Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 6
ISET 2024

Research on the Detection Algorithm of Elderly Falling Behavior Based on AlphaPose Model

Yang Chen #, Haitao Xu #, Jingfeng Li # Cihai Qiu *, Yitian Zhang

School of Mathematical, Physical and Computational Sciences, University of Reading, Whiteknights Rg67BE, UK

* Corresponding author: wv804542@student.reading.ac.uk

Abstract. This study aims to address the challenge of quickly and accurately detecting high-risk behaviors such as falls in elderly people in hardware environments with limited power consumption and cost. Therefore, this article proposes an optimized behavior anomaly detection algorithm based on the AlphaPose model. The algorithm first improved the human target detection and pose estimation models to achieve faster pedestrian detection and pose joint point inference; Then, using the adjusted AlphaPose model, efficiently obtain the image coordinate data of each joint point in the human body; Finally, by analyzing the relationship between the linear velocities of the human head and hip joints, as well as the changes in the angle between the vertical line of the human center and the X-axis of the image, it is determined whether a fall event has occurred. This study deployed the proposed algorithm on the JetsonNano embedded development board and compared its performance with mainstream fall detection algorithms based on human posture, including YOLOv3+Pose, YOLOv4+Pose, YOLOv5+Pose, trtpose, and NanoDet+Pose. The experimental results show that on the set embedded platform, when the image resolution is 320×240, this algorithm achieves a detection rate of 8.83 frames per second and an accuracy of 0.913, both of which are superior to the compared algorithms. In summary, this algorithm has high real-time performance and accuracy, and can effectively identify the falling behavior of elderly people in a timely manner.

Keywords: Real Time Fall Detection, Posture Joint Points, Embedded Platform, Deep Learning.

1. Introduction

Accurate detection of falls in elderly individuals is of great significance for timely treatment, which can significantly reduce the risk of fatal injury, especially given that by 2020, China's elderly population aged 60 and above will exceed 230 million, making it the country with the largest elderly population in the world. Falling is the leading cause of fatal injury in elderly people over 65 years old. Detection errors directly affect the effectiveness of timely interventions, which is crucial for reducing mortality and ensuring the safety of the elderly [1]. However, traditional detection methods have limitations in terms of accuracy and speed, often leading to delayed or missed detection of falls, thereby reducing their effectiveness. This article proposes an abnormal behavior detection algorithm based on the AlphaPose optimization model, aiming to quickly and accurately detect falls in the elderly. The algorithm enhances the optimization of pedestrian object detection and pose estimation models, swiftly obtaining the coordinates of human pose joint point images [2], and combines the characteristics of the fall process to determine the occurrence of falls. By innovatively analyzing the relationship between the linear velocities of key joints and the changes in the angle between the human body's vertical axis and the X-axis of the image, the algorithm further improves detection accuracy and efficiency.

2. Research on the Algorithm of Human Posture Theory

2.1. AlphaPose

AlphaPose is a deep learning model used for human pose estimation, aimed at detecting and tracking key points of the human body (such as the head, shoulders, elbows, and knees) from images or videos, in order to parse human poses and movements. This model is widely used in fields such as human behavior analysis, athlete training, and security monitoring. AlphaPose is based on deep neural network technology and can accurately detect and track human keypoints in complex scenes, achieving efficient pose estimation. Although deep learning models typically require a large amount of computing resources, AlphaPose is designed with real-time requirements in mind and can achieve real-time human pose detection and tracking to a certain extent. In addition, AlphaPose has the ability to detect and track multiple individuals simultaneously, not limited to a single individual. By increasing the dataset and improving the model structure, the performance and functionality of AlphaPose can be continuously improved.

2.2. Convolutional Neural Network

Convolutional Neural Network (CNN) [3] is inspired by the mammalian visual system and simulates a series of receptive fields contained in the visual cortex. These receptive fields are small and sensitive visual regions that act as local filters and are suitable for processing images with local correlations. The receptive field is commonly defined as the mapping area of pixels on the feature map output by each layer in a convolutional network in the input.

