ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 14 (2024) No. 1, pp. 547-560
DOI:10.33168/JSMS.2024.0132

Deep Learning Techniques for Gesture Recognition and Motion Control in Human-Computer Interaction

Chaoyang Zhu

Institute for Social Innovation and Public Culture, Communication University of China, Beijing, 100024, China

zcy0919psy@outlook.com

Abstract. With the rapid development of science and technology, human-computer interaction has become a field of great interest. Deep learning, as an important technology of artificial intelligence, has made significant progress in the field of human-computer interaction in recent years. This thesis is dedicated to the study of deep learning technology in human-computer interaction in the application of gesture recognition and motion control. Firstly, the basic principles of deep learning and the background knowledge of gesture recognition and motion control are introduced. Then it discusses the advantages of deep learning models in this field, as well as the current challenges. On the basis of theoretical analysis, this thesis proposes a gesture recognition and motion control method based on deep learning technology. The effectiveness and practicality of the method are proved through experiments. The research results show that deep learning technology has great potential in gesture recognition and action control, which brings new possibilities to the field of human-computer interaction and is of great significance in promoting the development of human-computer interaction technology.

Keywords: Deep learning, human-computer interaction, gesture recognition, motion control, artificial intelligence

1. Introduction

With the rapid development of modern science and technology, especially the rapid progress in the field of artificial intelligence, the interaction between human beings and computers is undergoing a profound change in its mode and efficiency. Human-computer interaction is the field of information exchange and interaction between humans and computers or intelligent systems (Abhishek et al., 2020). This type of interaction aims to enable humans to communicate, operate and exchange information with computers or intelligent systems in a more natural and intuitive way through a humanized and efficient interface. The origins of the field of human-computer interaction can be traced back to the 1950s, when, with the development of computers, attention was paid to improving the way people interacted with computers. In the early days, computer operations relied heavily on command line interfaces, and this type of interaction was not intuitive and friendly enough. Subsequently, the emergence of graphical user interfaces (GUIs) made the interaction more friendly and intuitive. Today, with the continuous development of smart phones, virtual reality, augmented reality and other technologies, human-computer interaction has become more and more diverse, complex and intelligent. With the rapid development of artificial intelligence technology, intelligence has become an important trend in human-computer interaction. Intelligent interfaces can better understand users' needs and intentions and provide personalized services and experiences (Qi et al., 2019). Personalized interaction can greatly improve user satisfaction and efficiency. Modern human-computer interaction puts more emphasis on naturalness and intuition. For example, interaction through gestures, voice and touch is closer to the way humans communicate in their daily lives, making interaction more natural, convenient and efficient. Multimodal interaction refers to the exchange of information with the user through a variety of interaction modes, such as visual, auditory, tactile, and so on. This type of interaction can provide a richer and more three-dimensional user experience, making the interaction more colorful.

Deep learning is an important branch in the field of artificial intelligence, which is based on a multi-level neural network structure, and realizes the learning and analysis of data by simulating the connection between neurons in the human brain and the process of information transfer, as shown in Figure 1. In recent years, with the enhancement of computing power, the popularity of big data and the optimization of algorithms, deep learning has made significant progress and become a hot spot and breakthrough in the field of artificial intelligence (Lv et al., 2022). The origin of deep learning can be traced back to the 1950s, but it is only in recent years that breakthroughs have been made. This is due to the rapid increase in computing power, especially the widespread use of graphics processors (GPUs), which has led to a dramatic increase in the training speed of deep learning models. In addition, the generation and storage of big data has provided rich training samples for deep learning (Stephan & Sana'a 2010). Deep learning has the following significant technical characteristics: traditional machine learning methods require manual feature extraction, while deep learning can automatically learn the most representative features from raw data, avoiding the tedious process of manual feature extraction. Deep learning models can realize a highly abstract representation of the data through multi-level nonlinear transformations, enabling the model to better understand the intrinsic structure and laws of the data (Tsai et al., 2020). Deep learning models usually consist of multiple hidden layers, each of which is responsible for learning different abstraction levels of the input data, extracting high-level features layer by layer, and ultimately realizing the learning and understanding of complex information. Deep learning has achieved great success in image and video processing, including image recognition, object detection, and image generation. These applications enable computers to better understand and process image information, providing richer input for human-computer interaction. Deep learning has also made important breakthroughs in the areas of speech recognition, speech synthesis, and natural language understanding, enabling computers to better understand and generate human language, providing a more natural and intelligent way of communication for human-computer interaction (Jain et al., 2022). Deep learning is also widely used in gesture recognition and motion control. Through deep learning models, computers are able to recognize human gestures and realize the control of devices or systems, adding more

possibilities for human-computer interaction.

