Fall detection algorithm for the elderly based on human posture estimation

Guangmin Sun
Faculty of Information Technology
Beijing University of Technology
Beijing, China
gmsun@bjut.edu.cn

Zhongqi Wang*
Faculty of Information Technology
Beijing University of Technology
Beijing, China
wangzq00806@163.com

***Abstract*—With the acceleration of the global aging process, the safety and health of the elderly has become a widespread concern, and falls have become the main health threat of the elderly. In this paper, a fall detection model based on OpenPose human posture estimation algorithm is proposed by using the fall detection method based on machine vision. On the basis of OpenPose human key point detection, combined with SSD - MobileNet object detection framework can remove the non human key points detected by OpenPose algorithm, reduce the false detection rate of the algorithm, improve the robustness of the algorithm in complex environment, and then extract the features of human joint points, use SVDD classification algorithm to classify, experiments show that this method can effectively detect the occurrence of falls, and the accuracy rate can reach more than 93%.**

Keywords—Human posture estimation, OpenPose, object detection

I. Introduction

In recent years, with the acceleration of the global aging process, the safety and health of the elderly has become a topic of widespread concern. Falls have become a major health hazard for the elderly over 60 years old. Falls can cause serious consequences, including fractures, superficial wounds and bruises of skin and soft tissues [1]. And about 47% [2] of the elderly can't restore their balance without help. If they can't help and cure the elderly in time, they will cause more damage. Therefore, it's very necessary to accurately and quickly identify the fall and inform the guardian or medical staff of the elderly. As an important technology of social public security, monitoring system has been widely used in different fields. In the intelligent monitoring system [3], the computer is required to process and analyze the collected image data in real time, detect, identify and track the target, and analyze the target's behavior. When the target's behavior is abnormal, the monitoring system will send out an alarm, save video data and a series of operations. The combination of intelligent monitoring system and fall detection can inform the guardian or alarm at the first time when the elderly fall, shorten the time from falling to getting treatment, reduce the possibility of secondary injury, and have great significance for improving the quality of life of the elderly.

II. Related Work

In recent years, researchers have proposed a variety of methods to realize automatic fall detection. Mubashir et al. Divide fall detection into three main types [4], which are based on wearable devices, environmental sensors and computer vision.

M. J. Mathie et al. Used accelerometers installed at the waist to obtain the acceleration of the human body, using the acceleration as the main feature of fall recognition, and the system can also be used to monitor a series of different movements, including gait, sitting and standing conversion, attitude conversion, etc. [5]. M. Kangas et al. Used a three-axis accelerometer installed at the waist to detect falls, and recognized falls by analyzing human posture [6].

The fall detection method based on environmental sensors is to use sensors or data acquisition devices installed in the detection environment area. When the human body falls, it will cause special signals to the surrounding environment. Through the analysis of these signals, fall detection can be realized. Alwan et al. Judged by the device with integrated piezoelectric sensor installed on the floor surface, monitored the vibration data on the floor in real time. When the human body falls, the data and features collected on the ground are compared with the data of human body falls in the experiment. If the features are consistent, the alarm mechanism will be triggered [7].

With the development of artificial intelligence, pattern recognition and other technologies, more and more fall detection methods based on machine vision have been proposed, which has become a research hotspot. Compared with the fall detection based on wearable devices, the fall detection based on machine vision has less interference, higher accuracy and robustness. And it can update and perfect its own system with the maturity of technology.

Rougier et al. Use a single camera to extract the three-dimensional motion track of the human head, analyze the speed extraction features of the head motion track, and judge whether the human body falls [8]. Nguyen et al. Combined with aspect ratio and drop angle to detect the body posture of the elderly. Through the classification of these two features to detect whether the human body falls [9]. Diraco et al. Used the camera to collect data, obtained the distance between the center of gravity of the human body and the ground by calibrating the camera parameters, and combined with the direction of the human spine to determine the occurrence of falls [10].

With the introduction and wide use of Microsoft Kinect somatosensory camera, more and more researchers and developers use Kinect to detect falls. Stone et al. [11] implemented a two-state fall detection technology, and verified the accuracy of the system on a large dataset collected in multiple apartments. Kawats et al. Make a fall judgment by detecting the height of the joint and the ground and the average

978-1-7281-6067-2/20/$31.00 ©2020 IEEE

Dalian, China
April 14-16, 2020

speed of each joint. After detection, they will confirm whether they fall or not by voice query or delayed automatic confirmation, and then send the alarm information to the designated mailbox or designated number [12]. Morel [13] et al. Defined and calculated three features (distance, angle, velocity) on several important joints. In order to distinguish falls from other activities, such as lying on your back, SVM algorithm is used for classification.

