Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 10, No. 1, 2024

Deep Learning-Based Recognition and Visualization of Human Motion Behavior

Guoqing Cai[1, *], Quan Zhang[2], Beichang Liu[3], Zhengyu Jin[4], Jili Qian[5]

[1] Information Studies, Trine University, Phoenix AZ, USA
[2] Information Studies, Trine University, Phoenix AZ, USA
[3] Information Studies, Trine University, salt lake city UT, USA
[4] Informatics, Univeristy of California, Irvine, CA, USA
[5] Information Studies, Trine University, Phoenix AZ, USA

[*]Corresponding author E-mail: cailiveinusa@gmail.com

Abstract: Human behavior recognition refers to the classification task of identifying the specific actions of human characters based on the characteristics of human body and the completed actions through a specific algorithm. It has a wide range of applications in intelligent surveillance, video retrieval and so on. The main challenge in this direction is to accurately extract the semantic information of each behavior to describe its dynamic changes in space and time. Therefore, this article introduces the latest research progress in the field of human behavior recognition. Through deep learning techniques, particularly convolutional neural networks and recurrent neural networks, human movements in video data can be effectively identified. However, deep learning models lack interpretability, which can be a challenge in practical applications. The researchers also introduce the application of traditional methods and deep learning-based methods to human behavior recognition, and explore the advantages of deep learning models in processing multi-time scale information and introducing attention mechanisms. Finally, the paper summarizes the potential of deep learning technology combined with multimodal data in behavioral analysis, and provides prospects for applications in smart fitness, health care and other fields.

Keywords: Deep learning; Human movement behavior; Visual recognition; RGB recognition.

1. Introduction

The research goal of human behavior recognition is to develop a visual system that mimics human behavior to understand and describe human behavior in a given scene through the recognition of human behavior, which has been widely used in intelligent security, virtual reality, human-computer interaction and other fields and has obtained high commercial value. In recent years, with the further research of low-cost wearable sensors and depth cameras, the type of data used for human behavior recognition research is not limited to RGB, but also new modes of data such as deep bone and infrared data. According to the type of data, the current popular human behavior recognition research methods include RGB based behavior recognition and bone based behavior recognition, both of which are hot directions in the field.Recognition plays a vital role across multiple domains, including surveillance, robotics, and healthcare. It involves the automatic identification and classification of human actions from video data, enabling applications such as identifying suspicious behavior in security footage, guiding robotic movements, and monitoring patient movements for healthcare purposes.Deep learning models have revolutionized human action recognition by effectively capturing temporal dependencies within action sequences. Unlike traditional approaches that rely on handcrafted features and shallow learning algorithms, deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), excel at learning complex patterns and temporal dynamics directly from raw data, such as video frames or motion sequences.

However, despite their high performance, deep learning models often lack interpretability, making it challenging to understand how they infer temporal relationships in human actions. This interpretability gap hinders the trustworthiness and applicability of these models in real-world scenarios, as stakeholders require insights into model decision-making processes for effective deployment and decision support.

2. Related Work

2.1. Traditional approaches to human action recognition

In traditional methods of human action recognition, hand-designed features and techniques based on machine learning algorithms are often used. For example, traditional motion recognition methods might use hand-designed features such as motion trajectories, color histograms, or optical flows to describe motion patterns in video sequences. For example, for the field of video surveillance, track-based features can be used to capture changes in direction and speed of human movement to distinguish between different types of actions, such as walking, running, or stopping. In addition, traditional machine learning algorithms, such as support vector machines (SVM) or Random Forest, are often used to classify and identify extracted features. In healthcare, for example, these algorithms can be used to classify the movement patterns of patients to help diagnose and monitor their health.

Traditional action recognition methods have some limitations when dealing with complex action patterns. These methods often rely on hand-designed features that may not fully capture the deep motivation and intent behind an action. For example, Lian Zhixian points out in his article that AI's lack of flexible recognition of human motivation suggests that traditional methods may not be able to effectively capture the deep motivations and intentions behind actions, thus limiting

its application in complex action patterns. Another problem is that traditional methods often adopt fixed models and algorithms to handle action recognition tasks, which limits their ability to adapt to new data and action patterns. In their research, Huan Zhanet al. mentioned the study of human activity recognition based on open set class incremental learning, which indicates the need for new action recognition and learning. Traditional methods may not be flexible enough to adapt to new modes of action, thus limiting their effectiveness in practical applications.

