Sensors and Actuators A 311 (2020) 112046

Contents lists available at ScienceDirect

Sensors and Actuators A: Physical

journal homepage: www.elsevier.com/locate/sna

sEMG-based recognition of composite motion with convolutional neural network

Shuhao Qi[a], Xingming Wu[a], Wei-Hai Chen[b,*], Jingmeng Liu[a], Jianbin Zhang[c], Jianhua Wang[a]

[a] School of Automation Science and Electrical Engineering, Beihang University, 100191 Beijing, China
[b] College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China
[c] School of Mechanical Engineering and Automation, Beihang University, 100191 Beijing, China

ARTICLE INFO

Article history:
Received 26 July 2019
Received in revised form 28 April 2020
Accepted 7 May 2020
Available online 24 May 2020

Keywords:
Surface EMG
Motion recognition
Convolution neural network
Transfer learning
Sensor fusion

ABSTRACT

Surface electromyography (sEMG) signals are expected to help recognize motions precisely and timely for its generation origins from muscular contractions. In most cases, existing researches about sEMG-based motion recognition cannot guarantee comprehensively excellent performance and apply flexibly to different types of motions. This paper proposes a new initiative via deep learning to recognize general composite motions, which processes sEMG signals as images. Inspired by several definitions of "sEMG Image" for static gestures, we define a novel "sEMG Image" to represent composite motions, which can make different cooperation of muscles reflected on image textures. With a well-designed convolutional neural network (CNN), this method can obtain effective filters automatically to extract features of texture by the training of considerable data. The results from two experiments of different composite motion recognition (including gentle writing motions and drastic sign language motions) indicate that this method embraces high accuracies and strong generalization ability for several influence factors. In addition, two techniques are proposed to further optimize this method: pre-train the network with an irrelevant dataset to reduce the demand for data and speed up the convergence; fuse multiple sensors with simple modifications for CNN to improve performance greatly.

1. Introduction

With the advancement of artificial intelligence, human–computer interaction (HCI) can be exploited to improve people's quality of life and work efficiency. Surface electromyography (sEMG) is a crucial type of physiologically electrical signal, representing real-time human motion intents. Because of this characteristic, sEMG-based HCI system exhibits a wide range of applications. In general, those applications can be split into two types, discrete classification (e.g., motion recognition [1] and diagnosis [2]) and continuous prediction (e.g., force estimation [3], trajectory prediction [4] and control of rehabilitation robots [5,6]). sEMG-based motion recognition has unique superiorities compared with those based on other sensors. On the one hand, unlike vision-based recognition that requires cameras mounted in the environment, sEMG-based recognition has not to suffer from space constraints because of the wearability. On the other hand, it is unnecessary to place sEMG electrodes at each joint like the data-glove based on Inertial Measurement Unit (IMU), because complex motions can be predicted with sEMG signals of several corresponding muscles. Therefore, the study about sEMG-based recognition possesses practical potentials by virtue of its wearability, rich information and noninterference to motions. So it has attracted the interest of researchers all over the world in recent decades.

In earlier studies, to improve the performance of recognition, researchers made many forces to explore discernible features and powerful classifiers. In order to improve the recognition accuracy, some researchers believed that ideal features can be acquired by extracting the useful information which is hidden in sEMG signals and removing the unwanted part and interferences. For instance, Phinyomark et al. [7] compared 37 time-domain and frequency-domain sEMG features for gesture recognition and analyzed their superiorities and weaknesses. In addition, various classifiers have been tested, including support vector machine (SVM) [8,9], random forest (RF) [10], linear discriminant analysis (LDA) [11], Gaus-

* Corresponding author.
E-mail addresses: qishuhao@buaa.edu.cn (S. Qi), wuxingming307@126.com (X. Wu), whchenbuaa@126.com (W.-H. Chen), jmliubuaa@126.com (J. Liu), jbzhangbuaa@163.com (J. Zhang), jhwangbuaa@126.com (J. Wang).

https://doi.org/10.1016/j.sna.2020.112046
0924-4247/

sian mixture model (GMM) [12], neuro-fuzzy system [13], etc. To determine a proper combination, Guo et al. [14] compared the performance of different recognition methods for 8 upper limb motions with 8 combinations of four feature extractions and two classification methods. Nevertheless, hand-crafted features always have their limitations, and traditional classifiers have no ability to represent complex mapping relations. Therefore, although plenties of combinations of features and classifiers were tested, the performance of sEMG-based recognition is far from the requirements of practical application.

In this decade, the usage of deep learning algorithms exhibits remarkable superiorities in supervised learning. With the increase of computing power and high-quality dataset, deep learning algorithms have achieved huge improvements in many application scenarios. In recent years, an increasing number of researchers have attempted to apply deep-learning algorithms on sEMG-based motion recognition. The network structure used by these studies is mostly convolutional neural network (CNN) and its variants. The study of Yang et al. [4] indicated that CNN-based method can effectively decode complex wrist movements with three degrees of freedom (DOF) from raw sEMG signals. Côté-Allard et al. [15] proposed to put spectrograms and continuous wavelet transform (CWT) of sEMG signals into CNN. Since CNN was originally designed for processing images, some researchers proposed several different concepts of "sEMG Image". Ding et al. [16] converted the sEMG signals to sEMG images by sliding window and established a parallel multiple-scale convolution architecture. The sEMG Image defined by Geng et al. [17] and Wei et al. [18] is a matrix of instantaneous sEMG signal offered by high-density electrodes where each electrode can be regarded as a pixel of sEMG images. Based on the work of Geng et al. [19] proposed a new sEMG image representation based on a traditional features vector and established an attention-based hybrid CNN and Recurrent neural network (RNN) architecture. All of the above methods have achieved more excellent performance than conventional methods. Atzori et al. [20] showed that CNN with a very simple architecture can produce accurate results comparable to the average classical classifiers. Therefore, deep-learning methods embrace a promising potential in the field of sEMG-based recognition.

In fact, as for motion recognition, there are two different types of targeted motions according to whether they are static or dynamic in motion units to be recognized, static gestures and composite motions. As shown in Fig. 1, static gestures (e.g., hand gestures) have no changes in motion units and composite motions (e.g., sign language motions, handwriting motions) are dynamic in motion units. In other words, composite motion consists of a dynamic sequence of gestures. There are no essential differences in the identification methods between these two types of motions, and the recognition for composite motions requires additional consideration of the time dimension compared to recognitions for static gestures. For example, Wu et al. [21] proposed a method that utilized IMU and sEMG signals to achieve sign language recognition with the selected features and SVM. Zhang et al.[22] proposed to utilize multimodal-CNN and long short term memory (LSTM) to achieve sentence-level recognition based on four types of wearable sensors (3-axis accelerometer, gyroscope, orientation, and sEMG). Huang et al.[23] proposed a method to recognize the handwritten numbers through comparison of sEMG samples and templates with dynamic time warping (DTW). However, whether for static gestures or composite motions, it is hard to apply the method of one task to another task because each method always is designed to specific tasks other than general tasks.

