Yuan et al.
Journal of NeuroEngineering and Rehabilitation (2024) 21:233
https://doi.org/10.1186/s12984-024-01526-3

Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

Exploring pattern-specific components associated with hand gestures through different sEMG measures

Yangyang Yuan[1], Jionghui Liu[2], Chenyun Dai[3], Xiao Liu[1], Bo Hu[1*] and Jiahao Fan[1*]

Abstract

For surface electromyography (sEMG) based human–machine interaction systems, accurately recognizing the users' gesture intent is crucial. However, due to the existence of subject-specific components in sEMG signals, subject-specific models may deteriorate when applied to new users. In this study, we hypothesize that in addition to subject-specific components, sEMG signals also contain pattern-specific components, which is independent of individuals and solely related to gesture patterns. Based on this hypothesis, we disentangled these two components from sEMG signals with an auto-encoder and applied the pattern-specific components to establish a general gesture recognition model in cross-subject scenarios. Furthermore, we compared the characteristics of the pattern-specific information contained in three categories of EMG measures: signal waveform, time-domain features, and frequency-domain features. Our hypothesis was validated on an open source database. Ultimately, the combination of time- and frequency-domain features achieved the best performance in gesture classification tasks, with a maximum accuracy of 84.3%. For individual feature, frequency-domain features performed the best and were proved most suitable for separating the two components. Additionally, we intuitively visualized the heatmaps of pattern-specific components based on the topological position of electrode arrays and explored their physiological interpretability by examining the correspondence between the heatmaps and muscle activation areas.

Keywords Surface electromyography, Gesture recognition, Feature projection, Auto-encode

Introduction

With the widespread adoption of electronic devices, human–machine interaction (HMI) systems have been extensively involved in our daily lives [1–3]. Gesture is one of the most natural interaction approaches for humans. Therefore, HMI systems [4] using gesture as the input commands have attracted significant attention from researchers. The human–machine interface, serving as the medium connecting humans and machines, plays an important role in HMI systems. With the in-depth exploration on gesture recognition systems, surface electromyography (sEMG) signal as a physiological interface has been widely used for the intuitive control of HMI systems [5], since it has a high signal-to-noise ratio and can be easily collected from the skin surface. As a result, sEMG-based gesture recognition systems, characterized by their anti-noise robustness and daily wearability, hold broad commercial prospects. Currently, sEMG-based wearable devices, such as prosthetic hand [6, 7] and wrist

*Correspondence:
Bo Hu
bohu@fudan.edu.cn
Jiahao Fan
fanjh18@fudan.edu.cn
[1] School of Information Science and Technology, Fudan University, Shanghai 200433, China
[2] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, China
[3] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200241, China

band [8, 9], have been used in clinical rehabilitation and daily activities.

To accurately control the devices according to the gesture commands, the system is required to establish a model to recognize the user's intents with high accuracy using sEMG. In practical use, most existing sEMG-based HMI systems oblige users to perform personalized calibration of the recognition model before use. However, the calibration process significantly reduces user's convenience and experience. Accordingly, the advanced systems are required to train a generalized model that can accommodate to all users using sEMG. However, affected by the difference of muscle contraction habits and physiological characteristics, sEMG presents different patterns across individuals, which could interfere with the decision-making of the motion intent recognition model [10]. We refer to the components varying from person to person as the subject-specific components of sEMG. The existence of this component brings significant challenges to establish such a generalized model suitable for all users. Fortunately, due to the similarity of human neurophysiological structures, the sEMG patterns of different individuals have many similar components when they perform the same gesture [11]. The common components are termed as pattern-specific components associated with a specific gesture. They reflect the common muscle contraction pattern in a wide population and are independent of individual characteristics, thus providing possibility to establish a generalized gesture recognition model based on sEMG.

To improve the performance of the model for new users, we need to increase the weights of pattern-specific components in model decision and reduce the influence of subject-specific components. The latest studies [12, 13] addressed this issue using the approach of transfer learning. The algorithm aims to construct a feature space to extract sEMG features with the smallest difference across subjects and the largest distance across gestures. Although transfer learning methods can effectively improve the performance of the cross-subject model, it requires a small amount of calibration data from the new user and additional model re-calibration step [14, 15], which still increase the user burden in practical use.

By contrast, our latest study [16] has firstly noticed that the subject-specific and pattern-specific components are orthogonal in sEMG. Accordingly, the two components can be disentangled from sEMG signals for with a multi-branch autoencoder (AE) and a decoder. After disentanglement, we can establish a generalized gesture recognition model using only the pattern-specific components. When any new user employs a HMI system with this model, the model can recognize their gesture intention by simply extracting pattern-specific components from their sEMG signals, without requiring any model re-calibration or any additional data from the new user. However, in that study [16], we only utilized common amplitude features in time-domain, including root mean square (RMS), wave length (WL), zero crossing (ZC), and slope sign change (SSC), lacking in-depth exploration of other sEMG measures. Besides, although this study proved the differences of the pattern-related components between different gestures and the similarities between different subjects, it did not provide insights into the disentangled components from a neurophysiological perspective. For better application of the disentangled components in HMI systems, a deeper understanding of its characteristics is necessary.

In our study, we explored the influence of different measures and their combinations besides of amplitude features. Specifically, we further compared the hand gesture recognition task accuracy of training the disentanglement model by original signal waveform, time-frequency features and other commonly used time-domain features. The validation was carried out on the pattern recognition dataset from Hyser [17], an open-access dataset available at the website (https://doi.org/10.13026/ym7v-bh53). The codes will be open sourced immediately once accepted. The novelty of the work are summarized as follows: (1) The disentanglement model is innovatively proposed inspired by a classical generative adversarial network (GAN). The GAN-based model can exact a more robust pattern-specific component against the individual difference of sEMG signals. (2) This study preliminarily investigated the similarities and differences between different patterns disentangled from different measures. The results provided physiological interpretability for the pattern-specific components in terms of the neuromuscular activation patterns of human body, promoting its application in HMI systems based on sEMG.

Materials

We validated our hypothesis with the pattern recognition subset of the open access high-density sEMG dataset Hyser [17], accessible at the website (https://doi.org/10.13026/ym7v-bh53). We selected the data of 10 gestures from the subset for this study. Here, a brief introduction is provided for the subject information and data acquisition in the following subsections.

Subjects

The experiment included 20 participants, consisting of 8 women and 12 men, aged between 22 and 34 years old, all intact and right-handed. Each participant received detailed information about the procedures and provided their signed informed consent before the experiment. The experiment was supervised and approved by the ethics committee of Fudan University (approval number: BE2035).

