A Channel-Fused Dense Convolutional Network for EEG-Based Emotion Recognition

Zhongke Gao, *Senior Member, IEEE*, Xinmin Wang, Yuxuan Yang, Yanli Li, Kai Ma, and Guanrong Chen, *Life Fellow, IEEE*

***Abstract*—Human emotion recognition could greatly contribute to human–computer interaction with promising applications in artificial intelligence. One of the challenges in recognition tasks is learning effective representations with stable performances from electroencephalogram (EEG) signals. In this article, we propose a novel deep-learning framework, named channel-fused dense convolutional network, for EEG-based emotion recognition. First, we use a 1-D convolution layer to receive weighted combinations of contextual features along the temporal dimension from EEG signals. Next, inspired by state-of-the-art object classification techniques, we employ 1-D dense structures to capture electrode correlations along the spatial dimension. The developed algorithm is capable of handling temporal dependencies and electrode correlations with the effective feature extraction from noisy EEG signals. Finally, we perform extensive experiments based on two popular EEG emotion datasets. Results indicate that our framework achieves prominent average accuracies of 90.63% and 92.58% on the SEED and DEAP datasets, respectively, which both receive better performances than most of the compared studies. The novel model provides an interpretable solution with excellent generalization capacity for broader EEG-based classification tasks.**

***Index Terms*—Brain–computer interface (BCI), convolutional neural network (CNN), deep learning (DL), electroencephalogram (EEG), emotion recognition.**

I. Introduction

HUMAN emotion has a great impact on our daily activities, which is concerned with various actions, such as relaxation, work, and entertainment. There are increasing research interests in the relationship between emotions and physiological functions [1]. It has been found that positive emotions could reflect pleasurable engagement, and are beneficial for human health and attitude [2]. However, accompanied with complaints of physical symptoms, negative emotions may adversely influence mental health and even cause serious psychological problems [3]. As the information explosion through social channels, it is quite challenging to reveal one's emotional clues. Recently, affective computing (AC) has emerged under the demand of deep knowledge and reasonable utilization for emotion [4], [5]. It is a promising area of research that has attracted increasing attentions from numerous cross-curricular fields, ranging from neuroscience to computer engineering. Emotion would be subtly influenced by multiple external and psychological factors, and is a combination of time, space, experience, and cultural background [6]. This aggravates the difficulties for emotion recognition research. Although great efforts have been made to explore the mechanisms and methods for emotion recognition [7], due to the intricate external patterns, effective emotion recognition methods are still in high demand for many technological applications.

There are several signal clues that are recorded to evaluate emotional states, such as facial expressions [8], speech signals [9], text messages [10], and physiological indexes like electrocardiogram (ECG) [11], electroencephalogram (EEG) [12]–[14], and dermal resistance [15]. The detection approaches, which merge facial expressions, speech signals, text messages, and other nonphysiological sources together, are fairly economical. However, these clues are varying from different human living habits and cultural backgrounds, thus not reliable to a certain degree. During collecting facial expressions, face images should be taken with high quality in good conditions. One may deliberately conceal the true feelings to trick the camera and the computer. Though these problems do not exist in some physiological indexes like dermal resistance, they could be affected by such factors as humidity and temperature, therefore they do not operate satisfactorily.

Among these clues, EEG is the most preferred source in emotion recognition research due to its good temporal resolution and information richness. Compared with other sources, EEG signal is more effective and authentic with the property of a strong unforgeability. Many studies have proved the correlations between emotional states and EEG signals in different brain regions [16]. Moreover, with the development and popularization of wearable EEG devices and dry electrode techniques [17], [18], EEG online systems can be easily implemented and are promising for practical applications in various tasks, such as sleep stage scoring [19], disease detection [20], and driver fatigue evaluation [21]. Therefore, EEG signal is

Manuscript received January 10, 2020; revised February 14, 2020; accepted February 20, 2020. Date of publication February 25, 2020; date of current version December 10, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61922062 and Grant 61873181, and in part by the Hong Kong Research Grants Council through GRF under Grant CityU-11200317. *(Corresponding author: Zhongke Gao.)*

Zhongke Gao, Xinmin Wang, Yuxuan Yang, and Yanli Li are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: zhongkegao@tju.edu.cn).

Kai Ma is with the Tencent Jarvis Lab, Shenzhen 518057, China (e-mail: kylekma@tencent.com).

Guanrong Chen is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: eegchen@cityu.edu.hk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCDS.2020.2976112.

Digital Object Identifier 10.1109/TCDS.2020.2976112

2379-8920 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

selected as the source for our emotion recognition study in this article.

It is noted that the recorded EEG signals are inevitably mixed with noises due to their low signal-to-noise ratios (SNRs), which makes it quite challenging to design computational algorithms for recognizing emotions. Actually, EEG signals are time-varying data, and recorded from multiple electrodes that are arranged with the standard 10–20 system [22]. These task-related signals contain the rich spatial–temporal information, which can reflect electrode correlations across the spatial dimension and contextual dependencies across the temporal dimension, respectively. The effective extraction of spatial–temporal information can help better recognize the emotions. Various methods have been developed to handle the temporal dependencies from EEG signals, including time-frequency analysis [23], complex network methods [24]–[27], and nonlinear analysis [28], [29]. Deserved to be mentioned, differential entropy (DE) and power spectral density (PSD) features have been proven valid for emotion recognition [12], [30], [31]. Meanwhile, principal component analysis (PCA) [32] and Fisher transform [33] are commonly used for feature selection and optimization. However, most of these feature-based methods mainly focus on extracting temporal features across a single channel while neglecting the information of electrode correlations. Additionally, the extraction of some features is quite time consuming, especially for nonlinear analysis, which cannot meet most demands online.

