Journal of Science and Engineering VOL. 1, NO. 1 (OCTOBER 2024) PP. 1-6

A Novel Gesture Recognition Network Based on LSTM

Lin Teng
College of Information and Communication Engineering
Harbin Engineering University
Harbin, China

Jiachi Wang
Department of Computer and Information Engineering
Youngsan University
Busan, South Korea

Editor: Mukesh Singhal

Abstract

Surface electromyography (sEMG) is a rich source of physiological data that reflects human movement intentions. The recognition of gestures through sEMG has garnered significant interest in the realms of human-computer interaction and rehabilitation. Currently, the majority of research on gesture recognition utilizing sEMG signals relies on discrete segmentation techniques, often overlooking the nuances of continuous and natural movements. A novel gesture recognition network based on LSTM for recognizing gestures from sEMG signals is prpopsed. The sEMG sensors are strategically positioned based on anatomical and muscular functions to optimize the capture of relevant physiological signals. In this study, finger curvature is employed to characterize gesture states, allowing each gesture at any given moment to be represented by a collection of varying finger curvatures, thereby enabling continuous gesture recognition. The experimental results indicate that this method significantly enhances the ability to mine representations from sEMG signals, offering valuable insights for deep learning models focused on human gesture recognition.

Keywords: surface electromyography (sEMG), gesture recognition, long short-term memory (LSTM) network

1 Introduction

As human-computer interaction technology continues to evolve and become widespread, gesture recognition, hailed for its natural and intuitive to interaction, has garnered increasing interest and focus. Within the domain of gesture recognition technologies, sEMG-based gesture recognition stands out for its practicality and broad range of potential applications (Chaves et al. (2024); Wang et al. (2024); Liu et al. (2024)). This method is capable of discerning muscle activities in various parts of the human body, including fingers, palms, and wrists, by capturing and analyzing surface electromyographic signals, thereby facilitating the identification and control of gestures.

In contrast to conventional gesture recognition methods that rely on visual or inertial sensors, technology that leverages sEMG for gesture recognition demonstrates superior resilience and consistency, enabling reliable and precise recognition of gestures across varying lighting and posture conditions. The application and proliferation of gesture recognition

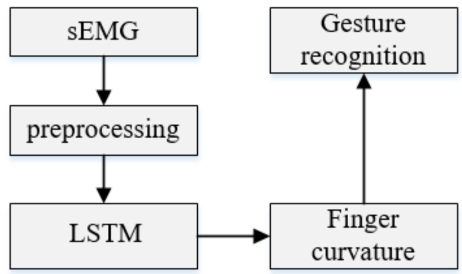

Figure 1: The comprehensive framework of the proposed approach

technology, underpinned by sEMG, have been notably evident in smart home technologies, advanced medical applications, and the control of prosthetic limbs. This approach has emerged as a significant area of exploration within the realm of human-computer interaction.

Nevertheless, the majority of research efforts have concentrated on the discrete classification of pre-defined gestures. Once signal patterns change, the recognition accuracy will fade. To address the limitation above, continuous motion estimation through sEMG is better suited to enhance practical application scenarios (Li et al. (2024)). Currently, the regression model of continuous motion estimation that leverages sEMG signals often results in redundant mappings between sEMG signals and their corresponding outputs.

To counter these challenges, deep learning based methodologies to accomplish continuous motion estimation are focused, which do not necessitate an intricate understanding of physiological mechanisms and are adept at capturing the spatial relationships within sEMG data. By targeting finger curvature, this study establishes the transformation from sEMG data to finger curvature using the LSTM network, thereby facilitating the continuous and accurate gesture recognition. The experimental outcomes confirm the validity of this approach.

2 Method

Figure 1 showcases the comprehensive framework of the proposed approach. Initially, the sEMG signals from the forearm are captured during the performance of a designated set of task gestures. Once the signals have undergone preliminary processing, they are fed into the LSTM network to derive the corresponding finger curvatures, ultimately yielding the continuous gesture data.

