



A force levels and gestures integrated multi-task strategy for neural decoding

Shaoyang Hua¹ · Congqing Wang¹ · Zuoshu Xie¹ · Xuewei Wu¹

Received: 7 January 2020 / Accepted: 29 March 2020 / Published online: 6 May 2020
© The Author(s) 2020

Abstract

This paper discusses the problem of decoding gestures represented by surface electromyography (sEMG) signals in the presence of variable force levels. It is an attempt that multi-task learning (MTL) is proposed to recognize gestures and force levels synchronously. First, methods of gesture recognition with different force levels are investigated. Then, MTL framework is presented to improve the gesture recognition performance and give information about force levels. Last but not least, to solve the problem that using the greedy principle in MTL, a modified pseudo-task augmentation (PTA) trajectory is introduced. Experiments conducted on two representative datasets demonstrate that compared with other methods, frequency domain information with convolutional neural network (CNN) is more suitable for gesture recognition with variable force levels. Besides, the feasibility of extracting features that are closely related to both gestures and force levels is verified via MTL. By influencing learning dynamics, the proposed PTA method can improve the results of all tasks, and make it applicable to the case where the main tasks and auxiliary tasks are clear.

Keywords Neural decoding · Multi-task learning (MTL) · Pseudo-task augmentation (PTA) · Convolutional neural network (CNN)

Introduction

Neural decoding based on electromyography (EMG) signals has attracted many researchers to explore [1]. It is a technology translating bioelectrical signals in muscles into corresponding instructions [2]. Compared with other human–computer interaction modes, neural decoding is more convenient and less constrained by the surrounding environment, resulting in tremendous development potential in medical, entertainment and military fields [3, 4].

There has been a lot of literature about neural decoding of gestures. Naik et al. [5] associated independent component analysis with Icasso clustering to extract features of surface electromyography (sEMG), then classified gestures by linear discriminant analysis (LDA). Lima et al. investigated relevance vector machines and fractal dimension to identify seven gestures [6]. Besides, convolutional neural network (CNN) has benefited from the success in the com-

puter vision field, and is applied in neural decoding [7–11]. Wei et al. [12] combined information detected from electrodes in different methods, and input them to a multi-stream CNN framework. Hu et al. [13] considered time-series information by recurrent neural network based on this work, which improved the recognition accuracy. In addition, Allard et al. [14] and Zhai et al. [15] proposed a novel method by calculating the feature matrices from time–frequency domain information and classified gestures with CNN models.

While all these methods usually recognize gestures under a fixed force level. The combination of data preprocessing and classifiers is not discussed in the various force levels situation, which means the neglect of strength information. Considering this factor, force myography was selected to recognize sixteen gestures at three force levels from nine subjects [16]. Different from the force sensing resistor (FSR) signals, sEMG signals cannot present information related to force directly. Jiang et al. [17] upgraded hardware with inertial measurement units (IMU) and sEMG sensors, analyzed information from both surface gestures and air gestures. But the force levels considered were only medium and low, the effect of high force level was not investigated. Besides,

✉ Shaoyang Hua
hsy_study@163.com

¹ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

EMG signals affected by different force levels may reduce the performance of the gestures recognition task. To eliminate the force influence, Al-Timemy et al. [18] adopted an energy-based feature set in the circumstance of six gestures at three force levels. Although this method obtained good results, it ignored the usefulness of force information in real life.

Intuitively, gestures recognition is always not an isolated problem. When a gesture is performed, the subject would provide different force levels according to the needs of the environment. As discussed in [19–21], these levels can also be decoded by analyzing the EMG signals. This naturally motivates us to explore the multi-task learning (MTL) framework to decode gestures and force levels in sEMG signals [22].

MTL is a method aiming to learn multiple related tasks simultaneously [23]. By sharing features representation between tasks, it makes models generalize better than learning independently one task [24]. Besides, the introduced inductive bias also leads to a risk reduction of over-fitting. However, most prior jobs either treat the importance of tasks equally or search them by greedy search, which usually cannot find the optimal parameters of models [25]. For conquering this problem, pseudo-task augmentation (PTA) is employed for influencing learning dynamics. As complementary, it is validated to result in performance gains for both single-task learning (STL) and MTL [26].

The main contributions of this paper can be summarized as follows:

1. Different from datasets that collect various gestures with constant force, a dataset containing eight gestures at three force levels is provided. On this basis, matches of data preprocessing and classifiers are compared to find appropriate methods for gestures recognition.
2. Different from a single task of gestures recognition or force estimation as most works do, neural decoding in this job is formulated as an MTL problem. The feasibility of whether gestures and force levels can be decoded synchronously from sEMG signals is explored, which boosts the performance of gestures recognition tasks and gives additional force information.
3. To modify the equal importance of tasks, a PTA strategy is followed with interest. Different from the method proposed in [26], a PTA strategy with weight coefficients is introduced, which considers the relationship between tasks and demonstrates efficient performance in experimental results.