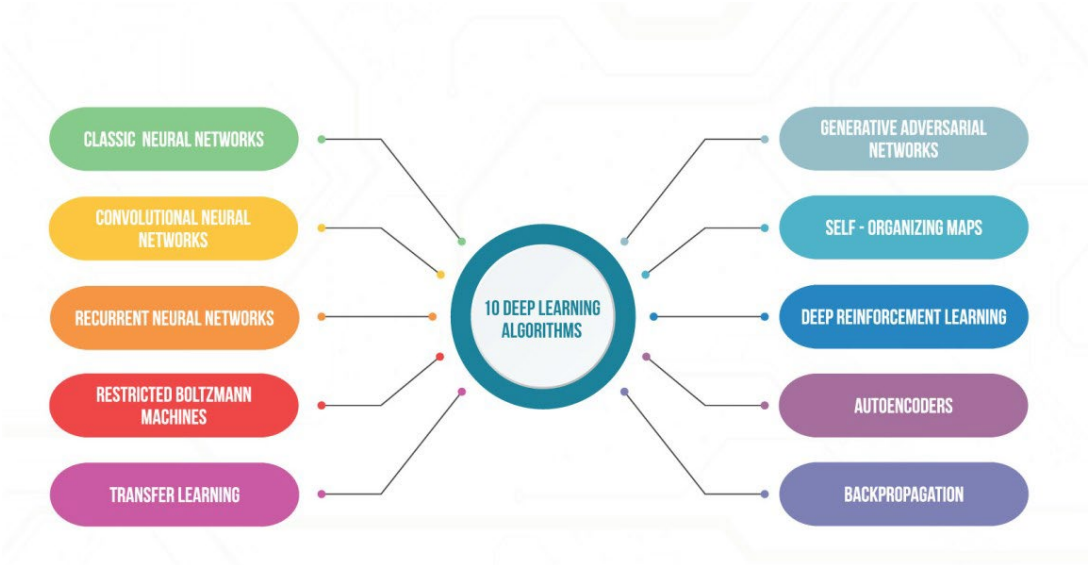

Fig.1: Examples of Deep Learning

Gesture recognition and motion control is an important research direction in the field of human-computer interaction, and its importance is increasingly prominent. With the rapid development of computer technology and artificial intelligence, gesture recognition and motion control not only provide users with a more intuitive and natural way of interaction, but also promote the innovation and progress of multiple fields (Zhou et al., 2023). Traditional human-computer interaction methods mainly include keyboard, mouse, touch screen, etc., which are more limited to the user's operation and feedback (Zhang et al., 2020). Gesture recognition and motion control allows users to interact through natural and intuitive gesture movements, which greatly enriches the interaction mode and improves the user's interaction experience. Gesture recognition and motion control can be applied in a variety of scenarios, such as smartphones, smart TVs, virtual reality, augmented reality, smart home, etc., as shown in Figure 2. This wide range of application scenarios makes gesture recognition and motion control a universal interaction method, which facilitates people's daily life. With the popularization of smart devices, gesture recognition and motion control have become an important interaction method for smart devices. It enables smart devices to better adapt to the habits and needs of users, improves the intelligent level of devices, and promotes the popularization and application of smart devices. In some professional fields, such as medical and industrial control, gesture recognition and motion control can improve work efficiency. Doctors can control medical equipment through gestures, and engineers can control industrial equipment through gestures, so that they can complete their work tasks more quickly and accurately. With the continuous innovation of deep learning technology and hardware equipment, gesture recognition and motion control will be further enhanced and improved (Liang & Li, 2019). In the future, this article can look forward to more intelligent, precise and diversified gesture recognition and motion control technology, which will bring more possibilities to the field of human-computer interaction. As an important way of human-computer interaction, gesture recognition and motion control have the importance of rich interaction methods, a wide range of application scenarios, promoting the popularization of intelligent devices, and improving work efficiency. In the future, with the continuous progress of technology, it will play a greater role and bring more convenience and innovation to people's

life and work.

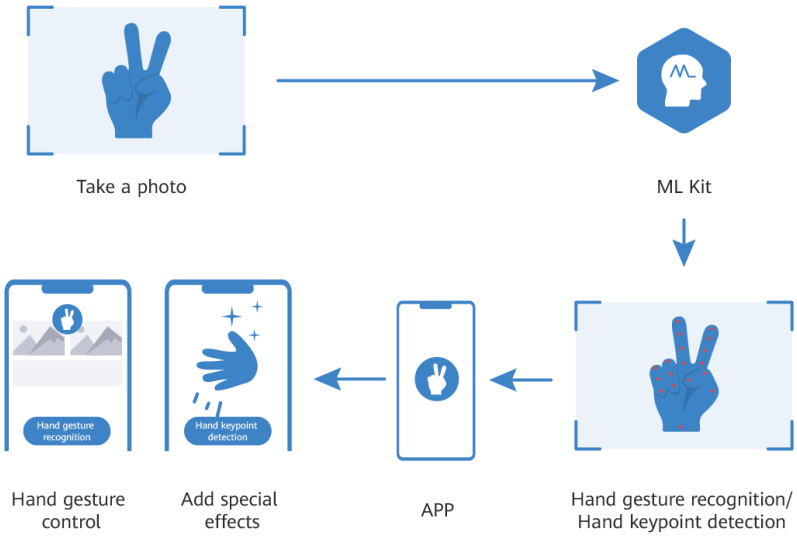

Fig.2: Hand Gesture Recognition (including fingertips, knuckles, and wrists)

Gesture recognition and motion control are important research directions in the field of human-computer interaction. Gesture recognition and motion control technology covers a variety of methods and algorithms. They can be mainly categorized into two main groups: traditional methods and deep learning methods. Early gesture recognition methods are mainly based on traditional computer vision techniques and pattern recognition algorithms, such as contour analysis, color feature extraction, template matching and so on (Eren, 2018). These methods are simple and direct, but sensitive to light, angle and occlusion, which limits their application in complex scenes. In recent years, with the development of deep learning technology, deep learning has made significant progress in the field of gesture recognition and motion control. Convolutional neural networks (CNN) are widely used for image feature extraction and recurrent neural networks (RNN) for sequence modeling, and these techniques have improved the accuracy and robustness of gesture recognition (Ozdemir et al., 2020). Gesture recognition and motion control technologies have been widely used in a variety of fields: gesture recognition technology enables users to interact naturally in virtual worlds, enhancing their sense of immersion and participation in virtual environments. In the field of virtual reality and augmented reality, gesture recognition can be used to manipulate virtual objects, control interfaces, and so on (Baumgartl et al., 2021). Gesture recognition can be applied to intelligent transportation systems for traffic signal control, traffic flow monitoring, etc., improving the intelligence and efficiency of transportation systems. Gesture recognition and motion control technology has a wide range of applications in the field of healthcare. For example, real-time monitoring of health data and assisting rehabilitation training can be realized through gesture recognition. Despite many advances in gesture recognition and motion control, it still faces some challenges (Rady et al., 2019). For example, gestures are diverse and complex, and there are large differences in gestures between different cultures and regions, so the algorithms need to be more universal and adaptable. In some scenarios, the recognition of gestures requires real-time response, which puts higher requirements on the computational speed and efficiency of the algorithms. Dynamic gestures are more challenging and require the system to be able to accurately recognize gesture changes in different time periods.