In recent years, deep-learning (DL) techniques have been developed rapidly, drawing a great deal of attentions in diverse research fields. Various network architectures have been proposed, such as deep belief networks (DBNs) [34], convolutional neural networks (CNNs) [35], and recurrent neural networks (RNNs) [36]. These models are superior in computational efficiency and model performance, and have exerted prodigious impacts in many fields, such as image classification [37], speech recognition [38], and time-series prediction [39], [40]. Note that EEG signals are of great importance in the extraction of spatial correlations and temporal dependencies, which slightly differs from 2-D images and speech signals. There are considerable explorations on EEG-based classification tasks, including motor imagery classification [41]–[43], fatigue driving evaluation [44], [45] and emotion recognition [46]–[48]. There is a detailed survey about CNNs in EEG analysis [49]. These works greatly enrich the explorations of DL-based EEG signal analyses. However, some challenges still remain to be solved. Most of the frameworks for EEG analysis are shallow networks, simply containing convolutional layers, pooling layers, and fully connected layers. This results in lacking the capability of nonlinear approximation, which may correspond to uncompetitive effects.

To address the above problems, we propose a novel deep architecture in this article, named channel-fused dense convolutional network (CDCN), for EEG-based emotion recognition tasks. This article has two contributions as follows.

1) We propose a CDCN framework to deal with contextual information and spatial correlations for emotion recognition, where the 1-D convolution and dense connections are integrated to enhance the performance. This method guarantees better performances compared with others competitive recognition methods for EEG-based emotion recognition on the SEED and DEAP datasets.
2) CDCN could elegantly handle the problems that shallow networks have weak capacity to deal with, where 2-D convolution requires more information of temporal dependencies and electrode correlations. In CDCN, the initial 1-D convolution layer plays the role of temporal feature selections, which outperforms other compared methods regarding both efficiency and accuracy, and could extend the application to broader EEG-based classification tasks.

To study the emotion recognition problem, the layout of this article is organized as follows. Section II briefly introduces the related work in emotion feature methods and DL techniques applied in EEG analyses. Section III presents the developed framework on emotion recognition tasks. Section IV reports the experimental results evaluated on two popular emotion datasets. Section V discusses the possibility of improvements for emotion recognition tasks. Section VI presents the conclusions.

II. Related Work

A. Emotion Features

EEG signal has been one of the most popular signal forms in emotion recognition tasks, which has attracted lots of attention and many feature methods have explored. The common features can be mainly associated with the following three categories: 1) time-domain features; 2) frequency-domain features; and 3) functional connectivity features [50].

Time-domain features aim to extract the temporal information through EEG signals, such as statistical features and fractal dimension features. The work [51] extracted statistical features, fractal dimension features, and other features of EEG signals as the inputs of support vector machine (SVM) for binary valence-arousal (VA) recognition on the DEAP dataset, which reached an average accuracy of 73.10%. Frequency-domain features mainly capture the spectral information from EEG signals, such as band power and DE [52], [53]. In [54], DE features were extracted as the inputs of a graph regularized extreme learning machine classifier, which received a mean accuracy of 91.07% on the SEED dataset. In particular, the DE feature is receiving increasing attention in emotion analysis tasks. Then, functional connectivity features focus on the correlation and synchronization information between sensors pairs, which serves as the spatial information of EEG signals. Moreover, there is a detailed survey of emotion recognition methods [55], which reviewed the main aspects involved in the recognition process, including subjects, studied features, and popular classifiers.

B. Deep Learning for EEG Analysis

Many methods have been proposed to investigate proper computational models for emotion recognition using EEG signals. Various DL architectures have been applied in classifying EEG signals to solve different recognition tasks. Typically,

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

most of the existing DL-based EEG studies can be summarized into two categories. The first one is based on using EEG signals as the input of a network. The second one is based on using the extracted features from EEG signals as the input of a network.

In EEG analysis tasks, it requires the developed model able to capture internal information from EEG signals. In [56], a compact fully convolutional network EEGNet was proposed to process EEG signals, which performed effectively in four different brain–computer interface (BCI) classification tasks. In our previous work [45], a spatial–temporal CNN was developed to deal with EEG signals in the fatigue detection task. RNNs are capable of extracting temporal information in EEG analyses. In [57], the model was based on a convolutional long short-term memory and a new temporal margin-based loss function, which achieved overall accuracies of 78.72% and 79.03% for recognizing valence and arousal emotions, respectively, on the DEAP dataset. A long short-term memory architecture was developed in [58] for cognitive workload estimation with accuracy up to 93%. These studies have proved that DL methods can learn effective representations from EEG signals.

Another direction toward better performance is to combine analysis with prior knowledge. In [59], EEG signals were transformed into multispectral images and a recurrent convolutional network was trained for mental load evaluation. EEG sequences were converted into 2-D graph matrixes with spectral filtering and dynamical graph CNNs were introduced for EEG emotion recognition in [60]. A hierarchical CNN was trained in [61] with 2-D maps organized from DE features in all channels, which was found efficient in emotion recognition tasks. A spatial–temporal RNN was proposed to integrate the feature learning from both spatial and temporal dimensions of signal sequences in [62], where the final accuracy for emotion recognition reached 89.5% on the SEED dataset. The above studies combined with prior knowledge to provide a good approach, which allows building specific frameworks for EEG-based classification tasks.

Although DL methods have achieved great progresses in EEG recognition tasks, there are still many challenges to be solved. For instance, the existing feature-based DL methods may attach little importance on the information of electrode correlations. To solve these problems, we extract the DE features from EEG signals, and then convert them into 2D matrixes as the input of the CDCN model, which allows effective processing of temporal dependencies and electrode correlations with series-specific modifications for EEG-based emotion recognition.

III. Methods

In this section, we first introduce the DE feature and use it as the input of the CDCN model, then present the developed framework with the model architecture and implementation details.

A. DE Feature

Various methods have been used to extract vital features from EEG signals. In the works reviewed, DE features have showed great abilities to measure the complexity of continuous random variables in emotion recognition tasks [31]. The calculation formula of DE feature is defined as

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

where X follows the Gaussian distribution $N(\mu, \sigma^2)$, π, and e are constants, and x is a variable. DE feature has been proven to be equivalent to the logarithmic PSD for fixed-length EEG segments in a certain frequency band. So, we extract the DE features in five main frequency bands (delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, and gamma: 31–50 Hz) for each channel. If the EEG signal has E channels, each feature matrix has the dimensions of [E, 5], which serves as the input of the CDCN model.

