

## Article

# Multidimensional Feature in Emotion Recognition Based on Multi-Channel EEG Signals

Qi Li <sup>1</sup>, Yunqing Liu <sup>1,\*</sup>, Quanyang Liu <sup>1</sup>, Qiong Zhang <sup>1</sup>, Fei Yan <sup>1</sup>, Yimin Ma <sup>1</sup> and Xinyu Zhang <sup>2</sup>

<sup>1</sup> Department of Electronics and Information Engineering, Changchun University of Science and Technology, Changchun 130000, China

<sup>2</sup> Economics School, Jilin University, Changchun 130000, China

\* Correspondence: mzlyq@cust.edu.cn

**Abstract:** As a major daily task for the popularization of artificial intelligence technology, more and more attention has been paid to the scientific research of mental state electroencephalogram (EEG) in recent years. To retain the spatial information of EEG signals and fully mine the EEG timing-related information, this paper proposes a novel EEG emotion recognition method. First, to obtain the frequency, spatial, and temporal information of multichannel EEG signals more comprehensively, we choose the multidimensional feature structure as the input of the artificial neural network. Then, a neural network model based on depthwise separable convolution is proposed, extracting the input structure's frequency and spatial features. The network can effectively reduce the computational parameters. Finally, we modeled using the ordered neuronal long short-term memory (ON-LSTM) network, which can automatically learn hierarchical information to extract deep emotional features hidden in EEG time series. The experimental results show that the proposed model can reasonably learn the correlation and temporal dimension information content between EEG multi-channel and improve emotion classification performance. We performed the experimental validation of this paper in two publicly available EEG emotional datasets. In the experiments on the DEAP dataset (a dataset for emotion analysis using EEG, physiological, and video signals), the mean accuracy of emotion recognition for arousal and valence is 95.02% and 94.61%, respectively. In the experiments on the SEED dataset (a dataset collection for various purposes using EEG signals), the average accuracy of emotion recognition is 95.49%.

**Keywords:** EEG; emotion recognition; multidimensional feature; depthwise separable convolution



**Citation:** Li, Q.; Liu, Y.; Liu, Q.; Zhang, Q.; Yan, F.; Ma, Y.; Zhang, X. Multidimensional Feature in Emotion Recognition Based on Multi-Channel EEG Signals. *Entropy* **2022**, *24*, 1830. <https://doi.org/10.3390/e24121830>

Academic Editor: Daya Shankar Gupta

Received: 27 October 2022  
Accepted: 8 December 2022  
Published: 15 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Emotion recognition is a multi-curricular field of study that combines cognitive science, social psychology, and electronic information science [1–3]. It is a complex problem and a hot spot in the cognitive science industry. Emotion identification methods are divided into two categories. One is emotion identification based on non-physiological data signals, such as words, language expressions, facial emotions, gestures, etc. [4]. However, this method is relatively subjective and cannot ensure the authenticity and effectiveness of emotions. Another emotion recognition is based on physiological data signals [5]. With the development of wearable and non-invasive physiological signal acquisition devices, the real-time performance and accuracy of the physiological signal collection have been greatly improved, promoting the development of physiological signals in emotion recognition. Nervous system physiology and social psychology studies have shown that brain electrical signals can reflect various brain wave thematic activities and effects of the brain. Since the occurrence of emotion is related to the activities of the cerebral cortex, the brain's electrical signals can also reflect the efficient information content of people's emotional state [6]. Because of the advantages of strong universality and high classification and identification accuracy, brain electronic signals have been gradually introduced into the field of emotion identification research.

Traditional EEG emotion identification methods usually capture handcrafted emotion characteristics, followed by shallow machine learning for emotion classification. The main methods include the K-nearest neighbor algorithm, mixed Gaussian algorithm, artificial neural network method, Bayesian algorithm, support vector machine, and improvements in these methods [7]. At this stage, the key characteristics of emotional discrimination of EEG signals are time-domain characteristics, frequency domain characteristics, and time frequency domain characteristics [8]. Time domain features are mainly used to directly extract the waveform features of EEG signals, including event related potential (ERP), signal statistics, energy, power, high order zero crossing analysis, non-stationary index (NSI), and fractal dimension (FD) [9]. Since the time-domain feature cannot display the frequency information of the signal, the researchers added frequency domain analysis. First, the original time domain signal is converted into the frequency domain to obtain the frequency spectrum. Then, the frequency band is decomposed into five sub-frequency bands ( $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$ ,  $\gamma$ ) that are closely related to human psychological activities [10]. Extract features from these five frequency bands. These features usually include power spectral density (PSD) [11], differential entropy (DE) [12,13], differential asymmetry (DASM) [14], and rational asymmetry (RASM) [15,16]. EEG signal is a non-stationary and non-linear random signal [17]. Since the scope of the Fourier transform is the entire time domain, it lacks local generalization ability. It is impossible to confirm the time corresponding to each frequency domain component of the non-stationary signal, so the time frequency domain is introduced to analyze the EEG signal [18]. Usually, the time frequency domain signal transformation mainly adopts methods, such as short time Fourier transform (STFT), wavelet transform (WT), and wavelet packet transform (WPT) or Hilbert–Huang [19]. However, these features need to be designed and selected according to specific purposes, which also limits their performance in EEG emotion classification.

The rapid development of deep learning technology in computer vision [20], speech [21], and natural language processing [22], methods based on deep neural networks, have begun to get the attention of researchers in this field. EEG usually contains much helpful information and has a wide range of applications. The related work is summarized in Table 1. Komolovaitè et al. [23] contributed to the application of Alzheimer’s disease. They used convolutional neural networks to learn discriminant features in EEG to classify familiarity and emotion. Thammasan used deep belief networks combined with fractal dimension, power spectrum, and discrete wavelet transform features to classify the emotional categories of EEG under music stimulation [24]. Tripathi gives the effect of the convolutional neural network in the two classifications (high/low) and three classifications (high/normal/low) of EEG emotions [25]. Elham scientifically investigated a three-dimensional convolutional neural network (3D-CNN) for emotion discrimination based on multichannel EEG data information, and the discrimination accuracy of valence and arousal were 87.44% and 88.49%, respectively [26]. Yang obtained the differential entropy characteristics of different frequency bands and applied the three-dimensional convolutional neural network to classify emotions on DEAP data at the level of valence and arousal. The mean accuracy is 89.45% and 90.24%, respectively [27].

Although the deep learning emotion recognition model has achieved higher accuracy than the shallow model, how to integrate more helpful EEG information to perform emotion recognition better and improve the validity and reliability of predictions made with machine learning remain challenging [28]. Firstly, regarding the relationship between EEG signal frequency bands and emotion types, Zheng et al. [29] applied a deep belief network model to select meaningful frequency bands of key channels by training the weight distribution of DBNs. The results showed that the Gamma and Beta bands exhibited better classification performance in emotion recognition. Secondly, EEG signal acquisition equipment generally includes 32 channels, 62 channels, and 128 channels. It may contain some EEG signals that have nothing to do with emotions, which can bring noise and interference to emotion recognition. In response to this problem, Meng-meng extracted the frequency domain features on each electrode and then combined the common

spatial patterns (CSP) algorithm with the spatial domain to calculate the spatial frequency domain features. The average accuracy rate in the three-category emotion recognition has reached 87.54% [30]. Finally, combining the temporal and spatial characteristics of EEG signals, Liu designed the RCNN LSTM hybrid deep neural network entity model, using the circulatory system neuron network to obtain the characteristics of brain electrical signals, then integrating the information content of the LSTM time concept characteristics to develop a fusion model [31].

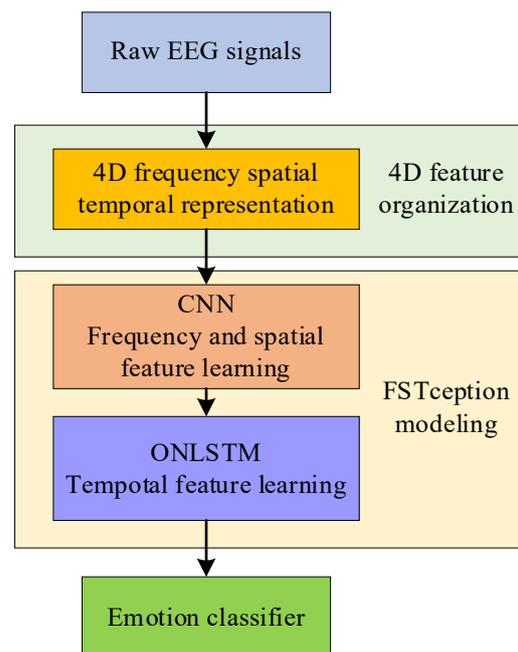
**Table 1.** The detailed comparison to other works.

Article	Method	Application	Dataset	Accuracy (%)
Komolovaitė et al. [23]	EEGNet SSVEP	Alzheimer research	Figshare website	50.2
Thammasan et al. [24]	Deep Belief Network (DBN)	Emotion recognition in music listening	15 recruited healthy students (valence/arousal emotion types)	82.42/88.24
Tripathi et al. [25]	Convolutional Neural Network (CNN)	Automatic emotion recognition	DEAP (valence/arousal emotion types)	73.36/81.41
Salama et al. [26]	3-Dimensional Convolutional Neural Networks (3D-CNN)	Emotion recognition	DEAP (valence/arousal emotion types)	88.49/87.44
Yang et al. [27]	Convolutional Neural Network (CNN)	Automatic emotion recognition	DEAP (valence/arousal emotion types)	90.24/89.45
Zheng et al. [29]	Deep Belief Network (DBN)	Emotion recognition	15 recruited subjects	86.65
Meng-meng et al. [30]	Common Spatial Pattern (CSP)	Emotion recognition	6 recruited healthy students	87.54
Liu et al. [31]	Recurrent Convolutional Neural Network and Long Short Term Memory (RCNN-LSTM)	Automatic emotion recognition	DEAP	96.63

In summary, the frequency, time, and space characteristics of EEG signals are critical for emotion identification. However, most of the literature only considers one or two of these three functions. Therefore, it is possible to adaptively capture important frequency, spatial and temporal characteristics in a unified network by integrating information from different fields, which is of crucial significance for improving the accuracy of EEG signal emotion recognition. The method proposed in this paper makes full use of frequency, temporal, and spatial information features while also making an outstanding contribution to the computational performance of the network, filling a gap in integrating information while improving computational performance in EEG signal emotion recognition.