Traditional action recognition methods also have problems in dealing with challenges under complex environmental conditions. Human motion recognition often needs to be carried out under complex environmental conditions, such as different lighting, background and occlusion. Traditional approaches may not be able to effectively address these challenges, resulting in reduced recognition accuracy. However, the traditional methods have limitations on the extraction of human motion features, and can not capture complex motion patterns and spatio-temporal relationships. For example, in video surveillance, methods based on manual features may not be able to effectively deal with motion recognition under different lighting conditions, resulting in reduced recognition accuracy. In addition, traditional machine learning algorithms often rely on manually selected feature sets and may not adequately mine the underlying patterns and information in the data, thus limiting the performance and generalization of action recognition. This can be drawn from the study by Huan et al., which may involve the challenge of action recognition in real-world situations. Therefore, in order to improve the accuracy and robustness of action recognition, it is necessary to explore new methods and techniques to overcome the limitations of traditional methods.

2.2. Motion recognition based on deep learning

The introduction of deep learning has brought a new development direction for human behavior recognition based on RGB. The original research methods of deep learning in the field of human behavior recognition focused on the feature extraction of RGB static images, and now it focuses on video images. Guo et al. investigated human behavior recognition technology based on static RGB images and discussed different methods of machine learning and deep learning for low-level feature extraction and high-level behavior representation. Ma et al. [5] discuss in detail the advantages and disadvantages of deep learning representations, while also introducing popular standard datasets. Pei Lishen et al. [6] focused more on summarizing the development of algorithms for understanding activity details in group behavior. In recent years, the advantages of combining bone data with deep learning have gradually become apparent. Many scholars gradually close Note Research on human behavior recognition based on bone. The investigation [7, 8] was not only detailed.

The structure and skeleton data of Graph Convolution Neural Networks (GCN) for human behavior recognition are introduced.The application of GCN in the field of human behavior recognition is emphasized. At present, the previous research does not include the comprehensive investigation of the two research methods based on RGB and bone, and lacks a macro and comprehensive introduction. Therefore, the above two popular methods are respectively carried out in this paper comprehensive classified survey was conducted.

Therefore, the advantages of deep learning motion recognition methods in capturing temporal dynamics also include their ability to handle multiple time scales. Deep learning models can capture information on different time scales in action sequences by using time Windows of different lengths to more fully understand and analyze the evolution of actions. In addition, deep learning models can also dynamically focus on important parts of action sequences by introducing attention mechanisms, improving the processing efficiency and accuracy of long time series. Overall, the advantage of deep learning motion recognition method in capturing temporal dynamics makes it one of the mainstream methods in the field of motion recognition, which provides strong support for realizing more accurate and robust motion recognition tasks.

2.3. RGB human body recognition and tracking

In recent years, many researchers have tried to integrate RGB video sequences Texture information in a column, depth information in a depth motion diagram, or bone diagram Bone position information. Spatiotemporal gradient autocorrelation vector features and gradient local autocorrelation feature (GLAC) were extracted using depth maps. At the same time, static pose model, motion model and dynamic migration model of skeletal frame images were used to represent the underlying features of the actions, and weight voting mechanism was used to integrate the above features. Imran et al. used RGB sequence and depth map sequence to calculate motion history map (MHI) and DMM, respectively. And input into four independent convolutional neural networks (CNNS) for training, and finally fuse the output of each CNN network

Score. Luo et al. extracted the central symmetric local operation of RGB sequences respectively

Dynamic feature (CS-Mltp) and three-dimensional position information feature of depth map sequence are described by sparse coding based time pyramid (ScTPM), and the recognition effect of feature layer fusion and fraction fusion is compared respectively.

Traditional RGB human behavior recognition feature extraction adopts manual annotation, while deep learning-based RGB human behavior recognition adopts deep architecture feature extraction. The following summarizes the methods of manual feature and depth architecture based on RGB respectively. 1.1 Manual Feature method Based on RGB human behavior recognition methods based on manual features generally include two main steps: behavior representation and behavior classification. The goal of behavior representation is to convert video information into feature vectors, extract representative and discriminative information of human behavior, and minimize changes to improve recognition performance. The methods of behavior representation can be roughly divided into global feature representation and local feature representation. The global representation method can capture the motion information of the whole body, but because the information capture area is a fixed rectangle, it will introduce irrelevant background information. Bobick, wait! Based on global representation, Motion Energy Image (MEI) and Motion History Image (MHI) are proposed to encode a single image of a dynamic human body.

