An Efficient LSTM Network for Emotion Recognition From Multichannel EEG Signals

Xiaobing Du, Cuixia Ma, Guanhua Zhang, Jinyao Li, Yu-Kun Lai, Guozhen Zhao, Xiaoming Deng, Yong-Jin Liu, *Senior Member, IEEE*, and Hongan Wang, *Senior Member, IEEE*

Abstract—Most previous EEG-based emotion recognition methods studied hand-crafted EEG features extracted from different electrodes. In this article, we study the relation among different EEG electrodes and propose a deep learning method to automatically extract the spatial features that characterize the functional relation between EEG signals at different electrodes. Our proposed deep model is called **AT**tention-based **LSTM** with **D**omain **D**iscriminator (ATDD-LSTM), a model based on Long Short-Term Memory (LSTM) for emotion recognition that can characterize nonlinear relations among EEG signals of different electrodes. To achieve state-of-the-art emotion recognition performance, the architecture of ATDD-LSTM has two distinguishing characteristics: (1) By applying the attention mechanism to the feature vectors produced by LSTM, ATDD-LSTM automatically selects suitable EEG channels for emotion recognition, which makes the learned model concentrate on the emotion related channels in response to a given emotion; (2) To minimize the significant feature distribution shift between different sessions and/or subjects, ATDD-LSTM uses a domain discriminator to modify the data representation space and generate domain-invariant features. We evaluate the proposed ATDD-LSTM model on three public EEG emotional databases (DEAP, SEED and CMEED) for emotion recognition. The experimental results demonstrate that our ATDD-LSTM model achieves superior performance on subject-dependent (for the same subject), subject-independent (for different subjects) and cross-session (for the same subject) evaluation.

Index Terms—Emotion recognition, multichannel EEG, LSTM, attention mechanism, domain adaptation

1 Introduction

Human emotions are complex psychological and physiological expressions, which are often related to subjective feelings, temperament, personality, motivational tendencies, behavioral reactions and physiological arousal [1], [2]. Both behavioral and physiological signals have been explored for human emotion recognition. The most commonly used behavioral signals include speech, facial expressions, as well as hand and body gestures [3], [4]. Compared to behavioral signals that are easy to disguise in emotion recognition, physiological measurements are more reliable to recognize human emotions [1]. Electroencephalography (EEG) is a physiological signal with an excellent temporal resolution, which can be directly used for emotion recognition through analyzing immediate brain activities elicited by emotional stimuli [5], [6]. In recent years, with the advance of brain-computer interface (BCI) techniques and the development of practical and precise emotion annotation tools (e.g., [7]), EEG-based applications have been flourishing, e.g., [8], [9], [10], [11].

Exploring practical EEG features for emotion recognition is vital. Although the EEG measurements usually have sufficient density to sample the brain electrical field (i.e., generally more than 30 electrodes are placed on the scalp), the spatial feature that optimally characterizes the functional relations among different EEG channels is rarely considered. Recently, a few pioneering works [12], [13], [14] have been proposed that explore such spatial features through multichannel EEG signals. Among these methods, the state of the art [14] introduced a dynamic graph convolutional neural network (DGCNN) to learn the optimal adjacency matrix M automatically. However, M can only represent linear relations, which characterizes the strengths of connections between pairs of EEG channels. In this paper, we propose ATtention-based LSTM with Domain Discriminator (ATDD-LSTM), a model based on Long Short-Term Memory (LSTM) for emotion recognition that can characterize *nonlinear* relations among multichannel EEG signals.

In the proposed ATDD-LSTM, the input for a given temporal sample is a channel sequence representing the EEG

- *Xiaobing Du, Jinyao Li, and Xiaoming Deng are with the Beijing Key Laboratory of Human Computer Interactions, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China. E-mail: {duxiaobing16, lijinyao19}@mails.ucas.ac.cn, xiaoming@iscas.ac.cn.*
- *Guanhua Zhang and Yong-Jin Liu are with BNRist, MOE-Key Laboratory of Pervasive Computing, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. E-mail: zgh17@mails.tsinghua.edu.cn, liuyongjin@tsinghua.edu.cn.*
- *Yu-Kun Lai is with the School of Computer Science and Informatics, Cardiff University, CF24 3AA Cardiff, Wales, U.K. E-mail: laiy4@cardiff.ac.uk.*
- *Guozhen Zhao is with the CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing 100101, China , and also with the Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China. E-mail: zhaogz@psych.ac.cn.*
- *Cuixia Ma and Hongan Wang are with the State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China, the University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Beijing Key Laboratory of Human Computer Interactions, International Joint Laboratory of artificial intelligence and emotional interaction, Beijing 100190, China. E-mail: {cuixia, hongan}@iscas.ac.cn.*

Manuscript received 4 Nov. 2019; revised 3 June 2020; accepted 22 July 2020. Date of publication 3 Aug. 2020; date of current version 6 Sept. 2022. (Corresponding author: Yong-Jin Liu and Cuixia Ma.) Recommended for acceptance by B. Hu. Digital Object Identifier no. 10.1109/TAFFC.2020.3013711

signal from different electrodes, and the output is the emotion label corresponding to the input EEG channel sequence. Unlike most previous research, we focus on addressing the following two challenges: (1) selecting effective emotion-related channels and (2) building domain-invariant features to ensure robust recognition across subjects and different sessions of a subject.

Channel Selection. Not all EEG signals are related to emotion. The EEG signals collected from different electrodes on the scalp reflect a variety of information, and it is well known that the electrodes located in the prefrontal cortex are associated with the emotional process [15], [16], [17], [18]. In our study, in addition to making use of existing neurophysiological research for establishing the relations among multi-channels, we expect that the data-driven approach can also help explore more subtle relations. To do so, in ATDD-LSTM we propose to use an attention mechanism to optimally search for emotion-related channels in response to a given emotion.

Domain-Invariant Features. Many previous studies build the emotion recognition model on the basis of each individual person's brain responses, due to the data distribution shift between different persons. Despite the popularity of subject-dependent models in EEG-based emotion recognition [19], [20], [21], some recent studies [22], [23], [24], [25], [26], [27], [28] suggest building models specially designed for subject-independent evaluation. To address the data distribution shift problem, we incorporate a domain discriminator in our model to constrain the features extracted from training (source) data and test (target) data to have similar distributions.

To sum up, our proposed ATDD-LSTM model not only extracts discriminative emotion-related features by learning the nonlinear relationships between EEG channels, but also constrains a domain-invariant data representation through a global domain discriminator. We consider a multi-channel EEG signal from different electrodes as a channel sequence, which allows us to use LSTM (typically applied for temporal feature extraction) to extract sequential features, taking nonlinear channel relations into account. Then we feed the sequential features to two network branches, namely a domain discriminator and an attention-based encoder-decoder. The domain discriminator is used to distinguish which domain (training data or test data) the input comes from, to narrow down the distribution shift. The attention-based encoder-decoder enables the representation of EEG features to focus on emotion-related features and maps the encoded features into a label space for classification. A reconstruction process (decoder) aims to improve the performance of the encoder. We demonstrate the effectiveness of our proposed ATDD-LSTM model on three main benchmark EEG emotional databases (DEAP [2], SEED [21] and CMEED [16], [17]). Besides, we also conduct ablation studies, which show the effectiveness of our domain discriminator module and attention mechanism. In particular, we make the following contributions:

- We apply the LSTM to the EEG channel sequence to characterize nonlinear relations among multi-channel EEG signals.
- We introduce the attention mechanism in our model to make the features focus on emotion-related aspects of the EEG signals when different emotion categories are concerned. Results show that the attention mechanism is effective in selecting emotion-related channels.
- We introduce a domain discriminator to constrain the feature distributions between training and test domains to be similar, which addresses the data distribution shift problem in cross-subject and cross-session scenarios, making the learned model more practically useful.

2 Related Work

2.1 Emotion Models

In general, two widely used emotion models exist for characterizing the emotional space: one is the discrete model and the other is the dimension model. In the discrete model, the emotional space is described by a few basic discrete emotions. Although there is no consensus reached for what emotions are "basic", many studies use at least six basic emotions: joy, sadness, surprise, fear, anger, and disgust [29]. In the dimension model (e.g., [30]), the emotional space is characterized with continuous coordinates in two or three dimensions, i.e., the valence-arousal or valence-arousal-dominance dimensions. Specifically, the valence dimension ranges from negative to positive, the arousal dimension ranges from calm to peaceful, then to active and finally to excited, and the dominance characterizes an individual's status ranging from *in control* to *being controlled*. The ATDD-LSTM method proposed in this paper can be used for both emotion models.

