Improving the Robustness and Adaptability of sEMG-Based Pattern Recognition Using Deep Domain Adaptation

Ping Shi, Xinran Zhang, Wei Li, and Hongliu Yu

***Abstract*—The pattern recognition (PR) based on surface electromyography (sEMG) could improve the quality of daily life of amputees. However, the lack of robustness and adaptability hinders its practical application. To realize the long-term reliability and user adaptability simultaneously, a novel multi-task dual-stream supervised domain adaptation (MDSDA) network based on convolutional neural network (CNN) was proposed. A long-term multi-subject sEMG signal acquisition was conducted to validate the performance of MDSDA, recruiting 12 able-bodied subjects. A total of thirty gestures were used for the acquisition, including one set of static gestures and two sets of dynamic gestures. The long-term multi-subject sEMG dataset is publicly available at the website. Four train-test estimations were designed to evaluate the robustness and adaptability of MDSDA. The results showed that MDSDA outperformed CNN and fune-tuning. Furthermore, we studied the divisibility between static and dynamic gestures that performed similar actions. The outcomes demonstrated that there existed high separability between them. This may be helpful to reduce the signal collection burden. Experimental results proved MDSDA has the potential to provide a robust and generalized PR system for the clinic applications.**

***Index Terms*—Adaptability, deep domain adaptation, electromyography (EMG), pattern recognition, robustness.**

I. Introduction

HUMAN hands play a great critical role in daily life. The function of the human hands accounts for 90% of upper limb function. As a result, upper limb amputation seriously impairs the capacity of amputees to conduct daily duties [1]. Intelligent upper limb prostheses enable help amputees to restore hand function. The surface electromyography (sEMG)-based pattern recognition (PR) holds promise for providing natural and fluent control for upper limb prostheses [2]. The core of PR is obtaining sufficient information from the residual muscles of the amputee and decoding the corresponding motion intention [3]. Despite the fact that large numbers of laboratory studies have yielded positive recognition outcomes, the abandonment rate of upper limb prostheses still remains high. The main obstacle to this phenomenon is that the recognition performance of PR system (PRs) decreases dramatically over time in clinical applications. At the same time, the system performance may vary a lot from one subject to the next. The inaccurate response and frequent recalibration of the recognition system will create a significant usage burden for the user. Thus, there is an urgent need to address the lack of robustness over time and adaptability across users of PRs [4], bridging the gap between laboratory researches and clinical applications [5].

The degradation of system performance is caused by the fact that sEMG signal is highly unstable in nature and easily affected by various external confounding factors [6]. During long-term usage, the performance of system drops sharply because of the influence of electrode shift, limb positions, force variance, muscle crosstalk, muscle fatigue, skin sweating as well as time-dependent characteristics on sEMG signals [7]. In addition, the inter-subject problem is more complex on account of individual differences, personal habits, muscular strength, fat thickness, and so on [8]. These lead to the lack of prolonged usability and user adaptability of PR classification system [9], [10]. Developing a clinically reliable PRs remains challenging due to the high specificity of sEMG signal.

Convolutional neural network (CNN) has achieved a lot of success as a deep learning (DL) network in fields including image processing, speech recognition, machine translation, computer vision and so on [11], [12], [13], [14], [15]. Researches show that CNN is well suitable for processing bioelectrical signals, especially sEMG signals [16]. Due to its strong and fast computing power, CNN has been increasingly popular in sEMG-based PR in recent years [17]. For example, Geng et al. compared the performance of their proposed CNN to other methods in three public databases, with the results showing that CNN outperformed linear discriminant analysis, support vector machines, K-Nearest neighbors, and random forest (RF) [18]. Although CNN can achieve impressive classification results, it is unable to resolve the severe performance deterioration caused by between-day and inter-subject problems [19], [20]. Because CNN assumes that the training and testing data follow the same probability distribution. However, due to domain shift, where the source and target domains have different feature spaces [21], [22], this assumption is rarely valid in practice.

Manuscript received 12 January 2022; revised 7 May 2022 and 25 June 2022; accepted 3 August 2022. Date of publication 10 August 2022; date of current version 7 November 2022. This work was supported by the National Key R&D Program of China under Grant 2018YFB1307200. *(Corresponding author: Hongliu Yu.)*

The authors are with the Institute of Rehabilitation Engineering and Technology, University of Shanghai for Science and Technology, Shanghai 200093, China (e-mail: rehabishi@163.com; zhangxinranoo@163.com; fbrcn1017@163.com; yhl98@hotmail.com).

Digital Object Identifier 10.1109/JBHI.2022.3197831

One potential approach to address the poor robustness and adaptability is adequate training time and large training data from multiple individuals, which can produce an enormous impact on enhancing myoelectric control performance [23]. Through this method, both electrode shift, muscle fatigue, time dependence and individual differences are involved in training protocol. Nevertheless, this method is time-consuming and unrealistic. Because the extensive training exercises and data collection can be severely exhausting on individuals. And realistically, it is extremely difficult to recruit subjects to participate in long-term sEMG signal acquisition.

One successful approach is transfer learning (TL) that has efficiently improved the robustness and adaptability of PRs [24]. Fine-tuning (FT) is the most common TL method [25], which could enhance the performance under non-ideal conditions. For instance, FT mentioned in [26], [27] outperformed the methods without FT when dealing with the performance reduction caused by the instability of sEMG signals. However, FT has a significant drawback in the form of catastrophic forgetting [28]. It causes the information learned in the source domain to be lost, destroying the reusability of FT model. Additionally, the traditional pre-train and fine-tuning paradigm is prone to overfitting when the target domain data are sparse [29].

Deep domain adaptation (DDA) is another prominent TL method, which is the application of DL methods to the domain adaptation (DA) problem. Its central idea is to align the data distribution between the source and target domains. Compared with traditional methods, DDA keeps not only the powerful feature learning capability of DL but also the migratory nature of DA. It provides a stable and superior classification capacity on different target domains with using less labeled target data [30]. In practice, lots of studies have been undertaken to utilize DDA to handle domain shift in sEMG-based PR classification tasks. Du et al. [31] proposed an unsupervised adaptive batch normalization (AdaBN) network with high density EMG (HD-EMG) signal as the inputs of their classifier to address the time-lapse problem. After domain adaptation, 82.3% in inter-session and 55.3% in inter-subject were obtained. Zhai et al. [32] designed a self-recalibrating unsupervised classifier based on CNN, which offered a much higher classification performance of 78.71% than RF in inter-session scenario. Côté-Allard et al. [33] applied a self-recalibrating unsupervised domain adversarial neural network for long-term gesture recognition. An offline accuracy of 77.65% was achieved for eleven gestures. And later, they first studied inter-user data in [34]. Inspired by progressive neural networks and multi-stream AdaBN network, an augmented DA network was proposed to solve domain shift in inter-subject scenario [35]. An increase of over 19.40% in inter-subject classification accuracy compared to standard training was generated with using the adaptive domain adversarial neural network (ADANN). Although performance improvement in inter-session/subject scenarios has been realized in above-mentioned works. However, none of these researches have proved that their network could simultaneously adapt to inter-session/subject tasks.

