IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 21, 2019, accepted December 13, 2019, date of publication December 24, 2019, date of current version January 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962085

Multimodal Emotion Recognition Based on Ensemble Convolutional Neural Network

HAIPING HUANG[1,3], ZHENCHAO HU[1,3], WENMING WANG[1,2,3], AND MIN WU[1,3]

[1]School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2]University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing 246011, China
[3]Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210023, China

Corresponding author: Haiping Huang (hhp@njupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672297, in part by the Key Research and Development Program of Jiangsu Province under Grant BE2017742, in part by the Sixth Talent Peaks Project of Jiangsu Province under Grant DZXX-017, in part by the National Key Research and Development Program under Grant 2018YFB0803403, in part by the Postgraduate Research and Practice Innovation Program of Jiangsu Province under Grant KYCX19_0908, and in part by the Key Project on Anhui Provincial Natural Science Study by Colleges and Universities under Grant KJ2019A0579.

ABSTRACT In recent years, emotional recognition based on Electrophysiological (EEG) signals has become more and more popular. But the researchers ignored the fact that peripheral physiological signals can also reflect changes in mood. We propose an Ensemble Convolutional Neural Network (ECNN) model, which is used to automatically mine the correlation between multi-channel EEG signals and peripheral physiological signals in order to improve the emotion recognition accuracy. First, we design five convolution networks and use global average pooling (GAP) layers instead of fully connected layers; and then the plurality voting strategy is adopted to establish the ensemble model; eventually this model divides emotions into four categories. Based on the simulations on DEAP dataset, the experimental results demonstrate the superiority of the ECNN compared with other methods.

INDEX TERMS ECNN, emotion recognition, physiological signal, plurality voting.

I. INTRODUCTION

Emotion plays an important role in interpersonal communication and medical research. Emotion recognition is an essential part of emotion research, and it is an interdisciplinary field of computer science, neuroscience, psychology and cognitive science [1]. Among them, "*Affective Computing*" is a relatively authoritative technology. Affective computing is that the computer can automatically identify, understand and reflect human emotions. For example, computers can recognize emotions from a person's facial expressions [2], voice [3], [4], blinking [5] and posture [6], etc. Most previous studies have focused on voice and facial expressions. But in some cases, people cannot accurately reflect their emotions into their facial expressions. For example, some people deliberately hide their true feelings for some special reasons, and patients who suffer from facial neuritis cannot express affection [7]. In order to solve the above problems, many researchers have proposed the emotion recognition method based on peripheral physiological signals or Electrophysiological (EEG) signals; and compared with the traditional methods, these two emotion recognition methods are more practical and reliable. Although some research developments regarding the above two methods have been achieved, there still exist some challenging problems. Firstly, most studies are only absorbed in emotion recognition based on peripheral physiological signals such as electrooculography (EOG) or only focus on EEG signals; however, the correlation between EEG signals and peripheral physiological signals is ignored. Secondly, most researchers use manual feature extraction method for EEG signal data; this method is not stable, and it would probably further affect the accuracy of emotion recognition.

In order to address the above challenging problems, our contributions can be summarized as follows:

(1) To improve the emotion recognition accuracy, the Ensemble Convolutional Neural Network (ECNN) is employed to dig out the correlation between multi-channel EEG signals and multiple peripheral physiological signals, which can extract the effective features for four categories of multimodal emotion recognition [8].

(2) The Global Average Pooling (GAP) strategy is adopted to replace the traditional fully connected layers [9] in order to address the overfitting problem, which also further

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

improve the accuracy and stability based on the effective utilization of information extracted by Convolutional Neural Network(CNN).

The rest of this paper is organized as follows. Section 2 introduces the relevant research on emotion recognition. Section 3 mainly describes our proposal. The experimental analysis and discussion are illustrated in section 4. Finally, section 5 concludes the whole paper.

