

GTGR-Net: Graph Attentional-Temporal Network for Surface-Electromyography-Based Gesture Recognition

Xiaoxu Jia¹, Hongbo Wang^{12345*}, Jingjing Luo¹, Zhiping Lai¹, Xueze Zhang¹, Weiqi Zhang¹, Xiuhong Tang¹

¹ Academy for Engineering & Technology, Fudan University, Shanghai 200433, China

² Parallel Robot and Mechatronic System Laboratory of Hebei Province, Yanshan University, Qinhuangdao, China

³ Intelligent Robot Engineering Research Center of Ministry of Education, Shanghai 200433, China

⁴ Shanghai Intelligent Robot Engineering Technology Research Center, Shanghai 200433, China

⁵ Shanghai Clinical Research Center for Aging and Medicine, Shanghai 200040, China

*Corresponding Author: Hongbo Wang. E-mail: Wanghongbo@fudan.edu.cn, 20210860113@fudan.edu.cn.

Abstract- In the process of active rehabilitation assisted by hand rehabilitation robot, the patient's hand motion intention, that is, the patient's gesture recognition, plays an important role. Gesture recognition based on sEMG signal is a hot research topic. Due to the spatial correlation and temporal non-stationary of sEMG signal, this research topic has many difficulties. To solve this problem, we propose a gesture recognition network GTGR-Net based on sEMG signal, which uses the combination of graph attention network and temporal convolutional network to extract the spatial and temporal information of sEMG signal. We verify the effect of our algorithm on three public data sets and achieve good results, which is better than other methods.

Keywords- Graph attention network, temporal convolutional network, sEMG, gesture recognition, graph structure.

I. INTRODUCTION

The hand rehabilitation robot needs to recognize the signals related to the rehabilitation movement of the patient, to accurately control the rehabilitation robot to take relevant actions to help the patient carry out active rehabilitation training. There are two kinds of signals based on vision and sensor as signals related to rehabilitation movement of patients [1]. However, it is very difficult to recognize hand motion intention by visual method, and all kinds of cameras are vulnerable to environmental factors such as illumination. In the method of passing through the sensor, the method based on sEMG signal has increasingly become a research hotspot because its acquisition equipment has the characteristics of non-invasive and easy disassembly, and sEMG signal acquisition is convenient and less affected by environmental factors.

Gesture recognition methods based on sEMG signals are mainly divided into traditional machine learning methods and deep learning methods. Because the former requires human experience for feature extraction, but human experience has limitations, it is easy to ignore the potential information in sEMG signals, while the deep learning method is an end-to-end method without human experience, so it is easier and more comprehensive to extract the information in sEMG signals. Graph neural network is a model of the neural network established on the graph. After it is proposed, deep learning methods have attracted more and more attention [2]. sEMG signals are superimposed by

the action potentials of muscle motor units. There will be a correlation between sEMG signals collected at sampling points in different arm parts. A graphical neural network can better reflect this relationship than other neural networks. By introducing the attention mechanism into the graph neural network and redefining the aggregation function of graph convolution, a graph attention network is obtained, which can better reflect the relationship between various nodes in the graph network [3]. In addition, the time information of sEMG signal, as a time-series signal, is also worthy of attention. Most studies only focus on spatial information or time information without considering them together, such as [4-5]. Therefore, combined with the actual situation, in the research of gesture recognition method based on sEMG signal, we propose graph attentional-temporal network for gesture recognition, which is called GTGR-net. We use the combination of graph attention network and temporal convolutional network to extract the spatial and temporal information of sEMG signal.

The main contributions of our work can be summarized as follows:

- We propose a neural network GTGR-net, which uses a graph attention network and temporal convolutional network to extract the spatial and temporal information of gesture recognition.
- To better obtain the high-level semantic information in sEMG signals, we use the GTGR module based on the combination of graph attention network and temporal convolutional network and then use multiple GTGR modules to extract the multi-level features of sEMG signals alternately in spatial and temporal dimensions.
- We use three public data sets to verify the effect of our algorithm and achieved good results.

