# A Novel Transferability Attention Neural Network Model for EEG Emotion Recognition

Yang Li, Boxun Fu, Fu Li*, Guangming Shi, Wenming Zheng



***Abstract*—The existed methods for electroencephalograph (EEG) emotion recognition always train the models based on all the EEG samples indistinguishably. However, some of the source (training) samples may lead to a negative influence because they are significant dissimilar with the target (test) samples. So it is necessary to give more attention to the EEG samples with strong transferability rather than forcefully training a classification model by all the samples. Furthermore, for an EEG sample, from the aspect of neuroscience, not all the brain regions of an EEG sample contains emotional information that can transferred to the test data effectively. Even some brain region data will make strong negative effect for learning the emotional classification model. Considering these two issues, in this paper, we propose a transferable attention neural network (TANN) for EEG emotion recognition, which learns the emotional discriminative information by highlighting the transferable EEG brain regions data and samples adaptively through local and global attention mechanism. This can be implemented by measuring the outputs of multiple brain-region-level discriminators and one single sample-level discriminator. We conduct the extensive experiments on three public EEG emotional datasets. The results validate that the proposed model achieves the state-of-the-art performance.**

***Index Terms*—EEG emotion recognition, transferable attention, brain region**

## I. Introduction

Emotion plays an important role in human daily life. It influences our rational decision-making, perception and cognition, and is essential in interpersonal communication [1]. Thus, it is necessary to make machines to understand human emotions in the field of human-computer interaction (HCI). To this end, the technology of emotion recognition provides a possible way for computers to capture human emotions, which is the first step to improve and humanize the interaction between humans and machines.

Generally, emotion recognition measures the emotional states by analyzing the data of bodily reactions under emotional conditions [2]. These reactions, including speech, facial expression and gesture, can adequately express our emotions under most circumstances. Nevertheless, these methods are subjective and cannot guarantee the authenticity of emotion [3]. Except for the above external methods, the internal physiological variables tend to be much close to the real emotions. Human brain, as the source of all the reactions, can reflect the mental activities including the emotion states. According to the studies of neurophysiology and psychology, EEG has the ability to record the brain neural activities, and can be used to decode the effective information of human emotional states [4], [5]. Consequently, EEG emotion recognition has received substantial attention from human-computer interaction and pattern recognition research communities in recent years [6], [7], [8].

Most EEG emotion recognition methods focus on two major tasks, i.e., EEG feature extraction and classification. The first task aims at seeking the discriminative emotion-related information from the raw EEG signals. EEG emotional signals usually consist of many neural processes and hence present a highly heterogeneous and nonstationary behavior [2]. Hence, how to extract the specific emotion information that contribute to the emotion recognition becomes a very important task. In [9], Jenke et al. summarized and evaluated all the existing EEG features extracted from time domain, frequency domain and time-frequency domain on their self-recorded EEG emotional dataset. The target of classification is modeling the correlation between the EEG emotional feature and the class labels, which leads to the interpretation of raw EEG emotional signals. Classification performance provides insight about how well a trained model can estimate the emotional state. Many advanced classification algorithms have been proposed over the years. For example, Zheng et al. [10] proposed a group sparse canonical correlation analysis method for simultaneous EEG channel selection and emotion recognition. Li et al. [8] fused the information propagation patterns and activation difference in the brain to improve emotional recognition. In [11], Alarcao and Fonseca summarized, reviewed and compared these works comprehensively.

Recently, many domain adaptation methods have been proposed to deal with EEG emotion recognition, especially in the subject-independent task, where the source and target data come from different subjects. These methods have significantly advanced the EEG emotion recognition task. For example, Zheng and Lu [12] evaluated four different domain adaptation approaches including Transfer component analysis (TCA) [13], Kernel Principle Analysis (KPCA) [14], Transductive Support Vector Machine (T-SVM) [15] and Transductive Parameter Transfer (TPT) [16] on SEED dataset, and find that the accuracy can be improved by 20% compared with the generic classifier. Lan et al. [17] made a comparative study on several state-of-the-art domain adaptation techniques on two EEG emotional datasets and the experiment results show that using domain adaptation technique can improve the accuracy significantly by 7.25% and 13.40% compared with

Yang Li, Boxun Fu, Fu Li and Guangming Shi are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, the School of Artificial Intelligence, Xidian University, Xian, 710071, China.*(*Corresponding author: Fu Li (E-mail: fuli@mail.xidian.edu.cn).)*

Wenming Zheng is with the Key Laboratory of Child Development and Learning Science (Ministry of Education), School of Biological Sciences and Medical Engineering, Southeast University, Nanjing, Jiangsu, 210096, China.

the baseline accuracy where no domain adaptation technique is used. In all the domain adaptation methods, the most well-established one is the domain adversarial neural network (DANN) [18], which constructs a two players mini-max game by using a domain discriminator that works adversarially with the feature extractor to generate the domain-invariable data representations. Li et al. adopted this setting and proposed a bi-hemisphere domain adversarial neural network (BiDANN) for EEG emotion recognition and achieved the state-of-the-art performance [19].

Nevertheless, we argue that the there are two issues need to be better addressed in EEG emotion recognition tasks. The first one is how to identify the positive EEG samples that consist of more emotion-related information. EEG emotional signals usually consist of many neural processes and are much vulnerable to negative effect of irrelevant knowledge, which incurs that some training EEG samples are significantly dissimilar with the test ones. Exploring how to highlight the positive EEG emotional samples and weaken the effect of negative samples will contribute more to emotion recognition. The second issue is how to weight the variability of different brain regions for EEG emotion recognition. Some studies of neuroscience have shown that different brain regions have different contributions for emotion expression [20]. In an EEG emotional sample, it is obvious that not all the brain regions contain the knowledge of emotion that can be transferred to the test samples. Making a strategy to distinguish the transferable and nontransferable brain regions is helpful to improve EEG emotion recognition.

