A Novel Transfer Learning Framework with Prototypical Representation based Pairwise Learning for Cross-Subject Cross-Session EEG-Based Emotion Recognition

Rushuang Zhou\textsuperscript{a,b,1,*}, Zhiguo Zhang\textsuperscript{a,b,c,d,1†}, Xin Yang\textsuperscript{a,b†}, Hong Fu\textsuperscript{a,b,§}, Li Zhang\textsuperscript{a,b,‡}, Linling Li\textsuperscript{a,b,∥}, Gan Huang\textsuperscript{a,b,**}, Yining Dong\textsuperscript{f,††}, Fali Li\textsuperscript{g,h,**}, and Zhen Liang\textsuperscript{a,b,††}

\textsuperscript{a}School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China
\textsuperscript{b}Guangdong Provincial Key Laboratory of Biomedical Measurements and Ultrasound Imaging, Shenzhen, China
\textsuperscript{c}Marshall Laboratory of Biomedical Engineering, Shenzhen, China
\textsuperscript{d}School of Data Science, City University of Hong Kong, Hong Kong
\textsuperscript{e}Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong
\textsuperscript{f}School of Life Science and Technology, Center for Information in Medicine, University of Electronic Science and Technology of China, China
\textsuperscript{g}School of Biomedical Engineering, Shenzhen, China
\textsuperscript{h}The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for Neuroinformation, University of Electronic Science and Technology of China, China
\textsuperscript{i}Peng Cheng Laboratory, Shenzhen, China

Email: *2018222087@szu.edu.cn, †zgzheng@szu.edu.cn, ‡yangxinknow@gmail.com, §hfu@eduhk.hk, ¶lijiang@szu.edu.cn, **huanggan@szu.edu.cn, ′′yinidong@cityu.edu.hk, ††fali.li@uestc.edu.cn, ‡‡janezliang@szu.edu.cn

\textbf{Abstract}—Affective brain-computer interfaces based on electroencephalography (EEG) gradually become an important branch in affective computing studies. However, the individual differences and the noisy labels limit the effectiveness and generalizability of EEG models in cross-subject cross-session emotion recognition tasks. In this paper, we propose a novel transfer learning framework with Prototypical Representation based Pairwise Learning (PR-PL) for EEG-based emotion recognition, which learns discriminative and generalized prototypical representations for emotion revealing across individuals and formulates emotion recognition as pairwise learning to alleviate the model reliance on precise label information. Prototypical learning based adversarial discriminative domain adaptation method is developed to encode the inherent emotion-related semantic structure of EEG data. Pairwise learning with an adaptive pseudo-labeling method is proposed to achieve a reliable and stable model learning with noisy labels. Through domain adaptation, besides aligning the feature representation of the source and the target on a shared feature space, the feature separability of both source and target domains is also considered. The characterized prototypical representations are evident with a high feature concentration within one single emotion category and a high feature separability across different emotion categories. Extensive experiments are conducted on two benchmark databases (SEED and SEED-IV) for recognizing three and four emotion categories using four cross-validation evaluation protocols. The experimental results demonstrate the superiority of the proposed PR-PL against the state-of-the-art methods on cross-subject cross-session, cross-subject within-session, within-subject cross-session, and within-subject within-session evaluation protocols, which shows the power of PR-PL in dealing with the ambiguity of neurophysiological responses in affective studies.

\textbf{Index Terms}—EEG; Emotion Recognition; Prototypical Representation; Pairwise Learning; Transfer Learning.

\section{INTRODUCTION}

AFFECTIVE computing is a growing interdisciplinary research field attracting researchers' attention from different areas including computer science, neuroscience, psychology, and signal processing \cite{1}. Recently, EEG-based emotion recognition studies have become an increasingly important topic for affective computing and human sentiment analysis \cite{2, 3}. A proper design of EEG-based emotion recognition models is helpful for facilitating the data processing, benefiting to discriminant feature characterization, and lightening the model performance. Currently, there exist two main critical issues in EEG-based emotion recognition models. One is individual differences eliminating: build a generalized affective computing model which could be tolerance enough to the individual differences in the simultaneously collected EEG signals; and another is noisy label learning: train a reliable and stable affective computing model which is less reliance on the subjective feedbacks.

In recent years, more and more researchers have begun to focus on applying transfer learning methods to alleviate the individual differences in EEG signals \cite{4–9}. Considering the individuals with and without labels (termed as source domain and target domain), the transfer learning methods try to minimize the distribution difference between the source and target domains, approximately satisfy the assumption of independent and identical distribution, and realize a higher recognition performance on the target domain. Through a domain-shifting strategy, the invariant feature representations across different domains are learned and the relationships

\footnote{Equal contributions.}
among the learned features, data distribution, and labels are explored. For example, Li et al. [5] proposed a multisource transfer learning method with two transfer learning stages. In the first stage, appropriate samples were selected from the existing source domain. In the second stage, a style transfer mapping was implemented to alleviate the differences between the selected source samples and the unknown target samples. The results showed that the proposed transfer method outperformed the non-transfer method with an improvement of 12.72% on the public SEED database [10] with a three-class classification problem (negative, neutral, and positive). Inspired by neuroscience findings that different emotions would lead to different brain reactions, Li et al. [6] proposed a novel R2G-STNN network to integrate the EEG spatial-temporal dynamics at the local and global brain areas and realize an efficient emotion recognition performance together with a domain shift learning. More details about current EEG-based emotion recognition models with the transfer learning algorithms are presented in Section II.

For video-evoking EEG-emotion experiments, subjects may not always be able to accurately react to the intended emotions, and at the same time may not be able to accurately describe and feedback their emotional changes. This would bring label noise to the emotional information annotation of EEG samples and further lead to a negative impact on the model performance [11]. To tackle this issue, Zhong et al. [8] developed an emotion-aware distribution learning method (RGNN), in which they blurred the label information by changing the one-hot label representation \( \{1, 0, 0\} \) to \( \{1 - \frac{2\epsilon}{3}, 0, 0\} \) and trained the model to be less sensitive label noise. However, the model performance would greatly rely on the selection of \( \epsilon \) value and an optimal \( \epsilon \) value selection could be different for different databases and different individuals. Current EEG-based emotion recognition models are mainly based on pointwise learning, which heavily rely on precisely labeled data.

