WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n2.01
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Falling Detection based on the Internet of Things

Zihao Lian *

College of mechanical and vehicle engineering, Changchun University, Changchun, 130022, China
*Corresponding Author: sukiwhisper1024@gmail.com

ABSTRACT

This article proposes a method for fall detection through computer vision. With the increasing aging population in China, falling has become a major form of hazard to the safety of the elderly. This method aims to utilize computer vision to detect and record falling behaviors. This article has designed a detection method using YOLOV5 as a tool. Compared to traditional detection methods, YOLO offers faster detection speed and more accurate detection accuracy, which will greatly enhance the speed of discovering falls among the elderly and facilitate timely rescue efforts.

KEYWORDS

Computer vision; Internet of Things; Fall detection

1. INTRODUCTION

With the development of technology, a series of human-machine interactions and artificial intelligence have gradually entered the lives of the general public. While enjoying the technological benefits of progress in our era, the health and elderly care issues of the elderly should also receive attention from society. In this age, young people are often busy and elderly individuals are left alone at home, inevitably leading to accidents. If accidents occur and are not promptly and properly handled, they can potentially lead to fatal consequences. However, with just one extra second, the outcome could be entirely different.

2. PREVIOUS RESEARCH

With the emergence of AI, there are now means to replace manual labor. The popularity of surveillance equipment and the improvement of detection accuracy have also made target detection more feasible in daily life.

In recent years, some researchers have proposed fall detection methods based on audio signals. For instance, Li's team detected falls using sound source phase methods [1]. Khan's team proposed comparing footstep signals with those of normal walking [2]. Yu's team proposed collecting sound waves from floors [3]. Fall detection designs based on audio signals are simple and easy to implement, but various interference in real life can easily lead to false detections.

Additionally, wearable devices have also been proposed for fall detection. Mathie's team compared collected acceleration data with preset thresholds [4]. Bourke's team used gyroscopes to measure pitch speed for detection [5]. Shibuya's team measured three-axis acceleration and angular velocity of the body to determine falls [6]. Fall detection using wearable devices has the advantages of high sensitivity and good real-time performance, but their high maintenance and manufacturing costs are

not suitable for the general population. Furthermore, interference from body movements, especially those involving the arms, can affect detection accuracy and lead to false detections.

The two previous detection methods require direct contact with individuals, which can easily lead to interference and have an impact on the daily lives of the elderly. In contrast, fall detection based on computer vision captures activity information through cameras and processes it using computers, compensating for the shortcomings of the above two methods. Additionally, it can monitor other abnormal situations. Many methods have been proposed both domestically and internationally for this approach.

Vaidehi's team designed an automatic human fall detection system by extracting features such as the aspect ratio and tilt angle of the human body. Zerrouki's team extracted human contour information from captured videos of human activities and trained a Markov model to distinguish between falls and non-fall behaviors [7]. Ma's team proposed using Kinect to extract human contours and construct an SVM to distinguish between daily activities and falls [8]. Stone's team proposed using Kinect to capture depth images of human motion and determining falls based on whether the trunk of the body is perpendicular to the ground [9]. Rougier's team proposed using a 3D monocular camera to track the head, calculate its movement in both horizontal and vertical directions, and compare it with preset thresholds to determine falls [10].

3. THE APPLICATION OF COMPUTER VISION INTELLIGENT ALGORITHMS

Computer vision encompasses three main tasks: classification, object detection, and instance segmentation. Among these, our focus is on object detection. Currently, the algorithmic models in computer vision are developing rapidly. For instance, YOLO, SSD, and Faster R-CNN are all excellent algorithms. In this article, we have adopted YOLO as the model for our detection algorithm.

YOLO is a commonly used object detection model. The essence of object detection is to locate the position of the desired object in an image or video and label its category. In the process of object detection, five key pieces of information need to be detected: the center position (x, y) of the object, its height and width, and its category. The principle of YOLO is to divide a photo or video into countless evenly sized and neatly arranged squares. It then detects the center of the object within these squares and utilizes the detected bounding boxes to perform a series of calculations and eliminations, ultimately presenting the answer.

4. YOLOV5

YOLOv5[11] is not the latest iteration in the series, but it possesses exceptional stability, which is why it has earned my preference. Object detection can be understood as a combination of object recognition and object tracking, and thus, it requires not only speed but also solid stability. However, like most things we are familiar with, achieving both is often a challenge, necessitating compromises.

Currently, object detection algorithms fall into two categories: One-stage, represented by YOLO, and Two-stage, represented by R-CNN. As their names suggest, Two-stage algorithms divide the detection process into two steps. First, they generate candidate regions (region proposals), and then classify these regions, often with fine-tuning of the locations. These algorithms typically achieve high recognition accuracy and low missed detection rates, but their downside is their slow speed, often limited to single-digit frames per second (Fps), making them impractical for video recognition.