There also exist many studies that seek to increase classification performance by designing a range of train-test strategies [36]. Asim et al. [37] proved that expanding the time span of training data improved the robustness of PRs over time. Evan et al. [38] extended a previously published ADANN to a cross-subject network. In emphasis, three dissimilar evaluation frameworks were performed to validate the network. At length, ADANN obtained 86.8–96.2% in accuracy for intact-limb, better than other methods. Yet these articles still only proved the long-term availability or adaptability.

In summary, although various DDA networks and training protocols were explored to boost the performance of PRs for realizing clinical application. Unfortunately, few individuals study a model that is both robust in long-term usage and generalized cross multiple users [39], [40], [41], which has a major impact on lowering the abandonment rate of commercial PR-based myoelectric prostheses.

Therefore, the first main aim of this paper was to design a gesture classification model with both long-term robustness and multi-subject adaptability. A novel multi-task dual-stream supervised domain adaptation network (MDSDA) was developed, a CNN-based DA network. It can reduce the need for frequent system recalibration and automatically adapt to temporal and inter-subject variations in sEMG signals. The second aim, and an innovation, was designing a long-term multi-subject sEMG signal acquisition experiment, contributing a multi-source dataset.[1] In [35], [42], [43], merely one class of static or functional grasping gesture was chosen for classification, and the quantity of selected gestures for experiment was small [25], [33], [37], [44]. In this paper, both static gestures and dynamic gestures were adopted to collect sEMG signals, in which dynamic gestures were grasps related to objects used to perform daily tasks, while static gestures were not. The dataset was used to test the robustness and adaptability of MDSDA. We tested the performance of MDSDA in four aspects, within-day, multi-day, inter-subject, and long-term inter-subject. After that, an investigation was proposed to compare the performance of FT and MDSDA in prolonged use and multi-user scenarios. Besides, we investigated the separability between static and dynamic gestures when performing similar hand actions. This study might generate fresh insight into the link between static and dynamic gestures.

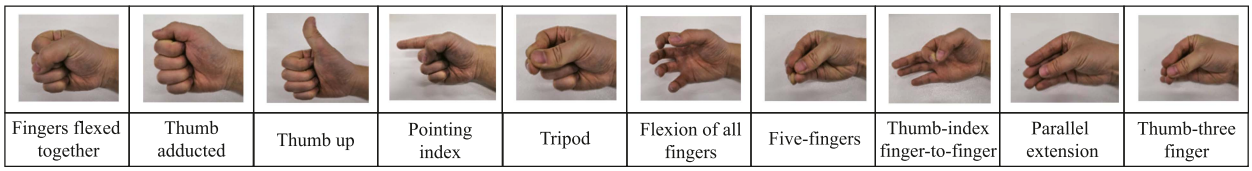
The rest of this paper is organized as follows: The Section II describes the design of long-term sEMG signal acquisition experiment and data processing process. The structures of CNN and MDSDA are shown in Section III. In Section IV, we introduce the designed evaluation experiments to evaluate the performance of MDSDA. The results are analyzed in Section V. Discussion and conclusion are performed in Section VI and Section VII.

II. Long-Term and Multi-Subject sEMG Dataset

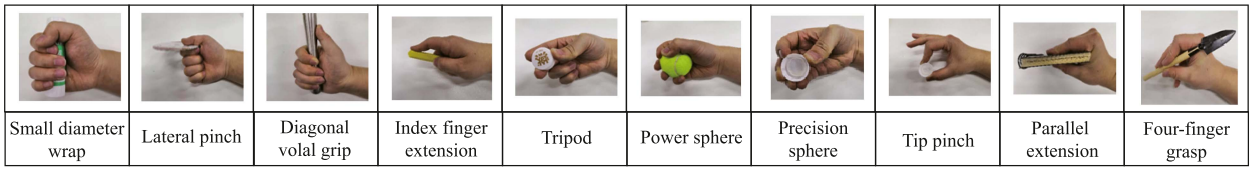
A. Subjects

We recruited 13 subjects, 7 females and 6 males, with an average age of 21.7 ± 1.6 years old. All subjects are healthy

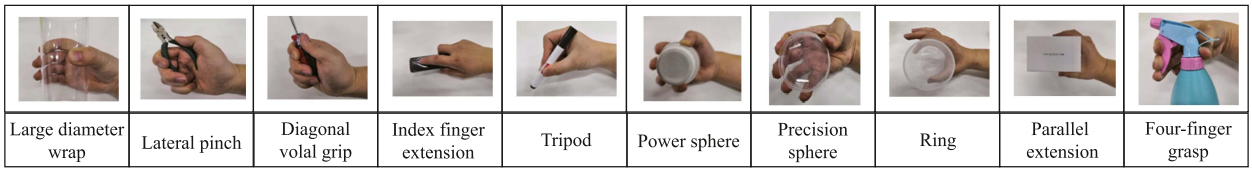
[1][Online]. Available: https://github.com/tinker1017/LONG-TERM-AND-MULTI-SUBJECT-SEMG-DATASET



(a) Static gestures: Basic hand movements



(b) Dynamic gestures1: Grasping and functional movements with the first group of objects



(c) Dynamic gestures2: Grasping and functional movements with the second group of objects

Fig. 1. The three sets of gestures used for recording.

and have no upper limb-related diseases. Before the experiment, we explained the acquisition to all subjects and recorded their names, gender, age, height, weight, forearm length, forearm circumference, dominant hand, exercise frequency and health status. Subjects have the right to withdraw from the experiment at any stage. Note that the ninth subject dropped out of the experiment due to personal reason. The recording experiment procedures conformed to the World Medical Association Declaration of Helsinki. All subjects were asked to sign a written informed consent. The study was approved by the Ethics Committee of Shanghai University of Medical & Health Sciences.

B. Hand Gestures

All the hand gestures were motivated by the functional actions with high frequency in daily life [45]. Three sets of hand gestures were selected for the acquisition: one set of static gestures (SG), one set of dynamic gestures 1 (DG1), and one set of dynamic gestures 2 (DG2), totaling 30 hand movements, which are shown in Fig. 1. SG set was 10 gestures which are frequently used in the daily life of people, selected from robotics and rehabilitation literature [19]. In particular, DG1 and DG2 evolved from SG by performing the same static action on two different objects. When presenting the performance verification results of the proposed model, SG, DG1, and DG2 were presented separately.

C. sEMG Acquisition Hardware

The sEMG signal was recorded using the commercial gForcePro+ armband, because it is non-intrusive, convenient, time-saving and more user-friendly. gForcePro+ is wearable sEMG sensors, which was produced by OYMotion Technologies (Shanghai, China, http://www.oymotion.com/product17/82). The armband has built-in 8-channel highly sensitive EMG sensor, 9-axis motion sensor, Bluetooth BLE4.2 and other modules. In this paper, only EMG data were used. It has a sampling frequency of 200 Hz–2000 Hz. In order to avoid the excessive loss of essential information of sEMG signals, 1000 Hz was used as the sampling frequency. It has an accompanying software to receive and save data via Bluetooth, allowing user-defined gestures to be captured.

During the experiment, the armband slid up along the forearm until there was no space between the electrodes and the skin, so the position of the armband was not strictly the same for each subject in each acquisition. The armband switch was oriented consistently upward.

