IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 30, 2020, accepted November 8, 2020, date of publication November 20, 2020,
date of current version December 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039542

Parallel Sequence-Channel Projection Convolutional Neural Network for EEG-Based Emotion Recognition

LILI SHEN, (Member, IEEE), WEI ZHAO, YANAN SHI, TIANYI QIN, AND BINGZHENG LIU

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

Corresponding author: Yanan Shi (shiyanan@tju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61671404 and Grant 61520106002.

ABSTRACT One of the challenges in emotion recognition is finding an effective way to represent spatial-temporal features from EEG. To fully utilize the features on multiple dimensions of EEG signals, we propose a parallel sequence-channel projection convolutional neural network, including temporal stream sub-network, spatial stream sub-network, and fusion classification block. Temporal stream extracts temporal continuity via sequence-projection layer while spatial stream captures spatial correlation via channel-projection layer. Both sequence-projection and channel-projection adopt length-synchronized convolutional kernel to decode whole time and space information. The size of length-synchronized convolutional kernel is equal to the length of transmitted EEG sequence. The fusion classification block combines the extracted temporal and spatial features into a joint spatial-temporal feature vector for emotion prediction. In addition, we present a baseline noise filtering module to amplify input signals and a random channels exchange strategy to enrich the baseline-removed emotional signals. Experimental evaluation on DEAP dataset reveals that the proposed method achieves state-of-the-art classification performance for the binary classification task. The recognition accuracies reach to 96.16% and 95.89% for valence and arousal. The proposed method can improve 3% to 6% than other latest advanced works.

INDEX TERMS Emotion recognition, multi-channel EEG, data augmentation, length-synchronized convolutional kernel, spatial-temporal feature.

I. INTRODUCTION

Human emotion is a complex psychological state involving both subjective experience and physiological reaction [1]–[3]. Emotion can be detected through facial expressions, music listening, eye gaze and physiological signals [4]–[7]. In the past few decades, researchers have proposed a variety of emotion description methods. For example, Ekman *et al.* divided emotion into six discrete basic categorizations: joy, sadness, surprise, fear, anger, and disgust [8]. On this basis, Parrott proposed the tree structure of emotion [9]. Russell presented a continuous valence-arousal scale [10], which took arousal and valence as horizontal and vertical axes to explain most emotional variations. The valence scale ranges from unhappy or sad to happy or joyful. The arousal scale ranges from calm or bored to stimulated or excited [11].

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqing Zhang.

Electroencephalography (EEG) is time-varying signal recorded by multiple electrodes in a standard 10-20 system [12]. The multi-channel EEG information reflects the contextual relevance on time dimension and the electrode correlation on space dimension. Therefore, how to extract effective spatial-temporal features from EEG is the key for emotion recognition. Previous studies mainly focused on hand-crafted feature extraction methods, such as double tree complex wavelet transform (DTCWT) [13], differential entropy (DE) [14] and power spectral density (PSD) [15]. However, hand-crafted features are designed based on a certain database, and only perform well for this database. What's more, hand-crafted feature extraction methods usually fail to capture more abstract EEG features.

In recent years, convolutional neural network (CNN) [16] has become prevalent in EEG emotion recognition. Song *et al.* applied a dynamical graph convolutional neural network (DGCNN) to analyze EEG data [17].

Li *et al.* proposed a hierarchical convolutional neural network [18]. These 2D-CNNs can capture spatial features, but it is difficult to extract temporal information. Therefore, Wang *et al.* utilized a 3D-CNN to decode spatial-temporal information [19]. However, the 3D-CNN cannot effectively extract both temporal and spatial features. To solve this problem, Lin *et al.* combined CNN with recurrent neural network (RNN) [20] to extract spatial-temporal features, and the method achieved satisfied results [21]. Nevertheless, RNN adopts sequential processing over time. Long-term information needs to traverse all units sequentially before entering the current unit. This structure easily lead to the gradient disappearance problem. Although the derivative long short-term memory (LSTM) [22] overcomes the problem, more complex linear layer requires a large amount of memory bandwidth to calculate weights.

Besides the design of feature extraction method, Yang *et al.* found that the emotional signals and the classification level have some correlation with the baseline signals [23]. The 3s baseline signals and the 60s emotional signals were divided into multiple slices, each slice is 1s duration. The average value of three baseline slices is subtracted from each emotional slice. Performance comparison showed baseline removal can significantly improve the classification task. Moreover, one of the problems that researchers must face is alleviating the overfitting problem caused by small dataset. Hence, Kang et al presented an independent component analysis (ICA) - evolution based data augmentation method [24]. Zhang *et al.* applied the empirical mode decomposition (EMD) on the EEG signals and mixed their intrinsic mode functions to create new artificial EEG signals [25].

To address these challenges, a parallel sequence-channel projection convolutional neural network (PSCP-Net) is proposed in this article. The network fully utilizes temporal continuity and spatial correlation of multi-channel EEG signals. In addition, a filter is designed to remove baseline noise, and the differences between emotional signals and filtered baseline signals are adopted as input of the network. In order to increase the training set scale, a data augmentation strategy by randomly exchange corresponding EEG channels between two homogeneous samples is presented. To sum up, the main contributions of this article are as follows:

1) An end-to-end PSCP-Net is proposed to synchronously decode temporal and spatial information, which are concatenated to a joint spatial-temporal feature vector. The network extracts temporal continuity by projecting whole time sequence on each channel, and captures spatial correlation by projecting all channels at same time point.
2) A baseline noise filtering (BNF) module is designed, which can amplify the differences between emotional signals and baseline ones. Moreover, a random channels exchange (RCE) data augmentation strategy is presented to enrich the differences.
3) The proposed method exhibits superior classification performance on DEAP dataset and outperforms other advanced methods for EEG-based emotion recognition.

The rest of this article is organized as follows: Section II summarizes the theoretical background and previous studies related to EEG emotion recognition. Section III describes the proposed method in detail. Section IV evaluates the method on DEAP database systematically. Section V is devoted to conclusion and discussion.

II. RELATED WORK

A. HAND-CRAFTED APPROACHES FOR EMOTION RECOGNITION

Research in EEG emotion recognition mainly involves two aspects [26]: feature extraction and emotion classification. Time-frequency (TF) analysis is a frequently-used feature extraction method in signal processing. In previous years, a variety of feature extraction methods had been proposed for emotion recognition in TF domain [27], such as DE [28] and PSD [29]. Zheng *et al.* demonstrated that DE is the most informative feature [30]. Zhang *et al.* indicated that EEG power and power asymmetry are related to emotional valence [31]. Zheng and Lu introduced deep belief network (DBN) to investigate critical frequency bands and channels [32]. However, single TF analysis method has limited feature representation ability. Mert *et al.* combined ICA with multivariate synchro squeezing transform (MSST) time-frequency analysis method to capture multiple features [33]. Furthermore, Chen *et al.* compared two feature extraction methods and four machine learning classifiers, then found that nonlinear dynamic features could improve the recognition accuracy [34]. Gao *et al.* introduced the complex network theory into time series analysis and achieved good results [35]. Features extracted from EEG signals need to be fed into classifier to realize classification tasks in these methods. Numerous researches demonstrated that support vector machine (SVM) [36], k-nearest neighbors (K-NN) [37], naive Bayes (NB) [38] and other evolutionary algorithms [39]–[41] have reliable ability to accomplish EEG classification task.

