From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition

Yang Li

Wenming Zheng

Lei Wang
University of Wollongong, leiw@uow.edu.au

Yuan Zong

Zhen Cui

Recommended Citation

From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition

Abstract

IEEE In this paper, we propose a novel Electroencephalograph (EEG) emotion recognition method inspired by neuroscience with respect to the brain response to different emotions. The proposed method, denoted by R2G-STNN, consists of spatial and temporal neural network models with regional to global hierarchical feature learning process to learn discriminative spatial-temporal EEG features. To learn the spatial features, a bidirectional long short term memory (BiLSTM) network is adopted to capture the intrinsic spatial relationships of EEG electrodes within brain region and between brain regions, respectively. Considering that different brain regions play different roles in the EEG emotion recognition, a region-attention layer into the R2G-STNN model is also introduced to learn a set of weights to strengthen or weaken the contributions of brain regions. Based on the spatial feature sequences, BiLSTM is adopted to learn both regional and global spatial-temporal features and the features are fitted into a classifier layer for learning emotion-discriminative features, in which a domain discriminator working corporately with the classifier is used to decrease the domain shift between training and testing data. Finally, to evaluate the proposed method, we conduct both subject-dependent and subject-independent EEG emotion recognition experiments on SEED database, and the experimental results show that the proposed method achieves state-of-the-art performance.

Disciplines

Engineering | Science and Technology Studies

Publication Details

From Regional to Global Brain: A Novel Hierarchical Spatial-Temporal Neural Network Model for EEG Emotion Recognition

Yang Li, *Student Member, IEEE,* Wenming Zheng*, *Senior Member, IEEE,* Lei Wang, *Senior Member, IEEE,* Yuan Zong, *Student Member, IEEE,* and Zhen Cui, *Member, IEEE*

Abstract—In this paper, we propose a novel Electroencephalograph (EEG) emotion recognition method inspired by neuroscience with respect to the brain response to different emotions. The proposed method, denoted by R2G-STNN, consists of spatial and temporal neural network models with regional to global hierarchical feature learning process to learn discriminative spatial-temporal EEG features. To learn the spatial features, a bidirectional long short term memory (BiLSTM) network is adopted to capture the intrinsic spatial relationships of EEG electrodes within brain region and between brain regions, respectively. Considering that different brain regions play different roles in the EEG emotion recognition, a region-attention layer into the R2G-STNN model is also introduced to learn a set of weights to strengthen or weaken the contributions of brain regions. Based on the spatial feature sequences, BiLSTM is adopted to learn both regional and global spatial-temporal features and the features are fitted into a classifier layer for learning emotion-discriminative features, in which a domain discriminator working corporately with the classifier is used to decrease the domain shift between training and testing data. Finally, to evaluate the proposed method, we conduct both subject-dependent and subject-independent EEG emotion recognition experiments on SEED database, and the experimental results show that the proposed method achieves state-of-the-art performance.

Index Terms—EEG emotion recognition, regional to global, spatial-temporal network

I. INTRODUCTION

Emotion plays an essential role in human life [1]. Positive emotions could be helpful to improve the efficiency of our daily work, while negative emotions may influence our decision-making, attention, or even health [2]. Although it is easier for us to identify the other people's emotion from their facial expressions or speeches, it is still difficult for machine to do such work. Nevertheless, the study of emotion recognition using computer had attracted more and more researchers during the past several years, and emotion recognition had become a hot research topic in the research community of affective computing and pattern recognition [3]. Basically, emotion recognition methods could be categorized into the verbal behavior based methods (e.g., emotion recognition based on speech signals) and the nonverbal behavior based methods (e.g., emotion recognition based on facial expression images or physiological signals [4]). As a typical physiological signal, Electroencephalograph (EEG) had been widely applied to dealing with emotion recognition in recent years.

In dealing with EEG based emotion recognition problem, we usually encounter two major technical challenges. One is how to extract discriminative emotional feature from EEG signals, and the other one is how to develop more effective computational model for emotion recognition. Typically, EEG features can be extracted from time domain, frequency domain, and time-frequency domain [5]. For example, Lin et al. [6] investigated the relationships between emotional states and brain activities, and extracted power spectrum density, differential asymmetry power, and rational asymmetry power as the features of EEG signals. For the computational model problem, researchers have proposed many methods and models to recognize emotions through EEG signals [7][8][9][10][11]. Garca-Martnez et al. [12] summarized the most recent works that have applied nonlinear methods in EEG signal analysis to emotion recognition. Among the various EEG emotion recognition methods, it is notable that the recent development of deep learning based methods are becoming dominant for improving the performance of EEG emotion recognition. For example, Zheng et al. [7] introduced deep belief networks (DBNs) to construct EEG-based emotion recognition models. Pallavi Pandey et al. [13] proposed a multilayer perceptron neural network for subject-independent emotion recognition. Song et al. [14] constructed a graph relation based on multi-channel EEG data and then performed graph convolution on it to extract feature for classification. Li et al. [15] considered the domain shift for EEG data, and utilized the difference between two brain hemispheres to decrease this shift and achieved the state-of-the-art performance.

Although various EEG emotion recognition methods had been proposed in the past several years, there are still some major issues that should be well investigated in order to further improve the EEG emotion recognition performance. The first one is to extract robust high-level semantic features from the EEG signal. The current EEG recognition methods usually employ some handcraft features, such as statistic

Yang Li is with the Key Laboratory of Child Development and Learning Science (Ministry of Education), and the school of Information Science and Engineering, Southeast University, Nanjing, Jiangsu, 210096, China.

Wenming Zheng and Yuan Zong are 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. (*Corresponding author: Wenming Zheng (E-mail: wenming_zheng@seu.edu.cn).)

Lei Wang is with the School of Computing and Information Technology, University of Wollongong, NSW, 2500, Australia.

Zhen Cui is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210096, China.

features in time domain, band power in frequency domain, and discrete wavelet transform in time-frequency domain [5]. It is desired to investigate more powerful discriminative deep features with both spatial and temporal information of EEG signals [16]. The second issue is about what brain regions are more contributive to the emotion recognition and how to make use of these topographical information of these brain regions to improve the EEG emotion recognition. The recent neuroscience researches showed that human's emotion is closely related to a variety of brain cortex regions [17], such as the orbital frontal cortex, ventral medial prefrontal cortex, amygdala [18][19][20]. Consequently, the EEG signals associated with different brain regions would contribute differently to emotion recognition [21][22][23][24] and hence making use of the spatial information of EEG signals would be helpful for emotion recognition and could provide a physiological explanation to understand human emotion. The third issue is how to utilize the temporal information of EEG signals across different brain regions to improve emotion recognition. This is because EEG signals are dynamical time series and the temporal information usually carries important emotion messages that are very helpful to identify different emotions.

Inspired by the neuroscience finding that the brain response to different emotions would be varied in different brain regions. In this paper we propose a novel neural network model, denoted by R2G-STNN, to address the aforementioned three major issues in EEG emotion recognition. The basic idea is to integrate the EEG spatial-temporal information of both local and global brain regions into the EEG features to boost the emotion recognition performance. Specifically, R2G-STNN consists of both spatial and temporal neural network layers with regional to global (R2G) hierarchies of feature learning process to capture the emotional responses and structural relation of different brain regions for learning the discriminative spatial-temporal EEG features. The R2G-STNN framework includes the following two major modules:

(1) **Feature extractor**. The feature extractor module aims to learn discriminative spatial and temporal EEG features using bidirectional long short term memory (BiLSTM) network within each brain region and between the different brain regions, respectively. Considering that different brain regions play different roles in the EEG emotion recognition, a region-attention layer is also introduced to learn a set of weights indicating the contributions of brain regions.

