Cross-Subject emotion recognition from EEG using Convolutional Neural Networks

Xiaolong Zhong[1], Zhong Yin[1,*], Jianhua Zhang[2]

1. Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Shanghai, 200093, P. R. China.
E-mail: 182560463@st.usst.edu.cn; yinzhong@usst.edu.cn

2. OsloMet Artificial Intelligence Lab, Department of Computer Science, Oslo Metropolitan University, Oslo, N-0130, Norway.
E-mail: jianhuaz@oslomet.no

Abstract: Using electroencephalogram (EEG) signals for emotion detection has aroused widespread research concern. However, across subjects emotional recognition has become an insurmountable gap which researchers cannot step across for a long time due to the poor generalizability of features across subjects. In response to this difficulty, in this study, the moving average(MA) technology is introduced to smooth out short-term fluctuations and highlight longer-term trends or cycles. Based on the MA technology, an effective method for cross-subject emotion recognition was then developed, which designed a method of salient region extraction based on attention mechanism, with the purpose of enhancing the capability of representations generated by a network by modelling the interdependencies between the channels of its informative features. The effectiveness of our method was validated on a dataset for emotion analysis using physiological signals (DEAP) and the MAHNOB-HCI multimodal tagging database. Compared with recent similar works, the method developed in this study for emotion recognition across all subjects was found to be effective, and its accuracy was 66.23% for valence and 68.50% for arousal (DEAP) and 70.25% for valence and 73.27% for arousal (MAHNOB) on the Gamma frequency band. And benefiting from the strong representational learning capacity in the two-dimensional space, our method is efficient in emotion recognition especially on Beta and Gamma waves.

Key Words: Emotional recognition, Physiological signals, Deep learning, Classification, Machine learning, Cross-subject.

1 Introduction

Affective Computing plays a significant role in various areas of daily life, emotion recognition has been an active theme of study for the last twenty years. The purpose of emotion classification is the retrieval of the affective status of human beings in a particular point in time given a relevant data recording. In our works we will concentrate on the physiological modalities of brain waves, i.e. electroencephalography (EEG), previous cognitive psychology researches have demonstrated the inner association between affective information of the human emotional state and the electrical activity of the cerebral cortex, and reported on EEG is direct reactions caused by emotions [1]. Yet, drawback of EEG signals are noisy recordings because of artifacts affected by eye movement. respiration or measurement equipment. Due to the relatively complexity of the EEG signals, EEG signals are intrinsic difficulty comprehended in regard to emotions. Therefore, EEG-based affect detection is a field of vigorous research for which a lot of comparisons of potential algorithms are still to be done.

Previous researches in the area of cognitive psychology indicate that there are statistical significance in the way individuals sense and expressing feelings [2]. People may express distinct feelings and exist distinct physiological responses modes when exposed to an identical affective stimulus, 'individual differences' was first put forward by Picard et al. and have been raised as a matter of considerable attention [3]. Nevertheless, the classifier could not make right judgement in respect of the context of various participants when physiological responses modes from distinct participants express obvious differences upon an identical emotion [4]. Accordingly, many studies tend to focus on individual subjects. Nevertheless, subject-dependent (intra-subject) classification model meet with two serious disadvantages. On the one hand, they need the gathering of a large deal of experiment samples to sufficiently model the relevance between the affective states and the EEG signals for each participant. On the other hand, they cannot be utilized for unseen subjects since they just utilize EEG signals relevant to the specific one person alone. These two severe flaws cause the technique impractical in a number of instances. Subject-dependent techniques invariably have been questioned as they fail in regard to general applicability. Accordingly, studies into affective classification have always focused on subject independent pattern (which signifies a categorization model set up by physiological signals mixed cross-subject). Plenty of representative patterns utilize a collection of training data to set up an ordinary, subject-independent (inter-subject) model, which is shared by every subject (e.g., [5-7]). Under the circumstances, a single classification model is set up by thinking about all samples as though it were originating from the same participant, have no taking the subject's distinctiveness into consideration. It has been exhibit that the cross-subject approach always performance poorly as contrasted to the subject-dependent approach. In spite of, sometimes the well prediction accuracy acquired in certain cases, these can be considerably increased by utilizing individual classification models adapted to each subject [8-9].

