Research article

Gesture recognition of continuous wavelet transform and deep convolution attention network

Xiaoguang Liu[1,2], Mingjin Zhang[1,2], Jiawei Wang[1,2], Xiaodong Wang[3], Tie Liang[1,2], Jun Li[1,2], Peng Xiong[1,2] and Xiuling Liu[1,2,*]

[1] College of Electronic and Information Engineering, Hebei University, Baoding, China
[2] Key Laboratory of Digital Medical Engineering of Hebei Province, Hebei University, Baoding, China
[3] Affiliated Hospital of Hebei University, Baoding, China

* Correspondence: Email: liuxiuling121@foxmal.com; Tel: +8613722296965.

Abstract: To solve the problem of missing data features using a deep convolutional neural network (DCNN), this paper proposes an improved gesture recognition method. The method first extracts the time-frequency spectrogram of surface electromyography (sEMG) using the continuous wavelet transform. Then, the Spatial Attention Module (SAM) is introduced to construct the DCNN-SAM model. The residual module is embedded to improve the feature representation of relevant regions, and reduces the problem of missing features. Finally, experiments with 10 different gestures are done for verification. The results validate that the recognition accuracy of the improved method is 96.1%. Compared with the DCNN, the accuracy is improved by about 6 percentage points.

Keywords: sEMG; gesture recognition; continuous wavelet transform; DCNN; SAM; residual module

1. Introduction

With the frequent occurrence of diseases, traffic accidents, and natural disasters, the number of patients with physical disabilities is gradually increasing. The most significant proportion of patients have some upper limb disabilities [1]. The rehabilitation robot for assisting movement using surface electromyography (sEMG) is becoming a hot issue of research. By exploring different recognition methods, sEMG is used to control intelligent prosthetics more efficiently and accurately. Intelligent prosthetics can help people complete simple tasks in many daily activities. For example, intelligent

prosthetics can help disabled patients finish hand rehabilitation training. SEMG reflects neuromuscular activity and contains some vital information with muscle activity [1–3]. SEMG is one of the most widely used signals for identifying gestures due to being easily acquired [3].

Although the original sEMG contains a large amount of information about the intention of hand movements, it cannot be directly applied to recognition due to the non-smooth and non-periodic nature of the signal [2]. Features of sEMG are very important, and they can determine the performance of the recognition method. It is not easy to extract significant features from sEMG. The traditional manual extraction method can only extract some low-level features of sEMG. So, extracting meaningful features of sEMG has been a huge challenge for gesture recognition. Many researchers are trying to explore new methods for extracting sEMG features [3]. sEMG contains a large amount of noise, due to the differences between individuals and the influence of the collection environment. Therefore, the raw sEMG needs to be preprocessed. In recent years, researchers have explored new networks to improve gesture recognition accuracy. As the depth of some networks increases, some factors affecting network performance appear [2–5]. This paper explores a new gesture recognition algorithm by improving CNN to reduce the adverse effects of network deepening.

CNN-based methods can capture local features in 2D or 3D spaces [3]. Na Duan from Fudan University preprocessed the sEMG using short-time Fourier transform, and extracted the spectrum diagram of sEMG, to extract the depth features [4]. However, the receptive field of convolutions limits CNNs [3]. The depth of a CNN can affect its ability to extract features. Especially, the gradient's disappearance and the increasing depth of the network can cause a lack of data feature information, and result in lower identification accuracy [5].

In order to solve the above problems, many researchers focused on various preprocessing methods for improved CNN models to enhance the recognition accuracy [6–9]. Liukai Xu et al. changed the energy kernel phase map of sEMG into grayscale images. The grayscale images were used as the input of a CNN to recognize different gestures [8]. In the latest study, Mehmet Akif Ozdemir et al. obtained time-frequency images of sEMG by applying short-time Fourier transform, continuous wavelet transform, and Hilbert-Yellow transforms. The pre-trained ResNet-50 network was used for gesture classification, and the best recognition accuracy was 93.75% [9]. Although the accuracy of gesture recognition has improved, it is still far from being usable for practical applications.

This paper proposes continuous wavelet transform combined with a deep convolutional neural network-spatial attention module (DCNN-SAM) model to realize gesture recognition. The main content of the paper is as follows.

● sEMG is preprocessed by filtering, and then the time-frequency graph of sEMG is extracted by using continuous wavelet transform.

● The DCNN-SAM model is designed, and the model can enhance the feature expression of crucial regions and reduce features loss of sEMG. The time-frequency graph of sEMG is used as the input of the model.

● An experiment is designed to verify the effectiveness of the proposed method using the sEMG collected from 10 subjects.

The structure of the paper is as follows: The second part mainly presents the construction of DCNN-SAM, including the spatial attention mechanism and residual module. In section 3, the experiment is designed, including the preprocessing and feature extraction, as well as results and analysis. Finally, we conclude the full text and discuss future work.

2. Materials and methods

In order to get the frequency components, sEMG is processed by continuous wavelet transform. DCNN-SAM is improved by embedding the residual module. Then, the time-frequency spectrogram of the sEMG is selected as input of the DCNN-SAM.

2.1. Spatial attention module

The spatial attention mechanism is used to enhance the feature representation of some critical regions [10]. It is essentially transforming the spatial information in the original image to another space through the module of spatial transformation and retains the essential information in it [11]. It generates a weight mask for each location and weights the output, and in turn it enhances the feature representation of some critical specific target regions. At the same time, features of irrelevant regions are also weakened [12]. Figure 1 shows the structural diagram of the spatial attention mechanism.
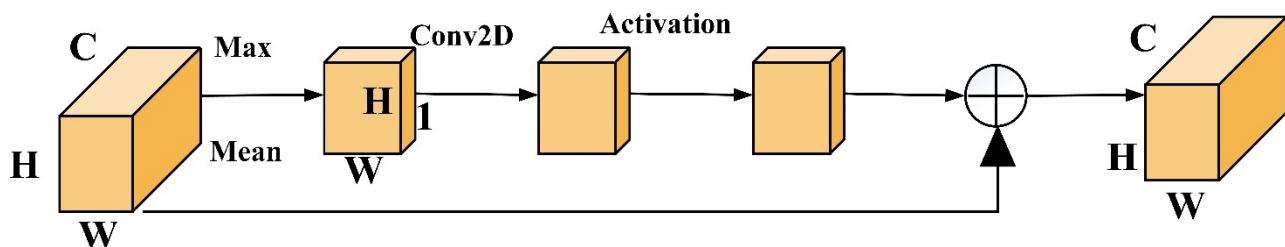

Figure 1. Spatial attention mechanism model.