B. DEEP LEARNING APPROACHES FOR EMOTION RECOGNITION

Deep learning [42] revealed excellent performance in many applications, such as image classification, video coding and visual saliency detection [43]–[48]. In EEG emotion classification task, some CNN-based methods show great advantages in feature extraction. Wang *et al.* proposed a 3D covariance shift adaptation-based CNN with a dense prediction layer [19]. Zhang *et al.* designed a graph convolutional broad network (GCB-Net) for exploring the deeper-level information of graph-structured data, and adopted broad learning system (BLS) to enhance feature representation [49]. Gao *et al.* proposed a channel-fused dense convolutional network [50]. The network adopted 1D convolutional layer to extract the contextual continuity along the time dimension, and utilized 1D dense structure to capture the electrode correlation along the space dimension. Wang *et al.* presented a concept

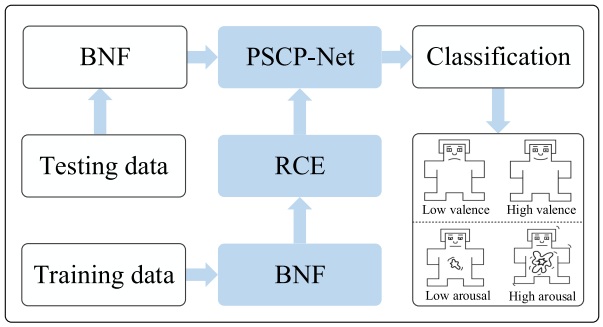


FIGURE 1. Flowchart of the proposed method.

of electrode-frequency distribution maps (EFDMs) with short-time Fourier transform (STFT) [51], and designed a residual block based on deep CNN for emotion classification.

Considering that RNN has great advantages in processing time information. Zhang *et al.* applied spatial-temporal RNN (STRNN) to integrate the learned spatial and temporal features [52]. Tao *et al.* proposed an attention-based convolutional recurrent neural network (ACRNN) to extract discriminative features from EEG signals [53]. Chen *et al.* transformed 1D chain-like EEG vector into a 2D mesh-like matrix sequence [54]. The 2D matrix sequence was divided into segments containing equal time points by using sliding window, and sent to both cascaded and parallel hybrid convolutional recurrent neural network for training. Based on a ConvLSTM network and a temporal margin-based loss function, Kim *et al.* formulated the emotion recognition task as a spectral-temporal sequence classification problem of bipolar EEG signals underlying brain lateralization and photoplethysmogram signals [55]. Wilaiprasitporn *et al.* proposed a cascaded model of CNN and RNN, and evaluated two types of RNNs both LSTM and gated recurrent unit (GRU) [56].

Although deep learning methods have obtained great progresses in EEG emotion recognition, there are still many problems to be solved. For instance, the existing feature-based deep learning methods pay less attention to contextual relevance and electrode correlation information. Therefore, we propose a PSCP-Net to extract time continuity via length-synchronized temporal convolutional kernel and capture space correlation via length-synchronized spatial convolutional filter.

III. PROPOSED METHOD

In this section, we will illustrate the proposed emotion recognition method in detail, shown in Fig. 1.

A. PARALLEL SEQUENCE-CHANNEL PROJECTION CONVOLUTIONAL NEURAL NETWORK

In this section, the proposed PSCP-Net will be introduced. The network is composed of temporal stream (TS) subnetwork, spatial stream (SS) sub-network and fusion classification block. Specifically, the TS and SS sub-networks constitute a parallel spatial-temporal network, which extracts temporal and spatial representation from EEG signals via sequence-projection layer and channel-projection layer, respectively. The fusion classification block is used to vectorize feature maps produced from the parallel network into a spatial-temporal vector, and then the vector is sent to the fully connected layers for classification. Fig. 2 is the structure of proposed PSCP-Net.

1) TS SUB-NETWORK BASED ON SEQUENCE-PROJECTION

The EEG sample $S_j = [C_1, C_2, \ldots, C_{32}]^T \in R^{32\times128}$ is fed into the sequence-projection layer to learn the temporal continuity on each channel, where $j \in [1, batchsize]$. Sequence-projection layer adopts temporal convolutional kernel whose size is equal to the length of transmitted EEG sequence. It is named length-synchronized convolutional kernel. Therefore, complete contextual relevance can be obtained via the length-synchronized convolutional kernel. In the first layer, the quantity of 256 temporal convolutional kernels with the size of (1, 128) is adopted to project each sequence, and move along one stride on space dimension. The shape of output maps is permutated from (32, 1, 256) to (32, 256) by transformation layer. Then, the number of 512 temporal convolutional kernels with the size of (1, 256) and the number of 1024 temporal convolutional kernels with the size of (1, 512) are used to learn higher-level temporal representation, respectively. Finally, we apply 64 temporal convolutional kernels with shape of (1, 1024) to reduce the length of outputs on time dimension. After four sequence-projection layers, the input segment S_j is resolved to a temporal feature vector TFV_j:

$$TFV_j = Conv1D(S_j), \quad TFV_j \in R^{2048} \tag{1}$$

2) SS SUB-NETWORK BASED ON CHANNEL-PROJECTION

The sample S_j is transposed to $S'_j = [D_1, D_2, \ldots, D_{128}]^T \in R^{128\times32}$ as the input of SS network, where channel-projection layer is applied to capture spatial correlation among all channels. The length-synchronized convolutional kernel can handle EEG signals in each channel simultaneously, and the electrode distribution does not need to be transformed to 2D mesh-like matrix. In the first layer, we utilize 64 spatial convolutional filters with the size of (1, 32) to project all channels at same time point, and move along one stride on time dimension. Then, the number of 128 spatial convolutional filters with the size of (1, 64) and the number of 256 spatial convolutional filters with the size of (1, 128) are adopted to integrate the spatial representation. In the last layer, the quantity of 16 spatial convolutional filters with shape of (1, 256) is employed to reduce the length of outputs on space dimension. After four channel-projection layers, the input segment S'_j is flatten to a spatial feature vector SFV_j:

$$SFV_j = Conv1D(S'_j), \quad SFV_j \in R^{2048} \tag{2}$$

3) FUSION CLASSIFICATION BLOCK

The fusion classification block is designed to tune the parameters by cross-validation and achieve the final emotion