2.2 EEG Features for Emotion Recognition

Feature extraction plays an important role in EEG-based emotion recognition [31]. A variety of feature extraction methods have been proposed and the obtained EEG features can be generally divided into three categories: time-domain features, frequency-domain features and time-frequency features.

Time-domain features mainly capture the temporal statistical information of EEG signals. Some representative time-domain EEG features include Hjorth feature [32], fractal dimension feature [33] and higher order crossing feature [34], etc. Frequency-domain features mainly capture the emotion information from the frequency domain perspective. For extracting frequency-domain features, it is essential to decompose the EEG signal into several frequency bands (e.g., δ band (1-3 Hz), θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz), and γ band (31-50 Hz)) [35], [36]. Then the EEG features can be extracted from each frequency band, respectively. Popular EEG frequency-domain feature extraction methods include Fourier transform (FT), power spectral density (PSD), wavelet transform (WT) [37] and differential entropy (DE) [36].

Time-frequency features capture both the temporal information and the information from the frequency domain, which are extracted from the unit time signal segmented by a sliding window. Based on time-frequency features, existing research achieved considerable success, e.g., Liu *et al.* [11] recognized emotions in real-time using the short-time Fourier transform (STFT) with a 2-second sliding time window

for feature extraction. Zheng *et al.* [38] extracted two types of features, namely, power spectral density and differential entropy, using STFT with a 4-second time window without overlapping for EEG-based emotion recognition. In this paper, we use the DE feature from the STFT with a 1-second sliding time window for subtle temporal analysis.

2.3 EEG-Based Emotion Recognition Models

Many emotion recognition methods have been proposed based on EEG signals. The reader is referred to [37] for a comprehensive overview. The majority of these research works utilize traditional machine learning algorithms to recognize/predict emotional states. For example, Liu *et al.* [11] adopted support vector machines (SVM) to recognize seven discrete emotions and neutrality. Piho and Tjahjadi [19] compared three supervised learning algorithms—SVM, K-nearest neighbors (KNN), and naive Bayes (NB)—and KNN achieves the best accuracy when recognizing valence.

Recently, deep neural networks have been successfully introduced into EEG-based emotion recognition and achieved the state-of-the-art performance. Zheng and Lu [21] fed PSD, DE, the differential asymmetry feature (DASM), the rational asymmetry feature (RASM) and the differential caudality feature (DCAU) into a Deep Belief Network (DBN) for extracting high-level emotional features, and the features are used for emotion classification. Tang *et al.* [39] proposed a bimodal deep denoising autoencoder and a bimodal-LSTM model that uses wavelet EEG features as input. A deep framework that adopted a convolutional neural network (CNN) kernel to extract emotional related features using the input of time, frequency, and electrode location features was proposed in [40].

While many existing deep models can perform well in EEG-based emotion recognition, less attention is paid to extract the EEG feature that optimizes the functional relations among different EEG channels/electrodes. A few pioneering research works [12], [13], [14] attempted to address this issue. Li *et al.* [12] proposed a preprocessing method which uses wavelet and scalogram transform to encapsulate the multi-channel EEG signals into grid-like frames, and hybrid CNN and recurrent neural network (RNN) to extract task-related features. By solving a group feature selection problem from raw EEG frequency features, Zheng [13] proposed a group sparse canonical correlation analysis for simultaneous selection from EEG multi-channels. A state-of-the-art work [14] characterized the 2D topographical map of EEG electrodes on the scalp with an adjacency matrix. The matrix was further fed into a graph convolutional neural network for optimizing weights in the matrix entities. However, the adjacency matrix can only characterize linear relations. In this paper, we exploit the LSTM network model [41] that has the ability to describe nonlinear relations.

2.4 Attention Mechanism and Domain Adaptation

The attention mechanism is widely used in various visual and natural language processing tasks (e.g., [42], [43], [44], [45], [46]), since it can for example locate the correct regions for image captioning or concentrate on the key part of a sentence given the aspect. Inspired by [46], we propose an attention mechanism for enforcing the model to attend to some automatically selected key information of multi-channel EEG signals in response to a specific emotion.

In practical applications related to EEG-based emotion recognition, data distribution shift exists between subjects and across sessions of the same subject. Due to the non-stationary characteristics of EEG and the wild environments, an EEG-based emotion recognition model trained with training data usually does not generalize well to the test data. This domain shift phenomenon has been addressed in general classification problems [47], [48], where domain adaptation is used to constrain the features to be invariant to the change of domains. For example, Tzeng *et al.* [49] proposed an effective Adversarial Discriminative Domain Adaptation (ADDA) method for cross-domain digit classification tasks. Long *et al.* [50] proposed Conditional Domain Adversarial Networks (CDANs) for aligning different domains of multimodal distributions in classification problems. For EEG-based emotion recognition, Luo *et al.* [51] proposed a novel Wasserstein Generative Adversarial Network Domain Adaptation (WGANDA) framework, which consists of GAN-like components and a two-step training procedure with pre-training and adversarial training to decrease the domain shift. An alternative way to address the domain shift is to use the domain discriminator to minimize the discrepancy between two probability distributions [25]. However, none of these research works can handle multi-channel EEG signals. In this paper, we incorporate a domain discriminator in our framework to generate domain-invariant data features.

3 Overview of the ATDD-LSTM Model

3.1 Input EEG Signal Representation

Five EEG features—differential entropy, power spectral density, differential asymmetry, rational asymmetry, differential caudality—are evaluated in [14] for multichannel EEG signal analysis, in which DE is reported to achieve the best overall performance in emotion recognition. Following [14], we use the DE features of multichannel EEG signals as input to the proposed ATDD-LSTM model.

DE characterizes the logarithm energy spectrum in a certain frequency band. Under the assumption that the EEG signal of a specific channel in a frequency band is approximately a Gaussian distribution, the computation of DE can be formulated as [21]

$$f(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} dx = \frac{1}{2}\log 2\pi e\sigma^2, \tag{1}$$

where X is a given segment of EEG signal with the Gaussian distribution $N(\mu, \sigma^2)$. Specifically, in the public EEG emotional databases (DEAP, SEED and CMEED) evaluated in Section 5, each subject has several recorded samples (called trials) and each sample is a length of multichannel EEG signals elicited under a specified emotion. For each sample, we use short-time Fourier transform with a non-overlapped Hanning window of one second to extract DE in each channel of five different frequency bands, i.e., δ band (1-3 Hz), θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz), and