A single feature extracted by traditional emotion research methods is not enough to describe the rich information contained in EEG, and a simple feature combination may cause redundant features in the feature space, thereby affecting model accuracy and increasing model complexity. This paper points out a new method of EEG emotion discrimination based on multi-domain feature fusion. The general flowchart of the proposed method is shown in Figure 1. Firstly, the EEG signal is converted into a 4D feature structure composed of several time slices, which contain frequency, spatial and temporal information. Secondly, because the EEG data training samples are limited and the feature dimension is large, we use the depthwise separable convolution for dimensionality reduction, which balances the width and depth of the network and can reduce the amount of computation. The frequency and spatial features of EEG signals are fully extracted by using the feature of

depthwise separable convolution. Finally, since the traditional LSTM does not consider the order information of updating neurons during feature learning, it means that each neuron is independent. Therefore, the knowledge learned from the perspective of spatial structure is discrete, which makes it difficult to distinguish the information level between neurons, which causes the loss of information at the spatial level. Therefore, this paper uses ON-LSTM to increase the induced bias towards structural information by sorting neurons, which enables the model to integrate the emotional hierarchy of EEG signals into LSTM through a specific neuron order without destroying its sequential form. Thereby, the feature information of the emotional hierarchy of the EEG signal is learned to express the hidden emotional information in the time series.



**Figure 1.** The general flowchart of the proposed method.

The key dedications in the text are as follows:

- We adopt a four-dimensional (4D) feature structure, including the frequency, space, and time information of EEG signals, as the feature input of EEG signal emotion recognition.
- We use the FSTception model based on a depthwise separable convolutional neural network to solve the problems of a few training samples of EEG data, large feature dimensions, and feature extraction.
- In particular, we adopt the ON-LSTM structure to deal with the deep emotional feature extraction hidden on time series in input features with a 4D structure.

Other parts of the organization are listed below. We will introduce the method proposed in this article in Section 2. Section 3 introduces the data set, experimental settings, results, and discussion. Finally, in Section 4, we conclude that we have achieved advanced performance on both the SEED and DEAP data sets.

## 2. Materials and Methods

This section mainly describes the 4-dimensional feature organization, FSTception model, and classification. The small details of each part will be introduced in detail.

### 2.1. 4D Frequency Spatial Temporal Representation

The original EEG signal can be divided into several different frequency patterns according to the in-band correlation of different behavioral states. The frequency modes and corresponding characteristics of EEG are listed in Table 2. The awareness increases

with the increase of the frequency band. We believe that emotions are generated when humans are highly awake. This statement also conforms to our intuitive understanding. If a person is unconscious, he/she is unlikely to have specific emotions. To extract the feature entirely, we constructed a 4D feature structure to integrate the three information features of EEG signals, as shown in Figure 2. In data processing, just like the previous work [26,32], because the Butterworth filter has the smoothest response curve in the passband and is simple in design and easy to implement, we use the Butterworth filter to divide the EEG signal into five frequency bands ( $\delta$ [1~4 Hz],  $\theta$ [4~8 Hz],  $\alpha$ [8~13 Hz],  $\beta$ [13~30 Hz],  $\gamma$ [30~50 Hz]). The frequency response of the filter is shown in Figure 3. After Butterworth filtering decomposition, the EEG data of a participant varied from  $40 \times 8064 \times 32$  (video  $\times$  sample  $\times$  channel) to  $40 \times 8064 \times 5 \times 32$  (video  $\times$  sample  $\times$  band  $\times$  channel). The EEG raw signal and its five frequency bands in the DEAP dataset are shown in Figure 4.

**Table 2.** The frequency modes and corresponding characteristics of EEG.

Patterns	Frequency	Brain State	Awareness
( $\delta$ ) Delta	1–4 Hz	Deep sleep pattern	Lower
( $\theta$ ) Theta	4–8 Hz	Light sleep pattern	Low
( $\alpha$ ) Alpha	8–13 Hz	Closing the eyes, relax state	Medium
( $\beta$ ) Beta	13–30 Hz	Active thinking, focus, high alert, anxious	High
( $\gamma$ ) Gamma	30–50 Hz	Mentally active and hypertensive	Higher

Since our focus is on identifying emotional states at the segmented level, the EEG signal is cut into  $l$  segments of degree  $n$ . Our goal is to categorize these fragments into the correct tags. The segmented EEG data is converted into  $40 \times l \times n \times 5 \times 32$  (video  $\times$  segment  $\times$  length  $\times$  band  $\times$  channel). When 0.5s is taken as the cutting window,  $l = 120$  is the number of segments to be cut, and  $n = 64$  is the sample points contained in a segment. Differential entropy can accurately reflect the response degree of the stimulus because of its ability to distinguish the balance between the low frequency band and the high frequency band of the EEG mode [33]. Therefore, the differential entropy algorithm is used in this experiment to represent the complexity of a fixed length EEG signal. Differential entropy is the generalized form of Shannon's information entropy on continuous variables, where  $X$  is a random variable, and its original expression is

$$H(X) = - \int f(x) \log[f(x)] dx \quad (1)$$

where  $f(x)$  is the probability density function of the  $X$ . Shi et al. [34] proved that the differential entropy of a random variable approximately obeying the Gaussian distribution  $N(\theta, \sigma^2)$  is equal to the logarithm of its energy spectrum in a specific frequency band. The specific expression is as follows

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

The original EEG signal is complex and does not obey the normal distribution, but the bandpass filtering of a specific range approximates the Gaussian distribution. In order to test whether the EEG after band-pass filtering follows the Gaussian distribution, 4800 EEG segments of 0.5 s length in each sub-band signal were randomly selected from 32 subjects. The Kolmogorov–Smirnov test was then applied to all 4800 EEG data segments to test whether each sub-band was subject to Gaussian distribution, and the significance level  $\alpha$  was set at 0.05. It can be proven that the probability of a sub-band signal satisfying the Gaussian distribution hypothesis is greater than 90%.

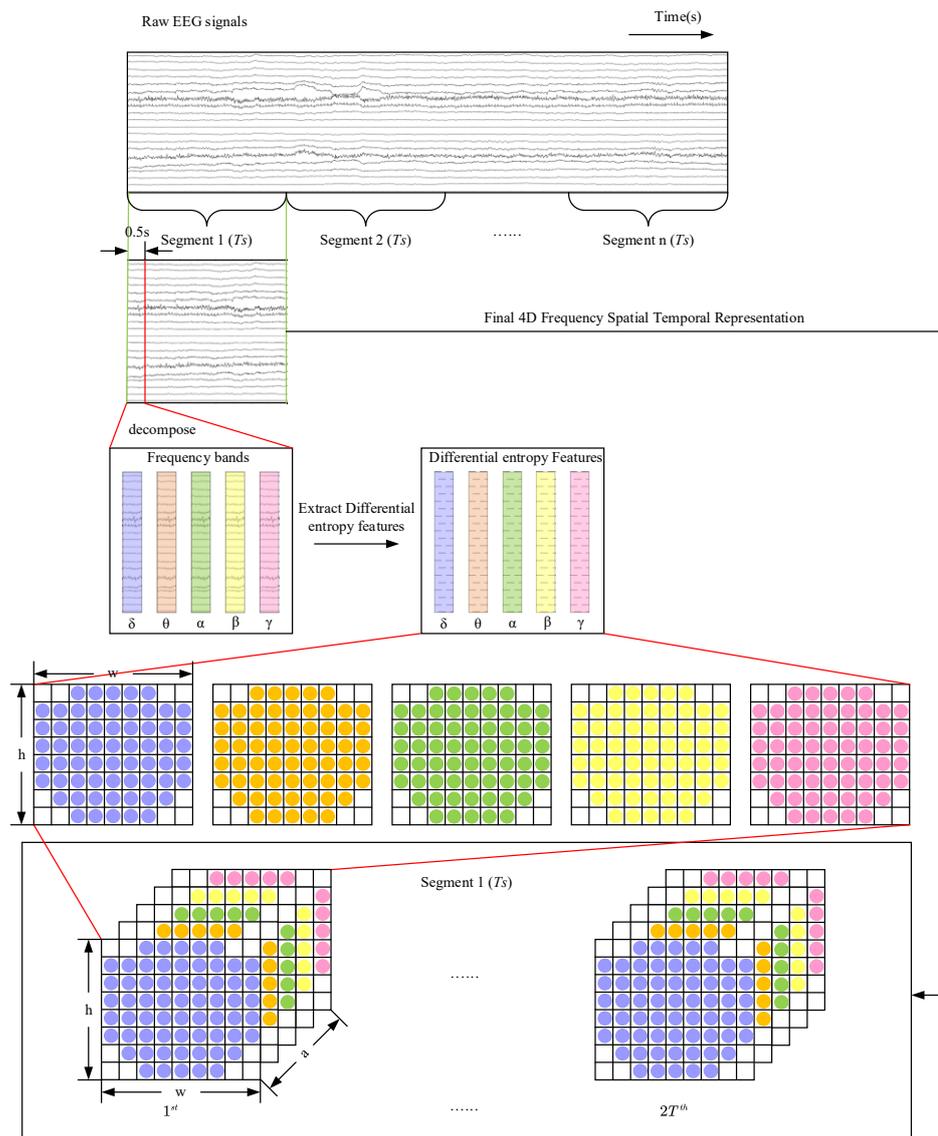


Figure 2. The generation of 4D inputs.