(2) **Classifier and discriminator**. This module aims to predict the emotion class information based on the spatial-temporal features obtained from the feature extractor module. It can also guide the overall neural network learning towards generating more discriminative EEG features for emotion classification. Moreover, we also introduce a discriminator to alleviate the domain shift between source and target domain data, which will enable the hierarchical feature leaning process to generate emotion discriminative but domain adaptive EEG features.

By combining the aforementioned three modules together, we can learn more discriminative and domain-robust EEG

features for improving emotion recognition performance. In summary, the major contributions of this paper include the following two major parts:

- Propose a novel neuroscience inspired hierarchical spatial-temporal EEG feature learning model, which is able to capture both spatial and temporal emotion information from EEG signals within each brain region and across different brain regions;
- Propose a weighting method to evaluate the different contributions of the brain regions to the EEG emotion recognition, which would be advantageous to further improve the EEG emotion recognition by enhancing the influence of the most contributive brain regions while alleviating the influence of the less contributive regions through the region weights.

The remainder of this paper is organized as follows: In section II, we overview the preliminary work of bidirectional long short term memory network. In section III, we specify the method of R2G-STNN as well as its application to EEG emotion recognition. In section IV, we conduct extensive experiments to evaluate the proposed method for EEG emotion recognition. Finally, in sections V, we discuss and conclude the paper.

II. PRELIMINARY

In this section, we will briefly overview the preliminary work of bidirectional long short term memory (BiLSTM) network [25] and then address how we can apply it to the EEG feature extraction task.

The conventional single directional long short term memory (LSTM) network [26] is a special type of recurrent neural network (RNN) [27], which usually consists of three gates and a cell state for dealing with the long-term dependence of data sample sequence. These gates allow LSTM to keep the important data information and forget unnecessary information of the data samples [28]. However, one shortcoming of conventional LSTM is that it only make use of the previous context. The BiLSTM module is able to process data using two directions with separate hidden layers, respectively [25]. As a result, compared with the traditional LSTM model, BiLSTM can access the long-range context in both input directions, and hence it could be better used to model time sequences.

When dealing with the EEG data processing problem, it is interesting to see that the EEG signals associated with the electrodes of each brain region can be treated as virtual sequence since the dimensions of all the electrodes are same. For this reason, the EEG data can be fed into BiLSTM module to extract high-level deep features, which will contain the spatial relation information. Moreover, benefiting from the less probable disturbance of electrode arrangement in the input sequence, we use BiLSTM rather than conventional single directional LSTM to model the electrodes' data in each brain region, then forward construct their spatial relations.

III. R2G-STNN FOR EEG-BASED EMOTION RECOGNITION

In this section, we will specify the R2G-STNN model as well as the method of applying this model to dealing with
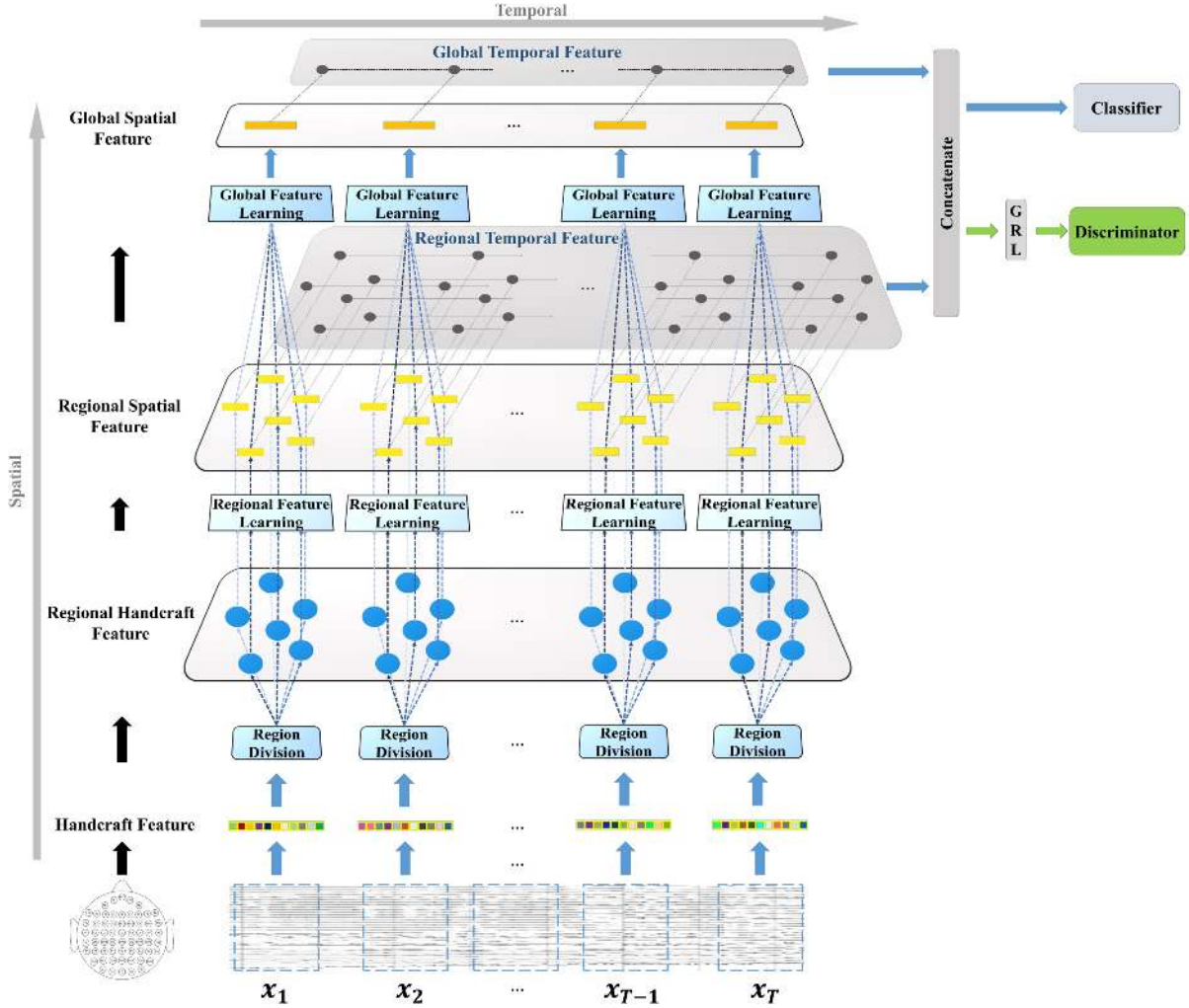
Fig. 1: The framework of R2G-STNN. The region to global (R2G) process includes two directions, i.e., spatial and temporal streams. The spatial stream constructs the relation in and among all the brain regions hierarchically, while the temporal stream captures the EEG signal's dynamic information as well as learning from the brain regions' time sequences.

