Hemispheric Asymmetry Measurement Network for Emotion Classification

Ruofan Yan, Na Lu(✉), Xu Niu, and Yuxuan Yan

School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China
lvna2009@xjtu.edu.cn

Abstract. Electroencephalogram (EEG) based emotion recognition has received considerable attention from many researchers. Methods based on deep learning have made significant progress. However, most of the existing solutions still need to use manually extracted features as the input to train the network model. Neuroscience studies suggest that emotion reveals asymmetric differences between the left and right hemispheres of the brain. Inspired by this fact, we proposed a hemispheric asymmetry measurement network (HAMNet) to learn discriminant features for emotion classification tasks. Our network is end-to-end and reaches the average accuracy of 96.45%, which achieves the state-of-the-art (SOTA) performance. Moreover, the visualization and analysis of the learned features provides a possibility for neuroscience to study the mechanism of emotion.

Keywords: Deep learning · Convolution neural network · EEG · Emotion recognition

1 Introduction

Emotion carries unique biological information, which integrates people's perceptions and behaviors. The analysis and recognition of emotions has become an interdisciplinary research topic, and received an increasing amount of applications in disease monitoring [1], task workload estimation [2] etc. Traditional emotion recognition methods based on facial expressions, voice tones, and body posture [3] can be camouflaged and thus are not reliable enough. In contrast, EEG signal is spontaneously generated by the human nervous system, which can very well reflect emotional states and is extremely hard to forge.

Deep learning approaches for EEG emotion classification have shown great potential in this field which [4, 5] can be divided into two categories: extracted features based methods and end-to-end methods. Zheng and Lu [6] extracted differential entropy (DE) [7] as EEG features, and used deep belief networks (DBN) for classification. These methods require handcrafted features and expert knowledge. On the other hand, end-to-end network has been proposed to promote the algorithm performance in real-time scenarios. In [5] a LSTM-based deep recursive neural network (RNN) was proposed to automatically learn the features from the original EEG signals. But the performance of end-to-end networks still needs to be improved. The interpretability of deep learning solutions is

W. Deng et al. (Eds.): CCBR 2022, LNCS 13628, pp. 307–314, 2022.
https://doi.org/10.1007/978-3-031-20233-9_31

also a concern and has not been solved in previous studies. To develop a more efficient end-to-end network solution for EEG emotion classification, a prior knowledge incorporated dual channel network is developed which measures the hemispheric asymmetry during emotion movement.

The main contributions of this study are as follows: (a) We proposed an end-to-end HAMNet, omitting the manual denoising and feature extraction in traditional methods, which obtains SOTA performance in emotion classification. (b) A dual channel mirror structure is applied to process EEG data to incorporate the spatial discrepancy between the two brain hemispheres. (c) We visualized the temporal and spatial EEG features, and spotted some inspiring discoveries in the field of neuroscience.

2 Related Work

In the past decade, correlations between emotion and EEG has been extensively researched and demonstrated. Neuroscience research shows that emotion-producing mechanisms are related to the asymmetry between the hemispheres of the brain [8]. The EEG signals in different frequency bands and channels have enclosed informative features for different emotions. Previous researches suggest that signals in gamma-band (roughly 30–100 Hz) are strongly related to emotional cognition [9, 10].

Duan et al. [11] implemented DE to represent emotional state representations. The combination of DE on symmetrical electrodes (Differential asymmetry, DASM; and rational asymmetry, RASM) [7] was also considered as emotional features. Besides, to smooth the feature sequence, the moving average filter and linear dynamic system (LDS) approach were applied.

Although researchers have focused on the asymmetry between the hemispheres of the brain and extracted features that contain mismatched information, these methods usually require prior knowledge, manual feature extraction and some feature smoothing methods. In [12], Niu et al. proposed a novel knowledge-driven feature component interpretable network (KFCNet) to solve the motor imagery classification problem. Band-pass linear-phase digital finite impulse response filters are applied to initialize the temporal convolution kernels. Inspired by the above methods, we designed filter banks in gamma band, and embed prior knowledge into convolution layer to train an emotion classification model.

