Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 1, No. 2, 2022

Yolov5-based fall detection algorithm for homebound people

Fumin Wang*

Queen Mary University of London Engineering School, NPU, Northwestern Polytechnical University, Xian, Shaanxi, 710072, China
* **Corresponding author**: Email: 664316139@mail.nwpu.edu.cn)

Abstract: According to the statistics from the official website of the National Bureau of Statistics, the population over 60 years old in China is close to 260 million, accounting for about 18.7% of the total population. As society moves towards aging, the safety problem of the elder people alone is gradually highlighted, and the accidental fall of the elderly at home has become a problem that cannot be ignored, and developing fall detection technology to mitigate the danger of falls of the elderly is imperative. So, this paper proposes a fall detection method which is based on deep learning. Specifically, the method is based on the Yolov3 method, and to enhance the accuracy and speed, the Yolov3 algorithm is updated to Yolov5s. The final experimental results also confirm that the above method has achieved the purpose.

Keywords: Yolov5s; Fall detection; Deep learning.

1. Introduction

At present, society is gradually aging, and the proportion of elderly people in the socio-demographic class is increasing, and the health and safety issues of the elderly are becoming more and more prominent. [1] Studies have shown that in China, nearly 20% of elderly people are seriously injured after a fall, even if they were very healthy before the fall, and most of them suffer secondary injuries because no one finds and sends them to the doctor in time after the fall, which is very common in elderly families living alone. [2] According to statistics, the number of elderly people living alone in China has exceeded 30 million, and it is a practical problem to detect falls and prevent secondary injuries caused by falls. Therefore, it is valuable to research and develop an accurate fall detection algorithm for the human body. It can detect the fall in time so that the elderly can get timely treatment and avoid causing secondary injuries. According to the current state of research at home and abroad [3,4], there are two kinds of the current mainstream algorithms on target detection. The first class of algorithms is to forecast the class and location of many targets directly and use only one convolutional neural network, which is represented by one-stage algorithms, Yolo and SSD are the representative algorithms of which. The second class, such as R-CNN, Fast R-CNN, Faster R-CNN and other two-stage R-CN algorithms which are derived from Region Proposal. In other words, the target candidate frame, i.e., the target location, needs to be generated before the candidate frame can be classified and regressed. In comparison, the second type of algorithm is slower but more accurate, while the first type is slightly less accurate but faster. Specifically, for fall detection, most of the existing algorithms are based on wearable devices [5], which are not ideal in terms of complexity and recognition capability. Therefore, studying the computer vision [6] in the field of fall detection may solve the above problems.

2. YOLO target detection algorithm family

2.1. Comparison of different algorithm series

Table 1. YOLO Target Detection Algorithm Series Comparison

	YOLO v1	YOLO v2	YOLO v3	YOLO v4	YOLO v5
Advantages	Detection speed is faster	Detection speed is better than FasterR-CNN, SSD, etc.	Fast detection speed	Efficient and powerful models	High flexibility
Disadvantages	The model relies on object recognition labeling data	Accuracy performance is not very good	The model complexity is high		Slightly inferior performance to YOLO v4

In the target detection field, YOLO v3 has been applied to this field and achieved good results. This paper aims to update the existing YOLO v3 fall detection research project to the YOLO v5 algorithm, which is more flexible and faster. In the next sections, I will present the innovations made in the network structure of YOLO v5 compared to YOLO v3, the creation of the dataset, the training of the model and the detection of the dataset. Based on the YOLO v5 algorithm, it makes the detection more accurate and faster.

2.2. Network structure diagram of Yolo v5

The structure of Yolo v5 is similar to that of Yolo v4, with more innovations compared to Yolo v3. There are four detection networks in Yolo v5, namely Yolo v5s, Yolo v5m, Yolo v5l, and Yolo v5x. The depth and the feature map width of Yolo v5s network is the smallest in the Yolov5 series, and its detection speed and AP accuracy are not excellent compared with the last three models, but it is a good choice if the monitoring target is mainly large targets and the speed is pursued. The model used in this paper is the Yolov5s

algorithm, and the following figure gives a view of the structure of Yolov5s network.

