Cross-User Electromyography Pattern Recognition Based on a Novel Spatial-Temporal Graph Convolutional Network

Mengjuan Xu, Xiang Chen, *Member, IEEE*, Yuwen Ruan, and Xu Zhang, *Member, IEEE*

***Abstract*— With the goal of promoting the development of myoelectric control technology, this paper focuses on exploring graph neural network (GNN) based robust electromyography (EMG) pattern recognition solutions. Given that high-density surface EMG (HD-sEMG) signal contains rich temporal and spatial information, the multi-view spatial-temporal graph convolutional network (MSTGCN)is adopted as the basic classifier, and a feature extraction convolutional neural network (CNN) module is designed and integrated into MSTGCN to generate a new model called CNN-MSTGCN. The EMG pattern recognition experiments are conducted on HD-sEMG data of 17 gestures from 11 subjects. The ablation experiments show that each functional module of the proposed CNN-MSTGCN network has played a more or less positive role in improving the performance of EMG pattern recognition. The user-independent recognition experiments and the transfer learning-based cross-user recognition experiments verify the advantages of the proposed CNN-MSTGCN network in improving recognition rate and reducing user training burden. In the user-independent recognition experiments, CNN-MSTGCN achieves the recognition rate of 68%, which is significantly better than those obtained by residual network-50 (ResNet50, 47.5%, $p < 0.001$) and long-short-term-memory (LSTM, 57.1%, p=0.045). In the transfer learning-based cross-user recognition experiments, TL-CMSTGCN achieves an impressive recognition rate of 92.3%, which is significantly superior to both TL-ResNet50 (84.6%, $p = 0.003$) and TL-LSTM (85.3%, $p = 0.008$). The research results of this paper indicate that GNN has certain advantages in overcoming the impact of individual differences, and can be used to provide possible solutions for achieving robust EMG pattern recognition technology.**

***Index Terms*— EMG, gesture recognition, graph neural network, transfer learning, user-independent.**

Manuscript received 4 September 2023; revised 28 November 2023; accepted 6 December 2023. Date of publication 12 December 2023; date of current version 12 January 2024. This work was supported by the National Natural Science Foundation of China under Grant 82272113 and Grant 61871360. *(Corresponding author: Xiang Chen.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Review Committee of the First Affiliated Hospital of Anhui Medical University under Application No. PJ 2014-08-04.

The authors are with the Department of Electronic Science and Technology, University of Science and Technology of China (USTC), Hefei 230027, China (e-mail: xch@ustc.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3342050

I. Introduction

ELECTROMYOGRAPHY(EMG) pattern recognition technology that recognizes EMG signals as meaningful action patterns or intentions, is crucial for the application of myoelectric control systems in fields such as prosthesis [1], rehabilitation engineering [2], human-computer interaction [3]etc. At present, EMG pattern recognition in user-specific mode (training and testing data from the same user) has shown significant progress. For gesture recognition tasks of 4-12 gestures, the recognition rates can reach 84.4% to 98.81% [4], [5], [6], [7], [8]. However, due to the need to train specific classifier for each user, user-specific mode will bring a heavy user training burden in practical applications. Therefore, robust EMG pattern recognition technology that works in user-independent mode has become a research hotspot in this field.

The goal of user-independent EMG pattern recognition research is to design universal classifiers for the application of myoelectric control systems. Unfortunately, due to the individual differences in anatomical structure (such as the relative position and spatial distribution of muscles), even for the same gesture, different users may have different patterns, which makes it difficult to ensure high user-independent recognition rate [9], [10], [11]. In fact, the obtained recognition rate of user-independent EMG pattern recognition is generally low in relevant researches. For instance, Matsubara et al. only achieved a 54% recognition rate on a 5-gesture recognition task by using support vector machine (SVM) [9]. Ketykó et al. proposed a 2-Stage Recurrent Neural Network (2SRNN) approach that achieved a recognition rate of 35.1% across 12-gesture recognition task [10]. The recognition rate of 65.03% was achieved on a 11-gesture recognition task by Côté-Allard et al. using a convolutional network (ConvNet) [11].

In recent years, scholars have made a lot of efforts and proposed a series of solutions based on domain generalization to improve the performance of user-independent EMG pattern recognition. In the domain generalization methods, common EMG features and patterns learned from the data of multiple users are used to improve user-independent recognition rate. Côté-Allard et al. introduced a multi-domain

learning algorithm named ADANN (Adaptive Domain Adversarial Neural Network) that enhanced the recognition rate of 11 gestures to 84.43% [11]. Li et al. proposed a mix-up and adversarial training for domain generalization (MAT-DG) framework and achieved a recognition rate of 91.85% for a 7-gesture recognition task [12]. Zhang et al. proposed a multi-source synchronized domain generalization (MS-DG) model and achieved a recognition rate of 73.04% for a 6-gesture recognition task [13].

In the meanwhile, as a compromise solution, some researchers have attempted to conduct cross-user EMG pattern recognition research based on transfer learning or domain adaptation [10], [14], [15], [16], [17]. The idea of transfer learning technology or domain adaptation technology is to extract knowledge from related fields or from the same fields that have been pre-trained to assist the training and classification of target tasks [18]. Although they introduced a little training burden for new users, the recognition rate can usually be improved even more. In the aforementioned study of Li et al. [12] and Zhang et al. [13], the recognition rate of the domain adaptation method MAT-DA and MS-DA reached 93.54% and 87.93% respectively, both of which were higher than those of the domain generalization methods (91.85% for MAT-DG and 73.04% for MS-DG). Wang et al. proposed a novel approach by integrating a transfer learning strategy into recurrent convolutional neural networks (RCNNs), improving the recognition rate of 49 gestures from 73.76% to 87.36% [14]. Campbell et al. presented an approach called adaptive domain adversarial neural network for cross-user myoelectric control, increasing 10-gesture recognition rate from 86.8% to 96.2% [17].

In terms of algorithm, both traditional machine learning methods [19], [20] and deep learning methods [5], [21] have been successfully applied to user-specific EMG pattern recognition. Specifically, it has been verified that deep learning models such as convolutional neural network (CNN) and recurrent neural network (RNN) can achieve higher accuracy of gesture recognition than traditional machine learning models. However, even deep learning models such as CNN and RNN failed to achieve satisfactory recognition performance in user-independent mode [13], [14]. Physiological studies have shown that regardless of the significant differences in muscle spatial distribution between individuals, the same gesture movements are controlled by the central nervous system issuing the same motor intention commands [22]. In other words, for the same gesture, regardless of whether there are anatomical differences in muscle distribution among different users, there are similar functional connections between the relevant muscles. If this functional connection between relevant muscles can be effectively captured, it may provide a solution for achieving high-accuracy user-independent EMG pattern recognition. However, RNN is a sequence model with poor ability to capture spatial information, and CNN primarily focus on local correlations between adjacent pixels, which makes it challenging for them to capture cross-muscle correlations.

Graph neural network (GNN) is a type of neural network architecture specifically designed to handle data represented as graphs, where nodes and edges represent entities and their relationships, respectively [23]. GNN has been successfully applied in many fields such as banking and finance [24], traffic management [25], bioinformatics [26] and computer vision [27] etc. In the aspect of electrophysiological signals, Jia et al. proposed a multi-view spatial-temporal graph convolutional network (MSTGCN) for sleep staging task based on electroencephalography (EEG) [28]. In order to better suit the characteristics of EEG signals, the EEG channels were mapped to nodes and the connections between channels are mapped to edges. By considering the functional connectivity of distinct brain regions, MSTGCN attained an impressive recognition rate of 89.5% in the 5-class sleep staging task. As mentioned earlier, different individuals may have similar functional connections in the muscles involved in performing the same gesture, so we speculate that GNN may have good application potential in user-independent EMG pattern recognition. In the context of EMG pattern recognition, GNN can map muscle channels to nodes in the graph representation and map their relationships to edges. In this way, it can learn the functional connection between non adjacent but related muscles, so as to overcome the influence of anatomical structure differences between individuals, and provide a possible solution for robust user-independent gesture recognition.

