

Reduce the User Burden of Multiuser Myoelectric Interface via Few-Shot Domain Adaptation

Bo Xue¹, Le Wu¹, *Member, IEEE*, Aiping Liu¹, *Member, IEEE*, Xu Zhang¹, *Member, IEEE*, Xiang Chen¹, and Xun Chen¹, *Senior Member, IEEE*

Abstract—Due to physiological and anatomical variations across users, myoelectric interfaces trained by multiple users cannot be adapted to the unique hand movement patterns of the new user. Most current work requires the new user to provide one or more trials per gesture (dozens to hundreds of samples), applying domain adaptation methods to calibrate the model and achieve promising movement recognition performance. However, the user burden associated with time-consuming electromyography signal acquisition and annotation is a key factor hindering the practical application of myoelectric control. As shown in this work, once the number of calibration samples is reduced, the performance of previous cross-user myoelectric interfaces will degrade due to the lack of enough statistics to characterize the distributions. In this paper, a few-shot supervised domain adaptation (FSSDA) framework is proposed to address this issue. It aligns the distributions of different domains by calculating the distribution distances of point-wise surrogates. Specifically, we introduce a positive-negative pair distance loss to find a shared embedding subspace where each scarce sample from the new user will be closer to the positive samples and away from the negative samples of multiple users. Thus, FSSDA allows every target domain sample to be paired with all source domain samples and optimizes the feature distance between each target domain sample and the source domain samples within the same batch, instead of direct estimation of the data distribution of the target domain. The proposed method is validated on two high-density EMG datasets, which achieves the averaged recognition accuracies of 97.59% and 82.78% with only 5 samples per gesture. In addition, FSSDA is also effective even when only one

sample per gesture is provided. The experimental results show that FSSDA greatly reduces the user burden and further facilitates the development of myoelectric pattern recognition techniques.

Index Terms—Electromyography (EMG), myoelectric pattern recognition, gesture recognition, few-shot domain adaptation.

I. INTRODUCTION

HAND movement recognition based on the surface electromyogram (sEMG) signals is a practical technique since it is easy to use and non-invasive [1]. The myoelectric interface enables interaction with the user and external devices and helps both healthy and disabled people with their everyday activities. The external devices consist of the power prostheses [2], rehabilitation robots [3] and Virtual Reality (VR) gaming interfaces [4]. As a key technology in myoelectric control, myoelectric pattern recognition allows dexterous movement of multiple degrees of freedom [5]. Typically, after collecting electromyogram (EMG) data from the user using surface electrodes, the hand motion recognition model can be configured using myoelectric pattern recognition techniques. Following the guided training protocol process, high gesture recognition accuracy could be achieved in a controlled laboratory environment.

However, there still exists a gap between the research studies and clinical implementations due to non-ideal factors, such as heavy user burden and the variability in the characteristics of sEMG signals [6]. The new user is required to perform the calibration process to train the classifier before using the prosthetic device, which is time-consuming and may take several days in practice. Additionally, the EMG signal also has a user-dependent nature, a factor that leads to differences in the signal measured when different users perform the same movement, despite the electrodes being worn in the same position [7]. This property is attributed to physiological and anatomical variations across users, e.g. fiber composition, skin resistance, muscle geometry, and fat content [8].

Many efforts are devoted to designing a cross-user model that is grounded in the background data of other users, with minimal effort for accommodating changes in the EMG signal characteristics from the new user [9], [10], [11]. Khushaba tried to find the correlation relationship between the pooled users and the new user [7]. All users' features were projected to a unified-style space associated with the expert set via canonical correlation analysis (CCA). The mapping from the

Manuscript received 21 September 2022; revised 27 November 2022 and 11 January 2023; accepted 12 January 2023. Date of publication 16 January 2023; date of current version 3 February 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFF0501601; in part by the National Natural Science Foundation of China under Grant 32271431, Grant 82272070, and Grant 61922075; in part by the Fundamental Research Funds for the Central Universities under Grant KY2100000123 and Grant WK2100000028; in part by the China Postdoctoral Science Foundation under Grant 2022M723055; and in part by the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-025. (Corresponding authors: Le Wu; Xun Chen.)

Bo Xue, Le Wu, Aiping Liu, Xu Zhang, and Xiang Chen are with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: lewu@ustc.edu.cn).

Xun Chen is with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China, and also with the Institute of Dataspace, Hefei Comprehensive National Science Center, Hefei 230000, China (e-mail: xunchen@ustc.edu.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3237181

new user to the expert set was learned with one trial of each gesture. This framework can overcome individual differences, achieving accuracies of $>83\%$ across multiple users. On this basis, Xue et al. [12] proposed a framework called CCA-OT (canonical correlation analysis-optimal transport) to further reduce the discrepancy in data distribution between the new user and other users via optimal transport. They achieved 8.49% improvement on thirteen Chinese sign language hand movements performed by ten intact-limb subjects compared to the CCA-only framework.

In recent years, convolutional neural networks (CNN) have been widely used in gesture recognition due to the development of deep learning techniques [13], [14], [15]. Atzori et al. have verified CNN with a simple architecture can achieve superior classification performance than conventional methods [16]. However, it is not very effective to use CNNs directly to solve cross-user gesture recognition problems. Therefore, some domain adaptation methods such as multi-stream Adaptive Batch Normalization (MS-AdaBN) [17], supportive model selection [18] and adversarial training [11] were combined with CNN to improve cross-user performance. Domain adaptation (DA) is a technique that uses labeled data from an associated source domain to solve the new task in the target domain [19]. DA is divided into supervised DA (SDA) and unsupervised DA (UDA) according to whether the target domain data is labeled or not. For the supervised DA, Kim et al. [18] firstly selected supportive CNN classifiers pretrained by several subjects, and then fine-tuned the classifiers with the first trial per gesture of the new subject. The final decision was based on the output that was most commonly classified by all supportive CNN classifiers. Campbell et al. [11] combined the adaptive batch normalization and domain adversarial training to CNN, which outperformed CCA in the single-repetition calibration cross-user setting. While the above studies have achieved great success for cross-user gesture recognition problems, they still require the new subject to provide one or more trials per gesture and hold one trial for 3-5 s, to achieve adaptation to the model. With an increasing number of gestures to be performed, the signal collection time has to be extended. What's worse, as a result of the non-stationary and sensitive nature of sEMG, the classifiers, as well as the prostheses need to be recalibrated every day to achieve optimal performance, which may become a heavy usage burden to the end-user [20]. Recently, Zhang et al. [21] attempted to cope with the cross-user gesture recognition problem using UDA method and achieved promising performance. They proposed a method to continuously update the parameters of the classifier by incorporating a self-guided adaptive sampling strategy for online learning. After several iterations, the performance of the classifier will gradually reach stability and realize a plug-and-play gestural interface. However, in the first few iterations, the classifier also needs a few hundred samples from the new user to align the margin distribution of the two domains. Therefore, the end-user has to put up with the poor performance of the model at the beginning of its use. Typically each trial is divided into dozens to hundreds of samples after the sliding window, and it would further reduce the usage burden on the new