The process is as follows: The input is an $H \times W \times C$ sEMG time-frequency spectrogram. First, a maximization and averaging operation is performed along the dimensional direction of the channel to produce two feature maps representing different information. A two-dimensional convolution with a convolution kernel size of 1×1 is used to perform the feature fusion operation [12,13]. Then, the two feature maps are combined to obtain a feature map of $H \times W \times 1$ size. Finally, the two-dimensional spatial attention feature map is output after activation by the sigmoid activation function. The original input time-frequency spectrogram is superimposed to obtain the spatial attention output feature map [13], and this in turn enables the features of the target region to be enhanced and retains the critical feature information that helps in gesture recognition classification.

2.2. Residual module

The idea of the residual module is that it changes the mapping from the initial feature X to the learned feature Y. Finally, the learned residual feature is added to the original feature [14]. The residual module can contain multiple convolutional layers. The original input feature is directly connected with the result of the output feature by skipping some convolutional network layers. Finally, all of them are taken as input of the activation function to get the learning output feature [15]. Figure 2 shows the structure diagram of the residual module. The residual module can optimize the deep learning network, alleviate the problem of network degradation, and ensure the integrity of signal features.
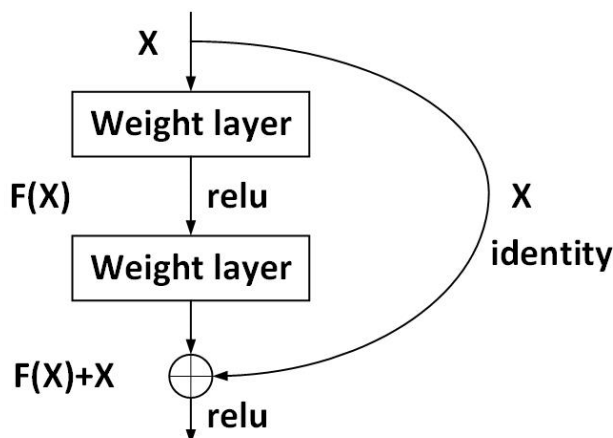
Figure 2. Residual module structure.

2.3. Network model

The CNN is one of the classic algorithms of deep learning [14–16]. In this section, four deep convolutional modules are designed to build the deep convolutional neural network (DCNN). The spatial attention mechanism module is added to construct the DCNN-SAM model. Figure 3 shows the model structure diagram.
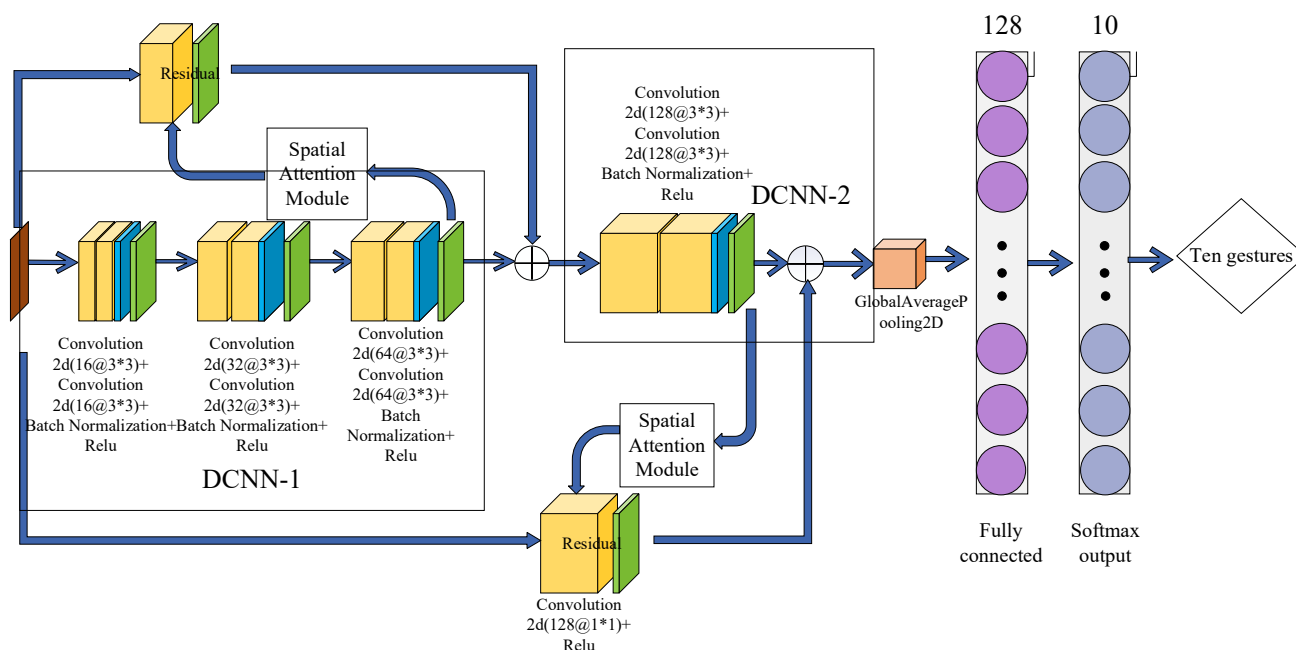

Figure 3. Network structure diagram.

The DCNN contains two parts, DCNN-1 and DCNN-2. DCNN-1 carries out the convolution operation on the time-frequency spectrogram of the sEMG. Then, the output of DCNN-1 is used as input for DCNN-2 to obtain more detailed sEMG features. DCNN-1 is composed of three

convolutional layers. DCNN-2 is composed of one convolutional layer. A batch-normalized BN layer exists between each convolutional block and activation layer. A residual module is embedded between two substructures, and a spatial attention mechanism is introduced to ensure the integrity of the features in the convolution operation. Finally, the full connection and softmax layer are used to get the result.

