Neuropsychologia 146 (2020) 107506

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Emotion recognition with convolutional neural network and EEG-based EFDMs☆

Fei Wang [a,*], Shichao Wu [a], Weiwei Zhang [a], Zongfeng Xu [b], Yahui Zhang [b], Chengdong Wu [a], Sonya Coleman [c]

[a] Faculty of Robot Science and Engineering, Northeastern University, Shenyang, 110169, China
[b] College of Information Science and Engineering, Northeastern University, Shenyang, 110819, China
[c] Intelligent Systems Research Centre, Ulster University, Londonderry, United Kingdom

ARTICLE INFO

Index Terms:
Emotion recognition
Electroencephalogram
Convolutional neural network
Electrode-frequency distribution maps
Gradient-weighted class activation mapping

ABSTRACT

Electroencephalogram (EEG), as a direct response to brain activity, can be used to detect mental states and physical conditions. Among various EEG-based emotion recognition studies, due to the non-linear, non-stationary and the individual difference of EEG signals, traditional recognition methods still have the disadvantages of complicated feature extraction and low recognition rates. Thus, this paper first proposes a novel concept of electrode-frequency distribution maps (EFDMs) with short-time Fourier transform (STFT). Residual block based deep convolutional neural network (CNN) is proposed for automatic feature extraction and emotion classification with EFDMs. Aim at the shortcomings of the small amount of EEG samples and the challenge of differences in individual emotions, which makes it difficult to construct a universal model, this paper proposes a cross-datasets emotion recognition method of deep model transfer learning. Experiments carried out on two publicly available datasets. The proposed method achieved an average classification score of 90.59% based on a short length of EEG data on SEED, which is 4.51% higher than the baseline method. Then, the pre-trained model was applied to DEAP through deep model transfer learning with a few samples, resulted an average accuracy of 82.84%. Finally, this paper adopts the gradient weighted class activation mapping (Grad-CAM) to get a glimpse of what features the CNN has learned during training from EFDMs and concludes that the high frequency bands are more favorable for emotion recognition.

1. Introduction

Human emotion plays an important role in the process of affective computing and human machine interaction (HMI) (Preethi et al., 2014). Moreover, many mental health issues are reported to be relevant to emotions, such as depression, attention deficit (Alkaysi et al., 2017), (Bocharov et al., 2017). Much information such as posture, facial expression, speech, skin responses, brain waves and heart rate are commonly used for emotion recognition (Liberati et al., 2015). There is some evidence that electroencephalogram (EEG) based methods are more reliable, demonstrating high accuracy and objective evaluation compared with other external features (Zheng et al., 2015). Although EEG has a poor spatial resolution and requires many sensors placed on the scalp, it provides an excellent temporal resolution, allowing researchers to study phase changes related to emotion. EEG is non-invasive, fast, and low-cost compared with other psychophysiological signals (Niemic, 2004). Various psychophysiological studies have demonstrated the relationship between human emotions and EEG signals (Sammler et al., 2007), (Mathersul et al., 2008), (Knyazev et al., 2010). With the wide implementation of machine learning methods in the field of emotion recognition, many remarkable results have been achieved. Sebe et al. summarized the studies of emotion recognition with single modality, described the challenging problem of multimodal emotion recognition (Sebe et al., 2005). Alarcao et al. presented a comprehensive overview of the existing works on EEG emotion recognition in recent years (Alarcao and Fonseca, 2019). A number of EEG datasets have been built with various emotions or scored in one continuous emotion space. However, the problem of modeling and

☆ This work was supported in part by the National Natural Science Foundation of China under Grant 61973065, Fundamental Research Funds for the Central Universities of China under Grant N172608005 and N182612002, Liaoning Provincial Natural Science Foundation of China under Grant 20180520007.

* Corresponding author.
E-mail address: wangfei@mail.neu.edu.cn (F. Wang).

https://doi.org/10.1016/j.neuropsychologia.2020.107506
Received 25 February 2020; Received in revised form 23 May 2020; Accepted 26 May 2020
Available online 1 June 2020
0028-3932/© 2020 Elsevier Ltd. All rights reserved.

detecting human emotions has not been fully investigated (Mühl et al., 2014). EEG based emotion recognition is still very challenging for the fuzzy boundary between emotion categories as well as the difference of EEG signals from kinds of subjects.

Various feature extraction, selection and classification methods have been proposed for EEG based emotion recognition (Zhuang et al., 2017). Friston modeled the brain as a large number of interacting nonlinear dynamical systems and emphasized the labile nature of normal brain dynamics (Friston, 2001). Several studies have suggested that the human brain can be considered as a chaotic system, i.e., a nonlinear system that exhibits particular sensitivity to initial conditions (Ezzatdoost et al., 2020). The nonlinear interaction between brain regions may reflect the unstable nature of brain dynamics. Thus, for this unstable and nonlinear EEG signals, a nonlinear analysis method such as sample entropy (Jie et al., 2014) is more appropriate than that of linear methods, which ignores information associated with nonlinear dynamics of the human brain. Time-frequency analysis methods are based on the spectrum of EEG signals. Power spectral density and differential entropy of sub-band EEG rhythms are commonly used as emotional features (Duan et al., 2013), (Ang et al., 2017). In the last decade, a large number of studies have demonstrated that the higher frequency rhythms such as beta and gamma outperform lower rhythms, i.e., delta and theta, for emotion recognition. Traditional recognition methods are mainly based on the combination of hand-crafted features and shallow models like *k*-nearest neighbor (KNN), support vector machines (SVM) and belief networks (BN) (Duan et al., 2012), (Sohaib et al., 2013), (Zubair and Yoon, 2018). However, EEG signals have a low signal-to-noise ratio (SNR) and are often mixed with noise generated in the process of data collection. Another much more challenging problem is that, unlike image or speech signals, EEG signals are temporally asymmetry and nonstationary, which has created significant difficulties for data preprocessing to obtain clean data for feature extraction. The nonstationary means the properties (mean, variance and covariance) of EEG signals varied with time partly or totally. Temporally asymmetric refers to the fact that the corresponding activation lobes and activation degree are different under various cognitive activities. Pauls has identified these two nonlinearity properties of EEG (Palus, 1996). Moreover, traditional manual feature extraction and selection methods are crucial to an affective model and require specific domain knowledge. The commonly used dimensionality reduction techniques for EEG signal analysis are principal component analysis (PCA) and Fisher projection. In general, the cost of these traditional feature selection methods increases quadratically with respect to the number of features that is included (Dash and Liu, 1997).

As a form of representation learning, deep learning can extract features automatically through model training (Zhang et al., 2018). Apart from the successful implementation in image and speech domains, deep learning has been introduced to physiological signals, such as EEG emotion recognition in recent years. Zheng et al. trained an efficient deep belief network (DBN) to classify three emotional states (negative, neutral, and positive) by extracting differential entropy (DE) of different frequency bands and achieved an average recognition of 86.65% (Zheng and Lu, 2015). As a typical deep neural network model, convolutional neural network (CNN) has achieved great progress in computer vision, image processing and speech recognition (Hatcher and Yu, 2018). Yanagimoto et al. built a CNN to recognize the emotional valence of DEAP and analyze various emotions with EEG (Yanagimoto and Sugimoto, 2016). Wen et al. rearranged the original EEG signals through Pearson Correlation Coefficients and fed them into the end-to-end CNN based model for the purposes of reducing the manual effort on features, which achieved an accuracy of 77.98% for Valence and 72.98% for Arousal on DEAP, respectively (Wen et al., 2017).

The mainly used feature extraction methods of EEG signals can mainly be divided into time domain, frequency domain, and time-frequency domain (Wang, 2011), (Chuang et al., 2014), (Li et al., 2017). Frequency analysis transformed the EEG signals into frequency domain for further feature extraction. Since many studies demonstrated that the frequency domain features have higher distinguishability, we proposed the novel concept of electrode-frequency distribution maps (EFDMs) firstly. With the successful application of CNN in speech recognition (Abdelhamid et al., 2014), we build a deep neural network for emotion recognition based on EFDMs. The EFDMs of EEG signals can be regarded as grayscale images. Therefore, with proposed EFDMs, we realized the purpose of constructing emotion recognition model based on CNN.

