EAI Endorsed Transactions
on e-Learning

Research Article

EEG Emotion Recognition based on Multi-scale Convolutional Self-Attention Networks

Hao Chao[1] and Yuan Fang[1,*]

[1]School of Computer Science and Technology, Henan Polytechnic University, Henan 454000, China

Abstract

A multi-view self-attention module is proposed and paired with a multi-scale convolutional model to build a multi-view self-attention convolutional network for multi-channel EEG emotion recognition. First, time and frequency domain characteristics are extracted from multi-channel EEG signals, and a three-dimensional feature matrix is built using spatial mapping connections. Then, a multi-scale convolutional network extracts the high-level abstract features from the feature matrix, and a multi-view self-attention network strengthens the features. Finally, use the multilayer perceptron for sentiment classification. The experimental results reveal that the multi-view self-attention convolutional network can effectively integrate the time domain, frequency domain, and spatial domain elements of EEG signals using the DEAP public emotion dataset. The multi-view self-attention module can eliminate superfluous data, apply attention weight to the network to hasten network convergence, and enhance model recognition precision.

Received on 15 August 2023; accepted on 06 September 2023; published on 06 September 2023

Keywords: Multi-Channel EEG Signal, Emotional Recognition, Multi-Scale Convolutional Network, Self-Attention Network

doi:10.4108/eetel.3722

1. Introduction

Emotion is a complicated mental state that manifests itself in bodily behavior and physiological functions. Emotions can be recognized through facial expressions, speech, behavior (gestures and postures), and physiological markers. Recognition based on non-physiological signs can be inefficient since people may intentionally or unconsciously conceal their genuine emotions. Physiological signals can provide more accurate and objective emotion recognition. Given that the central nervous system may govern and control the autonomic nervous system to participate in the emotional process, it is of tremendous scientific value to directly employ the electroencephalogram (EEG) to study the mechanism of emotion formation and detect emotional states [1, 2]. Emotional recognition is a pattern recognition technology application, and feature acquisition is a critical stage. Time domain feature analysis is primarily used to describe the waveform characteristics of EEG signals, such as mean value, first difference value, second difference value, variance, standard deviation, and so on [3], as well as Zero-crossing rate, slope transformation times, and Willison amplitude [4, 5]. Frequency-domain analysis, as opposed to time-domain analysis, can show the properties of each frequency component in a signal. Power spectral density, approximation entropy, and differential entropy are the three main features of the frequency domain. A popular technique for frequency-domain analysis is the Fourier transform. The time-frequency domain aspect of EEG signals refers to their temporal nature and simultaneous interpretation in the time and frequency domains. Time-frequency domain characteristics increase knowledge of the signal and broaden the scope of EEG signal analysis [6, 7]. The commonly used time-frequency analysis tools are wavelet transform, short-time Fourier transform and empirical mode decomposition.

Emotional recognition is frequently modeled using machine learning approaches. Camdra et al.[8] used the SVM model to classify emotions after extracting

*Corresponding author. Email: 6870961230@163.com

wavelet characteristics from small time segments. Mert and Akan et al.[9] proved that Hjordh parameters and correlation coefficients can be used as features to improve classification performance. Because deep learning models can automatically extract features, an increasing number of academics are preferring to use deep learning methods to improve classification quality[10]. Song et al.[11] introduced a dynamic graph convolutional neural network-based multi-channel EEG emotion identification technique. Wen et al.[12] suggested an end-to-end model based on a Convolutional Neural Network (CNN) that reconstructs the original EEG channels using the Pearson correlation coefficient and then sends the rearranged EEG signals to CNN for categorization. Emotion, as a high-level cognitive function of the human brain, is dependent on delicate coordination between local and even all brain regions. Zhong et al.[13] suggested a Regular Graph Neural Network (RGNN) for emotion identification in EEG. The RGNN considers the biological architecture of different brain areas to capture local and global interactions between multiple EEG channels.

Although the introduction of deep learning models in emotion recognition tasks can achieve good results, there are still some challenges. First, due to the volume conduction effect, the responses of the two electrodes adjacent to each other tend to be similar. Retaining the spatial position information of the EEG electrodes during feature extraction can provide a gain for emotion recognition, which is often ignored in previous studies. The number of sampling electrodes limits the resolution of EEG spatial feature maps, and low-resolution feature maps will affect the recognition results of the model. Secondly, the traditional CNN network model introduces a large amount of redundant data while extracting high-level abstract features. Because receptive fields of various sizes have limitations in learning spatial domain information, it is necessary to introduce a self-attention network to focus on the information that is more critical to the current task from a global perspective, which can effectively improve the efficiency and accuracy of task processing.

