

A CNN-based Method for Intent Recognition Using Inertial Measurement Units and Intelligent Lower Limb Prosthesis

Ben-Yue Su, Jie Wang, Shuang-Qing Liu, Min Sheng, Jing Jiang and Kui Xiang

Abstract—Powered intelligent lower limb prosthesis can actuate the knee and ankle joints, allowing transfemoral amputees to perform seamless transitions between locomotion states with the help of an intent recognition system. However, prior intent recognition studies often installed multiple sensors on the prosthesis, and they employed machine learning techniques to analyze time-series data with empirical features. We alternatively propose a novel method for training an intent recognition system that provides natural transitions between level walk, stair ascent/descent, and ramp ascent/descent. Since the transition between two neighboring states is driven by motion intent, we aim to explore the mapping between the motion state of a healthy leg and an amputee's motion intent before the upcoming transition of the prosthesis. We use inertial measurement units (IMUs) and put them on the healthy leg of lower limb amputees for monitoring its locomotion state. We analyze IMU data within the early swing phase of the healthy leg, and feed data into a convolutional neural network (CNN) to learn the feature mapping without expert participation. The proposed method can predict the motion intent of both unilateral amputees and the able-bodied, and help to adaptively calibrate the control strategy for actuating powered intelligent prosthesis in advance. Experimental results show that the recognition accuracy can reach a high level (94.15 % for the able-bodied, 89.23 % for amputees) on 13 classes of motion intent, containing 5 steady states on different terrains as well as 8 transitional states among the steady states.

Index Terms—Intent recognition, lower limb prosthesis, inertial measurement unit (IMU), convolutional neural networks (CNN), swing phase

I. INTRODUCTION

SINCE lower limb amputation is a major cause of disability, millions of transfemoral amputees over the world are suffering from difficulty in moving. The presence of prosthetic legs can help transfemoral amputees restore their locomotion and enable them to perform diverse daily activities [1]. However, using passive prosthesis may significantly impair the walk-

Manuscript received XX, 201X; revised XX, 201X. This work was supported by the National Nature Science Foundation of China (Nos. 11475003, 61603003, 11471093), the Science and Technology Major Project of Anhui Province (18030901021), the Leading Talent Team Project of Anhui Province, and the Anhui Provincial Department of Education outstanding top-notch talent-funded projects (No. gxbjZD26) in China.

Ben-Yue Su and Jie Wang are with the School of Computer and Information, Anqing Normal University, Anqing 246133, Anqing, Anhui, China. Email: bysu@aqnu.edu.cn

Min Sheng and Shuang-Qing Liu are with the School of Mathematics and Computational Science, Anqing Normal University, Anqing 246133, Anqing, Anhui, China.

Jing Jiang is with the School of Computer Science and Telecommunication Engineering, Jiangsu University, Jiangsu 212013, Zhenjiang, Jiangsu, China.

Kui Xiang is with the School of Automation, Wuhan University of Technology, Wuhan 430070, Wuhan, Hubei, China.

ing symmetry and metabolic energy efficiency of transfemoral amputees [2], [3]. The problem thereby stimulates a line of research on powered lower limb prosthesis with the aid of robotic techniques [4].

The powered intelligent prosthesis can automatically calibrate torque according to the gait speed and joint angle [5], [6]. A hierarchical control system allows for better control of the prosthesis [7]. As shown in Fig. 1, the hierarchical general control system generally consists of three components. The high-level controller comprehends human motion intent through activity mode recognition or direct volitional control while the middle-level controller perceives intent from the high-level controller and adjusts the control strategy. Then, the low-level controller eventually actuates the prosthesis based on the control strategy [4]. One major issue of research on intelligent lower limb prostheses is motion intent recognition, which influences the effectiveness of the high-level controller. The purpose of motion intent recognition is to accurately interpret the human intent information encoded by signals in the nerve center. An intent recognition algorithm with high performance makes the transitions between two neighboring locomotion modes become seamless, automatic, and natural [8].

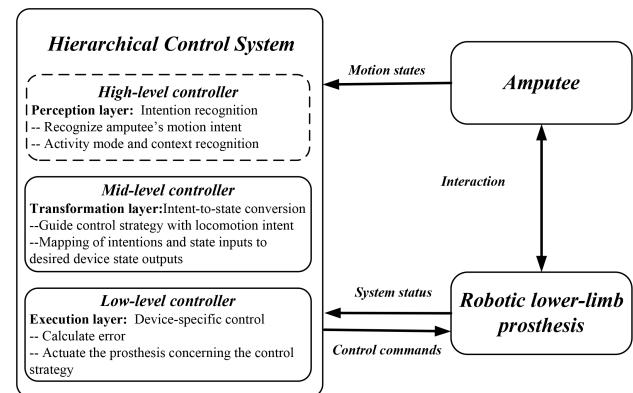


Fig. 1. Hierarchical control strategy of intelligent lower limb prostheses

The most commonly used sensors for intent recognition cover surface electromyogram (sEMG) sensors and mechanical sensors. Motion intent recognition based on the sEMG was first reported in [9]. Many studies then use pattern recognition and machine learning techniques to analyze the locomotion states and transitions of lower limb amputees. Huang *et al.* [10] first combined sEMG signals with pattern recognition to

classify seven daily motion states. They collected signals from eight able-bodied subjects and two subjects with transfemoral amputations while they were walking on different terrains. Moreover, two gait events were proposed (toe-off and heel-strike) to obtain the size of the analysis window. Linear discriminant analysis (LDA) and an artificial neural network (ANN) were used as classifiers, and the recognition accuracy reached 85.7 % [10]. They also used a combination of sEMG and load cell information from five transfemoral amputees [11]. The method could recognize six locomotion modes and five transitions with a high performance ($> 95\%$) on support vector machines (SVMs). Furthermore, they investigated the effects of locomotion mode recognition errors on the volitional control of lower limb prosthesis, and proved that not all of the mode recognition error affected gait stability in prosthesis users. They also defined and suggested "critical errors" as a new index to evaluate locomotion mode recognition [12]. Liu *et al.* [13] used three different adaptive pattern classifiers for locomotion mode recognition, and they indicated that the adaptive classifier could effectively maintain or restore the performance of locomotion mode recognition. Nevertheless, sEMG-based systems for prosthesis usually have poor performance in clinical trials, due to the presence of electrode shift, including the effects of nerve atrophy, variations in muscle contraction effort, and changes in electrode position [14].