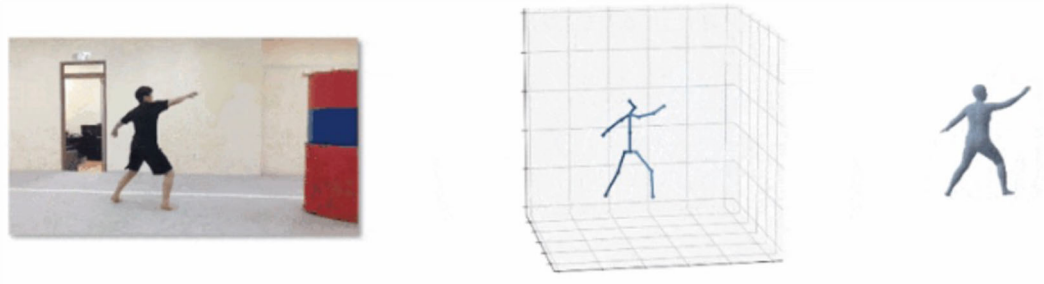

Figure 1. Example of RGB bone behavior recognition

Based on the previous review, this paper aims to explore deep learning methods in the field of human behavior recognition and apply them to realize yoga pose recognition. After introducing traditional methods and human behavior recognition methods based on deep learning, this paper will focus on human behavior recognition methods based on RGB images. Specifically, we will employ a deep learning architecture for feature extraction to improve recognition performance and accuracy. Traditional methods based on manual features have limitations, so we will employ deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) to directly learn complex patterns and temporal dynamics in video data. By discussing the advantages of deep learning models in capturing temporal dynamics, this paper will introduce their ability to handle multiple time scales and introduce attention mechanisms to dynamically focus on important action parts. Through this research, we hope to better understand and apply the potential of deep learning models in the field of human behavior recognition, and provide support for the realization of more accurate and reliable human behavior recognition technology.

3. Methodology

This research is based on deep learning technology, especially the MediaPipe and OpenCV software packages, aiming to realize yoga pose recognition. The goal of our research is to identify poses for the human body, limiting it to three common yoga poses: T-Pose, Tree-Pose and Warrior-Pose. The experimental process mainly includes the following steps:

1. Install the necessary packages, including MediaPipe and OpenCV, for image processing and pose recognition.

2. Import the necessary libraries, such as MediaPipe and OpenCV, for programming in a Python environment.

3. Initialize MediaPipe and Pose models to prepare for human pose recognition.

4. Read images and identify key points of the human body, and use MediaPipe for posture estimation.

5. Write functions to detect and identify postures in key points. In this step, we can use deep learning techniques to classify key points to determine the type of pose shown in the image.

6. Output predictive results, i.e. identified pose types, such as T-Pose, Tree-Pose, or Warrior-Pose.

In addition, we will also combine some RGB-based recognition methods and use deep learning architecture for feature extraction and posture classification to improve recognition performance and accuracy. Through this research, we aim to explore the feasibility of using deep learning technology to realize yoga pose recognition, and provide technical support for applications in related fields.

3.1. Experimental model

This part of the code mainly imports the necessary libraries and initializes the pose recognition model in MediaPipe. To be specific:

1.Import some necessary Python libraries such as math, cv2, numpy, time, mediapipe, and matplotlib.pyplot for mathematical operations, image processing, time calculation, posture recognition, and more.

2.The pose recognition model in MediaPipe is initialized, including the Settings of the pose recognition model, such as static image mode, minimum detection confidence and model complexity.

3.Initializes the drawing tool class in MediaPipe for annotating and drawing on images.

Figure 1. Experimental image

3.2. Data preprocessing

The MediaPipe library in conjunction with OpenCV to conduct pose detection on a provided image. It identifies and extracts keypoint landmarks for the first two detected points, showcasing their coordinates (x, y, z) and visibility in a tabular format. Additionally, the code overlays the identified landmarks on the input image to provide a visual representation of the pose detection process. This comprehensive approach aids in analyzing the spatial positioning of critical body points, facilitating tasks such as yoga pose recognition or movement tracking.

