Proceedings of the 2nd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/62/20240392

Analysis of gesture recognition applications based on deep learning

Yilin Liu

College of Information Science and Technology, Donghua University, Shanghai, 201602, China

bjones61696@student.napavalley.edu

Abstract. With the advent of artificial intelligence, deep learning has been continuously evolving. It has progressed from its initial stages, exemplified by AlphaGo, to the development of related technologies exclusively centred around deep learning for the game of Go. Subsequently, deep learning has found applications in an increasing number of domains, including healthcare, sports, and more. This paper delves into the profound influence and development of deep learning as the foundation for analysing its impact on gesture recognition. Deep learning has brought about significant advancements in this field, enabling more precise and versatile gesture recognition systems. Furthermore, we explore specific use cases and contexts where deep learning has been harnessed, such as in medical diagnostics and sports performance analysis. While scrutinizing these applications, a promising future for deep learning becomes evident, with the potential to revolutionize various industries and enhance technology interaction through the development of more intuitive and sophisticated gesture recognition systems. The continuous growth and evolution of deep learning offer a bright prospect for the future of human-machine interfaces and artificial intelligence-driven solutions.

Keywords: Artificial Intelligence, Deep Learning, Gesture Recognition.

1. Introduction

With the continuous development of deep learning, the application of artificial intelligence deep learning in the field of gesture recognition has made significant progress [1]. This technology has important applications in various domains, including computer vision, human-computer interaction, virtual reality, and augmented reality [2-4]. It not only enhances entertainment in daily life but also greatly advances the field of medical technology. However, gesture recognition also faces various technical challenges, such as the removal of complex backgrounds, variations in lighting conditions, diversity of gestures, real-time requirements, and accuracy issues. Addressing these challenges necessitates highly accurate algorithms and large-scale training data. This is where deep learning techniques, particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have played a pivotal role in gesture recognition. These deep learning models can automatically learn features and improve accuracy. This article aims to summarize various methods of deep learning and their relationship with gesture recognition to improve the recognition of details and differences within gestures. This leads to enhanced accuracy and robustness in gesture recognition, enabling reliable performance in both constrained environments and changing conditions. Furthermore, deep learning allows for the

© 2024 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/62/20240392

recognition of various types of gestures, including sign language, gesture control, and gesture actions. This versatility makes it widely applicable in multiple fields, ranging from assistive technology for people with disabilities to virtual reality and game control. Additionally, deep learning models exhibit adaptability, continuously improving their performance based on new data and usage contexts. Finally, deep learning models can be optimized for real-time requirements to meet the low-latency response needs of applications.

2. Fundamentals of Deep Learning Algorithms

2.1. Theoretical Analysis of the Convolutional Neural Network (CNN) Algorithm

CNN is a type of deep learning model primarily used for image recognition and processing. It consists of convolutional layers, which can capture image features and generate feature maps; pooling layers, which reduce the spatial dimensions of feature maps while preserving crucial information; activation functions connecting convolutional layers and fully connected layers; and fully connected layers and an output layer positioned between the convolutional layers and the output layer to classify information or transform it into a regression format. The CNN structure is shown in figure 1.

