Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning

Wentao Wei, Qingfeng Dai, Yongkang Wong, *Member, IEEE*, Yu Hu,
Mohan Kankanhalli, *Fellow, IEEE*, and Weidong Geng

***Abstract*—Gesture recognition using sparse multichannel surface electromyography (sEMG) is a challenging problem, and the solutions are far from optimal from the point of view of muscle–computer interface. In this paper, we address this problem from the context of multi-view deep learning. A novel multi-view convolutional neural network (CNN) framework is proposed by combining classical sEMG feature sets with a CNN-based deep learning model. The framework consists of two parts. In the first part, multi-view representations of sEMG are modeled in parallel by a multistream CNN, and a performance-based view construction strategy is proposed to choose the most discriminative views from classical feature sets for sEMG-based gesture recognition. In the second part, the learned multi-view deep features are fused through a view aggregation network composed of early and late fusion subnetworks, taking advantage of both early and late fusion of learned multi-view deep features. Evaluations on 11 sparse multichannel sEMG databases as well as five databases with both sEMG and inertial measurement unit data demonstrate that our multi-view framework outperforms single-view methods on both unimodal and multimodal sEMG data streams.**

***Index Terms*—Surface electromyography, muscle-computer interface, human-computer interface, multi-view learning, deep learning.**

I. Introduction

WHEN designing Human-Computer Interfaces (HCI), the objective is to create more natural and intuitive interfaces through which human users can interact with computers via speech, touch, or gesture [1]. Surface Electromyography (sEMG) is a technique that uses electrodes placed on skin to record the biosignal produced by skeletal muscles during contraction. The recorded biosignals provide sufficient information about human movements [2]; thus, they can be used to realize gesture-based intuitive HCI. Approaches that use sEMG-based HCI, which is also known as a Muscle-Computer Interface (MCI), have been widely used in rehabilitation robots [3], prostheses [4], and Sign Language Recognition (SLR) systems [5].

For MCIs, one key point is to precisely recognize user gestures from sEMG. The sEMG recording techniques can be categorized into sparse multichannel sEMG and high-density sEMG (HD-sEMG). HD-sEMG records both the temporal and spatial changes of muscle activities via hundreds of electrodes [6]; but greatly increases the amount of data collected as well as the complexity of the system. In contrast, sparse multichannel sEMG uses fewer electrodes to record sEMG data and has become the most widely used sEMG recording technique in wearable systems [7].

Recent studies have demonstrated the superior performance of using a deep learning approach such as a Convolutional Neural Network (CNN) in gesture recognition using HD-sEMG [6], [8], [9]. Nevertheless, the gesture recognition performance of deep learning approaches on sparse multichannel sEMG is still far from optimal. For instance, one end-to-end CNN model (i.e. GengNet) proposed by Geng *et al.* [6] was able to achieve a gesture recognition accuracy of 99.6% using HD-sEMG but achieved only 77.8% accuracy when tested with gesture recognition using sparse multichannel sEMG.

Conventional sEMG-based gesture recognition approaches are capable of achieving acceptable gesture recognition performances using sparse multichannel sEMG. The classifier inputs are usually composed of discriminative, human-selected, hand-crafted features known as *feature sets* [10]–[12]. These classical feature sets contain considerable heuristic knowledge; thus, integrating these classical features with deep learning approaches could improve the gesture recognition performance on sparse multichannel sEMG.Unfortunately, however, the knowledge captured in handcrafted features is often ignored in current deep learning approaches [6], [8], [9], [13]. Therefore, with the goal of achieving a better gesture recognition performance when using sparse multichannel sEMG, we were inspired to investigate how to combine classical sEMG feature sets with a CNN-based deep learning model.

Multi-view learning is usually defined as learning from multiple feature sets that reflect multiple views of data [14], [15]. From a multi-view learning viewport, each sEMG feature set can be considered as a view of sEMG. It is worth exploring how to integrate the classical feature sets into the deep

Manuscript received July 10, 2018; revised October 23, 2018 and January 17, 2019; accepted February 6, 2019. Date of publication February 13, 2019; date of current version September 18, 2019. This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFB1001302, in part by the National Natural Science Foundation of China under Grant 61379067, and in part by the National Research Foundation, Prime Minister's Office, Singapore, under its Strategic Capability Research Centres Funding Initiative. *(Corresponding author: Weidong Geng.)*

W. Wei, Q. Dai, and Y. Hu are with the State Key Laboratory of CAD&CG, College of Computer Science and Technology, Zhejiang University.

Y. Wong and M. Kankanhalli are with the School of Computing, National University of Singapore.

W. Geng is with the State Key Laboratory of CAD&CG, College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: gengwd@zju.edu.cn).

Digital Object Identifier 10.1109/TBME.2019.2899222

learning framework in the context of multi-view learning for gesture recognition using sparse multichannel sEMG.

This work provides three major contributions:

- We design a performance-based view construction strategy by selecting the most discriminative views in terms of their gesture recognition performance and turning them into signal images. The experimental results indicate a single-view architecture is capable of achieving better gesture recognition performance than are the state-of-the-art end-to-end deep learning approaches. In addition, we determine the optimal number of views for gesture recognition via an exhaustive search strategy.
- Based on the performance-based view construction strategy, we propose a multi-view deep learning framework that results in better gesture recognition performance. Specifically, we propose a novel CNN architecture with an embedded view aggregation network that aggregates multiple views of sEMG. The view aggregation network is composed of an early-fusion network and a late-fusion network; thus, it can take advantages of both early-fusion and late-fusion of learned multi-view deep features.
- Comprehensive evaluations performed on 11 sparse multichannel sEMG databases as well as 5 multimodal databases with sEMG and IMU data are used to validate the gesture recognition performance of our proposed multi-view deep learning framework. The experimental results show that the proposed multi-view deep learning framework achieves higher gesture recognition accuracy than do single-view deep learning methods on both unimodal and multimodal data streams.

The remainder of this paper is organized as follows. Section II reviews the related work; Section III formulates the learning problem; Section IV introduces the view construction process; Section V details the proposed multi-view CNN; Section VI demonstrates the experimental results; Section VII concludes the paper and discusses future work.