B. Model Architecture

DenseNet was proposed by Huang *et al.* [63], and compared with traditional CNNs, it has several compelling advantages: strengthen the feature propagation, alleviate the vanishing gradient problem, encourage feature reuse, and substantially reduce the number of parameters. However, DenseNet and its extensions are specially designed for 2-D image recognition tasks. Their structures for image processing actually contain 2-D convolutions and pooling operations, which are improper for multichannel EEG processing. Imposing the patterns in applications, it may lead to the information loss of electrode correlations and temporal dependencies in EEG signals. Taking the above issues into consideration, the proposed CDCN framework uses the 1-D convolution to deal with temporal information and the modified dense blocks to emphasize the importance of electrode correlations. Fig. 1 shows the pipeline of the proposed method for EEG-based emotion recognition.

Let matrix $X \in R^{E\times F}$ denote input samples with matching labels, where each signal sample has E electrodes and each electrode has F features, and C denotes the categories of emotional states. On the SEED dataset, the dimension of the input samples is 62-by-5; on the DEAP dataset, the dimension of the input samples is 32-by-5. The proposed CDCN contains convolutional layers, pooling layers, dense blocks, and transition blocks.

The dense block is used to improve the information flow and strengthen the feature reuse. It introduces direct connections from any layer to the subsequent layers, which is adopted to extract high-level features through the feature maps from the previous layer. Let x_l be the lth layer's output and the $(l+1)$th layer's input in the traditional neural network, which can be expressed as

$$x_l = H_l(x_{l-1}) \quad (2)$$

where H is the nonlinear transformation in the lth layer. Hence, for the lth layer in the dense block, it receives the feature maps from all preceding layers. Thus

$$x_l = H_l\big([x_0, x_1, \ldots, x_{l-1}]\big) \quad (3)$$

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

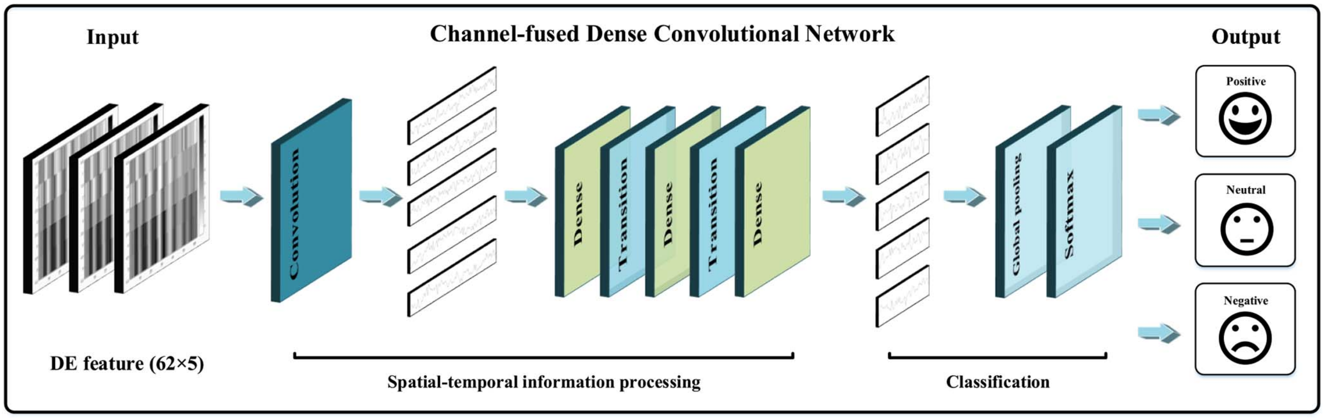

Fig. 1. Pipeline of the proposed method for EEG-based emotion recognition.

TABLE I
DETAILS OF THE DEVELOPED CDCN. THE SECOND COLUMN DENOTES THE OUTPUT SIZE OF THE CURRENT LAYER. THE SYMBOL [] CONTAINS THE KERNEL SIZE, THE NUMBER OF FEATURE MAPS AND THE TYPE OF LAYER, RESPECTIVELY

Layers	Output size	CDCN (k=12)
Input	62 × 5	–
Convolution	62 × 1	[1×5, map 24, stride 1, conv]
Dense block 1	62 × 1	[3×1, map 36: 96, conv] ×6
Transition block 1	62 × 1	[1×1, map 96, conv]
	31 × 1	[2×1, stride 2, maxpooling]
Dense block 2	31 × 1	[3×1, map 108:168, conv] ×6
Transition block 2	31 × 1	[1×1, map 168, conv]
	16 × 1	[2×1, stride 2, maxpooling]
Dense block 3	16 × 1	[3×1, map 180:240, conv] ×6
Classification	240	global average pooling
	3	softmax

where $[x_0, x_1, \ldots, x_{l-1}]$ denotes the cascade operation of $(l-1)$ layers' feature maps.

The transition block is employed to reduce the size of input feature maps, which needs some modifications to work for EEG signals. For practical use, the transition block consists of a batch normalization layer, a convolutional layer, and a max-pooling layer.

In our experiment, the depth of the CDCN is selected to 24 layers with details presented in Table I. The input dimensions and the number of emotions mentioned here corresponds to the SEED dataset. The grow rate k and the number of dense blocks are set to 12 and 3, respectively. Each dense block equally contains six convolution blocks, which orderly consists of a batch normalization [64], a rectified linear activation (ReLU) [65], and a 3×1 convolution with zero-padding to keep the size of feature maps fixed. In this case, the transition block is composed of a 1×1 convolution and a 2×1 max-pooling with stride 2.

In the framework, the $1 \times F$ convolution layer is initially performed as the input layer with twice the grow rate k feature maps, where N denotes the number of features per electrode. Three dense blocks are joined together. Meanwhile, two transition blocks are inserted between two contiguous dense blocks. A global average pooling is employed, followed by the last dense block, and then a Softmax classifier is appended. In the three dense blocks, the sizes of feature maps are 62×1, 31×1, and 16×1, respectively. If each convolution in the dense block produces k feature maps, the l th layer has $k_0 + k \times l$ output feature maps, where k_0 denotes the number of the layer's feature maps before the dense block. For the convolution layer in each transition block, the number of feature maps is the same as the previous convolution layer.