EEG emotion recognition. Fig. 1 illustrates the framework of the R2G-STNN method, from which we can see that the R2G-STNN method consists of three major parts, i.e., feature extractor part, the classifier part, and the discriminator part. In what follows, we will address these parts in details.

A. Spatial feature learning

Suppose that we are given a trial of raw EEG signals $\mathbf{S}$. Then, we divide $\mathbf{S}$ into several segments. For each segment of EEG signals, we extract a set of handcraft features (e.g., the differential entropy (DE) features [7] on five EEG frequency bands, i.e., δ band (1-3Hz), θ band (4-7Hz), α band (8-13Hz), β band (14-30Hz), γ band (31-50Hz)) from each EEG electrode. Moreover, to explore the dynamic temporal information of EEG signals, every $T = 9$ neighboring EEG segments are chosen to constitute an EEG sample. In this case, each EEG sample will correspond to a handcraft feature tensor.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T] \in \mathbb{R}^{d \times n \times T}$ denote an EEG sample, where $\mathbf{x}_i$ denotes a handcraft feature matrix extracted from the i-th segment of EEG signals (denoted by blue dashed rectangle box shown at the bottom of Fig. 1), d, n and T denote the number of features associated with each EEG electrode, the number of electrodes, and the number of EEG segments in one EEG sample, respectively. Then, from Fig. 1 we can see that, for each $\mathbf{x}_i$ ($i \in \{1, 2, \cdots, T\}$), the spatial feature learning for $\mathbf{x}_i$ consists of two feature extracting layers, which achieves a progressive feature learning from regional brain to global brain.

Fig. 2 illustrates the detailed learning procedure for the input handcraft feature matrix $\mathbf{x}_i$. Firstly, we group the columns of $\mathbf{x}_i$ into several clusters, where each cluster corresponds to a brain region determined by the spatial locations of EEG electrodes, resulting in a set of regional handcraft feature vectors in each brain region. Then, the regional handcraft feature vectors of each brain region are fed into the same amount of BiLSTM networks to learn regional deep features. After extracting the regional spatial feature, a region-attention layer is used to learn a set of weights indicating the importance of the divided regions and the weights are used to penalize the regional deep features. Finally, at the top of the spatial feature
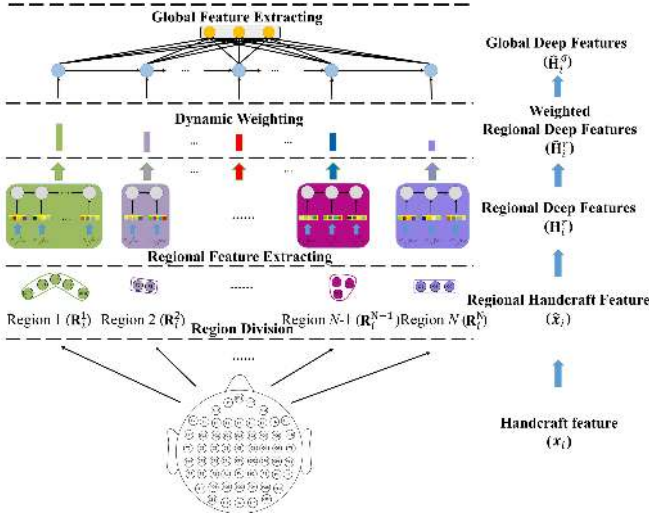
Fig. 2: The regional to global spatial feature learning process, which consists of three layers, i.e., regional feature learning layer, dynamic weight layer, and global feature learning layer.

learning layer, the weighted regional feature vectors are further fed into BiLSTM to learn global deep features. These detailed operations of spatial feature learning are described as follows:

1) Regional feature learning: Let $\mathbf{v}_{ij}$ denote the handcraft feature vector associated with the j-th EEG electrode. Let $\mathbf{x}_i = [\mathbf{v}_{i1}, \mathbf{v}_{i2}, \cdots, \mathbf{v}_{in}] \in \mathbb{R}^{d \times n}$. Now we group the n columns of $\mathbf{x}_i$ into several clusters according to the associated electrodes, where each cluster corresponds to a brain region, e.g.,

$$\text{brain region 1:} \quad \mathbf{R}_i^1 = [\mathbf{v}_{i1}^1, \mathbf{v}_{i2}^1, \cdots, \mathbf{v}_{in_1}^1],$$
$$\cdots$$
$$\text{brain region N:} \quad \mathbf{R}_i^N = [\mathbf{v}_{i1}^N, \mathbf{v}_{i2}^N, \cdots, \mathbf{v}_{in_N}^N],$$

where N is the number of brain regions, n_j denotes the number of electrodes in the j-th brain region, $n_1 + \cdots + n_N = n$.

In this case, we change the order of the columns of $\mathbf{x}_i$ and express it as a block matrix denoted by

$$\hat{\mathbf{x}}_i = [\mathbf{R}_i^1, \cdots, \mathbf{R}_i^N],$$

which is called the regional handcraft feature matrix. For each block matrix $\mathbf{R}_i^j$ ($j = 1, \cdots, N$), it is notable that each column corresponds to an EEG electrode. Consequently, we can model the spatial relationships of the electrodes by applying BiLSTM to the columns of each block matrix to obtain high-level deep features, which can be formulated as

$$\mathcal{L}(\mathbf{R}_i^1) = [\mathbf{h}_{i1}^1, \mathbf{h}_{i2}^1, \cdots, \mathbf{h}_{in_1}^1] \in \mathbb{R}^{2d_r \times n_1}, \tag{1}$$
$$\cdots$$
$$\mathcal{L}(\mathbf{R}_i^N) = [\mathbf{h}_{i1}^N, \mathbf{h}_{i2}^N, \cdots, \mathbf{h}_{in_N}^N] \in \mathbb{R}^{2d_r \times n_N}, \tag{2}$$

where $\mathcal{L}(\cdot)$ denotes the BiLSTM operation, $\mathbf{h}_{ik}^j \in \mathbb{R}^{2d_r}$ denotes the k-th forward and backward hidden units of BiLSTM, d_r is the hidden units dimension of regional spatial BiLSTM. After the above regional feature learning process, we concatenate the last hidden units of BiLSTM as the feature

representation. In this case, the local features of all the regions can be expressed as

$$\mathbf{H}_i^r = [\mathbf{h}_{in_1}^1, \cdots, \mathbf{h}_{in_N}^N] \in \mathbb{R}^{2d_r \times N}. \tag{3}$$

2) Attention-based brain region weights learning: The research of neuroscience indicated that different emotions are closely related to different brain regions such as orbital frontal cortex or ventral medial prefrontal cortex. Hence, the EEG signals associated with different brain regions would contribute differently to emotion recognition. For this reason, we introduce a weighting layer to emphasize the contributions of the electrodes of the brain regions in EEG emotion recognition.

Specifically, for each brain region, we introduce a dynamic weight, denoted by matrix $\mathbf{W} = \{w_{ij}\}$, to weight electrodes of each brain region, i.e.,

$$\hat{\mathbf{H}}_i^r = \mathbf{H}_i^r \mathbf{W}, \tag{4}$$
$$\mathbf{W} = (\mathbf{Q}\tanh(\mathbf{P}\mathbf{H}_i^r + \mathbf{b}^r \mathbf{e}^T))^{\mathrm{T}}, \tag{5}$$
$$w_{ij} = \frac{\exp(w_{ij})}{\sum_{k=1}^N \exp(w_{kj})}, \tag{6}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are learnable transformation matrices, $\mathbf{b}^r$ is the bias, and $\mathbf{e}$ denotes an N-dimensional vector with all elements 1, i.e., $\mathbf{e} = [1, 1, \cdots, 1]^T$. The columns of $\mathbf{W}$ are normalized and the elements are restricted to be nonnegative using Eq. (6). In this case, we obtain that the larger w_{ij} is, the more important i-th region is.