To conclude, there are two unresolved problems for existing researches which prevent sEMG-based motion recognition from achieving wide-range applications, imperfect recognition performance (including accuracy, generalization ability, adaptability, etc.) and incompatibility among different tasks. In response to the above two limitations, we hope to find a way which can achieve excellent performance for different recognition tasks. Considering limitations of hand-crafted features and conventional classifiers, deep learning is a promising way to breakthrough the limitations. Inspired by the definition of "sEMG Image" [17], we plan to define a new "sEMG Image" to present general composite tasks. The main contributions of this paper are as follows:

1. It is the first attempt to develop a general method to recognizing composite motions. We define a new "sEMG Image" to represent composite motions generally. Therefore, whether for drastic sign language motions and gentle writing motions, different cooperations of muscles can be reflected on the texture of sEMG images only if electrodes are placed at corresponding muscles. With a well-designed CNN, effective texture features would be extracted automatically through the training of considerable data, which will not be restricted by specific models. In this paper, the effectiveness for different composite motions is verified by two entirely different experiments, the recognition of handwritten digits and the recognition of 20 sign language motions.
2. By fully exploiting the potential of CNN, the present method exhibits excellent performance. Referring to results of the experiments, CNN achieves higher accuracies compared with classic methods. When it comes to different subjects and changes of writing speed and writing size, this method exhibits strong generalization ability. Additionally, we propose two means of deep learning to refine the proposed method: pre-train the network with an irrelevant dataset to speed up convergence and reduce the demand for data; fuse the information of others sensors with simple modification.
3. Without hand-crafted features and complex models, this method can be applied conveniently and timely in various scenarios for engineers and researchers.

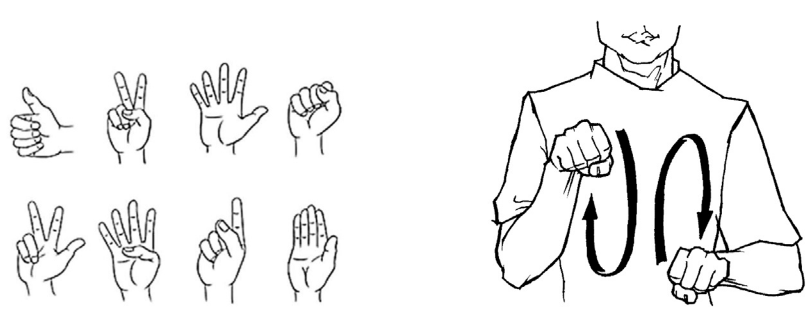

Fig. 1. Examples of static gesture (left) and composite motion (right).

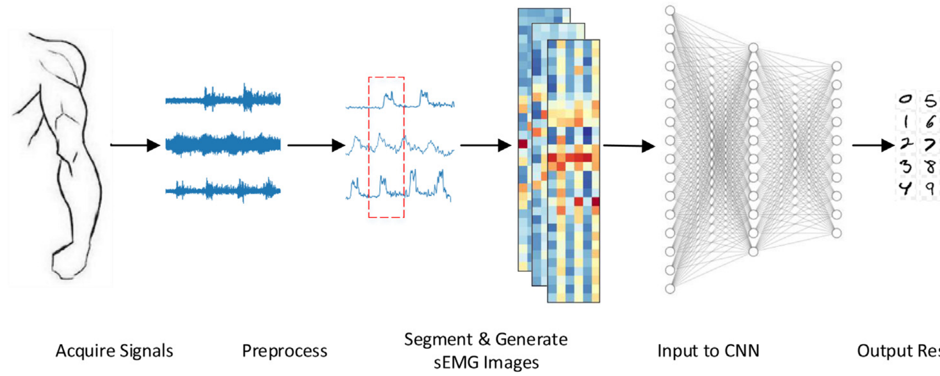

Fig. 2. Schematic presentation of the proposed method.

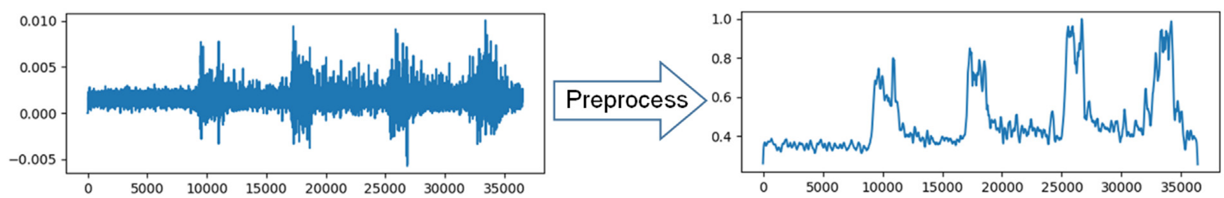

Fig. 3. Display of preprocessing.

2. Method

In this section, we will describe the proposed method in detail, which includes signals preprocessing, motion segmentation, sEMG Images generation, and establishment of CNN. The schematic of the proposed method is shown in Fig. 2. Subsequently, two optimization methods based on the proposed method are elaborated.

2.1. Signal preprocessing

EMG signals refer to potential differences caused by contraction of muscular fibers, which are controlled by the central nervous system. Compared with implanted electrodes that can pinpoint the corresponding muscle fibers, surface electrodes collect signals non-invasively, but collected signals are relatively weak and instable due to the superposition of signals from multiple muscles and the filtering by skin and fatty. Additionally, sEMG signals are susceptible to effects from external and internal factors (e.g., noise from external environment, friction between skin and electrodes, etc.). Therefore, sEMG signals are supposed to be preprocessed to improve signal-to-noise ratio. Potvin et al. [24] discovered “Less is more” phenomenon that removing 99% of the raw sEMG signal power will result in significant and substantial improvements in biceps force estimates. The reason for the phenomenon is that the low-frequency part contains redundant information, such as fatigue of muscles. In addition, Potvin et al. [25] stated that there was the non-liner relationship between the sEMG amplitudes and force observed for most muscles, which can be described by the following formula,

$$a(t) = \frac{e^{Ae(t)} - 1}{e^{A} - 1} \quad (1)$$

where $e(t)$ denotes raw sEMG value, $a(t)$ denotes the value after non-linear mapping and A is a constant. De Luca [26] stated that the amplitude of surface EMG signal is qualitatively related to the amount of torque (or force) measured about a joint, whereas an accurate quantitative relationship is elusive. Since our task aimed to classify the actions rather than estimate accurate numerical value of force, we set the parameters A empirically in the nonlinear transformation. The sampling frequency of the sEMG acquisition equipment that we used, the Trigno Wireless EMG System, is up to 2000Hz. Referring to signals processing of Potvin et al. [24], the preprocessing will be achieved in the following steps and the display of that is illustrated in Fig. 3:

1. The raw sEMG signals are low-pass filtered with a cutoff of 950 Hz using a third-order Butterworth filter.
2. This raw sEMG signals are high-pass filtered with a cutoff of 750 Hz using a third-order Butterworth filter.
3. The raw sEMG signals are full-wave rectified.
4. The filtered sEMG signals are linearly normalized as a percentage of the maximum sEMG amplitude of the trial.
5. This signals are then non-linearly normalized with Eq. (1) and proper down sampling.