This study suggests a multi-view self-Attention module to address the aforementioned issues and then integrates it with the multi-scale convolution model to create the multi-view self-attention convolutional Network (McM). EEG signals are first processed to extract their time domain and frequency domain features, after which a three-dimensional feature matrix is built using the relative position data of the collecting electrode. Hierarchical multi-scale information is extracted using a multi-scale convolutional neural network (MSCNN) under various receptive fields. The multi-view self-attention module (MvSA) then uses attention weight to strengthen spatial domain features by dividing low-resolution spatial feature maps into patches of the same size to determine whether there may be a relationship between the activity states of various brain regions and distinct emotional states. Multilayer perceptrons are used to classify emotions afterwards.

2. Related works

2.1. Model and Dataset

Discrete and continuous models are the two types of emotion classification models that are most common. According to the discrete emotion model, basic emotions make up complex emotions, which themselves are made up of other basic emotions. Ekman et al.[14] divided feelings into four categories: fear, rage, sadness, and liking. Emotions were categorized into eight primary states in the study by Plutchik et al.[15]: love, rage, sadness, happiness, expectancy, hatred, fear, and surprise. The continuous emotion model commonly has a two-dimensional emotion space composed of the "valence-arousal" dimension. Theoretically, the continuous emotion model can classify emotions more clearly and intuitively. On the arousal dimension, a score below 5 is defined as low arousal (LA), and a score above or equal to 5 is defined as high arousal (HA). Low valence (LV) and high valence (HV) are also divided using the same threshold.

The investigation used the DEAP dataset[16], which had 1280 recordings of the subjects' EEG and peripheral physiological signals. Based on the international 10/20 system, the EEG data was recorded from 32 active electrode channels at a sample rate of 512 Hz. Each participant rated their own level of arousal, valence, liking, and dominance using a self-assessment approach. In each experiment, participants selected a number between 1 and 9 to indicate their emotional condition.

2.2. Feature Extraction

The EEG signal time for each of the 1280 samples in the DEAP dataset is 60 seconds. Twenty segments were cut from the 60-second EEG signal so that 25,600 samples could be collected. For each sample, 32 electrode channels' time-domain and frequency-domain properties, including mean value, first difference value, second difference value, and approximative entropy, were retrieved.

A nonlinear dynamic quantity called approximate entropy (ApEn) is used to measure the predictability and regularity of time series fluctuations. In order to express the potential for the occurrence of new information in the time series, it employs a non-negative number to describe the complexity of a time series. The bigger the ApEn[17], the more complicated the time series. The reconstruction vector dimension and r are the parameters that establish the threshold of

similarity for the EEG signal $s(t)$. $m = 2$ and $r = 0.2$ were chosen for this study. Reset m-dimensional vectors $X(1)$, $X(2)$,..., $X(T-m+1)$, where $X(i) = [s(i), s(i+1), ...s(T+m-1)]$, count the number of vectors meeting the following conditions:

$$C_r^m(r) = \frac{\text{num}\left(d_m\left[X(i), X(j)\right] < r\right)}{T-m+1} \tag{1}$$

where $\mathrm{d}\left[X,X^*\right] = \max_a |\mathbf{s}(a) - \mathbf{s}^*(a)|$,$s(a)$ is an element of vector X, d represents the maximum difference in distance between the vector $X(i)$ and $X(j)$, The range of j is $[1, T-m+1]$, including $j = i$. The formula of the similar average rate is:

$$_m(r) = \frac{\sum_{i=1}^{T-m+1} \log\left(C_i^m(r)\right)}{T-m+1} \tag{2}$$

The APEn of EEG signals can be expressed as:

$$\text{ApEn} = \Phi_m(r) - \Phi_{m+1}(r) \tag{3}$$

2.3. 3D Feature Matrix

The mapping of the international 10/20 system to a 3D feature matrix is shown in Figure 1. The distribution of the 32 electrode channels utilized in the DEAP dataset is shown in Figure 1(a), which represents the international 10/20 system. A mapping matrix $\mathbf{M}^{9\times9}$ is shown in Figure 1(b), which was built based on the relative placements of electrodes on the brain to ensure as much completeness of spatial information as possible. The time domain and frequency domain information derived from separate electrode channels is inserted in the correct positions in the matrix, and unused channels are filled with 0. Each sample will extract four features, with each feature referenced to create a $\mathbf{R}^{1\times9\times9}$ feature map. In accordance with Figure 1(c), a 3D feature matrix $\mathbf{R}^{4\times9\times9}$ is created by combining four characteristics. Finally, use the 3D feature matrix as input to the model.

