RESEARCH ARTICLE

A Comparison of Four Neural Networks Algorithms on Locomotion Intention Recognition of Lower Limb Exoskeleton Based on Multi-source Information

Duojin Wang[1,2] [ID] **· Xiaoping Gu**[1] **· Hongliu Yu**[1,2]

Abstract

Lower Limb Exoskeletons (LLEs) are receiving increasing attention for supporting activities of daily living. In such active systems, an intelligent controller may be indispensable. In this paper, we proposed a locomotion intention recognition system based on time series data sets derived from human motion signals. Composed of input data and Deep Learning (DL) algorithms, this framework enables the detection and prediction of users' movement patterns. This makes it possible to predict the detection of locomotion modes, allowing the LLEs to provide smooth and seamless assistance. The pre-processed eight subjects were used as input to classify four scenes: Standing/Walking on Level Ground (S/WOLG), Up the Stairs (US), Down the Stairs (DS), and Walking on Grass (WOG). The result showed that the ResNet performed optimally compared to four algorithms (CNN, CNN-LSTM, ResNet, and ResNet-Att) with an approximate evaluation indicator of 100%. It is expected that the proposed locomotion intention system will significantly improve the safety and the effectiveness of LLE due to its high accuracy and predictive performance.

1 Introduction

In recent years, wearable robots have a promising future and can physically assist humans in locomotion [1]. In particular, LLEs are increasingly used and noticed by the public. Although wearable technology is promising, its control system needs to be further developed. Humans and robots need to work together to perform repetitive activities in applications, such as robot-assisted rehabilitation, leg exoskeletons, and others [2, 3]. The human–robot system may pose a serious problem. In fact, even an occasional wrong action while wearing a LLEs can cause irreversible damage to the human body [4]. Therefore, automatic recognition of the current state of human movement is a prerequisite for LLEs.

However, most current commercial exoskeletons, such as the Össur Power Knee prosthesis, ReWalk, and Indego exoskeletons [5, 6], generally communicate the intention to move by pressing the control button or performing abnormal body movements. This technique is not real-time capable and usually carries a risk of physical harm. By overcoming the interruptions from people and the environment, the metabolic cost of a person can be decreased by employing an effective control approach [7]. A complete intelligent intention detection system needs to be applied to LLEs to achieve safe performance.

Based on the data obtained from the wearer's movements, intention recognition can predict upcoming movements. Several works have completed intention recognition with sensor fusion to improve performance [8, 9]. This method relies on manual extraction, especially expert knowledge, to extract useful features.

DL can automatically extract features and is suitable for intent detection. In many researches, cameras are used to collect environmental data and apply a set of image classification algorithms [4, 10]. While this method can provide the desired classification, it wastes computational resources and requires expensive hardware and a long time to train

✉ Duojin Wang
duojin.wang@usst.edu.cn

1   Institute of Rehabilitation Engineering and Technology, University of Shanghai for Science and Technology, 516 Jungong Road, Shanghai 200093, China

2   Shanghai Engineering Research Center of Assistive Devices, 516 Jungong Road, Shanghai 200093, China

the model. However, time series data can respond quickly to classification problems and save memory. Therefore, in this study, time series data are adopted to predict human movement intentions, and the DL method is used to improve the prediction accuracy.

The contributions of this work include:

(1) Processing unbalanced label data.
(2) Applying the ResNet model to time series data sets.
(3) Comparing the performance of different models and identifying the best performance for classifying human motion intentions.

The rest of the paper is organized as follows: Section 2 describes the related work. The proposed methods and experiments are described in Sects. 3 and 4, respectively. The conclusion of this paper is presented in Sect. 5.

2 Related Work

Most of the research on human intention recognition involved using collected images for classification prediction and human motion signals for time series classification prediction, fusing these data to achieve high accuracy, and converting images into time series or translating the latter into the former.

Laschowski et al. [10] also collected the "ExoNet" image dataset, which contains real indoor and outdoor walking environments. They then trained and tested more than ten state-of-the-art deep CNN based on the dataset. Zhang et al. [24] develop an end-to-end unsupervised cross-subject adaptation with time series datasets. Based on the MCD [25], the feature generator aligns the features of the source and target domains to fool the domain classifier until it is unable to detect which domain the features originated from. To stabilize the point cloud of the environment, a depth camera and an IMU are used together. The original 3D point cloud was reduced to 2D and classified by a neural network [26]. Hur et al. [27] used a novel encoding technique to convert an inertial sensor signal into an image with minimal distortion and a CNN model for image-based intention classification.

A time series dataset (human motion signal) was used to study intention recognition of LLE. A variety of DL network models were used for training and testing. In general, LSTM or CNN-LSTM models are applied for prediction. Table 1 shows that LSTM is widely used for human motion signals, while ResNet classifies image data. We found that fewer studies have used ResNet in the time series datasets of LLE. Therefore, the following experiment was proposed to apply

ResNet to the intention recognition of LLE, and compare it with CNN and CNN-LSTM.

3 Proposed Methods

In this paper, we analyze several common network structures, including CNN, CNN-LSTM, ResNet, and the extension of ResNet-Att. Recently, CNN-LSTM seems to be the most widely used for time series processing [17, 28]. However, ResNet could also be applied to time series data. We would choose the highest prediction accuracy by comparing several algorithms.

3.1 Overall Framework

Our proposed framework is illustrated in Fig 1. We take the processed time series data as input. CNN, CNN-LSTM, ResNet, and ResNet-Att were selected as model objects for comparison. ResNet, although not commonly used for time series data, has surprisingly achieved excellent results. Recently, the attention mechanism has gained great popularity in DL. It refers to the human thinking mode that is able to scan data and focus on the desired part in the target field. However, we have found that the superposition of multilayer networks can lead to overfitting, and the attention mechanism is not suitable for all network models.

3.2 Data Processing

In this paper, the public datasets in [4] are processed. ZHONG B et al. organized seven healthy subjects and one trans-tibial amputee to participate in this study. Different locations for wearable cameras were chosen to collect images of the subject in the environment. IMU signals were also collected from sensors attached to the lower limbs. The lower limb device was attached to the shin area of the subjects. For the amputee subject, the device was attached to the top of the pants around the prosthetic socket of a passive lower limb prosthesis. A time series dataset containing accelerometer and gyroscope sensor readings and timestamps was used to predict human locomotion intentions.

In this dataset, we obtained four categories including flat ground, grass with special terrain, up and down stairs. These scenes are representative and have wide applicability, which can be used not only in LLE rehabilitation scenarios but also in assistive situations. In Table 2, we have labeled the terrains and described the distribution of the data labels.

