Journal of Computers Vol. 34 No. 3, June 2023, pp. 109-120
doi: 10.53106/199115992023063403008

EEG Emotion Recognition Method Based on 3D Feature Map and Improved DenseNet

Jing-Ran Su[1,2], Qiu-Sheng Li[1,2*], Qian-Li Zhang[1,2], Jun-Yong Hu[1,2]

[1] Research Center of Intelligent Control Engineering Technology, Gannan Normal University, Ganzhou, Jiangxi 341000, China
[2] School of Physics and Electronic Information, Gannan Normal University, Ganzhou, Jiangxi 341000, China
{sujingran, liqiusheng, zhangqianli, hujunyong}@gnnu.edu.cn

Received 17 June 2022; Revised 23 September 2022; Accepted 30 October 2022

Abstract. Emotion, as a high-level function of the human brain, has a great impact on people's mental health. To fully consider EEG signals' spatial information and time-frequency information, and realize human-computer interaction better. This paper proposes an improved DenseNet emotion recognition model based on 3D feature map. By extracting the differential entropy features of the θ, α, β and γ frequency bands of the EEG signals, and combining the position mapping relationship of the EEG channel electrodes, a three-dimensional feature map is constructed, and then the improved densely connected convolutional network (DenseNet) is used for secondary feature extraction and classification. To verify the effectiveness of this method, a classification experiment including positive, neutral and negative emotions is carried out on the SEED data set. The classification accuracy rates obtained in the single-subject experiment and the all-subject experiment are 98.51% and 98.68%, respectively. The experimental results show that the method of 3D feature map combined with feature reuse can get high-precision classification results, which provides a new direction for emotion recognition.

Keywords: EEG, electrode mapping, 3D feature map, feature reuse, multi-scale convolution kernel

1 Introduction

Emotion reflects people's attitudes and responses to objective things. It is generally believed that it is a psychological activity mediated by individual wishes and needs [1]. In recent years, problems such as fatigued driving and mental health have increasingly appeared in front of the public. Positive emotions can improve work efficiency and play a positive role in promoting human life and work, while negative emotions will lead to a decline in work efficiency. The long-term accumulated negative emotions are also a major source of depression flare-ups. Therefore, emotion recognition plays an important role in human life safety and mental health.

Emotion recognition is an interdisciplinary subject that integrates psychology, neuroscience, medicine and computer. At present, researchers from various countries have made some progress in the research of emotion recognition. From the analysis of data sources, the research on emotion recognition is mainly based on two aspects: non-physiological factors and physiological factors. Because people will deliberately conceal their emotions, it is impossible to effectively identify people's true emotional states from external features based on non-physiological factors, while physiological signals are not camouflaged and are not easily affected by people's subjective emotions. Common physiological signals include ECG, EEG, EMG, etc. EEG signals are widely used in emotion recognition due to their easy measurement and portable equipment. From the analysis of the use of classifiers, there are two methods: machine learning and deep learning. The machine learning method adopts the combination of manual feature extraction with support vector machine (SVM) [2], K-Nearest Neighbor (KNN), random forest [3], and other traditional machine learning methods. Among them, the EEG features used for emotion recognition mainly include Hjorth parameters [4], fractal dimension feature [5], differential entropy (DE) [6], power spectrum density, etc. Alazrai et al. [7] proposed a new feature extraction method based on time-frequency analysis, combined the extracted features with SVM for classification research, and verified it with the internationally public DEAP [8] data set. The final accuracy is 86.20%. In addition to extracting features directly from the original signals, some researchers use signal decomposition methods to extract features from

* Corresponding Author

the decomposed components, for example, Zhang et al. [9] used the method of empirical mode decomposition to extract the entropy features of samples and form feature vectors, which were used as the input of the SVM and verified by experiments on the DEAP data set, with an average accuracy of 93.20%.

Due to the non-stationarity and non-linearity of EEG signals and the correlation between different channels and different frequency bands, the traditional machine learning methods can not effectively use the spatial and time-frequency information of EEG signals to further improve the accuracy of EEG emotion classification, while the deep learning method can make up for this defect. The deep learning method mainly adopts an end-to-end training method, which can automatically extract features and complete classification tests without relying on manual experience to complete feature extraction. With the in-depth development of deep learning, researchers at home and abroad have paid more attention to the research of deep learning methods, such as deep belief network (DBN) [10], convolution neural network (CNN) [11], long short-term memory (LSTM) [12] and other deep neural networks, which have achieved good results in EEG emotion recognition [13-15]. However, the influence of shallow features is often ignored when using deep learning, shallow features have higher resolution.

In the aforementioned studies, whether using machine learning classification methods or deep learning classification methods, most of them are based on the full frequency range of EEG signals. Studies in psychology and neuroscience have shown that the five classical rhythms (δ (1~3Hz), θ (4~7 Hz), α (8~13Hz), β (14~30Hz), and γ (31~50Hz)) of EEG signals are closely related to various physiological and psychological activities of human beings. However, only a few researchers have taken this feature into account in the research process. Among them, Koelstra et al. [8] extracted the power spectral density of five EEG rhythms for emotion classification and recognition. Xin Li et al. [16] first decomposed and reconstructed the emotional EEG signals to obtain four rhythm waves, then extracted wavelet energy, wavelet entropy, approximate entropy and Hurst index for the four rhythm waves, and finally used PCA algorithm to reduce its dimension and send the processed results to the SVM classifier to obtain emotion recognition results. Based on the existing research, this paper will take the influence degree of the classical frequency bands on the recognition results as one of the key points of this paper.

