International Journal of Computer (IJC)

ISSN 2307-4523 (Print & Online)

http://ijcjournal.org/

The Elderly Fall Detection Algorithm Based on Human Joint Extraction and Object Detection

Haiguang Chen[a*], Susheng He[b], Mingxing Liu[c]

[a,b,c]*Shanghai Normal University, No.100 Haisi Road Fengxian District Shanghai, Shanghai 220000, China*

[a]*Email: chhg@shnu.edu.cn,* [b]*Email: 1319522956@qq.com,* [c]*Email: 1024585449@qq.com*

Abstract

Nowadays, the care of the elderly has become a social concern. The fall of the elderly has become one of the main factors threatening the health of the elderly. In this paper, we designed a fall detection algorithm based on human joint extraction and object detection.First,yolov4 was used to identify and detect the elderly. Then openpose was used to detect the human joint. Based on the human joint, this paper using Random Forest to classify the status of the elderly, there are three states of the elderly: falling down, lying down and other states. In the detection of a single old man, the accuracy of the model reached 99.3%, the sensitivity and specificity of the model reached 79.3% and 72.1%.

Keywords: Yolov4; Openpose; Random Forest; Human joint Extraction; Fall detection.

1. Introduction

According to the 2019 World Population Outlook [1], about 22.3% of the world's population is over the age of 65, and this rate will increase in future. Therefore, the safety of the elderly has become a more and more important problem. In the real scene, there are many problems that may lead to the fall of the old man. The old man cannot get help in time after the fall, and in serious cases, the old man may die. According to statistics, due to the elderly's body organs and other reasons, about 60% of the elderly have relatively serious consequences due to falling down, and about 10% of the elderly have died as a result of falling down. Although many falls can be prevented and avoided, such as outdoor activities try to avoid bumpy roads. The road, wear non-slip shoes, stand up slowly when sitting for a long time, etc. But there are no such ways. It's foolproof. After a fall, if you can do it to an elderly person who falls within an hour. Effective treatment can reduce the cost of treatment by nearly 60 percent. Therefore, in addition to effective prevention education, it is also important to monitor falls effectively.

---

* Corresponding author.

ISt is very important to design an algorithm for elderly fall detection with high sensitivity and specificity rate. Besides, the algorithm in this paper can run in Google colab, so it is more convenient. The basic flow of the algorithm was shown in figure 1. Our contribution are summarized as follows:

- We develop an efficient model to detect the falling down of the elderly.
- We can run the model in Google Colab instead of complex running environment. It is convenient for us to recurrence the result.

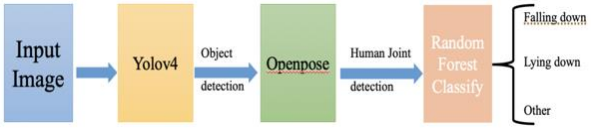

Figure 1: The basic flow of the algorithm

2. Related Work

In recent years, there have been a lot of papers on the elderly fall detection. According to literature, there are mainly three detection algorithms, namely, the wearable device based fall detection algorithm, the environment-based detection algorithm and the computer vision-based fall detection algorithm. The fall detection algorithm based on wearable devices mainly relies on some sensors. Currently, the most widely used fall detection sensors are accelerometers and gyroscopes, which are embedded into smart watches and smart phones. He [2] and his colleagues used accelerometer, gyroscope and Bluetooth module to embed these sensors on the circuit board, put the circuit board on the clothes of the old man, detect the daily acceleration and angular velocity of the old man, send data back to the mobile phone terminal through the Bluetooth module, and then use KNN algorithm to judge whether the old man falls or not. Based on this, Tamura [3] and his colleagues added airbags to the fall detection system, effectively reducing the risk of falls in the elderly. Lai [4] and his colleagues use multiple three-axis accelerator fall detection to the human body, the sensors distributed in the vulnerable part of the key points of the body, the model will use the sensor to obtain information, and the information transmission, this model also can be judged when more than normal acceleration, fall the probability value of events at the same time, the model can also detect fall caused by the level of consequences. Although wearable fall detection devices can detect falls of the elderly to a certain extent, there are two caveats because the system relies on wearable devices. First, some old people are not willing to carry the equipment, but the equipment may be damaged for other reasons and the data report may be wrong. Environment-based fall detection algorithm is also called scenario-based fall detection algorithm, which mainly uses the sound and vibration produced by falls to detect falls. Doukas and his colleagues proposed to detect falls in elderly people based on the acceleration information in the video stream. The idea of this algorithm is that when the acceleration in the vertical downward direction suddenly increases, the person will rapidly change from a normal state to a falling state, so that a fall event can be detected. Alwan [5] and his colleagues devised a fall detection system based on the ground vibration, the vibration frequency of the system through the ground to detect fall event occurs, the main ideas of the algorithm is considered as a result of normal walking vibration frequency and falls down the ground vibration frequency, are connected to the floor using a piezoelectric sensor to get the vibration information.

