Using YOLO-based pedestrian detection for monitoring UAV

Conference Paper · May 2019

DOI: 10.1117/12.2524219

7 authors, including:

Yanhua Shao
Southwest University of Science and Technology

Mei Yanying
Southwest University of Science and Technology

Hongyu Chu
Binzhou Medical University

Using YOLO-based Pedestrian Detection for Monitoring UAV

Depei Zhang[1], Yanhua Shao[1*], Yanying Mei[1], Hongyu Chu[1], Xiaoqiang Zhang[1], Huayi Zhan[2],Yunbo Rao[3]

1.School of Information and Engineering, Southwest University of Science and Technology, Mianyang Sichuan, China, 621010; syh@cqu.edu.cn.

2. Electrical Engineering and Computer Science, Northwestern University, Evanston 60208, USA;

3. Sch. Informat. & Software Engn. University of Electronic Science & Technology of China, Chengdu 610054, China.

ABSTRACT

Pedestrian detection (PD) is an important application domain in computer vision and pattern recognition. Conventional PD in real life scene is usually based on a fixed camera, which can only detect and track the pedestrians in the pre-defined monitoring region. However, when the pedestrian leaves the visible area of the fixed camera, it is usually difficult or impossible to monitor the pedestrian effectively. In response to the limitations of the conventional pedestrian detection scenarios in practical life, a four-rotor unmanned aerial vehicle (UAV) system equipped with a high-definition (HD) camera is designed and implemented to detect human targets. Considering the size of human body in aerial image is small and easily to be occluded, we draw on the advanced research results in the field of target detection and propose a robust pedestrian detection method based on YOLO (You Only Look Once) network. The flow of the proposed approach is as follows. Firstly, the HD camera, which is installed on the monitoring UAV, is used for capturing images of the designated outdoor area. Secondly, image sequences are collected and processed using the airborne embedded NVIDIA Jason TX1 and Ubuntu as the core and operating system, respectively. Finally, YOLO is used to train the pedestrian classifier and perform the pedestrian detection. Experimental results show that our method has good detection results under the complicated conditions of detecting small-scale pedestrians and pedestrian occlusion.

Keywords: YOLO, pedestrian detection, UAV, NVIDIA Jason TX1

1. INTRODUCTION

Pedestrian detection (PD) has become one of the most active topics in computer vision and machine learning, due to a wide range of promising applications [1]. However, in some areas of visual surveillance, for example, poachers detection in national nature reserves, search and rescue, due to the different application scenarios of human detection technology and the complexity of the environment. And because of the randomness of the moving human body, the diversity of postures, the difference of human body clothing, and the possible occlusion, the human body detection has certain difficulties and challenges in some specific scenarios. With the expansion of the application field of drones, the role of pedestrian detection based on drone platforms in monitoring, detection and other aspects is becoming more and more important. The small target scale of pedestrians is an important feature in aerial image. The problems caused by low target resolution and vulnerability to the environment pose serious challenges to existing pedestrian detection methods [2]. Therefore, finding a convenient and efficient pedestrian detection method is of great significance for pedestrian detection in special aerial scenes.

In order to cope with the problems of low detection efficiency, poor detection performance of fixed cameras in special scenes, and some scenes without a fixed camera, a pedestrian detection system based on You Only Look Once (YOLO) algorithm is designed [3]. The system acquires images through an on-board high-definition (HD) camera, and uses the YOLO model to detect and classify images to achieve detection of pedestrian targets for monitoring unmanned aerial vehicle (UAV).

The outline of the paper is listed as follows: Section 2 elaborates the details of the proposed framework and details the YOLO method of obtaining high-quality pedestrian detection. Section 3 shows the extensive experimental results on some real-world scenes. Finally, some conclusions of our work are drawn in Section 4.