2.2. Motion segmentation

As mentioned above, composite motions are comprised of sequences of gestures. Therefore, motion segmentation is necessary to discern the starting point and the terminal point of motion unit precisely and automatically. In this study, the method of sliding rectangular window is used to analyze consecutive signals. This method could distinguish whether the signals of unit window are belong to the process of motion according to the energy value Q. For multichannel signals, Q refers to the energy sum of the whole channels and its equation is as follows:

$$Q = \sum_{n=1}^{N} \sum_{t=0}^{W} (a_n(t) - B_n)^2 \quad (2)$$

where N denotes the number of channels, W denotes the length of unit window, $a_n(t)$ denotes the value of signals, B_n denotes the value of signal when the corresponding muscle is fully relaxed. Besides, threshold value to distinguish the states should be set in advance. To enhance adaptability, the threshold will be set according to the

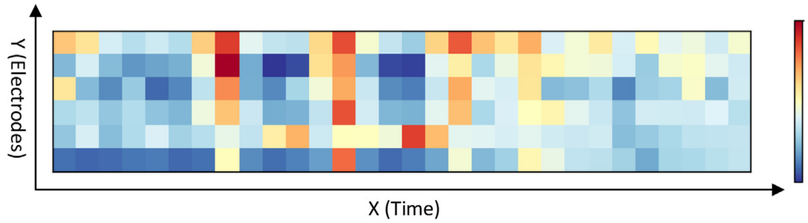

Fig. 4. Example of sEMG image.

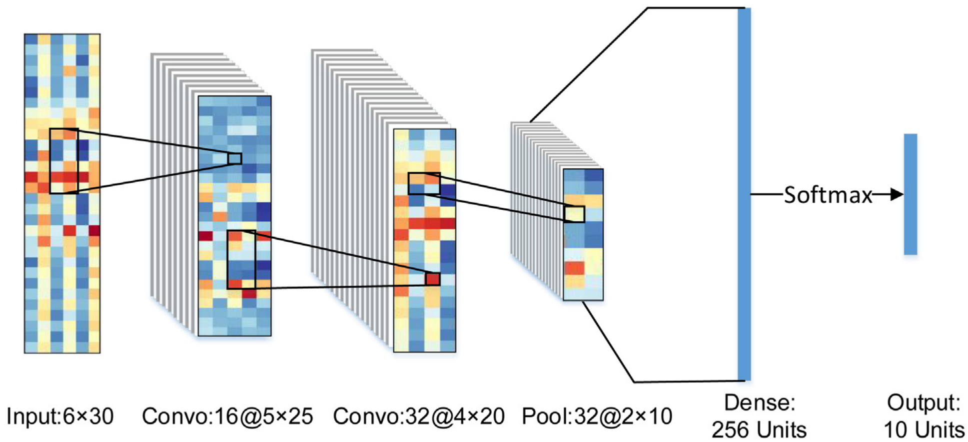

Fig. 5. Structure of the proposed CNN.

values of maximal voluntary contraction (MVC) and the potential of complete relaxation:

$$T = \eta \cdot W \sum_{n=1}^{N} (MVC_n - B_n)^2 \tag{3}$$

where MVC_n denotes the sEMG signal of the corresponding channel produced by the maximal voluntary contraction and η is an experiential constant. Motion is considered to begin when signals exceed the threshold for a certain period of time, and vice versa.

2.3. sEMG image

To represent static gestures, Geng et al.[17] defined signals from two-dimension high-density electrodes as "sEMG images" where each pixel represented instantaneous activations of corresponding electrodes. As for composite motion, there is a dynamic process comprised of a sequence of static gestures. So, in order to represent composite motion with image, the time dimension must be considered. Therefore, we define a new "sEMG image" as follows. *X*-axis of sEMG image represents time dimension, and Y-axis represents the multichannel of electrodes. According to the electrodes arrangement, *Y*-axis can also be considered as activation of different muscles. So each pixel of sEMG images represents the activation degree of corresponding electrode at a certain time. Take the sEMG image used in the experiment of handwritten digits for example. After motion segmentation, each signal sequence of 6 channels is split into 30 parts equally and each part would be averaged so that 6×30 matrix is obtained, which considered as "sEMG image". From an example of sEMG image in Fig. 4, stroke of digits means cooperation of different muscles and further unique array of muscular activation, which ultimately can be reflected in texture of sEMG image. In other words, texture features of sEMG images can reflect correlations between muscular activations and time sequences.

2.4. Convolution neural network

As a typical method of deep learning, CNN has been widely used in the field of image processing. The convolution kernels trained by considerable quantity of data correspond to operators which can extract specific features automatically. Furthermore, CNN [27] has the characteristics of translation invariance and scale invariance based on their shared-weights architecture, which can eliminate effects from motion difference in amplitude and trajectory.

Based on the structure of LeNet-5 [28], a convolutional neural network is designed in this study, adapted to sEMG images, as shown in Fig. 5. In the first layer, set 16 filters whose shapes are 2×6. As we all know, filters of CNN are often assumed to be square. In fact, it is because the horizontal and vertical dimension in the field of computer vision usually makes no difference. And convolution filters are not limited to square in other areas (e.g., NLP and radio signals). Considering the shape of sEMG image, convolution kernel is designed to the size of 2×6. Subsequently, rectified linear unit (ReLU) is selected as the activation function, which is expressed as Eq. (4). It effectively removes negative values from an activation map by setting them to zero, increasing the nonlinear properties of the network. ReLU is often prioritized compared with other functions because it facilitates the process of training without reducing accuracies. Similar to the first layer, the second layer is a convolution layer with 32 filters. Subsequently, set max pooling layer to reduce computational burden without reducing precision. Then, max pooling layer is connected to the dense layer and SoftMax function, which is Eq. (5). Finally, predicted results are generated according to probabilities.

$$f(x) = \begin{cases} 0, & x \le 0 \\ x, & x > 0 \end{cases} \tag{4}$$

$$S_i = \frac{e^i}{\sum_j e^j} \tag{5}$$

Furthermore, to prevent overfitting, the dropout strategy [29] is adopted in the fully-connected layer. At each training stage, individual nodes are either "dropped out" with probability $1 - p$ or keep with probability p and only the reduced network is trained on the data in that stage. Subsequently, the removed nodes are reinserted into the network with their original weights. In addition, the cross entropy, as shown in Eq. (6), characterizes the distance between the actual output probability p and the expected output probability q. It is general to utilize the cross entropy to evaluate training performance of the network. Adam [30], an optimization algorithm, is utilized to train the proposed network. This algorithm combines Adagrad [31] and RMSprop [32], which leverages the power of adaptive learning rates methods to find individual learning rates for each parameter.

$$H(p,q) = -\sum_{x}(p(x)\log q(x) + (1 - p(x))\log(1 - q(x))) \tag{6}$$

2.5. Transfer learning

As is vastly recognized, one of the major factors for accurate predictions of deep learning algorithms is the amount of training data available, i.e., deep learning algorithms are data-hungry. Additionally, the expense of sEMG signals' collection is huge because of its arduous process for subjects. Therefore, how to get excellent results based on small datasets should be considered. Transfer learning [33] seems a solution to this problem because it advocates transferring knowledge from other fields to improve performance of targeted task. To be specific, major advantages of transfer learning are to speed up convergence, improve accuracies and reduce demands for data. Through observation, there is abundant texture information in sEMG images. And MNIST is a classic dataset of handwritten number pictures, which also have abundant texture information. Therefore, the present paper attempts to adopt fine-tuning, a classic strategy of transfer learning, to learn capabilities of features extraction for texture information. The process of fine-tuning is presented as follows:

1. Pre-train the network by MNIST and save weight values of the network after training.
2. Fix weight values of first two layers with pre-trained values and train the fully-connected layer with the dataset of sEMG image.