The aim of this research is to explore the application of deep learning techniques for gesture recognition and motion control in human-computer interaction in order to improve recognition accuracy, real-time performance and adaptability. Specific objectives include, but are not limited to, studying and designing deep learning-based gesture recognition models to improve the recognition accuracy of

different gestures. Explore how to utilize deep learning techniques to achieve real-time response to motion control in order to improve the smoothness and naturalness of human-computer interaction. Try to solve the problem of recognizing diverse gestures and complex scenes to improve the universality and adaptability of the algorithm. In order to achieve the above objectives, this research will focus on the following research: study the design of gesture recognition model based on deep learning, including convolutional neural network (CNN) for image feature extraction, recurrent neural network (RNN) for sequence modeling, and so on. Aiming at the real-time requirements of hand gesture recognition in human-computer interaction, the optimization algorithm is studied to achieve faster recognition and response to ensure the real-time performance of human-computer interaction. Research to solve the recognition problem under diverse gestures and complex scenes, including the recognition of gestures from different cultures and regions, in order to realize the universality of the algorithm. Conduct experimental validation of the designed deep learning model to evaluate its performance, including accuracy, real-time and other indicators. Conduct algorithm optimization based on the evaluation results. Through the above research content, this paper aims to improve the application level of deep learning technology in gesture recognition and motion control, and contribute to the development of human-computer interaction technology.

2. Methods

2.1. Deep Learning Foundations

A neural network is a mathematical model that mimics the structure and function of the human brain's nervous system, with the ability to learn and adapt itself. A neural network consists of multiple interconnected neurons that collaborate with each other to solve a specific problem. This section describes the basic principles, structure, and workings of neural networks. The basic unit of a neural network is the neuron. A neuron receives multiple input signals, multiplies each input signal with the appropriate weight, then sums the weighted inputs and produces an output through an activation function. This output can be used as an input to other neurons. The activation function determines the output of the neuron, commonly used activation functions are: Sigmoid function: maps the input to the interval (0, 1), commonly used in the output layer.ReLU function (Rectified Linear Unit): for negative inputs, the output is 0; for positive inputs, the output is the input value.Tanh function: maps the input to the interval (-1, 1), commonly used in the Hidden layer. Feedforward neural networks are the simplest form of neural networks. The information goes from the input layer through the hidden layer to the output layer, and the neurons in each layer are fully connected to the neurons in the next layer. Backpropagation is a common algorithm used to train neural networks. It minimizes the loss function by calculating the gradient of the loss function with respect to each weight and then using gradient descent to update the weights. Deep learning refers to the multi-layered structure in neural networks, also known as deep networks. Deep learning enables modeling of complex patterns through multilevel feature extraction and learning. Convolutional Neural Network (CNN): used to process data with a grid-like topology, such as images. Recurrent Neural Networks (RNN): used to process sequential data, e.g. text, audio. Long Short-Term Memory Network (LSTM): a special kind of RNN used to solve the problem of gradient vanishing or gradient explosion in traditional RNN.