3. McM Emotional Recognition Model

3.1. Model Frame

The multi-scale convolutional model (MSCNN) and the multi-view self-attention module (MvSA) based on self-attention are the two separate components of our McM network model. A McM network is created by joining two modules in series, as shown in Figure 2. Feature abstraction is achieved by using the ascending feature of 1×1 convolution to augment the information in the feature dimension. Different convolutional kernel sizes in MSCNN can locally capture

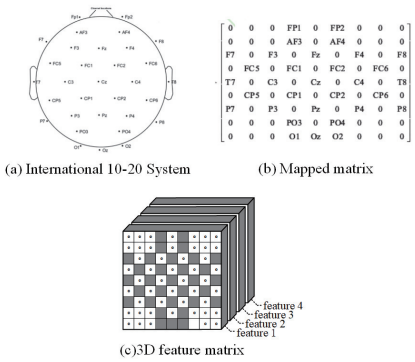

Figure 1. International 10-20 System and 3D feature matrix

the spatial correlation between similar electrode channels, which significantly facilitates network information sharing. Using self-attention processes to parameterize the weights between nodes, supervised augmentation of useful information, and suppression of unimportant information based on emotional labeling, MvSA can conduct electrode-by-electrode enhanced representation of multi-channel EEG signals. In the McM model, all convolutions use the ReLU activation function and residual connections to prevent network degradation. The dropout layer is used to randomly forget some network characteristics in order to prevent overfitting. The softmax function normalizes the final output emotion state.

3.2. Multi-view Self-attention Module

The MvSA calculates attention at the electrode channel level using a self-attention method. First, the 3D feature matrix $\mathbf{R}^{C\times H\times W}$ is partitioned into patches of size $\mathbf{r}^{C\times p_h\times p_w}$. The size of the patches in the spatial dimension is $P = p_h \times p_w$, and the number is $N = \left(\frac{H}{p_h}\right)\times\left(\frac{W}{p_w}\right)$. To compute these patches in parallel, the 3D feature matrix needs to be deformed into $\mathbf{R}^{'C\times N\times P}$. Then, the attention operation is performed from different views of the feature matrix $\mathbf{R}^{'}$, They are respectively Inner-Patch Self-Attention (IPSA), Among-Patch Self-Attention (APSA), and Blend-Patch Self-Attention (BPSA). Let the input of three kinds of self-attention operations are $I_1, I_2, I_3 \in I$, output $O_1, O_2, O_3 \in O$, weight matrix are W_q, W_k and W_v, then learning process of self-attention weight is as follows.

$$\mathbf{Q} = \mathbf{W}^q \cdot \mathbf{I} \tag{4}$$

$$\mathbf{K} = \mathbf{W}^k \cdot \mathbf{I} \tag{5}$$

$$\mathbf{V} = \mathbf{W}^v \cdot \mathbf{I} \tag{6}$$

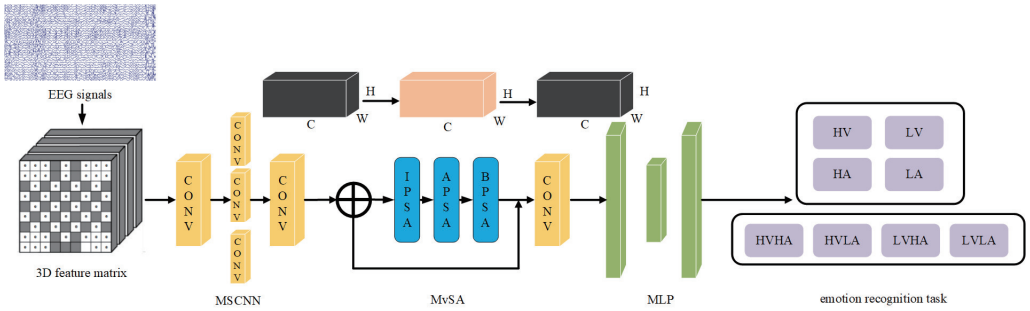

Figure 2. McM network structure diagram

$$\mathbf{O} = \text{softmax}\left(\frac{\mathbf{Q}\cdot\mathbf{K}^T}{\sqrt{\text{dim}}}\right)\cdot\mathbf{V} \tag{7}$$

Adding $\sqrt{\text{dim}}$ can prevent data extreme. This dim denotes the size of the final dimension of input I, The LayerNorm and BatchNorm layers need to be normalized after each attention computation, and then the MLP is used to combine the residual. The calculation process of MvSA is shown in Figure 3, with input as y_{n-1}, and output as y_n.

By decomposing $\mathbf{R}^{'}$ into different dimensions, different inputs can be obtained to achieve attention operations from different perspectives. Firstly, decompose $\mathbf{R}^{'}$ from the $z-axis$ direction to obtain the input $I_1 \in \mathbf{R}^{'C\times 1\times P}$, which can calculate attention weights in parallel within each patch, as shown in Figure 4 (a). Due to the limitation of patch size, the electrode channel weight calculated by IPSA is local, and information exchange between patches is needed to calculate the global attention weight. The input $I_2 \in \mathbf{R}^{'1\times N\times P}$ of APSA has complete global spatial domain information in a single feature dimension, and its self-attention calculation effect is similar to the multiplication of 2D matrices, as shown in Figure 4 (b). Each patch is calculated with other patches including it, and attention weights are learned from a global perspective.

