynamically with the sample, while others remain static.

One-dimensional convolutional neural networks (1D-CNNs) are used to capture high-level spatial frequency information between different channels based on a differential entropy (DE) feature vector with a 62×5 shape. For simplicity, the bias term of the convolution is ignored in the following description.

The original signal is first subjected to a 1×1 convolution operation, then sent to the static and dynamic network to extract features. The static part is composed of a 1×3 convolution operation, where W_0 is a $C_{out} \times C_{in} \times 1 \times 3$ weight matrix, C_{out} is the number of output channels, and C_{in} is the number of input channels. The static component serves as a common feature extractor, which explores and exploits similarities between different domains and maps the original feature space of different domains to a shared high-level hidden feature space. The dynamic component consists of multi-parallel 1×1 convolution kernels and an attention mechanism similar to SqueezeNet [19]. Building upon research

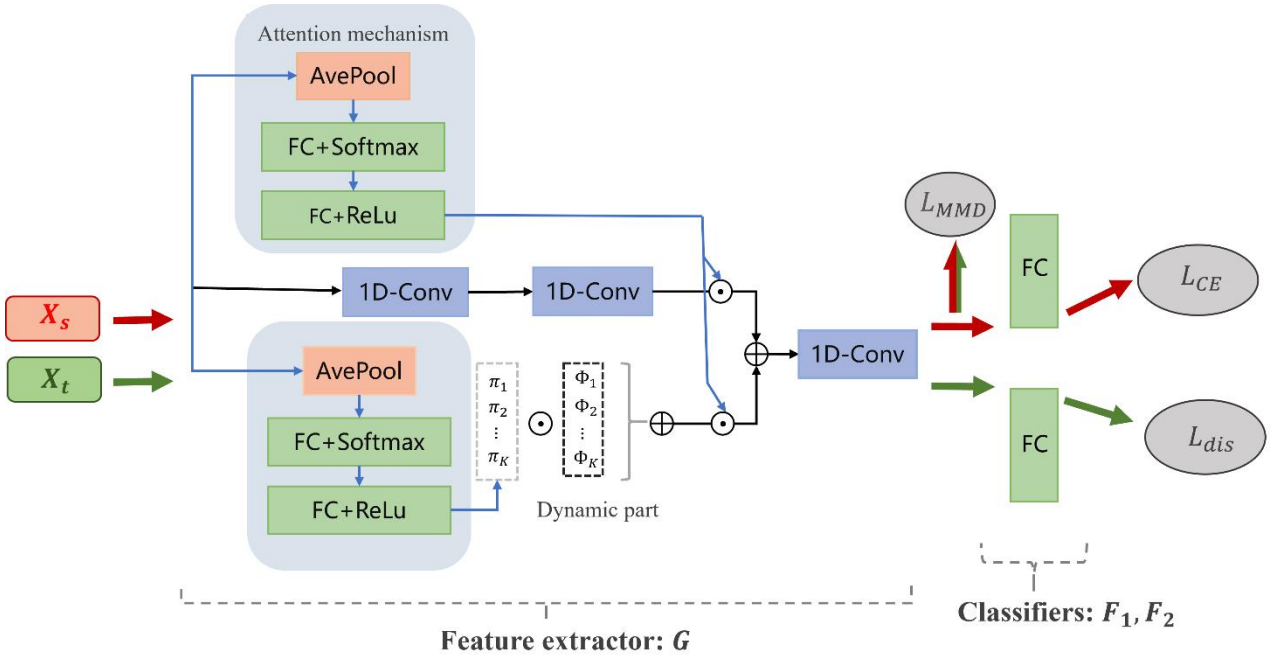

Fig. 1. Framework of proposed MS-ADRT. Whole structure can be divided into feature extractor, classifiers, and loss functions. Dynamic parameters of feature extractor are generated from input samples provided by the attention module.