Table 1 Previous recognition systems used algorithms for classification

| Dataset Category	References	Method	Description
Image	Kurbis et al. [11]	MobileNetV2	The development of a large-scale automated stair recognition system powered by MobileNetV2 was proposed to recognize indoor and outdoor real-world stair environments
	Kemaev et al. [12]	ResNet	ResNet architecture and its modifications produce state-of-the-art results in image classification problems
Human motion signal	Wang et al. [13]	Hidden Semi-Markov Models	The prediction method is based on Hidden Semi-Markov Models. The orientation of the dataset is collected by the Inertial Measurement Unit
	Patzer et al. [14]	Hidden Markov Model	Online classifications of a human wearing a LLE in two directions were extended
	Wu et al. [15]	GCN	The GCNM can recognize four leg phases between the foot and ground, including heel strike, foot flat, heel off, and swing
	Ren et al. [16]	LSTM	A wearable joint angle measurement device was designed for gait trajectory prediction, which can be used for predictive control through machine learning methods
	Chen et al. [17]	CNN—LSTM	The IMU is installed on the exoskeleton to collect motion information, which is used for gait pattern classification
	Su et al. [18]	LSTM	LSTM was tested for generalization across participants and designed to learn lower limb segment trajectories using training samples
	Li et al. [19]	LSTM	LSTM was designed for pattern recognition, such as level ground walking, stair ascending, stair descending, ramp ascending, and ramp descending
	Zhu et al. [20]	CNN-LSTM	The feasibility of the method was verified by analyzing the gait prediction results of different subjects
	Lu et al. [21]	CNN-LSTM	The model estimated the hip, knee, and ankle joint angles from sEMG signals in locomotion modes including walk, run, stair descent, stair ascent, stand-to-sit, sit-to-stand, and jump
Fusion	Guo et al. [22]	Machine Learning	Exploring a multi-source information fusion model for exoskeleton gait pattern transition recognition in terms of two aspects of multi-source information fusion strategy and multi-classifier fusion
	Zhang et al. [23]	CNN-LSTM	A precise fusion method of EEG and sEMG using the CNN-LSTM model was investigated to detect lower limb active movement
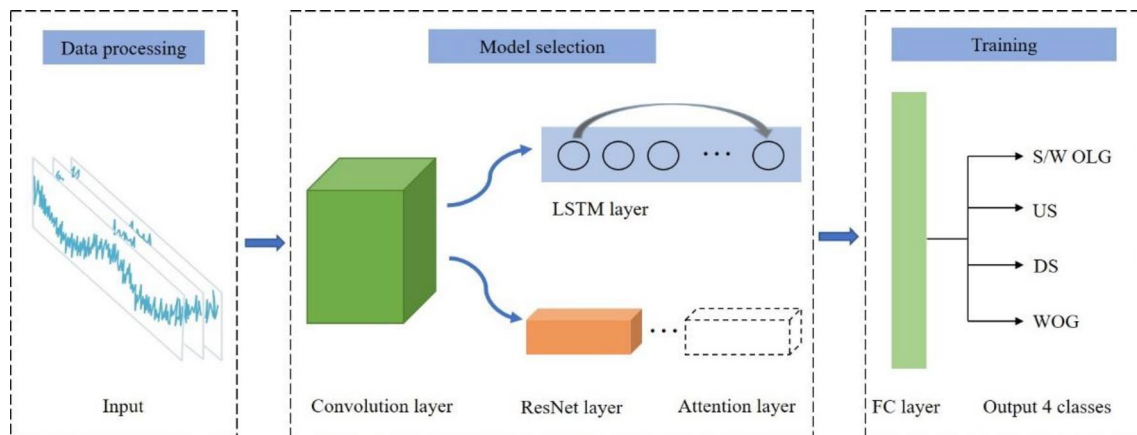
Fig. 1 Experimental network framework structure

Table 2 Labels of terrains and data points

Locomotion intention labels	Data points
0: Standing/walking on level ground (S/WOLG)	251,292
1: Up the stairs (US)	13,811
2: Down the stairs (DS)	18,266
3: Walking on grass (WOG)	51,609

Table 2 shows that the distribution of labels is unbalanced. S/WOLG labels have a high proportion, which could lead to dilution of some features and affect the experimental effect. To ensure that each label datapoint contains the same number (100,000), we performed a balanced resampling of the accelerometer and gyroscope signal data. The window size and the sliding window were set to 60 and 4, respectively. Thus, a single prediction is made for all four points.

From Fig. 2, we can see that the accelerometer and the gyroscope data of S/WOLG (Label 0) are relatively stable, and the data of US (Label 1) is the most volatile, which is also related to the increase in the motion amplitude when walking up the stairs. DS (Label 2) is much more stable than the US. Compared to S/WOLG, WOG (Label 3) produces noisy data when the road surface becomes uneven and unstable.

3.3 Model Proposed

(1) CNN-Based

The CNN consists of several different layers. We designed a structure that was suitable for this experiment. Three 1D convolutional layers were added. Given the input signal $x(n)$, the output $y(n)$ can be obtained by convolving the signal $x(n)$ with the convolution kernel $\omega(n)$ of size l [29].

$$y(n) = x(n) \times \omega(n) = \sum_{m=0}^{l-1} x(m) \cdot \omega(n-m) \tag{1}$$

To reduce overfitting, we added the dropout layer. The batch-normal layer could facilitate the stabilization of the network during training. The framework of the convolution layer is shown in Table 3.

