2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Cross-subject EEG-based Emotion Recognition Using Adversarial Domain Adaption with Attention Mechanism

Yalan Ye[1,*], Xin Zhu[1], Yunxia Li[2], Tongjie Pan[1], Wenwen He[1]

***Abstract*— Cross-subject EEG-based emotion recognition (ER) is a rewarding work in real-life applications, due to individual differences between one subject and another subject. Most existing studies focus on training a subject-specific ER model. However, it is time-consuming and unrealistic to design the customized subject-specific model for a new subject in cross-subject scenarios. In this paper, we propose an Adversarial Domain Adaption with an Attention Mechanism method for EEG-based ER, namely ADAAM-ER, to decrease the individual discrepancy. ADAAM-ER consists of a Graph Convolution Neural Networks with CNNs (GCNN-CNNs) and an Adversarial Domain Adaption with a Level-wise Attention Mechanism (ADALAM). Specifically, GCNN-CNNs as a feature extractor, which constructs a broader feature space, is designed to obtain more discriminative features. And ADALAM, which can decrease the individual discrepancy by alignment of the more transferable feature regions, is introduced to further obtain the discriminative features with higher transferability. Consequently, the proposed ADAAM-ER method can design a more transferable emotion recognition model with more discriminative features for a new subject via improving transferability. Experimental results on the SEED dataset have verified the effectiveness of the proposed ADAAM-ER method with the mean accuracy of 86.58%.**

I. INTRODUCTION

Emotion plays an essential role in our daily lives. With the continuous development of human-computer interaction, EEG-based emotion recognition (ER) has become an attractive research topic. However, individual differences between one subject and another subject lead to shift in data distribution and degrade the performance of cross-subject ER.

There have been several attempts to decrease the individual discrepancy. One technique is feature selection. For example, [1] and [2] select features that are common in the source and target domains, and construct a robust ER model based on these selected features. Another technique is the general domain adaption. Lan et al. [3] led a comparative study on several general domain adaptation techniques, such as transfer component analysis (TCA) [4], subspace alignment (SA) [5], etc. They found that the subject independent

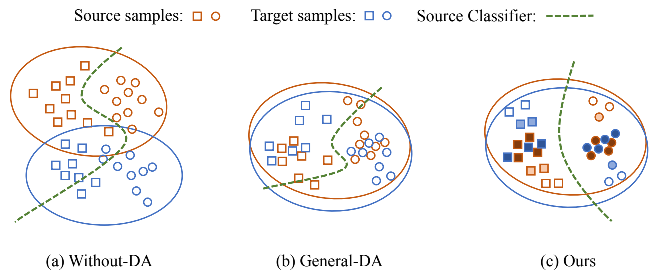


Fig. 1. The illustration of different domain adaptation strategies. (a) shows classification without domain alignment, (b) shows the general domain adaptation, which forces alignment of some completely dissimilar features may lead to negative transfer, and (c) our proposed ADAAM-ER pays more attention to more transferable feature regions (darker colors indicate higher transferability), and aligning those regions minimizes the discrepancy between source and target.

classification accuracy can be improved by around 10% through domain alignment. However, the two techniques have a limitation, that is, they need shallow handcrafted features whose discriminativeness may not strong enough to represent emotions.

Deep learning is helpful to obtain deep discrimimnative representation for ER. For example, Song et al. [6] generated deep discriminative representation through dynamical graph convolutional neural networks (DGCNN) to learn the intrinsic relationships between EEG channels. Zhang et al. [7] utilized broad concept to concatenate all hierarchical features to build a broad feature space. Although they maintain the feature discriminativeness, they did not explore transferability of the features.