III. Algorithm Design

A. OpenPose algorithm for human posture estimation

The current mainstream human pose estimation algorithm classification is shown in Figure 1. For human body posture detection It can be divided into two directions: the top-down detection method and the bottom-up detection method. For the top-down method, it is often to find all the people in the picture first, and then make attitude estimation for each person to find the key points of each person. The bottom-up detection method is the opposite. First, find all the key points in the picture, such as nose, left hand, right hand, hip joint, etc. Then reassemble these key points into everyone.

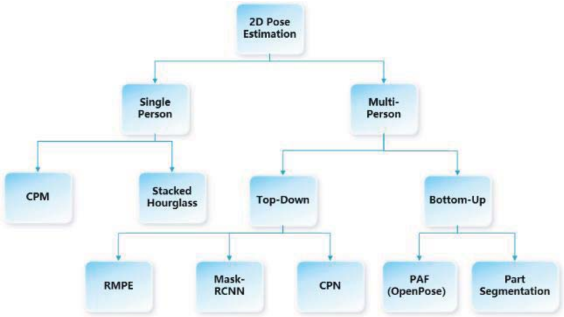


Fig. 1. Current mainstream human posture estimation algorithms

OpenPose [14] human posture estimation algorithm was published in 2017. Based on CPM, CMU team first finds the position of each joint point in the picture, and then proposes part affinity fields to assemble the key points. Firstly, this method uses VGG-19 network to extract feature map, as shown in Figure 2, which is initialized by the first 10 layers of VGG-19 and fine tuned to generate a set of feature map F input to the first stage of each branch. In the first stage, the network generates a set of prediction $S^1 = \rho^1$ (F) of confidence map and a set of partial vector relation field $L^1 = \varphi^1$ (F), where $\rho^1$and φ^1 is CNN for stage 1 inference. In the following stages, the prediction from the two branches of the previous stage and the original image feature F are used with the prediction of the next step.

$$S^t = \rho^t(\mathrm{F}, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (1)$$

$$L^t = \varphi^t(\mathrm{F}, S^{t-1}, L^{t-1}), \forall t \geq 2 \quad (2)$$

In the formula, ρ^t andφ^t are the prediction of stage t.

There may be confusion between the early body parts of the network, but through the later network iteration, The prediction of key points of human body is more and more accurate. The two branches of the network have two loss functions, one for each branch. L2 loss is used between the prediction and the maps and fields of the growth truth to prevent over fitting. The loss function is weighted.

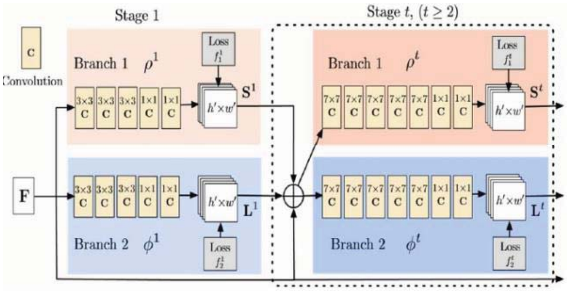


Fig. 2. Feature extraction network structure diagram

$$f = \sum_{t=1}^{C} (f_S^t + f_L^t) \quad (3)$$

f_S^t and f_L^t are L2 loss functions from 1 to C, respectively. f is the total loss function. Personal confidence maps are generated for each person, and each set of confidence is predicted by the network. In the process of network training, for a given type of key point J, for a certain person K, there is only one peak in his confidence graph, and the confidence label is the Gaussian distance from each point to the real point of the data. For multiple individuals, NMS is used to eliminate the low score data in the network.

The second step is to add vector coding to the predicted key points to connect different parts of the same person. Given a group of detected body parts, the confidence of each pair of body parts detection association is measured. One way of detection association is to detect the additional midpoint between each pair of limbs on the limb, and check the incidence of candidate parts detection. When the human body overlaps in order to solve these limitations, we use the PAFs feature representation method to retain the position and direction information of the limb support area. PAFs is the 2D vector relationship field of each limb. For each pixel in the region belonging to a specific limb, 2D vector coding is the direction from one part of the limb to another. Each type of limb has a corresponding PAFs area for connecting two related body parts. Non maximum suppression (NMS) is applied to the detection confidence map to obtain a set of discrete position candidates. For each part, there may be multiple candidate positions, which define a large group of possible limbs due to multiple people or false positives in the image. For the k-dimension matching problem known as NP-Hard, Hungary algorithm is used to obtain the best matching, and the results are shown in Figure 3.