* Corresponding Author: Zhong Yin. This work is sponsored by the National Natural Science Foundation of China under Grant No. 61703277 and the Shanghai Sailing Program (17YF1427000).

TABLE I

DATASET DESCRIPTION

Feature	DEAP/MAHNOB description
Number of participants	32 /24
Number of videos	40 /20
Number of EEG channels	32 /32
Rating scales	Valence and Arousal / Valence and Arousal
Rating values	1~9 /1~9
Sampling rate	128 Hz /28 Hz
Duration of experimental signals	60-s /60-s
Duration of pre-trial baseline signals	3-s /

In a word, researchers in the field have utilized two principal techniques to identify accurately the emotion. The features engineering-based techniques are included [10] and the second one Deep Learning(DL) is involved [11]. DL, i.e. convolutional neural networks(CNN), can automatically end-to-end learn the abstract features from deep scale original data, replacing feature engineering for affect recognition intentions. Therefore, in this work, the CNN is applied as one of the art classification algorithms which are trained using features based on automatically extraction. In our work, at first, we explore the applicability of subject-independent and cross-subject modeling methods in an EEG-based emotion classification case, by studying the available data in DEAP [12] and MAHNOB-HCI [13] public dataset for affective classification utilizing physiological signals. The DEAP and MAHNOB-HCI dataset include 32 and 24 participant's EEG signals and emotion tags, respectively. We extracted the Differential Entropy(DE) feature and then transform the DE feature vector to 2D-like frame, following made emotion classification on 2D-like frame using the CNN in the way like that in the field of computer vision. The content of the paper was organized as follows, In Sect. 2, we introduce and preprocess the DEAP and the MAHNOB-HCI datasets which are extensively utilized in EEG-based emotion classification domain. A particularly introduce of proposed method is presented in Sect. 3. The analysis of the results and the completion section of the study are given in Sect. 4 and Sect. 5, respectively.

2 Dataset and Preprocessing

2.1 Datasets Description and Data Acquisition

The DEAP dataset is a large open source dataset developed by a team of researchers at Queen Mary University of London, which is detecting and recording the multiple physiological signals with emotional evaluations generated by 32 volunteers under the selected video clips [12]. Specifically, it includes 32-channel EEG signals and some peripheral physiological signals. Each participant needs to watch 40 trials of music videos with different emotional stimuli for about one-minute for each video. The EEG signals and peripheral physiological signals of volunteers are recorded simultaneously during viewing. Then the subjects rated the videos on a scale of 1-9 in terms of Arousal, Valence, Liking, Dominance and Familiarity. In this article, we only focus on two-dimensional emotional model which the two dimensions are arousal (ranging from weak to strong) and valence (ranging from negative to positive). In this paper, we only applied the 32-channel EEG data in DEAP dataset which had been pre-processed by down sampling to 128 Hz, the data in the 4–45 Hz frequency bandwidth is preserved by band-pass filtering, and common average referencing and ocular artifacts removing by blind source separation algorithms.

The MAHNOB-HCI dataset is another recent database contained EEG, video, audio, gaze, and peripheral physiological recordings of 30 participants [13]. In this dataset, each participant watched 20 clips extracted from Hollywood movies and video websites, such as youtube.com and blip.tv. The stimulus videos ranged in duration from 35 to 117 s. After watching each stimulus, the participants used self-assessment manikins (SAMs) to rate their perceived arousal and valence on a discrete scale of 1 to 9. We then divided these selections into a high class (ratings 6–9) and a low class (ratings 1–5). Here, only the 24 participants for whom all 20 trials are available were used. Same as DEAP dataset, The MAHNOB-HCI dataset had been pre-processed by down sampling to 128 Hz. In the process of collecting EEG signals, the signals that the scalp electrodes can measure are actually potential differences. The potential difference here is the difference between the active electrode and the reference electrode. In order to obtain accurate values of these parameters, we need to use the least active electrode point as a reference in order to obtain the most ideal raw data. So it is unavoidable to set the reference electrode in the scalp recording measurement. We often place the reference electrode at the top of the head, between the electrodes CPz and Pz, which causes a large area of important brain signals to be affected by noise. To solve this problem, we introduced a common average reference processing technique. This technology actually uses the average potential of all potentials as a reference value, and the difference between the potential of each electrode and this reference value is used as the voltage value of the electrode. An important step in the pre-processing of EEG signals is to reduce noise and remove artifacts. For this purpose, we use high-pass filters to perform artifact processing. Finally, we selected the first 65 seconds of data and removed the first 5 seconds of data. TABLE I shows a summary of the DEAP and MAHNOB-HCI dataset.