Recently, deep domain adaptation embedding domain adaptation modules into deep learning has yielded satisfactory results in knowledge transferring in computer vision [11]. For example, Jin et al. [8] first applied adversarial domain adaption [12] into EEG-based ER, which archives a significant improvement by learning both discriminative and transferable features. Later, Li et al. [9] and Zhong et al. [10] extended the feature extractor of [12] to further improve feature discriminativeness. In Figure 1, compared with non-domain adaptation strategies (Fig. 1 (a)), the domain adaptation strategies (Fig. 1(b)) can decrease the distribution difference between a subject and another subject to some extent, but it may lead to negative transfer, which result in degradation the ER performance in cross-subject scenarios. Since there is only one domain discriminator, which means that all features are regarded as a whole to be transferred or not. However, features in the feature space do not have the same transferability, and taking advantage of this can

*This work was supported in part by a grant from the National Natural Science Foundation of China (No.61976047), in part by grants from Science & Technology Department of Sichuan Province of China (No. 2020YFG0087, 2020YFG0326 and 2021YFG0331), in part by a grant from the Fundamental Research Funds for the Central Universities (No. ZYGX2021YGLH016), and in part by a grant from the Open Innovation Fund of 55 Institute of China North Industries Group.

[1]Yalan Ye, Xin Zhu, Tongjie Pan, Wenwen He are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

[2]Yunxia Li is with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

*Corresponding author: Yalan Ye yalanye@uestc.edu.cn

TABLE I

A COMPREHENSIVE COMPARISON OF SOME EXISTING WORKS

Ref.	Strategies	Feature Extraction	Advantages		
			BFS	**DEF**	**TEF**
[1]	Feature Selection	Handcrafted Features	×	L	L
[2]	Feature Selection	Handcrafted Features	×	L	L
[4]	General-DA	Handcrafted Features	×	L	L
[5]	General-DA	Handcrafted Features	×	L	L
[6]	Without-DA	Deep Features	×	**H**	L
[7]	Without-DA	Deep Features	√	**H**	L
[8]	General-DA	Deep Features	×	M	M
[9]	General-DA	Deep Features	×	**H**	M
[10]	General-DA	Deep Features	×	**H**	M
Ours	ADALAM	Deep Features	√	**H**	**H**

* In the table, **BFS** denotes **B**roader **F**eature **S**pace, **DEF** denotes **D**iscriminativeness of **E**motion **F**eatures, **TEF** denotes **T**ransferability of **E**motion **F**eatures.
* We denote the discriminativeness and transferability of emotion features in the table as High(H), Middle(M) and Low(L).
* In the table, ADALAM denotes Adversarial Domain Adaption with a Level-wise Attention Mechanism.

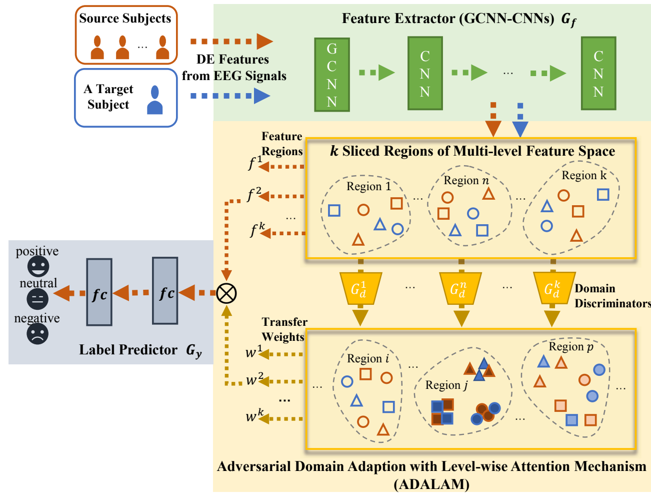


Fig. 2. The framework of the ADAAM-ER method, which consists of a Graph Convolution Neural Networks with CNNs (GCNN-CNNs) and an Adversarial Domain Adaption with a Level-wise Attention Mechanism (ADALAM). GCNN-CNNs as a feature extractor G_f, which constructs a broader feature space. And ADALAM, which can decrease the individual discrepancy by alignment of the more transferable feature regions.

improve the ER performance.

