

Intent-Aware Motion Analysis with Deep Learning: Methods, Applications, and Future Directions

Qingyu Rui

The High School Affiliated to
Renmin University of China,
Shenzhen

*Corresponding author:
ruiqingyu080405@gmail.com

Abstract:

Recent advances in intelligent sensors coupled with next-generation interface technologies have made motion intention recognition a foundational technology for assistive systems—including prosthetic controllers, exoskeleton walkers, and rehabilitation devices. Focusing on deep learning applications, this review examines current developments in motion intention recognition across different countries, covering both signal analysis and behavioral pattern identification. By comparing the advantages and disadvantages of different signal types (such as electromyographic signals, muscle pressure signals, inertial sensor signals, etc.) and recognition algorithms (traditional machine learning and deep neural networks), the significant advantages of deep learning in improving recognition accuracy and adaptability are revealed. Additionally, by reviewing recent representative achievements, the practical deployment of motion intention recognition technology in lower-limb exoskeletons and prosthetic systems is discussed. Finally, the main challenges in current research, such as signal heterogeneity, individual user differences, and model generalization ability, are summarized, and future development directions based on multimodal fusion and transfer learning are envisioned.

Keywords: Motion intention recognition; deep learning; convolutional neural networks; electromyographic signals; human-computer interaction.

1. Introduction

During the recent past, with the acceleration of an aging society and the growing demand for rehabilitation, how to efficiently recognize and respond to

human motion intentions through intelligent assistive devices has become a hot topic in the interdisciplinary research of biomedical engineering and artificial intelligence.

Motion intention recognition studies have progressed

from conventional machine learning approaches to fully integrated deep learning architectures, establishing a robust theoretical foundation for rehabilitative therapies and smart prosthetic systems. Domestic research started with multi-source data fusion. Cao Xianghong et al. (2017) constructed a multimodal gait recognition system based on hip joint angle, acceleration, and foot pressure. By using kernel principal component analysis and extreme learning machines, they achieved a recognition accuracy of 96.78% for six gait types with a response time of 0.52 seconds, although there is still room for improvement in feature extraction speed and computational complexity[1]. Ni Yu and Su Benyue (2020) proposed a lower-limb prosthesis intention recognition model that combines convolutional neural networks with LSTM. After removing manual feature selection, the accuracy surpassed traditional methods, highlighting the advantages of deep feature learning[2]. Liu Shuangqing (2019) developed an end-to-end LSTM framework using single inertial sensor data, dividing gait into steady-state and transitional-state categories. The model achieved over 96% accuracy in large-scale complex scenarios[3].

In terms of technology classification, Cao Menglin and others (2021) conducted a systematic review, dividing motion intention detection into three categories: skeletal model-driven, traditional statistical learning, and deep neural networks. They pointed out that surface electromyographic signals alone cannot accurately express intentions and emphasized the need to balance accuracy and low latency to meet clinical demands[4]. The latest review, written by Chen Weicong and Lai Changsheng (2025), provides a comprehensive review of the application of surface /electromyography and electroencephalography in rehabilitation robots. It deeply analyzes the challenges in data acquisition, model compression efficiency, and cross-scenario adaptability, pointing out directions for future research[5].

Significant achievements have been made abroad in the fields of exoskeletons and human-machine collaboration. Zhang et al. (2025) evaluated the performance of traditional feature extraction combined with deep learning strategies in sEMG signals, demonstrating that multi-source fusion can significantly improve stability and accuracy under noise[6]. Abidi (2024) combined multimodal data from EEG and EMG to construct an AH-CNN+LSTM model and introduced residual connections, significantly improving the intention prediction accuracy and anti-interference ability for patients with motor disorders[7]. Mínguez et al. (2018) used a Gaussian process dynamic model to predict pedestrian start-stop intentions, showing

prospective advantages with small sample sets. Long et al. (2018) combined sparse Gaussian process regression with fuzzy PID control to achieve precise tracking of exoskeleton devices on complex terrains, although real-time response performance still needs improvement[8].