A single feature dimension's APSA calculation results will have an impact on the prior IPSA results. Because of this, BPSA must modify attention weights in the global feature dimension while preserving communication across patches in the spatial domain. The electrode channels in various patches at the same location are combined to create a new patch block (blend-patch) by decomposing $\mathbf{R}^{'}$ from the $x-axis$, as shown in Figure 4(c). Input $I_3 \in \mathbf{R}^{'C\times N\times 1}$ of BPSA, I_3 not only keeps all feature dimensions, but also contains partial spatial information for all patches, allowing it to properly adapt the acquired attention weight from a global perspective.

The three components of MvSA are closely related. The first local attention calculation is carried out by IPSA, the results of the spatial attention calculation are strengthened by APSA, and the global attention weight adjustment is carried out by BPSA. MvSA can more effectively and efficiently distribute attention weight, filter electrode channels that are globally helpful for emotion recognition, and speed up the network's convergence rate.

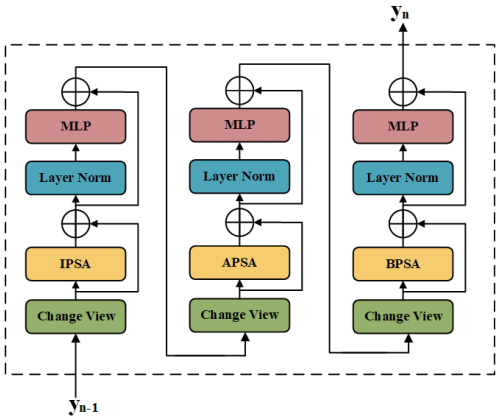

Figure 3. MvSA structure diagram

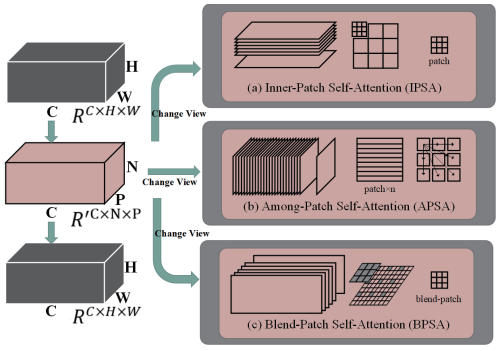

Figure 4. MvSA attention process

Table 1. The number of samples of different emotions

Emotion recognition task	Label	Count
Task of two categories	LV	11440
	HV	14160
	LA	10860
	HA	14740
Task of four categories	LVLA	5480
	HVLA	5380
	LVHA	5960
	HVHA	5780

Table 2. McM emotion recognition results

Dimension	McM-1	McM-2	McM-3
Valence	acc:0.8523	acc:0.8500	acc:0.8441
	f1:0.8655	f1:0.8644	f1:0.8607
	pre:0.8630	pre:0.8632	pre:0.8616
	recall: 0.8680	recall:0.8656	recall:0.8598
Arousal	acc:0.8523	acc: 0.8484	acc:0.8457
	f1:0.8727	f1:0.8690	f1:0.8672
	pre:0.8827	pre:0.8643	pre:0.8617
	recall:0.8630	recall:0.8737	recall:0.8728

4. Experimental Results and Analysis

The DEAP dataset contains 1280 EEG signals. To achieve an adequate number of experimental samples, the EEG signals were windowed with a 3-second Hanning window with no overlap, and each signal was separated into 20 subsegments. These subsegments are handled as independent samples and inherit the label of the original sample, and the sample distribution is shown in Table 1. The experimental results were validated using ten-fold cross-validation, which used the same data partitioning, feature extraction methods, hardware environments, software environments, and evaluation indicators for each experiment until all subsets were evaluated. The model is built with the PyTorch framework and runs on the GeForce RTX 3060 GPU. For training, the AdamW optimizer and cosine decay learning rate were utilized.

4.1. Emotion Recognition Results of McM

Three groups of network models (McM-1, McM-2, and McM-3) were put up in the experiment to examine how the McM network model affected the ability to recognize emotions. In comparison to McM-1, McM-2 has twice as many feature dimensions, while McM-3 is different from McM-1 in that the self-attention module is repeated more in McM-3. Table 2 displays the outcomes of emotion recognition tasks, while Table 3 displays the model settings. Accuracy, F1 score, Precision, and Recall are used to display the recognition performance indicators (Acc., Pre., Pre., and Recall.). In the binary classification challenge, the accuracy of the McM network is 85.23% in the valence dimension and 85.51% in the arousal dimension. The recognition accuracy rate remains nearly constant or even declines somewhat as the network size grows. McM-2 has more feature dimensions than McM-1. The convolutional network can learn more specifics as the convolutional layer is deepened, but it needs to be sufficiently deep to fully understand the input data. This method is limited in depth, and too deep of a convolution may provide similar and erroneous features. In addition, too many feature dimensions will result in an excessive number of model parameters, a lengthy training period, easy overfitting, and other issues. Adding more MvSA modules won't significantly change the recognition results, as shown by the comparison of McM-3 and McM-1, demonstrating the accuracy and effectiveness of MvSA's attention weighting.