The remaining parts of the paper are summarized as follows. Datasets and materials are discussed in Sect. 2. Proposed methods of multi-task CNN models associated with PTA are presented in Sect. 3. Experiments and results on two

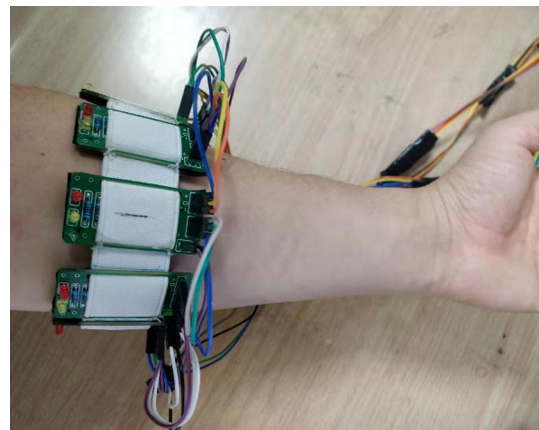


Fig. 1 sEMG signals collection device

datasets are demonstrated in Sect. 4. Finally, the conclusions of the paper are summarized in Sect. 5.

Dataset and materials

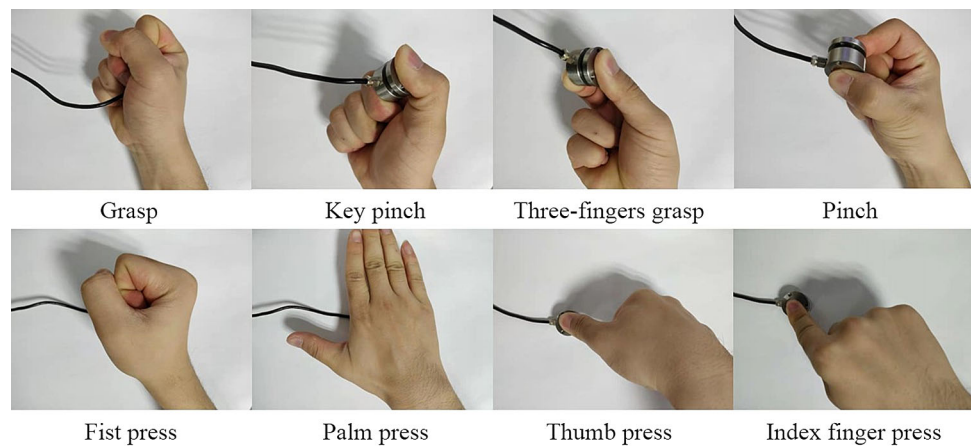
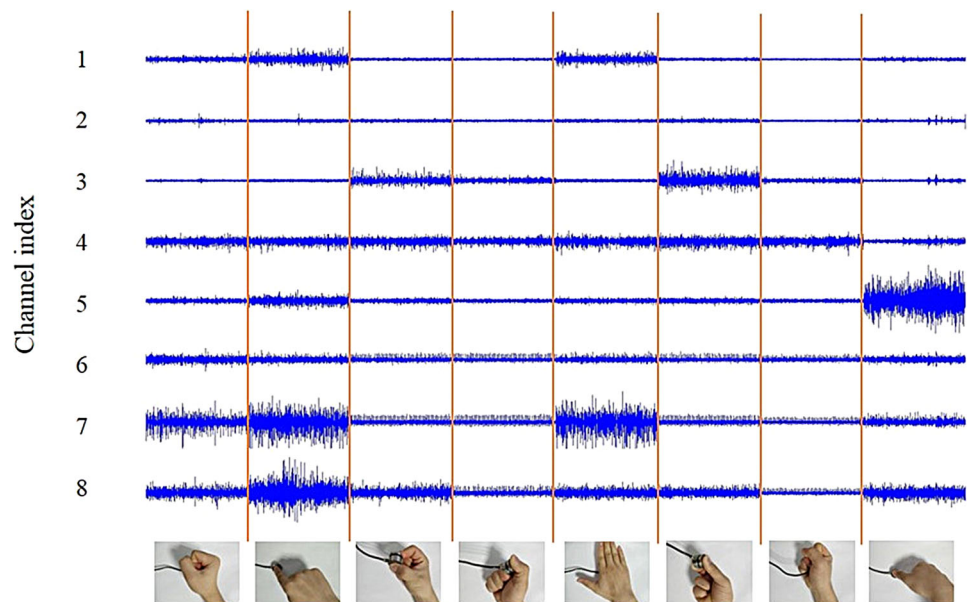
In this paper, two datasets are used in experiments. One represents amputees and the other represents subjects with healthy limbs.

Amputees dataset

In [18], nine amputees conducting six gestures with three force levels participated in the experiment. sEMG signals were sampled at a frequency of 2000 Hz. The gestures are separate thumb flexion, index flexion, fine pinch, tripod grip, hook grip and spherical grip. Each gesture was performed 5–9 trials and each trial lasts 2.5–20 s according to the amputees. The force levels are represented by high, median, and low. Following the protocol, the first eight EMG electrode channels are utilized in experiments. To solve sample imbalance, this paper first sorts the trials in each gesture according to the file size and then selects the first five largest ones for experiments. The first, third and fourth are used as the training set, the second and fifth are used as validation and testing set.