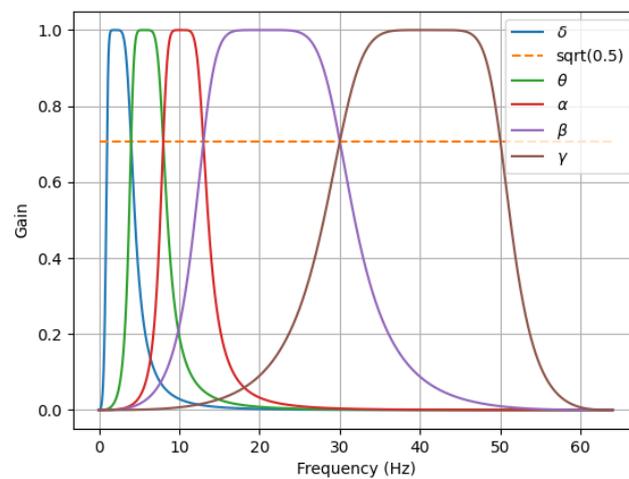
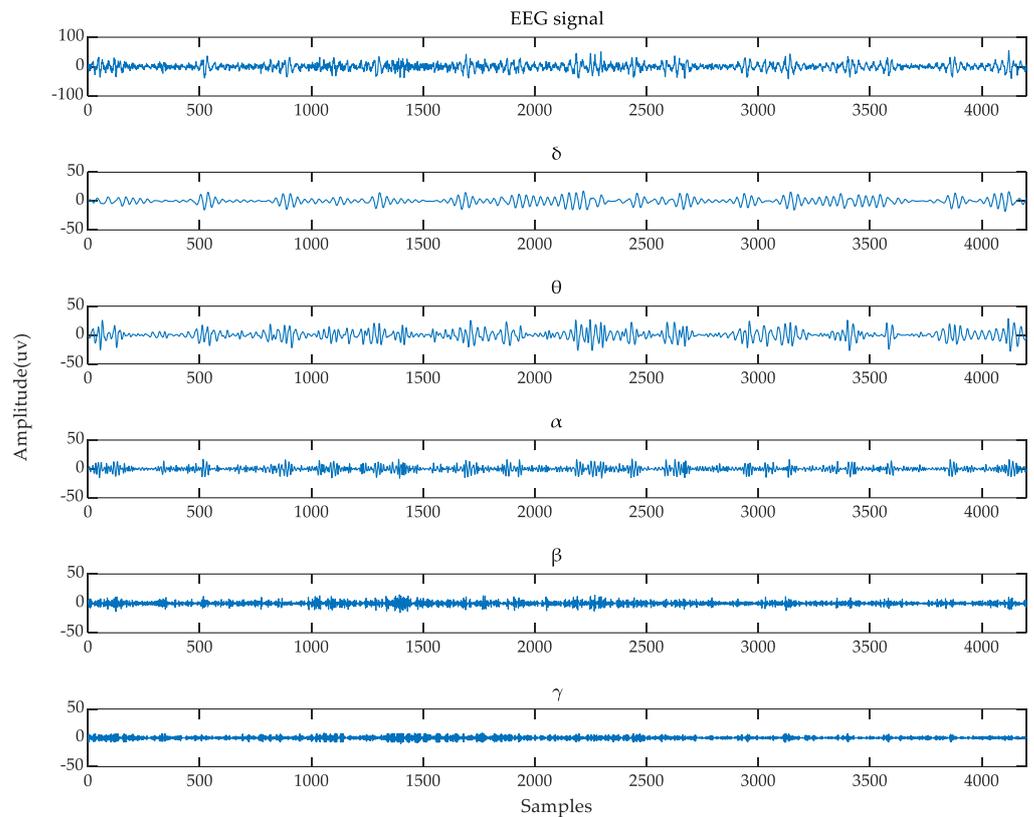


Figure 3. The frequency response of the filter.



**Figure 4.** The raw EEG signal and its five frequency bands in the DEAP dataset.

Therefore, in the fixed frequency band  $i$ , the DE feature of each EEG segment in each frequency band is respectively calculated as

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

where  $H_i$  and  $\sigma_i^2$  represent the differential entropy and signal variance of corresponding EEG signals in the frequency band  $i$ , respectively.

From Equation (3), where  $\pi, e$  are constants, we only need to know  $\sigma_i^2$  to get the differential entropy of  $X$ . For the frequency band of the EEG signal, as a result of the zero mean, the variance can be estimated as

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 \tag{4}$$

Equation (4) shows that the variance estimate of the signal sequence  $X$  is its average energy. Moreover, due to Parseval’s theorem, the average energy is related to the energy spectrum. It can be written as Equation (5) For the Discrete Fourier transform

$$\sum_{n=0}^{N-1} |x_n^2| = \frac{1}{N} \sum_{k=0}^{N-1} |X_k^2| = P_i \tag{5}$$

where  $X_k$  is the discrete Fourier transform coefficient of signal sequence  $X_i$ .  $N$  is the length of the fixed time windows.  $P_i$  can be thought of as the energy spectrum, equal to the value of the signal variance  $\sigma_i^2$  multiplied by a constant coefficient  $N$ .

Therefore, after the bandpass filter to a specific band  $i$ , its variance can be estimated by  $\frac{P_i}{N}$ . The relationship between the logarithmic energy spectrum and differential entropy can be expressed as

$$H_i(x) = \frac{1}{2} \log(2\pi e\sigma_i^2) = \frac{1}{2} \log(N\sigma_i^2) + \frac{1}{2} \log\left(\frac{2\pi e}{N}\right) = \frac{1}{2} \log(P_i) + \frac{1}{2} \log\left(\frac{2\pi e}{N}\right) \tag{6}$$

It can be seen that for EEG series of fixed length, the estimate of differential entropy is equivalent to the logarithmic energy spectrum of a specific frequency band.

EEG signals have characteristic frequency ranges and spatial distribution, which are often associated with different functional states of the brain. In the electroencephalogram, each electrode is physically adjacent to the other and records the electrical signals in one area of the brain. In order to maintain spatial information between multiple adjacent channels, according to the mapping method shown in Figure 5, the one-dimensional EEG signal is constructed into a two-dimensional matrix with a size of  $h \times w$ . The values of  $h$  and  $w$  are the maximum values of the electrodes in the vertical and horizontal directions, respectively. Among them, a value of zero indicates that the signal of the channel is not used. Then, the two-dimensional feature matrices of different frequency bands are superimposed at the same time to obtain a three-dimensional matrix  $h \times w \times a$ , where  $a$  indicates the total number of frequency bands. The 4D structure can be expressed as  $X_n \in R^{h \times w \times a \times 2T}$ ,  $n = 1, 2, \dots, S$ , where  $S$  is the total number of samples. In this work, the values of  $h$  and  $w$  were determined according to the two-dimensional matrix mapping of the distribution positions of EEG acquisition electrodes. Two methods were used to select a compact graph ( $8 \times 9$ ) and a sparse graph ( $19 \times 19$ ).

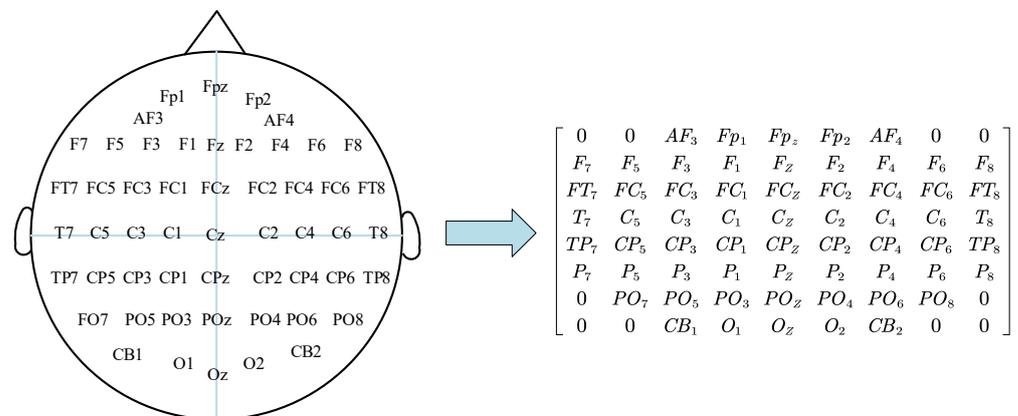


Figure 5. The 2D map of 62 channels.

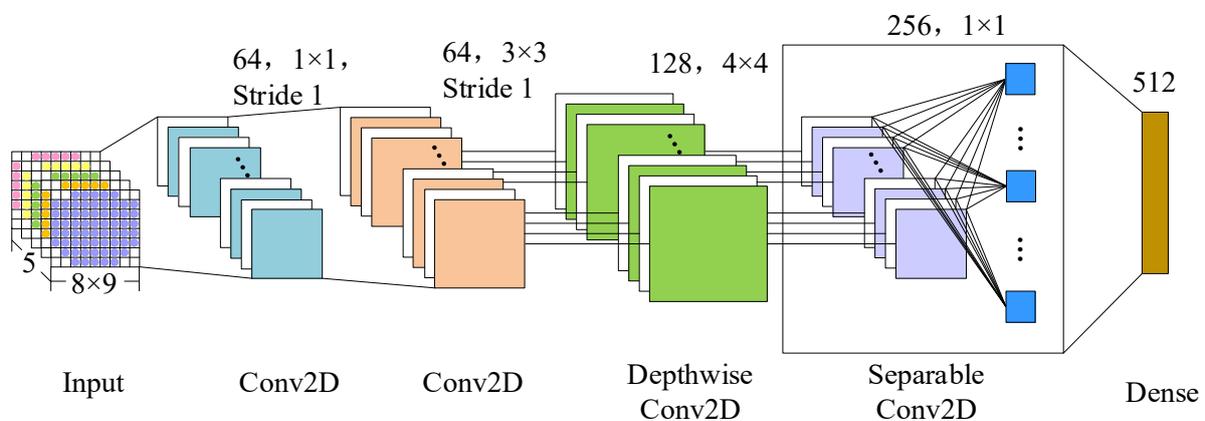
### 2.2. The Structure of FSTception

The structure of the FSTception model proposed in this paper consists of two parts, the first is that we use a deep separable convolutional network to further extract the frequency and spatial information of the 4D input feature structure. Secondly, the LSTM with ordered neuron structure is used to extract the contextual correlation of the time-series EEG signals, thus mining the deep emotional information of the EEG signals and improving the accuracy of emotion recognition.