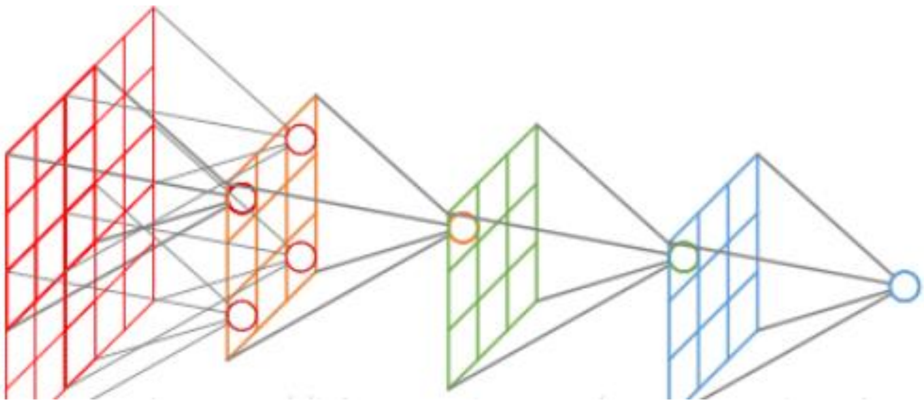

Figure 1. Schematic diagram of convolution process

As shown in Figure 1, the 3×3 matrix in the figure is the convolution kernel. For an N × N image, linear filtering is required for processing and analysis. A matrix can be used to process the image, and the output pixel is the product of the receptive field pixel and the corresponding element of the filter matrix. The stride in convolution operations is generally 1, and the size of the convolution kernel is usually 3×3 or 5×5. Odd convolution kernels have the concept of radius. The sum of all elements in the convolution kernel should be equal to 1 to maintain the brightness of the image after convolution; If the sum of elements is greater than 1, the convolved image will be brighter than the original image, otherwise it will become darker. After convolution, if each pixel value exceeds 255, it is directly truncated to 0 or 255. For black and white images, each pixel has only one channel and the grayscale value range is 0-255; For color images, each pixel has three channels to determine its RGB value.

The loss function (objective function) [4] is used to calculate the difference between the predicted value and the label value, while the optimizer is used to adjust the learning rate and momentum. Taking the simulation of linear equations as an example, W1 is the weight of the first iteration, and the thick line represents the real equation. The goal is to make W approach the true value through continuous iteration. The vertical line in the figure represents the difference between the predicted Y and the true Y after three iterations. In two-dimensional data, the difference value can be directly subtracted, while in multidimensional data, different formulas such as squared difference and mean square deviation need to be used to calculate the loss function. The loss function can be used to establish criteria for measuring the quality of decision functions and to evaluate the risk of decision functions using all training samples.

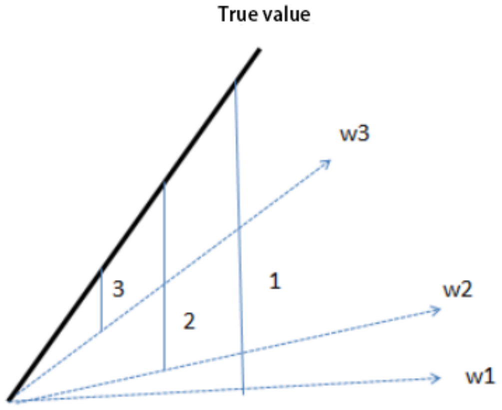

Figure 2. Loss Function Analysis Diagram

From Figure 2, it can be seen that the loss function analysis diagram assumes that the risk function $R(\theta)$ is calculated on known training samples (empirical data), as shown in expression (1):

$$R(\theta)=\frac{1}{N}\sum_{i=1}^{N}L(y^{(i)},\ f(x^{(i)},\ \theta)) \tag{1}$$

We need to find a value (θ) that minimizes the risk function. Common loss functions include 0-1 loss function:

$$L(y,\ f(x,\theta))=\begin{cases}1\\0\end{cases} \tag{2}$$

Square loss function $L(y,\hat{y})=(y-f(x,\theta))^2$. For classification problems, the cross entropy function is used

$$f_i(x,\theta)\in[0,1],\sum_{i=1}^{c}f_i(x,\theta)=1,\ f_i(x,\theta) \tag{3}$$

Compared to the likelihood function of the true category, the commonly used minimization method is the cross entropy loss function, which minimizes the negative logarithmic likelihood loss function.

2.3. AlphaPose human pose estimation

In order to ensure the efficiency and reliability of the fall detection algorithm, this study further developed the inference acceleration of the human pose estimation model and the design of the fall detection algorithm based on AlphaPose's high accuracy. Through this approach, it is hoped that the efficiency and accuracy of fall detection can be improved, enabling this technology to play a greater role in practical applications [5], as shown in Table 1.