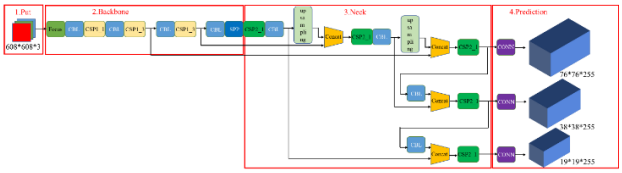

Figure 1. YOLO v5 network structure diagram

There are four aspects for the structure diagram of Yolo v5s: Input side, Backbone, Neck, and Prediction. Compared with the Yolo v3 algorithm, it has made innovations in all four parts：(1) Input: Enhance mosaic data and adaptive anchor frame calculation (2) Backbone: Focus structure and CSP structure (3) Neck: FPN structure and PAN structure (4) Prediction: GIOU_Loss

2.2.1. Inputs

Mosaic data enhancement

Yolo v5 carries out Mosaic data enhancement by randomly using 4 images, scaling randomly, and then randomly distributing them for stitching, the detection dataset was greatly supplemented, especially the random scaling added many small targets, which made the network richer. This improvement is very beneficial for enhancing the detection of small objects.

Adaptive anchor frame calculation

Yolov5 embeds calculation of initial anchor frame value into the code, the optimal anchor frame values in different training sets are updated in each training session with adaptive calculations.

2.2.2. Backbone

Focus structure

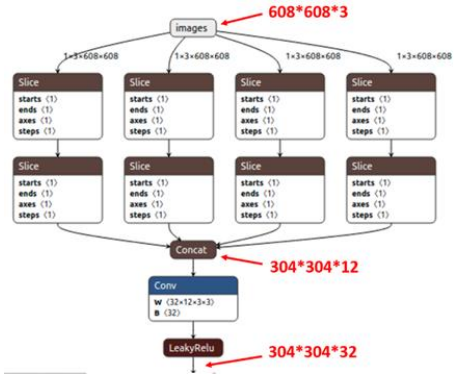

Figure 2. The Focus structure

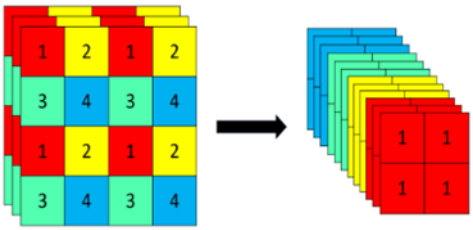

Figure 3. Slicing operation

In the Focus structure, Input the original 608×608×3 image to the Focus structure, use slicing operation to turn it into a 304×304×12 feature map first, and then a convolutional operation is then performed, which has 32 convolutional kernels, which are finally converted into 304×304×32 feature maps. The tile diagram on the right shows a 4×4×3 image with tiles as 2×2×12 feature maps.

2.2.3. Neck

Yolov5 now uses the same structure of FPN+PAN for Neck as in Yolov4. The FPN (Feature Pyramid Network) [7] is used as a reference, and the final output three feature maps (19∗19, 38∗38, 76∗76) have different scales by up-sampling and feature fusion. In order to improve the detection of large and small targets, the feature richness can be increased through multi-scale feature detection to achieve the purpose of detection [8].

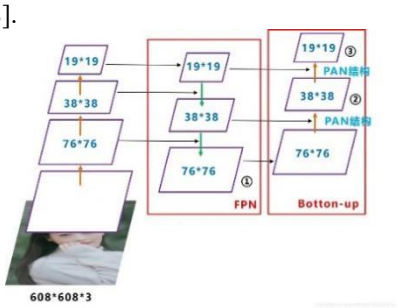

Figure 4. Neck structure

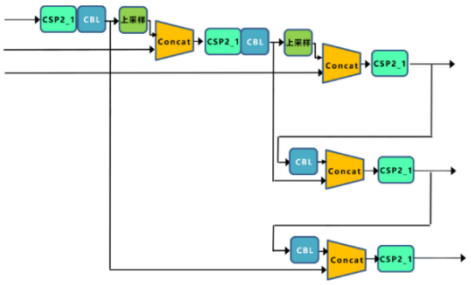

Figure 5. YOLOv5_Neck

Yolov5's Neck structure is enhanced with a CSP2 structure borrowed from the CSPNet design for network feature fusion.