Since mechanical signals directly record kinematics information, the studies that used mechanical sensors stand out from other research on intent recognition. In order to realize real-time recognition, Liu *et al.* [14] proposed to solely use a mechanical sensor system consisting of an accelerometer, gyroscope, and two pressure sensors. The intra-class correlation coefficient (ICC) was first calculated according to the Cartesian product of walking speed and terrain. Then, the sensor data were fused via Dempster–Shafer theory and classified with hidden Markov models (HMMs). They successfully predicted prosthesis users' intent with a rate of 95.8 % while the users were walking on different terrains (level ground, ascending stairs, descending stairs, and up/down ramps) [14]. However, the method does not take the transitional states into consideration. Young *et al.* [15] built an intent recognition system using mechanical sensors only (six-axis inertial measurement units (IMUs) and axial load). They collected steady-state and transitional data from six transfemoral amputees while five locomotion modes were performed. The system eventually recognized five steady states as well as eight transitional states, with an overall rate of 93.9 %. In their following works, dynamic Bayesian networks (DBNs) combined with time history information were used to improve the recognition accuracy to 94.7 % [16]; moreover, a user-independent intent recognition system was developed to reduce the training burden [17]. Furthermore, an ANN was used with a scaled conjugate gradient leaning algorithm to recognize the motion intent of amputees, and suitable and accurate results on off-line classification were acquired [18].

To summarize, prior studies usually used the mechanical or sEMG sensors that were embedded on the prosthesis for data collection, and selected handcrafted features for classification. However, manually selecting available features needs expert

participation and prior knowledge. A subjective experience will lead to unavoidable biases and limit the range of system application. Besides, conventional intent recognition methods often involve sensor fusion, which might increase complexities when considering multi-modal sensor signals of transitional states. The goal of this paper is to overcome these drawbacks. Therefore, we adopt a new implementation for sensor data collection, along with a novel redefinition of the transitional state.

In this paper, we propose a new method for training an intent recognition system that provides natural transitions between level walk, stair ascent/descent, and ramp ascent/descent. To reduce the computation cost, we solely use IMUs. Unlike the implementation of a conventional intent recognition system, we place three IMUs on the healthy leg of lower limb amputees, and analyze time series data from IMUs within the early swing phase. The setup allows us to monitor the locomotion of the healthy leg while exploring the mapping between its motion state and the transfemoral amputee's motion intent. Through the mapping, we expect that the upcoming motion state of the prosthesis can be predicted before the transition driven by motion intent occurs. To avoid manual feature selection, the convolutional neural network (CNN), which has made great strides in computer vision [19]–[21] and signal processing [22], is used in this paper. We employ the CNN to automatically learn the hybrid features without expert participation. Experimental results shows that the method can predict the motion intent of unilateral amputees, including the upcoming locomotion mode as well as the transition between two neighboring states.

The structure of this paper is as follows. Section 2 describes the related works on the CNN and our method in detail. Section 3 illustrates the experimental setup and results. Finally, concluding remarks and discussion are given in Section 4.

II. PROPOSED METHOD

This section first presents the concepts of the steady state/step and transitional state/step, which were first proposed in our previous work [23]. Their definitions can help readers comprehend the purpose and implementation of this paper. Second, we introduce the basic concepts of CNNs, and we review related works on motion state recognition based on CNNs. Additionally, we present the modified architecture designed for intent recognition. The mathematical model is also given to interpret the mechanism for processing intent signals.

A. Definitions of motion states

The motion states conducted by the lower limb has an inherent regularity. The regularity is often investigated within a single gait cycle, which consists of a swing phase and a stance phase [24]. Based on the principle, we put forward some new definitions of the motion states within a single gait cycle [23]. Furthermore, the steady and transitional states are exactly expressed by steady and transitional steps performed in the swing phase. The relevant definitions are given below.

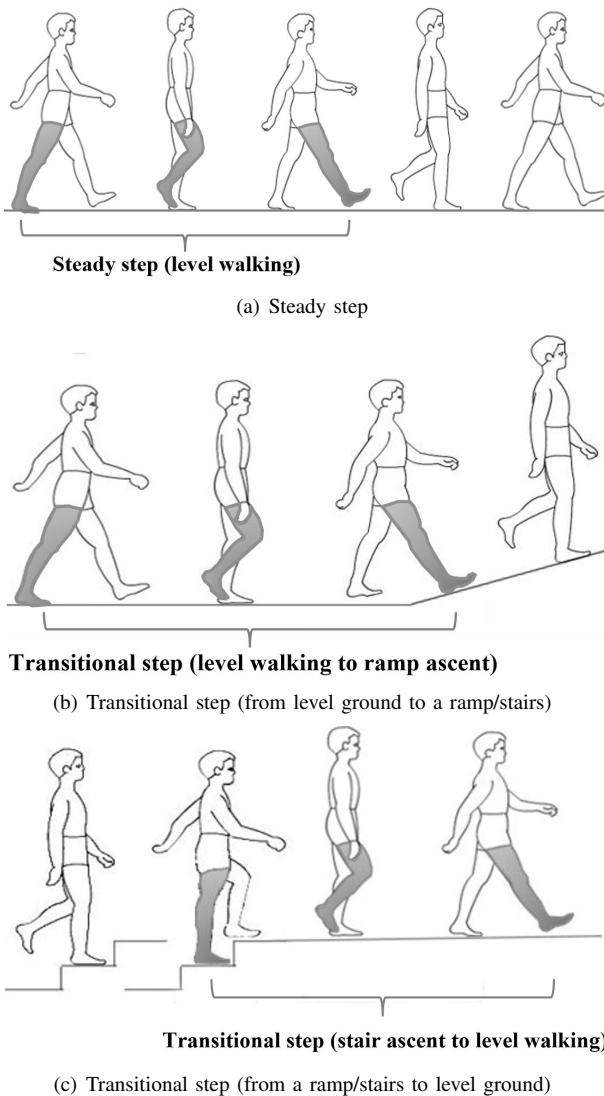


Fig. 2. Illustration of steady step transitional step

Figure 2 also illustrates the steady and transitional steps for better interpreting the proposed method.

The steady step begins at toe-off and ends with a heel-strike. The procedure within the swing phase is depicted in Fig. 2(a).

Second, as seen from Fig. 2(b), the transitional step begins at toe-off, where the foot rises from the level ground and continues until the heel touches the ramp or stair.

To express the transition step that starts from a ramp ascent/descent and ends at level ground, a third definition is provided. As shown in Fig. 2(c), the transitional step starts at toe-off, while the contralateral side begins to stand at the level ground, and ends at the heel-strike.

There are 13 motion states investigated in this study (five steady states and eight transitional states among the steady states on five different terrains): *level walking* (LW), *stair ascent* (SA), *stair descent* (SD), *ramp ascent* (RA), *ramp down* (RD), *level walking to stair ascent* (LW-SA), *level walking to stair descent* (LW-SD), *level walking to ramp ascent* (LW-RA), *level walking to ramp descent* (LW-RD), *stair ascent to level*

walking (SA-LW), *stair descent to level walking* (SD-LW), *ramp ascent to level walking* (RA-LW), and *ramp descent to level walking* (RD-LW).