user if only several samples per gesture were used to complete the calibration. Previous methods relied on calculating distances and similarities between distributions when learning differences among data distributions for the new user and multiple users, and those were difficult to represent with as few as one sample per gesture. The need to design a cross-user gesture recognition model that relies on less training data while attaining high performance has not been met.

In the few-shot scenario, i.e. when the size of the new user's samples is small, alignment of the distribution of the source domain and target domain becomes a challenge. To tackle this problem, inspired by [22], we introduce a few-shot supervised domain adaptation (FSSDA) framework based on positive-negative pair distance loss (P-N pair distance loss). This loss aims to learn a shared feature embedding so that each new user's sample will be closer to the positive samples and away from the negative samples of source domain. Our proposed model can adequately train each scarce target domain sample by pairing it with the source domain samples within the same minibatch to facilitate the classification of new user's samples. The strength of the point-wise surrogates approach is that it even allows a labeled target domain sample to be paired with all source domain samples, efficiently aligning the data distributions across domains. In this study, we explore whether the proposed multiuser myoelectric interface can rapidly adapt to the new user using background data from other users and significantly reduces the usage burden. Therefore, this work is of interest to promote the widespread commercial and clinical use of myoelectric control systems.

The rest of the paper is organized as follows. Section II presents the datasets and details the proposed framework FSSDA. In Section III, the design of experiments and results are included. An in-depth discussion is presented in Section IV. Finally, Section V concludes our work.

II. METHODS

A. Data Preparation

High-density surface electromyogram (HD-sEMG) is measured by a large two-dimensional array with close-spaced electrodes, which increases the possibility of extracting spatial information due to the increased density and coverage of the electrodes [23]. In the field of gesture recognition, combining it with CNN will achieve better performance [14], [15], [24]. Therefore, we conducted our experiments on two high-density EMG datasets. The first was collected by our group [21] and the second was the CapgMyo database collected by another research group [17].

1) *Dataset-1*: The first dataset contains nine intact-limb subjects (referred to as M1-to-M9), including five males and four females, all right-handed and aged between 24 and 35 years. All subjects were asked to execute six gestures, which included little finger extension, middle finger extension, index finger extension, extension of both index and middle fingers, extension of the last three fingers, and wrist extension. Each gesture was performed 10 times and one repetition of a gesture is called a trial. One trial lasted for about 5s at a comfortable medium-force level followed by 3s of rest to avoid muscle fatigue. HD-sEMG signal was collected using 100 mono-polar

channels which were arranged in a 10×10 grid placed on the surface of the extensor carpi ulnaris and the extensor digitorum with a sampling rate of 1,000 Hz. The signal was then processed by a two-stage amplifier with a total gain of 60 dB and filtered by a 20-500 Hz bandpass filter. More details on signal acquisition and electrode configuration were provided in [21].

2) *Dataset-II*: In addition to testing on our dataset, we further utilised the CapgMyo DB-c dataset collected and first used by Du et al. [17]. EMG signals were recorded from ten healthy subjects (referred to as N1-to-N10) ranging in age from 23 to 26 years with an 8×16 electrode array. The array consists of 8 acquisition modules, each containing a matrix (8×2) differential electrode array. The first module was placed on the extensor digitorum communis muscle and the other modules were arranged equidistantly around the surface of the right forearm. The 16-bit analogue-to-digital converter was used to sample the signal at 1000 Hz. The signal was then transferred to the PC via WIFI and filtered by a band-pass filter between 20 and 380 Hz. 12 classes of finger movements relevant to the activities of daily living were carried on during the experiments. Each gesture was required to be held for 3s to 10s and repeated 10 times. The subjects rested for 7s after completing one trial at a time to avoid muscle fatigue. To perfectly match the labels, only the static part of the movement (about 1s) was used. In the experiment, the processed data from N1-N9 were used since the data of N10 could not be downloaded from the website provided in the paper [17].

B. Data Preprocessing

For DATASET-I, in accordance with the typical settings described in the literature [5], [25], the sEMG signal stream was split into a series of overlapping windows (window length: 256 ms, step size: 128 ms). For DATASET-II, Following their earlier work [14], [17], the window length and the step size were set to 150 ms and 50 ms respectively. Then the windows of the resting segments were discarded using the amplitude threshold [26]. The threshold was empirically set to the mean plus three times the standard deviation of the baseline signal averaged over all channels. Thus, for both datasets, the EMG signal of one trial was split into approximately 40 and 20 samples respectively. For each analysis window of the active segment, three features, namely the waveform length from four time-domain feature set [27], f1 and f6 from the time-dependent power spectrum descriptors feature set [28] for each channel were empirically extracted. As a result, the analysis windows for the two datasets are converted into a $10 \times 10 \times 3$ and an $8 \times 16 \times 3$ feature matrix respectively. These processed feature matrices are considered as feature images equivalently, facilitating the training of CNN models [15], [16].