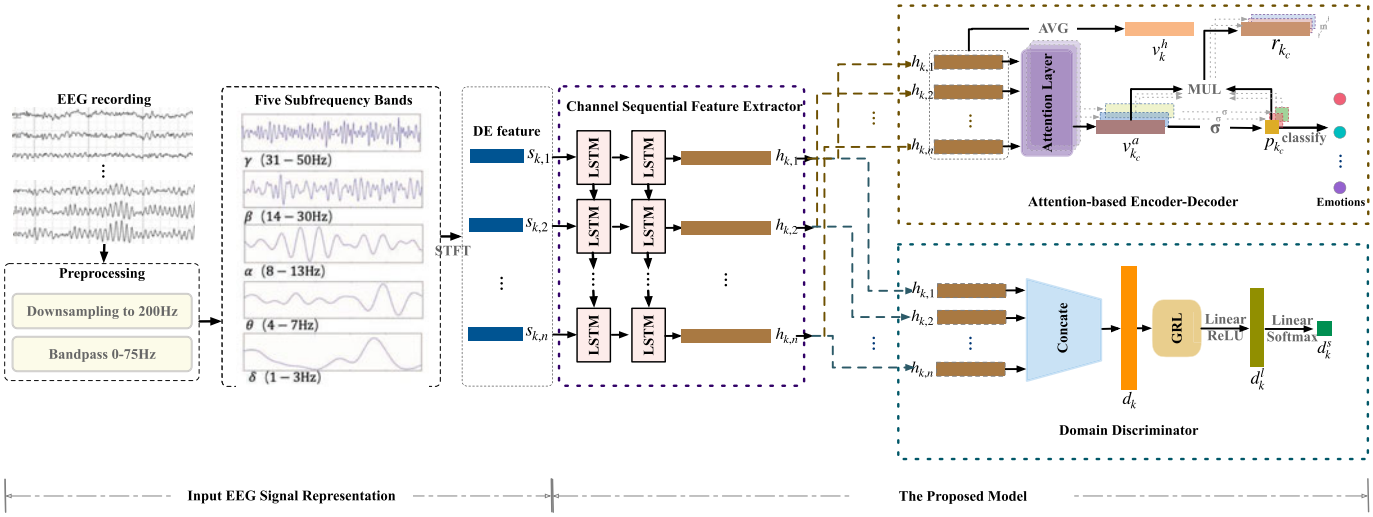


Fig. 1. Our ATDD-LSTM model for emotion recognition based on multichannel EEG data. It consists of a feature extractor, an attention-based encoder-decoder and a domain discriminator. The input to our model is a differential entropy (DE) matrix containing the signal representation extracted from each EEG channel separately. The channel sequential feature extractor aims at capturing sequential features from the channel sequence by using a 2-layer LSTM. The attention-based encoder-decoder learns the importance weights among different channels and emotion classes, and produces classification probabilities. A reconstruction process is also included to add a further constraint to improve learning. The domain discriminator is used to narrow down the feature distribution shift of training set and test set, and therefore helps generate domain-invariant features. See the text of the paper for detailed definition of symbols. In a nullshell, The output of the feature extractor is fed into the attention-based encoder-decoder and the domain discriminator in parallel, and these two modules help extract emotion-related and domain-invariant features through adversarial learning. The final classifier is included in the encoder-decoder module.

γ band (31-50 Hz). The sizes of DE features for 1-second windows are (32,4), (62,5) and (30,5) for DEAP, SEED and CMEED databases,[1] separately. We compute DEs of all windows and concatenate them to form a feature vector for representing one sample.

In the proposed ATDD-LSTM model, the input matrix is denoted as s_k, corresponding to the kth sample, $s_k = [s_{k,1}, s_{k,2}, \ldots, s_{k,n}]^T \in \mathbb{R}^{n \times d_x}$, where d_x is the dimension of DE feature vector containing DE features from different frequency bands, and $s_{k,i}$ is a d_x dimensional vector, which is a time-frequency feature corresponding to the ith channel of the kth sample.

3.2 Proposed Model

We illustrate the framework of the proposed ATDD-LSTM model in Fig. 1. Our model consists of three modules: a sequential feature extractor, a domain discriminator and an attention-based encoder-decoder. The sequential feature extractor utilizes a 2-layer LSTM to capture the sequential feature of multi-channel EEG recordings. The domain discriminator is designed to reduce the effect of the distribution shift between features of training and test sets and help the feature extractor to produce domain-invariant features. The attention-based encoder-decoder includes two parts. One uses the attention mechanism to focus on emotion-related channels and construct an integrated representation, and then predicts the classification probability. The other combines the above feature and probability to perform reconstruction, which leads to an encoder-decoder. The encoder-decoder adds further constraints to facilitate learning. By operating the encoder-decoder module m times, one for each emotion category, where m is the number of emotion categories, we can get a prediction probability for emotion recognition. The implementation details of these modules are presented below.

4 System Implementation Details

4.1 Channel Sequential Feature Extractor

To extract sequential features for channel sequences, we utilize a long short-term memory module to learn the sequential dependencies. Formally,

$$(c_{k,t}, h_{k,t}) = LSTM(c_{k,t-1}, h_{k,t-1}, s_{k,t}), \tag{2}$$

where memory cell $c_{k,t}$ and hidden state $h_{k,t}$ are functions of previous $c_{k,t-1}$, $h_{k,t-1}$ and input vector $s_{k,t}$ for channel t of the kth sample.

Although LSTM is typically used for temporal sequences, here we apply LSTM to EEG channel sequences. Note that we order the EEG channels based on the channel ordering methods of each independent database (refer to Fig. 2). Therefore, the LSTM is designed to receive the input as a series with d_x (the size of DE feature vector) variables and n (the number of channels) steps. Finally, we form a matrix[2] $H_k = [h_{k,1}, h_{k,2}, \ldots, h_{k,n}]$ and take H_k as the input to the discriminator and encoder-decoder. In this way, our model can capture the relationships between EEG channels, which is effective for EEG signal processing. Note that an alternative approach is to use the LSTM block to process a batch of n variables over d_x steps. This instead learns the relationship between different frequencies. By exploiting the between-channel dependencies, our proposed approach better extracts useful features; see the ablation study in Section 5.5 (Table 8).

1. In DEAP, δ band is not used, while all five bands are used in SEED and CMEED.

2. In our experiment, the dimension of each vector $h_{k,i}$ is 1,024.

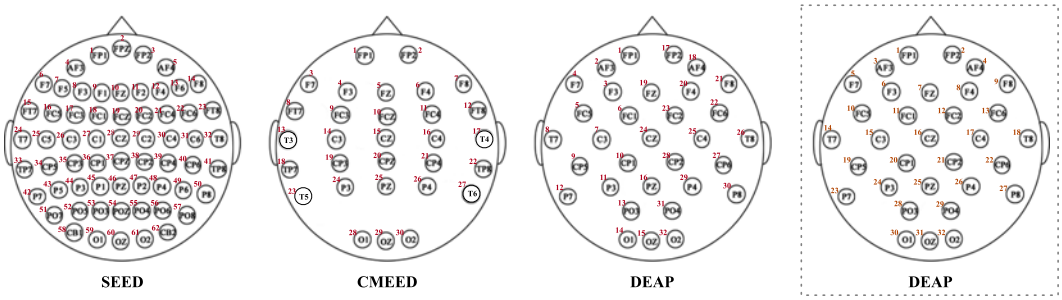


Fig. 2. Order and location of EEG electrodes in three main EEG databases used in this paper. The EEG signals of these databases were recorded according to the international 10-20 system. In the scalp map of each EEG database, we use numbers in red to show the used order of electrodes. These databases were recorded according to different orders of EEG electrodes based on different partition strategies of the brain. SEED and CMEED EEG electrodes are ordered following the rule from left to right and front to back depending on the partition of five brain function regions, and DEAP EEG electrodes are ordered based on the distance between electrodes with the hemisphere restriction. To demonstrate the insensitivity of our method to the order of electrodes, we also show the order of electrodes sorted based on the same rule as SEED and CMEED in the scalp map of the DEAP database inside the dashed box.

Our framework can learn the nonlinear relationships between EEG channels regardless of the ordering method. To demonstrate this, we reorder the channel sequence of DEAP using an ordering method similar to SEED (see the DEAP scalp in the dashed box of Fig. 2). Specifically, the ordering method depends on the partition of five brain function regions: Frontal (F), Temporal (T), Parietal (P), Occipital (O) and Central (C) [52]. We then performed experiments on the DEAP database following the subject-independent protocol after electrodes were reordered, and achieved the two binary classification accuracies (mean accuracy and standard deviation (%)) of $68.36/06.14$ for valence, $70.63/07.21$ for arousal. Compared with the results of DEAP with default channel ordering (valence: $69.06/06.37$, arousal: $72.97/06.57$), the accuracies are similar, showing that our ATDD-LSTM model is robust to different EEG electrode orders. More details of experimental setup are given in Section 5.

4.2 Attention-Based Encoder-Decoder

The attention-based encoder-decoder (AT) module consists of an encoder and a decoder. As a specific channel may be informative to identify whether the arousal dimension is calm or active but worthless for recognizing whether the valence dimension is negative or positive. The encoder takes advantage of the attention mechanism to capture the relationship between EEG channels and different emotion classes, and construct a vector representation for the EEG sample. Then, it maps the newly-formed feature to a probability that indicates how likely the sample falls into this emotion class. The decoder mainly combines the outputs of the encoder to reconstruct an EEG representation. By forming an encoder-decoder, this reconstructed vector can optimize the function of the encoder during back propagation.