In the field of human-machine collaboration, Smith et al. (2023) discussed the algorithmic fairness and privacy protection in assistive walking technologies at the macro level, advocating for the establishment of multimodal datasets and enhanced algorithm transparency to reduce system bias[9]. Petković (2019) combined HMM and theory of mind to analyze worker behavior patterns and design dynamic obstacle avoidance strategies, improving the safety and efficiency of warehouse robots, although its reliance on augmented reality devices limits its widespread adoption[10]. Liu et al. (2019) proposed a video-driven CNN+LSTM hybrid architecture, achieving 83% action recognition accuracy in scenarios without wearable devices. Park (2019) integrated offline pretraining and online decision-making with the I-planner, realizing dynamic behavior trajectory modeling and safe path planning[11]. Park (2019) integrated offline pretraining and online decision-making with the I-planner, realizing dynamic behavior trajectory modeling and safe path planning[12]. Wang (2021) designed a multimodal learning architecture for hand interaction tasks, integrating voice commands with sensor data and using extreme learning machines to achieve real-time model updates[13]. Grabhof (2023) used LSTM to process low-quality videos and generate motion sequences, providing solutions for low-cost devices, though the robustness of long-term predictions still requires further validation[14].

Overall, scholars both domestically and internationally have conducted extensive research on the acquisition and modeling of multimodal signals such as bioelectric, visual, and linguistic signals, striving to strike a balance between performance optimization and application value. Future efforts should focus on improving cross-scenario adaptability, model compression and acceleration, as well as generalization ability with small samples, to further advance the development of human-machine collaborative intelligence.

2. Signal sources and acquisition methods Section Headings

The realization of motion intention recognition relies on the accurate acquisition of human movement, physiological, or neural signals. The most frequently employed signal types are summarized in table 1.

Table 1. Comparing the four Signal Types

Signal Type	Typical Sensors/Devices	Acquisition Principle	Advantages	Limitations
sEMG	Adhesive dry/wet electrodes	Electrodes detect potential changes caused by muscle fiber discharge	Good timeliness in reflecting muscle activation intentions; suitable for real-time control	Easily affected by skin resistance
Muscle Pressure	Force-sensitive resistors (FSR)	Monitors small changes in pressure on the skin surface due to muscle expansion	Comfortable to wear; high tolerance to position; relatively low noise	Not sensitive to small movements;
IMU	Three-axis accelerometer + Gyroscope	Captures angular velocity, linear acceleration, and posture angle changes	High accuracy with large movement amplitudes; easy to integrate into wearable devices	Difficult to directly capture intention
EEG	Multi-channel dry/wet electrode helmets	Records oscillations in different frequency bands of the electroencephalogram	Can capture intention signals before movement; rich high-dimensional information	Signals are easily disturbed by noise
fNIRS	Head-worn optical fiber probes	Measures changes in brain oxygen concentration	Slightly better spatial localization than EEG; sensitive to cognitive load	Higher response delay

Among them, sEMG and IMU are currently the most widely used combination. Previous studies (e.g., Niu Miaohu et al. [15]) have achieved high-precision gait state classification by using convolutional neural networks for multi-scale feature extraction of sEMG time-series signals. The combination of muscle pressure signals and IMU (e.g., Du Yancheng et al. [16]) has shown good adaptability and practicality.

3. Deep learning methods in motion intention recognition

Owing to their exceptional capacity for feature extraction and nonlinear representation, deep learning approaches have exhibited outstanding efficacy in decoding movement intentions. The following discussion will focus on three aspects: model types, structural design, and fusion methods. Fig. 1 shows the flow chart of the Deep Learning Methods.