Data acquisition

Figure 1 illustrates the electrode setup during the data acquisition. The high-density sEMG signals with 256-channels were collected using four 8 × 8 electrode arrays positioned on the forearm, with two arrays each on the flexor and the extensor muscles. The configuration of each electrode array includes 8 × 8 gelled electrodes spaced 10 mm apart (center-to-center). Each electrode is an ellipse with 5-mm major axis and 2.8-mm minor axis. A right leg drive electrode and a reference electrode were placed on the head of the ulna and the olecranon respectively. The Quattrocento system (OT Bioelettronica in Torino, Italy) sampled the data at a frequency of 2048 Hz with a 150-fold amplification gain and a 16-bit resolution.

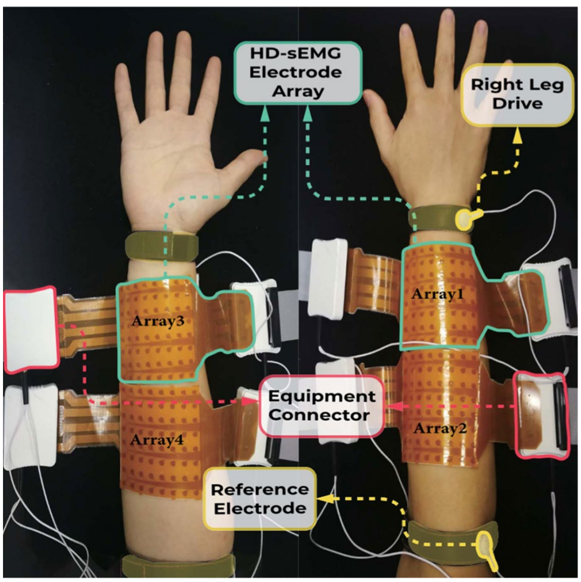


Fig. 1 Electrode setup in experiment

The experiment was conducted in a quite room. During data acquisition, subjects sat in a comfortable position in front of a computer screen, performing the required gestures following the instructions shown on it. In the whole experiment, 34 gestures were involved. For each gesture, the subject was required to perform two repeated trials, with each trial comprised of three 1-s dynamic tasks (from relaxing state to the required gesture followed with returning to relaxing state) and one 4-s maintenance task (from relaxing state to the required gesture followed with maintenance at that gesture). To avoid muscle fatigue, a 2-s inter-task resting period and a 5-s inter-trial resting period were provided. Subjects repeated the experiment on two separate days with an interval of 3 to 25 days, averaging around 8.5 ± 6 days. When performing a wrong task or missing a task, they were asked to inform the experiment assistant. These tasks would be removed from the final dataset. On average, 2.30 ± 2.71 dynamic tasks and 0.85 ± 1.05 maintenance tasks in each experiment were removed from the final dataset.

Gesture selection

In this study, 10 gestures (illustrated in Fig. 2) were selected for validation, namely (1) wrist flexion, (2) wrist extension, (3) wrist radial, (4) wrist ulnar, (5) wrist pronation, (6) wrist supination, (7) hand close, (8) hand open, (9) thumb and index fingers pinch, (10) thumb and middle fingers pinch. We selected these gestures because they are most commonly used in daily life and easily completed for the subjects.

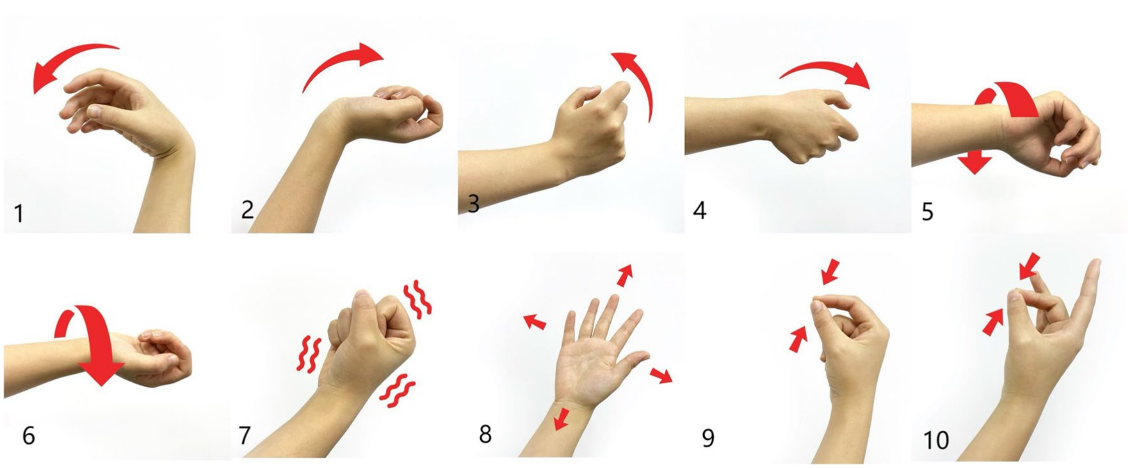


Fig. 2 The selected 10 gestures in this study

In practical scenarios, users are more inclined to completing a gesture at a comfortable speed (usually within 1 s) instead of maintaining the gesture for a period of time. Therefore, we only used dynamic tasks for further analysis in this study.

Methods

Signal preprocessing

We decimated the sampling rate of raw EMG signal into 1024 Hz. Then, the raw signals were filtered with a 10–500 Hz band-pass filter. Sequentially, a series of notch filters at 50 Hz and its harmonics up to 400 Hz were applied to attenuate the power-line interference. After filtering, we segmented the sEMG signals according to the trigger recorded during the acquisition. Each segment contains 1024 data points (1024 Hz × 1 s). Considering that most of the subjects have a certain reaction time before performing the gesture, we selected the last 0.5 s of each data segment as one sample.

EMG measures

The information carried in sEMG signals can be reflected from many views, such as waveform, time-domain features and frequency-domain features. To compare the similarities and differences between the pattern-specific components disentangled from distinct types of feature information, we selected different sEMG measures as the input of the network [18], including raw signal, sEMG envelope, short-time Fourier transformation (STFT), root mean square (RMS), wave length (WL), zero crossing (ZC), and slope sign change (SSC). For these measures, raw signal and sEMG envelope reflect the waveform characteristics of sEMG signals, STFT contains frequency-domain information in sEMG, and the last four measures represent the time-domain features of sEMG. In the following subsections, these measures are introduced in detail.

For clearer explanation, each sample is denoted as $\{x_i^j\} \in \mathbb{R}^d$. Specifically, x denotes to the sEMG features with dimension of d, which can be varied across different sEMG measures. Since the sEMG signals used in this study have 256 channels, each sample can be referred as a $d \times 256$ matrix. $i \in \{1, N_s\}$ and $j \in \{1, N_p\}$ denote the indexes of subject identity and gesture respectively, where $N_s = 20$ and $N_p = 10$ since we have 20 subjects and 10 gestures.

Raw signal

To control the size of data for model training due to the memory size of GPU, we decimated the segmented data three times. Accordingly, the size of raw signal measure for each sample is 171×256 ($d = 171$).

sEMG envelope

For the sEMG envelope, we smoothed the signal to reduce the significant vibration in the raw signal. In detail, we calculated the root mean square (RMS) of each sample with a window length of 31.25 ms and a step length of 1.95 ms. Accordingly, the size of sEMG envelope measure for each sample is 240×256 ($d = 240$).