Toreyin [6] and his colleagues used sensors to perform wavelet transform on sound, vibration, infrared and other data, and then extracted features. Then, they used samples in normal state (positive samples) and samples in fall state (negative samples) to conduct HMM training, and finally integrated all information for fall detection. Therefore, this kind of algorithm has many inconvenient factors in the actual application scene, just like the fall detection algorithm based on wearable devices. Compared with the previous two algorithms, the advantages of the computer vision-based fall detection algorithm are obvious. First of all, we don't need to ask the elderly to carry other sensors or devices to sense the state of a fall. Secondly, there are now cameras almost everywhere people are seen, making computer-vision-based fall detection more widely used. Wang [7] and his colleagues proposed to extract video frames from different video sequences and use computer vision technology to carry out fall detection algorithm. The algorithm uses PCANet trained by different video frames to obtain the prediction results of each frame according to the PCANet model (including standing, falling, and falling). Finally, the prediction results of the continuous frame PCANet model are combined with SVM for fall detection. Wang kun [8] is proposed using multistage fall behavior identification of the SVM discriminant model, the specific algorithm is as follows: the first level vector machine (SVM) is used to identify the fall behavior of single frame image, the second vector machine (SVM) is used to identify more consecutive frames of motion behavior recognition, finally through the two level vector machine (SVM) to get the final result of the fall behavior of judgment. Yanran [9] and his colleagues designed a fall detection algorithm based on the Kinect network. The core idea is to use Kinect to extract the skeleton of the human body, and then use the motion characteristics of the center of the spine to recognize the fall. Secondly, there are many algorithms that use the method of human joint point extraction to detect falling behavior. But in general, these algorithms still have some problems. For example, considering the way of human joint point extraction, fall recognition will extract some key points from non-character objects. This is because the background of the characters is more complicated, which makes the computer unable to Identify the problem. The elderly fall detection algorithm based on human joint points proposed in this paper first uses the Yolov4 model that is more effective in the field of target detection to detect the elderly, and then uses the current Openpose model that is more effective in the field of joint detection to perform human skeleton joint The point is extracted, and then the ensemble learning-random forest algorithm is used to classify the human pose. Therefore, the method proposed in this paper can effectively solve the problems encountered by the aforementioned algorithms.

3. Algorithm

In the process of fall detection algorithm, people's state is generally classified. Fan [10] and his colleagues divided people into four states: standing, falling, already falling, and not moving. This paper also adopts the idea of state classification. This paper divides the state of the elderly into three states: falling state, lying down state and other states. The reason for this design idea is that we mainly need to detect the falling state of the elderly. Considering the flat state, it is because the normal flat state and the flat state will occur after falling. Therefore, it is necessary to judge whether the flat state is after falling according to the relevant information of the video frames before and after falling. This method further enhances the fitting degree of the model. In this paper, Yolov4 was first used to detect the staff. After the elderly were detected, Openpose was used to extract the nodes of the elderly in the image, and then random forest algorithm was used to classify the extracted features.