in dynamic convolutions [20], dynamic parallel convolutions are used to enhance the representation ability of the feature extractor with negligible additional floating-point operations per second. To achieve personalized integration, an attention mechanism that generates weights $\lambda(x)$ is also used for the combination of static and dynamic networks. Let $W_{\theta(x)}$ denote the feature extractor parameters, W_0 denote the static component, and $\Delta W_{\theta(x)}$ denote the dynamic component as a residual block. $W_{\theta(x)}$ can be defined as:

$$W_{\theta(x)}=\lambda(x)W_0+(1-\lambda(x))\Delta W_{\theta(x)} \quad (1)$$

Regarding the attention mechanism architecture, squeeze-and-excitation [18] is adopted to dynamically aggregate multi-parallel convolution kernels. Specifically, the input features are first squeezed by average pooling. The output is then processed through a sequence of operations, i.e., Multi-Layer Perception (MLP), ReLU activation layer, MLP, and softmax activation layer, as depicted in the blue box in Fig. 1, to obtain a normalized attention score $\pi(x)$, defined as:

$$\pi(x)=\text{Softmax}(W_2(\text{ReLU}(W_1\text{avgpooling}(x)+b_1))+b_2) \quad (2)$$

The squeeze-and-excitation technique is used to dynamically adjust the weights of K parallel 1×1 convolution modules with full use of global frequency and EEG channel information. The 1×1 convolution kernels $\{\Phi_1,\ldots,\Phi_i,\ldots,\Phi_K\}$ are linearly weighted and summed as follows:

$$\Delta W_{\theta(x)}=\sum_{i=1}^{k}\pi_i(x)\Phi_i \quad (3)$$

For a given input x, the convolution $x*\Phi_i$ (* represents convolution operation) is a linear model, while the combined model $x*\Delta W_{\theta(x)}$ is a non-linear function. Therefore, the dynamic residual network possesses more presentation power than the static network.

The output Φ_i of each convolution kernel can be considered as a potential feature subspace. The attention score $\pi_i(x)$ assigned for optimal integration of feature subspaces is dependent on different values of x and plays a key role in enhancing beneficial information and reducing useless knowledge. Parallel dynamic convolution blocks in the feature extractor allow the model to adapt to different samples, and implicitly realize multi-source domain adaptation. In this case, we transformed EEG emotion recognition from a multi-source domain adaptation problem to a single-source problem.

Different from dynamic convolutions [20], the squeeze-and-excitation-based attention mechanism dynamically merges static and dynamic networks. The attention calculation is the same as (2), although the output is not attention over convolution kernels, but rather a two-dimensional attention vector $\lambda(x)$ over the static and dynamic modules. Parameter $\lambda(x)$ is used to determine whether to focus on the static or dynamic parts.

C. Domain Adaptation Scheme

As the dynamic feature extractor proposed above converts a multi-source domain adaptation problem to a single-source domain adaptation problem, there is no need to perform domain alignment for each pair of domains separately. Let $X_s=\{x_{s1},x_{s2},\cdots,x_{si},\cdots,x_{sn}\}$ and $X_t=\{x_{t1},x_{t2},\cdots,x_{tj},\cdots,x_{tn}\}$ represent the samples of the source and target domains in training, respectively, where n is the number of samples of the source domain and X_s, m is the amount of samples of the target domain X_t. Let H_k denote the reproduced kernel Hilbert space (RKHS) with characteristic kernel k. φ is the mapping function of kernel k. MMD is defined as:

$$L_{MMD}(X_s,X_t)=\left\|\frac{1}{n}\sum_{i=1}^{n}\phi(x_{si})-\frac{1}{m}\sum_{j=1}^{m}\phi(x_{tj})\right\|_H^2 \quad (4)$$

Let x denote one sample and y denote the emotion label. The total model is composed of feature extractor G , as described above, and classifiers F_1 and F_2 (green box in Fig. 1). Let $p_1(y|x_t)$ represent the C-dimensional probability distribution output of F_1 when input is x_t , and $p_2(y|x_t)$ represent the output of F_2. Let p_{1c}, p_{2c} represent the probability that the input sample is classified as the cth category, and C is the number of categories of the emotion. To align marginal probability distribution p(x) and conditional probability distribution p(y|x) , we propose a three-stage adversarial training mechanism referring to MCD.

Step A First, feature extractor G and classifiers F_1 and F_2 are trained to ensure that the source samples are correctly classified. MMD is utilized to encourage feature representations of the source and target domains extracted by the feature extractor to exhibit the same marginal distributions. The objective function can be presented as (5):

$$\min_{G,F_1F_2} L_{CE}(X_s,Y_s)+\gamma L_{MMD}(X_s,X_t) \quad (5)$$

$$\gamma=\frac{2}{1+e^{-\frac{r}{100}}}-1$$

where γ is the weight, which gradually increases from 0 as training iteration r progresses and controls the tradeoff between capturing emotion-discriminative features and domain-invariant features.