Note that we focus on the transitional step performed by the healthy leg for exploring the mapping between the transitional state/step and amputees' intent. Thus the subjects have to adjust their step length to make sure the transitions occur at the swing phase of the sound leg (the details of the step sequence are shown in Table I). Through perceiving the current motion state of the healthy leg, the embedded intent recognition system can first analyze the amputee's intent and then modify the control strategy in advance for providing natural and seamless transitions. Besides, transitions are constrained to occur at the stance phase of the affected side in real application. This can be viewed as a safety measure to eliminate potential hazards associated with changing motion states while the prosthesis is swinging or approaching the ground.

B. CNNs for intent recognition

CNN has received significant attention, leading technological advances in various domainsimage classification [25], semantic segmentation [26], and object detection [27]. The CNN architecture was historically inspired by the structure of the mammalian visual system [28], *i.e.*, the hierarchical structure of the visual cortex [29]. The engineering success of the CNN compared to the fully connected neural network is largely due to the strong regularization imposed by the convolution structure and pooling layer: with far fewer weights to learn, the network can generalize better [30].

Due to the attractive qualities of the CNN, and the relative efficiency of its local architecture, the network has been proven to be particularly powerful when addressing signals and time series, given the ability of this type of network to automatically extract local features from raw data while remaining aware of the data structure [31]. It is commonplace to employ CNN for motion state recognition, *i.e.*, human activity recognition (HAR) [32], especially HAR based on wearable sensors [33]–[35]. Generally, HAR based on wearable sensors aims to classify multiple classes of human motion states through multichannel sensor signals. Moreover, some previous studies even realized automatic recognition, which in essence benefits from the hierarchical structure of the CNN. This line of study stimulates our research on intent recognition.

Traditional intent recognition studies predefined a set of handcrafted features based on the prior knowledge, and selected a subset of features by a trial and error method. To this end, we therefore design a CNN structure to automatically learn the high-level and mid-level features from the sensor signals without any prior knowledge. Before illustrating the proposed CNN architecture, we first give explanations of its component layers.

Input layer. One pattern with respect to locomotion state, *i.e.*, IMU data, is shown in Eq. (1).

$$s = [a(t_1), a(t_2), \dots, a(t_i), \dots, a(t_n)]^T \quad (1)$$

Here, $a(t_i) = (a_1(t_i), a_2(t_i), \dots, a_{M \times N}(t_i))$ is the observation vector at time t_i . M is the number of measuring units of a single IMU, and N is the total number of IMUs.

TABLE I
MOTION STATES AND CORRESPONDING STEP SEQUENCE

| Motions | Details of step sequence |
|--|---|
| Five steady states | Sound leg – affected side – sound leg – affected side – sound leg |
| Level walking to stair ascent (LW-SA) | Sound leg (LW) – affected side (LW) – sound leg (LW-SA) – affected side (LW-SA) – sound leg (SA) |
| Level walking to stair descent (LW-SD) | Sound leg (LW) – affected side (LW) – sound leg (LW-SD) – affected side (LW-SD) – sound leg (SD) |
| Level walking to ramp ascent (LW-RA) | Sound leg (LW) – affected side (LW) – sound leg (LW-RA) – affected side (LW-RA) – sound leg (RA) |
| Level walking to ramp descent (LW-RD) | Sound leg (LW) – affected side (LW) – sound leg (LW-RD) – affected side (LW-RD) – sound leg (RD) |
| Stair ascent to level walking (SA-LW) | Affected side (SA) – sound leg (SA) – affected side (SA-LW) – sound side (SA-LW) – affected side (LW) |
| Stair descent to level walking (SD-LW) | Affected side (SD) – sound leg (SD) – affected side (SD-LW) – sound side (SD-LW) – affected side (LW) |
| Ramp ascent to level walking (RA-LW) | Sound leg (RA) – affected side (RA) – sound leg (RA-LW) – affected side (RA-LW) – sound side (LW) |
| Ramp descent to level walking (RD-LW) | Sound leg (RD) – affected side (RD) – sound leg (RD-LW) – affected side (RD-LW) – sound side (LW) |

As is known, the image that is fed into the input layer for image classification typically possesses pixel matrices from multiple channels [36]. For example, the colored images in the CIFAR-10 dataset have 32×32 pixels and RGB channels. We borrow this idea and resize s as input. The sensor signal from the j th channel is firstly expressed by a $1 \times n$ vector as below:

$$x_j = (a_j(t_1), a_j(t_2), \dots, a_j(t_n))^T \quad (2)$$

where $j = 1, 2, \dots, M \times N$. Thus, each pattern X is considered as a sample with $1 \times n$ pixels and $M \times N$ channels.

Convolution layers. Convolution layers are mainly used for feature extraction. The proposed CNN involves convolution operators with different sizes of kernels, namely, convolution kernels, which are responsible for learning relevant features from input data. The calculation of the convolution layer l is shown in Eq. (3).

$$x_j^l = f\left(\sum_{i=1}^C x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (3)$$

Here, x_i^{l-1} ($i = 1, 2, \dots, C$) is the feature map of the i th channel when pattern X goes through the layer $l - 1$. x_j^l is the output of the j th channel within the convolution layer l . The symbol $*$ stands for the convolution operator, and $f(\cdot)$ is the activation function. k_{ij}^l and b_j^l are the learnable convolution kernel matrix and bias, respectively. The output of the convolution layer l is considered as the input of the next layer in the flow.

Pooling. The purpose of pooling is to reduce the input dimensions by performing down-sampling (which also prevents overfitting of the model). More formally,

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l), \quad (4)$$

where $\text{down}(\cdot)$ represents a sub-sampling function (max pooling, average pooling, to name a few) of the j th channel within the pooling layer l . Typically this function will operate over each distinct n -by- n block in the input map so that the output map is n -times smaller along both spatial dimensions. However, in this paper, since the input map is a one-dimensional feature map in the corresponding channel, its output is also a one-dimensional vector but has a smaller length.

Fully connected layers. Ordinary, fully connected layers are composed of numerous neurons. These neurons compute the weight sum of the inputs with assigned weights, and output the activations as below:

$$x_j^l = f(\omega^l x^{l-1} + b^l), \quad (5)$$

where ω^l is the weight coefficient of fully connected layers l .

Once the topology is determined, the network parameters must be learned based on the backpropagation principle. Because the error over the whole training dataset is a sum over the individual errors on each pattern, we consider minimizing the objective function as follows:

$$\text{Minimize } J(k, b, w) = \frac{1}{2} \sum_{n=1}^N \|t_n - y_n\|^2, \quad (6)$$

where t_n indicates the actual output in the form of a 1×13 vector containing the probability of each class, and y_n is the ideal output of the n th sample. For minimizing the above function, stochastic gradient descent (SGD), which slightly updates weights in the direction of the gradient, is adopted as the learning rule of the CNN.

Moreover, discovering an optimal CNN architecture usually involves a large number of parameters that significantly influence the network's performance. The most relevant parameters to determine the architecture of the CNN are the number of layers, the number and size of kernels, and the activation function in each layer. There are also some hyperparameters, such as learning rate and the mini-batch size; these are not directly related to the topology, but rather to the learning process. To summarize, the proposed CNN architecture is depicted in Fig. 3, and the parameters are listed in Table II.