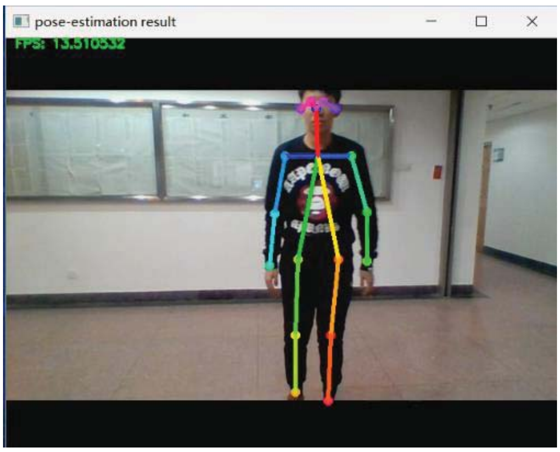


Fig. 3. Human posture detection result

B. SSD-MobileNet object detection model

Because OpenPose uses a bottom-up detection method, it first detects the key points of human body, and then checks them.

The combination of key points into a single person may appear in the area where no one appears to detect the key points of human body. The solution to this situation is to use SSD MobileNet object detection model to detect the key points of human body and remove the non-human part.

The SSD-MobileNet network structure is similar to the VGG_ SSD network structure. Eight volume layers are added after the conv13 volume layer, six of which are used to detect the target. SSD-MobileNet network model uses SSD model as the basic model, combined with the characteristics of MobileNet which uses fewer parameters and reduces the amount of computation. The network structure is shown in Figure 3. Because the model combines the advantages of MobileNet and SSD networks, on the basis of good accuracy, it uses small-scale parameter networks to reduce the amount of calculation, shorten the training time, improve the detection speed, and is widely used in the task of object detection.

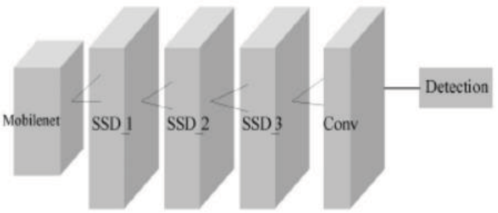
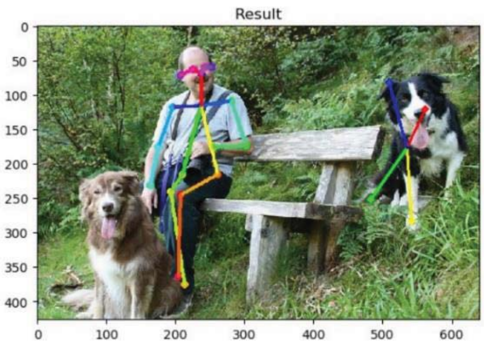
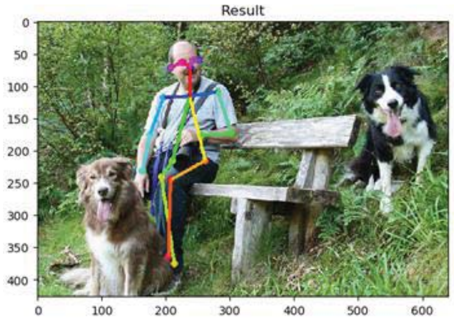


Fig. 4. Structure of SSD-MobileNet convolutional neural network

The purpose of transfer learning is to apply the knowledge learned in one application scenario to another. The knowledge learned by neural network is mainly reflected in the weight parameters trained on specific tasks, so the essence of "shift learning" is "shift of weight". Traditional machine learning is mainly based on the assumption of the same distribution, which requires researchers to provide a large number of labeled data. However, in the actual application process, different data sets may have some problems, such as the difference of data distribution, which will cause the problems that the existing data sets cannot continue to use, and the distribution of some data will change over time. The above problems can be solved by using the migration learning method, which allows the existing knowledge to be applied to the new and only a few sample data areas. It only needs to retrain the part of a model, and it does not need too long training time to get a better network model. As shown in Figure 5, redundant human joints detected by OpenPose can be removed before and after using SSD-MobileNet.