2. THE PROPOSED ALGORITHM

A diagram of our proposed system is shown in Fig. 1. Apparently, our HD YOLO-based pedestrian detection for outdoor monitoring UAV is mainly composed of four steps:

1). Get HD Image Sequence through the onboard camera, 2). Image preprocessing, such as median filtering and so on, 3). Image recognition by YOLO, 4). The detection information transfer. In this design, the detection information can be transmitted to the remote controller and the ground station separately.

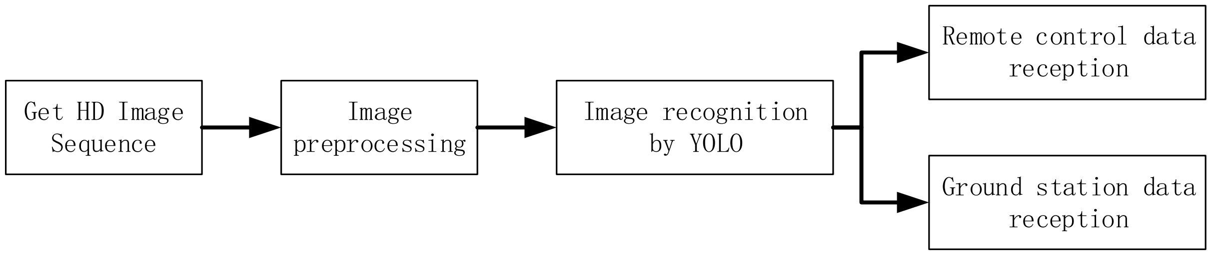


Figure 1. Processing stages of the proposed model.

2.1 The design of Monitoring UAV

The system configuration of our monitoring UAV is presented in as Fig. 2. By the figure below, the UAV is made up of such three main modules as a HD camera, NVIDIA TX1, and flying control system from theirs function view.

The YOLO detection is implemented on an NVIDIA Jetson TX1 microcomputer due to its light weight and good GPU based computational capability. As for the UAV, considering the platform payload capacity and implementation simplicity, a four-rotor frame with a self-designed autopilot is selected. The camera is mounted to the UAV directly. When performing the detection mission, the UAV is set to position-hold flight mode, where it uses the onboard GPS, barometer, and IMU to maintain its position and altitude at a set of given value. In this letter, the altitude of the UAV is set to be 7m; this value is set based on the image resolution and camera field of view (FOV) [4].

