Real-time fall detection system based on deep learning and infrared array sensors

Tianyi Gong[1], *Xinyu* Yin[1], *Shicheng* Yan[2], *Junhao* Pan[3,4], *Yifan* Yang[3,4], and *Jianyang* Liu[2,*]

[1]School of Information Science and Technology, Southwest Jiaotong University, China
[2]School of Mechanical Engineering, Southwest Jiaotong University, China
[3]School of Public Administration, Southwest Jiaotong University, China
[4]National Interdisciplinary Institute on Aging, Southwest Jiaotong University, China

Abstract. In this paper, a novel fall detection algorithm based infrared image is proposed. Firstly, the RetinexNet algorithm is adopted for the infrared image pre-processing and enhancement, then the YOLOv3 algorithm is improved by adding three bounding boxes to achieve the task of falling posture detection and recognition, finally a fall data set collected by ourselves is utilized to train and test the algorithm. The experimental results shows that our proposed algorithm achieves excellent fall detection accuracy result and outperforms the traditional YOLOv3 algorithm, the average accuracy of our proposed algorithm is more than 90.86%, which meets the requirements of the fall detection task quite well.

1 Introduction

With the progress of the times, the problem of aging population is intensifying. According to surveys, falling is a major threat to the health of older people, with nearly 34% of people over 60 years of age falling at least once a year, with 64.4% of these falls occurring at home[1]. The National Chronic Disease Surveillance System (NCDS) which is a Chinese organization shows that falls account for 22.57% of all deaths among older people[2]. There are currently two main types of fall detection methods internationally: acceleration sensor recognition systems and vision recognition systems. In terms of acceleration sensors, they are further divided into human body sensors and environmental sensors. Among the human body pressure and acceleration sensors are mainly the human posture recognition system designed by Zhou Boxiang et al. in 2015 using a three-axis acceleration sensor and CC2430[3] and the wearable acceleration sensor recognition system designed by Chen Weimei et al. in 2017 using acceleration sensor data collection and K-means clustering algorithm to recognise elderly people's movements[4], in terms of environmental sensor aspect vibration sensors and microphones were installed in the indoor living environment. The detection and judgement of the elderly posture is mainly determined by a comprehensive analysis of various environmental information. And the main studies in this area are Zhuang Xiaodan et al[5]. designed a fall detection system based on microphone to collect sound information and Alwan M et al. used a shock sensor to determine the fall pose and, in their study, they used

[*] Corresponding author: manchest@swjtu.edu.cn

floor-mounted vibration sensors to determine whether an elderly person had fallen[6]. In visual fall posture detection started relatively late, in 2007 Rougier Caroline et al. determined whether an elderly person had fallen by extracting the contour points of the human body[7]. in 2009 Foroughi Homa et al. proposed to circle the human body with an ellipse, which is considered to describe the contour of the human body[8]. This would allow the elderly person to determine if they have fallen based on the ratio of the long axis to the short axis of the detected ellipse. However, these solutions often face problems such as high cost, lack of protection of personal privacy and the fact that they are not easily portable in life using wearable devices.

2 Proposed method

We propose a method to detect falls in the elderly by using infrared matrix sensor and computer vision. This method not only protects the privacy of users, but also reduces the cost of equipment. However, due to the low pixel count of the matrix sensor, the acquired images are characterised by poor clarity, more noise and contain less information. We use the following solutions: (1) The acquired images are enhanced using the RetinexNet algorithmic network, so that the acquired images contain more information and less noise. (2) Training with the improved YOLOv3 algorithm improves the detection of micro targets and blurred objects in the YOLOv3 network, increasing the accuracy of the model. This way the infrared matrix module can reduce costs while protecting personal privacy and obtain a high accuracy rate, addressing the shortcomings of the previous recognition schemes.