3.1 Yolov4 object detection

Yolov4 is a target detection algorithm improved by Alexey and his colleagues based on Yolo [11]. Its core idea is still to take the entire graph as the input of the network and directly return the location of the boundary box and the category of the target at the output layer. Yolov4 algorithm, the image is divided into $K \times K$ grid, if an object in the center of the grid, the grid will predict the target, is responsible for each grid predicting M bounding box, in addition, each bounding box in return to their own position coordinates (x, y, w, h) at the same time, also need to give the corresponding confidence. Bounding Box Confidence refers to the confidence of the predicted objects to be detected, as well as the bounding box's accuracy information, and the calculation formula is:

$$confidence = Pr(object) * IOU_{pred}^{truth} \quad (1)$$

(x,y) is the upper left coordinate of the bounding box, (w, h) is the width and height of bounding box. When an object to be detected falls within the grid, its Pr(Object) value is 1, and when an object to be detected does not fall within the grid cell, its Pr(Object) value is 0. The IOU_{pred}^{truth} values are the predicted IOU values of bounding box and actual bounding boxes. In this paper, K is 7 and M is 2. So each Bounding box can return five values, (x, y, w, h) and confidence. Besides, each grid cell should predict a category information, so the final loss function of Yolov4 is

$$loss(object) = \lambda_{coord} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} (2 - w_i \times h_i)[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} (2 - w_i \times h_i)[(w_i - \hat{w}_i)^2 + (h_i - \hat{h}_i)^2] - \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{obj} [\hat{C}_i log(C_i) + (1 - \hat{C}_i) log(1 - C_i)] - \lambda_{noobj} \sum_{i=0}^{K\times K} \sum_{j=0}^{M} I_{ij}^{noobj} [\hat{C}_i log(C_i) + (1 - \hat{C}_i) log(1 - C_i)] - \sum_{i=0}^{K\times K} I_{ij}^{obj} \sum_{c\in classes} [\hat{p}_i(c) log(p_i(c)) + (1 - \hat{p}_i(c)) log(1 - p_i(c))] \quad (2)$$

I_i^{obj} is used to determine whether the target falls in grid I, I_{ij}^{obj} is used to determine whether the jth box in the ith grid is responsible for predicting this target, λ_{coord} and λ_{noobj} represent the weight values of bounding boxes containing and excluding targets respectively. Whole bounding Box aimed at people's bounding box of images is clustered through k-means algorithm and measured with IOU distance, whose calculation formula is

$$d(box, centerpoint) = 1 - IOU(box, centerpoint) \quad (3)$$

Box refers to the pre-selected Anchor box and centerpoint refers to the center of bounding Box.

3.2 Improved Openpose Human Joint Detection

Openpose is a bottom-up posture recognition algorithm. In short, it first identifies all the human body nodes, and then identifies the skeleton information. The previous Openpose algorithm adopts CPM+PAF. During training, a PAF connection is to connect two known bone points. If one of the two bone points does not exist, the PAF tag

is not generated. But in the real scenario, a number of factors could lead to a part of the human body joint point has not been captured, so this article adopted the tag correction method was improved, algorithm of the specific process is to use the current state of the art model (for example, has trained CMU - POSE) to a tag to generate data sets, and then at the time of training, the ground way and the fusion of the label, get new labels as the current training groud way. For keypoint label, take Max (groundtruth label, generate_label) directly on the corresponding label, while the label of PAF takes the large one in groundtruth label and generate label. The main network structure of the Openpose algorithm adopts VggNet [12] as the skeleton network, and then adopts two branches to respectively return the position S of the joint node and the trend L of pixel points in the skeleton. And the results of the subsequent two branch networks are multi-stage iterative. The loss function is calculated once for each stage, and then L, S and the original image features extracted by vggNet are connected to continue the training in the next stage. Its network structure is shown in Figure 2.