(2) CNN-LSTM-Based

The CNN-LSTM is composed of CNN and LSTM. The gate structure of LSTM [30] mainly consists of a forgetting gate, an input gate, and an output gate in Fig. 3. The forgetting gate simulates the action of the human brain and represents the discarded information.
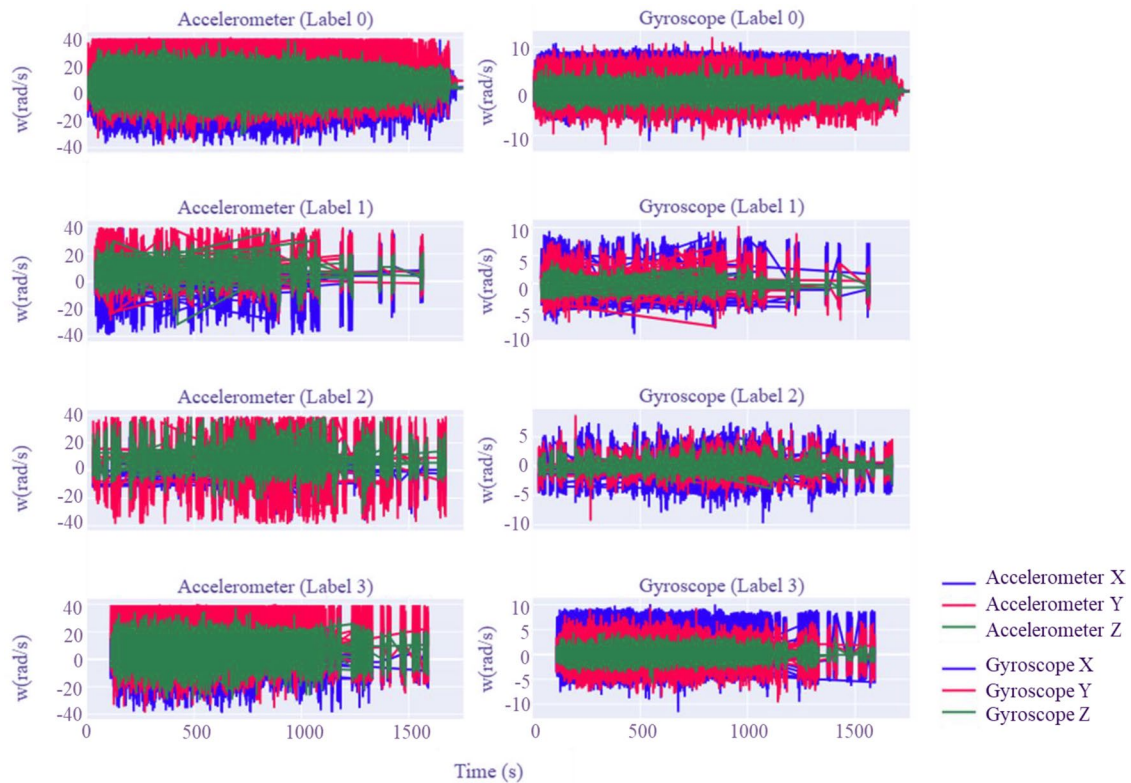

Fig. 2 The raw data related to detailed presentations

Table 3 Layer parameters of the CNN architecture

Name	In_channels	Out_channels	Kernel size	Stride
1D convolution 1	6	12	5	1
1D BatchNorm 1	–	–	–	–
Relu1	12	–	–	–
1D convolution 2	12	24	5	1
1D BatchNorm 2	–	–	–	–
Relu2	24	–	–	–
1D convolution 3	24	48	5	1
1D BatchNorm 3	–	–	–	–
Relu3	48	–	–	–
Full connection 1	2304	256	–	–
1D BatchNorm 4	–	–	–	–
Relu4	256	–	–	–
Full connection 2	256	64	–	–
1D BatchNorm 5	–	–	–	–
Dropout	–	–	–	–
Full connection 3	64	4	–	–

"In_channels" means the number of channels in the input. "Out_channels" is the number of channels produced by the convolution. "Kernel size" is the size of the convolution kernel. "Stride" is the convolution stride. A "–" indicates that the specified parameter is not available

Using the input parameters, each layer is calculated by the following functions:

$$i_t = \sigma(\omega_{ii}x_t + b_{ii} + \omega_{hi}h_{t-1} + b_{hi}) \tag{2}$$

$$f_t = \sigma(\omega_{if}x_t + b_{if} + \omega_{hf}h_{t-1} + b_{hf}) \tag{3}$$

$$\widetilde{c}_t = tanh(\omega_{\widetilde{ic}}x_t + b_{\widetilde{ic}} + \omega_{h\widetilde{c}}h_{t-1} + b_{h\widetilde{c}}) \tag{4}$$

$$o_t = \sigma(\omega_{io}x_t + b_{io} + \omega_{ho}h_{t-1} + b_{ho}) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{6}$$

$$h_t = o_t \odot \tanh\left(c_t\right) \tag{7}$$

CNN-LSTM takes advantage of CNN to extract spatial features, and LSTM promotes the extraction of input information [31]. The network structure we designed is shown in Table 4.

(3) ResNet-Based

The ResNet [32] is widely used for feature extraction. With the continuous depth of CNN, this phenomenon may cause the convergence of the network to be degraded, the accuracy to deteriorate, and overfitting to occur. To solve this problem, ResNet has been proposed. In this work, we selected ResNet-50 to realize the recognition of human locomotion intention.

Attention mechanisms [33] have been used by many researchers to improve the performance of the network. Therefore, in this paper, we also propose the possibility that attention mechanisms can improve the model accuracy. However, we also found that the additional attention mechanisms would cause an overfitting phenomenon when the accuracy of the original model reached a high level, which is not applicable to this experiment. The corresponding evidence is provided in the following part.

The proposed framework is described in Fig. 4.

The Channel Attention Module (CAM) [34] was added to ResNet, its description can be found in Fig. 5, which is composed of ResNet-Att. Channel attention refers to a
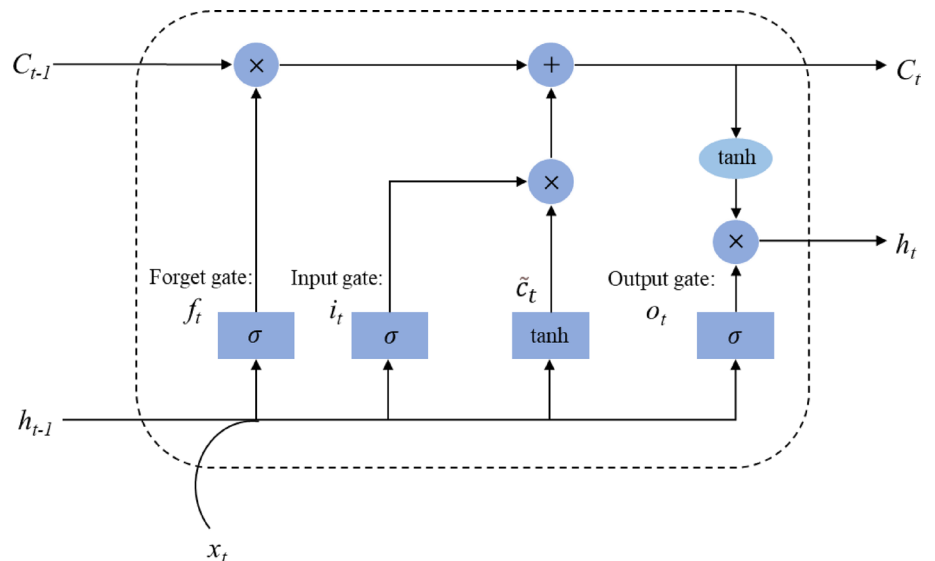
Fig. 3 LSTM cell structure