2.2.4. Outputs

Bounding box loss function

CIOU_Loss is used as the loss function of the Bounding box in Yolov5:

$$CIOU\_Loss = 1 \sim CIOU = 1 - (IOU - \frac{Distance\_2^2}{Distance\_C^2} - \frac{v^2}{(1-IOU)+v})$$

Where v is a parameter measuring the consistency of the aspect ratio, which we can also define as:

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w^{p}}{h^{p}})^2$$

The introduction of this influence factor v takes the aspect ratio of both the prediction box and the target box into account. Therefore, CIOU_ Loss takes into account three important geometric factors that should be considered in the target frame regression function: overlap area, centroid distance, and aspect ratio. This makes the regression of the prediction frame faster and more accurate.

3. Experimental results and analysis

3.1. Dataset

For deep learning, the final detection effect is closely related to the quality of the dataset. A prerequisite for the network to be able to fully understand the features of the detection target is to have a sufficient number of samples for the network to learn. The dataset used in this experiment is made from open datasets available on the Internet. In the model training for this experiment, it is planned to divide the dataset into training sets and validation sets in an 8:2 ratio.

The xml-formatted label file is converted to a txt-formatted label file.

3.2. Model training

In this experiment, the training environment settings are: python programming, TensorFlow open-source framework 1.14, keras open-source framework 2.2.4, cuda10.1 cudnn7.4, video card 2070s. The training of model: that is, fitting the network parameters in the model to make the predicted values more and more accurate. This experiment uses the loss function to calculate the error of the network. Then the obtained error inversion can adjust the internal network parameters, and the adjusted internal network parameters make the error smaller and smaller. The above process uses pre-trained weights downloaded from the Internet - yolov5s.pt with training epochs of 50. The training process and internal parameters can be viewed through the visualization tool tensor board, as shown below.

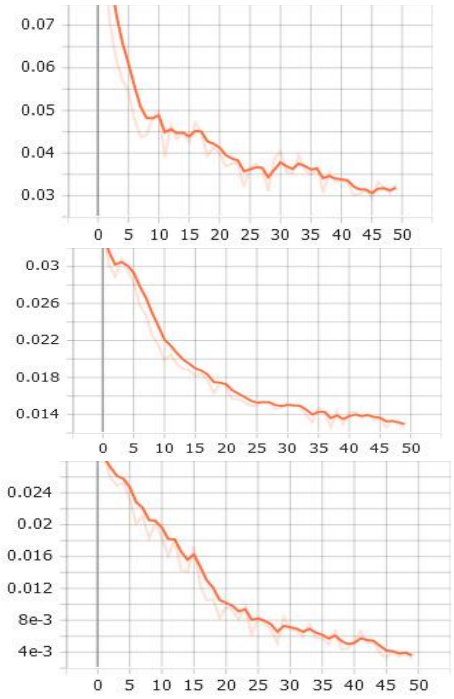

Figure 6. Training loss curves

3.3. Evaluation metrics

The evaluation metric chosen for this experiment is the average accuracy, that is, the mAP index. The average (average accuracy) of the AP for all object classes is called the mAP metric, and different recall values correspond to the APs of different object classes. These recall values are equal to or less than the values of maximum precision, whereby the value of maximum precision is obtained and used as an independent variable, and then the area under the curve is plotted as the AP for that target, as shown in Figure 7. [9] The miss rate of identified objects is called recall and the accuracy rate of identified objects is called accuracy. The calculation formula is shown below (1) and (2). Number of properly identified object classes – true positives, number of misidentified object classes – false positives, number of misidentified or identified target classes – false negatives.