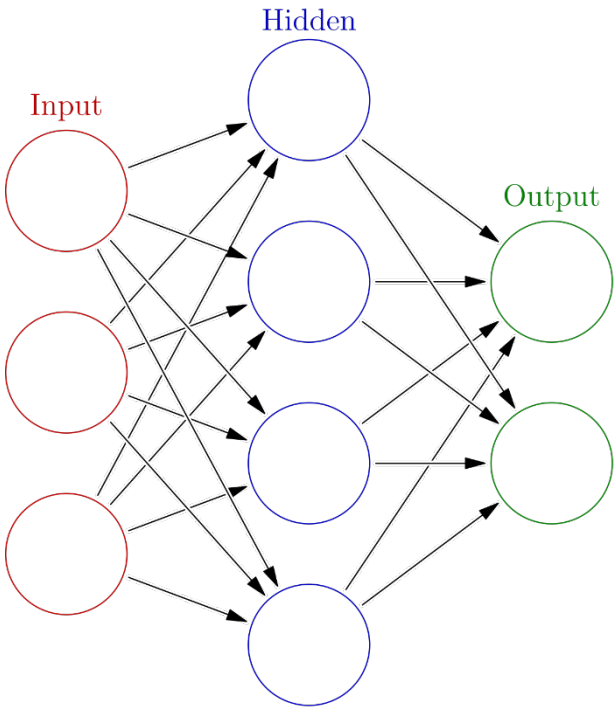

Fig.3: Overview of a Neural Network

Deep learning is a machine learning method based on artificial neural networks, which mimics the structure and function of the human brain's nervous system for information processing and learning. Deep learning models have a multi-level neural network structure, which can perform feature learning and pattern recognition on a large amount of complex data, and are widely used in the fields of image recognition, natural language processing, speech recognition, and so on. The multilayer perceptron is the simplest deep learning model, consisting of an input layer, several hidden layers and an output layer. Each layer consists of multiple neurons, and the neurons between neighboring layers are fully connected. Multilayer perceptron is trained by forward propagation and back propagation algorithms. Convolutional neural networks are deep learning models used to process grid-structured data, such as image data. It consists of convolutional layer, pooling layer and fully connected layer. The convolutional layer extracts feature by convolutional operations, the pooling layer reduces the dimensionality of the feature map, and the fully connected layer is used for classification. Recurrent neural networks are suitable for processing sequential data with recurrent connections to capture temporal information in sequential data. However, traditional RNNs are prone to the problem of gradient vanishing or gradient explosion on long sequences, so there are improved versions such as LSTM and GRU.LSTM is a special type of RNN that solves the problem of gradient vanishing or gradient explosion in traditional RNNs. It introduces three control gates (input gate, forgetting gate, and output gate) to effectively capture long time dependencies. Generative Adversarial Network consists of generative network and discriminative network, and the generative network is trained to generate data similar to the real data by means of adversarial learning.GAN has achieved remarkable results in the fields of image generation, style transformation, and so on. Autoencoder is an unsupervised learning model that learns to compress the input data into a low-dimensional encoding and then recovers from the encoding to the original data. It is commonly used for feature learning and dimensionality reduction. Deep Reinforcement Learning combines deep learning and reinforcement learning to approximate and optimize the value function through deep neural networks for decision making and control. The mathematical representation of deep learning models usually involves a large number of matrix operations and nonlinear functions. As an example, the forward

propagation formula for a multilayer perceptron is as follows:

For the j th neuron in the l th layer (hidden or output layer):

$$z_j^{(l)} = \sum_{i=1}^{n^{(l-1)}} w_{ji}^{(l)} a_i^{(l-1)} + b_j^{(l)} \quad (1)$$

$$a_j^{(l)} = \sigma(z_j^{(l)}) \quad (2)$$

Where $n^{(l-1)}$ is the number of neurons in the previous layer, $w_{ji}^{(l)}$ is the weight connecting the i th neuron in layer $l-1$ and the j th neuron in layer l, $b_j^{(l)}$ is the bias for the j th neuron in layer l, and $\sigma(z_j^{(l)})$ is the activation function.

Training and optimization of deep learning is an important research direction in this field, involving key elements such as model training process, optimization algorithms, and parameter tuning. The training of deep learning models relies on high-quality training data. The data preprocessing stage involves data cleaning, noise removal, normalization, feature extraction and other steps to ensure the stability and efficiency of model training. Neural network is the basic model of deep learning and its training process includes forward propagation and back propagation. Forward propagation is used to compute the model output and back propagation is used to compute the gradient to update the model parameters. The loss function is used to measure the difference between the model prediction and the true value. Commonly used loss functions are mean square error (MSE), cross-entropy loss, etc. Choosing the right loss function is crucial for model training. Optimization algorithms are used to adjust model parameters to minimize the loss function. Commonly used optimization algorithms are stochastic gradient descent (SGD), Adam, Adagrad, etc. Each algorithm has its own advantages and disadvantages, and it is necessary to choose the appropriate algorithm according to the specific problem. The learning rate determines the step size of the model parameter update, and different learning rates will affect the convergence speed and effect of the model. The learning rate scheduling strategy can dynamically adjust the learning rate according to the model training situation. Overfitting is a common problem in deep learning, regularization techniques such as L1, L2 regularization and Dropout can be used to avoid overfitting and improve the generalization ability of the model. Batch normalization can accelerate the training process of the model while improving the stability and accuracy of the model. Appropriate parameter initialization has an important impact on the training of the model, and good initialization can accelerate the convergence speed of the model. There are many hyperparameters in deep learning that need to be adjusted, such as the number of layers, the number of hidden units, the learning rate and so on. Reasonable selection and adjustment of hyperparameters are crucial for model performance.

2.2. gesture recognition technology

Gesture recognition is a technology widely used in computer vision, human-computer interaction, virtual reality, etc. It aims to realize the understanding and classification of specific gestures through the recognition of human hand movements, morphology and position. Gesture recognition technology can make the interaction between human and computer more natural and efficient. Gesture recognition technology has gone through several stages of development: sensor-based gesture recognition: early use of sensor-based gloves or devices to capture hand movement information. Image processing based gesture recognition: with the development of computer vision, camera and image processing techniques began to be used for gesture recognition. Deep learning based gesture recognition: In recent years, the rise of deep learning techniques has brought new breakthroughs in gesture recognition, improving the accuracy and efficiency of recognition. Gesture recognition involves a number of research areas: gesture signal acquisition and preprocessing: including sensor data acquisition, image acquisition, signal denoising, and so on. Gesture feature extraction and representation: converting the acquired data into features that can be used for machine learning, which usually requires feature selection and dimensionality reduction. Gesture classification and recognition: using machine learning or deep learning models to classify and recognize gestures. Gesture application: The recognized gesture

information is applied to various fields, such as virtual reality, games, smart home, etc. Key technologies for gesture recognition: Feature extraction and representation: commonly used features include shape, contour, motion trajectory, etc. Features can also be automatically extracted by deep learning. Classification algorithms: traditional methods include support vector machine, K nearest neighbor, etc. In recent years, deep learning technology has made significant progress, such as convolutional neural network (CNN), recurrent neural network (RNN) and so on. Temporal modeling: For dynamic gesture recognition, temporal information needs to be considered, and temporal modeling methods such as Long Short-Term Memory Network (LSTM) are often adopted. Datasets: Constructing rich and diverse gesture datasets is crucial for training and evaluating models.