Frequency-domain features

We selected short-time Fourier transformation (STFT) [19] as the frequency-domain feature of sEMG signal because it can preserve time-series information as well. Specifically, the window length of STFT was set as 125 ms and the overlap was zero, generating 4 windows for each 0.5-s sample. To reduce memory occupation, we downsampled the spectrum by four times. Accordingly, the size of frequency-domain measure for each sample is 256 (4 windows $\times$ 64 elements per window) $\times$ 256 (d=256).

Time-domain features

For time-domain feature extraction, we selected four representative time-domain features of sEMG [20], namely root mean square (RMS), wave length (WL), zero crossing (ZC), and slope sign change (SSC). The entire sample in each channel was used to calculate the value of each feature. The vectors of four time-domain features were concatenated into one matrix. Accordingly, the size of time-domain measure for each sample is 4×256 ($d = 4$).

Combination of different EMG measures

The frequency-domain and time-domain features of sEMG carry two types of information in two representative but distinct perspectives, both of them performing well in gesture recognition tasks. Therefore, we assumed that the combination of the two measures may provide a more complete description for the characteristics of sEMG signals to achieve a better disentanglement effect.

Therefore, we also investigated the performance of the combination of different EMG measures. However, considering the disentanglement model used in this study (see Fig. 3, and more details can be found in the next section), there are multiple options for combining different features at different positions in the network:

Combining before encoder

In this case, the STFT and all time-domain features were combined into a $d \times 16 \times 16$ array ($d = 260$), and then

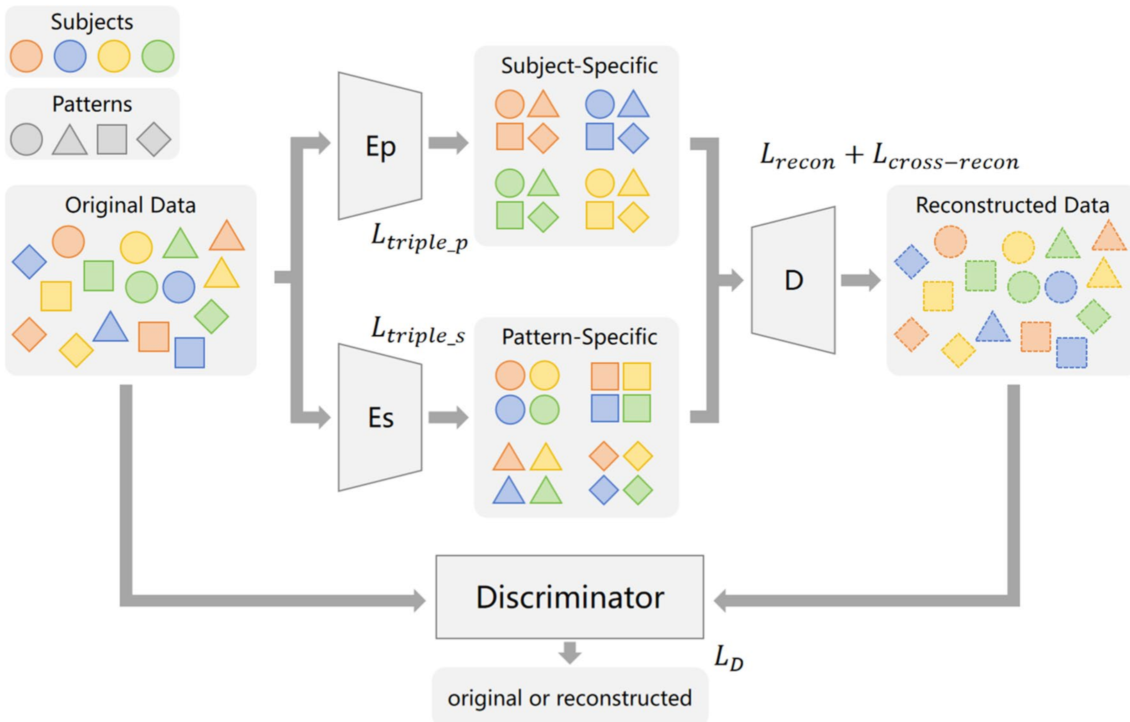


Fig. 3 The framework of proposed model

served as the model input. In other words, they shared the same encoders and decoder during training.

Combining before decoder
In this case, the STFT and all time-domain features were separately input into the encoder and then combined before the decoder. It means that they had independent encoders for disentanglement and shared a decoder for reconstruction.

Combining before classifier
In this case, the STFT and all time-domain features have independent encoders and decoder from each other. Their pattern-specific components were combined after the model training and then used for gesture recognition.

Disentanglement network model
To enable the disentanglement model to learn the spatial information of array electrodes, we remapped the data into 16×16 according to the electrode topological position during acquisition. Accordingly, the sEMG measures were fed into the disentanglement model in the shape of $d \times 16 \times 16$.

In our original study which firstly proposed the sEMG disentanglement model, a multiple encoders and one decoder model structure was proposed to separate the subject-specific and pattern-specific components [16]. In this study, the disentanglement model is innovatively proposed inspired by a classical generative adversarial network (GAN) [21], combined with the original structure. The modification can further enhance the robustness of the model against the individual difference of sEMG signals. The generator is constructed based on a multi-encoder and single-decoder architecture [22, 23], and the discriminator is composed of a series of fully connected layers. The whole architecture of the network is illustrated in Fig. 3. The two encoders, E_p and E_s, share the same architecture based on convolutional neural networks (CNN), but are trained independently. They respectively serve to disentangle the pattern-specific and subject-specific components from the the sEMG measures. Accordingly, the decoder takes the responsibility to reconstruct the original inputs with the two components disentangled by the encoders. For better reconstruction of the generator, the task of the discriminator *Dis* is to distinguish between real samples and reconstructed samples.

The detailed network parameters of the model are listed in Table 1. In the table, Conv, IN, LRLU, UpS, RP and DO denote Convolution, Instance Normalization, Leaky ReLU, Upsample, Reflection Pad and Dropout layers, respectively. k, s and p respectively denote the kernel, stride and padding size of the Convolution Layer. In/Out denotes the input/output channel number of the

Table 1 The detailed parameters of the proposed network

Module	Layers	k	s	p	In/Out
Encoder	Conv + IN + LRLU	3	2	1	d/512
	Conv + IN + LRLU	3	2	1	512/256
	Conv + IN + LRLU	3	2	1	256/128
Decoder	UpS + RP + Conv + DO + LRLU	3	1	1	256/128
	UpS + RP + Conv + DO + LRLU	3	1	1	128/64
	UpS + RP + Conv + DO + LRLU	3	1	1	64/d
Discriminator	Linear + LRLU	–	–	–	d*256/512
	Linear + LRLU	–	–	–	512/256
	Linear + LRLU	–	–	–	256/1

Convolution Layer. The slope of Leaky ReLU and the probability of Dropout are both 0.2.