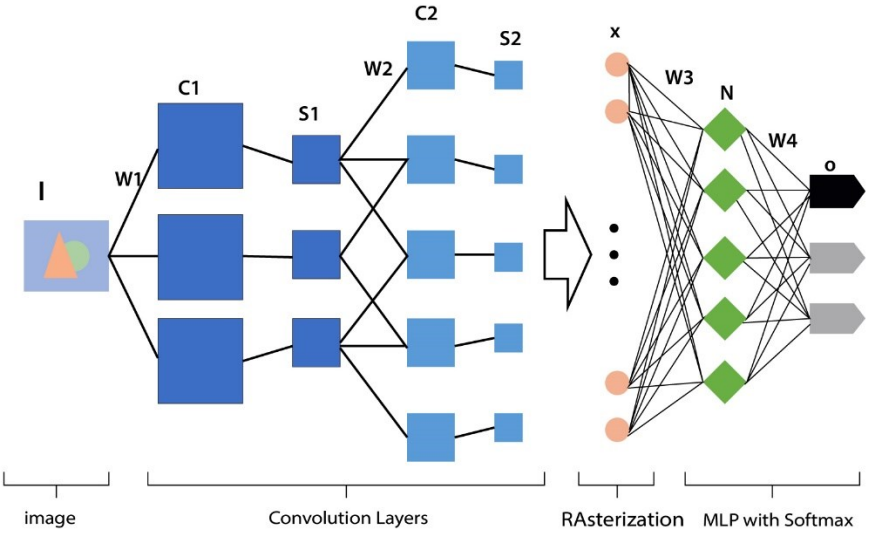

Figure 1. The CNN structure (Photo/Picture credit: Original).

2.2. The connection between CNN algorithm and gesture recognition

The connection between CNN and artificial intelligence gesture recognition is closely intertwined. CNN is a deep learning model specifically designed for processing image data, while artificial intelligence gesture recognition is an application domain aimed at identifying and understanding the meaning of human gestures, actions, or postures [5, 6].

Feature extraction: CNN is widely employed in image processing for feature extraction. Convolutional layers and pooling layers can automatically learn and extract features from images, which are crucial for identifying key information in gestures, such as the position of fingers or the shape of the palm.

Image classification: CNN can be used for image classification tasks, which is essential for gesture recognition. Well-trained CNN models can compare input images with different gesture or posture categories, thus enabling gesture classification.

Real-time processing: Artificial intelligence gesture recognition typically requires real-time processing in applications like user interfaces controlled by gestures or gesture recognition in games.

DOI: 10.54254/2755-2721/62/20240392

CNN models can be optimized for real-time performance, ensuring gesture recognition in low-latency environments.

Data augmentation: Gesture recognition demands a substantial amount of training data for accurate recognition of various gestures. CNN can be combined with data augmentation techniques to enhance model performance by transforming and augmenting training data, increasing data diversity.

Transfer learning: Transfer learning is a method of applying pre-trained CNN models to new tasks. In gesture recognition, pre-trained CNN models can be used and fine-tuned for specific gesture recognition tasks, saving significant training time and resources. Gu further elaborates on the advantages of CNN in the field of gesture recognition. Additionally, it combines CNN with LSTM and channel attention, showcasing the excellent ability of CNN to extract information in both the temporal and spatial domains [7].

2.3. The advantages of Recurrent Neural Network (RNN)

There is indeed a connection between Recurrent Neural Network (RNN) and artificial intelligence gesture recognition. Although Convolutional Neural Network (CNN) are more common in practical applications, RNNs can also be used for gesture recognition tasks in certain scenarios. The RNN structure is shown in figure 2.

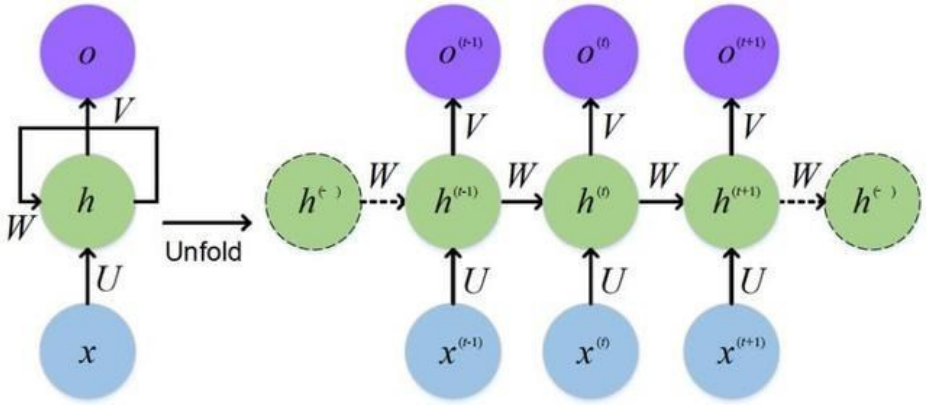

Figure 2. The RNN structure (Photo/Picture credit: Original).