At present, studies on EEG emotion recognition mainly focus on subject-dependent emotion recognition tasks. For engineering applications, it's obviously impossible to collect a huge amount of subjects' EEG signals in advance to build a universal emotion recognition model to identify the emotions of every person. Therefore, how to realize the subject- dependent pattern classification is one tough issue in the practical application of emotion recognition. Traditional emotion recognition models are usually established for a specific task on a small dataset, thus they often fail to achieve good effect under new tasks, due to the possible differences in stimulus paradigm, subjects and EEG acquisition equipment. In addition, the learning process of deep neural networks is vitally important, and generally requires a large amount of labeled data, while the acquisition of EEG signals is not as easy as that of image, speech and text signals. Accordingly, how to achieve a highly effective classifier through the training process based on a small number of labeled samples is another issue that needs to be considered. In this paper, transfer learning is employed to solve those problems highlighted above. Among various transfer learning methods, one is to reuse the pre-trained model from source domain to target domain, dependent on the similarities of data, tasks and models between them (Pan and Yang, 2010). Transfer learning accelerates the training process by transferring the pre-trained model parameters to a new domain task. Since Yosinski et al. published an article on how to transfer the features in deep neural network, it has achieved a rapid development in the field of image processing (Yosinski et al., 2014).

We firstly proposed a novel concept of EFDMs based on multiple channel EEG signals. Then four residual blocks based CNN was built for automatic feature extraction and emotion classification with EFDMs as input. We mainly set up two experiments in this paper. One is to evaluate the effectiveness of the proposed method on SEED. Second, based on the deep model transfer learning strategy, the pre-rained CNN from the first experiment is applied to DEAP for the cross-datasets emotion recognition. At the last, we have given more neuroscience interpretation by revealing the key EEG electrodes and frequency bands corresponding to each emotion category based on the attention mechanism of deep neural network and the proposed EFDMs.

2. Methods

In this section, we will detail the general framework of the EFDMs based CNN for emotion recognition, including a short description of short-time Fourier transform (STFT), the structure and key parameters of the proposed CNN as well as a brief introduction to Grad-CAM.

2.1. Short-time Fourier Transform

Fourier transform (FT) is often used to analyze the frequency features of time series. It provides the frequency information averaged over the entire signal time interval, and does not know the time when each frequency component appears. Therefore, the spectrum of two signals with large difference in time domain may be the same in frequency domain. That is to say, FT assumes that the time sequences are stationary, which is a false hypothesis for EEG signals apparently. For these nonstationary signal analyze, the time series should be cut into minor segments, and within each segment, the signal waves can be approximately considered as stationary signals used for FT. The idea is called STFT. It is a sequence of Fourier transforms of a windowed signal, used to analyze how the

frequency content of a nonstationary signal changes. Provides the time-localized frequency information for situations in which frequency components of a signal vary over time. The calculation of STFT is defined as:

$$X(\tau, w) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-jwt}dt \quad (1)$$

where $x(t)$ represents original signal and $\omega(t)$ indicates the window function such as the Hanning window as shown in (2). It's a linear combination of modulated rectangular windows, and usually emerges in applications that require low aliasing and less spectrum leakage.

$$w(n) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \quad (2)$$

in which n represents the window length and N is the sampling number.

As for discrete time series, the data could be broken up into segments. Each segment is Fourier transformed, and the complex result is added to a matrix, which records magnitude and phase for each point in time and frequency. The calculation of STFT for discrete time series can be expressed as:

$$X(m, w) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-jwn} \quad (3)$$

where $x[n]$ is a time series and $\omega[n]$ is window function. With a normalization of $X(m,w)$, we got the corresponding EFDMs.

2.2. The proposed model for EEG emotion recognition

In image processing, convolution operations can effectively filter image information, and CNN make use of these characteristics to achieve automatic feature extraction from images. In order to apply the CNN for automatic feature extraction and pattern classification in EEG-based emotion recognition, we proposed a novel concept of EFDMs based on multiple channel EEG signals. These EFDMs can be treated as grayscale images to apply two dimensional convolution operation.

A CNN with four residual blocks is proposed for EEG emotion recognition with EFDMs as input. The general network structure is shown in Fig. 1. The network consists of 1 convolution layer, 4 residual blocks, 4 max pooling layers, 2 fully connected layers, and finally the Softmax layer. The network also includes 5 batch normalization and 4 dropout layers for over-fitting consideration. The size of the max pooling window is 2 × 2, and the window slide step is 2. In addition, all intermediate layers use the Rectified Linear Unit (ReLU) as an activation function. The detailed structure of the residual block is shown in the dashed box. The size of the convolution kernel in the residual block is 3, 3, 1, the sliding step is 1, and one batch normalization layer is included after each convolution layer.

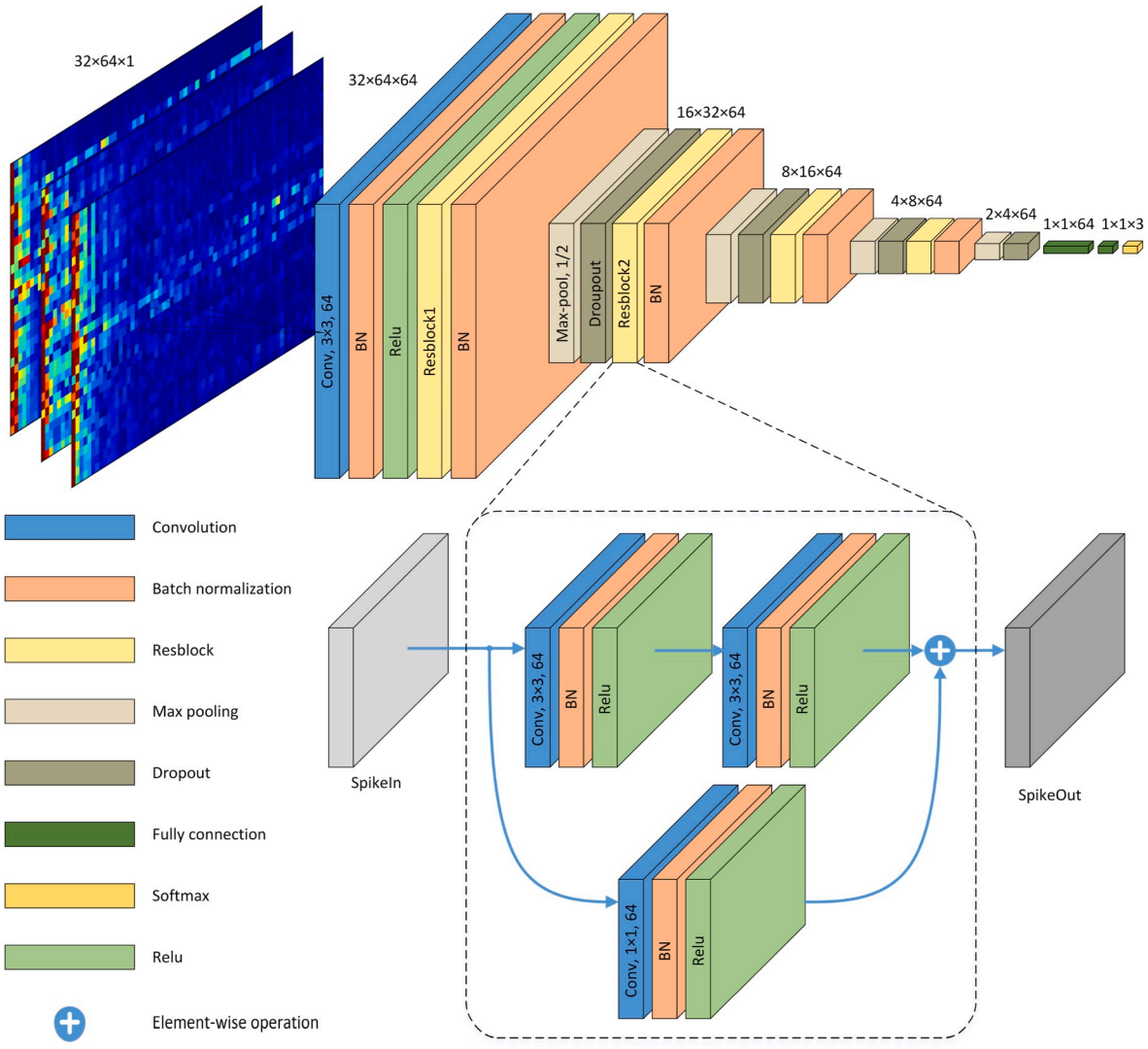

Fig. 1. The proposed residual block based CNN for EEG emotion recognition.