As seen from Fig. 3, the architecture mainly consists of two repeated basic blocks. Each block has a convolution layer with a filter size of 1×3 at first. The stride is set as 2 so that the spatial resolution of the signal is halved through the convolution. Inspired by the Inception block proposed in [37], we employ a similar structure with 1×3 and 1×5 convolutions in each block since the signals can be processed at various scales. Then, features concatenated along the channels are the input for the next layer. Additionally, to make the width and height of the information flow remain unchangeable, the stride

is fixed as 1 with padding for every node within the block. Finally, we adopt global pooling [38] over feature maps in the classification layer, which is less prone to overfitting than traditional fully connected layers. For the sake of simplicity, Table III shows the details of the data flow.

Overall, the framework of our proposed method for motion intent recognition is presented in Algorithm 1.

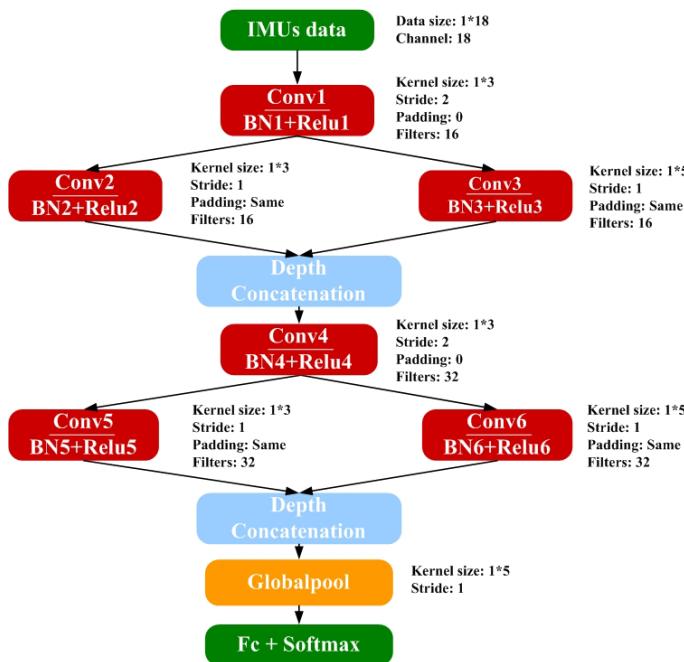


Fig. 3. Structure of the proposed CNN model

TABLE II
HYPERPARAMETERS OF THE PROPOSED CNN MODEL

| Parameter | Value |
|---------------------|--------|
| Optimizer | SGD |
| Mini-batch size | 200 |
| Number of epochs | 50 |
| Learning rate | 0.01 |
| Learning rate decay | 0.1 |
| Weight decay | 0.0005 |
| Momentum | 0.9 |

III. EXPERIMENTS

A. Experimental protocol

The experimental configurations are shown in Fig. 4. An 8-stair staircase with a height of 16 cm was used for stair ascent/descent tests. Meanwhile, a 10° ramp approximately 6 m long was used for ramp ascent/descent tests (see Fig. 4(a)). For the devices, multiple IMUs from Noitom Perception Legacy¹ were employed. Each IMU consists of single triaxial accelerometer and single triaxial gyroscope with a sampling

TABLE III
DETAILS OF THE DATA FLOW WITHIN THE MODIFIED CNN

| Layer | Depth | Input size | Output size |
|---------------|-------|-------------------------|-------------------------|
| Input | 1 | $1 \times 18 \times 18$ | $1 \times 18 \times 18$ |
| Conv1 | 1 | $1 \times 18 \times 18$ | $1 \times 9 \times 16$ |
| Conv2 + Conv3 | 1 | $1 \times 9 \times 16$ | $1 \times 9 \times 32$ |
| Conv4 | 1 | $1 \times 9 \times 32$ | $1 \times 5 \times 32$ |
| Conv5 + Conv6 | 1 | $1 \times 5 \times 32$ | $1 \times 5 \times 64$ |
| Global pool | 1 | $1 \times 5 \times 64$ | $1 \times 1 \times 64$ |
| Fc + Softmax | 1 | $1 \times 1 \times 64$ | $1 \times 1 \times 13$ |
| Output | 1 | $1 \times 1 \times 13$ | 1 |

Algorithm 1 The CNN-based motion intent recognition

Input:

The set of hyperparameters, Θ ; The set of training set, $\{X_{train}, Y_{train}\}$; The set of testing set, $\{X_{test}, Y_{test}\}$;

Output:

The recognition accuracy, acc ;

- 1: Initialize the parameters set $\Theta' = \{k, b, \omega\}$ of CNN, whose architecture is predetermined;
- 2: **while** the maximum epoch is not reached **do**
- 3: Calculate the output of the CNN through forward computation associated with input $\{X_{train}, Y_{train}\}$;
- 4: Update the Θ' by optimizing the objective function $J(k, b, \omega)$ using SGD and Θ ;
- 5: **end while**
- 6: Calculate the output of the CNN through forward computation associated with input X_{test} ;
- 7: Compute the recognition accuracy acc based on the test label Y_{test} and the network prediction T_{test} ;
- 8: **return** acc ;

frequency of 96 Hz. We only used three IMUs, and placed them at the thigh, shank, and ankle of the healthy leg (see Fig. 4(b)). Additionally, all of the subjects were supervised by a doctor during the experiments.

Ten able-bodied subjects (five males and five females) and one transtibial amputee participated. The recruited able-bodied subjects, varied in age (18–30), height (1.58–1.83 m) and weight (40–86 kg). Similar to what Liu *et al.* did in [13], participants were required to wear a special hands-free crutch (iWalk 2.0²; see Fig. 6) for imitating the walking posture of a transfemoral amputee. The transfemoral subject, aged 67 (an amputee for 12 years), wore his own prosthesis (Teh Lin, made in Taiwan; see Fig. 5), and followed the experimental instructions (see Fig. 7).

The able-bodied subjects were instructed to walk with a comfortable speed. They completed 10 trials (every trial contained at least 5 steps) of 5 steady states, as well as 8 transitional states among the steady states on 5 different terrains.