To further improve EEG-based model effectiveness and generalizability in the emotion recognition task and eliminate the negative effects from individual differences and label noises, in this paper, we formulate the emotion recognition tasks as a pairwise learning problem and propose a novel transfer learning framework with prototypical representation based pairwise learning (which is termed as PR-PL below). Here, we model the relative relationship between pairs of EEG samples in terms of prototypical representations, which is advantageous to pointwise learning when the labeling task is difficult and even the provided labels are wrong labels [12]. The major novelties of the proposed PR-PL are summarized as follows. (1) We formulate emotion recognition as pairwise learning to replace the classifier and greatly alleviate the label dependence on emotion labels. The pairwise learning provides us an alternative way to measure whether two EEG signals belong to the same emotion category without the reliance on the precise labeling information. The extensive experimental results on two well-known emotional databases (SEED [10] and SEED-IV [13]) prove the proposed PR-PL is the most accurate model in solving the EEG-based emotion recognition tasks under different application environments (cross-subject cross-session, cross-subject single-session, within-subject cross-session, and within-subject single-session). (2) We propose a novel prototypical learning based adversarial domain adaptation method to explore latent variables of emotion categories, encode the semantic structure of EEG data, and learn subject-generalized prototypical representations for emotion revealing across individuals. The characterized prototypical representations show a high feature concentration within one single emotion category and a high feature separability across different emotion categories. (3) Different from the existing transfer learning methods that only focus on feature separability in the source domain, we consider the feature separability of both source and target domains through the end-to-end domain adversarial training to further enhance the model effectiveness and generalizability.

II. RELATED WORK

The existing EEG-based emotion recognition models with the transfer learning algorithms can be generally categorized into two types. (a) Non-deep transfer learning models. Pan et al. [14] proposed a transfer component analysis (TCA) algorithm to reduce the marginal distribution difference between the source and target domains, in which the transfer information was learned in a reproducing kernel Hilbert space through maximizing mean discrepancy. Zheng and Lu [15] introduced two types of subject-to-subject transfer methods to deal with the challenge of the individual differences in EEG signal processing. One was to explore a shared common feature space underlying source and target domains using TCA and kernel principal analysis (KPCA), and another was to construct multiple personalized classifiers on the source domain and map the classifier parameters to the target domain using transductive parameter transfer (TPT). These non-deep transfer learning strategies show the possibility to bridge the discrepancy across two domains with improved performance on the target domain. However, due to the small capacity and low complexity, the model accuracy and stability are still limited, which fails to satisfy the requirements of affective brain-computer interfaces (aBCI) in practical applications. (b) Deep transfer learning models. Most of the existing affective models are based on deep transfer learning methods built with domain-adversarial neural network (DANN) proposed in [16]. The main idea of DANN is to find a shared feature representation for source and target domains with indistinguishable distribution differences and also maintain the predictive ability of the estimated features on the source samples for a specific classification task. Li et al. [17] was the first to introduce DANN in aBCI. Benefitting from the powerful feature representation ability of deep networks and the high efficiency of adversarial learning in distributed adaptation, the results showed DANN based aBCI system was significantly superior compared to other methods. The following aBCI systems could be considered as a series of DANN based models, which generally start from two directions to improve the DANN performance in solving EEG-based emotion recognition tasks. (1) Incorporating the prior knowledge of neuroscience and brain anatomy with DANN. Inspired by the neuroscience findings of the asymmetry property of the left and right hemispheres in emotional...
responses, Yang et al. [18] proposed a bi-hemisphere domain adversarial neural network (BiDANN), in which a global and two local domain discriminators were designed to learn discriminant features from each cerebral hemisphere related to emotion perception and also improve the feature stability to the variation of different domains. The experiments on the SEED database demonstrated BiDANN achieved higher emotion recognition performance than DANN. Considering the emotional responses from different brain regions would be varied, Yang et al. [6] proposed an R2G-STNN (regional to global-spatial-temporal neural network) to integrate the spatial-temporal information from local and global brain regions under importance guidance and characterize hierarchical feature representations. Similarly, under an assumption that not all EEG channels are same important in emotion recognition tasks, Du et al. [19] integrated attention mechanism, long-short-term memory (LSTM), and DANN to propose an attention-based LSTM with domain discriminator (ATDD-LSTM) and characterize the nonlinear relations among different EEG channels in a data-driven approach and optimally select informative emotion-related channels. (2) 
Incorporating the probability distribution with DANN. To deal with the training instability of DANN, Luo et al. [20] introduced WGAN-GP (Wasserstein generative adversarial network with gradient penalty) to narrow down the distance between the marginal probability distributions of different subjects. The results showed the model stability was improved and a better cross-subject EEG-based emotion recognition was achieved. However, current DANN based models only consider the marginal distribution differences but ignore the joint distribution differences of different domains. To address this problem, Li et al. [21] introduced a joint domain adaptation network (JLNN), where the joint distribution adaptation (JDA) method was incorporated with a unified framework of task-invariant features (MDA) and task-specific features (CDA). Although the above models have achieved higher accuracies compared to the original model with DANN in emotion recognition tasks, there still exist three major technical challenges. First, the learned feature representation is susceptible to noise interference from both source and target domains and would further affect the model generalizability [22, 23]. Second, the existing models only focus on the feature separability in the source domain but ignore the feature separability in the target domain. The current DANN based models mainly concern the emotion classification loss in the source domain, which would lead to the over-fitting of source domain data and the decrease of classification ability on target domain data. Third, the existing algorithms largely rely on a large amount of labeled source domain data. However, in practical EEG applications, it is difficult to collect accurate labels for each single EEG trial.

III. METHODOLOGY

Suppose the EEG trials in the source domain \( S \) and target domain \( T \) are given as \( D_s = \{ X_s, Y_s \} \) and \( D_t = \{ X_t, Y_t \} \), where \( X_s, Y_s = \{ (x_{si}, y_{si}) \}_{i=0}^{N_s} \) and \( X_t, Y_t = \{ (x_{ti}, y_{ti}) \}_{i=0}^{N_t} \). Here, \( X_s \) and \( X_t \) are EEG samples, and \( Y_s^t \) and \( Y_t^s \) are the corresponding emotion labels. To make the narrative clearer, the frequently used notations are summarized in Table I. As shown in Fig. 1, the proposed PR-PL includes three losses (domain adversarial loss, pairwise learning loss on source domain, and pairwise learning loss on target domain) and two main parts (prototypical representation and pairwise learning). In the prototypical representation, three types of features are defined. The characterized features \( f(X_s) \) or \( f(X_t) \) from the EEG samples are termed as sample features, which are forced to be as indistinguishable as possible from the source and target domains. Under an assumption that any emotion could be represented by a prototype via prototypical learning, the prototype features of each emotion category are learned based on the sample features \( f(X_s) \) and \( Y_s \) from the source domain. The interaction relationships between the sample features and prototype features are measured and the interaction features are characterized which will be used in the following pairwise learning. In the pairwise learning, the pair relationships on both source and target domains are explored. As the information about \( Y_t \) is unknown during model training, an adaptive thresholding method is developed for valid pair selection and pseudo label generation.