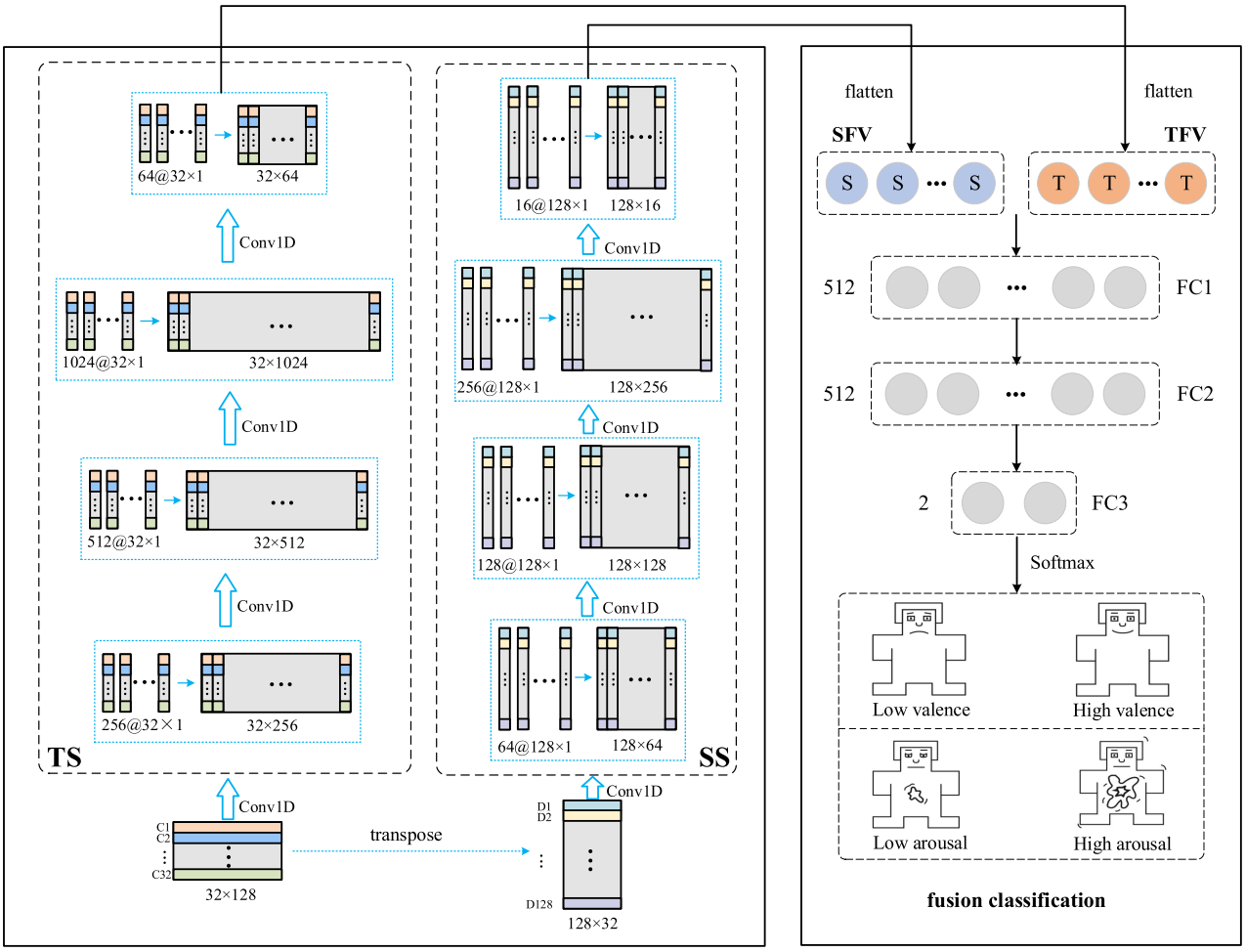


FIGURE 2. Schematic illustration of the PSCP-Net architecture.

classification. The flattened spatial and temporal feature vectors are concatenated to a joint spatial-temporal feature vector $S-TFV_j$:

$$S-TFV_j = concat[SFV_j, TFV_j] \in R^{4096} \tag{3}$$

Then, the fully connected layers receive $S-TFV_j$ as an input to predict human emotional state:

$$y_j = Soft\max[FC(S-TFV_j)], \quad y_j \in R^2 \tag{4}$$

Particularly, the cross-entropy objective function [57] is employed as the loss function of model optimization, which can be expressed as:

$$\hat{\theta} = \arg\min_{\theta}(\sum_{j=1}^{n}\sum_{k=1}^{K} -\log(p_k)\delta(y_j = l_k) + \alpha||\theta||) \tag{5}$$

where $\hat{\theta}$ and θ denote the parameters of well-trained model and current one, n represents the number of training samples which contain K class labels, p_k is the k-th prediction probability of model outputs, δ symbolizes the indicator function, y_j and l_k respectively mean predicted label and true one, α is the trade-off regularization weight.

B. BASELINE NOISE FILTERING MODULE

The DEAP dataset is composed of 3s baseline signals and 60s emotional signals, and the sampling frequency is down-sampled to 128 Hz. The differential signals between emotional signals and baseline ones are adopted to replace emotional signals as the input of model. In order to amplify the differences, a filter is designed to remove the baseline signals with violent fluctuation.

The 3s baseline signals on the 1^{st} channel are transformed into the form of dictionary (i, p_i). The key i is used to record initial order of sampling point p_i. Then, the dictionaries are sorted in ascending order according to the p_i, as in Fig.3 (b). The middle 2s baseline are intercepted from the sorted baseline sequence, as shown in Fig.3 (c). Finally, the intercepted baseline dictionaries are sorted in ascending order based on the key i to restoring intercepted p_i to the initial order, as illustrated in Fig. 3 (d). The restored baseline sequence is the final filtered baseline sequence F_1 of the 1^{st} channel. The violent signals of baseline signals are removed. For 32 EEG channels, a filtered baseline vector (FBV) is formulated as:

$$FBV = [F_1, F_2, \ldots, F_{32}]^T \in R^{32\times256} \tag{6}$$

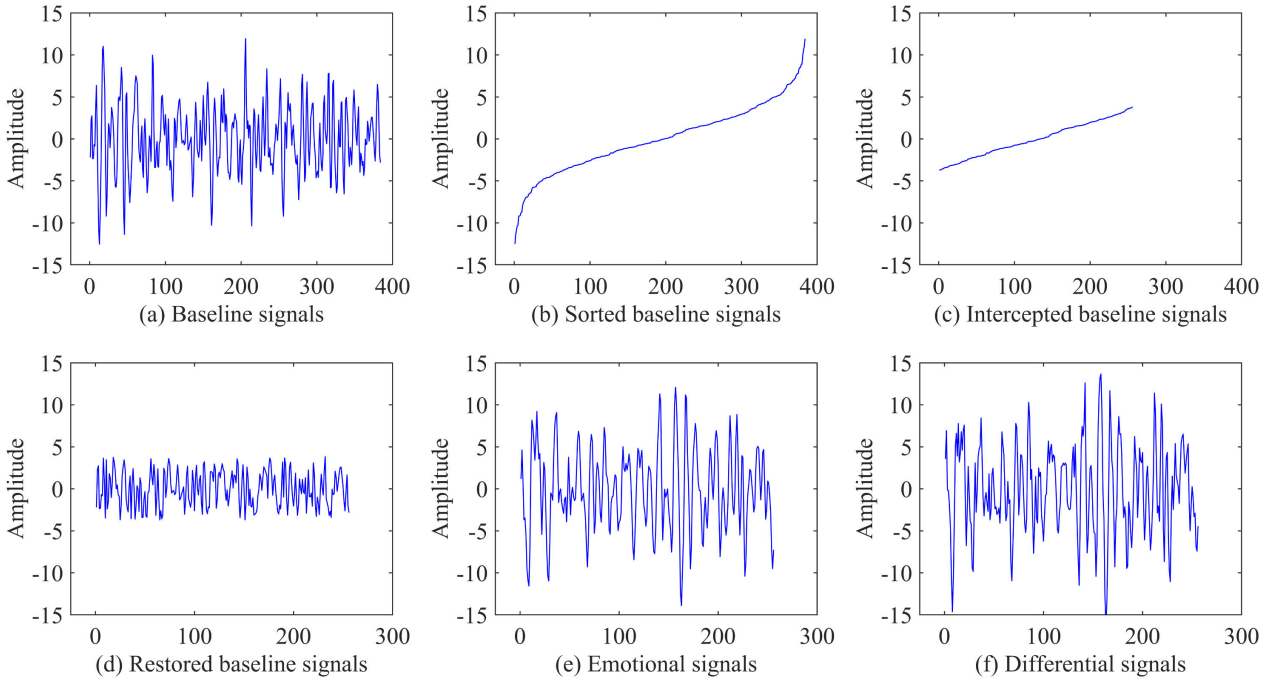


FIGURE 3. Flowchart of baseline removal.

Next, we segment the emotional signals into N small matrixes (32×256) and then minus the FBV for each matrix. All small differential matrixes are concatenated into a big matrix, which is the differential representation with the same size as the emotional signals. As shown in Fig. 3 (e) and Fig. 3 (f), the differential signals have stronger fluctuation than the emotional ones. The characteristics of emotional signals are amplified by BNF module, which can help the network to decode information.

Let the matrix $X \in R^{C \times L}$ denotes the differential representation, where $C = 32$ and $L = 7680$ (60s $\times$ 128Hz) represent EEG channels and differential signals. We can obtain 60 samples (32×128) by slicing the matrix X with non-overlapping. There are 40 emotional videos, so a total of 2400 (40 videos $\times$ 60 samples) samples can be gotten for each subject. Then, each sample is normalized across the non-zero elements using Z-score normalization by the following equation:

$$z = \frac{x - \mu}{\sigma} \tag{7}$$

where x is the non-zero element, μ denotes the average value of all non-zero elements and σ represents the standard deviation of sample elements.

C. DATA AUGMENTATION STRATEGY BASED ON RANDOM CHANNELS EXCHANGE

Similar EEG signals are produced when subjects face to the similar emotional stimulus. Hence a random channels exchange strategy is proposed to augment training set. Without changing the EEG data on channel, the training set can be expanded by randomly exchanging the corresponding EEG channels between the same kind of emotional samples. In order to ensure that there are enough differences between exchanged sample and original one, the number of exchanged channels should have a lower limit (LL) and an upper limit (UL).

The training data are augmented with training epochs online. As shown in Fig. 4, each training batch contains two kinds of samples. These samples are divided into two classes, named H class and L class. The number of each class is recorded as HighNum and LowNum. A random seed T generated from [LL, UL] represents the number of exchanged channels. To guarantee that a batch has at least half of original samples, HighNum and LowNum are divided by four to obtain exchanged iterations, named HighTimes and LowTimes. Two samples are randomly selected from a certain class, and then T channels are randomly picked from 32 channels for exchange.

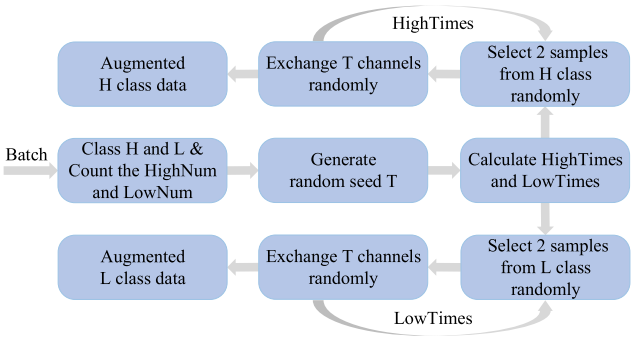


FIGURE 4. Flowchart of random channels exchange.

It is worth noting that the EEG signals produced by different subjects are diverse due to personal reasons. Therefore, the proposed data augmentation strategy cannot be cross-used among subjects.

D. MODEL IMPLEMENTATION

The batch normalization (BN) [58] layer is implemented to follow each convolutional layer, and projects the input to a normal distribution and tunes the optimal parameters of the network. In the PSCP-Net, behind each convolutional layer and fully connected layer, rectified linear unit (ReLU) is inserted as activation function. The L2 regularization strategy with a weight 10^{-4} is adopted to combat the overfitting problem. The Adam optimizer [59] is utilized with a learning rate 10^{-4} to minimize the cross-entropy loss function. The exponential decay algorithm is applied with a decay rate 0.997 to accelerate the convergence rate. The batch size is always maintained at 32. The mixed data of 32 subjects are divided into training set and testing set according to the ratio of 7: 3. The average accuracies of 10-fold cross-validation after 1000 training epochs are used as the final comparative values.

IV. EXPERIMENT RESULTS

In this section, the public DEAP dataset is first introduced. Then, our method is compared with other competitive studies. Finally, the performance of proposed method is analyzed on this dataset.

A. DATASET

The experiments are performed on public DEAP dataset [11], which is a benchmark for emotion classification research. The multi-channel dataset is often adopted to analyze various emotions from EEG signals and peripheral physiological signals. In this article, only EEG signals are applied for emotion recognition.

The DEAP dataset is composed of the responses of 32 healthy participants (50 percent females), whose mean age is 26.9 years. There are 40 music videos of 60s duration for each participant. At the end of each video, a self-assessment was performed for the levels of valence, arousal, dominance and liking on a continuous scale between 1 and 9. The sampling frequency was down-sampled from 512 Hz to 128 Hz, and EOG artifacts were removed with ICA. A band-pass frequency filter from 4.0-45.0 Hz was applied. The preprocessed EEG data for each subject consists of 40 trials and corresponding labels. Each trial contains 60s emotional signals and 3s pre-trial baseline signals. The data format is illustrated in Table 1. The value 5 is set as the threshold to divide the videos into 2 classes according to the rated levels. Then the task is transformed into two binary classification problem of high or low valence and arousal.

TABLE 1. Data format for each subject.

Array	Array shape	Array content
data	$40 \times 32 \times 8064$	video/trial $\times$ channel $\times$ data
labels	40×2	video/trial $\times$ label (valence, arousal)

B. SELECTION FOR THE RANDOM SEED T OF RCE

The proposed data augmentation strategy is described in section III, where the random seed T is generated from [LL, UL]. DEAP dataset has 32 EEG channels. In order to determine the best LL and UL values, the PSCP-Net is used to test T value from 0 to 31, where 0 means no data augmentation.

As shown in Table 2, the recognition accuracy rises first and then declines with the increase of T value. This is because the differences between new sample and original one become subtler when T value is small or large. The recognition accuracy on valence is the highest when T is in [14, 23]. For arousal, the best is [12, 21]. Therefore, the two closed intervals are chosen as the range of T value.