To sum up, there are two major aspects of the limitations of research process, the first is some researchers in the use of deep network failed to fully consider the features of shallow affect the performance of classification, followed by most researchers ignore the characteristics of EEG signal itself when extracting features, its classical frequency band and the spatial structure of EEG channel contain a lot of emotional information. To solve the above problems, this paper proposes a EEG emotion recognition method based on three-dimensional feature map and densely connected convolutional network (DenseNet) [17]. The main contributions of this paper can be summarized as follows:

(1) 3D feature map. Firstly, the position information between electrodes of the EEG channel is mapped into two dimensions, so that the relative position information between electrodes is preserved as much as possible, that is, the spatial information of the EEG signals. Then, the differential entropy features of the four rhythms (θ, α, β and γ) are extracted to form a three-dimensional feature map, which is used as the input of the neural network.

(2) MSC-DenseNet. Combining the advantages of shallow feature reuse of the DenseNet network and multi-scale convolution kernel, a deep learning network is constructed to classify and recognize positive, neutral, and negative emotional states.

2 Data Acquisition and Feature Extraction

2.1 SEED Data Set and Its Preprocessing

It doesn't make much sense to compare the accuracy rate obtained by using different data sets. In this paper, the SJTU emotion EEG data set (SEED) [6, 14] provided by the Center for Brain-like Computing and Machine Intelligence (BCMI) of Shanghai Jiao Tong University is used to study the EEG emotion recognition including three emotions (positive, neutral and negative).

The SEED data set records the EEG signals on 62 channels. 15 movie clips (5 positives, 5 neutral, and 5 negatives) are selected as emotion-evoking material in the data set. The detailed information of the 15 movie clips are shown in Table 1. Each movie clip maximally reflects the emotional states of the subjects. In order to avoid fatigue caused by excessively long clips, each movie clip is about 4 minutes long. Each subject is required to watch 15 movie clips in each experiment, including 7 males and 8 females, with a total of 15 subjects participating. Three experiments are carried out, with an interval of about one week. A total of 45 groups (15×3) of EEG data are collected.

In order to carry out subsequent recognition experiments, certain preprocessing of the data set is required. Firstly, the collected EEG signals are down-sampled, the original 1000 Hz EEG data are down-sampled to 200 Hz, and then a 0-75 Hz band-pass filter is applied to perform denoise and de-artifact processing to obtain the preprocessed data set. In addition, in order to obtain enough data samples for deep learning training, each sample is segmented without overlapping using a 1s sliding window, so that each subject will get 3394 data samples for one experiment.

Table 1. Source information of movie clips

Number	Label	Source of film clip
1	negative	Tangshan Earthquake
2	negative	Back to 1942
3	positive	Lost in Thailand
4	positive	Flirting Scholar
5	positive	Just Another Pandora's Box
6	neutral	World Heritage in China

2.2 Differential Entropy Feature Extraction and Feature Smoothing

In information theory, entropy is called information quantity, which is a measure of uncertainty. Today, the concept of entropy is also used in the analysis of EEG signals, and its original calculation formula is as follows:

$$h(X) = -\int_X f(x)\log[f(x)]dx. \tag{1}$$

Where, X is a random variable, $f(x)$ is a probability density function of X. For the time series X subject to gaussian distribution $N(\mu,\sigma^2)$, its differential entropy can be defined as (2):

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}})dx \\ &= \frac{1}{2}\log(2\pi e\sigma^2). \end{aligned} \tag{2}$$

In a fixed frequency band i, differential entropy can be defined as (3):

$$h_i(X) = \frac{1}{2}\log(2\pi e\sigma_i^2). \tag{3}$$

Where, $h_i(X)$ and σ_i^2 respectively represent the differential entropy and signal variance of the corresponding EEG signal in the frequency band i.

In order to explore the time-frequency information of EEG signals, this paper extracts the differential entropy features from the four rhythms, and uses the linear dynamic system (LDS) [18] to smooth the features, so as to filter out some information unrelated to emotion from the differential entropy features and reduce the presence of noise.

2.3 EEG Channel Electrode Mapping

Fig. 1 and Fig. 2 describe the electrode mapping process of the EEG channels and the conversion process from 2D features to 3D features, respectively, and a 3D differential entropy feature map with a size of 9×9×4 can be obtained. Mapping the complex electrode position relationship in 3D space to a 2D map with a size of 9×9 to simulate the relative position between electrodes, which provides convenience for convolution neural network to explore the spatial position relationship between channels, that is, to extract spatial information. In the 2D map of 9×9, the used EEG channels, i.e. the orange part, are filled with differential entropy features, and the unused EEG channels, i.e. the blue part, are filled with 0.

The EEG signals are first mapped to four rhythms (θ, α, β and γ), and then the differential entropy features corresponding to each frequency band are extracted respectively. The first step in Fig. 2 is to extract the differential entropy features of each frequency band. The second step is to construct a 2D feature map according to the mapping principle in Fig. 1. Finally, the 3D feature map is formed by stacking the 2D feature maps of the four frequency bands. This 3D feature map fully integrates the spatial information and time-frequency information of the EEG signals, which lays a solid foundation for the subsequent use of deep learning for secondary feature extraction and better emotion classification.