II. Related Works

The sEMG-based gesture recognition approaches can be broadly categorized into deep learning-based approaches and conventional machine learning approaches. For conventional machine learning approaches, raw sEMG is generally thought to be impractical for gesture recognition due to its noisy, random and nonstationary nature [16]. Therefore, human experts have proposed a number of classical sEMG feature sets. We summarized 11 classical sEMG feature sets as follows:

- **Hudgins's Feature Set (Hudgins_FS) [17]**: Mean Absolute Value (MAV), Waveform Length (WL), Slope Sign Change (SSC), and Zero Crossing (ZC).
- **Du's Feature Set (Du_FS) [18]**: Integrated EMG (IEMG), Variance (VAR), Willison Amplitude (WAMP), WL, SSC, ZC.
- **Temporal-Spatial Descriptors (TSD) [19]**: A newly proposed feature set that consists of 7 Time Domain Descriptors (TDD) extracted from individual and combined sEMG channels.
- **Atzori's Feature Set (Atzori_FS) [20]**: Root Mean Square (RMS), Marginal of Discrete Wavelet Transform (mDWT), Histogram of EMG (HEMG) and Hudgins_FS.
- **Phinyomark's Feature Set 1 (Phin_FS1) [10]**: MAV, WL, WAMP, ZC, Mean Absolute Value Slope (MAVS), Autoregressive Coefficients (ARC), Mean Frequency (MNF) and Power Spectrum Ratio (PSR).
- **Phinyomark's Feature Set 2 (Phin_FS2) [21]**: Sample Entropy(SampEn), Cepstrum Coefficients (CC), RMS and WL.
- **Doswald's Feature Set (Doswald_FS) [22]**: 58 statistics of Hilbert-Huang Transform (HHT), 29 statistics of autoregressive residue, mean frequency from all Intrinsic Mode Functions (IMF) in HHT, together with Phin_FS.
- **Time Domain features combined with Autoregressive model coefficients (TDAR) [23]**: MAV, SSC, WL, VAR, WAMP, ARC, and ZC.
- **Discrete Wavelet Transform Coefficients (DWTC) [24]**
- **Discrete Wavelet Packet Transform Coefficients (DWPTC) [25]**
- **Continuous Wavelet Transform Coefficients (CWTC) [26]**

Recently, deep learning based approaches such as CNNs have been merged into sEMG-based gesture recognition systems [6], [13]. While CNNs have proven to be effective for gesture recognition using HD-sEMG [6], [8], [9], for gesture recognition using sparse multichannel sEMG, their performance is still far from optimal. For instance, Atzori *et al.* [13] compared a CNN with various conventional classifiers on their proposed sparse multichannel NinaPro databases. The average gesture recognition accuracies achieved by their CNN on the first and second subdatabases were 66.6% and 60.3%, respectively–considerably lower than those achieved by conventional classifiers such as random forests. Zhai *et al.* [27] fed a spectrogram of sEMG into a CNN for gesture recognition, but the gesture recognition accuracy achieved on the second subdatabase of NinaPro database was only 78.7%.

A number of reviews on multi-view learning have been published in recent years [14], [15], [28]. According to Zhao *et al.* [15], multi-view deep learning is one of the important problems for practical applications of multi-view learning because deep learning has recently demonstrated outstanding performance in a variety of tasks. Thus far, multi-view deep learning has been successfully applied to computer vision-based pattern recognition tasks. For example, Ge *et al.* [29] projected a 3D depth image onto three orthogonal planes to acquire its multi-view representations for hand pose estimation. They trained three CNNs in parallel to map a projected image of each view to its corresponding heat maps. Su *et al.* [30] rendered 2D images from multiple views for CNN-based 3D shape recognition, where each multi-view representation was separately passed through one CNN, and all the views were aggregated via a view-pooling layer.

Motivated by the aforementioned multi-view learning methods, in this paper we present a novel multi-view deep learning framework to improve the performance of sEMG-based gesture recognition.

III. Problem Statement

The problem of sEMG-based gesture recognition by deep learning can be formulated as follows:

$$\boldsymbol{y} = h_{\boldsymbol{\omega}}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x}$ denotes the input to the CNN, $h_{\boldsymbol{\omega}}$ denotes the CNN model with the learned parameters $\boldsymbol{\omega}$, and $\boldsymbol{y}$ represents classification results in the form of output softmax scores from the network. As most of the existing deep learning approaches for sEMG-based gesture recognition are based on end-to-end learning [6], [8], [9], [13], $\boldsymbol{x}$ usually takes the form of raw sEMG signals. Some studies have also attempted to use sEMG features such as the sEMG spectrogram as CNN input [27].

From the multi-view deep learning viewpoint of sEMG-based gesture recognition, a view construction process f_{vc} is usually needed to construct multiple views of sEMG $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N$ from raw sEMG signals, $\boldsymbol{x}$. This view contruction process can be formulated as follows:

$$\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N = f_{vc}(\boldsymbol{x}) \tag{2}$$

After view construction, the constructed views are fed into a multi-view CNN, denoted as $h_{\boldsymbol{\omega}_v}$, for gesture recognition:

$$\boldsymbol{y} = h_{\boldsymbol{\omega}_v}(\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_N) \tag{3}$$

In the following sections, we will delineate the details of our proposed view construction process and multi-view CNN architecture.

IV. Performance-Based View Construction

To construction multi-view representations of sEMG, we first extract the 11 classical sEMG feature sets listed in Section II. To extract Hudgins_FS, Du_FS, Phin_FS1 and TDAR, we discarded the ZC feature because it counts the number of times that the amplitude values cross zero [17]. However, the sEMG signals in some sEMG databases (e.g., the first subdatabase of NinaPro) contain only positive values [20].

After feature extraction, the extracted feature sets are subsequently transformed into their *signal images*. A signal image is a special permutation of sEMG channels that can cause a CNN to better capture the correlations among sEMG collected by different electrodes [31]. The generation algorithm of the signal image used in this paper is based on Algorithm 1 from [31]; however, we modified this algorithm slightly because it was unable to generate the desired signal image for sEMG with an even number of channels. Specifically, for sEMG with an even number of N_C channels, the input number of channels N_s is set to $N_C + 1$ instead of N_C.

The 11 extracted feature sets can be considered as 11 different views of sEMG. However, this approach may lead to two major problems. First, the total dimensionality of all 11 sEMG feature sets is high, and such a high dimensionality may increase the computational complexity and reduce the generalization ability of the model. Second, the 11 sEMG feature sets contain redundant information. For instance, Phin_FS1 is a subset of Doswald_FS, and Hudgins_FS is a subset of Atzori_FS. According to Kumar *et al.* [32], the success of multi-view learning relies on either the consensus principle or the complementary principle. The consensus principle means that the correlations among all views should be maximized, and the complementary principle means that each view should contain some information that other views do not contain.

TABLE I
Gesture Recognition Accuracies of the 11 Classical Feature Sets on NinaPro DB1, DB2, DB5 and BioPatRec DB2. The Bold Entries Indicate the Top5 Feature Sets

Feature set	Classification accuracy			
	NinaPro DB1	NinaPro DB2	NinaPro DB5	BioPatRec DB2
Du_FS	82.4%	**78.4%**	83.7%	88.2%
TSD	81.0%	59.6%	81.9%	87.7%
Atzori_FS	83.7%	8.1%	83.94%	87.6%
Phin_FS1	**85.4%**	**74.4%**	**85.8%**	**90.9%**
Phin_FS2	84.3%	60.0%	84.4%	88.5%
Doswald_FS	**85.3%**	61.7%	**85.2%**	**90.6%**
TDAR	**85.0%**	**75.6%**	84.5%	88.5%
DWTC	**85.7%**	**67.6%**	**86.0%**	**90.0%**
DWPTC	**85.9%**	57.2%	**85.8%**	**90.3%**
CWTC	84.4%	**61.8%**	**85.3%**	**89.6%**
Hudgins_FS	82.5%	37.9%	83.4%	88.2%

To reduce the computational complexity and construct more discriminative views that fulfill the complementary principle of multi-view learning, we conducted a performance-based view construction process to select the most discriminative views based on their gesture recognition performance. We extracted the 11 sEMG feature sets from the first and the fifth subdatabases of NinaPro [20], [33] using 200 ms time windows, transformed them into signal images, and then individually input these feature sets into a single-stream CNN for gesture recognition. Details of the classified gestures can be found in [33], [34]: the evaluation metric is the same as that used in previous studies [6], [20], [33] on these two databases. Specifically, on NinaPro DB1, the CNN was trained by the 1st, 3rd, 4th, 6th, 7th, 8th and 9th trials and tested by the remaining trials for each subject, and on NinaPro DB5, the CNN was trained by the 1st, 3rd, 4th and 6th trials and tested by the remaining trials for each subject. The gesture recognition accuracy was averaged over all subjects.