Table 1. Performance Comparison of Mainstream Human Pose Detection Models Unit:%

Model	AP@0. 5：0. 95	AP@0. 5	AP@0. 75
OpenPose	61.73	84.87	66.97
Detectron	66.97	88.23	72.98
AlphaPose	73.29	89.33	79.23

3. Human fall detection method

The aim of this study is to develop a human fall detection method, which includes two core aspects: firstly, to accelerate and optimize the operation of the posture estimation model to achieve fast detection of human targets and efficient inference of joint positions, thereby ensuring that the algorithm can run with low latency and high throughput on embedded platforms; Secondly, using the obtained coordinate data of human joint point images, an algorithm was constructed to determine whether an individual has fallen. This algorithm integrates the instantaneous posture changes of the human body during the fall process and the short-term stability characteristics of the state after the fall [6].

3.1. Inference acceleration optimization of attitude estimation model

The AlphaPose model adopts a staged attitude estimation method. Therefore, the work of model inference acceleration mainly revolves around the optimization of human object detection models and posture joint detection models. After accelerating the inference of the human target detection model, we serialized the human target images to improve the efficiency of data exchange between models [7].

The original AlphaPose model used YOLOv3 [8] for human object detection. With the launch of YOLOv4, its detection accuracy and speed are significantly better than YOLOv3, and it can adapt to more complex detection environments (such as different lighting conditions, occlusion, etc.) [8]. Therefore, this study chose the lightweight model YOLOv4-tiny-416 (with a parameter size of 24.3MB) from the YOLOv4 series as the optimization object, achieving a good balance between detection accuracy and speed.

3.2. Human Fall Detection Algorithm

In daily life, an individual's falling process usually follows a pattern of transitioning from a walking or standing state to a falling process, ultimately reaching a falling state. Due to the rapid and brief nature of the fall process, it is difficult to capture the characteristics of falls during this stage. Falling incidents in specific populations, especially the elderly, are often accompanied by prolonged periods of falling, and their characteristics are more obvious and easy to capture. Therefore, achieving rapid and accurate detection of fall status during the process of falling has significant practical value.

The optimized and accelerated human pose estimation model can quickly and accurately locate the position information of the main joints of the human body, as shown in the 18 key pose nodes in Figure 3 (a). With the help of these posture joint point data, a specialized judgment algorithm for human falls was constructed. This algorithm combines the posture change characteristics at the moment of falling with the stability characteristics of the state in a short period of time after falling to monitor and identify fall events. Specifically, this algorithm includes: (1) instantaneous fall feature analysis, which identifies through the mathematical relationship between the linear velocity of the head joint points and the linear velocity of the hip joint points; (2) Analysis of fall state characteristics, based on the change in the angle between the human body's midline and the X-axis of the image, further subdivided into the angle θ u between the upper body's midline and the X-axis and the angle θ d between the lower body's midline and the X-axis. Figures 3 (b) and (c) show the relevant fall feature analysis process. The concept of instantaneous fall characteristics originates from the physical dynamics of a person during the initial fall, where the ankle position usually does not move significantly, and the rotation of the human torso around the ankle results in points far from the center of the circle having greater linear velocity. As for the characteristics of falling states, the balance characteristics of the human body when standing naturally are utilized. When the angle between the human torso and the horizontal plane exceeds a certain threshold, it indicates a high risk of falling[9].