Healthy subjects dataset

For experiments of subjects with healthy limbs, a wearable device is developed to collect the sEMG signals (see Fig. 1). The device consists of 16 acquisition modules. Each module contains a pair of electrodes with a vertical distance of 10 mm. All the 16 sEMG signals were amplified with gain of 960 and filtered with a bandwidth of 20–500 Hz. Besides, signals were sampled at 1000 Hz with an analog-to-digital converter, resulting in an 8-bit digital signal in each module.

Fig. 2 Eight gestures with variable force levels**Fig. 3** sEMG signals from eight channels

For comparison, signals from acquisition modules with odd number indexes are chosen in experiments.

As for the collection period, seven subjects aged from 25 to 30 volunteered, including two females and five males. During collection, a force sensor was first used to test the maximum voluntary contraction (MVC) of each subject. Then, subjects were asked to perform eight gestures with three force levels, representing 20%, 40% and 60% of MVC, respectively. These gestures are separately palm press, thumb press, three-finger grasp, grasp, pinch, fist press and key pinch (see Fig. 2). All of them are selected from the commonly used gestures with strength, such as grasp or contact with a surface. For each gesture, five trials were collected and each trial continued for five seconds. Considering muscle fatigue, there was a rest for several seconds between every two trials. sEMG signals detected from the selected eight channels are shown in Fig. 3.

MTL and PTA strategies

Data preprocessing and CNN models

Given the raw sEMG signals, we divide them into small segments by the sliding window method. Considering the large amount of dataset required by the CNN model, overlapped windowing scheme is utilized. Drawing on the past experiments, window length should be shorter than 300 ms to satisfy subjects imperceptible in real-life applications [27, 28]. In this work, it is set to 200 ms for both datasets, with an overlapped window size of 140 ms.

Before inputting to CNN, segmented data in each electrode was transformed into frequency information by Fast Fourier Transform. Considering the majority of sEMG energy ranged from 0 to 500 Hz, for the dataset of amputees,

the first 100 spectrums were used as input [14, 29]. And the first frequency band was removed to reduce baseline drift and motion artifact. With respect to the dataset of healthy subjects, the first five frequency bands were removed due to filters of the hardware.

To make full use of information in electrode channels, a multi-stream CNN model is designed as seen in Fig. 4. In this way, spectrums from each electrode are used as the input of each stream. And there are three blocks in the stream, which are batch normalization (BN) layers, convolutional layers, and max-pooling layers. The convolutional layer has 32 kernels sized 1×3 , and the max-pooling layer is sized 1×2 with a stride of 2. By stacking blocks, CNN model extracts features through a hierarchy of spectrum abstractions. All the streams will converge into fully connected (FC) layers for classification. There are three FC layers in the model, and the nodes of the first two are separately 512 and 256, with a dropout probability of 0.5. While the last FC layer consisted of two parts connecting to the second FC layer independently, namely the number of gestures and force levels.

MTL framework

As described before, two tasks are finished by the last FC layer of the model simultaneously, which means they share the same features as most MTL frameworks do. The specific information of MTL framework is introduced below.

Given a training set with T samples $D = \{s_i, y_i\}_{i=1}^T$, where s_i is the i th signal segment, y_i are the corresponding labels made up of gesture label (y_i^g) and force level label (y_i^f). For clarity, the index i is eliminated, then the shared feature vector $\mathbf{x} \in \mathbb{R}^{C \times 1}$ of the last max-pooling layer can be formulated as:

$$\mathbf{x} = f(s; \mathbf{k}_c, \mathbf{b}_c, \boldsymbol{\gamma}, \boldsymbol{\beta}) \quad (1)$$

where f donates the non-linear function from the input signals to features. \mathbf{k}_c and \mathbf{b}_c are the parameters of kernels and bias vectors of convolutional layers. $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are the set of scales and shifts in the BN layers.

After feature representation from the last pooling layer, three FC layers are employed for classification. Suppose $\mathbf{W}_i \in \mathbb{R}^{D_{i-1} \times D_i}$ and $\mathbf{b}_i \in \mathbb{R}^{D_i \times 1}$ are the weight matrices and bias vectors of the i th FC layer with output number of D_i ($D_0 = C$), then the prediction score of the i th FC layer \mathbf{y}_i^p is as follows:

$$\mathbf{y}_i^p = \mathbf{W}_i^T \mathbf{y}_{i-1}^p + \mathbf{b}_i \quad (2)$$

where $i = 1, 2, 3$, $\mathbf{y}_0^p = \mathbf{x}$. Specifically, the third FC layer contains both gestures and force levels. Let $\mathbf{W}_{3g}, \mathbf{W}_{3f}, \mathbf{b}_{3g}, \mathbf{b}_{3f}$ donate the weight matrices and bias vectors of the two-part in the last layer, the outputs \mathbf{y}^{pg} and \mathbf{y}^{pf} can be represented as:

$$\mathbf{y}^{pg} = \mathbf{W}_{3g}^T \mathbf{x} + \mathbf{b}_{3g} \quad (3)$$

$$\mathbf{y}^{pf} = \mathbf{W}_{3f}^T \mathbf{x} + \mathbf{b}_{3f} \quad (4)$$

The probabilities of \mathbf{x} belonging to gestures (\hat{y}^{pg}) and force levels (\hat{y}^{pf}) are calculated by feeding \mathbf{y} into a softmax function.