C. SDA With Scarce Target Domain Samples

Suppose that we have two domains: one is the source domain (the training set made of pairs $\mathcal{D}^s = \{(\mathbf{x}_i^s, \mathbf{y}_i^s)\}_{i=1}^M$ representing the samples of existing users), the other is the target domain ($\mathcal{D}^t = \{\mathbf{x}_j^t\}_{j=1}^N$ representing samples of the new

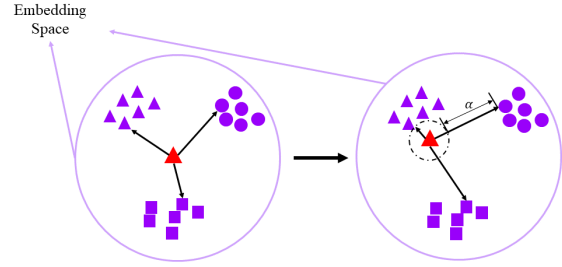


Fig. 1. Illustration of P-N pair distance loss effect on a target domain sample in the embedding space. Different shapes represent different classes. Different colours represent samples from different domains, where purple represents the samples from source domain and red represents the samples from target domain.

user). The features $\mathbf{x}_i^s \in \mathcal{X}^s$ and $\mathbf{x}_j^t \in \mathcal{X}^t$ are realizations of the random variables \mathbf{X}^s and \mathbf{X}^t respectively, while the labels $\mathbf{y}_i^s \in \mathcal{Y}^s$ and $\mathbf{y}_j^t \in \mathcal{Y}^t$ are realizations of the random variables \mathbf{Y}^s and \mathbf{Y}^t respectively. In our issue, the feature space and the label space of two domains are the same, i.e. $\mathcal{X}^s = \mathcal{X}^t, \mathcal{Y}^s = \mathcal{Y}^t$. Furthermore, the target domain \mathcal{D}^t contains partially labeled data $\{\mathbf{x}_k^t\}_{k=1}^{N_l}$ (the calibration set \mathcal{D}^{tl}) and unlabelled data $\{\mathbf{x}_q^t\}_{q=1}^{N_u}$ (the testing set \mathcal{D}^{tu}), i.e. $\mathcal{D}^t = \mathcal{D}^{tl} \cup \mathcal{D}^{tu}$. It is assumed that there is a covariate shift [29] between \mathbf{X}^s and \mathbf{X}^t and thus a difference between the probability distributions $p(\mathbf{X}^s)$ and $p(\mathbf{X}^t)$. SDA is utilized to train a robust classifier $f: \mathcal{X} \rightarrow \mathcal{Y}$ capable of classifying \mathcal{D}^{tu} by using the labeled data from \mathcal{D}^s and \mathcal{D}^{tl} . In this work, our particular concern is that only several labeled samples per class of target domain are available, or even only one sample per class.

D. P-N Pair Distance Loss

Typically, f can be decomposed into two functions, i.e. $f = k \circ h$, where $k: \mathcal{X} \rightarrow \mathcal{Z}$ is a mapping from the input space \mathcal{X} to an embedding feature space \mathcal{Z} , and $h: \mathcal{Z} \rightarrow \mathcal{Y}$ is a function for making predictions. For each sample \mathbf{x}_i^t in the target domain \mathcal{D}^{tl} , we can find a positive sample \mathbf{x}_j^s of the same class as it in the source domain and a negative sample \mathbf{x}_k^s of the different class from it in the source domain. As shown in Fig. 1, to make the distance between the target domain samples and the positive sample (P-pair distance) as small as possible and the distance between them and the negative sample (N-pair distance) as far as possible, positive and negative pair distance loss (P-N pair distance loss) is defined as follows:

$$\begin{aligned} \mathcal{L}_{P-N} = & \sum_i^{n_t} \sum_{j,k}^{n_s} \left[\left(\|\mathbf{k}(\mathbf{x}_i^t) - \mathbf{k}(\mathbf{x}_j^s)\|_F^2 \right. \right. \\ & \left. \left. - \|\mathbf{k}(\mathbf{x}_i^t) - \mathbf{k}(\mathbf{x}_k^s)\|_F^2 + \alpha \right)_+ \right] \\ \text{s.t. } & \mathbf{y}_i^t = \mathbf{y}_j^s \neq \mathbf{y}_k^s \end{aligned} \quad (1)$$

where $[\cdot]_+ = \max(\cdot, 0)$, $\|\cdot\|_F$ denotes the Frobenius norm, n_t and n_s represent the size of target domain samples and source domain samples in a batch, and α is the margin distance of separability between the P-N pairs in the embedding space. It is noted that without considering the sample selection from two domains, P-N pair distance loss becomes the well-known triplet loss [30].

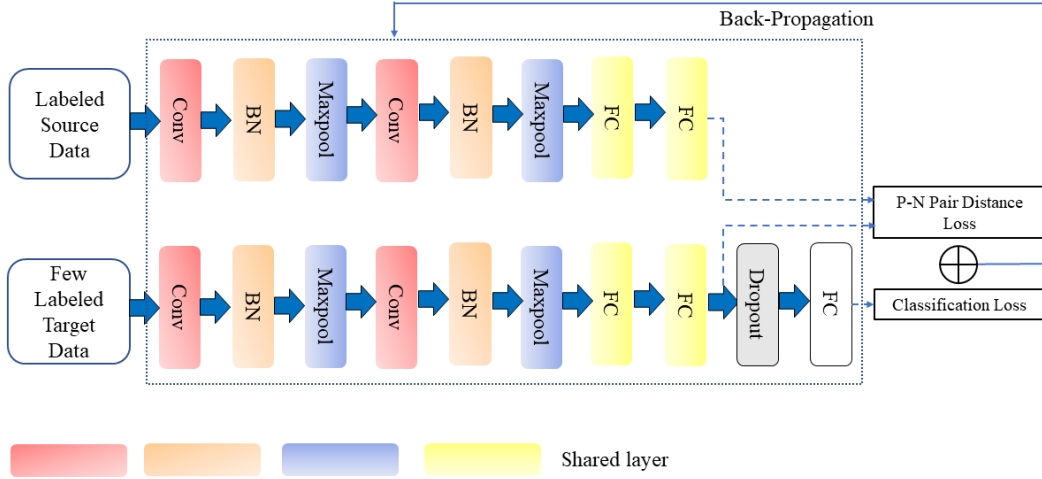


Fig. 2. Illustration of the proposed cross-user model FSSDA.