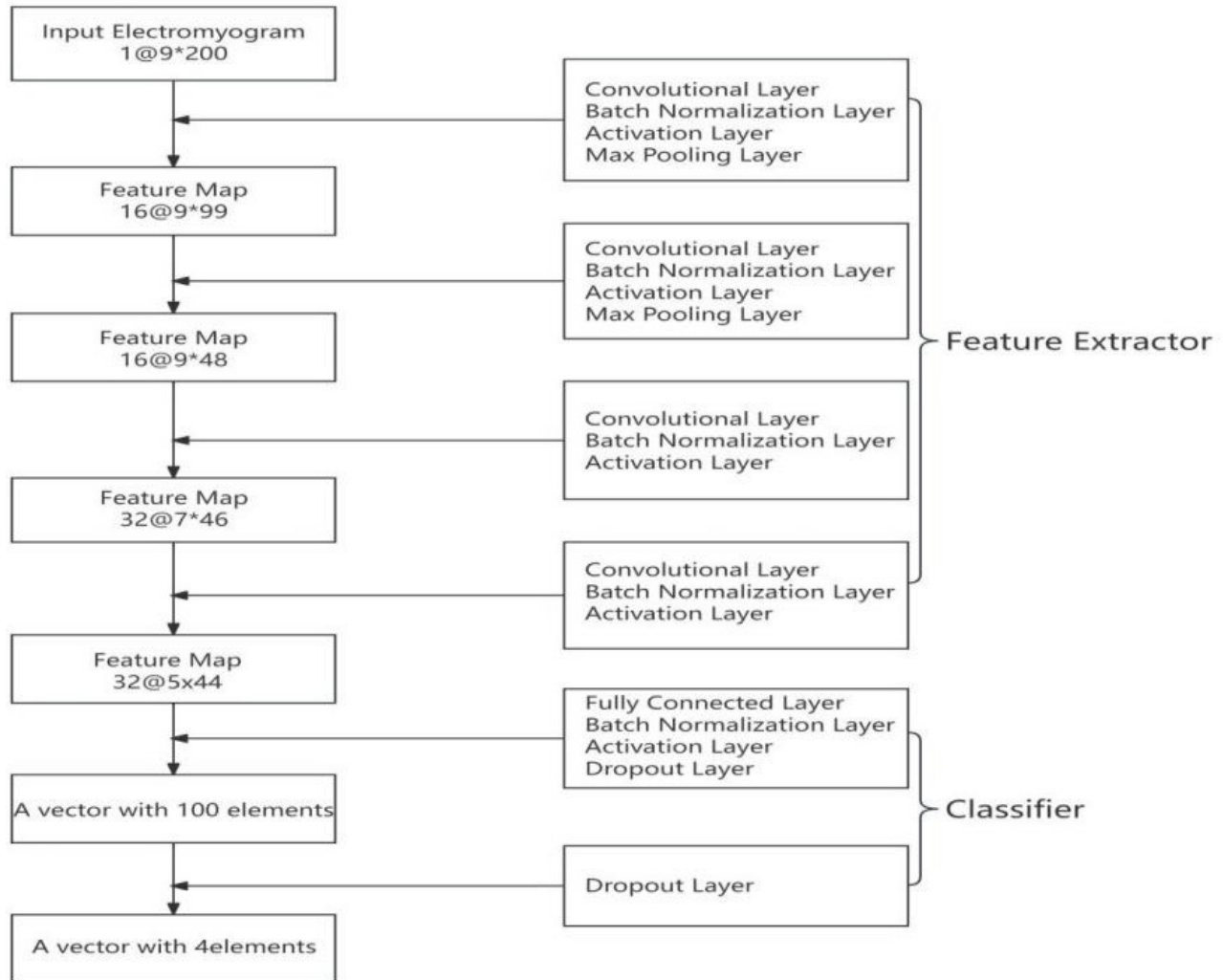


Fig 1. The flow chart of the Deep Learning Methods

3.1 Model Types and Structures

Common deep learning models include:

Convolutional Neural Networks (CNN): CNNs have strong local feature extraction capabilities for 2D or time-series signals, making them suitable for electromyography (EMG) or sensor matrix signals.

Long Short-Term Memory Networks (LSTM): LSTMs are superior for modeling time series, as they can capture the temporal relationships of movements, making them highly effective for motion prediction tasks.

Transformer Architecture: In recent years, transformers have shown superior capabilities in sequence modeling, surpassing LSTM, especially for modeling long-term dependencies. This approach is still in the introduction phase for motion intention recognition.

Multichannel Fusion Networks: These networks are used to fuse multiple heterogeneous signals, such as EMG and

IMU data, in both spatial and temporal dimensions.