3) Global feature learning: This layer is used to capture the potential structural information based on the above obtained weighted regional spatial features $\hat{\mathbf{H}}_i^r$. Here, we employ another BiLSTM network to capture the global structure of all EEG brain regions, which can be formulated as

$$\mathcal{L}(\hat{\mathbf{H}}_i^r) = [\mathbf{h}_{i1}^g, \mathbf{h}_{i2}^g, \cdots, \mathbf{h}_{iN}^g] \in \mathbb{R}^{d_g \times N}, \tag{7}$$

where $\mathbf{h}_{ik}^g$ denote the k-th hidden unit of BiLSTM, d_g is dimension of global spatial BiLSTM.

For the feature vector sequence $\mathbf{h}_{i1}^g, \cdots, \mathbf{h}_{iN}^g$, we use them as input data and learn another feature vector sequence according to the following rule:

$$\hat{\mathbf{h}}_{ik}^g = \sigma(\sum_{i=1}^N G_{jk}^g \mathbf{h}_{ij}^g + \mathbf{b}^g), \quad k = 1, 2, \cdots, K, \tag{8}$$
$$\hat{\mathbf{H}}_i^g = [\hat{\mathbf{h}}_{i1}^g, \hat{\mathbf{h}}_{i2}^g, \cdots, \hat{\mathbf{h}}_{iK}^g], \tag{9}$$

where $\mathbf{G}^g = [G_{jk}^g]_{N \times K}$ is a project matrix, $\mathbf{b}^g$ is a bias and K is the length of the compressed sequence, $\sigma(\cdot)$ denotes a nonlinear mapping function. In this case, we finally get the global-level features associated with the i-th EEG segmentation handcraft feature matrix $\mathbf{x}_i$.

B. *Temporal feature learning*

Let $\mathbf{h}_i^j$ ($j = 1, \cdots, N$) denote the last hidden unit of the j-th brain region associated with the i-th EEG handcraft feature matrix $\mathbf{x}_i$. Let

$$\mathbf{H}^{r1} \triangleq [\mathbf{h}_1^1, \mathbf{h}_2^1, \cdots, \mathbf{h}_T^1], \tag{10}$$
$$\cdots$$
$$\mathbf{H}^{rN} \triangleq [\mathbf{h}_1^N, \mathbf{h}_2^N, \cdots, \mathbf{h}_T^N]. \tag{11}$$

Then, the columns of the feature matrix $\mathbf{H}^{rj}$ $(j = 1, \cdots, N)$ constitute a feature vector sequence associated with the j-th brain region. Hence, we can apply BiLSTM to learning temporal information among the feature vector sequence, which results in the following regional temporal feature vectors:

$$
\begin{aligned}
\mathbf{Y}^{rt} &= [\mathcal{L}(\mathbf{H}^{r1}), \cdots, \mathcal{L}(\mathbf{H}^{rN})] \\
&= [(\mathbf{y}^{rt}_{11}, \cdots, \mathbf{y}^{rt}_{1T}), \cdots, (\mathbf{y}^{rt}_{N1}, \cdots, \mathbf{y}^{rt}_{NT})] \\
&= [\mathbf{Y}^{rt}_1, \cdots, \mathbf{Y}^{rt}_N],
\end{aligned} \tag{12}
$$

where $\mathbf{Y}^{rt}_j = [\mathbf{y}^{rt}_{j1}, \cdots, \mathbf{y}^{rt}_{jT}] \in \mathbb{R}^{d_{rt} \times T}$ denotes the regional temporal feature matrix associated with the j-th brain region, d_{rt} is the hidden units dimension of regional temporal BiLSTM.

On the other hand, to learn the temporal information based on the global feature matrices $\hat{\mathbf{H}}^g_i$, we concatenate the columns of $\hat{\mathbf{H}}^g_i$ into a longer feature vector, denoted by $\hat{\mathbf{h}}^g_i$, i.e.,

$$
\hat{\mathbf{h}}^g_i = [(\hat{\mathbf{h}}^g_{i1})^{\mathrm{T}}, \cdots, (\hat{\mathbf{h}}^g_{iK})^{\mathrm{T}}]^{\mathrm{T}}. \tag{13}
$$

Let $\mathbf{Y}^g = [\hat{\mathbf{h}}^g_1, \cdots, \hat{\mathbf{h}}^g_T]$. Then, the global temporal feature $\mathbf{Y}^{gt}$ can be computed as

$$
\mathbf{Y}^{gt} = \mathcal{L}(\mathbf{Y}^g) = [\mathbf{y}^{gt}_1, \cdots, \mathbf{y}^{gt}_T] \in \mathbb{R}^{d_{gt} \times T}, \tag{14}
$$

where d_{gt} is the hidden units dimension of global temporal BiLSTM.

The final feature vector, denoted by $\mathbf{y}^{rg}$, of the EEG sample $\mathbf{X}$ (consisting of T segments) containing both regional and global information is expressed as

$$
\mathbf{y}^{rg} = [(\mathbf{y}^{rt}_{1T})^{\mathrm{T}}, (\mathbf{y}^{rt}_{2T})^{\mathrm{T}}, \cdots, (\mathbf{y}^{rt}_{NT})^{\mathrm{T}}, (\mathbf{y}^{gt}_T)^{\mathrm{T}}]^{\mathrm{T}}. \tag{15}
$$

C. Classifier and discriminator

Based on the final feature vector $\mathbf{y}^{rg}$, we can predict the class label of the input EEG sample $\mathbf{X}$ by using the simple linear transform approach, which can be formulated as

$$
\mathbf{O} = \mathbf{G}\mathbf{y}^{rg} + \mathbf{b}_c = [o_1, o_2, \cdots, o_C], \tag{16}
$$

where $\mathbf{G}$ and $\mathbf{b}_c$ are respectively the transform matrix and bias, C is the number of class. The elements of the output $\mathbf{O}$ are then fed into a softmax function for emotion class prediction, i.e.,

$$
P(c|\mathbf{X}) = \max \left\{ \frac{\exp(o_k)}{\sum_{i=1}^C \exp(o_i)} | k = 1, \cdots, C \right\}, \tag{17}
$$

where $P(c|\mathbf{X})$ denotes the probability for the input $\mathbf{X}$ being predicted as the c-th class.

Now suppose that we have M data samples from source domain, which are expressed as M matrices $\mathbf{X}^S_i$ $(i = 1, \cdots, M)$. Then, the loss function of the classifier can be formulated as

$$
L_c(\mathbf{X}^S_1, \cdots, \mathbf{X}^S_M; \theta_f, \theta_c) = \sum_{i=1}^M \sum_{c=1}^C -\tau(l_i, c) \times \log P(c|\mathbf{X}^S_i), \tag{18}
$$

where l_i denotes the ground-truth label of $\mathbf{X}^S_i$, θ_f and θ_c denote the parameters of feature extractor and classifier, and $\tau(l_i, c)$ is expressed as

$$
\tau(l_i, c) = \begin{cases} 1, & \text{if } l_i = c, \\ 0, & \text{otherwise.} \end{cases} \tag{19}
$$

Consequently, from (18) and (19), by minimizing the loss function $L_c(\mathbf{X}^S_1 \cdots, \mathbf{X}^S_M; \theta_f, \theta_c)$, we would be able to achieve the maximal probability of correctly predicting the emotion class of each training sample.