Therefore, this project opted for a One-stage algorithm. Unlike Two-stage, One-stage algorithms do not require the generation of candidate regions. Instead, they directly output the probability of an object's category and its location coordinates. This single-step detection process allows them to

achieve speeds ranging from 45Fps to 155Fps, surpassing the 24Fps required for human eyes to perceive continuous images. Even with such speeds, YOLOv1 achieved a mean Average Precision (mAP) of 50, which is a remarkable accomplishment.

As mentioned earlier, to comprehensively evaluate the accuracy of object detection algorithms, this project needs a metric to represent it. This is where the concept of Intersection over Union (IoU) comes in. IoU is calculated as the ratio of the intersection area between the true value and the predicted value to their union area.

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{1}$$

Subsequently, this article introduce the concept of mean Average Precision (mAP), which represents the overall performance of the detection results. This metric is composed of Precision and Recall.

$$Pr\ e\ cision = \frac{TP}{TP+FP} \tag{2}$$

$$Re\ c\ all = \frac{TP}{TP+FN} \tag{3}$$

In this equation, TP (true positive) represents correctly identifying a positive instance as positive, FP (false positive) represents incorrectly identifying a negative instance as positive, and FN (false negative) represents incorrectly identifying a positive instance as negative. There is also TN (true negative), which is not included in the calculation, representing correctly identifying a negative instance as negative. In the context of object detection, achieving TP and FN can be considered as successfully achieving the objective, while FP and TN indicate errors. It's worth noting that while FN also counts as achieving the objective, the ultimate goal of object detection is to detect desired objects, rather than excluding unwanted ones, therefore FN is not included in the calculation.

In practical object detection tasks, this article also introduces the concept of confidence, which has two aspects. Firstly, it represents the probability of an object existing within the detection range. Secondly, it represents the potential IOU value in the case of an object being present within the range.

Through certain methods, which can plot the PR curve, calculate the AP value, and further derive the mAP value.

Compared to YOLOV4, the most significant upgrade in V5 lies in the existence of four different weight versions: YOLOV5l, YOLOV5m, YOLOV5s, and YOLOV5x. Here, I refer to the paper published by the Wang team, who found in their testing that YOLOV5s has the fastest speed, with a response time of 2.2 milliseconds and an FPS of up to 455. However, its mAP value is relatively low, around 14. On the other hand, YOLOV5x achieves an mAP of around 170, but its response time is 6 milliseconds, and its FPS is only around 170. Although there is a significant numerical difference, considering the usage scenarios, YOLOV5x is undoubtedly the most suitable choice. The reaction speed of humans, without training, is approximately 300 milliseconds, and even with training, the limit is generally no less than 100 milliseconds. Similarly, although the FPS of YOLOV5x is only around 170, it far exceeds the 24 frames recognizable by the human eye. Therefore, from a human perspective, it is difficult to distinguish the difference between YOLOV5s and YOLOV5x through one's own senses without technological means. However, the difference in mAP is evident, which is the reason for choosing YOLOV5x.

5. EXPERIMENTAL DESIGN AND RESEARCH

In our experiment, we prepared three sets of data as subjects, with one set used for training and the other two sets for actual detection. In the application of fall detection, we need to train the YOLO

model to target human fall actions. Firstly, we collected a large number of images related to falling behaviors, mainly focusing on falls in football or other sports out of ethical considerations. After annotation, we obtained a set of training datasets. Once we had enough images for model training, we could proceed with the detection tests.

5.1. Training of YOLO Model

Since YOLOv5 is open-source, it doesn't need to spend much time on debugging and can instead focus our efforts on annotation and training. In YOLO, when we input a photo, it automatically generates a grid of n*n square bounding boxes. After we annotate the targets, YOLO calculates the five key data points mentioned earlier based on the presence and weights of targets in each box: the center point coordinates (x, y), the width and height (w, h), and the target label. By guiding the YOLO model to learn these data points, it completes a round of training. With an increasing number of effective training iterations, it gradually achieves higher accuracy rates (mAP).

5.2. The Distinction Between Falling Behavior and Lying Down Behavior

During the model training process, we intentionally included the behavior of lying down, but it did not provide valuable positive feedback for the training. Therefore, after discussion, this article decided to introduce a new detection target.