The process is divided into three stages, the input stage of data, the deep convolution stage, and the feature output stage. The input of the model is a time-frequency spectrogram of size *40 × 40 × 1*. The features are extracted and learned at the deep convolution stage. Finally, the trained model is used to recognize the corresponding gesture.

The convolutional module consists of two-dimensional convolutional kernels, a batch normalization (BN) layer, and an activation layer. The size of the convolutional kernel is *3 × 3*, the moving step is 1, and the numbers of convolutional kernels in the four layers are 16, 32, 64 and 128, respectively. The BN layer improves the speed of network training by inhibiting network gradient disappearance or gradient explosion during network training. The activation function can better extract sparse features and improve the learning speed and learning accuracy by using a rectified linear unit (Relu). The spatial attention mechanism is added between the third and fourth convolutional modules to strengthen the features in specific regions. A residual module is embedded for optimizing DCNN-SAM. The numbers of convolutional kernels of two residual structures are 64 and 128, respectively. The residual structure can effectively alleviate the network degradation problem by connecting the initial input with the output feature. Then, a global average pooling layer is used to calculate the average of all pixels of each channel to obtain a new *1 × 1* channel map. It can suppresses overfitting and reduces the redundant network parameters. Finally, a fully connected layer of size 128 is used to connect the data and collect the information related to the sample feature space. A softmax function is used as the output layer to get the final classification results.

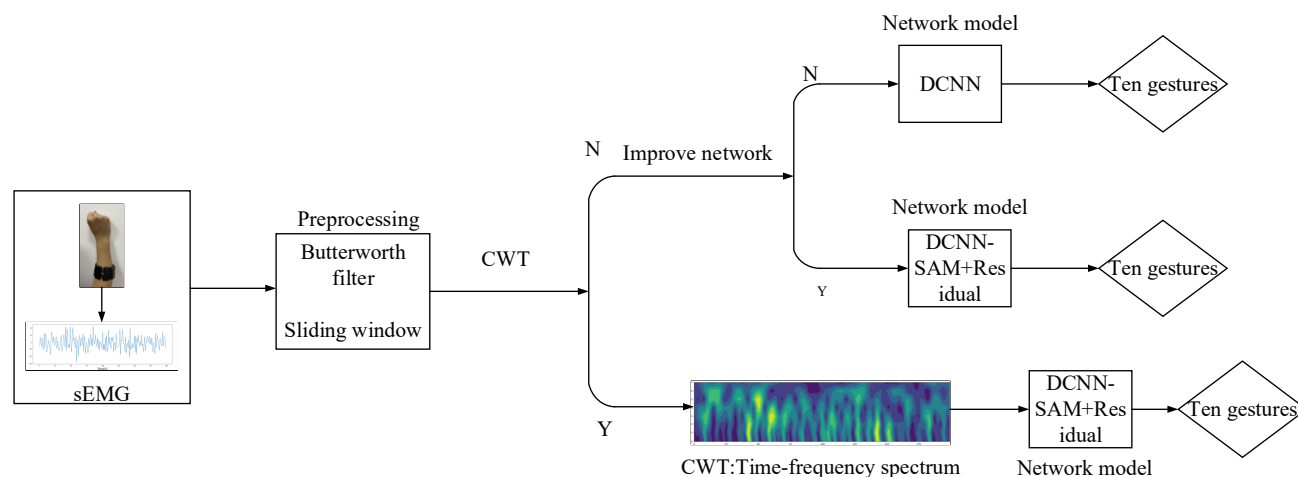## 3. Experiments and results analysis

Figure 4. Experimental flow chart.

A dataset containing ten different gestures was acquired by using the Myo EMG armband with ten subjects. The experimental included preprocess sliding window segmentation [15,16], continuous

wavelet transform, and the DCNN-SAM model training and testing [17]. Figure 4 shows the process diagram of the experiment.

3.1. Acquisition of sEMG

The Myo EMG armband was used to acquire sEMG. It has eight sEMG channels with a sampling frequency of 200 Hz [18,19]. The Myo EMG armband is worn on the subject's right forearm, as shown in Figure 5. Before the signal acquisition, the subject is told not to do strenuous exercise to avoid the effects of muscle fatigue during the experiment. The skin surface is wiped with alcohol pads to remove the influential oil and dead skin to ensure good contact between the subject's skin and the electrodes [20].

Figure 5. Wearing position of the Myo EMG armband.

The data was collected from 10 healthy subjects, including 5 males and 5 females, all aged between 20 and 30 years. Table 1 shows the details of the subjects. All subjects were informed of the details of the experiment and voluntarily completed the relevant written informed consent form.

Table 1. The information of the subjects.

Number of subjects	Male to female ratio	Average height (cm)	Average weight (kg)	Average BMI (kg/m^2)
10	1:1	170.5 ± 9.2	60.17 ± 3.12	21.35 ± 3.18