Temporal Information Processing: Artificial intelligence gesture recognition often involves processing time series data because gestures typically contain dynamic movements. RNNs are well-suited for handling time series data in deep learning models and can capture temporal features of gestures, such as hand trajectories and gesture changes [8, 9].

Recurrent Structure: RNNs have a recurrent structure that allows information to propagate within the network, making them suitable for processing sequences of varying lengths. In gesture recognition, the temporal information of gestures can be modelled using the recurrent structure of RNNs.

Sequence Labelling: Gesture recognition tasks typically involve mapping a time series to a sequence of labels, where each label represents the gesture category at a specific time. RNNs can be used for sequence-to-sequence tasks and, therefore, can be employed to map gesture time series to corresponding label sequences.

Handling Long-Term Dependencies: RNNs excel at handling long-term dependencies, which can be crucial for some gesture recognition tasks. For instance, recognizing the dynamic evolution of gestures may require considering longer contextual information.

Combining CNNs and RNNs: Sometimes, a combination of CNNs and RNNs is used to leverage their respective strengths. For example, CNNs can extract spatial features from static gesture images, while RNNs can process the temporal aspects of gestures, resulting in more accurate gesture recognition.

It's worth noting that RNNs have certain drawbacks, such as training difficulties and the vanishing gradient problem. Therefore, the choice of using CNNs, RNNs, or their combinations depends on the specific requirements of the gesture recognition task and the characteristics of the dataset. In practical

DOI: 10.54254/2755-2721/62/20240392

applications, the selection of an appropriate model architecture is typically based on task demands and performance criteria.

However, when RNN and CNN are combined for use in gesture recognition, their impact remains significant. This also underscores that contemporary technology cannot progress in isolation and must leverage a combination of ideas to achieve greater effectiveness [10].

2.4. Long Short-Term Memory (LSTM) inherits the advantages

LSTM was originally designed to address the long-term dependency problem commonly found in standard RNNs. LSTM is capable of effectively transmitting and maintaining information over long sequences without forgetting crucial information from earlier time steps. Moreover, LSTM can mitigate the vanishing gradient and exploding gradient problems encountered in RNNs.

There is a close relationship between LSTM networks and artificial intelligence gesture recognition. LSTM is a variant of recurrent neural networks (RNNs) with powerful capabilities for handling sequential data. Here are the connections between LSTM and artificial intelligence gesture recognition:

Modelling Temporal Information: Artificial intelligence gesture recognition often involves considering the temporal aspects of gestures, including dynamic changes in gestures, hand trajectories, and motion patterns. LSTM networks are designed to capture temporal information, making them well-suited for processing the time sequences associated with gestures. The LSTM structure is shown in figure 3.