C. CLASSIFICATION PERFORMANCE COMPARISON

A number of emotion classification methods are introduced in the related work. In this subsection, some of them are compared with the proposed method on DEAP dataset. The average accuracy of 10-fold cross-validation is adopted as comparison value.

According to Table 3, average accuracies of the proposed method on valence and arousal are 96.16% and 95.89%. The performance of other seven methods fluctuate between 72.1% and 93.72%. The results indicate that the proposed method is superior to other seven methods. Specifically, our method is 12% points higher than Mohammadi, 24% points higher than Wang, 6% points higher than Yang, 11% points higher than Kang and 3% points higher than Gao, and also surpasses the methods proposed by Chen and Tao. The main reason is that Mohammadi adopts hand-crafted feature extraction method which fail to capture more deep information. Wang utilizes 3D-CNN to extract the spatial-temporal information from EEG signals. However, 3D-CNN cannot integrate effective spatical-temporal representation. Yang, Kang, Chen and Tao adopt hybrid neural network combined with CNN and RNN to capture spatial-temporal features. Nevertheless, these complex network structures are prone to overfitting due to the limited dataset scale. Gao employs CNN based on small size 1D convolutional kernels, which lack the ability of extracting time continuity. Compared with the others, the PSCP-Net applies sequence-projection and channel-projection to jointly decode spatial-temporal information from EEG signals. Moreover, the input signal is amplified via BNF and enriched via RCE. Hence, the proposed method achieves excellent performance.

D. PSCP-NET ARCHITECTURE EVALUATION

1) EMOTION CLASSIFICATION EXPERIMENT

TS network, SS network and proposed PSCP-Net are evaluated on DEAP database. Fig. 5 and Fig. 6 demonstrate the

TABLE 2. Performance comparison of PSCP-Net using different T value on valence and arousal.

Recognition Accuracy (%) Comparison for Different T Value on "Valence" (mean ± std. dev.)							
T	Acc	T	Acc	T	Acc	T	Acc
0	89.40 ± 0.72	8	92.57 ± 0.80	16	**94.29 ± 0.40**	24	93.06 ± 0.29
1	89.37 ± 1.37	9	92.92 ± 0.44	17	93.92 ± 0.30	25	92.81 ± 0.35
2	90.05 ± 0.78	10	93.05 ± 0.62	18	**94.15 ± 0.32**	26	92.58 ± 0.33
3	91.87 ± 0.44	11	93.51 ± 0.58	19	**94.29 ± 0.22**	27	92.03 ± 0.63
4	91.03 ± 0.68	12	93.46 ± 0.94	20	**94.01 ± 0.31**	28	91.95 ± 0.43
5	91.04 ± 0.66	13	93.79 ± 0.38	21	93.51 ± 0.18	29	90.11 ± 0.81
6	91.64 ± 0.54	14	**94.04 ± 0.19**	22	93.75 ± 0.49	30	89.48 ± 0.44
7	91.33 ± 0.68	15	**94.11 ± 0.51**	23	**94.24 ± 0.16**	31	89.23 ± 1.73
Recognition Accuracy (%) Comparison for Different T Value on "Arousal" (mean ± std. dev.)							
T	Acc	T	Acc	T	Acc	T	Acc
0	89.54 ± 1.28	8	92.95 ± 0.68	16	**94.71 ± 0.43**	24	93.44 ± 0.52
1	89.07 ± 0.52	9	93.16 ± 0.27	17	**94.20 ± 0.44**	25	93.00 ± 0.73
2	90.23 ± 0.77	10	93.26 ± 0.33	18	**94.36 ± 0.25**	26	92.80 ± 0.94
3	90.07 ± 0.48	11	93.65 ± 0.57	19	**94.19 ± 0.48**	27	92.69 ± 0.33
4	90.91 ± 0.50	12	**94.41 ± 0.32**	20	**94.34 ± 0.28**	28	91.12 ± 1.04
5	90.96 ± 0.43	13	93.78 ± 0.60	21	**94.18 ± 0.34**	29	91.22 ± 0.42
6	90.60 ± 1.14	14	**94.25 ± 0.40**	22	93.85 ± 0.54	30	90.46 ± 0.76
7	92.53 ± 0.49	15	**94.23 ± 0.65**	23	93.92 ± 0.38	31	90.29 ± 0.81

T denotes the number of channels to be exchanged between the two homogeneous samples.

TABLE 3. Performance comparison of different methods on valence and arousal.

Study	Year	Method Description	Experimental Details	Accuracy (%)	
				Valence	Arousal
Mohammadi [28]	2017	Wavelet features of theta, alpha, beta, gamma and noises, classification with KNN	Set 4.5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	84.05	86.75
Wang [19]	2018	Covariance shift and the unreliability of emotional ground truth, 3D convolutional kernel, EmotioNet	Set 5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	72.1	73.3
Yang [23]	2018	Baseline signals pre-processing, 2D-like frame, parallel convolutional recurrent neural network	Set 5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	90.08	91.03
Kang [24]	2019	ICA - evolution based data augmentation method, CNN-LSTM	Set 5 as the threshold, binary classification of positive/negative status. Subject dependent	84.92	84.92
Gao [50]	2020	Weight combinations of contextual features, channel-fused dense convolutional network	Set 4.8 and 5.2 as the low and high threshold, binary classification of high/low valence and arousal. Subject dependent	92.24	92.92
Chen [54]	2020	2D mesh-like matrix sequences, cascaded hybrid convolutional recurrent neural network, parallel hybrid convolutional recurrent neural network	Set 5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	93.64	93.26
Tao [53]	2020	Attention-based convolutional recurrent neural network (ACRNN)	Set 5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	93.72	93.38
Proposed	- -	Pre-processing of baseline noise filtering (BNF), data augmentation of random channels exchange (RCE), PSCP-Net	Set 5 as the threshold, binary classification of high/low valence and arousal. Subject dependent	**96.16**	**95.89**

results through 10-fold cross-validation. TS network and SS network denote single stream as the input of classification block.

Comparing TS network, SS network and PSCP-Net with each other, it can be found that the same network performs individually for different subjects. For example, the PSCP-Net is the best-performing network for subject 1, but it is the worst-performing network for subject 4. This indicates that the personal reasons have a great impact on the network. In order to compare the performance of three networks on different subjects fairly, same hyperparameters are used in the experiment. In fact, adjusting network hyperparameters for different subjects can improve network performance. Comparing TS network with SS network, it can be found that the performance of SS network is always better than TS network, and the average recognition accuracy difference is

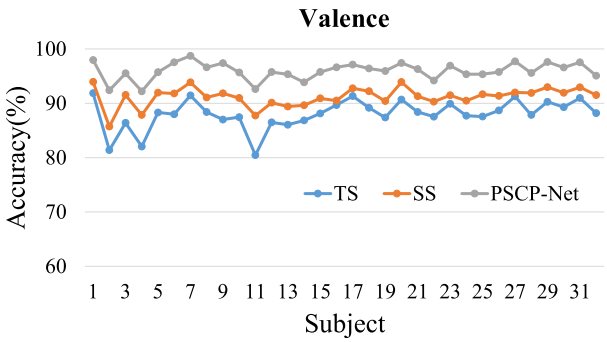


FIGURE 5. Performance comparison of each subject using three networks for valence.