C. Implementation Details

The backpropagation through time (BPTT) algorithm is used to optimize the network parameters until receiving the optimal solutions or reaching the maximum of epochs. Besides, the cross-entropy objective function is employed as the loss function for model optimization [66], which is expressed as

$$\text{Loss} = \text{cross\_entropy}\big(y, y^p\big) \tag{4}$$

where y and y^p denote the ground truth of train samples and the predicted ones, respectively. During the training process, the function *Loss* is aimed at reducing the loss value to improve the prediction accuracy of the model. We train the CDCN framework using the Adam algorithm [67] with a learning rate of 10^{-3}. The batch size is set to 64. The ground truth of the validation set is coded by one-hot states. The model is implemented with the Keras library,[1] which is extended from Google Tensorflow.[2] Then, we save the best model by monitoring on the validation set, and the elapsed time for model training with a GeForce Titan X for about twenty minutes.

IV. EXPERIMENTS

In this section, we first introduce the SEED dataset and provide a detailed explanation of dataset preprocessing, which is used for evaluating the performance of the proposed method. Then, we analyze the model performance on the SEED dataset and show performance comparisons with other competitive methods used previously for similar researches. Finally,

[1] https://keras.io/
[2] https://www.tensorflow.org/

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

TABLE II
PERFORMANCE ON DIFFERENT FREQUENCY BANDS

Frequency band	Delta	Theta	Alpha	Beta	Gamma	All (%)
CDCN	65.19/6.83	69.84/8.37	72.16/8.49	80.83/7.64	82.63/8.01	**90.63/4.34**

we conduct experiments to evaluate the performance of the developed CDCN model on the DEAP dataset with method comparisons.

A. SEED Dataset and Preprocessing

The SEED dataset,[3] contributed by Duan *et al.* [31] and Zheng and Lu [68], focuses on EEG-based emotion recognition tasks. Fifteen emotion-evoking film clips with audios and scenes chosen as stimulus materials from the SEED dataset, could help collect high-quality signals. Three categories of emotions (positive, neutral and negative) were considered in the experiments. Each of the clips lasted for about four minutes and each emotion corresponds to five film clips.

Fifteen volunteering undergrads were selected to perform experiments with evaluation pasted on the Eysenck personality questionnaire (EPQ), which could assess one's personality traits [69]. Before each experiment, the subjects were advised to follow the procedure and refrained from unnecessary body movements. Each scalp EEG signal was collected by a 62-channel recording cap (ESI Neuroscan) and downsampled to the 200-Hz sampling rate. All electrodes were arranged in accordance with the standard 10–20 system [22]. The EEG signals were processed with a band-pass filter of 0.3–50 Hz to remove physiological and power frequency noises. Besides, face videos were recorded by a frontal camera. In each trial, the orders of executions were a 5 s tip, a film clip, a 45 s self-evaluation, and a 15 s rest. Each subject completed 15 trials for three sessions with an interval of one week or more between two sessions.

For each subject, the time durations are equal and fixed. We extract the EEG samples according to the duration of each movie, and then divide each channel of the EEG signal into the same-length segments of 1s without overlapping. The numbers of samples in three categories remain 1120, 1054, and 1070, respectively. Then, we compute the DE features on each EEG sample. One-hot state codes the sample representations for three categories. We take the same experimental protocol in [62], [68], and [70] to evaluate the performance of the CDCN model on the SEED dataset. For all the 15 trials from each session, the first nine trials are taken as the training set and the remaining six trials are taken as the testing set. Note that, the training data and the testing data are from different trials of the same session. The final recognition accuracy is averaged with the recognition accuracy of all the 15 subjects.

B. Overall Performance

The task can be regarded as a three-classification problem and CDCN is trained for EEG-based emotion recognition by the dataset. In our experiments, the input sample is a 2-D feature matrix extracted from a 1s segment and the label prediction is exported for model evaluation. Besides, the individual performance is obtained by averaging the effects of three sessions. Fig. 2 presents the recognition accuracies of the CDCN framework for each subject using the SEED dataset.

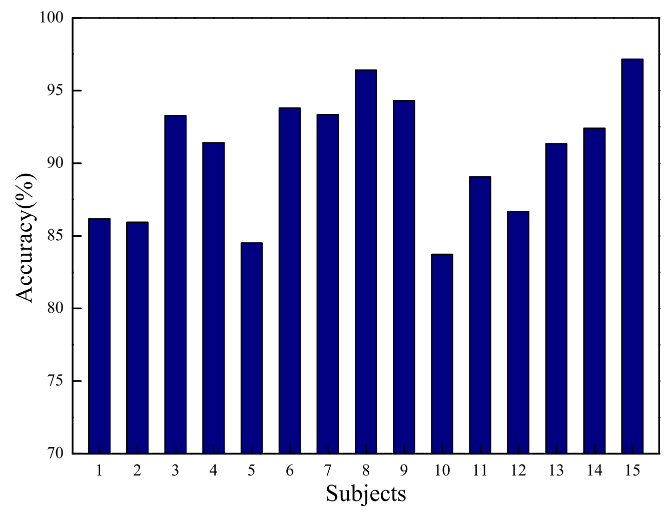

Fig. 2. Performances of the CDCN framework using the SEED dataset.

From Fig. 2, we find that the CDCN algorithm for within subject is stably effective on the SEED dataset and all surpass 84%. The mean classification accuracy achieves 90.63% with a standard deviation 4.34%. The performance of eight subjects are over the mean accuracy, while seven are below it. Definite individual variations indeed exist, possibly caused by experimental situations and the subjects' physical conditions.

To reveal the relationship between different frequency bands and emotion states, the performance of the CDCN framework on different frequency bands (delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, and gamma: 31–50 Hz) are calculated using the SEED dataset, which is shown in Table II. Note that *All* in Table II denotes the performance uses the features of all the five frequency bands.