$$\text{softmax}(\mathbf{y}^{pg})_m = p(\hat{y}^{pg} = m | \mathbf{x}) = \frac{\exp(y_m^{pg})}{\sum_i \exp(y_i^{pg})}, \quad (5)$$

$$\text{softmax}(\mathbf{y}^{pf})_n = p(\hat{y}^{pf} = n | \mathbf{x}) = \frac{\exp(y_n^{pf})}{\sum_j \exp(y_j^{pf})}, \quad (6)$$

where y_i^{pg} and y_j^{pf} are the i th element in \mathbf{y}^{pg} and the j th element in \mathbf{y}^{pf} . The softmax function converts the output \mathbf{y}^{pg} and \mathbf{y}^{pf} into a probability distribution over respective labels. Finally, the predicted gesture \hat{y}^{pg} and force level \hat{y}^{pf} are obtained via:

$$\hat{y}^{pg} = \underset{m}{\operatorname{argmax}} \text{softmax}(\mathbf{y}^{pg})_m, \quad (7)$$

$$\hat{y}^{pf} = \underset{n}{\operatorname{argmax}} \text{softmax}(\mathbf{y}^{pf})_n, \quad (8)$$

The cross-entropy losses are employed:

$$L_g = - \sum_{m=1}^M y_m^g \log(p(\hat{y}_m^{pg} = y_m^g) | \mathbf{x}, \mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2, \mathbf{W}_{3g}, \mathbf{b}_{3g}) \quad (9)$$

$$L_f = - \sum_{n=1}^N y_n^f \log(p(\hat{y}_n^{pf} = y_n^f) | \mathbf{x}, \mathbf{W}_1, \mathbf{b}_1, \mathbf{W}_2, \mathbf{b}_2, \mathbf{W}_{3f}, \mathbf{b}_{3f}) \quad (10)$$

where M and N are the number of gestures and the number of force levels, respectively. Let the parameters of the whole model donate as Θ , compared with STL, the loss function consisted of two parts is as follows.

$$\min_{\Theta} (L_g + \alpha L_f) \quad (11)$$

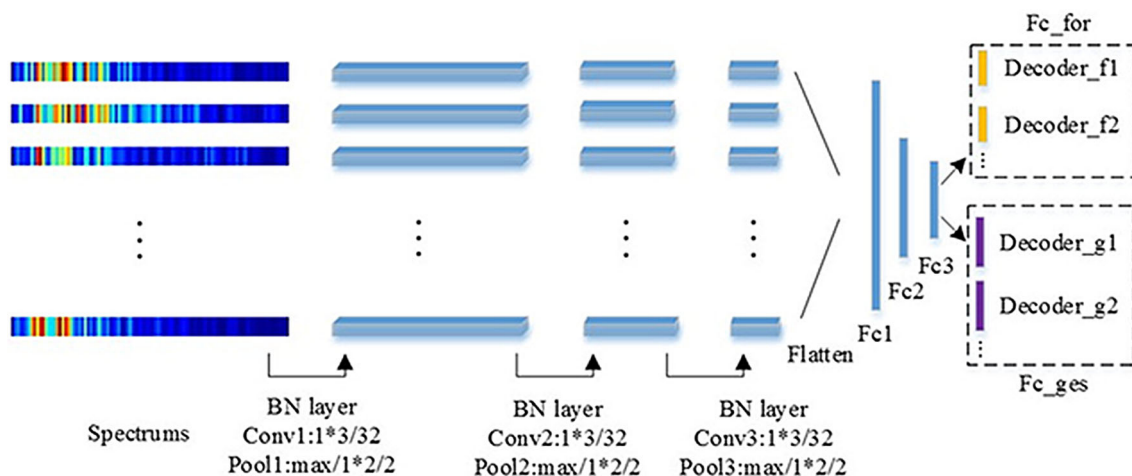


Fig. 4 Multi-stream CNN structure

where α represents the importance of the auxiliary task. In real life, the force levels prediction will be valuable on condition that gestures are recognized correctly, so α ranges from 0 to 1.

PTA strategy

PTA strategy adopts the idea from MTL that training related tasks drawn from the same feature space [30]. If the last layer of the MTL framework is seen as a decoder for each task, the PTA mean numbers of distinct decoders are made within the task. It has been proven that the PTA strategy has a fundamental effect on learning dynamics, which leads to further improvements in both STL and MTL. For MTL with T tasks and D decoders in each, the learning problem of PTA strategy can be expressed as follows:

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \frac{1}{TD} \sum_{t=1}^T \sum_{d=1}^D L(y^t, \hat{y}^{td}) \tag{12}$$

where y^t is the true label of the t th task, and \hat{y}^{td} is the predicted score of the d th decoder in the t th task.