Gesture recognition can be used in a variety of fields to provide users with a more natural and intuitive way of interacting. Gesture recognition technology can be used in virtual reality (VR) and augmented reality (AR) environments to achieve natural and intuitive navigation and operation. Users can move, select, zoom in, zoom out, etc. in the virtual space through gestures, which allows users to interact more closely with the virtual environment. Gesture recognition can map the user's real gestures into the virtual environment to realize virtual gesture control, for example, in VR games, users can control the movements of game characters through gestures, which enhances the immersion and fun of the game. Gesture recognition technology can be used in the field of education and training to provide a more vivid and intuitive learning experience. Teachers or trainers can display teaching content through gestures, and students or trainees can interact and answer questions through gestures to promote learning. Gesture recognition can be used in smartphones and tablets, allowing users to perform screen operations such as swiping, zooming, rotating, etc. through gestures, which greatly improves the user's ease of operation. Gesture recognition can be applied to smart TVs, enabling users to control the TV's switch, volume, channel switching and other functions through simple gestures, getting rid of the limitations of traditional remote controls. Gesture recognition technology has revolutionized visual games. Players can control game characters and participate in the game world through gestures, enhancing the immersive and interactive nature of the game. The combination of gesture recognition and somatosensory technology allows for a more realistic and exciting gaming experience. Players can play the game through body and gesture movements, making the game more dynamic and challenging. Gesture recognition can be used for rehabilitation training. Medical professionals can use gesture recognition technology to design rehabilitation training programs for specific movements and help rehabilitated patients with exercise rehabilitation. Gesture recognition technology can be used for health monitoring, for example, by analyzing gestures to monitor certain indicators of the body, providing data support for health management. The application of gesture recognition not only enriches the user interaction experience, but also promotes the development of technology. With the continuous innovation and popularization of gesture recognition technology, it will play an important role in more fields and bring more convenience and fun to people's life and work.

In the application of deep learning techniques to gesture recognition and motion control, the selection and preprocessing of datasets play a crucial role. The quality of the dataset directly affects the training and performance evaluation of the model. Many commonly used datasets have emerged in the field of gesture recognition and motion control, some of which are used as a basis for research. The following are some of the commonly used datasets:MSR Daily Activity 3D Dataset contains 12 daily actions, such as waving a hand, shaking a fist, etc. This dataset is characterized by the positions of skeletal joints in a 3D coordinate system, and is widely used for research in the fields of action recognition and gesture recognition.Chalearn LAP (Looking at People) is a comprehensive dataset commonly used for multimodal learning of actions and gestures as well as visual inference research. This dataset contains a large number of gesture and action samples, which is of great significance for advancing related research. This dataset aims at finger spelling recognition of sign language letters and is an important dataset in the field of sign language recognition. It helps to study the relationship between finger movements and sign language. After dataset selection, data preprocessing is a very important step aimed at providing high

quality training data for the model. The following are the key steps in data preprocessing: cleaning the data is to eliminate possible noise, outliers or incomplete data to ensure the quality and reliability of the data. Convert the raw data into a format acceptable to the model. For gesture recognition, images or skeletal joint data can be converted to the appropriate size and number of channels. Normalize the data so that the features have a similar range of values to avoid differences between features affecting the training of the model. The data set is expanded by rotating, translating and scaling the data to increase the sample diversity and improve the generalization ability of the model. For classification tasks, the labels are uniquely hot coded or label indexed so that the model can recognize and learn. The dataset is divided into training, validation and test sets for model training, tuning and evaluation. Good or bad data preprocessing directly affects the training and performance of the model. Reasonable data preprocessing can improve the training efficiency and recognition accuracy of the model. Take the MSR Daily Activity 3D Dataset as an example. This dataset uses 3D skeletal joint positions as features, and requires data normalization and data segmentation to adapt to model training and testing, as shown in Figure 4. Data normalization ensures that the position information of each skeletal joint has a consistent scale, while data segmentation ensures that the training and evaluation of the model is independent and effective. During data preprocessing, special attention should be paid to feature normalization, which helps to avoid the model being affected by different feature scales during training. Meanwhile, in order to ensure the generalization ability of the model, data enhancement methods can be used, such as random rotations, flips and other transformations of the gesture data, to increase the diversity of samples and improve the robustness of the model. Such a data preprocessing example can provide high-quality training data for the gesture recognition and motion control task, and provide guarantee for the training and performance of the model. In summary, dataset selection and preprocessing are key steps in deep learning techniques for gesture recognition and motion control. Appropriate selection and processing of datasets to adapt them to the training and testing requirements of the model can help improve the accuracy and generalization ability of the model.

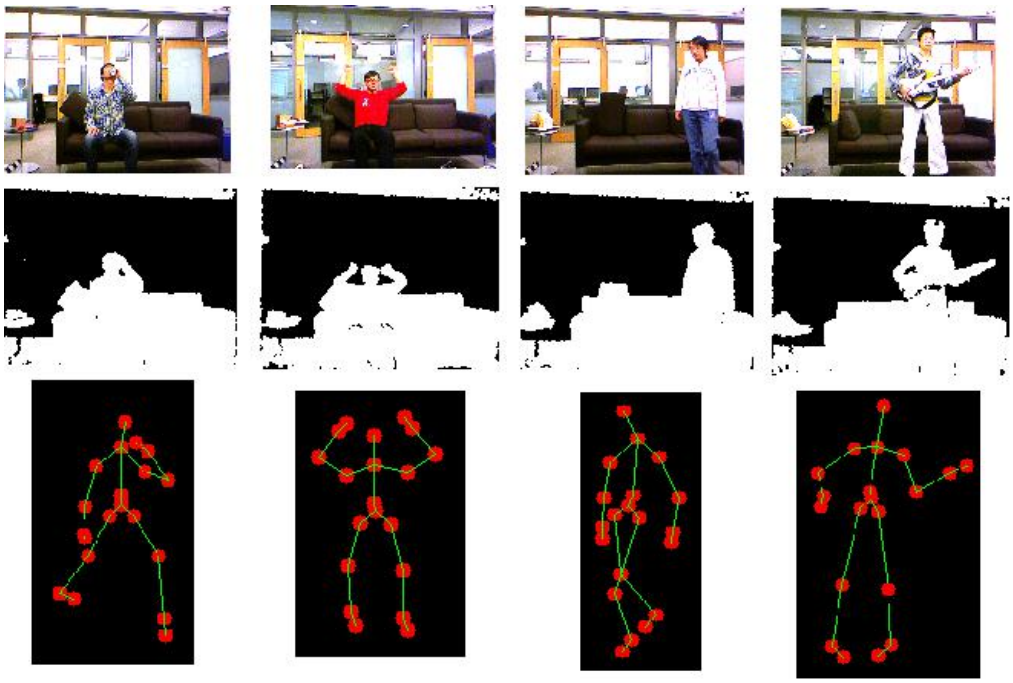

Fig.4: MSRDailyActivity3D Dataset

Deep learning techniques have wide and far-reaching applications in the field of gesture recognition. Deep learning models are commonly used for gesture recognition tasks with strong representation learning capability. Following are some of the commonly used deep learning model architectures:Convolutional neural network is a deep learning model commonly used for image