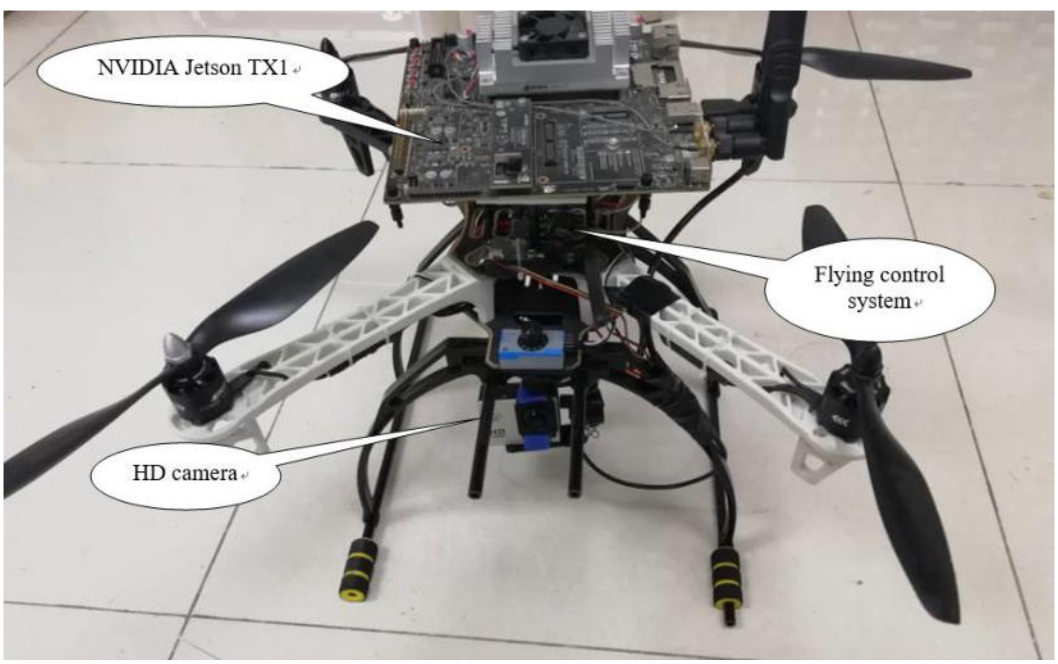


Figure 2. The four-rotor monitoring UAV.

2.2 The basic idea of YOLO

You Only Look Once (YOLO) is one of the emerging convolutional neural network configuration for object detection in recent years. Different from previous deep learning object detection approaches, YOLO treats the detection task as a one-step regression problem instead of a classification problem. The name YOLO comes from which the algorithm can localize the object and extract the object class within one scan [5-6].

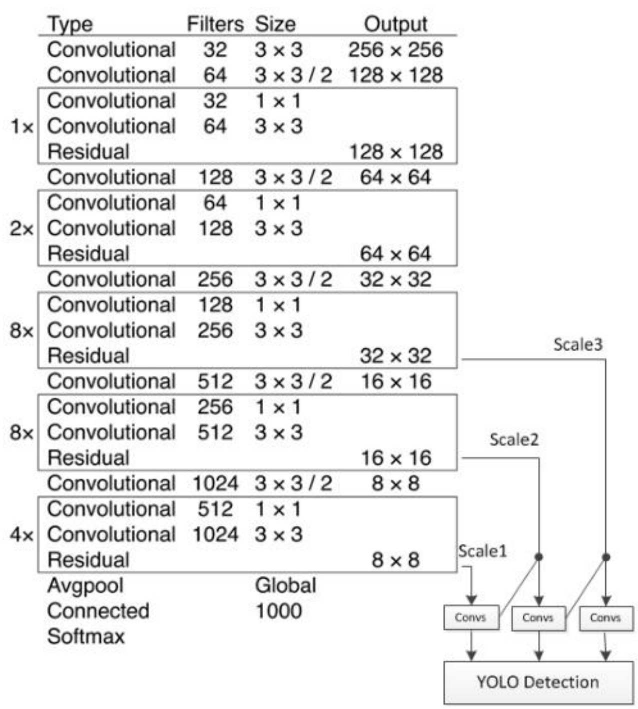
Compared with the traditional methods, the biggest feature of YOLO is that it does not use this kind of scanning based on Suggestions box, it scans the entire training image [7-8]. Once it can predict the location and category of all boxes and can better distinguish between goals and background. It direct return to the bounding box position and the bounding box belongs to the category in the finally output layer. This mechanism of YOLO makes that compared with most of the deep learning based object detection algorithms, the speed of target detection is faster, the accuracy of image detection is higher, but the positioning accuracy is not high. However, for our HD YOLO-based pedestrian detection for Outdoor monitoring UAV, the main performance index is good real time, and it can quickly find out if there is the target. So the program only need to consider the processing speed of the program and the accuracy of identification. Considering all these factors, this paper chooses the YOLO network [9].

As for the logic flow of YOLO, it first divides the image into small grids, then performs five bounding boxes and class probability prediction based on each grid. Then, it analyses distributions of all box and object class based on probabilities. Finally, according to a set threshold value, detection results fulfill the threshold value are considered as the output value. The loss function is designed with the combination of bounding box position, non-object bounding box confidence, with-object bounding box confidence, and object class.

2.3 YOLOv3

The YOLO algorithm was first proposed in CVPR 2016 and has been developed so far, the latest version is YOLOv3 which was proposed in 2018. This article uses the latest YOLOv3 for pedestrian detection. Compared to the previous version, there are three main innovations [10].