#### 2.2.1. Frequency Spatial Characteristic Learning

Both spatial and temporal information is vital for time-series EEG signals, which represent activity relationships at different locations in the brain and the dynamics of brain patterns over time. In order to better represent the spatial features of EEG information in limited training samples, we propose the FSTception model, a CNN structure based on depthwise separable convolutions. For the 4D frequency space time structure expressed as  $X_n$ , we apply the FSTception model to obtain frequency and spatial information from

each time slice. The structure is shown in Figure 6. First, since the size of the 2D feature maps in the input samples is  $8 \times 9$ , the two-dimensional map size is relatively small. In order to better preserve all the information, we set up two convolutional layers, the first convolutional layer has 64 feature maps, and the filter size is  $1 \times 1$ , the second convolutional layer has 128 feature maps, the filter size is  $3 \times 3$ , and the deep features are extracted using convolution kernels of different scales. Then, we use deep convolution to learn spatial features. The main benefit of deep convolution is to reduce the total number of parameters in the network training process because the convolution is not fully connected to the previous feature map but are connected to each feature map separately to learn spatial features of specific frequencies. Importantly, when used in specific EEG applications, this operation provides a direct method to learn the spatial characteristics of each time slice so that the spatial characteristics of specific frequencies can be effectively extracted. Finally, we used separable convolution, which is a combination of deep convolution and pointwise convolution. The main advantages are: (1) reduce the number of fitting parameters; (2) summarize its feature maps separately by learning each core, and then merge the optimal results and output, thereby expressing the relationship between the internal and cross-feature maps of the feature map. When performing emotion recognition on EEG, this operation is used to summarize a single feature map (depthwise convolution) and optimize the combined feature map (pointwise convolution), because different feature maps can represent information at different time scales. For the convolutional layers of this network, zero-filling and rectified linear unit (ReLU) activation functions and normalization operations are used. To better reduce overfitting and improve the robustness of the network, we employ a max pooling layer with a size of  $2 \times 2$  and a step size of 2 after the operation of the convolution neural network. Finally, the output of the max-pooling layer is unrolled and sent to a fully connected layer with 512 units. The frequency and spatial features of the final output EEG are expressed as  $Q_n = R^{512 \times 2T}$ .



**Figure 6.** The structure of CNN module for frequency and spatial feature learning.

### 2.2.2. Temporal Characteristic Learning

Since the EEG signal contains dynamic content, the changes between time slices in the 4-dimensional structure may hide other information, which may help perform more accurate sentiment classification. Considering that the long short-term memory (LSTM) network has better modeling capabilities for time series information, this paper uses ON-LSTM to integrate neurons into the LSTM through specific sorting, thereby allowing ON-LSTM to learn the hierarchical structure information automatically [35]. The detailed structure of ON-LSTM is shown in Figure 7. ON-LSTM is actually a change in the overall design of the LSTM cell state update. The cell state is done in a hierarchical operation, and the traditional LSTM is to update all the cell states. The cell state is out of order, while ON-LSTM to cell state is updated by two integers for the hierarchy and then hierarchical for computation. This paper is the first innovative use of ON-LSTM in the temporal analysis of EEG emotional features. ON-LSTM is an improved network structure based on LSTM,

given a CNN output encoding sequence  $Q_n = (q_1, q_2, q_{2T})$ , where  $q_t = R^{512}$  and  $t = 1, 2, 2T$ . We apply an ON-LSTM layer with 128 data stores to discover the temporal dependence of internal structural segments. The gate structure and output structure of ON-LSTM are the same as those of general LSTM:

$$f_t = \sigma(W_f q_t + U_f h_{t-1} + b_f) \tag{7}$$

$$i_t = \sigma(W_i q_t + U_i h_{t-1} + b_i) \tag{8}$$

$$o_t = \sigma(W_o q_t + U_o h_{t-1} + b_o) \tag{9}$$

$$\hat{c}_t = \tanh(W_c q_t + U_c h_{t-1} + b_c) \tag{10}$$

$$h_t = o_t \circ \tanh(c_t) \tag{11}$$

where historical information  $h_{t-1}$  and current information are input, the input gate  $i_t$  determines the amount of change of the input vector  $q_t$  to the information in the storage unit at the current moment. Output gates  $o_t$  manipulate the output of information contained in the current storage unit. The forget gate  $f_t$  determines the degree of influence of the historical information  $h_{t-1}$  at the last moment on the information in the current storage.  $h_t$  is the final output of the ON-LSTM unit at a time  $t$ .  $\sigma(\cdot)$  is the logistic sigmoid function,  $\tanh(\cdot)$  is the hyperbolic tangent function,  $W$  and  $U$  are the weight matrix corresponding to each gate.  $b$  is the bias term corresponding to each gate.  $\circ$  is the corresponding element multiplication.

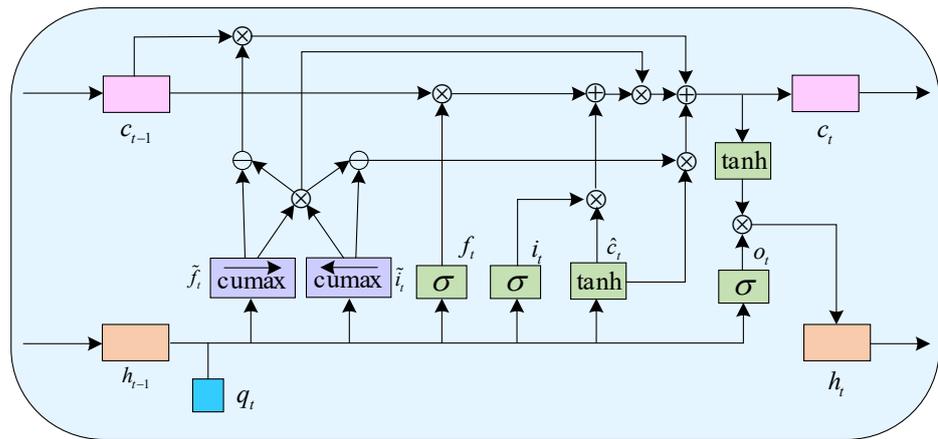


Figure 7. The detailed structure of ON-LSTM.

The improvement of ON-LSTM relative to LSTM is different in the update mechanism from  $\hat{c}_t$  to  $c_t$ , thus introducing a hierarchical structure of based on LSTM.

$$\tilde{f}_t = \vec{cs}(\text{softmax}(W_j q_t + U_j h_{t-1} + b_f)) \tag{12}$$

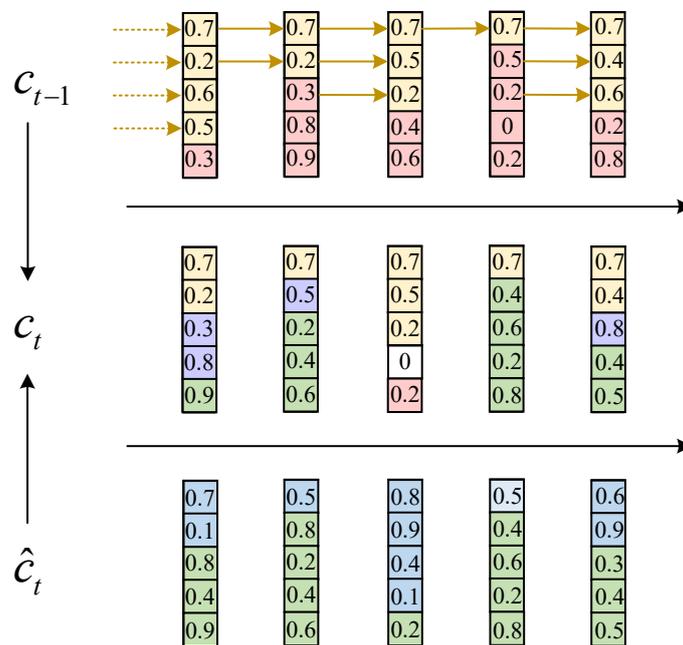
$$\tilde{i}_t = \overleftarrow{cs}(\text{softmax}(W_i q_t + U_i h_{t-1} + b_i)) \tag{13}$$

$$w_t = \tilde{f}_t \circ \tilde{i}_t \tag{14}$$

$$c_t = w_t \circ (f_t \circ c_{t-1} + \tilde{i}_t \circ \tilde{c}_t) + (\tilde{f}_t - w_t) \circ c_{t-1} + (\tilde{i}_t - w_t) \circ \tilde{c}_t \tag{15}$$

where  $\vec{cs}$  is cumsum from left to right, and  $\overleftarrow{cs}$  is cumsum from right to left. The main forget gate  $\tilde{f}_t$  determines the level of influence to the hierarchical data in the current data storage from the hierarchical information content included in the historical information  $h_{t-1}$  at the last moment. The main input gate  $\tilde{i}_t$  determines the amount of change of the level information contained in the input vector  $q_t$  at the current moment to the level information in the storage unit. After ON-LSTM sorts the neurons, the information level

is represented by the front and back of the position. Then, when updating the neuron, first predict the historical level  $h_{t-1}$  and the input level  $q_t$ , respectively, and update the neurons between partitions through these two levels. ON-LSTM partition update is shown in Figure 8. In this way, the high-level information may retain a considerable distance because the high-level will directly copy the historical information so that the historical information is continuously copied without being changed, and the low-level information may be updated at each step of input. Because the lower layer directly copies the input, and the input is constantly changing, a hierarchical structure is formed through information classification. More generally speaking, it is a group update. Higher group information is transmitted farther (larger span), and lower group span is more minor. These different spans form the hierarchical structure of the input sequence. By learning the hierarchical structure information, it can better mine the hidden connections between each time slice and then achieve more accurate sentiment classification.



**Figure 8.** ON-LSTM partition update. The numbers are randomly generated, with historical information at the top, current input at the bottom, and currently integrated output in the middle. The yellow part at the top is the historical information level (the main forgetting gate), the green part at the bottom is the input information level (the main input gate), the yellow part in the middle is the directly copied historical information, the green is the directly copied input information, the purple is the intersection information fused according to the LSTM, and the white is the unrelated “blank area”.