To this end, in this paper, we propose a transferable attention neural network (TANN) to deal with the above tranferability learning problem for EEG emotion recognition. This transferability of data can be measured by calculating from the outputs of domain discriminators. Specifically, for the domain adversarial neural network [18], the output of domain discriminator is the probability of input data belongs to source or target domain. When the probability approaches 0, it represents the input data belongs to source domain, while approaching 1 indicates that it belongs to the target domain. Therefore, TANN takes advantages of the domain discriminator to measure the transferability from the training data to test data. Concretely, the framework of TANN includes the following three major modules:

- **Feature extractor.** The goal of feature extractor is to extract the high-level discriminative deep feature from raw EEG data for classification. EEG data is made of several electrodes that are set under the coordinates on the scalp, which are predefined referring to the locations of different brain regions. In the feature learning procedure, we should well retain this intrinsic structural information that will be helpful for classification. To achieve this, TANN employs two directional recurrent neural networks (RNN) that traverse all the electrodes from horizontal and vertical directions, which will construct a complete relationship and generate discriminative deep features for all the EEG electrodes.
- **Attention module.** The attention module aims to weight the input training data according to the level of transferability. For EEG emotional data, there is a large distribution gap between training and test data, resulting that some training EEG data are significantly dissimilar with the test. Moreover, from the aspect of neuroscience, not all the brain regions of an EEG sample contains emotional information that can transferred to the test data effectively. Therefore, TANN employs multiple brain-region-level and one sample-level discriminators to assess the transferability of EEG sample and the inside brain region data, then strengthen or weaken the contributions of these brain regions and samples for emotion classification.
- **Classifier.** Like most supervised learning methods, we introduce a classifier to predict the emotion class label based on the deep features obtained by the feature extractor. It will guide the feature extracting process towards generate more discriminative EEG features for emotion classification.

To the best of our knowledge, this is the first work to exploit the global and local transferability of EEG signals for emotion recognition. The experimental results verify the proposed TANN method can achieve the state-of-the-art performance on three public datasets.

## II. Preliminary

In this section, we briefly overview the preliminary of transferable attention and then address how we can apply it to EEG emotion recognition.

Most attention based methods focus on how to highlight or weaken different parts in an image according to their contribution for classification but neglect the evaluation for each training sample [21]. It is known that not all the training samples are similar with the test. It will be a negative influence in the learning process if we feed the model with all the training samples forcefully. Transferable attention (TA) is designed to deal with this problem [22]. When a training sample is much easier to be transferred to the test, it will be rewarded with more attention due to the high similarity with the test data, which is called transferable attention. Inspired by adversarial learning methods, this attention can be realized by calculating the outputs of the discriminator, which can reflect the similarity between training and test data.

Since in EEG emotion recognition tasks, not all the training EEG data are useful in the process of learning a model, exploring the transferability of EEG data will be meaningful and can further improve EEG emotion recognition.

## III. The proposed model for EEG emotion recognition

To specify the proposed method clearly, we illustrate the framework of the proposed TANN model in Fig. 1. TANN aims to distinguish which training samples are easy or hard to be transferred to test samples. Through penalizing these training samples, it can further improve EEG emotion recognition. Besides, considering not all the brain regions have the equal transferability, as well as measuring the similarity across EEG samples, TANN also focuses on the brain regions with

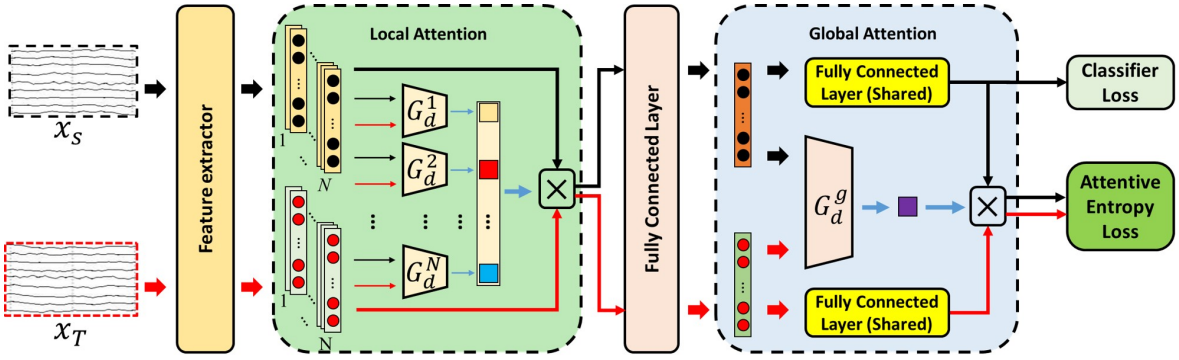



Fig. 1: The framework of TANN. TANN consists of two major modules, i.e., local and global attentions, that can make the model focus on the brain regions and samples with higher transferability.

high transferability. To achieve this goal, we adopt local and global attentions to the EEG emotion sample and its inside brain regions' data, respectively. These attention weights can be obtained from the outputs of multiple local and one global domain discriminators. Concretely, TANN consists of three major modules, i.e., feature extractor, attention layers, and classifier. In the following, we illustrate these parts detailedly.

### A. Feature extractor

The process of feature extraction is depicted in Fig. 2, and the goal is to represent the EEG emotional data in a more discriminative feature space so as to improve the EEG classification performance. The EEG deep features are extracted by two directional RNN modules that traverse the spatial regions under two predefined stacks, which are determined with respect to horizontal and vertical directions. These two directional RNNs are complementary to construct a complete relationship of electrodes locations that avoid losing the intrinsic structural information of EEG data. By doing this, we can obtain the high-level features for each EEG electrode that facilitate to construct the brain regions' features.

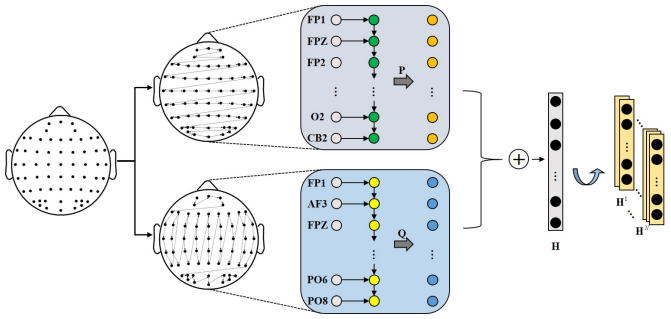



Fig. 2: The process of feature extraction. We first extract the deep feature for each electrode, and then rearrange them to form the data representation of brain regions.