First， feature extraction uses a new network. As shown in figure 3, new network is a hybrid approach between the network used in YOLOv2, Darknet-19, and that newfangled residual network stuff with a total of 53 convolutional layers called Darknet-53.



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

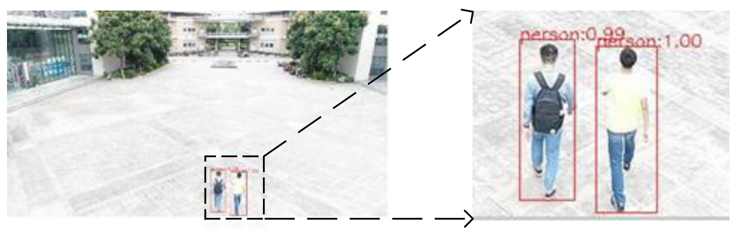
Figure 3. The network architecture of YOLOv3.

Second, prediction across scales. YOLOv3 draws on the FPN (Feature Pyramid Network) to predict boxes on three different scales. Small features are fused by upsampling and large scale. System extracts features from those scales using a similar concept to feature pyramid networks

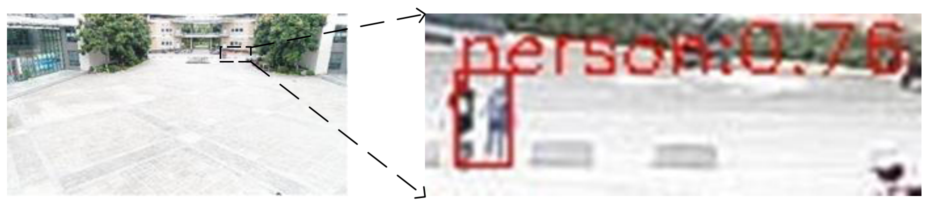
Third, class prediction. Instead of using softmax, binary cross-entropy loss is used for class prediction, which can support multi-tag prediction.

3. EXPERIMENTS AND RESULTS

In this section, we demonstrate the experimental results in some real-life scenarios, such as square scene as shown in Fig. 4a, 4b, garden scene as shown in Fig. 5a, 5b. Our evaluation monitoring UAV system includes a UAV flight control system, which adopted the STM32F427 processor with the embedded real-time operating system μC/OS-III, and an image processing system based on the NVIDIA TX1 running Ubuntu 16.04 and OpenCV, processing speed is 2 fps. The detection effect is shown in Fig 4 and 5.

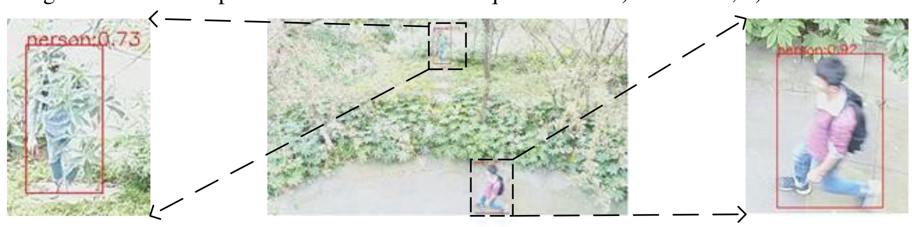


a)

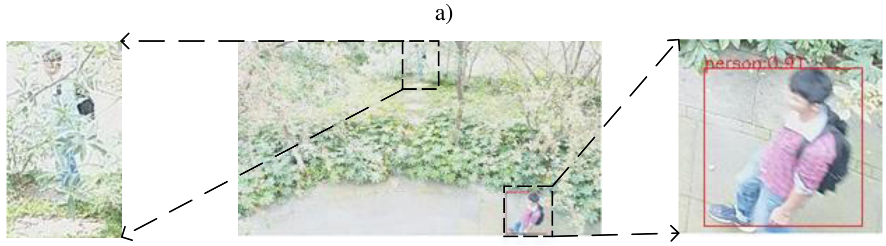


b)

Figure 4. Results of pedestrian detection in the square scene. a) close shot, b) distant view.



a)



b)

Figure 5. Results of pedestrian detection in the garden. a) Front and occlusion, b) sideway and occlusion.