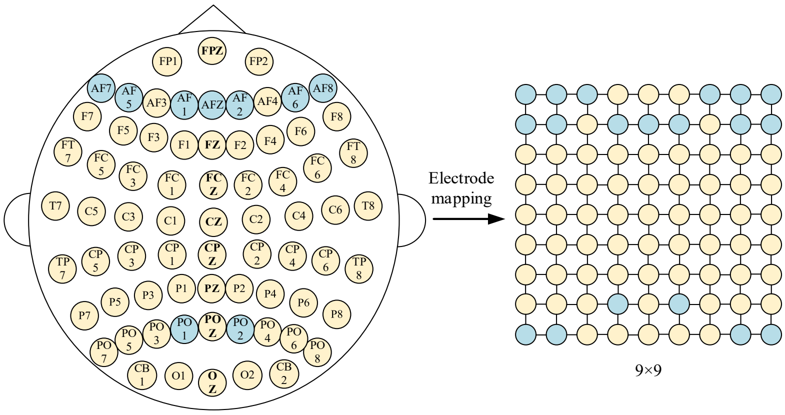


Fig. 1. EEG channel electrode mapping

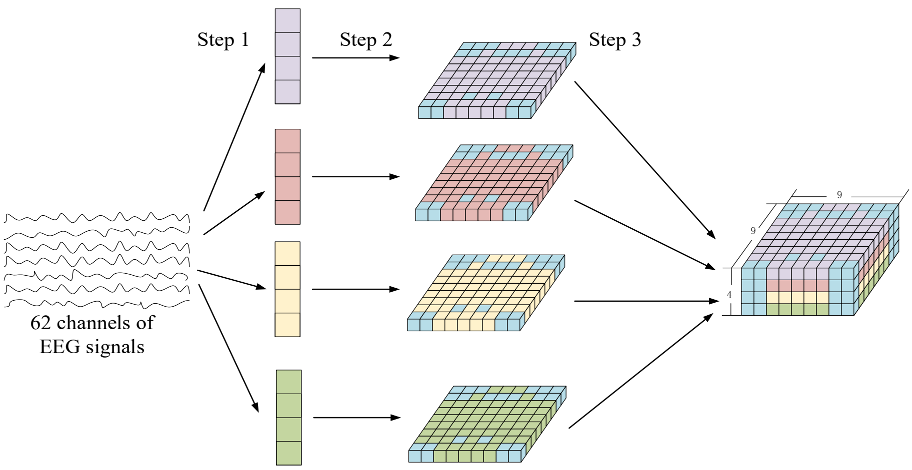


Fig. 2. Transformation process

3 Methods and Models

CNN is one of the important representative algorithms of deep learning, which has been widely used in image recognition and classification [19]. Due to its mature performance in image recognition and classification, many researchers have applied CNN to the research field of emotion recognition, and achieved good results. Based on the research of other researchers, this paper introduces the densely connected convolutional network (DenseNet) into the research of emotion recognition. The characteristics of shallow feature reuse of DenseNet points out a new way for the research of emotion recognition.

3.1 DenseNet

Densely Connected Convolutional Network (DenseNet) [17], different from other deep neural networks, does not seek features from the depth or width of the network, but maximizes the ability of the network through feature reuse. For each of these layers, the features of all previous layers are used as input, and their own features are used as input for all subsequent layers. For each of these layers, the features of all previous layers are used as input, and their own features are used as input for all subsequent layers. In addition, it allows cross-layer connections between any two non-adjacent layers to connect features learned from different layers and enhance the ability of feature representation. DenseNet is mainly composed of dense block, transition layer and bottleneck layer, wherein the dense block realizes the feature reuse function.

In the dense block, the feature map x_l of layer l is calculated by using the feature map $x_0, x_1, \ldots, x_{l-1}$ of the previous layer $l-1$, which can be expressed as (4):

$$x_1 = H_l([x_0, x_1, \ldots, x_{l-1})]. \quad \textbf{(4)}$$

where, $[x_0, x_1, \cdots, x_{l-1}]$ represents the tensor splicing of the feature maps generated from the 0th layer to the $(l-1)$th layer, and $H_l(\cdot)$ is a composite function consisting of 3 sequential operations: Batch Normalization (BN), Rectified Linear Unit (ReLU) and 3×3 Convolution (Conv).

3.2 MSC-DenseNet

Inspired by the Inception structure of GoogLeNet [20], here we introduce the multi-scale convolution kernel parallel merging module (MSC) into the DenseNet network. The MSC module is shown in Fig. 3.

