Research on an Elderly Indoor Fall Detection System Based on IoT and Computer Vision Technology: Integration and Performance Analysis of YOLOv5 and OpenPose

Yaokai Wang *

Xiamen institute of international education college, Xiamen University of Technology, Xiamen, China

* Corresponding Author Email: 1812010902@stu.hrbust.edu.cn

Abstract. The rapid growth of the global elderly population has led to an increased need for efficient and reliable fall detection systems to ensure timely medical assistance. Traditional monitoring methods, such as wearable devices and environmental sensors, often face challenges in accuracy, reliability, and user compliance. This study proposes an indoor fall detection system for the elderly, integrating Internet of Things (IoT) and computer vision technologies, specifically utilizing YOLOv5 and OpenPose algorithms. YOLOv5 is employed for rapid and accurate human body detection, providing position information by computing the center of objects in frames or videos. OpenPose is then used for detailed real-time pose estimation, detecting 135 key points on the human body—including hands, face, and feet—to assess whether a fall has occurred based on posture analysis. The system was tested using two datasets, GMDCSA and URFD, achieving sensitivity values of 0.9412 and 0.9583, respectively, indicating high accuracy in fall detection. The false positive rates were 0.0588 and 0.2857, while the false negative rates were 0.0588 and 0.0417, demonstrating the system's reliability in minimizing detection errors. The integration of YOLOv5 and OpenPose leverages their combined strengths in object detection and pose estimation, resulting in a robust solution suitable for real-time applications.

Keywords: Fall Monitoring, Yolov5, Openpose, Posture Assessment, Computer Vision Technology.

1. Introduction

The rapid growth of the global elderly population has led to an increased incidence of falls, which can result in serious health complications or even death. Timely detection of falls is crucial to provide immediate medical assistance and reduce the risk of further injuries [1]. Traditional fall monitoring methods often rely on smart wearable devices, vision sensors, and environmental sensors. However, these devices may suffer from accuracy and reliability issues, and the requirement for the elderly to carry devices can lead to non-compliance or loss, resulting in monitoring errors [2]. Therefore, there is a pressing need for more reliable and non-intrusive fall detection systems that can effectively monitor the elderly without the drawbacks of conventional methods.

Recent advancements in computer vision technology have opened new avenues for developing efficient fall detection systems. Algorithms like YOLOv5 have demonstrated high accuracy and fast detection speeds in recognizing human bodies in complex scenes, providing essential data for subsequent pose analysis [3] [4] [5]. OpenPose is another advanced deep learning framework capable of real-time human keypoint detection and full-body pose estimation, accurately capturing changes in human posture by recognizing multiple key points simultaneously [6]. Integrating these technologies can enhance the robustness and reliability of fall detection systems. Previous studies have explored various models for fall detection, such as trajectory-weighted convolutional features using modified VGG-16 networks [8], systems combining cameras and cloud servers for keypoint recognition [9], and LSTM-based techniques utilizing skeleton data from OpenPose [10]. However, challenges remain in handling complex scenarios like occlusions, varying lighting conditions, and overlapping individuals.

This study proposes an indoor fall detection system for the elderly by integrating Internet of Things (IoT) and computer vision technologies, specifically utilizing YOLOv5 and OpenPose algorithms.

The system first employs YOLOv5 for rapid and accurate human body detection, computing the center of objects in frames or videos to provide position information [7]. The detected human body regions are then analyzed using OpenPose for detailed real-time pose estimation, detecting 135 key points on the human body—including the face, hands, and feet—to assess whether a fall has occurred based on posture analysis. Parameters such as model size and input resolution are adjusted for both YOLOv5 and OpenPose, and error handling mechanisms are designed to improve robustness in complex scenarios like occlusions and multi-person overlaps [7]. The system was tested using two datasets, GMDCSA and URFD, achieving sensitivity values of 0.9412 and 0.9583, respectively, indicating high accuracy in fall detection. The study demonstrates that the integration of YOLOv5 and OpenPose can effectively enhance the accuracy and real-time performance of fall detection systems, offering a reliable solution for monitoring the elderly and potentially reducing the risk of injury or death due to falls.