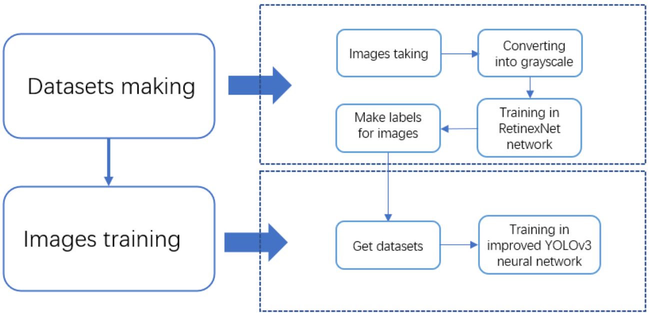
Datasets making
Images training
Images taking
Converting into grayscale
Make labels for images
Training in RetinexNet network
Get datasets
Training in improved YOLOv3 neural network

Fig. 1. Introduction to our solution.

3 Image pre-processing

3.1 Feature extraction method

The Retinex theory was developed by Land et al. in 1971 as a simulation of the human eye's ability to perceive colour and luminance. The Retinex theory estimates the intensity of the incident light from the image and reduces the effect of the incident light on the reflected light to enhance the image. According to the Retinex theory[9] the image consists of two parts, denoted as the incident light and the reflected object, which can be expressed by the following equation The image is expressed as:

$$S(x,y) = R(x,y) \times I(x,y) \tag{1}$$

In this formula, (x,y) is the coordinates of the image pixels, $S(x,y)$ is the original image i.e. the processed image, $R(x,y)$ is the reflectance image i.e. after greyscale processing image, and $I(x, y)$ is the incident light image i.e. the original image which is captured by the IR matrix module.

3.2 RetinexNet network structure

In this paper, we use RetinexNet algorithm to enhance low contrast images and low light images. The RetinexNet algorithm consists of a three-step artificial neural network: decomposition, enhancement and reconstruction. In the decomposition step, the DecomNet network takes advantage of the fact that the grayscale image and the original infrared image have the same reflectance to decompose the grayscale image to estimate the incident light image. In the adjustment step, the incident light image is enhanced with the EnhanceNet network, which uses a holistic framework of encoding-decoding to maintain consistent global contrast and background information in a multi-scale cascade, while focusing on local colour distribution information. In addition, there is often a large amount of noise in low-light environments, so reflected images need to be denoised. Finally, incident light and reflectance images are reconstructed by element-by-element point multiplication. We apply RetinexNet to infrared image enhancement. The IR image captured by the IR matrix module is processed as a greyscale image, then the greyscale image is used as the lowest light image and the IR image is fed into the RetinexNet network as a high light image for training, resulting in a contrast-enhanced, noise-reduced IR image.