D. Acquisition Protocol

Subjects were asked to sit in front of the table, with their forearm comfortably on the desktop via adjusting the chair height. The armband was not attached to the desktop to prevent the signal recording from being affected due by the extrusion of the desktop. As depicted in Fig. 2, there was a gesture action video instruction on the computer as a guide. Subjects imitated gestures in the video under the guidance of prompt information. The data were labeled, which were obtained through regular exercise-rest. Every gesture was repeated ten times. For all gestures, subjects held the gesture for 5 seconds, then rested for 2 seconds. After recording ten repetitions of one gesture, rest for 2 minutes. Completing the acquisition of thirty gestures in one day was a session. Each subject was to undergo ten sessions over ten consecutive days. The time interval between two sessions must be greater than 24 hours. Before formally collecting data, we explained the experimental protocol to all subjects and gave them simple recording exercises. All subjects had not previously participated in similar sEMG signal acquisition experiments.

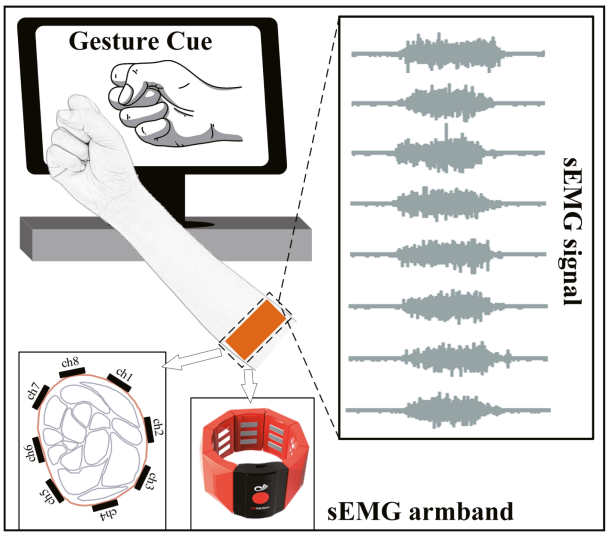


Fig. 2. The schematic of sEMG signal acquisition experiment.

E. sEMG Signal Processing

It is time-consuming to transform sEMG signals into new formats, like time domain and frequency domain conversions and so on, so the raw sEMG signals were used as the input of PRs. CNN can extract high-level and intrinsic characteristics automatically from the original data. The raw sEMG data contain a lot of noise, therefore, a fourth-order butterworth bandpass filter (20 Hz, 495 Hz) was used to remove motion artifacts and high-frequency noise, keeping the useful motion information unchanged. Subsequently, the signals were standardized using min-max normalization algorithm. The raw sEMG signals were directly used as the input of the classifier after simple standardization and filtering.

The sliding window method was used as data enhancement technology to segment the sEMG data into a series of window samples. Studies suggest that the optimal delay in sEMG-based PRs is between 100 ms and 250 ms [46]. Therefore, a window with length $W = 200$ ms slided across the data with an incremental length $I = 50$ ms. This window length can ensure continuous classification. The windows of input data were set to 8 $\times$ 200 (channels $\times$ window length) dimensional input matrix M.

$$M = \begin{bmatrix} d_{1,1} & \cdots & d_{1,200} \\ \vdots & \ddots & \vdots \\ d_{8,1} & \cdots & d_{8,200} \end{bmatrix} \tag{1}$$

III. Methodology

In this section, we give a detailed description of the proposed CNN and MDSDA architecture.

A. CNN Structure

The CNN structure proposed in this paper consisted of three convolutional blocks and two fully connected blocks, with a total of 18 layers. There were 6, 16, and 64 filters in the three convolution layers, and all convolutional kernels had a size of 3×3. The input boundary of each convolution layer was filled with a size of 1×1. Following each convolution layer, there was a batch normalization (BN) layer, a ReLU layer, and a MaxPool2d layer. The input data of each layer in the network were batch normalized so that the distribution of input data in each layer was relatively stable and the learning speed of the model was accelerated. For a batch of data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{m}$, BN will normalize each sample into the follow form

$$\widehat{\mathbf{x}}^{(j)} = \frac{\mathbf{x}^{(j)} - \mu^{(j)}}{\sqrt{\sigma^2(j) + \epsilon}}, \quad y^{(j)} = \gamma^{(j)}\widehat{\mathbf{x}}^{(j)} + \beta^{(j)} \tag{2}$$

where j represents the channel and μ, σ^2 denote the mean and variance, respectively. The number of hidden units in the full connection layer was 120 and 84, respectively. There was an activation function ReLU and a dropout layer behind each full connection layer, with a discard rate of 0.5. Both the BN layer and dropout layer have the effect of preventing overfitting and can reduce the complexity of the model. The Adam optimizer with an initial learning rate of 0.001 was utilized to reduce errors and update parameters.

The batch size B of the input was 256, so a sample of a batch was $M_b = \{M_1, M_2, \ldots, M_B\}_{i=1}^{N}$, where N was the number of all batches in a training iteration. The number of training iterations was 100. PyTorch 1.9.0 was used to build the classification and migration network, and the training and evaluation of the network were based on GeForce RTX 3080 GPU.

B. Deep Domain Adaptation Network

Generally speaking, DA can be categorized into supervised DA (SDA) and unsupervised DA (UDA) [47]. In UDA, there are labeled data in source domain data D_S and no labels in target domain data D_T. The major task of DA is concentrated on the coordination of feature distribution between domains. In contrast, the D_T of SDA has a small amount of labeled data for model recalibration, which can be used to build a bridge from source domain to target domain. Therefore, SDA is more suitable for dealing with the domain shift problem in sEMG, which could more accurately adapt to the instability of sEMG signals.

According to the rules of SDA, we defined three types of data: $D_S = \{(\mathbf{X}_m^S, y_m^S)\}_{m=1}^{M}$ is the source domain data, $D_T^{cal} = \{(\mathbf{X}_n^T, y_n^T)\}_{n=1}^{N}$ and $D_T^{test} = \{(\mathbf{X}_l^T, y_l^T)\}_{l=1}^{L}$ are from the target domain data. They have the same dimensionality of features and label spaces, but the data length is inconsistent. M, N, and L stand for the length of data, where $N \ll M \ll L$. $\mathbf{X}$ represents a window of sEMG signal input, and y stands for the gesture label of $\mathbf{X}$.

Combining SDA, supervised learning and multi-task learning, a CNN-based MDSDA network was proposed, where the two streams had the same structure with CNN in the previous section, but independent in calculation. As depicted in Fig. 3, the MDSDA scheme consisted of two sub-networks: the source network and the target network. This operation effectively aligned

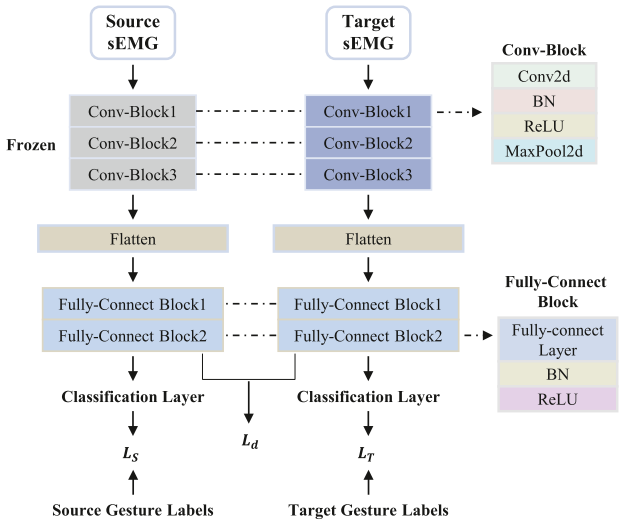


Fig. 3. The proposed MDSDA architecture.

each target domain data D_T with source domain data D_S for recalibration with a small amount of target data. Meanwhile, by pairing the samples of the two domains for model pre-training, the load of calculation can be further reduced.