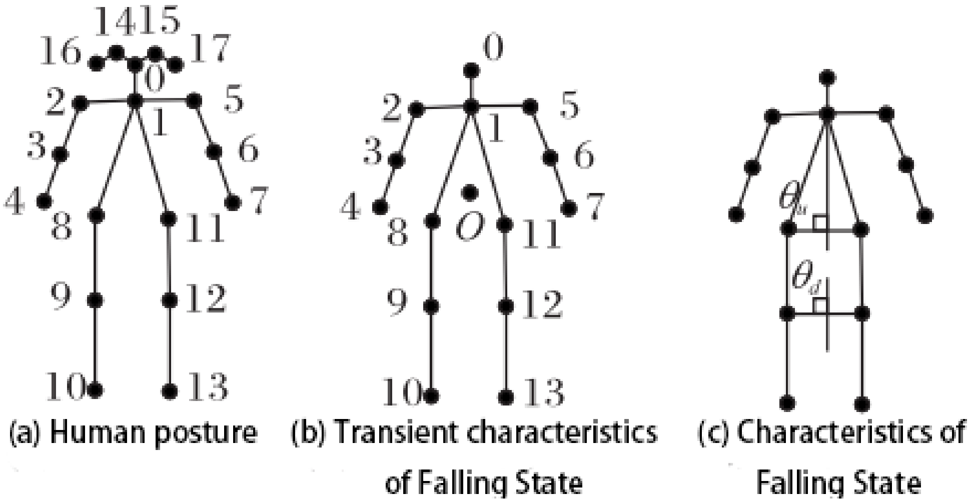

Figure 3. Human posture diagram and fall characteristics

4. Experimental Results and Analysis

4.1. Quantitative analysis

To verify the performance and accuracy of the algorithm proposed in this study, it was compared with other advanced human pose estimation and object detection algorithms, including YOLOv3 spp-416+Pose, YOLOv4 tiny 416+Pose, trtpose, YOLOv5s+Pose, and NanoDet-m+Pose. The evaluation indicators mainly focus on the detection frame rate and detection accuracy, and count the number of missed and false detections in the test videos. In addition, this article also analyzed the contributions of instantaneous fall characteristics and fall state characteristics to fall event detection.

The accuracy is calculated by the ratio P=Tc/Ta, where Tc represents the number of videos correctly detected as falls, and Ta is the total number of videos in the test video set. Missed detection refers to the number of fall events that were not identified in the test video, while false detection refers to the number of non fall behaviors incorrectly identified as falls.

According to the comparative data in Table 2, it can be observed that the algorithm proposed in this study is significantly better than the control algorithm in terms of detection frame rate and accuracy. Specifically, the algorithm achieved a high-speed processing capability of 8.83 frames per second, significantly improving real-time performance. At the same time, the algorithm performs excellently in reducing the number of false positives and missed detections, with an accuracy of over 91%, which to some extent confirms its applicability in practical application scenarios. Although YOLOv5s+Pose is close in accuracy to the algorithm proposed in this study, YOLOv5s has a slower detection speed on the Jetson Nano platform, making it difficult to meet real-time requirements. Although other control algorithms have higher accuracy in some cases, they are also limited by lower detection frame rates and cannot balance timeliness [10].

In terms of human pose estimation, trtpose adopts a bottom-up method to detect keypoints, but its overall performance is unsatisfactory, with accuracy ranking last among the compared algorithms. However, the detection frame rate of trtpose is higher than YOLOv3-spp-416+Pose, YOLOv5s+Pose, and NanoDet-m+Pose.

From the statistical data, it can also be found that compared to instantaneous fall features, fall state features have more effective triggering times, indicating that fall state features are more effective in detecting fall events and have higher practical value.

Table 2. Quantitative analysis of algorithm performance using fall test videos

Comparison algorithm	Detecting frame rate (frame/s)	Accuracy (P)	Number of missed detections	Number of false positives	Real time performance	Timeliness	Effectiveness
This article's algorithm	8.83	91.3%	least	least	have	have	high
YOLOv3-spp-416+Pose	-	-	-	-	nothing	nothing	-
YOLOv4-tiny-416+Pose	-	-	-	-	nothing	nothing	-
trt_pose	higher than YOLOv3-spp-416+Pose、YOLOv5s+Pose、NanoDet-m+Pose	minimum	-	-	have	nothing	low
YOLOv5s+ Pose	Slower than other comparison algorithms on JetsonNano	Approaching the algorithm presented in this article	-	-	nothing	nothing	-
NanoDet-m+Pose	-	-	-	-	nothing	nothing	-