II. RELATED WORK

Among the gesture recognition methods based on sEMG signals, the deep learning method has more research potential than traditional machine learning methods because it does not need human experience to extract features. The convolution neural network method and cyclic neural network method are the main methods of deep learning. Many algorithms are improved on their basis. What they have in common is that they describe each frame of sEMG data as an image or vector sequence to extract spatial and

temporal features [6-8]. Although these methods have achieved good results, they ignore the correlation between the sEMG signals collected at the sampling points in different parts of the arm and the intimate relationship between these correlations.

Graph neural networks can be classified into two types: spectral-domain and spatial domain. Graph attention network is the main representative of the airspace category. Graph attention network is a new neural network architecture that can operate the graph data structure. As a kind of graph neural network model, graph attention network, compared with graph convolution network, aggregates the information of neighbor vertices by using adjacency matrix to update the state. It redefines the aggregation function by introducing an attention mechanism, and the calculation of association weight depends on the feature expression of nodes [3]. An attention model is proposed by referring to the human attention mechanism, and then different weights are assigned to the different contributions of different parts of the data. It can deal with variable-size inputs and focus attention on the most relevant parts of the input data.

III. METHOD

When making gestures, the muscles on the arm generate action potentials, and the sensors on the arm sense collect these action potentials [9]. Usually, a gesture contains a series of frames about sEMG signals, so it has spatial and temporal characteristics. In the action recognition task based on the human skeleton, the spatial and temporal features of human skeleton action can be extracted by combining graph convolution network and temporal convolutional network [10-11]. Therefore, referring to this method, graph convolution network and temporal convolutional network are introduced into gesture recognition based on sEMG signal [12]. Inspired by them, we introduce graph attention networks and temporal convolutional networks. By introducing the attention mechanism into the graph neural network, we distribute different weights according to the contribution of different parts of the data. Different from graph convolution network, it just aggregates the information of neighbor vertices through adjacency matrix. Among them, the temporal convolutional network is a special convolution neural network for time series [13].

In this part, we introduce our proposed network architecture. The network combines graph attention network and temporal convolutional network to build GTGR module and then uses multiple GTGR modules to alternately extract multi-level features of sEMG signals in spatial and temporal dimensions, to obtain high-level semantic information that can be used for gesture recognition, and train a *softmax* classifier to realize gesture classification. Because the combination of graph attention network and temporal convolutional network can better extract spatial and temporal information, our network architecture has better results.

A. GTGR module

We designed a GTGR module, which is used to extract spatial and temporal information from sEMG, as shown in

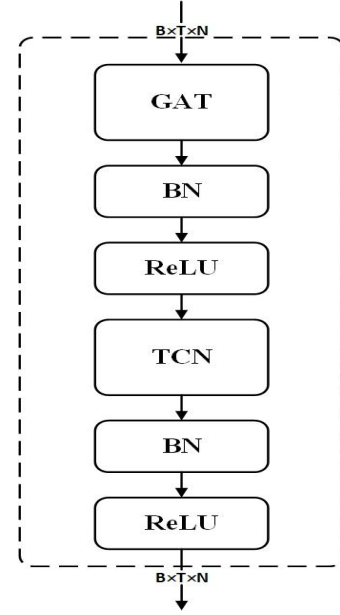


Figure 1. GTGR module

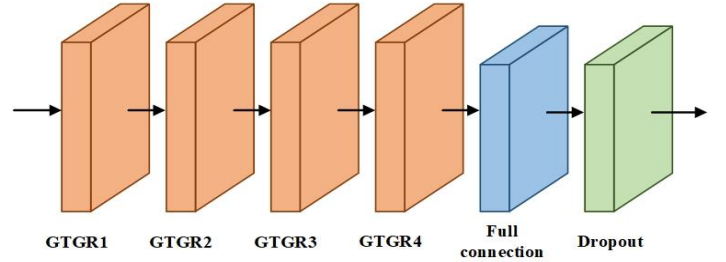


Figure 2. GTGR-net

Figure 1. After data input, the spatial information of the data will be extracted through a graph attention network layer, and then regularized and activated through a batch normalization layer and rectified linear units, which can speed up the training process and enhance the representation ability of the network. After passing through the neural network layer of graph attention, each vertex of sEMG signal data of each frame has undergone a more comprehensive feature fusion in the spatial dimension. Then the data will pass through a temporal convolutional module so that the time feature can be extracted in the time dimension.