However, this method gives equal importance to all tasks by default, which is usually unreasonable in MTL. For exam-

ple, only when the gestures are predicted correctly, can the prediction of force levels be meaningful. Therefore, an approximate range of weight is first determined through grid search, then learning dynamics is further affected by PTA strategy. The modified PTA strategy during the training period is conducted as follows:

$$\Theta_{tr} = \underset{\Theta}{\operatorname{argmin}} \frac{1}{D} \sum_{d=1}^D \{L(y^g, \hat{y}^{gd}) + \alpha \cdot L(y^f, \hat{y}^{fd})\} \tag{13}$$

where \hat{y}^{gd} and \hat{y}^{fd} are prediction scores of the d th decoder for the gestures and force levels, respectively. For validation, the best performing decoder for each task is selected as follows:

$$\Theta_{eval} = \underset{\Theta}{\operatorname{argmin}} \{L(y^g, \hat{y}^{gd_1}) + \alpha L(y^f, \hat{y}^{fd_2})\} \tag{14}$$

where $d_1, d_2 \in [1, D]$. Because the two parts of the loss function are independent, the weight parameter α does not influence the final result. So it is ignored during the experiments. To distinguish it from the original PTA strategy, the algorithm in this paper is expressed as a weighted PTA (WPTA). The specific implementation of WPTA is shown in Algorithm 1.

Algorithm 1 WPTA Strategy

```

1: Given the CNN framework with  $T$  tasks and one decoder of each task
2: Select  $\alpha$  which makes best results with Eq. 11 by grid search
3: Given the parameters  $\Theta$  of CNN framework with  $T$  tasks and  $D$  decoders of each task
   (parameters of decoders are initialized independently in  $\Theta$ )
4: while training is not done do
5:   for  $i = 1$  to  $M$ (training meta-iteration) do
6:     Freezing all decoders except the first one in all tasks
7:     Update parameters  $\Theta$  with Eq. 13
8:   end for
9:   for  $t = 1$  to  $T$  do
10:    for  $d = 1$  to  $D$  do
11:      Select the best performance of decoders of each task for validation with Eq. 14
12:    end for
13:  end for
14:end while
15: return all parameters  $\Theta$ 

```

In this way, the parameters of each decoder are initialized independently to ensure the learning dynamics work. Furthermore, by updating Θ while freezing parameters of decoders except the first one of all tasks every iteration, the optimal model can still be learned for each task. Finally, for validation and testing progress, the best performing decoder of each task is selected, which contributes to improving computational efficiency.

Experiments and analysis

In this section, the proposed methods are evaluated with a dataset of amputees and healthy subjects, respectively. With each dataset, a range of settings are conducted as follows: (1) gestures recognition with different methods; (2) multi-task recognition for gestures and force levels; (3) MTL with WPTA strategy. All these mentioned experiments are conducted three times and the average performance is shown below.

Recognition results of amputees

Gestures recognition with different methods

Gestures recognition under different force levels is an important topic. Combining with previous experience, data processing and algorithm are firstly discussed in this paper. To

verify the rationality of the proposed method, methods in [12, 17] are used for comparison. The former uses downsampling and low-pass Butterworth filter for amplitude estimation and then proposes a multi-stream CNN based on the processed data to classify sEMG signals, which has a strong contrast with the proposed method. The latter solves similar tasks of this paper. It first extracts features including mean absolute value, zero crossing, slope sign changes and waveform length from the raw signal, and then adopts LDA to recognize gestures with two force levels. For the dataset of amputees, the results of nine patients using three methods separately are shown in Fig. 5.

It shows that the CNN model based on frequency domain information makes the best result (about 5.57% higher than CNN with time-domain information, and 2.14% higher than the traditional method). The phenomenon means that compared with the time domain information used in [12], frequency domain information is more sensitive to the variance of force. And different from the traditional method, the CNN model can extract implicit information in the data more effectively, which verifies the effectiveness of the proposed method.

Multi-task recognition for gestures and force levels

On the basis of multi-stream CNN with spectrums as inputs, an MTL framework is employed as described in Sect. 3. Considering the particularity of the task, gestures recognition

Fig. 5 Result in different methods

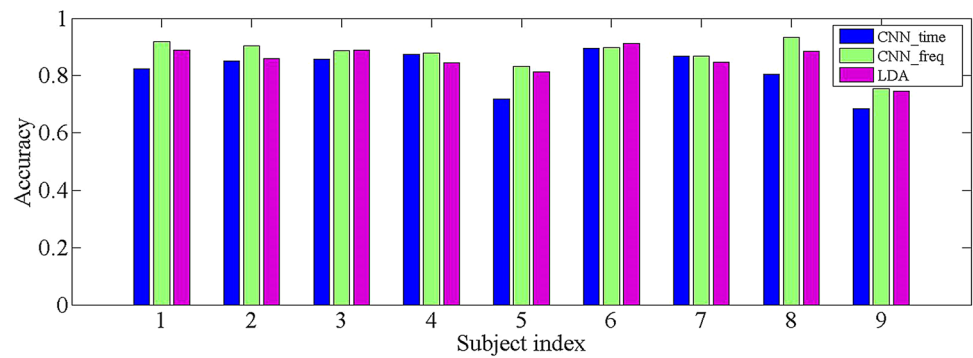


Table 1 Gesture recognition results in different weight coefficients (%)