2. Methodology

2.1. YOLOv5

YOLOv5 is very fast in detection while maintaining a high accuracy rate, and is able to process images from video streams in real time for application scenarios that require fast response [3]. Adaptive anchor frame computation will be performed at the input layer, which means that the network can automatically adjust the size and proportion of the anchor frame according to the content of the input image to better match targets of different sizes and shapes, in addition to fixed-size scaling [4]. YOLOv5 also supports adaptive image scaling, which means that the Internet can automatically adjust dimensions of the input image based on its aspect ratio to maintain the target's proportion and position in the image [5].

2.2. OpenPose

OpenPose is capable of detecting and tracking key elements of the human physical in real time, including the positions and postures of the head, arms, legs and other parts of the body, and is able to detect and track the postures of multiple people simultaneously in complex scenes, with a high degree of robustness and accuracy [6].

2.3. Integration Approach

Firstly, YOLOv5 is used for fast human body detection, and then OpenPose is used for detailed pose estimation of the detected human body region, and then the parameters such as model size and input resolution are adjusted for YOLOv5 and OpenPose, and for complex scenarios such as occlusion and multi-people overlapping, the corresponding error handling mechanism necessary to be designed to make better the robustness of the system [7].

Reference model.

Model 1: A drop detection system based on trajectory-weighted deep convolutional ranking Pool Descriptor (TDRD) using a slightly modified VGG-16 network to extract CNN feature maps from the input RGB video, the extracted feature maps are normalized using a spatio-temporal normalization method, and then improved dense trajectories are computed based on the input RGB video to obtain the trajectory attention maps; the trajectory attention maps can locate the human region [8]. Then, for each frame, a trajectory-weighted convolutional feature is computed from the extracted normalized feature map and the trajectory attention map. This trajectory-weighted convolutional feature maps the sequence of all frames and can give information about the dynamics of human movement. To reduce redundancy, cluster pooling is applied to the sequence of trajectory-weighted convolutional features. Finally, rank pooling is applied to the cluster pooled sequence to obtain the TDRD.

Model 2: The system consists of three components i.e. (i) two cameras, (ii) cloud server, and (iii) smartphone. The cameras are used to capture the video of the subject and convert it into a high

frequency image. Raspberry Pi 3 B-type board is used for the cameras [9]. The cloud server was used to process the data and the smartphone was used to receive alerts about falls. A deep neural network model is used to recognize 14 key points of the human body. The detected key point data map is fed into the deep neural network for fall detection.

Model 3: An LSTM-based fall detection technique for manufacturing. OpenPose is used to obtain skeleton data. Two datasets, URFD and SDUFall are utilized. LSTM with 2 layers of stacked and 256 layers of hidden features is used [10]. The raw skeleton data is processed to obtain the human body centerline coordinate (SHCLC) velocity. The raw data of SHCLC was used and more accurate results were obtained.

3. Analysis

Experimental Setup：First, human detection is performed using YOLOv5 to identify the position of the human body in the video or image, and second, pose estimation is performed using OpenPose to analyze the specific pose of the human body and to determine whether it is in a state of falling. The pre-trained YOLOv5 model is loaded and the video frames or images are preprocessed to fit the input requirements of the model (e.g., resizing, normalization, etc.). Process the YOLOv5 model to obtain the coordinates and confidence level of the human detection frame. Filter the detection results according to the confidence threshold and keep only the human body detection frames with high confidence. Load the pre-trained OpenPose model, and for the human body regions detected by YOLOv5, crop or re-localize them so that they are used as inputs to OpenPose. Then the OpenPose model is used to determine whether the human body is in a falling state by performing pose estimation, get the coordinates of the significant points of the every human body, and analyzing the positional relationships of the primary points, such as the relative position of the head to the ground and the posture of the limbs.