E. Description of FSSDA Framework

Fig. 2 shows our proposed FSSDA method for cross-user gesture recognition based on CNN. For DATASET-I, the CNN architecture was designed based on Lenet-5 [31], which has been widely used in gesture recognition and shown good performance [15], [21], [32], [33]. Each scarce target domain sample is paired with all source domain samples, resulting in $M * N_t$ pairs. This ensures that the total number of samples in the source and target domain is equal, to facilitate the design of the batch sampler. The sampler then allows for the same number of samples from existing users and the new user in each batch. The training architecture mainly has two streams, one for source domain and the other for target domain. For DATASET-I, a network structure similar to [21] was used. Each stream contains two CNN blocks and two fully connected (FC) layers. Within each CNN block, there is a convolutional layer to extract features, a batch normalization layer to accelerate model convergence and a maximum pooling layer to reduce feature dimensionality while retaining the more important information. The number of filters in the convolutional layers of two convolutional blocks is 64 and 128 respectively, and kernel size is $3*3$ with padding = 1. The next two FC layers contain 128 and 64 units respectively. Except for the last FC layer in the stream of target domain, all the previous layers share the same weight, which can be seen as the embedding function k . A rectified linear unit is equipped after the convolutional layers and the FC layers to avoid vanishing gradient and enhance non-linear fitting capability of the model [34]. A separate FC layer containing 6 units is designed as a classification layer h for the target domain. Additionally, a dropout layer with a probability of 0.5 is adopted before the final FC layer to reduce overfitting over training. The output of h is the predicted probabilities of the gesture categories. The overall loss of FSSDA is as follows:

$$\mathcal{L} = \mathcal{L}_{P-N} + \mathcal{L}_C \quad (2)$$

where \mathcal{L}_{P-N} is defined in (1), and the loss \mathcal{L}_C is the classification loss calculated from the target domain. In this

experiment, this classification loss function is chosen as the CE (cross-entropy) function.

$$\mathcal{L}_C = -\frac{1}{n_t} \sum_{i=1}^{n_t} \sum_{c=1}^C y_{ic}^t \log f(\mathbf{x}_i^t)_c \quad (3)$$

where n_t is the size of the target samples in a batch. C represents the number of categories of gestures. y_{ic}^t represents a binary indicator function: it equals 1 if the sample \mathbf{x}_i^t truly belongs to class c , otherwise it equals 0. $f(\mathbf{x}_i^t)_c$ is the predicted probability that the sample \mathbf{x}_i^t belongs to class c .

At each training iteration, the samples of both domains enter the network at the same time and pass through the same defined layer to achieve the shared weight of these network layers. The P-N pair distance loss is calculated after the second FC layer of both domains. The classification layer for the target domain stream is defined separately and allows the calculation of the CE loss. Finally, the total loss is calculated and the parameters of the model are updated by gradient back propagation. For DATASET-II, the same network structure as [17] is adopted for both streams of the source domain and the target domain, except that the stream of the source domain does not contain a final classification layer. The other parameter settings are the same for both networks except that the batch size is empirically set to 32 and 64 for two networks. Within each batch, the number of samples in two domains is equal (half of batch size). We use Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.99$, $weight_decay = 0.0001$) [35], an epoch number of 20 in all experiments. The learning rate is initially set to 0.0001 and is divided by 10 after the 8th and 16th epochs. The models are implemented in PyTorch 1.12.0. and all experiments are conducted on four GeForce RTX 2080 Ti GPUs.

F. Performance Evaluation and Statistics Analysis

As with other works in multi-user gesture recognition [7], [11], [21], the leave-one-subject-out cross-validation scheme was used to test our proposed framework, which means that one subject was randomly selected as the new user from all subjects and the rest of subjects were utilized as the training

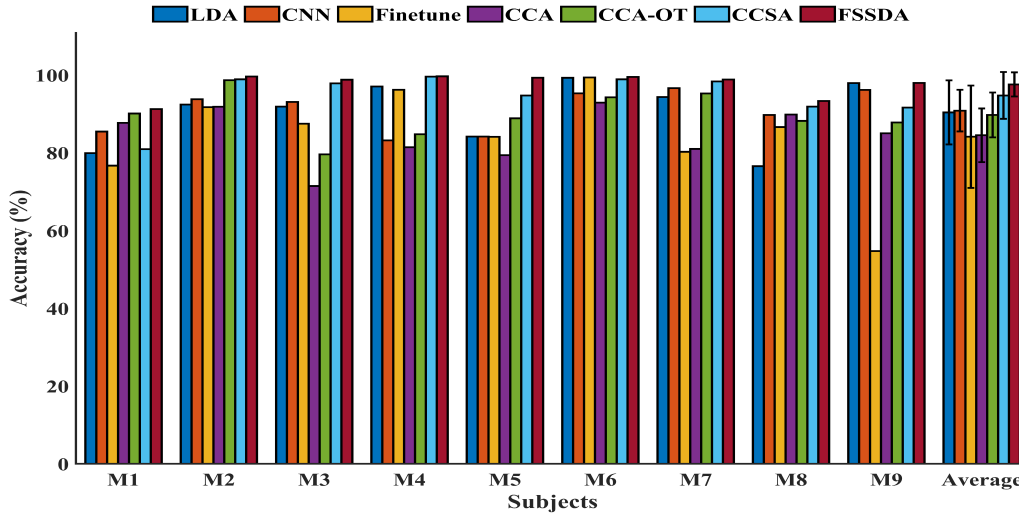


Fig. 3. The classification accuracies using the proposed method and 6 other cross-user gesture recognition methods on DATASET-I.

set until each user was tested once as a new user. For all subjects, the first trial of each gesture was utilized as the calibration set and the rest of trials were utilized as the testing set or training set depending on whether it was from a new user or background users. The performance of the classifier is measured with classification accuracy, which is defined as the ratio between the number of samples correctly classified and the total number of samples for the new user in each cross-validation.

Many cross-user gesture recognition methods were employed for comparison purposes. (1) LDA (linear discriminant analysis) [36]: a classifier that is easy to implement and has efficient performance. (2) CNN: the same network structure as the stream of target domain in the proposed method, without any domain adaptation operations. (3) CCA [7]: the most classic CCA-based cross-user gesture recognition framework. (4) CCA-OT [12]: this framework further reduces the discrepancy in data distribution between the source and target domain by optimal transport based on CCA. (5) Finetune [15]: a technique that uses a few samples of the new user to fine-tune the parameters of the FC layer of a model pretrained by other users. (6) CCSA [22]: an approach to solving the problem of domain adaptation with few samples is used for comparison as it fits better with the problem we are trying to solve. This method uses classification and contrastive semantic alignment loss to achieve the alignment of the distribution between two domains. (7) MS-AdaBN [17]: the first approach to calibrate the cross-user gesture recognition model using UDA. (8) SGDA (self-guided domain adaptation) [21]: an UDA approach with a self-guided adaptive sampling strategy. In this paper, to make a fair comparison, we used the networks adopted in the papers proposing the datasets respectively [17], [21]. The settings of training conditions for the different methods are shown in Table I. For the deep learning methods, other parameters (e.g. the learning rate, the optimizer) are set to be the same as FSSDA. For the traditional methods, the parameters are set according to their previous work.