The designed ten gestures included fist, one, two, OK, open hand, praise, six, up, down, and eight. Figure 6 shows the gestures. Subjects wore the Myo EMG armband on the right forearm and extended the arm straight and perpendicular to the ground. Subjects were trained on the corresponding gestures before formal experiments. All gestures were repeated 15 times, and each action was recorded about 16 s. Each gesture yields 3200 sampling points under the sampling frequency of 200 Hz.
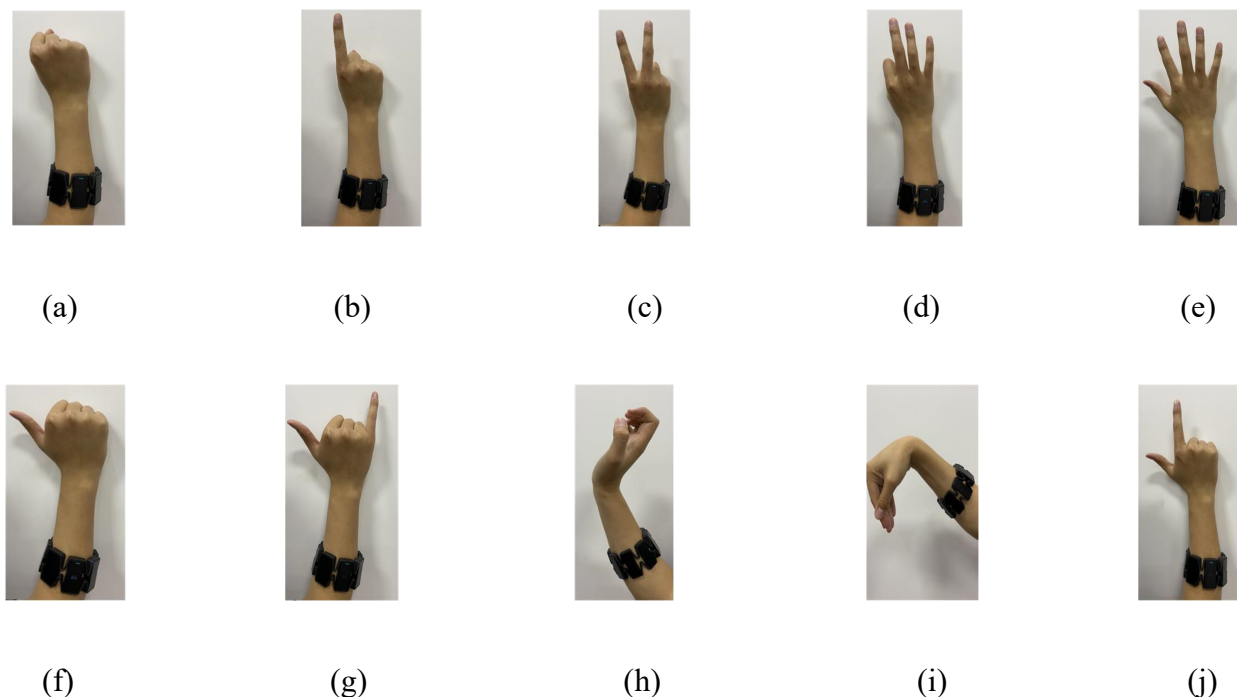
Figure 6. Classified gestures: (a) fist; (b) one; (c) two; (d) OK; (e) open hand; (f) praise; (g) six; (h) up; (i) down; (j) eight.

3.2. sEMG data filtering and denoising

Effective information of sEMG is mainly distributed between 20 and 200 Hz [21]. SEMG will be interfered with by the industrial frequency signal at 50 Hz and the low-frequency signal below 20 Hz [20–22]. So, the Butterworth high-pass and band-pass filters were selected to process sEMG. The low-frequency noise below 20 Hz was filtered using the high-pass filter, and the industrial frequency noise at 50 Hz was filtered using the band-pass filter [23].

The Butterworth filter has been widely used due to being simple and easy to design [24]. The Butterworth high-pass filter is shown in Eq (1), where w represents the frequency, w_c represents the cut-off frequency, $|H(w)|$ represents the amplitude, and n represents the order of the filter. We selected a third-order high-pass filter to filter out low-frequency noise below 20 Hz.

$$|H(w)|^2 = \frac{1}{1+\left(\frac{w}{w_c}\right)^{2n}} \tag{1}$$

A band-pass filter can block the passage of specific frequency. Equation (2) expresses an ideal filter, where w_0 = 50 Hz. It can filter out 50 Hz power-frequency noise.

$$\left|H(e^{jw)}\right| = \begin{cases} 1, & w \neq w_0 \\ 0, & w = w_0 \end{cases} \tag{2}$$

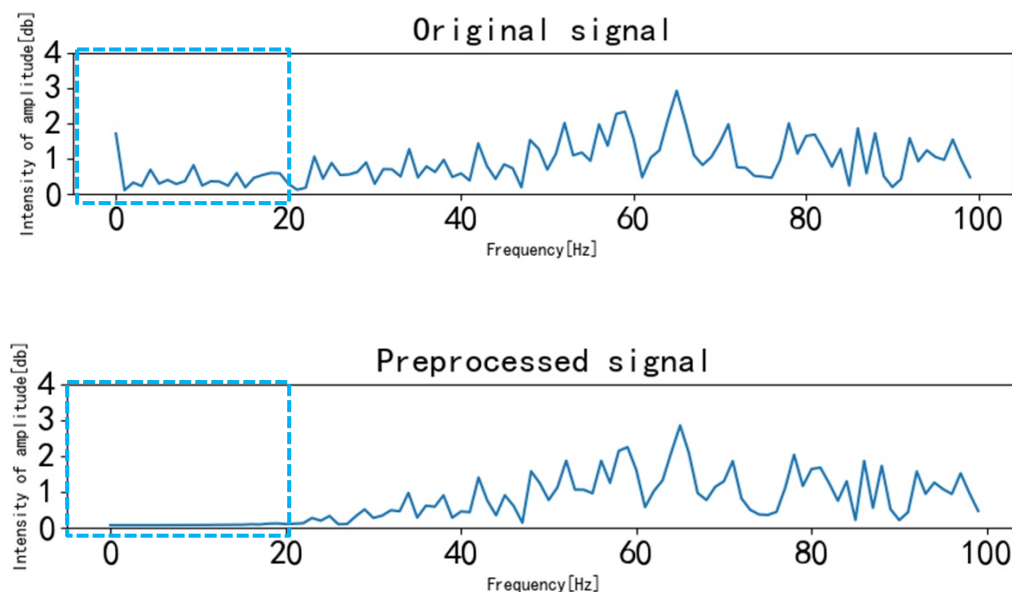Figure 7 shows the spectrum of sEMG after filtering.

Figure 7. sEMG before and after filtering.

3.3. Sliding window data segmentation

SEMG must be segmented before continuous wavelet transform. The sliding window segmentation method is used to cut the sEMG to ensure signal continuity [25]. The number of overlapping windows is shown in Eq (3) [26].

$$W = \frac{n-k}{s} + 1 \tag{3}$$

W is the number of windows, n is the number of sampling points, k is the window size, and s is the sliding step length. The sliding window and sliding step size are 200 and 100, respectively. Figure 8 shows the schematic diagram of the sliding window segmentation.
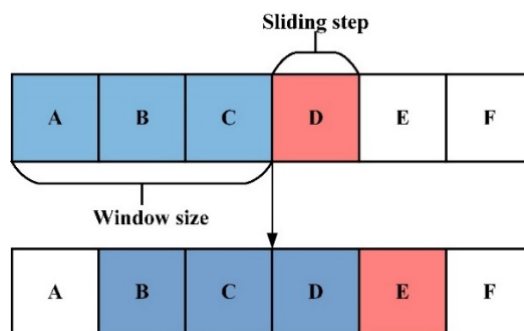

Figure 8. Schematic diagram of sliding window segmentation.