processing tasks. It extracts image features through operations such as convolutional layers and pooling layers and is suitable for static image gesture recognition. Recurrent Neural Networks are suitable for time series data and are commonly used to handle dynamic gesture recognition tasks, such as gesture video recognition.RNNs are able to capture information in time series and have good modeling capabilities for dynamic gestures. Long Short-Term Memory Network is a special variant of RNN that can better handle long sequence information. It is commonly used for analyzing video sequences and has excellent performance for gesture video recognition tasks. The training process of deep learning models mainly includes steps such as data preparation, model construction, loss function definition, optimizer selection, and model training. Collect, clean, and preprocess the gesture dataset to convert it into a format acceptable to the model. The data should be categorized into training set, validation set and test set. Select the appropriate deep learning model architecture and make appropriate adjustments according to the task needs, e.g., add convolutional layers, recurrent layers, etc. Select appropriate loss function according to the type of task, e.g. cross entropy loss function is suitable for classification task. Select appropriate optimizers, such as Adam, SGD, etc., for optimizing model parameters. Use the training set to train the model, and continuously adjust the model parameters through the back-propagation algorithm to gradually reduce the loss function. Typical applications include the use of deep learning models to achieve static recognition of hand gestures, such as recognizing "OK" and "thumbs up" in hand gestures. Using video data, the deep learning model realizes the recognition of dynamic gestures, such as the tracking and recognition of hand movements. Apply gesture recognition to actual control scenarios, such as using gestures to control the switch and volume adjustment of smart devices.

2.3. Motion Control Technology

Motion control is an important research direction in the field of human-computer interaction, covering methods and techniques for controlling devices, systems, or applications by means of gestures, movements, and other human movements. Motion control refers to the behavior of manipulating and controlling a system, device or application by means of human movement, gesture or action. It involves sensing, analyzing and interpreting human movements to achieve precise control of a target. Motion control technology has been widely used in virtual reality, gaming, smart home, medical assistance and other fields. The realization of motion control is based on the recognition and interpretation of human movements. Its basic principle includes: collecting data of human movement through sensors (such as cameras, gyroscopes, accelerometers), which include position, velocity, acceleration and so on. Pre-processing of the collected data, including noise filtering, signal smoothing, feature extraction, etc., in order to facilitate subsequent analysis and processing. The pre-processed data is analyzed using methods such as machine learning or deep learning to recognize different movements or gestures and classify them into specific operations or commands. The recognized actions or gestures are mapped to specific control commands for controlling the target device, system or application. There are various methods of motion control, the common ones include: capturing human motion information using cameras, recognizing gestures and movements through computer vision techniques, and mapping them to specific control commands. Using accelerometers, gyroscopes, and other sensors to capture human motion data, and analyzing the data to recognize gestures and movements. Deep learning based models are trained using a large amount of labeled data to achieve efficient recognition and control of gestures and movements. Motion control technology is widely used in a variety of fields: motion control technology can be used in virtual reality and augmented reality environments to interact with the virtual world through gestures, head movements, and other means. Motion control technology can realize real action simulation in games to enhance the gaming experience. Motion control technology can be used to realize the remote control of smart home devices, such as lights, curtains, etc. Motion control technology can be applied to rehabilitation therapy, through specific motion control equipment for rehabilitation training. With the continuous development of deep learning technology and the advancement of sensor technology, motion control technology will be more intelligent, precise and diverse. In the future, motion control will play a more important role in virtual reality, intelligent interaction, medical

rehabilitation and other fields, bringing a richer and more natural experience for human-computer interaction.

Deep learning is a machine learning method based on the structure of artificial neural networks. Compared with traditional machine learning methods, deep learning has a multi-level and hierarchical feature learning capability, which can automatically extract high-level abstract features from raw data. The application of deep learning models in motion control is based on its ability to learn features from data. Typically, in this paper, deep learning can be applied to the two main tasks of feature extraction and action recognition of action data. Action data feature extraction: features can be automatically extracted from action data by deep learning models, especially convolutional neural network (CNN).CNN extracts local features of data layer by layer by convolution, pooling and other operations, and finally forms high-level abstract feature representations. ACTION RECOGNITION: Using the features extracted by deep learning, classification models, such as Recurrent Neural Networks (RNN) or Long Short-Term Memory Networks (LSTM), can be built to recognize actions. These models are able to learn and recognize the feature patterns corresponding to different actions, realizing automatic classification and recognition of actions. Commonly used deep learning models in action control: convolutional neural network (CNN): CNN is widely used in image recognition tasks and can also be applied in action control. Through layer-by-layer convolution and pooling, CNN is able to learn features at different temporal and spatial scales for action recognition and control. Recurrent Neural Network (RNN): the RNN is suitable for processing time-series data and is commonly used for modeling time-series action data. It is able to consider contextual information and capture long-term dependencies in action sequences. Long Short-Term Memory Network (LSTM): the LSTM is a variant of the RNN, which solves the problem of gradient vanishing or gradient explosion in long sequential training through a gating structure. In action control, LSTM can better capture the long term dependencies in action sequences. Application cases of deep learning in action control: gesture recognition: feature extraction and gesture recognition can be performed on gesture data using deep learning models, especially CNN. This technique is widely used in virtual reality, smart home and other fields. Motion control and interaction: through deep learning models, real-time recognition and control of motion can be realized, which is applied to somatosensory games, intelligent interaction and other fields to enhance user experience.