Concretely, for an EEG sample $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, where $d$ and $n$ are the dimension and number of EEG electrode, the above process can be formulated as

$$\mathbf{s}_i^h = \sigma(\mathbf{U}^h \mathbf{x}_i^h + \sum_{j=1}^{n} e_{ij}^h \mathbf{V}^h \mathbf{h}_j^h + \mathbf{b}^h) \in \mathbb{R}^{d_f}, \quad e_{ij}^h = \begin{cases} 1, & \text{if } \mathbf{x}_j^h \in \mathcal{N}(\mathbf{x}_i^h), \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

$$\mathbf{s}_i^v = \sigma(\mathbf{U}^v \mathbf{x}_i^v + \sum_{j=1}^{n} e_{ij}^v \mathbf{V}^v \mathbf{h}_j^v + \mathbf{b}^v) \in \mathbb{R}^{d_f}, \quad e_{ij}^v = \begin{cases} 1, & \text{if } \mathbf{x}_j^v \in \mathcal{N}(\mathbf{x}_i^v), \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{s}_i^{\cdot}$ is the hidden unit of the RNN module as well as the data representation for the electrode $\mathbf{x}_i$, and $d_f$ is its dimension; $\{\mathbf{U}^{\cdot} \in \mathbb{R}^{d_f \times d}, \mathbf{V}^{\cdot} \in \mathbb{R}^{d_f \times d_f}, \mathbf{b}^{\cdot} \in \mathbb{R}^{d_f \times 1}\}$ are the learnable transformation matrices of RNN module; $\sigma(\cdot)$ denotes the nonlinear operation such as Sigmoid function; and $\mathcal{N}(\mathbf{x}_i^{\cdot})$ denotes the set of predecessors of node $\mathbf{x}_i^{\cdot}$.

Due to that TANN consists of horizontal and vertical directional RNNs to represent EEG electrode, we can obtain the data representations that not only contain the information of the electrodes itself but also the nearby relationship. Specifically, it can be expressed as $\mathbf{S}^h = \{\mathbf{s}_i^h\}$ that contains the information from left and right electrodes, and $\mathbf{S}^v = \{\mathbf{s}_i^v\}$ that includes the information from up and down electrodes. To integrate these spatial information into a overall representation, we arrange the order of the columns of $\mathbf{S}^h$ and $\mathbf{S}^v$, and use two transformation matrices $\mathbf{P}$ and $\mathbf{Q}$ to obtain the deep features $\mathbf{H} = \{\mathbf{h}_k\}$ for all the electrodes, in which

$$\mathbf{h}_i = \mathbf{P}\mathbf{s}_i^h + \mathbf{Q}\mathbf{s}_i^v + \mathbf{b} \in \mathbb{R}^{d_f'}, i \in \{1, \cdots, n\}. \tag{3}$$

Here $\mathbf{h}_i$ is the deep representation of electrode $\mathbf{x}_i$ that kepdf the location structural relation, $d_f$ is the dimension.

### B. Attention layers

For EEG emotion samples, there is a large distribution gap between training and test data. Some training samples are very dissimilar with the test ones. Therefore, to avoid training a model with all the source samples indiscriminately, TANN measures the transferability of all the training samples and then strengthen or weaken them in the learning process of the model. Besides, as we know, for emotion recognition, not all the brain regions of an EEG sample contains emotional

information that can transferred to the test data effectively. Some brain regions are more transferable than the others. Due to this, TANN not only employs a global attention layer to weight the sample-level transferability but also a local attention layer as a complement to focus on the brain-region-level transferability. Specifically, the transferability is quantified by the entropy of the outputs of domain discriminator. The domain discriminator can generate the probability of confusion between source (training) and target (test) data. When the probability approaches 0.5, it indicates that the input has good ability to confuse the domain discriminator, which nicely meet our need to highlight the data with positive transferability. In the following, we will demonstrate how to achieve the local and global attentions by transferability learning.

*1) Local transferable attention on brain-region-level:* After obtaining the data representation $\mathbf{h}_i$ of each electrode of $\mathbf{X}$, TANN employs local attention to highlight the brain regions with high transferability. Here we first group the electrodes into several clusters according to the associated brain region locations, which can be formulated as

$$\begin{aligned} &\text{brain region 1: } \mathbf{H}^1 = [\mathbf{h}_1^1, \mathbf{h}_2^1, \cdots, \mathbf{h}_{n_1}^1], \\ &\cdots \qquad\qquad \cdots \\ &\text{brain region N: } \mathbf{H}^N = [\mathbf{h}_1^N, \mathbf{h}_2^N, \cdots, \mathbf{h}_{n_N}^N], \end{aligned} \tag{4}$$

where $N$ is the number of brain regions, $n_c$ denotes the number of electrodes in the $c$-th brain region, $n_1 + \cdots + n_N = n$. In this case, the reordered deep feature can be expressed as

$$\hat{\mathbf{H}} = [\mathbf{H}^1, \cdots, \mathbf{H}^N]. \tag{5}$$

Based on the above process, we can obtain the deep features of all the brain regions from source and target EEG samples, which can be denoted as $\hat{\mathbf{H}}_S = [\mathbf{H}_S^1, \cdots, \mathbf{H}_S^N]$ and $\hat{\mathbf{H}}_T = [\mathbf{H}_T^1, \cdots, \mathbf{H}_T^N]$. Then they are fed to $N$ local discriminators to calculate the transferability. Concretely, let $\mathbf{d}^{N_i} = \{d_s^{N_i}, d_t^{N_i}\}$ denote the output probability of one discriminator for brain region $N_i$, where $d_s^{N_i}$ and $d_t^{N_i}$ are the probabilities that the input belongs to the source and target data, respectively. Then we can quantity the transferability of this brain region through the entropy function in information theory [22], which is defined as

$$H(\mathbf{d}^{N_i}) = -d_s^{N_i} \cdot \log(d_s^{N_i}) - d_t^{N_i} \cdot \log(d_t^{N_i}). \tag{6}$$

Then the higher transferability of a brain region has, the more attention value is.

However, for an EEG signal, the emotion information is the most difficult component to be transferred. Due to this, we reverse the attention values for the brain regions to make the model pay attention on the difficult transferred brain regions. Thus the attention value for brain region $N_i$ is defined as

$$w^{N_i} = 1 - H(\mathbf{d}^{N_i}). \tag{7}$$

Besides, to mitigate the negative effect of wrong attentions, we adopt the residual attention mechanism to make the model more robust. Thus, after local attention layer, the data representations for EEG sample $\mathbf{X}$ can be formulated as

$$\hat{\mathbf{H}}' = [(1 + w^1)\mathbf{H}^1, \cdots, (1 + w^N)\mathbf{H}^N] \in \mathbb{R}^{d_{f'} \times n}. \tag{8}$$

Here the loss function of the local discriminators for all the brain regions can be formulated as

$$L_d^l = \frac{1}{N} \sum_{N_i=1}^{N} L_d^{l_{N_i}}(\mathbf{X}^S, \mathbf{X}^T | \theta_d^{l_{N_i}}), \tag{9}$$

where

$$L_d^{l_{N_i}} = -\sum\nolimits_{t=1}^{M_1} \log p(0|\mathbf{X}_t^{S_{N_i}}) - \sum\nolimits_{t'=1}^{M_2} \log p(1|\mathbf{X}_{t'}^{T_{N_i}}) \tag{10}$$

denote the loss of the local discriminator for brain region $N_i$; $p(0|\mathbf{X}_t^{S_{N_i}})$ and $p(1|\mathbf{X}_{t'}^{T_{N_i}})$ are the probabilities of the input data belongs to source and target domains respectively; $\theta_d^{l_{N_i}}$ is the parameter of the local attention network; $\mathbf{X}_t^{S_{N_i}}$ and $\mathbf{X}_{t'}^{T_{N_i}}$ represent the $N_i$ brain region data of the $t$-th and $t'$-th source and target sample, respectively; $M_1$ and $M_2$ are the number of the source and target data.