In this paper, we propose an Adversarial Domain Adaption with an Attention Mechanism method for EEG-based ER (ADAAM-ER) to decrease the individual discrepancy. ADAAM-ER consists of two parts, one is a Graph Convolution Neural Networks with CNNs (GCNN-CNNs), which is used to extract multi-level discriminative representations as a feature extractor. Inspired by GCB-net [7], multiple regular CNNs after graph convolution are stacked to learn high-level emotion representations. The other one is an Adversarial Domain Adaption with a Level-wise Attention Mechanism (ADALAM), which is used to decrease the distribution shift between source subjects and target subject. Technically, we divide the multi-level feature space into regions, and each region corresponds to a domain discriminator. After the training, the regions, which are hard to distinguish by domain discriminators, are considered to have better transferability relatively. By assigning greater weight to those regions, the training concentrates on alignment of the more transferable regions (Fig. 1 (c)). So the proposed ADAAM-ER method can design a more transferable ER model with more discriminative features via improving transferability. Table I compares our proposed method with the above EEG-based ER classification methods [1], [2], [4]–[10], and it can be seen that our proposed method more advantageous in terms of feature discriminativeness as well as transferability.

In summary, the contributions of this paper are as follows:

- We propose an Adversarial Domain Adaption with Attention Mechanism ADAAM-ER method to recognize emotions based on EEG. Compared with the existing cross-subject ER methods, the proposed ADAAM-ER method can extract both discriminative and transferable features, which improves ER performance in cross-subject scenarios.
- GCNN-CNNs as a feature extractor, which is constructed a broader feature space, aims to obtain more discriminative features. And to further obtain the discriminative features with higher transferability, an Adversarial Domain Adaption with a Level-wise Attention Mechanism (ADALAM) is designed to decrease the individual discrepancy by alignment of the more transferable feature regions.
- Comprehensive experiments on the SEED dataset clearly demonstrated the effectiveness of the proposed method and achieves the mean accuracy of 86.58%.

II. THE PROPOSED METHOD

The framework of the proposed method is shown in Fig. 2. The data from source subjects (training subjects) and a target subject (test subject) input to the feature extractor G_f by GCNN-CNNs to obtain multi-level discriminative emotion representations. Then, the output of feature extractor fed into a label predictor G_y and multiple domain discriminators G_d in parallel, resulting in classification loss and transfer loss. After backpropagation, the label predictor G_y guarantees the discriminativeness of features, and the domain discriminators G_d ensure the domain-invariance via the Adversarial Domain Adaption with a Level-wise Attention Mechanism. More details are demonstrated as follows.

A. Graph Convolution Neural Networks with CNNs

ADAAM-ER applies the graph convolution neural networks with stacked regular CNNs (GCNN-CNNs) to extract discriminative features. The differential entropy (DE) [13] matrix extracted from each EEG channel separately of source (training data) and target (test data) will be input to GCNN to capture structure information between different EEG channels, then use multiple regular CNNs for further abstraction of features.

1) Graph Representation: The DE matrix input to the GCNN needs to be converted into a representation of the graph structure. Motivated by DGCNN [6], each EEG channel is considered as a vertex node in graph, and the connection between two different vertex nodes corresponds to an edge of the graph.

The undirected graph can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{A})$, where $\mathcal{V}$ is the set of n nodes and $\mathcal{E}$ is the set of edges connecting these nodes. $\mathbf{A} \in \mathbb{R}^{n\times n}$ is the weighted adjacency matrix to represent the edge set $\mathcal{E}$, each entry A_{ij} indicates the connection between ith node and jth one.

2) Spectral Graph Filtering: Spectral graph filtering uses graph Fourier analysis, which depends on symmetric normalized Laplacian matrix $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ of graph, where $\mathbf{D} \in \mathbb{R}^{n\times n}$ is a diagonal matrix and $\mathbf{I}$ is an identity matrix. $\mathbf{L}$ can be decomposed as $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{U}$ is the orthonormal eigenvector matrix of $\mathbf{L}$ and $\mathbf{\Lambda} = diag(\lambda_0, \lambda_1, ..., \lambda_{n-1}) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with corresponding eigenvalues. Given graph signal $\mathbf{X}$, the graph Fourier transform of $\mathbf{X}$ is $\hat{\mathbf{X}} = \mathbf{U}^T\mathbf{X}$, and its inverse transform is $\mathbf{X} = \mathbf{U}\hat{\mathbf{X}}$. Therefore, the graph convolution between $\mathbf{X}$ and a filter $\mathbf{G}$ is defined as