For example, in literature [15], A deep learning model utilizing convolutional networks is introduced to perceive motion intentions at multiple scales, as shown in the figure below. This model effectively extracts muscle activity features at different time scales, enhancing the robustness of recognition.

3.2 Signal fusion methods

Deep learning is also widely used in the design of signal fusion strategies, including:

Early Fusion (Data Level): The signals from various sensors are concatenated into a unified input. **Mid Fusion (Feature Level):** Features are extracted from each signal individually and then combined. **Late Fusion (Decision Level):** Each model independently classifies the signals, and then weighted voting or decision-making is per-

formed. Table 2 shows the comparison of common fusion strategies.

Table 2. The comparison of common fusion strategies

Fusion Method	Advantages	Disadvantages
Data-level Fusion	Simple to implement, complete data	Difficulty in coordinating scale differences
Feature-level Fusion	Rich in information, stable performance	Complex network design
Decision-level Fusion	Independent optimization of each model	Significant information loss

Currently, feature-level fusion methods dominate the deepest learning research.

4. Application and case analysis of typical systems and fusion methods

4.1 Detection of movement intent in lower-extremity exoskeletons

As a classic human-robot interaction system, the lower-extremity exoskeleton's effectiveness hinges on precise interpretation of the user's movement intentions. Niu Miaohu et al. [15] constructed a recognition framework based on multi-channel sEMG and convolutional neural networks (CNN), significantly enhancing the stability of gait classification. This system achieves active responses for states such as standing, starting, and walking at a constant speed through high-precision real-time motion recognition, demonstrating good practical deployment value.

4.2 Motion Intention recognition in prosthetics for above-knee amputees

To address the rehabilitation needs of above-knee amputees, Du Yancheng et al. [16] designed a fusion system of muscle pressure signals and IMU, combining traditional feature extraction with a Bayesian classifier. This system successfully distinguished five types of lower-limb movements (such as kicking, stepping, and knee lifting). The method not only reduces the burden on the wearer but also enhances signal robustness, making it suitable for long-term wearable deployment.

4.3 Rehabilitation assistance training platform

Some rehabilitation training platforms have incorporated LSTM models to predict muscle activation patterns and intervene in the assistive mechanism control in advance. For example, a research team[17] embedded an IMU+LSTM network structure into an intelligent gait belt, allowing the system to adjust gait support before the patient is about to take a step. This enables a proactive rehabilitation strategy of "intention prediction rather than response."

4.4 Typical application of data-level fusion

Data-level fusion of sEMG and IMU (raw concatenation) involves directly concatenating the raw data collected by the two sensors within the same time window into a high-dimensional vector, which is then used for subsequent model training or recognition tasks. For example, in gesture recognition, sEMG signals from 8 channels and IMU data from 6 axes (accelerometer + gyroscope) are typically time-aligned, flattened, and concatenated before being input into a neural network for classification. This method effectively combines muscle activity and movement dynamics, significantly improving recognition accuracy.

4.5 Typical applications of feature-level fusion

Typical application a typical multi-channel CNN parallel structure, sEMG and IMU signals are input into separate convolutional neural network channels to extract their respective spatial or temporal features. These features are then concatenated or fused at the feature level and passed through shared fully connected layers for classification. This structure preserves the independent feature representation capability of each signal while enabling deep information fusion. It demonstrates higher robustness and accuracy in tasks such as action recognition and gait analysis, making it particularly suitable for scenarios involving multi-modal signal fusion.

4.6 ons of decision-level fusion

In a typical ensemble learning system, sEMG and IMU each train independent classifiers (such as SVM, KNN, Random Forest, etc.), and the decision outputs of the two models are ultimately fused through methods like majority voting or weighted averaging. This approach offers advantages in modularity and flexibility, allowing full utilization of each sensor model's decision-making ability. It is commonly used in multi-modal action recognition and robust enhancement scenarios, especially suited for handling heterogeneous sensor data and cross-domain generalization problems.