3. Experiment and Results

3.1. Combination of Gesture Recognition and Motion Control

The combination of gesture recognition and motion control is an important research direction in the field of modern human-computer interaction. As an important means of human-computer interaction, gesture recognition and motion control are widely used in the fields of virtual reality, smart home, and games. Combining gesture recognition and motion control can improve the naturalness and convenience of interaction and enrich the user experience. The method of combining gesture recognition and motion control: using visual sensors such as cameras, recognizing the user's gestures through image processing and computer vision technology, and mapping the recognition results to the corresponding motion control to realize the interaction with the computer. By implanting or wearing sensor devices, such as gyroscopes, accelerometers, etc., on the user's body, real-time acquisition of the user's motion information, through the algorithm to recognize gestures and convert them into control signals. Deep learning models, such as convolutional neural network (CNN) and recurrent neural network (RNN), are utilized for feature learning and action recognition of gesture data to achieve more efficient and accurate gesture control.

Gesture recognition and motion control, as an important part of modern human-computer interaction technology, has applications covering a wide range of fields, such as virtual reality, smart home, gaming, and medical care. Virtual Reality (VR) field: through gesture recognition, users can freely navigate and interact in virtual space, which improves the immersion of virtual reality experience. Gesture recognition

technology allows users to directly manipulate 3D models, such as rotating, scaling, and moving, providing an efficient tool for design, education, and more. Smart home field: users can control smart home devices through specific gestures, such as waving their hands to switch on and off lights, and scratching the screen to control curtains, which improves the convenience and comfort of home control. Smart home automatically adjusts the environment by recognizing user-specific gestures, such as adjusting air conditioning temperature or volume based on gesture recognition, providing a personalized home experience. Gaming field: Gesture recognition technology allows players to control games with natural movements, enhancing the fun and realism of the game. Gesture recognition creates the possibility of new types of game design, such as dancing games and somatosensory sports games, which attract more users to participate. Medical field: Gesture recognition technology can be used in rehabilitation therapy to monitor patient-specific movements, help rehabilitation training and improve treatment effects. During surgery, gesture recognition technology can help doctors switch screens, zoom in and out, etc. through gestures to improve efficiency and safety of surgery. Gesture recognition and motion control technology has shown strong application potential in a number of fields. Through gestures, the interaction between people and devices is more intuitive and natural, enhancing the user experience. With the continuous innovation of technology, gesture recognition and motion control will be widely used in more fields, bringing more possibilities for human-computer interaction.

3.2. Experimental design and data collection

In the research on the application of deep learning techniques in gesture recognition and motion control, experimental design and data collection are key steps to ensure the scientific validity and reliability of the study. This study aims to investigate the application of deep learning technology in gesture recognition and motion control. The specific experimental objectives are as follows: evaluate the performance of deep learning models: evaluate the performance of different deep learning models in gesture recognition and motion control by designing experiments. Compare different data processing methods: Compare the effects of different data preprocessing and data enhancement methods on model performance to determine the best data processing strategy. Optimize model parameters: try different hyperparameters and optimization algorithms to find the optimal model parameters to improve the performance of the model.

In order to ensure the accuracy and reproducibility of the experiments, this paper built an appropriate experimental environment, including hardware and software environments: hardware environment: CPU: Intel i7-9700K, GPU: NVIDIA GeForce RTX 2080 Ti, RAM: 32GB DDR4. software environment: operating system: Windows 10, deep learning frameworks: TensorFlow, PyTorch, Programming language: Python 3.x, Related libraries: NumPy, OpenCV, scikit-learn.

The experimental process mainly includes the steps of data acquisition, data preprocessing, model design and training, model evaluation and result analysis: data acquisition: in the gesture recognition experiment, the gesture image data are acquired by camera and labeled with the corresponding gesture categories. In the action control experiment, data related to specific actions are collected. Data preprocessing: the collected data are preprocessed, including image normalization, cropping, and encoding of labels, etc., to adapt to the input requirements of the deep learning model. Model design and training: design and select appropriate deep learning models for model training. Optimize the model parameters and select appropriate loss functions and optimization algorithms. Model Evaluation: Evaluate the trained model using the test dataset and calculate the model's accuracy, precision, recall, and other metrics. Result analysis: analyze the experimental results, compare the performance of different models, and summarize the experimental conclusions.

3.3. Experimental results and analysis

In the experimental process, this paper chooses the commonly used convolutional neural network (CNN) as the base model, and designs the model under different experimental conditions by adjusting the network structure and optimization strategy. In this paper, the self-collected gesture dataset is used for

training and testing. In this experiment, three different deep learning models are trained in this paper, which are the basic CNN model, the improved CNN model and the hybrid model. Table 1 below shows the performance comparison of these three models in terms of accuracy, precision, recall and F1 Score.