*2) Global transferable attention on sample-level:* Although the above local attention for all the brain regions can make a fine-grained transfer learning between the source and target domain data, there is a possible that the local domain discriminator find fewer brain regions to transfer. Meanwhile, due to the distribution difference, there are some negative samples in the source data that are very dissimilar with the target data. It will weak the efficiency If we force training the model with these negative samples equaling with the other positive samples. Hence, after weighting the transferability of brain regions with local attention, we adopt the global transferable attention on the sample-level to transfer the knowledge from source to target domain.

Concretely, after local attention module, the input feature can be expressed as

$$\tilde{\mathbf{H}} = \hat{\mathbf{H}}'\mathbf{S} \in \mathbb{R}^{d_{f'} \times n'}, \tag{11}$$

where $\mathbf{S}$ is a learnable transformation matrix. Then it is sent to a global discriminator

$$\begin{aligned} L_d^g(\mathbf{X}^S, \mathbf{X}^T | \theta_d^g) = &-\sum\nolimits_{t=1}^{M_1} \log p(0|\mathbf{X}_t^S) \\ &-\sum\nolimits_{t'=1}^{M_2} \log p(1|\mathbf{X}_{t'}^T), \end{aligned} \tag{12}$$

to highlight the EEG samples with higher transferability, where $\theta_d^g$ is the parameter of the global attention network. Concretely, let $\mathbf{d} = \{d_s, d_t\}$ denote the output probability of the global discriminator, where $d_s$ and $d_t$ are the probabilities that the input belongs to the source and target data respectively. The global attention value $w$ can be calculated as

$$w = 1 + H(\mathbf{d}), \tag{13}$$

$$H(\mathbf{d}) = -d_s \cdot \log(d_s) - d_t \cdot \log(d_t). \tag{14}$$

Here we also adopt the residual mechanism to avoid the wrong attention. In this case, we obtain that the more transferability is, the larger attention value $w$ is.

Inspired by Long et al. [23], the entropy minimization principle can refine the classifier adaptation, which can increase the confidence of the classifier prediction. Thus, we utilize the global domain discriminator to generate the global attention values acting on the label entropy to enhance the certainty of the source samples that are more similar with the target

samples. Then $w$ is embedded into the label entropy loss to achieve the function for global attention. Hence the loss function of the label entropy, which is called attentive entropy loss, can be written as

$$L_e(\mathbf{X}^S, \mathbf{X}^T|\theta_e) = \sum_{k=1}^{M_1+M_2}\sum_{c=1}^{C} -w \cdot p(c|\mathbf{X}_k) \cdot \log p(c|\mathbf{X}_k), \quad (15)$$

where $\mathbf{X}_k$ is the $k$-th sample in $\{\mathbf{X}^S, \mathbf{X}^T\}$; $w$ is the global attention value for EEG sample $\mathbf{X}_k$; and $C$ is the number of emotion classes.

### C. Classifier

To enhance the discriminative ability of the model, we add the classifier to TANN model. Concretely, based on the final feature vector $\tilde{\mathbf{H}}$ in Eq. (11), we first arrange the matrix $\tilde{\mathbf{H}}$ into a vector $\tilde{\mathbf{h}}$, and then use the simple linear transform approach to predict the class label, which can be formulated as

$$\mathbf{O} = \mathbf{G}\tilde{\mathbf{h}} + \mathbf{b}_c = [o_1, \cdots, o_C], \quad (16)$$

where $\mathbf{G}$ and $\mathbf{b}^c$ are the transformation matrices. Finally, the output vector $\mathbf{O}$ is fed into the softmax layer for emotion classification, which can be written as

$$p(c|\mathbf{X}_t) = \exp(o_c) / \sum_{i=1}^{C} \exp(o_i), \quad (17)$$

where $p(c|\mathbf{X}_t)$ denotes the predicted probability that the input sample $\mathbf{X}_t$ belongs to the $c$-th class. As a result, the label $\tilde{l}$ of sample $\mathbf{X}_t$ is predicted as

$$\tilde{l} = arg\max_{c} p(c|\mathbf{X}_t). \quad (18)$$

Hence, the loss function of the classifier can be expressed as

$$L_c(\mathbf{X}^S|\theta_c) = \sum_{t=1}^{M_1}\sum_{c=1}^{C} -\tau(l,c) \cdot \log p(c|\mathbf{X}_t), \quad \tau(l,c) = \begin{cases} 1, & \text{if } l = c, \\ 0, & \text{otherwise}, \end{cases} \quad (19)$$

where $\theta_c$ denotes the parameter of the classifier.

### D. The optimization

In summary, the overall loss function includes four parts, i.e., local and global discriminator losses, classifier loss and the attentive entropy loss. Concretely, the loss function of the proposed TANN method can be formulated as

$$L(\mathbf{X}^S, \mathbf{X}^T|\theta_c, \theta_e, \theta_d^l, \theta_d^g) = L_c(\mathbf{X}^S|\theta_c) + \alpha L_e(\mathbf{X}^S, \mathbf{X}^T|\theta_e) - \beta\left(\frac{1}{N}\sum_{N_i=1}^{N} L_d^{l_{N_i}}(\mathbf{X}^S, \mathbf{X}^T|\theta_d^{l_{N_i}}) + L_d^g(\mathbf{X}^S, \mathbf{X}^T|\theta_d^g)\right), \quad (20)$$

where $\alpha$ and $\beta$ are the hyper-parameters, $L_d^{l_{N_i}}$ and $L_d^g$ represent the losses of local and global attention discriminators. Then we iteratively optimize the classifier, attentive entropy, local and global attention discriminators. Concretely, the parameters can be found through minimizing and maximizing

$$(\hat{\theta}_f, \hat{\theta}_c) = \arg\min_{\theta_f, \theta_c} L_c(\mathbf{X}^S|\theta_f, \theta_c, \hat{\theta}_e, \hat{\theta}_d^l, \hat{\theta}_d^g), \quad (21)$$

$$\hat{\theta}_e = \arg\min_{\theta_e} L_e(\mathbf{X}^S, \mathbf{X}^T|\hat{\theta}_f, \hat{\theta}_c, \theta_e, \hat{\theta}_d^l, \hat{\theta}_d^g), \quad (22)$$

$$\hat{\theta}_d^{l_{N_i}} = \arg\max_{\theta_d^{l_{N_i}}} L_d^{l_{N_i}}(\mathbf{X}^S, \mathbf{X}^T|\hat{\theta}_f, \hat{\theta}_c, \hat{\theta}_e, \theta_d^{l_{N_i}}, \hat{\theta}_d^g), \quad (23)$$

$$\hat{\theta}_d^g = \arg\max_{\theta_d^g} L_d^g(\mathbf{X}^S, \mathbf{X}^T|\hat{\theta}_f, \hat{\theta}_c, \hat{\theta}_e, \hat{\theta}_d^l, \theta_d^g). \quad (24)$$