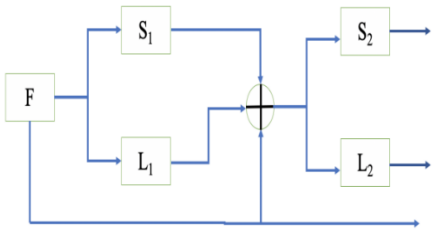

Figure 2: The Basic Structure of Openpose

Where F represents the original image features extracted through vggNet, 1 and 2 represent the first and second stages respectively. For the whole network, the calculation process of S_i and L_i is as follows:

$$\begin{cases} S_i = \rho_i(F, S_{i-1}, L_{i-1}), \ \forall i \geq 2 \\ L_i = \sigma_i(F, S_{i-1}, L_{i-1}), \ \forall i \geq 2 \end{cases} \tag{4}$$

Where ρ_i and σ_i respectively represent the convolutional neural network of L and S in stage i. At the same time, there are still some problems with the improved Openpose algorithm. Joint information may be marked where there is no human being, which may lead to the failure of the random forest classifier to correctly identify the current posture state of the elderly. Therefore, after the target detection using Yolov4, the information of the human body's closing nodes can be detected through Openpose, which can greatly improve the accuracy.

3.3 Random Forest Classifier

In this paper, the random forest algorithm is used to classify the image after the openpose algorithm is used to extract the joint nodes, and the state of the characters is divided into three categories: falling state, lying down state, and other states. The random forest algorithm USES support vector machine (SVM) based on radial basis kernel function as the basic classifier. In the offline stage, Bootstrap sampling is used for training data, and SVM classifier with the best attribute selected based on random attributes is trained. Because SVM has the ability to process high-dimensional data, it can easily handle the key node features extracted by Openpose without reducing the accuracy and generalization ability of the classifier.

The workflow of SVM is shown in Figure 3:

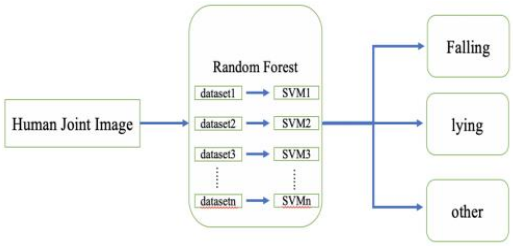

Figure 3: The Workflow of SVM

3.4 Analysis of State

In order to determine whether there are falls in the whole video sequence, it is necessary to further analyze the classified states of random forest. The main purpose of this paper is to extract the key state. The goal of this paper is to detect whether there is a fall event in the scene. Therefore, it only needs to extract the key state in one frame of the image. Since each state is of different importance to the fall detection algorithm, different weight values are assigned to them, and the fall state is $S_{falling}$, the flat state is S_{lying} and the other state is S_{other}. According to the difference of each state, the importance of each state needs to be distinguished. The importance is set as follows:

$$S_{falling} > S_{lying} > S_{\text{other}} \quad (5)$$

Therefore, when three or two states simultaneously appear in the video frame image, the key state will be taken as the state of the video frame image, and the weights of the three states will be assigned respectively:$w_{falling} > w_{lying} > w_{other}, and\ w_{falling} = 3, w_{lying} = 2, w_{other} = 1$. In this way, it is possible to distinguish the state of the characters in the multi-player scene, and only need to find $S_{falling}$ to find whether there is a fall phenomenon in the video sequence.

4. Experiments

We test the influence of different training improvement techniques on accuracy of the classifier on UR Fall Detection dataset and High quality fall simulation dataset, using the GPU provided by Google Colab as a training machine.

4.1 Introduction to Dataset

This dataset contains 70 (30 falls + 40 activities of daily living) sequences. Fall events are recorded with 2 Microsoft Kinect cameras and corresponding accelerometric data. ADL events are recorded with only one device (camera 0) and accelerometer. Sensor data was collected using PS Move (60Hz) and x-IMU (256Hz) devices. The dataset is organized as follows. Each row contains sequence of depth and RGB images for camera 0 and camera 1 (parallel to the floor and ceiling mounted, respectively), synchronization data, and raw accelerometer

data. Each video stream is stored in separate zip archive in form of png image sequence.

4.2 Analysis of experimental results

There are two indicators that are widely used in the field of fall detection, namely sensitivity and specificity.

$$sensitivity = \frac{TP}{TP+FN} \quad (6)$$

$$Specificity = \frac{TN}{TP+FN} \quad (7)$$

The confusion matrix is shown in the table1.