II. RELATED WORKS

People's EEG signals are objective and cannot be concealed [10], which can accurately and intuitively reflect people's psychological status and emotional characteristics. So it has attracted the attention of many researchers. Atkinson and Campos [11] uses minimum Redundancy-Maximum-Relevance (mRMR) to extract features from multi-channel EEG signals, and then Support Vector Machine (SVM) is used to classify emotions. Verma and Tiwary [12] preprocessed EEG signals using kernel Principal Component Analysis (K-PCA), and they classified emotions using k-Nearest Neighbors (KNN) and Support Vector Machine based on Radial Basis Function (RBF-SVM). But the accuracy of emotional recognition based on SVM is not high when dealing with big data and multi-class classification problem. In order to solve this problem, Zheng *et al.* [13] trained a Deep Belief Network (DBN) by using the Differential Entropy (DE) feature extracted from the multi-channel EEG as the input, where Hidden Markov Model (HMM) is proposed as an auxiliary method, to obtain a more reliable emotional transition state. In addition, Zhuang *et al.* [14] adopted Empirical Mode Decomposition (EMD) method to extract features of multi-channel EEG signals, and EMD method can get multiple Intrinsic Mode Functions (IMFs) and a single residual signal. They took the statistical features of IMFs as the final feature selection, and then used the SVM for emotion recognition. However, the effective information of EEG signals cannot be fully extracted because of the existence of residual signal. All the methods mentioned above only used EEG signal to identify emotions, and they did not take the effects of peripheral physiological signals on emotions into account.

In order to improve the accuracy of emotion recognition, researchers proposed multimodal emotion recognition and began to use peripheral physiological signals to assist EEG signals, so that various physiological changes of human body can be fully considered in feature extraction. Yin *et al.* [15] performed emotional recognition on single channel EEG signal and single peripheral physiological signal. They first normalized each type of signal and used ensemble classifier of stacked auto-encoder for emotion recognition. Chen *et al.* [16] used K-Nearest Neighbor and Random Forest to perform emotion recognition on multi-channel EEG signals and a variety of peripheral physiological signals.

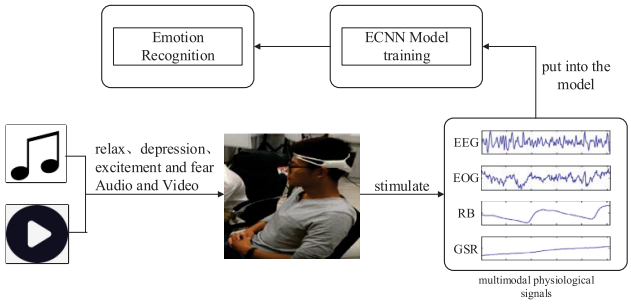


FIGURE 1. Flowchart of multimodal emotion recognition.

III. METHODS

Our proposed methodology addresses two challenging problems. The first one is how to find the correlation between multiple peripheral physiological signals and multi-EEG signals; the second one is how to improve the accuracy of emotion recognition based on the extracted features. The emotion recognition framework for dealing with multimodal physiological signals is illustrated in Fig. 1.

A. ENSEMBLE CONVOLUTIONAL NEURAL NETWORK

We design an ensemble convolutional neural network (ECNN) which is devoted to conduct emotion recognition tasks, where CNN is used to capture the correlation between multiple channel signals and Ensemble Learning (EL) is employed to improve classification ability. Furthermore, we will adopt the GAP layer to improve the effect of emotion classification. Next, we will introduce the two components CNN and ensemble strategy of our model, respectively.

1) CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional neural network is one of the most successful models in deep learning, which is generally composed of multiple convolution layers and pooling layers. The convolution layer simulates the biological mechanism of simple cells with local receptive fields and extracts the primary characteristics of signals by means of sparse connectivity and parameter sharing [17].