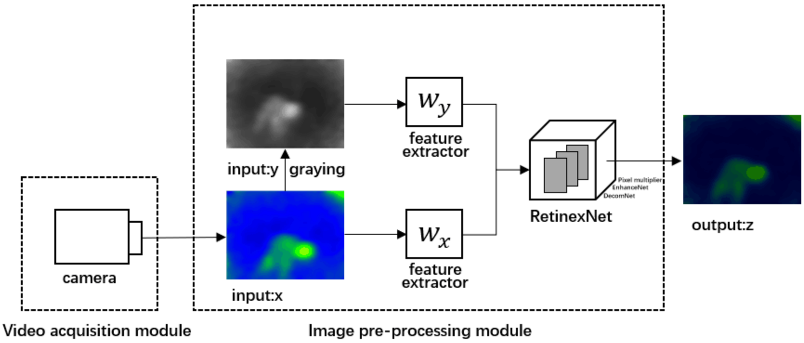

Fig. 2. Structure of our RetinexNet algorithm**.**

4 Improved YOLOv3 algorithm

4.1 Features of YOLOv3 scale

The YOLO (You only look once) family of algorithms is one of the faster methods of current target detection algorithms[10]. YOLO algorithms sacrifice some of its accuracy but greatly compensates for the lack of real-time performance of existing algorithms and can efficiently perform real-world commercial items in the field of industrial target detection. YOLOv3 uses a Darknet-53 network structure containing 53 convolutional layers and there is also a link between the convolution network and the residual network, and YOLOv3 uses three different scales of feature maps for target detection[11]. However, due to problems with the network structure itself and the size segmentation of the Bounding Box, the YOLOv3 algorithm has some shortcomings when detecting targets with complex backgrounds, distant objects to be measured and not high enough pixels. So, the detection of micro and fuzzy targets still has shortcomings. In this paper, the algorithm is improved by adding 8×8, 4×4 and 2×2 convolutional scales to the original feature scales, and the activation function of the neural network units is modified to improve the adaptability of the model, while some of the network layer parameters are reduced to avoid the problem of excessive training volume and long training time, where the increased convolutional scales are shown in Figure 3.

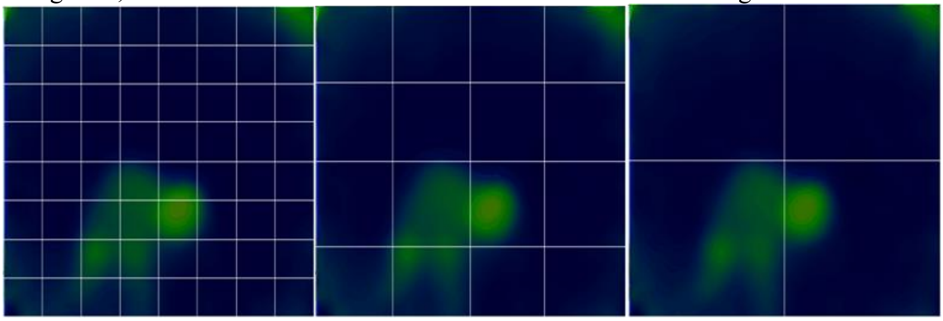

Fig. 3. The characteristic scale map added.

4.2 Multiscale feature fusion with YOLOv3

The YOLOv3 algorithm initially has 3 scales of feature maps. The first scale (13×13) is obtained by performing several convolution operations at layer 79, with a high down sampling multiplier and a relatively large perceptual field, suitable for detecting objects of large size in images. The second scale (26×26) feature map is obtained by up-sampling the above results and concatting them with the results of the 61st layer, and then by several convolution operations. It has a medium-scale field of perception and is suitable for detecting medium-scale objects. The third scale (52×52) is obtained by up sampling the results of layer 91, concatting them with the results of layer 36, and performing several convolution operations [12].

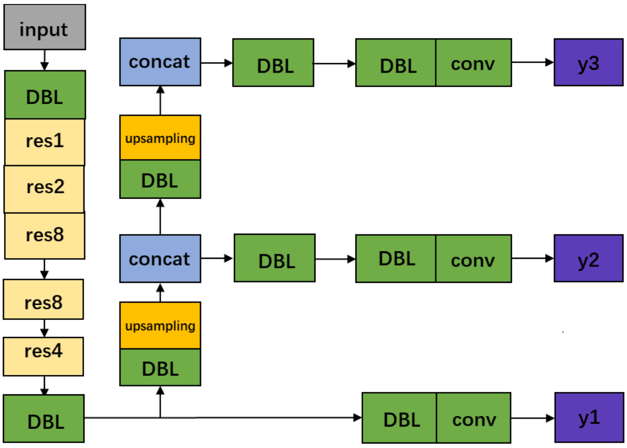

Fig. 4. The structure of our YOLOv3 algorithm.