5. Conclusion

Motion intention recognition technology is the core foundation for the development of intelligent exoskeletons, prosthetic control systems, and rehabilitation training platforms. This paper systematically reviews the development trajectories of various signal sources (such as sEMG, IMU, muscle pressure, etc.) and recognition methods, with a particular focus on the application of deep learning models in this field in recent years. From multi-scale feature extraction using convolutional neural networks, to temporal relationship modeling with LSTM, and the exploration of Transformer and signal fusion architectures, all these approaches demonstrate the immense potential of deep learning in improving the accuracy and adaptability of intention recognition.

References

- [1] X. Cao, L. Liu, P. Yang, et al., „Human motion intention recognition using multi-source information and extreme learning machine,“ J. Sensor Technol., vol. 30, no. 8, pp. 1171-1177, 2017.Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.
- [2] Y. Ni and B. Su, “Application of deep learning in motion intention recognition of powered lower-limb prostheses,” Anqing Normal Univ. J. (Nat. Sci. Ed.), vol. 26, no. 4, pp. 80-84, 2020.Fangfang. Research on power load forecasting based on Improved BP neural network. Harbin Institute of Technology, 2011.
- [3] S. Liu, “Motion intention recognition of intelligent lower-limb prostheses based on LSTM deep learning model,” Hefei Univ. J. (Comprehensive Ed.), vol. 36, no. 5, pp. 96-104, 2019.
- [4] M. Cao, Y. Chen, J. Wang, et al., “Research progress on human motion intention recognition based on surface electromyography,” Chin. J. Rehabil. Theory Pract., vol. 27, no. 5, pp. 595-603, 2021.Ma Kunlong. Short term distributed load forecasting method based on big data. Changsha: Hunan University, 2014.
- [5] W. Chen and C. Lai, “Research progress on motion intention recognition based on deep learning of sEMG and EEG,” Sci. Innov. Appl., vol. 15, no. 2, pp. 1-10, 2025.
- [6] Zhang X, Qu Y, Zhang G, et al. Review of sEMG for Exoskeleton Robots: Motion Intention Recognition Techniques and Applications[J]. Sensors, 2025, 25(8): 2448.
- [7] Abidi M H. Multimodal data-based human motion intention prediction using adaptive hybrid deep learning network for movement challenged person[J]. Scientific Reports, 2024, 14(1): 30633.
- [8] Long Y, Du Z, Wang W, et al. Human motion intent learning-based motion assistance control for a wearable exoskeleton[J]. Robotics and Computer-Integrated Manufacturing, 2018, 49: 317-327.
- [9] Smith E M, Graham D, Morgan C, et al. Artificial intelligence and assistive technology: risks, rewards, challenges, and opportunities[J]. Assistive Technology, 2023, 35(5): 375-377.
- [10] Petković T, Puljiz D, Marković I, et al. Human intention estimation based on hidden Markov model motion validation for safe flexible robotized warehouses[J]. Robotics and Computer-Integrated Manufacturing, 2019, 57: 182-196.
- [11] Liu Z, Liu Q, Xu W, et al. Deep learning-based human motion prediction considering context awareness for human-robot collaboration in manufacturing[J]. procedia cirp, 2019, 83: 272-278.
- [12] Park J S, Park C, Manocha D. I-planner: Intention-aware motion planning using learning-based human motion prediction[J]. The International Journal of Robotics Research, 2019, 38(1): 23-39.
- [13] Wang W, Li R, Chen Y, et al. Predicting human intentions in human–robot hand-over tasks through multimodal learning[J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(3): 2339-2353.
- [14] Graßhof S, Bastholm M, Brandt S S. Neural network-based human motion predictor and smoother[J]. SN Computer Science, 2023, 4(6): 760.
- [15] M. Niu and F. Lei, “Motion intention recognition algorithm for lower-limb exoskeletons,” J. Intell. Syst., vol. 20, no. 2, pp. 407-415, 2025.
- [16] D. Du, L. Yang, X. Wang, et al., “Lower-limb motion intention recognition based on muscle pressure signals and inertial measurement units,” Inf. Control, vol. 53, no. 6, pp. 761-773, 2024
- [17] Yuan, Ruyu, et al. “Deep Feature Fusion Model for Lower Limb Intention Recognition Using Multimodal Sensors.” Univ Shanghai Sci & Technol, 2024.