Multiple convolution kernels are used for convolution operations on features obtained from the previous layer, and the combination of the results of these operations can get the features for the next output via the activation function, as shown in Eq. 1.

$$x_j^m = f\left(\sum_{i \in M_j} x_i^{m-1} * w_{ij}^m + b_j^m\right) \tag{1}$$

Here x_j^m is the j^{th} feature of layer m, w_{ij}^m is the connection weight between the j^{th} feature of layer m and the i^{th} feature of layer $m-1$. M_j represents a selection of input features. b_j^m is the corresponding offset parameter. Symbol $*$ is the convolution operation. $f(\circ)$ is the activation function of the

network which adopts Rectified Linear Units (Relu) in order to enhance the network performance [18]. The expression of Relu function is as follows:

$$f(x) = \max(0, x) \tag{2}$$

The pooling layer is a further optimization of the convolution layer, which can retain the main features and reduce the number of parameters in the next layer to prevent overfitting [19].

The training of CNN model mainly consists of two stages. The first stage is forward propagation, which describes the process of data transmitted into convolutional layers and pooling layers, then passing through the activation function, and finally getting the corresponding output. The second stage is back propagation and the weight update. The error function of each network layer is obtained by back propagation based on the error between predicted value and actual value, and then the loss function can be obtained. As shown in Eq. 3, the multi-class classification cross entropy loss function [20] is adopted in order to cooperate with softmax layer.

$$L = -\frac{1}{N}\sum_{n=0}^{N-1}\sum_{k=0}^{K-1} y_n^k \log p_n^k \tag{3}$$

In Eq. 3, N represents the total number of samples, K is the number of labels, the probability that the n^{th} sample is predicted to be the s^{th} label is p_n^k, and y_n^k is the real data of the k^{th} label in the n^{th} sample. Weight updating uses Root Mean Square prop (RMSprop) optimization algorithm, which is not easy to fall into local optimum compared with Stochastic Gradient Descent (SGD) algorithm, and this algorithm has the fast convergence rate and is more suitable for deep convolution network. The weight update equations are as follow:

$$s_{dw_t} = \beta s_{dw_{t-1}} + (1-\beta)\, dW_t^2 \tag{4}$$

$$s_{db_t} = \beta s_{db_{t-1}} + (1-\beta)\, db_t^2 \tag{5}$$

$$W_t = W_{t-1} - \alpha \frac{dW_t}{\sqrt{s_{dw_t}} + \varepsilon} \tag{6}$$

$$b_t = b_{t-1} - \alpha \frac{db_t}{\sqrt{s_{db_t}} + \varepsilon} \tag{7}$$

Here t is the current iteration times, and the gradient obtained by back propagation are dW and db. α is the network learning rate and β is the exponential of gradient accumulation. The gradient momentum of the t^{th} iteration are s_{dw_t} and s_{db_t}. In Eq. 6 and 7, weights W_t and b_t are updated by the gradient momentum. In order to prevent the denominator from being zero, a very small value ε is used for smoothing. We set $\varepsilon = 10^{-8}$, $\alpha = 0.001$, $\beta = 0$.

2) ENSEMBLE LEARNING(EL)

The main idea of ensemble learning is to first achieve independent training and learning based on several base learners and then flexibly combine some of them according to the learning effect. The effect of combination is usually better than that of an individual learner. Each convolutional network can be considered as a weak learner, and we need to ensemble them into a strong learner according to certain strategies because the learning effect of single CNN layer is usually poor. Furthermore, in this way, the problem of overfitting caused by the single CNN model will be overcome.