Table 1. Extraction point data table

Keypoint	x	y	z	Visibility
NOSE	254.24	125.57	-143.61	0.9999
LEFT_EYE_INNER	263.94	110.65	-124.76	0.9999

The marked output is as follows:

Figure 3. Original picture labeling behavior

*# Plot Pose landmarks in 3D.*mp_drawing.plot_landmarks(results.pose_world_landmarks, mp_pose.POSE_CONNECTIONS)

This code aims to visualize the detected pose landmarks on a provided image. Begins by creating a copy of the sample image to draw the landmarks on. If any landmarks are found in the image, it proceeds to draw them using the `mp_drawing.draw_landmarks` function from the MediaPipe library. The resulting image with overlaid landmarks is displayed using matplotlib, providing a visual representation of the detected pose.

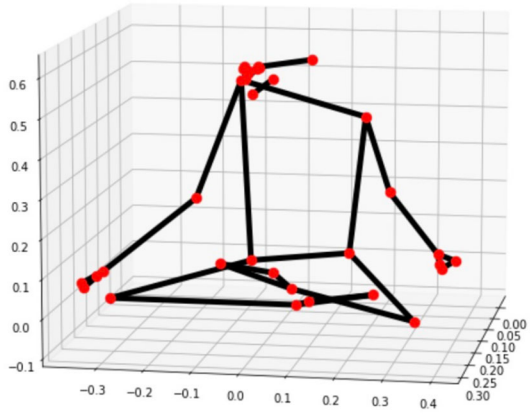

Figure 4. Visualization of the initial action

Additionally, there's a mention of detecting poses and recognition, which implies that this code segment may be part of a larger system for detecting and recognizing poses. It also suggests the potential application of the code for identifying different yoga poses such as T-pose, Tree-pose, and Warrior-Pose, among others. However, the specific implementation details for pose detection and recognition are not provided in this snippet.

3.3. Experimental result

From this short experiment, we can draw the following conclusions:

1. Posture recognition accuracy: By using deep learning technology and MediaPipe library, we successfully realized the recognition of human posture in images. The results showed that we were able to accurately detect key points in the human body and identify common yoga poses such as T-Pose, Tree-Pose and Warrior-Pose.

2. Key point extraction and visualization: In the experiment, we used MediaPipe and OpenCV libraries to extract key points in images, and marked and visualized them on the original images. This detailed key-point extraction and visualization helps us understand the spatial position of human poses and provides a basis for subsequent yoga pose recognition.

3. Application of deep learning technology: In this experiment, we used deep learning technology for posture recognition, demonstrating its potential and effectiveness in the field of human behavior recognition. By combining the MediaPipe and OpenCV libraries, we were able to quickly implement the pose recognition task and achieve good performance in accuracy and efficiency.

In summary, this experiment shows that deep learning technology has a broad application prospect in human posture recognition, and can provide important support and application value for intelligent fitness, medical care and other fields.

4. Conclusion

Compared to traditional RGB data, multimodal RGB-D data has many advantages in terms of behavior analysis. RGB image data is easily affected by external factors unrelated to behavior, such as shooting environment, lighting, and the texture of the actor's clothing. It is a very challenging task to infer the skeleton posture, contour information and some key action information of the actor directly from the RGB video image, which makes many video analysis and behavioral action analysis techniques difficult to be widely used in real life. In contrast, in depth video images, because pedestrians and the surrounding shooting scene usually have a high degree of recognition, and the depth data obtained is not affected by clothing, it is more simple, convenient and accurate to obtain pedestrian contour and skeleton information. In addition, color information in RGB video can depict the apparent texture features of objects in more detail, which is particularly important when dealing with behaviors involving human interaction with objects. To sum up, multi-modal RGB-D data provides richer information for behavior analysis and is expected to promote the development and application of behavior and action analysis technology in practical applications.

Therefore, the application of deep learning technology in the field of human motion recognition provides a strong support for achieving more accurate and reliable motion recognition tasks. This technology can directly extract features from raw data, no longer rely on manually designed features, thus improving the recognition performance and generalization ability. By processing information on multiple time scales and introducing attention mechanisms, deep learning models are able to more fully understand and analyze the evolution of actions, improving the efficiency and accuracy of processing long time series. Deep learning technology has a broad application prospect and is expected to be widely used in fields such as surveillance, robotics and healthcare to provide more efficient solutions in these fields.