The final node output of ON-LSTM is used to represent the highest level of the EEG segment,  $Y_n \in R^{128}$ , which integrates the frequency, time, and space features of the EEG segment.

### 2.3. Classification

According to the final characteristics representation  $Y_n$ , we apply the fully connected layer and the softmax activation function to predict and analyze the sample label of the EEG segment  $X_n$ . The calculation is as follows

$$Out = AY_n + b = [out_1, out_2, \dots, out_c] \tag{16}$$

where  $A$  represents the transform matrix,  $b$  represents the bias, and  $m$  represents the number of emotion categories. Then, according to the output result, it is sent to the softmax classifier to obtain the final emotion discrimination result, which can be described as

$$P(c|X_n) = \frac{\exp(out_j)}{\sum_{i=1}^{i=c} \exp(out_i)} (j = 1, \dots, c) \quad (17)$$

where represents the normalized prediction results of each category. In back propagation, the Adam gradient algorithm is used.

### 3. Results

#### 3.1. Experimental Setup

The tasks in this paper are done on TensorFlow on a suitable GPU (Nvidia Quadro P5000) using Python 3.7 and Keras. For each trial, we completed a five-fold cross validation and measured the mean classification accuracy (ACC) and standard deviation (STD) for each experimenter. The final characteristics of everyone's mode are shown as the mean ACC and STD of all subjects. Applying the Adam boost loss function, we set the learning rate to 0.003, the batch size to 120, and the epoch to 100. Codes are available at <https://github.com/lixiaonian2022/code.git> (accessed on 14 July 2022).

All EEG trials are divided into five groups randomly, then four folds of trials are combined as training data, and the left 1-fold is considered as test data. The experiments are repeated five times so that every fold has a chance to be the test data. The result is calculated as the average of five experiments.

#### 3.2. Dataset

All subjects were in good physical and psychological health, and 32 conductor AgCl electrodes were used in the collection of data, with scalp electrodes positioned in accordance with the international standard 10–20 system.

##### 3.2.1. SEED Dataset

The SEED data set [36] has a total of 15 physical and mental health subjects (seven men, eight women, mean age 23.27). Each subject performed three sets of experiments, and each experiment watched 15 movie clips (five positive segments, five negative segments, five neutral segments). The viewing process of each segment can be divided into four stages, including a 5-s start prompt, 4-min movie playback time, 45-s self-assessment, and 15-s rest time. During the film screening, two films with the same emotion will not be displayed consecutively. Each volunteer carried out three experiments, each time separated by one week, for a total of 45 experiments.

##### 3.2.2. DEAP Dataset

The DEAP data set [37] contains 32 subjects, and each subject has 32 channels of EEG signals and 8 channels of peripheral physiological signals. The content of the DEAP dataset is shown in Table 3. The 32 channels EEG signal is used as the experimental data in this paper. The EEG data signal was first sampled at a sampling frequency of 512 Hz. Then, the sampling frequency was reduced to 128 Hz, and then filtered using a 4.0–45.0 Hz filter to remove EOG artifacts. Each subject watched 40 one-minute emotional music videos. After watching each short video, the subjects scored the level of arousal, valence, liking, and controlling influence. In this experiment, only two dimensions of arousal and potency were selected for testing.

#### 3.3. EEG Data Preprocessing

In order to eliminate the influence of artificial interference, such as ECG, EMG, and power frequency interference, on emotion recognition, we retained more practical com-

ponents of emotional EEG. The raw multichannel EEG data were first filtered using a Butterworth bandpass filter, using 1 Hz and 50 Hz limits to divide the frequency bands into five bands containing emotional information. These include the  $\delta$  band (1 to 4 Hz), the  $\theta$  band (4–8 Hz), the  $\alpha$  band (8–13 Hz), the  $\beta$  band (13–30 Hz), and the  $\gamma$  band (30–50 Hz). Each trial contains an EEG signal of 63 s. Among them, the first 3 s is the baseline signal, and the 60 s is the emotional signal of the EEG.

**Table 3.** DEAP dataset content.

Name	Size	Contents
Data	$40 \times 40 \times 8064$	videos $\times$ channels $\times$ data
Labels	$40 \times 4$	videos $\times$ labels (valence, arousal, dominance, liking)

The EEG of 63 s was divided into 126 segments with a length of 0.5 s on average, the first 3 s was the baseline signal, and the eigenvalue of each frequency band of each segment was calculated. In the experiment, in order to improve the accuracy of the experiment, the difference between the characteristic value of EEG recorded by the subject under the stimulus and the mean characteristic value of the baseline signal was used as the experimental data, and the difference at the moment was recorded as the average segmentation of the 63 s EEG into 126 segments with the length of 0.5 s, in which the first 3 s was the baseline signal. The eigenvalue of each frequency band of each segment is calculated. In the experiment, in order to improve the accuracy of the experiment, the difference between the characteristic value of EEG recorded by the subject under the stimulus and the mean characteristic value of the baseline signal was used as the experimental data, and the difference at time  $t$  was denoted as  $EEG\_removed_t$ . The calculation formula is as follows

$$EEG\_removed_t = feature_t - \frac{base\_feature_1 + base\_feature_2 + \dots + base\_feature_6}{6} \quad (18)$$

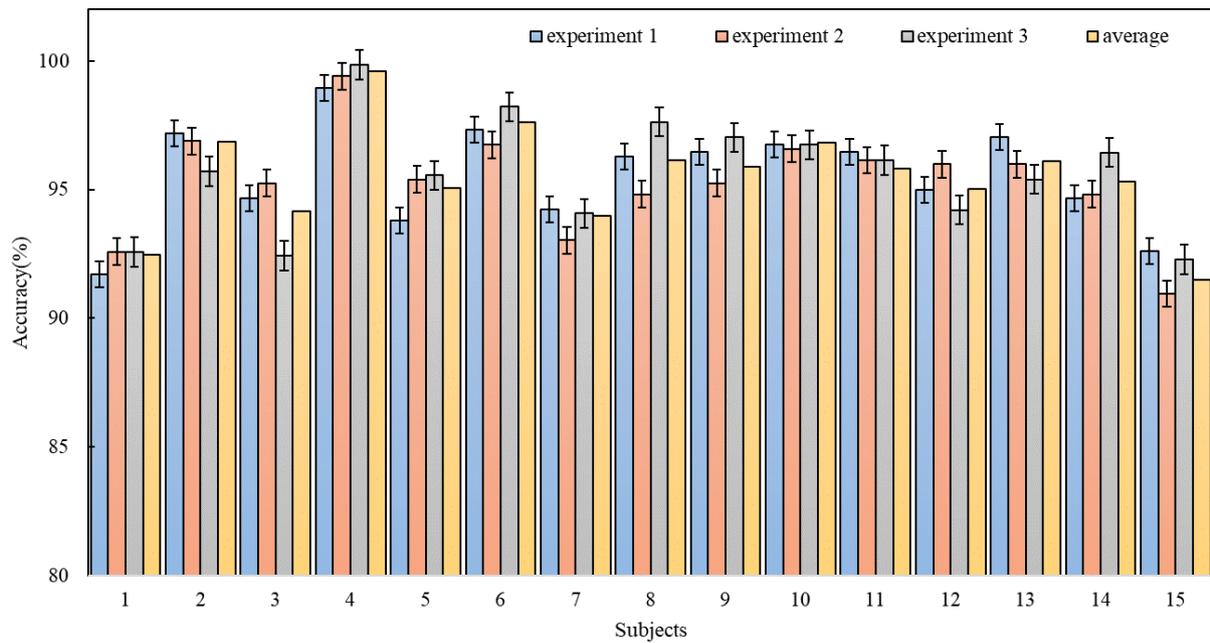
where  $feature_t$  is the EEG characteristic value extracted at time  $t$  and  $base\_feature_i$  ( $i = 1, 2, \dots, 6$ ) is the baseline signal characteristic value extracted at time  $i$ .

### 3.4. Results and Discussion

The research results of the 4D-FSTception method proposed in this article on the SEED data set is shown in Table 4, and as shown in bar Figure 9, numerical changes can be observed and compared more easily. The accuracy rate of all subjects exceeded 90%, and the average accuracy rate reached 95.49%. For different subjects, the classification accuracy rate is different. For subject 1, the average classification accuracy rate in the three time periods is 91.5% (the lowest), and for subject 4, the classification accuracy rate in the three time periods is 91.5%. The average classification accuracy rate is 99.59% (the highest), which reflects individual differences. However, for the same subject, the classification accuracy of different time periods fluctuates. For example, for subject 8, the classification accuracy rate in time period 2 is 94.81% (lowest), and the classification accuracy rate in time period 3 is 97.63% (highest). Among 15 subjects, eight items (#2, #4, #6, #8, #9, #10, #11, #13) are better than the mean classification accuracy.