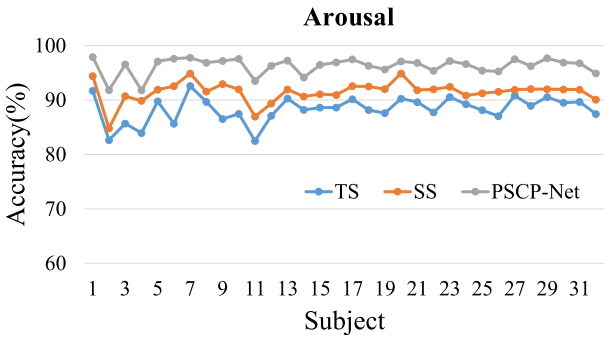


FIGURE 6. Performance comparison of each subject using three networks for arousal.

about 3%. Spatial features are easier to extract than temporal features, and the convolutional network has more advantages in capturing spatial information.

The average recognition accuracies of TS and SS networks are 88.16% and 91.34%. The experimental results also show that the performance of single TS or SS network is good, which may benefit from the proposed BNF and RCE modules. On the whole, the PSCP-Net can always significantly outperform TS and SS networks on all subjects and average accuracies reach to 95.96% and 96.24% for valence and arousal. The PSCP-Net is more robust than other two networks because it can extract abundant spatial-temporal features.

2) LEARNING PROGRESS VISUALIZATION

In order to get an inside view in the learning process, three networks (PSCP-Net, TS network and SS network) losses (the Negative log-likelihood cost) of the training set and testing set are monitored for 100 training epochs. The results are shown in Fig. 7 and Fig. 8.

As can be seen, the training losses of the three networks are similar. However, the testing loss of the PSCP-Net is the smallest. The testing loss of SS network is slightly lower than that of TS network. There is a big gap between the training loss and the testing loss of TS and SS networks, so it is deduced that overfitting occurs in training. In addition, the testing loss of PSCP-Net declines slightly faster than that of TS and SS networks, which demonstrate powerful ability of PSCP-Net.

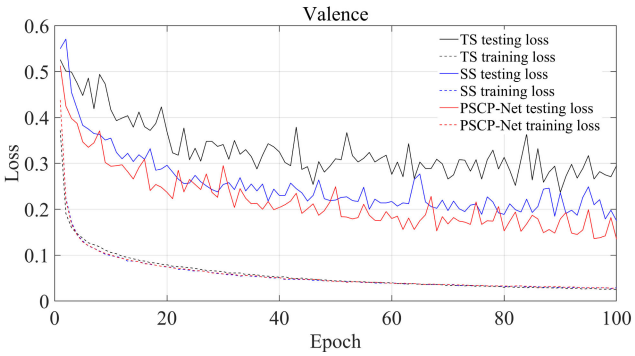


FIGURE 7. Testing and training losses of three networks on valence.

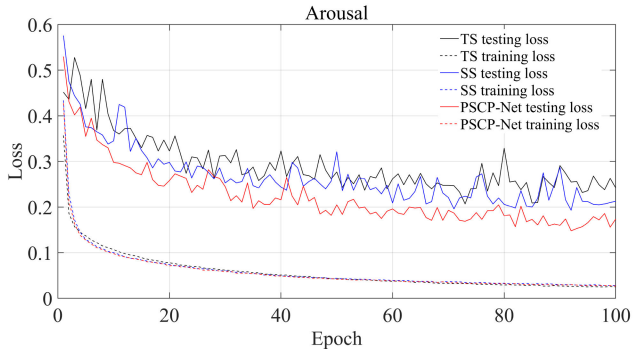


FIGURE 8. Testing and training losses of three networks on arousal.

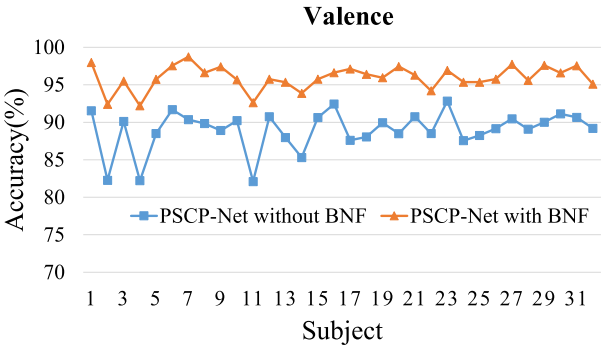


FIGURE 9. Performance comparison of PSCP-Net w/o BNF for valence.

E. INFLUENCE OF PROPOSED BNF AND RCE MODULES

1) EFFECT OF BASELINE NOISE FILTERING

The differences between emotional signals and filtered baseline signals are used as input of the network. In order to verify the effectiveness of proposed BNF module, two baseline preprocessing experiments are conducted on the PSCP-Net. In the first one, the 3s baseline signals are divided into three segments to calculate mean value [23], and then the averaged baseline signals are subtracted from emotional signals. The second one adopts the proposed BNF module. RCE is performed in both two experiments. The experimental results are shown in Fig. 9 and Fig. 10 through 10-fold cross-validation.

The average recognition accuracies with BNF and without BNF are 96% and 89%, respectively. The average accuracy with BNF improves by nearly 7%, which demonstrates that

TABLE 4. Average accuracy and F_1 score under different validation method.

Experimental validation	Module combination	Accuracy (%)		F_1 Score (%)	
		Valence	Arousal	Valence	Arousal
Subject dependent	BNF + RCE (100 epochs) + TS	88.96 ± 0.48	89.73 ± 0.29	73.09 ± 2.54	73.61 ± 3.58
	BNF + RCE (100 epochs) + SS	91.78 ± 0.32	91.57 ± 0.23	75.92 ± 2.35	75.47 ± 3.40
	BNF + RCE (100 epochs) + PSCP-Net	94.49 ± 0.62	94.70 ± 0.24	78.65 ± 2.14	78.63 ± 3.06
	RCE (100 epochs) + PSCP-Net	89.16 ± 0.32	89.78 ± 0.36	73.29 ± 2.53	73.66 ± 3.58
	BNF + 1000 epochs + PSCP-Net	91.48 ± 0.12	91.19 ± 0.14	75.62 ± 2.37	75.08 ± 3.44
	BNF + RCE (1000 epochs) + PSCP-Net	**96.16 ± 0.20**	**95.89 ± 0.18**	**80.33 ± 2.01**	**79.84 ± 2.93**
Subject independent	BNF + RCE (1000 epochs) + PSCP-Net	62.98 ± 1.09	63.72 ± 1.71	50.12 ± 3.53	51.61 ± 5.06

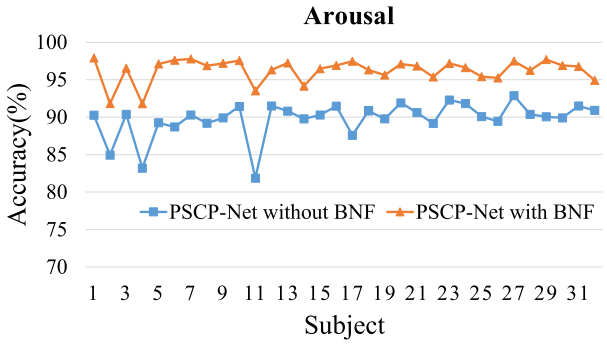


FIGURE 10. Performance comparison of PSCP-Net w/o BNF for arousal.