If the head position is detected to be significantly lower than it would be in normal standing, it may be a sign of a fall.2. If the posture of the limbs suggests that the body is out of balance (e.g., arms outstretched in an attempt to support the body), it may be in the process of falling.3. Combined with the posture information from the multiple frames, a fall event is confirmed if consecutive frames show a tendency to fall or if a fall has already taken place.

As show in the table 1. Two datasets, GMDCSA (created by conducting falls and ADL functions with individual subjects wearing different clothes) and URFD (containing a sequence of 70 (30 falls and 40 functions of daily life)) were called and correlation formulas were applied based on these two datasets:

Table1. Performance Metrics of the Fall Detection System on GMDCSA and URFD Datasets

Norm	Formulas	Datasets	
		GMDCSA	URFD
Sensitivity	TP / (TP + FN)	0.9412	0.9583
Specificity	TN / (FP + TN)	0.9416	0.7143
Precision	TP / (TP + FP)	0.9416	0.6571
False positive rate	FP / (FP + TN)	0.0588	0.2857
False Negative	FN / (FN + TP)	0.0588	0.0417
Accuracy	(TP + TN) / (P + N)	0.9412	0.8030
F1 Score	2TP / (2TP + FP + FN)	0.9416	0.7797

TP: Size of the samples that were actually positive and were anticipated to be positive.
TN: Size of the samples that were actually negative and anticipated to be negative.
FP: Size of the samples that were actually negative but anticipated to be positive (false positives).
FN: Size of the samples that were actually positive but anticipated to be negative (underreporting).

Note: After using GMDCSA and URFD datasets, the sensitivity reaches 0.9412 and 0.9583 respectively which are close to 1. The false positive rate is 0.0588 and 0.2857 respectively, and the false negative rate is 0.0588 and 0.0417 respectively, which are close to 0. This indicates that the final results are more accurate.

4. Application Scenarios Analysis

As a deep-rooted learning based target detection algorithm, YOLOv5 is known for its high accuracy, fast detection speed and easy deployment. It is able to accurately recognize target objects, including the human body, in complex scenes, providing basic data for subsequent pose analysis. OpenPose is an advanced deep learning framework for real-time human keypoint detection and full-body pose estimation. It is able to recognize and localize multiple key points of the human body simultaneously, such as the face, hands, feet, etc., so as to accurately capture the changes in the human body's posture. This is particularly important for fall detection, as falls are usually accompanied by significant changes in human posture. Falls occur in the home, especially for families with small children. By combining YOLOv5 with OpenPose, an intelligent fall monitoring system can be developed. The system can monitor video data in the home in real time, quickly identify children using YOLOv5, and subsequently perform key point localization and posture analysis using OpenPose. Once a fall is detected, the system can immediately send out an alert to notify parents or guardians to take emergency measures, thus effectively avoiding or minimizing injuries caused by falls. By installing a fall monitoring system based on YOLOv5+OpenPose in schools and other places where children gather, comprehensive monitoring of children's activity areas can be realized. The system is able to automatically detect and warn of falls, providing timely feedback to teachers and staff so that they can quickly intervene and take the necessary rescue measures. In addition, the system can be used to acquire and analyze fall data, providing a scientific basis for improving campus safety measures.