A. Sample feature extraction

To make both source and target data satisfy the assumption of independent and identical distribution and obey the same distribution, we characterize the sample features based on a domain adversarial training introduced in DANN. Here, the distribution difference between the source domain and the target domain is alleviated and the sample features with domain invariant properties are characterized. Through this process, the individual differences in EEG signals could be alleviated and the generalization ability of the models could be improved [6, 18–20]. Specifically, the domain difference between the source domain sample feature \( f(X_s) \) and the target domain sample feature \( f(X_t) \) are minimized by adopting domain adversarial training. Here, \( f(\cdot) \) is the designed feature extractor for extracting the sample features from EEG signals. A discriminator network \( d(\cdot) \) with the parameter \( \theta_d \) is introduced to distinguish whether the characterized sample features \( (f(X_s) \) or \( f(X_t) \)) come from the source domain \( (S) \) or target domain \( (T) \).
or the target domain (T). Its loss function is a standard two-category cross-entropy loss function, given as

$$L_{\text{disc}}(\theta_f, \theta_d) = - \sum_{i=0}^{N_s} \log d(f(x_i^s)) - \sum_{i=0}^{N_t} \log (1 - d(f(x_i^t)))$$

(1)

In the training process, we adopt the end-to-end training method [16], and implement the domain adversarial training by introducing a gradient reversal layer. The feature extractor $f(\cdot)$ maximizes the classification ability to enhance emotion recognition performance and at the same time, the discriminator $d(\cdot)$ minimizes the domain discrimination to reduce the distribution difference between the source domain and target domain. The final domain adversarial training objective function is defined as

$$\min_{\theta_f} \max_{\theta_d} L_{\text{classifier}}(\theta_f) - \lambda L_{\text{disc}}(\theta_f, \theta_d),$$

(2)

where $L_{\text{classifier}}(\theta_f)$ is the classification loss to measure the classification ability in source domain. In this paper, $L_{\text{classifier}}(\theta_f)$ will be realized by pairwise learning on source and target domains introduced in Section III-D and III-E below. $L_{\text{disc}}(\theta_f, \theta_d)$ is the adversarial loss for the discriminator to be trained to distinguish the sample features characterized from source and target domains. $\theta_f$ and $\theta_d$ are the parameters of $f(\cdot)$ and $d(\cdot)$. $\lambda$ is a balanced hyperparameter for ensuring the stability of domain adversarial, which is given by an exponential growth method as

$$\lambda = \frac{2}{1 - \exp(-p)} - 1,$$

(3)

Here, $p$ is a factor related to the training round, given by a ratio of the current training round to the maximum training round.

### B. Prototype feature extraction

We assume that there exists a prototype for each emotion category. Through prototypical learning, the prototype features are learned to indicate the representation property of every single emotion category. Based on the sample features extracted from different subjects under different emotions, we could consider these sample features are distributed around the prototype features. In other words, for each emotion category, the prototype features could be considered as the “center of mass” of all the sample features. From the perspective of a probability distribution, the prototype feature of an emotion category can be regarded as the mean value of the sample feature distribution of the emotion, and the variance of the distribution is caused by the non-stationary of EEG, including but not limited to individual differences. Assume that the sample features under an emotion category $c$ obey the Gaussian distribution $N(\mu_c, \sigma_c^2)$. The prototype features of the emotion category $c$ could be calculated as the mean vector $\mu_c$ of the distribution. For the source domain data $\{X_s, Y_s\} = \{(x_i^s, y_i^s)\}_{i=0}^{N_s}$, the corresponding sample features are characterized by the feature extractor $f(\cdot)$, given as $f(x_i^s)$ (defined in Section III-A). The prototype feature vector of the emotion category $c$ can be calculated by averaging all the sample features that belong to this category, given as

$$\mu_c = \frac{1}{|X_s^c|} \sum_{x_i^s \in X_s^c} f(x_i^s),$$

(4)

where $X_s^c = \{(x_i^s, y_i^s = c)\}_{i=0}^{N}$ are a collection of source domain data belonging to the emotion category $c$. $|X_s^c|$ are the corresponding sample size in this emotion category. In other words, $\mu_c$ could be expressed as the centroid of the sample features of $X_s^c$. The mean value calculation is a widely used,
simple, and effective noise reduction strategy, which can make the prototype features stronger than the sample features and help to alleviate the problem of the traditional DANN network being susceptible to related noise interference \cite{24}. It is worth noting that since the calculation of the prototype features needs to use emotional label information, we only use the source domain data here for prototype feature extraction.

C. Interaction feature extraction

The traditional DANN extracted domain-invariant shared feature representations could be easily contaminated by the shared and related noises in source and target domains. In this paper, we introduce a bilinear interaction to measure the interaction relationships between the sample features and prototype features and extract the interaction features for the following pairwise learning. For a given sample feature \( f(x_i) \), its interaction relationship to a certain prototype feature \( \mu_i(c) \) can be measured by a bilinear transformation \( h(\cdot) \), defined as

\[
h(f(x_i), \mu_i) = f(x_i)^T S \mu_i, \tag{5}
\]

where \( S \in R^{n \times n} \) is a bilinear transformation matrix with trainable parameters and not restricted by symmetry or positive definiteness. Suppose there are total \( n \) emotion categories, the interaction measurement between a certain sample feature \( f(x_i) \) and different prototype features can be represented as

\[
l_i = \text{softmax} \left( [h(f(x_s^i), \mu_1), ..., h(f(x_s^i), \mu_n)] \right), \tag{6}
\]

where \( \{\mu_i\}_{i=1}^n \) are all the prototype features. An introduction of the activation function (softmax) here is to add nonlinear advantages to the interaction measurement, enhance the feature representation ability, and at the same time allow the feature vector \( l \) to have category prediction capabilities.

D. Pairwise learning on source domain

Traditional pointwise learning algorithms often regard the feature vector \( l \) as the predicted label of the sample feature \( x_i \) and use the cross-entropy loss function to match the label \( y_i \) of \( x_i \) based on supervised training. Then, if the sample features of one EEG trial match the prototype features of \( j \)th emotion category the most, the EEG trial will be assigned to the \( j \)th emotion category. However, this type of pointwise learning only focuses on the relationship between sample features and prototype features and ignores the relationship between different sample features. For example, the samples features belonging to different emotion categories should be separated from each other, and the sample features belonging to the same emotion category should be gathered together. To tackle this issue, we introduce pairwise learning to capture the inherent relationship of samples.