2.2 Feature Extracting

Recently, a great quantity of different entropy estimators have been applied to measure the complexity of continuous EEG signals. Shi et al. proposed Differential entropy (DE) for EEG-based vigilance estimation and used it to measure the complexity of EEG signals [14]. In this study, differential entropy(DE) is presented to characterize the level of vigilance, as a complexity quantify for EEG signals, which has been proven to be appropriate for emotion analyzing in previous studies. Its calculation formula can be defined as,

$$h(X) = -\int_S f(x)\log(f(x))dx \tag{1}$$

where S is the support set of the random variable. X is a continuous random variable. f(x) is the probability density function of X. For the series $X \sim \phi(x) = (1/\sqrt{2\pi\sigma^2}) \times e^{-x^2/2\sigma^2}$. Then calculating the differential entropy expressed as:

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

According to the summary made by Zhang et al. [15], five frequency bands rhythm of EEG signals, such as Delta(0.5Hz~3Hz), Theta (4Hz~7Hz), Alpha (8Hz~13Hz), Beta (14Hz~30Hz) and Gamma (above 31Hz~45Hz) were closely associated with emotional and other psychological activities. To confirm our point of view and study the influence of distinct frequency patterns on mood, we utilized a Butterworth filter to decompose the raw data into 4 frequency bands $(\theta,\ \alpha,\ \beta,\ \gamma)$.

2.3 Data Preprocessing

According to the summary made by Yang et al. [42], the DE features of baseline EEG signals can help to improve emotion recognition accuracy significantly. Thence. we use baseline signals to measure the differences between signals which are recorded while subjects are under stimulation and baseline signals. For each subject, each channel of all 32 channels is first filtered in $(\theta,\ \alpha,\ \beta,\ \gamma)$ frequency bands from 4 Hz to 45 Hz by Butterworth filter, and a 128-point segment window with nonoverlap is used to divide each frequency band signal to pre-trial baseline data of 3-s length. Apply the feature extraction method defined earlier ,the DE feature is calculated over each window 128-point data for all of 3 window data and transform each of them into 4 DE feature vector $X_{DE}^{base(\theta,\ \alpha,\ \beta,\ \gamma)} \in R^{32}$. The next step is concatenating all of these 4 DE feature vector $X_{DE}^{base(\theta,\ \alpha,\ \beta,\ \gamma)} \in R^{32}$ according to the order of $(\theta,\ \alpha,\ \beta,\ \gamma)$ frequency bands into a big DE feature vector $X_{DE}^{base} \in R^{128}$ .Finally, the mean DE feature value of these three window data are computed to represent the DE feature of pre-trial baseline signals. This process can be showed as:

$$X_{DE}^{base} \leftarrow \frac{\sum_{i=1}^{i=3} X_{DE}^{base(\theta,\ \alpha,\ \beta,\ \gamma)}(i)}{3} \quad (3)$$

$X_{DE}^{base(\theta,\ \alpha,\ \beta,\ \gamma)}(i)$ represents the DE feature are calculated over each window 128-point data for all of 3 window data, respectively. X_{DE}^{base} denotes the mean DE feature value of these three window data.