3 Method

In this section we will introduce the architecture of HAMNet, data mirroring, and the two-step training strategy of pre-traning and fine-tuning. Architecture of HAMNet is shown in Fig. 1 where BN means batch normalization and PA means power calculation. Parameters of each layer are illustrated in the figure. Specifically, HAMNet includes a mirrored data input layer, two parallel convolution layers, a power calculation layer, a relational metric layer, a classification layer, and a softmax layer. Layers in the orange dotted box is the single channel network for pre-training, and the module in the blue dotted box is the parallel network for fine-tuning.

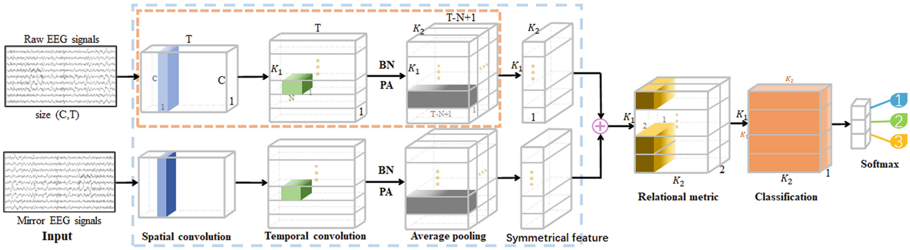

Fig. 1. Architecture of the HAMNet

3.1 Network Structure

Firstly, the raw EEG data trial $X_R \in R^{C\times T}$ is mirrored in the channel dimension, denoted as $X_M \in R^{C\times T}$. C represents the number of electrodes, and T represents the number of sampling points. They are both used as the parallel depthwise input samples. The details of mirror processing will be discussed in Sect. 3.2. The first layer of HAMNet implements spatial filtering and achieves the effect of spatial source separation through spatial convolution.

The second parallel convolutional layer shares the same parameters, which are initialized by the FIR filter bank designed by windowed inverse Fourier transform, in analogy to temporal filtering. The pass band of FIR filters ranges from 40 Hz to 76 Hz, and each filter bandwidth is 2 Hz. A total of 18 sets of filters are used.

For a discrete signal sequence $\{x_1, x_2, ...x_n\}$ in a specific frequency band, the Fourier transform of each point is denoted as $X_k, k = \{1, 2, ...n\}$, the average power spectral density (PSD) can be formulated as

$$PSD = \frac{\sum_{k=1}^{n} |X_k|^2}{n}. \tag{1}$$

Through batch normalization (BN), average power calculation (PA) and pooling layer, it is equivalent to obtain the PSD of the samples and the symmetric mirror samples. Then the obtained symmetric features are concatenated and fed to the following relational metric learning layer.

The relational metric layer performs a convolution operation on the concatenated input to automatically measure the relationship between the two symmetric samples. This layer can implicitly learn the difference between the original EEG distribution and its symmetric distribution. Therefore, the asymmetry of motion EEG distribution can be considered. The output of the relation between the symmetric samples are further fed to the classification layer. The final emotion labels are given by the classification layer and the softmax layer. Cross-entropy loss is used to update the network.

3.2 Mirror Data Processing

The input to HAMNet is composed of the original EEG sequence and its mirror version. Based on the EEG mirror processing, a symmetric version of a specific emotion signal distribution can be obtained. The relation comparison between the original distribution and its mirror distribution can shed new light on the emotion classification considering the hemispheric asymmetry of emotion EEG signal.

We symmetrically flip the original samples to get the mirror samples according to the electrode position. Electrodes are configured according to ESI NeuroScan System with 62 channels in total as shown in Fig. 2. The data in the middle column (FPZ, FZ, FCZ, CZ, CPZ, PZ, POZ, OZ) does not change, and the 27 pairs of electrode channels (FP1, F7, F3, FT7, FC3, T7, P7, C3, TP7, CP3, P3, O1, AF3, F5, F7, FC5, FC1, C5, C1, CP5, CP1, P5, P1, PO7, PO5, PO3, and CB1 of the left hemisphere, and FP2, F8, F4, FT8, FC4, T8, P8, C4, TP8, CP4, P4, O2, AF4, F6, F8, FC6, FC2, C6, C2, CP6, CP2, P6, P2, PO8, PO6, PO4, and CB2 of the right hemisphere) are mirrored respectively. The amount of input data has been doubled in this way.