We summarized the characteristics of human lying-down behavior. Firstly, when falling, humans may close their eyes due to pain. Secondly, the environment where lying-down occurs is different from that of falling. Lying down typically happens on sofas, beds, and other places, making it easy to capture items like beds, pillows, blankets, and other surrounding objects. These latter items are referred to as peripheral objects. We included features such as facial features and peripheral objects as detection targets. To visually analyze the detection data, this article proposed a calculation formula:

$$p(n) = a * f(n) - a * g(n) - a * h(n) \tag{4}$$

In this formula, this article introduce the sensitivity factor 'a'. In practical use and subsequent development, the value of 'a' can be adjusted. A higher value of 'a' makes the model more cautious, while a lower value makes it more sensitive. 'f(n)' represents the number of detected falling behaviors, typically 0 or 1. 'g(n)' stands for the number of detected facial features, and 'h(n)' for the number of detected surrounding objects.

Utilizing this formula, this article summarizes the detection results into a single data point, 'p(n)'. When this data point is greater than 0, it is considered a detected falling behavior, and a warning is issued. When the data falls between -1 and 0, it is regarded as a suspected falling behavior, and a warning is also triggered. When the data is less than -1, it is considered that no falling behavior has been detected, and no warning is issued.

5.3. Practical Application and Effectiveness Evaluation

In modern life, the frequency of home camera usage has increased, especially in households with elderly and young children. Combining the YOLO model with home cameras and adding an alarm system can play a role in reporting falls as soon as they are detected.

The core advantage of fall detection using computer vision is to offload the work of human monitoring, especially when connected to smart devices. This enables remote monitoring, remote operation, remote debugging, and remote maintenance, significantly reducing labor costs.

By adjusting the sensitivity, the detection system can meet the conditions of more users. Under high sensitivity conditions, any suspected fall behavior will be reported to prevent minor issues from

escalating. Under low sensitivity conditions, most useless detections can be filtered out. Additionally, while detecting, the entire system is gradually learning the user's habits and actual situations to complete self-upgrades.

6. CONCLUSION AND FUTURE RESEARCH

Given the current state of human development, whether scientifically or ethically, absolute trust in machine vision has not yet been achieved.

From a technical perspective, computer vision has relatively high environmental requirements. For instance, when designing formulas, we considered introducing more parameters to standardize the entire detection process. However, as annotations and training progressed, the more parameters there were, the more uncertainty arose. Eventually, after omitting some detection targets, the relatively reasonable formula mentioned earlier emerged.

While some may hold great hope for artificial intelligence, in practical use, AI rarely meets everyone's expectations. No matter how advanced the instrument is, there will always be some degree of malfunction or misjudgment. The question of whether AI can replace humans remains the biggest challenge facing AI in the future.

In the future, AI is bound to play a significant role in human life, especially in daily living. Computer vision has great potential and market due to its ability to liberate humans and its lower cost. For instance, by applying the method proposed in this article and connecting it with smart devices, we can expand many convenient functions.

REFERENCE

[1] Li, Y., Ho, K. C., & Popescu, M. (2012). A microphone array system for automatic fall detection. IEEE Transactions on Bio-medical Engineering/IEEE Transactions on Biomedical Engineering, 59(5), 1291–1301.

[2] Khan, M. S., Yu, M., Feng, P., Wang, L., & Chambers, J. A. (2015). An unsupervised acoustic fall detection system using source separation for sound interference suppression. Signal Processing, 110, 199–210.

[3] Shuo, Y., & Chen, H. (2017). Fall Detection with Orientation Calibration Using a Single Motion Sensor. In Springer eBooks, pp. 233–240.

[4] Mathie, M., Coster, A. C., Lovell, N. H., & Celler, B. G. (2004). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiological Measurement, 25(2), R1–R20.

[5] Bourke, A. K., & Lyons, G. (2008). A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Medical Engineering & Physics, 30(1), 84–90.

[6] Shibuya, N., Nukala, B., Rodriguez, A., Tsay, J., Nguyên, T., Zupancic, S., & Lie, D. Y. C. (2015). A real-time fall detection system using a wearable GAIT analysis sensor and a support vector machine (SVM) classifier. Eighth International Conference on Mobile Computing and Ubiquitous Networking.

[7] Vaidehi, V., Ganapathy, K., Mohan, K., Aldrin, A., & Nirmal, K. S. J. (2011). Video based automatic fall detection in indoor environment. In International Conference on Recent Trends in Information Technology (ICRTIT). IEEE.

[8] Ma, X., Wang, H., Xue, B., Zhou, M., Ji, B., & Li, Y. (2014). Depth-Based human fall detection via shape features and improved extreme learning machine. IEEE Journal of Biomedical and Health Informatics, 18(6), 1915–1922.

[9] Stone E E, Skubic M. (2014) Fall detection in homes of older adults using the Microsoft Kinect. IEEE journal of biomedical and health informatics, 19(1): 290-301

[10] Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2007). Fall detection from human shape and motion history using video surveillance. In International Conference on Advanced Information Networking and Applications Workshops (AINAW'07). IEEE.

[11] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE conference on computer vision and pattern recognition.