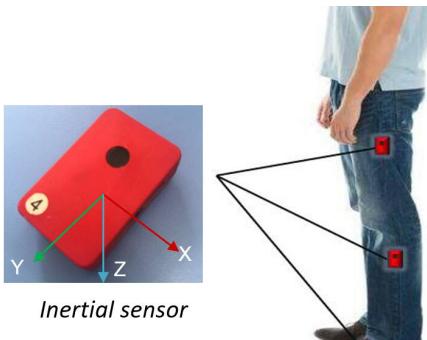
The amputee subject also completed 10 repetitions of each locomotion mode (level ground, stairs, and ramp) and the

¹Refer to <http://www.noitom.com/>

²Refer to <https://www.peglegs.co.uk/>



(a) Experimental platform



(b) IMUs and their placement

Fig. 4. Experimental configurations



Fig. 5. Transtibial amputee and his prosthesis

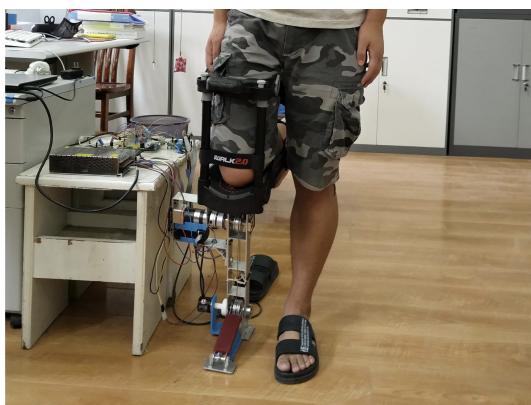


Fig. 6. An able-bodied subject wearing a hands-free crutch

necessary transitions between these locomotion modes. For the stair circuit, the subject first walked on level ground for a few strides, and then walked up stairs and transitioned back to level walking for about two strides. Then the subject stopped and turned around, walked back to the stairs, and descended. Finally, the subject continued to walk on level ground. Likewise, for the ramp circuit, the subject first walked on level ground, then walked up the ramp, and transitioned back to level walking for about two strides. Then the subject stopped and turned around, walked back to the ramp, and went down it. Finally, the subject continued walking on level ground for a few strides before stopping. We obtained 1300 samples from the able-bodied subjects and 130 samples from the transfemoral amputee.

We did not require the amputee to follow a step sequence, and thereby the transfemoral subject had a variable number of steps during the trials. Note that transtibial amputees tend to adjust their step length before performing transition for safety. They always have the transitions from level walking to going up stairs to occur at the sound side, while transitions from level walking to going down stairs occur at the prosthesis side.



Fig. 7. Experiments of stair ascent and stair descent

B. Data preprocessing

The original signals were processed by a Butterworth filter with a cut-off frequency of 40 Hz. Then, IMU signals were normalized to account for the side of amputation: able-bodied subjects imitated left-sided amputees, and the transtibial amputee was a right-sided one. The side-to-side (left/right) accelerometer and the front and horizontal-plane gyroscope signals were transformed (reversed sign) for subjects with left-sided amputations.

Additionally, the signals corresponding to the steady step and transitional step were extracted with a fixed window. The starting point of the window was decided by the time when the steady/transitional step began. The window size was set as 47 (about 0.5 s) while considering the sampling frequency (96 Hz). This imposes a strict restriction on the window employment, requiring the range of the extracted data to stay within the duration of the steady/transitional step. In other words, the healthy leg would be in the swing phase while the prosthesis is in the stance phase during the window. Note

TABLE IV
 USER-INDEPENDENT AND USER-DEPENDENT CLASSIFICATION STRATEGIES

| Classification type | Training data | Testing data | Cross validation |
|---------------------|--|---|---------------------------------|
| User independent | Nine able-bodied subjects | One novel able-bodied subject | Across subjects ($\times 10$) |
| User dependent | 90 % pooled trials from ten able-bodied subjects | 10 % withheld trial from ten able-bodied subjects | Across subjects ($\times 10$) |
| User dependent | 80 % trials from the transtibial amputee | 20 % withheld trial from the transtibial amputee | Within subject ($\times 5$) |

that the window with a size of 47 only occupies half of a whole swing phase.

We normalized the window data into $[0, 1]$ and resized the data as mentioned in Section II. Since we used three IMUs in the implementation ($M = 3$ in Eq. 1) and each IMU consists of a triaxial accelerometer and triaxial gyroscope ($N = 6$), the number of channels was fixed as 18 (computed by $M \times N$). We rescaled every data point to a 1×18 vector with 18 channels, as stated in Table III.

The sliding window method was carried out for data augmentation. The size of the sliding window was set to 18, and the stride was limited to 1 (as shown in Table 8). Thus, we imposed time history information to each processed data, which can help improve classification performance [16]. Consequently, each sample was divided into 30 segments. Note that the data augmentation was only for training data. For matching the size of training data, we randomly extracted a window with a size of 18 from the original test sample as the final test data. Details about how to divide the dataset into training and testing data will be introduced in the following content.

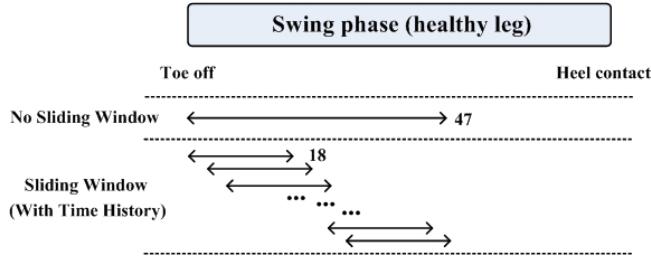


Fig. 8. Sliding window method for training data

C. Classification type

User-independent and user-dependent classification cases were tested and compared in Table IV. For user-independent classification, the proposed CNN was trained on the data from nine able-bodied subjects, and tested on the data from the tenth able-bodied subject. This was repeated 10 times so that each subject was withheld once (*i.e.*, leave-one-out cross validation). To test the robustness of the algorithm, user-dependent classification was tested on both the able-bodied and amputee subjects. For able-bodied subjects, the processed sample set was randomly divided into 10 parts, where 9 parts

were selected as the training set and other part was selected as the testing set. For the transtibial amputee, 20 % of samples were randomly selected as testing data, and the rest were used as training data. Cross validation was done across all 10 (5) trials such that each trial was left out once.

D. Experimental results

1) *User-dependent classification.* Figure 9 shows the performance on user-dependent classification with respect to able-bodied subjects. The testing accuracy reached $94.15\% \pm 3.04\%$. The confusion matrix in Fig. 9 describes how many samples are misclassified within the testing set. Additionally, the numbers on the diagonal represent the correctly classified testing samples. Correctly classifying the transitional steps is much more difficult than classifying the steady steps, as reported in our previous work [23]. However, our CNN-based method addresses this disadvantage, and it can recognize both transitional steps and steady steps with high performance.

Since there is only one transtibial amputee involved, the user-dependent classification was conducted on limited samples. The dataset collected from the transtibial amputee was randomly split into training and testing data with a ratio of 8 : 2. The sliding window method was carried out on the training data; there were 26 testing data and 3120 training data in total. Figure 10 shows the confusion matrix for this case. The testing accuracy reached $89.23\% \pm 4.21\%$ on 13 motion states. The results from able-bodied subjects were basically consistent with that from the transtibial amputee, although a slight difference was observed.