**Table 4.** Recognition Accuracy for each subject on SEED dataset.

Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy
1	92.45%	5	95.06%	9	95.89%	13	96.12%
2	96.86%	6	97.63%	10	96.83%	14	95.32%
3	94.17%	7	93.99%	11	95.83%	15	91.50%
4	99.59%	8	96.15%	12	95.03%		



**Figure 9.** Overall performance of the 4D-FSTception model on SEED dataset.

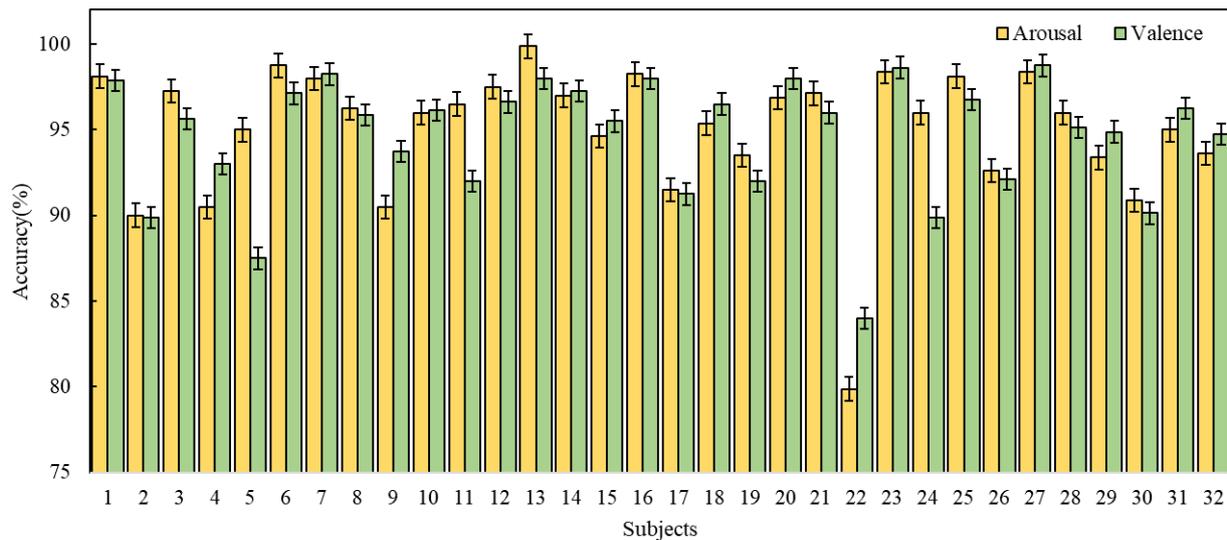
The experimental results of the 4D-FSTception method on the DEAP dataset are shown in Tables 5 and 6, respectively, and as shown in bar Figure 10. The average accuracy of arousal classification is 95.02%, and the standard deviation is 1.85%. Only two subjects (#2 and #22) had an accuracy rate less than 90%. In particular, the valence and arousal accuracy rates of tester #22 were 79.88% and 84%, respectively, which were lower than other testers. The reason for this may be that the subjects were inattentive during the test operation, or the level of subjective experience was not effectively reported after the test.

**Table 5.** Recognition Accuracy for each subject on “Arousal”.

Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy
1	98.13%	9	90.50%	17	91.50%	25	98.13%
2	90.00%	10	96.00%	18	95.38%	26	92.63%
3	97.25%	11	96.50%	19	93.50%	27	98.38%
4	90.50%	12	97.50%	20	96.88%	28	96.00%
5	95.00%	13	99.86%	21	97.13%	29	93.38%
6	98.75%	14	97.00%	22	79.88%	30	90.88%
7	98.00%	15	94.63%	23	98.38%	31	95.00%
8	96.25%	16	98.25%	24	96.00%	32	93.63%

**Table 6.** Recognition Accuracy for each subject on “Valence”.

Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy	Subjects	Accuracy
1	97.88%	9	93.75%	17	91.25%	25	96.75%
2	89.88%	10	96.13%	18	96.50%	26	92.13%
3	95.63%	11	92.00%	19	92.00%	27	98.75%
4	93.00%	12	96.63%	20	98.00%	28	95.13%
5	87.50%	13	98.00%	21	96.00%	29	94.88%
6	97.13%	14	97.25%	22	84.00%	30	90.13%
7	98.25%	15	95.50%	23	98.63%	31	96.25%
8	95.88%	16	98.00%	24	89.88%	32	94.75%



**Figure 10.** Overall performance of the 4D-FSTception model on DEAP dataset.

The accuracy, precision, recall, and F1-score metrics are used to evaluate the performance of the 4D-FSTception. The evaluation index is calculated based on all the predicted results in DEAP and SEED datasets, as shown in Tables 7 and 8. The confusion matrix of different data sets is shown in Figures 11 and 12. The superiority of the model in this paper in EEG signal emotion recognition was fully demonstrated.

**Table 7.** The score of 4D-FSTception evaluation index is on SEED dataset.

Classes	Accuracy	Precision	Recall	F1-Score
Negative	96.67%	94.87%	92.50%	93.67%
Neutral	93.42%	94.29%	93.94%	96.12%
Positive	96.38%	98.96%	95.00%	96.94%

**Table 8.** The score of 4D-FSTception evaluation index is on DEAP dataset.

Valence/Arousal	Class	Accuracy	Precision	Recall	F1-Score
Valence	High	94.90%	98.48%	93.94%	96.47%
	Low	93.42%	98.41%	93.18%	96.12%
Arousal	High	96.32%	98.02%	91.67%	94.74%
	Low	93.72%	93.30%	95.37%	96.71%

In order to verify the computational performance of the methods in this paper, we will choose different methods for time cost analysis in both SEED and DEAP datasets. To be able to view and compare maximum and minimum values in a more straightforward way. Figures 13 and 14, respectively, show the performance of accuracy and time cost by different methods in different datasets. As shown in Table 9, for the SEED dataset, a comparison between the HCNN method with a sparse input map of  $19 \times 19$  and the 4D-CRNN method with the same input map of  $8 \times 9$  as in this paper was used. After training, it can be seen that the accuracy of the compact graph is 95.49%, which is 6.89% higher than that of the sparse graph. Furthermore, in terms of their time cost, the compact map spends only 660 s per topic in the five-fold cross-validation, which is almost a quarter of the time cost of the sparse map. Moreover, it can be found that the accuracy of our compact map is close to that of the 4D-CRNN method, which is also a compact map. However, in terms of time cost, the method in this paper is 143 s faster than the 4D-CRNN method. On the DEAP

dataset, our compact map outperforms the sparse map and outperforms the 4D-CRNN method with the same compact map in terms of accuracy and training time cost. For valence classification, the average accuracy of our method is 95.49%, which is 0.42% higher than that of the 4D-CRNN method with the same input map. In addition, the time cost of this paper’s method is even 15 s faster than that of the 4D-CRNN method. For arousal classification, we can obtain a similar conclusion. The reason for obtaining similar results may be that adding a sparse map of zeros between adjacent electrodes does not add any useful information. The lower time cost of the compact map may be due to its smaller size, which involves fewer convolution filters to compute. We obtained the model parameter quantities as well as the floating-point operations to be 12.7 M and 20.61 G, respectively, in the DEAP dataset. Compared with other methods, the algorithm in this paper has better performance. The results for both datasets lead us to believe that the method in this paper exhibits better computational performance when classifying emotions.

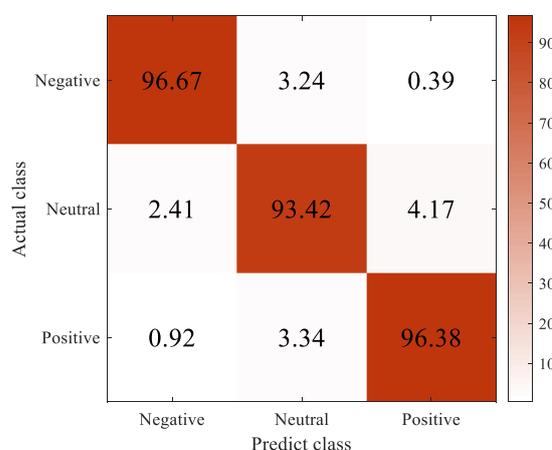


Figure 11. Confusion matrices of our method on SEED dataset.

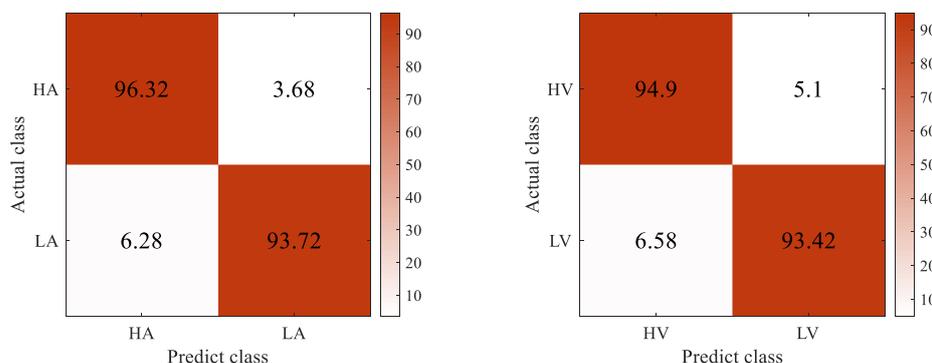


Figure 12. Confusion matrix: (left) arousal, (right) valence on DEAP dataset.

To verify the importance of the combined structure of our proposed CNN model and ON-LSTM, we performed two different combinations of CNN + LSTM (4D-FSTception (LSTM)) and CNN + ON-LSTM (4D-FSTception) in the SEED and DEAP datasets. The comparative experimental results are shown in Figure 15. Structure experiment. The average classification accuracy of 4D-FSTception in the SEED dataset and the DEAP dataset is 0.4% higher than the average classification accuracy of 4D-FSTception (LSTM). The results show that the ON-LSTM structure contributes to modeling time-series information in EEG signal emotion recognition. The hidden connection between each time slice can be better mined by learning the hierarchical structure information. In particular, the output results of the network proposed in this paper, after the EEG frequency and spatial information modeling are used as the input of the ON-LSTM structure, which can better capture the

continuity of emotion changes in time for the emotion recognition of the EEG signal with the 4D structure, is of great significance.

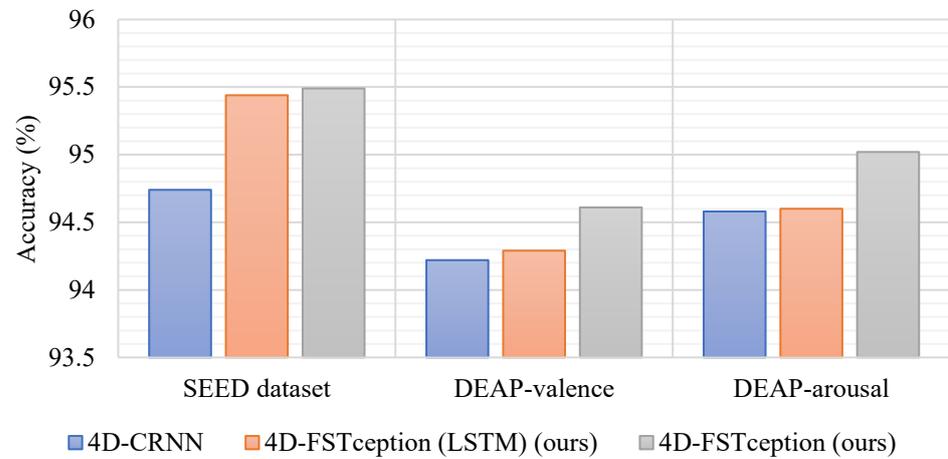


Figure 13. Accuracy performance of different methods in different datasets.