2.6. Sensors fusion

Sensors fusion is a common strategy to improve performance when considering limitations of individual sensor. Since sEMG signals directly represent information about muscular activation instead of positions, many researches [34,22] of motion recognition prefer to combine information from multiple sensors (e.g., gyroscope, accelerometer, etc.) to achieve excellent results. Although this paper focus on sEMG-based motion recognition, the proposed method can fuse multiple sensors with simple modification for the network. Based on the definition of sEMG image, we proposed a new 3-dimension data frame, including signals of sEMG and accelerometers. Delsys myoelectric electrodes we used are integrated with 3-axis accelerometers so that acceleration information can be recorded simultaneously with the sEMG signals. Like the processing of sEMG signals, 3D acceleration signals would be segmented and converted into matrixes. Thus, the input data is the matrix of size $4 \times 6 \times 30$, including sEMG data of size $1 \times 6 \times 30$ and acceleration data of size $3 \times 6 \times 30$. Referring to the structure of 2D-CNN mentioned above, the structure of 3D-CNN is designed to fit 3D data frame, as shown in Fig. 6. In the first layer, set 32 filters whose shapes are $2 \times 2 \times 6$ and set ReLU as the activation function. The second layer is the convolution layer that has 64 filters of size $2 \times 2 \times 6$. Subsequently, set max pool so as to reduce computational burden without reducing precision. Finally, set the dense layer and then output the results normalized by SoftMax function. The optimization algorithm, loss function, drop strategy is identical as 2D-CNN.

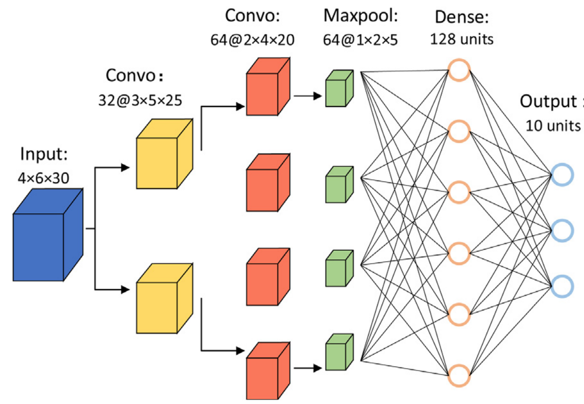

Fig. 6. Structure of 3D-CNN.

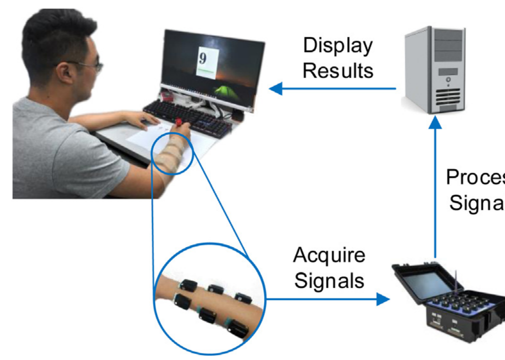

Fig. 7. Process of data collection.

3. Experiments and results

In order to testify the universality of this method, the proposed method is applied to two different tasks of composite motion recognition, the recognition of handwriting motions and sign language motions. Handwriting motions are slight and intricate, while sign language motions are wide-range and multi-joints. Both experiments conduct full and comprehensive verifications and also apply classic methods to compare with the proposed method. In addition, the superiority of Transfer Learning will be testified in the first experiment, the handwriting motions recognition; And the superiority of sensors fusion will be testified in the second experiment, the sign language motions recognition.

Although recently some sEMG datasets were published, such as NinaPro [35], there are no open-source sEMG data sets available for the composite motions targeted by the proposed method. So we were supposed to establish a dataset. The equipment we used is Trigno Wireless sEMG System, which is high-performance and easy-to-use. The wireless electrodes send signals to base station with the frequency of 2000 Hz, and computer acquires multichannel signals from the base station by Universal Serial Bus (USB). Furthermore, we designed a program to collect labeled signals based on Python. In order to facilitate the one-to-one correspondence between data and labels, set the 3-seconds acquisition interval when the subjects must complete motions according to on-screen instructions. Fig. 7 exhibits the process of data collection in the experiment of recognizing handwritten digits. In both exper-

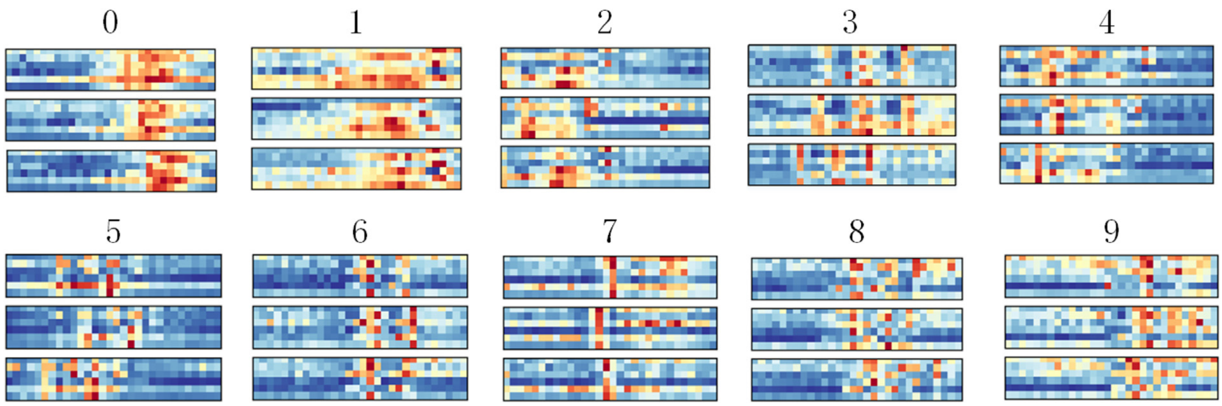

Fig. 8. sEMG images of writing digits.

iments, 3 healthy subjects helped to complete the data collection for different motions. Before collection process, the values of totally relaxing and maximal voluntary contraction for each subject were recorded in advance. During the process, subjects were required to avoid extra motions and unnecessary muscular contractions.

Subsequently, CNN and transfer learning (TransL) method are established with the help of TensorFlow [36], an end-to-end open-source platform for deep learning. In addition, we select three classical and general methods to compare with the proposed method, including support vector machine (SVM), random forest (RF) and dynamic time warping (DTW). Referring to the work of Chen et al. [8] and Su et al. [10], SVM and RF are built based on scikit-learn [37], a free software machine learning library. Classic features will be extracted and sent into classifiers for training, including mean absolute average, zero-crossing numbers, variance, autoregressive coefficient, and median frequency. In addition, referring to the method proposed by Huang et al. [23], DTW will be applied to compare with the proposed method in the experiment of handwriting numbers. With the help of the open-source project [38] in the Github, DTW algorithm is established to measure the similarity between two temporal sequences. Compared with the template, the most similar action in multiple channels is considered to be the result of the prediction. Finally, to derive a more accurate estimate of model prediction performance, fivefold cross-validation [39] is adopted.