Model training

The training loss is composed of the generator loss and the discriminator loss, respectively denoted by $\mathcal{L}_G$ and $\mathcal{L}_D$. In the generator loss $\mathcal{L}_G$, four main parts are involved, namely triplet loss on subject $\mathcal{L}_{trip\_s}$, triplet loss on gesture pattern $\mathcal{L}_{trip\_p}$, self reconstruction loss $\mathcal{L}_{recon}$ and cross reconstruction loss $\mathcal{L}_{corss\_recon}$. During training, the generator and the discriminator are updated by $\mathcal{L}_G$ and $\mathcal{L}_D$ in turn independently.

The triplet losses of subject identity can minimize the distance between samples from the same subject in the latent space, and maximize that between samples from the different subjects. It can be described as following formulas:

$$\mathcal{L}_{trip\_s} = \mathbb{E}[\|E_s(x_{i,j}) - E_s(x_{i,l})\| - \|E_s(x_{i,j}) - E_s(x_{m,k})\| + \alpha]_+ \tag{1}$$

Accordingly, the triplet losses of gesture pattern can cluster the samples of the same gesture in the latent space as close as possible, and scatter those of different gestures as far away as possible. It can be described as following formulas:

$$\mathcal{L}_{trip\_p} = \mathbb{E}[\|E_p(x_{i,j}) - E_p(x_{l,j})\| - \|E_p(x_{i,j}) - E_p(x_{m,k})\| + \alpha]_+ \tag{2}$$

To ensure that the disentangled components can reconstruct the original signal, we add two reconstruction loss terms to the entire training loss, namely $\mathcal{L}_{recon}$ and $\mathcal{L}_{corss\_recon}$. The former encourages the subject-specific and pattern-specific components extracted from the real sample to be reconstructed as close to itself as possible. The latter utilizes the subject-specific and pattern-specific components from two different samples to reconstruct a real sample, enhancing the independence between the two components. They are respectively formulated as:

$$\mathcal{L}_{recon} = \mathbb{E}[\|D(E_p(x_{i,j}), E_s(x_{i,j})) - x_{i,j}\|] \tag{3}$$

$$\mathcal{L}_{cross\_recon} = \mathbb{E}[\|D(E_s(x_{i,l}), E_p(x_{m,j})) - x_{i,j}\|] \tag{4}$$

Eventually, we obtain the total loss of the generator by summing the above four terms:

$$\mathcal{L}_G = \mathcal{L}_{recon} + \mathcal{L}_{cross\_recon} + \lambda_1\mathcal{L}_{trip\_p} + \lambda_2\mathcal{L}_{trip\_s} \tag{5}$$

To balance the contribution of different parts in training, we multiply L_{trip_p} and L_{trip_s} by two balance weights, respectively termed as λ_1 and λ_2. Considering that the subject-specific and pattern-specific components are of equal importance in this study, λ_1 and λ_2 are both set to 0.5.

To distinguish real samples and reconstructed samples, we use x_n and $y_n \in 0, 1$ to denote the sample and its label, where the sample is real if $y_n = 1$ and is reconstructed if else. $n \in \{1, N_s\}$ denotes the index of the sample, where N denotes the number of all samples. Therefore, the loss of the discriminator can be described as:

$$\mathcal{L}_D = -[y_n \cdot log(Dis(x_n)) + (1 - y_n) \cdot log(1 - Dis(x_n))] \tag{6}$$

Validation protocols

For the validation, we used different sEMG measures as the model input. Considering that the proposed model is a generic model oriented to cross-subject scenarios, we trained the model on data from 15 of the 20 subjects, and tested it on data from the rest. Accordingly, a five-fold cross-validation was conducted on the 20 subjects.

Since the final purpose of this study is to identify the user's gesture, we employed the recognition accuracy to evaluate the effectiveness of the extracted pattern-specific components. Additionally, to get a more concrete understanding of the pattern-specific components extracted from different types of inputs, we visualized them in the form of 2-D heatmap based on the topological position of electrode arrays. In this way, we further compared the similarities and differences of pattern-specific components from different sEMG measures and different hand gestures.

Gesture recognition accuracy

For gesture recognition accuracy, we directly input the components extracted by the pattern-specific encoder into three typical classifiers, namely Support Vector Machine with linear kernel (SVM), K-Nearest Neighbor (KNN) and Random Forest (RF). The classification accuracy of 10 gestures were recorded. These three classifiers

Table 2 Gesture recognition accuracy of different sEMG measures

Input		KNN (%)	SVM (%)	RF (%)
Waveform	Raw	52.15 ± 12.37	52.61 ± 11.57	50.96 ± 12.57
	Envelope	67.52 ± 13.27	66.94 ± 14.15	66.10 ± 14.69
Frequency-Domain	STFT	78.33 ± 11.35	79.41 ± 10.13	77.52 ± 10.55
Time-Domain	SSC	62.44 ± 15.49	62.42 ± 16.27	61.96 ± 15.04
	RMS	74.32 ± 13.23	74.41 ± 13.08	73.02 ± 13.77
	WL	76.36 ± 15.53	76.23 ± 14.82	75.01 ± 15.86
	ZC	57.25 ± 14.96	58.45 ± 13.68	57.31 ± 15.04
	ALL	81.84 ± 12.06	82.51 ± 11.82	81.20 ± 11.99

Table 3 Gesture recognition accuracy of combining frequency-domain and time-domain measures at different layers

Concatenate layer	KNN (%)	SVM (%)	RF (%)
Encoder	80.38 ± 10.41	81.35 ± 10.21	78.81 ± 9.77
Decoder	79.18 ± 12.44	80.02 ± 11.97	76.20 ± 15.00
Classifier	81.80 ± 11.34	84.30 ± 10.42	80.99 ± 11.25

are chosen for that they do not perform additional non-linear transformations on the extracted components, thus making the recognition accuracy more dependent on the quality of pattern-specific components, instead of the classifiers.

Gesture pattern visualization

For gesture pattern visualization, we reconstructed the pattern-specific components into $d \times 16 \times 16$, then plotted its 16×16 heatmap according to the average value of all feature dimensions. For reconstruction, the pattern-specific components matrix was concatenated with an all-zero matrix shaped like the subject-specific components, and then processed by the decoder. The output can be considered as the sEMG features with only pattern-related information, arranged in a way consistent with the topological position of electrode arrays. Therefore, we can find the muscle groups capturing the model attention with the highlighted areas of the heatmap. In this way, we can further explore the relationship between the model attention area and the muscle activation pattern, providing neurophysiological interpretation for the pattern-specific components extracted by the model.