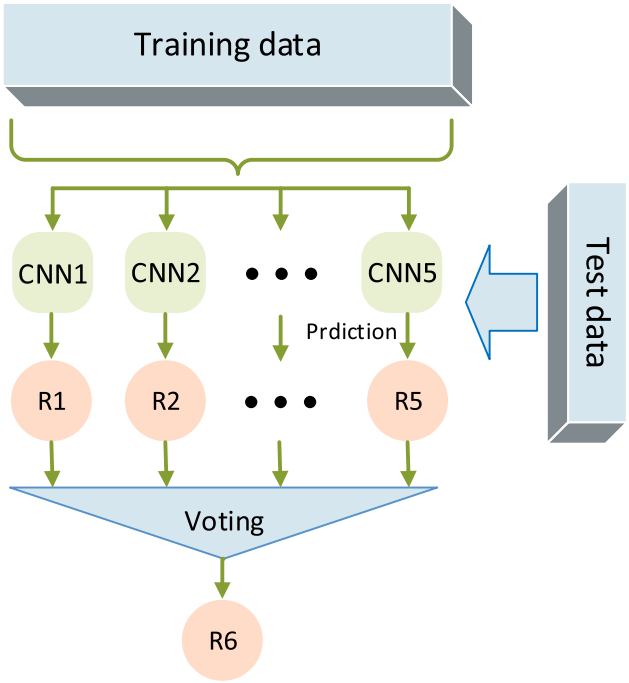


FIGURE 2. Framework of ECNN.

The proposed ECNN framework is depicted in Fig. 2. This paper designs five CNN classifiers, which contains CNN1∼CNN5 respectively. We regard each CNN classifier as a base learner. On the basis of five CNN learners, a strong learner is formed by the strategy of plurality voting. Each learner y_i will predict a result $Y(x)$ from the category set $\{c_1, c_2, \ldots, c_N\}$. The ensemble strategy is formulated as follows:

$$Y(x) = c_{\arg\max_j \sum_{i=1}^{T} y_i^j(x)} \tag{8}$$

wherein, N-dimensional vector $\left(y_i^1(x), y_i^2(x), \ldots y_i^N(x)\right)$ represents the predicted output of y_i on sample x, and $y_i^j(x)$ represents the output of y_i on category c_j. T is the number of learners. j is the number of categories, and its value is 4.

B. GAP

Conventional CNNs always add fully connected layers behind convolution layers to achieve specific applications. However, the parameters of the fully connected network layer are so many and prone to overfitting, which will affect the generalization ability of the network. There is a regularization method to alleviate the overfitting, however continuous tests on the effects of different dropout will bring extra time consumption using this method.

In this paper, we adopt the global average pooling (GAP) layers to replace the traditional fully connected layers in CNN. The last convolution network layer generates a feature map for each category, and then the feature maps are transmitted into the global average pooling layer to take the average value of each feature map which is finally passed to the softmax layer. The advantage of GAP is that there are no additional parameters to be optimized, which can effectively avoid overfitting. In addition, GAP is actually a regularizer, which directly converts the number of features of the previous network into the number of classifications for better effect.

IV. EXPERIMENTS AND DISCUSSIONS

In this section, we will demonstrate the effectiveness of ensemble learning; and meanwhile the effects of GAP and other methods on emotion recognition will be measured; and then, the availability of different physiological signals on emotion recognition will be compared.

A. EXPERIMENTAL DATASET

In this paper, we use the DEAP dataset [21] to validate our proposed approach, which is an open source dataset based on physiological signal of emotion recognition. In this dataset, 32 subjects were selected. Each subject wore a data acquisition device to watch video, which could collect his/her EEG signals from 32 different channels of the brain and 8 kinds of peripheral physiological characteristics. They selected 60 seconds of EEG signals for downsampling to 128Hz, and each channel generated 8064 discrete sampling points. After watching videos, subjects gave feedback on their emotions, and finally we obtained the evaluation values of different emotions on the four metrics of arousal, valence, liking and dominance. Based on the two-dimensional emotional space proposed by Russell [22]: the arousal (it ranges from relaxed to aroused) and the valence (it ranges from pleasant to unpleasant), our experimental results can be divided into four categories.

In order to obtain accurate emotional labels, we adopt a simple and fast algorithm k-means [23] to cluster arousal and valence degree on DEAP dataset. The k-means algorithm divides the sample set into K clusters according to the distance between samples. The aim of this algorithm is to make the distance between samples within clusters as small as possible and that between clusters as large as possible. The clustering results are shown in Fig. 3, where the vertical axis represents "*valence*" and the horizontal axis represents "*arousal*". These cluster results $\{c_1 = Relax, c_2 = Depression, c_3 = Excitement, c_4 = Fear\}$ will be considered as sample labels which can describe the emotional state more objectively and clearly.