In addition to the model structure, the loss function is also crucial to MDSDA. In this framework, gesture classification loss and domain variance loss were combined to adjust the weights of CNN. The first was used to classify gestures by supervised learning, and the second was to minimize the distribution divergence by aligning the feature distribution of source domain and target domain. The overall loss of MDSDA is as follows:

$$L(\boldsymbol{\theta} \mid \mathbf{X}^S, y^S, \mathbf{X}^T, y^T) = L_S + L_T + L_d \tag{3}$$

where L_S and L_T were the classification losses calculated from the source domain D_S and the calibration data D_T^{cal}. This classification loss function was obtained by optimizing the most commonly used cross entropy cost function. The formula is as follow:

$$L_S = \frac{1}{n_s} \sum_{i=1}^{n_s} c\left(\boldsymbol{\Theta}_c^S \mid \mathbf{X}_i^S, y_i^S\right) \tag{4}$$

$$L_T = \frac{1}{n_t} \sum_{j=1}^{n_t} c\left(\boldsymbol{\Theta}_c^T \mid \mathbf{X}_j^T, y_j^T\right) \tag{5}$$

The optimization of domain variance loss used the maximum mean difference (MMD), which was used to measure the average distance between two domains in Reproducing Kernel Hilbert Space (RKHS).

$$L_d = \lambda_u r_u(\boldsymbol{\Theta}_f^S, \boldsymbol{\Theta}_f^T \mid \mathbf{X}^S, \mathbf{X}^T) \tag{6}$$

Among them, the empirical estimation of square MMD was calculated by the following formula:

$$MMD^2(\mathbf{X}^S, \mathbf{X}^T) = \left\| \sum_{i=1}^{n_s} \frac{\varphi(\mathbf{x}_i^S)}{n_s} - \sum_{j=1}^{n_t} \frac{\varphi(\mathbf{x}_j^T)}{n_t} \right\|_{\mathcal{H}}^2 \tag{7}$$

where $\mathbf{x}_i^S$ and $\mathbf{x}_j^T$ represent the output of the subsequent fully-connected block in the source stream and target stream, respectively.

The model was tested with the help of adaptive BN, which is helpful to realize domain-specific information encoding. During model training, adaptive BN can associate a set of parameters with each domain, but there are common weights and deviations across all domains. When recalibration and testing, all convolution layers and linear layers were frozen, only the mean and variance were updated, and the model weight remained unchanged. Unlike conventional batch standardization, domain related adaptive batch standardization parameters are retained to adapt to the activation after training. This enables the pre-training model to adapt to the invisible domain by using a small amount of recalibration data to learn its BN parameters.

C. Statistical Analysis

The performance of the classifier was represented using the overall accuracy, that was, the ratio of the correct number of correct classifications to the total number of samples.

$$Accuracy = \frac{Number_{correct}}{Number_{total}} \times 100\% \tag{8}$$

A one-way analysis of variance (ANOVA) with factor classifiers (CNN, FT, and MDSDA) was used to compare the overall performace in multi-day, inter-subject, and long-term inter-subject scenarios respectively. P-values less than 0.05 were considered significant.

IV. Evaluation Experiments Design

In this section, we explained how the evaluation experiments were designed to access the performance of MDSDA from multiple aspects. These experiments were designed to test the long-term usability and user adaptability of MDSDA, with CNN as the baseline for no DA method. When testing CNN, the odd number of repetitions were used for training and the even number of repetitions were used for testing. And when testing MDSDA, the 1^{st} and 2^{nd} repetitions were used for retraining and the remaining eight repetitions were used for testing. The pre-training set of MDSDA under each strategy was the same data as the CNN training set, and the testing set was different.

A. Domain Shift Effects on Inter-Gesture Estimation

This experiment was leveraged to investigate the sorted capacity between static gesture and two dynamic gestures. They are similar in kinematics. Therefore, in this paper, we experimented to verify the relationship between the three sets of gestures. Besides, no attempt has been made to study whether one set of gestures can be used to predict another. The CNN network was used as the classifier for this experiment, where no TL method was applied. This study was performed independently for each static gesture with two corresponding dynamic gestures. Data from the tenth session were randomly selected for this experiment. For each set of gestures, the odd number of repetitions were used for training and the even number of repetitions for testing.

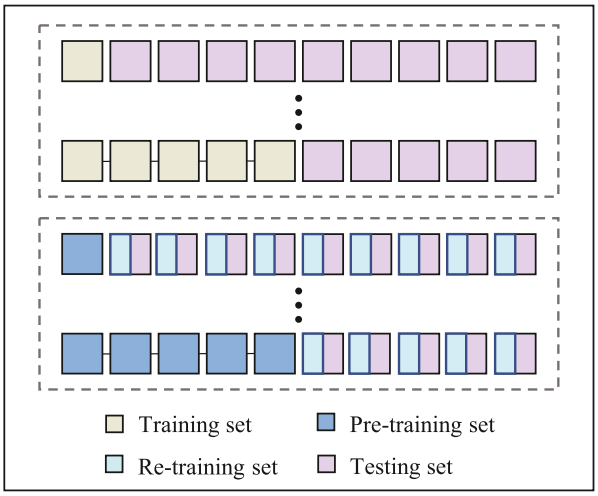


Fig. 4. The dataset division under MD strategy. A square represents data collected from one session. The above part is applicable to CNN, and the bottom part is for training MDSDA.

B. Within-Day Estimation (WD)

It was used to validate the performance of the benchmark model CNN. The twelve subjects were trained independently. For each subject, the training set and the testing set were from the same session. The results were shown as the mean of the ten sessions accuracies.

C. Multi-Day Estimation (MD)

It was used to test whether CNN and MDSDA are robust in long-term use. In this scenario, the training and testing data belonged to different time distributions. Moreover, the effect of training set size on classification performance was also studied.

For CNN, the training set was sequentially increased from one session to five sessions in an increment of one session each. And after the size of the training set was determined, every remaining session would be used as a separate testing set in turn. The pre-training set for MDSDA was the same as the training set for CNN, and the testing set for CNN was used for retraining and testing MDSDA. The results are shown as the mean value of the accuracies of all subjects on the testing sessions. In detail, the ceiling of training sets was five sessions because the remaining five sessions were needed to verify the long-term stability. The detailed dataset division in this strategy is shown in Fig. 4.

D. Inter-Subject Estimation (IS)

It was used to validate whether CNN and MDSDA could be adapted to different subjects. For each subject, ten sessions were used independently for training and testing, and the results were averaged over ten sessions. Twelve-fold cross-validation was used for each session, with twelve divisions of twelve subjects, each subject taking turns as the testing set.

For CNN, when a subject was used for testing, the remaining eleven subjects were used as the training set. For MDSDA, 20% of the data of the left subject were used for calibration and the rest were used for testing.

E. Long-Term Inter-Subject Estimation (LIS)

It was used to validate whether the MDSDA could maintain stable inter-subject performance during long-term usage. The difference from inter-subject was that both user-specific and time-dependent variations were involved in the training protocol. The recalibrated MDSDA model was needed to predict gestures performed by the testing subject in ten sessions.