With regard to the 60-s EEG experimental signals whilesubjects are under stimulation. Wang et al. [16] reported that time window with size 1 s is suitable for emotion recognition and then the 7680-point data of each channel is divided 60 epochs, which each epoch the length of data is 128-point. The whole number of EEG epochs from 40 trials of each participant was 40∗60 = 2400, and the dimension of this EEG experimental signals after segmented was 128(time points) ∗32(channels)∗4(bands)∗2400 (epochs). Then the DE feature is calculated, we can get 4 DE feature vector $X_{\mathrm{DE}}^{\exp er(\theta,\ \alpha,\ \beta,\ \gamma)} \in R^{32}$ of each epoch. At the same , the last step is concatenating all of these 4 DE feature vector $X_{\mathrm{DE}}^{\exp er(\theta,\ \alpha,\ \beta,\ \gamma)} \in R^{32}$ according to the order of $(\theta,\ \alpha,\ \beta,\ \gamma)$ frequency bands into a big DE feature vector $X_{DE}^{\exp er} \in R^{128}$.

EEG signals are non-stationary random time series signals. As same, in order to maintain time information hidden in the 2-D maps at a time interval of 1-s, we introduced the moving average(MA) technology to smooth out short-term fluctuations and highlight longer-term trends or cycles. Processing time series data after extracting differential entropy features according to the basic idea of the MA, we consider that the information at the current moment is related to the information at all previous moments for each second of data per video. This process can be showed as:

$$X_{DE}^{\exp er}(t) = \frac{X_{DE}^{\exp er}(t) + X_{DE}^{\exp er}(t-1) + \ldots + X_{DE}^{\exp er}(1)}{t} = \frac{\sum_{i=1}^{t} X_{DE}^{\exp er}(i)}{t} \quad (4)$$

Where, $X_{DE}^{\exp er}(i)$ denotes the DE feature value at time i of each video. The value of t is an integer value ranges from 1 to 60. Finally, the DE deviation between experimental EEG and pre-trial baseline EEG is calculated to stand for the emotional state feature of this epoch. This process can be showed as:

$$X_{DE} \leftarrow X_{DE}^{\exp er} - X_{DE}^{base} \quad (5)$$

Where $X_{DE}^{\exp er}$ denotes the DE feature are calculated over each window 128-point data for all of 60 window data, X_{DE} represents the DE deviation between experimental EEG and pre-trial baseline EEG. Based on the emotional rating grade of each stimulus video in the range of 1-9 in valence and arousal domain, the median 5 was simply adopted as the threshold to divide the rating grade into two classes: less than or equal to 5 labeled with 0 meaning low valence/arousal, more than 5 labeled with 1 meaning high valence/arousal.

In order to keep the spatial structure information among multiple adjacent channels, We transform 1D DE feature vector $X_{DE}^{\exp er}$ to 2D plane $(h \times l)$. Where h and l is the no more than the number of the vertical and horizontal used electrodes. For our DEAP dataset, h=l=9. Hence, the corresponding 2D plane of feature vector $X_{DE}^{\exp er}$ of frequency bands i $(i \in (\theta,\ \alpha,\ \beta,\ \gamma))$ is denoted as $\mathrm{FM}^{i} \in R^{h \times l}$. Zeroes are used to fill the DEs from channels that are unused in DEAP dataset. As yet, we have acquired four 2D planes for each EEG epoch. This process is depicted in Figs. 1.

3 Classification Using a Convolution Neural Network

Deep learning can automatically end-to-end learn the abstract features from deep scale original data, replacing feature engineering for affect recognition intentions. The CNN is one of the most typical and extensively employed classifier for deep learning. Our CNN classifiers are based on the LeNet-5 framework, a well-known CNN classifier