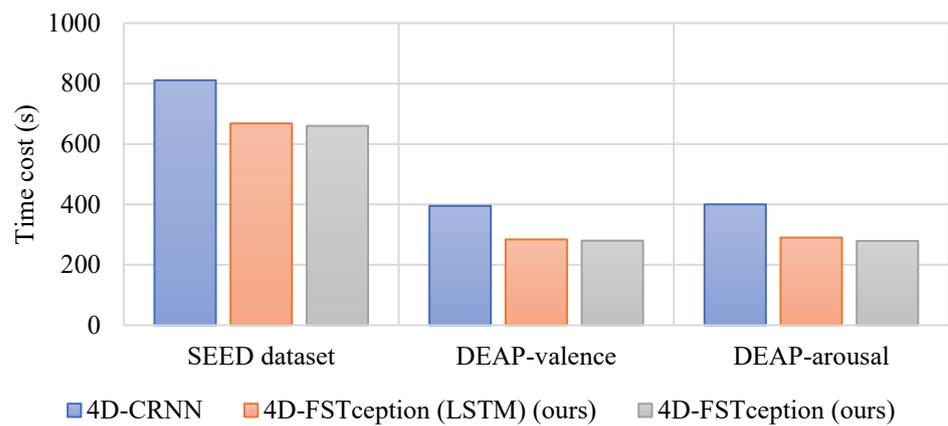


Figure 14. Time cost performance of different methods in different datasets.

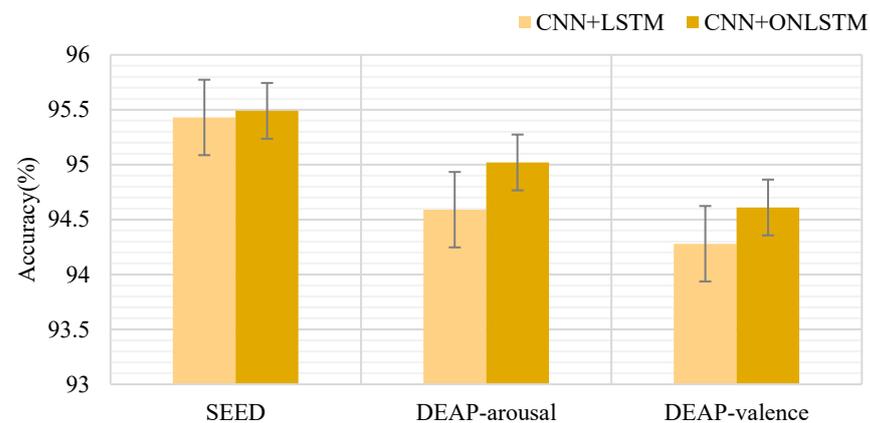


Figure 15. CNN + LSTM and CNN + ON-LSTM compare the experimental results.

### 3.5. Method Comparison

In order to show that the 4D-FSTception method is superior to other classification methods, this article chooses the following EEG emotion classification methods for comparison.

1. HCNN [38]: It uses a hierarchical CNN for EEG emotion classification and recognition and uses the differential entropy features of two-dimensional EEG as the input of the

neural network model, which proves that the band and band are more suitable for emotion recognition. The method considers the spatial information and frequency information of the EEG signal.

2. RGNN [39]: It uses the adjacency matrix in the graph neural network to simulate the inter-channel relationship in the EEG signal and realizes the simultaneous capture of the relationship between the local channel and the global channel. The connection and sparseness of the adjacency matrix are determined by humans. Supported by the neurological theory of brain tissue, this method shows that the relationship between global channels and local channels in the left and right hemispheres plays an important role in emotion recognition.
3. BDGLS [40]: It uses differential entropy features as input data. By combining the advantages of dynamic graph convolutional neural networks and generalized learning systems, emotion recognition accuracy can be improved over the full frequency band of EEG features. This method considers the frequency information and spatial information of the EEG signal at the same time.
4. PCRNN [32]: It first uses the CNN module to obtain space characteristics from each 2D EEG topographic map, then uses LSTM to obtain time characteristics from the EEG vector sequence, and finally integrates space and time characteristics to carry out emotional classification.
5. 4D-CRNN [41]: It first extracts features from EEG signals to construct a four-dimensional feature structure, then uses convolutional recurrent neural networks to extract EEG signals to obtain spatial features and frequency features, and uses LSTM to extract time from EEG vector sequences features, and finally carry out EEG emotion classification.

**Table 9.** Comparison of computational performance with other methods.

Method	Map Shape	SEED		DEAP-Valence		DEAP-Arousal			
		Acc (%)	Time Cost (s)	Acc (%)	Time Cost (s)	Acc (%)	Time Cost (s)	FLOPS (G)	Params (M)
HCNN [38]	19 × 19	88.60	3600	-	-	-	-	-	-
4D-CRNN [42]	8 × 9	94.74	811	94.22	395	94.58	400	18.63	15.8
4D-FSTception (LSTM) (ours)	8 × 9	95.44	668	94.29	284	94.60	290	19.94	14.4
4D-FSTception (ours)	8 × 9	95.49	660	94.61	280	95.02	279	20.61	12.7

6. 4D-aNN (DE) [42]: It uses 4D space-spectrum-time representations containing the space, frequency spectrum, and time information of the EEG signal as input. An attention mechanism is added to the CNN module and the bidirectional LSTM module. This method also considers the time, space, and frequency of EEG information.
7. ACRNN [43]: It adopts the convolutional recurrent neural network method based on the attention mechanism. It first uses the attention mechanism to distribute the weights between channels, then uses the CNN to extract the spatial information of the EEG, and finally uses the RNN to integrate and extract the temporal information features of the EEG. The method considers spatial information and temporal information for the emotion classification of EEG signals.

The above methods considered in the statistical validation experiments in this paper mainly include HCNN, RGNN, and BDGLS methods that used only frequency and spatial information for sentiment classification and recognition, as well as PCRNN and ACRNN methods that used spatial and temporal information features for sentiment classification. We also worked with 4D-CRNN and 4D-aNN methods that used frequency, spatial, and temporal features for sentiment recognition at the same time.

We reproduce these methods described above based on the structural parameters exploited in the original paper. We conduct five-fold cross-validation experiments for these approaches on the SEED and DEAP data. We list the mean accuracy and standard deviation

for each method on the SEED data set and DEAP data set in Table 10. For the SEED data set, the ACC of 4D-FSTception is 95.49%, which is 6.89% higher than the HCNN method that only considers the spatial information of the EEG signal, and it is also 1.25% higher than the RGNN. For the frequency information of the EEG signal simultaneously and the BDGLS method of spatial information, this method is 1.83% higher than this method. Both the 4D-CRNN and 4D-aNN (DE) methods consider the frequency, space, and time features at the same time. We choose the same input feature structure, and the differential entropy feature is selected in the frequency feature extraction. The 4D-FSTception is 0.75% higher than 4D-CRNN and 0.6% higher than 4D-aNN. For the DEAP data set, the experimental method in this paper achieves good results in EEG emotion recognition; its valence and arousal are 94.61% and 95.02%, respectively, which are 0.39% and 0.44% higher than 4D-CRNN, respectively. It is 4.35% and 4.04% higher than PCRNN, which only considers frequency and temporal information, and 0.89% and 1.64% higher than ACRNN, which only considers spatial and temporal information. For the 4D-CRNN method that simultaneously considers frequency, spatial, and temporal features, our method is 0.39% and 0.44% higher than 4D-CRNN in valence and arousal, respectively.

**Table 10.** The performances (average ACC  $\pm$  STD (%)) of the compared methods on SEED and DEAP datasets.

Method	Information	ACC $\pm$ STD (%)		
		SEED	DEAP-Valence	DEAP-Arousal
HCNN [38]	Frequency + spatial	88.60 $\pm$ 2.60	-	-
RGNN [39]	Frequency + spatial	94.24 $\pm$ 5.95	-	-
BDGLS [40]	Frequency + spatial	93.66 $\pm$ 6.11	-	-
PCRNN [32]	Spatial + temporal	-	90.26 $\pm$ 2.88	90.98 $\pm$ 3.09
ACRNN [43]	Spatial + temporal	93.72	3.21	93.38
4D-CRNN [41]	Frequency + spatial + temporal	94.74 $\pm$ 2.32	94.22 $\pm$ 2.61	94.58 $\pm$ 3.69
4D-aNN (DE) [42]	Frequency + spatial + temporal	95.39 $\pm$ 3.05	-	-
4D-FSTception (LSTM) (ours)	Frequency + spatial + temporal	95.44 $\pm$ 0.32	94.29 $\pm$ 1.89	94.60 $\pm$ 2.08
4D-FSTception (ours)	Frequency + spatial + temporal	95.49 $\pm$ 3.01	94.61 $\pm$ 2.83	95.02 $\pm$ 2.85

To sum up, the method proposed in this paper shows the best emotion recognition and classification performance in any combination of frequency, time, and space information or in the combination of the three characteristics.

#### 4. Conclusions

The paper points out a 4D-FSTception method for EEG emotion identification. The 4D input features include the frequency features of different frequency bands of the EEG signal, as well as the continuous-time features, while also maintaining the spatial information between channels. This method first extracts the differential entropy feature vectors of different frequency bands of the EEG signal at the same time. Secondly, according to the distribution position of the electrodes, the multichannel one-dimensional feature vector is converted into a two-dimensional feature matrix, which fully considers the spatial position information of the EEG channel. Then, the two-dimensional feature matrix is superimposed to form a three-dimensional matrix, and finally, the three-dimensional EEG sequence time segments divided into fixed lengths constitute 4D input features. Next, we introduce the FSTception model closely combined with CNN and ON-LSTM in detail. CNN is used to solve frequency and space data, and ON-LSTM is used to obtain time correlations from CNN output to reveal the underlying deep emotional characteristics. The proposed method achieves better performance on both SEED and DEAP datasets. By comparing with seven excellent studies, the 4D-FSTception model can effectively improve the accuracy of emotion discrimination in EEG signals. We hope to enhance our paper by building our own data set and verifying the validity of our approach in a real-world scenario. In future work, we

can consider combining DE features with more features to further enhance the emotion recognition performance of EEG signals through feature complementarity.