5. Conclusion

YOLOv5 is able to accurately detect the human body in most cases, but may misdetect or miss in extreme lighting conditions, occlusion, or dense crowds. OpenPose is able to estimate the postural keypoints better over the detected human body area, but is again affected by occlusion, complex backgrounds, or multiple overlapping people. Relying on head position alone may not be accurate enough, as head movements during a fall can be affected by multiple factors (e.g., hand support, etc.). Combining pose information from multiple frames can improve the accuracy of the judgment, but may also introduce delays and requires dealing with smoothing and filtering of the pose data. YOLOv5 quickly recognizes the human target and provides accurate input to OpenPose. OpenPose then analyzes the pose of the recognized the human bodies and extracts the positional particulars of the key points. Finally, by comparing and analyzing the positional changes of these key points, it is possible to accurately determine if a fall has occurred. The combined application of YOLOv5 and OpenPose exhibits high precision and real-time. The fast detection capability of YOLOv5 ensures that the system can process images from the video stream in real-time, whereas the precise gesture recognition capability of OpenPose ensures fall detection accuracy. In addition, the performance and system accuracy can be further improved by continuously optimizing the algorithm and model parameters. It is an effective and reliable method to realize fall detection through the fusion application of YOLOv5 and OpenPose. The method not only has high accuracy and real-time performance, but also can be widely used in a variety of scenarios that require fall detection, such as families, kindergartens, schools, hospitals and so on. In the future, with the continuous development and improvement of the technology it can be expected that the method will be applied and promoted in more fields.

Optimize the models of YOLOv5 and OpenPose to adapt to more complex scenes and lighting conditions, and use more advanced algorithms or techniques to deal with occlusion and multi-person overlapping problems, such as introducing an attention mechanism or using graph-convolutional networks to improve the robustness of detection. Consider multiple gesture features (e.g., angle and position of the head, limbs, and torso) to determine a fall, and introduce machine studying or deep-rooted studying models to automatically learn the features and patterns of a fall and optimize them based on the training data, as well as to consider preparatory actions (e.g., body imbalance, etc.) prior to a fall, and to warn of the risk of a fall in advance to improve the logic of the fall judgment. Optimize algorithms and models to reduce computation and memory usage, and use more efficient hardware or distributed processing systems to accelerate the process to improve real-time performance and efficiency.

References

[1] Alam E., Sufian A., Dutta P., Leo M. Vision-based human fall detection systems using deep learning: A review. Computers in Biology and Medicine, 146 (2022) 105626.

[2] Prateek M., Sharma D., Tiwari R., Sharma R., Kumar K., Kumar N. (Eds.). Next Generation Computing Technologies on Computational Intelligence: 4th International Conference, NGCT 2018, Dehradun, India, November 21–22, 2018, Revised Selected Papers. Vol. 922. Springer Nature, 2019.

[3] Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., Wang, R.: 'A Review of Key Technologies for Emotion Analysis Using Multimodal Information', Cognitive Computation, 2024, 1, (1), pp. 1 - 27.

[4] Paul P.K. Information technology and computing programs in the universities of West Bengal: An analytical case study. International Journal of Applied Science and Engineering, 7 (2) (2019) 135 - 144.

[5] Rahaman S. Use of Internet by the undergraduate and postgraduate students at the University of Gour Banga, Malda, West Bengal. Library Philosophy and Practice, (2021) 1 - 11.

[6] Zhu, X., Huang, Y., Wang, X., Wang, R.: 'Emotion recognition based on brain-like multimodal hierarchical perception', Multimedia Tools and Applications, 2024, 83, (18), pp. 56039 - 56057.

[7] Kwolek B., Kepski M. Human fall detection on embedded platform using depth maps and wireless accelerometer. Computer Methods and Programs in Biomedicine, 117 (3) (2014) 489 -501.

[8] Gao L., Sun J. Research on the application of 5G technology in the teaching and practice of health management of special populations. Teaching and Research Office of Health Management, Naval Medical University, Shanghai 200433.

[9] Zhu, X., Xu, H., Zhao, Z., Wang, X., Wei, X., Zhang, Y., Zuo, J.: 'An environmental intrusion detection technology based on WiFi', Wireless Personal Communications, 2021, 119, (2), pp. 1425 - 1436.

[10] Gaya-Morey F.X., Manresa-Yee C., Buades-Rubio J.M. Deep learning for computer vision-based activity recognition and fall detection of the elderly: A systematic review. Applied Intelligence, (2024) 1 - 26.