The assignment of subjects used for training and testing was the same as for the inter-subject scenario. The difference was that subjects' ten-session data would be used together for training. 20% of the data from each of the ten sessions were used to recalibrate the MDSDA, and 80% of the remaining data from each session were used for testing.

V. Results

A. Results on Inter-Geature Estimation

In this experiment, the separability between each static gesture and its corresponding two dynamic gestures was studied. Fig. 5 presents the confusion matrix of classification results on ten gestures. A high classification rate was generated from three similar gestures, which indicates that the sEMG signals of these three gestures are distinctly different. Each group of static gesture and dynamic gesture could be well distinguished, which means that a model trained on static gestures data can be applied to identify dynamic gestures with slight dynamic gesture data for DA. This interesting finding had a potential impact on diminishing sEMG signal recording burden and the burden of users in real life.

B. Results on Within-Day Estimation

In this scenario, the armband was worn only once and would not be removed during the experiment. So the difference in sEMG signals mainly came from the muscle fatigue, repetitive action, and peripheral muscle crosstalk of the same subject. Fig. 6 shows the accuracy in within-day scenario. The classification accuracies of twelve participants were all above 90%. As the baseline classifier for this paper, CNN exhibited excellent performance. One remarkable finding is that the accuracy of S5 for SG is $95.82\% \pm 8.46\%$, with relatively large error, but this is not because the difference in ten-day accuracies is relatively obvious. Fig. 7 shows the accuracies of S5 and S12 for SG in ten days. It can be seen that for S5 only the accuracy on the first day is 70.8%, in contrast, the remaining accuracies are all more than 95%. This led to the obvious overall distribution difference. It is speculated that this subject did not adapt and adjust well when performing the acquisition on the first day. While the reason for large error of S12 is that the accuracies over ten days are not grossly stable. As shown in the Fig. 6, the maximum and minimum accuracy of SG, DG1, and DG2 appeared in different subjects, and the accuracies of the 12 subjects were different, which just reflects the user-specific of sEMG signal.

C. Results on Multi-Day Estimation

The results of SG, DG1, and DG2 are shown in Fig. 8. First of all, in terms of long-term usability, it can be seen from the figure

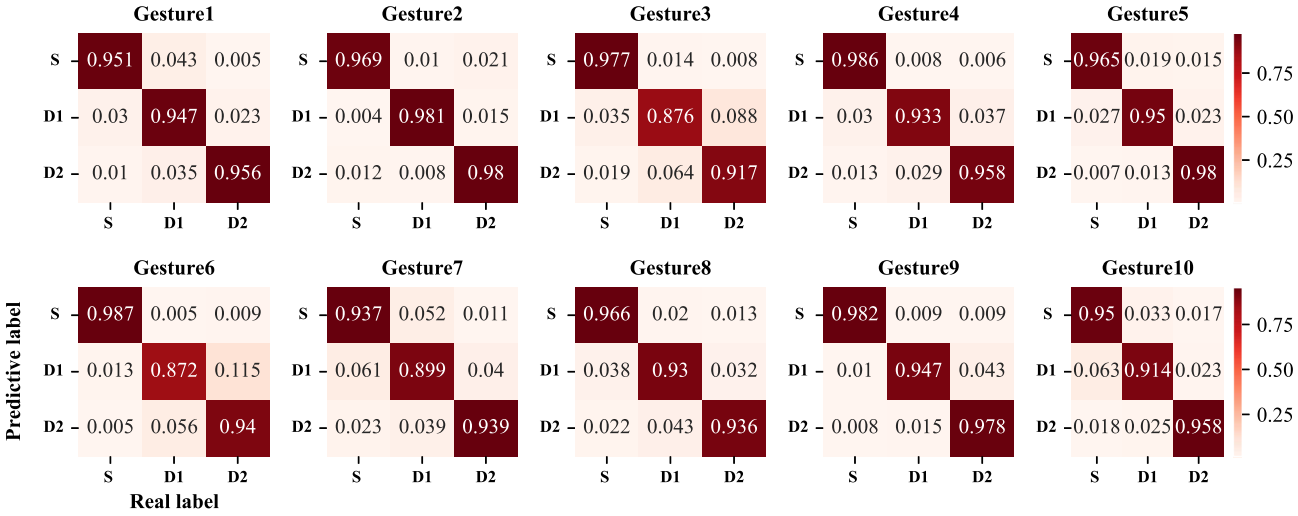


Fig. 5. The confusion matrix of classification on ten gestures.

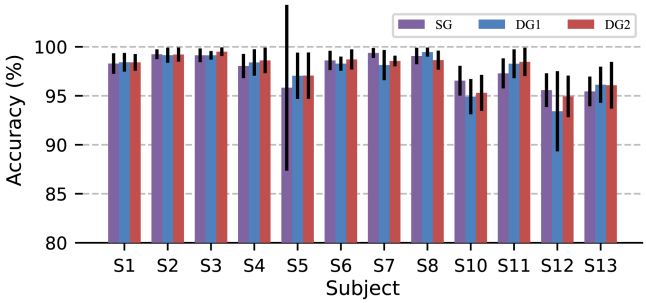


Fig. 6. Classification accuracy in within-day scenario with using CNN.

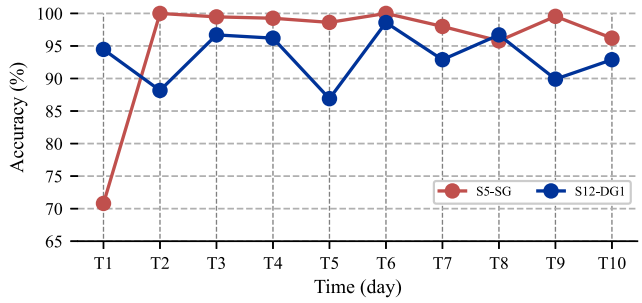


Fig. 7. Ten sessions accuracies of S5 for SG and S12 for DG1.

that the performance of CNN decreased slightly over time. For example, when the training set size in SG was '1d', the testing accuracy was 30.5% ± 7.47% on the second day, by contrast, the classification accuracy was only 24.62% ± 7.91% on the tenth day, which dropped by 5.88%. The CNN algorithm just has slight robustness to time variation.

Moreover, as the training set became larger, the classification accuracy went up significantly. For CNN method, the accuracies of SG with the training set size from one to five days were $28.67\% \pm 2.48\%$, $36.7\% \pm 2.62\%$, $44.6\% \pm 3.59\%$, $48.39\% \pm 2.79\%$, $52.46\% \pm 2.07\%$, respectively. The classification accuracy gradually rose with the increase of the training set size. The accuracies of DG1 and DG2 also had an increase of 22.57% and 20.01%, respectively. With MDSDA, the effect of the training set size on model performance is reduced. When the training set size was one day and five days, the accuracies of SG were respectively $96.57\% \pm 0.59\%$ and $98\% \pm 0.38\%$. The accuracies of DG1 were respectively $96.03\% \pm 0.63\%$ and $97.83\% \pm 0.5\%$. The accuracies of DG2 were respectively $96.57\% \pm 0.61\%$ and $97.6\% \pm 0.42\%$. It can be seen that the quantity of training set was five times different, but the gesture recognition accuracy just had a difference of less than 2%. Using data of one session can achieve the same classification effect to 5 sessions. These results confirm that MDSDA achieves supremely better improvement than the method of increasing training set size.