In fact, GNN has been preliminarily applied in EMG pattern recognition [29], [30], [31], [32], and has achieved satisfactory recognition rates in user-specific mode. Lai et al. proposed a spatial-temporal convolutional network (STCN-GR) based on EMG graph, achieving an average recognition rate of 99.8% for 8 gestures [29]. Lee et al. designed a self-attention-based graph neural network and the average recognition rate of 18 gestures was about 97% [30]. Yang et al. proposed a multi-modal fusion strategy for EEG and surface EMG based on graph theory, and obtained the average recognition rate of 93.86% for 4 gestures [31]. However, in the few studies conducted on more challenging user-independent EMG pattern recognition, the performance of GNN is not satisfactory. Zou et al. proposed an EMG graph based multi-label hand gesture recognition (MLHG) model. For 7 gestures, the average recognition rate was only 53.52% [32].

With the goal of promoting the development of myoelectric control technology, this paper focuses on exploring GNN-based EMG pattern recognition solutions. Specifically, its main innovations and contributions lie in: 1) Given that high-density surface EMG (HD-sEMG) signal contains rich spatial and temporal information, the multi-view spatial-temporal graph convolutional network MSTGCN [28] with spatial-temporal attention mechanism is introduced as the basic classifier. On this basis, a feature extraction CNN module is designed to improve feature representation ability of HD-sEMG signal and integrated into MSTGCN to generate a new model called CNN-MSTGCN; 2) Ablation experiments are conducted on 17 gestures in user-specific mode to explore the role of the functional modules including CNN, multi-view and attention mechanism modules in CNN-MSTGCN; 3) Through user-independent EMG pattern recognition experiments and transfer learning-based cross-user EMG pattern recognition experiments, it is verified that the proposed CNN-MSTGCN has certain advantages in improving

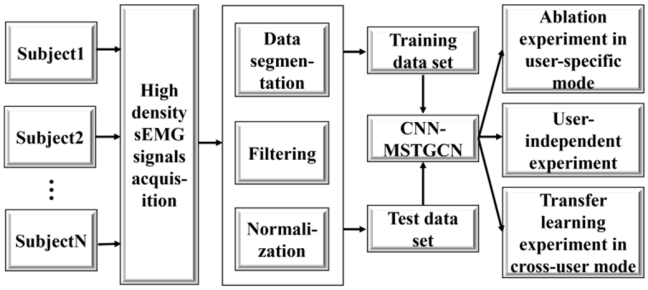

Fig. 1. Block diagram of proposed gesture recognition framework.

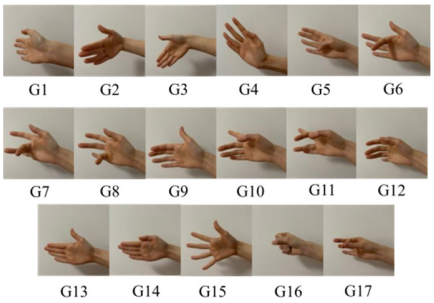

Fig. 2. Target gestures: wrist extension (G1), wrist flexion (G2), wrist ulnar deviation (G3), wrist radial deviation (G4), flexion of thumb (G5), flexion of index finger (G6), flexion of middle finger (G7), flexion of ring finger (G8), extension of thumb (G9), extension of index finger (G10), extension of middle finger (G11), extension of little finger (G12), thumb abduction (G13), thumb retraction (G14), spread of five fingers (G15), clenching of fist (G16) and thumb-index finger pinch (G17).

recognition rate and reducing user training burden, and can be used to provide possible solutions for achieving robust EMG pattern recognition technology.

II. Methods

As shown in Fig. 1, for a gesture recognition task, HD-sEMG signals are acquired from N subjects, and then are preprocessed and sent to the proposed CNN-MSTGCN network. Three kinds of gesture recognition experiments are carried out to verify the feasibility and superiority of the proposed CNN-MSTGCN in achieving robust EMG pattern recognition. The specific technical details and network design are elaborated in the following sections.

A. Gesture HD-sEMG Database

In this study, a gesture HD-sEMG database established by our team is targeted as the research object. This database consists of HD-sEMG data of C (C=17) gestures (as shown in Fig. 2) from N (N=11) healthy, non-disabled subjects. The subjects range in age from 22 to 25 years and are right-handed. All subjects have been informed of the experimental protocol before the experiments and signed the informed consent (No. PJ 2014-08-04) approved by the Ethics Review Committee of the First Affiliated Hospital of Anhui Medical University.

The signal acquisition device is composed of signal preprocessing hardware (as shown in Fig. 3(a)) and 4 high-density electrode arrays (as shown in Fig. 3(b)), with a total of M (M=128) electrode channels. Each electrode is 3.5 mm in diameter. Two of the electrode arrays, each consisting of

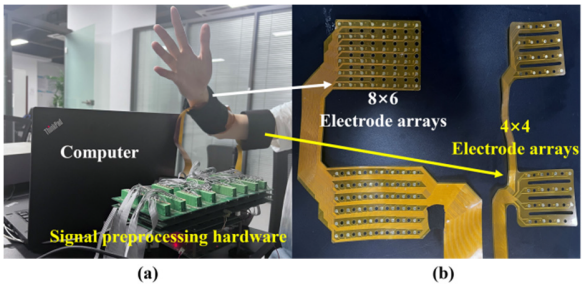

Fig. 3. HD-sEMG signal acquisition device.

48 electrodes in the shape of 8×6 and separated by 14 mm, collect HD-sEMG signals from the extensor and flexor muscles of the forearm respectively. The other two electrode arrays, each consisting of 16 electrodes in the shape of 4×4 and separated by 14 mm, collect HD-sEMG signals from the biceps and triceps respectively. Prior to placement of the electrode array, the target muscles are wiped with alcohol for disinfecting and conductive adhesives are applied to the electrodes to reduce impedance. When wearing the electrode array, ensure that its center line is aligned with the center line on both sides of the arm to minimize electrode array shift due to the difference in the thickness of subjects' arm. Finally, the electrode array is fixed with a flexible silicone sheet with strong adhesion and a telescopic brachial band to minimize the electrode array shift during gesture operation. The HD-sEMG signal is sampled at a rate of 1 kHz, amplified 1371.1 times by an amplifier, then filtered through a 20-500 Hz bandpass filter and converted into a digital signal by a 16-bit AD converter (ADS1198). The resulting digital signal is displayed on the computer screen in real time and stored on disk for subsequent off-line data processing and analysis.

In the process of data acquisition, the subjects sit on a chair of moderate height and relax their right forearm on the table in a neutral posture, that is, the rest state. When performing a gesture, it takes about 1s to move from the rest state to the gesture, about (1-3)s to hold the gesture, and about 1s to return to the rest state. Data acquisition experiment is conducted on 11 subjects respectively. Each subject performs 17 target gestures in sequence, each gesture is repeated 8 times, and 5s rest is taken between two repetitions to prevent muscle fatigue.

B. Signal Preprocessing

Since the acquired HD-sEMG signals include active segments and rest segments, the rest segments need to be removed because only the active segments are required for EMG pattern recognition. An amplitude thresholding approach [33] is adopted to remove the rest segments in the HD-sEMG signals. As a rule of thumb, the threshold is preset to be the mean plus three times the standard deviation of the resting signals averaged over all channels.

After obtaining an active segment, the sliding window method is used to segment it to obtain gesture samples. Due to the varying length of time that different subjects maintain their gestures, the (2-3)s of the active segment are uniformly selected to divide the samples. Since the sampling rate is 1 kHz, the 1s active segment contains 1000 timing points.