Specifically, we perform the encoder through

$$v_k^h = \frac{1}{n}\sum_{i=1}^{n} h_{k,i} \tag{3}$$

$$e_{i,c}^k = h_{k,i} w_{a,c} \tag{4}$$

$$a_{i,c}^k = \frac{exp(e_{i,c}^k)}{\sum_{j=1}^{n} exp(e_{j,c}^k)} \tag{5}$$

$$v_{k_c}^a = \sum_{i=1}^{n} h_{k,i} a_{i,c}^k, \tag{6}$$

where $h_{k,i}$ is the feature of the ith channel in the kth EEG sample, and $w_{a,c} \in \mathbb{R}^{d_h}$ is the parameters of cth emotion class for the attention layer. $a_{i,c}^k$ is the attention importance score for each channel, corresponding to the cth emotion class, which is obtained by multiplying the hidden vector matrix with the weight matrix and normalizing it to a probability distribution $a_c^k = [a_{1,c}^k, a_{2,c}^k, \ldots, a_{n,c}^k]$ over the channels (see Section 5.5 in which we visualize them for better understanding). $v_k^h \in \mathbb{R}^{d_h}$ is the average of input feature vectors over channels, and $v_{k_c}^a \in \mathbb{R}^{d_h}$ is a weighted sum over features of all channels, which de-emphasizes the irrelevant channels specific to emotion c by the attention mechanism. Therefore, the attention module helps the network to learn discriminative features to make better prediction for emotion classification.

After we obtain the feature $v_{k_c}^a$, the probability that sample k belongs to the emotion class c can be computed as follows

$$p_{k_c} = \text{sigmoid}(W_c^p \cdot v_{k_c}^a + b_c^p), \tag{7}$$

where W_c^p and b_c^p are learnable parameters corresponding to the cth emotion class. During the decoding stage, the reconstruction representation of an input instance is obtained by multiplying probability p_{k_c} and weighted representation $v_{k_c}^a$, i.e., $r_{k_c} = p_{k_c} v_{k_c}^a$. Consequently, v_k^h and r_{k_c} share the same space and they are further compared to constrain the learning (see more details in Eq. (14) in Section 4.4).

By performing the whole encoder-decoder process m times, one for each emotion class, we can form a probability vector $P_k = [p_{k_1}, p_{k_2}, \ldots, p_{k_m}]$, which is the primary criterion for classification.

4.3 Domain Discriminator

The domain discriminator (DD) aims to judge whether the EEG samples come from source (training) or target (test) domain. Through weakening the classification ability of DD, the feature extractor is updated in the direction of generating more domain-invariant features. As a result, the model is able to solve the feature distribution shift problem.

Specifically, we begin with concatenating the input hidden vector matrix H_k according to the channel dimension to

form vector d_k. Since the overall aim is to maximize DD loss, we apply a gradient reversal layer (GRL) [53] before feeding d_k to a linear transformation and ReLU activation to extract the domain-related feature d_k^l in Eq. (8), where W^l and b^l are the weight matrix and bias vector, respectively.

$$d_k^l = \text{relu}(W^l \cdot d_k + b^l) \tag{8}$$

$$d_k^s = \text{softmax}(W^s \cdot d_k^l + b^s). \tag{9}$$

The function of GRL is to change the gradient sign and pass the gradient backward during back propagation (BP). Therefore, it achieves the reversion of updating direction.

Finally, we obtain the probability of being training samples or test samples by mapping d_k^l to a two-dimensional space and applying softmax function in Eq. (9), where W^s and b^s are the weight matrix and bias vector learned during the training process.

4.4 Training Objective

Denote $X_R = [X_S, X_T]$ to be the entire data with X_S as a set of labeled (training) data and X_T as a set of unlabeled (test) data. Let Y_S be the labels associated with X_S. We refer to $D_S = \{(X_S, Y_S)\}$ as the source (training) domain and $D_T = \{X_T\}$ as the target (test) domain. In order to train the domain discriminator, we build a binary label vector $Y_R^D = [Y_S^D, Y_T^D]$, where elements of Y_S^D are set to 1, and those of Y_T^D are set to 0.

Algorithm 1. Training of ATDD-LSTM

Input:
Training data set X_S and its ground-truth label set Y_S;
Test data set X_T;
Source (training) domain labels Y_S^D, $\forall y_S^D \in Y_S^D, y_S^D = 1$, and target (test) domain labels Y_T^D, $\forall y_T^D \in Y_T^D, y_T^D = 0$;
Learning rate α.

Output:
Parameters: $\hat{\theta}_f$, $\hat{\theta}_a$, $\hat{\theta}_d$.

1: Use X_S and Y_S to update the parameters of attention-based encoder-decoder and feature extractor:
$\theta_a \leftarrow \theta_a - \alpha \frac{\partial L_a}{\partial \theta_a}$, $\theta_f \leftarrow \theta_f - \alpha \frac{\partial L_a}{\partial \theta_f}$;
2: Use X_S, X_T, Y_S^D and Y_T^D to update the parameters of the domain discriminator and feature extractor:
$\theta_d \leftarrow \theta_d - \alpha \frac{\partial L_d}{\partial \theta_d}$, $\theta_f \leftarrow \theta_f - \alpha \frac{\partial L_d}{\partial \theta_f}$;
3: Go to step 1 until convergence, or the algorithm reaches the maximum number of iterations (set to 256 in our experiments);
4: **return**$(\theta_f, \theta_a, \theta_d)$ as trained parameters $(\hat{\theta}_f, \hat{\theta}_a, \hat{\theta}_d)$.

The overall training objective of the ATDD-LSTM model can be formulated as

$$L(X_R; \theta_f, \theta_a, \theta_d) = L_a(X_S; \theta_f, \theta_a) - L_d(X_R; \theta_f, \theta_d), \tag{10}$$

where θ_f, θ_a and θ_d are parameters of the sequential feature extractor, attention-based encoder-decoder and domain discriminator, respectively, and L_a and L_d are loss functions of the attention-based encoder-decoder module and domain discriminator module. We can optimize the objective function by

$$(\hat{\theta}_f, \hat{\theta}_a) = \arg\min_{\theta_f, \theta_a} L_a(X_S; (\theta_f, \theta_a), \hat{\theta}_d) \tag{11}$$

$$\hat{\theta}_d = \arg\max_{\theta_d} L_d(X_R; \hat{\theta}_f, \hat{\theta}_a, \theta_d). \tag{12}$$

The loss function of the attention-based encoder-decoder module L_a aims to minimize the reconstruction error and maximize the probability for the emotion corresponding to the true label, while maximizing the reconstruction error and minimizing the probability for any other emotion class. To achieve these, we adopt the contrastive max-margin loss which has been used in many studies [54], [55]. The loss L_a consists of the probability objective function $J(\theta_f, \theta_a)$ and the reconstruction objective function $U(\theta_f, \theta_a)$.

Given an EEG sample s_k and its true label y_k, the loss functions of the attention-based encoder-decoder module can be formulated as

$$J(\theta_f, \theta_a) = \sum_k \max(0, 1 + \sum_{i=1}^{m} y_{k_i} p_{k_i}) \tag{13}$$

$$U(\theta_f, \theta_a) = \sum_k \max(0, 1 + \sum_{i=1}^{m} y_{k_i} v_k^h r_{k_i}), \tag{14}$$

where y_{k_i} is set to -1 only if i is equal to y_k. Otherwise, it is set to 1. The probability objective function $J(\theta_f, \theta_a)$ guides the model to perform correct classification, and the reconstruction objective function $U(\theta_f, \theta_a)$ ensures that the reconstructed vector $r_{k_{y_k}}$ is similar to the instance representation v_k^h, while making $r_{k_i} (i \neq y_k)$ distinct from v_k^h.

For the loss of domain discriminator module L_d, we adopt cross-entropy as the loss function

$$L_d(X_R; \theta_f, \theta_d) = -\sum_k q_k \log d_k^s, \tag{15}$$

where q_k is the one-hot encoding of domain label in Y_R^D. By maximizing L_d, the feature extractor is optimized to discard the domain-specific components of the input. Therefore, we can decrease the distribution shift between source and target domains.

We train the encoder-decoder and the discriminator iteratively, updating θ_f, θ_a by minimizing L_a and maximizing L_d, and updating θ_d by minimizing L_d. We convert this min-max goal to a minimum loss function $L = L_a + (-L_d)$ using a gradient reversal layer (refer to Fig. 1), which changes the gradients to have an opposite sign during back-propagation. Thus the feature extractor can generate data representations that minimize the loss of encoder-decoder and maximize the loss of the discriminator. The pseudo-code of overall training process is summarized in Algorithm 1.