3.1. Experiment I: handwritten digits recognition

Writing motions are mainly manipulated by the cooperation of forearm muscles. Therefore, place six wireless electrodes in the position of subject's forearm as shown in Fig. 7. To prevent electrodes from shaking, the medical self-adhesive strap was used to fix the electrodes. Subsequently, acquisition process mentioned above was repeated 100 times for writing one digits and signals were recorded simultaneously. In sum, $3 \times 10 \times 100$ labeled signal sequences have been collected. By motion segmentation algorithm, the average time of extracted signals is about 1.2 s. And then, signals are converted into sEMG images. Fig. 8 exhibits Subject-1's 3×10 sEMG images of writing digits 0–9. It is obvious that different strokes lead to different texture features and different samples of the same digit possess uniform textures, which indicates the potential of separability.

With validation set and training set split by cross-validation, classifiers of this method and comparative methods are trained and the results are listed in Table 1, which exhibits accuracies of different classifiers for different subjects. It is noted that the accuracies in Table 1 are obtained based on the dataset from the same experimental session of the same subject. Results of analysis of variance (ANOVA) presented in Table 2 indicate that there are significant differences ($p < 0.05$) among subjects and methods. Through Tukey test, the results presented in Table 3 indicate five methods can be divided into three homogeneous subsets: performances of CNN, TransL and DTW are significantly superior to RF and SVM; performance of DTW reaches the level of CNN, but there is still significantly inferior to TransL. Additionally, from Table 4, the

Table 1
Accuracies (%) of handwritten digits recognition experiment.

Method	Subject		
	Subject-1	Subject-2	Subject-3
SVM	56.2	51.2	44.1
RF	58.4	62.7	50.5
DTW	75.8	82.4	65.4
CNN	79.2	86.5	75.3
TransL	83.3	89.1	80.2

Table 2
Results of the two-way ANOVA for recognition accuracy.

Source	Type III sum of squares	df	Mean square	F	Sig.
Corrected model	0.293	6	0.049	52.760	0.000
Intercept	7.207	1	7.207	7798.406	0.000
Subject	0.033	2	0.017	17.863	0.001
Method	0.260	4	0.065	70.208	0.000
Error	0.007	8	0.001		
Total	7.506	15			

Table 3
Results of Tukey HSD for method.

Method	N	Subset		
		1	2	3
SVM	3	0.5050		
RF	3	0.5720		
DTW	3		0.7453	
CNN	3		0.8033	0.8033
TransL	3			0.8400
Sig.		0.140	0.227	0.602

Table 4
Results of Tukey HSD for subject.

Method	N	Subset	
		1	2
Subject-3	5	0.6306	
Subject-1	5		0.7052
Subject-2	5		0.7436
Sig.		1.000	0.175

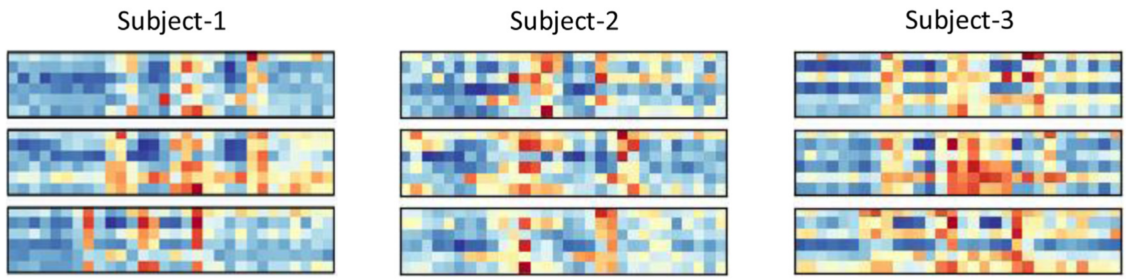

Fig. 9. sEMG images of writing digits “3” among three subjects.

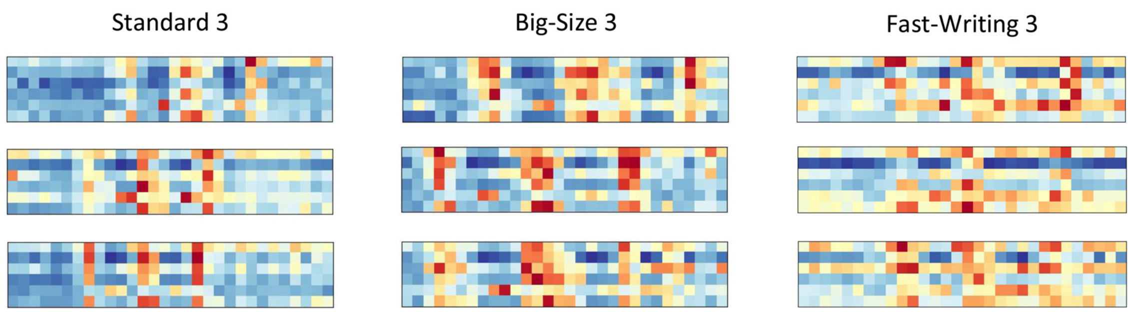

Fig. 10. sEMG images of writing digits “3” under different writing modes.

Table 5
Accuracies (%) of different writing modes.

Method	Standard	Different Size	Different speed
SVM	56.2	42.3	44.5
RF	58.4	45.8	49.2
DTW	75.8	47.4	54.7
CNN	79.2	71.6	75.3
TransL	83.3	76.9	80.1

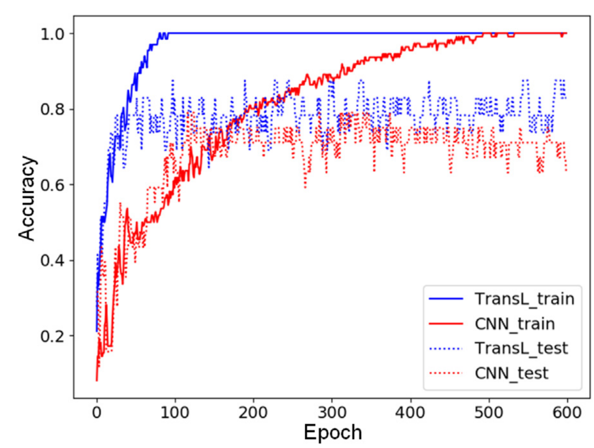

Fig. 11. Training process of CNN and TransL.

performance of Subject-3 is significant inferior to those of Subject-2 and Subject-3. Through the observation and comparison of writing habits and sEMG Images of three subjects, which are presented in Fig. 9, we found that writing motions of Subject-3 were more stable and forceful so that his muscles kept high-level activation during writing. It can be concluded that writing habits of individuals is a considerable influence factor to the performance of recognition.

To further test the generalization ability of this method, effects of different writing habits are considered. Specifically, sEMG signals of Subject-1 are recorded under different writing modes, including writing big-size digits and writing with faster speed. Fig. 10 exhibits the sEMG images of writing digits “3” under different writing modes. By the way, the reason for choosing digit “3” is that its texture is representative and easy to compare. Compared with sEMG images of standard mode, textures of writing big-size digits are stretched and reinforced because amplitudes of strokes are larger. As for fast-writing mode, image textures are more indistinct and fuzzy than those of standard mode due to the increase of motion intensity. Statistically, the average time of extracted signals of fast-writing model is about 0.6 seconds, which is almost half of standard mode. With signals of writing big-size digits as validation set, the generalization ability with respect to written digit size is expected to be tested; similarly, the generalization ability with respect to writing speed will be tested through making signals of fast-writing mode as validation set. Subsequently, the results of this method and comparative methods are listed in Table 5. As the results show, the performance of these methods has declined in various degree when considering different speeds of writing and digit sizes, while the declining degree of DTW is more serious than CNN because DTW relies on fixed templates. Through Tukey test, the performance of CNN and TransL is significantly more superior than three other methods. One possible explanation to excellent performance of deep-learning methods is the invariance of CNN improved by the max pool. Thus, in order to verify this point, max pools were removed so that the accuracy decreased by 3–8%. Therefore, it can be concluded that CNN and TransL possesses stronger generalization ability to reduce the effects from different writing habits.