Let $\mathbf{X}_{test}$ be a test sample. Then, we use the following formula to determine the emotion class label of $\mathbf{X}_{test}$:

$$
l_{test} = \arg\max_c \{P(c|\mathbf{X}_{test})|c = 1, \cdots, C\}, \tag{20}
$$

where l_{test} is the predicted label of the testing sample $\mathbf{X}_{test}$.

In dealing with EEG emotion recognition, it is notable that the training and testing EEG data samples may come from different domains, e.g., the training and testing data samples come from different subjects. In this case, the emotion recognition model trained based on the training data may not be well suitable for the testing data. To solve this problem, we introduce a discriminator that works corporately with the classifier to produce emotion-discriminative and domain-invariant features.

Specifically, suppose that we are given two data sets $\mathbf{X}^S = \{\mathbf{X}^S_1, \cdots, \mathbf{X}^S_{M_1}\}$ and $\mathbf{X}^T = \{\mathbf{X}^T_1, \cdots, \mathbf{X}^T_{M_2}\}$ from source domain and target domain, respectively, where M_1 and M_2 denote the number of source data set and target data set, respectively. To alleviate the domain difference, we introduce the following loss function

$$
L_d(\mathbf{X}^S_i, \mathbf{X}^T_j; \theta_f, \theta_d) = -\sum_{i=1}^{M_1} \log P(0|\mathbf{X}^S_i) - \sum_{j=1}^{M_2} \log P(1|\mathbf{X}^T_j), \tag{21}
$$

where $P(0|\mathbf{X}^S_i)$ denotes the probability that $\mathbf{X}^S_i$ belongs to source domain while $P(1|\mathbf{X}^T_i)$ denotes the probability that $\mathbf{X}^T_i$ belongs to target domain, respectively, θ_d denotes the parameter of discriminator. By maximizing the above loss function of discriminator, the feature extracting process would results in domain-invariable features to alleviate the domain difference in emotion recognition.

D. The optimization of R2G-STNN

In the aforementioned section, we pointed out that by minimizing the loss function of (18) we would achieve the better emotion class prediction for the training data samples, while by maximizing the loss function of (21) we would be able to achieve the domain-invariable features to alleviate the domain difference in emotion recognition. Consequently, by simultaneously minimizing the loss function of (18) and maximizing the loss function of (21), we would be able to achieve better emotion classification. For this reason, we define the overall loss function of R2G-STNN as

$$
\begin{aligned}
L(\mathbf{X}^S, \mathbf{X}^T | \theta_f, \theta_c, \theta_d) &= L_c(\mathbf{X}^S; \theta_f, \theta_c) \\
&\quad - L_d(\mathbf{X}^S, \mathbf{X}^T; \theta_f, \theta_d), \tag{22}
\end{aligned}
$$

and our target is to find the optimal parameters to minimize the loss function $L(\mathbf{X}^S, \mathbf{X}^T | \theta_f, \theta_c, \theta_d)$.

The optimal parameters of $L(\mathbf{X}^S, \mathbf{X}^T | \theta_f, \theta_c, \theta_d)$ can be solved by iteratively minimizing $L_c(\mathbf{X}^S; \theta_f, \theta_c)$ and maximizing $L_d(\mathbf{X}^S, \mathbf{X}^T; \theta_f, \theta_d)$. Specifically, we adopt stochastic

gradient descent (SGD) algorithm [29] to find the optimal model parameters, i.e.,

$$(\hat{\theta}_f, \hat{\theta}_c) = \arg\min_{\theta_f, \theta_c} L_c(\mathbf{X}^S, (\theta_f, \theta_c), \hat{\theta}_d), \quad (23)$$

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

Through minimizing the loss function L_c, the feature extractor will able to learn the emotion discriminative features. On the other hand, by maximizing the loss function L_d, it can extract domain-invariant features. Consequently, by simultaneously minimizing L_c while maximizing L_d, we can finally obtain the emotion-discriminative while domain-invariant features for emotion recognition.

Additionally, in solving the optimal parameters of R2G-STNN, we also introduce a gradient reverse layer (GRL) for the discriminator to change the maximizing problem into a minimizing problem, such that the parameters can be optimized by using SGD approach, where GRL acts as an identity transform in the forward-propagation but reverses the gradient sign while performing the back-propagation operation. In this case, the update of the parameters can be formulated as

$$\theta_d \leftarrow \theta_d - \alpha\frac{\partial L_d}{\partial \theta_d}, \quad \theta_f \leftarrow \theta_f + \alpha\frac{\partial L_d}{\partial \theta_f}, \quad (25)$$

where α is the learning rate.

IV. EXPERIMENTS

In this section, we will conduct extensive experiments on SEED database [7] to evaluate the proposed R2G-STNN method. The SEED database was built by using 62-channels' ESI NeuroScan system to record EEG signals, in which the of electrodes are located according to the 10-20 system [21]. We group the 62 electrodes into 16 clusters, i.e., the cluster number $N = 16$, based on the spatial locations of the electrodes. Fig. 3 shows the divisions of the 62 electrodes into 16 clusters whereas Table I summarizes the EEG electrodes in each cluster (brain region) as well as the handcraft EEG feature set size in each brain region. In the SEED database, there are 15 subjects and each subject contains the EEG data recorded from three sessions. For each subject, every session contains 15 trials of EEG samples and each trial contains 185-238 samples covering three emotion classes, i.e., positive, negative and neutral emotions. As a result, there are totally 3200 samples in every session, in which each emotion contains about 1060 samples. Additionally, for each sample, the number of EEG segments T is fixed at $T = 9$ such that each EEG sample corresponds to a $5\times62\times9$ handcraft feature tensor. In addition, the parameters of d_r, d_g, d_{rt}, and d_{gt} are respectively fixed at 100, 150, 200, and 250 throughout the experiments.

A. Subject-dependent EEG Emotion Recognition Experiment

In this experiment, we adopt the similar subject-dependent EEG emotion recognition protocol used in [7] and [15] to evaluate the proposed method, where both training and testing data come from the same subject but different EEG trials. Specifically, we choose 9 trials of EEG signals in every session to serve as training data set and use the other 6 trials from
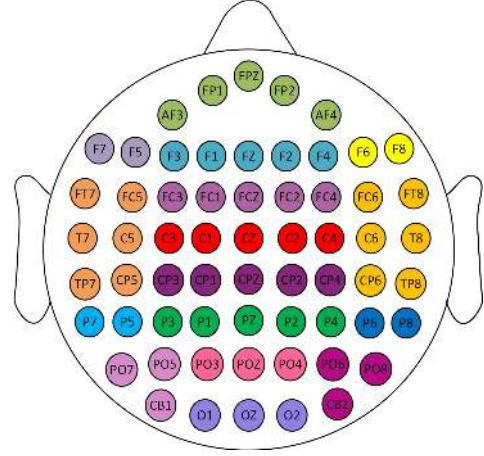

Fig. 3: An illustration of the divisions of the 62 electrodes into 16 clusters, where the same color denotes the electrodes are grouped into the same brain region.