For the source domain data \( \{X_s, Y_s\} = \{x_{s,i}, y_{s,i}\}_{i=1}^{N_s} \), the corresponding loss function for pairwise learning is defined as

\[
\mathcal{L}_{\text{class}}(\theta) = \sum_{i,j} L \left( r_{ij}, g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \right), \tag{7}
\]

where \( g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \) is the similarity measurement of the samples \( x_{s,i}^* \) and \( x_{s,j}^* \) with the parameter of \( \theta \). According to the assumption of pairwise learning, if \( y_{s,i}^* = y_{s,j}^* \), then \( r_{ij} = 1 \); otherwise \( r_{ij} = 0 \). The loss function \( L(\cdot) \) is a difference calculation of \( r_{ij} \) and \( g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \), given by a two-category cross-entropy loss as

\[
L \left( r_{ij}, g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \right) = - r_{ij} \log \left( g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \right) - (1 - r_{ij}) \log \left( 1 - g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \right). \tag{8}
\]

In the training process, the label information of source domain data is used to define \( r \) in a supervised manner. In other words, based on the given information of \( Y_s \), if two samples belong to the same emotion category, then \( r = 1 \), otherwise \( r = 0 \). A supervised \( r \) can ensure the stability of the training process and the generalization ability of the model. The next key question is how to define a proper \( g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) \) to compute the similarity between \( x_{s,i}^* \) and \( x_{s,j}^* \) in terms of the characterized interaction features (termed as \( l_i \) and \( l_j \)). To make the similarity results locate in the range of \([0, 1]\) and extract better and more robust feature representations for subsequent emotion recognition, we add a norm restriction on \( l \) as

\[
l_{norm} = \frac{l}{\|l\|_2}. \tag{9}
\]

The similarity of \( l_{norm} \) and \( l_{norm} \) is calculated as the cosine similarity, given as

\[
g \left( x_{s,i}^*, x_{s,j}^*; \theta \right) = l_{norm} \cdot l_{norm} = \frac{l_{s,i}^* \cdot l_{s,j}^*}{\|l_{s,i}^*\|_2 \|l_{s,j}^*\|_2}, \tag{10}
\]

where \( \cdot \) refers to inner product operation. As stated in Chang et al. \cite{25}, the above-mentioned norm restriction can make the vector \( l \) have a clustering function, and the elements in the vector represent the probability that the feature belongs to a certain category cluster. Overall, the objective function of pairwise learning on the source domain is defined as

\[
\mathcal{L}_{\text{pairwise}}(\theta) = \sum_{i,j} L \left( r_{ij}, l_{norm} \cdot l_{norm} \right) + \beta \mathcal{R}, \tag{11}
\]

where \( \theta = \{\theta_f, S\} \), Here, \( \theta_f \) is the parameter of feature extractor \( f(\cdot) \), and \( S \) is the defined bilinear transformation matrix for interaction feature extraction. Besides, to avoid redundant feature extraction, a soft regularization \( \mathcal{R} \) is introduced with a weight parameter of \( \beta \), which is defined as

\[
\mathcal{R} = \left\| P^T P - I \right\|_F. \tag{12}
\]

Here, each row of the matrix \( P \) refers to the prototype feature belonging to one emotion category, \( \left\| \cdot \right\|_F \) is a \( F \) norm of the matrix, and \( I \) is an identity matrix. The above loss function (Eq. \ref{eq:pairwise_loss}) could be interpreted as a clustering loss, instead of emotion category classification loss. The main optimization goal is to gather the EEG samples that may belong to the same emotion category and separate the EEG samples that do not belong to the same category. The vector \( l \) is the characterized interaction feature for mapping to a non-linear informative feature space by measuring the interaction relationship between the sample features and all the available prototype features.
E. Pairwise learning on target domain

In this paper, besides of source domain, we also introduce the pairwise learning on the target domain to improve the feature separability in the target domain, as

\[
L_{\text{pairwise}}^{t}(\theta_f, S) = \sum_{i,j} L\left(r_{ij}^{t}, \frac{l_i^{t} \cdot l_j^{t}}{||l_i^{t}||_2 ||l_j^{t}||_2}\right), \quad (13)
\]

where \(l_i^{t}\) is the interaction features of the target domain data characterized in Eq. [6]. The scalar \(r_{ij}^{t}\) symbolizes the pairing relationship of the samples in the target domain. Since the label information of the target domain is completely missing in the training process, we cannot accurately obtain the pairing relationship as the source domain. To address this issue, we introduce an adaptive thresholding method to generate the valid pseudo labels and define the pairing relationship of the target domain. Suppose that \(r_{ij}^{t}\) is defined as

\[
r_{ij}^{t} := \begin{cases} 1, & \text{if } l_i^{t} \cdot l_j^{t} \geq \tau_u, \\ 0, & \text{if } l_i^{t} \cdot l_j^{t} < \tau_l, \quad i, j = 1, \ldots, n \end{cases} \quad (14)
\]

where \(\tau_u\) and \(\tau_l\) are the upper and lower bounds for valid pair selection. Here, if the calculated pairwise similarity is higher or equal to the defined upper bound \((\tau_u)\), then the corresponding pseudo label \(r_{ij}^{t}\) would be assigned to 1; while if the calculated pairwise similarity is lower than the defined lower bound \((\tau_l)\), then the corresponding pseudo label \(r_{ij}^{t}\) would be assigned to 0. For the other pairs that do not meet the threshold requirement, we would consider that the model is uncertain about whether the sample pair is paired or not and consider these pairs as invalid results. In order to prevent incorrect optimization, the invalid pairs would be temporarily excluded and will not participate in training and loss calculations at the current training round.

In the early training stage, the classification performance in the target domain is not good enough. In order to ensure the training stability, we set a strict upper threshold \((\tau_u)\) and a lower threshold \((\tau_l)\) and exclude most of the pair results in the target domain. Along with the enhancement of model performance in the target domain, we can gradually lower the upper threshold and raise the lower threshold and allow more samples to participate in the training part. In other words, along with the increase of training step, more pair samples in the target domain will be included for model learning. Here, we form a non-linear dynamic update for thresholding, as

\[
\tau_{h}^{t} = \tau_{h}^{t-1} - \frac{\tau_{h}^{t-1} - \tau_{l}^{t-1}}{\text{maxepoch}} \quad \tau_{l}^{t} = \tau_{l}^{t-1} + \frac{\tau_{h}^{t-1} - \tau_{l}^{t-1}}{\text{maxepoch}} \quad (15)
\]

where \(\tau_{h}^{t}\) represents the upper threshold of the current training round \(t\), \(l_i^{t}\) represents the current lower threshold, and \(\text{maxepoch}\) is the maximum training round. Based on the given initial values \(\tau_{h}^{0}\) and \(\tau_{l}^{0}\), the calculations of \(\tau_{h}\) and \(\tau_{l}\) could be considered as non-linear changes with respect to the training rounds as

\[
\begin{align*}
\tau_{h}^{t} = \tau_{h}^{0} - \frac{\tau_{h}^{0} - \tau_{l}^{0}}{2} \left(1 - \frac{t}{\text{maxepoch}}\right)^{\gamma} \\
\tau_{l}^{t} = \tau_{l}^{0} + \frac{\tau_{h}^{0} - \tau_{l}^{0}}{2} \left(1 - \frac{t}{\text{maxepoch}}\right)^{\gamma}
\end{align*} \quad (17)
\]

In all, combining the domain adversarial loss (Eq. [2]), the pairwise learning loss in the source domain (Eq. [11] and the pairwise learning loss in the target domain (Eq. [13]), the final objective function of PR-PL could be given as follows:

\[
\min_{\theta_f, S} \max_{\theta_d} L_{\text{pairwise}}^{s}(\theta_f, S) + \gamma L_{\text{pairwise}}^{t}(\theta_f, S) - \lambda L_{\text{disc}} (\theta_f, \theta_d), \quad (18)
\]

where \(\gamma\) is a hyperparameter to control the importance of the pairwise loss of the target domain, given as \(\gamma = \delta \frac{\text{epoch}}{\text{maxepoch}}\).

