sEMG Motion Intention Recognition Based on Wavelet Time-Frequency Spectrum and ConvLSTM

Yixin Liu, Daxin Xin, Jin Hua and Mengzhao Liu
College of Electronic Information Engineering, Xi’an Technological University, Xi’an 710021, China
Email: 974795316@qq.com

Abstract. The electromyographic signal (EMG) is a kind of bioelectrical signal, which can predict human motion intention through signal analysis. Multiple classification models are used to predict the motion intention. It has been found that the classification accuracy is closely related to the feature information extracted from the signal. Traditionally, features are designed manually through prior knowledge. In this paper, a kind of EMG signal classification method based on convolutional neural network and convolutional long-term memory network (CNN-ConvLSTM) is proposed. ConvLSTM, with its global feature extraction capability, is designed to extract the sEMG features of each channel, and due to its strong local feature extraction capability, CNN was designed to further extract the fused feature information and realize end-to-end classification. The experimental results show that this algorithm has better classification performance than the existing classification methods. The electromyographic signal (EMG) is a kind of bioelectrical signal to recognize human movement intention.

1. Introduction
The electromyographic signal applications, due to the electromyographic signal acquisition process is easily affected by noise, and electromyographic signal itself is a kind of weak electrical signals of the acquisition to the poor quality of the electromyographic signal of itself [1, 2], is difficult to identify, how to accurately extracted from the complex noise electromyographic signal, and complete the follow-up action recognition, become the research in this field.

Since sEMG signals are complex and non-stationary and random, it is difficult to obtain satisfactory classification and recognition effect simply by taking sEMG signals as the input of a classifier [3]. Therefore, feature extraction of signals [4, 5] is carried out to represent signal types. In this paper, the sEMG signal is extracted by two-dimensional time-frequency graph transformation and the sEMG signal is recognized by deep learning to judge the user’s motion intention.

2. Data Processing Method

2.1. Pretreatment
In the process of motion intention recognition, more biological information can be extracted by dividing the data into frames in order to achieve stable recognition. Sliding window segmentation is the most common and simple method. In this paper, the data length of one frame is intercepted to 256, and 128 is moved on the original data to make certain overlap between Windows, and then the data of the next frame is intercepted, as shown in figure 1. This approach is widely used because it increases the stability of the data per frame and is faster and simpler than simply dividing the number of non-overlapping
frames. But this way to the window length and overlap length is difficult to determine. For sEMG signals, window length is usually selected between 100 and 300ms, and window overlap length is half of window length.

In order to reduce spectral energy leakage, window function can be used to truncate. In this project, hanning window is used to add window processing to sEMG signals. As shown in the second figure of figure 2, hanning window smooths the discontinuous changes at the truncation point to reduce spectral leakage and realize time-frequency localization.

Figure 1. Frame 10 to 13 sEMG signal.  
Figure 2. Frame and window sEMG signal.

2.2. Deal with the Noise

In the process of data collection, there is a small shift between the muscle contraction and the electrode, which is called baseline drift, and a linear filter is needed to remove the baseline drift. Baseline drift noise is a low-frequency signal, and its energy is mainly around 0.1Hz. The processing results are shown in figure 3.

Figure 3. De-baselining drift.

Wavelet transform can analyze non-stationary signals such as electromyographic signals and decompose a signal into a series of wavelets, including both frequency information and time information. The average amplitude of the wavelet over the whole time range is 0.

A process description of denoising based on wavelet analysis is as follows:

1. To remove the baseline drift noise signal after 6 layer of wavelet decomposition, get a different layer of high frequency coefficient.
2. The unbiased likelihood estimation method to determine the breakdown threshold value of high frequency coefficients of every layer, then use soft threshold method was carried out on the high frequency coefficients quantization threshold processing, high frequency information.
3. Reconstruct the low-frequency coefficients of the 6th layer after wavelet decomposition and the
high-frequency information of each layer after thresholding to obtain sEMG signal after filtering the high-frequency noise.

In this paper, signals of one movement and one channel were selected for denoising comparison of the sEMG signals of the upper limbs. The comparison before and after denoising is shown in figure 4. It can be found that the de-noised signal has an obvious smoothing effect and retains the waveform form in sEMG signal, indicating that the wavelet analysis scheme adopted has a good filtering effect.