The network's focus area and the outcome of McM's attention weight distribution can both be seen clearly in the heat map. To do this, the same subject's four data points, each designated HVHA, HVLA, LVHA, and LVLA, were utilized as inputs to plot the McM network's interest regions for each of the four emotion categories. A 3D feature matrix has been used to represent the left side of Figure 5 for reference. White represents the dot filled with 0, and the intensity of the red and blue colors shows how active the electrode is. In these 3D feature matrices, McM can first identify the underutilized electrode channels and decrease their attention weight, which is especially evident in the four corners of the spatial domain. There are several interleaved electrode channels in the centre of the three-dimensional feature matrix. These places that are filled with 0 after the convolution calculation will also contain some information about nearby electrodes. Second, the frontal-parietal lobe junction has electrodes that are constantly active, as seen by the three-dimensional eigenmatrix, and McM adds some weight to this region. An essential part of the brain that controls emotion and auditory perception is the left and right temporal lobes, which also have a high attention weight. In DEAP data sets, the occipital lobe performs poorly and is frequently disregarded as a visual perception and processing area.

It is clear from the heat map's condition that different emotional states result in varied EEG signal strengths being recorded by electrodes in various locations. There may be a connection between a brain region and an emotion state. The suggested McM network model may prioritize and screen out electrode channel input that is useful for emotion identification, increasing the model's training efficiency and recognition precision.

Table 3. McM emotion recognition model

Model	McM-1	McM-2	McM-3
MSCNN	convolution 36 1×1 filters	convolution 72 1×1 filters	convolution 36 1×1 filters
	$\begin{bmatrix} conv\ 36\ 3\times3 \\ conv\ 36\ 5\times5 \\ conv\ 36\ 7\times7 \end{bmatrix}$	$\begin{bmatrix} conv\ 72\ 3\times3 \\ conv\ 72\ 5\times5 \\ conv\ 72\ 7\times7 \end{bmatrix}$	$\begin{bmatrix} conv\ 36\ 3\times3 \\ conv\ 36\ 5\times5 \\ conv\ 36\ 7\times7 \end{bmatrix}$
	concatenation-layer: axis=1		
	convolution 36 1×1 filters	convolution 72 1×1 filters	convolution 36 1×1 filters
MvSA	$\begin{bmatrix} IPSA\ head=2 \\ APSA\ head=1 \\ BPSA\ head=2 \end{bmatrix}$	$\begin{bmatrix} IPSA\ head=2 \\ APSA\ head=1 \\ BPSA\ head=2 \end{bmatrix}$	$\begin{bmatrix} IPSA\ head=2 \\ APSA\ head=1 \\ BPSA\ head=2 \end{bmatrix} \times 3$
	convolution 36 1×1 filters	convolution 72 1×1 filters	convolution 36 1×1 filters
MLP	flatten-layer:axis=1		
	in:2916 hidden:729 out:1024 droprate:0.4	in:5832 hidden:1458 out:1024 droprate:0.4	in:2916 hidden:729 out:1024 droprate=0.4

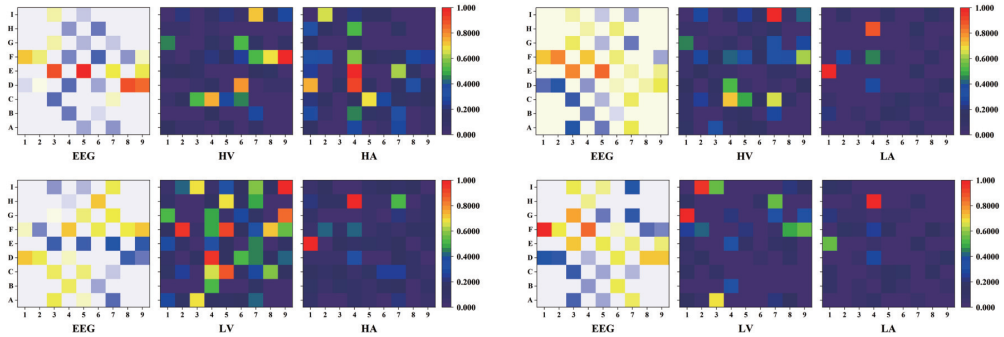

Figure 5. Areas of concern for deep learning networks

4.2. Multi-View Self-Attention Network

Seven sets of comparative experiments were carried out on the MvSA module by itself, three attention computing units (IPSA, APSA, and BPSA), and their pairwise combination of attention modules (IA-PSA, IB-PSA, and AB-PSA) in order to examine the performance of MvSA in the area of EEG emotion recognition. The ViT network model and the Swin network model were examined simultaneously with the MvSA module. A three-dimensional feature matrix is used as the input for each group of networks, and all hyperparameters are left at their default values. The binary emotion classification tasks for valence and arousal employ the same MLP layer.