IV. EXPERIMENTAL RESULTS

A. Emotional Databases and Data Preprocessing

To have a fair comparison with the state-of-the-art methods, we validate our proposed model on two well-known public databases: SEED [10] and SEED-IV [13]. In SEED database [10], a total of 15 subjects were invited. Each subject performed three sessions on different days and each session contained 15 trials. Total three emotions were elicited (negative, neutral, and positive). In SEED-IV database [13], a total of 15 subjects participated in the experiment. For each subject, a total of 3 sessions were performed on different days and each session contained 24 trials. Total four emotions were elicited (happiness, sadness, fear, and neutral). The EEG signals of both SEED and SEED-IV databases were simultaneously collected using the 62-channel ESI Neuroscan system.

For EEG preprocessing, the data sampling rate was first downsampled to 200Hz, and the contaminated noises (e.g. EMG and EOG) were manually removed. Then, the data were filtered by a band-pass filter of 0.3 Hz to 50Hz. For each trial, the data was divided into a number of segments with a length of 1s. Based on the pre-defined five frequency bands: Delta (1-3 Hz), Theta (4-7 Hz), Alpha (8-13 Hz), Beta (14-30 Hz), and Gamma (31-50 Hz), the corresponding differential entropy (DE) features were extracted to represent the logarithm energy spectrum in a specific frequency band and total 310 features (5 frequency band \(\times\) 62 channels) were obtained for one EEG segment. Then, all the features were smoothed with the linear dynamic system (LDS) method, which can utilize the time dependency of emotion changes and filter out emotion unrelated and noisy EEG components [26].

B. Implementation Results

In our experiments, the feature extractor \(f\) and discriminator \(d\) are both made up of multilayer perceptron (MLP) with the Relu activation function. All the parameters are randomly initialized from a uniform distribution. The bilinear operator matrix \(S\) is also randomly initialized. In the model architecture, the feature extractor structure is designed as 310 (input
layer)-64 (hidden layer 1)-Relu activation-64 (output feature layer). The discriminator structure is designed as 64 (input layer)-64 (hidden layer 1)-Relu activation-dropout layer-64 (hidden layer 2)-1 (output layer)-Sigmoid activation. The size of matrix $S$ given in Eq. 5 is $64 \times 64$. Besides, we adopt an RMSprop optimizer for network training, which shows a better performance than the other classic optimizers. The learning rate is set to $1e^{-3}$ and the mini-batch size for training is 96. To avoid overfitting problems, we use $L2$ regularizes $(1e-5)$ in the networks. The regularization coefficient $\beta$ in Eq. 11 is 0.01. The balance parameter $\gamma$ for pairwise learning on the target domain in Eq. 18 is controlled by a constant factor $\delta$ of 2. The threshold $\tau^0_h$ and $\tau^i_0$ are given to 0.9 and 0.5 respectively. All the models are trained on an NVIDIA GeForce RTX 2080 GPU, with CUDA 10.0 using the Pytorch API.

C. Experiment Protocols

To fully evaluate the robustness and stability of the proposed model and compare it with the existing literature, we validate PR-PL using four different validation protocols. (1) **Cross-subject cross-session leave-one-subject-out cross-validation.** We evaluate the model with a strict cross-subject cross-session leave-one-subject-out to fully estimate the model robustness on the unknown subject(s) and session(s). One subject’s all sessions data are used as the target and the remaining subjects’ all sessions are used as the source. We repeat the training validation until each subject’s all sessions are treated as the target for once. Due to the variants in individuals and sessions, this evaluation protocol is the greatest challenge to the model effectiveness in the EEG-based emotion recognition tasks. (2) **Cross-subject single-session leave-one-subject-out cross-validation.** It is the most widely used validation scheme in the EEG-based emotion recognition tasks [5], [20], [21], [27]. One subject’s one-session data is treated as the target and the other remaining subjects are used as the source. We repeat the training validation process until each subject is treated as the target for once. Same as the other studies, we only consider the first session in this type of cross-validation. (3) **Within-subject cross-session leave-one-session-out cross-validation.** Similar to the existing methods, a time-series cross-validation method is adopted here, where the past data is used to predict current or future data. For one subject, the first two sessions are used as the source, and the latter session is used as the target. The average accuracies and standard deviations across subjects are calculated as the final results. (4) **Within-subject single-session cross-validation.** Following the validation protocol presented in the existing studies [10], [13], for each session of one subject, we use the first 9 (SEED) or 16 (SEED-IV) trials as the source and the rest 6 (SEED) or 8 (SEED-IV) trials as the target. The results are reported as the average performance across all the subjects. In the following performance comparison across four different validation protocols, the model results reproduced by us are indicated by ‘*’.

### Table II: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using cross-subject cross-session leave-one-subject-out cross-validation. Here, the model results reproduced by us are indicated by ‘*’.

| Methods          | $P_{acc}$ ± $P_{std}$ (%) |
|------------------|---------------------------|
| Traditional machine learning methods |                   |
| RF* [28]         | 69.60±0.0764              |
| SVM* [20]        | 68.15±0.0738              |
| TCA* [14]        | 64.02±0.0796              |
| SA* [37]         | 61.41±0.0975              |
| DCORAL* [35]     | 81.97±0.0516              |
| DCC* [36]        | 82.17±0.0496              |
| PR-PL            | **85.56±0.0478**          |

| Deep learning methods |                   |
|----------------------|-------------------|
| DGCNN [38] (58)      | 52.82±0.0923      |
| RGNN [8]             | 73.84±0.0802      |
| BiDANN [18] (65)     | 65.59±0.1039      |
| PR-PL                | 74.92±0.0792      |

D. **Cross-subject cross-session leave-one-subject-out cross-validation results**

To verify the model efficiency and stability on both cross-subject and cross-session conditions, we verify the proposed PR-PL using cross-subject cross-session leave-one-subject-out cross-validation on both SEED and SEED-IV databases. As reported in Table II and Table III the results show our proposed model achieves the highest results, where the emotion recognition performance of PR-PL is 85.56%±4.78% for three-class classification task on SEED and 74.92%±7.92% for four-class classification task on SEED-IV. Compared to the existing studies, the proposed PR-PL increases the classification accuracy to 3.39% and 1.08% for SEED and SEED-IV, with smaller standard deviations. These results demonstrate the proposed PR-PL has better affective effectiveness with higher recognition accuracy and better generalizability.