![Figure 4. Contrast of waveform before and after denoising.](image)

2.3. Time-Frequency Spectrum

In this paper, using continuous wavelet transform, the algorithm steps of generating time-frequency graph by sEMG transformation are as follows:

Step 1: Set $a$ as the scale (stretching factor), $f_s$ as the sampling frequency, and $F_c$ as the wavelet center frequency, then the actual frequency of $a$ is:

$$F_a = F_c \times f_s / a$$

(1)

Step 2: According to equation (1), in order to the converted frequency sequence to be an isometric sequence, the scale sequence must take the following form:

$$c/\text{totalscal}, \ldots, c(\text{totalscal} - 1), c/4, c/2, c$$

(2)

where: $\text{totalscal}$ is the length of scale sequence used in wavelet transform of signal (preset to 256 in this paper), and C is a constant.

Step 3: It can be seen from equation (1) that the actual frequency corresponding to scale $c/\text{totalscal}$ should be $f_s/2$, so it can be obtained:

$$c = 2 \times F_c \times \text{totalscal}$$

(3)

Substitute equation (3) into equation (2) to obtain the required scale sequence $t$.

Step 4: After determining the wavelet base and scale, the wavelet coefficient $W_f(a,b)$ is obtained by using the principle of continuous wavelet transform. Then the scale sequence is converted into the actual frequency sequence $F$ according to the principle of equation (1). Finally, combining with the time series $t$, the wavelet time-frequency spectrum can be drawn to obtain the characteristic information.

The x-coordinate of the spectrum is time, the y-coordinate is frequency, and the coordinate point value is the myoelectric signal data energy, the energy value is expressed by color. The upper end of the color band indicates that the energy of the point is stronger, while the lower end of the color band indicates that the energy of the point is weaker [6, 7].

It can be seen from figures 5a and 5b that with the same force, different degrees of force are exerted on different muscles, and the force of the total extensor digitalis muscle is greater than that of the thumb metacarpal muscle. It can be seen from figures 5a and 5c that the total extensor of finger is more
powerful than the downward movement of palm. Therefore, the time-frequency diagram can be used to extract the time-frequency characteristics of upper limb movements based on sEMG signals.

![Time-Frequency spectrum](image)

**Figure 5.** Time-Frequency spectrum. (a) time-frequency spectrum of make a fist in extensor digitorum; (b) time-frequency spectrum of make a fist in the thumb of muscle; (c) time-frequency spectrum of Palm down in extensor digitorum; (d) time-frequency spectrum of Palm down in the thumb of muscle.

3. Network structure
Electromyographic signals are bioelectrical signals produced by muscle movements. Different muscles produce different electromyographic signals under the same movement. In this paper, eight-channel sEMG sensors are used to collect sEMG signals to obtain more biological information representing this action. Since there is a strong correlation between the sEMG signals of different channels and they represent the same action, the sEMG signals of different channels can be regarded as sequence data [8, 9].

The network identification flow chart is shown in figure 6. The network consisted of a CNN, three ConvLSTM layers, a full ConvLSTM layer, and a Softmax layer. After pretreatment, sEMG signals from different channels were sent to CNN for spatial feature extraction. The spatial features of the
extracted channels were fed into ConvLSTM as sequence data. In ConvLSTM layer, convolution kernel was introduced to extract spatial features, thus reducing the dimension of the structure. The spatial information of sEMG signals is extracted from different channels by convolution kernel sliding data. At the same time, the output of the previous cell and the input of the current cell are merged into the input of the current cell. The features extracted by convolution operation are transferred to the next layer. Thus, ConvLSTM could use not only the time series features of the sEMG signals, but also the convolution operation to extract the spatial features. In the last layer, a Softmax layer was introduced to classify actions as a classifier.

4. Experiment and Analysis

Eight-channel electromyographic signal acquisition equipment is used to collect surface electromyographic signal (sEMG). The sampling resolution is 10 bit and the sampling frequency was 500 Hz. Eight surface sticking electrodes are used for the acquisition electrode. Eight upper limb movements including extending palm, clenching, palm up, palm down, radial bending, ruler bending, arm up and arm down are collected, as shown in figure 7. The subjects are 5 healthy subjects aged 23-25 years. Each subject is asked to make approximately 3s for each gesture, followed by stretching five fingers to relax and rest for 3s. Each gesture intention is collected for 20 times, and the collected signals are stored in the form of a table. A total of 800 sets of data are obtained. The experimental data set is composed of eight-channel time-frequency sEMG atlas of eight movements. Eighty percent of the samples for each movement are training data, and twenty percent of the samples for each movement are test data.