D. Results on Inter-Subject Estimation

In this experiment, the training set and testing set belonged to different people but from the same day. From Fig. 9 we can see that MDSDA was greatly superior to CNN. An increase of over 60% in accuracy was achieved.

E. Results on Long-Term Inter-Subject Estimation

Fig. 10 is the result of the long-term inter-subject validation experiment. For LIS strategy, the effect of time variable is introduced on the basis of IS. The recognition accuracy of all subjects were just about 50% when using CNN. Even in the case of large training data sets, it was still unable to obtain ideal gesture recognition accuracy. After DA, the accuracies basically reached above 80%, increasing by about 30%.

VI. Discussion

To the best of the authors' knowledge, the investigation on distinguishability between static gesture and dynamic gesture was first performed in this paper. In our experiment, dynamic gesture evolved from static gesture. They have similarities in kinematics, thus we desire to understand the relationship between them.

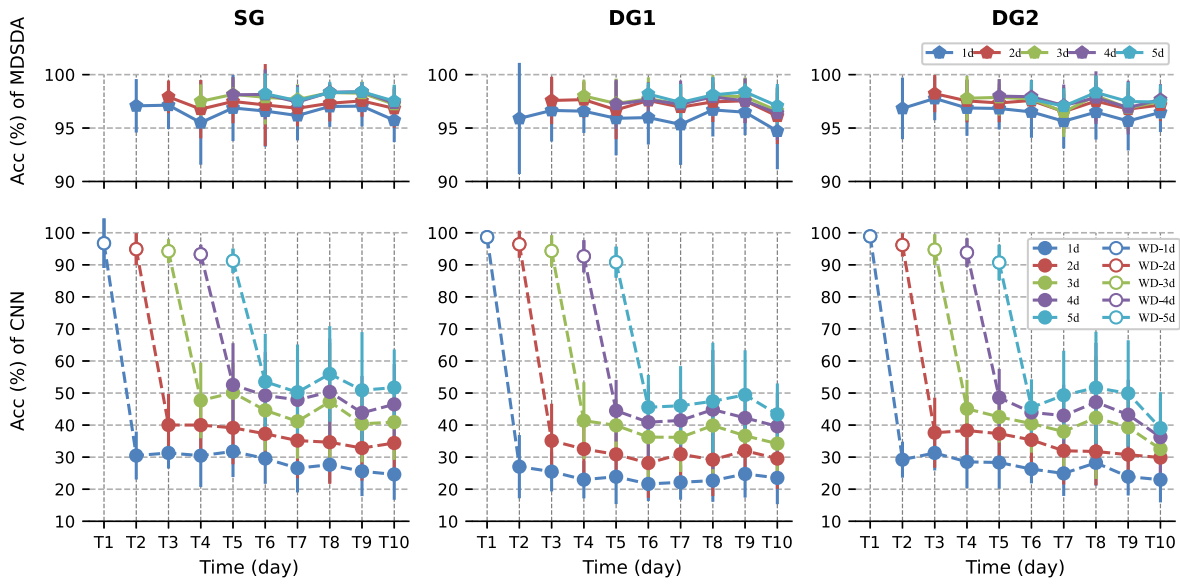


Fig. 8. Classification accuracy of multi-day strategy. 'xd' means the size of training set is x, ranging from one to five days. 'WD-xd' refers to the classification results of the x day in within-day scenario, which is used to compare with multi-day scenario. A session was completed in one day.

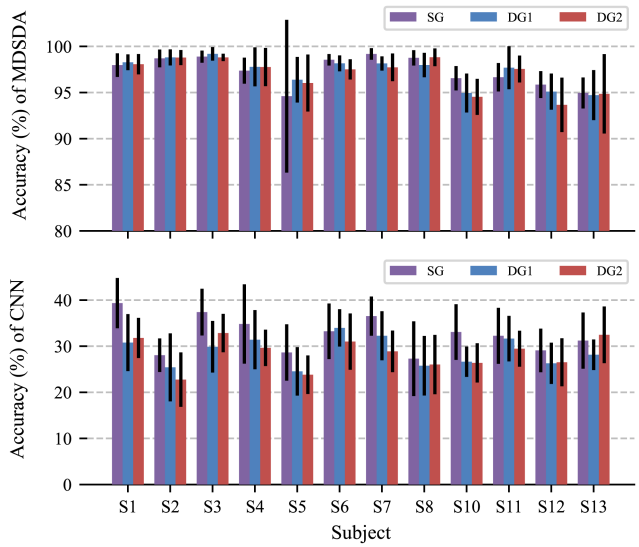


Fig. 9. Accuracy of inter-subject.

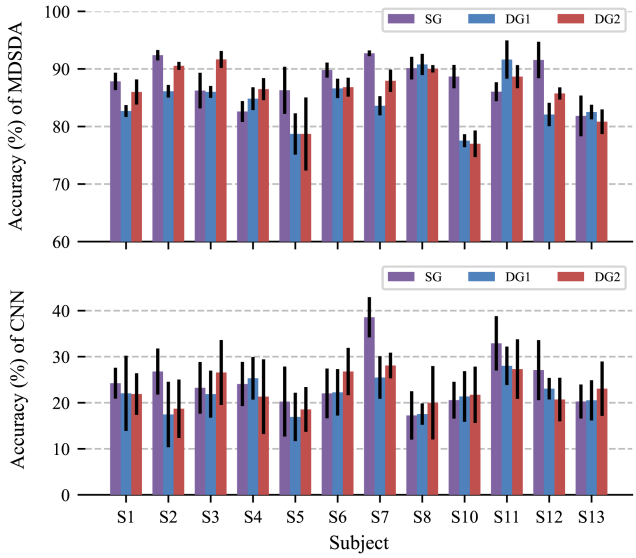


Fig. 10. Accuracy of long-term inter-subject.

As is known, recording data for the long-term and multi-subject experiment is a big project, engendering a great burden on user. Static gestures and functional grasping gestures are often used for classification separately, considered as independent. However, no attempt has been made to study whether one set of postures can be used to predict another. Results shown in Fig. 5 proved that the three seemingly identical hand gestures were actually three different actions. They differ significantly in sEMG signals. This means the three types of gestures can be considered as different domains, just like different days and different subjects. In this way, static gestures can be treated as source domain and dynamic gestures as target domain. Via DA, the network trained on static gestures could be used to predict dynamic gestures in case of insufficient dynamic gestures data, vice versa. It is expected to achieve more hand motions by capturing a smaller number of gestures, reducing the usage burden. Further work is needed to be conducted in order to verify whether static gestures can be used to predict dynamic gestures precisely, and vice versa. This experiment helps demonstrate that the proposed algorithm is robust across multiple subjects on multiple days and even multiple gestures. Of course, the current experiment has limitations and does not fully achieve the proof of robustness in multiple cases, and further experimental proofs are needed subsequently.