B. GTGR-Net

The overall architecture of GTGR-net is shown in Figure 2. The overall network architecture can be composed of any GTGR module. In our work, according to the experimental results, we selected four GTGR modules to form our overall network architecture and connected a full connection layer behind the last GTGR module to flatten our features, and then input them to the Dropout layer. Our final output feature is obtained by a *softmax* classifier.

IV. EXPERIMENTS

A. Data sets and Experimental configurations

The first data set comes from the sEMG data set published by Zhejiang University. This data set is a high-density sEMG data set [14]. In the data collection process, a total of 23 healthy subjects were recruited for testing. The electrode array used was 8x16. The interval density of the electrodes was relatively large, belonging to high-density surface electromyography. The collected gesture set belongs to a subset of nine pro databases. Capgmyo database has three sub-datasets, namely DB-a, DB-b and DB-c. DB-a and DB-b contain eight gestures, and our experiment uses these two data sets.

Bandmyo is a sparse sEMG data set [15]. In the data collection process, a total of 6 healthy subjects were recruited for testing, including 4 men and 2 women. The collection equipment used MYO armbands with 8 channels for collection, and the collected gestures included 15.

Our training model uses PyTorch 1.8.1, the optimizer used in the model is Adam optimizer, and the learning rate is 0.01. Our loss function uses the cross-entropy loss function. Each time we use the sliding window strategy to get a set of electromyographic signals with a fixed step size.

B. Results and Comparison

We will compare with other work on three data sets. The comparison of experimental results is shown in Table I. The results of other comparison methods come from various literature. From the comparison of experimental results, it can be seen that the experimental results of our algorithm on these three data sets are better than other methods. Even if the results are the same, our network is simpler and the amount of parameters is less.

On the three data sets of our experiment, compared with the method with the best experimental results before, our method has the same accuracy on DB-a and DB-b data set, but our network is simpler and has fewer parameters. For the BandMyo data set, our accuracy is improved by 3.5%.

TABLE I. THE EXPERIMENTAL RESULTS OF OUR METHOD ARE COMPARED WITH THOSE OF OTHER METHODS ON THREE PUBLIC DATA SETS. THE EXPERIMENTAL RESULTS OF OTHER METHODS ARE FROM THEIR PAPERS.

	Accuracy (%)		
	CapgMyo DB-a	CapgMyo DB-b	BandMyo
GengNet [16]	99.5	98.6	57.8
STCN-GR [12]	99.8	99.4	75.8
SVM [17]	71.0	70.8	59.4
RF [17]	83.2	76.2	68.1
STF-GR [17]	91.7	90.3	71.7
GTGR-Net (ours)	99.8	99.4	79.3

C. Ablation study

Our network is composed of multiple GTGR modules, and each GTGR module is composed of a graph attention network and temporal convolutional network. Therefore, our ablation experiment needs to study the number of GTGR

modules and the effect of graph attention network and temporal convolutional network on the overall performance. Our experimental results are shown in Figure 3.

From the analysis of our experimental results, it can be concluded that the experimental results of our model graph attention network and the model temporal convolutional network have more advantages compared with the DB-a data set and DB-b data set. We can conclude that compared with feature extraction in the time dimension, feature extraction in the space dimension can bring more gain to gesture recognition tasks. With the increase in the number of GTGR modules, we can see that feature extraction is more and more sufficient, and the effect of gesture recognition is improving. When we increase the number of GTGR modules to 4, our results on these three data sets reach the best. If we increase the number of GTGR modules, there will be overfitting.

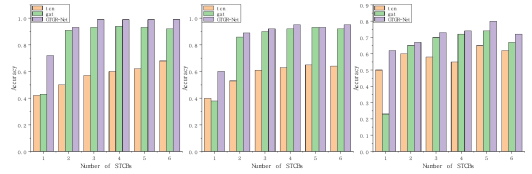


Figure 3. Ablation experiments on three data sets. The “GTGR-Net” is the proposed complete model, the “tcn” and the “gat” are the models that remove GAT and TCN.