We chose datasets containing more gesture categories and subjects for our feature set selection experiments because the results are more informative. Therefore, DB1, DB2 and DB5 from the NinaPro dataset and DB2 from the BioPatRec dataset were chosen to evaluate the abovementioned sEMG feature sets. The gesture recognition accuracies are presented in Table I. The gesture recognition accuracies achieved by the Phin_FS1, Phin_FS2, Doswald_FS, TSD, CWTC, DWTC and DWPTC feature sets are higher than those achieved by other feature sets. Because these datasets are collected from different muscle areas via different types of sEMG electrodes, it is difficult to achieve a consistent recognition accuracy ranking for the evaluated 11 feature sets. The selection of feature sets is highly relevant to the dataset, and we provide a method for selecting views for multi-view learning.

To determine the optimal number of views for multi-view learning, we carried out an exhaustive search strategy. Specifically, we first selected 2 best-performing feature sets (those with

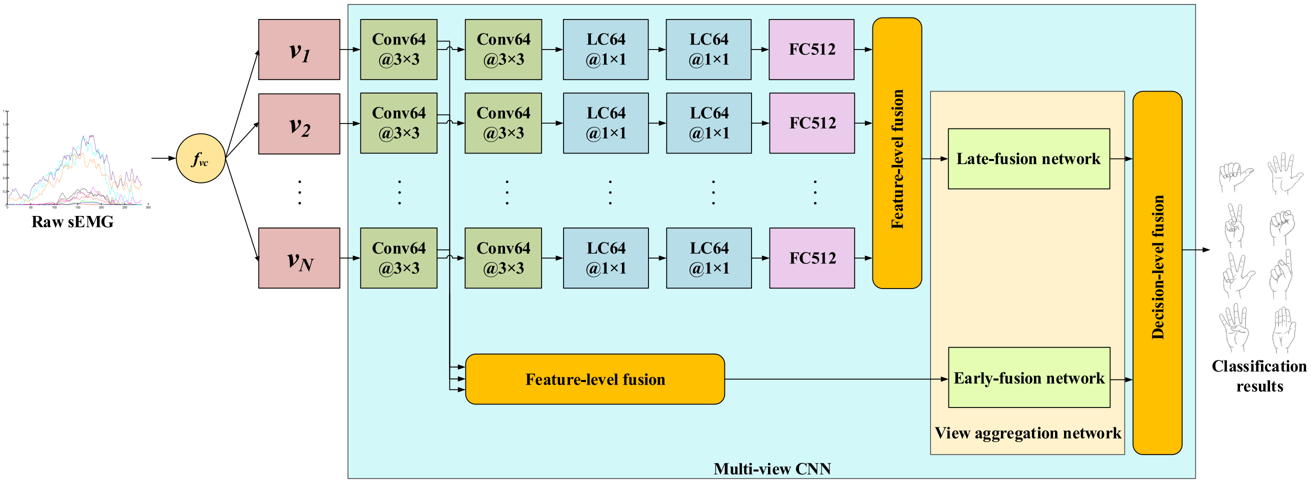

Fig. 1. Illustration of the proposed multi-view deep learning framework for sEMG-based gesture recognition. Conv, LC, FC and BN respectively denote the convolution layer, locally-connected layer, fully-connected layer and batch normalization, respectively. The number following the layer name denotes the number of filters, and the numbers after the ampersand (@) denote the convolution kernel size.

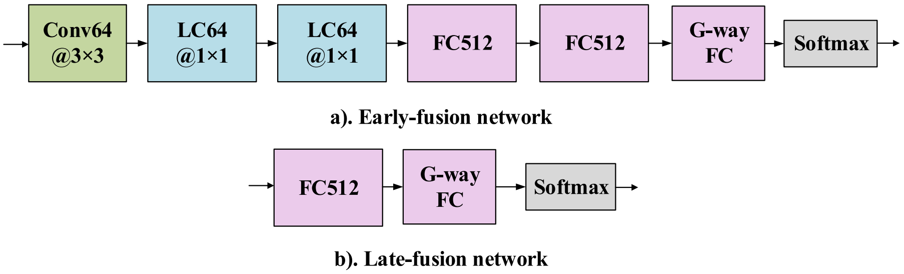

Fig. 2. Illustration of the early-fusion network and the late-fusion network from Fig. 1.

the top 2 gesture recognition accuracies) as 2 different views for multi-view gesture recognition, then we successively added the 3rd, 4th and 5th best feature sets. Fig. 3 shows the correlation between the number of selected feature sets and the gesture recognition accuracy. Increasing the number of feature sets improves the recognition accuracy. When the number of selected feature sets exceeds 3, accuracy growth is limited, but the computational cost increases dramatically. Considering application scenarios with high real-time requirements, such as embedded systems, we selected 3 views to balance recognition accuracy and computational cost.

V. Multi-View CNN

Motivated by multi-view deep learning approaches that input multiple views into multiple CNNs [29], [30], in this work, multiple views of sEMG are first modeled in parallel by a multistream CNN. As shown in Fig. 1, each CNN stream is composed of 2 convolutional layers, 2 locally connected (LC) layers and 1 fully connected (FC) layer. Each convolutional layer consists of 64 3 × 3 filters with a stride of 1 and a zero padding of 1. Each LC layer consists of 64 1 × 1 filters. The FC layer consists of 512 hidden units. Batch normalization and the ReLU nonlinearity function [38] are applied to each layer, and dropout [39] is applied to the FC layer and the last LC layer to prevent overfitting. To avoid potential interference caused by magnitude variations between the sEMG features extracted from different channels, we apply batch normalization before the first convolutional layer to normalize the CNN input.

The outputs of these multiple CNNs must be fused to obtain the final classification results. Generally, multistream fusion approaches can be divided into feature-level fusion and decision-level fusion [40]. The feature-level fusion approaches can be further subdivided into early fusion and late fusion [41]. Early fusion fuses the low-level deep features learned by the CNNs, while late fusion fuses the high-level deep features learned by the CNNs [42]. For multistream fusion problems in deep learning, determining the optimal multistream fusion approach is challenging.