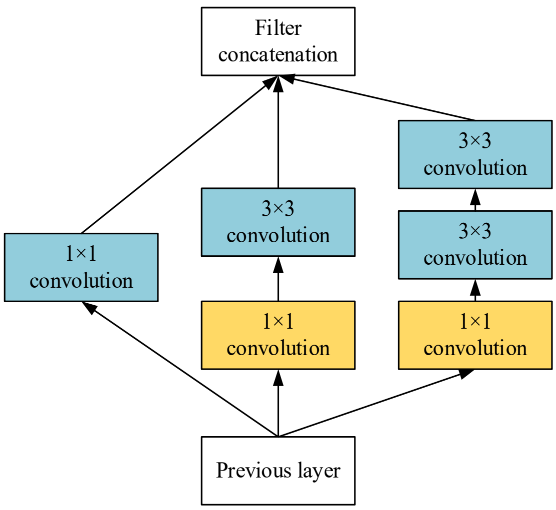


Fig. 3. Multi-scale convolution kernel parallel merging module

MSC adopts convolution kernels of three sizes (1×1, 3×3 and 5×5). Small convolution kernels can obtain more detailed features, while large convolution kernels have larger receptive fields, which can fully consider the relative position relationship between EEG channels and extract the spatial information of EEG signals. Among them, the 5×5 convolution kernel can be replaced by two 3×3 convolution kernels. On the one hand, the stacked 3×3 convolution kernel provides a larger number of activation functions, which increases the nonlinear characteristics of the network. On the other hand, under the condition of the same receptive field, the effect of the neural network can be improved to a certain extent, and the calculation amount of network parameters can be reduced. Each stacked 3×3 convolution kernel is preceded by a 1×1 convolution operation, which can linearly combine the information between different channels to realize cross-channel information interaction and information integration. The overall network structure of MSC-DenseNet is shown in Fig. 4.

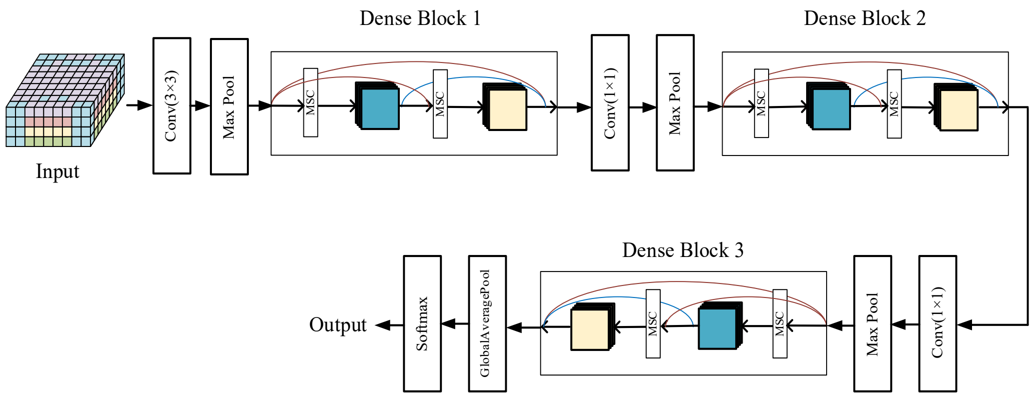


Fig. 4. Network structure of MSC-DenseNet

The whole network is composed of three dense blocks modules, two transition layers and multi-size convolution parallel merging modules. As can be seen from the figure (Fig. 4), 16 convolution kernels of 3×3 size are first used for convolution operation to perform the initial extraction of features; then Batch Normalization (BN) is used, BN can not only play a role in regularization, but also it can effectively increase the learning rate, and normalize the data inside each mini-batch to normalize the output to a normal distribution of N (0,1); finally, use 16 3×3 maximum with a step size of 2 Pooling for down-sampling.

After convolution and pooling, the input data enters the first dense block, that is, the core part of the network. After the processing of the first dense block, it can not only realize the reuse of shallow features, but also realize multi-layer feature extraction with the help of multi-scale convolution in the MSC module. After the first dense block, there is a transition layer, which consists of a 1×1 convolutional layer and a 2×2 maximum pooling layer. After several experimental verifications, it is found that the experimental effect obtained by using 3 dense block modules and 2 transition layers is the best. After the third dense block, global average pooling is used to replace the fully connected layer, and finally the softmax classifier is used to classify the three emotional states of positive, negative and neutral.

3.3 Experimental Environment

The MSC-DenseNet model in this paper is implemented through Python programming under the Tensorflow framework. The experimental environment is Intel Core i5-8500 CPU @ 3.00GHz 16 GB memory, Intel UHD Graphics 630 graphics card, 64-bit Windows10 system.

Table 2 gives the detailed parameter setting information of the MSC-DenseNet structure. Among them, the learning rate determines the step size update in the entire optimization process. If a larger learning rate is used in the early stage of training, the forward step size will be longer, and the gradient descent can be performed at a faster speed. In the later stage of training, gradually reducing the learning rate is conducive to the convergence of the algorithm, making it close to the optimal value. In this paper, piecewise constant attenuation is adopted. The learning rate used in the first 3600 batch sizes during training is 0.00001. Using a larger learning rate in the initial stage not only speeds up the gradient descent, but also saves a certain amount of training time and increases the training rate. After that, the learning rate is gradually reduced. The learning rate of 3600-5000 batch size is set to 0.000005, the learning rate of 5000-7000 batch size is set to 0.0000025, and finally the convergence of the algorithm is achieved with a learning rate of 0.000001.