The residual block based CNN can effectively alleviate the problem of gradient disappearance and gradient explosion through the shortcut connections between layers. The network embedded with max pooling layer has a certain translation and rotation invariance to the input. Moreover, since the emotion-related features of EEG signals mainly reflected in the sub-frequency rhythms, the pooling operation in the frequency direction can make the neural network more effective for extracting emotion-relevant features from EFDMs. Finally, two fully connected layers are used for emotion classification based on the features extracted by the former convolution layers.

2.3. Grad-CAM

Gradient-weighted class activation mapping (Grad-CAM) is used to make CNN-based models more transparent by producing visual explanations (Selvaraju et al., 2020). This can be used to understand the importance of input data with respect to a target class of interest. In order to obtain the class-discriminative localization map Grad-CAM for any class c ($L^c_{Grad-CAM}$), the gradient of the score for class c was first computed (y^c), with respect to the feature maps A^k of a convolutional layer. These gradients flowing back are global-average-pooled to obtain the neuron importance weights: α^c_k

$$\alpha^c_k = \frac{1}{Z}\sum_{i\in w}\sum_{j\in h}\frac{\partial y^c}{\partial A^k_{ij}} \tag{4}$$

in which Z represents the number of pixels in the feature map.

Through performing a weighted combination of forward activation maps followed by a ReLU, the Grad-CAM can be expressed as:

$$L^c_{Grad-CAM} = \mathrm{Re}LU\left(\sum_k \alpha^c_k A^k\right) \tag{5}$$

The output $L^c_{Grad-CAM}$ indicates which parts the proposed neural networks have paid more attention to, and we denote them as attention heat maps. For each emotion, we use (6) to calculate the average heat maps of all samples to understand what's the difference when classify different emotions.

$$L_{AVE} = \frac{1}{N}\sum L^c_{Grad-CAM} \tag{6}$$

3. Dataset description and analysis

In this section, we make a description on two EEG emotion recognition datasets, i.e. SEED and DEAP. Then some data preprocessing methods are presented to prepare samples for cross-datasets emotion recognition. Finally, data distribution between different subjects are analyzed.

3.1. SEED dataset description

SEED dataset contains three categories of emotions, i.e., negative, neutral, and positive. Fifteen subjects (7 males and 8 females) participated in the experiments. EEG signals were recorded using an ESI NeuroScan System at a sampling rate of 1000 Hz from a 62-channels active AgCl electrode cap according to the international 10–20 system while they were watching emotional film clips. There are 15 trials (film clips watching test) in one experiment. Each subject participated in the experiment 3 times at an interval of one week or longer.

For EEG signal processing, the raw EEG data were first down-sampled to 200 Hz. In order to filter the noise and remove most artifacts, a bandpass filter of 0.5Hz–70Hz was performed (Zheng et al., 2019).

3.2. DEAP dataset description

DEAP is a multimodal dataset consisting of EEG recordings collected while watching the selected video clips to analyze human affective states. The EEG and peripheral physiological signals of 32 participants were recorded using a Biosemi ActiveTwo system as each watched 40 1-min long excerpts of music videos. The experiments were performed in two laboratory environments with controlled illumination. The EEG signals were recorded at a sampling rate of 512Hz from 32 active electrodes according to the international 10–20 system. Each participant assesses their levels of arousal, valence, dominance and liking using self-assessment manikins (SAM). Participants selected the numbers 1–9 for emotional state for each clip. The arousal scales extend from passive to active, and valence ranges from negative to positive.

Some preprocessing operations have been applied to the raw data, such as down sampling the recordings to 128 Hz; EOG artifacts were removed; a bandpass filter of 4Hz–45Hz was applied; averaged to the common reference. After that, the data were segmented into 60 s trials and a 3 s pre-trial baseline removed (Koelstra et al., 2012).

3.3. Data preprocessing

For the SEED dataset since the length of the EEG signals acquired under various stimulus differs greatly, we firstly count the lengths of all EEG trials, then the EEG signals are truncated (taking the first 37,000 sampling points for subsequent analysis) to ensure that every kind of emotion has the same number of samples.

Due to the big differences in the experimental protocol, the composition of subjects, and the configuration of signal acquisition system between DEAP and SEED dataset. In order to ensure that the classification task using both SEED and DEAP is similar, the emotional space of DEAP is divided into discrete parts according to the valence score similar to the approach in (Lan et al., 2019). Samples with a score in valence greater than 7 were classified as positive, samples with scores less than or equal to 7 and greater than 3 were classified as neutral, and samples with scores no more than 3 were treated as negative. Based on this classification criterion, the number of subjects with different types of emotion was counted in each film stimulus, and the frequency results are shown in Fig. 2. The horizontal axis represents the film clip index from 1 to 40, and the vertical axis represents the number of subjects with one specific emotion corresponding to each stimulus. We then look for the trials that have the most participants who reported to have successfully induced positive, neutral and negative emotion, respectively. These trials are: #18 for positive emotion, #16 for neutral emotion, and #38 for negative emotion, each having 27, 28 and 19 subjects respectively. Fourteen subjects in DEAP (numbered 2, 5, 10, 11, 12, 13, 14, 15, 19, 22, 24, 26, 28 and 31) that successfully induced all three types of emotions under these three trials (#18, #16 and #38) were selected for subsequent experiments.

After that, the EEG signals in each channel are divided into a number of samples with a 1s long non-overlapping Hanning window in two datasets. Hence, we obtain 185 samples of one trail corresponding to one film clip, and 41625 samples are obtained under different emotions in SEED. For DEAP, since each trial lasts for 63 s and the first 3 s are baseline recording without emotion elicitation, we only use the segment from the 4th second to the end. Thus, 60 samples were obtained in each trial, the total number of samples is 2520. Finally, Fourier Transform and normalization are performed on the samples to get the EFDMs.