In this paper, we propose a novel network architecture to take advantage of both early and late fusion, called a *view aggregation network*. The view aggregation network is composed of two subnetworks: the early-fusion network and the late-fusion network. As shown in Fig. 2, the early-fusion network fuses the outputs of the first layers of all the CNNs together, and the late-fusion network fuses the outputs of the final layers of all the CNNs. Suppose the outputs of the j_{th} layer of the i_{th} stream in the multistream CNN is $\boldsymbol{H}_i^j$. The inputs of the two subnetworks $\boldsymbol{H}_{\texttt{early}}$ and $\boldsymbol{H}_{\texttt{late}}$ are

$$\begin{aligned}\boldsymbol{H}_{\texttt{early}} &= fuse_{\texttt{F}}(\boldsymbol{H}_i^1, i = 1, 2, 3)\\ \boldsymbol{H}_{\texttt{late}} &= fuse_{\texttt{F}}(\boldsymbol{H}_i^5, i = 1, 2, 3),\end{aligned} \tag{4}$$

where $fuse_{\texttt{F}}(\cdot)$ denotes the feature-level fusion.

TABLE II
SPECIFICATIONS OF THE sEMG DATABASES USED IN THIS PAPER

Name	Number of gestures	Number of gestures to be classified	Intact subjets	Amputated subjets	Number of sEMG channels	Number of IMU channels			Number of trials	Trials for training	Trials for testing	Sampling rate
						Accelerometer	Gyroscope	Magnetometer				
NinaPro DB1 [20]	53	52	27	0	10	-	-	-	10	1,3,4,6,7,8,9	2,5,10	100Hz
NinaPro DB2 [20]	50	50	40	0	12	36	-	-	6	1,3,4,6	2,5	2000Hz
NinaPro DB3 [20]	50	50	0	11	12	36	-	-	6	1,3,4,6	2,5	2000Hz
NinaPro DB4 [33]	53	53	10	0	12	-	-	-	6	1,3,4,6	2,5	2000Hz
NinaPro DB5 [33]	53	53	10	0	16	3	-	-	6	1,3,4,6	2,5	2000Hz
NinaPro DB6 [35]	14	7	10	0	16	48	-	-	120	1,3,...,119	2,4,...,120	2000Hz
NinaPro DB7 [36]	41	41	20	2	12	36	36	36	6	1,3,4,6	2,5	2000Hz
BioPatRec DB1 [37]	10	10	20	0	4	-	-	-	3	1	2	2000Hz
BioPatRec DB2 [37]	26	26	17	0	8	-	-	-	3	1	2	2000Hz
BioPatRec DB3 [37]	10	10	8	0	4	-	-	-	3	1	2	2000Hz
BioPatRec DB4 [37]	8	8	8	0	16	-	-	-	3	1	2	2000Hz

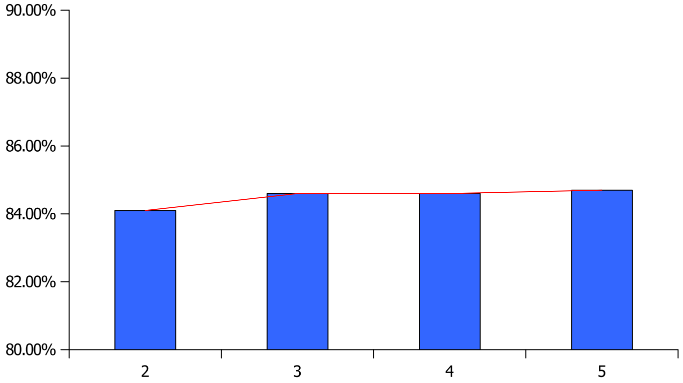

Fig. 3. Correlation between the number of selected feature sets and the gesture recognition accuracy.

Following the view aggregation network, the softmax scores of the two subnetworks are fused via decision-level fusion to obtain the final classification results; this process can be described as follows:

$$\boldsymbol{y}_{\texttt{final}} = fuse_{\texttt{S}}(\boldsymbol{y}_{\texttt{early}}, \boldsymbol{y}_{\texttt{late}}), \tag{5}$$

where $\boldsymbol{y}_{\texttt{early}}$ and $\boldsymbol{y}_{\texttt{late}}$ are the softmax scores of the early-fusion network and the late-fusion network, respectively, $fuse_{\texttt{S}}$ denotes decision-level fusion, and $\boldsymbol{y}_{\texttt{final}}$ denotes the final softmax scores.

VI. EXPERIMENTS AND RESULTS

In our work, the experiments are divided into three parts. First, the gesture recognition performance of the proposed view aggregation network is compared with that of its two subnetworks, the view-pooling, and the decision-level fusion based view aggregation approaches. Second, we conduct a performance comparison with state-of-the-art methods. Finally, we compare the proposed multi-view deep learning framework and single-view methods on both unimodal sEMG streams and multimodal sEMG and IMU data streams.

A. Database

Our evaluations were carried out on 7 subdatabases of the NinaPro database [20], [33], [35], [36] (denoted as NinaPro DB1-DB7) and 4 subdatabases of the BioPatRec database [37] (denoted as BioPatRec DB1-DB4). Among these databases, 5 contain IMU data, including Ninapro DB2, Ninapro DB3, Ninapro DB5, Ninapro DB6, and Ninapro DB7; these are denoted as sEMG^{plus} databases. The specifications of the databases mentioned above can be found in Table II.

NinaPro DB5 contains 16-channel sEMG collected by two Thalmic Myo armbands. The author of NinaPro DB5 [33] provided baseline accuracies for the double Myo setup (denoted as NinaPro DB5) and two single Myo setups (the upper Myo armband is denoted as NinaPro DB5-1, and the lower Myo armband is denoted as NinaPro DB5-2) when classifying 41 hand movements using the mDWT feature using a SVM classifier [33]. The same Myo setups were also adopted here.

B. Experimental Setup

The proposed multistream CNNs were implemented with MxNet [43], and trained using Stochastic Gradient Descent (SGD). For the experiments on NinaPro, the dropout was set to 0.5 during the pretraining stage, and was changed to 0.65 during the training stage. The total number of training epochs was set to 28. The learning rate was initialized to 0.1 and divided by 10 after the 16th and 24th epochs. For the experiments on BioPatRec DB, the dropout during both pretraining and training was set to 0.5. During the training stage, only one training epoch was used for each round of the cross-validation. To ensure sufficient training samples, the training sets of all subjects were used for pretraining, and for each subject, the CNN was initialized by the pretrained model. Because BioPatRec DB is a relatively small-scale database in which each movement was collected in only 3 trials, to ensure convergence we used all the available data in this database for pretraining. For fair comparisons, we set the same hyperparameters in the experiments on each dataset.

Following the experimental configuration used in [19], for the experiments on BioPatRec DB, we discarded the beginning and end of the predefined 70% contraction time for each trial. Due to memory limitations, for the experiments on NinaPro DB2, DB3, DB4, DB6 and DB7 we downsampled the sEMG data from 2000 Hz to 100 Hz.