### Table III: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED-IV database using cross-subject cross-session leave-one-subject-out cross-validation. Here, the model results reproduced by us are indicated by ‘*’.

| Methods          | $P_{acc}$ ± $P_{std}$ (%) |
|------------------|---------------------------|
| Traditional machine learning methods |                   |
| RF* [28]         | 40.83±0.0728              |
| SVM* [20]        | 53.44±0.0912              |
| TCA* [14]        | 49.44±0.0909              |
| SA* [37]         | 45.89±0.0827              |
| KPCA [17]        | 49.35±0.0974              |
| DGCNN [38]       | 58.87±0.0813              |
| RGNN [8]         | 69.03±0.0866              |
| BiDANN [18]      | 54.63±0.0803              |
| PR-PL            | **74.92±0.0792**          |

E. **Cross-subject single-session leave-one-subject-out cross-validation results**

Table IV summarizes the model results in cross-subject single-session leave-one-subject-out recognition task and com-
TABLE IV: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using cross-subject single-session leave-one-subject-out cross-validation. Here, the model results reproduced by us are indicated by ‘*’.

| Methods     | $P_{acc}$ | Methods     | $P_{acc}$ |
|-------------|-----------|-------------|-----------|
| **Traditional machine learning methods** |
| TKL [18]   | 63.54±15.47 | T-SVM [18] | 72.53±14.00 |
| TCA [14]   | 63.64±14.88 | TPT [17]  | 75.17±12.83 |
| KPCA [17]  | 61.28±14.62 | GFK [37]  | 71.31±14.09 |
| SA [37]    | 69.00±10.89 | DICA [19] | 69.40±07.80 |
| DNN [17]   | 61.01±12.38 | SVM [17]  | 58.18±13.85 |
| **Deep learning methods** |
| DGCNN [38] | 79.95±09.02 | DAN [17]  | 83.81±08.56 |
| RGNN [8]   | 85.30±06.72 | BiHDM [37] | 85.40±07.53 |
| WGAN-GP [20] | 87.10±07.10 | MMD [21] | 80.88±10.10 |
| ATDD-DANN [19] | 90.92±01.05 | JDA-Net [21] | 88.28±11.44 |
| R2G-STNN [6] | 84.16±07.63 | SimNet* [29] | 81.58±05.11 |
| BiDANN [18] | 83.28±09.60 | DResNet [39] | 85.30±08.00 |
| ADA [21]    | 84.47±10.65 | DANN [21] | 81.65±09.92 |
| **PR-PL**    | 93.06±05.12 |

TABLE V: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using within-subject cross-session cross-validation. Here, the model results reproduced by us are indicated by ‘*’.

| Methods     | $P_{acc}$ | Methods     | $P_{acc}$ |
|-------------|-----------|-------------|-----------|
| **Traditional machine learning methods** |
| RF* [28]   | 69.57±06.55 | KNN* [29] | 75.68±13.82 |
| TCA* [14]  | 74.27±12.88 | CORAL* [32] | 84.18±09.81 |
| SA* [23]   | 69.84±09.46 | GFK* [34] | 78.79±09.39 |
| **Deep learning methods** |
| DAN* [17]  | 89.16±07.90 | SimNet* [24] | 86.88±07.83 |
| DDC* [36]  | 91.14±05.61 | ADA [21] | 89.13±07.13 |
| DANN [16]  | 89.45±06.74 | MMD [21] | 84.38±12.05 |
| JDA-Net [21] | 91.17±08.11 | DCORAL* [35] | 88.67±06.25 |
| **PR-PL**   | 93.18±06.55 |

TABLE VI: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED-IV database using within-subject cross-session cross-validation. Here, the model results reproduced by us are indicated by ‘*’.

| Methods     | $P_{acc}$ | Methods     | $P_{acc}$ |
|-------------|-----------|-------------|-----------|
| **Traditional machine learning methods** |
| RF* [28]   | 60.27±16.36 | KNN* [29] | 54.18±16.28 |
| TCA* [14] | 51.88±15.84 | CORAL* [32] | 66.06±15.13 |
| SA* [23] | 52.81±09.53 | GFK* [34] | 56.14±12.15 |
| **Deep learning methods** |
| DGCORAL [40] | 65.10±13.20 | DAN [40] | 60.20±10.20 |
| DDC [40] | 68.30±16.60 | MEERNet [40] | 72.10±14.10 |
| **PR-PL** | 74.62±14.15 |

F. Within-subject cross-session cross-validation results

By calculating the average and standard deviation of the experimental results of each subject, the final within-subject cross-session cross-validation results are reported in Table VI for SEED database and Table VII for SEED-IV database. For both databases, our proposed PR-PL achieves the highest recognition performance compared with the state-of-the-art methods (including both traditional machine learning methods and deep learning methods), where the results are 93.18±6.55% and 74.62±14.15% for SEED (three-class emotion recognition) and SEED-IV (four-class emotion recognition), respectively.

G. Within-subject single-session cross-validation results

Consistent with the evaluation method presented in the previous studies that only consider the first two sessions of the SEED database for experiments, we present the within-subject single-session results in Table VII. It shows our proposed model obtains the best recognition performance of 94.84%. Comparing the recognition results between cross-subject single-session (Table IV) and within-subject single-session (Table VII) emotion recognition tasks, the proposed PR-PL achieves the highest accuracies and at the same time perform the closest results on the two cross-validation methods (cross-subject single-session: 93.06±05.12; within-subject single-session: 94.84±09.16). For the other models, such as DGCNN [38], BiDANN [18], R2G-STNN [6], BiHDM [37], BiDANN [18], and WGAN-GP [20], the proposed PR-PL with pairwise learning can avoid the inherent defects of DANN design (e.g. only considers feature separability on source domain) and well address the individual differences and noisy labeling issues in aBCI applications.
TABLE VII: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED database using within-subject single-session cross-validation. Here, the model results reproduced by us are indicated by ‘*’.  

| Methods          | $P_{\text{acc}}$ | Methods          | $P_{\text{acc}}$ |
|------------------|------------------|------------------|------------------|
| **Traditional machine learning methods**       |                  | **Traditional machine learning methods**       |                  |
| SVM [37]         | 83.99±0.97       | GRSRL [41]       | 87.39±0.86       |
| RF [37]          | 78.46±1.17       | GSCCA [42]       | 82.96±0.99       |
| CCA [37]         | 77.63±1.32       | DBN [10]         | 86.08±0.83       |
| **Deep learning methods**                        |                  | **Deep learning methods**                        |                  |
| DANN [16]        | 63.07            | BiHDM [37]       | 93.12±10.84      |
| DANN [16]        | 91.36            | BiHDM [37]       | 74.35±13.12      |
| RGNN [8]         | 94.24            | DBN [10]         | 91.36±0.83       |
| GSCCA [42]       | 82.96±0.99       | DBN [10]         | 91.36±0.83       |
| SimNet* [24]     | 71.38            | PR-PL            | 93.06±0.09       |
| **PR-PL**        | 89.92            | PR-PL            | 93.06±0.09       |

TABLE VIII: The mean accuracies (%) and standard deviations (%) of emotion recognition on SEED-IV database using within-subject single-session cross-validation. Here, the model results reproduced by us are indicated by ‘*’.  