2.1 LSTM

The LSTM model is architected in two main segments. Initially, the input sEMG data undergoes feature extraction through the CNN layer, yielding a set of feature representations. Subsequently, these features are sequentially conveyed to the LSTM layer in accordance with their temporal order. The LSTM integrates the output from the preceding CNN layer with the current input, synthesizing a coherent time series. Such a design enhances the

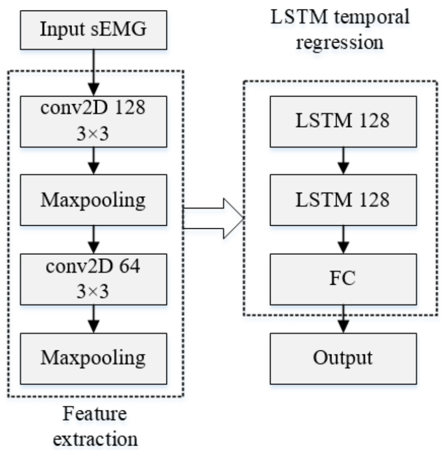

Figure 2: Overall architecture of LSTM model

LSTM's capacity to manage time series data processing tasks effectively. Figure 2 illustrates the overall architecture of the LSTM model.

Initially, the sEMG data is segmented into multiple parts using a sliding window technique, with each segment being reshaped into an $L \times C$ format for input into the CNN. Here, the input dimensions are 20×5, where the number 20 signifies a 200ms active window width, and 5 corresponds to the five channels of sEMG data acquisition. The CNN architecture comprises two convolutional layers, two pooling layers, and two fully connected layers. The first and second convolutional layers include 128 and 64 3×3 filters, respectively. Both utilize ReLU as their activation function. The first and second pooling layers perform 2×2 max pooling. Following the initial pooling layer, a Dropout layer with a parameter of 0.5 is incorporated to mitigate the risk of model overfitting. The addition of a Batch Normalization layer subsequent to each convolutional layer serves to normalize the layer outputs, thus enhancing the model's convergence rate and overall generalization capabilities. The output features of the CNN are represented as the vector $v = [v_1, v_2, \ldots, v_n]$.

The output vector from the CNN layer serves as the input for the LSTM layer, which is responsible for predicting the finger curvature. The fundamental unit of an LSTM layer is composed of a forget gate, an input gate, and an output gate. These gates enable the LSTM to more efficiently decide which information to discard and which to preserve. Given v and h, the update process of the LSTM unit at time step t can be articulated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, v_t] + b_f). \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, v_t] + b_i). \tag{2}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, v_t] + b_c). \tag{3}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t. \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, v_t] + b_o). \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t). \tag{6}$$

where f_t denotes the forget gate, i_t signifies the input gate, o_t indicates the output gate, c_t represents the cell state, σ is the sigmoid activation function, tanh refers to the hyperbolic tangent activation function, and W along with b stand for the weights and biases, respectively. The LSTM network comprises two layers of LSTM units followed by a fully connected layer.

2.2 Finger Curvature

Fingers are pivotal in defining gestures, suggesting that gesture recognition can be effectively reduced to the recognition of finger states. The movement of the fingers can be essentially viewed as the motion occurring at the metacarpophalangeal joints. These joints serve as the primary control points for finger movement, with the interphalangeal joints bending and extending around them, a concept captured by the term "finger curvature". The finger curvature for each finger, denoted as C_i where $i = 1, 2, 3, 4, 5$, represents the individual curvature of the five fingers respectively, with C_i ranging within the interval $[0, 1]$. A gesture G is thus represented by the collective curvatures of the fingers, expressed as $G = [C_1, C_2, C_3, C_4, C_5]$. This set of values encapsulates the gesture through the curvatures of each individual finger. The output of the LSTM network is reshaped to 5-dimensional vector and its elements are normalized to the interval $[0, 1]$, as the predicted finger curvature.