5 Experiments

In this section, we present experimental results on three EEG databases, i.e., the DEAP database [2], the SJTU Emotion EEG Database (SEED)[21] and CAS Movie-induced Emotion EEG Database (CMEED) [16], [17] which are commonly used for emotion recognition evaluation. We compare our proposed ATDD-LSTM model with four representative methods: one classic machine learning method (SVM), one classic

deep network model called deep belief network (DBN) [21], two state-of-the-art deep network models called graph convolutional neural network (GCNN) [56] and dynamic GCNN [14]. SVM and DBN can only process EEG signals channel by channel, while GCNN and DGCNN can handle multichannel EEG signals. In addition, we also present an ablation study, showing the effectiveness of the feature aggregation scheme in ATDD-LSTM.

5.1 EEG Databases for Emotion Recognition

The DEAP database [2] contains EEG data of 32 participants, which was collected via 32 EEG electrodes from the subjects when they were watching 40 one-minute long excerpts of music videos. Participants rated each video in terms of the levels of arousal, valence, like/dislike, dominance and familiarity ranging from 1 to 9. In our study, the 8 peripheral channels were removed and only EEG signals were used for emotion recognition. The EEG signals were recorded from 32 EEG electrodes according to the international 10-20 system. We segment the valence dimension to positive/negative and the arousal dimension to high/low arousal, both are binary classification problems. To balance the classes, we follow the same threshold and partition strategy in [2]. In DEAP, a bandpass filter with frequency range 4.0-45.0 Hz was applied. We first decompose the EEG signals into 4 frequency bands, including θ band (4-7 Hz), α band (8-13 Hz), β band (14-30 Hz) and γ band (31-45 Hz). Then for every frequency band, we extract DE features from each channel.

The SEED database contains EEG data of 15 Chinese subjects, which was collected via 62 EEG electrodes from the subjects when they were watching 15 Chinese film clips (the duration of each film clip is about 4 minutes). Three emotions (positive, neutral and negative) were elicited. A self-assessment was conducted for each subject after the film clip was watched. One key difference between SEED and DEAP is that for the same subject, SEED contains three sessions at a time interval of one week or longer. Within each session, the same 15 movie clips were watched by every subject. These settings make it possible for us to conduct cross-session experiments, thereby validating the stability and generalizability of our model. In SEED, a bandpass filter with frequency range 1.0-75.0 Hz was applied. We extract the DE features in five frequency bands (δ, θ, α, β and γ bands).

The CMEED database contains EEG data of 37 Chinese subjects, which was collected via 30 EEG electrodes from the subjects when they were watching 16 two-minute Chinese film clips. After watching, participants rated each video in terms of the levels of arousal and valence ranging from 1 to 9. The EEG signals in one trial were recorded from 30 EEG electrodes according to 10-20 system with a sampling rate of 128 Hz. We further segment the valence dimension to positive/negative and the arousal dimension to high/low arousal, both are binary classification problems. To balance the classes, we set the threshold of valence to 4 (i.e., ≥ 4 is positive) and arousal to 6 (i.e., ≥ 6 is high arousal). In CMEED, a bandpass filter with frequency range 1.0-45.0 Hz was applied. In the same way as the SEED database, 5 frequency bands (δ, θ, α, β and γ) are extracted on CMEED. We extract DE features on different frequency bands and channels respectively.

TABLE 1
Mean and Standard Deviation (%) of Accuracies Achieved by SVM, DBN, GCNN, DGCNN and ATDD-LSTM Using 9-vs-6 Subject Dependent Protocol on the SEED Database

Method	Accuracy (mean/std)
SVM	83.99/ 09.72
DBN	86.08/ 08.34
GCNN	87.40/09.20
DGCNN	90.40/ 08.49
ATDD-LSTM	**91.08/06.43**

Experiment Protocols. Based on the three databases, we conduct two kinds of experiments to evaluate the performance of EEG-based emotion recognition: subject-dependent and subject-independent. On the SEED database, an additional cross-session evaluation is also performed.

5.2 Evaluation on SEED Database

The EEG data in the SEED database is associated with three emotional labels, i.e., positive, neutral and negative. Therefore, we perform a three-class classification task for evaluation.

Subject-Dependent Evaluation. We follow the protocol of [14], [21] to evaluate different methods. Specifically, for each subject, there are 15 trials of EEG data in one session. These trials are divided into two parts: the first 9 trials of EEG data are used as the training data, and the remaining 6 trials serve as the test data. After the recognition accuracy corresponding to each subject is obtained, the average classification accuracy and standard deviation over sessions of all 15 subjects are computed. The results are summarized in Table 1, indicating that ATDD-LSTM has the best performance compared with SVM, DBN, GCNN and DGCNN. Note that DGCNN, GCNN and DBN methods can also achieve competitive results and are better than SVM. Moreover, the DGCNN method is specially designed as an improved version of GCNN for multichannel EEG signal analysis [14]. Thereafter, we only compare DBN and DGCNN with our ATDD-LSTM methods.

Cross-Session Evaluation. Compared with the other two databases, a unique characteristic of the SEED database is that it consists of three sessions for each participant (recorded at different times). We use this characteristic to investigate the stability of different recognition methods across sessions. To the best of our knowledge, very few research works perform cross-session experiments on the SEED database. However, this experiment setting is much more challenging and useful for practical applications; in other words, it predicts the emotion of the same subject at different times when the same stimulus is received, and its recognition accuracy shows the stability of different methods over time. To do the cross-session evaluation, we perform leave-one-session-out cross validation for each subject. Specifically, for each subject, its two sessions of EEG data are used as the training data, and the remaining one session serves as the test data. The recognition accuracy corresponding to each subject is obtained by averaging on three-fold cross validation study. Finally, the average recognition accuracy and standard deviation over all

TABLE 2
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Session-Out Cross-Validation Subject-Dependent Protocol on the SEED Database

Method	Accuracy (mean/std)
DBN	56.82/ 10.97
DGCNN	73.06/ 10.36
ATDD-LSTM	**79.26**/12.79

TABLE 3
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Subject-Out Cross-Validation Subject Independent Protocol on the SEED Database

Method	Accuracy (mean/std)
DBN	53.91/ 11.15
DGCNN	79.95/ 09.02
ATDD-LSTM	**90.92**/**01.05**

TABLE 4
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Clip-Out Cross-Validation Subject Dependent Protocol on the DEAP Database

Method	Accuracy	
	Valence (mean/std)	Arousal (mean/std)
DBN	60.69/ 7.20	64.63/ 10.19
DGCNN	86.06/ 02.61	85.61/ 02.44
ATDD-LSTM	**90.91**/12.95	**90.87**/11.32

15 subjects are computed. The results are summarized in Table 2, showing that ATDD-LSTM achieves the best recognition performance.

Subject-Independent Evaluation. We apply the leave-one-subject-out cross validation strategy to evaluate the emotion recognition. In each fold, the EEG data of 14 subjects is used for training and the remaining one subject's EEG data is used as the test data. After repeating 15 folds on 15 subjects, the classification accuracy and standard deviation are computed by averaging on all subjects. The results are presented in Table 3, showing that ATDD-LSTM outperforms DBN and DGCNN.

5.3 Evaluation on DEAP Database

Following the same threshold and partition scheme in [2], we classify the emotion labels on the DEAP database into high/low arousal and positive/negative valence. Then we perform a binary classification task for evaluation.

Subject-Dependent Evaluation. We utilize leave-one-clip-out cross validation for each subject to evaluate different methods. Specifically, for all the 40 trials of one subject, we select 39 trials of EEG data as the training data and leave 1 trial as the test data. The experiments are repeated until the samples of all stimuli for the same subject are used once as the test data. The average classification accuracy and standard deviations of 40×32 experiments are computed as the final performance measure. The results are summarized in Table 4, showing that ATDD-LSTM achieves the best performance.