B. EXPERIMENTAL RESULTS AND ANALYSIS

We take 32 EEG channel signals and three kinds of peripheral physical signals including GSR (Galvanic skin esponse), RB (Respiration belt) and EOG (Electrooculogram). We place the designed ensemble convolution model with the input data of $60 \times 128 \times 35$ for training. A total of 1280 samples are used in the experiments, where 80% of the samples of each subject were randomly selected as the training set, and the remaining 20% as the test set.

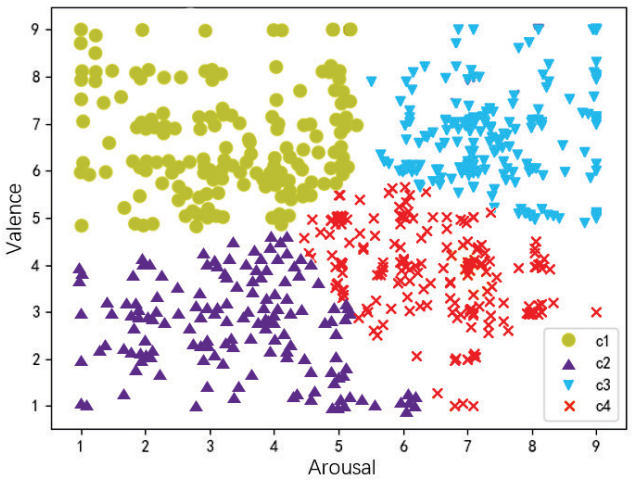


FIGURE 3. Clustering results.

The specific parameters of each layer is shown in Table 1. In CNN1 and CNN2, the convolution layer is responsible to initially extract more relevant features and enable the subsequent convolution layers easier to extract effective information. CNN3 network structure adds a convolution layer of 3×3-30 on CNN2 to reduce the number of features. CNN4 and CNN5 can further reduce the number of features so that the convolution layer can automatically extract the features and send them to the GAP layer. We set the maximum iteration times of the CNN model to be 500. In order to reduce the training time of the model, the training will be terminated once the loss function exceeds 25 iterations without decreasing. The pooling layer adopts the average pooling, and the softmax layer is selected as the classifier.

Generally speaking, the classification effect of CNN will be improved with the increase of network depth, but more layers of CNN will easily lead to overfitting, which will greatly reduce the classification effect instead. Specific experimental results are shown in Table 2.

CNN1 has the worst effect due to its shallow depth; the depth of CNN3 is suitable, and the average accuracy is 78.26%; CNN4 has reduced the classification effect. Through ensemble learning, the average accuracy of plurality voting strategy reaches 82.92%, while the lowest accuracy reaches 71.37%, which is significantly improved compared with other single CNN model.

Because the plurality voting strategy sets the best choice of most classifiers, it is more accurate than single classifier and reduces the miscalculation. Although the single model with appropriate depth can also achieve desirable classification effect, the ensemble model achieves the best classification accuracy.

In order to illustrate the advantages of GAP layer in multimodal emotion recognition [24], we compare it with the fully

TABLE 1. Designed structural parameters of each CNN.

Structure	CNN1	CNN2	CNN3	CNN4	CNN5
Convolutional layer	3×3-35	3×3-35	3×3-35	3×3-35	3×3-35
Convolutional layer	-	5×5-35	5×5-35	5×5-35	5×5-35
Convolutional layer	-	-	3×3-30	3×3-30	3×3-30
Pooling layer	-	-	2×2	2×2	2×2
Convolutional layer	-	-	-	3×3-20	3×3-20
Pooling layer	-	-	-	3×3	3×3
Convolutional layer	-	-	-	-	3×3-10
Pooling layer	-	-	-	-	2×2
Convolutional layer	1×1-4	1×1-4	1×1-4	1×1-4	1×1-4
GAP	1	1	1	1	1

TABLE 2. Table of accuracy of each model.