3.4. EFDMs

The Fourier transform is applied to each EEG channel of all samples obtained above, and then the transformed results are normalized to produce the input data that are suitable for CNN. The normalized results in two dimensions were known as EFDMs. The EFDMs of the EEG signals can be represented as grayscale images, and the normalized results can

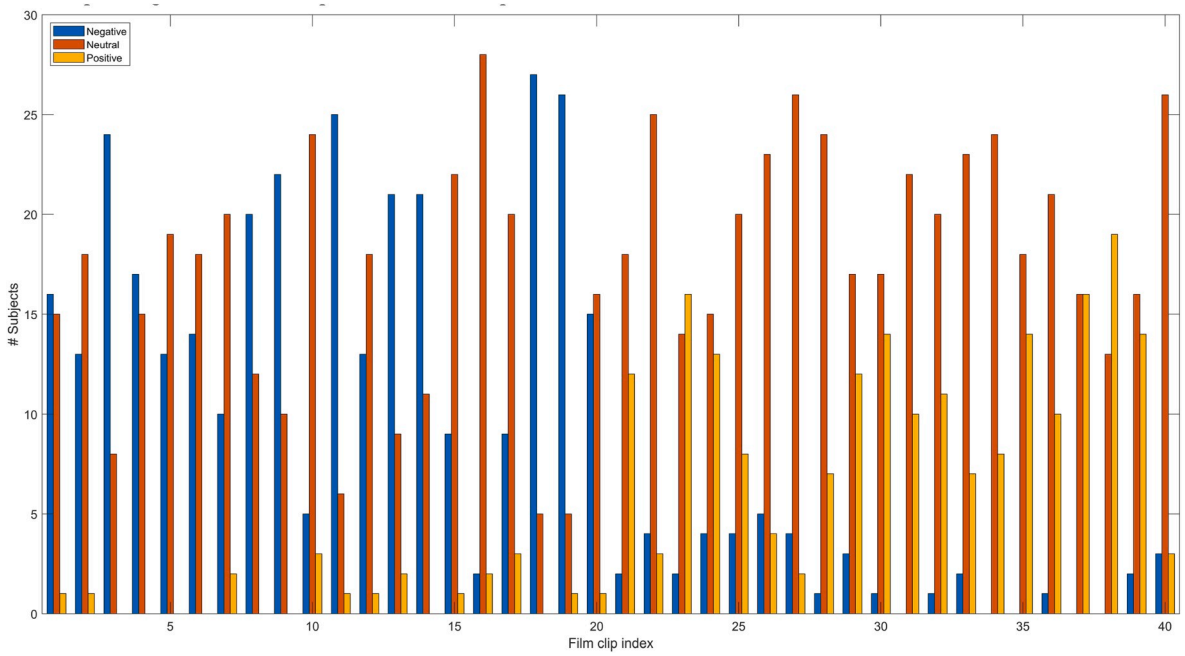

Fig. 2. DEAP emotion space discrete results with valence score.

be compared to the gray pixel value. Therefore, we can build a CNN for EEG-based emotion recognition with EFDMs. Fig. 3 shows the EFDMs under different emotions.

3.5. Data distribution analysis

In order to illustrate the difference in EEG signals between different subjects, we use the SEED dataset as an example and randomly select 50 samples of five subjects under three emotional states for analysis. Firstly, the DE of five sub-band EEG signals in all channels are extracted, and the feature vectors are formed. Then the PCA is used to reduce the dimensionality of the features, the two components with the largest eigenvalues are retained for data distribution analysis.

As can be seen from Fig. 4, the data distribution among different subjects is quite different, which does not satisfy the independent and identical assumption between training and test samples in traditional machine learning. In addition, the feature differences among three kinds of emotions of the same subject are not obvious. Therefore, in this case, traditional machine learning methods often fail to achieve good recognition results. The recently proposed transfer learning is specifically designed to solve this problem. Such methods usually carried out within one dataset, which has some similar parts among different subjects, such as EEG signal acquisition equipment and experimental process, this is helpful for knowledge transfer from source to target domain. However, in cross-dataset emotion recognition task, the differences introduced by different signal acquisition equipment and experimental environments need also to be considered. Therefore, it is more difficult to realize the transfer learning of emotion recognition model with cross-datasets.

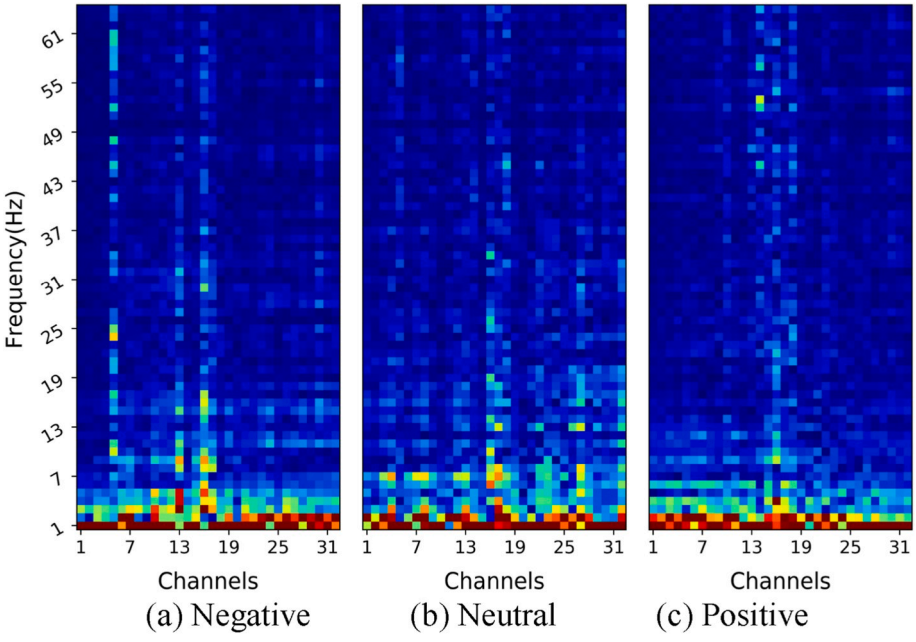

Fig. 3. EFDMs under different emotions.

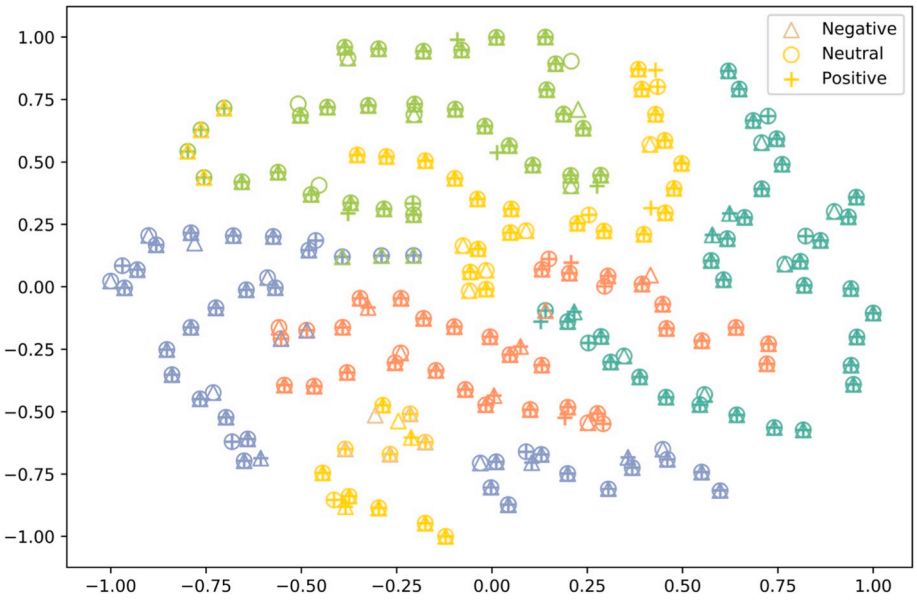

Fig. 4. Two-dimension visualization of features selected from different subjects in SEED.

4. Experiments and results analysis

We set up two experiments. First, the effectiveness of the proposed method for EEG-based emotion recognition is verified using SEED. Then, based on the deep neural network transfer learning strategy, the pre-trained model is applied to DEAP with 12 training samples of each emotion class.

4.1. SEED based emotion recognition

Over the past few years, many scholars have conducted notable research on EEG based emotion recognition with SEED. To compare the proposed approach with (Zheng and Lu, 2015), (Lu et al., 2015), (Liu et al., 2016), (Yang et al., 2017), in this experiment, we strictly obey the protocol of Zheng et al. (Zheng and Lu, 2015). Specifically, for all 15 trials of EEG data associated with one session of one subject, the first 9 trials are used to serve as the training set and the remaining 6 are the testing set. Then, the recognition accuracy corresponding to each period is obtained for each subject. Finally, the average classification accuracy over three sessions for all 15 subjects is calculated.

The training and testing processes are implemented using Pytorch framework with Adam algorithm as an optimizer; the learning rate is set as 0.0001, and the loss function is a cross entropy loss function.