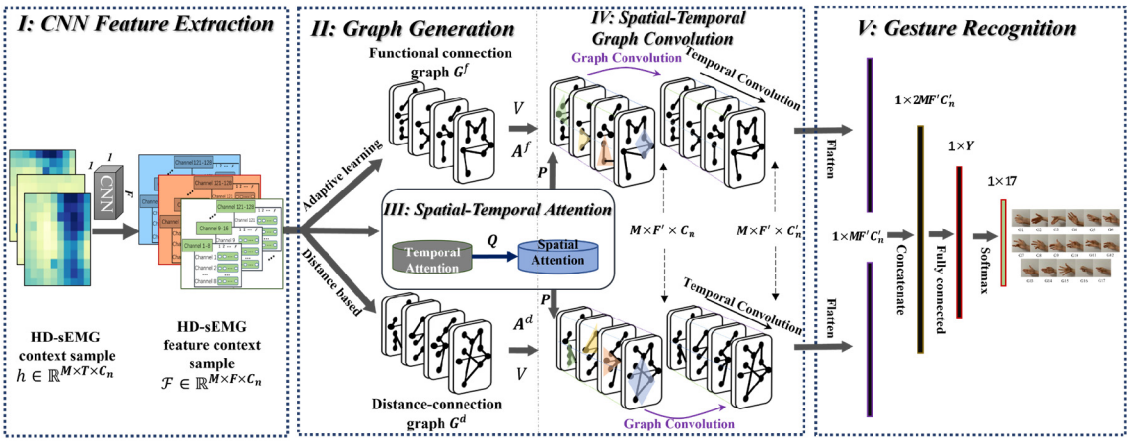

Fig. 4. The network architecture diagram of CNN-MSTGCN proposed in this study.

The window length T is set as 100 timing points, and the step length l is set as 50 timing points, so the shape of the independent sample generated is $M\times T$, that is 128×100. Since each gesture repetition can be divided into 19 samples, a total of $S=19\times 8(repetitions)\times 17(gestures)\times 11(subjects)=28424$ independent gesture samples are generated.

Considering that the obtained gesture samples still contain baseline drift, noise, motion artifact, etc., additional data preprocessing is performed. First, the mean value of each channel is subtracted to eliminate baseline drift; Second, data points whose absolute value is greater than the absolute mean value +8 times the standard deviation are removed as extreme values and replaced with the absolute mean value; Third, considering that the main energy of HD-sEMG signals are concentrated in the range of 20-150 Hz, a 50-order high-pass filter with a cutoff frequency of 20 Hz is adopted for removing motion artifacts; Fourth, the signal envelope is extracted through the rectification operation; Finally, signals of each time point of each sample are normalized by Min-Max Scaling method.

After all the above data preprocessing operations, a dataset $O = (o_1, o_2, \ldots, o_S) \in \mathbb{R}^{S\times M\times T}$ is constructed. Where $o_i = \left(x_1^i, x_2^i, \ldots, x_M^i\right) \in \mathbb{R}^{M\times T}$, $(i\in 1, 2, \cdots, S)$ is an independent sample. Then, to accommodate the input format required by graph neural networks, a new dataset $H = (h_1, h_2, \ldots, h_K) \in \mathbb{R}^{K\times M\times T\times C_n}$, called HD-sEMG context dataset, is created. The specific construction method is that the adjacent C_n samples in dataset O are combined into one sample in dataset H. Therefore, an independent sample of dataset H is $h_i = (o_{i-n}, \ldots, o_i, \ldots, o_{i+n}) \in \mathbb{R}^{M\times T\times C_n}$, $h_i \in H$ $(i\in 1, 2, \cdots, K)$. Where K is the total number of independent samples in dataset H, $C_n=2n+1$ represents the length of the HD-sEMG context sample, and $n\in N+$ denotes the time context coefficient. For fair comparison, dataset H is also used as input of comparison classifiers ResNet50 and LSTM.

C. EMG Pattern Recognition Algorithms

1) CNN-MSTGCN: CNN-MSTGCN proposed in this study is based on the state-of-the-art MSTGCN network proposed by Jia et al. [28]. The overall framework of CNN-MSTGCN, as illustrated in Fig. 4, consists of five modules, namely, CNN feature extraction module, graph generation module, spatial-temporal attention module, spatial-temporal graph convolution module, and gesture recognition. CNN feature extraction module is set up at the front end of the network to convert original HD-EMG signal into a feature representation with more discriminative ability, so as to improve the classification performance. In the graph generation module, two types of views, namely functional connection graph and distance-connection graph, are employed to capture more diverse topological information contained in HD-sEMG signal. Due to the presence of both spatial and temporal information in HD-sEMG signal, the spatial-temporal attention module is used to capture more valuable spatial information through spatial attention mechanism and more valuable temporal information through temporal attention mechanism. In the spatial-temporal graph convolution module, spatial graph convolution is employed to learn the similar spatial muscle connectivity relationships among different subjects, and a temporal convolution is used to extract the temporal dimension information. Finally, the features extracted by different spatial-temporal graph convolution modules are concatenated for gesture recognition. The detailed procedure is outlined in the pseudocode described in Algorithm 1.

a) CNN Feature Extraction Module: CNN is a type of artificial neural network known for its powerful feature extraction capabilities [34]. Its unique local connection and weight sharing structure make it particularly advantageous in the fields of image processing, computer vision, even EMG pattern recognition. In the proposed CNN-MSTGCN network, a standard CNN module is implanted into the front of the graph neural network. It is expected that the powerful feature extraction ability of CNN will be beneficial for improving the accuracy of EMG pattern recognition. In particular, a 1×1 standard convolution operation is employed with a convolution kernel size of (1, 1) and a stride of (1, 1). The number of filters is adjusted as hyper-parameter to determine the optimal network performance. Through CNN feature extraction module, the HD-sEMG context sample $h_i = (o_{i-n}, \ldots, o_i, \ldots, o_{i+n}) \in \mathbb{R}^{M\times T\times C_n}$ changes to HD-sEMG feature context sample $\mathcal{F}_i =$

Algorithm 1 CNN-MSTGCN Model Architecture

Input: HD-sEMG context dataset $\boldsymbol{H} = (\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_K}) \in \mathbb{R}^{\boldsymbol{K \times M \times T \times C_n}}$, Number of training epochs I.

Output: Gesture classification predictions for C classes of gestures.

```
Initialize network parameters.
for i = 1 to I do
  for k = 1 to K do
    Extract features by CNN feature extraction module;
    Generate the distance-connection graph via Eq. (1), and create the functional connection graph via Eq. (2)-(3);
    Capture valuable temporal information through the temporal attention mechanism via Eq. (5), and capture valuable spatial information through the spatial attention mechanism via Eq. (6);
    Extract spatial information through spatial graph convolution module via Eq. (7);
    Extract temporal information through temporal convolution module via Eq. (8);
    Obtain the C-class gesture classification results through fully connected and softmax layers;
    Calculate loss function via Eq. (4);
    Back propagation and update the parameters;
  end for
end for
```

$(\mathrm{f}_{i-n}, \ldots, \mathrm{f}_i, \ldots, \mathrm{f}_{i+n}) \in \mathbb{R}^{M \times F \times C_n}$, where F represents the number of features extracted. As a result, HD-sEMG context dataset $H = (h_1, h_2, \ldots, h_K) \in \mathbb{R}^{K \times M \times T \times C_n}$ changes to HD-sEMG feature context dataset $\mathbb{F} = (\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_K) \in \mathbb{R}^{K \times M \times F \times C_n}$, $\mathcal{F}_i \in \mathbb{F}\ (i \in 1, 2, \cdots, K)$.

b) Graph Generation Module: The function of graph generation module is to convert the sample $\mathcal{F}_i = (\mathrm{f}_{i-n}, \ldots, \mathrm{f}_i, \ldots, \mathrm{f}_{i+n}) \in \mathbb{R}^{M \times F \times C_n}$ into HD-sEMG graph context sample $\mathbb{G}_i = (G_{i-n}, \ldots, G_i, \ldots, G_{i+n})$ for input into the subsequent graph convolution module. Specifically, each G_i is independently generated by the corresponding $\mathrm{f}_i = \left(f_1^i, f_2^i, \ldots, f_M^i\right) \in \mathbb{R}^{M \times F}$. $G_i = (V_i, \boldsymbol{A_i})$ is an undirected graph, representing an independent HD-sEMG graph sample in $\mathbb{G}_i$. G_i includes V_i nodes and each node corresponds to an electrode channel and has F features, $|V_i| = M$ is the number of nodes in HD-sEMG graph sample; $\boldsymbol{A_i} \in \mathbb{R}^{M \times M}$ is the adjacency matrix of the HD-sEMG graph sample, and represents the connection relation between each node. Up to this point, the nodes and their associated features have been determined, that is, V_i corresponding to $\mathrm{f}_i = \left(f_1^i, f_2^i, \ldots, f_M^i\right) \in \mathbb{R}^{M \times F}$ is fixed, while $\boldsymbol{A_i}$ is variable depending on different graph generation methods. In this study, $\boldsymbol{A_i}$ is generated through two ways: one is to generate distance connections based on node distance, and the other is to generate functional connections based on adaptive graph learning.