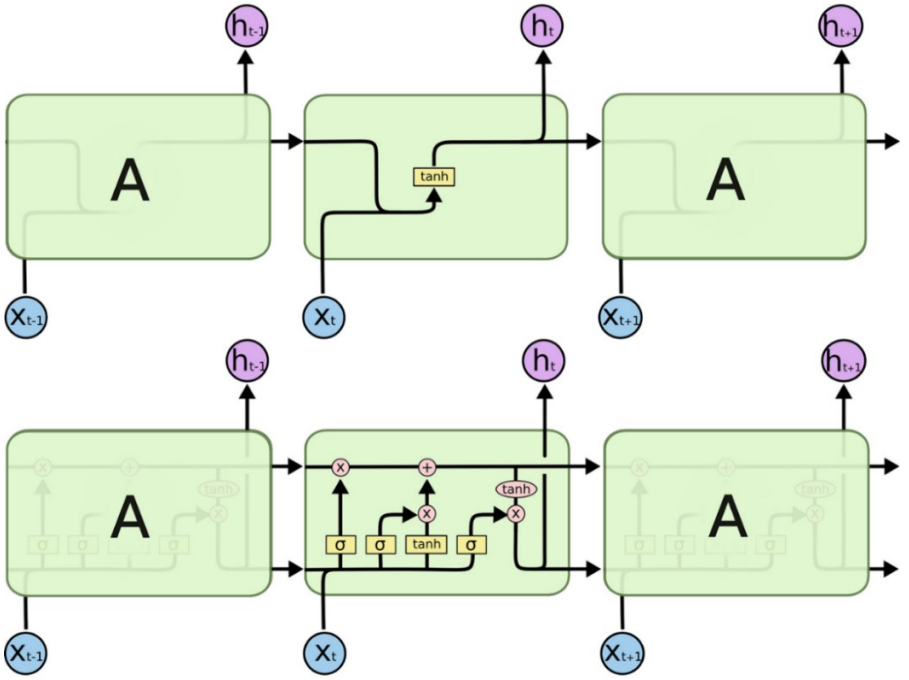

Figure 3. The LSTM structure (Photo/Picture credit: Original).

Handling Long-Term Dependencies: LSTM networks have the ability to model long-term dependencies in data. This is crucial in gesture recognition, as dynamic features of gestures may evolve over longer time spans. LSTM can capture these extended dependencies, contributing to more accurate gesture recognition.

Sequence-to-Sequence Tasks: Gesture recognition tasks can often be framed as mapping input gesture time sequences to corresponding label sequences. LSTM can be used for sequence-to-sequence learning tasks, making it applicable to gesture recognition where sequential information is vital.

DOI: 10.54254/2755-2721/62/20240392

Combining with Convolutional Neural Networks (CNNs): Typically, LSTM is combined with CNNs to leverage their respective strengths. CNNs are effective at extracting spatial features from static gesture images, while LSTM excels at processing the dynamic temporal information of gestures. This combination enhances the overall performance of gesture recognition systems [11].

3. The application of deep learning in gesture recognition

3.1. The application of deep learning in gesture translation

Translation software is proliferating nowadays, but specialized sign language translation for people with disabilities is not yet widespread. The reason behind this is that sign language translation is primarily carried out by individuals who have learned sign language, and their efficiency is far from that of machine translation. Additionally, sign languages are region-specific, meaning that sign languages can vary between different regions. It is challenging for one person to comprehensively learn various regional sign languages. Therefore, utilizing deep learning for gesture recognition can enhance the precision of sign language recognition and enrich the content of remembered sign language.

In the article by Raffort J, it is mentioned that wearable devices can collect a substantial amount of raw data from various physiological signals [1]. Subsequently, machine learning algorithms can be employed to process this data, establishing a new data analysis paradigm that contributes to the advancement and practicality of intelligent wearable devices. The selection of appropriate machine learning algorithms enables the extraction of valuable information regarding various signal attributes from the raw data, enhancing the intelligence of these wearable devices. The correct choice of algorithms for different types of raw data is crucial in establishing an accurate and reliable correlation between sensing signals and physiological status.

Furthermore, by incorporating electromyography (EMG) sensors to record muscle activity, the Word Error Rate (WER) for continuous sentence recognition has been reduced to 9.6%, which is significantly lower than isolated methods. Moreover, the detection of sign language and recognition of phrases containing six sign words typically take less than 0.9 seconds. This progress will facilitate the integration of this technology into various wearable artificial intelligence biosensors in the future [1].

Similar research is also being conducted, combining CNN and LSTM with surface electromyography. After processing through convolution, pooling, gradient descent, and dropout layers, they obtained a classification model for eight different gestures. The model was evaluated using test dataset, achieving an impressive accuracy rate of up to 96%, demonstrating its high feasibility [12].

3.2. The application of deep learning in gesture recognition in the medical field.

In the medical field, both for doctors and patients, hand gestures play a crucial role. Utilizing deep learning for standard hand gestures can help prevent and correct issues related to hand posture, thereby enhancing the precision of surgical procedures performed by doctors and improving the risk prediction for common individuals. Additionally, it allows for timely correction of physiological problems in patients.