For convenience during the performance comparison, we adopted the same evaluation metrics for intra-subject gesture recognition as those used in previous studies on these sEMG databases [6], [19], [20], [33], [35], [36]; these metrics can be found in Table II. For inter-subject gesture recognition on NinaPro DB1, we used Leave-One-Subject-Out Cross-Validation (LOSOCV): the data from each subject were used in turn as the test data, and the CNN was trained using the data of the remaining subjects. For the NinaPro DB2-DB7 and BioPatRec databases, we conducted fourfold cross-validation for inter-subject gesture recognition.

The inter-subject gesture recognition performance may be affected by the distribution difference between the training and the test data. To address this problem, our previous work [8] considered such a distribution difference as a domain adaptation problem and proposed a multistream AdaBN deep domain adaptation technique for sEMG-based gesture recognition. In this work, we also employed the multistream AdaBN deep domain adaptation framework [8] to reduce the negative effect introduced by distribution differences between the training and the test data, by using a small amount of the test data for adaptation. Specifically, for experiments on NinaPro DB1, the trials numbered 1, 3, 4, 5 and 9 of all 27 subjects were used for training and adaptation, and the trials numbered 2, 6, 7, 8 and 10 were used for testing in each fold of the cross-validation. More details on the multistream AdaBN framework and experimental configurations can be found in our published paper [8].

In all the experiments, we evaluated the gesture recognition accuracy, which is defined as

$$\text{Classification Accuracy} = \frac{\text{Number of correct classifications}}{\text{Total number of test samples}} * 100\%, \tag{6}$$

and averaged the gesture recognition accuracy over all subjects.

C. Evaluation of the Proposed View Aggregation Network

This subsection presents the evaluations conducted to validate the effectiveness of our proposed view aggregation network. First, compared the performance of the proposed view aggregation network and its two subnetworks. Then, we compared the gesture recognition performance achieved by the view aggregation network with the view-pooling and decision-level fusion based view aggregation approaches. The experiments in this section were performed on the NinaPro database. We adopted 200 ms time windows to extract the sEMG feature sets during view construction.

The gesture recognition accuracies achieved by the proposed view aggregation network and its two subnetworks are shown in Table III. The view aggregation network achieved higher gesture recognition accuracy than did its two subnetworks, which validated our hypothesis that the hybrid of early and late fusion takes advantages of both and improves the gesture recognition performance.

View-pooling was proposed by Su *et al.* [30] for view aggregation; it is essentially an elementwise maximum operation. According to Su *et al.* [30], the view-pooling layer is closely related to the max-pooling layer of a CNN. Su *et al.* considered performing view-pooling in different locations of the network. Following their study, we also evaluated view pooling at different network locations.

In addition to view-pooling based view aggregation, we also evaluated decision-level fusion based view aggregation, which fuses the softmax scores produced by the classifier for each view to obtain the final classification results. Suppose the softmax scores of the i_{th} stream are $\boldsymbol{y}_i$ and the final softmax scores are $\boldsymbol{y}_{\texttt{final}}$; here, we evaluated two decision-level fusion approaches for view aggregation:

TABLE III
GESTURE RECOGNITION ACCURACIES ACHIEVED BY THE VIEW AGGREGATION NETWORK AND ITS TWO SUBNETWORKS ON NINAPRO. THE RESULTS IN BOLD TEXT INDICATE THE BEST PERFORMANCES

Method	Database	Accuracy
View aggregation network	NinaPro DB1	**88.2%**
Early-fusion network	NinaPro DB1	87.5%
Late-fusion network	NinaPro DB1	87.9%
View aggregation network	NinaPro DB2	**83.7%**
Early-fusion network	NinaPro DB2	83.3%
Late-fusion network	NinaPro DB2	82.5%
View aggregation network	NinaPro DB5	**90.0%**
Early-fusion network	NinaPro DB5	89.5%
Late-fusion network	NinaPro DB5	89.7%

Score summation fusion: Elementwise summation of the softmax scores of all streams, i.e.,

$$\boldsymbol{y}_{\texttt{final}} = \sum_{i=1,2,3} \boldsymbol{y}_i \tag{7}$$

Score maximum fusion: Elementwise maximum of the softmax scores of all streams, i.e.,

$$\boldsymbol{y}_{\texttt{final}} = \max(\boldsymbol{y}_i, i = 1, 2, 3) \tag{8}$$

The network architectures of view-pooling and decision-level fusion based view aggregation are shown in Fig. 6. We discarded the pretraining step for all the experiments on NinaPro DB2 to reduce the computational cost and accelerate our experiments.

The gesture recognition accuracies are shown in Table IV, which shows that on NinaPro DB1 and DB2, our proposed view aggregation network achieves higher gesture recognition accuracy than do view-pooling and decision-level fusion based view aggregation approaches. On NinaPro DB5, their gesture recognition performances are quite similar.

D. Comparison With the State-of-the-Art Gesture Recognition Approaches

In this subsection, we present a performance comparison with state-of-the-art sEMG-based gesture recognition approaches to demonstrate the advantages of our proposed multi-view deep learning framework. This comparison was conducted on 4 sparse multichannel sEMG databases: NinaPro DB1, NinaPro DB2, NinaPro DB5, and BioPatRec DB2. The specifications of these databases, the experimental setups used and the evaluation metrics applied to these databases are described in Sections VI-A and VI-B.

As previous studies have noted that the controller delay of a real-time myoelectric control system should be kept below 300 ms [17], [45], we consider only 50 ms, 100 ms, 150 ms and 200 ms time windows to satisfy this constraint. The

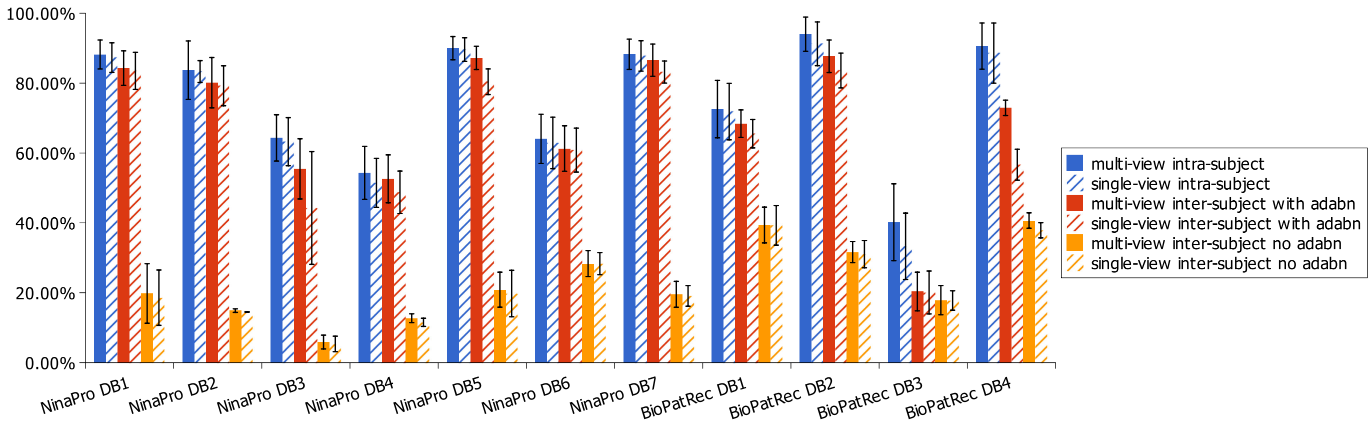

Fig. 4. Performance comparison of the multi-view and single-view deep learning approaches. The height of each column represents the average accuracy, while the error bars represent the standard deviation.