In this paper, after adding 8×8, 4×4 and 2×2 scales for super-feature fusion, 512 3×3 convolutional layers are added to the corresponding data elements based on the layered features, so that the dimensionality of the layered feature channels remains consistent. In order to solve the interference caused by the different distribution of each layer in the training process and accelerate the training of the dataset, we add a Batch Normalization (BN) layer, and process the data elements through conv, BN, deconv, and pool layers in multiple dimensions, and finally superimpose them to generate the super-feature fusion. We also improve the activation function of YOLOv3 to improve the generalization ability of the new model and reduce the probability of overfitting.

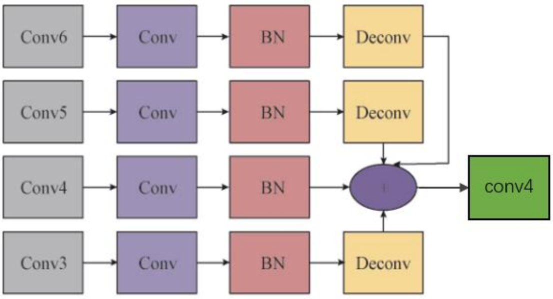

Fig. 5. Hyper feature fusion flow chart.

5 Experiments and results

All experimental tests were performed on a laptop and its parameters are as follows: GPU is Nvidia RTX2060, CPU is Inter CORE i7 9750H and memory is Hynix 16GB. The infrared camera module uses the MLX90640 32x24 IR array module using USB for serial communication.

5.1 Data set

The data set used in this paper is a self-made data set. We selected two experimenters to simulate the daily life actions of the elderly, mainly including walking, standing and falling, and used the infrared matrix sensor to connect with the computer through serial port to collect the human posture. In this paper, some duplicate and similar infrared images are deleted to form a data set with 3000 images. We selected non-falling actions such as standing and walking in the data set as a negative example and falling actions as a positive example. As shown in Figure 6, the data set includes falling images and non-falling images, including 1500 falling images and 1500 non-falling images.

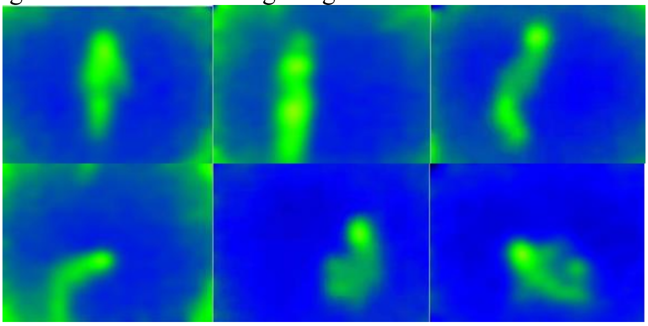

Fig. 6. Sample images of Data set.

5.2 Image processing result

In this paper, we use the IR matrix module to create our own IR dataset, greyscale the dataset, feed the greyscale image and the original IR image into the RetinexNet network for image enhancement, and the enhanced image is shown in Figure 7. The images in the acquisition dataset shown here all contain significant lighting variations and blurring issues. After processing with the RetinexNet network, the illumination variation of the images diminishes, which reduces the training error due to the image illumination problem and improves the adaptability of the algorithm.

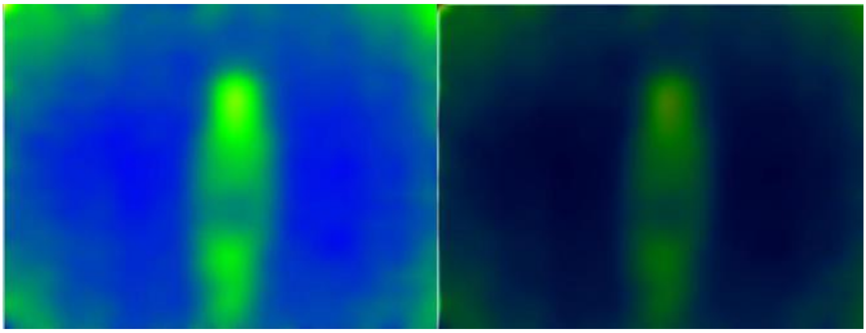

Fig. 7. Display graph of image enhancement effects.