Table 2. Parameter Settings of MSC-DenseNet

Description	Parameter
Batch size	128
Activation function	ReLU
Classifier	softmax
Loss function	cross-entropy loss
Optimizer	Adam [21]
Iterations	1000
Pooling	maximum pooling, global average pooling
Learning rate	[0.00001, 0.000005, 0.0000025, 0.000001]

4 Results and Discussion

This section uses the data of 4 rhythms to conduct experiments separately to find the rhythm that can best represent emotional state, and compare it with the effect obtained by using all the data of the 4 rhythms in combination.

4.1 Evaluation Indicators

In this paper, the accuracy rate (Accuracy), precision rate (Precision), recall rate (Recall), F1-score and confusion matrix are used to evaluate the model performance, which are defined as follows:

1) Accuracy: Accuracy refers to the ratio of the number of correctly classified samples to the total number of samples

$$A = \frac{TP + TN}{TP + FP + TN + FN}. \quad \textbf{(5)}$$

2) Precision: Precision refers to the proportion of correctly predicted positive data to the total predicted positive data

$$P = \frac{TP}{TP + FP}. \quad \textbf{(6)}$$

3) Recall: Recall refers to the ratio of correctly predicted positive data to the actual positive data

$$R = \frac{TP}{TP + FN}. \quad \textbf{(7)}$$

4) F1-score: F1-score refers to the harmonic mean of precision and recall. The macro average is used in this paper. Firstly, the Precision and Recall of each category are calculated to calculate their respective F1-scores, and then average the F1-scores obtained by the three categories to obtain the final F1- score. Its maximum value is 1, and its minimum value is 0. The larger the value, the better the model.

$$F = \frac{2 \times P \times R}{P + R}. \quad \textbf{(8)}$$

In Equation (5) to Equation (7), for positive emotions, TP represents the number of samples predicted by the model to be positive emotions, but actually positive emotions; FP represents the number of samples predicted by the model to be positive emotions, but actually negative and neutral; TN represents the number of samples predicted by the model is the number of samples that are negative and neutral, but are actually negative and neutral; FN represents the number of samples predicted by the model to be negative and neutral, but actually positive. Neutral emotions, negative emotions are similar to positive emotions.

5) Confusion matrix: Also known as likelihood table or error matrix, is a standard format for accuracy evaluation, which is expressed in matrix form of N rows and N columns. It is not only limited to the analysis of the accuracy rate, but can bring more analysis, present the visual effect of the algorithm performance, and intuitively see the classification and recognition errors.

4.2 Single-Subject Experiment

Before the start of the experiment, the data samples of each subject are randomly divided into three parts: training set, validation set and test set, of which the training set account for 70%, the validation set account for 20%, and the test set account for 10%.

Table 3. Results of single-subject experiment

Subject	θ		α		β		γ		All band	
	F1-score	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score	Accuracy
1	0.9761	0.9765	0.9643	0.9647	0.9766	0.9765	0.9758	0.9765	0.9885	0.9882
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0.9740	0.9735	0.9733	0.9735	0.9733	0.9735	0.9759	0.9765	0.9788	0.9794
4	0.9790	0.9794	0.9469	0.9471	0.9821	0.9824	0.9883	0.9882	0.9883	0.9882
5	0.9915	0.9912	0.9763	0.9765	0.9883	0.9882	0.9883	0.9882	0.9942	0.9941
6	0.9885	0.9882	0.9857	0.9853	0.9729	0.9735	0.9701	0.9706	0.9788	0.9794
7	0.9621	0.9618	0.9736	0.9735	0.9762	0.9765	0.9819	0.9824	0.9859	0.9853
8	0.9640	0.9647	0.9942	0.9941	0.9759	0.9765	0.9788	0.9794	0.9847	0.9853
9	0.9589	0.9588	0.9644	0.9647	0.9764	0.9765	0.9760	0.9765	0.9762	0.9765
10	0.9829	0.9824	0.9827	0.9824	0.9734	0.9735	0.9734	0.9735	0.9788	0.9794
11	0.9640	0.9647	0.9294	0.9294	0.9859	0.9853	0.9883	0.9882	0.9790	0.9794
12	0.9644	0.9647	0.9576	0.9588	0.9816	0.9824	0.9816	0.9824	0.9911	0.9912
13	0.9942	0.9941	0.9847	0.9853	0.9703	0.9706	0.9792	0.9794	0.9791	0.9794
14	0.9801	0.9794	0.9692	0.9706	0.9732	0.9735	0.9820	0.9824	0.9788	0.9794
15	0.9883	0.9882	0.9827	0.9824	0.9853	0.9853	0.9915	0.9912	0.9913	0.9912
Average	0.9779	0.9778	0.9723	0.9725	0.9794	0.9796	0.9821	0.9824	0.9849	**0.9851**

Table 3 presents the F1-score and accuracy rate of the 4 rhythms used alone and all rhythms used together for a single subject. Among them, no matter which rhythm, the accuracy rate of the 2th subject can reach 100%, indicating that the model in this paper can obtain ideal classification accuracy in the 2th subject. Judging from the average F1-score and average accuracy rate of the 15 subjects, the F1-scores of the 4 rhythms are all above 0.97, especially the F1-score of the γ rhythm reaches 0.9821, and its accuracy also reaches 98.24%, which verified the superiority of shallow feature reuse and multi-scale convolution kernel. Compared with the other three rhythms, the classification accuracy of the α rhythm is relatively low, and in the 4th subject, the accuracy is only 94.71%, while the lowest accuracy of θ, β and γ rhythm is 95.88%, 97.06% and 97.06%, respectively, which are higher than α rhythm.