TABLE IV
PERFORMANCE COMPARISON OF THE VIEW AGGREGATION NETWORK, THE VIEW-POOLING BASED VIEW AGGREGATION APPROACHES, AND THE SCORE FUSION BASED VIEW AGGREGATION APPROACHES. THE RESULTS IN BOLD ENTRIES INDICATE THE BEST PERFORMANCES

Method	Database	Accuracy
View aggregation network	NinaPro DB1	**88.2%**
View-pooling at 1st Conv	NinaPro DB1	87.2%
View-pooling at 2nd Conv	NinaPro DB1	87.4%
View-pooling at 1st LC	NinaPro DB1	87.4%
View-pooling at 2nd LC	NinaPro DB1	87.6%
View-pooling at 1st FC	NinaPro DB1	87.5%
View-pooling at 2nd FC	NinaPro DB1	88.0%
Score summation fusion	NinaPro DB1	87.8%
Score maximum fusion	NinaPro DB1	87.7%
View aggregation network	NinaPro DB2	**81.4%**
View-pooling at 1st Conv	NinaPro DB2	80.7%
View-pooling at 2nd Conv	NinaPro DB2	80.1%
View-pooling at 1st LC	NinaPro DB2	80.7%
View-pooling at 2nd LC	NinaPro DB2	77.7%
View-pooling at 1st FC	NinaPro DB2	77.2%
View-pooling at 2nd FC	NinaPro DB2	76.3%
Score summation fusion	NinaPro DB2	77.0%
Score maximum fusion	NinaPro DB2	77.5%
View aggregation network	NinaPro DB5	90.0%
View-pooling at 1st Conv	NinaPro DB5	89.5%
View-pooling at 2nd Conv	NinaPro DB5	89.7%
View-pooling at 1st LC	NinaPro DB5	89.3%
View-pooling at 2nd LC	NinaPro DB5	89.7%
View-pooling at 1st FC	NinaPro DB5	89.7%
View-pooling at 2nd FC	NinaPro DB5	89.4%
Score summation fusion	NinaPro DB5	**90.1%**
Score maximum fusion	NinaPro DB5	90.0%

experimental results reported in previous studies as well as those achieved by our proposed multi-view deep learning framework (i.e., MV-CNN) on the evaluated sEMG databases are summarized in Table V. For the various window lengths below 300 ms, the gesture recognition accuracies achieved by MV-CNN on different sEMG databases are higher than those achieved by the state-of-the-art methods, and the performance gap is significant. For example, the baseline gesture recognition accuracy proposed by the author of NinaPro DB5 was 69.0%, which was achieved by an SVM classifier using mDWT features. Using the same evaluation metric, our proposed multi-view deep learning framework achieved a gesture recognition accuracy of 90.0% on NinaPro DB5.

E. Comparison Between Multi-View and Single-View Learning

Finally, we validated the effectiveness of multi-view deep learning (denoted as MV-CNN) compared to single-view deep learning approach (denoted as SV-CNN), which input the concatenation of multi-view representations into a single-stream CNN.

In this section, the comparisons between multi-view and single-view learning are divided into two parts: one part used sEMG databases and the other used sEMG^{plus} databases. The SV-CNN network architecture is illustrated in Fig. 5. For Ninapro DB5, the window length is 200 ms, and the window step is 100 ms. For the other NinaPro subdatabases, the window length and step are 200 ms and 10 ms, respectively. For all the BioPatRec subdatasets, the window length and step are 150 ms and 50 ms, respectively. All the hyperparameter settings are the same.

First, for the sEMG databases, the gesture recognition accuracies achieved by MV-CNN and SV-CNN are shown in Fig. 4. MV-CNN outperformed SV-CNN on a unimodal data stream of sEMG regarding both intra- and inter-subject gesture recognition.

Second, to demonstrate the superiority of our multi-view framework on multimodal data streams, we performed comparisons between multi-view and single-view methods on the sEMG^{plus} databases. For sEMG^{plus}, the SV-CNN input was the concatenation of multi-view representations of sEMG signals and the activity image [31] of the IMU signals, while the activity image of the IMU signals was inserted into MV-CNN as a new view. We adopted intra-subject gesture recognition accuracy as our evaluation metric. The experimental results on sEMG^{plus} are shown in Table VI.

These experimental results show that SV-CNN and MV-CNN performed significantly better on sEMG^{plus} data than on

TABLE V

GESTURE RECOGNITION ACCURACIES OF THE PROPOSED MV-CNN COMPARED WITH THE ACCURACIES OF EXISTING WORKS ON 4 sEMG DATABASES. THE BOLD ENTRIES INDICATE THE SCORES OF THE PROPOSED METHOD

Method	Database	Number of movements to be classified	Window length			
			50ms	100ms	150ms	200ms
random forests [20]	NinaPro DB1	50	-	-	-	75.3%
GengNet [6] (end-to-end CNN)	NinaPro DB1	52	-	-	-	77.8%
AtzoriNet [13] (end-to-end CNN)	NinaPro DB1	50	-	-	66.6±6.4%	-
multi-stream CNN [44]	NinaPro DB1	52	81.7%	83.4%	84.4%	85.0%
MV-CNN	NinaPro DB1	52	**85.8%**	**86.8%**	**87.4%**	**88.2%**
random forests [20]	NinaPro DB2	50	-	-	-	75.3%
AtzoriNet [13] (end-to-end CNN)	NinaPro DB2	50	-	-	60.3±7.7%	-
ZhaiNet [27] (single-view CNN)	NinaPro DB2	50	-	-	-	78.7%
MV-CNN	NinaPro DB2	50	**80.6%**	**81.1%**	**82.7%**	**83.7%**
SVM [33]	NinaPro DB5 (all Myo armbands)	41	-	-	-	69.0%
MV-CNN	NinaPro DB5 (all Myo armbands)	41	-	-	-	**90.0%**
SVM [33]	NinaPro DB5-1 (upper Myo armband)	41	-	-	-	55.3%
MV-CNN	NinaPro DB5-1 (upper Myo armband)	41	-	-	-	**83.9%**
SVM [33]	NinaPro DB5-2 (lower Myo armband)	41	-	-	-	54.8%
MV-CNN	NinaPro DB5-2 (lower Myo armband)	41	-	-	-	**83.1%**
LDA [19]	BioPatRec DB	26	86.3%	-	92.9%	-
MV-CNN	BioPatRec DB	26	**90.9%**	**-**	**94.0%**	**-**