In order to verify superiorities of TransL, training processes of CNN and TransL were recorded in Fig. 11. TransL converged in less than 200 epoches, whereas CNN basically converged after 500 epoches. It is clear that the convergence of TransL is faster and the accuracy is higher, which indicates that the knowledge obtained from pre-training by MNIST is valuable for this task. Subsequently, to compare degree of data demand between these two methods, convergent test accuracy of different-size datasets were recorded, as shown in Fig. 12. This Figure suggests that TransL performed better with the same-size dataset. Overall, it can be concluded that the pre-training by MNIST enhanced performance, sped up convergence and reduced the demand for data. The video in the adjunct

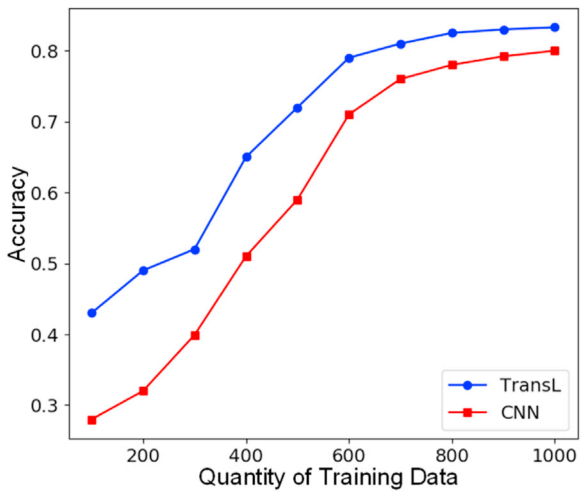

Fig. 12. Accuracy curve with respect to data quantity.

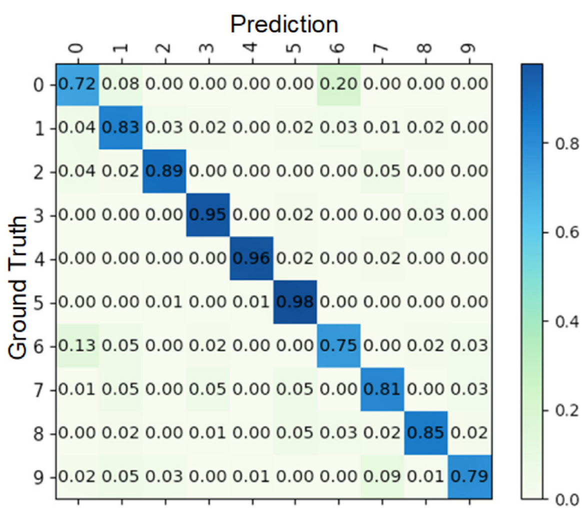

Fig. 13. Confusion matrix of handwritten digits.

exhibits the process of data collection and online recognition of handwriting numbers. The online recognition of number 0-9 does not always work and the performance is slightly lower than offline recognition. From confusion matrix, as shown in Fig. 13, recognition accuracies of the numbers 3, 4 and 5 were very high, while 0 and 6 were easy to confuse each other. After excluding confusing digits 6–9, the accuracy of online recognition for 0–5 numbers is about 85%, as shown in the video.

3.2. Experiment II: sign language recognition

Sign language was created to help hearing-impaired people to express their thoughts, while most of hearing people cannot understand the meaning of sign language motions. Thus, automatic sign language recognition is significant to break the barriers between the deaf community and hearing people. Similar to handwritten numbers, 20 signs of American Sign Language (ASL) used in daily conversations are selected in our paper. Three subjects, who were beginners of ASL, imitated the motions of ASL from the video in the Internet and repeats each sign 30 times. In addition, not only sEMG signals but also acceleration signals are recorded simultaneously. Six electrodes were placed on the positions as shown in Fig. 14. Similarly, the medical self-adhesive strap was used to fix the electrodes.

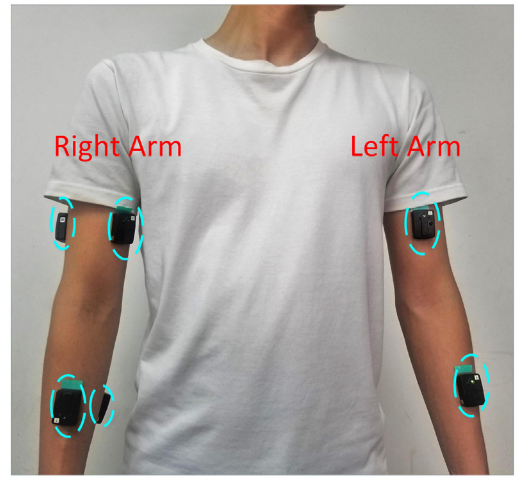

Fig. 14. Electrode positions of sign language experiment.

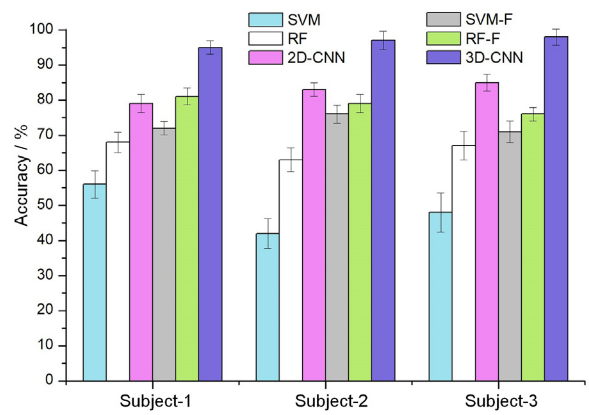

Fig. 15. Accuracies of sign language motion recognition.

sEMG signals reflect activation of corresponding muscles other than accurate position, so pure sEMG signals seem unsubstantial and deficient. Unlike writing motions, sign language motions are wide-range and multi-joint movements, which are appropriate to fuse information of acceleration signals and sEMG signals. Thus, acceleration signals are introduced to identify the sign language motions with the sEMG signals. In this experiment, we also selected SVM and RF as comparable methods. Referring to the studies of Zhang et al. [34] and Su et al. [10], in addition to features of pure sEMG signals, features of acceleration signals (e.g., absolute average, variance, and autoregressive coefficient) are fed into SVM and RF simultaneously, which are named as SVM-F and RF-F to distinguish from SVM and RF based on pure sEMG signals. As mentioned above, 3D-CNN is established by slight modifications of 2D-CNN. Subsequently, sEMG signals and acceleration signals are sent into the 3D-CNN. Finally, the results of the proposed methods and comparative methods are exhibited in Fig. 15. In general, methods which fuse acceleration and sEMG signals exhibit better performance than corresponding methods which rely on pure sEMG signals. Whether or not the acceleration signal is fused, CNN has obvious advantages over SVM and RF. Tukey test also reveals that 3D-CNN is significantly superior than other methods including 2D-CNN. And 3D-CNN has reached accuracies of 94–97% on three different subjects, which reaches the applicable level in real life.