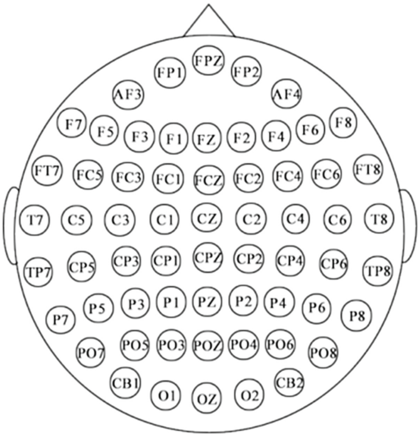

Fig. 2. ESI NeuroScan system

3.3 Pre-training and Fine-Tuning

Two stages of training is adopted. Pre-training only takes the raw EEG data as input through single-branch convolutional network (branch in the orange dotted box), and directly sends the output to the pooling and classification layers. Then fine-tuning operation feeds both the original and mirrored data into the overall HAMNet network with the two branches altogether. The parameters of pre-training result for spatial convolution are loaded as its initialization. The temporal convolutional layers are frozen in the fine-tuning stage.

4 Experiments

In this section, we firstly introduce the emotion dataset and the details of the experiment settings. Then we compare our results with baselines, visualize the learned features and discuss the results.

4.1 Dataset

SEED [6] from Shanghai Jiaotong University is a collection of EEG emotion data provided by BCMI Labs. The dataset contains EEG signals of subjects watching movie clips labeled as positive, negative, and neutral. The experiment provides 15 movie clips of four minutes each. Each experiment has 15 trials. 15 subjects each performed three experiments with a one-week interval between each experiment. There are raw data and feature extraction data in the dataset, we choose the raw data in this research. EEG signals are downsampled to 200 Hz, and cropped by second to generate the samples.

4.2 Experimental Setting

We adopt five-fold cross-validation method on each experiment[13]. There are about 680 samples per fold. We take the average classification accuracy (ACC) of 3 experiments across all subjects as a metric to evaluate model performance. In the network, K_1 is set as 16, K_2 is set as 18, and N is set as 41. Adam optimization is employed to minimize the loss function. The learning rate is set as 10^{-4} in pre-training and changes to 0.5×10^{-4} while fine-tuning. We set the first-moment decay term and the second-moment decay term to 0.9 and 0.999 respectively. 800 epochs are iterated in pre-training and 600 epochs are iterated in fine-tuning with a batch size of 100. Pytorch is used for coding. A computer server with a CPU E5–2643 of 3.40 GHz and a GPU of NVIDIA Quadro NVS315 was used.

4.3 Experiment Results

The performance comparison of our method with traditional methods and multiple deep network-based solutions were conducted on SEED dataset. The method in [10] was the first one using DE and DASM as emotion-related features, and training a SVM as the classifier. The method in [6] proposed a deep belief network (DBN) with DE features extracted from part of channels. In [14] which was published in 2022, the raw EEG signals were converted into a four-dimensional space-spectral-temporal representation, and then the 4D-ann network adopted spectral and spatial attention mechanisms to adaptively assign weights to different brain regions and frequency bands. Li et al. [15] proposed a novel bi-hemispheric discrepancy model (BiHDM) in 2020 which considered the asymmetric differences between two hemispheres and achieved excellent performance. We use ACC to compare our model with the above four baseline models, as shown in Table 1.

Table 1. The classification accuracy of different methods on SEED.