By combining deep learning techniques with multimodal RGB-D data, we can obtain richer information, providing more perspectives and possibilities for behavioral analysis. This comprehensive approach is expected to drive the development of behavior and motion analysis technology in practical applications, providing more possibilities for smart fitness, healthcare and other fields. Deep learning motion

recognition technology is expected to overcome the limitations of traditional methods, provide more possibilities for achieving more accurate and reliable motion recognition tasks, and provide important support for the development and progress of related fields.

Acknowledgment

In the process of writing this article, we were deeply inspired by the research of Xiang et al., especially their [4]"Machine Learning-Based Vehicle Intention Trajectory Recognition and Prediction for Autonomous Driving." work on intelligent e-commerce and other recommendations such as healthcare. By reading their papers, we gained many valuable insights that have had a profound impact on our research. We would like to give special thanks to Xiang et al and their team, as their work has provided us with a direction and a way forward. At the same time, we also want to thank a paper, their hard work and research results for the whole field to bring new thinking and inspiration. We sincerely thank them for their contributions and hope to continue to cooperate with them in this field in the future to jointly promote the development of intelligent recommendation systems.

References

[1] Akbar, A., Peoples, N., Xie, H., Sergot, P., Hussein, H., Peacock IV, W. F., & Rafique, Z. . (2022). Thrombolytic Administration for Acute Ischemic Stroke: What Processes can be Optimized?. McGill Journal of Medicine, 20(2).

[2] Xiao, J., Chen, Y., Ou, Y., Yu, H., & Xiao, Y. (2024). Baichuan2-Sum: Instruction Finetune Baichuan2-7B Model for Dialogue Summarization. arXiv preprint arXiv:2401.15496.

[3] Huo, Shuning, et al. "Deep Learning Approaches for Improving Question Answering Systems in Hepatocellular Carcinoma Research." arXiv preprint arXiv:2402.16038 (2024).

[4] Yu, Hanyi, et al. "Machine Learning-Based Vehicle Intention Trajectory Recognition and Prediction for Autonomous Driving." arXiv preprint arXiv:2402.16036 (2024).

[5] Xiang, Yafei, et al. "Text Understanding and Generation Using Transformer Models for Intelligent E-commerce Recommendations." arXiv preprint arXiv:2402.16035 (2024).

[6] Zhu, Mengran, et al. "Utilizing GANs for Fraud Detection: Model Training with Synthetic Transaction Data." arXiv preprint arXiv:2402.09830 (2024).

[7] Gong, Yulu, et al. "Enhancing Cybersecurity Resilience in Finance with Deep Learning for Advanced Threat Detection." arXiv preprint arXiv:2402.09820 (2024).

[8] Zhang, Y., & Zhang, H. (2023). Enhancing robot path planning through a twin-reinforced chimp optimization algorithm and evolutionary programming algorithm. IEEE Access.

[9] Chen, Jianhang, et al. "One-stage object referring with gaze estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[10] Duan, Shiheng, et al. “Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model Based on Machine Learning”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 132-6, https://doi.org/10.54097/ObMPjw5n.

[11] Chen , Jianfeng, et al. “Implementation of an AI-Based MRD Evaluation and Prediction Model for Multiple Myeloma”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 127-31, https://doi.org/10.54097/zJ4MnbWW.

[12] “Implementation of Computer Vision Technology Based on Artificial Intelligence for Medical Image Analysis”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 69-76, https://doi.org/10.62051/ijcsit.v1n1.10.

[13] Cai, Guoqing et al. “A deep learning-based algorithm for crop Disease identification positioning using computer vision.” International Journal of Computer Science and Information Technology (2023): n. pag.

[14] “Machine Learning Model Training and Practice: A Study on Constructing a Novel Drug Detection System”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 139-46, https://doi.org/10.62051/ijcsit.v1n1.19.

[15] “Unveiling the Future Navigating Next-Generation AI Frontiers and Innovations in Application”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 147-56, https://doi.org/10.62051/ijcsit.v1n1.20.