We compared the proposed models with other state-of -the-art approaches (Tang et al., 2017), (Li et al., 2018), (Song et al., 2018) and the baseline method, which uses DBN directly as the classifier. As shown in Table 1, Bimodal-LSTM achieved the best accuracy (93.97%) among (Lu et al., 2015), (Liu et al., 2016), (Yang et al., 2017), (Tang et al., 2017) with 4 s of EEG as well as eye movement information. Based on single EEG, Li et al. (Lu et al., 2015) obtained the best recognition rate of 92.38% with 9s EEG. The result of the proposed model based on EFDMs and CNN is 90.59%, which is 4.51% higher than the baseline results with differential entropy and DBN. Compared with other methods, the data samples of 1s used in this paper are shorter and the process to produce EFDMs through STFT is simpler compared to DE. That is to say, the EEG based emotion recognition method combined with EFDMs and CNN is effective.

To see the results of recognizing each emotion, we depict the confusion matrix corresponding to the experiments using SEED, as shown in Fig. 5. Each row of the confusion matrix represents the target class and each column represents the predicted class that a classifier outputs. The element (i,j) is the percentage of samples in class i that were classified as class j. From the results we can see that, in general, positive emotion can be recognized with high accuracy (93%), while negative emotion is more difficult to recognize, and very easy to be confused with neutral emotion.

Table 1
Some notable works on SEED dataset.

Method	Feature	Classifier	Signal	Accuracy (%)
Zheng and Lu (2015)	DE	DBN	EEG(1s)	86.08
Lu et al. (2015)	DE (EEG)	Fuzzy integral fusion strategy	EEG (4s) + Eye movement	87.59
Liu et al. (Liu et al., 2016), 2016	DE	BDAE + SVM	EEG (4s) + Eye movement	91.01
Yang et al. (Yang et al., 2017), 2017	DE	hierarchical network with subnetwork nodes	EEG (4s) + Eye movement	91.51
Tang et al. (Tang et al., 2017), 2017	PSD, DE, Mean, SD	Bimodal-LSTM	EEG (4s) + Eye movement	93.97
Li et al. (Li et al., 2018), 2018	DE	BiDANN	EEG (9s)	92.38
Song et al. (Song et al., 2018), 2018	DE	DGCNN	EEG (1s)	90.40
Ours	EFDMs	CNN	EEG (1s)	90.59

4.2. DEAP based emotion recognition

The goal of machine learning is to build a model that is as general as possible to meet the requirements of different user groups and different environments. However, such an ideal model often fails to meet the expected requirements in practical applications. Therefore, how to establish a universal model to tackle the possible differences between

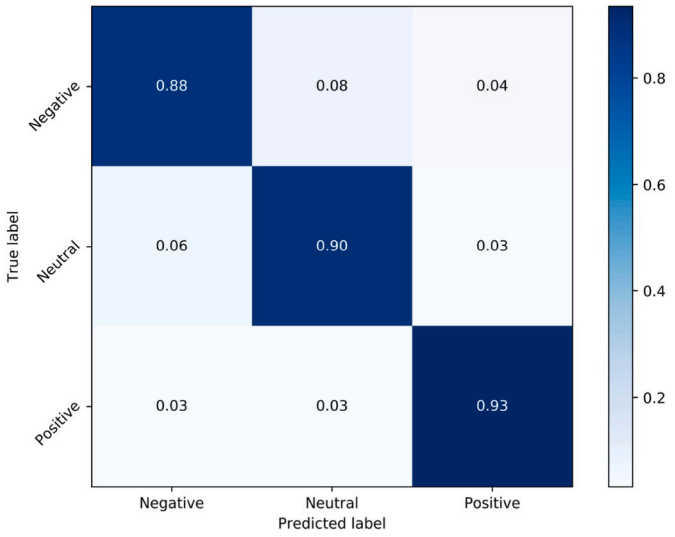

Fig. 5. Confusion matrix on SEED.

subjects and signal acquisition devices under different classification tasks, as well as realizing few-shot learning, is a problem that needs to be taken into consideration in CNN-based emotion recognition system. Various studies on CNN have shown that shallow convolution layers are designed to extract common basic features from the input, while deeper convolution layers can extract more abstract and task related features. Therefore, it is possible to get an accurate classification result based on partial fine-tuning of the pre-trained CNN with a few training samples. Generally, the accuracy is positively correlated with the number of fine-tuned layers. To this end, through two deep neural network transfer learning strategies, i.e. just fine-tune the fully connected layers or fine-tune all layers, the pre-trained CNN with SEED is transformed for another emotion recognition task based on DEAP.

In order to produce EFDMs with the same attributes (including channel order, frequency range and size) for deep model transfer learning between two datasets, we take following different preprocessing methods. For SEED, 32 EEG channels (Fp1, Fp2, AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, PO4, O1, Oz, O2) are shared with DEAP and the first 64 frequency points are selected to build EFDMs with a size of 32×64. For DEAP, the EEG channels are rearranged according to the former presented electrode order to ensure that they are consistent with SEED.

Based on the review of relevant works, we found that some scholars have conducted research on emotion recognition with transfer learning across two datasets (Lan et al., 2019). However, there are some differences between the research focus. The main research focus of this paper is to realize emotion recognition based on a deep model transfer strategy with a few training samples. While the latter aims to use a domain adaptation method to transfer the classification knowledge learned using SEED, to DEAP. There are also differences in experimental settings. In this paper, a small amount of data of the target subject is used for training, while the latter uses the leave-one- subject-out cross-validation strategy for classification on DEAP. (The data of each session in SEED were used as source samples, and each subject in DEAP was set as a target sample for testing.) The recognition results using the DEAP dataset with domain adaptation from (Lan et al., 2019) are shown in Table 2.

It can be seen from the table that Transfer component analysis (TCA) achieved the best recognition accuracy under the three experimental settings. However, the recognition accuracy of all domain adaptation methods is very low (no more than 43%), and the recognition result for Information theoretical learning (ITL) is even lower than that of the baseline method which did not adopt transfer learning.

From here on, we will carry out deep model transfer learning based on a small number of training samples. For DEAP, we randomly divide the samples of each subject into a training and testing dataset with a training versus testing ratio of 1:4. (Which means 20% for training, 80% for testing, the training sample size of each emotion is 12.) Then, two different deep model transfer learning strategies are used to fine-tune the pre-trained CNN with the training dataset, after which the network is tested on the testing samples. During model fine-tuning, Adam is used as an optimizer, the learning rate is 0.00002 with cross-entropy as a loss function. The recognition accuracy of the proposed method using the DEAP dataset based on a few training samples is shown in Fig. 6.

Table 2
Cross-datasets emotion recognition results with leave-one-subject-out cross-validation strategy.

Method	SEEDI→DEAP	SEEDII→DEAP	SEEDIII→DEAP
Baseline	34.57 (7.98)	32.99 (3.44)	32.51 (6.73)
MIDA	40.34 (14.72)	39.90 (14.83)	37.46 (13.11)
TCA	**42.60** (14.69)	**42.40** (14.56)	**39.76** (15.15)
SA	36.73 (10.69)	37.36 (7.90)	37.27 (10.05)
ITL	34.50 (13.17)	34.10 (9.29)	33.62 (10.53)
GFK	41.91 (11.33)	40.08 (11.53)	39.53 (11.31)
KPCA	35.60 (6.97)	34.69 (4.34)	35.11 (10.05)

The average recognition accuracy and standard deviation of the baseline, fine-tune fc layers and fine-tune all layers are 32.94% (3.80), 70.14% (9.81), and 82.84% (10.74), respectively. The baseline indicates that the pre-trained model using SEED is directly applied to DEAP without using a transfer learning strategy. It can be seen that the recognition accuracy of all 15 subjects with this method is very low (less than 40%), which means that the data distribution of DEAP is quite different from that of SEED. The data distribution between these two datasets doesn't satisfy the independent and identical assumption of traditional machine learning. It is worth noting that the baseline result is similar to the baseline recognition accuracy in (Lan et al., 2019). Fine-tune fc layers and fine-tune all layers represent the recognition results obtained by using 12 training samples to fine-tune the pre-trained CNN with fully connected layers or all layers, respectively. For each subject, the recognition results of these two methods changed synchronously. Compared with the baseline results, these two transfer learning strategies improved the recognition accuracy significantly for every subject. The method of fine-tune all layers achieved the best classification results among 15 subjects with an average accuracy of 82.84%, the highest result is 96.60% on subject 11, while a lower result is 61.18% and 63.13% for subject 7 and 9, respectively. The best recognition accuracy using fine-tune fc layers is 84.86% with subject 11. However, its performance on subject 5, 7, 9, and 14 is poor, all of them are lower than 60%.