In the early versions, YOLO has not performed well on the detection of small targets. However, with new multi-scale predictions, YOLOv3 has relatively high Average Precisions (APS) performance.

Fig. 4 show the test results on the square scene with different pedestrian sizes. According to the test results, it is known that in an open scene, even if the pedestrian target is small in the image, it can still be detected as shown in Fig.4b. But, by comparing the results of the test, when the pedestrian size is relatively large as shown in Fig.4a, the confidence obtained by the detection is higher. At the same time, Fig.4.b shows that the proximity detection feature of the algorithm is poor, when more than two small objects appear in one small cell or multiple different objects appear in one small cell, the effect is not so good. For example, in Fig.4.b, although there are two pedestrians, only one is detected. The reason is that YOLO divides the picture into small cells of S*S, and at the same time defaults all the same bounding boxes in the same grid as the same kind of objects.

Fig. 5 show the test results in the garden scene with different pedestrian gestures and direction. According to the test results, even in the case of occlusion, the pedestrian facing the lens can still be detected, but the confidence level has decreased. When the pedestrian is sideways, the detection effect is better when there is no occlusion, and it is difficult to detect when there is occlusion and at the same time it is a sideways situation as shown in Fig. 5b.

4. CONCLUSION

In this paper, an algorithm for a robust pedestrian detection method for outdoor monitoring UAV with embedded system based on the YOLO has been proposed. Our method shows promising results in some complex real-world scenes.

An open problem is still precision and efficiency of the algorithm. Future work will focus on more comprehensive sample training and optimization of the YOLO network in embedded system. For example, using google mobelnet or Tencent ncnn network, which is a high-performance neural network inference framework optimized for the mobile platform, can improve the processing speed of the algorithm in the embedded system.

ACKNOWLEDGMENT

The authors would like to thank all reviewers for very helpful comments and suggestions. This work is supported by the National Natural Science Foundation of China Fund No. 61601382, the Doctoral Fund of Southwest University of Science and Technology No. 16zx7148, the Scientific Research Fund of Sichuan Provincial Education Department No. 17ZB0454, the Fund of Robot Technology Used for Special Environment Key Laboratory of Sichuan Province under Grant No. 15kftk05, and Longshan academic talent research supporting program of SWUST (18LZX632).

REFERENCES

[1] Andreopoulos A, Tsotsos J K. 50 Years of object recognition: Directions forward [J]. Computer Vision & Image Understanding, 2013, 117(8):827-891.

[2] Mao, H. Z.,Yao, S.,Tang, T. Q., etc. Towards Real-Time Object Detection on Embedded Systems[J]. Ieee Transactions on Emerging Topics in Computing, 2018, 6 (3): 417-431.

[3] N. Tijtgat, W. V. Ranst, B. Volckaert, T. Goedemé, and F. D. Turck, "Embedded Real-Time Object Detection for a UAV Warning System," in 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 2110-2118.

[4] S. Xu, A. Savvaris, S. He, H. Shin, and A. Tsourdos, "Real-time Implementation of YOLO+JPDA for Small Scale UAV Multiple Object Tracking," in 2018 International Conference on Unmanned Aircraft Systems (ICUAS), 2018, pp. 1336-1341.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788

[6] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, 2017.

[7] R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440-1448.

[8] Ren, S., He, K., Girshick, R., and Sun, J.: 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', IEEE T Pattern Anal Mach Intell, 2017, 39, (6), pp. 1137-1149.

[9] Peng Q, Luo W, Hong G, et al. Pedestrian Detection for Transformer Substation Based on Gaussian Mixture Model and YOLO[C]. International Conference on Intelligent Human-Machine Systems and Cybernetics. IEEE, 2016:562-565.

[10] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv preprint arXiv:1804.02767,2018.