3.4. Continuous wavelet transform (CWT)

The wavelet transform is used for time-frequency analysis [27,28]. It can analyze information on

time-frequency and adapt to the requirements of time-frequency analysis automatically [29]. Equation (4) shows the wavelet transform.

$$WT(a, \ \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) * \varphi\left(\frac{t-\tau}{a}\right) dt \tag{4}$$

In Eq (4), a represents the scale parameter, τ represents the time parameter, $f(t)$ represents the original signal, $WT(a,\tau)$ represents the transform value, and $\varphi(t)$ represents the wavelet transform function. Equation (4) shows that the frequency w and time T correspond to the scale [30]. The wavelet transform can obtain the frequency components of the signal and the moments when each frequency component appears [31]. Through continuous wavelet transform, the time-frequency spectrogram can be extracted to obtain information on the signal frequency and the time. Figure 9 shows a channel's sEMG map and the time-frequency spectrum map.
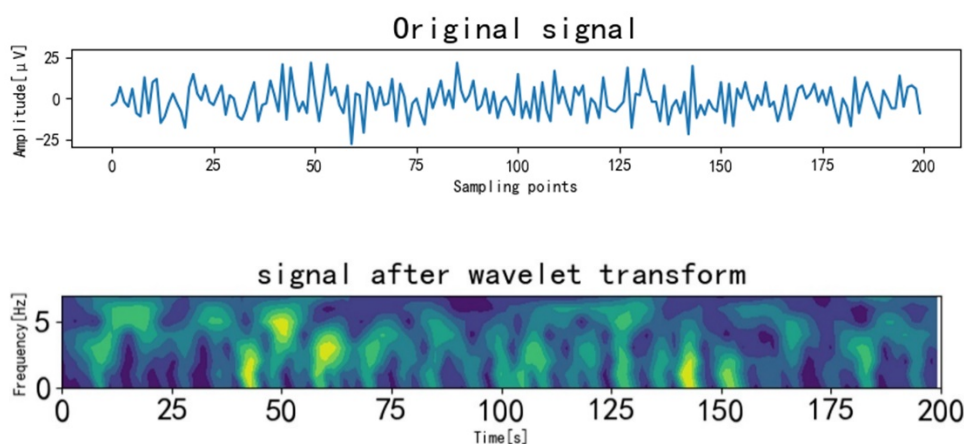

Figure 9. sEMG map with the time-frequency spectrum map.

3.5. Analysis of experimental results

The sEMG for the experiment was divided into three parts: 70% of the data are used for training, 20% of the data are used for testing, and 10% of the data are used for validation. DCNN-SAM uses an adaptive optimizer (Adam), the loss function uses an inter-class cross-entropy function, the random deactivation is 0.5, the initial learning rate is 0.001, and the number of iterations for network training is 100.

One subject is evaluated individually to verify the feasibility of the experiment. In order to demonstrate the superiority of this method, the result is compared with the classification models commonly used.

Figure 10 shows the accuracy and loss curves for the classification recognition with 100 iterations of training. From the graph, we can see that the accuracy value of the set increases rapidly, and the loss value decreases rapidly in the first 10 iterations. The accuracy value is greater than 90%, and the loss value is less than 0.2 after 10 iterations. The model can converge rapidly. The test set's accuracy and loss value remain stable after the number of iterations reaches 18.
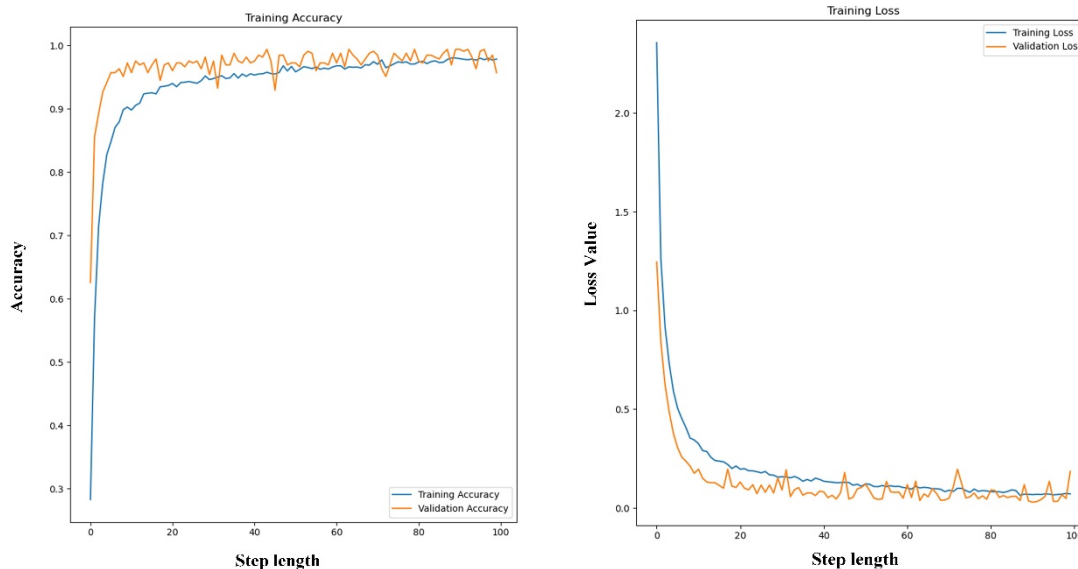
Figure 10. Curves of accuracy and loss functions for training and testing sets.

The result is compared with other models, including DCNN-SAM without continuous wavelet transform, DCNN and CNN. We chose four measures to evaluate the recognition results: Accuracy, Recall, Precision, and F1-score.