Structure	Minimum accuracy	Maximum accuracy	Average accuracy
CNN1	56.49%	69.31%	61.2%
CNN2	61.58%	81.69%	68.53%
CNN3	62.56%	91.16%	78.26%
CNN4	59.39%	90.38%	76.33%
CNN5	60.49%	90.46%	78.88%
Voting	**71.37%**	**92.34%**	**82.92%**

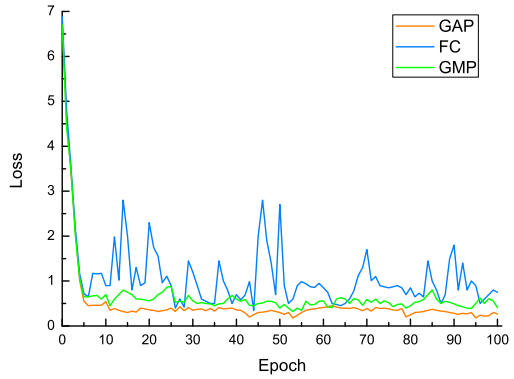


FIGURE 4. Training convergence rate of different network layers.

connected layer (FC) and the global maximum pooling layer (GMP). Fig. 4 shows the effect of different network layers following with convolutional neural network layers. We can find that the convergence speed of GAP is about the same as that of GMP and FC at the beginning of the training process, but the convergence speed of FC becomes unstable in the later training process. GMP convergence is stable but the accuracy is not as high as that of GAP, and the convergence effect of GAP is the best and most stable. This is because GAP represents the mean value of all the feature maps, and GMP is the maximum value, it is not accurate and stable.

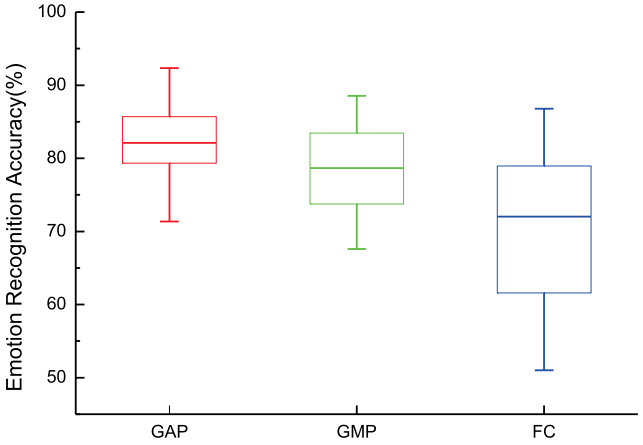


FIGURE 5. The effect of different network layers on accuracy.

The experimental result is illustrated in Fig. 5. It can be found that the GAP acquire the highest average accuracy and the smallest variance, while the FC has the worst precision and the GMP is somewhere in between. This is because the GAP greatly reduces the number of parameters and makes the feature extraction of convolutional neural network more deliberate, so the effect is better than that of the FC.

After determining the ensemble model of 5 convolution layers, we consider the average accuracy of emotion recognition as the metric, and our proposal is compared with other 3 similar methods [11], [14], [15]. The experimental results are illustrated in Fig. 6.

The comparison shows the superiority of our proposal. Both the methods adopted by Zhuang *et al.* [14] and Atkinson and Campos [11] have limited accuracy because of the insufficient and imprecise of features extracted from EEG signals. Compared with the above two methods, Yin *et al.* [15] use EOG signals to assist single EEG signals, which improved the emotion recognition accuracy. Compared with [15], our ensemble convolutional neural network can preferably mine the correlation between different signals and select better features, so as to effectively improve the accuracy of emotion recognition.