Additionally, statistical significance is assessed using a one-way repeated-measure analysis of variance (ANOVA) test with a significance value of $p = 0.05$. All statistical analyses were conducted using SPSS (v.24.0, SPSS Inc. Chicago, IL, USA).

III. EXPERIMENTS AND RESULTS

A. Classification Performance of FSSDA on Two EMG Datasets

In the few-shot learning scenario, each new user provides 5 samples of each gesture (1/8 of one trial for DATASET-I, 1/4 of one trial for DATASET-II). It should be noted that the data used for training LDA and CNN is only the calibration samples of the new user, as we have found that this works better than using other users' data or a combination of both. For CCA and CCA-OT, due to the small number of calibration samples, a replication operation was performed on these samples to facilitate the transformation of features into a unified-style space. The other experimental conditions for the seven methods were set equal to make a fair comparison. Fig. 3 and Fig. 4 show the classification accuracies achieved by six methods on two datasets. The accuracies of nine subjects in DATASET-I using our proposed method (FSSDA) and LDA, CNN, Finetune, CCA, CCA-OT, and CCSA were $97.59 \pm 3.10\%$, $90.39 \pm 8.24\%$, $90.83 \pm 5.36\%$, $84.13 \pm 13.16\%$, $84.50 \pm 6.90\%$, $89.73 \pm 5.77\%$, and $94.75 \pm 6.02\%$, respectively. Also, the proposed method (FSSDA), LDA, CNN, Finetune, CCA, CCA-OT, and CCSA, yielded accuracies of $82.78 \pm 5.17\%$, $73.29 \pm 7.87\%$, $72.80 \pm 6.60\%$, $49.41 \pm 3.38\%$, $52.15 \pm 6.97\%$, $56.45 \pm 7.18\%$, and $76.60 \pm 7.70\%$ on DATASET-II, respectively. The results of the comparison with the other two UDA approaches are presented in Table I. The ANOVA revealed the proposed method performed significantly better for cross-user gesture recognition than all other comparison methods on both datasets ($p < 0.05$ for any comparison).

B. Experiments on Different Hyperparameter α

Fig. 5 shows the averaged accuracy of nine subjects in DATASET-I under different α settings. The margin α controls the distance between samples in the target domain and samples

TABLE I
COMPARISON OF TRAINING CONDITIONS AND AVERAGED CLASSIFICATION ACCURACIES (%) OF DIFFERENT METHODS

Methods	Types	number of calibration samples	Batchsize on DATASET-I / DATASET-II	Accuracies (%) on DATASET-I / DATASET-II
Traditional	LDA [36]	-	-	90.39±8.24 / 73.29±7.87
Traditional	CCA [7]	SDA	-	84.50±6.90 / 52.15±6.97
Traditional	CCA-OT [12]	SDA	-	89.73±5.77 / 56.45±7.18
Deep Learning	CNN	5 samples per gesture	16 / 32	90.83±5.36 / 72.80±6.60
Deep Learning	Finetune [15]	SDA	16 / 32	84.13±13.16 / 49.41±3.38
Deep Learning	CCSA [22]	SDA	32 / 64	94.75±6.02 / 76.60±7.70
Deep Learning	FSSDA	SDA	32 / 64	97.59±3.10 / 82.78±5.17
Deep Learning	MS-AdaBN [17]	UDA	100 / 100	67.99±21.06 / 47.64±4.88
Deep Learning	SGDA [5]	UDA	100 / 100	90.40±14.44 / 55.27±5.54

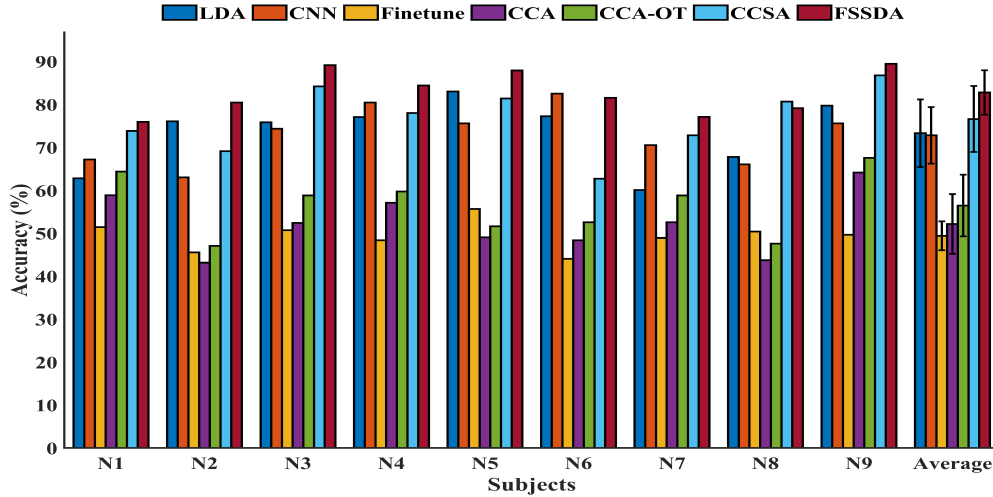


Fig. 4. The classification accuracies using the proposed method and 6 other cross-user gesture recognition methods on DATASET-II.

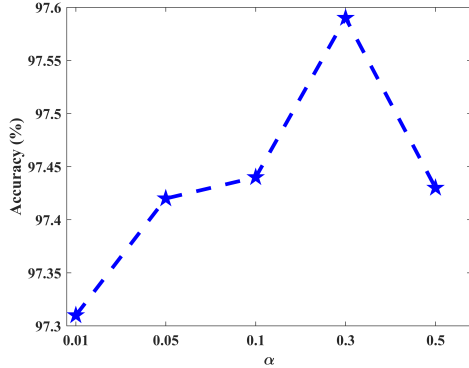


Fig. 5. The impact of the parameters of α on the classification accuracy.

of the same and different categories in the source domain. Note that the accuracy gradually grows as α increases from 0.01 to 0.5, and reaches the best at $\alpha = 0.3$. However, when α is larger (e.g. $\alpha = 0.5$), the accuracy starts to drop. It is due to the fact that larger margins increase the difficulty of feature learning. For DATASET-I, we set α to 0.3. The experimental results on DATASET-II have a similar trend, and we selected the optimal $\alpha = 0.05$ for it.