Method	SVM [10]	DBN [6]	BiHDM [15]	4D-ANN [14]	**HAMNet**
Feature	DE	DE	Raw data	Raw data	Raw data
ACC(%)	84.22	86.65	93.12	96.10	**96.45**

The ACC of our model reaches 96.45%, which achieves the state-of-the-art performance on the SEED dataset comparing with the baseline models. The ACC of the pre-training network was 93.53%, which increased by 2.92% after fine-tuning, which has well demonstrated the effectiveness of our two-step training strategy. Figure 3 shows the average accuracy on each subject. An obvious improvement with fine-tuning can be observed.

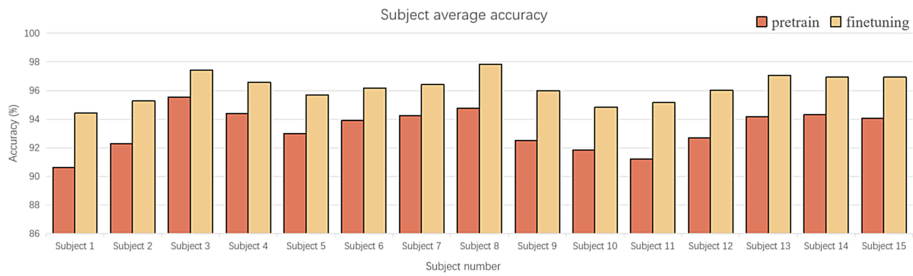

Fig. 3. ACC of HAMNet on each subject

4.4 Visualization of the Learned Temporal Features

To visualize the features learned from the time convolution layer based on FIR bandpass filters, we visualized the convolution kernels trained on Subject 4, as shown in Fig. 4. Compared with the initial filter curve, the learned filters with axis offset shows the adaptability of time domain learning, indicating effective information in narrow bands.

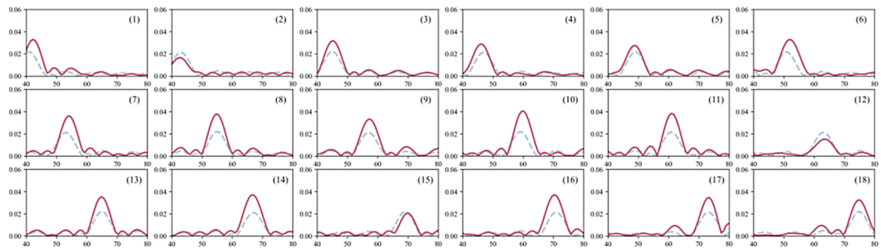

Fig. 4. Illustration of the amplitude-frequency maps of 18 filter banks after Fourier transform. The blue dotted line represents the initial filter state, and the red solid line depicts the filters learned by the temporal layer.

For some filters, the main lobe amplitude of the curve increases, and the side lobe amplitude decreases. Therefore the spectrum leakage phenomenon is reduced, indicating the performance has been automatically improved.

4.5 Visualization of the Learned Spatial Features

Figure 5 shows the brain topographic map of the kernels in the parallel spatial convolution layers of Subject 4. The number of spatial convolution kernels was set to 16. The two edge electrodes of CB1 and CB2 were removed due to insignificance. By comparing the brain maps corresponding to the original data and the mirror data, we can observe that the two features are almost symmetric, and the phenomenon of event-related synchronization and event-related desynchronization occurs. From the dark areas, it can be observed that the frontal lobe, temporal lobe and central area of the brain contain important information related to emotion recognition.

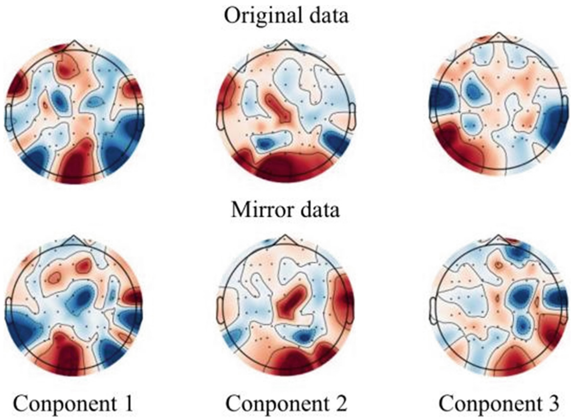

Fig. 5. Illustration of the brain topographic map. The black dots in the figure represent the positions of 60 electrodes, and the red and blue colors represent the positive and negative weights respectively. The darker the color, the greater the absolute value is.