TABLE I: The EEG electrodes associated with each brain region and the data size in each brain region in the experiment.

Brain region	Electrode name	EEG data size $(d \times n_j)$
Pre-Frontal	AF3, FP1 FPZ, FP2, AF4	5×5
Frontal	F3, F1 FZ, F2, F4	5×5
Left Frontal	F7, F5	5×2
Right Frontal	F8, F6	5×2
Left Temporal	FT7, FC5, T7 C5, TP7, CP5	5×6
Right Temporal	FT8, FC6, T8 C6, TP8, CP6	5×6
Frontal Central	FC3, FC1 FCZ, FC2, FC4	5×5
Central	C3, C1, CZ, C2, C4	5×5
Central Parietal	CP3, CP1 CPZ, CP2, CP4	5×5
Left Parietal	P7, P5	5×2
Right Parietal	P8, P6	5×2
Parietal	P3, P1, PZ, P2, P4	5×5
Left Parietal Occipital	PO7, PO5, CB1	5×3
Right Parietal Occipital	PO8, PO6, CB2	5×3
Parietal Occipital	PO3, POZ, PO4	5×3
Occipital	O1, OZ, O2	5×3

the same session as testing data. Then, we use the proposed R2G-STNN method to conduct the EEG emotion recognition experiments and calculate the average recognition accuracy and standard deviation of all the 15 subjects as the final recognition result. For comparison purpose, we also conduct the same experiments using several commonly used methods, which are listed as follows:

- Three baseline methods: Support Vector Machine (SVM) [30], Random Forest (RF) [31], and Canonical Correlation Analysis (CCA) [32];

TABLE II: The mean accuracies (ACC) and standard deviations (STD) of the various EEG emotion recognition methods in the subject-dependent experiment.

Method	SVM [30]	RF [31]	CCA [32]	GSCCA [4]	DBN [7]	GRSLR [9]
ACC/STD (%)	83.99/9.72	78.46/11.77	77.63/13.21	82.96/9.95	86.08/8.34	87.39/8.64
Method	GCNN [33]	DGCNN [14]	DANN [34]	BiDANN [15]	R2G-STNN	
ACC/STD (%)	87.40/9.20	90.40/8.49	91.36/8.30	92.38/7.04	**93.38/5.96**	

- Two subspace learning methods: Group Sparse Canonical Correlation Analysis (GSCCA) [8] and Graph Regularized Sparse Linear Regression (GRSLR) [9];
- Five deep learning methods: Deep Belief Networks (DBN) [7], Graph Convolutional Neural Networks (GCNN) [33], Dynamical Graph Convolutional Neural Networks (DGCNN) [14], Domain Adversarial Neural Network (DANN) [34], and Bi-hemisphere Domain Adversarial Neural Network (BiDANN) [15].

Table II shows the experimental results of the various methods, from which we can see that the proposed R2G-STNN method achieves the recognition accuracy as high as 93.38%, which is the best recognition result among the various recognition methods. The better recognition results of R2G-STNN may largely attribute to the fact that R2G-STNN not only utilizes the temporal information of the EEG signal but also the spatial information of the brain regions, which would be in favor of the learning of more discriminative features related to emotion. Moreover, to visualize the confusion among the three emotions, we also depict the confusion matrix of the emotion recognition results in Fig. 4, from which we observe that both positive and neutral emotions are less confused than the negative emotion.

Fig. 4: The EEG emotion recognition confusion matrix of the R2G-STNN method in the subject-dependent experiment.

In addition, to investigate the impacts of different frequency bands of EEG signals on the emotion recognition, we also conduct additional experiments by adopting the similar approach used in [8]. Specifically, we firstly extract DE features from the raw EEG signals with respect to five different frequency bands, i.e., δ, θ, α, β and γ frequency bands. Then, we conduct EEG emotion recognition experiments based on the DE features extracted from the five frequency bands, respectively. The EEG emotion recognition results are shown in Table III, from which we can observe that the higher frequency bands such as β and γ frequency bands, achieve better performance than the lower ones such as δ, θ and α frequency bands. This observation coincides with the neurophysiology research findings about emotion [35].

TABLE III: The mean accuracies (ACC) and standard deviations (STD) of the various EEG emotion recognition methods with respect to five frequency bands in the subject-dependent experiment.

Methods	The results (%) of ACC/STD				
	δ	θ	α	β	γ
SVM [30]	60.50 (14.14)	60.95 (10.20)	66.64 (14.41)	80.76 (11.56)	79.56 (11.38)
RF [31]	64.56 (8.32)	65.27 (11.64)	65.67 (13.94)	73.35 (14.35)	74.48 (12.80)
CCA [32]	55.30 (12.02)	55.75 (10.99)	64.96 (12.05)	69.16 (11.45)	70.67 (14.06)
GSCCA [4]	63.92 (11.16)	64.64 (10.33)	70.10 (14.76)	76.93 (11.00)	77.98 (10.72)
DBN [7]	64.32 (12.45)	60.77 (10.42)	64.01 (15.97)	78.92 (12.48)	79.19 (14.58)
GRSLR [9]	63.90 (11.83)	62.61 (10.73)	71.11 (9.04)	81.18 (10.74)	81.91 (10.36)
GCNN [33]	72.75 (10.85)	74.40 (8.23)	73.46 (12.17)	83.24 (9.93)	83.36 (9.43)
DGCNN [14]	74.25 (11.42)	71.52 (5.99)	74.43 (12.16)	83.65 (10.17)	85.73 (10.64)
DANN [34]	72.13 (11.22)	68.75 (7.40)	70.27 (10.84)	83.35 (11.46)	87.89 (11.35)
BiDANN [15]	76.97 (10.95)	75.56 (7.88)	81.03 (11.74)	**89.65** **(9.59)**	88.64 (9.46)
R2G-STNN	**77.76** **(9.92)**	**76.17** **(7.43)**	**82.30** **(10.21)**	88.35 (10.52)	**88.90** **(9.97)**

B. Subject-independent EEG Emotion Recognition Experiment

In this experiment, we will investigate the subject-independent EEG emotion recognition problem, in which the training EEG data samples and the testing ones come from different subjects [15][46]. To this end, we adopt leave-one-subject-out (LOSO) cross validation strategy to conduct the experiment, in which we circularly take one subject's EEG signals as testing data and the EEG signals of all the other subjects as training data. The average result of all the recognition accuracies is then calculated after each subject has been used once as testing data. For comparison purpose, we also conduct the same experiment using several methods, which are listed as follows:

- Two baseline methods: SVM [30] and kernel principal component analysis (KPCA) [39];
- Nine transfer subspace learning methods: Kullback-Leibler importance estimation procedure (KLIEP) [36], unconstrained least-squares importance fitting (UL-SIF) [37], selective transfer machine (STM) [38], transfer component analysis (TCA) [40], transfer kernel learning

TABLE IV: The mean accuracies (ACC) and standard deviations (STD) of the various EEG emotion recognition methods in the subject-independent experiment.