The confusion matrix of the proposed method is shown in Fig. 7. As can be seen from the figure, the baseline method classified almost all samples (about 84%) as neutral emotion, only a small number of samples are recognized as negative, while fewer samples are classified as positive (less than 1%). It shows that there is a big difference in the EFDMs between the two datasets, and the pre-trained CNN learned from SEED cannot be directly used for DEAP. With the proposed deep model transfer learning strategy of fine-tuning the fully connected layers of the pre-trained CNN, the classification accuracy has been greatly improved. The best result is obtained in neutral emotion recognition (77%), followed by positive, while the result of negative emotion recognition is not as good as the former (62%). With the method of fine-tuning all layers, the recognition results have been further enhanced, the positive emotion recognition is the best (86%), the neutral is second, and the accuracy of the negative emotion recognition has reached 79%. It is worth noting that the emotions achieved with the best classification accuracy of fine-tune fc layers and fine-tune all layers are neutral and positive, respectively. That is to say, through fine-tuning the weights of the convolution layers in the pre-trained CNN, it has helped to learn more emotion

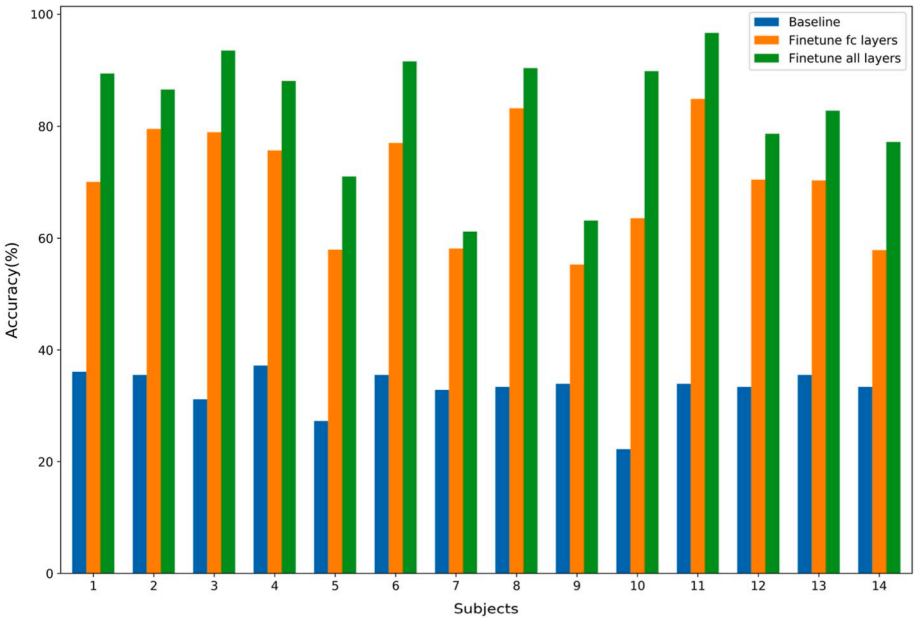

Fig. 6. The recognition accuracy of our proposed method on DEAP.

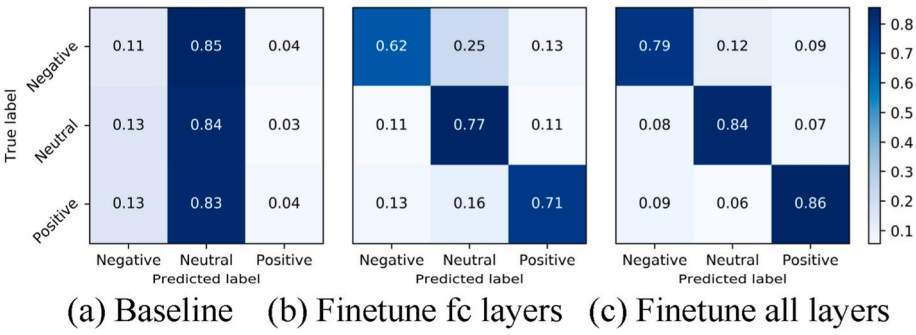

Fig. 7. The confusion matrix on DEAP.

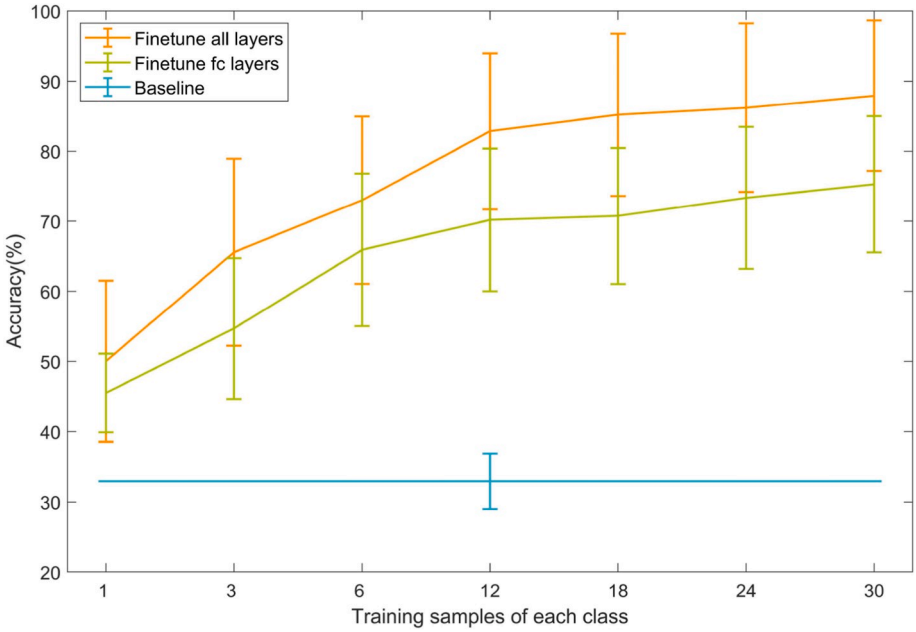

Fig. 8. Classification accuracy with varying number of training samples.

related features in DEAP. The average recognition accuracy for fine-tuning all layers is 82.84%.

We set up six comparative experiments to illustrate the effectiveness of the proposed methods for EEG based emotion recognition, including one-shot learning (taking one sample from each emotion to form a training set, and using remaining samples for testing), as well as the experiments with a training data proportion of 0.05, 0.1, 0.3, 0.4 and 0.5, respectively. For the training dataset, we randomly selected a number of samples from all types of emotions according to the training data proportion, and the remaining samples are used to form the testing dataset. In order to avoid the problem that a small number of randomly selected training samples are not representative enough, the comparative experiment under every experimental setting was repeated five times, and the average value was used as the final result. The average recognition accuracy and standard deviation of the proposed methods with a different number of training samples are shown in Fig. 8.