Table 1: Confusion matrix

TP	True Positive: The number of video frames in which a fall event occurs and a fall event occurs is detected
TN	True Negative: The number of video frames in which no falls occurred and no falls were detected
FP	False Positive: The number of video frames in which no falls actually occur but are detected as having falls
FN	False Negative: The actual number of video frames in which a fall event occurs but is not detected

The advantages and disadvantages of the algorithm are compared on the two data sets respectively. The results are shown in the table2 below:

Table 2: Comparison of Experimental Results(single person dataset)

Methods	Sensitivity	specificity
Charfi and his colleagues [13]	99.61%	98.00%
Yaxiang Fan and his colleagues [14]	98.43%	100.00%
Method in the paper	99.3%	99.3%

Table 3: Comparison of Experimental Results(Multi-person dataset)

Methods	Sensitivity	specificity
Charfi and his colleagues [13]	62.2%	41%
Yaxiang Fan and his colleagues [14]	74.2%	68.6%
Method in the paper	79.3%	72.1%

5. Conclusions

In this paper, we design a new fall detection algorithm for the elderly. It can run in Google Colab, this is very convenient for students to do some research. Besides the sensitivity and specificity has reached a new step. The accuracy of the model reached 99.3%, the sensitivity and specificity of the model reached 79.3% and 72.1%. Anyway, this model can be used in some real life scenarios. Due to the limitations of the training data set, there is room for improvement if the model is to run in cross-domain scenarios as well as real-world scenarios.

References

[1]. 2019 World Population Outlook, The United Nations, 2019

[2]. He J, Hu C, Wang X. A smart device enabled system for autonomous fall detection and alert[J]. International Journal of Distributed Sensor Networks, 2016, 12(2): 2308 183.

[3]. Tamura T, Yoshimura T, Sekine M, et al. A wearable airbag to prevent fall injuries[J]. IEEE Transactions on Information Technology in Biomedicine, 2009, 13(6): 910-914.

[4]. LAI C F, CHANG S Y, CHAO H C，et al.Detection of cog- nitive injured body region using multiple triaxial accelerome- ters for elderly falling[J]. IEEE Sensors Journal, 2011, 11 (3) : 763-770.

[5]. Alwan M, Rajendran P J, Kell S, et al. A smart and passive floor-vibration based fall detector for elderly[C]. Information and Communication Technologies, 2006. ICTTA'06. 2nd, 2006: 1003- 1007.

[6]. Toreyin B U, Soyer A B, Onaran I, et al. Falling person detection using multi-sensor signal processing[J]. EURASIP Journal on Advances in Signal Processing, 2007, 2008(1): 149304.

[7]. Wang C-C, Chiang C-Y, Lin P-Y, et al. Development of a fall detecting system for the elderly residents[C]. Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on, 2008: 1359-1362.

[8]. Wang Kun. Research on human fall detection based on deep learning features[D]. East China Normal University, 2017.

[9]. Yan Ran. Fall detection for the elderly based on Kinect sensor [J]. Electronic Technology and Software Engineering, 2017, (22): 83-83.

[10]. Fan Yu-yu, Li Li-pin, Dang Rui-rong based on stochastic resonance with any large frequency weak Research on signal Detection Methods 0]. Acta Instrumentation, 2013, 34(3) : 566-572.

[11]. REDMON J, FARHADI A. YOLO9000: better, faster, stron- ger[C]/ /2017 IEEE conference on computer vision and pattern recognition. Honolulu, Hawaii: IEEE, 2017: 6517 - 6525.

[12]. CAOZ, SIMONT, WEISE, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]/ /2017 IEEE conference on computer vision and pattern recognition. Hon- olulu, Hawaii: IEEE, 2017: 1302-1310.

[13]. CHARFI I, MITERAN J, DUBOIS J,et al. Definition and performance evaluation of a robust SVM based fall detection solution[C]/ /Eighth international conference on signal image technology and internet based systems. Naples, Italy: IEEE, 2012: 218-224.

[14]. FAN Yaxiang, LEVINE M D, WEN Gongjian, et al. A deep neural network for real-time detection of falling humans in naturally occurring scenes[J]. Neurocomputing, 2017, 260: 43-58.