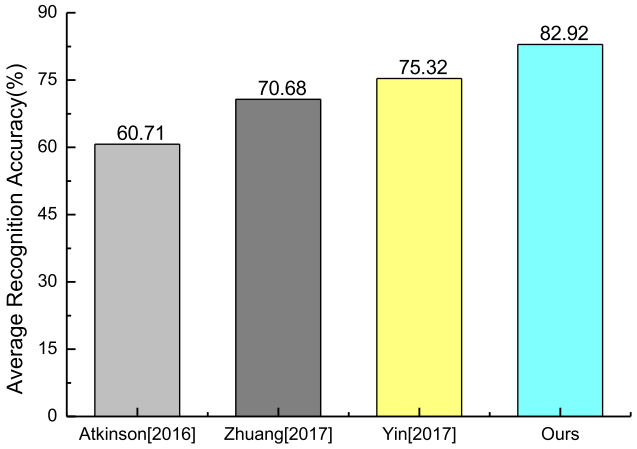


FIGURE 6. Performance comparison between relevant methods.

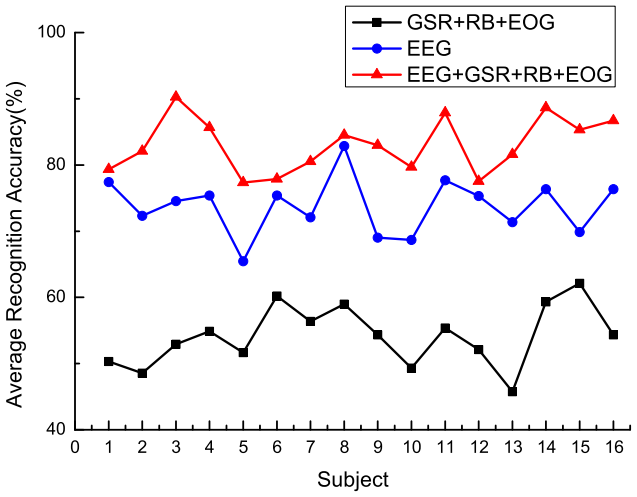


FIGURE 7. The average accuracy of single modal signal and multimodal signal obtained by emotional classification training of ECNN.

Next, our proposal is used to conduct emotional recognition experiments on three peripheral physiological signals (GSR, RB and EOG), single EEG signals and their combination, respectively. And 16 subjects are randomly selected for experimental analysis. Fig. 7 shows the difference between multimodal physiological signal recognition and single modal one. The average classification accuracy of emotion recognition based on three kinds of peripheral physiological signals is 54.15% (the lowest). The classification accuracy of EEG signals is better than that of "GSR+RB+EOG" with an average accuracy of 73.76%. However, the classification effect of multimodal physiological signals combining with EEG signals (EEG+GSR+RB+EOG) is obviously higher than that of single EEG signals.

V. CONCLUSION

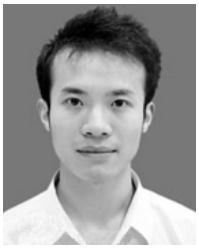
In this paper, we proposed a new ensemble convolutional neural network model (ECNN) for multimodal emotion recognition. In order to improve the stability and accuracy of emotion recognition, the CNN is responsible for mining the inter-channel information from EEG and peripheral physiological signals and extracting the effective features; and the GAP layers can address the overfitting problem by substituting the fully connected layers. And meanwhile, we adopt the plurality voting strategy to establish the ensemble model for achieving the four categories of emotions. Experimental results on DEAP dataset show that the average accuracy of emotion recognition is improved to 82.92% with the ECNN, which is superior to the single CNN model. Furthermore, compared with the single EEG signals and single peripheral physiological ones, the accuracy of multimodal emotion recognition has been significantly increased by 9.16% and 28.77%, respectively.

However, the training time of our ECNN model is still unsatisfactory, and how to further reduce time consumption has become one of issues that we need to address in the future. Apart from this, for future works it is interesting to use the ensemble recurrent neural network to identify emotions because of EEG and peripheral physiological signals are time series data.