4.2. qualitative analysis

During the experiment, five types of missed detections were observed. Figure 4 shows a detailed analysis of five representative cases. Among them, the circular symbol represents the detected fall event The first scenario involves the test subject being in a curled up falling state, with their body area close to the edge of the image, their arms and legs obscured by the body, resulting in severe deformation of the human body image and loss of the head. These factors collectively affect the extraction and recognition of human features, resulting in the inability of the proposed algorithm, YOLOv3 spp-416+Pose, YOLOv4 tiny 416+Pose, and YOLOv5s+Pose to detect human targets and perform pose estimation. Although NanoDet-m+Pose can recognize the human body, the confidence of the joint points is low, and the detected posture does not match the actual situation, resulting in incorrect detection (as indicated by the arrow). Trtpose can only recognize the left shoulder joint point, with insufficient information, and is considered invalid detection In the second scenario, the lighting distribution is uneven, and the test subject falls in a lateral position. The head and arms are partially covered by the body, and the area is dimly lit, resulting in unclear gradient features of the human body. The legs are further covered by tables and chairs, making it difficult to detect human targets, and the reference algorithms cannot recognize them properly In the third scenario, the tester is dressed in black clothing and adopts a side lying posture, with legs glued together and missing ankle images, resulting in a reduction in body contour features. The algorithm proposed in this article and other comparative algorithms have failed to continuously recognize human targets and cannot stably perform fall detection. Although trtpose can recognize the main joint points above the knee joint, there is a significant difference from the actual posture, and the joint point positions are offset, which is considered invalid detection. NanoDet-m+Pose failed to detect human targets, resulting in posture detection failure In the fourth scenario, when the tester lies on their side, their head and arms cover themselves, their legs are covered by a table, and the human body image features are not obvious. In addition, the background color is similar to the color of the clothes worn by the personnel, and the image contrast is low, which is not conducive to human target detection. All comparison algorithms failed to detect human targets properly The last scenario is when the tester falls in a prone position with their legs covered by their body. Due to the low shooting height of the camera, the human body area in the image presents a distortion effect of "near large, far small", making human target detection difficult. YOLOv3 spp-416+Pose only successfully detected human bodies in some frames of the video, and the detection results were unstable. The detected human body posture had a significant deviation from the true posture, resulting in low credibility. Other comparative algorithms failed to detect human targets properly.