**Author Contributions:** Material preparation, data collection, and analysis were performed by Q.L. (Qi Li), Q.L. (Quanyang Liu), Y.M. and X.Z. The first draft of the manuscript was written by Q.L. (Qi Li) and all authors commented on previous versions of the manuscript. Y.L., Q.Z. and F.Y. made critical revisions to work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Science and Technology Department Project of Jilin Province, grant number YDZJ202101ZYTS064.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Datasets are available in a publicly accessible repository that does not issue DOIs. These data can be found at the following address: <http://www.eecs.qmul.ac.uk/mmv/datasets/deap/index.html> (accessed on 20 July 2022), and <https://bcmi.sjtu.edu.cn/home/seed/> (accessed on 16 March 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Bota, P.J.; Wang, C.; Fred, A.L.N.; Da Silva, H.P. A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals. *IEEE Access* **2019**, *7*, 140990–141020. [CrossRef]
2. Huang, Y.A.; Dupont, P.; Van de Vliet, L.; Jastorff, J.; Peeters, R.; Theys, T.; Van Loon, J.; Van Paesschen, V.; Van Den Stock, J.; Vandenbulcke, M. Network level characteristics in the emotion recognition network after unilateral temporal lobe surgery. *Eur. J. Neurosci.* **2020**, *52*, 3470–3484. [CrossRef]
3. Egger, M.; Ley, M.; Hanke, S. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [CrossRef]
4. Luo, J.; Tian, Y.; Yu, H.; Chen, Y.; Wu, M. Semi-Supervised Cross-Subject Emotion Recognition Based on Stacked Denoising Autoencoder Architecture Using a Fusion of Multi-Modal Physiological Signals. *Entropy* **2022**, *24*, 577. [CrossRef]
5. Yao, L.; Wang, M.; Lu, Y.; Li, H.; Zhang, X. EEG-Based Emotion Recognition by Exploiting Fused Network Entropy Measures of Complex Networks across Subjects. *Entropy* **2021**, *23*, 984. [CrossRef]
6. Keshmiri, S.; Shiomi, M.; Ishiguro, H. Entropy of the Multi-Channel EEG Recordings Identifies the Distributed Signatures of Negative, Neutral and Positive Affect in Whole-Brain Variability. *Entropy* **2019**, *21*, 1228. [CrossRef]
7. Pan, L.; Wang, S.; Yin, Z.; Song, A. Recognition of Human Inner Emotion Based on Two-Stage FCA-Relieff Feature Optimization. *Inf. Technol. Control* **2022**, *51*, 32–47. [CrossRef]
8. Gao, Z.; Cui, X.; Wan, W.; Gu, Z. Recognition of Emotional States Using Multiscale Information Analysis of High Frequency EEG Oscillations. *Entropy* **2019**, *21*, 609. [CrossRef]
9. Chao, H.; Dong, L.; Liu, Y.; Lu, B. Emotion recognition from multiband EEG signals using CapsNet. *Sensors* **2019**, *19*, 2212. [CrossRef]
10. Catrambone, V.; Greco, A.; Scilingo, E.P.; Valenza, G. Functional Linear and Nonlinear Brain–Heart Interplay during Emotional Video Elicitation: A Maximum Information Coefficient Study. *Entropy* **2019**, *21*, 892. [CrossRef]
11. Krishnan, P.T.; Joseph Raj, A.N.; Rajangam, V. Emotion classification from speech signal based on empirical mode decomposition and non-linear features. *Complex Intell. Syst.* **2021**, *7*, 1919–1934. [CrossRef]
12. Danelljan, M.; Robinson, A.; Shahbaz Khan, F.; Felsberg, M. Beyond correlation filters: Learning continuous convolution operators for visual tracking. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 472–488. [CrossRef]
13. Zhang, J.; Min, Y. Four-Classes Human Emotion Recognition Via Entropy Characteristic and Random Forest. *Inf. Technol. Control* **2020**, *49*, 285–298. [CrossRef]
14. Cao, R.; Shi, H.; Wang, X.; Huo, S.; Hao, Y.; Wang, B.; Guo, H.; Xiang, J. Hemispheric Asymmetry of Functional Brain Networks under Different Emotions Using EEG Data. *Entropy* **2020**, *22*, 939. [CrossRef]
15. Liu, Y.; Sourina, O. Real-time fractal-based valence level recognition from EEG. In *Transactions on Computational Science XVIII*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 101–120. [CrossRef]
16. Lin, Y.P.; Wang, C.H.; Jung, T.P.; Wu, T.L.; Jeng, S.K.; Duann, J.R.; Chen, J.H. EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 1798–1806. [CrossRef]
17. Altan, G.; Yayık, A.; Kutlu, Y. Deep learning with ConvNet predicts imagery tasks through EEG. *Neural Process. Lett.* **2021**, *53*, 2917–2932. [CrossRef]
18. Subha, D.P.; Joseph, P.K.; Acharya, U.R.; Lim, C.M. EEG signal analysis: A survey. *J. Med. Syst.* **2010**, *34*, 195–212. [CrossRef]
19. Zou, X.; Feng, L.; Sun, H. Compressive Sensing of Multichannel EEG Signals Based on Graph Fourier Transform and Cosparsity. *Neural Process. Lett.* **2020**, *51*, 1227–1236. [CrossRef]

20. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [[CrossRef](#)]
21. Khalil, R.A.; Jones, E.; Babar, M.I.; Jan, T.; Zafar, M.H.; Alhussain, T. Speech emotion recognition using deep learning techniques: A review. *IEEE Access* **2019**, *7*, 117327–117345. [[CrossRef](#)]
22. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **2018**, *13*, 55–75. [[CrossRef](#)]
23. Komolovaitė, D.; Maskeliūnas, R.; Damaševičius, R. Deep Convolutional Neural Network-Based Visual Stimuli Classification Using Electroencephalography Signals of Healthy and Alzheimer’s Disease Subjects. *Life* **2022**, *12*, 374. [[CrossRef](#)]
24. Thammasan, N.; Fukui, K.I.; Numao, M. Application of deep belief networks in eeg-based dynamic music-emotion recognition. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 881–888. [[CrossRef](#)]
25. Tripathi, S.; Acharya, S.; Sharma, R.D.; Mittal, S.; Bhattacharya, S. Using deep and convolutional neural networks for accurate emotion classification on deap dataset. In Proceedings of the Twenty-ninth IAAI Conference, San Francisco, CA, USA, 6–9 February 2017. [[CrossRef](#)]
26. Salama, E.S.; El-Khoribi, R.A.; Shoman, M.E.; Shalaby, M.A.W. EEG-based emotion recognition using 3D convolutional neural networks. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 329–337. [[CrossRef](#)]
27. Yang, Y.; Wu, Q.; Fu, Y.; Chen, X. Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In *International Conference on Neural Information Processing*; Springer: Cham, Switzerland, 2018; pp. 433–443. [[CrossRef](#)]
28. Riezler, S.; Hagmann, M. *Validity, Reliability, and Significance*; Springer: Cham, Switzerland, 2022. [[CrossRef](#)]
29. Zheng, W.L.; Lu, B.L. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [[CrossRef](#)]
30. Meng-meng, Y.A.N.; Zhao, L.V.; Wen-hui, S.U.N. Extraction of spatial features of emotional EEG signals based on common spatial pattern. *J. Graph.* **2020**, *41*, 424.
31. Liu, C.Y.; Li, W.Q.; Bi, X.J. Research on EEG emotion recognition based on RCNN-LSTM. *Acta Autom. Sin.* **2020**, *45*, 1–9. [[CrossRef](#)]
32. Yang, Y.; Wu, Q.; Qiu, M.; Wang, Y.; Chen, X. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.
33. Duan, R.N.; Zhu, J.Y.; Lu, B.L. Differential entropy feature for EEG-based emotion classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; pp. 81–84. [[CrossRef](#)]
34. Shi, L.C.; Jiao, Y.Y.; Lu, B.L. Differential entropy feature for EEG-based vigilance estimation. In Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, 3–7 July 2013; pp. 6627–6630.
35. Shen, Y.; Tan, S.; Sordani, A.; Courville, A. Ordered neurons: Integrating tree structures into recurrent neural networks. *arXiv* **2018**, arXiv:1810.09536.
36. Zheng, W.L.; Liu, W.; Lu, Y.; Lu, B.L.; Cichocki, A. Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Trans. Cybern.* **2018**, *49*, 1110–1122. [[CrossRef](#)]
37. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [[CrossRef](#)]
38. Li, J.; Zhang, Z.; He, H. Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cogn. Comput.* **2018**, *10*, 368–380. [[CrossRef](#)]
39. Zhong, P.; Wang, D.; Miao, C. EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1290–1301. [[CrossRef](#)]
40. Wang, X.H.; Zhang, T.; Xu, X.M.; Chen, L.; Xing, X.F.; Chen, C.P. EEG emotion recognition using dynamical graph convolutional neural networks and broad learning system. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018; pp. 1240–1244. [[CrossRef](#)]
41. Shen, F.; Dai, G.; Lin, G.; Zhang, J.; Kong, W.; Zeng, H. EEG-based emotion recognition using 4D convolutional recurrent neural network. *Cogn. Neurodynamics* **2020**, *14*, 815–828. [[CrossRef](#)]
42. Xiao, G.; Shi, M.; Ye, M.; Xu, B.; Chen, Z.; Ren, Q. 4D attention-based neural network for EEG emotion recognition. *Cogn. Neurodynamics* **2022**, *16*, 805–818. [[CrossRef](#)]
43. Tao, W.; Li, C.; Song, R.; Cheng, J.; Liu, Y.; Wan, F.; Chen, X. EEG-based emotion recognition via channel-wise attention and self attention. *IEEE Trans. Affect. Comput.* **2020**. [[CrossRef](#)]