In addition, it can be seen from Table 3 that the F1-score and accuracy obtained by the γ rhythm are similar to the results obtained by using all rhythms together, indicating that better classification results can also be obtained when only the γ rhythm is used, which is consistent with the basic knowledge of EEG signals. The frequency range of the γ rhythm is about 31-50 Hz, which is the highest frequency component in the EEG signals. When a signal in this frequency band appears, people are usually in a very excited state, or are strongly stimulated. So γ rhythm can capture effective information under both positive and negative stimuli, and the accuracy rate is the highest when γ rhythm is used. The α rhythm is more common in adults with eyes closed and restful, therefore, the accuracy rate obtained by it is relatively low. The θ rhythm is more prominent in adults who are frustrated or depressed and mentally ill.

In conclusion, under normal circumstances, γ rhythm contains more emotional state information. In other words, in order to save training time, γ rhythm can be used to complete the emotion recognition task.

4.3 All-Subject Experiment

The data samples of all subjects are taken as experimental data. There are 3394 data samples for each subject, a total of 15 subjects, and the data set is divided in the same way as that of the previous single-subject experiment.

Table 4 shows the results of the experiment. Among the four rhythms, the β rhythm has the highest accuracy of 98.47%, which is similar to 98.45% of the γ rhythm. Generally, the β rhythm occurs when a person is nervous or emotional, and it is similar to γ rhythm in that they occur when people are emotionally excited. With all rhythms used, the resulting accuracy is 98.68%, slightly higher than that of the β rhythm. This further confirms that a single rhythm can be used to replace all the rhythms mentioned above.

Table 5 presents the confusion matrix for each rhythm, with each column representing the predicted category and each row representing the actual category. Assume that Label 0 represents the neutral emotion, Label 1 represents the positive emotion, Label 2 represents the negative emotion, and assume that P represents the predicted emotion and A represents the actual emotion. Of which, the θ rhythm has the worst prediction result for neutral emotions, and its error rate is the highest. Compared with other rhythms, the α rhythm has a higher error rate in predicting positive emotions and negative emotions, and most of the positive emotions and negative emotions are wrongly predicted as neutral emotions. This is consistent with the theory mentioned above that α rhythm usually occurs in the resting state of the adult.

Table 4. Results of all-subject experiment

	θ	α	β	γ	All band
F1-score	0.9797	0.9725	0.9847	0.9845	0.9868
Accuracy	0.9798	0.9725	**0.9847**	0.9845	**0.9868**

Table 5. Confusion matrix

(a) θ rhythm

A P	0	1	2
0	1643	18	17
1	17	1729	19
2	26	6	616

(b) α rhythm

A P	0	1	2
0	1653	19	6
1	41	1706	18
2	43	13	1592

(c) β rhythm

A P	0	1	2
0	1661	11	6
1	26	1729	10
2	18	7	1623

(d) γ rhythm

A P	0	1	2
0	1664	6	8
1	29	1725	11
2	21	4	1623

(e) all band

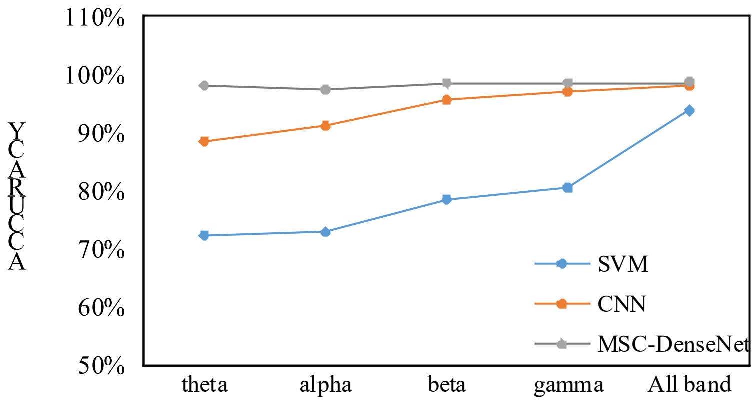
A P	0	1	2
0	1662	3	13
1	22	1720	23
2	6	0	1642

The prediction results of using the β rhythm alone, the γ rhythm alone, and using all rhythms are similar, which also indicates that if less parameters and workload are required, either β rhythm or γ rhythm alone can be used for emotion recognition.

It can be seen from the above experimental results that the MSC-DenseNet model proposed in this paper shows obvious advantages. The average accuracy of the single-subject experiment can reach 98.51%, and the accuracy of the all-subject experiment can reach 98.68%, which fully shows that the 3D feature map combined with feature reuse has a good classification performance in the research on EEG emotion recognition.