| Methods          | $P_{\text{acc}}$ | Methods          | $P_{\text{acc}}$ |
|------------------|------------------|------------------|------------------|
| **Traditional machine learning methods**       |                  | **Traditional machine learning methods**       |                  |
| SVM [37]         | 56.61±20.05      | GRSRL [41]       | 69.32±19.57      |
| RF [37]          | 50.97±16.22      | BiHDM [37]       | 93.12±10.84      |
| CCA [37]         | 54.47±18.48      | BiHDM [37]       | 74.35±14.09      |
| **Deep learning methods**                        |                  | **Deep learning methods**                        |                  |
| DGCNN [43]       | 65.16±16.29      | BiHDM [37]       | 74.35±14.09      |
| ATDD-DANN [19]   | 91.08±0.63       | SimNet [24]      | 71.38±13.12      |
| R2G-STNN [6]     | 93.38±0.56       | STRNN [44]       | 89.50±0.73       |
| BiDANN [18]      | 92.38±0.74       | DANN [16]        | 63.07±12.66      |
| GCNN [37]        | 87.40±0.90       | PR-PL            | 93.06±0.09       |
| **PR-PL**        | 94.84±0.09       | PR-PL            | 93.06±0.09       |

A. Ablation Study

We conduct the ablation study to systematically explore the effectiveness of different components in the proposed model and examine the corresponding contributions to the overall performance. As shown in Table IX it is found that the introduction of domain adversarial training can greatly enhance the emotion recognition performance on the target domain. When the model is without discriminator and target domain information, the recognition accuracy reduces from 93.06% to 83.30%. Such a significant drop shows the significant impact of individual differences problem on model performance and highlights the great potential of transfer learning in bBCI applications. Besides, the results show a combination of pairwise learning on the source and target domain benefits to the model performance, where the recognition accuracy is increased by 6.35% (from 87.5% to 93.06%). For the pseudo-labeling method, the corresponding accuracy increases from 89.92% to 92.46% when the pseudo-labeling method changes from fixed to linear dynamic. The accuracy further increases to 93.06%, when a nonlinear dynamic based adaptive pseudo-labeling method is adopted. The results show a non-linear dynamic pseudo-labeling could be helpful to screen out the valid paired samples and improve the model trainability. For the final loss function given in Eq. 18 instead of using a fixed weight for the pairwise loss on the target domain, we propose to update the weight gradually along with the training epochs to prevent model learning failures in the early training stage and balance the relationships among different losses. The benefit of a dynamic $\gamma$ is also reflected in the ablation study, where the recognition accuracy increases from 89.47% and 93.06%.

B. Effect of Noisy Labels

To further verify the model robustness during noisy label learning, we randomly contaminate the source labels with $\eta$% noises and test the corresponding model performance on unknown target data. Specifically, we replace $\eta$% real labels in $Y_s$ using randomly generated labels and train the model in supervised learning. Then, we test the trained model performance on the target domain. Note here that the noisy contamination is only conducted on the source domain, as the target domain needs to be used for model evaluation. In the implementation, $\eta$% value is adjusted to 10%, 20%, and 30%, respectively. The corresponding model accuracies with the standard deviations are 89.22±6.05%, 88.39±6.73%, and 87.71±5.02%. It shows, with an increase of label noise ratio from 10% to 30%, the model performance decreases slightly, with a decrease rate of 1.69%. These results demonstrate the proposed PR-PL is a reliable model which has a higher tolerance to noisy labels.

C. Confusion Matrices

To qualitatively study the model performance in each emotion class, we visually analyze the confusion matrices and compare results with the latest models [8], [37]. As shown in Fig. it shows all the models are good at distinguishing positive emotions from other emotions (the recognition rates are all above 90%), but it is relatively poor at distinguishing negative and neutral emotions. For example, the recognition rate of neural in BiDANN [18] is even lower than 80% (76.72%). Compared to the existing methods (a), (b), and (c), our proposed model can enhance the model’s recognition ability, especially for distinguishing neutral and negative emotions. As shown in (d), the recognition rates for negative, neutral, and positive emotions are 92.10%, 90.39%, and 96.50%.
D. Visualization of Learned Representation

To verify the effectiveness of the proposed model from a more intuitive perspective, we visualize the characterized sample and interaction features of source and target domains using T-SNE [45] in Fig. 3. Here, we randomly select 500 samples from the source and 500 samples from the target for visualization of the learned feature representation. Compared to representation learned by the other model settings (w/o pairwise learning on the source and target and w/o pairwise learning on the target), the representation learned by the proposed PR-PL forms more separated emotional clusters. Comparing the extracted sample features (c) and interaction features (f) by the proposed PR-PL, the separability of the extracted interaction features from different emotion classes is further enlarged and at the same time, the concentration of the feature distribution for each emotion is also improved.

E. Conclusion

The paper proposes a novel transfer learning framework with prototypical representation based pairwise learning (PR-PL), that characterizes EEG data with prototypical representations and formulates the EEG-based emotion recognition task as pairwise learning. We evaluate our proposed model on two well-known emotional databases (SEED and SEED-IV) under four cross-validation protocols (cross-subject single-session, within-subject single-session, within-subject cross-session, and cross-subject cross-session) and compare with the existing state-of-the-art methods. Our extensive experimental results show PR-PL achieves the best results on all four cross-validation protocols and demonstrate the advantage of PR-PL in tackling individual differences and noisy labeling issues in abCI systems.

VI. Conflicts of Interest

The authors declare that they have no conflicts of interest.

VII. Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61906122, in part by Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions (2021SHIBS0003), in part by the Tencent “Rhinoceros Birds”-Scientific Research Foundation for Young Teachers of Shenzhen University, and in part by the High Level University Construction under Grant 000002110133.

References

[1] S. Siddharth, T.-P. Jung, and T. J. Sejnowski, “Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing,” IEEE Transactions on Affective Computing, 2019.
[2] X. Hu, J. Chen, F. Wang, and D. Zhang, “Ten challenges for eeg-based affective computing,” Brain Science Advances, vol. 5, no. 1, pp. 1–20, 2019.
[3] W. Hu, G. Huang, L. Li, L. Zhang, Z. Zhang, and Z. Liang, “Video-triggered eeg-emotion public databases and current methods: A survey,” Brain Science Advances, vol. 6, no. 3, pp. 255–287, 2020.
[4] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, “Transfer learning in brain-computer interfaces,” IEEE Computational Intelligence Magazine, vol. 11, no. 1, pp. 20–31, 2016.
[5] 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.
[6] 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.
[7] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, “Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network,” Knowledge-Based Systems, vol. 205, p. 106243, 2020.
[8] P. Zhong, D. Wang, and C. Miao, “Eeg-based emotion recognition using regularized graph neural networks,” IEEE Transactions on Affective Computing, 2020.
[9] X. Gu, Z. Cao, A. Jolfaei, P. Xu, D. Wu, T.-P. Jung, and C.-T. Lin, “Eeg-
based brain-computer interfaces (bcs): A survey of recent studies on
signal sensing technologies and computational intelligence approaches
and their applications,” IEEE/ACM transactions on computational biol-
ogy and bioinformatics, 2021.