Table 4 Layer parameters of the LSTM-CNN structure

Name	In_channels	Out_channels	Kernel size	Stride
1D convolution 1	6	12	5	1
1D BatchNorm 1	–	–	–	–
Relu 1	12	–	–	–
1D convolution 2	12	24	5	1
1D BatchNorm 2	–	–	–	–
Relu 2	24	–	–	–
Dropout 1	–	–	–	–
1D convolution 3	24	48	5	1
1D BatchNorm 3	–	–	–	–
Relu 3	48	–	–	–
Dropout 2	–	–	–	–
LSTM	48	128	–	–
Full connection 1	7680	256	–	–
1D BatchNorm 4	–	–	–	–
Relu 4	256	–	–	–
Dropout 3	–	–	–	–
Full connection 2	256	64	–	–
1D BatchNorm 5	–	–	–	–
Relu 5	64	–	–	–
Dropout 4	–	–	–	–
Full connection 3	64	4	–	–

mechanism that allows a network to weight feature mappings according to context in order to achieve better performance [35, 36]. We restrict ourselves to channel attention in this work since its execution typically requires less computation.

As shown in Fig. 5, given an input $F \in R^{C \times H \times W}$, M_C and M_S are obtained through MaxPool, AvgPool, and then through Shared MLP, which are calculated by the following functions:

$$M_c(F) = BN(MLP(MaxPool(F)))$$
$$= BN(\omega_1(\omega_0 MaxPool(F) + b_0) + b_1)) \tag{8}$$

$$M_s(F) = BN(MLP(AvgPool(F)))$$
$$= BN(\omega_1(\omega_0 AvgPool(F) + b_0) + b_1)) \tag{9}$$

$$M(F) = \sigma(M_c(F) + M_s(F)) \tag{10}$$

where $\omega_0 \in R^{\frac{C}{r} \times C}, b_0 \in R^{\frac{C}{r}}, \omega_1 \in R^{C \times \frac{C}{r}}, b_1 \in R^C$, r is the reduction ratio, BN is defined as a batch normalization operation, and σ is a sigmoid function.

The input F is the feature extracted by ResNet, CAM directly weighted F. The overall process can be concluded:

$$F\prime = M_c(F) \otimes F \tag{11}$$

$$F'' = M_s(F) \otimes F\prime \tag{12}$$

where $\otimes$ represents element-wise multiplication., F'' is the final refined output.

4 Experiments

4.1 Experimental Setting

In contrast to Zhong et al., we did not use images for dataset [4], but decided to use the IMU signal and the time stamp. The train-test split is 8:2. Data from healthy individuals were used for the training set, while the test set included transtibial amputees. The performance of CNN, CNN-LSTM, ResNet, and ResNet-Att was compared.

The network was trained with an Adam optimizer. The epoch and the batch sizes were 20 and 64, respectively, and
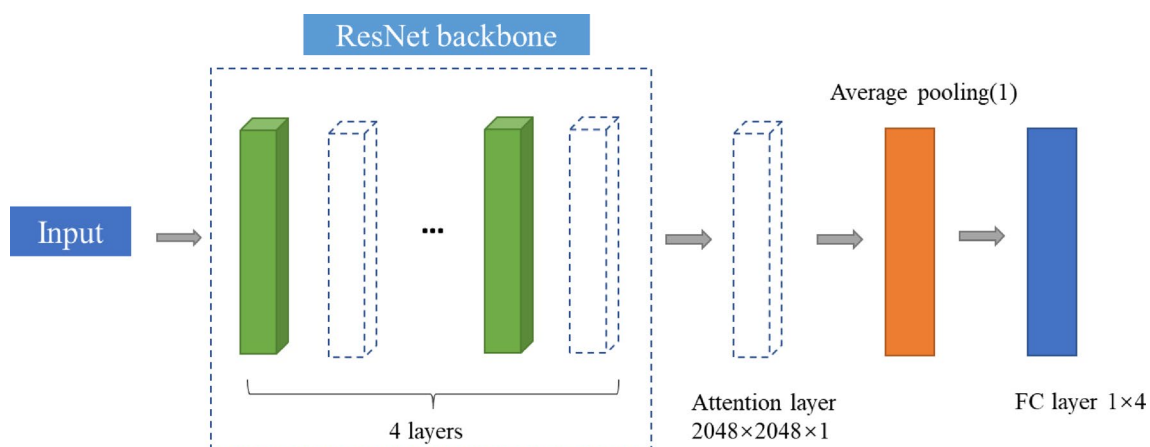