$$\mathbf{X} * \mathbf{G} = \mathbf{U}((\mathbf{U}^T\mathbf{G}) \odot (\mathbf{U}^T\mathbf{X})) = \mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^T\mathbf{X}, \quad (1)$$

where $\odot$ is the Hadamard product and $g(\mathbf{\Lambda}) = diag(g(\lambda_0), g(\lambda_1), ..., g(\lambda_{n-1}))$ denotes a diagonal matrix with n spectral filter coefficients.

Directly computing $g(\mathbf{\Lambda})$ is too expensive. K-order Chebyshev polynomial is adopted to approximate $g(\mathbf{\Lambda})$ as follows

$$g(\mathbf{\Lambda}) = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{\Lambda}}), \quad (2)$$

where θ_k is the Chebyshev polynomial coefficient and $T_k(x)$ can be recursively calculated according to the following expressions

$$\begin{cases} T_0(x) = 1, T_1(x) = x \\ T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), \quad k \geq 2 \end{cases}, \quad (3)$$

Hence, the filtering operation can be converted as follows

$$\mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^T\mathbf{X} = \mathbf{U}(\sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{\Lambda}})))\mathbf{U}^T\mathbf{X} = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}))\mathbf{X}, \quad (4)$$

where $\tilde{\mathbf{L}} = \frac{2\mathbf{L}}{\lambda_{max}} - \mathbf{I}$. So the output of K-order Chebyshev graph convolution can be represented as

$$GCNN(\mathbf{X}) = ReLU(((\sum_{k=0}^{K-1} \theta_k T_k(\tilde{\mathbf{L}}))\mathbf{X}) \cdot \theta_{gcnn}), \quad (5)$$

where θ_{gcnn} is the matrix of all parameters learned in GCNN. $ReLU$ is used to increase nonlinearity. The abstract representations computed by $GCNN(x)$ are fed into the following regular CNNs to obtain higher-level features. After several regular convolution, features of all layers are concatenated as follows:

$$G_f(\mathbf{X}) = [GCNN, CNN_1, CNN_2 ..., CNN_l], \quad (6)$$

where G_f represents the GCNN-CNNs feature extractor to extract discriminative emotion-related features.

B. Adversarial Domain Adaption with Level-wise Attention Mechanism

The idea of adversarial learning has a wide range of applications in deep learning and is usually used to improve the robustness of learning methods [14]. Adversarial domain adaptation is the alignment of source and target domain feature spaces through adversarial learning of feature extractors and discriminators.

As mentioned before, not all features extracted by G_f are equally transferable and using only one domain discriminator does not make better use of those features that have higher transferability. Therefore, to better lower the influence of individual difference, we extend the adversarial domain adaption by proposing adversarial domain adaption with attention mechanism. Specifically, we divide the features of each level into two regions, each region corresponding to a domain discriminator. As shown in Fig. 2, there are K domain discriminators $G_d^k, k = 1, 2, ..., K$. Applying this to all K domain discriminators $G_d^k, k = 1, 2, ..., K$ yields

$$L_d = \frac{1}{Kn} \sum_{k=1}^{K} \sum_{x_i \in (\mathcal{D}_s \cup \mathcal{D}_t)} L_d^k(G_d^k(\boldsymbol{f_i^k}), d_i), \quad (7)$$

where $\boldsymbol{f_i^k} = (G_f(x_i))^k$ is the concatenated representation of GCNN-CNNs in region k, $d_i \in 0, 1$ serves as the domain label for the i-th sample x_i, L_d^k is the cross-entropy loss of the domain discriminator G_d^k. The output $\hat{d}_i^k = G_d^k(\boldsymbol{f_i^k})$ of each domain discriminator G_d^k is the predicted probability of the region k of sample i belonging to the source domain. When the probability is close to 1, it means the region k belongs to the source domain, and 0 indicates it belongs to the target domain. We focus on the more transferable regions by giving more weight to those. We apply the entropy criterion $E(p) = -\sum_j p_j \cdot log(p_j)$ to give the transfer weight for each region k as:

$$w_i^k = 1 - E(\hat{d}_i^k), \quad (8)$$

Therefore, the $\boldsymbol{f_i^k}$ are transformed using transfer weight as

$$\boldsymbol{m}_i^k = (1 + w_i^k) \cdot \boldsymbol{f_i^k}, \quad (9)$$

In this way, the feature region with strong transferability will get more attention.

C. Optimization of ADAAM-ER

After the transformed features $\boldsymbol{m}_i$ are obtained, the probability that sample i belongs to the emotion class (positive, neutral, negative) can be computed as follows

$$G_y(\boldsymbol{m}_i) = softmax(fc(\boldsymbol{m}_i)), \quad (10)$$

where $fc(\boldsymbol{m}_i) = \boldsymbol{m}_i \cdot \theta_y$ is the output of fully connected layer.

TABLE II

MEAN ACCURACIES AND STANDARD DEVIATIONS (%) OF DIFFERENT METHODS IN THE SUBJECT-INDEPENDENT EXPERIMENT ON SEED

Method	Accuracy (mean/std)
SVM [15]	56.73/16.29
TCA [4]	63.64/14.88
SA [5]	69.00/10.89
DAN [8]	79.19/13.14
DGCNN [6]	79.95/09.02
GCB-net [7]	-
R2G-STNN [9]	84.16/07.63
RGNN [10]	85.30/**06.72**
Ours	**86.58**/08.41

Integrating all things together, the objective of ADAAM-ER is

$$C_0(\theta_f,\theta_y,\theta_d|_{k=1}^{K}) = \frac{1}{n_s}\sum_{x_i\in\mathcal{D}_s} L_y(G_y(\boldsymbol{m}_i),y_i) - \frac{\lambda}{Kn}\sum_{x_i\in\mathcal{D}} L_d^k(G_d^k((G_f(x_i))^k),d_i), \tag{11}$$

where $n = n_s + n_t$, $\mathcal{D} = \mathcal{D}_s \cup \mathcal{D}_t$, $G_y(\boldsymbol{m}_i)$ and y_i are the vectors of predicted labels and true labels respectively, $L_y(\cdot,\cdot)$ is the loss of class label prediction and λ is a hyper-parameter that balances the two objectives in the unified optimization problem. The minimax optimization problem is to find the network parameters $\hat{\theta}_f, \hat{\theta}_y$ and $\hat{\theta}_d^k(k = 1, 2, ..., K)$ that jointly satisfy

$$\begin{aligned}(\hat{\theta}_f,\hat{\theta}_y) &= \arg\min_{\theta_f,\theta_y} C(\theta_f,\theta_y,\theta_d^k|_{k=1}^{K}),\\ (\hat{\theta_d^1},...,\hat{\theta_d^K}) &= \arg\max_{\theta_d^1,...,\theta_d^K} C(\theta_f,\theta_y,\theta_d^k|_{k=1}^{K}).\end{aligned} \tag{12}$$

where θ_f are the parameters of feature extractor, θ_d^k are the parameters of domain discriminator G_d^k, θ_y are the parameters of label predictor.

The proposed method focus more on transferable regions between different levels to achieve a better domain alignment via attention mechanism.

III. MATERIALS

A. EEG Dataset

In this paper, we evaluate the proposed method on a public dataset called SJTU Emotion EEG Dataset (SEED) [16]. The SEED dataset consisted of 15 subjects who collected 62-channel EEG data, each subject watched multiple well-edited movie clips of positive, neutral, and negative categories.

B. Signal Processing and Feature Extraction

For signal processing, the preprocessed EEG data provided by SEED was down-sampled to 200 Hz. A bandpass frequency filter from 0-75 Hz was applied. A 512-point short-time Fourier transform with a non-overlapped Hanning window of 1s was used to extract the frequency domain features for each channel [13].