[10] 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.

[11] Y. Jia, M. Salzmann, and T. Darrell, “Factorized latent spaces with
structured sparsity,” in Advances in Neural Information Processing
Systems 23: 24th Annual Conference on Computational Intelligence,
ser. AAI16. AAAI Press, 2010, pp. 982–990.

[12] L. Yao, S. Li, Y. Li, M. Huai, J. Gao, and A. Zhang, “Representation
learning for treatment effect estimation from observational data,” Ad-
vances in Neural Information Processing Systems, vol. 31, 2018.

[13] 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.

[14] 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.

[15] W.-L. Zheng and B.-L. Lu, “Personalizing eeg-based affective models
with transfer learning,” in Proceedings of the Twenty-Fifth International
Joint Conference on Artificial Intelligence, ser. IJCAI’16. AAAI Press,
pp. 2732–2738, 2016.

[16] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Lavi-
olette, M. Marchand, and V. Lempitsky, “Domain-adversarial training
of neural networks,” The journal of machine learning research, vol. 17,
no. 1, pp. 2096–2030, 2016.

[17] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, “Cross-subject emotion
recognition using deep adaptation networks,” in Neural Information
Processing, L. Cheng, A. C. S. Leung, and S. Ozawa, Eds. Cham:
Springer International Publishing, 2018, pp. 403–413.

[18] 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.

[19] X. Du, C. Ma, G. Zhang, J. Li, Y.-K. Lai, G. Zhao, X. Deng, Y.-J. Liu,
and H. Wang, “An efficient lstm network for emotion recognition from
multichannel eeg signals,” IEEE Transactions on Affective Computing,
pp. 1–1, 2020.

[20] Y. Luo, S. Y. Zhang, W. L. Zheng, and B. L. Lu, “Wgan domain adap-
tation for eeg-based emotion recognition,” International Conference
on Neural Information Processing, 2018.

[21] 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, 2020.

[22] M. Salzmann, C. H. Ek, R. Urtasun, and T. Darrell, “Factorized
orthogonal latent spaces,” Journal of Machine Learning Research, vol. 9,
pp. 701–708, 2010.

[23] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan,
“Domain separation networks,” ser. NIPS’16, 2016, p. 343–351.

[24] P. O. Pinheiro, “Unsupervised domain adaptation with similarity learn-
ing,” in 2018 IEEE/CVF Conference on Computer Vision and Pattern
Recognition, 2018, pp. 8004–8013.

[25] J. Chang, L. Wang, G. Meng, S. Xiang, and C. Pan, “Deep adaptive
image clustering,” in 2017 IEEE International Conference on Computer
Vision (ICCV), 2017, pp. 5880–5888.

[26] L.-C. Shi and B.-L. Lu, “Off-line and on-line vigilance estimation based
on linear dynamical system and manifold learning,” in 2010 Annual
International Conference of the IEEE Engineering in Medicine and
Biology, 2010, pp. 6587–6590.

[27] W.-L. Zheng, Y.-Q. Zhang, J.-Y. Zhu, and B.-L. Lu, “Transfer com-
ponents between subjects for eeg-based emotion recognition,” in 2015
International Conference on Affective Computing and Intelligent Inter-
action (ACII), 2015, pp. 917–922.

[28] Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32,
2001.

[29] “Alternative k-nearest neighbour rules in supervised pattern recogni-
tion: Part 1. k-nearest neighbour classification by using alternative
voting rules,” Analytica Chimica Acta, vol. 136, pp. 15–27, 1982.
[Online]. Available: https://www.sciencedirect.com/science/article/pii/
S0003267001953590

[30] J. Suykens and J. Vandewalle, “Least squares support vector machine
classifiers,” Neural Processing Letters, vol. 9, no. 3, pp. 293–300, 1999.

[31] J. Zhu, A. Arbor, and T. Hastie, “Multi-class adaboost,” Statistics & Its
Interface, vol. 2, no. 3, pp. 349–360, 2006.

[32] B. Sun, J. Feng, and K. Saenko, “Return of frustratingly easy domain
adaptation,” in Proceedings of the Thirtieth AAAI Conference on Artifi-
cial Intelligence, ser. AAAI’16. AAAI Press, 2016, p. 2058–2065.

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

[34] B. Gong, Y. Shi, F. Sha, and K. Grauman, “Geodesic flow kernel for
unsupervised domain adaptation,” in 2012 IEEE Conference on Com-
puter Vision and Pattern Recognition, 2012, pp. 2066–2073.

[35] B. Sun and K. Saenko, “Deep coral: Correlation alignment for deep
domain adaptation,” in Computer Vision – ECCV 2016 Workshops,
G. Hua and H. Jégou, Eds. Cham: Springer International Publishing,
2016, pp. 443–450.

[36] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep
domain confusion: Maximizing for domain invariance,” CoRR, vol.
abs/1412.3474, 2014. [Online]. Available: http://arxiv.org/abs/1412.3474

[37] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal
recurrent neural network for emotion recognition,” IEEE Transactions
on Cybernetics, 2019, pp. 839–847, 2019.

[38] T. Song, W. Zheng, P. Song, and Z. Cui, “Eeg emotion recognition
using dynamical graph convolutional neural networks,” IEEE Transac-
tions on Affective Computing, vol. 11, no. 3, pp. 532–541, 2018.

[39] B.-Q. Ma, H. Li, W.-L. Zheng, and B.-L. Lu, “Reducing the subject
variability of eeg signals with adversarial domain generalization,” in
Neural Information Processing, T. Gedeon, K. W. Wong, and M. Lee,
Eds. Cham: Springer International Publishing, 2019, pp. 30–42.

[40] H. Chen, Z. Li, M. Jin, and J. Li, “Meer-net: Multi-source eeg-based
emotion recognition network for generalization across subjects and
sessions,” in 2021 43rd Annual International Conference of the IEEE
Engineering in Medicine & Biology Society (EMBC). IEEE, 2021, pp.
6094–6097.

[41] Y. Li, W. Zheng, Z. Cui, Y. Zong, and S. Ge, “Eeg emotion recogni-
tion based on graph regularized sparse linear regression,” Neural Process-
ing Letters, vol. 49, p. 1–17, 04 2019.

[42] 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.

[43] T. Song, W. Zheng, P. Song, and Z. Cui, “Eeg emotion recognition
using dynamical graph convolutional neural networks,” IEEE Transac-
tions on Affective Computing, vol. 11, no. 3, pp. 532–541, 2020.

[44] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial-temporal
recurrent neural network for emotion recognition,” IEEE Transactions
on Cybernetics, 2019, pp. 839–847, 2019.

[45] V. D. M. Laurens and G. Hinton, “Visualizing data using t-sne,” Journal
of Machine Learning Research, vol. 9, no. 2605, pp. 2579–2605, 2008.