Step B Second, classifiers F_1 and F_2 are trained on fixed feature extractor G to maximize classification discrepancy of the target domain between the two classifiers. L1-distance is exploited to measure the discrepancy between the two classifiers, consistent with MCD.

$$\min_{F_1F_2} L_{CE}(X_s,Y_s)- L_{dis}(X_t) \quad (6)$$

$$L_{dis}(X_t)=\|P_1(X_t)-P_2(X_t)\|_1 \quad (7)$$

Step C Third, feature extractor G is trained to minimize the discrepancy between the two fixed classifiers. During experiments, it was difficult to correct the target domain samples whose features were distributed outside the source domain, so hyperparameter n was set to denote n repetitions of feature extractor G training in the final step. The objective function is shown in (8):

$$\min_{G} L_{dis}(X_t) \quad (8)$$

For the overall training, each epoch continues the three-stage training process sequentially according to its respective loss function until the model converges.

III. Experiments

A. Datasets and Preprocessing

Two public EEG datasets, i.e., SEED [21] and SEED-IV [22], were used to evaluate our proposed method. The SEED dataset contains EEG data from 15 subjects asked to watch 15 Chinese film clips labeled as negative, neutral, and positive emotions. Each subject participated in three experiments, with a one-week interval between each session. The SEED-IV dataset contains EEG data from 15 subjects asked to watch 24 Chinese film clips categorized as happy, sad, calm, and fearful. Each subject participated in three experiments, with time intervals between experiments of 1 day to 2 months. For both datasets, EEG data were recorded using a 62-channel ESI Neuroscan system, with the international 10–20 system layout and sample rate of 1 000 Hz. After EEG data collection, the signals were preprocessed to improve the signal-to-noise ratio. For SEED and SEED-IV, the raw EEG signals were first down-sampled to a 200-Hz sampling rate, then filtered with a 1–75 Hz band-pass filter and segmented to 1 s in SEED dataset and 4 s in SEED-IV dataset, respectively. The DE feature [23] was extracted from each segment in five frequency bands: i.e., delta (1–4 Hz), theta (4–8 Hz), alpha (8–14 Hz), beta (14–30 Hz), and gamma (31–50 Hz).

Zheng et al [21] demonstrated that the DE feature is effective and reliable in emotion classification. The DE feature can be used to measure the complexity of a continuous random variable, calculated with the following formula based on signal x drawn from Gaussian distribution $N(\mu, \delta^2)$:

$$h(x)=-\int_{-\infty}^{\infty} p(x)\log(p(x))=\frac{1}{2}\log 2\pi e\delta^2 \quad (9)$$

Before feeding into the model, we normalize all DE data in an electrode-wise way. Chen et al. [17] investigated the impact of different normalization strategies and verified that electrode-wise normalization significantly outperforms sample-wise and global-wise normalization.

B. Parameter Settings

As shown in Fig. 1, the whole feature extractor is composed of 1 × 1, 1 × 3, and 1 × 1 convolutions, with [128, 128, 62] filters, respectively. The dynamic parallel 1 × 1 convolution blocks are only added on the 1 × 3 convolution kernel. Following Resnet [24], batch normalization is performed before the ReLU activation function and after each convolution operation. The two classifiers are both composed of three-layer MLPs, reshaping the feature from 310D-256D-128D to the size of the emotion categories (three in SEED, four in SEED-IV). The final layer is followed by a softmax classification. An Adam optimizer is adopted for the training process, with the learning rate initially set to 0.01 with a weight decay of 0.0005 for all experiments, and all other parameters set to default: β_1= 0.9, β_2= 0.99, ϵ= 10e-8. The size of the mini-batch is 256.

C. Experimental Settings

We take only the first session of SEED and SEED-IV in the cross-subject experiment to avoid the effect of time. In detail, the prediction accuracy of the cross-subject task is the average of the leave-one-subject-out cross-validations. We perform cross-session experiments for each subject separately to prevent the influence of different individuals, and the cross-session accuracy is the average of all subjects.