To illustrate the concept of four measures, we introduce several definitions: TP, TN, FP, and FN. TP is the number of correctly classified samples, TN is the number of incorrectly classified samples, FP is the number of incorrectly classified samples as correct samples, and FN is the number of correctly classified samples as incorrect samples. The accuracy is the ratio of the number of correctly predicted samples to the total number of samples, the precision is the ratio of the number of correctly predicted samples to all of the number of correctly predicted samples, the recall is the ratio of the number of correctly predicted samples to all of the number of correctly predicted samples, and the F1-score can be regarded as a weighted average of the accuracy and recall. A higher F1 score means a better performance of the network model. Equations (5)–(8) shows the formulas for the four measures.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{8}$$

Table 2 shows the results of the experiment of ten subjects.

Table 2. Results of ten gestures categories.

Classifier	Accuracy	Recall	Precision	F1-score
CWT+DCNN-SAM	0.961	0.963	0.973	0.958
DCNN-SAM	0.929	0.931	0.937	0.928
DCNN	0.900	0.904	0.911	0.898
CNN	0.820	0.822	0.828	0.817

From the experimental results, we can see that the method proposed significantly improved accuracy compared to other model methods, with an accuracy of 96.1%. Using continuous wavelet transform to analyze the sEMG has a more significant impact on the classification. We can see that continuous wavelet transform extracted the useful frequency information, helping the model improve the accuracy of gesture recognition by about 4 percentage points. This indicates that the improvement method of the Spatial Attention Module and the residual module achieve a significant effect, and it not only improves the recognition accuracy of the model but also avoids the problems that affect the classification accuracy, such as gradient disappearance due to the increased of network depth. Thus the model can achieve better performance.

Recently, some SOTA algorithms have been developed in gesture recognition, including gesture recognition based on the postural graph convolutional network (GCN) method. The network takes related gestures of hand and body as the input, and then the GCN, the residual connection, and the residual module structure are used for classification. The pyramid structure of multi-scale features is extracted by using the multi-scale multi-attention time-frequency converter network (MSMHA-VTN), the multi-scale head attention model of the transformer. The model adopts different attention dimensions for each head of the transformer, and it can provide an attention mechanism at the multi-scale level. In some recent studies, researchers proposed a new gesture recognition method based on AlexNet transfer learning and Adam optimizer. The classification of different configuration transfer learning is tested. Others proposed some new methods based on CNN, building a 12-layer CNN as the backbone. The improved 20-channel data enhancement method was used to avoid overfitting. CNN was used to recognize different gestures. We propose a new gesture recognition method of this present paper is compared with the SOTA algorithms mentioned above. Table 3 shows the comparison results. According to the data, the accuracy of our method is improved compared with the SOTA algorithms, indicating the feasibility of the algorithm proposed.

Table 3. The recognition accuracy of different SOTA classification algorithms.

Classifier	Accuracy	Recall	Precision	F1-score
CWT+DCNN-SAM	0.961	0.963	0.973	0.958
GCN	0.926	0.929	0.931	0.926
MSMHA-VTN	0.912	0.912	0.918	0.910
AlexNet+Adam	0.915	0.914	0.921	0.916
CNNSP+CNN	0.908	0.908	0.913	0.906

Figure 11 shows the confusion matrix of classification results, and we can see the recognition accuracy of each gesture. Most of the gestures have a relatively high recognition accuracy, with an overall concentration on the diagonal. The recognition accuracy of individual gestures is relatively low,

such as the two gestures of one and down. The reason may be that there are some differences in the strength of signals in the process of acquisition, and it affects the subsequent data processing and classification. In addition, the low degree of differentiation between gestures is also a reason for the different recognition accuracy values of different gestures, as it is easy for them to be mistaken for other similar gestures. The magnitude of the action varies from gesture to gesture, so the part of the arm muscles involved also varies, causing different recognition accuracies simultaneously.
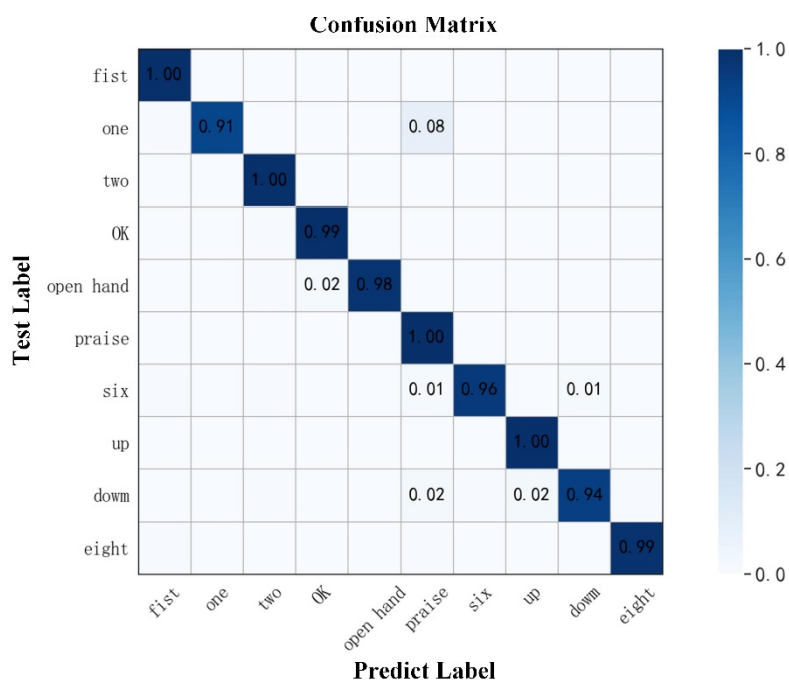

Figure 11. Confusion matrix of classification and recognition of ten gestures.

We selected five healthy subjects without muscle diseases for a separate experimental evaluation of gesture recognition. The five subjects included three males and two females. Table 4 shows information of the five subjects. Figure 12 shows the evaluation results.

It can be seen that the accuracy of the proposed method is significantly improved compared to other classification algorithms. The recognition accuracy can reach 98.5% in the experiments with a single subject. This shows that the algorithm also applies to the gesture recognition of a single subject.

Table 4. Information of five healthy subjects.