Due to postural imbalances resulting from strokes, the quality of life for patients is significantly compromised. This is especially true when symptoms such as stiff fingers directly impair the gripping and stretching functions of the hands. Hence, timely correction of finger posture in stroke patients is of utmost importance. Moreover, intelligent recognition of finger postures can facilitate barrier-free communication. The research objective of this study is to investigate the correction of finger postures in stroke patients based on principles of artificial learning. Furthermore, gesture recognition serves as the foundation for intelligent posture recognition, laying the groundwork for the development of intelligent rehabilitation devices.

The research methodology involves several steps. Firstly, preprocessing, noise reduction, and edge detection are applied to the original gesture images to obtain images of the hand's edges. Subsequently, two input channels are chosen for a Convolutional Neural Network (CNN), one for gesture images and the other for hand edge images. Each channel consists of an equal number of convolutional layers with individual parameters. Finally, feature fusion is conducted in the fully connected layer, and a softmax

DOI: 10.54254/2755-2721/62/20240392

classifier is employed to categorize the output results. Experimental results demonstrate that the dual-channel DC-Net algorithm effectively enhances the recognition rate for rehabilitation gestures, achieving a training accuracy of 99.6% with a reduced loss value of 0.06. This boosts the CNN's generalization ability [2].

Building upon the research in intelligent posture recognition, this paper further explores the application of visual detection in gesture recognition. It establishes an integrated system for gesture recognition and visual detection, enabling functionalities such as gesture data collection, deep learning network training, and network testing. The data collection module is employed to gather training and testing data, the deep learning training module is utilized for parameter adjustment and training result visualization, and the deep learning testing module evaluates the network's performance.

In the future, building upon this research foundation, the development of intelligent interaction methods based on posture recognition will be expanded. This includes the extension of rehabilitation training for various body parts, ultimately creating a comprehensive system of intelligent rehabilitation devices for stroke patients [2].

3.3. The Kinematic Study of Deep Learning Gesture Recognition

Gesture recognition also has significant applications in kinematics because making gestures is a form of motion. While the recognition of static gestures has been mentioned earlier, here, we mainly elaborate on the recognition of dynamic and three-dimensional gestures. The significance of recognizing these gestures lies in enhancing a machine's ability to preprocess gestures [13].

Yueqin Wang combines gesture recognition with motion sensors, specifically MEMS (Micro-Electro-Mechanical Systems) inertial sensors. In this approach, motion sensors serve as the input, continuously collecting real-time motion information from user gestures. This information includes data from 3-axis accelerometers, 3-axis gyroscopes (angular velocity), and 3-axis magnetometers (magnetic field intensity). Using quaternion mathematics, the system calculates three attitude angles (pitch, roll, and yaw) from this data. Subsequently, a recognition algorithm is designed by analyzing and mining these 12 sets of data [3].

Currently, common gesture recognition algorithms include Hidden Markov Models (HMM) and template matching. However, since this research focuses on a set of distinct gestures that require real-time and accurate recognition, it proposes a "feature analysis" method. This approach involves selecting distinctive and representative feature values to establish a pre-classifier for gestures.

The study categorizes gestures into six types: motion, tap, rotation, shake, hook, and cross. For the rotation and motion categories, specific directions are further identified. Through the observation and analysis of a large number of samples, five recognition features are determined to build the pre-classifier: gesture length, energy, number of peaks (angular velocity peaks, acceleration peaks, and attitude angle peaks), unilateral angular velocity, and the axis with maximum angular velocity energy. Using these features, the system first identifies the major category to which a gesture belongs and then determines the specific motion direction.

Experimental validation shows that the "feature analysis" method achieves an average recognition accuracy of 89.2%. While ensuring high recognition accuracy, it significantly improves the real-time performance of 3D dynamic gesture recognition, making it highly practical [3].