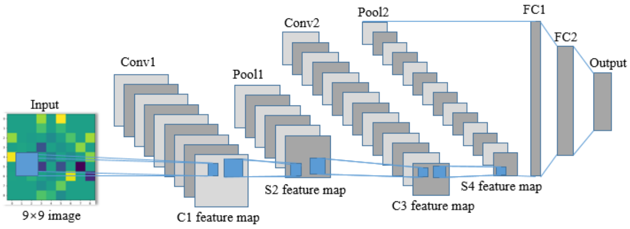

Fig. 1. The basic architecture of the LeNet-5 network

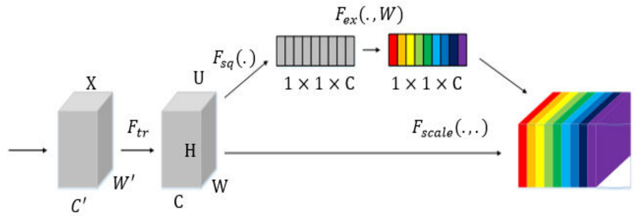

Fig. 2. The structure of the SE building block

first presented by LeCun et al. in [17] is the essential model for all kinds of CNN applications such as image classification. The LeNet-5 basic architecture is as shown in the Fig. 1.

Dataset for our experiment, we use the swish activation (Ramachandran et al., 2017) [18] function instead of the ReLU activation function. Swish is a smooth, non-monotonic function defined as $f(x)=1/(1+e^{-x})$.

The classification performance is closely related to the electrodes at different spatial positions. Therefore, we introduce the Squeeze-excitation (SE) block (Hu et al., 2018) [19], with the purpose of enhancing the capability of representations generated by a network by modelling the interdependencies between the channels of its informative features. The structure of the SE building block is depicted in Fig. 2.

At any given transformation $F_{tr}: X \rightarrow U, X \in \mathbb{R}^{W' \times H' \times C'}, U \in \mathbb{R}^{W \times H \times C}$,e.g. a convolution, we would be able to produce a relevant SE block to fulfil feature recalibration. Firstly, the features U are performed a squeeze action, which generates a channel descriptor by fusing feature maps span their space dimensions $(H \times W)$. The intention of this descriptor is to generate an embedding of the global spatial information of channel-wise feature responses, so that the information from the global receptive field of the network can be used by its lower layer. The information aggregated in the squeeze operation is followed by an excitation action. The excitation operation is in the form of a simple self-gating mechanism, which embeds as input and generates a set of modulation weights for each channel. Applying these weights to the feature map U to produce the output of the SE block, which can be fed immediately to the network subsequent layers.

For our model, name it SE_CNN. The first convolutional layer filters the 9×9×1 input 2D array corresponding to each frequency band of $(\theta, \ \alpha, \ \beta, \ \gamma)$ with 100 kernels of 3 rows and 3 columns. The next layer is a MaxPooling layer with pooling size 2×2 with a stride of 2 pixels. Then goes on a layer is another convolutional layer which again with 100 filters and 3*3 size kernel takes as input the output of the MaxPooling layer. Followed by is a MaxPooling layer with pooling size 2×2 with a stride of 2 pixels. The first fully connected layer has 120 neurons. And the second fully connected layer has 120 neurons. Finally, our final fully connected dense neural layer has 2 neurons.

Due to the limited number of samples, we consider that even simple convolutional networks can be overfitting. Therefore, we introduce some main techniques into the SE_CNN networks in which we reduce overfitting. We introduce L2 regularization technique into every convolution layer and the dropout technology is introduced into the full connection layer. To improve network performance, we use the batch normalization (BN) technology and the SE block between the convolutional layer and the MaxPooling layer.

4 Experiments and Results

4.1 Experiments Setup

After the data processing of the DEAP and the MAHNOB-HCI datasets, the complex and low signal-to-noise raw signals was simplified, and the following was obtained:

$$\begin{aligned} M_1 &= 76800(trials)\times 9(rows)\times 9(cols)\times 1(channels) \\ M_2 &= 28800(trials)\times 9(rows)\times 9(cols)\times 1(channels) \end{aligned} \tag{6}$$

Then, M1 and M2 were inputted into the SE_CNN network. Each dataset was normalized previous to the utilize of SE_CNN modeling for emotion recognition, which contributed to ameliorate the convergence performance and classification accuracy of the model. In the SE_CNN program, we have applied a leave-one-response-out cross-validation technique, in which a single participant obtained from the entire readings is used as the test participant while the remaining readings is used in the training process.