Method	KLIEP [36]	ULSIF [37]	STM [38]	SVM [30]	KPCA [39]
ACC/STD (%)	45.71/17.76	51.18/13.57	51.23/14.82	56.73/16.29	61.28/14.62
Method	TCA [40]	TKL [41]	SA [42]	GFK [43]	T-SVM [44]
ACC/STD(%)	63.64/14.88	63.54/15.47	69.00/10.89	71.31/14.09	72.53/14.00
Method	TPT [45]	DGCNN [14]	DANN [34]	BiDANN [15]	R2G-STNN
ACC/STD (%)	76.31/15.89	79.95/9.02	75.08/11.18	83.28/9.60	**84.16/7.63**

(TKL) [41], subspace alignment (SA) [42], geodesic flow kernel (GFK) [43], transductive SVM (T-SVM) [44], and transductive parameter transfer (TPT) [45];
- Three recent deep learning methods: DGCNN [14], DANN [34] and BiDANN [15].

The experimental results of the various methods are shown in Table IV, from which we can again see that R2G-STNN achieves the state-of-the-art performance among the various methods. To visualize the confusion among the three emotions recognized by R2G-STNN, we depict the confusion matrix according to the recognition results. Fig. 5 shows the results of the confusion matrix, from which we can observe that the positive emotion is less confused than both negative and neutral emotions.

Fig. 5: The EEG emotion recognition confusion matrix of the R2G-STNN method in the subject-independent experiment.

Similar to the subject-dependent experiment, in this experiment we also conduct additional experiments to investigate the impacts of different frequency bands of EEG signals on the emotion recognition. Table V shows the experimental results, from which we can also observe that the higher frequency bands achieve better recognition results than the lower ones. Moreover, from Table V we can see that R2G-STNN achieves the best recognition results for all the five frequency bands among the various methods.

To visualize the impact of the different brain regions on emotion recognition, we also depict the weight distribution over brain regions based on the weighted matrix $\mathbf{W}$ defined in (5), in which the sum of each row of $\mathbf{W}$ is used to demonstrate the contribution of the corresponding brain region to the emotion recognition. Fig. 6 shows the topographical map of the sum of each row of $\mathbf{W}$, in which the areas with deeper red color mean significant contributions of the corresponding brain regions. From Fig. 6, we can see that the frontal brain regions are important in EEG emotion recognition experiment, which coincides with the cognition observations of biological psychology [47].

TABLE V: The mean accuracies (ACC) and standard deviations (STD) of the various EEG emotion recognition methods with respect to five frequency bands in the subject-independent experiment.

Methods	The results (%) of ACC (STD)				
	δ	θ	α	β	γ
KLIEP [36]	39.22 (11.31)	35.98 (7.50)	33.31 (6.60)	44.47 (12.89)	42.05 (12.65)
ULSIF [37]	41.32 (11.30)	36.27 (6.84)	38.94 (8.30)	41.87 (13.64)	41.02 (11.65)
STM [38]	44.16 (9.60)	40.89 (8.22)	40.37 (9.82)	42.09 (13.34)	47.97 (12.43)
SVM [30]	43.06 (8.27)	40.07 (6.50)	43.97 (10.89)	48.63 (10.29)	51.59 (11.83)
TCA [40]	44.10 (8.22)	41.26 (9.21)	42.93 (14.33)	43.93 (10.06)	48.43 (9.73)
TKL [41]	48.36 (10.31)	52.60 (11.84)	52.89 (11.07)	55.47 (9.80)	59.81 (12.41)
SA [42]	53.23 (7.47)	50.60 (8.31)	55.06 (10.60)	56.72 (10.78)	64.47 (14.96)
GFK [43]	52.73 (11.90)	54.07 (06.78)	54.98 (11.49)	59.29 (10.75)	66.92 (10.97)
DGCNN [14]	49.79 (10.94)	46.36 (12.06)	48.29 (12.28)	56.15 (14.01)	54.87 (17.53)
DANN [34]	56.66 (6.48)	54.95 (10.45)	59.37 (10.57)	67.14 (7.10)	71.30 (10.84)
BiDANN[15]	62.04 (6.64)	62.13 (7.37)	63.31 (11.46)	73.55 (8.83)	73.25 (9.21)
R2G-STNN	**63.34 (5.31)**	**63.78 (7.53)**	**64.27 (10.88)**	**74.85 (8.02)**	**74.54 (8.41)**

V. DISCUSSIONS AND CONCLUSIONS

In this paper, we proposed a novel R2G-STNN method inspired by the neuroscience findings that the human brain regions have different responses to emotion and applied it to EEG emotion recognition, in which a hierarchical feature learning procedure from regional brain to global brain is proposed to extract the spatial-temporal EEG features. Extensive experiments on SEED EEG emotional database demonstrated that the proposed R2G-STNN method achieves the state-of-the-art performance in both subject-dependent and subject-independent EEG emotion recognition. The better recognition performance of R2G-STNN may largely attribute to the fact the R2G-STNN makes use of weighting layer and both regional and global temporal layers. To verify these points, we also conduct additional experiments by removing some of the layers, which include the following three reduced models:
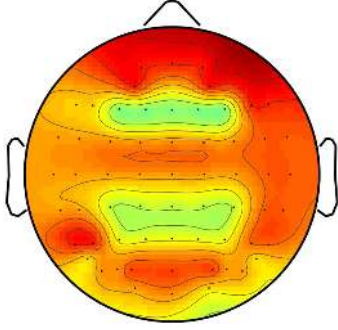
(1) R2G-STNN-R1, which removes both dynamic weighting

Fig. 6: The topographical map of the sum of each row of $\mathbf{W}$, where deeper red color denotes more significant contribution of the corresponding brain region.

layer and regional temporal feature;

(2) R2G-STNN-R2, which neglects the regional temporal feature and only uses the global temporal feature as the final vector for classification;

(3) R2G-STNN-R3, which treats all the brain regions equally, but has the same spatial-temporal structure with R2G-STNN.
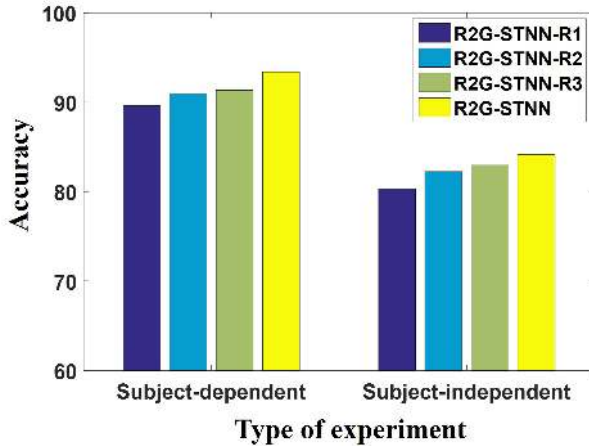

Fig. 7: The comparison of EEG emotion recognition results among four methods: (1) R2G-STNN-R1; (2) R2G-STNN-R2; (3) R2G-STNN-R3; (4) R2G-STNN.

The experimental results are shown in Fig. 7, from which we can see that, for both subject-dependent and subject-independent experiments, the recognition accuracies of R2G-STNN-R1, R2G-STNN-R2, R2G-STNN-R3, and R2G-STNN have the following relationship:

$$
\begin{aligned}
\text{R2G-STNN-R1} &< \text{R2G-STNN-R2} \\
&< \text{R2G-STNN-R3} < \text{R2G-STNN.} \quad (26)
\end{aligned}
$$

The relationship demonstrated in (26) the importance of the dynamic weighting layer and regional temporal feature learning layer, each of which can improve the classification performance in both subject-dependent and subject-independent EEG emotion recognition experiments.