V. CONCLUSION

In this paper, we propose GTGR-net for gesture recognition based on the sEMG signal. We introduce the use graph attention network, which redefines the aggregation function by introducing the attention mechanism. The calculation of the association weight depends on the feature expression of the node and then assigns different weights to the different contributions of different parts of the data. It can deal with variable-size inputs and focus on the most relevant parts of the input data. In this way, the connection weight between our different sEMG signal acquisition channels changes dynamically. The data through the graph attention network is input to the sequential convolution neural network for time feature extraction. Through the experimental results, we can find that our model has certain advantages in accuracy, network architecture and parameter quantity.

ACKNOWLEDGMENT

This work was supported in part by the National Key R&D Program of China under Grant 2019YFB1312500, 2021YFC0122702, and 2019YFB1311701, in part by National Natural Science Foundation of China under Grant

U1913216, in part by Shanghai 2019 Science and Technology Innovation Action Plan under Grant 19441908200, and in part by Fudan-ZiMaoYiHao Medical Device Joint Experimental Center Project under Grant SGH2310020/007.

REFERENCES

- [1] CHEOK M J, OMAR Z, JAWARD M H. A review of hand gesture and sign language recognition techniques[J]. *International Journal of Machine Learning and Cybernetics*, 2019, 10(1):131-153.
- [2] GORI M, MONFARDINI G, SCARSELLI F. A new model for learning in graph domains[C]//*Proceedings. 2005 IEEE International Joint Conference on Neural Networks*, 2005.: volume 2. [S. l.]: IEEE, 2005: 729-734.
- [3] Velikovi P, Cucurull G, Casanova A, et al. Graph Attention Networks[J]. 2017.
- [4] Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., Li, J.: Gesture recognition by instantaneous surface emg images. *Scientific reports* 6(1), 1–8 (2016).
- [5] Ketyk'o, I., Kovacs, F., Varga, K.Z.: Domain adaptation for semg-based gesture recognition with recurrent neural networks. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. pp. 1–7. IEEE (2019)
- [6] Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., Li, J.: Gesture recognition by instantaneous surface emg images. *Scientific reports* 6(1), 1–8 (2016)
- [7] Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., Geng, W.: A novel attention-based hybrid cnn-rnn architecture for semg-based gesture recognition. *PloS one* 13(10), e0206049 (2018)
- [8] Tsinganos, P., Cornelis, B., Cornelis, J., Jansen, B., Skodras, A.: Improved gesture recognition based on semg signals and tcn. In: *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 1169–1173. IEEE (2019)
- [9] Rahimian, E., Zabihi, S., Atashzar, S.F., Asif, A., Mohammadi, A.: Xceptiontime: independent time-window xceptiontime architecture for hand gesture classification. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 1304–1308. IEEE (2020)
- [10] Shi, L., Zhang, Y., Cheng, J., Lu, H.: Two-stream adaptive graph convolutional networks for skeleton-based action recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 12026–12035 (2019)
- [11] Yan, S., Xiong, Y., Lin, D.: Spatial temporal graph convolutional networks for skeleton-based action recognition. In: *Proceedings of the AAAI conference on artificial intelligence*. vol. 32 (2018)
- [12] Lai Z. et al. (2021) STCN-GR: Spatial-Temporal Convolutional Networks for Surface-Electromyography-Based Gesture Recognition. In: Mantoro T., Lee M., Ayu M.A., Wong K.W., Hidayanto A.N. (eds) *Neural Information Processing. ICONIP 2021*.
- [13] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. *arXiv preprint arXiv:1803.01271*, 2018.
- [14] DU Y, JIN W, WEI W, et al. Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation[J]. *Sensors*, 2017, 17(3):458.
- [15] ZHANG Y, CHEN Y, YU H, et al. Learning effective spatial-temporal features for semg armband-based gesture recognition[J]. *IEEE Internet of Things Journal*, 2020, 7(8):6979-6992.
- [16] Geng, W., Du, Y., Jin, W., Wei, W., Hu, Y., Li, J.: Gesture recognition by instantaneous surface emg images. *Scientific reports* 6(1), 1–8 (2016)
- [17] Zhang, Y., Chen, Y., Yu, H., Yang, X., Lu, W.: Learning effective spatial-temporal features for semg armband-based gesture recognition. *IEEE Internet of Things Journal* 7(8), 6979–6992 (2020)