[16] W. Sun, W. Wan, L. Pan, J. Xu, and Q. Zeng, “The Integration of Large-Scale Language Models Into Intelligent Adjudication: Justification Rules and Implementation Pathways”, Journal of Industrial Engineering & Applied Science, vol. 2, no. 1, pp. 13–20, Feb. 2024.

[17] Zhou, Yanlin, et al. "Utilizing AI-Enhanced Multi-Omics Integration for Predictive Modeling of Disease Susceptibility in Functional Phenotypes." Journal of Theory and Practice of Engineering Science 4.02 (2024): 45-51.

[18] Q. Cheng, M. Tian, L. Yang, J. Zheng, and D. Xin, “Enhancing High-Frequency Trading Strategies with Edge Computing and Deep Learning”, Journal of Industrial Engineering & Applied Science, vol. 2, no. 1, pp. 32–38, Feb. 2024.

[19] Liang, Penghao, et al. "Enhancing Security in DevOps by Integrating Artificial Intelligence and Machine Learning." Journal of Theory and Practice of Engineering Science 4.02 (2024): 31-37.

[20] Chen, J. (2022). The Reform of School Education and Teaching Under the “Double Reduction” Policy. Scientific and Social Research, 4(2), 42-45. (Feb 2022)

[21] Ji, Huan, et al. "Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising." Academic Journal of Science and Technology 9.2 (2024): 215-220.

[22] Zhang, Chenwei, et al. "SegNet Network Architecture for Deep Learning Image Segmentation and Its Integrated Applications and Prospects." Academic Journal of Science and Technology 9.2 (2024): 224-229.

[23] Wang, Yong, et al. "Autonomous Driving System Driven by Artificial Intelligence Perception Fusion." Academic Journal of Science and Technology 9.2 (2024): 193-198.

[24] Zhou, Y., Osman, A., Willms, M., Kunz, A., Philipp, S., Blatt, J., & Eul, S. (2023). Semantic Wireframe Detection.

[25] Zhang, Y., Gono, R., & Jasiński, M. (2023). An Improvement in Dynamic Behavior of Single Phase PM Brushless DC Motor Using Deep Neural Network and Mixture of Experts. IEEE Access.

[26] Qian, Wenpin, et al. "Clinical Medical Detection and Diagnosis Technology Based on the AlexNet Network Model." Academic Journal of Science and Technology 9.2 (2024): 207-211.'

[27] Zhang, Quan, et al. "Application of the AlphaFold2 Protein Prediction Algorithm Based on Artificial Intelligence." Journal of Theory and Practice of Engineering Science 4.02 (2024): 58-65.

[28] Wang, H., Bao, Q., Shui, Z., Li, L., & Ji, H. (2024). A Novel Approach to Credit Card Security with Generative Adversarial Networks and Security Assessment.

[29] Wu, Jiang, et al. "Case Study of Next-Generation Artificial Intelligence in Medical Image Diagnosis Based on Cloud Computing." Journal of Theory and Practice of Engineering Science 4.02 (2024): 66-73.

[30] Zhu, Mingwei, et al. "Enhancing Collaborative Machine Learning for Security and Privacy in Federated Learning." Journal of Theory and Practice of Engineering Science 4.02 (2024): 74-82.

[31] Zhang, Y., Abdullah, S., Ullah, I., & Ghani, F. (2024). A new approach to neural network via double hierarchy linguistic information: Application in robot selection. Engineering Applications of Artificial Intelligence, 129, 107581.

[32] Yang, Le, et al. "Research and Application of Visual Object Recognition System Based on Deep Learning and Neural Morphological Computation." International Journal of Computer Science and Information Technology 2.1 (2024): 10-17.

[33] Qian, Jili, et al. "A Liver Cancer Question-Answering System Based on Next-Generation Intelligence and the Large Model Med-PaLM 2." International Journal of Computer Science and Information Technology 2.1 (2024): 28-35.

[34] Bao, Qiaozhi, et al. "Exploring ICU Mortality Risk Prediction and Interpretability Analysis Using Machine Learning." (2024).

[35] K. Xu, X. Wang, Z. Hu and Z. Zhang, "3D Face Recognition Based on Twin Neural Network Combining Deep Map and Texture," 2019 IEEE 19th International Conference on Communication Technology (ICCT), Xi'an, China, 2019, pp. 1665-1668, doi: 10.1109/ICCT46805.2019.8947113.