One noteworthy point is that the SG recognition accuracies of all subjects are higher than that of DG1 and DG2. The reason may be that the execution of the static gestures is relatively simple, the difference in execution time between subjects is

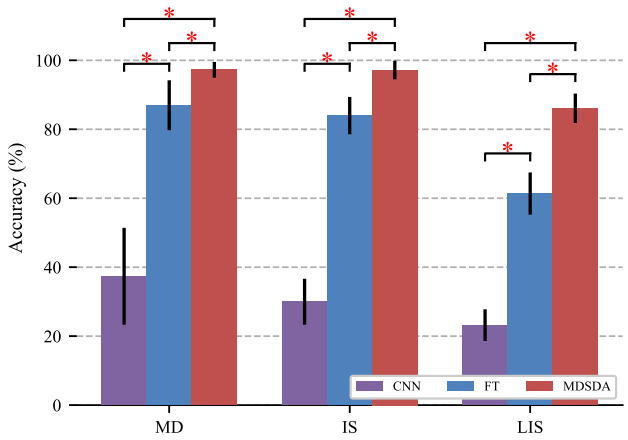


Fig. 11. The classification accuracy of CNN, FT and MDSDA in multi-day, inter-subject and long-term inter-subject scenarios. The red * indicates a significant difference ($p<0.05$).

relatively small, and the dynamic gestures need to grasp the corresponding objects.

Domain shift in multi-day and inter-subject scenarios is a major obstruction to achieve clinical use of sEMG-based gesture recognition [1]. However, the results of validation experiment showed that the proposed MDSDA network mitigated the bad effect of the instability of sEMG signals. In multi-day, inter-subject, and long-term inter-subject scenarios, due to acquisition time changing and individual differences, CNN just obtained average accuracies of 37.37%, 29.98%, and 23.2%, respectively. After applying MDSDA, promotions of 59.88%, 67.22%, and 62.9% had been produced. Another relatively common TL method, FT, was also used for MD, IS, and LIS evaluation experiments. The classification results of CNN, FT, MDSDA are shown in Fig. 11. One-way ANOVA test showed significant difference ($p<0.05$) between three classifiers in three situations. The accuracies of FT in MD and IS scenarios were only $86.97\% \pm 7.24\%$ and $83.97\% \pm 5.42\%$, which are 9.81% and 10.04% lower than MDSDA, respectively. As expected, MDSDA remarkably outperformed FT. Unlike FT with catastrophic forgetting, the well pre-trained network was not be discarded, but integrated layer by layer with a new network for the new target domain. The average accuracy of 97.25% and 97.2% in multi-day and inter-subject scenarios suggest that MDSDA has the capacity to address time-lapse and subject changes individually. This means that MDSDA is already robust and generalized enough. Many current studies have also achieved such success. However, far too little attention has been paid to improving both robustness and adaptability of PRs [40]. In practical applications, time and subject variables are present simultaneously. It is unrealistic for users to perform tedious training protocols without professional guidance. And the cost of training a classifier specifically for each user on clinical applications is huge. The long-term inter-subject strategy was designed to solve this problem. In LIS scenario, the training set contained the sEMG signals of all subjects for ten sessions. Then, the trained MDSDA was used to predict the gestures of the test subject for ten sessions. It is worth noting that, a 24.75% gap was found between MDSDA and FT in LIS estimation. MDSDA obtained an average accuracy of $86.1\% \pm 4.26\%$, however, FT just achieved an accuracy of $61.35\% \pm 6.13\%$. In summary, MDSDA excellently adapted to time-lapse and changes across subjects, achieving eximious and stabilized robustness and adaptability.

On the one hand, the accuracies over time obtained in [31], [32], [33], [37] were 82.3%, 77.65%, 78.71%, and 94.22% respectively, which realized nice time robustness but still poorer than our result 97.25% in MD scenario. On the other hand, in [48], efforts were made to get a generalized model for multiple users. And finally, they obtained an accuracy of 80.17% on there private dataset. Different from these studies above, the improvement of MDSDA is more successful as shown by its impressive performance on MD and IS. Moreover, LIS, which focused on both time robustness and subject adaptability, also performed well. The main reason may attribute to the improvement of MDSDA algorithm. Both gesture classification loss and domain variance loss were used to adjust the network weights. The domain variance loss allows MDSDA to align the feature distribution of source domain and target domain, adapting different target domains more accurately and fast.

An investigation was performed to examine the effect of the amount of training set on system performance in multi-day strategy. Fig. 8 was the results of multi-day strategy, showing that increasing the size of training set can improve the robust performance of PRs gradually. However, when the training set size was five sessions, the accuracies did not exceed 60%. Increasing the size of training set does improve the classification accuracy, but it is unrealistic. The huge training data will bring a heavy burden to model training. And in practice, it is unrealistic for users to use much more than five sessions to recalibrate the PRs. Because frequent recalibration training happens to be one of the reasons for the high rejection rate of PR control-based myoelectric prostheses. As shown in Fig. 8, promotions of 22.12% and 1.42% were observed, which represented the increasing accuracy by CNN and MDSDA with the training size from 1 to 5 sessions. These results suggested that MDSDA decreased the influence of training set size on the improvement in the system performance. MDSDA achieved near-perfect classification accuracy using a short-time recalibration. This finding may have important implications for developing a PRs with a lower training burden in real long-term usage for amputees.

In conclusion, there are still some deficiencies in this paper. We addressed the problem of inadvertent electrode displacement, a small shift produced by subjects when they took off the armband. However, our study did not involve a larger shift, such as 2 cm. Typically, a shift of 1 cm could result in severe performance degradation [26], [49]. Besides, our study does not address the impact of limb position. Under the specific conditions of experimental research, the subjects were asked to place their forearms comfortably flat on the table for each acquisition. Nonetheless, in real life, the users' limb position is uncontrollable and random, which possibly leads to a performance degradation compared with the well-performed PRs in the experimental environment [50]. Furthermore, the effect of limb strength also remains to be explored. Therefore, it is necessary to train the PRs by combining the training environment with daily life scenarios. Another important problem is the lack

of amputees. MDSDA has not been tested on amputees, staying at the stage of experimental research. And also, a study about testing the performance of MDSDA using NinaPro database 6 should be conducted.

VII. Conclusion

This study set out to improve the robustness and adaptability of PRs meanwhile. A long-term multi-subject sEMG recording was performed. Verification results have proved that MDSDA showed a superior performance. Furthermore, a high classification accuracy between each static gesture and relevant two dynamic gestures may be of assistance to reducing sEMG signal recording burden. Although it will be very challenging to include all confounding factors in the training protocol, providing a robust and generalized PRs without strict control conditions is meaningful. Therefore, further work will be carried out to access the model in complex real-life scenarios to realize real-time usability.

Acknowledgment

The authors would like to acknowledge all subjects for their active participation. The authors would like to thank Qing-Yun Meng, a senior engineer, for supporting to this study.

References

[1] J. L. Nawfel, K. B. Englehart, and E. J. Scheme, "A multivariate approach to predicting myoelectric control usability," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1312–1327, 2021, doi: 10.1109/TNSRE.2021.3094324.

[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, doi: 10.1109/TNSRE.2012.2196711.

[3] D. Farina, A. Holobar, R. Merletti, and R. M. Enoka, "Decoding the neural drive to muscles from the surface electromyogram," *Clin. Neurophysiol.: Official J. Int. Federation Clin. Neurophysiol.*, vol. 121, no. 10, pp. 1616–1623, Oct. 2010, doi: 10.1016/j.clinph.2009.10.040. PMID: 20444646.