3 Experiment

3.1 Settings

The experiment data set includes 183560 samples of 52 kinds of actions, and each gesture has the same number of samples. The accuracy (Acc) index is utilized to evaluate the recognition of gestures. The comparison methods in the experiment include existing researches as follows: Compact CNN (Chen et al. (2020)), MV-CNN (Wei et al. (2019)), CNN+LSTM (Khushaba et al. (2020)), CNN+CviT (Shen et al. (2022)), and CNN+transfer learning (CNN+TL) (Côté-Allard et al. (2019)).

3.2 Results

As shown in Table 1, although the above methods have achieved good results, the model proposed in this paper can provide more accurate gesture recognition results than them.

4 Conclusion

This study introduces a gesture recognition approach that leverages an LSTM network, with a CNN for extracting spatiotemporal features, achieving state-of-the-art gesture recognition classification accuracy on a vast dataset. Subsequent research will aim to minimize the quantity of sEMG sensors used without compromising the model's efficacy, as the reduction

Table 1: Gesture recognition results of different models

Method	Window size(ms)	Acc(%)
Compact CNN	200	70.10
MV-CNN	200	75.90
CNN+LSTM	260	84.57
CNN+CviT	200	76.93
CNN+TL	260	69.08
Proposed	200	88.50

of sensors can lead to a loss of signal information. Thus, accurately identifying gestures with a limited number of sEMG sensors continues to pose a significant challenge. Moreover, enhancing the model's robustness is also a complex issue that needs to be addressed. It is proposed that future work could involve altering the dataset to reflect divisions among different individuals rather than repetitive actions within the same subject. Additionally, the incorporation of generative adversarial networks may be explored to further augment the data and improve the model's performance.

Acknowledgments and Disclosure of Funding

The authors declare that there is no conflict of interest and no funding.

References

Gabriel S. Chaves, Anderson S. Vieira, and Markus V. S. Lima. semg-based gesture classifier through DTW and enhanced muscle activity detection. *IEEE Access*, 12:117595–117607, 2024.

Lin Chen, Jianting Fu, Yuheng Wu, Haochen Li, and Bin Zheng. Hand gesture recognition using compact CNN via surface electromyography signals. *Sensors*, 20(3):672, 2020.

Ulysse Côté-Allard, Cheikh Latyr Fall, Alexandre Drouin, Alexandre Campeau-Lecours, Clément Gosselin, Kyrre Glette, François Laviolette, and Benoit Gosselin. Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 27(4):760–771, 2019.

Rami N. Khushaba, Angkoon Phinyomark, Ali H. Al-Timemy, and Erik J. Scheme. Recursive multi-signal temporal fusions with attention mechanism improves EMG feature extraction. *IEEE Transactions on Artificial Intelligence*, 1(2):139–150, 2020.

Jun Li, Chunlong Zou, Dalai Tang, Ying Sun, Hanwen Fan, Boao Li, and Xinjie Tang. Multi-stream fusion network for continuous gesture recognition based on semg. *International Journal of Wireless and Mobile Computing*, 26(4):374–383, 2024.

Xiaoguang Liu, Lijian Hu, Liang Tie, Li Jun, Xiaodong Wang, and Xiuling Liu. Integration of convolutional neural network and vision transformer for gesture recognition using semg. *Biomedical Signal Processing and Control*, 98:106686, 2024.

Shu Shen, Xuebin Wang, Fan Mao, Lijuan Sun, and Minghui Gu. Movements classification through semg with convolutional vision transformer and stacking ensemble learning. *IEEE Sensors Journal*, 22(13):13318–13325, 2022.

Shurun Wang, Hao Tang, Feng Chen, Qi Tan, and Qi Jiang. Integrated block-wise neural network with auto-learning search framework for finger gesture recognition using semg signals. *Artificial Intelligence in Medicine*, 149:102777, 2024.

Wentao Wei, Qingfeng Dai, Yongkang Wong, Yu Hu, Mohan S. Kankanhalli, and Weidong Geng. Surface-electromyography-based gesture recognition by multi-view deep learning. *IEEE Transactions on Biomedical Engineering*, 66(10):2964–2973, 2019.