Correlation coefficient

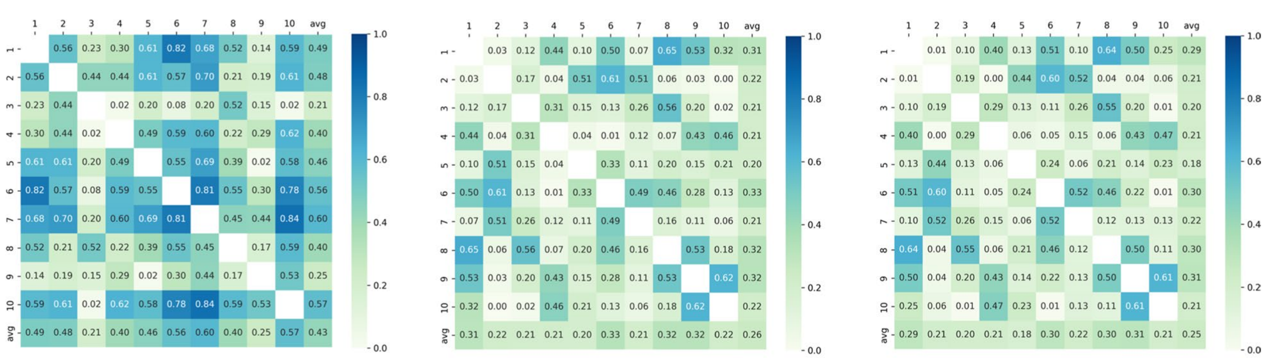
To further quantify the correlation of different gesture patterns after decoding, we calculated the correlation coefficients between the reconstructed heatmaps of one gesture and that of the others in pair. According to the gesture recognition accuracy (shown in Tables 2 and 3), we selected three best-performing EMG measures for result presentation (Fig. 4).

Ablation experiment

To evaluate the impact of GAN on the performance of the disentanglement model, we conducted an ablation experiment, training and testing the model without GAN. In the ablation experiment, the loss function of the model only have one item $\mathcal{L}_G$.

Statistical analysis

We conducted the Shapiro-Wilk test on the accuracies of all measures. The results showed that the accuracy values did not follow Gaussian distribution. Therefore, we employed non-parametric tests for statistical analysis. Specifically, the Wilcoxon signed-rank test was selected



(a) Four time-domain features. (b) STFT. (c) STFT and four time-domain features.

Fig. 4 Correlation coefficient between gestures for different EMG measures. Each row/column in the figure corresponds to the number of each gesture, with the last row/column representing the mean correlation coefficient of that row/column as the statistical method for comparisons in this study. A significant difference of $p < 0.05$ was used in this study.

Results

Gesture recognition accuracy

Table 2 presents the gesture recognition accuracy of the pattern-specific components extracted from different measures. Three representative linear classifiers were employed for performance comparison. As illustrated in Table 2, the pattern-specific components disentangled from raw sEMG signals showed the poorest performance in gesture recognition. In contrast, sEMG envelope has improved the classification accuracy by 14.95% on average. The combination of four classical time-domain features reached the highest classification accuracy of 82.51%, outstanding in the gesture classification task (with $p < 0.05$ for almost all single time-domain or waveform measures). STFT, as a frequency-domain feature, achieved a maximum accuracy of 79.41% when using signal sEMG measure, only 3% lower than the four time-domain feature combination. This results indicated that STFT and the combination of time-domain features are two equivalent measures with none significant difference ($p > 0.05$). Furthermore, for each time-domain features, the performance of RMS and WL was much higher than that of SSC and ZC ($p < 0.05$), exhibiting distinct variability across features.

Table 3 shows the gesture recognition accuracy of combining STFT and the four time-domain measures at different layers of the model. On average, the pattern-specific components extracted by different encoders independently and combined before input into the classifiers achieved the best performance. However, there was no significant difference between the recognition accuracies with different combination layers ($p > 0.05$). In the case of concatenating time-domain and frequency-domain features as EMG measure, the recognition accuracy was between using only frequency-domain feature and only time-domain features. Additionally, the recognition accuracies for the three types of sEMG measures did not show significant difference ($p > 0.05$).

Gesture pattern visualization

Fig. 5 shows the reconstructed 16×16 heatmap of the pattern-specific components extracted from different sEMG measures. To compare the characteristics of the pattern-specific components across subjects, we presented the heatmap of two representative subjects. In general, the heatmap of the pattern-specific components presents similarities between different users and differences between different gestures.

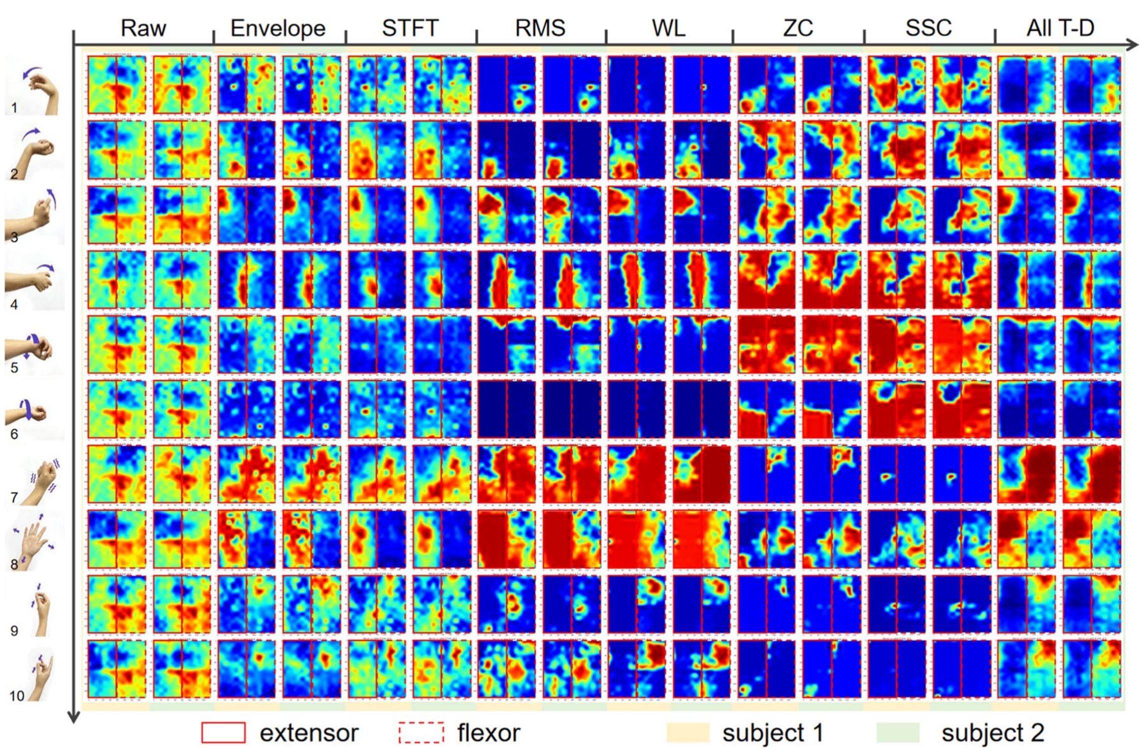


Fig. 5 The reconstructed heatmap of disentangled pattern-specific components for 10 gestures from two representative subjects. T-D is the abbreviation for time-domain