As observed in Table II, the accuracies of Beta and Gamma bands achieved 80% while the accuracies of three frequency bands (Delta, Theta, and Alpha) were lower than 80%. These results suggest that Beta and Gamma bands of brain activity are more related to emotional states than other frequency bands. The performance on All column works much better than these on five separate frequency bands. It shows that the features from all the five frequency bands contribute to the performance of the CDCN model on the SEED dataset, which indicates that the information related to emotion recognition tasks is distributed in different frequency bands of EEG signals.

[3]http://bcmi.sjtu.edu.cn/home/seed/

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

TABLE III
PERFORMANCE COMPARISONS ON THE SEED DATASET

Method	Description	Data type	Subject number	Accuracy (%)
SyncNet [71]	Convolutional neural network with Gaussian Process adapter	EEG	15	77.9
GSCCA [72]	Group sparse canonical correlation analysis with frequency features	EEG	15	83.72
DBN [68]	Deep belief network with DE features	EEG	15	86.08
HCNN [61]	Hierarchical convolutional neural network with DE features	EEG	4	86.2
SVM [68]	Support vector machine with DE features from 12 channels	EEG	15	86.65
MNN [73]	Minimalist neural network with reinforced gradient coefficients from 12 channels	EEG	15	88.23
STRNN [62]	Two-layer recurrent neural network with DE features	EEG	15	89.50
BDAE [70]	Bimodal deep autoencoder with DE features	EEG + EOG	9	91.01
GELM [54]	Graph regularized extreme learning machine with DE features	EEG	15	91.07
CDCN	Channel-fused dense convolutional network with DE features	EEG	15	**90.63**

The confusion matrix of the CDCN framework evaluated on the SEED dataset is presented in Fig. 3, which shows the classification accuracy of each emotion. The value (i, j) denotes the percentage of samples in class i that is classified as class j. As shown in Fig. 3, our method shows a good performance in recognizing all three types of emotions and the accuracies of them exceed 82%. Positive emotion is recognized with higher accuracies, while negative emotion is relatively difficult to recognize as it is partly confused with the neutral.

These results reflect that our CDCN model provides an effective relationship between emotional states and EEG signals. But, from another perspective, we suggest taking the factors of interclass variations into consideration to establish a more robust emotion recognition framework, which could be our future research work.

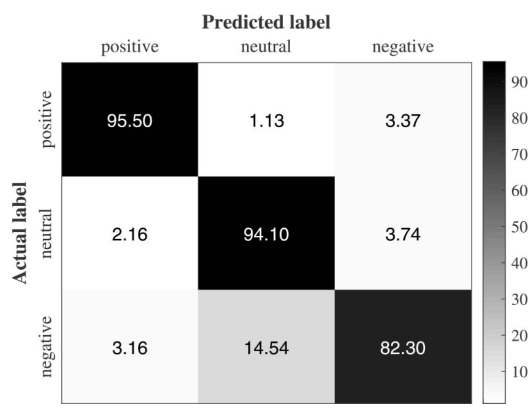

Fig. 3. Confusion matrix of the CDCN framework evaluated on the SEED dataset.

C. Comparison With Previous Studies

In recent years, there are growing interests in the publicly available dataset SEED, based on which various studies have been conducted to explore the challenging task of emotion recognition. Despite some differences in detailed treatments, the existing approaches provide valuable research ideas and findings. Here, we select some competitive studies focusing on the SEED dataset for comparisons.

Table III shows the performances and other details of the following methods: SyncNet [71], GSCCA [72], DBN [68], HCNN [61], SVM [68], MNN [73], STRNN [62], BDAE [70], and GELM [54]. Most of these methods employed all the subjects from the SEED dataset and take their EEG signals for analysis, while HCNN and BDAE considered four and nine subjects, respectively. Note that BDAE gets additional terms from eye movement combined with EEG signals. Therefore, we also took these differences into consideration in the performance comparisons.

DE features are employed in a large part of the above studies, except SyncNet. SyncNet gets the mean accuracy of 77.9% but it performs worse than other DE-related methods, which reflects the crucial correlation between DE features and emotional states. The performance of these methods combined with DE features changes between 83.72% and 91.07%, while GELM achieves a higher accuracy of 91.07%. Moreover, various neural network frameworks are used to recognize different emotional states, except GSCCA and SVM. STRNN with accuracy of 89.50% benefits a lot from the spatial–temporal structure, while BDAE and GELM methods also manage DE features well, and both have performances above 91%. Note that, the effects of BDAE and GELM are slightly better than our framework, with higher standard deviations of 8.91% and 7.54%, respectively. Besides, BDAE employs the recorded data of 9 subjects and additional eye movement signals.

Among these ten methods, CDCN shows an excellent performance for emotion recognition tasks through EEG signals with considerable enhancement compared with other

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
PERFORMANCE COMPARISONS ON THE DEAP DATASET

Study	Description	Accuracy (%)	
		Valence	Arousal
RVM [74]	Relevance vector machine with graph-theoretic features	69	67
C-RNN [75]	Convolutional recurrent neural network with wavelet transform-based features	72.06	74.12
CNN [76]	Convolutional neural network with 101 extracted features per channel	81.41	73.36
MDL [70]	SVM with the pre-trained features from the bimodal deep autoencoder	85.2	80.5
WT-SVM [77]	SVM with the extracted features from discrete wavelet transform method	84.95	84.14
DBN [78]	Deep belief network with power spectral density features	88.33	88.59
CDCN	Channel-fused dense convolutional network with DE features	**92.24**	**92.92**

methods. This indicates that our CDCN framework can robustly capture valid information from EEG signals owing to its good handling of electrode correlations and temporal dependencies.