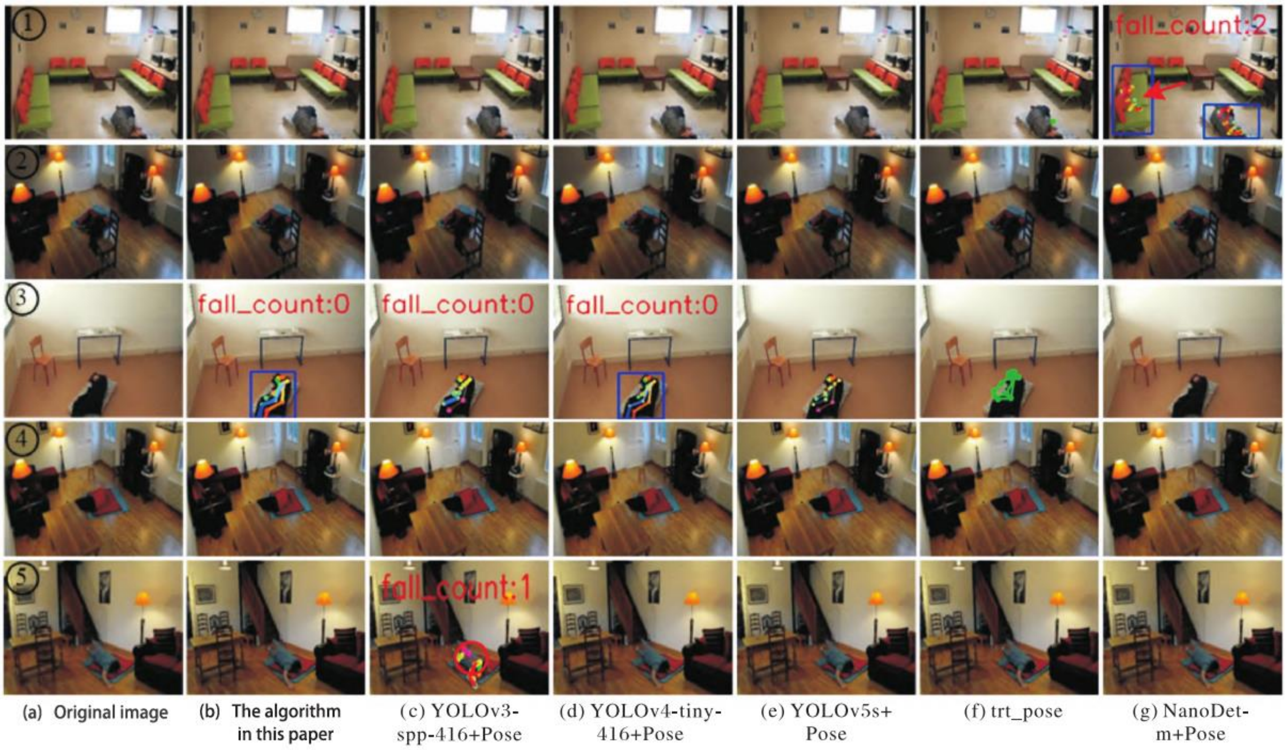

Figure 4. Comparison of Test Results of Different Algorithms

5. Conclusions

In summary, in order to promote the practical application of fall detection systems and improve the behavioral safety level of elderly people in home environments or independent activities, this study proposes an abnormal behavior detection algorithm based on the AlphaPose optimization model. This study achieved an efficient fall detection algorithm by optimizing the human detection and posture estimation models. This algorithm can be quickly deployed on cost-effective and energy-efficient embedded devices. The experiment has verified the excellent performance of the algorithm in real-time and accuracy, especially suitable for rapid online monitoring of elderly people's fall behavior, demonstrating significant application potential. However, after comparative analysis, this technology still has limitations:

(1) Performance degradation under non-uniform lighting conditions; Variations in lighting, such as day-night changes or indoor adjustments, can degrade image quality, affecting the detection algorithm's accuracy in recognizing human contours and posture, and complicating fall detection.

(2) Reduced efficiency at low monitoring angles:

Low camera angles can cause image distortion, especially with wide-angle lenses, leading to misinterpretation of posture and reduced fall detection efficiency.

(3) If the direction of the fall is parallel to the Y-axis of the image and the position is close to the midpoint of the X-axis, and the head coordinates exceed the ankle coordinates, it is difficult to distinguish between the falling and standing states.

When an elderly person falls in a direction parallel to the Y-axis of the camera's image and the position is close to the midpoint of the X-axis, the algorithm may struggle to distinguish between a falling and a standing state. This situation typically occurs when the fall motion aligns with the camera’s line of sight, and due to the head coordinates surpassing the ankle coordinates, the posture of the person may appear very similar to a standing position. This confusion is particularly evident in low-resolution or low-frame-rate images. Although this issue can be alleviated by increasing the number of cameras or using multi-angle monitoring, it also implies higher costs and more complex system deployment.

To enhance the robustness of the algorithm, future research will focus on the following aspects:

(1) Develop and integrate a lighting adaptive compensation mechanism to enhance the system's adaptability to lighting changes.

(2) Explore multi perspective camera collaborative detection strategies to alleviate occlusion problems caused by a single perspective.

(3) Given the current lack of video materials targeting abnormal behavior among the elderly in public datasets, this study plans to systematically collect relevant data, expand the dataset, and conduct experiments across a broader spectrum of human behavior and testing scenarios. These efforts aim to comprehensively evaluate and improve the performance of the proposed algorithm.

References

[1] Fang Hao-Shu, Li Jiefeng, Tang Hongyang, et al. AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time [J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022,1-10 .

[2] He Haoyang, Liu Guoliang, Zhu Xianglai, et al. Human Pose Estimation Based on Multiple Models Using Distributed 3D Camera Network Interaction [J/OL]. IEEE Sensors Journal, 2019,1-8 .

[3] Hao You, Ping Hu, Shirui Li, et al. Gradient-Aligned Convolutional Neural Network [J/OL]. Pattern Recognition, 2021,1-9 .

[4] Tan Matthias H. Y. A method for bounded loss function and feature function inversion [J/OL]. Quality Technology and Quantitative Management, 2016,1-10 .

[5] Liang Ruiyan, Yang Hui A lightweight fall detection algorithm framework based on RPApose and XJ-GCN [J/OL] Computer Applications, 2024,1-10 .

[6] Zhu Yan, Zhang Yaping A fall remote monitoring system based on LabVIEW and MATLAB [J] Computer Applications and Software, 2023, 40 (12): 2022-208

[7] Chen Chen, Xu Huiying, Zhu Xinzhong, etc Improved indoor pedestrian fall detection algorithm based on YOLOv8: FDW-YOLO [J/OL] Computer Engineering and Science, 2024,1-14

[8] Liu Baokai, He Fengjie, Du Shiqiang, et al. An advanced YOLOv3 method for small object detection [J/OL]. Journal of Intelligent & Fuzzy Systems, 2023,1-9 .

[9] Wang Xin, Yang Xiumei Personnel fall detection based on YOLOv5s and improved centroid tracking [J] Electronic Measurement Technology, 2023, 46 (24): 172-178

[10] Feng Mingxu Research on Key Issues in Multi time Series Classification, Recognition, and Modeling of Wearable Health Monitoring [D] Nanchang University, 2023