2) *User-independent classification.* The confusion matrix among 13 motion states in the case of user-independent classification is shown in Fig. 11. The testing accuracy of user-independent classification could only reach $82.23\% \pm 7.46\%$. Correctly predicting the steady states is of great importance to the intent recognition system since the steady states are more likely to be performed in real life. Thus, we divided the experiment into two parts. One part aimed to evaluate the performance of the classifier trained by only steady state patterns, and the other evaluated the classification results that only involved transition state patterns. The results for steady states and transition states were $92.20\% \pm 4.76\%$ (Fig. 12) and $88.75\% \pm 6.68\%$, respectively, which are higher than the overall result mentioned before. The reason for this is that a large proportion of the errors are misclassifications of data from RD and RD-LW due to pattern similarity. Compared

| Confusion Matrix for Test Data | | | | | | | | | | | | |
|--------------------------------|-------|----|----|----|----|----|----|----|----|----|----|----|
| True class | LW-SA | 96 | 3 | | | | | | 2 | | | |
| | LW-SD | | 95 | 1 | 4 | | 1 | | | 2 | | |
| | LW-RA | 2 | | 94 | 2 | 1 | | | | | 1 | |
| | LW-RD | 2 | 1 | 2 | 92 | | | | 2 | | | |
| | RA-LW | | | | 94 | | 1 | | | | 3 | |
| | RD-LW | | | | | 92 | 1 | 1 | | | | 10 |
| | SA-LW | | | | | 1 | 95 | | 1 | | | 1 |
| | SD-LW | | 1 | | 2 | 1 | | 93 | | 3 | | |
| | LW | 2 | | 1 | | | 1 | 1 | 96 | | | |
| | SA | | | | | | 1 | | | 98 | | |
| | SD | 1 | | | 1 | 1 | 1 | 4 | | | 95 | |
| | RA | | | | 2 | | | | | | 96 | 1 |
| | RD | | | | 1 | 4 | | 1 | 1 | | | 88 |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |
| | | | | | | | | | | | | |

with Fig. 11, the classification accuracy presented in Fig. 12 increased because only 5 classes instead of all 13 classes were classified. Therefore, if independently considering steady states in the recognition system, the proposed method will make it more effective.

| Confusion Matrix for Test Data | | | | | | | | |
|--------------------------------|-------|-------|-------|-------|-----------------|-------|-------|-------|
| True class | LW | 93 | | 1 | 2 | 7 | 90.3% | 9.7% |
| | SA | | 97 | | 8 | | 92.4% | 7.6% |
| | SD | 1 | 2 | 96 | | 3 | 94.1% | 5.9% |
| | RA | 1 | 1 | 1 | 87 | 2 | 94.6% | 5.4% |
| | RD | 5 | | 2 | 3 | 88 | 89.8% | 10.2% |
| | 93.0% | 97.0% | 96.0% | 87.0% | 88.0% | | | |
| | | 7.0% | 3.0% | 4.0% | 13.0% | 12.0% | | |
| LW SA SD | | | RA | RD | Predicted class | | | |

Fig. 12. Confusion matrix of steady states in user-independent classification

It seems that the misclassification rate of the user-independent system is higher than that of the user-dependent system. Thus, using subject-specific training data is highly valuable for reducing intent recognition error rates. Compared to the user-dependent strategy, the relatively high error rate on the user-independent classifier demonstrates the difficulty of generalizing training data to a fully novel user.

E. Experimental analysis

1) *Network architecture search*. We incorporate a greedy-wise method for searching the optimal CNN architecture. We test 3 models, which have different numbers of convolution blocks. In Fig. 13, the left side shows the architecture of the convolution block while the right side depicts the respective performances on three cases. We can find that the model with $N = 2$ (N is the number of convolution blocks) used in this paper achieves the highest performance on user-dependent classification. This proves that adding an additional convolutional layer can lead to a better result because the testing accuracy on $N = 1$ is slightly lower than that on

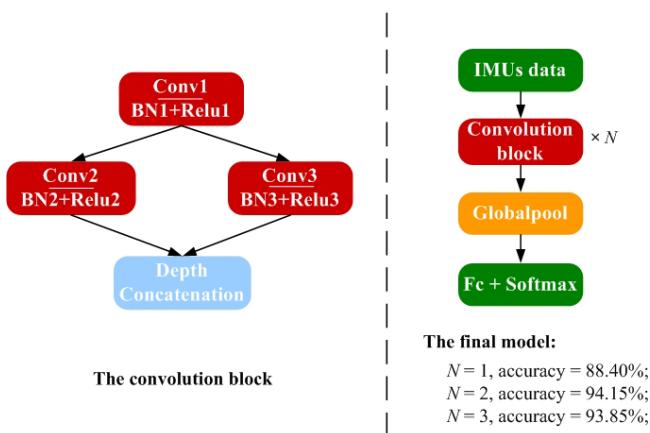


Fig. 13. Testing accuracies of the CNN with different convolution blocks

$N = 2$. Besides, to minimize computation cost, we select two convolution blocks as the final solution since the two models have similar results.

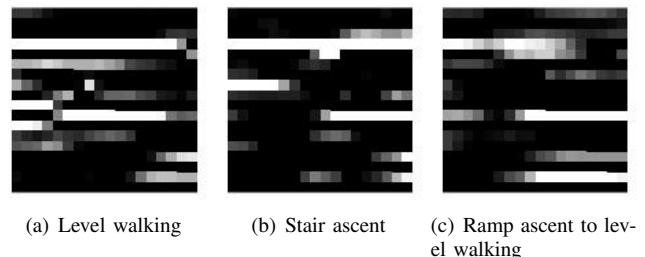


Fig. 14. Visualization of the three patterns from different classes

2) *Feature visualization*. We demonstrate the validity of the feature learning procedure for intent recognition through visualization. First, we randomly select three patterns in the testing set, and visualize these samples first. Since the original 1×18 input images with 18 channels cannot be clearly described, the size of patterns is thereby set as 18×18 . The processed images are shown in Fig. 14. For any input image in the figure, the pixels in each row express the corresponding values within a specific channel. From Fig. 14, we can see that the input images are different from each other since these patterns belong to different classes. In spite of intuitive indiscernibility, these images can be learned by the deep neural network.