Table 4 Gesture recognition accuracy of disentangled pattern-specific components extracted by models with and without GAN

Inputs		All T-D (%)	STFT (%)	All T-D + STFT (%)
KNN	With GAN	81.84 ± 12.06	78.33 ± 11.35	80.38 ± 10.41
	Without GAN	77.15 ± 14.24	72.99 ± 12.29	74.12 ± 12.42
SVM	With GAN	82.51 ± 11.82	79.41 ± 10.13	81.35 ± 10.21
	Without GAN	78.75 ± 13.23	74.03 ± 12.47	75.06 ± 12.86
RF	With GAN	81.20 ± 11.99	77.52 ± 10.55	78.81 ± 9.77
	Without GAN	75.57 ± 14.85	71.60 ± 13.88	72.81 ± 12.99

More specifically, the heatmap reconstructed by different EMG measures presented different characteristics. For raw signal, the contrast between different areas of the heatmap across gestures is relatively low. In contrast, the heatmaps from frequency domain and time-domain features show significant differences between different gestures. Moreover, their highlighted areas are more compactly clustered.

For the selected time-domain features, those directly reflecting amplitude information, such as RMS and WL, exhibited similar patterns. By contrast, their patterns differed significantly from those of features that do not directly reflect amplitude information, such as ZC and SSC.

Correlation coefficient

Fig. 4 illustrates the correlation coefficient between the heatmaps of 10 gestures, reconstructed by pattern-specific components, in pairs. A smaller correlation coefficient value indicates that the sEMG feature pattern generated by one gesture is less similar with that by other gestures, making that gesture easier to be recognized. Overall, when concatenating STFT and time-domain features as EMG measure, the correlation coefficient values between gestures were the smallest. For using STFT only, the values were comparable. However, when using only time-domain features, the average correlation coefficient value increased distinctly from 0.25 to 0.43.

Comparison between models with and without GAN

To evaluate the impact of GAN on the proposed model, we compared the heatmaps and recognition accuracy of the pattern-specific components disentangled by models with and without GAN. Considering that the only STFT, the combination of four time-domain features, and the combination time-domain features and STFT (combined before encoder) performed significantly better than other measures ($p < 0.05$), we selected the three types of measures for further comparison in this ablation experiment. As shown in Fig. 6, the highlighted areas in the heatmaps of pattern-specific components extracted by the model with GAN are more concentrated than that without GAN. Additionally, with the inclusion of GAN, the pattern-specific differences between different gestures are more pronounced, which helps the classifier better recognize different patterns. Accordingly, Table 4 shows the recognition accuracies of pattern-specific components extracted by models with and without GAN. The recognition accuracy of the model with GAN is significantly higher than that without GAN for STFT and the combination time-domain features and STFT ($p < 0.05$), which is consistent with the results shown in the heatmaps.

Discussion

Previous research found that sEMG-based cross-subject gesture recognition models are still effective when applied to new users, although the accuracy moderately decreases [10]. Based on this conclusion, we innovatively hypothesize that two disentangled components exist in EMG signals, namely pattern-specific and subject-specific components. On one hand, the pattern-specific components indicate that different users produce a large amount of similar EMG signals when performing the same gesture tasks, which explains why cross-subject models can work when applied to a new user. Previous neurophysiological studies using high-density EMG have also confirmed this conclusion, showing that a wide range of populations generate similar spatial features of EMG signals when performing the same gesture tasks [24, 25]. On the other hand, the subject-specific components represent the variations in EMG signals produced by different users due to differences in individual neuromuscular structures, personal exertion habits, and signal acquisition environments (such as noise levels or electrode placement), even when performing the same gesture tasks. This provides the reason why the precision of cross-user models decreases when applied to new users.

Therefore, our hypothesis that EMG signals contain pattern-specific and subject-specific components is well established. Further exploration reveals that these two components are orthogonal to each other. The encoder-decoder-based architecture in deep learning is extremely suitable for decoupling two orthogonal components. Accordingly, the overall network architecture is naturally proposed based on these two components. The encoder that encodes the pattern-specific components to cluster EMG signal samples generated by different users performing the same gesture as closely as possible, while preserving samples generated by different gestures as far apart as possible. Conversely, the encoder that encodes the subject-specific components to cluster EMG signal samples generated by the same user performing different gestures as closely as possible, while preserving samples from different users performing gestures as far apart as possible. Additionally, the decoder ensures that the two disentangled components can be reconstructed back into the original EMG measures, further guaranteeing the correctness of the disentangled information. In the disentanglement model used in this study, we further integrated GAN into the entire model. The presence of the adversarial network forces the reconstructed EMG measures closer to the original EMG measures, improving the gesture recognition accuracy of the model.

In this study, we compared the effects of three different categories of EMG measures on the disentanglement effect. For each single measure, we found that the