In our experiments, cross-subject consists of a multi-source domain adaptation task with 14 source domains, while cross-session contains two source domains. As the cross-subject task is more difficult to transfer than the cross-session task, we used a three-stage training method combining MMD and MCD as described in Section 2. In the cross-session task, we removed the MMD method. We conducted several experiments with K set to 2–5. Results showed that MS-

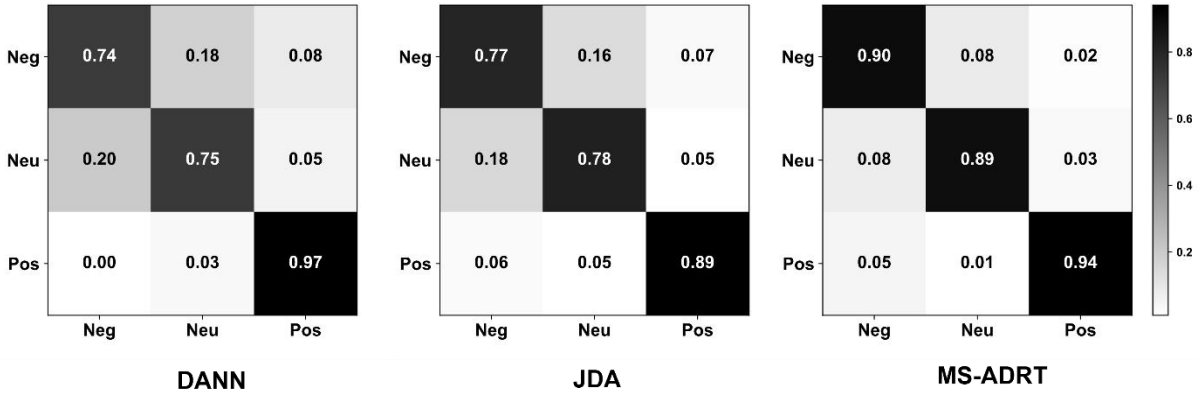

Fig. 2. Confusion matrix of EEG emotion recognition of SEED dataset using MS-ADRT, DANN, and JDA.

ADRT is insensitive to the number of parallel convolution kernels. Considering accuracy and computational effort, we used four parallel convolution kernels for the cross-subject task and two parallel convolution kernels for the cross-session task. All experiments were performed using an NVIDIA RTX 3090 GPU with 24 Gb of memory.

IV. Result and Analysis

A. Performance Across Subjects and Sessions

TABLE I shows the cross-subject and cross-session emotion recognition results based on the SEED and SEED-IV datasets. Mean and variance of prediction accuracy for all individuals were used as evaluation metrics, considering overall level and local variability.

As seen in TABLE I, the MDA-only algorithms, such as MMD, DANN, and DAN, showed lower classification accuracy than the algorithms accounting for CDA, such as MCD and JDA. For the SEED dataset, MS-ADRT demonstrated the highest accuracy across subjects (90.81%) and sessions (91.60 %) and showed the highest accuracy for the SEED-IV dataset. Thus, these results indicated that dynamic neural networks are effective in emotion recognition.

Furthermore, we found that the prediction accuracy of cross-session transfer was generally higher than that of cross-subject transfer in both datasets, regardless of the algorithm used. These findings suggest that the distribution differences in EEG signals between subjects are greater than the distribution differences in the same subject across sessions, signifying that knowledge transfer across sessions is easier than transfer across subjects. Our approach showed greater improvement in accuracy across subjects where transfer was more difficult, and weaker improvement in accuracy across sessions.

B. Performance Across Emotions

To explore the classification results of the three emotions under different algorithms, we constructed a confusion matrix for the three emotions across subject tasks (Fig. 2). Compared with DANN and JDA, our method showed high prediction accuracy for all three emotions. Comparison showed that the classification of neutral and negative emotions was relatively difficult. Previous studies using SEED data have reported that positive emotions are easier to classify for video-stimulated emotions [25, 26]. In music-induced EEG data, classification of pleasure is more accurate than that of fear [27], which is inconsistent with our findings. Thus, the patterns of brain arousal for different emotions need to be further explored.