D. Experiments on DEAP Dataset

In this part, we conduct experiments to evaluate the performance of the developed CDCN model on the DEAP[4] dataset. The DEAP dataset contained EEG and peripheral physiological signals of 32 subjects (50 percent females), which were collected by watching 40 one-minute music videos. The subjects were asked to perform self-assessment of arousal, valence, liking, and dominance on a score from 1 to 9 for each video. The EEG signals were recorded using 32 active AgCl electrodes and downsampled to the 128-Hz sampling rate. Then, a band-pass filter from 4 Hz to 45 Hz was applied. These two preprocessing steps were pregiven in the DEAP dataset [13]. The data is segmented into 1 s samples without overlapping. To compare the performance with previous studies, we construct two classification tasks based on the VA model: low/high valence (task1) and low/high arousal (task2). Besides, for the labels of these two tasks, the self-assessment score between 1 and 4.8 is low and the value between 5.2 and 9 is high. We take the same experimental protocol in [70] to evaluate the performance of the CDCN model on the DEAP dataset. For all the 40 trials of each subject, the samples from 36 trials are taken as the training set and the samples from the remaining 4 trials are taken as the testing set, which can avoid dependency between the training and testing set. The final recognition accuracy is averaged with the recognition accuracy of all the 32 subjects.

Table IV shows the experimental results of the CDCN method with respect to the emotion dimensions (including valence and arousal). We compare the CDCN results with the effects of six existing studies on the DEAP dataset, including RVM [74], C-RNN [75], CNN [76], MDL [70], WT-SVM [77], and DBN [78]. These existing studies used different extracted features and different classifiers. Note that GELM [54] was also evaluated on the DEAP dataset, but using a subject-independent validation scheme with sample shuffling. Therefore, we do not include GELM [54] in the method comparisons on the DEAP dataset.

From Table IV, the average accuracies of the developed CDCN model are 92.24% and 92.92% for two different tasks, respectively. The performance of six compared methods changes between 67% and 89%. Results show that the CDCN model performs much better than the other six compared methods on the DEAP dataset. Moreover, the trained model can be efficiently applied in online BCI recognition tasks.

V. DISCUSSION

The above comparison analysis indicates that the CDCN framework performs better than most of the existing methods, which combines prior knowledge with the developed model to co-train this task. This can be primarily attributed to the 1-D convolution along the temporal dimension in the CDCN framework. It provides an interpretable sense for efficient weighted combinations of contextual features using trained filters. For example, a $1 \times F$ convolution could model the input sample of size $[E, F]$, with E electrodes and F features. Several feature maps of size $[E, 1]$ are obtained after proper training, which denotes adaptive contextual features. There is one component for handling the temporal information, which performs well among the compared methods with different DL frameworks.

Another observation is the employment of 1-D dense block. Dense block encourages feature reuse across the whole network and is sufficient in feature maps with low dimensions. It gains electrode correlations along the spatial dimension through EEG signals. Notably, our CDCN receives excellent performance compared with other DL-based methods, which emphasizes the importance of spatial–temporal information. Especially, the feature reuse of 1-D dense block can help investigate the information of electrode correlations faster, which is the key point of the CDCN model performs notably well among these compared methods. Furthermore, the design principles of the CDCN model can be employed by broader EEG-based recognition tasks. Existing studies have developed some novel methods to conduct channel selection [68], [79], [80]. In [79], the extracted features like synchronization likelihood were used to reduce the number of EEG channels, which led to a slight loss of the classification accuracy rate for emotion assessment. In [68], the critical channels were found through analyzing the weight distributions of the trained DBNs. Our future work will focus on using fewer critical channels to train the recognition networks.

VI. CONCLUSION

In this article, a novel CDCN model is proposed to provide robust representations for emotion recognition from EEG

[4]http://www.eecs.qmul.ac.uk/mmv/datasets/deap/

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

signals. Dense block is used to encourage feature reuse on image-based classification tasks, and the 1-D dense block is reformulated from the 2-D dense block so as to collect electrode correlations from EEG signals. On the other hand, inspired by the basic components of convolution, a 1-D convolution is employed to receive proper weighted combinations of contextual features to deal with temporal information. Consequently, with the above two advances, the proposed CDCN is used to target multichannel series, making it quite proper for managing the spatial–temporal information from EEG signals. The experimental results demonstrate that the CDCN model has excellent performances compared with other existing methods when evaluating on the SEED and DEAP datasets.

One possible improvement of our framework is to integrate signal sequence with prior knowledge. Although the CDCN model outperforms several DL methods with extracted features, prior knowledge could be combined with EEG sequences to further strengthen the model for even better performances with positive effects. In addition, 1-D convolution plays the role of EEG weighted feature combinations, which greatly contributes to the computational efficiency. Especially, if one examines the weights of the trained networks, more crucial information about the weighted feature combinations could be used for EEG feature selection. It is promising that fewer extracted features can gain better performance and deeply discover the contribution between task labels and the features in different frequency bands. The method has good task-adaptive ability and can be applied to other different EEG classification tasks, including fatigue recognition and motor imagery. Therefore, following-up studies could be conducted to further improve the performance on different EEG classification tasks. Looking forward, with excellent real-time performances, the CDCN model could be effectively employed to health monitoring tasks, and extended for broader BCI recognition tasks.

References

[1] R. Pandey and A. K. Choubey, "Emotion and health: An overview," *SIS J. Projective Psychol. Mental Health*, vol. 17, no. 2, pp. 135–152, 2010.

[2] S. D. Pressman and S. Cohen, "Does positive affect influence health?" *Psychol. Bull.*, vol. 131, no. 6, pp. 925–971, 2005.

[3] E. Mumford, H. J. Schlesinger, and G. V. Glass, "The effect of psychological intervention on recovery from surgery and heart attacks: An analysis of the literature," *Amer. J. Public Health*, vol. 72, no. 2, pp. 141–151, 1982.

[4] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.

[5] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," *IEEE Trans. Affect. Comput.*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[6] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[7] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, May 2017.

[8] E. Bekele, D. Bian, J. Peterman, S. Park, and N. Sarkar, "Design of a virtual reality system for affect analysis in facial expressions (VR-SAAFE); Application to Schizophrenia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 739–749, Jun. 2017.

[9] A. Mencattini, E. Martinelli, F. Ringeval, B. Schuller, and C. D. Natale, "Continuous estimation of emotions in speech by dynamic cooperative speaker models," *IEEE Trans. Affect. Comput.*, vol. 8, no. 3, pp. 314–327, Jul.–Sep. 2017.