The above maximization problem, i.e., Eq. (23) and (24), can be transferred to a minimization problem through adopting a gradient reversal layer (GRL) [18] before the discriminator, which will act as an identity transform in the forward-propagation but reverse the gradient sign while performing the back-propagation operation. Then we can use the stochastic gradient decent (SGD) algorithm to solve the parameter optimization process easily. Specifically, the parameters can be updated by the rules below

$$\theta_c \leftarrow \theta_c - \frac{\partial L_c}{\partial \theta_c}, \quad \theta_e \leftarrow \theta_e - \alpha \cdot \frac{\partial L_e}{\partial \theta_e}, \quad (25)$$

$$\theta_d^{l_{N_i}} \leftarrow \theta_d^{l_{N_i}} - \beta \cdot \frac{\partial L_d^{l_{N_i}}}{\partial \theta_d^{l_{N_i}}}, \quad \theta_d^g \leftarrow \theta_d^g - \beta \cdot \frac{\partial L_d^g}{\partial \theta_d^g}, \quad (26)$$

$$\theta_f \leftarrow \theta_f - \left(\frac{\partial L_c}{\partial \theta_f} + \alpha \cdot \frac{\partial L_e}{\partial \theta_f} - \beta \cdot \frac{\partial L_d^{l_{N_i}}}{\partial \theta_f} - \beta \cdot \frac{\partial L_d^g}{\partial \theta_f}\right). \quad (27)$$

## IV. Experiments

### A. Datasets and settings

To evaluate the proposed TANN method adequately, we conduct the experiments on three public EEG emotion datasets, namely,

(1) **SEED** [7] dataset is a standard benchmark for EEG emotion recognition. It contains three types of emotions, i.e., *happy*, *neutral* and *sad*, from 15 subjects' EEG emotional signals.

(2) **SEED-IV**[1] [24] dataset includes four types of emotions from 15 subjects. Compared with SEED, it contains an extra emotion *fear*.

(3) **MPED**[1] [25] dataset includes seven refined emotion types, i.e., *joy*, *funny*, *neutral*, *sad*, *fear*, *disgust* and *anger* from 30 subjects.

On these datasets, we design two kinds of EEG emotion recognition experiments including the subject-dependent and subject-independent ones. Table I summarizes the number of training and test samples, and the experimental protocols used in the experiments. The concrete protocols are described as follows:

- **The subject-dependent experiment** - In this experiment, the training and test data come from the same subject but different trials. We adopt the same protocols as [7], [24] and [26]. Namely, for SEED, we use the former nine trials of EEG data per session of each subject as source

[1]Note that both SEED-IV and MPED are multi-modal datasets. MPED consists of 30 subjects' EEG data, among which 23 subjects contain multi-modal data. In this experiment, we only use the EEG modal data.

(training) domain data while using the remaining six trials per session as target (test) domain data; for SEED-IV, we use the first sixteen trials per session of each subject as the training data, and the last eight trials containing all emotions (each emotion with two trials) as the test data; for MPED, we use twenty-one trials of EEG data as training data and the rest seven trials consisting of seven emotions as test data for each subject. The mean accuracy (ACC) and standard deviation (STD) are used as the evaluation criteria for all the subjects in the dataset.

- **The subject-independent experiment** - In this experiment, the training and test data come from different subjects, which is a harder task than the above subject-dependent one but more conductive to practical applications. We adopt the leave-one-subject-out (LOSO) cross-validation strategy [12] to evaluate the proposed TANN model. LOSO strategy uses the EEG signals of one subject as test data and the rest subjects' EEG signals as training data. This procedure is repeated such that the EEG signals of each subject will be used as test data once. Again, the mean accuracy (ACC) and standard deviation (STD) are used as the evaluation criteria.

Besides, we use the released handcraft features, namely, the differential entropy (DE) in SEED and SEED-IV, and the Short-Time Fourier Transform (STFT) in MPED, as the input to feed our model. Thus the sizes $d \times n$ of the input sample $\mathbf{X}_t$ are $5 \times 62$, $5 \times 62$ and $1 \times 62$ for these three datasets, respectively. Moreover, in the experiment, we respectively set the dimension $d_f$ and $d'_f$ of the feature extractor to 32; the number of brain region $N$ to 16[2]; the dimension $n'$ of the input for the global attention layer to 6; the hyper-parameters $\alpha$ and $\beta$ are both set to 0.1 throughout the experiment. Specifically, we implemented TANN using TensorFlow[3] on one Nvidia 1080Ti GPU. The learning rate, momentum and weight decay rate are set as 0.003, 0.9 and 0.95, respectively. The network is trained using SGD with batch size of 200.

### B. Experiment results

To validate the classification superiority of TANN, we also conduct the same experiments using various existed methods. Recall that the distribution gap in the subject-independent task is much larger than that in the subject-dependent one. In this case, domain adaptation methods shall be properly employed in order to achieve promising performance. Therefore, in the experiment on subject-independent task, we include many domain adaptation methods in the comparison. By doing so, we can effectively validate the state-of-the-art performance of our method. The comparable methods are listed as follows:

- Two baseline methods: linear support vector machine (SVM) [28], and random forest (RF) [29];

[2]Concretely, the brain regions include Pre-Frontal (AF3, FP1, FPZ, FP2, AF4), Frontal (F3, F1, FZ, F2, F4), Left Frontal (F7, F5), Right Frontal (F8, F6), Left Temporal (FT7, FC5, T7, C5, TP7, CP5), Right Temporal (FT8, FC6, T8, C6, TP8, CP6), Frontal Central (FC3, FC1, FCZ, FC2, FC4), Central (C3, C1, CZ, C2, C4), Central Parietal (CP3, CP1, CPZ, CP2, CP4), Left Parietal (P7, P5), Right Parietal (P8, P6), Parietal (P3, P1, PZ, P2, P4), Left Parietal Occipital (PO7, PO5, CB1), Right Parietal Occipital (PO8, PO6, CB2), Parietal Occipital (PO3, POZ, PO4), Occipital (O1, OZ, O2) lobes.