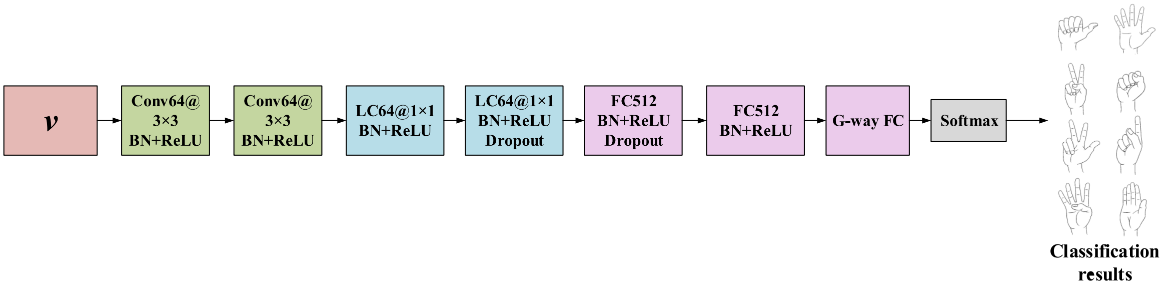

Fig. 5. Illustration of the network architecture of SV-CNN.

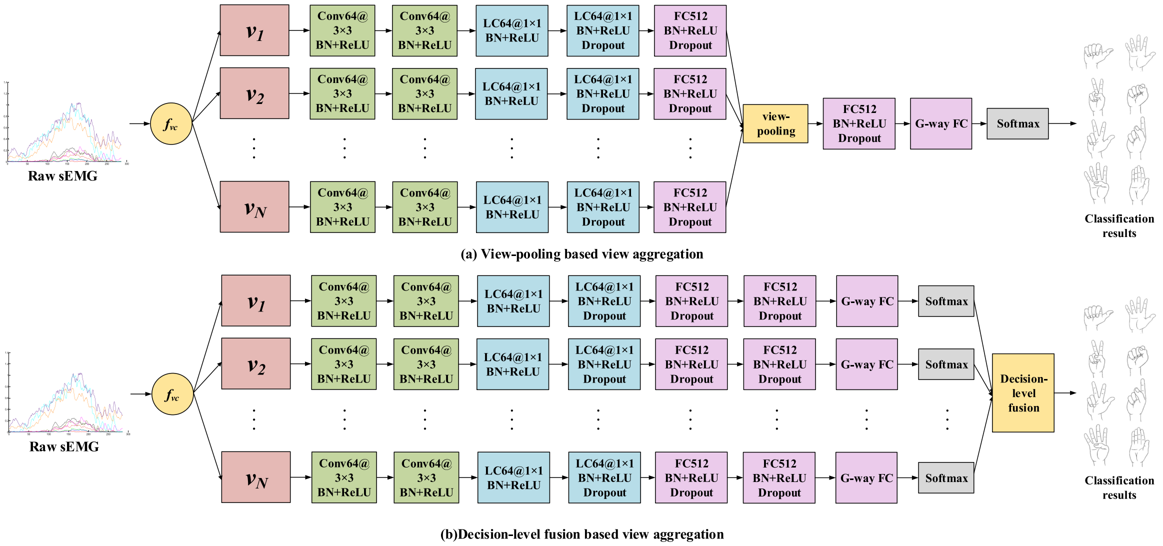

Fig. 6. Illustration of the view-pooling based view aggregation and score fusion based view aggregation approaches. Conv, LC, FC and BN respectively denote the convolution layer, locally-connected layer, fully-connected layer and batch normalization. The numbers following the layer name denote the number of filters, and the numbers after the ampersand (@) denote the convolution kernel size.

TABLE VI
PERFORMANCE COMPARISON BETWEEN MULTI-VIEW AND SINGLE-VIEW DEEP LEARNING APPROACHES FOR BOTH WITH AND WITHOUT IMU DATA. THE RESULTS IN BOLD TEXT INDICATE THE BEST PERFORMANCES. WE APPLIED THE SAME HYPERPARAMETERS IN THE EXPERIMENTS ON EACH DATASET

Database	Method	Intra-subject gesture recognition accuracy	
		With IMU	Without IMU
Ninapro DB2	multi-view	**94.40%**	83.70%
	single-view	92.84%	83.30%
Ninapro DB3	multi-view	**87.06%**	64.30%
	single-view	84.97%	63.30%
Ninapro DB5	multi-view	**91.31%**	90.00%
	single-view	90.22%	89.60%
Ninapro DB6	multi-view	**77.10%**	64.10%
	single-view	73.99%	62.90%
Ninapro DB7	multi-view	**94.54%**	88.30%
	single-view	92.48%	87.80%

sEMG data due to the presence of the modal IMU data stream. Table VI shows that MV-CNN outperformed SV-CNN overall on the sEMG^{plus} data; its average improvement reaches approximately 1.98%. Fig. 4 shows that MV-CNN performed better than did SV-CNN on the sEMG databases. Therefore, multi-view learning is an effective way to improve the CNN gesture recognition preformance on both unimodal sEMG data streams and multimodal sEMG and IMU data streams.

VII. CONCLUSION

To achieve better gesture recognition performance of sEMG-based HCI using sparse multichannel sEMG, in this paper, we presented a novel multi-view deep learning framework that extracts multiple classical sEMG feature sets and transforms them into multi-view representations of sEMG. To reduce the computational complexity and ensure the complementary principle, the view construction process is based on gesture recognition performance: the most discriminative multi-view representations were selected and subsequently input into a multi-view CNN for gesture recognition.For view selection, we evaluated 11 feature sets on 4 datasets and evaluated the correlations between the number of selected views and the gesture recognition accuracy.

We also performed experiments on both sEMG data and sEMG^{plus} data to demonstrate the superiority of the proposed multi-view deep learning framework. For sEMG data, we compared our multi-view framework with the state-of-art methods on 11 sparse multichannel sEMG databases, including 7 subdatabases from the NinaPro databases and 4 sub-databases from the BioPatRec databases, using both intra- and inter-subject evaluation metrics. The experimental results show that the proposed network achieved higher gesture recognition accuracies than those achieved by either the single-view deep learning framework or the state-of-the-art methods. For sEMG^{plus} data, our multi-view framework was compared with the single-view deep learning framework on 5 databases that contain IMU data, including NinaPro DB2, NinaPro DB3, NinaPro DB5, NinaPro DB6, and NinaPro DB7, using an intra-subject evaluation metric. The results showed that our multi-view framework outperformed the single-view framework by an average of 1.98%, which is a convincing demonstration of the superiority of multi-view framework.

Our future work will focus on the application of more representative feature sets and the mutual information (MI)-based feature selection techniques to discard redundant features and build up more discriminative multi-view representations. We also plan to exploit the correlations between different fingers during hand movements and employ state-of-the-art multilabel learning techniques to achieve better gesture recognition performances. Moreover, we will also focus on utilizing a Generative Adversarial Network(GAN) [46] to improve the gesture recognition preformance.