Figure 6 displays the recognition outcomes using valence and arousal as emotional cues. In the valence dimension, the network utilizing only the MvSA module outperforms the network using only a single-view self-attention unit by 7.34%, 14.19%, and 13.62% in the training set and by 2.31%, 3.95%, and 5.16% in the test set, respectively. Comparing the network using only IPSA, APSA, and BPSA units, the network utilizing only MvSA modules improves the training set by 8.70%, 13.15%, and 13.42%, and the test set by 2.93%, 4.30%, and 5.16%, respectively, in terms of the arousal dimension. While generally superior to the single attention unit network, the network using the combination of two attention units performs worse than the MvSA module. The MvSA module enhanced the valence dimension by 3.82%, 3.75%, and 12.08%, respectively, and the arousal dimension by 3.27%, 5.23%, and 7.53%, when compared to the combined attention network. On the test set, it likewise performs better than IA-PSA, IB-PSA, and AB-PSA. On the test set, the MvSA module outperforms the ViT and Swin models. The recognition accuracy of MvSA in two dimensions was enhanced by 7.58% and 5.43% when compared to ViT and by 3.36% and 3.99%, respectively, when compared to Swin in the value and arousal dimensions.

It is clear that the attention weight acquired by utilizing a single attention unit has some limits,

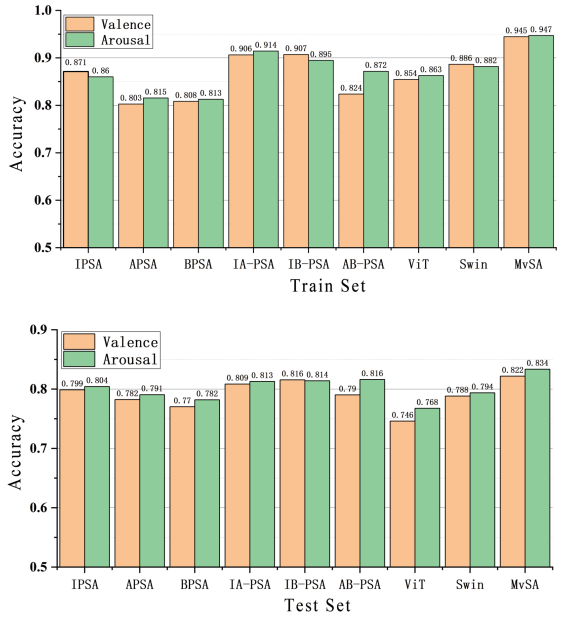

Figure 6. Training results of different self-attention networks

and the learning capabilities of the network model can be enhanced by merging the findings of the diverse viewpoints on attention. The experimental findings support the usefulness of the MvSA module's attention calculation from various angles. The model's recognition accuracy can be increased by combining these attention outcomes to produce attention weights that are more accurate.

4.3. Self-Attention Networks and Convolution

CNN does local modeling in the spatial domain, whereas the self-attention network's domain of application is global, making it simpler to capture the intricate relationships between electrode channels in various spatial locations and accelerating the network's convergence. CNN's capacity for inductive bias is absent from the self-attention network. The self-attention network's power to generalize and build models can both be enhanced by CNN's substantial prior information. Faster convergence speed and better model capacity can be attained by the McM network model created by fusing MSCNN and MvSA. Three sets of single-scale CNN models (CNN1, CNN2, and CNN3) with convolution kernels of 3×3, 5×5, and 7×7, as well as a set of MSCNN, were created in order to demonstrate the aforementioned result. Combine these CNN networks with the MvSA module to get CNN1-MvSA, CNN2-MvSA, CNN3-MvSA, and McM. Figure 7 depicts the training procedure. The convergence speed and recognition accuracy are increased when using the MvSA module after the convolutional network in comparison