TABLE 5
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Subject-Out Cross-Validation Subject Independent Protocol on the DEAP Database

Method	Accuracy	
	Valence (mean/std)	Arousal (mean/std)
DBN	56.76/ 07.26	58.98/ 13.60
DGCNN	59.29/ 06.83	61.10/ 12.28
ATDD-LSTM	**69.06**/**06.37**	**72.97**/**06.57**

TABLE 6
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Clip-Out Cross-Validation Subject-Dependent Protocol on the CMEED Database

Method	Accuracy	
	Valence (mean/std)	Arousal (mean/std)
DBN	53.69/ 17.68	71.79/ 17.53
DGCNN	74.47/ 08.46	81.28/ 09.02
ATDD-LSTM	**91.53**/09.00	**91.55**/11.32

TABLE 7
Mean and Standard Deviation (%) of Accuracies Achieved by DBN, DGCNN and ATDD-LSTM Using Leave-One-Subject-Out Cross-Validation Subject-Independent Protocol on the CMEED Database

Method	Accuracy	
	Valence (mean/std)	Arousal (mean/std)
DBN	60.58/ 13.37	56.79/ 25.76
DGCNN	66.86/ 10.24	65.52/ 15.80
ATDD-LSTM	**94.21**/**05.88**	**88.03**/**06.32**

Subject-Independent Evaluation. The same as on SEED, we apply the leave-one-subject-out cross-validation strategy for evaluation. After repeating 32 folds on 32 subjects, the classification accuracy and standard deviation are computed by averaging on all subjects. The results are presented in Table 5, showing that ATDD-LSTM outperforms DBN and DGCNN.

5.4 Evaluation on CMEED Database

Similar to DEAP, we set thresholds to classify the emotion labels on the CMEED to high/low arousal and positive/negative valence. Then we perform a binary classification task for evaluation.

Subject-Dependent Evaluation. The same leave-one-clip-out cross validation was applied as for DEAP. The results are summarized in Table 6, showing that ATDD-LSTM achieves the best performance.

Subject-Independent Evaluation. The same leave-one-subject-out cross validation was applied as for SEED and for DEAP. The results are summarized in Table 7, showing that ATDD-LSTM achieves the best performance.

TABLE 8
Mean and Standard Deviation (%) of Accuracies Achieved by LSTM, ATDD-LSTM, ATDD-LSTM^{freq}, AT-LSTM and DD-LSTM Using Leave-One-Subject-Out Cross-Validation Subject Independent Protocol on SEED Database

Method	Accuracy (mean/std)
LSTM	77.48/ 06.78
AT-LSTM	84.15/ 06.65
DD-LSTM	87.11/ 07.22
ATDD-LSTM^{freq}	88.15/06.04
ATDD-LSTM	**90.92/01.05**

5.5 Ablation Study

We conduct an ablation study on the SEED database to validate the effectiveness of each component in our proposed ATDD-LSTM model. ATDD-LSTM is built upon LSTM with two additional modules: attention-based encoder-decoder and domain discriminator. Specifically, we compare our complete ATDD-LSTM model with three variant models:

- LSTM: only the basic LSTM model without the attention-based encoder-decoder and the domain discriminator;
- AT-LSTM: the LSTM model with only attention-based encoder-decoder;
- DD-LSTM: the LSTM model with only the domain discriminator.

Recall that to capture the relationship among different channels in our proposed ATDD-LSTM model, the LSTM module uses a series with d_x (the size of DE feature vector) variables and n (the number of channels) steps as input. An alternative approach is to use the LSTM block to process a batch of n variables over d_x steps. This instead learns the relationship between different frequencies. We call this alternative approach ATDD-LSTM^{freq}. The results of ablation study are summarized in Table 8, showing that ATDD-LSTM is indeed the best models among these variant models.

Visualization. To better understand how attention mechanism learns to solve emotion recognition tasks, methods to visualize functional aspects of attention mechanism and feature maps can be helpful. The visualization of attention allocation in Fig. 3 shows the qualitative results. In order to analyze the effectiveness of attention mechanism more intuitively, we map the attention weights to the electrode location on scalp. For the binary classification task on the arousal dimension or valence dimension, different channels may not contribute equally. From the scalp map, we observe that the frontal lobe and occipital lobe are correlated with the arousal dimension of emotion, and the parietal lobe and temporal lobe reflect obvious lateral partial phenomenon for the valence dimension of emotion. Moreover, we can also find that the right hemisphere activity is more related to the emotional states on the valence dimension. This visualization result demonstrates the effectiveness of attention mechanism in ATDD-LSTM for capturing emotion-related channels.

We further visualize the feature vectors recorded in the leave-one-subject-out cross-validation experiments on DEAP for demonstrating the effectiveness of attention mechanism. Specifically, v_k^h is a simple average of hidden vectors from LSTM, while v_{kc}^a is a weighted sum of them. To make the visualization clearer, we extract the principal components using a Kernel PCA (KPCA) with a Radial Basis Function (RBF) kernel. The dimension is reduced from 1,024 to 37 for preserving 99 percent of differences between features. As shown in Fig. 4,

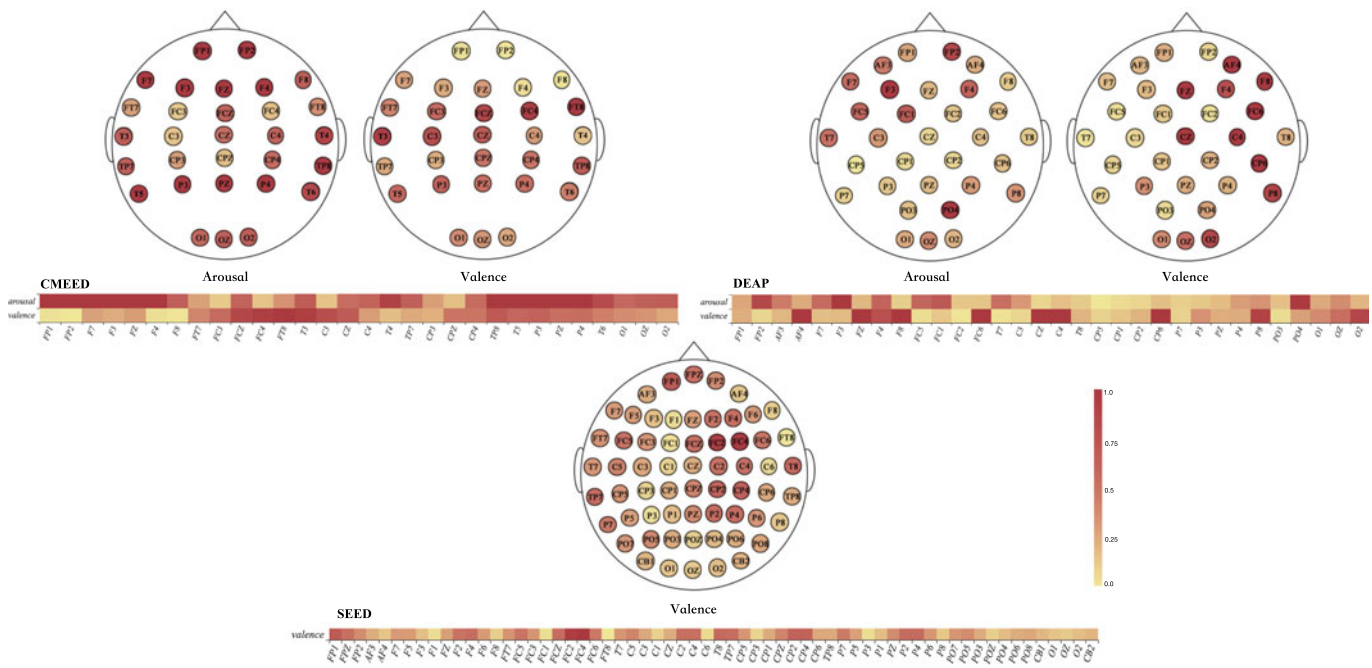


Fig. 3. Visualization of attention allocation. In the heat map, the horizontal axis represents the number of channels and each grid represents the attention weight on this channel, and the vertical axis shows arousal dimension and valence dimension. The scalp map represents the contribution of all electrodes to the binary classification on the arousal dimension or the valence dimension of emotion, with the electrode location corresponding to the grid in the heat map. Notably, since the SEED database is only annotated with the valence dimension of emotion (positive, neutral and negative), only the scalp map associated with the valence dimension is shown.