Distance-connection Graph: Like many of the existing works in the field of EMG pattern recognition, the 128 channel electrodes are viewed as a 16×8 HD-sEMG image, so that each electrode corresponds to one of the pixels. The main idea of generating the distance-connection adjacency matrix $\boldsymbol{A_i^d}$ is as follows: the more distant pixels in the image, the smaller the connection value between them. The specific implementation method is to take the reciprocal of block distance between pixels as the connection value, for example, the connection value between adjacent pixels is 1. As the diagonal elements of the adjacency matrix represent the connection values between each pixel and itself, according to the definition, the connection values are even greater than that of adjacent pixels, so they are set to 2. Specifically, it can be summarized by Eq. (1), where $\boldsymbol{A_i^d}$ is the distance-connection adjacency matrix, m and n $(m, n \in \{1, 2, \cdots, M\})$ respectively represent the indexes of row and column and also represent ergodic pixels in HD-sEMG image, $m_{\mathrm{coord\_}x}$ and $m_{\mathrm{coord\_}y}$ respectively represent horizontal and vertical coordinate values of pixel m, and the same is true for n. Finally, the adjacency matrix is normalized by row.

$$\boldsymbol{A_i^d}(m, n) = \begin{cases} \dfrac{1}{\sqrt{\left(m_{\mathrm{coord\_}x} - n_{\mathrm{coord\_}x}\right)^2 + \left(m_{\mathrm{coord\_}y} - n_{\mathrm{coord\_}y}\right)^2}} & m \neq \mathrm{n} \\ 2 & m = n \end{cases} \tag{1}$$

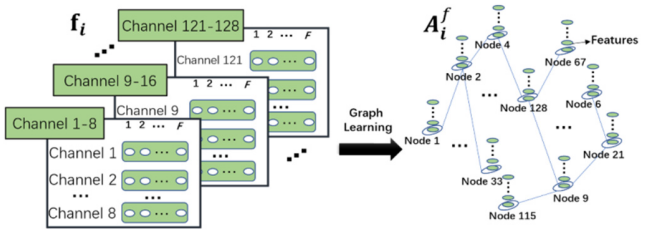

Fig. 5. Generate functional connection adjacency matrix by adaptive graph learning.

Functional Connection Graph: Since muscles are uniformly controlled by the central nervous system and functional connections may exist between non-adjacent muscles, it is often insufficient to represent the information contained in HD-sEMG signals by distance connections alone. Therefore, adaptive graph learning method [35] is adopted to dynamically generate functional connection adjacency matrix $\boldsymbol{A_i^f}$ as shown in Fig. 5. $\boldsymbol{A_i^f}$, as shown in Eq. (2), is derived by leveraging neural network to learn the relationships between the features of each channel within HD-sEMG feature sample $\mathbf{f_i}$, where f_m and f_n represent the features of different nodes in sample $\mathrm{f}_i = \left(f_1^i, f_2^i, \ldots, f_M^i\right) \in \mathbb{R}^{M \times F}$, $m, n \in \{1, 2, \cdots, M\}$, $\boldsymbol{w} = (w_1, w_2, \ldots, w_F) \in \mathbb{R}^F$ represents the weight vector of the neural network. The activation function ReLU guarantees that $\boldsymbol{A_i^f}$ is non-negative. Finally, $\boldsymbol{A_i^f}$ is normalized by softmax operation. The weight vector $\boldsymbol{w}$ is updated by minimizing the loss functions L_{GL} described in Eq. (3), where $\lambda \geq 0$ is a regularization parameter.

$$\boldsymbol{A_i^f} = \frac{\exp\left(\mathrm{ReLU}\left(\boldsymbol{w} | f_m - f_n |\right)\right)}{\sum_{n=1}^{M} \exp\left(\mathrm{ReLU}\left(\boldsymbol{w} | f_m - f_n |\right)\right)} \tag{2}$$

$$L_{GL} = \sum_{m,n=1}^{M} || f_m - f_n ||_2^2 \boldsymbol{A_i^f} + \lambda | \boldsymbol{A_i^f} |_F^2 \tag{3}$$

Since gesture recognition is a multi-classification task, cross entropy loss is adopted, as shown in Eq. (4).

$$L_{CE} = -\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{C} y_{i,j}\log \hat{y}_{i,j} \tag{4}$$

where K is the number of samples, C is the number of gesture categories, $y_{i,j}$ is the real category, and $\hat{y}_{i,j}$ is the prediction category. Therefore, the total loss function is $L_{total} = L_{CE} + L_{GL}$.

c) Spatial-temporal Attention Module: The temporal attention mechanism is adopted to capture valuable temporal information in HD-sEMG feature context data $\mathcal{F}_i = (\mathrm{f}_{i-n}, \ldots, \mathrm{f}_i, \ldots, \mathrm{f}_{i+n}) \in \mathbb{R}^{M\times F\times C_n}$, as shown in Eq. (5).

$$\boldsymbol{Q_i} = \text{softmax}\left(\boldsymbol{B_q}\cdot\sigma\left(\left((\mathcal{F}_\mathrm{i})^{\mathrm{T}}\boldsymbol{U_1}\right)\boldsymbol{U_2}\,(\boldsymbol{U_3}\mathcal{F}_\mathrm{i}) + \boldsymbol{b_q}\right)\right) \tag{5}$$

where $\boldsymbol{B}_q$, $\boldsymbol{b}_q \in \mathbb{R}^{C_n\times C_n}$, $\boldsymbol{U}_1 \in \mathbb{R}^{M}$, $\boldsymbol{U}_2 \in \mathbb{R}^{F\times M}$, $\boldsymbol{U}_3 \in \mathbb{R}^{F}$ are trainable parameters, σ indicates the sigmoid activation function. $\boldsymbol{Q_i}$ is the temporal attention matrix of the shape $C_n \times C_n$, representing the correlation between the context signals, and it is obtained by dynamically calculating the input $\mathcal{F}_i$. Finally, the attention matrix is normalized by softmax function.

$$\boldsymbol{P_i} = \text{softmax}\left(\boldsymbol{B}_p\cdot\sigma\left(\left(\hat{\mathcal{F}}_i\boldsymbol{V}_1\right)\boldsymbol{V}_2\left(\boldsymbol{V}_3\hat{\mathcal{F}}_i\right)^{T} + \boldsymbol{b}_p\right)\right) \tag{6}$$

In order to automatically extract more valuable spatial information, the spatial attention mechanism defined as Eq. (6) is adopted. The input of the spatial attention module is tuned by the temporal attention, namely $\hat{\mathcal{F}}_i = \left(\hat{\mathrm{f}_{i-n}}, \ldots, \hat{\mathrm{f}_i}, \ldots, \hat{\mathrm{f}_{i+n}}\right) = (\mathrm{f}_{i-n}, \ldots, \mathrm{f}_i, \ldots, \mathrm{f}_{i+n})\,\boldsymbol{Q_i} \in \mathbb{R}^{M\times F\times C_n}$. $\boldsymbol{B}_p, \boldsymbol{b}_p \in \mathbb{R}^{M\times M}$, $\boldsymbol{V}_1 \in \mathbb{R}^{C_n}$, $\boldsymbol{V}_2 \in \mathbb{R}^{F\times C_n}$, $\boldsymbol{V}_3 \in \mathbb{R}^{F}$ are trainable parameters, $\boldsymbol{P_i}$ is the spatial attention matrix of the shape $M \times M$, representing the correlation between nodes.