We implemented the SE_CNN with Keras framework libraries in Python and trained it on a Tesla T4 GPU based on Google cloud platform. The truncated normal distribution function was used to initialize the weight of kernels and the Adam optimizer was adopted to minimize the crossentropy loss function, the initial learning rate was 1.0e-05 for the SE_CNN. The keep probability of dropout operation was 0.6. L2 regularization was added to avoid overfitting and improve generalization capability, the penalty strength of L2 was 0.6 for the SE_CNN. For our experiments we use 100 epochs and train our model using batches of 32 experiments each. Finally, we fine-tuned the network to get the final classification result.

4.2 Effectiveness of different frequency band for emotion recognition

For the sake of evaluate the validity of different frequency band for emotion recognition, in our research, we train a SE_CNN model for each 2D plane produced by frequency band of $(\theta, \ \alpha, \ \beta, \ \gamma)$. We first employ only one single frequency band for classification each time. Table 4 present a comparison of performance between the valence and the arousal classification recognition accuracies of the SE_CNN and those of CNN model with various cases (e.g. different activation functions, whether the CNN architecture

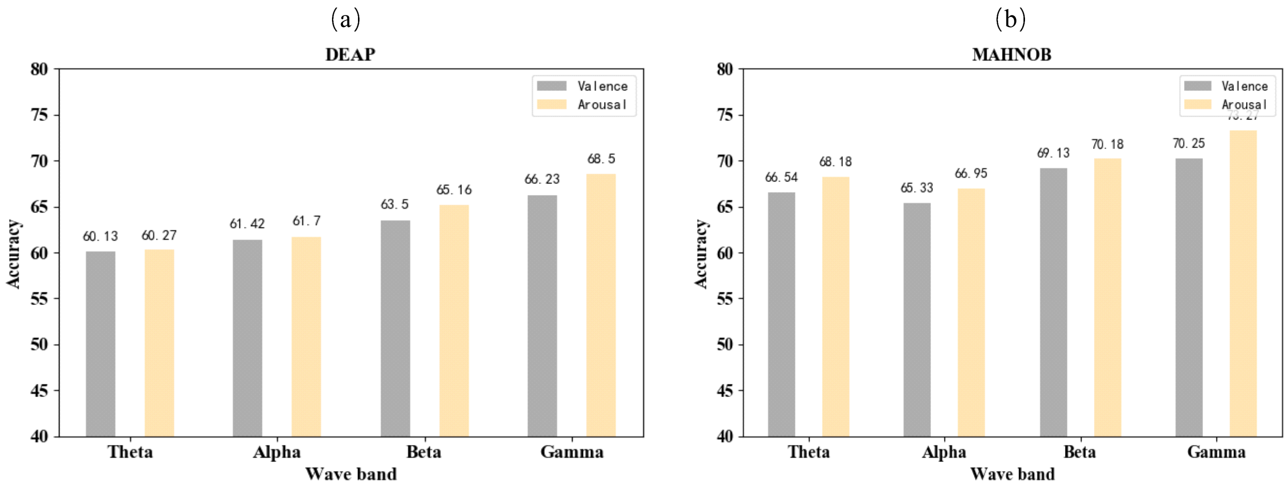

Fig. 3. Subgraphs (a) and (b) represent the emotions classification accuracy applied in the different research frequency band of EEG-based deep learning for binary classification into low/high valence and arousal for cross subject emotion recognition on the DEAP and the MAHNOB databases, respectively.

introduce SE block) using the DEAP and the MAHNOB for each single frequency band. With MA means the data has been processed by MA process and other operations are similar. For the conformity of the analysis of the two datasets, only cross-subject affective classification was carried out for the valence and arousal classification.