5.3 Image recognition result

The self-made datasets shown in Figure 7 contains infrared images of people in various poses in various scenes with a sample size of 3000. The pose of the target in the images is labelled with labelImg software, where the label information is "walk, fall". The dataset processed by the RetinexNet network was then fed into a modified YOLOv3 algorithm model, of which 80% was used as training data, 15% as validation data and 5% as test data. The adaptability and accuracy of the algorithm are improved by using image enhancement network and additional Bounding Box. The final detection results obtained from the best weighted files are shown in Figure 8.

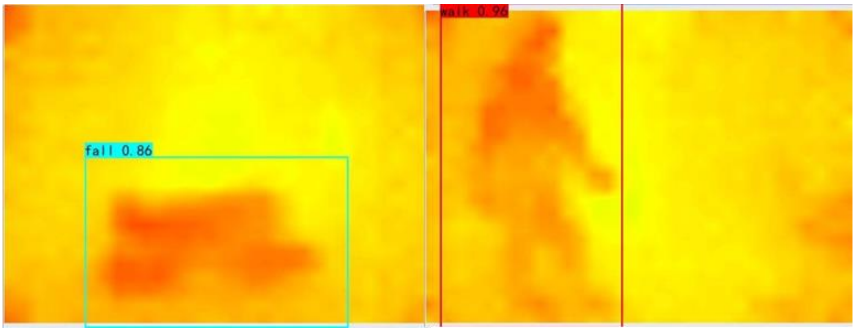

Fig. 8. Display diagram of detection effect.

5.4 Evaluation Criteria

In the training evaluation index, we use the loss function and mAP (mean average precision) indicator. The mAP indicator is the average value of all target category AP (average precision), and the AP of each target category are different recall values. Next, select the maximum precision when it is greater than or equal to these recall values, take recall value as the independent variable, the maximum precision under this recall as the dependent variable, draw the curve, and the area under the curve as the AP of this target, as shown in Figure 9. TP is the sample that correctly detects the state of the person, FP is the number of incorrectly detected human states, FN is the number of elderly targets not detected.

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

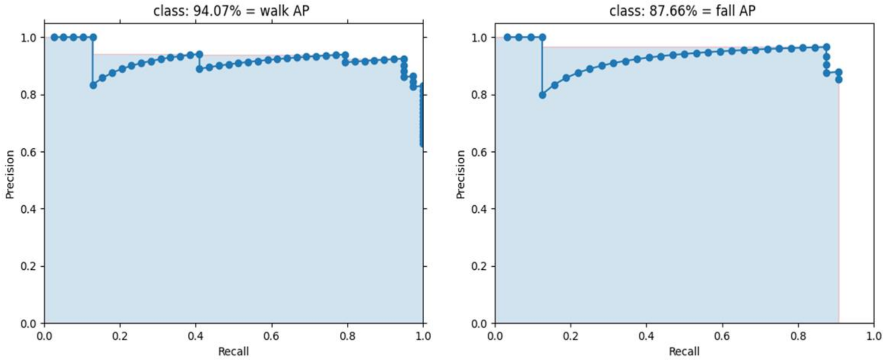

Fig. 9. AP calculation diagram.

And the loss curve and mAP graphics is shown in the figure 10.

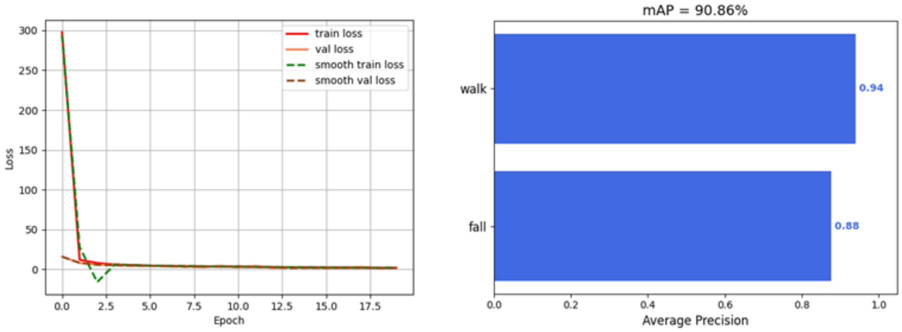

Fig. 10. Loss curve and mAP graph of YOLOv3 training results.