For feature extraction, SEED provides several extracted EEG features on five frequency bands. [13] has shown that DE feature on Gamma band and total five frequency bands are more relevant to emotion than other sub-bands, like Delta, Theta, Alpha, and Beta. [7] also demonstrates that DE features on total five frequency bands achieve the best performance. Hence, we choose DE features on total five frequency bands as the input of our method.

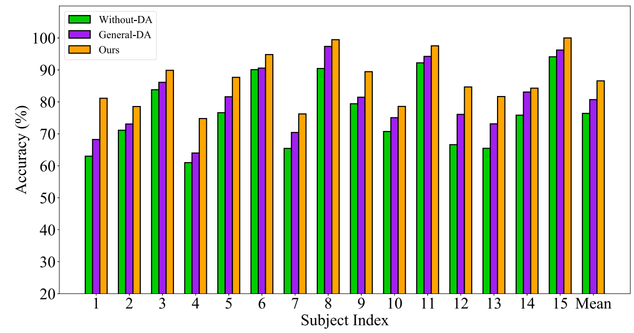


Fig. 3. The accuracies of three models (Without-DA, General-DA, Ours) for 15 subjects and the mean accuracy. Our ADAAM-ER outperforms the other models on each subject.

IV. RESULTS AND DISCUSSION

In this section, we will discuss the experiment results of the proposed method on the SEED dataset.

A. Classification Performance

We adopt subject-independent to evaluate the proposed method, in which the training EEG data and the testing ones come from different subjects.

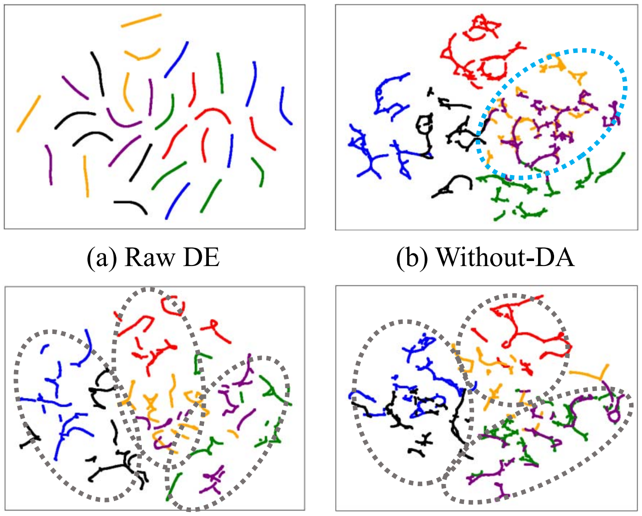
Table II presents the classification accuracies and standard deviations of the state-of-the-art methods. We can discover the proposed method outperforms others with a mean accuracy of 86.58%. The standard deviation of ADAAM-ER is not the smallest, because the result of subject 4 is extremely low. The major performance improvement can be attributed to two factors: 1) the proposed method uses features of multiple levels to form a broader feature space, so more distinguishing representation can be extracted; 2) the proposed method focuses on the regions of the feature space with good transferability, and aligning these regions can make knowledge transfer better.

B. Ablation Study

We conduct ablation studies on the SEED dataset to illustrate the effectiveness of each component in the proposed method. Specially, we compare the following three variant models:

- Without-DA: only GCNN with stacked regular CNNs, and without domain adaption;
- General-DA: the GCNN with stacked regular CNNs, and with the global adversarial domain adaption;
- Ours: the GCNN with stacked regular CNNs, and with the level-wise attention-based adversarial domain adaption;

Fig. 3 shows the accuracies of three variant models for 15 subjects. Compared with the other two models, Without-DA model has the lowest accuracy. General-DA model uses a



(a) Raw DE (b) Without-DA (c) General-DA (d) Ours

Fig. 4. The t-SNE of feature distribution of variant models in the source domain and target domain. (red/green/blue dots: the negative/neutral/positive emotion in source domain, orange/purple/black dots: the negative/neutral/positive emotion in target domain, gray dashed circle dots: an emotion category in source and target domains.)

global domain discriminator, which assumes that the whole feature space has the same transferability by default, which may lead to negative migration. ADAAM-ER applies the attention mechanism that allows the model to focus more on those domain-invariant features to archive a significant improvement in subject transferring.