We can see that the accuracy direction of these two deep model transfer learning is consistent, both increase with the number of training samples used. When the number of training samples from each emotion increased to 12 (the training data proportion is 0.25), the growth trend of these two methods was significantly slower. Under each experimental setup, the result of fine-tune all layers is always better than that of fine-tune fc layers. This shows that there is a difference in EFDMs between SEED and DEAP, which needs to be adjusted through the fine-tuning with the weights of the convolution layers. Additionally, under the 'one-shot learning' experimental setup, which uses only one sample from each emotion kind for training, the accuracy of these two methods (e.g., fine-tune fc layers 45.50% (5.60), fine-tune all layers 50.02% (11.48)) is much higher than that of the baseline method (32.94% (3.80)). This also illustrated by the effectiveness of the proposed methods in emotional recognition by fine-tuning the CNN with a few samples.

4.3. What did our network learn?

The existing CNN based EEG emotion recognition studies take the original EEG signals or time frequency maps as the input. However, the original EEG signals cannot represent its frequency feature, and the time frequency maps cannot reflect the position relationship between EEG channels. The EFDMs proposed in this paper can simultaneously give expression to the frequency distribution as well as the EEG electrodes position information. Based on the attention mechanism of deep neural network, we adopt Grad-CAM to analysis what information the CNN has learned from EFDMs. Investigate the key EEG electrodes as well as frequency bands corresponding to each emotion category automatically and simultaneously.

Fig. 9 shows the 'attention maps' generated with Grad-CAM of different emotion categories. The brighter the color is, the more important the information contained in this area is to emotion recognition. Similarly, the darker the color is, the less important this area is. From the figure, we can see that the EEG channels and frequency bands that the CNN focused on are quite different. The average attention level for all channels are shown in the right histogram, which represent the average of Grad-CAM value across each channel. From these attention heat maps and histograms, we can find that there is a large similarity between negative and neutral emotion. That's why the network misclassified 6% of neutral emotions into negative (with 3% into positive), while the proportion of negative samples misclassified into neutral is 8% (with 4% into positive) as shown in Fig. 5.

From Fig. 9 (a), (d), the key frequency bands related to negative emotion recognition mainly concentrated in 25–57 Hz, and the key channels distributed around FC2, FC6. From Fig. 9 (b), (e), the key EEG frequency bands and channels of neutral emotion are 27–55Hz, and T8, CP5, CP1, CP2. Similarly, the critical information of positive emotion from Fig. 9 (c), (f) are 24–59Hz, and FC2, FC6. Although the distribution of key frequency bands under three emotions is highly coincident, the

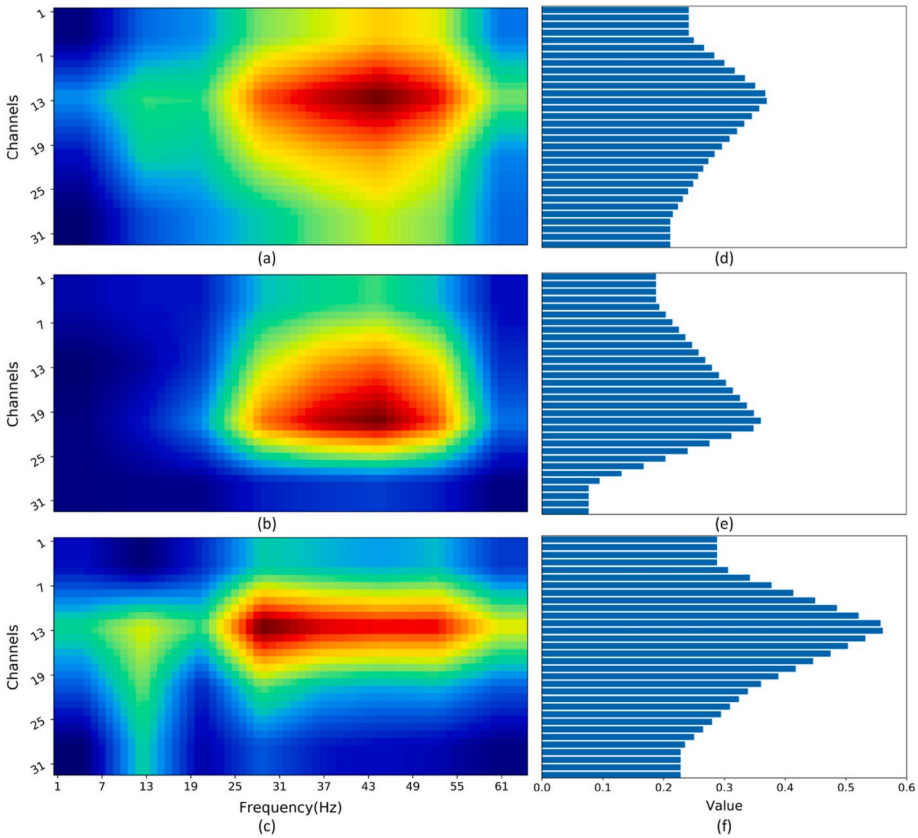

Fig. 9. Heat maps and average attention level for every channel obtained through Grad-CAM on SEED. (a), (d) Negative. (b), (e) Neutral. (c), (f) Positive.

key point of positive (29Hz) is quite different from that of negative (44Hz), and neutral (44Hz). This means the high frequency feature components contain more distinguishing information for EEG based emotion recognition. In addition, the alpha band (8–13Hz) of some channels is helpful for negative and positive emotion classification, while not for neutral emotion. We can draw the conclusion that CNN pays more attention to the high frequency bands of the EEG signals (24–59Hz), which is consistent with the conclusion in (Zheng et al., 2015), (Zheng and Lu, 2015), (Wang et al., 2014), (Zheng et al., 2019). Therefore, the CNN can be trained to automatically discover the EEG channels and features that are conducive to emotion recognition. It is worth noting that the range of key EEG channels and frequency bands obtained in this paper is a little wider than the true information due to the influence of two-dimensional convolution operation. That is to say, the real key EEG channels and frequency bands related to emotion recognition should be concentrated in several channels and frequency bands with the highest brightness in the attention maps.

5. Conclusion

In this paper, we have provided a solution to tackle the challenge of differences in individual emotions with deep model transfer learning. Aims to build a robust emotion recognition model independent of stimulus, subjects, and EEG collection device etc. We have mainly set up two experiments, within and cross-datasets emotion recognition. First, the effectiveness of the proposed approach is valid on SEED with an average accuracy of 90.59%. After that, the pre-rained CNN from the first experiment is applied to DEAP with the deep model transfer learning method. Experiments show that when 12 training samples of each emotion are used for deep model fine-tune, a high accuracy can be achieved with a few samples. At the last, based on the attention mechanism of deep neural network, we adopt Grad-CAM to analysis what information the CNN has learned from EFDMs, obtained the key EEG electrodes and frequency bands corresponding to each emotion category automatically and simultaneously. The results show that the high frequency bands (24–59Hz) are more helpful for emotion recognition. he key channels of neutral are T8, CP5, CP1, CP2, which is different from that of negative and positive (FC2, FC6).

From Table 1, we can see that the proposed approach hasn't achieved the best performance, this may due to the 1s signal used is shorter than that of others with 4s and 9s, or due to the lack of eye movement data. We will consider the issue of EEG data length as well as the multimodal data fusion method for emotion recognition in the future. Moreover, we only studied the transfer learning method of fine-tuning deep neural networks to tackle the challenge of individual difference between subjects with the cross-datasets emotion recognition experiment at present. More and more advanced deep transfer learning methods have emerged recently. Therefore, more attempts should be tried with these algorithms. Furthermore, the source and target domain included in this paper is the same. Concentrate on the EEG emotion recognition issue with insufficient samples and different source and target domain is another work worth studying.

CRediT authorship contribution statement

Shichao Wu: Methodology, Software, Writing - original draft, Writing - review & editing. **Weiwei Zhang:** Data curation, Validation. **Zongfeng Xu:** Resources. **Yahui Zhang:** Visualization. **Chengdong Wu:** Investigation. **Sonya Coleman:** Formal analysis.

References

Abdelhamid, O., Mohamed, A., Jiang, H., Deng, L., Penn, G., Yu, D., 2014. Convolutional neural networks for speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 22 (10), 1533–1545.