Subject	α										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
S1	91.95	92.27	92.65	92.93	92.77	92.97	92.91	92.47	93.17	92.79	92.71
S2	90.46	92.52	92.11	92.50	92.28	92.76	92.28	91.94	91.98	92.07	92.37
S3	88.78	89.84	91.35	91.41	91.24	91.12	91.19	90.52	90.00	90.92	90.35
S4	87.92	91.33	91.14	91.09	91.04	91.20	91.28	91.25	91.44	91.01	91.06
S5	83.33	83.97	84.79	84.51	84.15	84.77	84.03	84.31	84.24	84.00	84.58
S6	89.74	92.24	91.75	92.07	92.17	91.47	91.91	91.91	91.83	91.71	91.27
S7	86.73	88.25	87.62	88.20	87.60	88.02	88.25	88.60	88.74	88.28	87.81
S8	93.45	93.79	93.88	93.88	94.09	93.82	93.98	94.00	93.85	93.66	93.72
S9	75.48	77.34	77.75	78.63	78.44	77.59	78.89	78.05	78.54	77.08	78.59
AVG	87.70	89.12	89.31	89.56	89.41	89.37	89.52	89.33	89.40	89.15	89.24

Table 2 Force levels recognition results in different weight coefficients (%)

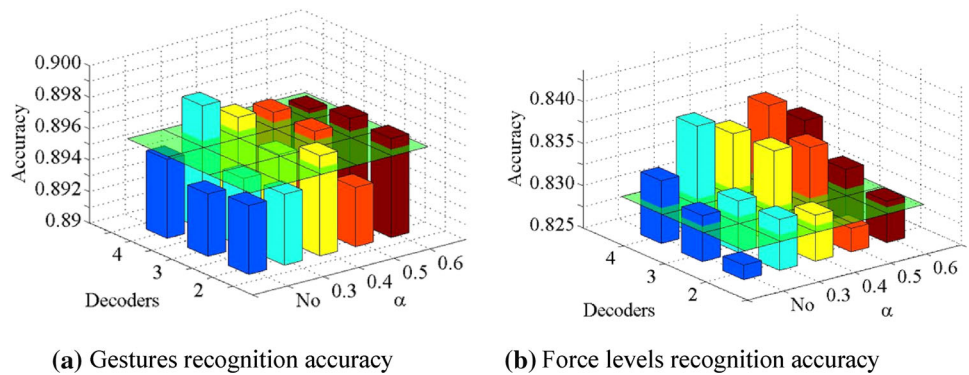
Subject	α										
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
S1	–	86.37	87.55	88.19	88.27	88.33	88.39	88.67	88.11	88.31	88.65
S2	–	86.43	87.00	86.74	86.41	86.87	86.89	86.93	86.15	87.00	86.76
S3	–	82.15	83.94	83.84	83.67	82.96	83.39	83.89	83.35	83.28	83.80
S4	–	80.11	81.67	81.19	82.42	80.49	81.91	81.16	80.92	81.59	80.97
S5	–	76.73	78.36	78.43	78.19	77.24	78.50	78.57	77.36	77.62	78.57
S6	–	71.66	72.58	72.54	72.02	72.06	72.02	72.10	72.42	72.82	72.26
S7	–	69.88	71.77	73.95	73.34	73.86	73.35	73.79	74.00	73.65	73.88
S8	–	87.18	87.48	88.09	87.64	87.55	87.86	87.83	87.42	88.37	87.51
S9	–	88.34	88.91	89.08	89.47	89.21	89.34	89.32	89.90	88.93	89.42
AVG	–	81.17	82.29	82.67	82.55	82.30	82.58	82.68	82.41	82.62	82.63

task is more important than force levels recognition task, the weight coefficient is searched by setting α from 0.0 to 1.0, with an interval of 0.1. Results of both gestures and force levels prediction are shown in Tables 1 and 2.

Notably, α is 0.0 means that a single task is performed without force levels prediction. So there is no result. From Table 1, it is seen that compared with the single gestures recognition, MTL can effectively improve the accuracy of gestures. This suggests that the gestures and force levels share a uniform feature space, which verifies the feasibility that both tasks can be classified within a MTL framework.

Specifically, features learned from force levels recognition task maybe helpful for gestures recognition task, and vice versa. Thus both tasks can have a preferable performance with MTL. Besides, this method provides additional information about force and makes it more practical in real life. On this basis, MTL using a grid search can improve the accuracy of gestures recognition task, which further proves that there is a correlation between the two tasks, and the variable force levels affect the gestures recognition.