REFERENCES

[1] R. J. Dolan, "Emotion, cognition, and behavior." *Science*, vol. 298, no. 5596, pp. 1191–1194, 2002.
[2] P. D. Purnamasari, A. A. P. Ratna, and B. Kusumoputro, "Development of filtered bispectrum for eeg signal feature extraction in automatic emotion recognition using artificial neural networks," *Algorithms*, vol. 10, no. 2, p. 63, 2017.
[3] R. W. Picard and R. Picard, *Affective computing*. MIT press Cambridge, 1997, vol. 252.
[4] 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.
[5] 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.
[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] 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.
[9] 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.
[10] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring eeg features in cross-subject emotion recognition," *Frontiers in neuroscience*, vol. 12, p. 162, 2018.
[11] H. Cai, X. Zhang, Y. Zhang, Z. Wang, and B. Hu, "A case-based reasoning model for depression based on three-electrode eeg data," *IEEE Transactions on Affective Computing*, 2018.
[12] 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.
[13] P. Pandey and K. Seeja, "Emotional state recognition with eeg signals using subject independent approach," in *Data Science and Big Data Analytics*. Springer, 2019, pp. 117–124.
[14] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Transactions on Affective Computing*, 2018.
[15] 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 *27th International Joint Conference on Artificial Intelligence (IJCAI)*. AAAI Press, 2018, pp. 1561–1567.
[16] Z. G. Doborjeh, M. G. Doborjeh, and N. Kasabov, "Attentional bias pattern recognition in spiking neural networks from spatio-temporal eeg data," *Cognitive Computation*, vol. 10, no. 1, pp. 35–48, 2018.
[17] E. Lotfi and M.-R. Akbarzadeh-T, "Practical emotional neural networks," *Neural Networks*, vol. 59, pp. 61–72, 2014.
[18] J. C. Britton, K. L. Phan, S. F. Taylor, R. C. Welsh, K. C. Berridge, and I. Liberzon, "Neural correlates of social and nonsocial emotions: An fmri study," *Neuroimage*, vol. 31, no. 1, pp. 397–409, 2006.
[19] A. Etkin, T. Egner, and R. Kalisch, "Emotional processing in anterior cingulate and medial prefrontal cortex," *Trends in Cognitive Sciences*, vol. 15, no. 2, pp. 85–93, 2011.
[20] K. A. Lindquist and L. F. Barrett, "A functional architecture of the human brain: emerging insights from the science of emotion," *Trends in Cognitive Sciences*, vol. 16, no. 11, pp. 533–540, 2012.
[21] R. Oostenveld and P. Praamstra, "The five percent electrode system for high-resolution eeg and erp measurements," *Clinical neurophysiology*, vol. 112, no. 4, pp. 713–719, 2001.
[22] W. Heller and J. B. Nitscke, "Regional brain activity in emotion: A framework for understanding cognition in depression," *Cognition & Emotion*, vol. 11, no. 5-6, pp. 637–661, 1997.
[23] R. J. Davidson, "Affective style, psychopathology, and resilience: brain mechanisms and plasticity." *American Psychologist*, vol. 55, no. 11, p. 1196, 2000.
[24] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review," *Behavioral and brain sciences*, vol. 35, no. 3, pp. 121–143, 2012.

[25] Z. Yu, V. Ramanarayanan, D. Suendermann-Oeft, X. Wang, K. Zechner, L. Chen, J. Tao, A. Ivanou, and Y. Qian, "Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech," in *Automatic Speech Recognition and Understanding (ASRU), 2015 IEEE Workshop on*. IEEE, 2015, pp. 338–345.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3118–3125.

[28] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[29] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

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

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

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

[33] 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.

[34] 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.

[35] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognition and emotion*, vol. 23, no. 2, pp. 209–237, 2009.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] M. Long, J. Wang, J. Sun, and S. Y. Philip, "Domain invariant transfer kernel learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1519–1532, 2015.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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 (IJCAI)*. AAAI Press, 2016, pp. 2732–2738.

[47] J. A. Coan and J. J. Allen, "Frontal eeg asymmetry as a moderator and mediator of emotion," *Biological psychology*, vol. 67, no. 1-2, pp. 7–50, 2004.

Yang Li received the B.S. degree in electronic information and science technology from School of Physics and Electronics, Shandong Normal University, China, in 2012, the M.S. degree in electronic and communication engineering from School of Electronic Engineering, Xidian University, China, in 2015. Currently, he is pursuing the Ph.D. degree in information and communication engineering in Southeast University, China. His researches focus on Pattern Recognition and Machine Learning.
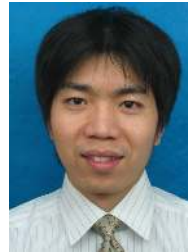
Wenming Zheng received the B.S. degree in computer science from Fuzhou University, Fuzhou, China, in 1997, the M.S. degree in computer science from Huaqiao University, Quanzhou, China, in 2001, and the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 2004. Since 2004, he has been with the Research Center for Learning Science, Southeast University. He is currently a Professor with the Key Laboratory of Child Development and Learning Science, Ministry of Education, Southeast University. His research interests include affective computing, pattern recognition, machine learning, and computer vision. He is an associated editor of IEEE Transactions on Affective Computing and an editorial board member of The Visual Computer.

Lei Wang received his PhD degree from Nanyang Technological University, Singapore. He is now Associate Professor at School of Computing and Information Technology of University of Wollongong, Australia. His research interests include machine learning, pattern recognition, and computer vision. Dr. Wang has published 120+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IJCV, CVPR, ICCV and ECCV, etc. He was awarded the Early Career Researcher Award by Australian Academy of Science and Australian Research Council. He served as the General Co-Chair of DICTA 2014 and on the Technical Program Committees of 20+ international conferences and workshops. Lei Wang is senior member of IEEE.

Yuan Zong received the BS and MS degrees in electronics engineering from Nanjing Normal University, Nanjing, China, in 2011 and 2014, respectively. He is currently working toward the PhD degree in the Key Laboratory of Child Development and Learning Science of Ministry of Education, School of Biological Sciences and Medical Engineering, Southeast University, Nanjing, China. From 2016 to 2017, he was working as a visiting student with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. His research interests include affective computing, pattern recognition, and computer vision.

Zhen Cui received Ph.D. degrees from Institute of Computing Technology (ICT), Chinese Academy of Sciences in 2014. He was a Research Fellow in the Department of Electrical and Computer Engineering at National University of Singapore (NUS) from 2014 to 2015. He also spent half a year as a Research Assistant on Nanyang Technological University (NTU) from Jun. 2012 to Dec. 2012. Currently he is a Professor of Nanjing University of Science and Technology, China. His research interests cover computer vision, pattern recognition and machine learning, especially focusing on deep learning, manifold learning, sparse coding, face detection/alignment/recognition, object tracking, image super resolution, emotion analysis, etc. He has published several papers in the top conferences NIPS/CVPR/ECCV and some journals of IEEE Transactions. More details can be found in http://aip.seu.edu.cn/zcui/.