Fig. 4 The basic structure of Resnet50 and attention mechanisms have been added. The green module is the bottleneck of the ResNet, and each layer is constructed by a number of blocks. The dotted box is where we would insert the attention mechanism
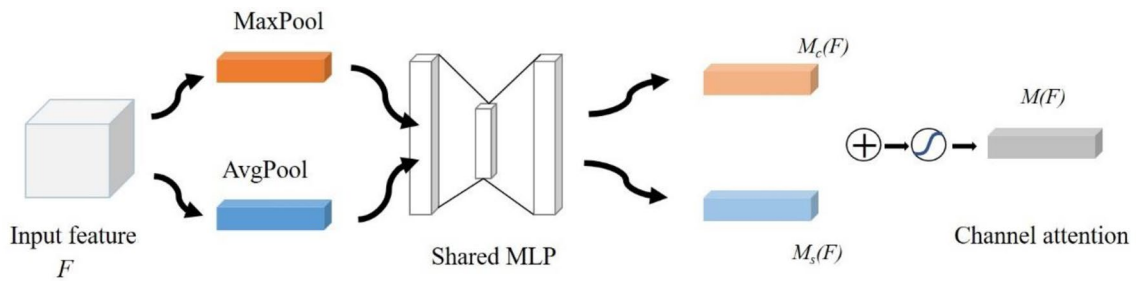
Fig. 5 The structure of the CAM. The input features are subjected to maximum and average pooling, respectively. Then, the results are used as input to the shared MLP layer. The processed results are added and activated with sigmoid. The weight of each CAM is obtained, which would be multiplied by the features
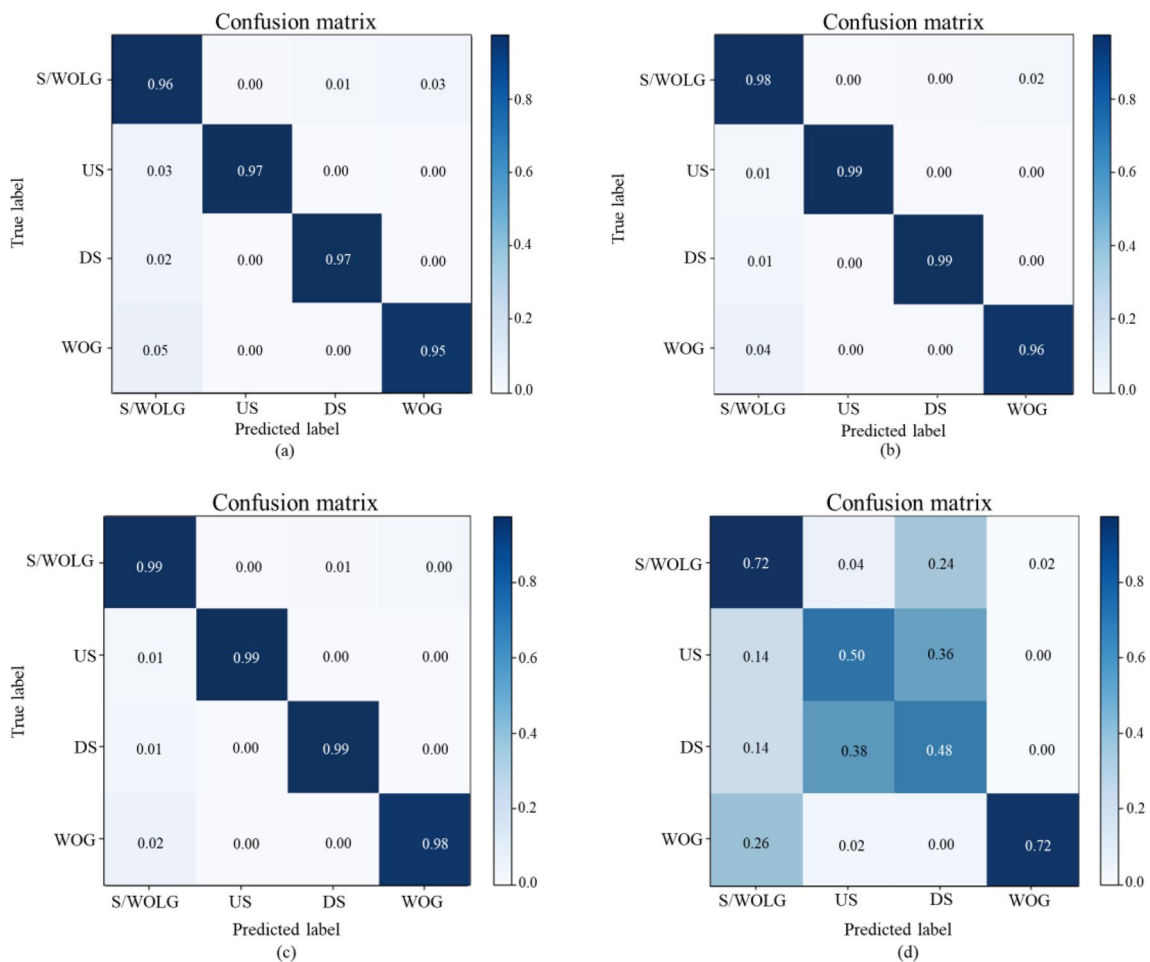

Fig. 6 The confusion matrix test accuracy of the CNN (**a**), LSTM-CNN (**b**), ResNet (**c**), and ResNet-Att (**d**)
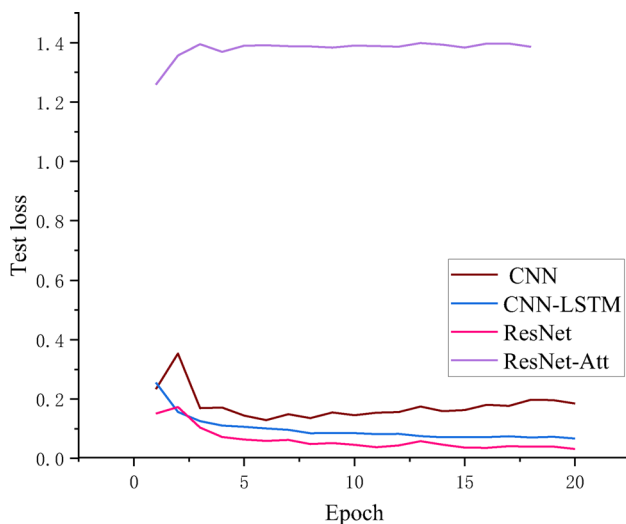
Fig. 7 The test loss of CNN, CNN-LSTM, ResNet, and ResNet-Att

Table 5 Classification report of models

Model	Precision	Recall	F1 score	Accuracy
CNN	0.96	0.92	0.94	0.96
CNN-LSTM	0.98	0.96	0.97	0.98
ResNet	**0.99**	**0.98**	**0.98**	**0.99**
ResNet-Att	0.61	0.60	0.61	0.61

Bold values show our experimental results, which are intended to attract readers' attention and represent the outstanding performance of the algorithm

the dropout rate was set to 0.1. The network was implemented using PyTorch and tested on a computer with an AMD Ryzen 7 5800H with Radeon Graphics, a 16 GB memory chip, and a graphics card (NVIDIA GeForce RTX3070).

4.2 Results on Datasets

A comparative experiment was performed with CNN, CNN-LSTM, ResNet, and ResNet-Att. The accuracy of the confusion matrix is shown in Fig. 6a–d, which is high in Fig. 6a–c. ResNet outperformed the competition in the four algorithms, achieving 99%, 99%, 99%, and 98%, respectively. WOG can be easily identified as S/WOLG, which was selected because some research [37] pointed out that the special terrain has not been studied yet. We believe that it is necessary for LLE to detect in special terrain, which is of great help for auxiliary. The performance of recognition in special terrain can help LLE to adapt to different walking speeds. Accuracy, precision, recall, F1 score and loss were chosen as evaluation criteria.

As shown in Fig. 6d, the ResNet-Att performed worse, which made the network and the optimization process more complex. The ResNet performed well, and the additional attention mechanism could lead to overfitting.

From Fig. 7, it can be seen that the loss curve of ResNet decreases compared to the other models and has the best performance. The loss is almost 0. However, the loss of ResNet-Att is around 1.4, so it is not suitable to add an attention mechanism to this experiment. ResNet has achieved a higher classification in the intention recognition of LLE. Therefore, after the experimental demonstration, we would use ResNet network for the subsequent deployment on the lower computer.

As shown in Table 5, precision is the positive category that takes into account the prediction of all samples. Recall is the correctly predicted positive categories with all actual positive samples and measures the number of actual positive cases that can be recalled. The F1 score considers both precision and recall. It is high only when both precision and recall rates are in high proportion.