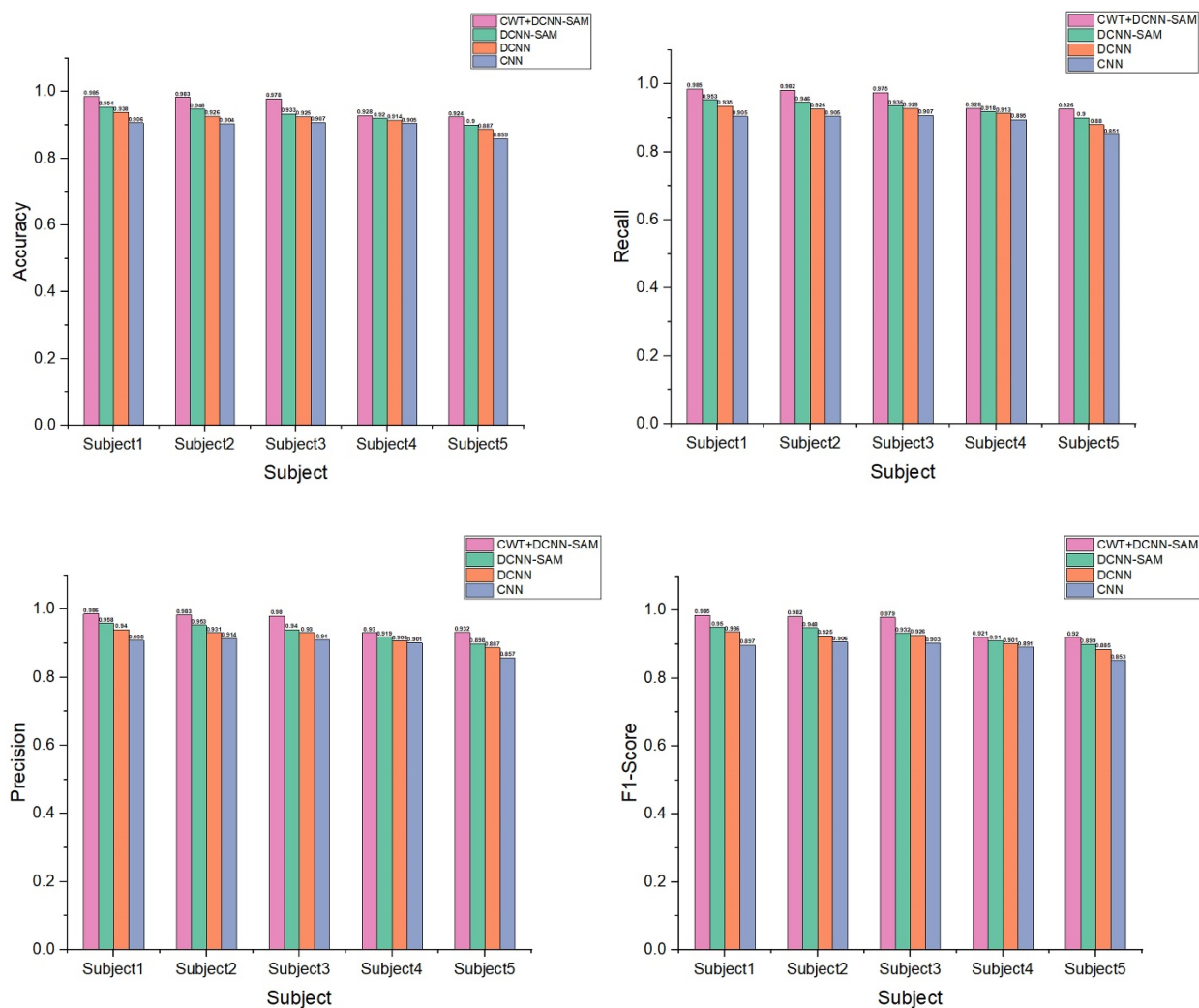| Subject	Gender	Age	Height [cm]	Weight [kg]
Subject 1	Male	24	174	118
Subject 2	Male	25	176	138
Subject 3	Male	25	183	130
Subject 4	Female	26	168	110
Subject 5	Female	25	165	106

Figure 12. Measurement indexes of five subjects under different classification models.

4. Conclusions

We propose a DCNN-SAM model for sEMG gesture recognition. First, time-frequency analysis of sEMG is done using continuous wavelet transform. Then, the spatial attention mechanism is applied to construct the DCNN-SAM model. Finally, DCNN-SAM is optimized by embedding the residual module to avoid the gradient disappearance and local information loss. The experiment demonstrated that the method has a good recognition and classification. Although the proposed DCNN-SAM model improves the recognition accuracy, the training time of the model is slightly increased due to the deepening and increasing complexity of the network. In future research, the model can be further improved to reduce the training time.

Acknowledgments

Conflict of interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper. We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled "Gesture recognition based on continuous wavelet transform and deep convolution attention network".

Data availability

The programs' data used to support the findings of this study are available from the corresponding author upon request.

References

1. X. Jiang, Y. H. Li, K. Zou, X. D. Yuan, A multi-channel correlation feature gesture recognition method for electromyographic signals, *Comput. Eng. Appl.*, **2023** (2023), 1–9.

2. L. Liu, H. Y. Pu, LSTM-based multi-dimensional feature gesture recognition in real time, *Comput. Sci.*, **48** (2021), 328–333.

3. Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI, *Inf. Fusion*, **91** (2023), 376–387. https://doi.org/10.1016/j.inffus.2022.10.022

4. N. Duan, L. Z. Liu, X. J. Yu, Q. Li, S. C. Yeh, Classification of multichannel surface-electromyography signals based on convolutional neural networks, *J. Ind. Inf. Integr.*, **15** (2019), 201–206. https://doi.org/10.1016/j.jii.2018.09.001

5. F. Milletari, N. Navab, S. A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, (2016), 565–571. https://doi.org/10.1109/3DV.2016.79

6. L. M. Luo, Z. Y. Xu, X. H. Xie, L. Li, Convolutional neural network-based gesture recognition for surface EMG signals in Chinese, *Comput. Program. Skills Maint.*, **2021** (2021), 137–138+163. https://doi.org/10.16184/j.cnki.comprg.2021.01.049

7. V. Shanmuganathan, H. R. Yesudhas, M. S. Khan, M. Khari, A. H. Gandomi, R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals, *Neural Comput. Appl.*, **32** (2020), 16723–16736. https://doi.org/10.1007/s00521-020-05349-w

8. L. K. Xu, K. Q. Zhang, Z. H. Xu, G. Yang, Convolutional neural network human gesture recognition algorithm based on energy kernel phase map of surface EMG signals, *J. Biomed. Eng.*, **38** (2021), 621–629.