a



b

Fig. 5. Before using SSD-MobileNet (a) after using SSD-MobileNet (b)

C. SVDD classification algorithm

Although the trajectory of joint points is obtained, there is still a problem of data imbalance. In daily life, due to the diversity of fall types, it not only needs a lot of manpower and resources, but also is difficult to obtain enough actual sample drop, which leads to sample imbalance, so how to use imbalance data to detect falls has become a basic problem. In this method, SVDD anomaly detection method is used, and grid search method is used to optimize the parameters. The normal domain of positive sample training was formed, and the fall behavior was detected by the normal domain. SVDD is primarily used for anomaly detection and image classification. Unlike classifier support vector machine (SVM), SVDD is a single class SVM algorithm. The main idea is to train positive samples to construct normal domain hypersphere, while the samples outside the hypersphere are abnormal samples. For a given positive sample set X, it contains n positive samples (Xi = 0,1,2,..., n), so as to reduce the influence of abnormal data points included in the normal domain when constructing the hypersphere. The penalty factor C and the relaxation variable ξ_i, when the center of the hypersphere of the hemisphere R, make the positive samples completely surrounded by the sphere. The corresponding optimization equation is:

$$min|R^2 + C\sum_i^N \xi_i| \tag{4}$$

For a typical quadratic programming problem, Lagrange multiplier is introduced and the corresponding function is solved. Through Often, after deleting outliers, the data will not be distributed in a spherical way. Therefore, the kernel function K is introduced to transform the nonlinear problems in low dimensional space into high dimensional linear problems. Compared with KNN and SVM, this method has significant effect in reducing false detection.

IV. Experimental Result

The main configuration of the computer used in the experiment is shown in Table I. Its operating system is windows 10. The main software package needed mainly includes PyCharm, Python 3.6.4, CUDA 9.0, CUDNN7.1, tensorflow 1.11, OpenCV, etc.

TABLE I. EXPERIMENTAL HARDWARE SYSTEM

Hardware	Hardware type
CPU	Intel(R) Core(TM) i7-7700HQ
GPU	NVIDIA GeForce GTX1060
Graphic memory	6 GB
RAM	16 GB
Hard disk	512GB SSD
System	Windows 10

When training the human key point detection model, in order to make the model more suitable for the application scenario of this method, we collected 1500 RGB images of human body's various actions, including daily human behavior and fall images. After labeling, we used tensorflow framework to train the collected images. With the increasing number of iterations, the loss function value does not decrease, and set the number of iterations to 10000 When it reaches 10000, the loss value is 0.0925, and the model is used to test the test set, the accuracy is 91.2%, which proves that the algorithm meets the use requirements.

When training the human key point detection model, in order to make the model more suitable for the application scenario of this method, we collected 1500 RGB images of human body's various actions, including daily human behavior and fall images. After labeling, we used tensorflow framework to train the collected images. With the increasing number of iterations, the loss function value does not decrease, and set the number of iterations to 10000 When it reaches 10000, the loss value is 0.0925, and the model is used to test the test set, the accuracy is 91.2%, which proves that the algorithm meets the use requirements. Using MobileNet network to pre train COCO data set, this data set includes 80 kinds of targets, including 12 kinds of targets, such as human body, animals, vehicles, household appliances, etc. the parameter model of pre training is obtained.

Using the self built human dataset collected before, use the LabelImg annotation tool to annotate the human body. After the samples are tagged, the corresponding XML file is generated, then the CSV form is generated by calling xml_to_csv.py, then the CSV file is transformed into tensorflow format recognized tfrecord format by using generate_tfrecord.py file. The fusion feature of the data pre training model is transferred into SSD network, and the training is stopped when the loss reaches 1.8, the total loss is 1.8, and the average accuracy is 78.4%. After the implementation of this step, redundant key points can be removed and only the correct key points of human skeleton can be retained.

$$sensitivity = \frac{TP}{TP+FN} \quad (5)$$

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

TP (real case) is the number of videos with falling events and falling events detected in the video; FN (false negative case) is the number of videos with falling events but not detected; FP (false positive case) is the number of videos with no falling events but falling events; TN (true negative case) is the number of videos without falling events in the video The number of videos detected as not falling.

The comparison of experimental results is shown in Table II.