4.4 Comparison of Different Algorithms

In order to verify the effectiveness of the proposed method, a representative machine learning algorithm-SVM and a representative deep learning algorithm-CNN are selected to compare and analyze the performance of the proposed method in the all-subject experiment. Among them, the SVM kernel function uses the Gaussian kernel function, and the penalty factor is set to 1. CNN selects three convolution kernels of different sizes, 1×1, 3×3, and 5×5. 3×3 adopts maximum pooling, and globally average pooling replaces the fully connected layer, and finally softmax outputs the classification result. Fig. 5 visually shows the accuracy results of SVM, CNN and our method in the all-subject experiment.



	theta	alpha	beta	gamma	All band
SVM	72.30%	72.76%	78.45%	80.42%	93.69%
CNN	88.39%	91.00%	95.54%	96.98%	97.84%
MSC-DenseNet	97.98%	97.25%	98.47%	98.45%	98.68%

Fig. 5. Comparison results of different algorithms

As can be seen from Fig. 5, the accuracy obtained by the method in this paper is the highest, followed by CNN, and SVM is the worst, with the accuracy of only 93.69%, indicating that compared with machine learning, deep learning has certain advantages in the field of EEG emotion recognition. Using the CNN model with multi-scale convolution kernel, the accuracy can reach 97.84% when all rhythms are used, but the accuracy of the proposed method on each rhythm has certain advantages, and the highest accuracy can reach 98.68%, indicating the effectiveness of multi-scale convolution kernel combined with feature reuse.

4.5 Comparison of Similar Studies

At present, many EEG emotion recognition methods have been proposed. In this section, we compare our method with other existing methods that also use the SEED data set. The pre-processing of EEG signals in reference [14] is the same as this paper, extracting the differential entropy features of five frequency bands of EEG signals, and then selecting the EEG channel that can best express emotional state, and using deep belief network to achieve classification. As can be seen from Table 6, the accuracy obtained by using DBN model is more than 10 percentage points lower than the results in this paper, indicating the superiority of the improved DenseNet model in this paper. Reference [22] converts one-dimensional EEG time domain information into two-dimensional differential entropy frequency domain features according to the electrode positions of EEG channels, and uses hierarchical convolutional neural network (HCNN) for secondary feature extraction and classification. Reference [22] takes into account the information contained in the location of the channel electrode as in this paper, but the accuracy obtained is not as high as in this paper. Reference [23] first uses wavelet packet transform (WPT) to decompose the signal to construct a two-dimensional structure sample, and then votes and weights the 6 designed CNN models with different depths to establish an integrated model, which achieves a classification accuracy of 93.12% in a single-subject experiment. Reference [24] proposes a new deep learning classification model—3D convolutional attention neural network (3DCANN), which consists of a spatiotemporal feature extraction module and an EEG channel attention weight learning module, and obtains a classification Accuracy of 97.35%. Reference [25] proposes a new adaptive optimization space-frequency differential entropy (AOSFDE) feature, which firstly selects channels according to the lead importance, and then optimizes the differential entropy feature by sparse regression algorithm. For the 15 subjects, the average Accuracy rates of the three emotion binary classification scenarios of positive/negative, positive/neutral, and neutral/negative reaches 91.80%, 93.30%, and 85.10%, respectively. The data processing method and classifier design of Reference [24-25] are different from ours, but the model in this paper can get relatively good results, which further verifies the advantages of the method in this paper. Table 6 presents the experimental comparison results between the method in this paper and these methods mentioned above.

In a word, the classification accuracy of the MSC-DenseNet method on the SEED data set is higher than that of other literature methods, indicating that the method in this paper achieves better results on the SEED data set and has a particular value in emotion recognition.

Table 6. Comparison of similar studies

Method	Input	Model	Accuracy
Reference [14]	DE	DBN	86.08%
Reference [22]	DE	HCNN	88.20%
Reference [23]	WPT	integrated network	93.12%
Reference [25]	AOSFDE	SVM	93.30%
Reference [24]	denoising, outlier processing	3DCANN	97.35%
Proposed method	DE	MSC-DenseNet	**98.51%**

5 Conclusion

With the successful application of deep learning algorithms in many fields, more and more researchers try to apply deep learning to the field of EEG emotion recognition, and have achieved a series of research results. In this regard, we have also made some attempts. In this article, a three-dimensional differential entropy feature map is constructed by combining the spatial characteristics and time-frequency characteristics of EEG signals, and then based on the classical DenseNet, we propose an improved DenseNet model by adding a multi-scale convolution kernel module into the classical DenseNet. An average classification accuracy of 98.51% is obtained in the single-subject experiment, and a classification accuracy of 98.68% is obtained in the all-subject experiment. The experiment results show that the technology has certain advantages for better feature extraction and classification recognition. In addition, the introduction of feature reuse in the DenseNet provides a new idea for further research on EEG emotion recognition, and has good research potential. In order to prove the stability of the model, we conducted experiments on 15 subjects respectively, and achieved satisfactory results. At the same time, we also conducted comparative experiments with traditional machine learning, and the experimental results show that the deep learning method has better performance in the field of emotion recognition.

However, only extracting differential entropy feature is too simple, which is not enough to fully reflect the emotional information contained in the EEG signals. We will consider the combination of multiple features and change the input format of the network to better improve the performance of the classifier. Besides, some studies have shown that there is a certain redundancy of EEG channels. In future research, we will try to use a certain algorithm to find the key EEG channels, so as to extract more discriminative information and further improve the emotion recognition rate.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 61561004) and the Science and Technology Project of Jiangxi Provincial Department of Education of China (No. GJJ201408).