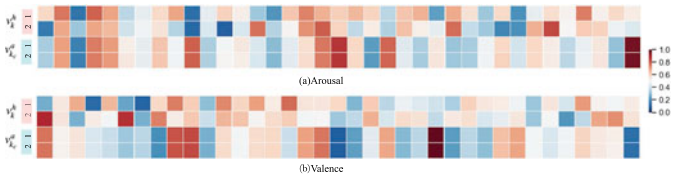


Fig. 4. Visualization of feature vectors v_k^h (without attention layer) and $v_{k_c}^a$ (with attention layer) for two samples on the binary classification task of the arousal dimension and the valence dimension. Each column represents one feature dimension.

each row corresponds to one sample, and each column represents a feature dimension. We show the visualization of feature vectors for two samples corresponding to the arousal dimension and valence dimension of emotion. The ground-truths of two samples are low arousal dimension and low valence dimension (negative), respectively. Obviously, the use of the attention layer makes a difference to $v_{k_c}^a$ by balancing the weight of each channel. Specifically, our framework with the attention layer can extract more consistent feature vectors for predicting the arousal dimension and the valence dimension of emotion. In addition, Fig. 5 further confirms this finding from the overall perspective.

Finally, we visualize the feature distributions in 2-dimensional space using t-SNE to show the effectiveness of domain discriminator in reducing feature distribution shift. In Fig. 6, we present the training feature distribution and test feature distribution from ATDD-LSTM and AT-LSTM in the same 2D space, respectively. We observe that our model with the domain discriminator can ensure the feature distribution coverage is closer and the data representations are invariant, with respect to the training or test domain.

6 Discussion

In this paper, we have proposed the ATDD-LSTM model for EEG-based emotion recognition. Both subject-dependent experiments and subject-independent experiments on DEAP, SEED, and CMEED databases have demonstrated that our model can achieve state-of-the-art results. Furthermore, we used the ablation study results to show that the integration of attention mechanism and domain discriminator in our base model is beneficial to emotion recognition.

For EEG-based emotion recognition, most previous studies have shown that the emotional states can be distinguished by using all the electrodes on the scalp [35]. Based on the

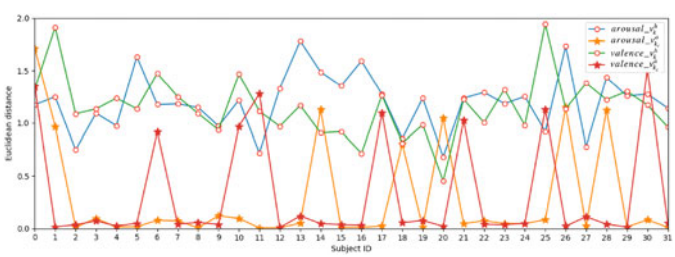
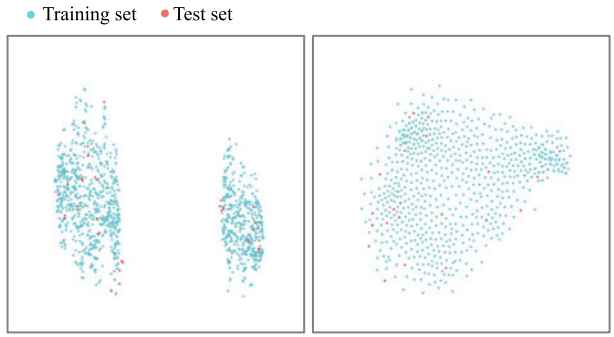


Fig. 5. Quantitative evaluation of the similarity between feature vectors, according to the leave-one-subject-out cross-validation experiments on DEAP. The horizontal axis represents individual subjects, and the vertical axis shows the euclidean distance between v_k^h (or $v_{k_c}^a$) of samples belonging to the same category, which were selected randomly.



(a) AT-LSTM (b) ATDD-LSTM

Fig. 6. Visualization of feature embedding using t-SNE on the training (source) set and test (target) set of DEAP. (a) shows the feature space of using AT-LSTM (the method without the domain discriminator). (b) shows the feature space of using ATDD-LSTM.

previous research of the regions related to emotion recognition in the brain [57], Zheng *et al*. [38] selected six symmetrical temporal electrode locations, excluding the frontal lobe, for emotion recognition. However, the frontal asymmetry has been considered to be correlated with the arousal dimension and the valence dimension of emotion in [58], [59], [60]; therefore, we applied the attention mechanism to select emotion-related electrodes dynamically. The results in our study have demonstrated that the frontal lobe and occipital lobe are correlated with the arousal dimension of emotion, and the parietal lobe and temporal lobe reflect obvious lateral partial phenomenon for the valence dimension of emotion (refer to Fig. 3). According to the ablation study results in this paper, the attention mechanism can focus on selecting electrodes that are useful for emotion recognition.

Due to the noise from the changing environments and the non-stationary characteristics of EEG, the critical factor restricting the establishment of the subject-independent model is the difference in data distribution among subjects. From a practical perspective, we try to embed domain adaptation and feature learning within one training process to optimize the subject-independent model, inspired by the domain adaptation method [53]. According to the cross-session results and subject-independent results in the ablation study, the domain discriminator is able to constrain the feature extractor to obtain domain invariant features.

7 Conclusion

In this paper, we propose an effective attention-based LSTM with domain discriminator (ATDD-LSTM) model, which is a deep neural network model to recognize human emotions from multichannel EEG signals. Aiming at extracting dynamic and domain-invariant features, we design a global domain discriminator for narrowing the distribution difference between training and test domains during the process of sequential feature extraction from LSTM. Using an ablation study, we demonstrate that the attention mechanism can significantly improve the emotion recognition performance. Finally, the subject-dependent and subject-independent cross validation experiments on SEED, DEAP, CMEED databases are conducted and experimental results show that the proposed ATDD-LSTM model achieves the state-of-the-art performance on emotion recognition.

ACKNOWLEDGMENTS

This work was partially supported by the National Key Research and Development Program of China (Grant No. 2016YFB1001200) and the Natural Science Foundation of China (U1736220, 61725204, 61872346).

REFERENCES

[1] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 10, pp. 1175–1191, Oct. 2001.

[2] S. Koelstra *et al.*, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[3] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans. Affective Comput.*, vol. 4, no. 1, pp. 15–33, Jan. 2013.

[4] R. Adolphs, "Recognizing emotion from facial expressions: Psychological and neurological mechanisms," *Behav. Cogn. Neurosci. Rev.*, vol. 1, no. 1, pp. 21–62, 2002.

[5] M. Y. V. Bekkedal, R. John, and P. Jaak, "Human brain EEG indices of emotions: Delineating responses to affective vocalizations by measuring frontal theta event-related synchronization," *Neurosci. Biobehav. Rev.*, vol. 35, no. 9, pp. 1959–1970, 2011.

[6] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.

[7] K. Sharma, C. Castellini, F. Stulp, and E. L. Van den Broek, "Continuous, real-time emotion annotation: A novel joystick-based analysis framework," *IEEE Trans. Affective Comput.*, vol. 11, no. 1, pp. 78–84, Firstquarter 2020.

[8] Y. M. Chi, Y. T. Wang, Y. Wang, C. Maier, T. P. Jung, and G. Cauwenberghs, "Dry and noncontact EEG sensors for mobile brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 2, pp. 228–235, Mar. 2012.

[9] Y.-J. Huang, C.-Y. Wu, A. M.-K. Wong, and B.-S. Lin, "Novel active comb-shaped dry electrode for EEG measurement in hairy site," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 256–263, Jan. 2015.

[10] X. Li, B. Hu, S. Sun, and H. Cai, "EEG-based mild depressive detection using feature selection methods and classifiers," *Comput. Methods Prog. Biomed.*, vol. 136, no. C, pp. 151–161, 2016.

[11] Y. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from EEG signals," *IEEE Trans. Affective Comput.*, vol. 9, no. 4, pp. 550–562, Fourth-quarter 2018.

[12] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2016, pp. 352–359.

[13] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017.

[14] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affective Comput.*, to be published, doi: 10.1109/TAFFC.2018.2817622.

[15] J. E. Harmon, "Contributions from research on anger and cognitive dissonance to understanding the motivational functions of asymmetrical frontal brain activity," *Biol. Psychol.*, vol. 67, no. 1, pp. 51–76, 2004.