TABLE II. EXPERIMENTAL COMPARISON RESULTS

algorithm	sensitivity(%)	Specificity(%)
SVM	92.5	93.7
KNN	93.8	92.3
Article method	94.6	93.8

The results in the actual scenario are shown in Figure 6

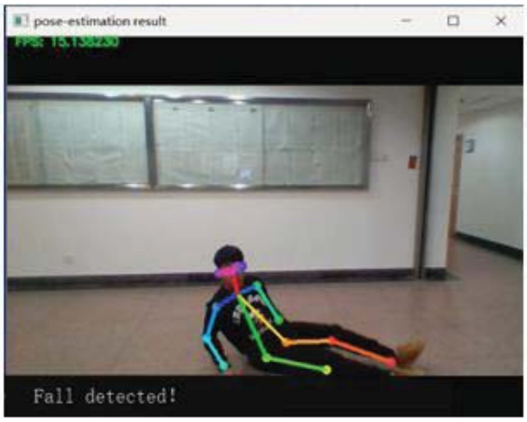


Fig. 6. Test results in actual scenarios

V. CONCLUSION

This paper presents a fall detection model based on OpenPose algorithm. Based on the detection of human key points in OpenPose, in view of the insufficient data of fall type in COCO dataset, the data set is collected for migration learning to increase the accuracy of the model, combined with SSD-MobileNet object detection framework, the non human key points detected by OpenPose algorithm are removed, and the key point features are extracted and the abnormal values are classified by SVDD classification algorithm. Experiments show that this method can reduce the false detection rate and improve the robustness in complex environment.

REFERENCES

[1] L. Yardley, H. Smith, “A prospective study of the relationship between feared consequences of falling and avoidance of activity in community-living older people,” The Gerontologist, vol. 42, no. 1, pp. 17-23, 2002.

[2] M E, Tinetti, W L, Liu, E B, Claus, “Predictors and prognosis of inability to get up after falls among elderly persons,” JAMA, vol. 269, no. 1, pp. 65-70, 1993.

[3] A. Yilmaz, O. Javed, M. Shah, “Object tracking: a survey. ACM Comput Surv,” ACM Computing Surveys, vol. 38, no. 4, pp. 13, 2006.

[4] M. Mubashir, L. Shao, L. Seed, “A survey on fall detection: Principles and approaches,” Neurocomputing, vol. 100, pp. 144–152, 2013.

[5] M. J. Mathie, A. C. F. Coster, N. H. Lovell, et al., “Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement,” Physiological Measurement, vol. 25, no. 2, pp. R1-R20, 2004.

[6] M. Kangas, I. Vikman, J. Wiklander, et al., “Sensitivity and specificity of fall detection in people aged 40 years and over,” Gait & Posture, vol. 29, no. 4, pp. 0-574, 2009.

[7] M. Alwan, P. J. Rajendran, S. Kell, et al., “A Smart and Passive Floor-Vibration Based Fall Detector for Elderly. Information and Communication Technologies,” ICTTA '06. 2nd. IEEE, 2006.

[8] C. Rougier, J. Meunier, A. St-Arnaud, et al., “Monocular 3D head tracking to detect falls of elderly people,” Conf Proc IEEE Eng Med Biol Soc, vol. 1, no. 1, pp. 6384-6387, 2006.

[9] V. D. Nguyen, M. T. Le, A. D. Do, et al., “An efficient camera-based surveillance for fall detection of elderly people,” 2014 IEEE 9th Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2014.

[10] G. Diraco, A. Leone, P. Siciliano, “An active vision system for fall detection and posture recognition in elderly healthcare,” Design, Automation & Test in Europe Conference & Exhibition(DATE), IEEE, pp. l536-1541, 2010.

[11] E. E. Stone, M. Skubic, “Fall Detection in Homes of Older Adults Using the Microsoft Kinect,” IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1, pp. 290-301, 2015.

[12] C. Kawatsu, J Li, C Chung, “Development of a Fall Detection System with Microsoft Kinect,” Robot Intelligence Technology and Applications. IEEE, 2012.

[13] T. Tran, T. Le, J.Morel, “An analysis on human fall detection using

skeleton from Microsoft Kinect," 2014 IEEE Fifth International Conference on Communications and Electronics(ICCE). IEEE, pp. 484-489, 2014.

[14] Z. Cao, T. Simon, Wei S E, et al., "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," CVPR 2016. 1611. 08050, 2016.