$$recall = \frac{TruePositives}{TruePositives + FalseNegatives} (1)$$

$$precision = \frac{TruePositives}{TruePositives + FalsePositives} (2)$$

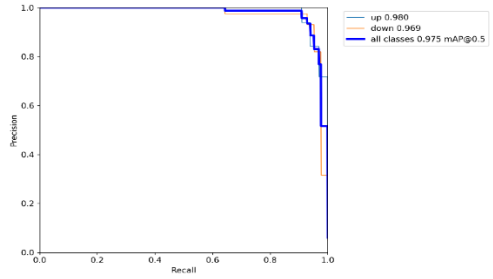

Figure 7. AP Calculation Diagram

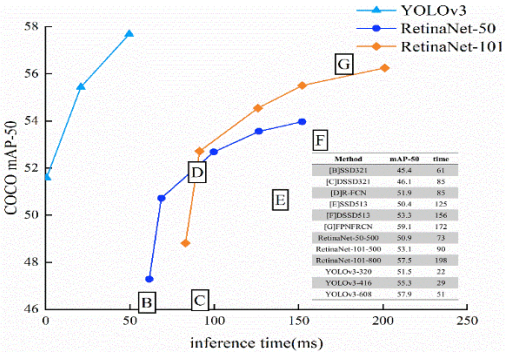

Figure 8. Comparison of common detection algorithms

3.4. Detection results

Currently, a comparison of commonly used object detection algorithms is shown in Figure 8 [10]. The detection results of the dataset show that the mAP-50 value of yolov3 algorithm is about 55, while the Yolov5 algorithm has a value of mAP-50 around 60 as obtained by Figure 5 above. This indicates that the YOLOv5 algorithm has a higher mAP. There are two detection results set up here: the up class and the down class. Figures 9 and 10 are the results of the corresponding test dataset.

Figure 9. No drop test results

Figure 10. Drop test results

The AP for the upward class is 0.980, the downward class of AP is 0.969, while the generated mAP is 0.0.975. The detection results obtained from this network model are compared with the detection results of other algorithms on the VOC dataset, it can be seen that our algorithm has further improved in terms of accuracy.

4. Conclusion

This experiment introduced a fall detection method which

is derived from deep learning, and this experiment updates the algorithm in the previous Yolov3 fall detection method to Yolov5s model, and the fall detection proceed with the latest Yolov5 network. The final results can confirm that the method introduced in this experiment makes the fall detection more accurate. In the next study research, I will work on applying semi-supervised deep learning networks to the study of fall detection in order to use unlabeled images and videos comprehensively.

References

[1] T. Xu, J. Chen, Z. Li and Y. Cai (2021) Fall Detection Based on Person Detection and Multi-target Tracking In: 2021 11th International Conference on Information Technology in Medicine and Education (ITME), pp. 60-65

[2] W J Choi, K Lim, S Kim et al. (2021) Science of Falling and Injury in Older Adults-Do All Falls Lead to Death? Literature Review. J. Physical Therapy Korea, vol. 28, no. 3, pp. 161-167.

[3] H Ramirez, S A Velastin, I Meza et al. (2021) Fall detection and activity recognition using human skeleton features. J. IEEE Access, vol. 9, pp. 33532-33542.

[4] M S Islam, H Shahriar, S Sneha et al. (2020) Mobile Sensor-based fall detection framework. In: Proceedings of 2020 IEEE 44th Annual Computers Software and Applications Conference (COMPSAC), pp. 693-698.

[5] A Núñez-Marcos, G Azkune and I. Arganda-Carreras (2017) Vision-Based Fall Detection with Convolutional Neural Networks. J. Wireless Communications and Mobile Computing, vol. 2017, pp. 1-16.

[6] K Chaccour, R Darazi, A H El Hassani et al. (2016) From fall detection to fall prevention: A generic classification of fall-related systems. J. IEEE Sensors Journal, vol. 17, no. 3, pp. 812-822.

[7] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He and Bharath Hariharan (2017) Feature Pyramid Networks for Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 936-944.

[8] Zhihua Zhou (2016) Machine Learning. In: Tsinghua University Press. Beijing.

[9] X. Wang and K. Jia (2020) Human Fall Detection Algorithm Based on YOLOv3. In: 2020 IEEE 5th International Conference on Image, Vision and Computing (ICIVC). pp. 50-54.

[10] E Casilari, R Lora-Rivera and F. García-Lagos (2020) A study on the application of convolutional neural networks to fall detection evaluated with multiple public datasets. J. Sensors, vol. 20, no. 5, pp. 1466.