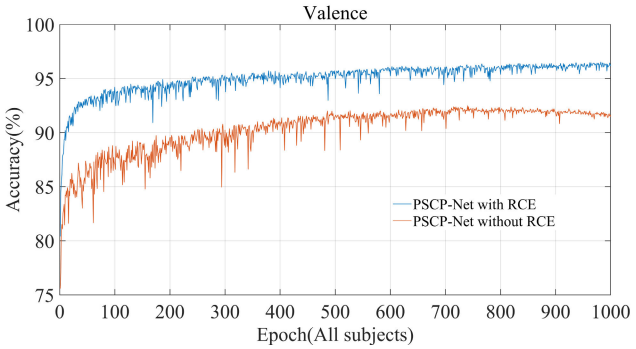


FIGURE 11. Performance comparison of PSCP-Net w/o RCE for valence.

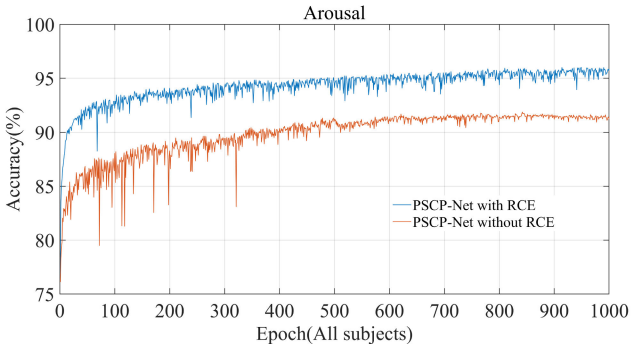


FIGURE 12. Performance comparison of PSCP-Net w/o RCE for arousal.

the EEG signals with BNF can better reveal the emotional states of subject. It is difficult for subject to achieve ideal calm state due to personal reasons, so some useless noises will inevitably be mixed in the process of recording baseline signals. The proposed BNF module can effectively remove these noises.

2) EFFECT OF RANDOM CHANNELS EXCHANGE

The proposed data augmentation strategy is validated in this experiment. In Fig. 11 and Fig. 12, the prediction accuracy with RCE is slowly rising even at around the 1000th epoch, and the accuracy is always higher than that without. However, the recognition accuracy without RCE reaches peak near the 700th epoch. As iterations increase, the accuracy begins to decline slightly due to overtraining. The main reason is that the proposed RCE strategy can enrich the diversity of training set.

F. COMBINATION EXPERIMENT

In the proposed emotion recognition method, PSCP-Net, BNF and RCE play an important role in performance improvement. In this subsection, subject dependent experiment is conducted to measure their contribution while independent experiment is executed to test generalization performance. Table 4 shows the average accuracy and F_1 score.

In the subject dependent experiment, the training set and testing set are mixed data of 32 subjects. As can be seen in Table 4, the PSCP-Net has the best performance in terms of accuracy and F_1 score, followed by SS network and TS network. The BNF module can improve the accuracy by about 5%. The RCE strategy needs to train numerous epochs to fully show its advantage.

In the subject independent experiment, the data corresponds to 30 subjects are selected for training and 2 subjects are selected for testing. From this experiment, it can be found that our model captures effective features in spite of the individual differences.

V. CONCLUSION AND DISCUSSION

In this article, the PSCP-Net is designed to learn the temporal continuity and spatial correlation from EEG signals. In addition, a baseline signals filtering module and a data augmentation strategy based on RCE are proposed. The experimental results on DEAP dataset indicate that the proposed method achieves excellent performance in emotion classification tasks. The average recognition accuracies of valence and arousal are 96.16% and 95.89%, which is significantly higher

than other advanced methods. The advantages of proposed method can be summarized in following three points:

1) The proposed PSCP-Net includes TS sub-network, SS sub-network and fusion classification block. Specifically, the network utilizes length-synchronized convolutional kernel to extract temporal-spatial representation from EEG signals via sequence-projection layer and channel-projection layer.
2) This article fully take advantage of the baseline signals, and design a filter to remove baseline noise. The baseline noise filtering module can effectively reduce the interference caused by baseline noise.
3) Considering the characteristics of multi-channel EEG signals, corresponding EEG channels are randomly exchanged to expand training set. Large dataset provides more learning space for the network and improves the robustness.

Both peripheral physiological signals and EEG signals belong to time series, so there are similarities in regard to feature extraction. In future work, we will consider adding peripheral physiological signals to emotion recognition task. Another worth studying issue is the combination of facial expressions and physiological signals. Compared with physiological signals, facial expressions are easier to collect. Hence, facial expressions can be used to supplement the deficiencies of physiological signals. At the same time, we also find that there is a lack of a database which integrates facial expressions and physiological signals in emotion recognition. Therefore, we hope to establish such a database in the future.

REFERENCES

[1] 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. 2019.

[2] H. A. Vu, Y. Yamazaki, F. Dong, and K. Hirota, "Emotion recognition based on human gesture and speech information using RT middleware," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Taipei, Taiwan, Jun. 2011, pp. 787–791.

[3] G. Castellano, S. D. Villalba, and A. Camurri, "Recognising human emotions from body movement and gesture dynamics," in *Proc. 2nd Int. Conf. Affect. Comput. Intell. Interact.*, Sep. 2007, pp. 71–82.

[4] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 36, no. 1, pp. 96–105, Feb. 2006.

[5] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[6] P. Lakhan, N. Banluesombatkul, V. Changniam, R. Dhithijaiyratn, P. Leelaarporn, E. Boonchieng, S. Hompoonsup, and T. Wilaiprasitporn, "Consumer grade brain sensing for emotion recognition," *IEEE Sensors J.*, vol. 19, no. 21, pp. 9896–9907, Nov. 2019.

[7] P. Sawangjai, S. Hompoonsup, P. Leelaarporn, S. Kongwudhikunakorn, and T. Wilaiprasitporn, "Consumer grade EEG measuring sensors as research tools: A review," *IEEE Sensors J.*, vol. 20, no. 8, pp. 3996–4024, Apr. 2020.

[8] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, and K. Scherer, "Universals and cultural differences in the judgments of facial expressions of emotion," *J. Personality Social Psychol.*, vol. 53, no. 4, pp. 712–717, Oct. 1987.

[9] W. G. Parrott, *Emotions in Social Psychology: Essential Readings*. Philadelphia, PA, USA: Psychology, 2001, pp. 26–56.

[10] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980.

[11] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

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

[13] X. Xu, Y. Zhang, M. Tang, H. Gu, S. Yan, and J. Yang, "Emotion recognition based on double tree complex wavelet transform and machine learning in Internet of Things," *IEEE Access*, vol. 7, pp. 154114–154120, 2019.

[14] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, San Diego, CA, USA, Nov. 2013, pp. 81–84.