9. M. A. Ozdemir, D. H. Kisa, O. Guren, A. Akan, Hand gesture classification using time–frequency images and transfer learning based on cnn, *Biomed. Signal Process. Control*, **77** (2022), 103787. https://doi.org/10.1016/j.bspc.2022.103787

10. K. Feng, S. Dong, D. B. Liu, Surface myoelectric signal gesture recognition based on empirical modal decomposition-wavelet packet transform, *Chin. J. Med. Phys.*, **38** (2021), 461–467.

11. X. Xi, W. Jiang, X. Hua, H. Wang, C. Yang, Y. B. Zhao, et al., Simultaneous and continuous estimation of joint angles based on surface electromyography state-space model, *IEEE Sens. J.*, **21** (2021), 8089–8099. https://doi.org/10.1109/JSEN.2020.3048983

12. Y. H. Li, X. Jiang, K. Zou, X. D. Yuan, A multi-stream convolutional myoelectric gesture recognition network with fused attention mechanism, *Comput. Appl. Res.*, **38** (2021), 3258–3263. https://doi.org/10.19734/j.issn.1001-3695.2021.04.0100

13. L. H. Shi, Research on dynamic gesture recognition based on attentional convolutional neural network in Chinese, *Opt. Technol.*, **46** (2020), 750–756. https://doi.org/10.13741/j.cnki.11-1879/o4.2020.06.019

14. C. Yan, P. A. Mu, Research on static gesture recognition in complex context, *Software Guide*, **21** (2022), 171–176.

15. L. K. Xu, K. Q. Zhang, Z. H. Xu, G. K. Yang, Human gesture recognition algorithm based on convolution neural network based on energy kernel phase diagram of surface EMG signal, *J. Biomed. Eng.*, **38** (2021), 621–629.

16. W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, W. Geng, A Multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface, *Pattern Recognit. Lett.*, **119** (2017), 131–138. https://doi.org/10.1016/j.patrec.2017.12.005

17. H. X. Cheng, K. Cheng, L. Cheng, Z. Q. Jiang, A gesture recognition method based on residual fusion biflow graph convolutional network in Chinese, *Electron. Meas. Technol.*, **45** (2022), 20–24. https://doi.org/10.19651/j.cnki.emt.2108617

18. C. Liu, X. F. Feng, A gesture recognition method based on wireless signal and improved TCN in Chinese, *Comput. Eng. Des.*, **43** (2022), 2317–2324. https://doi.org/10.16208/j.issn1000-7024.2022.08.029

19. C. Tepe, M. Erdim, Classification of surface electromyography and gyroscopic signals of finger gestures acquired by Myo armband using machine learning methods, *Biomed. Signal Process. Control*, **75** (2022),103588. https://doi.org/10.1016/j.bspc.2022.103588

20. X. J. Zhang, C. Y. Li, A gesture segmentation recognition algorithm based on deep learning multi-feature fusion in Chinese, *J. Jinan Univ. (Nat. Sci. Ed.)*, **36** (2022), 286–291. https://doi.org/10.13349/j.cnki.jdxbn.20220110.001

21. F. Kong, J. Deng, Z. Fan, Gesture recognition system based on ultrasonic FMCW and ConvLSTM model, *Measurement*, **190** (2022), 110743. https://doi.org/10.1016/j.measurement.2022.110743

22. H. F. Hassan, S. J. Abou-Loukh, I. K. Ibraheem, Teleoperated robotic arm movement using electromyography signal with wearable Myo armband, *J. King Saud Univ. Eng. Sci.*, **32** (2019), 378–387. https://doi.org/10.1016/j.jksues.2019.05.001

23. J. M. Fajardo, O. Gomez, F. Prieto, EMG hand gesture classification using handcrafted and deep features, *Biomed. Signal Process. Control*, **63** (2020), 102210. https://doi.org/10.1016/j.bspc.2020.102210

24. C. Tepe, M. C. Demir, The effects of the number of channels and gyroscopic data on the classification performance in EMG data acquired by Myo armband, *J. Comput. Sci.*, **51** (2021), 101348. https://doi.org/10.1016/j.jocs.2021.101348

25. J. O. Pinzón-Arenas, R. Jiménez-Moreno, A. Rubiano, Percentage estimation of muscular activity of the forearm by means of EMG signals based on the gesture recognized using CNN, *Sens. Bio-Sens. Res.*, **29** (2020), 100353. https://doi.org/10.1016/j.sbsr.2020.100353

26. L. Xu, K. Zhang, G. Yang, J. Chu, Gesture recognition using dual-stream CNN based on fusion of sEMG energy kernel phase portrait and IMU amplitude image, *Biomed. Signal Process. Control*, **73** (2022), 103364. https://doi.org/10.1016/j.bspc.2021.103364

27. X. Z. Wang, P. Yue, J. W. Wang, Z. Li, Q. H. Tian, Wavelet transform low-frequency information with Xception network for static gesture recognition, *Software Guide*, **20** (2021), 12–19.

28. C. Q. Hu, N. Qu, S. Zhang, Z. Jiang, Application of continuous wavelet transform and deep residual shrinkage network with attention mechanism to low-voltage series arc fault detection, *Power Grid Technol.*, **2022** (2022), 1–10.

29. J. X. Li, L. Shen, C. Cai, R. N. Yang, K. Luo, Improved frequency slicing wavelet transform and convolutional neural network for myoelectric signal recognition of hand gestures, *J. Nanchang Univ*, **43** (2021), 401–408.

30. Y. Jiang, C. Chen, X. Zhang, C. Chen, Y. Zhou, G. Ni, et al., Shoulder muscle activation pattern recognition based on sEMG and machine learning algorithms, *Comput. Methods Programs Biomed.*, **197** (2020), 105721. https://doi.org/10.1016/j.cmpb.2020.105721

31. Z. C. Hu, Y. T. Zhou, B. J. Shi, H. He, A static gesture recognition algorithm combining attention mechanism and feature fusion in Chinese, *Comput. Eng.*, **48** (2022), 240–246. https://doi.org/10.19678/j.issn.1000-3428.0060912