[3]https://www.tensorflow.org/

TABLE I: The number of training and test samples, and the experimental protocols used in the experiment.

(a) The subject-dependent experiment

| Dataset | | Training | Test | Protocol |
|---|---|---|---|---|
| SEED | | 2010 | 1384 | [Zheng and Lu][7] |
| SEED-IV | Session 1 | 561 | 290 | [Zheng et al.][24] |
| | Session 2 | 550 | 282 | |
| | Session 3 | 576 | 246 | |
| MPED | | 2520 | 840 | [Song et al.][27] |

(b) The subject-independent experiment

| Dataset | | Training | Test | Protocol |
|---|---|---|---|---|
| SEED | | 47516 | 3394 | [Zheng et al.][12] |
| SEED-IV | Session 1 | 11914 | 851 | LOSO* |
| | Session 2 | 11648 | 832 | |
| | Session 3 | 11508 | 822 | |
| MPED | | 97440 | 3360 | LOSO* |

$*$ LOSO denotes the leave-one-subject-out strategy.

- Three subspace learning methods: canonical correlation analysis (CCA) [30], group sparse canonical correlation analysis (GSCCA) [31], and graph regularization sparse linear regression (GRSLR) [32];
- Six transfer subspace learning methods: Kullback-Leibler importance estimation procedure (KLIEP) [33], unconstrained least-squares importance fitting (ULSIF) [34], selective transfer machine (STM) [35], transfer component analysis (TCA) [13], subspace alignment (SA) [36], and geodesic flow kernel (GFK) [37];
- Seven recent deep learning methods: deep believe network (DBN) [7], graph convolutional neural network (GCNN) [38], dynamical graph convolutional neural network (DGCNN) [25], domain adversarial neural networks (DANN) [18], bi-hemisphere domain adversarial neural network (BiDANN) [39], EmotionMeter [24], and attention-long short-term memory (A-LSTM) [26].

All the methods are representative ones in the previous studies of emotion recognition. We directly quote (or reproduce) their results from the literature to ensure a convincing comparison with the proposed method.

The results are summarized in Table II and III. Note that the subspace based methods, such as TCA, SA and GFK, are problematic to handle a large amount of EEG data due to the computer memory limitation and computational issue. Therefore, to compare with them, we have to randomly select 3000 EEG feature samples from the training data set to train these methods. Besides, the comparable methods adopting domain adaptation technique train the model with labeled training data and unlabeled test data as TANN does. From Table II and III, we have three observations:

(1) The proposed TANN model outperforms all the comparable methods on all the three datasets. Especially on SEED-IV dataset, the mean improvement is about 3.4% and 2.5% over the state-of-the-art methods A-LSTM and BiDANN. It verifies the learned transferable data representation are useful for EEG emotion recognition.

(2) The proposed TANN is superior to the recent domain adaptation methods. TANN has an improvement of 1.0%, 3.7% and 2.1% for subject-dependent task in Table II, and 1.2%, 2.4% and 2.5% for subject-independent task in Table III than the BiDANN method, which also adopts domain adversarial learning strategy to train the model. This reveals that the local and global attention structures are helpful to learn the discriminative information for emotion recognition.

(3) Even under the same classification models, the performance of the subject-independent tasks are quite lower than the subject-dependent ones. It is clear to see the gaps on three datasets are about 13%, 5% and 12%, respectively. This reveals that the individual difference is a negative influence on EEG emotion recognition, and should be mitigated in the subject-independent task.

TABLE II: The classification performance for subject-dependent EEG emotion recognition on SEED, SEED-IV and MPED datasets.

| Method | ACC / STD (%) | | |
|---|---|---|---|
| | SEED | SEED-IV | MPED |
| SVM [28] | 83.99/09.72 | 56.61/20.05* | 32.39/09.53* |
| RF [29] | 78.46/11.77 | 50.97/16.22* | 23.83/06.82* |
| CCA [30] | 77.63/13.21 | 54.47/18.48* | 29.08/07.96* |
| GSCCA [31] | 82.96/09.95 | 69.08/16.66* | 36.78/07.76* |
| DBN [7] | 86.08/08.34 | 66.77/07.38* | 35.07/11.25* |
| GRSLR [32] | 87.39/08.64 | 69.32/19.57* | 34.58/08.41* |
| GCNN [38] | 87.40/09.20 | 68.34/15.42* | 33.26/06.44* |
| DGCNN [25] | 90.40/08.49 | 69.88/16.29* | 32.37/06.08* |
| DANN [18] | 91.36/08.30 | 63.07/12.66* | 35.04/06.52* |
| BiDANN [39] | 92.38/07.04 | 70.29/12.63* | 37.71/06.04* |
| EmotionMeter [24] | – | 70.59/17.01 | – |
| A-LSTM [27] | 88.61/10.16* | 69.50/15.65* | 38.99/07.53* |
| TANN | **93.34/06.64** | **73.94/13.65** | **39.82/07.98** |

∗ indicates the experiment results obtained are based on our own implementation.
– indicates the experiment results are not reported on that dataset.

## C. Discussion

*1) The confusion of different emotions based on TANN model:* To better understand the confusion of TANN in recognizing different emotions, we depict the confusion matrices of subject-dependent and subject-independent EEG emotion recognition experiments in Fig. 3 and 4, respectively, from which we have the following observations:

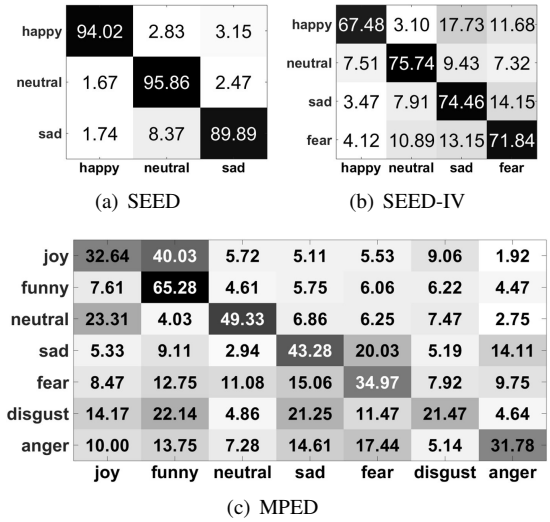
(1) In Fig. 3, for SEED, the classification accuracies for three emotions are about 90%, and the happy and neutral emotions are easier to be recognized than the sad emotion; for SEED-IV, which consists of four emotions, we can see the negative emotions, i.e., sad and fear, are confused by the classifier with higher possibility; and for MPED, the confusion is more complex because it has more emotions than the other two datasets. It is obvious to see that the funny emotion is the easiest to be recognized and has 16% more than the neutral emotion on the second place. Except this, we can find that the funny and joy are easier to be confused maybe because both of them are positive emotions.