C. Experiments on Different Numbers of Calibration Samples

The impact of the number of samples utilized in the calibration process on classification accuracy when utilizing different methods was also analyzed. A certain number of samples were randomly selected from the first trial of each gesture performed by the new user. The process was repeated 3 times and the

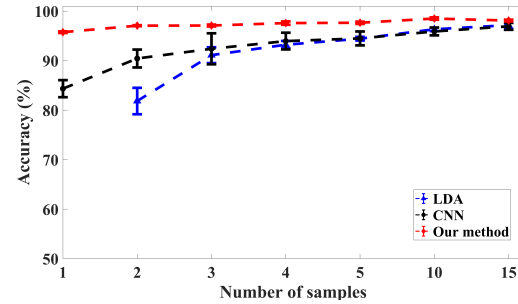


Fig. 6. Effect of the number of samples provided by the subjects on the classification accuracy while using different methods on DATASET-I.

averaged recognition accuracy was calculated using different approaches. Fig. 6 and Fig. 7 show the averaged recognition accuracy of the different methods on both datasets as the number of samples per gesture grows. Additionally, note that as the number of training samples is equal to the number of classes, the LDA classifier cannot be trained at this point, so this curve starts when the number of samples is 2. It can be observed that the performance of LDA and CNN is poor when the number of calibration samples is small, and with increasing numbers of calibration samples, the final results of the three methods are almost equal. Additionally, FSSDA still works well compared to the other two methods even when only one sample is provided for each gesture, achieving 95.72% and 59.46% accuracy on the two datasets respectively.

D. Visualization of Few-Shot Domain Adaptation

Fig. 8 shows the visualization results of embedding features using t-distributed stochastic neighbor embedding (t-SNE)

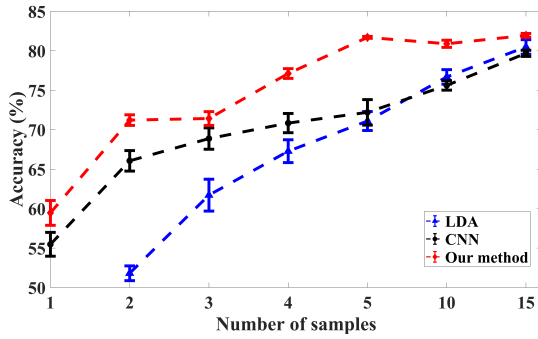


Fig. 7. Effect of the number of samples provided by the subjects on the classification accuracy while using different methods on DATASET-II.

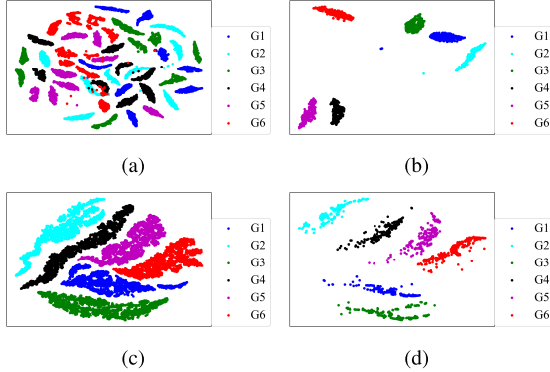


Fig. 8. The t-SNE visualization of sample distributions of six gesture tasks from a representative subject. (a) Source domain before FSSDA (b) Target domain before FSSDA (c) Source domain after FSSDA (d) Target domain after FSSDA. Dots of different colours indicate different patterns of samples.

technology [37]. Fig. 8(a) and Fig. 8(b) demonstrate the visualization of the embedding of high-dimensional features extracted from two domains into a two-dimensional space before using FSSDA. It can be observed that the distribution of samples in the same category varies considerably across users, and some of the samples in different categories even overlap significantly with each other. After applying FSSDA, the distribution of samples in the source (Fig. 8(c)) and target domain (Fig. 8(d)) is close, with samples from the same category close to each other and samples from different categories of different users far away from each other, showing good separability and thus facilitating the classification of the gesture tasks for the new user.

E. Computational Complexity Analysis

The number of parameters (Params), floating point operations (FLOPs) and the running time of FSSDA on two datasets are presented in Table II. The model used in DATASET-II has more convolutional blocks, resulting in a larger number of params than the model used in DATASET-I. We also test the running time by inputting the testing samples of the subjects from two datasets into the models one by one. Therefore, the average processing time of one sample is considered as the running time of FSSDA. Although the running time of the model used in DATASET-II is slightly longer than the model used in DATASET-I, they are both far less than the time requirement (300 ms) of real-time gesture recognition in myoelectric control systems [36].

TABLE II

THE COMPUTATIONAL COMPLEXITY OF FSSDA ON TWO DATASETS			
	Params (M)	FLOPs (G)	Running time (ms / sample)
DATASET-I	0.15	0.0042	0.064±0.0081
DATASET-II	4.57	0.021	0.084±0.012

IV. DISCUSSIONS

Even with the rapid development of EMG pattern recognition technology in recent years, severe user burden is a significant barrier to the development of cross-user gesture recognition technology. Due to the non-stationary and variability in the characteristics of the EMG signal, direct application of classifiers trained with other user's data to the new user is poor, as has been demonstrated in the literature [11], [21]. The difficulty of this problem lies in adapting the cross-user model to learn the user-specific nature of the new user with minimal training data, while taking advantage of the similarity of patterns between different users. In this paper, a cross-user gesture recognition framework FSSDA based on the P-N pair distance loss is proposed to alleviate the usage burden of the end-user by requiring only several samples per gesture. First, we pair the scarce samples of the new user with the sufficient samples of the background users and feed them into the network together, so that there is the same number of both domain samples in each batch. Then, P-N pair distance loss is employed to find a shared embedding space that finds the most appropriate location for each new user's sample, which is close to the positive samples and far from the negative samples in the source domain. Additionally, CE loss is calculated on the samples of the new user to build a gesture classification model more suitable for it. By leveraging the distribution distances of point-wise surrogates for the new user and other users, FSSDA enables adaptive stability, reducing sEMG signal recording time by more than four times, without sacrificing accuracy in gesture recognition.