According to the above evaluation criteria, ResNet performs outstandingly in this dataset, which confirms the future applications of intention recognition in LLE. Compared to most current studies using CNN and CNN-LSTM networks, ResNet also has a great advantage in recognition.

However, future studies need to make more improvements. We should not use only the kinematic data, which may have limitations in the complex real-world conditions. Research is needed in the area of multi-sensor data fusion, where data from the vision system can be used to complement automatic motion pattern control decisions based on mechanical, inertial, and/or neuromuscular sensors. This is because the single environmental feature does not clearly express the user's motion intentions in real life. Fusion of camera data with kinematic data could improve performance. Although data fusion is currently insufficient, we will investigate this aspect further in future.

5 Discussion

In this work, an offline dataset was used to train and evaluate the framework. The offline dataset was collected from seven healthy subjects and one trans-tibial amputee. The training and testing procedure is described as follows: (1) Dividing the offline dataset into a training, and a test dataset; (2) Balancing the training dataset of data points with different labels to obtain the same number of 100,000 each; (3) Setting the window size (60) and sliding window (4). The data is normalized at the interval of (0, 0.5); (4) Training the locomotion prediction network with the training dataset. The input of the network is the human motion signal, and the output of the prediction network is the locomotion category; (5) Perform dropout sampling to obtain predictions from the

Table 6 Compared with previous research that used different neural networks for human locomotion classification

Reference	Datasets	Model	Performance			
			Precision	Recall	F1	Accuracy
[38]	UCI-HAR[39]	CNN	0.93	0.93	0.93	0.93
		LSTM	0.89	0.89	0.89	0.89
		MLP	0.87	0.87	0.87	0.87
	Pamap2[40]	CNN	0.92	0.91	0.91	0.91
		LSTM	0.87	0.85	0.85	0.86
		MLP	0.83	0.82	0.82	0.82
[41]	UCI-HAR	CNN	–	–	0.92	–
		CNN-LSTM	–	0.96	0.96	0.96
	WISDM[42]	CNN	–	–	0.93	–
		CNN-LSTM	–	0.96	0.96	0.96
	OPPORTUNITY[43]	CNN	–	–	0.85	–
		CNN-LSTM	–	0.93	0.92	0.93
Ours	[4]	ResNet	**0.99**	**0.98**	**0.98**	**0.99**

Bold values show our experimental results, which are intended to attract readers' attention and represent the outstanding performance of the algorithm

trained terrain prediction network for the test dataset; (6) Evaluate the trained framework with the test datasets.

We performed a balanced resampling of the data to avoid the large impact on the accuracy after the softmax layer caused by the label imbalance problem, and carried out a series of normalizations to map the data to the same range.

As shown in Fig. 6, the highest classification accuracy for the classification algorithm presented in this paper can reach almost 99% in the experiments, which is about 3% higher than the accuracy of CNN classifier. Compared with CNN, ResNet added a shortcut connection, namely the residual unit, which makes the network not too deep. However, ResNet-Att made the networks more complex, which reduced the accuracy by approximately 30% compared to ResNet. CNN-LSTM was often used to classify a time series dataset into different motion patterns, and performed 1% lower than ResNet. ResNet appears to have great potential.

According to Table 6, when compared to [38, 41], CNN performs similarly to our experimental part in recognizing human motion intention, and the present CNN-LSTM fusion algorithm in the state of detection, which has a good performance. Its performance is indeed significantly improved compared to CNN. However, the performance is still inferior to ResNet. Therefore, our proposed application of ResNet for human activity recognition is successful.

We believe that the application of ResNet in the field of intention recognition is feasible. Compared to CNN, ResNet is an improved algorithm with better performance than CNN. However, most current studies focus on the improved performance of CNN-LSTM, which undoubtedly complicates the network. This is an undesirable measure. Considering that the real-time performance of intention recognition is significant, the complexity of the model will greatly reduce the real-time performance of the recognition system. Therefore, we proposed that ResNet is more suitable for intention recognition than CNN and CNN-LSTM, and our results also show that ResNet will have better performance. At the same time, we also assumed that the attention mechanism would improve the performance. Contrary to our expectation, the attention module by ResNet would reduce the accuracy and make the results overfitting. In future experiments, we will reconsider and further discuss the addition of attention mechanisms. In this work, we conclude that ResNet has the best performance in the time series of intention recognition. We will classify ResNet as the algorithm of our intention recognition system in an actual prototype experiment.

In addition, the four common types of locomotion intentions were accurately estimated in this paper. Most of the estimation errors in this paper were less than 4%. Data fusion could enable improvement of the system. In [4], the lower limb camera is combined with an on-glasses camera, which can facilitate the prediction of distant terrain. The result showed that the accuracy was significantly improved. However, since the camera is not self-contained in the wearable robots, adding an on-glasses camera may increase the number of frames to be processed, resulting in lower system efficiency. Therefore, in [44], two cameras were used to capture the images simultaneously. The feature vectors of the images from both cameras were concatenated by feature-level fusion, but this is not the best solution. The two cameras can be activated asynchronously to dynamically combine the advantages of both cameras for different scenarios. However, the camera can cause privacy issues, so most current research does not provide an effective processing method in data fusion. Some studies

simply spliced datasets to achieve low-level data fusion, while the mainstream is feature-level fusion, which may cause relatively large feature loss in real processing, so decision-level fusion is increasingly perceived by the public compared to the previous two situations. In future, we will conduct a series of comparative experiments on these three fusion methods to verify the appropriate method and improve the performance of the algorithm. We believe that data fusion will have a positive impact on recognition accuracy. We will continue to explore the limitations of this aspect.

6 Conclusion

A time series-based locomotion recognition was developed for LLE. In this study, we used the dataset of seven healthy subjects and one trans-tibial amputee. Four locomotion modes, including S/WOLG, US, DS, and WOG, were analyzed during the experiment. To facilitate comparative experiments, four models were proposed in this study and an attention mechanism was added. We conclude that ResNet has great potential for processing time series datasets. The promising results are expected to significantly improve the decision making in locomotion recognition of LLE. The high classification accuracy in this work provides a good theoretical illustration for the intention recognition of LLE. In the subsequent experiments of the lower computer, we will also use ResNet for experimental demonstration. Although it has not yet been verified on an actual prototype, we will use it in follow-up experiments to prove its performance in further studies.

We are concerned that in the field of intention recognition, the realistic environment of the exoskeleton is complex, and it is not enough to process only homogeneous data. The reality is composed of multi-source, heterogeneous data [45]. Therefore, in future, we will develop a series of multi-source information acquisition devices in the design laboratory, which are not limited to kinematic data. In addition, in-depth research on multi-source heterogeneous fusion [46, 47] methods and algorithms will be conducted, which is also lacking in many current studies.