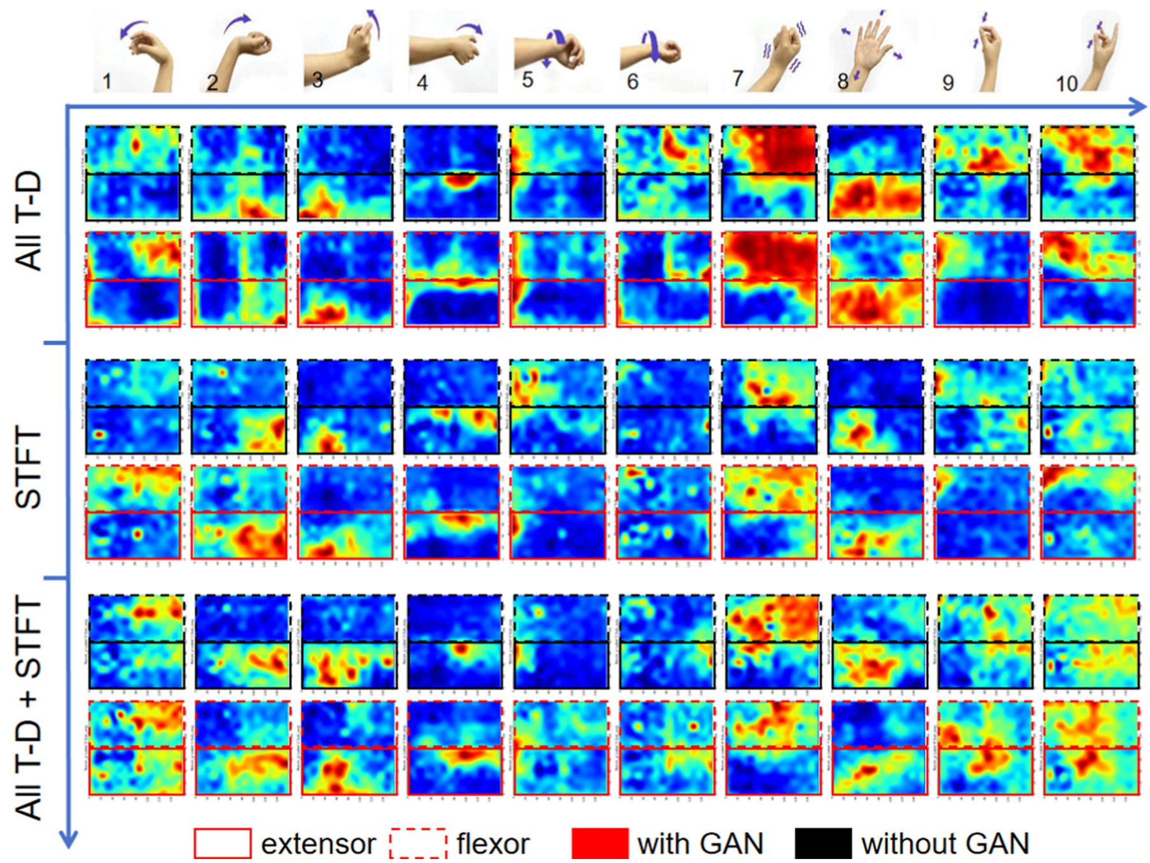


Fig. 6 The reconstructed heatmap of disentangled pattern-specific components for 10 gestures extracted by models with and without GAN. T-D is the abbreviation for time-domain. Note that each heatmap is the average of that from 20 subjects

frequency domain measure STFT had the best performance, followed by time-domain measures, and waveform information the worst. First, the poor performance of the original waveform may be due to the excessive details in the original waveform, making it difficult to fully reconstruct such detailed signals with the sample size of only 20 subjects. Additionally, since the sEMG is a colored Gaussian process, their waveforms exhibit a certain randomness at the macro level, further increasing the difficulty of disentanglement. Therefore, the entire model training may be underfitting. Smoothing the signals by extracting the EMG envelope can significantly improve the disentanglement effect, yet it is still not ideal. Second, STFT yielded the highest recognition accuracy. The inspiration for establishing the disentanglement model in this work originates from style transfer learning [26] in the field of computer vision. In the image recognition task, the same content or object but with different painting styles in images still needs to be identically recognized [27, 28]. Similarly, the pattern-specific and subject-specific components in EMG can be regarded respectively as "content" and "style" in EMG. In the field of computer vision, recent studies have found that content and style components in frequency domain information are much easier to be orthogonalized [29]. By comparing different EMG measures, we found that neurophysiological signals show similar conclusions as images. Thirdly, the overall performance of time-domain measures was slightly lower than STFT, but the performance of gesture recognition had a large gap across different measures. We found that RMS and WL, which directly represent sEMG amplitude feature, performed the best. These two measures are also the most commonly used and intuitive EMG feature metrics. By contrast, SSC and ZC, which cannot directly reflect the feature of sEMG amplitude, did not perform well. However, when all time-domain measures were combined together as the model inputs, the performance of the disentanglement model significantly improved, indicating that SSC and ZC can complement the information for common amplitude measures, although they provide limited information individually.

Interestingly, concatenating time-domain and frequency-domain information did not significantly improve performance, except that when concatenated them before the classifier using SVM. This might be because the sample size of STFT frequency domain features is much larger than that of time-domain features, leading to a feature extraction bias mostly depending on STFT measures. The time-domain features may have little impact on the final gesture recognition accuracy, supplementing limited information. In addition, combining the two measures before the encoder or decoder shares the same network for feature extraction, which further causes the features extracted from the frequency and time-domain measures to be more similar. However, when concatenating them before the classifier, the independent encoders and decoders ensure relatively sufficient extraction of both time-domain and frequency-domain information, thus contributing to a considerable improvement in the performance when using SVM as the classifier.

This study aims to use the disentanglement model to extract pattern-specific components to improve cross-subject gesture recognition accuracy. This requires the model to ensure that the EMG feature heatmaps generated by different gestures are as distinct as possible, while the heatmap for each gesture is as clustered as possible. By visualizing the heatmaps generated with different sEMG measures, we found that STFT strikes a balance between these two aspects, achieving the highest recognition accuracy when using one single feaure. Although time-domain features differentiate well between different gestures on the heatmap, the heatmaps of many gestures (e.g., gesture No. 5, 6, 7, and 8) using one time-domain feature covers a large area, which nearly occupies the entire space. This resulted in high correlation coefficients between these gestures and other gestures, making them difficult to be distinguished and thus leading to a decline in accuracy. When combining all the time-domain features, this issue was substantially relieved through observing the heatmaps shown in Fig. 5. In addition, the heatmap region of feature extraction using disentanglement model is very close to the activation area of muscle contraction when performing the same hand gestures [24, 25]. However, the heatmaps generated by different gestures using the original waveform are very similar, yielding the poor recognition performance. The classifiers used in this study were the simplest (e.g., KNN, SVM, and RF), as the primary focus was to investigate the effectiveness of the disentanglement model in extracting features of the pattern-specific and subject-specific components. Under the same database and classifier conditions, the classification accuracy has greatly improved compared to our previous studies, which only extracted common handcrafted features or used CNN networks for feature extraction [10]. Therefore, exploring the use of more complex classifiers to further improve gesture recognition accuracy is beyond the scope of this study.

In our study, we explored the effectiveness of different sEMG measures in extracting pattern-recognition components through a disentanglement model. We have concluded that time-frequency domain features, such as STFT, outperform traditional amplitude features for gesture recognition tasks. It is noteworthy that in certain EMG applications, such as prosthetic control [30], proportional control is more commonly used than gesture recognition. In proportional control, regulating the amplitude of prosthetic movement is essential. Traditional amplitude features have intrinsic advantages, as they are linearly related to prosthetic amplitude, where time-frequency domain features lack this linear correlation. Therefore, when applying the disentanglement model proposed in this study to proportional control, combining different sEMG measures as control inputs is important. For example, pattern-specific components of traditional amplitude features can be directly used to control the amplitude of prosthetic movement. In contrast, the frequency domain features are not simply linearly correlated with muscle contraction levels, making them difficult to be directly applied in proportional control. However, neural networks or highly nonlinear disentanglement decoders may be able to effectively extract the information embedded in the frequency domain features. For instance, Fig. 5 shows that time-frequency domain features provide better physiological interpretability, with disentangled features closely associated with muscle activations. Accordingly, time-frequency domain features can potentially serve as a reference for channel selection, assisting proportional control by selecting channels with higher muscle activity and lower noise levels, thereby improving the accuracy of proportional control. Future research should further investigate how to leverage different sEMG measures in combination with the disentanglement model to enhance the precision of proportional control.