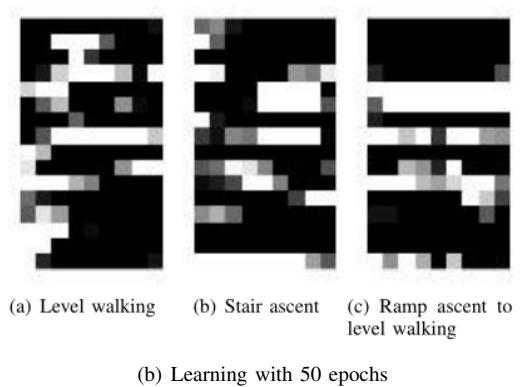
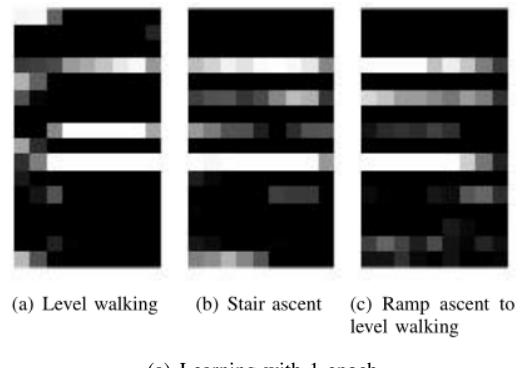


Fig. 15. Visualization of the feature maps of the first convolution layer

We further visualize the feature maps output by the first convolution layer, as shown in Fig. 15. Since the number of kernels is 16 and the size of each kernel is 1×3 in the first convolution layer, the size of output maps is designed

as $1 \times 9 \times 16$. Figures 15(a) and 15(b) exhibit the convolution result with 1 epoch and 50 epochs training of three patterns, respectively. Although we cannot find a regular information in both results, the learning process is able to continuously update the parameters for exploring complex features. The observations show that the outputs in Fig. 15(b) have a large number of "white squares," which demonstrates that the features become sparse with the epochs.

IV. CONCLUSIONS AND DISCUSSIONS

With the help of an intent recognition system, a powerful prosthesis allows a transfemoral amputee to perform natural transitions between neighboring locomotion modes. To improve the classical intent recognition studies, we propose a novel approach based on the CNN and IMUs data. Theoretically, it can well recognize both steady states and transitions, and provide real-time feedback to powerful prostheses. Experimental results also verify its validity and effectiveness. Table V shows the comparison between our method and state-of-the-art methods in the field of motion intent recognition. The main contributions of our proposed method can be summarized as follows.

First, we only used three IMUs rather than multi-modal sensors as conventionally used [6], [14], [16], [17]. This strategy can avoid the increased computation cost due to fusion of multi-modal signals.

Second, our approach investigated both transition states and steady states, yet some traditional works classified only steady states or transition states [6], [14], [16].

Third, we collected the IMU data from the healthy leg, which is quite different from the methodology used in past studies [6], [14], [16], [17]. Furthermore, IMU data within the early swing phase of the healthy leg was extracted and analyzed before the transition of the prosthesis occurred. This strategy was selected to explore the mapping between the motion state of the healthy leg and the transfemoral amputee's motion intent. Through the mapping found by machine

learning, the upcoming motion state of the prosthesis can be predicted before the transition that is driven by motion intent. The difficulty of our method lies in the fact that the time series data provide little feature information for classification due to the short duration of the swing phase. As is known, the stance phase accounts for approximately 60 % and the swing phase accounts for 40 % within a gait cycle. Most previous studies thereby focused on the data obtained within the stance phase of the prosthesis for fully utilizing richer information. Our proposed method proves that analyzing the swing phase of the healthy leg can also realize intent recognition with high performance.

Finally, prior works often manually selected the most predictive features of signals. By contrast, our approach adopted a hierarchical CNN for automatic feature learning. This made feature engineering able to be carried out without expert participation. As a result, the new algorithm is highly competitive and superior to that in our previous work ([23]) in terms of accuracy. Note that 95.12 % is the best accuracy in our previous work ([23]) while the best accuracy in this paper is 97.19 %. To the authors' best knowledge, deep learning has not yet been applied in the research field of intent recognition.

Although this study provided an alternative intent recognition system with respect to lower limb prosthesis, several limitations are present. First, this study was conducted in a laboratory environment and the able-bodied subjects were instructed to follow a pre-defined step sequence. However, a step sequence is unpredictable in real life. Furthermore, when confronted with a ramp or stairs, transfemoral amputees often transition onto the ramp with either leg (sound or prosthetic side), begin ascending stairs with their sound side, and begin descending stairs with their prosthetic side [17]. Second, whether the effect of walking speed variations can influence classification performance was not discussed in this paper. In fact, all of the subjects were instructed to walk with a comfortable speed. Finally, we tested generalization of the proposed method to only one novel amputee. Due

TABLE V
 COMPARISON WITH STATE-OF-THE-ART METHODS

| Reference | Subjects | Sensors | Feature extraction | Classifier | Position of sensors | Number of motion intents | Accuracy |
|-------------------------|-------------------------------------|--|--------------------|------------|---------------------|--------------------------|--|
| Young et al., 2014 [16] | 6 transfemoral amputees | IMU, axial load cell, motor potentiometers, encoders | Manual | DBN | Prosthesis | 5 | 94.7 % |
| Young et al., 2014 [17] | 6 transfemoral amputees | IMU, axial load cell, motor current | Manual | LDA | Prosthesis | 13 | 93.9 % |
| Liu et al., 2017 [14] | 3 able-bodied, 2 amputees | Accelerometer, gyroscope, pressure sensor | Manual | HMM | Prosthesis | 5 | 95.8 % |
| Zheng et al., 2017 [6] | 6 transfemoral amputees | IMU, load cell | Manual | QDA +SVM | Prosthesis | 8 | 94.9 % |
| Su et al., 2018 [23] | 10 able-bodied | IMUs | Manual | SVM | Healthy side | 13 | 95.12 % |
| Our method | 10 able-bodied 1 amputee | IMUs | Automatic | CNN | Healthy side | 13 | $94.15 \% \pm 3.04 \%$ $89.23 \% \pm 4.21 \%$ |

to the limited experimental condition, the user-independent classification was not conducted on their motion data. Future work will consider more characteristics of the step sequence of amputees, and be devoted to designing an improved intent recognition algorithm that can generalize to a range of ramp slopes and stair grades.