[4] D. Yang, Y. Gu, N. V. Thakor, and H. Liu, "Improving the functionality, robustness, and adaptability of myoelectric control for dexterous motion restoration," *Exp. Brain Res.*, vol. 237, no. 2, pp. 291–311, Feb. 2019. [Online]. Available: https://doi.org/10.1007/s00221-018-5441-x

[5] C. Castellini et al., "Proceedings of the first workshop on peripheral machine interfaces: Going beyond traditional surface electromyography," *Front. Neurorobot.*, vol. 8, 2014, Art. no. 22.

[6] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Develop.*, vol. 48, no. 6, pp. 643–659, 2011.

[7] J. Liu, X. Sheng, D. Zhang, J. He, and X. Zhu, "Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 1, pp. 166–176, Jan. 2016, doi: 10.1109/JBHI.2014.2380454.

[8] 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, doi: 10.1109/TNSRE.2015.2420654.

[9] J. He, D. Zhang, N. Jiang, X. Sheng, D. Farina, and X. Zhu, "User adaptation in long-term, open-loop myoelectric training: Implications for EMG pattern recognition in prosthesis control," *J. Neural Eng.*, vol. 12, no. 4, 2015, Art. no. 046005. doi: 10.1088/1741-2560/12/4/046005.

[10] D. Yang, L. Jiang, Q. Huang, R. Liu, and H. Liu, "Experimental study of an EMG-controlled 5-DoF anthropomorphic prosthetic hand for motion restoration," *J. Intell. Robotic Syst.*, vol. 76, no. 3, pp. 427–441, 2014.

[11] 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.

[12] D. Silver et al., "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[13] G. Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[16] 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, 2018, Art. no. 21.

[17] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019.

[18] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, pp. 1–8, 2016.

[19] 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. 2019.

[20] F. Palermo, M. Cognolato, A. Gijsberts, H. M. üller, B. Caputo, and M. Atzori, "Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data," in *Proc. Int. Conf. Rehabil. Robot.*, 2017, pp. 1154–1159.

[21] C. Yu et al., "Learning to match distributions for domain adaptation," 2020, *arXiv:2007.10791*. [Online]. Available: http://arxiv.org/abs/2007.10791

[22] Y.-P. Lin and T.-P. Jung, "Improving eeg-based emotion classification using conditional transfer learning," *Frontiers Hum. Neurosci.*, vol. 11, 2017, Art. no. 334.

[23] A. Tabor, S. Bateman, and E. Scheme, "Evaluation of myoelectric control learning using multi-session game-based training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 9, pp. 1680–1689, Sep. 2018.

[24] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[25] Y. Bengio, "Deep learning of representations for unsupervised and transfer learning," in *Proc. ICML Workshop Unsupervised Transfer Learn. JMLR Workshop Conf. Proc.*, 2012, pp. 17–36.

[26] A. Ameri, M. A. Akhaee, E. Scheme, and K. Englehart, "A deep transfer learning approach to reducing the effect of electrode shift in EMG pattern recognition-based control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 2, pp. 370–379, Feb. 2020.

[27] W. Wang, B. Chen, P. Xia, J. Hu, and Y. Peng, "Sensor fusion for myoelectric control based on deep learning with recurrent convolutional neural networks," *Artif. Organs*, vol. 42, no. 9, pp. E272–E282, Sep. 2018. [Online]. Available: https://doi.org/10.1111/aor.1315

[28] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 12, pp. 2935–2947, Dec. 2018.

[29] A. A. Rusu et al., "Progressive neural networks," 2016, *arXiv:1606.04671*.

[30] G. Wilson and D. J. Cook, "A survey of unsupervised deep domain adaptation," *ACM Trans. Intell. Syst. Technol*, vol. 11, no. 5, pp. 1–46, 2020.

[31] 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, no. 3, 2017, Art. no. 458.

[32] X. Zhai, B. Jelfs, R. H. Chan, and C. Tin, "Self-recalibrating surface emg pattern recognition for neuroprosthesis control based on convolutional neural network," *Frontiers Neurosci.*, vol. 11, 2017, Art. no. 379.

[33] U. Côté-Allard et al., "Unsupervised domain adversarial self-calibration for electromyography-based gesture recognition," *IEEE Access*, vol. 8, pp. 177941–177955, 2020.

[34] U. Côté-Allard et al., "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 760–771, Apr. 2019.

[35] U. Côté-Allard, C. L. Fall, A. Campeau-Lecours, C. Gosselin, F. Laviolette, and B. Gosselin, "Transfer learning for sEMG hand gestures recognition using convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2017, pp. 1663–1668.

[36] A. Waris, I. K. Niazi, M. Jamil, K. Englehart, W. Jensen, and E. N. Kamavuako, "Multiday evaluation of techniques for EMG-based classification of hand motions," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 4, pp. 1526–1534, Jul. 2019.

[37] A. Waris, I. Mendez, K. Englehart, W. Jensen, and E. N. Kamavuako, "On the robustness of real-time myoelectric control investigations: A multiday fitts' law approach," *J. Neural Eng.*, vol. 16, no. 2, 2019, Art. no. 026003.

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

[39] H. Xu and A. Xiong, "Advances and disturbances in sEMG-based intentions and movements recognition: A review," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13019–13028, Jun. 2021.

[40] O. W. Samuel et al., "Intelligent EMG pattern recognition control method for upper-limb multifunctional prostheses: Advances, current challenges, and future prospects," *IEEE Access*, vol. 7, pp. 10150–10165, 2019.

[41] Y. Gu, D. Yang, Q. Huang, W. Yang, and H. Liu, "Robust EMG pattern recognition in the presence of confounding factors: Features, classifiers and adaptive learning," *Expert Syst. Appl.*, vol. 96, pp. 208–217, 2018.

[42] Y. Zhang, Y. Chen, H. Yu, X. Yang, and W. Lu, "Learning effective spatial–temporal features for sEMG armband-based gesture recognition," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 6979–6992, Aug. 2020.

[43] T. Ruan, K. Yin, and S. Zhou, "Convolutional neural network based human movement recognition using surface electromyography," in *Proc. IEEE 1st Int. Conf. Micro/Nano Sensors AI, Healthcare, Robot.*, 2018, pp. 68–72.

[44] U. Côté-Allard et al., "A transferable adaptive domain adversarial neural network for virtual reality augmented EMG-based gesture recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 546–555, 2021.

[45] M. Atzori et al., "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Sci. Data*, vol. 1, no. 1, pp. 1–13, 2014.

[46] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 186–192, Apr. 2011.

[47] M. Wang and W. Deng, "Deep visual domain adaptation: A. survey," *Neurocomputing*, vol. 312, pp. 135–153, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523121830668

[48] Y. Zhang, Y. Chen, H. Yu, X. Yang, and W. Lu, "Dual layer transfer learning for sEMG-based user-independent gesture recognition," *Pers. Ubiquitous Comput.*, pp. 1–12, Apr. 2020, doi: 10.1007/s00779-020-01397-0.

[49] L. Hargrove, K. Englehart, and B. Hudgins, "The effect of electrode displacements on pattern recognition based myoelectric control," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 2203–2206.

[50] A. Fougner, E. Scheme, A. D. C. Chan, K. Englehart, and Ø. Stavdahl, "Resolving the limb position effect in myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 6, pp. 644–51, Dec. 2011.