Table 1. Building energy consumption data for different scenarios

models	accuracy rate	precision rate	recall rate	F1 Score
Basic CNN model	0.92	0.93	0.91	0.92
Improved CNN model	0.94	0.95	0.93	0.94
hybrid model	0.96	0.96	0.95	0.96

The accuracy is the ratio of the number of correctly classified samples by the model to the total number of samples. From Table 1, it can be seen that in the experiments of this paper, the hybrid model achieved the highest accuracy rate of 0.96, which is slightly higher than the improved CNN model (0.94), and in turn higher than the basic CNN model (0.92). This indicates that the hybrid model is able to make classification decisions more accurately over the entire sample set. Precision rate is the proportion of actual positive cases that are judged by the model to be positive, recall rate is the proportion of actual positive cases that are judged by the model to be positive, and F1 Score is the reconciled average of precision rate and recall rate. Observing the experimental results of this paper, this paper finds that the hybrid model achieves the highest values for Precision Rate, Recall Rate and F1 Score, which are 0.96, 0.95 and 0.96, respectively. the Improved CNN model is the next highest, which is 0.95, 0.93 and 0.94, respectively. the Precision Rate, Recall Rate and F1 Score for the Basic CNN model are 0.93, 0.91 and 0.92, respectively. The high and low levels of these metrics indicate that the hybrid model is able to identify true positive examples more accurately and tries to avoid misjudging negative examples as positive ones. By comparing the experimental results, this paper can conclude that the hybrid model has the best performance in gesture recognition and action control, probably because it combines the advantages of different models. The improved CNN model is significantly better than the basic CNN model, but still slightly inferior to the hybrid model. The deep learning model is promising for a wide range of applications in gesture recognition and action control, but further optimization and improvement are still needed. Combining the above analysis, this paper can conclude that the hybrid model is the best performance among these three models. The hybrid model combines the advantages of the basic CNN model and the improved CNN model, and is able to strike a balance between precision and recall, resulting in a better F1 Score. The improved CNN model also performs well, followed by the basic CNN model. This suggests that improved model design and structural adjustments can positively affect the performance of the model.

4. Conclusion

This study aims to explore and compare the application of deep learning techniques in gesture recognition and motion control. This paper designed and experimented the basic CNN model, the improved CNN model and the hybrid model, and evaluated their performances through multiple metrics. The following is the summary and conclusion of the study: model performance comparison: through the analysis and comparison of the experimental results, this paper finds that the hybrid model performs optimally in terms of accuracy, precision, recall and F1 Score. It is followed by the improved CNN model, while the basic CNN model is slightly inferior in each index. Hybrid Model Advantage: The hybrid model fully utilizes the advantages of the basic CNN model and the improved CNN model, integrates their performance, and achieves the best overall performance. This verifies the effectiveness of synthesizing different models. Experimental significance: the results of this study are important for guiding the design and development of practical gesture recognition and motion control systems. In this paper, the performance of different models is comprehensively evaluated, which provides a basis for selecting the appropriate model. FUTURE OUTLOOK: Future research can further optimize the hybrid

model and explore more effective model fusion methods. Meanwhile, the introduction of advanced deep learning techniques from more fields can be considered to further improve the performance and accuracy of gesture recognition and motion control. In summary, this study is of positive significance for promoting the application of deep learning techniques in the field of gesture recognition and motion control. This paper expects to see more innovations and breakthroughs based on deep learning in the future to meet diverse human-computer interaction needs.

References

Abhishek, B., Krishi, K., Meghana, M., Daaniyaal, M., & Anupama, H. S. (2020). Hand gesture recognition using machine learning algorithms. *Computer Science and Information Technologies*, 1(3), 116-120.

Baumgartl, H., Sauter, D., Schenk, C., Atik, C., & Buettner, R. (2021). Vision-based hand gesture recognition for human-computer interaction using MobileNetV2. *In 2021 IEEE 45th Annual Computers, Software, and Applications Conference* (COMPSAC) (pp. 1667-1674). IEEE.

Eren, M. (2018). Forecasting of the fuzzy univariate time series by the optimal lagged regression structure determined based on the genetic algorithm. *Economic Computation & Economic Cybernetics Studies & Research*, 52(2).

Jain, R., Jain, M., Jain, R., & Madan, S. (2022). Human computer interaction–Hand gesture recognition. *Advanced Journal of Graduate Research*, 11(1), 1-9.

Liang, P. P., & Li, C. W. (2019). Impact of cooperation uncertainty on the robustness of manufacturing service system. *Advances in Production Engineering &. Management*, 14(2), 189-200.

Lv, Z., Poiesi, F., Dong, Q., Lloret, J., & Song, H. (2022). Deep learning for intelligent human–computer interaction. *Applied Sciences*, 12(22), 11457.

Ozdemir, M. A., Kisa, D. H., Guren, O., Onan, A., & Akan, A. (2020). EMG based hand gesture recognition using deep learning. *In 2020 Medical Technologies Congress* (TIPTEKNO) (pp. 1-4). IEEE.

Qi, J., Jiang, G., Li, G., Sun, Y., & Tao, B. (2019). Intelligent human-computer interaction based on surface EMG gesture recognition. *Ieee Access*, 7, 61378-61387.

Rady, M. A., Youssef, S. M., & Fayed, S. F. (2019). Smart gesture-based control in human computer interaction applications for special-need people. *In 2019 novel intelligent and leading emerging sciences conference* (NILES) (Vol. 1, pp. 244-248). IEEE.

Stephan, J. J., & Sana'a Khudayer. (2010). Gesture Recognition for Human-Computer Interaction (HCI). Int. *J. Adv. Comp. Techn.*, 2(4), 30-35.

Tsai, T. H., Huang, C. C., & Zhang, K. L. (2020). Design of hand gesture recognition system for human-computer interaction. *Multimedia tools and applications*, 79, 5989-6007.

Zhang, Z. J., Wang, P., Wan, M. Y., Guo, J. H., & Luo, C. L. (2020). Interactive impacts of overconfidence and fairness concern on supply chain performance. *Advances in Production Engineering & Management*, 15(3), 277-294.

Zhou, H., Wang, D., Yu, Y., & Zhang, Z. (2023). Research Progress of Human–Computer Interaction Technology Based on Gesture Recognition. *Electronics*, 12(13), 2805.