Fig. 3. show the classification accuracy of SE_CNN on four frequency bands on the DEAP and the MAHNOB database, respectively. On Gamma, its accuracy was 66.23% for valence and 68.50% for arousal (DEAP) and 70.25% for valence and 73.27% for arousal (MAHNOB). On Beta, its accuracy was 63.50% for valence and 65.16% for arousal (DEAP) and 69.13% for valence and 70.18% for arousal (MAHNOB). Regardless of the kind of classifier used, Beta frequency wave and Gamma frequency wave are superior to the others for classification. Therefore, it shows that the replacement of the RELU activation function with the switch activation function in the CNN-based the MA processing technology architecture and the introduction of SE-block can help classify emotion.

4.3 Results comparison with other classification methods

We also compare our model with other different approaches on DEAP dataset. Li et al. adopted the SVM approach and the “leave-one-subject-out” verification strategy to evaluate recognition performance. Using automatic feature selection methods, the highest mean recognition accuracy of 59.06% on the DEAP dataset [20]. Pandey et al. proposed a subject independent emotion recognition technique from EEG signals using Variational Mode Decomposition (VMD) as a feature extraction technique and Deep Neural Network as the classifier, the highest mean recognition accuracy of 61.25% for valence and 62.50% for arousal [21]. Pandey et al. proposed a subject independent emotion recognition technique from EEG signals using Empirical Mode Decomposition (EMD) as a feature extraction technique and the SVM as the classifier, the highest mean recognition accuracy of 59.22% for valence and 55.70% for arousal [21]. Rayatdoost et al. adopted a subject independent emotion recognition technique from EEG signals using automatic feature selection methods and the RF as the classifier, the highest mean recognition accuracy of 58.4% for valence and 57.6% for arousal [22]. In [23], two deep learning models based on the deep and convolutional neural networks method wasused for classifying low/high valence and arousal based on the feature extraction from EEG raw data from the DEAP dataset, with State of the Art classification accuracies of 81.40% and 73.36%. But we think there are problems with the experiments done in this document, they extracted experiment and participant number features for classification. The use of experiment number features as classification features is meaningless because the trained model has no generalization ability.

The contrastive results on the DEAP dataset are shown in Fig. 4. The comparison shows the effectiveness of our model. The proposed model outperforms the EEG-based only approaches significantly, which is about 7% points higher than Li [2018] for valence, 5% points higher for valence and 6% points higher for arousal than Pandey [2019] ， 7% points higher for valence and 13% points higher for valence than Pandey [2019] and 8% points higher

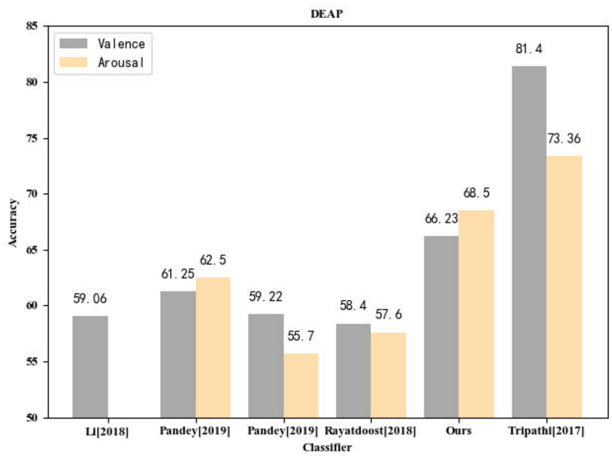

Fig. 4. Performance comparison between relevant approaches.

for valence and 11% points higher for valence than Rayatdoost [2018] . While compared to the Tripathi [2017] approaches, the recognition accuracy of their method is much higher than our method. But, we think there are problems with the experiments done in this document, they extracted experiment and participant number features for classification. The use of experiment number features as classification features is meaningless because the trained model has no generalization ability.

5 Conclusions

In our work, we employ various models to recognize two-category emotional states. In comparison to recent similar classification methods, the method proposed in this research is valid for cross-subject sentiment classification. Benefitting from the strong abstract representational capabilities, deep models are superior to the shallow models. We identified the high-frequency brain waves, i.e., Beta and Gamma waves, are more geared to sentiment classification. And the SE-block can enhance the capability of representations generated by a network by modelling the interdependencies between the channels of its informative features.