Fig. 6 Results of WPTA method



Multi-task recognition with WPTA method

Table 1 shows that when α is equal to 0.3–0.6, the gestures recognition accuracy performs best. So the weight coefficients in Eqs. 13 and 14 are set to 0.3, 0.4, 0.5 and 0.6, separately. For comparing the method in [25], training with Eq. 12 is also performed, which donates as No on α axis in Fig. 6.

The green plane donates the best results obtained in MTL methods with grid search (α is 0.3, number of decoders is 1). It is seen that with MTL and WPTA, improvements are achieved in both tasks, demonstrating the superiority of the proposed method. This is reasoned by the WPTA characteristics which change the learning dynamics adaptively, which can be regarded as a process of fine-tuning importance of auxiliary tasks. So it improves the disadvantage of fixed weight coefficient among all subjects in a grid search. By increasing the number of decoders, the results get better, which draws the same conclusion with [26]. Theoretically, the original PTA method may achieve the same results when there are enough decoders. However, it is a waste of computing resources of hardware.

Recognition results of healthy subjects

As for the dataset of healthy subjects, following the above conduct, we have successively carried out experiments on different settings. For further analysis, experiments on segmented data with length of 150 ms are also carried out. The best performance of gestures prediction and corresponding force levels prediction scores are shown in Table 3.

For MTL and MTL with WPTA in Table 3, the α is equal to 0.3 in the data segment length of 150 ms and 0.4 in 200 ms. It is seen that the algorithm proposed still has certain advantages compared with other methods. As the spectrums in 200 ms contain detailed information in the frequency domain, it shows better performance compared with 150 ms.

In particular, a comparison between STL and MTL with WPTA for each subject based on frequency information is demonstrated in Table 4. It is seen that for most subjects,

Table 3 Results of healthy subjects (%)

Methods		150 ms	200 ms	
Gestures	Traditional methods [17]	89.43	90.84	
	CNN with time information [12]	93.63	94.20	
	Single task	93.81	94.61	
	Multi-task	94.16	95.40	
	Multi-task and WPTA			
	Decoder number = 2	94.71	95.65	
Decoder number = 3	94.85	95.53		
Decoder number = 4	95.03	95.73		
Force levels	Traditional methods [17]	–	–	
	CNN with time information [12]	–	–	
	Single task	–	–	
	Multi-task	89.00	90.17	
	Multi-task and WPTA			
	Decoder number = 2	89.18	90.39	
Decoder number = 3	89.12	90.44		
Decoder number = 4	89.18	90.34		

the proposed method makes performable results, no matter what the time window size is. This is the reason that with MTL, features learned from force levels recognition task improve the main task performance. By introducing an additional loss, MTL also acts as a regularizer preventing over-fitting. Besides, via WPTA, the disadvantage of equal tasks importance is overcome, and a reasonable importance weights distribution is obtained. Rather than one decoder of each task, WPTA provides choices by more decoders, thus further improving the performance. In addition, due to the different physiologic states according to the subjects, the comparison also differs. In detail, stronger subjects can better resist to muscle fatigue during collection, resulting in a more stable EMG signal at each force level. While for others, EMG signals will be affected by muscle fatigue, and the proposed method cannot make much improvement in this situation.

In addition, compared with the experimental results of amputees, results of healthy subjects perform better although

Table 4 Comparison of STL and MTL with WPTA (%)

Window length (ms)	Method	S1	S2	S3	S4	S5	S6	S7
150	STL	96.41	94.56	98.29	88.51	93.32	92.51	93.06
	MTL and WPTA	96.85	96.64	98.87	88.86	94.04	94.59	93.55
200	STL	97.88	95.47	98.73	90.44	89.62	92.48	94.49
	MTL and WPTA	97.96	97.71	99.31	90.44	94.49	95.79	94.44

there are more gestures under the same electrode channels condition, and accuracy differences among methods are smaller than that of amputees. This is reasoned that without damage to the forearm, sEMG signals of healthy subjects are more regular.

Conclusions and future work

Gestures recognition based on sEMG signals has been greatly investigated in recent years. Compared with existing methods, neural decoding under variable force levels situation is still a difficult problem. This paper firstly explores the combinations of data preprocessing and classifiers, proving that frequency domain information is more sensitive to strength. Considering the importance of force information in real life, MTL framework is leveraged to decode the gestures and force levels simultaneously. Experimental results validate the efficiency of MTL. Considering that grid search is so rough that may lead to local optimal solutions, PTA strategy is proposed here. However, it cannot improve the performance because all tasks are assigned as the same weight by default. By combining the above two methods, WPTA technology is applied, which boosts the performance of all tasks in MTL. It is worth noting that similar optimization forms are not only suitable for sEMG signals decoding, but also for any other MTL tasks.