TABLE III: The classification performance for subject-independent EEG emotion recognition on SEED, SEED-IV and MPED datasets.

| Method | ACC / STD (%) | | |
|---|---|---|---|
| | SEED | SEED-IV | MPED |
| KLIEP [33] | 45.71/17.76 | 31.46/09.20* | 18.92/04.54* |
| ULSIF [34] | 51.18/13.57 | 32.99/11.05* | 19.63/03.81* |
| STM [35] | 51.23/14.82 | 39.39/12.40* | 20.89/03.62* |
| SVM [28] | 56.73/16.29 | 37.99/12.52* | 19.66/03.96* |
| TCA [13] | 63.64/14.88 | 56.56/13.77* | 19.50/03.61* |
| SA [36] | 69.00/10.89 | 64.44/09.46* | 20.74/04.17* |
| GFK [37] | 71.31/14.09 | 64.38/11.41* | 20.27/04.34* |
| A-LSTM [27] | 72.18/10.85* | 55.03/09.28* | 24.06/04.58* |
| DANN [18] | 75.08/11.18 | 47.59/10.01* | 22.36/04.37* |
| DGCNN [25] | 79.95/09.02 | 52.82/09.23* | 25.12/04.20* |
| DAN [40] | 83.81/08.56 | 58.87/08.13 | – |
| BiDANN [39] | 83.28/09.60 | 65.59/10.39* | 25.86/04.92* |
| TANN | **84.41/08.75** | **68.00/08.35** | **28.32/05.11** |

∗ indicates the experiment results obtained are based on our own implementation.
– indicates the experiment results are not reported on that dataset.

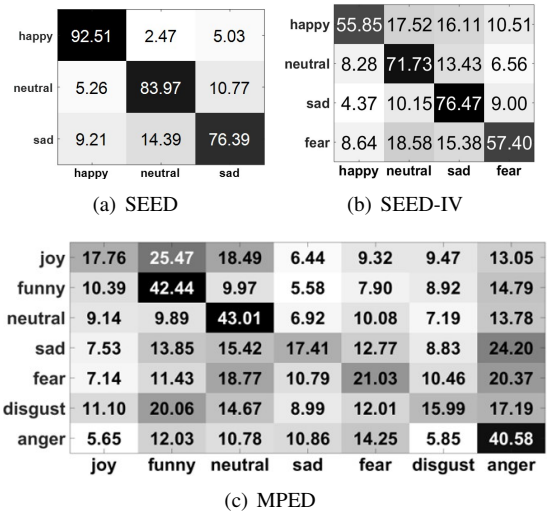
(2) From the results of subject-independent EEG emotion recognition experiment in Fig. 4, we can observe that, for SEED, which has three types of emotions, the happy emotion is much easier to be recognized than neutral and sad; for SEED-IV, the neutral and sad emotions are much easier to be recognized; for MPED, which is a hard seven classification problem, the accuraries of funny, neutral and anger emotions overpass that of the other emotions, and this reveals that we should focus on the joy, sad, fear and disgust emotion data in the task of classifying seven emotions.




(a) SEED (b) SEED-IV

(c) MPED

Fig. 3: The confusion matrices based on the subject-dependent experimental results on three datasets.




(a) SEED (b) SEED-IV

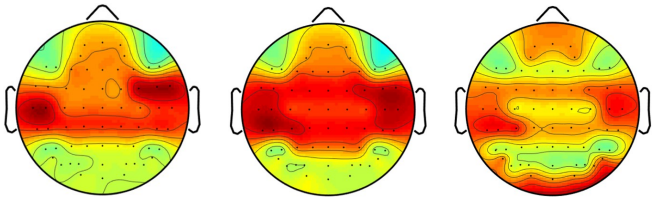
(c) MPED

Fig. 4: The confusion matrices based on the subject-independent experimental results on three datasets.

*2) The transferability of different brain regions:* To investigate the transferability of different brain regions for EEG emotion recognition, we visualize all the brain regions by mapping the local attention values $w$ in Eq. (7) into the corresponding electrodes. The obtained results are shown in Fig. 5, from which we have two observations:

(1) The left and right temporal lobes make more important contribution for emotion recognition in all the three datasets, which coincides with the previous EEG emotion studies [6], [7]. This also reveals that, as well as the proposed model can adaptively give attention to different brain regions, it is still effective to capture the most important ones.

(2) The activation areas are slightly different across datasets. For example, there is a broader activation to the temporal lobes for SEED-IV compared with SEED. And for MPED, which consists of more types of emotions, the occipital lobe, as well as the temporal lobe, contributes more for emotion expression.

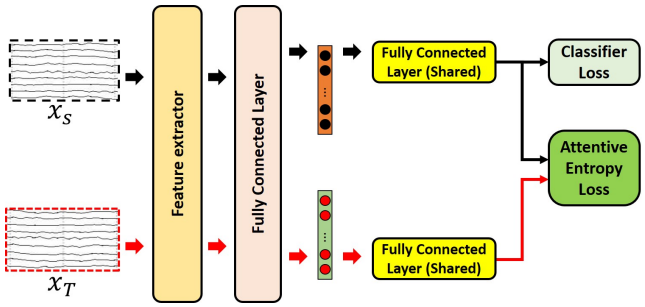


(a) SEED (b) SEED-IV (c) MPED

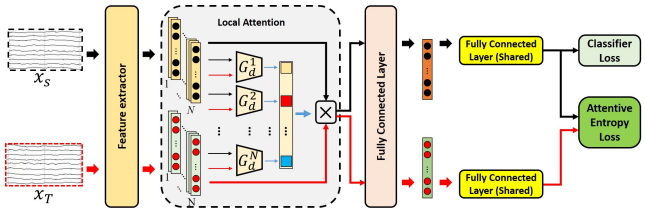
Fig. 5: The transferability of different EEG brain regions.