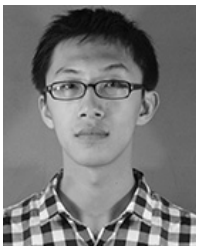
REFERENCES

[1] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2017.
[2] J. Guo, Z. Lei, J. Wan, E. Avots, N. Hajarolasvadi, B. Knyazev, A. Kuharenko, J. C. S. Jacques, Jr., X. Baró, H. Demirel, S. Escalera, A. Allik, and G. Anbarjafari, "Dominant and complementary emotion recognition from still images of faces," *IEEE Access*, vol. 6, pp. 26391–26403, 2018.
[3] S. Zhang, S. Zhang, T. Huang, and W. Gao, "Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching," *IEEE Trans. Multimedia*, vol. 20, no. 6, pp. 1576–1590, Oct. 2017.
[4] I. Tautkute, T. Trzcinski, and A. Bielski, "I know how you feel: Emotion recognition with facial landmarks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1878–1880.
[5] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr. 2012.
[6] J. Yan, G. Lu, X. D. Bai, H. Li, N. Sun, and R. Liang, "A novel supervised bimodal emotion recognition approach based on facial expression and body gesture," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. 101, no. 11, pp. 2003–2006, Jul. 2018.
[7] P. C. Petrantonakis and L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 5, pp. 737–746, Sep. 2011.
[8] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[9] M. Lin, Q. Chen, and S. Yan, "Network in network," Dec. 2013, *arXiv:1312.4400*. [Online]. Available: https://arxiv.org/abs/1312.4400
[10] W. B. Cannon, "The james-lange theory of emotions: A critical examination and an alternative theory," *Amer. J. Psychol.*, vol. 39, nos. 1–4, pp. 106–124, Dec. 1927.
[11] 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.
[12] G. K. Verma and U. S. Tiwary, "Affect representation and recognition in 3D continuous valence–arousal–dominance space," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2159–2183, Jan. 2017.
[13] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.

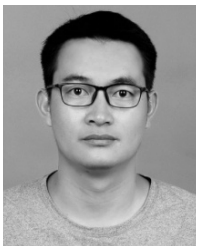
[14] N. Zhuang, Y. Zeng, L. Tong, C. Zhang, H. Zhang, and B. Yan, "Emotion recognition from EEG signals using multidimensional information in EMD domain," *Biomed Res. Int.*, vol. 2017, pp. 1–9, Aug. 2017.
[15] Z. Yin, M. Zhao, Y. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Comput. Methods Programs Biomed.*, vol. 140, pp. 93–110, Mar. 2017.
[16] J. Chen, B. Hu, Y. Wang, Y. Dai, Y. Yao, and S. Zhao, "A three-stage decision framework for multi-subject emotion recognition using physiological signals," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2016, pp. 470–474.
[17] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Technol.*, to be published.
[18] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, Apr. 2011, pp. 315–323.
[19] J. Zhang, S. Li, and Z. Yin, "Pattern classification of instantaneous mental workload using ensemble of convolutional neural networks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 14896–14901, Jul. 2017.
[20] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.
[21] 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 analysisusing physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Mar. 2012.
[22] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, p. 1161, Dec. 1980.
[23] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 881–892, Jul. 2002.
[24] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.



HAIPING HUANG received the B.Eng. and M.Eng. degrees in computer science and technology from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2002 and 2005, respectively, and the Ph.D. degree in computer application technology from Soochow University, Suzhou, China, in 2009. From May 2013 to November 2013, he was a Visiting Scholar with the School of Electronics and Computer Science, University of Southampton, Southampton, U.K. He is currently a Professor with the School of Computer Science, Nanjing University of Posts and Telecommunications. His research interests include wireless sensor networks and the Internet of Things.



ZHENCHAO HU received the B.Eng. degree from the Tongda College, Nanjing University of Posts and Telecommunications, Nanjing, China, in 2016, where he is currently pursuing the M.S. degree with the School of Computer Science. His research interests include machine learning and deep learning in electroencephalogram.



WENMING WANG received the M.S. degree from the College of Information Science and Technology, Jinan University, Guangzhou, China, in 2014. He is currently pursuing the Ph.D. degree with the School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing, China. He is currently a Lecturer with the School of Computer and Information, Anqing Normal University. His research interests include wireless sensor networks and information security.

MIN WU received the Ph.D. degree in information network from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2011. Her research interests include wireless sensor networks and the Internet of Things.