REFERENCES

- [1] H. F. Pernot, L. P. De Witte, E. Lindeman, and J. Cluitmans, "Daily functioning of the lower extremity amputee: an overview of the literature," *Clinical Rehabilitation*, vol. 11, no. 2, pp. 93–106, 1997.
- [2] T. Schmalz, S. Blumentritt, and B. Marx, "Biomechanical analysis of stair ambulation in lower limb amputees," *Gait Posture*, vol. 25, no. 2, pp. 267–278, 2007.
- [3] R. L. Waters, J. Perry, D. Antonelli, and H. Hislop, "Energy cost of walking of amputees: the influence of level of amputation," *Journal of Bone & Joint Surgery-American Volume*, vol. 58, no. 1, pp. 42–46, 1976.
- [4] M. R. Tucker, J. Olivier, A. Pagel, H. Bleuler, M. Bouri, O. Lambercy, J. D. R. Milln, R. Riener, H. Vallery, and R. Gassert, "Control strategies for active lower extremity prosthetics and orthotics: a review," *Journal of Neuroengineering & Rehabilitation*, vol. 12, no. 1, pp. 1–29, 2015.
- [5] P. Cherelle, V. Grosu, A. Matthys, B. Vanderborght, and D. Lefeber, "Design and validation of the ankle mimicking prosthetic (AMP) foot 2.0," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 22, no. 1, pp. 138–148, 2013.
- [6] E. H. Zheng and Q. N. Wang, "Noncontact capacitive sensing-based locomotion transition recognition for amputees with robotic transtibial prostheses," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 25, no. 2, pp. 161–170, 2016.
- [7] F. Sup, H. A. Varol, J. Mitchell, T. J. Withrow, and M. Goldfarb, "Preliminary evaluations of a self-contained anthropomorphic transfemoral prosthesis," *IEEE/ASME Transactions on Mechatronics*, vol. 14, no. 6, pp. 667–676, 2009.
- [8] J. D. Miller, M. S. Beazer, and M. E. Hahn, "Myoelectric walking mode classification for transtibial amputees," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 10, pp. 2745–2750, 2013.
- [9] K. Nazarpour, A. R. Sharafat, and S. M. Firoozabadi, "Application of higher order statistics to surface electromyogram signal classification," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1762–1769, 2007.
- [10] H. Huang, T. A. Kuiken, and R. D. Lipschutz, "A strategy for identifying locomotion modes using surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 1, pp. 65–73, 2009.
- [11] H. Huang, F. Zhang, L. J. Hargrove, Z. Dou, D. R. Rogers, and K. B. Englehart, "Continuous locomotion-mode identification for prosthetic legs based on neuromuscular-mechanical fusion," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2867–2875, 2011.
- [12] F. Zhang, M. Liu, and H. Huang, "Effects of locomotion mode recognition errors on volitional control of powered above-knee prostheses," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 23, no. 1, pp. 64–72, 2015.
- [13] M. Liu, F. Zhang, and H. Huang, "An adaptive classification strategy for reliable locomotion mode recognition," *Sensors*, vol. 17, no. 9, p. 2020, 2017.
- [14] Z. J. Liu, W. Lin, Y. L. Geng, and P. Yang, "Intent pattern recognition of lower-limb motion based on mechanical sensors," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 651–660, 2017.
- [15] A. J. Young, A. M. Simon, and L. J. Hargrove, "A training method for locomotion mode prediction using powered lower limb prostheses," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 22, no. 3, pp. 671–677, 2014.
- [16] A. J. Young, A. M. Simon, N. P. Fey, and L. J. Hargrove, "Intent recognition in a powered lower limb prosthesis using time history information," *Annals of Biomedical Engineering*, vol. 42, no. 3, pp. 631–641, 2014.
- [17] A. Young and L. Hargrove, "A classification method for user-independent intent recognition for transfemoral amputees using powered lower limb prostheses," *IEEE Transactions on Neural Systems & Rehabilitation Engineering*, vol. 24, no. 2, pp. 217–225, 2016.
- [18] R. B. Woodward, J. A. Spanias, and L. J. Hargrove, "User intent prediction with a scaled conjugate gradient trained artificial neural network for lower limb amputees using a powered prosthesis," *International Conference of the IEEE Engineering in Medicine & Biology Society*, pp. 6405–6408, 2016.
- [19] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2015, pp. 770–778.
- [20] F. Y. Liu, G. S. Lin, and C. H. Shen, "CRF learning with CNN features for image segmentation," *Pattern Recognition*, vol. 48, no. 10, pp. 2983–2992, 2015.
- [21] C. Fernando, D. Banarse, M. Reynolds, F. Besse, D. Pfau, M. Jaderberg, M. Lanctot, and D. Wierstra, "Convolution by evolution: Differentiable pattern producing networks," in *Genetic & Evolutionary Computation Conference*, 2016, pp. 109–116.
- [22] M. Sheng, J. Jiang, B. Y. Su, A. A. Yahya, and G. J. Wang, "Short-time activity recognition with wearable sensors using convolutional neural network," in *ACM SIGGRAPH Conference on Virtual-Reality Continuum & Its Applications in Industry*, 2016, pp. 413–416.
- [23] B. Y. Su, J. Wang, S. Q. Liu, M. Sheng, and K. Xiang, "An improved motion intent recognition method for intelligent lower limb prosthesis driven by inertial motion capture data," *Acta Automatica Sinica*, 2018, Accepted in Chinese.
- [24] S. K. Au, P. Dilworth, and H. Herr, "An ankle-foot emulation system for the study of human walking biomechanics," in *IEEE International Conference on Robotics & Automation*, 2006, pp. 2939–2945.
- [25] D. S. Kermany, M. Goldbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, and F. Yan, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
- [26] K. M. He, G. Gkioxari, P. Dollr, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision*, 2017, pp. 2980–2988.
- [27] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016.
- [28] D. J. Felleman and D. C. Van Essen, "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.
- [30] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [31] A. Baldominos, Y. Saez, and P. Isasi, "Evolutionary convolutional neural networks: an application to handwriting recognition," *Neurocomputing*, vol. 283, pp. 38–52, 2017.
- [32] S. W. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2012.
- [33] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [34] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *International Joint Conference on Artificial Intelligence*, 2015, pp. 3995–4001.
- [35] W. C. Jiang and Z. Z. Yin, "Human activity recognition using wearable sensors by deep convolutional neural networks," in *ACM International Conference on Multimedia*, 2015, pp. 1307–1310.
- [36] K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2015.
- [38] M. Lin, Q. Chen, and S. C. Yan, "Network in network," in *International Conference on Learning Representations*, 2014.



Ben-Yue Su received the Ph.D. degree from the School of Computer and Information, Hefei University of Technology in 2007. He is currently a professor at the School of Computer and Information, Anqing Normal University, and the University Key Laboratory of Intelligent Perception and Computing of Anhui Province. His research interests cover pattern recognition and machine learning, image processing, and computer graphics. He is the corresponding author of this paper.



Kui Xiang received the Ph.D. degree from control theory and control engineering, Zhejiang University in 2006. He is a professor at the School of Automation, Wuhan University of Technology. His research interests include human walking mechanisms and bio-inspired intelligent joints.



Jie Wang received the B.S. degree from the School of Computer and Information, Anqing Normal University in 2016. She is currently working toward the M.A. degree at the School of Computer and Information, Anqing Normal University. Her research interests cover machine learning and motion intent recognition.



Shuang-Qing Liu received the B.S. degree from the School of Mathematics and Computational Science, Anqing Normal University in 2016. He is currently working toward the M.A. degree at the School of Mathematics and Computational Science, Anqing Normal University. His research interests cover machine learning and motion intent recognition.



Min Sheng received the Ph.D. degree from the School of Computer and Information, Hefei University of Technology in 2009. She is a professor at the School of Mathematics and Computational Science, Anqing Normal University. Her research interests include pattern recognition and image and video processing.



Jing Jiang received the M.A. degree in the School of Computer and Information, Anqing Normal University in 2017. He is currently pursuing the Ph.D. degree at the School of Computer Science and Telecommunication Engineering, Jiangsu University. His research interests include multiobjective optimization and neuroevolution.