4. Conclusion

In conclusion, the application of deep learning techniques has undeniably revolutionized the field of gesture recognition, yielding remarkable advancements in accuracy. By leveraging sophisticated deep learning architectures such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), gesture recognition systems have become adept at capturing intricate gesture features, enabling more precise and reliable recognition.

However, it's important to acknowledge that this progress is not without its challenges. Differing lighting conditions can cause variations in the appearance of gestures, thereby increasing the difficulty of recognition. Particularly in outdoor environments or under unstable lighting sources, the performance of gesture recognition may deteriorate. Additionally, people use a variety of different gestures to convey

DOI: 10.54254/2755-2721/62/20240392

distinct intentions. This implies that the system needs to be capable of handling diverse gestures and accurately categorizing them, which presents a challenge for training and testing models. In some cases, gesture recognition involves monitoring and recording users' actions, raising privacy concerns. Appropriate privacy policies and technical safeguards must be established to address these issues. One significant challenge lies in the need for vast amounts of labelled data for training deep learning models, which can be time-consuming and resource-intensive to acquire. Additionally, ensuring the robustness and generalizability of gesture recognition systems across diverse environments and user demographics remains an ongoing challenge. Addressing these issues will be crucial for further improving the efficacy of gesture recognition.

Looking ahead, the future of gesture recognition holds exciting possibilities. Researchers are exploring the fusion of multiple modalities, including vision, audio, and motion, to enhance robustness and accuracy further. This multi-modal approach could enable gesture recognition systems to better understand context and user intent, opening new avenues for applications in fields such as healthcare, gaming, and human-computer interaction.

References

[1] Raffort J, Adam C, Carrier M, et al. Application of Artificial Intelligence for Automatic Vascular Segmentation: Development of a Hybrid Method Combining Expert System with Deep Learning[J]. European Journal of Vascular and Endovascular Surgery, 2022, 63(2): e42-e43.

[2] Gao Y. Research on the Application of Artificial Intelligence Technology in the Development of Computer Vision[J]. Highlights in Science, Engineering and Technology, 2022, 9: 80-84.

[3] Zhang Daozhou. Research on Finger-Shaped Flexible Correction and Posture Intelligent Recognition in Stroke Patients. Yangzhou University, 2021.

[4] Wang Youqin. Gesture Recognition Based on Motion Sensors. Harbin Institute of Technology, 2015.

[5] Jin, Ke. Key Technologies Research on Gesture Recognition in Complex Scenes Based on Depth Information. Shanghai Jiao Tong University, 2018.

[6] Han, Lei. Research on FMCW Radar Gesture Recognition Algorithm Based on Deep Learning. Nanjing University of Posts and Telecommunications, 2022.

[7] Gu Y,Li J. A Novel WiFi Gesture Recognition Method Based on CNN-LSTM and Channel Attention. The International Society for Applied Computing.Proceedings of 2021 3rd International Conference on Advanced Information Science and System (AISS 2021).ACM, 2021:497-500.

[8] Gan, Ji. Algorithm Research on Handwritten Text Recognition and Related Issues. University of the Chinese Academy of Sciences (School of Computer Science and Technology, Chinese Academy of Sciences), 2021.

[9] Liu, Man. Design and Intelligent Application Research of Deep Learning Methods under the Introduction of Uncertain Information and Difference Structure. Yunnan University of Finance and Economics, 2023.

[10] Prakash K B, Eluri R K, Naidu N B, et al. Accurate hand gesture recognition using CNN and RNN approaches. International Journal, 2020, 9(3).

[11] Song Tianyu. Research on Hill Flood Forecasting Based on Long Short-Term Memory Network. Dalian University of Technology, 2022.

[12] Bo L, Banghua Y, Shouwei G, et al. Hand Gesture Recognition using sEMG Signals Based on CNN. 2021 40th Chinese Control Conference (CCC). IEEE, 2021: 7180-7184.

[13] Yang Haihong. Research on Human-Computer Interaction Application of Gesture Recognition Based on Deep Learning. Journal of Baoshan College, 2022, 41(05): 83-91.