C. Study on Alignment Loss Function

As shown in TABLE II, we compared the effects of using three different domain alignment losses on the proposed dynamic residual feature extractor: i.e., marginal distribution loss of MMD, conditional distribution loss of MCD, and joint distribution loss of both. In the absence of MCD, L_{dis} in Fig. 1 needs to be removed. For both cross-subject and cross-session, there was a significant improvement in accuracy when using conditional distribution MCD. It should be noted that alignment loss is crucial for the multi-source domain adaptation problem. As shown in Table II, without the alignment loss function, although the model can adapt to all source domains, migration to the target domain is difficult due to the large distribution differences between the target and source domains.

The dynamic feature extractor aims to improve the representational power of the model [20, 30], so the model can fully learn from multi-source data, while alignment loss can constrain the feature extractor to extract statistically similar and emotionally sensitive features in the target domain [31].

For cross-subject testing, using both MMD and MCD produced higher prediction accuracy than using either alone. However, for cross-session testing, using MCD alone resulted in better outcomes compared to using both, which led to negative transfer. The choice of alignment algorithm needs to be adapted to the difficulty of the migration problem. For problems with small migration gaps, overly powerful alignment can lead to negative migration and degradation of results. The emotion recognition classification results indicated that cross-individual tasks are more difficult than cross-session tasks, consistent with the findings of [25].

D. Comparison Between Static And Dynamic Feature Extractors

In Table III, we compared the classification results between the static feature extractor with the dynamic feature extractor

employed by MS-ADRT. For the static feature extractor, the two attention mechanisms and the parallel convolution blocks in Fig. 1 need to be removed, and the rest of the parameter settings were consistent with MS-ADRT. As seen in TABLE III, MS-ADRT achieves nearly 1% improvement in the cross-subject task and 0.5% improvement in the cross-session task compared to the static feature extractor. It confirmed our claim that MS-ADRT simplifies domain alignment by dynamically adapting to all source domains. Moreover, this dynamic feature extractor can be easily grafted onto existing domain adaptation algorithms, providing a new research paradigm for the multi-source domain adaptation problem of EEG emotion recognition.

TABLE I

COMPARISON OF DIFFERENT ALGORITHMS ON SEED AND SEED-IV DATASETS

Dataset	Method	Cross-subject	Cross-session
SEED	MMD	80.88 ± 10.1	84.38 ± 12.05
	MCD	85.63 ± 6.04	87.65 ± 9.23
	DAN [17, 28]	65.84 ± 2.25	79.93 ± 7.06
	DANN [28]	79.19 ± 13.14	83.15 ± 12.01
	JDA [29]	88.28 ± 11.44	91.17 ± 8.11
	MS-ADRT	**90.81 ± 6.98**	**91.60 ± 4.57**
SEED-IV	DAN [28]	32.44 ± 9.02	55.14 ± 12.79
	DDC [17]	37.41 ± 6.36	57.63 ± 11.28
	MS-ADRT	**68.98 ± 6.80**	**76.11 ± 13.42**

TABLE II

COMPARISON OF DOMAIN ALIGNMENT LOSSES ON SEED AND SEED-IV DATASETS

Dataset	Method	Cross-subject	Cross-session
SEED	MMD and MCD	**90.81 ± 6.98**	91.38 ± 5.45
	MCD	90.59 ± 7.19	**91.60 ± 4.57**
	MMD	78.79 ± 4.67	86.36 ± 9.04
	w/o MMD and MCD	76.57 ± 7.30	85.29 ± 8.50
SEED-IV	MMD and MCD	**68.98 ± 6.80**	74.24 ± 14.27
	MCD	67.86 ± 9.43	**76.11 ± 13.42**
	MMD	63.34 ± 11.14	75.77 ± 12.03
	w/o MMD and MCD	61.85 ± 11.18	71.41 ± 12.11

TABLE III

COMPARISON BETWEEN STATIC AND DYNAMIC FEATURE EXTRACTORS

Dataset	Method	Cross-subject	Cross-session
SEED	MS-ADRT	**90.81 ± 6.98**	**91.60 ± 4.57**
	Only static	89.54 ± 6.94	90.89 ± 7.78
SEED-IV	MS-ADRT	**68.98 ± 6.80**	**76.11 ± 13.42**
	Only static	67.96 ± 11.91	75.51 ± 12.14