As demonstrated in Fig. 3 and Fig. 4, in the few-shot scenario, some previous methods did not provide a satisfactory solution for cross-user gesture recognition. The recognition accuracy of CCA, CCA-OT and Finetune is even lower than the recognition accuracy of LDA trained with only a few samples from the new user, dropping by 6.26%, 5.89% and 0.66% on the first dataset and by 23.88%, 21.14% and 16.84% on the second dataset, respectively. CNN achieved similar recognition rates as LDA in this case. In contrast, Finetune achieved the worst results, probably because the number of samples of the new user was too small to allow the network to learn the differences in movement patterns between the new user and other users. Although CCA-OT further reduces the discrepancy between the distribution of two domains on top of CCA, it does not exceed the strong baseline of LDA. Inaccurate statistical information about the data distribution of the new user is calculated, leading to negative migration, i.e. knowledge learned from background users in the source domain can have a negative effect on the prediction of the new user's samples in the target domain. In this case, training an LDA or CNN with only calibration samples would produce better results than a cross-user model with negative migration, although prone to overfitting. Thus, the above approaches

lack sufficient data from the target domain for learning the differences between distributions, and do not make full use of the scarce samples of the new user, resulting in a cross-user model that is not well suited to the new user. Nevertheless, CCSA achieves better results than the previous approach, with improvements of 4.36% and 3.31% over LDA on the two datasets respectively, which employs contrastive loss to bound the relationship between the scarce target samples and large source domain samples for the separation and alignment of the distribution. However, the recognition rate achieved by this method is still lower than that of FSSDA because it requires the P-pair distance = 0, which can lead to model collapse, i.e. all samples of the same category will output the same embedding features, thus affecting the feature representation ability. Instead, P-N pair distance loss is to require that for each sample in the target domain, the distance between it and samples of a different category from source domain is larger than the distance between it and samples of the same category from the source domain, which ensures the distinguishability of the embedding features and avoids the difficulty of model learning caused by overly strong constraints.

A noteworthy point is that our group has recently used unsupervised domain adaptation to address the cross-user problem [21], implementing a plug-and-play myoelectric interface. This approach designs an adaptive domain adaptation model that uses an adaptive sampling strategy to gradually align the margin distribution across users, continuously updating the model until the new user are satisfied. Although this approach does not require the new user to provide labels of the gestures, it still requires several hundred samples to update the model parameters (to ensure gesture diversity) to achieve optimal performance, whereas we only need five samples of each gesture. As shown in Table I, the final recognition accuracies of this method on the two datasets are $90.41 \pm 14.44\%$ and $55.27 \pm 5.54\%$, which are still poorer than our result $97.59 \pm 3.10\%$ and $82.78 \pm 5.17\%$. Therefore, for the second dataset with poorly discriminated gesture samples, the learning of margin distributions is difficult. With a few labeled samples, SDA achieves better cross-user gesture recognition performance than UDA.

Fig. 6 and Fig. 7 also investigate the effect of the number of calibration samples on two classifiers and FSSDA for two datasets. The trend of the curves with FSSDA on the two datasets is similar, except that the curve on the DATASET-II changes more significantly. This is because in DATASET-I, the number of gesture categories is small and the samples are well differentiated, so promising recognition accuracies are achieved when only one calibration sample per gesture is provided. In DATASET-II, the number of gesture categories is large and the samples are poorly differentiated. Therefore, the recognition accuracy is lower when there is one calibration sample per gesture. With each additional calibration data provided, the network learns more accurate classification boundaries and the recognition accuracy improves. Nevertheless, when the number of calibration samples is small, the results of all methods on DATASET-II are not very promising. Although the proposed method achieves the optimal results, no obvious improvement in recognition rate is achieved

when only 2 and 3 calibration samples are provided. This can be attributed to the difficulty of the proposed method to adequately align the distributions of two domains when the input information is limited. However, it is worth noting that although the curves for the two datasets start out with different levels of growth, they can reach stability once the calibration samples reach five. Furthermore, the results demonstrate that increasing the size of calibration samples could gradually enhance the robust performance of the cross-user gesture recognition method. However, it certainly places an increasing burden on the new user to calibrate. The time-consuming and frequent recalibration is one of the reasons for the high rejection rate of myoelectric pattern recognition-based prostheses. Our method achieves stable and good performance with as few as 5 calibration samples on both datasets, which confirms that FSSDA reduces the effect of the number of the calibration samples on the cross-user model performance. FSSDA achieves the best classification results with a short signal collection time, which may have significant implications for the design of prosthetic devices with a low user burden.

The visualization of t-SNE in Fig. 8 further validates the effectiveness of FSSDA. The data distribution of the new user always differs from that of existing users. Using the positive and negative pair relationship between samples in two domains, the samples in the target domain are brought closer together with samples of the same category in the source domain and further apart with samples of a different category in the source domain. We can intuitively observe that the decision boundaries between different gesture patterns in two domains become similar after using FSSDA, indicating that the model learns the inherent correlation patterns between the new user and other users, consistent with the hypothesis that the EMG patterns of specific muscle activation are similar for different users. Overall our proposed method leverages minimal calibration data from the new user and combines additional sources of variability provided by other users to build a cross-user gesture recognition model, which is superior to the previous cross-subject techniques and may be a viable option for improving EMG control based on pattern recognition.

Finally, the current study still has several limitations. This work did not take into account in the experiments the non-ideal factors encountered in practice, such as electrode shift [5] and limb position [38], which may lead to performance degradation in myoelectric control. Another issue is that while FSSDA reduces the size of calibration samples for the new user, it still requires samples for each gesture. If it was possible to calibrate all gestures using a small number of gestures or to use zero retraining as in the cross-day gesture recognition problems [39], [40], this would further reduce the usage burden on the end-user, which would be our future work.