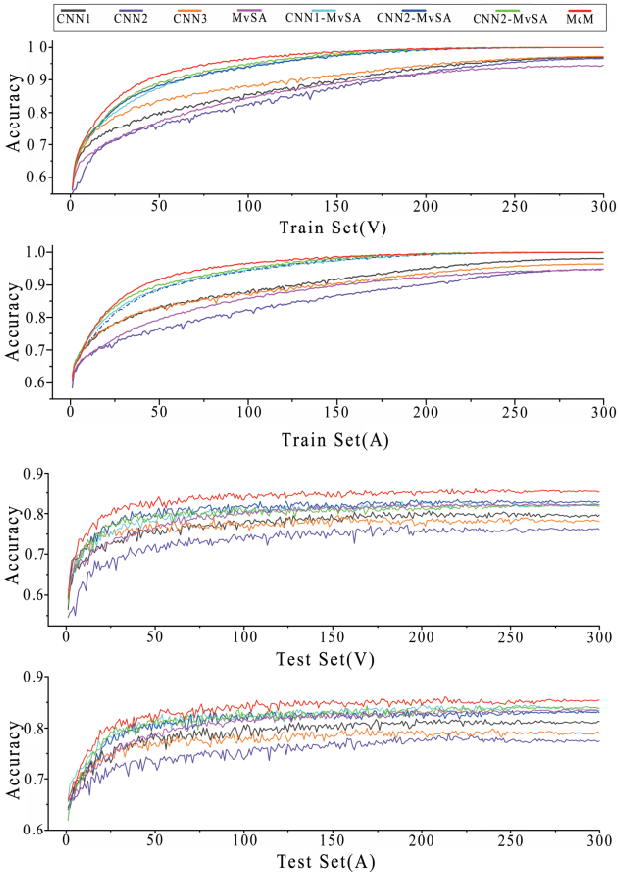

Figure 7. The training process of different networks

to the baseline CNN model. The accuracy of CNN1, CNN2, and CNN3 in the value dimension and arousal dimension on the training set increases with the training epoch when it is less than 250, and it increases slowly when the training epoch is larger than 250. Similar to this, once the training period had reached 175 when the attention module MvSA was added, the training data had already been fitted. On the test set, the model's recognition accuracy was seen. The CNN network with MvSA module increased by 2.47%, 6.80%, and 8.66% in the value dimension and 2.58%, 5.51%, and 8.63% in the arousal dimension when compared to the network without MvSA module, respectively.

MSCNN overcomes the drawback of utilizing a single convolution kernel for learning and is better able to handle high-dimensional time-domain and frequency information than the single-scale CNN model. When utilizing MSCNN, recognition accuracy in the valence dimension is 1.80%, 5.35%, and 8.30% greater than when using CNN1, CNN2, and CNN3, respectively, which use single-scale convolution. It increased by 1.99%, 5.78%, and 11.39% in terms of the arousal component, respectively. The valence dimension and

the arousal dimension both increased after MvSA was applied to the MSCNN model, by 3.75% and 2.23%, respectively. Table 4 displays the specific experimental outcomes.

Table 4. Self-attention network and CNN network training results

Model	Valence		Arousal	
	Acc.	F1.	Acc.	F1.
CNN1	0.7968	0.8152	0.8129	0.8417
CNN2	0.7613	0.7867	0.7750	0.8076
CNN3	0.7318	0.7509	0.7189	0.7520
MSCNN	0.8148	0.8364	0.8328	0.8559
CNN1-MvSA	0.8215	0.8393	0.8387	0.8607
CNN2-MvSA	0.8293	0.8469	0.8301	0.8524
CNN3-MvSA	0.8184	0.8327	0.8383	0.8582
MvSA	0.8219	0.8424	0.8371	0.8604
McM	0.8523	0.8655	0.8551	0.8727

4.4. Recent Research

This method's recognition performance under two affective labeling schemes is compared below to recent studies, all of which use DEAP data sets to conduct binary classification tasks in the two dimensions of valence and arousal. Because accuracy is the most widely utilized criterion, recognition accuracy is chosen as the model's index.

The 3D feature matrix is a great input option for emotion detection since it is straightforward to build, contains a plethora of time domain, frequency domain, and spatial characteristics, and works well with multi-channel EEG inputs. In order to increase the model's convergence speed and recognition accuracy, the method uses the McM network to integrate three-dimensional features and to carry out the attention operation on the three-dimensional feature matrix from various viewpoints. Table 5 describes the methodologies utilized in the comparison studies as well as the retrieved features and accuracy. The proposed McM network outperforms existing methods in terms of accuracy.

Table 5. Comparison of McM model with recent research

Method	Accuracy	
	valence	arousal
Tugba ERGIN and Mehmet Akif OZDEMIR[18]	67.93%	71.60%
Yuzhe Zhang, Huan Liu et al.[19]	72.89%	77.03%
Rabiul Islam, Milon Islam et al.[20]	78.22%	74.92%
Ante Topic, Mladen Russo[21]	74.91%	75.44%
Qi Li, Yunqing Liu et al.[22]	75.9%	79.3%
Magdiel Jiménez-Guarneros and Roberto Alejo-Eleuterio et al.[23]	Avg:79.36	
Yue Gao, Xiangling Fu et al.[24]	81.77%	81.95%
McM	85.23%	85.50%

5. Conclusion

In this paper, we propose a multi-channel EEG emotion recognition system that uses a 3D feature matrix as input and is based on multi-scale convolutional networks and self-attention networks. The suggested method may combine the properties of multi-channel EEG signals in the time domain, frequency domain, and spatial domain to recognize emotions. Additionally, the relationship between various emotional states and electrode spatial positions is discovered in the low-resolution EEG spatial domain through the attention mechanism. The information about useful positions is emphasized, and the information about useless positions is suppressed, resulting in a more targeted training of the network. Additionally, by contrasting the effectiveness of the suggested method with certain previous studies, the effectiveness of the proposed network in the field of emotion recognition is demonstrated.