*3) Ablation study:* To see the importance of each module of TANN for EEG emotion recognition, we conduct an ablation study by removing the local and global attention layers both and separately. These reduced models are depicted in Fig. 6, which includes

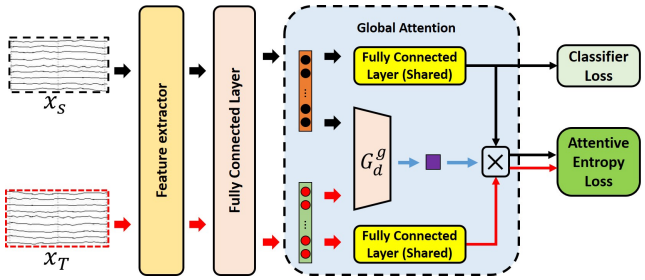
- TANN-R1, which removes both the local and global attention modules;
- TANN-R2, which neglects the global transferability for EEG samples;
- TANN-R3, which employs the same structure of TANN model except the local attention layer.




(a) TANN-R1




(b) TANN-R2




(c) TANN-R3

Fig. 6: The frameworks of the reduced models of TANN: (a) TANN-R1, (b) TANN-R2, (c) TANN-R3.

The experimental results are shown in Table IV, from which we can have three observations:

(1) It is effective for the structure of the feature extractor in TANN. From the results of TANN-R1, we can see it achieves comparable performance on three datasets. This verifies the obtained deep data representation by two directional recurrent neural networks is discriminative for emotion recognition.

(2) Either the local or global transferable attention modules can enhance emotion recognition. In contrast to TANN-R1, TANN-R2 and TANN-R3 improve the accuracy, on average, by 1.8% and 1.5% on three datasets, respectively.

(3) By assembling the feature extractor, local and global attention modules, TANN achieves the best performance. We can see TANN has a further improvement of 3% compared with TANN-R2 and TANN-R3.

The above results verify the effectiveness of the three important modules in TANN.

TABLE IV: The comparison of EEG emotion recognition results among four methods: (1) TANN-R1, (2) TANN-R2, (3) TANN-R3; (4) TANN.

| Method | ACC / STD (%) | | |
|---|---|---|---|
| | SEED | SEED-IV | MPED |
| TANN-R1 | 87.06/09.45 | 68.28/14.28 | 37.92/07.80 |
| TANN-R2 | 89.73/07.53 | **70.82/14.65** | **38.10/07.98** |
| TANN-R3 | **91.03/07.63** | 68.72/13.30 | 38.06/08.21 |
| TANN | **93.34/06.64** | **73.94/13.65** | **39.82/07.98** |

## V. Conclusion

In this paper, we propose a transferable attention neural network (TANN) to deal with EEG emotion recognition problem, which is motivated by the finding that not all the training samples have the equal contribution for emotion recognition, which also happens for the importance of different brain regions in this sample. TANN has the ability to learn the positive and negative information from the sample-level and brain-region-level, which can improve EEG emotion recognition. The proposed framework is easy to implement and the extensive experiments on three public EEG emotion datasets demonstrated that the proposed TANN method achieves the state-of-the-art performance. Besides, based on TANN, we also investigate the transferability of different brain regions in EEG emotion recognition and find that the temporal lobe and occipital lobe contribute more for emotion expression. In the future work, we will further investigate more operations for learning the transferability information to explore the potential efficacy of transferable attention for EEG emotion recognition.

## References


[1] R. W. Picard, *Affective computing*. MIT press, 2000.

[2] B. García-Martínez, A. Martinez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, "A review on nonlinear methods using electroencephalographic recordings for emotion recognition," *IEEE Transactions on Affective Computing*, 2019.

[3] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang, "Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44 317–44 328, 2019.

[4] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.

[5] D. Mathersul, L. M. Williams, P. J. Hopkinson, and A. H. Kemp, "Investigating models of affect: relationships among eeg alpha asymmetry, depression, and anxiety," *Emotion*, vol. 8, no. 4, pp. 560–572, 2008.

[6] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "Eeg-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[7] 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.

[8] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang *et al.*, "Eeg based emotion recognition by combining functional connectivity network and local activations," *IEEE Transactions on Biomedical Engineering*, 2019.

[9] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[10] W. Zheng, "Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 3, pp. 281–290, 2017.

[11] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using eeg signals: a survey," *IEEE Transactions on Affective Computing*, 2017.

[12] W.-L. Zheng and B.-L. Lu, "Personalizing eeg-based affective models with transfer learning," in *International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI Press, 2016, pp. 2732–2738.

[13] 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.

[14] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[15] R. Collobert, F. Sinz, J. Weston, and L. Bottou, "Large scale transductive svms," *Journal of Machine Learning Research*, vol. 7, no. Aug, pp. 1687–1712, 2006.

[16] E. Sangineto, G. Zen, E. Ricci, and N. Sebe, "We are not all equal: Personalizing models for facial expression analysis with transductive parameter transfer," in *Proceedings of the 22nd ACM international conference on Multimedia (MM)*. ACM, 2014, pp. 357–366.

[17] 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, 2018.

[18] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[19] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for eeg emotion recognition," *IEEE Transactions on Affective Computing*, 2018.

[20] P. A. Kragel and K. S. Labar, "Decoding the nature of emotion in the brain," *Trends in Cognitive Sciences*, vol. 20, no. 6, pp. 444–455, 2016.

[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," pp. 5998–6008, 2017.

[22] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, "Transferable attention for domain adaptation," 2019.

[23] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 136–144.

[24] 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, pp. 1110–1122, 2019.

[25] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, 2018.

[26] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "Mped: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12 177–12 191, 2019.

[27] ——, "Mped: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12 177–12 191, 2019.

[28] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[30] B. Thompson, "Canonical correlation analysis," *Encyclopedia of Statistics in Behavioral Science*, 2005.

[31] W. Zheng, "Multichannel eeg-based emotion recognition via group sparse canonical correlation analysis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 3, pp. 281–290, 2017.

[32] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, "Eeg emotion recognition based on graph regularized sparse linear regression," *Neural Processing Letters*, pp. 1–17, 2018.

[33] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *Advances in Neural Information Processing Systems (NIPS)*, 2008, pp. 1433–1440.

[34] T. Kanamori, S. Hido, and M. Sugiyama, "A least-squares approach to direct importance estimation," *The Journal of Machine Learning Research*, vol. 10, pp. 1391–1445, 2009.

[35] W.-S. Chu, F. De la Torre, and J. F. Cohn, "Selective transfer machine for personalized facial expression analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 3, pp. 529–545, 2017.

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

[37] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2012, pp. 2066–2073.

[38] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in Neural Information Processing Systems (NIPS)*, 2016, pp. 3844–3852.

[39] Y. Li, W. Zheng, Z. Cui, T. Zhang, and Y. Zong, "A novel neural network model based on cerebral hemispheric asymmetry for eeg emotion recognition." in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 1561–1567.

[40] 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*. Springer, 2018, pp. 403–413.