We further visualize the feature distributions in 2-dimensional space using t-SNE to indicate the effectiveness of the proposed method. Fig. 4 shows the visualization of the raw DE features, the embedded features of Without-DA, General-DA and ADAAM-ER, respectively. We make the following observations: 1) the raw DE features are scattered in the source domain and the target domain; 2) after learning by the Without-DA, features of different emotion categories in the source domain are distinguishable. However, since the domain adaptation has not yet been carried out, the distribution difference of same category between the source and target domain is obvious, so the performance of classifier in the target domain is not good, we can see from Fig. 4 (b) that the distributions of negative and neutral in the target domain are mixed together; 3) after learning by the General-DA, the difference in the distribution of source and target domains is reduced to some extent, but still unsatisfactory; 4) our ADAAM-ER restricts the distribution difference between the source domain and the target domain, which make source classifier more suitable for target data. Therefore, our ADAAM-ER focuses on features with high transferability to achieve a better domain alignment to improve the ER performance in cross-subject scenarios.

V. CONCLUSION

In this paper, an Adversarial Domain Adaption with an Attention Mechanism method for EEG-based ER is proposed to design a more transferable emotion recognition model with more discriminative features for a new subject by decrease the individual discrepancy in cross-subject scenarios.

The proposed method extracts more discriminative features by the Graph Convolution Neural Networks with CNNs as the feature extractor. And the Adversarial Domain Adaption with a Level-wise Attention Mechanism is introduced to pay more attention on alignment of the more transferable feature regions to decrease the individual difference between one subject and another subject. Experimental results on the SEED dataset have verified that the our proposed method is able to effectively improve the cross-subject EEG-based ER performance.

REFERENCES

[1] Z. Yin, Y. Wang, L. Liu, W. Zhang, and J. Zhang, "Cross-subject eeg feature selection for emotion recognition using transfer recursive feature elimination," *Frontiers in Neurorobotics*, vol. 11, p. 19, 2017.

[2] L. lan Chen, A. Zhang, and X. guang Lou, "Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning," *Expert Systems with Applications*, vol. 137, pp. 266–280, 2019.

[3] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for eeg-based emotion recognition: A comparative study on two public datasets," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 1, pp. 85–94, 2019.

[4] 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, 2011.

[5] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE International Conference on Computer Vision*, pp. 2960–2967, 2013.

[6] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, 2020.

[7] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "Gcb-net: Graph convolutional broad network and its application in emotion recognition," *IEEE Transactions on Affective Computing*, pp. 1–1, 2019.

[8] Y.-M. Jin, Y.-D. Luo, W.-L. Zheng, and B.-L. Lu, "Eeg-based emotion recognition using domain adaptation network," in *2017 International Conference on Orange Technologies (ICOT)*, pp. 222–225, 2017.

[9] 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*, pp. 1–1, 2019.

[10] P. Zhong, D. Wang, and C. Miao, "Eeg-based emotion recognition using regularized graph neural networks," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[11] Y. Ye, Z. Huang, T. Pan, J. Li, and H. T. Shen, "Reducing bias to source samples for unsupervised domain adaptation," *Neural Networks*, 2021.

[12] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on Machine Learning* (F. Bach and D. Blei, eds.), vol. 37 of *Proceedings of Machine Learning Research*, (Lille, France), pp. 1180–1189, PMLR, 07–09 Jul 2015.

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

[14] Y. Ye, Y. He, T. Pan, J. Li, and H. T. Shen, "Alleviating domain shift via discriminative learning for generalized zero-shot learning," *IEEE Transactions on Multimedia*, 2021.

[15] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, p. 293–300, June 1999.

[16] 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.