d) Spatial-temporal Graph Convolution Module: Spatial Graph Convolution: Graph convolution is adopted to extract spatial information from HD-sEMG feature context signal. Specifically, convolution kernels containing F' filters are used to perform Chebyshev graph convolution operation [36] on input $\mathcal{F}_i = (\mathrm{f}_{i-n}, \ldots, \mathrm{f}_i, \ldots, \mathrm{f}_{i+n}) \in \mathbb{R}^{M\times F\times C_n}$ to get output $Z_i = (\mathrm{z}_{i-n}, \ldots, \mathrm{z}_i, \ldots, z_{i+n}) \in \mathbb{R}^{M\times F'\times C_n}$. Specifically in graph convolution, Laplacian matrix is used to represent the link graph. The Laplacian matrix, denoted as $\boldsymbol{L}$, is defined as the difference between the degree matrix ($\boldsymbol{D}$) and the adjacency matrix ($\boldsymbol{A}$) of the graph. It can be expressed mathematically as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, and its normalized form is $\boldsymbol{L} = \boldsymbol{I_N} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}} \in \mathbb{R}^{N\times N}$, where $\boldsymbol{I_N}$ is a unit matrix, and the degree matrix $\boldsymbol{D} \in \mathbb{R}^{N\times N}$ is a diagonal matrix, consisting of node degrees, $\boldsymbol{D_{ii}} = \sum_{j\neq i}^{N}\boldsymbol{A_{ij}}$. Through the Laplacian matrix $\boldsymbol{L}$, each node can aggregate state information from its neighbors. The operation for spatial correlation extraction in graph-based models is defined as Eq. (7).

$$g_{\theta *G}\mathcal{F}_i = g_\theta(\boldsymbol{L})\mathcal{F}_i = \sum_{k=0}^{\eta-1}\theta_k T_k\tilde{(\boldsymbol{L})}\mathcal{F}_i \tag{7}$$

where g_θ is the convolution kernel, $*_G$ is the graph convolution operation, each node aggregates information from η order neighbors. The recursive definition of the Chebyshev polynomial is $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$, $T_0(x) = 1$, $T_1(x) = x$. $\theta \in \mathbb{R}^{\eta}$ is a Chebyshev polynomial coefficients vector, $\tilde{\boldsymbol{L}} = \frac{2}{\lambda_{\max}}\boldsymbol{L} - \boldsymbol{I}_N$, $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{L}$.

To dynamically adjust the correlation between each node, for each term of the Chebyshev polynomial, $T_k(\tilde{\boldsymbol{L}})$ is multiplied by the spatial attention matrix $\boldsymbol{P}$, resulting in $T_k(\tilde{\boldsymbol{L}})\odot\boldsymbol{P}$. Here, $\odot$ represents Hadamard product. Thus the graph convolution equation above is converted to $Z_i = \sum_{k=0}^{\eta-1}\theta_k(T_k(\tilde{\boldsymbol{L}})\odot\boldsymbol{P})\mathcal{F}_i$.

Temporal Convolution: A temporal CNN is adopted to perform convolution operation in temporal dimension. Specifically, a temporal convolution operation is performed on the output $Z_i = (\mathrm{z}_{i-n}, \ldots, \mathrm{z}_i, \ldots, z_{i+n}) \in \mathbb{R}^{M\times F'\times C_n}$ of the graph convolution operation to obtain $S_i = (\mathrm{s}_{i-n}, \ldots, \mathrm{s}_i, \ldots, s_{i+n}) \in \mathbb{R}^{M\times F'\times C'_n}$, as shown in Eq. (8).

$$S_i = \text{ReLU}\left(\Theta * (\text{ReLU}\,(Z_i))\right) \in \mathbb{R}^{M\times F'\times C'_n} \tag{8}$$

where $*$ represents the temporal convolution operation, and Θ represents the temporal convolution kernel.

e) Gesture Recognition Module: After the operations of the aforementioned modules are performed, the features obtained from the two views are flattened, concatenated, and sent to a fully connected layer for final gesture classification. The fully connected layer is configured with Y neurons. Y is a hyper-parameter that can be adjusted to achieve the best recognition performance. Finally, the softmax layer is used to obtain the 17-class gesture classification results.

2) Comparison Classifiers:

a) ResNet50: ResNet50 [37] is a deep convolutional neural network model, which adopts the residual learning method to solve the problem that the accuracy decreases when the network depth increases. Therefore, it stands out among many CNNs and performs well in many fields. Based on its high efficiency, ResNet50 is adopted as a comparison classifier in this study. In particular, the complete ResNet50 network architecture is employed, and slight modifications are made to accommodate HD-sEMG input and output format specific to this task. The input shape of ResNet50 is adjusted from the traditional image format (i.e., $224\times224\times3$) to $M \times T \times C_n$ format. Additionally, the output of the network is modified to consist of C *(C=17)* neurons corresponding to C gesture classification.

b) LSTM: LSTM [38] has been widely and efficiently applied in EMG pattern recognition due to its excellent sequence modeling ability and gating mechanism, which can handle long time series data and has a good ability to extract temporal features effectively. Therefore, it is also adopted as a comparison classifier in this study. To accommodate the input format of LSTM network, the time dimension and context dimension of the HD-sEMG context dataset $H = (h_1, h_2, \ldots, h_K) \in \mathbb{R}^{K\times M\times T\times C_n}$ are merged. This merging process results in a reshaped new dataset $H' = \left(h'_1, h'_2, \ldots, h'_K\right) \in \mathbb{R}^{K\times M\times TC_n}$, where TC_n represents the combined dimension of time and context and corresponds to the time step in the LSTM input form.

Ultimately, a specific LSTM network structure is employed. The input shape is defined as $M \times TC_n$, where TC_n represents

the number of time steps in the input data, and M represents the dimensionality of input features at each time step. The LSTM network is configured with 4 layers. Additionally, a hidden size of 64 is chosen, indicating that each LSTM layer consists of 64 hidden states. The output shape is set as C, corresponding to the number of gesture categories.

c) Performance Evaluation: The recognition rate is used as the performance evaluation criterion of the classifiers, which is defined as the ratio of the number of correctly classified samples to the number of all input samples. In order to verify the significance of the result differences among all the methods, one-way repeated-measure ANOVA is adopted for significance test and the Bonferroni method is employed for post hoc multiple comparisons tests. The significance level is set at 0.05, and all analyses are performed using SPSS (v. 25.0, SPSS Inc. Chicago, IL, USA).

III. Experimental Results

In this study, four distinct experiments are conducted. First, the parameter setting experiment is performed to determine the optimal parameters of networks and time context length; Then, the ablation experiment is conducted to explore the role of the key modules in CNN-MSTGCN; Finally, the user-independent experiment and the cross-user experiment based on transfer learning strategy are carried out to evaluate the performance of the proposed CNN-MSTGCN applied to EMG pattern recognition. In the experiments, the following hardware and software configurations are employed: a Tesla T4 GPU with 15109MiB of GPU memory, an Intel(R) Xeon(R) Gold 6240R CPU @ 2.40GHz, operating on Ubuntu 20.04.1, and utilizing TensorFlow version 1.15.0.

A. Parameter Setting Experiment

1) Network Parameters Setting: To optimize the performance of each network, adjustments are made to the network structures and hyper-parameters individually.Network parameters (including structures and hyper-parameters) are determined according to the loss function value and recognition rate of the validation set. When the training error converges and the generalization error is reduced to a relatively low level, the parameters are considered as the optimal choices. Using the proposed CNN-MSTGCN and comparison networks ResNet50 and LSTM, gesture recognition experiments are conducted on Subject 1 in user-specific mode. The training set and validation set are divided in a ratio of 7:3. As a result, the optimal parameters of graph neural network, ResNet50 and LSTM are listed in Table I-III respectively.