As can be seen from the figure, the training model for this experiment is better, and the trend of the decreasing curve changes from rapid to smooth, while also maintaining a high average accuracy rate.

6 Conclusion

This paper proposes an improved algorithm based on YOLOV3 to better detect the fall posture of elderly people and adds pre-processing of the images using RetinexNet to reduce the image loss problem that exists when the infrared matrix module acquires photographs. Furthermore, this paper compares the algorithm with the classical YOLOV3 algorithm. Combining the data in Table 1, it is easy to find that the YOLOV3 algorithm is slightly lacking in accuracy in recognition due to blurred boundaries and loss of picture information. This algorithm uses RetinexNet for image pre-processing and then uses the improved YOLOv3 algorithm for recognition. The algorithm of this paper can process the data better and detect the target more accurately after a comprehensive comparison; from the final data comparison table, it can be obtained that the improved algorithm has better accuracy and speed with or without image pre-processing. However, the data object studied in this paper is a single-frame image, and the realistic single-frame phenomenon does not necessarily reflect the actual movement of the elderly accurately. In order to make the algorithm more realistic and convincing, the algorithm model needs to be improved when studying objects with multiple consecutive frames or videos.

Table 1. Comparison of parameters and performance of the algorithm in this paper with YOLOV3.

name	backbone	AP		F1		Precision		Recall	
		fall	walk	fall	walk	fall	walk	fall	walk
Our-YOLOv3	Darknet-53	**87.66%**	**94.07%**	**90%**	**94%**	96.43%	**92.50%**	84.38%	**94.87%**
YOLOv3	Darknet-53	86.85%	93.78%	89%	91%	**96.62%**	90.48%	**84.89%**	92.64%

This work was supported by the The matic Research Project of Southwest Jiaotong University under 2682021ZTPY093, The National Social Science Fund Project under 2018G050008 and The Soft Science Research Project of Sichuan Provincial Department of Science and Technology under Q114720S01006

References

1. MHPRCCIPRB. People's Medical Publishing House(2007)(in Chinese)
2. Wang, H. , S. Zhao , and E. Zeng . SDCSCAP. Chinese Journal of Social Medicine. (2014)
3. Liu J, Yang W W, Wang Y, et al. OMVBAAPANN. *International Journal of Food Engineering*. (2011)
4. Deng H L, Xiao Y G. DGEIMSTC. *Applied Mechanics & Materials*.(2012)
5. Zhuang, Xiaodan , et al. AFDUGMMAGS. *IEEE International Conference on Acoustics, Speech and Signal Processing IEEE*.(2009)
6. Alwan, M. , et al. ASPFVBFDE. *International Conference on Information & Communication Technologies IEEE*(2006)
7. Rougier, Caroline , et al. FDHSMHUVS. *Advanced Information Networking and Applications Workshops*. (2007)
8. Foroughi, Homa , B. S. Aski , and H. Pourreza . IVSMFDEIHE. *Computer and Information Technology*. (2008)
9. Fang Shao-Mei, Guo Chang-Hong, Wu Pei, Lei Jian-Ping. MSREACI. (2009)
10. Joseph Redmon Ali Farhadi. YII. (2018)
11. C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, and A. C. Berg. DSSD. (2017)
12. M. Everingham, L. Van Gool, C. K. Williams, J. Winn, A. Zisserman. PVOCC. *International journal of computer vision*. (2010)