[16] G. Zhao, Y. Zhang, and Y. Ge, "Frontal EEG asymmetry and middle line power difference in discrete emotions," *Front. Behav. Neurosci.*, vol. 12, 2018, Art. no. 225.

[17] G. Zhao, Y. Zhang, Y. Ge, Y. Zheng, X. Sun, and K. Zhang, "Asymmetric hemisphere activation in tenderness: Evidence from EEG signals," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 8029.

[18] M. K. Ahirwal and M. R. Kose, "Audio-visual stimulation based emotion classification by correlated EEG channels," *Health Technol.*, vol. 10, no. 1, pp. 7–23, 2020.

[19] L. Piho and T. Tjahjadi, "A mutual information based adaptive windowing of informative EEG for emotion recognition," *IEEE Trans. Affective Comput.*, to be published, doi: 10.1109/TAFFC.2018.2840973.

[20] W. Zheng, J. Zhu, and B. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 417–429, Third Quarter 2019.

[21] W. Zheng and B. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[22] W. Zheng, Y. Zhang, J. Zhu, and B. Lu, "Transfer components between subjects for eeg-based emotion recognition," in *Proc. Int. Conf. Affect. Comput. Intell. Interaction*, 2015, pp. 917–922.

[23] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4746–4752.

[24] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[25] 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 *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 1561–1567.

[26] P. Pandey and K. Seeja, "Subject-independent emotion detection from EEG signals using deep neural network," in *Proc. Int. Conf. Innovative Comput. Commun.*, 2019, pp. 41–46.

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

[28] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[29] D. Keltner and P. Ekman, "The psychophysiology of emotion," in *Handbook of Emotions*, New York, NY, USA: Guilford Publications, 2000, pp. 236–249.

[30] H. Schlosberg, "Three dimensions of emotion," *Psychol. Rev.*, vol. 61, no. 2, pp. 81–88, 1954.

[31] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[32] H. Bo, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[33] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from EEG," in *Proc. Trans. Comput. Sci.* 18th, 2013, pp. 101–120.

[34] P. Petrantonakis and L. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–97, Mar. 2010.

[35] D. Nie, X. Wang, L. Shi, and B. Lu, "EEG-based emotion recognition during watching movies," in *Proc. Int. IEEE/EMBS Conf. Neural Eng.*, 2011, pp. 667–670.

[36] L. Shi, Y. Jiao, and B. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2013, pp. 6627–6630.

[37] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Third Quarter 2019.

[38] W. Zheng, W. Liu, Y. Lu, B. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

[39] H. Tang, W. Liu, W. Zheng, and B. Lu, "Multimodal emotion recognition using deep neural networks," in *Proc. 24th Int. Conf. Neural Inf. Process.*, 2017, pp. 811–819.

[40] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, 2017, Art. no. 016003.

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

[42] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.

[43] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2204–2212.

[44] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[45] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2015, pp. 1412–1421.

[46] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2016, pp. 606–615.

[47] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *The J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.
[48] J. Hoffman *et al.*, "CyCADA: Cycle-consistent adversarial domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1989–1998.
[49] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2962–2971.
[50] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 1647–1657.
[51] Y. Luo, S. Zhang, W. Zheng, and B. Lu, "WGAN domain adaptation for EEG-based emotion recognition," in *Proc. 25th Int. Conf. Neural Inf. Process.*, 2018, pp. 275–286.
[52] R. Adolphs, "Neural systems for recognizing emotion," *Curr. Opinion Neurobiol.*, vol. 12, no. 2, pp. 169–177, 2002.
[53] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proc. 32nd Int. Conf. Mach. Learn., Res.*, F. Bach and D. Blei, Eds., Lille, France: PMLR, vol. 37, 2015, pp. 1180–1189.
[54] M. Iyyer, A. Guha, S. Chaturvedi, J. L. Boyd-Graber, and H. D. III, "Feuding families and former friends: Unsupervised learning for dynamic fictional relationships," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol.*, 2016, pp. 1534–1544.
[55] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An unsupervised neural attention model for aspect extraction," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguist.*, 2017, pp. 388–397.
[56] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3837–3845.
[57] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
[58] R. E. Wheeler, R. J. Davidson, and A. J. Tomarken, "Frontal brain asymmetry and emotional reactivity: A biological substrate of affective style," *Psychophysiology*, vol. 30, no. 1, pp. 82–89, 1993.
[59] J. A. Coan and J. J. Allen, "Frontal EEG asymmetry as a moderator and mediator of emotion," *Biol. Psychol.*, vol. 67, no. 1/2, pp. 7–50, 2004.
[60] B. D. Poole and P. A. Gable, "Affective motivational direction drives asymmetric frontal hemisphere activation," *Exp. Brain Res.*, vol. 232, no. 7, pp. 2121–2130, 2014.

Xiaobing Du received the bachelor's degree from the School of software, Shandong University, Shandong, China, in 2016. She is currently working toward the PhD degree in the State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China. Her research interests include affective computing and human-computer interaction.

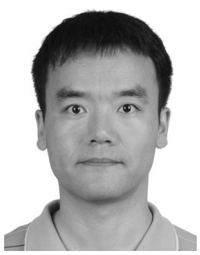
Cuixia Ma received the BS and MS degrees from Shandong University, China, in 1997 and 2000, respectively, and the PhD degree from the Institute of Software, Chinese Academy of Sciences, China, in 2003. She was a research associate with the Department of Computer Science, Naval Postgraduate School, Monterey, California, from 2005 to 2006. She is currently a professor with the Institute of Software, Chinese Academy of Sciences, China. Her research interests include human computer interaction and multimedia computing.

Guanhua Zhang received the BEng degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2017. She is currently working toward the master's degree with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. Her research interests include cognitive computing and human-computer interaction.

Jinyao Li received the bachelor's degree from the College of Information Science and Technology, Beijing Normal University, Beijing, China, in 2019. She is currently working toward the master's degree at the University of Chinese Academy of Sciences, Beijing, China. Her research interests include affective computing and human-computer interaction.

Yu-Kun Lai received the bachelor's and PhD degrees in computer science from Tsinghua University, China, in 2003 and 2008, respectively. He is currently a reader of visual computing with the School of Computer Science and Informatics, Cardiff University, United Kingdom. His research interests include computer graphics, geometry processing, image processing and computer vision. He is on the editorial boards of the *Computer Graphics Forum* and the *The Visual Computer*.

Guozhen Zhao received the BS degree in industrial engineering from Tianjin University, Tianjin, China, in 2007, and the MS and PhD degrees in industrial and systems engineering from the State University of New York, Buffalo, New York, in 2009 and 2011, respectively. Since 2012, he is an associate professor with the Institute of Psychology, Chinese Academy of Sciences, Beijing, China. His current research interests include human cognition and performance, human computer interaction, and neuroergonomics and their applications in intelligent system design.



Xiaoming Deng received the BS and MS degrees from Wuhan University, China, in 2001 and 2004, respectively, and the PhD degree from the Institute of Automation, Chinese Academy of Sciences (CAS), China, in 2008. After a two-year postdoctoral at the Institute of Computing Technology, Chinese Academy of Sciences, China, he joined the Institute of Software, Chinese Academy of Sciences, China, in 2010, where he is currently an associate professor. He was a research fellow with the National University of Singapore, Singapore from 2012 to 2013. His main research interests are in computer vision, and specifically related to 3D reconstruction, human motion tracking and synthesis, and deep learning.

Yong-Jin Liu (Senior Member, IEEE) received the BEng degree from Tianjin University, Tianjin, China, in 1998, and the MPhil and PhD degrees from the Hong Kong University of Science and Technology, Hong Kong, China, in 2000 and 2004, respectively. He is currently a professor with the Department of Computer Science and Technology, Tsinghua University, Beijing, China. His research interests include computational geometry, computer vision, cognitive computation, and pattern analysis.

Hongan Wang (Senior Member, IEEE) received the PhD degree from the Institute of Software, Chinese Academy of Sciences, Beijing, China, in 1999. He is a professor with the Institute of Software, Chinese Academy of Sciences, China. He is currently the director of Intelligence Engineering Laboratory. His research interests include human-computer interaction, real-time intelligence, and real-time active database.