2) Time Context Length Parameter Setting: Since time context length C_n has an impact on the result of gesture recognition, before all experiments, an optimal context length is first selected as a fixed parameter through the following experiment. Using CNN-MSTGCN network, the gesture recognition experiment is conducted on all the 11 subjects in user-specific mode. The training and test data are divided in a ratio of 7: 3.C_n is set to 1, 3, 5, 7, and 9. The experimental results are shown in Fig. 6. It can be observed that the average recognition rate achieves its highest value (i.e., 96.9%) when C_n is set to 5. When C_n is less than 5, the recognition rate tends to decrease, possibly due to the insufficient inclusion of time information. Conversely, when C_n exceeds 5, the recognition rate also declines, potentially due to the inclusion of redundant or irrelevant information. Based on these findings, a C_n of 5 is selected in all subsequent experiments. This choice strikes a balance between capturing sufficient time

TABLE I
THE PARAMETERS OF GRAPH NEURAL NETWORK

Parameters	CNN-MSTGCN
Learning rate	0.0002
Batch size	64
Optimizer	Adam
Standard convolution filters	64
Block number of (M)STGCN	1
Graph convolution filters	64
Chebyshev polynomial Ƞ	3
Regularization parameter	0.0001
Temporal convolution filters	64
Dropout	0.5
Fully connected layer	64

TABLE II
THE PARAMETERS OF RESNET50

Parameters	ResNet50
Input shape	128×100×5
Preprocess conv layer	7×7, 64
Block1 × 3	1×1, 64 // 3×3, 64 // 1×1, 256
Block2 × 4	1×1, 128 // 3×3, 128 // 1×1, 512
Block3 × 6	1×1, 256 // 3×3, 256 // 1×1, 1024
Block4 × 3	1×1, 512 // 3×3, 512 // 1×1, 2048
Output shape	17
Learning rate	0.01
Batch size	64
Optimizer	Adam

TABLE III
THE PARAMETERS OF LSTM

Parameters	LSTM
Input shape	500
Input size	128
Hidden layers	4
Hidden size	64
Output shape	17
Learning rate	0.1
Batch size	64
Optimizer	Adam

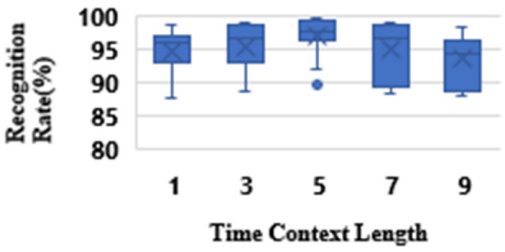

Fig. 6. Gesture recognition results with different time context length.

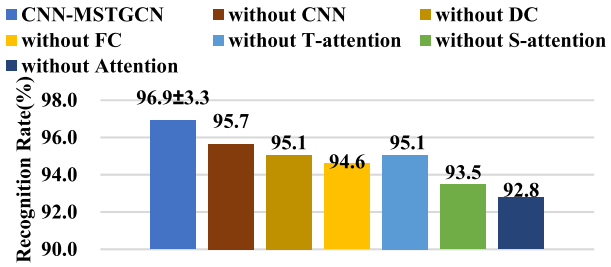

Fig. 7. Ablation experiment results.

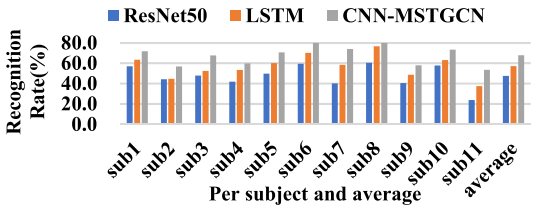

Fig. 8. 11-fold cross-validation experiment in user-independent mode.

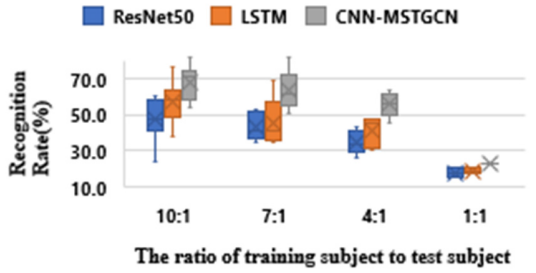

Fig. 9. Experimental results under different training test ratios.

information for accurate recognition and avoiding the inclusion of excessive or unnecessary temporal data.

B. Ablation Experiment on CNN-MSTGCN in User-Specific Mode

In order to explore the role of each functional modules in the proposed CNN-MSTGCN, ablation experiment is conducted. In specific, gesture recognition experiments are conducted on 11 subjects in user-specific mode with a training test ratio of 7:3. Seven classifiers are utilized in the experiments, including: CNN-MSTGCN, "without CNN" which does not have the CNN feature extraction module, "without DC" which does not utilize distance-connection graph, "without FC" which does not utilize functional connection graph, "without T-attention" which does not include the temporal attention mechanism, "without S-attention" which does not include the spatial attention mechanism, "without Attention" which does not include both attention mechanisms. The average experimental results with standard deviations are shown in Fig. 7.

The experimental results show that the average recognition rates of CNN-MSTGCN, "without CNN", "without DC", "without FC", "without T-attention", "without S-attention" and "without Attention" are 96.9%, 95.7%, 95.1%, 94.6%, 95.1%, 93.5%, 92.8% respectively. The average recognition rate of CNN-MSTGCN is 1.2% higher than that of "without CNN", indicating that CNN feature extraction module plays a certain role in improving recognition rate. The average recognition rate of CNN-MSTGCN is 1.8% and 2.3% higher than that of "without DC" and "without FC", suggesting that the inclusions of the distance-connection graph and functional connection graph are also beneficial. Further, the functional connection graph is more important than the distance-connection graph. In terms of the attention mechanism, CNN-MSTGCN outperforms the "without T-attention" network by 1.8%, the "without S-attention" network by 3.4%, and the network without any attention mechanism by 4.1%.

C. User-Independent Gesture Recognition Experiment

1) 11-Fold Cross-Validation Experiment: In user-independent gesture recognition experiment, data from one of the 11 subjects is taken as the test set, and data of the remaining 10 subjects are taken as the training set (i.e., at the training test ratio of 10:1). A total of 11 experiments are carried out using CNN-MSTGCN, ResNet50 and LSTM respectively, and the mean recognition rate of each subject are shown in Fig. 8. The average recognition rates achieved by CNN-MSTGCN (68.0%) are significantly higher compared to ResNet50 (47.5%, $p < 0.001$) and LSTM (57.1%, p=0.045). These experimental results verify that the proposed CNN-MSTGCN has comparative advantages in user-independent EMG pattern recognition.

2) Experiments Under Different Ratios of Training Subject to Test Subject: Similar to the 11-fold cross-validation experiment, 8, 5 and 2 subjects are randomly selected to perform 8-fold, 5-fold and 2-fold cross-validation experiments respectively (i.e., at the training test ratio of 7:1, 4:1 and 1:1 respectively), and the final results are demonstrated in Fig. 9 as a boxplot. As the number of training subject decreases, the user-independent recognition rate of each classifier will decrease. Under the same ratio, the recognition rate of CNN-MSTGCN is significantly higher than that of ResNet50 and LSTM. At the training test ratio of 10:1, the average recognition rate of CNN-MSTGCN (68.0%) is significantly higher ($p < 0.05$) than that of ResNet50 (47.5%) and LSTM (57.1%). Similarly, at ratios of 7:1 and 4:1, CNN-MSTGCN (63.4% and 55.7%, respectively) outperforms ResNet50 (47.5% and 43.3%, respectively) and LSTM (57.1% and 45.5%, respectively) with statistical significance (p=0.001 and $p < 0.001$, respectively). However, at the ratio of 1:1, the difference in performance between the three classifiers is not statistically significant (p=0.245). When the training test ratio is 4:1, the average recognition rate of CNN-MSTGCN (55.7%) is still higher than or close to that of ResNet50 (47.5%) and LSTM (57.1%) at the ratio of 10:1. The experimental results show that in user-independent mode, CNN-MSTGCN has a lower training burden.

D. Cross-User Experiments Based on Transfer Learning

In cross-user gesture recognition experiment, the transfer learning strategy is integrated on the basis of the proposed CNN-MSTGCN, i.e., TL-CMSTGCN. TL-CMSTGCN is also compared with ResNet50 and LSTM that integrate transfer learning strategy. The data of 10 subjects is used as the training

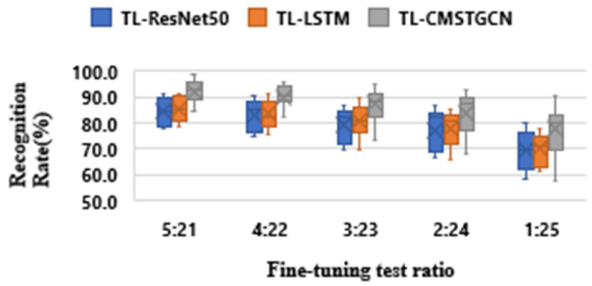

Fig. 10. Cross-user experiments with various fine-tuning test ratio.

set, and the data of the remaining test subject is divided into fine-tuning set and test set. Because the data of each test subject can be equally divided into 26 parts, the ratio of the fine-tuning set to the test set is progressively reduced at the ratios of 5:21->4:22->3:23->2:24->1:25 to evaluate the robustness of TL-CMSTGCN to different amounts of fine-tuning data.