References

[1] Kim M K, Kim M, Oh E, et al. A review on the computational methods for emotional state estimation from the human EEG[J]. Computational and mathematical methods in medicine, 2013, 2013.

[2] Gross J J, John O P. Revealing feelings: facets of emotional expressivity in self-reports, peer ratings, and behavior[J]. Journal of personality and social psychology, 1997, 72(2): 435.

[3] Picard R W, Vyzas E, Healey J. Toward machine emotional intelligence: Analysis of affective physiological state[J]. IEEE transactions on pattern analysis and machine intelligence, 2001, 23(10): 1175-1191.

[4] Chen J, Hu B, Wang Y, et al. Subject-independent emotion recognition based on physiological signals: a three-stage decision method[J]. BMC medical informatics and decision making, 2017, 17(3): 167.

[5] Wang X W, Nie D, Lu B L. Emotional state classification from EEG data using machine learning approach[J]. Neurocomputing, 2014, 129: 94-106.

[6] Hadjidimitriou S, Charisis V, Hadjileontiadis L. Towards a practical subject-independent affective state recognition based on time-domain EEG feature extraction[J]. International Journal of Heritage in the Digital Era, 2015, 4(2): 165-177.

[7] Li X, Song D, Zhang P, et al. Exploring EEG features in cross-subject emotion recognition[J]. Frontiers in neuroscience, 2018, 12: 162.

[8] Salmeron-Majadas S, Arevalillo-Herráez M, Santos O C, et al. Filtering of spontaneous and low intensity emotions in educational contexts[C]//International Conference on Artificial Intelligence in Education. Springer, Cham, 2015: 429-438.

[9] Ayesh A, Arevalillo-Herráez M, Ferri F J. Cognitive reasoning and inferences through psychologically based personalised modelling of emotions using associative classifiers[C]//2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing. IEEE, 2014: 67-72.

[10] Ali M, Mosa A H, Al Machot F, et al. EEG-based emotion recognition approach for e-healthcare applications[C]//2016 eighth international conference on ubiquitous and future networks (ICUFN). IEEE, 2016: 946-950.

[11] Kim Y, Lee H, Provost E M. Deep learning for robust feature generation in audiovisual emotion recognition[C]//2013 IEEE international conference on acoustics, speech and signal processing. IEEE, 2013: 3687-3691.

[12] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis; using physiological signals[J]. IEEE transactions on affective computing, 2011, 3(1): 18-31.

[13] Soleymani M, Lichtenauer J, Pun T, et al. A multimodal database for affect recognition and implicit tagging[J]. IEEE transactions on affective computing, 2011, 3(1): 42-55.

[14] Shi L C, Jiao Y Y, Lu B L. Differential entropy feature for EEG-based vigilance estimation[C]//2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2013: 6627-6630.

[15] Zhang X, Yao L, Kanhere S S, et al. MindID: Person identification from brain waves through attention-based recurrent neural network[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(3): 1-23.

[16] Wang X W, Nie D, Lu B L. Emotional state classification from EEG data using machine learning approach[J]. Neurocomputing, 2014, 129: 94-106.

[17] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[18] Ramachandran P, Zoph B, Le Q V. Swish: a self-gated activation function[J]. arXiv preprint arXiv:1710.05941, 2017, 7.

[19] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[20] Li X, Song D, Zhang P, et al. Exploring EEG features in cross-subject emotion recognition[J]. Frontiers in neuroscience, 2018, 12: 162.

[21] Yang F, Zhao X, Jiang W, et al. Multi-method Fusion of Cross-Subject Emotion Recognition Based on High-Dimensional EEG Features[J]. Frontiers in computational neuroscience, 2019, 13.

[22] Rayatdoost S, Soleymani M. Cross-corpus eeg-based emotion recognition[C]//2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2018: 1-6.

[23] Tripathi S, Acharya S, Sharma R D, et al. Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset[C]//Twenty-Ninth IAAI Conference. 2017.