Furthermore, with simple modification of 3D-CNN, the accuracy based on only acceleration signals is obtained, which is compared with accuracies of pure sEMG and sensor fusion in Table 6. Accuracies based on separate sensor are significantly inferior to that

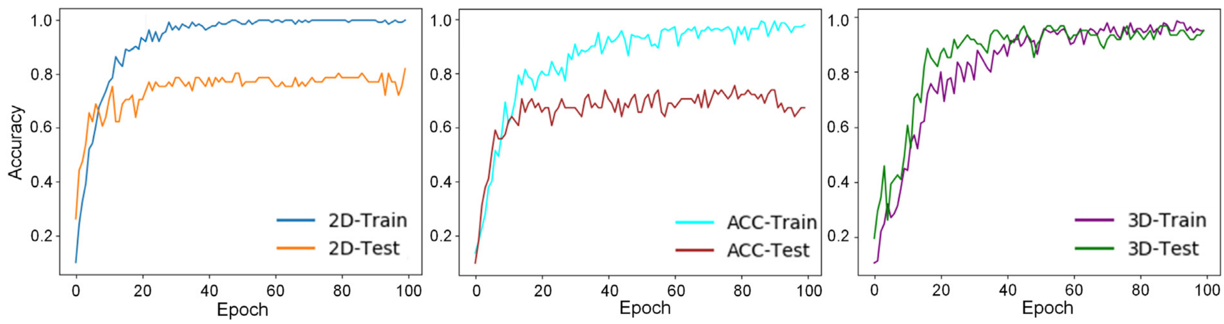

Fig. 16. Training processes of different information source.

Table 6
Sensors fusion comparative experiment.

Information source	Pure sEMG	Pure Acc	Sensors Fusion
Test Acc	0.78 ± 0.03	0.69 ± 0.03	0.94 ± 0.02

of sensors fusion, which is also verified in the training process in Fig. 16. Unlike tendencies of curves based on sensors fusion, test accuracy of methods based on separate sensor cannot continue to ascend with training accuracy, which reveals the limitations of information resource. Therefore, it is advisable to say that sensors fusion provides sufficient extractable information and helps network convergent to more excellent performance.

4. Discussion

As for gentle writing movements that needs complex cooperation of forearm muscles, plenty of uncontrollable factors will influence the activation of each muscle, such as changes of sitting posture, muscle fatigue and changes of stroke order. These factors determine that this task is challenging. From final results, although sEMG-based recognition of handwriting movement still cannot go outside the laboratory, the proposed method exhibits excellent performance compared with conventional methods. As for wide-range and multi-joint sign language motions, this method has reached the applicable level in real life. The excellent performance on both different tasks verify universal effectiveness for dynamic composite motions. In addition, this method has no complex model, no complex concept and no complex network. Therefore, it is convenient for researchers and engineers to apply this method to their scenarios which need to recognize composite motions.

In both experiments, deep-learning methods are superior to conventional classifiers in general. The reasons to the inferiority of conventional classifiers (i.e., SVM and RF) are lack of capacity that extracts correlations among various features, and limitation from handcrafted features. Through training, convolutional filters in CNN are adept to extract texture information of sEMG image which represents correlations between muscular activation and time sequence. And the results also have revealed that generalization ability of CNN can reduce the effects from different writing habits. Therefore, it can be demonstrated that this method fully take advantage of superiorities of CNN to achieve excellent performance. However, sEMG signals [40] are highly subject-specific and vary considerably even between experimental sessions of the same subject within the same experimental paradigm. From results of the inter-subjects experiments we tried, inter-session accuracies have significant inferiority compared with intra-session accuracies and inter-subject differences exceeds the generalization ability of CNN. In further research, model adaptation techniques will be introduced to address the inter-session and inter-subject problems.

Furthermore, this paper proposes two techniques to explore the potential of deep-learning methods. First, to reduce the considerable demand for data, it is practical to transfer knowledge from an irrelevant dataset. Second, this method can fuse multiple sensors conveniently with simple modifications, which can improve the performance significantly. The successes of these two technology applications remind us that there are still huge potentials for deep-learning methods in the fields of sEMG-based recognition. In the future, we hope we can utilize deep-learning method to achieve sentence-level recognition of sign language motions with the help of semantic knowledge.

5. Conclusion

In this paper, we proposed an easy-to-use and high-performance method to recognize universal composite motions based on sEMG signals. This method proposes initiatively to make correlation of time sequence and muscular activation reflected in texture of sEMG images, and utilize CNN to extract texture information. Through the experiments which applied this method to the recognition for two different composite motions, excellent performances on both different tasks verified the universal effectiveness for various composite motions. By exploiting the potential of CNN, this method is demonstrated to breakthrough limitations of conventional methods and possesses stronger generalization ability to reduce effect from influence factors. The successes of two techniques (transfer learning and sensor fusion) make us believe that there are still huge potentials for deep-learning methods in the fields of sEMG-based recognition.

Authors' contribution

Shuhao Qi: methodology, software, validation, writing – original draft. Xingming Wu: supervision, writing – review & editing. Weihai Chen: supervision, project administration. Jingmeng Liu: supervision, writing – reviewing and editing. Jianbin Zhang: writing – reviewing and editing, funding acquisition. Jianhua Wang: writing – review & editing, funding acquisition.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant Nos. 51675018 and 61773042, and the Special Funding for Top Talents of Shandong Province.

References

[1] M. Shafivulla, V. Rajesh, H. Khan, SEMG based human computer interface for robotic wheel, 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI) (2012) 1–5.