Author contributions

Conceptualization, J.F. and C.D.; methodology, J.F. and Y.Y.; software, J.F. and Y.Y.; validation, all authors; formal analysis, Y.Y. and J.L.; investigation, Y.Y. and J.L.; resources, B.H., J.F. and C.D.; data acquisition, J.F. and C.D.; writing-original draft preparation, Y.Y. and C.D.; writing-review and editing, all authors; visualization, all authors; supervision, B.H. and J.F.; project administration, B.H. and X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by STI 2030 - Major Projects 2022ZD0208900 and the National Natural Science Foundation of China under Grant 62173094.

Availability of data and materials

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The experiment was supervised and approved by the ethics committee of Fudan University (approval number: BE2035). Each participant received detailed information about the procedures and provided their signed informed consent before the experiment.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 16 September 2024 Accepted: 4 December 2024
Published online: 31 December 2024

References

1. Yin R, Wang D, Zhao S, Lou Z, Shen G. Wearable sensors-enabled human-machine interaction systems: from design to application. Adv Funct Mater. 2021;31(11):2008936.
2. Cross ES, Ramsey R. Mind meets machine: Towards a cognitive science of human-machine interactions. Trends Cognit Sci. 2021;25(3):200–12.
3. Sharma T, Sharma KP, Veer K. Decomposition and evaluation of sEMG for hand prostheses control. Measurement. 2021;186: 110102.
4. Guo L, Lu Z, Yao L. Human-machine interaction sensing technology based on hand gesture recognition: a review. IEEE Trans Hum Mach Syst. 2021;51(4):300–9.
5. Moin A, Zhou A, Rahimi A, Menon A, Benatti S, Alexandrov G, Tamakloe S, Ting J, Yamamoto N, Khan Y, et al. A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. Nat Electron. 2021;4(1):54–63.
6. PonPriya P, Priya E. Design and control of prosthetic hand using myoelectric signal. In: 2017 2nd International Conference on Computing and Communications Technologies (ICCCT). IEEE; 2017. pp. 383–7.
7. Fang B, Wang C, Sun F, Chen Z, Shan J, Liu H, Ding W, Liang W. Simultaneous sEMG recognition of gestures and force levels for interaction with prosthetic hand. IEEE Trans Neural Syst Rehabil Eng. 2022;30:2426–36.
8. Shin S, Kang M, Jung J, Kim YT. Development of miniaturized wearable wristband type surface EMG measurement system for biometric authentication. Electronics. 2021;10(8):923.
9. Su H, Kim T-H, Moeinnia H, Kim WS. A 3-D-printed portable EMG wristband for the quantitative detection of finger motion. IEEE Sens J. 2023;23(7):7895–901.
10. Meng L, Jiang X, Liu X, Fan J, Ren H, Guo Y, Diao H, Wang Z, Chen C, Dai C, et al. User-tailored hand gesture recognition system for wearable prosthesis and armband based on surface electromyogram. IEEE Trans Instrum Meas. 2022;71:1–16.
11. Abrams RA, Ziets RJ, Lieber RL, Botte MJ. Anatomy of the radial nerve motor branches in the forearm. J Hand Surg. 1997;22(2):232–7.
12. Wang Z, Wan H, Meng L, Zeng Z, Akay M, Chen C, Chen W. Optimization of inter-subject sEMG-based hand gesture recognition tasks using unsupervised domain adaptation techniques. Biomed Signal Process Control. 2024;92:106086.
13. Liu Y, Peng X, Tan Y, Oyemakinde TT, Wang M, Li G, Li X. A novel unsupervised dynamic feature domain adaptation strategy for cross-individual myoelectric gesture recognition. J Neural Eng. 2024;20(6):066044.
14. Chen X, Li Y, Hu R, Zhang X, Chen X. Hand gesture recognition based on surface electromyography using convolutional neural network with transfer learning method. IEEE J Biomed Health Inform. 2020;25(4):1292–304.
15. Wang K, Chen Y, Zhang Y, Yang X, Hu C. Iterative self-training based domain adaptation for cross-user sEMG gesture recognition. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023.
16. Fan J, Jiang X, Liu X, Meng L, Jia F, Dai C. Surface EMG feature disentanglement for robust pattern recognition. Expert Syst Appl. 2024;237:121224.
17. Jiang X, Liu X, Fan J, Ye X, Dai C, Clancy EA, Akay M, Chen W. Open access dataset, toolbox and benchmark processing results of high-density surface electromyogram recordings. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1035–46.
18. Kumar S, Veer K, Kumar S. Current trends in feature extraction and classification methodologies of biomedical signals. Curr Med Imaging. 2024;20(1):090323214502.
19. Griffin D, Lim J. Signal estimation from modified short-time Fourier transform. IEEE Trans Acoust Speech Signal Process. 1984;32(2):236–43.
20. Li J, Jiang X, Liu X, Jia F, Dai C. Optimizing the feature set and electrode configuration of high-density electromyogram via interpretable deep forest. Biomed Signal Process Control. 2024;87:105445.
21. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. Adv Neural Inform Process Syst. 2014;27.
22. Aberman K, Wu R, Lischinski D, Chen B, Cohen-Or D. Learning character-agnostic motion for motion retargeting in 2D. arXiv preprint arXiv: 1905.01680. 2019.
23. Gu X, Guo Y, Deligianni F, Lo B, Yang G-Z. Cross-subject and cross-modal transfer for generalized abnormal gait pattern recognition. IEEE Trans Neural Netw Learn Syst. 2020;32(2):546–60.
24. Hu X, Suresh NL, Xue C, Rymer WZ. Extracting extensor digitorum communis activation patterns using high-density surface electromyography. Front Physiol. 2015;6:279.
25. Dai C, Hu X. Extracting and classifying spatial muscle activation patterns in forearm flexor muscles using high-density electromyogram recordings. Int J Neural Syst. 2019;29(01):1850025.
26. Gupta V, Sadana R, Moudgil S. Image style transfer using convolutional neural networks based on transfer learning. Int J Comput Syst Eng. 2019;5(1):53–60.
27. Deng Y, Tang F, Dong W, Ma C, Pan X, Wang L, Xu C. Stytr2: image style transfer with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022. pp. 11326–36.
28. Kwon G, Ye JC. Clipstyler: image style transfer with a single text condition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022. pp. 18062–71.
29. Yoo J, Uh Y, Chun S, Kang B, Ha J-W. Photorealistic style transfer via wavelet transforms. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. pp. 9036–45.
30. Veer K, Vig R. Comparison of surface electromyogram signal for prosthetic applications. Curr Signal Transduct Therapy. 2018;13(2):168–72.

Publisher's Note