As shown in Fig. 10, TL-CMSTGCN achieves a significantly higher average recognition rate (92.3%) compared to TL-ResNet50 (84.6%, p = 0.003) and TL-LSTM (85.3%, p = 0.008) at the ratio of 5:21, which demonstrates the superiority of TL-CMSTGCN in cross-user EMG pattern recognition. Moreover, TL-CMSTGCN maintains a relative high recognition rate even when the number of fine-tuning samples is reduced, indicating a lighter user training burden. In comparison, the other two classifiers experience a more significant decline in recognition rates. Notably, when the fine-tuning test ratio is set to 4:22, the average recognition rate of TL-CMSTGCN remains above 90%. At the ratio of 3:23, the average recognition rate of TL-CMSTGCN (87.2%) surpasses that of TL-ResNet50 (84.6%) and TL-LSTM (85.3%) at the ratio of 5:21. These results demonstrate that TL-CMSTGCN can achieve higher recognition rates with fewer fine-tuning samples, thereby reducing the training burden on users.

IV. Discussion

Considering that GNN can learn functional connections between non adjacent but related muscles, this study focuses on exploring GNN-based EMG pattern recognition solutions. Specifically, using MSTGCN with spatial-temporal attention mechanism as basic classifier, a new CNN-MSTGCN network is proposed, and its advantages are verified by ablation experiments, user-independent pattern recognition experiments and cross-user pattern recognition experiments. Based on the state-of-the-art in the field of EMG pattern recognition, the research results of this study can be discussed as follows:

A. The Role of Key Functional Modules in CNN-MSTGCN

The proposed CNN-MSTGCN comprises multiple functional modules including CNN feature extraction module, graph generation module, and spatial-temporal attention mechanism module. The results of the ablation experiment show that each functional module has played a more or less role in improving the performance of EMG pattern recognition. At first, CNN-MSTGCN outperforms MSTGCN without CNN feature extraction module in terms of average recognition rate, demonstrating that the CNN feature extraction module is able to capture discriminative features from HD-sEMG data; Second, both the two graph generation methods play certain role in improving network performance, and the functional connection graph is more effective than the distance-connection graph. The possible reason is that the functional connection graph can not only reflect the spatial proximity of HD-sEMG signal, but also can learn the potential relationship between signals in different regions; Third, CNN-MSTGCN networks without the temporal attention mechanism or spatial attention mechanism (as well as without both attention mechanisms) exhibit inferior performance compared to the complete CNN-MSTGCN. Specifically, the recognition rate of CNN-MSTGCN without spatial attention mechanism decreases more significantly than that of CNN-MSTGCN without temporal attention mechanism. This result suggests that the spatial attention mechanism plays a more crucial role than the temporal attention mechanism.

TABLE IV
Comparison of Relevant Studies on User-Independent EMG Pattern Recognition

Authors	Year	Gest-ures	Classifier	Recognition Rate (%)
Saponas et al. [39]	2008	5	SVM	57
Orabona et al. [40]	2009	3	SVM	89.2
Matsubara et al. [9]	2013	5	SVM	54
Ketyko et al. [10]	2019	12	2SRNN	35.1
Côté-Allard et al. [11]	2020	11	ConvNet	65.03
Zhang et al. [41]	2020	5	SMO	41.5
			KNN	41.05
			RF	44.43
			PCA	35.18
Yu et al. [15]	2021	12	CNN	58
Zou et al. [32]	2023	7	MLHG	53.52
Ours	**2023**	**17**	ResNet50	47.5
			LSTM	57.1
			CNN-MSTGCN	**68.0**

B. The Superior Performance of CNN-MSTGCN in User-Independent EMG Pattern Recognition

As shown in Table IV, although many classifiers such as SVM [9], [39], [40], 2SRNN [10], ConvNet [11], Sequential minimal optimization (SMO), K-nearest neighbor (KNN), random forest (RF) and principal component analysis (PCA) [41], CNN [15]etc., have been used in relevant studies for user-independent EMG pattern recognition, relatively low recognition rates were reported.

In this study, the proposed CNN-MSTGCN achieves a recognition rate of 68% in 17-gesture recognition task. Observing Table IV, it can be seen that our work involves the most types of gestures, with a recognition rate only lower than 89.2% of the 3-gesture task in the work of Orabona et al. [40]. Comparing the two works based on graph neural networks, the recognition rate of CNN-MSTGCN is 68% for 17 gestures, while the MLHG model proposed by Zou et al. [32] reached only 53.52% in a 7-gesture task. Additionally, CNN-MSTGCN performs significantly better than ResNet50 and LSTM in

TABLE V
COMPARISON OF RELEVANT STUDIES ON CROSS-USER EMG PATTERN

Authors	Year	Gest -ures	Strategies	Recognition Rate(%)
Ketyko et al. [10]	2019	12	2SRNN-DA	67
Zhang et al. [41]	2020	5	dualTL	80.17
Yu et al.[15]	2021	12	TL-CNN	71.16
Li et al. [12]	2023	7	MAT-DA	93.54
Zhang et al. [13]	2023	6	MS-DA	87.93
Ours	**2023**	**17**	TL-ResNet50	84.6
			TL-LSTM	85.3
			TL-CMSTGCN	**92.3**

TABLE VI
TRAINING AND TEST TIMES OF THREE CLASSIFIERS

Classifiers	training time (s/2562 sample)	test time (ms/sample)
CNN-MSTGCN	15.99	2.61
ResNet50	3.10	0.11
LSTM	2.60	0.10

user-independent mode. All of above results indicate that the proposed CNN-MSTGCN has better application potential than traditional methods and existing graph neural networks in user-independent EMG pattern recognition tasks.

C. The Superior Performance of the Proposed CNN-MSTGCN in Cross-User EMG Pattern Recognition

As summarized in Table V, transfer learning (TL) and domain adaptation (DA) strategies are usually adopted in relevant studies on cross-user EMG pattern recognition. In this study, transfer learning strategy is integrated into CNN-MSTGCN to get the cross-user EMG pattern recognition scheme TL-CMSTGCN. Observing Table V, it can be seen that our work involves the most types of gestures and TL-CMSTGCN obtains a relatively high recognition rate of 92.3%, which is only lower than 93.54% of the 7-gesture task in the work of Li et al. [12]. This result demonstrates that the proposed TL-CMSTGCN scheme has great advantages in improving cross-user gesture recognition rate. Compared with CNN-MSTGCN in user-independent mode, only requiring a small amount of data for fine-tuning, TL-CMSTGCN increases the average recognition rate by more than 20%. In the meanwhile, the TL-CMSTGCN scheme has the advantage of low training burden. Due to the fact that most users can generally accept low training burden, the high recognition rate of TL-CMSTGCN exceeding 90% makes it possible to apply it in myoelectric control systems.

D. The Evaluation of Real-Time Performance

To offer a more comprehensive evaluation and gain better insights into the performance and practical feasibility of our proposed method, a statistical analysis of training and test times is conducted, as depicted in Table VI. The results of CNN-MSTGCN indicate a training time of 15.99s for 2562 training samples and a test time of 2.61ms per test sample. Although this may not reach the speeds of ResNet50 and LSTM, it still meets the real-time requirements.

E. Limitations and Further Work

We would like to point out the limitations of this study. First, the research results are obtained through offline data processing. Although it takes only 2.61ms to recognize a gesture in the offline program, which well meets the requirement of real-time myoelectric control, real-time online testing is not conducted in this study, and the specific real-time performance needs to be verified through follow-up work; Second, although the proposed CNN-MSTGCN performs better in user-independent mode than previous works, the recognition rate of 68% is not satisfactory. Therefore, future work should focus on improving network structure to enhance recognition rate; Third, although TL-CMSTGCN improves the cross-user recognition rate to a high level of 92.3% at the expense of a small amount of user training burden, more endeavors should be made to optimize transfer learning process and minimize the training burden. Finally, additional public datasets can be utilized to further support the feasibility and superiority of the proposed approach.