[2] P. Palmes, W. Tech Ang, F. Widjaja, L.C.S. Tan, W.-L. Au, Pattern mining of multichannel SEMG for tremor classification, IEEE Trans. Bio-med. Eng. 57 (2010) 2795–2805.
[3] C. Wu, H. Zeng, A. Song, B. Xu, Grip force and 3d push-pull force estimation based on SEMG and GRNN, Front. Neurosci. 11 (2017) 1–15.
[4] W. Yang, D. Yang, Y. Liu, H. Liu, Decoding simultaneous multi-dof wrist movements from raw emg signals using a convolutional neural network, IEEE Trans. Human–Mach. Syst. 49 (5) (2019) 411–420.
[5] L. Peternel, T. Noda, T. Petri, A. Ude, J. Morimoto, J. Babi, Adaptive control of exoskeleton robots for periodic assistive behaviours based on EMG feedback minimisation, PLOS ONE 11 (2) (2016) 1–26.
[6] K.-R. Mun, Z. Guo, H. Yu, Restriction of pelvic lateral and rotational motions alters lower limb kinematics and muscle activation pattern during over-ground walking, Med. Biol. Eng. Comput. 54 (2016) 1621–1629.
[7] A. Phinyomark, P. Phukpattaranont, C. Limsakul, Feature reduction and selection for EMG signal classification, Expert Syst. Appl. 39 (2012) 7420–7431.
[8] X. Chen, Z. Wang, Pattern recognition of number gestures based on a wireless surface EMG system, Biomed. Signal Process. Control 8 (2013) 184–192.
[9] A. Alkan, M. Günay, Identification of EMG signals using discriminant analysis and SVM classifier, Expert Syst. Appl. 39 (1) (2012) 44–47.
[10] S. Ruiliang, C. Xiang, C. Shuai, Z. Xu, Random forest-based recognition of isolated sign language subwords using data from accelerometers and surface electromyographic sensors, Sensors 39 (2016) 1–15.
[11] A.J. Young, L.H. Smith, E.J. Rouse, L.J. Hargrove, Classification of simultaneous movements using surface EMG pattern recognition, IEEE Trans. Biomed. Eng. 60 (5) (2013) 1250–1258.
[12] Z. Ju, H. Liu, A generalised framework for analysing human hand motions based on multisensor information, IEEE Int. Conf. Fuzzy Syst. (2012) 1–6.
[13] B. Alexandre, F. Gabriela, A neuro-fuzzy system for characterization of arm movements, Sensors 13 (2) (2013) 2613–2630.
[14] S. Guo, M. Pang, B. Gao, H. Hirata, H. Ishihara, Comparison of SEMG-based feature extraction and motion classification methods for upper-limb movement, Sensors 15 (2015) 9022–9038.
[15] U. Côté-Allard, C.L. Fall, A. Drouin, A. Campeau-Lecours, B. Gosselin, Deep learning for electromyographic hand gesture signal classification using transfer learning, IEEE Trans. Neural Syst. Rehabil. Eng. 99 (2018) 1–15.
[16] Z. Ding, C. Yang, Z. Tian, C. Yi, Y. Fu, F. Jiang, SEMG-based gesture recognition with convolution neural networks, Sustainability 10 (2018) 1–12.
[17] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, J. Li, Gesture recognition by instantaneous surface EMG images, Sci. Rep. 6 (36571) (2016) 1–8.
[18] W. Wei, Y. Wong, Y. Du, Y. Hu, M. Kankanhalli, W. Geng, A multi-stream convolutional neural network for SEMG-based gesture recognition in muscle-computer interface, Pattern Recogn. Lett. 119 (2017) 131–138.
[19] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, W. Geng, A novel attention-based hybrid cnn-RNN architecture for SEMG-based gesture recognition, PLOS ONE 13 (2018) 1–18.
[20] M. Atzori, M. Cognolato, H. Müller, Deep learning with convolutional neural networks applied to electromyography data: a resource for the classification of movements for prosthetic hands, Front. Neurorobot. 10 (2016) 9.
[21] J. Wu, L. Sun, R. Jafari, A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors, IEEE J. Biomed. Health Informatics 20 (5) (2016) 1281–1290.
[22] Q. Zhang, D. Wang, R. Zhao, Y. Yu, Myosign: Enabling End-To-End Sign Language Recognition With Wearables, 2019, pp. 650–660.
[23] G. Huang, D. Zhang, X. Zheng, X. Zhu, An EMG-based handwriting recognition through dynamic time warping, Proc. IEEE Eng. Med. Biol. Soc. (2010) 1–4.
[24] J.R. Potvin, S.H.M. Brown, Less is more: high pass filtering, to remove up muscle force estimates, J. Electromyogr. Kinesiol. 14 (3) (2004) 389–399.
[25] J.R. Potvin, R.W. Norman, S.M. McGill, Mechanically corrected EMG for the continuous estimation of erector spinae muscle loading during repetitive lifting, Eur. J. Appl. Physiol. Occup. Physiol. 74 (1) (1996) 119–132.
[26] C.J. De Luca, The use of electromyography in biomechanics, J. Appl. Biomech. 13 (1997) 135–163.
[27] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016 http://www.deeplearningbook.org.
[28] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.
[30] D. Kingma, J. Ba, Adam: a method for stochastic optimization, International Conference on Learning Representations (2014) 1–15.
[31] J. Duchi, E. Hazan, Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Mach. Learn. Res. 12 (7) (2011) 257–269.
[32] T. Tieleman, G. Hinton, Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude, Coursera: Neural Netw. Mach. Learn. 4 (2) (2012) 26–31.
[33] T. Lisa, S. Jude, Transfer learning, Handb. Res. Mach. Learn. Appl. Trends (2010) 242–264.
[34] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, J. Yang, A framework for hand gesture recognition based on accelerometer and EMG sensors, IEEE Trans. Syst. Man Cybern. Part A Syst. Humans 41 (6) (2011) 1064–1076.
[35] M. Atzori, A. Gijsberts, et al., Electromyography data for non-invasive naturally controlled robotic hand prostheses, Sci. Data 1 (140053) (2014) 1–13.
[36] M. Abadi, P. Barham, et al., Tensorflow: a system for large-scale machine learning, 12th USENIX Symposium on Operating Systems Design and Implementation (2016) 265–283.
[37] F. Pedregosa, G. Varoquaux, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[38] P. Rouanet, DTW (Dynamic Time Warping) Python Module, 2019 https://github.com/pierre-rouanet/dtw.
[39] G. Seni, J.F.E. Iv, Ensemble methods in data mining: improving accuracy through combining predictions, Synth. Lect. Data Mining Knowl. Disc. 2 (2010) 1–126.
[40] Y. Du, W. Jin, W. Wei, Y. Hu, W. Geng, Surface EMG-based inter-session gesture recognition enhanced by deep domain adaptation, Sensors 17 (3) (2017) 458.

Biographies

Shuhao Qi obtained the B.Eng. degree from University of Electronic Science and Technology of China in 2017 and the M.Eng. degree from Beihang University, China, in 2020. His main research directions are pattern recognition, bio-electrical signal processing and robotics.

Xingming Wu obtained his B.Eng. and M.Eng. degrees from Zhejiang University, China, in 1985 and 1988, respectively. He has been with the School of Automation Science and Electronic Engineering at Beihang University, China, as an Associate Professor from 1992 and as a Professor since 2017. His main research directions are intelligent sensor systems, embedded systems, autonomous mobile robots, and image processing.

Weihai Chen received the B.Eng. degree from Zhejiang University, China, in 1982, and the M.Eng. and Ph.D. degrees from Beihang University, China, in 1988 and 1996, respectively. He has been with the School of Automation, Beihang University, as an Associate Professor from 1998 and as a Professor since 2007. He has published over 200 technical papers in referred journals and conference proceedings and filed 18 patents. His research interests include bio-inspired robotics, computer vision, image processing, precision mechanism, automation, and control.

Jingmeng Liu received the B.S. degree from Anhui Polytechnic University, Wuhu, China, in 1991 and the M.S. and Ph.D. degrees from Beihang University, Beijing, China, in 2000 and 2004, respectively. He is currently an Associate Professor with the School of Automation Science and Electrical Engineering, Beihang University. His current research interests include actuators, precision control, and mechatronics.

Jianbin Zhang obtained his B.Eng. and M.Eng. degrees from Beihang University, China, in 1983 and 1986, respectively. He is currently a Professor in the School of Mechanical Engineering and Automation, Beihang University, Beijing, China. His research interest is mainly focused on the engineering manufacturing and robotics.

Jianhua Wang received his M.Eng. degree from Beihang University, China, in 1987. He is currently an Associate Professor in the School of Automation Science and Electronic Engineering at Beihang University, China. His research interests include robot control, machine learning, and autonomous navigation.