Data Availability All data and materials related to the study can be obtained through contacting the author at 213332822@st.usst.edu.cn.

Declarations

Conflict of interest The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

References

1. Mooney, L. M., Rouse, E. J., & Herr, H. M. (2014). Autonomous exoskeleton reduces metabolic cost of human walking during load carriage. *Journal of Neuroengineering and Rehabilitation, 11*, 1–11. https://doi.org/10.1186/1743-0003-11-80
2. Yang, J. T., Sun, T. R., Cheng, L., & Hou, Z. G. (2022). Spatial repetitive impedance learning control for robot-assisted rehabilitation. *IEEE/ASME Transactions on Mechatronics, 28*, 1280–1290. https://doi.org/10.1109/TMECH.2022.3221931
3. Mokhtari, M., Taghizadeh, M., & Mazare, M. (2021). Impedance control based on optimal adaptive high order super twisting sliding mode for a 7-DOF lower limb exoskeleton. *Meccanica, 56*, 535–548. https://doi.org/10.1007/s11012-021-01308-4
4. Zhong, B. X., Da Silva, R. L., Li, M., Huang, H., & Lobaton, E. (2020). Environmental context prediction for lower limb prostheses with uncertainty quantification. *IEEE Transactions on Automation Science and Engineering, 18*, 458–470. https://doi.org/10.1109/TASE.2020.2993399
5. Tucker, M. R., Olivier, J., Pagel, A., Bleuler, H., Bouri, M., Lambercy, O., Millán, J. D. R., Riener, R., Vallery, H., & Gassert, R. (2015). Control strategies for active lower extremity prosthetics and orthotics: a review. *Journal of Neuroengineering and Rehabilitation, 12*, 1–30. https://doi.org/10.1186/1743-0003-12-1
6. Young, A. J., & Ferris, D. P. (2016). State of the art and future directions for lower limb robotic exoskeletons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25*, 171–182. https://doi.org/10.1109/TNSRE.2016.2521160
7. Mokhtari, M., Taghizadeh, M., & Mazare, M. (2021). Hybrid adaptive robust control based on CPG and ZMP for a lower limb exoskeleton. *Robotica, 39*, 181–199. https://doi.org/10.1017/S0263574720000260
8. Hu, B., Rouse, E., & Hargrove, L. (2018). Fusion of bilateral lower-limb neuromechanical signals improves prediction of locomotor activities. *Front Robot AI, 5*, 78. https://doi.org/10.3389/frobt.2018.00078
9. Huang, H., Zhang, F., Hargrove, L. J., Dou, Z., Rogers, D. R., & Englehart, K. B. (2011). Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion. *IEEE Transactions on Biomedical Engineering, 58*, 2867–2875. https://doi.org/10.1109/TBME.2011.2161671
10. Laschowski, B., McNally, W., Wong, A., & McPhee, J. (2022). Environment classification for robotic leg prostheses and exoskeletons using deep convolutional neural networks. *Frontiers in Neurorobotics, 15*, 1–17. https://doi.org/10.3389/fnbot.2021.730965
11. Kurbis, A. G., Laschowski, B., & Mihailidis, A. (2022). Stair recognition for robotic exoskeleton control using computer vision and deep learning. *IEEE International Conference on Rehabilitation Robotics, Rotterdam, Netherlands, 2022*, 1–6. https://doi.org/10.1109/ICORR55369.2022.9896501
12. Kemaev, I., Polykovskiy, D., & Vetrov, D. (2018). Reset: learning recurrent dynamic routing in resnet-like neural networks. *The 10th Asian Conference on Machine Learning*, Beijing, China, 95, 422–437. https://doi.org/10.48550/arXiv.1811.04380
13. Wang, M., Wu, X. Y., Liu, D. X., & Wang, C. (2016). A human motion prediction algorithm based on HSMM for SIAT's exoskeleton. *The 35th Chinese Control Conference*, Chengdu, China, 3891–3896. https://doi.org/10.1109/ChiCC.2016.7553959
14. Patzer, I., & Asfour, T. (2019). Minimal sensor setup in lower limb exoskeletons for motion classification based on multi-modal sensor data. *IEEE International Conference on Intelligent Robots and Systems*, Macau, China, 8164–8170. https://doi.org/10.1109/Humanoids43949.2019.9035014

15. Wu, X. Y., Yuan, Y., Zhang, X. K., Wang, C., Xu, T. T., & Tao, D. C. (2022). Gait phase classification for a lower limb exoskeleton system based on a graph convolutional network model. *IEEE Transactions on Industrial Electronics, 69*, 4999–5008. https://doi.org/10.1109/tie.2021.3082067

16. Ren, B., Zhang, Z. Q., Zhang, C., & Chen, S. L. (2022). Motion trajectories prediction of lower limb exoskeleton based on long short-term memory (LSTM) networks. *Actuators, 11*, 1–15. https://doi.org/10.3390/act11030073

17. Chen, C. F., Du, Z. J., He, L., Shi, Y. J., Wang, J. Q., & Dong, W. (2021). A novel gait pattern recognition method based on LSTM-CNN for lower limb exoskeleton. *Journal of Bionic Engineering, 18*, 1059–1072. https://doi.org/10.1007/s42235-021-00083-y

18. Su, B. B., & Gutierrez-Farewik, E. M. (2020). Gait trajectory and gait phase prediction based on an LSTM network. *Sensors, 20*, 1–17. https://doi.org/10.3390/s20247127

19. Li, J. X., Gao, T., Zhang, Z. H., Wu, G. H., Zhang, H., Zheng, J. B., Gao, Y. F., & Wang, Y. (2022). A novel method of pattern recognition based on TLSTM in lower limb exoskeleton in many terrains. *The 4th International Conference on Intelligent Control, Measurement and Signal Processing*, Hangzhou, China, 733–737. https://doi.org/10.1109/ICMSP55950.2022.9859005

20. Zhu, M., Guan, X. R., Li, Z., He, L., Wang, Z., & Cai, K. S. (2023). sEMG-based lower limb motion prediction using CNN-LSTM with improved PCA optimization algorithm. *Journal of Bionic Engineering, 20*, 612–627. https://doi.org/10.1007/s42235-022-00280-3

21. Lu, Y. Z., Wang, H., Zhou, B., Wei, C. F., & Xu, S. Q. (2022). Continuous and simultaneous estimation of lower limb multi-joint angles from sEMG signals based on stacked convolutional and LSTM models. *Expert Systems with Applications, 203*, 1–20. https://doi.org/10.1016/j.eswa.2022.117340