REFERENCES

[1] M. Turk and G. Robertson, "Perceptual user interfaces (introduction)," *Commun. ACM*, vol. 43, no. 3, pp. 32–34, 2000.

[2] J. Wei *et al.*, "Classification of human hand movements using surface EMG for myoelectric control," in *Proc. Adv. Comput. Intell. Syst., Contributions Presented 16th U.K. Workshop Comput. Intell.*, 2017, pp. 331–339.

[3] Y. Fan and Y. Yin, "Active and progressive exoskeleton rehabilitation using multisource information fusion from EMG and force-position EPP," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3314–3321, Dec. 2013.

[4] W. Guo *et al.*, "An enhanced human-computer interface based on simultaneous sEMG and NIRS for prostheses control," in *Proc. IEEE Int. Conf. Inf. Autom.*, 2014, pp. 204–207.

[5] Y. Li *et al.*, "A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2695–2704, Oct. 2012.

[6] W. Geng *et al.*, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, 2016, Art. no. 36571.

[7] M. Ergeneci *et al.*, "An embedded, eight channel, noise canceling, wireless, wearable sEMG data acquisition system with adaptive muscle contraction detection," *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 1, pp. 68–79, Feb. 2018.

[8] Y. Du *et al.*, "Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation," *Sensors*, vol. 17, no. 3, 2017, Art. no. E458.

[9] Y. Du *et al.*, "Semi-supervised learning for surface EMG-based gesture recognition," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 1624–1630.

[10] A. Phinyomark *et al.*, "Feature reduction and selection for EMG signal classification," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7420–7431, 2012.

[11] M. F. Lucas *et al.*, "Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization," *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 169–174, 2008.

[12] M. A. Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, 2007.

[13] M. Atzori *et al.*, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers Neurorobot.*, vol. 10, pp. 9–18, 2016.

[14] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, no. 7, pp. 2031–2038, Dec. 2013.

[15] J. Zhao *et al.*, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, 2017.

[16] D. Farina and R. Merletti, "Comparison of algorithms for estimation of EMG variables during voluntary isometric contractions," *J. Electromyography Kinesiol.*, vol. 10, no. 5, pp. 337–349, 2000.

[17] B. Hudgins *et al.*, "A new strategy for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 40, no. 1, pp. 82–94, Jan. 1993.

[18] Y. C. Du *et al.*, "Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4283–4291, 2010.

[19] R. N. Khushaba *et al.*, "A framework of temporal-spatial descriptors-based feature extraction for improved myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1821–1831, Oct. 2017.

[20] M. Atzori *et al.*, “Electromyography data for non-invasive naturally-controlled robotic hand prostheses,” *Sci. Data*, vol. 1, 2014, Art. no. 140053.

[21] A. Phinyomark *et al.*, “EMG feature evaluation for improving myoelectric pattern recognition robustness,” *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, 2013.

[22] A. Doswald *et al.*, “Advanced processing of sEMG signals for user independent gesture recognition,” in *Proc. Mediterranean Conf. Med. Biol. Eng. Comput.*, 2013, pp. 758–761.

[23] H. Huang *et al.*, “Ant colony optimization-based feature selection method for surface electromyography signals classification,” *Comput. Biol. Med.*, vol. 42, no. 1, pp. 30–38, 2012.

[24] F. Duan *et al.*, “sEMG-based identification of hand motion commands using wavelet neural network combined with discrete wavelet transform,” *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1923–1934, Mar. 2016.

[25] K. Kiatpanichagij and N. Afzulpurkar, “Use of supervised discretization with PCA in wavelet packet transformation-based surface electromyogram classification,” *Biomed. Signal Process. Control*, vol. 4, no. 2, pp. 127–138, 2009.

[26] J. Kilby and H. G. Hosseini, “Extracting effective features of SEMG using continuous wavelet transform,” in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 1704–1707.

[27] X. Zhai *et al.*, “Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network,” *Frontiers Neurosci.*, vol. 11, pp. 379–389, 2017.

[28] S. Sun, “A survey on multi-view learning,” *Neural Comput. Appl.*, vol. 23, no. 7–8, pp. 2031–2038, 2013.

[29] L. Ge *et al.*, “Robust 3D hand pose estimation in single depth images: From single-view CNN to multi-view CNNs,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3593–3601.

[30] H. Su *et al.*, “Multi-view convolutional neural networks for 3D shape recognition,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 945–953.

[31] W. Jiang and Z. Yin, “Human activity recognition using wearable sensors by deep convolutional neural networks,” in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1307–1310.

[32] V. Kumar and S. Minz, “Multi-view ensemble learning: An optimal feature set partitioning for high-dimensional data classification,” *Knowl. Inf. Syst.*, vol. 49, no. 1, pp. 1–59, 2016.

[33] S. Pizzolato *et al.*, “Comparison of six electromyography acquisition setups on hand movement classification tasks,” *PLoS ONE*, vol. 12, no. 10, pp. 1–17, Oct. 2017.

[34] M. Atzori *et al.*, “Building the NinaPro database: A resource for the biorobotics community,” in *Proc. IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics*, 2012, pp. 1258–1265.

[35] F. Palermo *et al.*, “Repeatability of grasp recognition for robotic hand prosthesis control based on sEMG data,” in *Proc. Int. Conf. Rehabil. Robot.*, 2017, pp. 1154–1159.

[36] A. Krasoulis *et al.*, “Improved prosthetic hand control with concurrent use of myoelectric and inertial measurements,” *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 71–84, 2017.

[37] M. Ortiz-Catalan *et al.*, “BioPatRec: A modular research platform for the control of artificial limbs based on pattern recognition algorithms,” *Source Code Biol. Med.*, vol. 8, no. 1, pp. 11–28, Apr. 2013.

[38] A. Krizhevsky *et al.*, “ImageNet classification with deep convolutional neural networks,” in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[39] N. Srivastava *et al.*, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[40] P. K. Atrey *et al.*, “Multimodal fusion for multimedia analysis: A survey,” *Multimedia Syst.*, vol. 16, no. 6, pp. 345–379, 2010.

[41] J. Wagner *et al.*, “Multispectral pedestrian detection using deep fusion convolutional neural networks,” in *Proc. Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn.*, 2016, pp. 509–514.

[42] J. Liu *et al.*, “Multispectral Deep Neural Networks for Pedestrain detection,” in *Proc. British Machine Vision Conf.*, pp. 73.1–73.13, Sep. 2016.

[43] T. Chen *et al.*, “MXNet: A flexible and efficient machine learning library for heterogeneous distributed systems,” in *Proc. Neural Inf. Process. Syst., Workshop Mach. Learn. Syst.*, 2015, pp. 1–6.

[44] W. Wei *et al.*, “A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface,” *Pattern Recognit. Lett.*, vol. 119, pp. 131–138, 2017.

[45] 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.

[46] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 27, pp. 2672–2680.