The model in this paper is implemented on TensorFlow and Keras frameworks. Each sEMG signal is divided into 20 frames. The input space size is adjusted to 256×256. Convolution kernel sizes were 5×5×5, 3×3×3 and 1×1×1, and three ConvLSTM networks with 256 hidden units each were used. The model was trained using the stochastic gradient descent method, and the Objective loss function was optimized using the Adam network optimizer. The initial learning rate was 0.001, the momentum was set at 0.9, and the batch size was set at 32.
Table 1 compares the recognition effects of the eight actions. The experimental results show that the network can effectively extract the features of the eight sEMG time-frequency spectrums and complete the feature classification, without the need to manually design feature vectors and classifiers according to the characteristics of the actions to be recognized.

| Motion          | Accuracy |
|-----------------|----------|
| Extend palm     | 98.2%    |
| Make fist       | 97.3%    |
| Palm up         | 92%      |
| Palm down       | 93.2%    |
| Radial bend     | 96%      |
| Ruler bend      | 96.3%    |
| Arm up          | 87.3%    |
| Arm down        | 84.8%    |
| All motion      | 93.13%   |

Table 2 shows the recognition rates of different methods. The results show that the test results of this method in the existing data set are superior to other identification and classification methods. LSTM can only extract the temporal features but not the spatial features, thus obtaining the lowest recognition rate. CNN can extract the spatial details of myoelectric time-frequency spectrum, and its recognition rate is good. Since both convolution operation and feedback mechanism could be used to extract temporal features, ConvLSTM performed better than LSTM which could only extract temporal features and CNN which could only extract spatial features, with an recognition rate of 92.5%. ConvLSTM not only had the temporal modeling capability of LSTM, but also the ability to extract global features, so to speak, of space-time properties. On the basis of using ConvLSTM to extract global space-time features, finally using CNN to further extract local features, which can capture inter-frame spatial information and inter-frame time information at the same time. Experimental results show that the ConvLSTM and CNN combined method has better recognition performance than other methods.

| Methods        | Accuracy |
|----------------|----------|
| LSTM           | 90.8%    |
| CNN            | 91.22%   |
| ConvLSTM       | 92.5%    |
| Ours           | 93.13%   |

5. Conclusion
In this paper, eight kinds of motions were selected for sEMG signal collection. In combination with the characteristics of time-frequency spectra, the relevant theories and methods of convolutional neural network and convolutional short-time memory network were studied, and the sEMG signal ConvLSTM-CNN identification model was established to classify the eight motions. In the preprocessing, wavelet transform is used to denoise sEMG signals. Existing algorithms cannot combine the sEMG characteristics of different channels. LSTM can well extract the correlation characteristics of sEMG signals from different channels. Meanwhile, CNN can effectively extract local spatial features through convolution operation. Therefore, combining the advantages of CNN and LSTM, this paper introduced a ConvLSTM network which could simultaneously extract the space-time characteristics. The sEMG signals of the different channels were thought to be sequence inputs of ConvLSTM. Compared with the
four networks, the experimental results show that ConvLSTM combined with CNN algorithm is a feasible sEMG signal recognition method.

Acknowledgement
This project is supported by National Fund Youth Project (11804263).

References
[1] Lin K, Wu C, Huang X, et al. 2015 A robust gesture recognition algorithm based on surface EMG Proceeding of International Conference on Advanced Computational Intelligence (ICACI 2015) pp 131-136.
[2] Huang D and Chen B 2019 Surface EMG decoding for hand gestures based on spectrogram and CNN-LSTM 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI 2019) (IEEE) pp 123-126.
[3] Manfredo A, Arjan G, Claudio C, et al. 2014 Electromyography data for non-invasive naturally-controlled robotic hand prostheses Sci. Data 1 140053.
[4] Feng J, Wu X, Chen J, et al. 2019 Joint multilayer spatial-spectral classification of hyperspectral images based on CNN and ConvLSTM IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (IEEE) pp 588-591.
[5] Hu W S, Li H C, Pan L, et al. 2020 Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image Classification IEEE Transactions on Geoscience and Remote Sensing 58 (6) 4237-4250.
[6] PukhovaV, Gorelova E, Ferrini G, et al. 2017 Time-frequency representation of signals by wavelet transform Young Researchers in Electrical & Electronic Engineering (IEEE) pp 715-718.
[7] Luo K, Du K, Cai Z, et al. 2017 A modified frequency slice wavelet transform for physiological signal time-frequency analysis 2017 Chinese Automation Congress (CAC) pp 3441-3444.
[8] Chao H, He Y, Zhang J, et al. 2018 GaitSet: Regarding gait as a set for cross-view gait recognition arXiv:1811.06186v4.
[9] Chen W, Zhang L, Xi Y, et al. 2019 Polarimetric HRRP target recognition based on ConvLSTM IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (IEEE) pp 334-337.