22. Guo, C. Y., Song, Q. Z., & Liu, Y. L. (2022). Research on the application of multi-source information fusion in multiple gait pattern transition recognition. *Sensors (Basel), 22*, 1–12. https://doi.org/10.3390/s22218551

23. Zhang, X. D., Li, H. Z., Dong, R. L., Lu, Z. F., & Li, C. X. (2022). Electroencephalogram and surface electromyogram fusion-based precise detection of lower limb voluntary movement using convolution neural network-long short-term memory model. *Frontiers in Neuroscience, 16*, 1–21. https://doi.org/10.3389/fnins.2022.954387

24. Zhang, K. E., Wang, J., De Silva, C. W., & Fu, C. L. (2020). Unsupervised cross-subject adaptation for predicting human locomotion intent. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*, 646–657. https://doi.org/10.1109/TNSRE.2020.2966749

25. Kuniaki Saito, Watanabe, K., Ushiku, Y., & Harada, T. (2018). Maximum classifier discrepancy for unsupervised domain adaptation. *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 3723–3732. https://doi.org/10.1109/CVPR.2018.00392

26. Zhang, K. E., Xiong, C. H., Zhang, W., Liu, H. Y., Lai, D. Y., Rong, Y. M., & Fu, C. L. (2019). Environmental features recognition for lower limb prostheses toward predictive walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27*, 465–476. https://doi.org/10.1109/TNSRE.2019.2895221

27. Hur, T., Bang, J., Huynh-The, T., Lee, J. W., Kim, J. I., & Lee, S. Y. (2018). Iss2Image: A novel signal-encoding technique for CNN-based human activity recognition. *Sensors (Basel), 18*, 1–19. https://doi.org/10.3390/s18113910

28. Khatun, M., Yousuf, M., Ahmed, S., Uddin, M. Z., Alyami, S., Al-Ashhab, S., Akhdar, H., Khan, A., Azad, A. K. M., & Moni, M. A. (2022). Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE Journal of Translational Engineering in Health and Medicine, 10*, 1–1. https://doi.org/10.1109/JTEHM.2022.3177710

29. Zhao, J. F., Mao, X., & Chen, L. J. (2018). Learning deep features to recognise speech emotion using merged deep CNN. *IET Signal Processing, 12*, 713–721. https://doi.org/10.1049/iet-spr.2017.0320

30. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

31. Zhou, X., Wu, X. T., Ding, P., Li, X. G., He, N. H., Zhang, G. Z., & Zhang, X. X. (2019). Research on transformer partial discharge UHF pattern recognition based on CNN-LSTM. *Energies, 13*, 1–13. https://doi.org/10.3390/en13010061

32. He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J., (2016). Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 770–778. https://doi.org/10.1109/CVPR.2016.90

33. Liu, T. L., Luo, R. H., Xu, L. Q., Feng, D. C., Cao, L., Liu, S. Y., & Guo, J. J. (2022). Spatial channel attention for deep convolutional neural networks. *Mathematics, 10*, 1–10. https://doi.org/10.3390/math10101750

34. Woo, S., Park, J., Lee, J.Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. *The 15th European Conference on Computer Vision*, Munich, Germany, 11211, 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

35. Zhang, H., Wu, C. R., Zhang, Z. Y., Zhu, Y., Lin, H. B., Zhang, Z., Sun, Y., He, T., Mueller, J., & Manmatha, R. (2022). Resnest: Split-attention networks. *Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 2736–2746. https://doi.org/10.1109/CVPRW56347.2022.00309

36. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

37. Pinto-Fernandez, D., Torricelli, D., del Carmen Sanchez-Villamanan, M., Aller, F., Mombaur, K., Conti, R., Vitiello, N., Moreno, J. C., & Pons, J. L. (2020). Performance evaluation of lower limb exoskeletons: A systematic review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*, 1573–1583. https://doi.org/10.1109/TNSRE.2020.2989481

38. Wan, S. H., Qi, L. Y., Xu, X. L., Tong, C., & Gu, Z. H. (2020). Deep learning models for real-time human activity recognition with smartphones. *Mobile Networks and Applications, 25*, 743–755. https://doi.org/10.1007/s11036-019-01445-x

39. Reyes-Ortiz, J. L., Oneto, L., Samà, A., Parra, X., & Anguita, D. (2016). Transition-aware human activity recognition using smartphones. *Neurocomputing, 171*, 754–767. https://doi.org/10.1016/j.neucom.2015.07.085

40. Reiss, A., & Stricker, D. (2012). Introducing a new benchmarked dataset for activity monitoring. *The 16th International Symposium on Wearable Computers*, Newcastle, England, 108–109. https://doi.org/10.1109/ISWC.2012.13

41. Xia, K., Huang, J. G., & Wang, H. Y. (2020). LSTM-CNN architecture for human activity recognition. *IEEE Access, 8*, 56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

42. Kwapisz, J. R., Weiss, G. M., & Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter, 12*, 74–82. https://doi.org/10.1145/1964897.1964918

43. Roggen, D., Calatroni, A., Rossi, M., Holleczek, T., Förster, K., Tröster, G., Lukowicz, P., Bannach, D., Pirkl, G., & Ferscha, A. (2010). Collecting complex activity datasets in highly rich networked sensor environments. *The 7th International Conference on Networked Sensing Systems*, Kassel, Germany, 233–240. https://doi.org/10.1109/INSS.2010.5573462

44. Zhong, B. X., Silva, R. L. D., Tran, M., Huang, H., & Lobaton, E. (2022). Efficient environmental context prediction for lower limb prostheses. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52*, 3980–3994. https://doi.org/10.1109/TSMC.2021.3084036

45. Zhang, L. L., Xie, Y. X., Xidao, L., & Zhang, X. (2018). Multi-source heterogeneous data fusion. *International Conference on Artificial Intelligence and Big Data*, Chengdu, China, 47–51. https://doi.org/10.1109/ICAIBD.2018.8396165

46. Jiang, M. M., Wu, Q., & Li, X. T. (2022). Multisource heterogeneous data fusion analysis of regional digital construction based on machine learning. *Journal of Sensors, 2022*, 1–11. https://doi.org/10.1155/2022/8205929

47. Zhang, F., Yang, J., Sun, C., Guo, X., & Wan, T. T. (2021). Research on multi-source heterogeneous data fusion technology of new energy vehicles under the new four modernizations. *Journal of Physics: Conference Series, 1865*, 1–15. https://doi.org/10.1088/1742-6596/1865/2/022034