5 Conclusion

In this paper, inspired by the research on emotion generation mechanism in neuroscience, we proposed an emotion classification method based on symmetric channel deep learning model. The proposed HAMNet is an end-to-end network, which does not need to extract features manually, saves computation, and has a better performance than the current mainstream emotion classification methods. The processing speed of emotion recognition reaches about 1 s, providing a possible solution for real-time classification. In addition, our model uses band-pass filters to embed emotional band prior knowledge into the convolution layer to initialize parameters. Based on the results, we visualized and analyzed the temporal and spatial learned features, and verified the effectiveness of asymmetric differences in the cerebral hemisphere in emotion classification tasks. Besides, the features learned by our model are interpretable after visualization, which provides the possibility for neuroscience scientists to study the mechanism of emotion.

Acknowledgement. This work is supported by National Natural Science Foundation of China grant 61876147.

References

1. Ali, M., Mosa, A.H., Machot, F.A., Kyamakya, K.: EEG-based emotion recognition approach for e-healthcare applications. In: 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), pp. 946–950. IEEE Press (2016)
2. Kothe, C.A., Makeig, S.: Estimation of task workload from EEG data: new and current tools and perspectives. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6547–6551. IEEE Press (2011)
3. Kessous, L., Castellano, G., Caridakis, G.: Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. J. Multimodal User Interf. **3**(1), 33–48 (2010)
4. Yang, Y., Wu, Q., Qiu, M., Wang, Y., Chen, X.: Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE Press (2018)
5. Alhagry, S., Fahmy, A.A., El-Khoribi, R.A.: Emotion recognition based on EEG using LSTM recurrent neural network. Emotion **8**(10), 355–358 (2017)
6. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Autonom. Mental Dev. **7**(3), 162–175 (2015)
7. Shi, L.C., Jiao, Y.Y., Lu, B.L.: Differential entropy feature for EEG-based vigilance estimation. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6627–6630. IEEE Press (2013)
8. Alarcao, S.M., Fonseca, M.J.: Emotions recognition using EEG signals: a survey. IEEE Trans. Affect. Comput. **10**(3), 374–393 (2017)
9. Keil, A., Müller, M.M., Gruber, T., et al.: Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. J. Clin. Neurophysiol. **112**(11), 2057–2068 (2001)
10. Balconi, M., Lucchiari, C.: Consciousness and arousal effects on emotional face processing as revealed by brain oscillations. a gamma band analysis. Int. J. Psychophysiol. **67**(1), 41–46 (2008)
11. Duan, R.N., Zhu, J.Y., Lu, B.L.: Differential entropy feature for EEG-based emotion classification. In: 6th International IEEE/EMBS Conference on Neural Engineering. IEEE 2013, pp. 81–84 (2013)
12. Niu, X., Lu, N., Kang, J.H., Cui, Z.Y.: Knowledge-driven feature component interpretable network for motor imagery classification. J. Neural Eng. **19**(1), 016032 (2022)
13. Zheng, W.L., Zhu, J.Y., Lu, B.L.: Identifying stable patterns over time for emotion recognition from EEG. J. IEEE Trans. Affective Comput. **10**(3), 417–429 (2017)
14. Xiao, G., Shi, M., Ye, M., et al.: 4D attention-based neural network for EEG emotion recognition. J. Cogn. Neurodyn. 2022, 1–14 (2022). https://doi.org/10.1007/s11571-021-09751-5
15. Li, Y., Wang, L., Zheng, W., et al.: A novel bi-hemispheric discrepancy model for EEG emotion recognition. IEEE Trans. Cogn. Dev. Syst. **13**(2), 354–367 (2020)