V. CONCLUSION

In this paper, we propose a novel multiuser myoelectric interface FSSDA. This framework could generalize well after observing very few samples per class of the new user. P-N pair distance loss was proposed to perform domain adaptation on the sufficient samples from the source domain and the scarce samples from the target domain, despite the mismatch in

distribution between them. We found that using the point-wise surrogates of distribution distances to solve the problem of the distribution alignment across users was very effective when the number of calibration samples was small, even if there was only one sample per gesture. FSSDA will substantially reduce the heavy usage burden required, reducing the sEMG signal recording time by more than four times. The results on both HD-sEMG datasets verify the impressive performance of FSSDA in comparison to the prior cross-user models in the few-shot scenario. This work has important implications for reducing the usage burden, improving acceptance, and increasing the base of potential users using EMG gesture recognition. This will facilitate the widespread use of EMG control systems in consumer and industrial applications.

REFERENCES

- [1] C. Castellini, A. E. Fiorilla, and G. Sandini, "Multi-subject/daily-life activity EMG-based control of mechanical hands," *J. NeuroEngineering Rehabil.*, vol. 6, no. 1, pp. 1–11, Dec. 2009.
- [2] A. Fougner, Ø. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, Sep. 2012.
- [3] L. Liu, X. Chen, Z. Lu, S. Cao, D. Wu, and X. Zhang, "Development of an EMG-ACC-based upper limb rehabilitation training system," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 3, pp. 244–253, Mar. 2017.
- [4] A. Dash and U. Lahiri, "Design of virtual reality-enabled surface electromyogram-triggered grip exercise platform," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 444–452, Feb. 2020.
- [5] X. Zhang, L. Wu, B. Yu, X. Chen, and X. Chen, "Adaptive calibration of electrode array shifts enables robust myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 7, pp. 1947–1957, Jul. 2020.
- [6] E. Rahimian, S. Zabih, A. Asif, D. Farina, S. F. Atashzar, and A. Mohammadi, "FS-HGR: Few-shot learning for hand gesture recognition via electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1004–1015, 2021.
- [7] R. N. Khushaba, "Correlation analysis of electromyogram signals for multiuser myoelectric interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 745–755, Jul. 2014.
- [8] L. Bi, A. G. Feleke, and C. Guan, "A review on EMG-based motor intention prediction of continuous human upper limb motion for human–robot collaboration," *Biomed. Signal Process. Control*, vol. 51, pp. 113–127, May 2019.
- [9] T. Tommasi, F. Orabona, C. Castellini, and B. Caputo, "Improving control of dexterous hand prostheses using adaptive learning," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 207–219, Feb. 2012.
- [10] 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.
- [11] E. Campbell, A. Phinyomark, and E. Scheme, "Deep cross-user models reduce the training burden in myoelectric control," *Frontiers Neurosci.*, vol. 15, p. 595, May 2021.
- [12] B. Xue et al., "Multiuser gesture recognition using sEMG signals via canonical correlation analysis and optimal transport," *Comput. Biol. Med.*, vol. 130, Mar. 2021, Art. no. 104188.
- [13] K.-H. Park and S.-W. Lee, "Movement intention decoding based on deep learning for multiuser myoelectric interfaces," in *Proc. 4th Int. Winter Conf. Brain-Computer Interface (BCI)*, Feb. 2016, pp. 1–2.
- [14] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, and W. Geng, "A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface," *Pattern Recognit. Lett.*, vol. 119, pp. 131–138, Mar. 2017.
- [15] 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.
- [16] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers Neuroinformatics*, vol. 10, p. 9, Sep. 2016.
- [17] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, p. 458, Mar. 2017.
- [18] K.-T. Kim, C. Guan, and S.-W. Lee, "A subject-transfer framework based on single-trial EMG analysis using convolutional neural networks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 94–103, Jan. 2020.
- [19] W. Mei and D. Weihong, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Jul. 2018.
- [20] X. Zhai, B. Jelfs, R. H. M. Chan, and C. Tin, "Self-re-calibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network," *Frontiers Neurosci.*, vol. 11, p. 379, Jul. 2017.
- [21] X. Zhang, X. Zhang, L. Wu, C. Li, X. Chen, and X. Chen, "Domain adaptation with self-guided adaptive sampling strategy: Feature alignment for cross-user myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 1374–1383, 2022.
- [22] S. Motiian, M. Piccirilli, D. A. Adjeroh, and G. Doretto, "Unified deep supervised domain adaptation and generalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5715–5725.
- [23] A. Phinyomark and E. Scheme, "EMG pattern recognition in the era of big data and deep learning," *Big Data Cogn. Comput.*, vol. 2, no. 3, p. 21, 2018.
- [24] W. Geng et al., "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, 2016.
- [25] D. Joshi, B. H. Nakamura, and M. E. Hahn, "High energy spectrogram with integrated prior knowledge for EMG-based locomotion classification," *Med. Eng. Phys.*, vol. 37, no. 5, pp. 518–524, May 2015.
- [26] 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.
- [27] B. Hudgins, P. Parker, and R. N. Scott, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.
- [28] A. Al-Timemy, R. Khushaba, G. Bugmann, and J. Escudero, "Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 6, pp. 650–661, Jun. 2015.
- [29] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Statist. Planning Inference*, vol. 90, no. 2, pp. 227–244, 2000.
- [30] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2015, pp. 815–823.
- [31] 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.
- [32] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 14, no. 2, pp. 232–243, Apr. 2020.
- [33] W. Yang, D. Yang, Y. Liu, and H. Liu, "Decoding simultaneous multi-DOF wrist movements from raw EMG signals using a convolutional neural network," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 5, pp. 411–420, Oct. 2019.
- [34] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, Jun. 2011, pp. 315–323.
- [35] D. P. Kingma and J. Ba, "ADAM: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–15.
- [36] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.
- [37] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.
- [38] R. N. Khushaba, M. Takruri, J. V. Miro, and S. Kodagoda, "Towards limb position invariant myoelectric pattern recognition using time-dependent spectral features," *Neural Netw.*, vol. 55, pp. 42–58, Jul. 2014.
- [39] J. Liu, X. Sheng, D. Zhang, N. Jiang, and X. Zhu, "Towards zero retraining for myoelectric control based on common model component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 4, pp. 444–454, Apr. 2016.
- [40] H. Wang, P. Huang, T. Xu, G. Li, and Y. Hu, "Towards zero retraining for multiday motion recognition via a fully unsupervised adaptive approach and fabric myoelectric armband," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 217–225, 2022.