References

[1] Houssein E. H., Hammad A. and Ali A A., Human emotion recognition from EEG-based brain-computer interface using machine learning: a comprehensive review [J]. *Neural Computing*, & Applications, vol. 2022, 34(15): 12527-57.

[2] Li X., Zhang Y. Z. and Tiwari P. et al., EEG Based Emotion Recognition: A Tutorial and Review [J]. *Acm Computing Surveys*, vol. 2023, 55(4).

[3] Takahashi K., Remarks on emotion recognition from multi-modal bio-potential signals [C]. *IEEE International Conference on Industrial Technology*, IEEE, 04 2004 International Conference on, F, 2005.

[4] Riedl M., Müller A. and Wessel N., Practical considerations of permutation entropy [J]. *European Physical Journal Special Topics*, vol. 2013, 222(2): 249-62.

[5] Namazi H., Aghasian E. and Ala T. S. et al., Complexity-based classification of EEG signal in normal subjects and patients with epilepsy [J]. *Technology and health care : official journal of the European Society for Engineering and Medicine*, vol. 2019, 28(1): 1-10.

[6] Toole JO., Discrete quadratic time-frequency distributions: Definition, computation, and a newborn electroencephalogram application [J]. *algorithms*, vol. 2013.

[7] Alazrai R., Homoud R. and Alwanni H. et al., EEG-Based Emotion Recognition Using Quadratic Time-Frequency Distribution [J]. *Sensors*, vol. 2018, 18(8).

[8] Candra H., Yuwono M. and Chai R. et al., Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine [C].

Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015.

[9] Mert A. and Akan A., Emotion recognition from EEG signals by using multivariate empirical mode decomposition [J]. *Pattern Analysis and Applications*, vol. 2016, 21(1)

[10] Rahman M. M., Sarkar A. K. and Hossain M. A. et al., Recognition of human emotions using EEG signals: A review [J]. *Comput Biol Med*, vol. 2021, 136(2): 104696.

[11] Song T., Zheng W. and Song P. et al., EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks [C]. *IEEE Transactions on Affective Computing*, vol. 2020, 11, no. 3, pp. 532-541, 1 July-Sept.

[12] Wen Z., Xu R. and Du J., A novel convolutional neural networks for emotion recognition based on EEG signal [C]. *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, IEEE, 2017.

[13] Zhong P., Wang D. and Miao C., EEG-Based Emotion Recognition Using Regularized Graph Neural Networks [J]. *Transactions on Affective Computing*, vol. 2022, 13, no. 3, pp. 1290-1301, 1 July-Sept.

[14] Woo S., Park J. and Lee J. Y. et al., Universals and cultural differences in the judgments of facial expressions of emotion [J]. *Pers Soc Psychol*, vol. 1987, 53(4): 712.

[15] Plutchik R., The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice [J]. *American Scientist*, vol. 2001, 89(4): 344-50.

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

[17] Pincus S. M., Approximate entropy as a measure of system complexity [J]. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 1991, 88(6): 2297-301.

[18] Ergin T., Ozdemir M. A. and Guren O., Emotion detection using EEG signals based on Multivariate Synchrosqueezing Transform and Deep Learning [C]. *Medical Technologies Congress (TIPTEKNO)*, IEEE, 2021, F.

[19] Zhang Y., Liu H. and Zhang D. et al., EEG-based emotion recognition with emotion localization via hierarchical self-attention [J]. *IEEE Transactions on Affective Computing*, vol. 2022.

[20] Islam R., Islam M. and Rahman M. M. et al., EEG channel correlation based model for emotion recognition [J]. *Computers in Biology and Medicine*, vol. 2021, 136: 104757.

[21] Topic A. and Russo M., Emotion recognition based on EEG feature maps through deep learning network [J]. *Engineering Science and Technology an International Journal*, vol. 2021, 24(6): 1442-54.

[22] Li Q., Liu Y. and Liu C. et al., EEG signal processing and emotion recognition using Convolutional Neural Network [C]. *2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)*, IEEE, 2021, F.

[23] Jiménez-Guarneros M. and Alejo-Eleuterio R., A Class-Incremental Learning Method Based on Preserving the Learned Feature Space for EEG-Based Emotion Recognition [J]. *Mathematics*, vol. 2022, 10(4): 598.

[24] Gao Y., Fu X. and Ouyang T. et al., A Class-Incremental Learning Method Based on Preserving the Learned Feature Space for EEG-Based Emotion Recognition [J]. *Adv. Eng. Informatics*, vol. 2022, 29: 1574-8.