In the future work, we will continue to study the adaptive weight method, and replace the grid-search with it.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arjunan SP, Kumar DK (2010) Decoding subtle forearm flexions using fractal features of surface electromyogram from single and multiple sensors. *J Neuroeng Rehabil* 7:53. <https://doi.org/10.1186/1743-0003-7-53>
- Wang S, Zhang J, Wang H, Lin N, Zong C (2020) Fine-grained neural decoding with distributed word representations. *Inf Sci* 507:256–272
- Li X, Samuel OW, Zhang X, Wang H, Fang P, Li G (2017) A motion-classification strategy based on sEMG-EEG signal combination for upper-limb amputees. *J Neuroeng Rehabil* 14:2. <https://doi.org/10.1186/s12984-016-0212-z>
- Zhuojun X, Yantao T, Yang L (2015) sEMG pattern recognition of muscle force of upper arm for intelligent bionic limb control. *J Bionic Eng* 12(2):316–323
- Naik GR, Al-Timemy AH, Nguyen HT (2015) Transradial amputee gesture classification using an optimal number of sEMG sensors: an approach using ICA clustering. *IEEE Trans Neural Syst Rehabil Eng* 24(8):837–846
- Lima CA, Coelho AL, Madeo RC, Peres SM (2016) Classification of electromyography signals using relevance vector machines and fractal dimension. *Neural Comput Appl* 27(3):791–804
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Boston, MA, pp 1–9
- Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. *European conference on computer vision*, pp 818–833
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Las Vegas, NV, pp 770–778
- Kollias D, Tagaris A, Stafylopatis A, Kollias S, Tagaris G (2018) Deep neural architectures for prediction in healthcare. *Complex Intell Syst* 4(2):119–131
- Wei W, Wong Y, Du Y, Hu Y, Kankanhalli M, Geng W (2017) A multi-stream convolutional neural network for sEMG-based gesture recognition in muscle-computer interface. *Pattern Recogn Lett* 119(1):131–138
- Hu Y, Wong Y, Wei W, Du Y, Kankanhalli M, Geng W (2018) A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0206049>
- Cote-Allard U, Fall CL, Drouin A, Campeau-Lecours A, Gosselin C, Glette K, Laviolette F, Gosselin B (2019) Deep learning for electromyographic hand gesture signal classification using transfer learning. *IEEE Trans Neural Syst Rehabil Eng* 27(4):760–771
- Zhai X, Jelfs B, Chan RH, Tin C (2017) Self-recalibrating surface EMG pattern recognition for neuroprosthesis control based on convolutional neural network. *Front Neurosci*. <https://doi.org/10.3389/fnins.2017.00379>

16. Jiang X, Merhi LK, Menon C (2017) Force exertion affects grasp classification using force myography. *IEEE Trans Human Mach Syst* 48(2):219–226
17. Jiang S, Lv B, Guo W, Zhang C, Wang H, Sheng X, Shull PB (2017) Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU sensing. *IEEE Trans Ind Inf* 14(8):3376–3385
18. Al-Timemy AH, Khushaba RN, Bugmann G, Escudero J (2015) Improving the performance against force variation of EMG controlled multifunctional upper-limb prostheses for transradial amputees. *IEEE Trans Neural Syst Rehabil Eng* 24(6):650–661
19. Kim S, Kim J, Kim M, Kim S, Park J (2019) Grasping force estimation by sEMG signals and arm posture: tensor decomposition approach. *J Bionic Eng* 16(3):455–467
20. Luo J, Liu C, Yang C (2019) Estimation of EMG-based force using a neural-network-based approach. *IEEE Access* 7:64856–64865
21. Zhang S, Guo S, Gao B, Huang Q, Pang M, Hirata H, Ishihara H (2016) Muscle strength assessment system using sEMG-based force prediction method for wrist joint. *J Med Biol Eng* 36(1):121–131
22. Yin X, Liu X (2017) Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans Image Process* 27(2):964–975
23. Ouyang X, Xu S, Zhang C, Zhou P, Yang Y, Liu G, Li X (2019) A 3D-CNN and LSTM based multi-task learning architecture for action recognition. *IEEE Access* 7:40757–40770
24. Ruder S (2017) An overview of multi-task learning in deep neural networks. [arXiv:1706.05098](https://arxiv.org/abs/1706.05098)
25. Pang C, Liu H, Li X (2019) Multitask learning of time-frequency CNN for sound source localization. *IEEE Access* 7:40725–40737
26. Meyerson E, Miiikkulainen R (2018) Pseudo-task augmentation: from deep multitask learning to intratask sharing—and back. [arXiv:1803.04062](https://arxiv.org/abs/1803.04062)
27. Englehart K, Hudgins B (2003) A robust real-time control scheme for multifunction myoelectric control. *IEEE Trans Biomed Eng* 50(7):848–854
28. Farrell TR, Weir RF (2007) The optimal controller delay for myoelectric prostheses. *IEEE Trans Neural Syst Rehabil Eng* 15(1):111–118
29. Khezri M, Jahed M (2010) A neuro-fuzzy inference system for sEMG-based identification of hand motion commands. *IEEE Trans Ind Electron* 58(5):1952–1960
30. Lu X, Shen P, Tsao Y, Kawai H (2016) A pseudo-task design in multi-task learning deep neural network for speaker recognition, 10th International Symposium on Chinese Spoken Language Processing, 1–5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.