Alarcao, S.M., Fonseca, M.J., 2019. Emotions recognition using EEG signals: a survey. IEEE Transactions on Affective Computing 10 (3), 374–393.

Alkaysi, A.M., Alani, A., Loo, C., Powell, T.Y., Martin, D., Breakspear, M., Boonstra, T.W., 2017. Predicting tDCS treatment outcomes of patients with major depressive disorder using automated EEG classification. Journal of Affective Disorders 208, 597–603.

Ang, A., Yeong, Y., Wee, W., 2017. Emotion classification from EEG signals using time-frequency-DWT features and ANN. Journal of Computer and Communications 5, 75–79.

Bocharov, A.V., Knyazev, G.G., Savostyanov, A.N., 2017. Depression and implicit emotion processing: an EEG study. Neurophysiologie Clinique-clinical Neurophysiology 47 (3), 225–230.

Chuang, C., Ko, L., Lin, Y., Jung, T., Lin, C., 2014. Independent component ensemble of EEG for brain–computer interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering 22 (2), 230–238.

Dash, M., Liu, H., 1997. Feature selection for classification. In intelligent data analysis 1, 131–156.

Duan, R., Wang, X., Lu, B., 2012. EEG-based emotion recognition in listening music by using support vector machine and linear dynamic system. In international conference on neural information processing. Springer, Berlin, Heidelberg, pp. 468–475.

Duan, R., Zhu, J., Lu, B., 2013. Differential entropy feature for EEG-based emotion classification. In international ieee/embs conference on neural engineering. IEEE, pp. 81–84.

Ezzatdoost, K., Hojjati, H., Aghajan, H., 2020. Decoding olfactory stimuli in EEG data using nonlinear features: a pilot study. Journal of Neuroscience Methods 341, 108780.

Friston, K.J., 2001. Book Review: Brain function, nonlinear coupling, and neuronal transients. The Neuroscientist 7 (5), 406–418.

Hatcher, W.G., Yu, W., 2018. A survey of deep learning: platforms, applications and emerging research trends. IEEE Access 6, 24411–24432.

Jie, X., Cao, R., Li, L., 2014. Emotion recognition based on the sample entropy of EEG. Biomedical Materials and Engineering 24 (1), 1185–1192.

Knyazev, G.G., Slobodskojplusnin, J.Y., Bocharov, A.V., 2010. Gender differences in implicit and explicit processing of emotional facial expressions as revealed by event-related theta synchronization. Emotion 10 (5), 678–687.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I., 2012. Deap: a database for emotion analysis; using physiological signals. IEEE Transactions on Affective Computing 3 (1), 18–31.

Lan, Z., Sourina, O., Wang, L., Scherer, R., Mullerputz, G., 2019. Domain adaptation techniques for EEG-based emotion recognition: a comparative study on two public datasets. IEEE Trans. on Cognitive and Developmental Systems 11 (1), 85–94.

Li, M., Chen, W., Zhang, T., 2017. Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble. Biomedical Signal Processing and Control 31, 357–365.

Li, Y., Zheng, W., Cui, Z., Zhang, T., Zong, Y., 2018. A novel neural network model based on cerebral hemispheric asymmetry for EEG emotion recognition. In international joint conference on artificial intelligence. Morgan Kaufmann, pp. 1561–1567.

Liberati, G., Federici, S., Pasqualotto, E., 2015. Extracting neurophysiological signals reflecting users' emotional and affective responses to BCI use: a systematic literature review. NeuroRehabilitation 37 (3), 341–358.

Liu, W., Zheng, W., Lu, B., 2016. Emotion recognition using multimodal deep learning. In international conference on neural information processing. Springer, Cham, pp. 521–529.

Lu, Y., Zheng, W., Li, B., Lu, B., 2015. Combining eye movements and EEG to enhance emotion recognition. In international conference on artificial intelligence. Morgan Kaufmann, pp. 1170–1176.

Mathersul, D., Williams, L.M., Hopkinson, P.J., Kemp, A.H., 2008. Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety. Emotion 8 (4), 560–572.

Mühl, C., Allison, B., Nijholt, A., Chanel, G., 2014. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. Brain-Computer Interfaces 1 (2), 66–84.

Niemic, C., 2004. Studies of emotion: a theoretical and empirical review of psychophysiological studies of emotion, 1. jur: Journal of Undergraduate Research, pp. 15–18.

Palus, M., 1996. Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. Biological Cybernetics 75 (5), 389–396.

Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22 (10), 1345–1359.

Preethi, J., Sreeshakthy, M., Dhilipan, A., 2014. A survey on EEG based emotion analysis using various feature extraction techniques. International Journal of Science, Engineering and Technology Research (IJSETR) 3 (11), 3113–3120.

Sammler, D., Grigutsch, M., Fritz, T., Koelsch, S., 2007. Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. Psychophysiology 44 (2), 293–304.

Sebe, N., Cohen, I., Gevers, T., Huang, T.S., 2005. Multimodal approaches for emotion recognition: a survey. Electronic Imaging 56–67.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-cam: visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision 128 (2), 336–359.

Sohaib, A.T., Qureshi, S., Hagelbäck, J., Hilborn, O., Jercic, P., 2013. Evaluating classifiers for emotion recognition using EEG. In international conference on augmented cognition. Springer, Berlin, Heidelberg, pp. 492–501.

Song, T., Zheng, W., Song, P., Cui, Z., 2018. EEG emotion recognition using dynamical graph convolutional neural networks. IEEE Transactions on Affective Computing. https://doi.org/10.1109/TAFFC.2018.2817622.

Tang, H., Liu, W., Zheng, W., Lu, B., 2017. Multimodal emotion recognition using deep neural networks. In international conference on neural information processing. Springer, Cham, pp. 811–819.
Wang, H., 2011. Optimizing spatial filters for single-trial EEG classification via a discriminant extension to CSP: the Fisher criterion. Medical & Biological Engineering & Computing 49 (9), 997–1001.
Wang, X., Nie, D., Lu, B., 2014. Emotional state classification from EEG data using machine learning approach. Neurocomputing 129, 94–106.
Wen, Z., Xu, R., Du, J., 2017. A novel convolutional neural networks for emotion recognition based on eeg signal. In 2017 international conference on security, pattern analysis, and cybernetics. IEEE, pp. 672–677.
Yanagimoto, M., Sugimoto, C., 2016. Convolutional neural networks using supervised pre-training for EEG-based emotion recognition. In the 8th international workshop on biosignal interpretation. IEEE, pp. 72–75.
Yang, Y., Wu, Q.M., Zheng, W., Lu, B., 2017. EEG-based emotion recognition using hierarchical network with subnetwork nodes. IEEE Transactions on Cognitive and Developmental Systems 10 (2), 408–419.
Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In neural information processing systems. MIT Press, pp. 3320–3328.
Zhang, Q., Yang, L.T., Chen, Z., Li, P., 2018. A survey on deep learning for big data. Information Fusion 42, 146–157.
Zheng, W., Liu, W., Lu, Y., Lu, B., Cichocki, A., 2019. Emotionmeter: a multimodal framework for recognizing human emotions. IEEE Transactions on Systems, Man, and Cybernetics 49 (3), 1110–1122.
Zheng, W., Lu, B., 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Transactions on Autonomous Mental Development 7 (3), 162–175.
Zheng, W., Guo, H., Lu, B., 2015. Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network. In international ieee/embs conference on neural engineering. IEEE, pp. 154–157.
Zheng, W., Zhu, J., Lu, B., 2019. Identifying stable patterns over time for emotion recognition from EEG. IEEE Transactions on Affective Computing 10, 417–429.
Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., Yan, B., 2017. Emotion recognition from EEG signals using multidimensional information in EMD domain. BioMed Research International 1–9.
Zubair, M., Yoon, C., 2018. EEG based classification of human emotions using discrete wavelet transform. IT Convergence and Security 2017. Springer, Singapore, pp. 21–28.