References

[1] T. Wilaiprasitporn, A. Ditthapron, K. Matchaparn, T. Tongbuasirilai, N. Banluesombatkul, E. Chuangsuwanich, Affective EEG-Based Person Identification Using the Deep Learning Approach, IEEE Transactions on Cognitive and Developmental Systems, 12(3)(2020) 486-496.

[2] C.-H. Wang, Z.-H. Wen, H.Wang, M. Yu, P.-C. Yang, L. Yang, J. Dai, X.-S. Cao, Research on emotional state recognition algorithm based on EEG signal, Chinese Medical Equipment Journal 41(2)(2020) 20-25.

[3] L. Breiman, Random Forests, Mach Learning 45(1)(2001) 5-32.

[4] R.-M. Mehmood, H.-J. Lee, Emotion Classification of EEG Brain Signal Using SVM and KNN, in: Proc. 2015 IEEE International Conference on Multimedia and Expo Workshops (icmew), 2015.

[5] T. Tuncer, S. Dogan, A. Subasi, A New Fractal Pattern Feature Generation Function Based Emotion Recognition Method Using EEG, Chaos, Solitons & Fractals 144(2021) 110671.

[6] R.-N. Duan, J.-Y. Zhu, B.-L. Lu, Differential Entropy Feature for EEG-based Emotion Classification, in: Proc. 2013 6th International IEEE/EMBS Conference Neural Engineering (ner), 2013.

[7] R. Alazrai, R. Homoud, H. Alwanni, M.-I. Daoud, EEG-based Emotion Recognition Using Quadratic Time-frequency Distribution, Sensors 18(8)(2018) 2739.

[8] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, Deap: a Database for Emotion Analysis; using Physiological Signals, IEEE Transactions on Affective Computing 3(1)(2012) 18-31.

[9] Y. Zhang, X. Ji, S. Zhang, An Approach to EEG-based Emotion Recognition Using Combined Feature Extraction Method, Neuroscience Letters 633(2016) 152-157.

[10] W.-L. Zheng, H.-T. Guo, B.-L. Lu, Revealing Critical Channels and Frequency Bands for Emotion Recognition From EEG with Deep Belief Network, in: Proc. 2015 7th International IEEE/EMBS Conference on Neural Engineering (ner), 2015.

[11] C.-L. Liu, W.-H. Hsaio, Y.-C. Tu, Time Series Classification with Multivariate Convolutional Neural Network, IEEE Transactions on Industrial Electronics 66(6)(2019) 4788-4797.

[12] M. Ramzan, S. Dawn, Fused CNN-LSTM Deep Learning Emotion Recognition Model Using Electroencephalography Signals, International Journal of Neuroscience 133(6)(2023) 587-597.

[13] T. Song, W. Zheng, P. Song, Z. Cui, EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks, IEEE Transactions on Affective Computing 11(3)(2020) 532-541.

[14] W.-L. Zheng, B.-L. Lu, Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks, IEEE Transactions on Autonomous Mental Development 7(3)(2015) 162-175.

[15] M.-A. Asghar, M.-J. Khan, M. Rizwan, R.-M. Mehmood, S.-H. Kim, An Innovative Multi-model Neural Network Approach for Feature Selection in Emotion Recognition Using Deep Feature Clustering, Sensors 20(13)(2020) 3765.

[16] X. Li, E.-J. Cai, Y.-X. Tian, X.-Q. Sun, M.-D. Man, An improved electroencephalogram feature extraction algorithm and its application in emotion recognition, Journal of Biomedical Engineering 34(4)(2017) 510-517.

[17] G. Huang, Z. Liu, L. Maaten, K.-Q. Weinberger, Densely Connected Convolutional Networks, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[18] X. Wang, D. Nie, B.-L. Lu, Emotional State Classification From EEG Data Using Machine Learning Approach, Neurocomputing 129(2014) 94-106.

[19] T. Phan, S. Kim, H. Yang, G. Lee, EEG-based Emotion Recognition By Convolutional Neural Network with Multi-scale Kernels, Sensors 21(15)(2021) 5092.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going Deeper with Convolutions, in: Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[21] D. Kingma, J. Ba, Adam: a Method for Stochastic Optimization. <https://arxiv.org/abs/1412.6980>, 2014 (accessed 31.10.2021).

[22] J. Li, Z. Zhang, H. He, Implementation of EEG Emotion Recognition System Based on Hierarchical Convolutional Neural Networks, in: Proc. International Conference on Brain Inspired Cognitive Systems, 2016.

[23] C. Wei, L.-L. Chen, A. Zhang, Emotion Recognition of EEG Based on Ensemble Convolutional Neural Networks, Journal of East China University of Technology: Natural Science Edition 45(4)(2019) 614-622.

[24] S. Liu, X. Wang, L. Zhao, B. Li, W. Hu, J. Yu, Y.-D. Zhang, 3DCANN: a Spatio-temporal Convolution Attention Neural Network for EEG Emotion Recognition, IEEE Journal of Biomedical and Health Informatics 26(11)(2022) 5321-5331.

[25] M.-M. Miao, B.-G. Xu, W.-J. Hu, A. Wang, A. Song, Emotion EEG recognition based on the adaptive optimized spatial-frequency differential entropy, Chinese Journal of instrumentation 42(3)(2021) 221-230.