[10] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," *Comput. Intell.*, vol. 29, no. 3, pp. 436–465, Aug. 2013.

[11] F. Agrafioti, D. Hatzinakos, and A. K. Anderson, "ECG pattern analysis for emotion detection," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 102–115, Jan.–Mar. 2012.

[12] K. Tanaka, M. Tanaka, T. Kajiwara, and H. Q. Wang, "A practical SSVEP-based algorithm for perceptual dominance estimation in binocular rivalry," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 476–482, Jun. 2018.

[13] 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.–Mar. 2012.

[14] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, Apr. 2014.

[15] P. Buitelaar *et al.*, "MixedEmotions: An open-source toolbox for multimodal emotion analysis," *IEEE Trans. Multimedia*, vol. 20, no. 9, pp. 2454–2465, Sep. 2018.

[16] A. R. Damasio *et al.*, "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nat. Neurosci.*, vol. 3, no. 10, pp. 1049–1056, 2000.

[17] Y. M. Chi, Y. T. Wang, Y. J. 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.

[18] Y. Tu, Y. S. Hung, L. Hu, G. Huang, Y. Hu, and Z. Zhang, "An automated and fast approach to detect single-trial visual evoked potentials with application to brain–computer interface," *Clin. Neurophysiol.*, vol. 125, no. 12, pp. 2372–2383, Mar. 2014.

[19] M. H. Silber *et al.*, "The visual scoring of sleep in adults," *J. Clin. Sleep Med.*, vol. 3, no. 2, pp. 121–131, 2007.

[20] J. W. Cao, J. H. Zhu, W. B. Hu, and A. Kummert, "Epileptic signal classification with deep EEG features by stacked CNNs," *IEEE Trans. Cogn. Develop. Syst.*, early access, doi: 10.1109/TCDS.2019.2936441.

[21] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012.

[22] G. H. Klem, H. O. Luders, H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," *Electroencephalography Clin. Neurophysiol.*, vol. 52, no. 3, pp. 3–6, 1999.

[23] V. Vanitha and P. Krishnan, "Time-frequency analysis of EEG for improved classification of emotion," *Int. J. Biomed. Eng. Technol.*, vol. 23, nos. 2–4, pp. 191–212, 2017.

[24] Z. K. Gao, Q. Cai, Y. X. Yang, N. Dong, and S. S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptiform EEG," *Int. J. Neural. Syst.*, vol. 27, no. 4, Jun. 2017, Art. no. 1750005.

[25] Z. K. Gao *et al.*, "An adaptive optimal-kernel time-frequency representation-based complex network method for characterizing fatigued behavior using the SSVEP-based BCI system," *Knowl. Based. Syst.*, vol. 152, pp. 163–171, Jul. 2018.

[26] L. Cheng *et al.*, "Principal states of dynamic functional connectivity reveal the link between resting-state and task-state brain: An fMRI study," *Int. J. Neural Syst.*, vol. 28, no. 7, Apr. 2018, Art. no. 1850002.

[27] H. Zhu *et al.*, "Abnormal dynamic functional connectivity associated with subcortical networks in Parkinsons disease: A temporal variability perspective," *Front. Neurosci.*, vol. 13, p. 80, Feb. 2019, doi: 10.3389/fnins.2019.00080.

[28] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.

[29] Z. K. Gao, W. D. Dang, M. X. Liu, W. Guo, K. Ma, and G. R. Chen, "Classification of EEG signals on VEP-based BCI systems with broad learning," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, doi: 10.1109/TSMC.2020.2964684.

[30] G. L. Ahern and G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis," *Neuropsychologia*, vol. 23, no. 6, pp. 745–755, 1985.

[31] R. N. Duan, J. Y. Zhu, and B. L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. IEEE/EMBS Conf. Neur. Eng.*, San Diego, CA, USA, 2013, pp. 81–84.

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

[32] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *Sci. World J.*, vol. 2014, p. 10, Sep. 2014, doi: 10.1155/2014/627892.

[33] Y. H. Liu, C. T. Wu, W. T. Cheng, Y. T. Hsiao, P. M. Chen, and J. T. Teng, "Emotion recognition from single-trial EEG based on Kernel Fisher's emotion pattern and imbalanced quasiconformal kernel support vector machine," *Sensors*, vol. 14, no. 8, pp. 13361–13388, Aug, 2014.

[34] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[35] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1–9.

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

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[38] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[39] X. Shi *et al.*, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 5622–5632.

[40] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, Las Vegas, NV, USA, 2016, pp. 961–971.

[41] 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, Feb. 2017, Art. no. 016003.

[42] T. Uktveris and V. Jusas, "Application of convolutional neural networks to four-class motor imagery classification problem," *Inf. Technol. Control*, vol. 46, no. 2, pp. 260–273, 2017.

[43] N. Lu, T. Li, X. Ren, and H. Miao, "A deep learning scheme for motor imagery classification based on restricted Boltzmann machines," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 566–576, Jun. 2017.

[44] M. Hajinoroozi, Z. Mao, T.-P. Jung, C.-T. Lin, and Y. Huang, "EEG-based prediction of driver's cognitive performance by deep convolutional neural network," *Signal Process. Imag.*, vol. 47, pp. 549–555, Sep. 2016.

[45] Z. K. Gao *et al.*, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019.

[46] Y. X. Yang *et al.*, "A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG," *Chaos*, vol. 28, no. 8, Aug. 2018, Art. no. 085724.

[47] J. P. Li, S. Qiu, C. D. Du, Y. X. Wang, and H. G. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cogn. Develop. Syst.*, early access, doi: 10.1109/TCDS.2019.2949306.

[48] Y. M. Yang, Q. J. Wu, W.-L. Zheng, and B.-L. Lu, "EEG-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Develop. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2018.

[49] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Map.*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

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

[51] J. Atkinson and D. Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers," *Expert. Syst. Appl.*, vol. 47, pp. 35–41, Apr. 2016.

[52] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. IEEE/EMBS Conf. Neural Eng. (EMBC)*, Osaka, Japan, 2013, pp. 6627–6630.