V. CONCLUSION

In order to promote the development of myoelectric control technology, this study focuses on exploring GNN-based EMG pattern recognition solutions. Specifically, a new graph neural network CNN-MSTGCN is proposed for high-density EMG signal pattern recognition. The characteristics of the proposed CNN-MSTGCN lie in: 1) Using a graph neural network MST-GCN with spatial-temporal attention mechanism as the basic classifier to adaptively learn the cross muscle connections; 2) The front-end is embedded with a CNN feature extraction module to improve the feature representation ability of HD-sEMG signals. Through user-independent pattern recognition experiments and cross-user pattern recognition experiments based on transfer learning strategy, the advantages of the proposed CNN-MSTGCN in improving recognition rate and reducing user training burden have been verified. The research results of this paper confirm that GNN has certain advantages in overcoming the impact of individual differences, and can provide possible solutions for achieving robust EMG pattern recognition.

REFERENCES

[1] A. Krasoulis, S. Vijayakumar, and K. Nazarpour, "Multi-grip classification-based prosthesis control with two EMG-IMU sensors," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 508–518, 2019.
[2] X. Song, S. S. Van De Ven, L. Liu, F. J. Wouda, H. Wang, and P. B. Shull, "Activities of daily living-based rehabilitation system for arm and hand motor function retraining after stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 621–631, 2022.
[3] B. Fang et al., "Simultaneous sEMG recognition of gestures and force levels for interaction with prosthetic hand," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2426–2436, 2022.
[4] J. Behrenbeck et al., "Classification and regression of spatio-temporal signals using NeuCube and its realization on SpiNNaker neuromorphic hardware," *J. Neural Eng.*, vol. 16, no. 2, Apr. 2019, Art. no. 026014.
[5] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.
[6] G. Li, A. E. Schultz, and T. A. Kuiken, "Quantifying pattern recognition—Based myoelectric control of multifunctional transradial prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 185–192, Apr. 2010.

[7] G. R. Naik and H. T. Nguyen, "Nonnegative matrix factorization for the identification of EMG finger movements: Evaluation using matrix analysis," *IEEE J. Biomed. Health Informat.*, vol. 19, no. 2, pp. 478–485, Mar. 2015.

[8] A. J. Young, L. H. Smith, E. J. Rouse, and L. J. Hargrove, "Classification of simultaneous movements using surface EMG pattern recognition," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 5, pp. 1250–1258, May 2013.

[9] T. Matsubara and J. Morimoto, "Bilinear modeling of EMG signals to extract user-independent features for multiuser myoelectric interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2205–2213, Aug. 2013.

[10] I. Ketykó, F. Kovács, and K. Z. Varga, "Domain adaptation for sEMG-based gesture recognition with recurrent neural networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–7.

[11] U. Cote-Allard, E. Campbell, A. Phinyomark, F. Laviolette, B. Gosselin, and E. Scheme, "Interpreting deep learning features for myoelectric control: A comparison with handcrafted features," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 158, Mar. 2020.

[12] X. Li, X. Zhang, X. Chen, X. Chen, and L. Zhang, "A unified user-generic framework for myoelectric pattern recognition: Mix-up and adversarial training for domain generalization and adaptation," *IEEE Trans. Biomed. Eng.*, vol. 70, no. 8, pp. 2248–2257, Aug. 2023, doi: 10.1109/TBME.2023.3239687.

[13] X. Zhang, L. Wu, X. Zhang, X. Chen, C. Li, and X. Chen, "Multi-source domain generalization and adaptation toward cross-subject myoelectric pattern recognition," *J. Neural Eng.*, vol. 20, no. 1, Feb. 2023, Art. no. 016050.

[14] W. Wang, B. Chen, P. Xia, J. Hu, and Y. Peng, "Sensor fusion for myoelectric control based on deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 9, pp. 272–282, Sep. 2018.

[15] Z. Yu, J. Zhao, Y. Wang, L. He, and S. Wang, "Surface EMG-based instantaneous hand gesture recognition using convolutional neural network with the transfer learning method," *Sensors*, vol. 21, no. 7, p. 2540, Apr. 2021.

[16] X. Chen, Y. Li, R. Hu, X. Zhang, and X. Chen, "Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1292–1304, Apr. 2021.

[17] E. Campbell, A. Phinyomark, and E. Scheme, "Deep cross-user models reduce the training burden in myoelectric control," *Frontiers Neurosci.*, vol. 15, May 2021, Art. no. 657958.

[18] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[19] J.-U. Chu and Y.-J. Lee, "Conjugate-prior-penalized learning of Gaussian mixture models for multifunction myoelectric hand control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 287–297, Jun. 2009.

[20] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Feature reduction and selection for EMG signal classification," *Exp. Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, Jun. 2012.

[21] M. Simão, P. Neto, and O. Gibaru, "EMG-based online classification of gestures with recurrent neural networks," *Pattern Recognit. Lett.*, vol. 128, pp. 45–51, Dec. 2019.

[22] R. F. Schmidt and G. Thews, *Human Physiology*. Berlin, Germany: Springer, 1983.

[23] A. Micheli, "Neural network for graphs: A contextual constructive approach," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, pp. 498–511, Mar. 2009.

[24] X. Mao, M. Liu, and Y. Wang, "Using GNN to detect financial fraud based on the related party transactions network," *Proc. Comput. Sci.*, vol. 214, pp. 351–358, Jan. 2022.

[25] Y. Xie, Y. Xiong, and Y. Zhu, "SAST-GNN: A self-attention based spatio-temporal graph neural network for traffic prediction," in *Proc. 25th Int. Conf. Database Syst. Adv. Appl. (DASFAA)*, Jeju, South Korea. Cham, Switzerland: Springer, Sep. 2020, pp. 707–714.

[26] M. Reau, N. Renaud, L. C. Xue, and A. M. Bonvin, "DeepRank-GNN: A graph neural network framework to learn patterns in protein–protein interfaces," *Bioinformatics*, vol. 39, no. 1, 2023, Art. no. btac759.

[27] W. Shi and R. Rajkumar, "Point-GNN: Graph neural network for 3D object detection in a point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1708–1716.

[28] Z. Jia et al., "Multi-view spatial–temporal graph convolutional networks with domain generalization for sleep stage classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1977–1986, 2021.

[29] Z. Lai et al., "STCN-GR: Spatial–temporal convolutional networks for surface-electromyography-based gesture recognition," in *Neural Information Processing*. Berlin, Germany: Springer, 2021, pp. 27–39.

[30] H. Lee et al., "Stretchable array electromyography sensor with graph neural network for static and dynamic gestures recognition system," *npj Flexible Electron.*, vol. 7, no. 1, p. 20, Apr. 2023.

[31] S. Yang, M. Li, and J. Wang, "Fusing sEMG and EEG to increase the robustness of hand motion recognition using functional connectivity and GCN," *IEEE Sensors J.*, vol. 22, no. 24, pp. 24309–24319, Dec. 2022.

[32] Y. Zou, L. Cheng, L. Han, Z. Li, and L. Song, "Decoding electromyographic signal with multiple labels for hand gesture recognition," *IEEE Signal Process. Lett.*, vol. 30, pp. 483–487, 2023.

[33] E. J. Scheme, B. S. Hudgins, and K. B. Englehart, "Confidence-based rejection for improved pattern recognition myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1563–1570, Jun. 2013.

[34] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[35] Z. Jia et al., "GraphSleepNet: Adaptive spatial–temporal graph convolutional networks for sleep stage classification," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 1324–1330.

[36] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[39] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2008, pp. 515–524.

[40] F. Orabona, C. Castellini, B. Caputo, A. E. Fiorilla, and G. Sandini, "Model adaptation with least-squares SVM for adaptive hand prosthetics," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 2897–2903.

[41] Y. Zhang, Y. Chen, H. Yu, X. Yang, and W. Lu, "Dual layer transfer learning for sEMG-based user-independent gesture recognition," *Pers. Ubiquitous Comput.*, vol. 26, no. 3, pp. 575–586, Jun. 2022.