E. Feature Visualization

To clearly represent how the distribution of features captured by the feature extractor varies, we used t-distributed stochastic neighbor embedding (t-SNE) [32] to map the

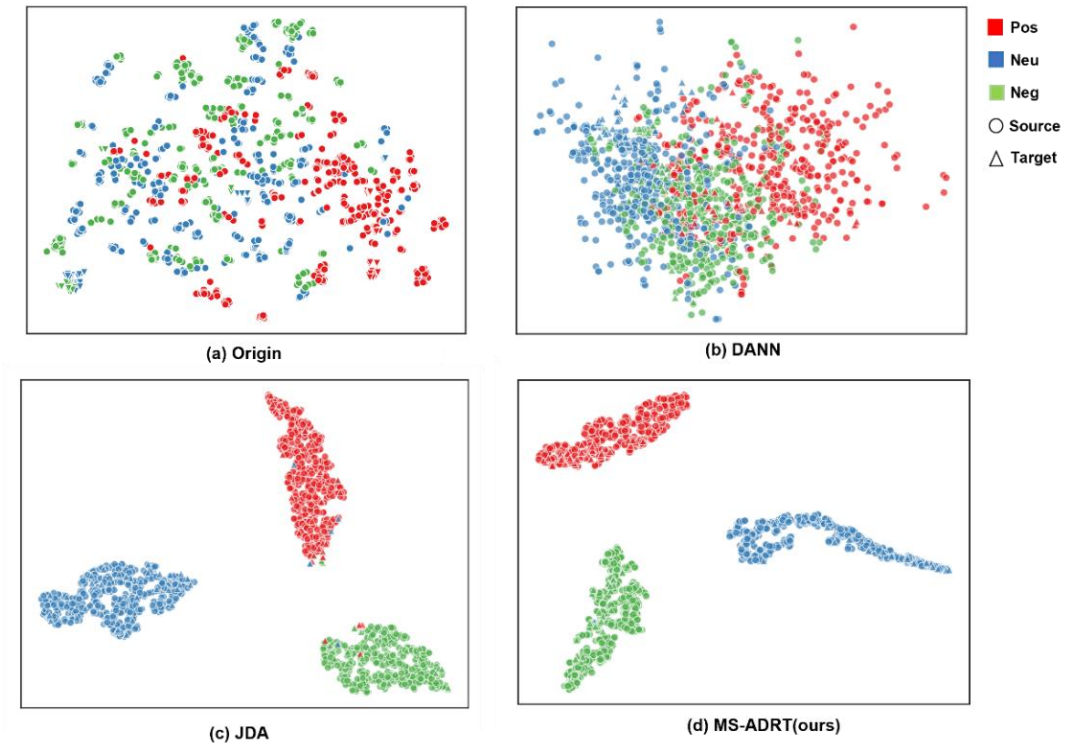

Fig. 3. Visualization of latent representations using t-SNE. For simplicity, we only used SEED as an example for illustration. (a) Denotes original DE feature distribution, and (b), (c), and (d) are the visualization results of the features extracted by DANN, JDA, and MS-ADRT, respectively.

features extracted by the last fully connected layer to a 2D plane after principal component analysis (PCA) [33]. The t-SNE non-linear downscaling method can map the structure of high-dimensional feature space to low-dimensional space for feature visualization, and the t-SNE algorithm can be accelerated by removing redundant crosstalk features through PCA dimensionality reduction.

For simplicity, we randomly selected the second subjects as target domains and the rest as source domains, and randomly selected 100 samples from each domain. Embedding distributions using different algorithms are shown in Fig. 3. The embedding distributions of neutral and negative emotions were harder to separate, leading to lower classification accuracy than that of positive emotion. The boundaries of the three emotion categories in DANN were relatively blurry as DANN has no alignment constraints for different categories. Compared with JDA, our method achieved clearer intra-class compression and inter-class separation, resulting in higher classification accuracy in target domains (points of the triangle in Fig. 3). Thus, these results suggest that our approach can flexibly consider the distribution discrepancies of multi-source domains to enhance the transfer effect.