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

[54] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul.–Sep. 2019.

[55] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul.–Sep. 2019.

[56] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, 2018, Art. no. 056013.

[57] B. H. Kim and S. Jo, "Deep physiological affect network for the recognition of human emotions," *IEEE Trans. Affect. Comput.*, early access, doi: 10.1109/TAFFC.2018.2790939.

[58] R. G. Hefron, B. J. Borghetti, J. C. Christensen, and C. M. S. Kabban, "Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation," *Pattern Recognit. Lett.*, vol. 94, pp. 96–104, Jul. 2017.

[59] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from EEG with deep recurrent-convolutional neural networks," in *Proc. Int. Conf. Learn. Represent.*, San Juan, Puerto Rico, 2016, pp. 1–15.

[60] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, early access, doi: 10.1109/TAFFC.2018.2817622.

[61] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for EEG-based emotion recognition," *Cognit. Comput.*, vol. 10, pp. 1–13, Dec. 2017.

[62] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.

[63] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[64] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, 2015, pp. 448–456.

[65] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, Haifa, Israel, 2000, pp. 807–814.

[66] D. M. Kline and V. L. Berardi, "Revisiting squared-error and cross-entropy functions for training neural network classifiers," *Neural Comput. Appl.*, vol. 14, no. 4, pp. 310–318, 2005.

[67] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.

[68] 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.

[69] S. B. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Pers. Ind. Differ.*, vol. 6, no. 1, pp. 21–29, 1985.

[70] W. Liu, W. L. Zheng, and B. L. Lu, "Emotion recognition using multimodal deep learning," in *Proc. Int. Conf. Neural Inf. Process.*, Kyoto, Japan, 2016, pp. 521–529.

[71] Y. Li *et al.*, "Targeting EEG/LFP synchrony with neural nets," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 4623–4633.

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

[73] S. Keshmiri, H. Sumioka, J. Nakanishi, and H. Ishiguro, "Emotional state estimation using a modified gradient-based neural architecture with weighted estimates," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2017, pp. 4371–4378.

[74] R. Gupta, K. U. R. Laghari, and T. H. Falk, "Relevance vector classifier decision fusion and EEG graph-theoretic features for automatic affective state characterization," *Neurocomputing*, vol. 174, pp. 875–884, Jan. 2016.

[75] 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. Bioinformat. Biomed. (BIBM)*, Guangdong, China, 2016, pp. 352–359.

[76] 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. AAAI Conf. Innov. Appl.*, San Francisco, CA, USA, 2017, pp. 4746–4752.

[77] M. Ali, A. H. Mosa, F. Al Machot, and K. Kyamakya, "EEG-based emotion recognition approach for e-healthcare applications," in *Proc. IEEE Conf. Ubiquitous Future Netw.*, Vienna, Austria, 2016, pp. 946–950.

[78] H. Xu and K. N. Plataniotis, "Affective states classification using EEG and semi-supervised deep learning approaches," in *Proc. IEEE Workshop Multimedia Signal Process.*, 2016, pp. 1–6.

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.

[79] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *Proc. IEEE Eur. Signal Process. Conf.*, 2007, pp. 1241–1245.

[80] X. Jia, K. Li, X. Li, and A. Zhang, "A novel semi-supervised deep learning framework for affective state recognition on EEG signals," in *Proc. IEEE Int. Conf. Bioinformat. Bioeng.*, 2014, pp. 30–37.

Zhongke Gao (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees from Tianjin University, Tianjin, China, in 2007 and 2010, respectively.

He is currently a Full Professor with the School of Electrical and Information Engineering, Tianjin University, where he is the Director of the Laboratory of Complex Networks and Intelligent Systems. His work has been cited more than 2600 times according to Google Scholar. His research interests include deep learning, EEG analysis, complex networks, brain–computer interface, and wearable intelligent devices. He has published more than 100 journal papers in the above fields.

Prof. Gao has been serving as an Editorial Board Member for *Scientific Reports* and an Associate Editor of IEEE ACCESS, *Neural Processing Letters*, and *Royal Society Open Science*.

Xinmin Wang received the bachelor's degree in automation from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the master's degree in control science and engineering with the School of Electrical and Information Engineering.

His research interests include affective computing, EEG analysis, and machine learning.

Yuxuan Yang received the bachelor's degree in automation from Anhui University, Hefei, China, in 2014, the master's degree in automation from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2017, where she is currently pursuing the Ph.D. degree with the School of Electrical and Information Engineering.

Her research interests include affective computing, brain–computer interface, machine learning, and complex networks.

Yanli Li received the bachelor's degree in automation from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2017, where he is currently pursuing the master's degree in control engineering with the School of Electrical and Information Engineering.

His research interests include affective computing, brain–computer interface, and machine learning.

Kai Ma received the Ph.D. degree from the University of Illinois at Chicago, Chicago, IL, USA, in 2014.

He worked with Siemens Medical Solution, Malvern, PA, USA, for more than five years. He is currently working as a Principal Researcher with Tencent, Shenzhen, China.

His research interests include medical image analysis, deep learning, computer vision, and brain–computer interface.

Guanrong Chen (Life Fellow, IEEE) received the M.Sc. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 1981, and the Ph.D. degree in applied mathematics from Texas A&M University, College Station, TX, USA, in 1987.

He was a Tenured Full Professor with the University of Houston, Houston, TX, USA. Since 2000, he has been a Chair Professor and the Founding Director of the Center for Chaos and Complex Networks, City University of Hong Kong, Hong Kong.

Prof. Chen received the 2011 Euler Gold Medal in Russia and the conferred Honorary Doctorate by the Saint Petersburg State University, Russia, in 2011, and by the University of Le Havre, Romandie, France, in 2014. He is a Highly Cited Researcher in Engineering as well as in Mathematics according to Thomson Reuters. He is a member of the Academia of Europe and a fellow of the World Academy of Sciences.

Authorized licensed use limited to: HANGZHOU DIANZI UNIVERSITY. Downloaded on December 19,2024 at 01:13:49 UTC from IEEE Xplore. Restrictions apply.