[15] F. Al-Shargie, U. Tariq, M. Alex, H. Mir, and H. Al-Nashash, "Emotion recognition based on fusion of local cortical activations and dynamic functional networks connectivity: An EEG study," *IEEE Access*, vol. 7, pp. 143550–143562, 2019.

[16] Y. Le Cun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard, "Handwritten digit recognition: Applications of neural network chips and automatic learning," *IEEE Commun. Mag.*, vol. 27, no. 11, pp. 41–46, Nov. 1989.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul. 2020.

[18] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for EEG-based emotion recognition," *Cognit. Comput.*, vol. 10, no. 2, pp. 368–380, Apr. 2018.

[19] Y. Wang, Z. Huang, B. McCane, and P. Neo, "EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489715.

[20] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, Jun. 1989.

[21] S. Lin and G. C. Runger, "GCRNN: Group-constrained convolutional recurrent neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4709–4718, Oct. 2018.

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

[23] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Rio de Janeiro, Brazil, Jul. 2018, pp. 1–7, doi: 10.1109/IJCNN.2018.8489331.

[24] J.-S. Kang, S. Kavuri, and M. Lee, "ICA-evolution based data augmentation with ensemble deep neural networks using time and frequency kernels for emotion recognition from EEG-data," *IEEE Trans. Affect. Comput.*, early access, Sep. 20, 2019, doi: 10.1109/TAFFC.2019.2942587.

[25] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, and Z. Sun, "A novel deep learning approach with data augmentation to classify motor imagery signals," *IEEE Access*, vol. 7, pp. 15945–15954, 2019.

[26] 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.

[27] 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. 2014.

[28] Z. Mohammadi, J. Frounchi, and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 1985–1990, Aug. 2017.

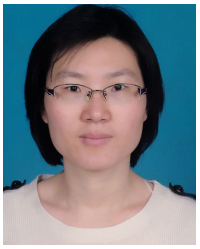
[29] M. Alsolamy and A. Fattouh, "Emotion estimation from EEG signals during listening to quran using PSD features," in *Proc. 7th Int. Conf. Comput. Sci. Inf. Technol. (CSIT)*, Amman, Jordan, Jul. 2016, pp. 1–5.

[30] 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. 2019.

[31] Q. Zhang, S. Jeong, and M. Lee, "Autonomous emotion development using incremental modified adaptive neuro-fuzzy inference system," *Neurocomputing*, vol. 86, pp. 33–44, Jun. 2012.

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

[33] A. Mert and A. Akan, "Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform," *Digit. Signal Process.*, vol. 81, pp. 106–115, Oct. 2018.
[34] P. Chen and J. Zhang, "Performance comparison of machine learning algorithms for EEG-signal-based emotion recognition," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, Oct. 2017, pp. 208–216.
[35] Z. Gao, K. Zhang, W. Dang, Y. Yang, Z. Wang, H. Duan, and G. Chen, "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.
[36] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[37] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
[38] H. J. Yoon and S. Y. Chung, "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2230–2237, Dec. 2013.
[39] C. L. P. Chen, T. Zhang, L. Chen, and S. C. Tam, "I-Ching divination evolutionary algorithm and its convergence analysis," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 2–13, Jan. 2017.
[40] Y. Shen, S. Lou, and X. Wang, "Estimation method of point spread function based on Kalman filter for accurately evaluating real optical properties of photonic crystal fibers," *Appl. Opt.*, vol. 53, no. 9, pp. 1838–1845, Mar. 2014.
[41] Z. Ma, Z.-H. Tan, and J. Guo, "Feature selection for neutral vector in EEG signal classification," *Neurocomputing*, vol. 174, pp. 937–945, Jan. 2016.
[42] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
[44] Z. Ma, D. Chang, J. Xie, Y. Ding, S. Wen, X. Li, Z. Si, and J. Guo, "Fine-grained vehicle classification with channel max pooling modified CNNs," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3224–3233, Apr. 2019.
[45] Z. Pan, X. Yi, Y. Zhang, B. Jeon, and S. Kwong, "Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC," *IEEE Trans. Image Process.*, vol. 29, pp. 5352–5366, Mar. 2020.
[46] Z. Pan, X. Yi, Y. Zhang, H. Yuan, F. L. Wang, and S. Kwong, "Frame-level bit allocation optimization based on video content characteristics for HEVC," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 1, pp. 1–20, Apr. 2020.
[47] R. Cong, J. Lei, H. Fu, J. Hou, Q. Huang, and S. Kwong, "Going from RGB to RGBD saliency: A depth-guided transformation model," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3627–3639, Aug. 2020.
[48] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 10, pp. 2941–2959, Oct. 2019.
[49] T. Zhang, X. Wang, X. Xu, and C. L. P. Chen, "GCB-net: Graph convolutional broad network and its application in emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Aug. 27, 2019, doi: 10.1109/TAFFC.2019.2937768.
[50] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, early access, Feb. 25, 2020, doi: 10.1109/TCDS.2020.2976112.
[51] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, Art. no. 107506.
[52] 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.
[53] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "EEG-based emotion recognition via channel-wise attention and self attention," *IEEE Trans. Affect. Comput.*, early access, Sep. 22, 2020, doi: 10.1109/TAFFC.2020.3025777.
[54] J. Chen, D. Jiang, Y. Zhang, and P. Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset," *Comput. Commun.*, vol. 154, pp. 58–65, Mar. 2020.
[55] B. Hyung Kim and S. Jo, "Deep physiological affect network for the recognition of human emotions," *IEEE Trans. Affect. Comput.*, vol. 11, no. 2, pp. 230–243, Jun. 2020.
[56] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, and E. Chuangsuwanich, "Affective EEG-based person identification using the deep learning approach," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 3, pp. 486–496, Sep. 2020.
[57] 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, Dec. 2005.
[58] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[59] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015, pp. 1–15.



LILI SHEN (Member, IEEE) received the Ph.D. degree in communication and information system from Tianjin University, Tianjin, China, in 2010. She was a Visiting Scholar with the Centre for Vision Research, York University, Toronto, Canada, in 2014. She is currently an Associate Professor with the School of Electrical and Information Engineering, Tianjin University, China. Her research interests include 2-D/3-D image processing, computer vision, and multimedia communications.

WEI ZHAO received the B.Eng. degree from the College of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China, in 2019. He is currently pursuing the M.S. degree with the School of Electrical and Information Engineering, Tianjin University, China. His research interests include machine learning and deep learning in electroencephalogram.

YANAN SHI received the B.S. degree in communication engineering from Tianjin University, Tianjin, China, in 2018, where she is currently pursuing the M.S. degree with the School of Electrical and Information Engineering. Her research interests include video coding and deep learning.

TIANYI QIN received the B.S. degree in communication engineering from the Ocean University of China, Qingdao, China, in 2019. He is currently pursuing the M.S. degree with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include computer vision and image retrieval.

BINGZHENG LIU received the B.Eng. degree in electronic information engineering from Northeastern University at Qinhuangdao, Qinhuangdao, China, in 2015. He is currently pursuing the M.S. degree in electronics and communication engineering with the School of Electrical and Information Engineering, Tianjin University, Tianjin, China. His research interests include view synthesis and image classification in computer vision.