V. Conclusions

In the current paper, we propose and discuss the MS-ADRT model, which not only considers large cross-subject and cross-session variation challenges in EEG-based emotion recognition, but also large discrepancies in the distribution between source domains. Specifically, we introduce a dynamic feature extractor, where the model parameters of the dynamic residual block are not fixed but vary with the sample. The model achieves multi-source domain adaptation by adapting to samples, thus eliminating the need to create a domain adaptation branch for each source domain. Experimentally, the model shows superior adaptation to multi-source domains and transfer performance compared to other state-of-the-art EEG-based emotion recognition methods.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 61971420), Science Frontier Program of the Chinese Academy of Sciences (Grant No. QYZDJ-SSW-SMC019), and Science and Technology Innovation 2030 – Brain Science and Brain-Inspired Intelligence Project (Grant No. 2021ZD0200200).

References

[1] M. Valstar et al., "Avec 2016: Depression, mood, and emotion recognition workshop and challenge," in Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge, 2016, pp. 3-10.

[2] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges," Brain-Computer Interfaces, vol. 1, no. 2, pp. 66-84, 2014.

[3] X. Li, B. Hu, T. Zhu, J. Yan, and F. Zheng, "Towards affective learning with an EEG feedback approach," in Proceedings of the first ACM International Workshop on Multimedia Technologies for Distance Learning, 2009, pp. 33-38.

[4] S. Sanei and J. A. Chambers, EEG signal processing. John Wiley & Sons, 2013.

[5] X.-W. Wang, D. Nie, and B.-L. Lu, "EEG-based emotion recognition using frequency domain features and support vector machines," in International Conference on Neural Information Processing, 2011, pp. 734-743: Springer.

[6] A. Bhardwaj, A. Gupta, P. Jain, A. Rani, and J. Yadav, "Classification of human emotions from EEG signals using SVM and LDA Classifiers," in 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 2015, pp. 180-185: IEEE.

[7] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," Advances in Neural Information Processing Systems, vol. 19, 2006.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345-1359, 2009.

[9] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," IEEE Transactions on Neural Networks, vol. 22, no. 2, pp. 199-210, 2010.

[10] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu, "Equivalence of distance-based and RKHS-based statistics in hypothesis testing," The Annals of Statistics, pp. 2263-2291, 2013.

[11] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in International Conference on Machine Learning, 2015, pp. 97-105: PMLR.

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[13] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3723-3732.

[14] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject EEG emotion recognition," IEEE Transactions on Cybernetics, vol. 50, no. 7, pp. 3281-3293, 2019.

[15] Y. Li, L. Yuan, Y. Chen, P. Wang, and N. Vasconcelos, "Dynamic transfer for multi-source domain adaptation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10998-11007.

[16] K. Zhang, M. Gong, and B. Schölkopf, "Multi-source domain adaptation: A causal view," in Twenty-ninth AAAI Conference on Artificial Intelligence, 2015.

[17] 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.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132-7141.

[19] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," arXiv preprint arXiv:1602.07360, 2016.

[20] Y. Chen, X. Dai, M. Liu, D. Chen, L. Yuan, and Z. Liu, "Dynamic convolution: Attention over convolution kernels," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11030-11039.

[21] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Transactions on Autonomous Mental Development, vol. 7, no. 3, pp. 162-175, 2015.

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

[23] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 6627-6630: IEEE.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[25] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition," IEEE Transactions on Affective Computing, 2019.

[26] Y. Wang et al., "EEG-based emotion recognition with similarity learning network," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1209-1212: IEEE.

[27] N. Masood and H. Farooq, "Comparing Neural Correlates of Human Emotions across Multiple Stimulus Presentation Paradigms," Brain Sciences, vol. 11, no. 6, p. 696, 2021.

[28] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, "Cross-subject emotion recognition using deep adaptation networks," in International Conference on Neural Information Processing, 2018, pp. 403-413: Springer.

[29] 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.

[30] B. Yang, G. Bender, Q. V. Le, and J. Ngiam, "Condconv: Conditionally parameterized convolutions for efficient inference," Advances in Neural Information Processing Systems, vol. 32, 2019.

[31] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, pp. 766-785, 2019.

[32] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of machine learning research, vol. 9, no. 11, 2008.S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemometrics and Intelligent Laboratory Systems, vol. 2, no. 1-3, pp. 37-52, 1987

[33] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," Chemometrics and Intelligent Laboratory Systems, vol. 2, no. 1-3, pp. 37-52, 1987