Computer Vision Studies
Volume 1, Issue 1, No. 1
Published online August 29 2022

Sanderman Publishing House

Recognition of tarmac from UAV view based on Yolo V3

Ming-peng Cai, Geng Yang, Jia-chun Zhou, Qin Li, Hong Lai
Shenzhen Institute & Information Technology, Shenzhen, China

Abstract. In this paper, training and testing of YOLO V3 in-depth learning object detection model were realized through acquisition and labelling of tarmac image data during flying process of unmanned aerial vehicle (UAV), and precise recognition of tarmac from UAV view was achieved at last, laying a good foundation for application of precise UAV landing. In this project, the most advanced visual in-depth learning object detection technology at present was applied, to recognize location of tarmac during landing process of UAV, thus enabling UAV to realize rough to precise positioning during the landing process, and realize civil UAV's function of making precise recognition of tarmac and precise landing in an innovative way, which can also be operated and realized in lost-cost onboard small edge computing module.

Key words. unmanned aerial vehicle (UAV), YOLO, recognition of tarmac, image data.

1. Introduction

UAV hardware equipment has already been mature and stable at present, and there are many mature hardware manufacturers, and among top 10 UAV hardware brands, there are 4 in Shenzhen and 6 in Guangdong-Hong Kong-Macao Greater Bay Area. UAV has begun to enter our life, and substituted artificial operation on site in various aspects of all walks of life with an important role, including ecological environment monitoring, agricultural modernization/agriculture, forestry and plant protection, geographical mapping, power line patrol, logistics distribution and movie and television photography and so on. In November 2016, the *"13th Five-Year" Plan for Development of National Strategic Emerging Industries* was issued by the State Council, requiring that: development of multi-purpose UAV, new-configuration aircrafts and other strategic aviation equipment shall be accelerated. One of the key generic technologies in *Development Planning for the New Generation of Artificial Intelligence* issued by the State Council in July 2017 is the intelligent technology of autonomous unmanned system, including studying on computer-vision-based location, navigation and recognition under complex environment.

Precise landing of civil UAV mainly relies on satellite positioning technology, however, satellite positioning especially GPS positioning technology is of insufficient precision in China, and the differential positioning equipment is expensive due to the limitations of differential station layout, besides, GPS chip and positioning technology is not conducive to wide promotion and application of precise landing of civil UAV in China due to limited by U.S. GPS technology. Regarding this problem, the most advanced visual in-depth learning technology at present can be applied, to enable UAV to realize rough to precise positioning during the landing process, and realize civil UAV's function of making precise recognition of tarmac and precise landing in an innovative way, which can also be operated and realized in lost-cost onboard small edge computing module. Visual recognition technology has developed from traditional filter image processing method to in-depth learning method, and at present, the leading in-depth learning object detection technologies in this field mainly include SSD [1], Faster R-CNN [2] and YOLO [3]. In this paper, advantages and disadvantages of three in-depth learning object detection technologies were briefly analyzed, and YOLO V3 technology with relatively excellent performance was selected to complete data acquisition, data labelling, data training and data verification tasks gradually, thus realizing precise recognition of tarmac from UAV view.

2. Technical principle

2.1. Model comparison and selection

YOLO V3 is a single-stage-structure object detection model, and its network structure has 3 characteristic patterns of different sizes, which correspond to features of deep layer, middle layer and shallow layer from top to bottom, respectively. The deep layer is of small feature size, with greater feeling, thus better for detection of large-scale objects, while the shallow layer is on the contrary, and is better for detection of small-scale objects. Compared with the previous YOLO Version, YOLO V3 has improved its detection effect of mAP and small objects by adjusting the feature extraction network and making use of multi-scale feature for object analysis. SSD is a method that uses single deep neural network to detect objects in the image, discretizes output space of the bounding box into a group of default boxes, and enables each feature map position to have different aspect ratios and scales. It is more accurate than the previously most advanced single excitation detector (YOLO). Faster R-CNN consists of two modules, where, the first module is deep convolutional network used for generating proposed area, and the second module is Fast R-CNN detector. The whole system is a single and uniform object detection network; due to existence of RPN, Faster R-CNN has greater accuracy than YOLO and SSD, but slower than them.

Data Acquisition →	Test sets: 3,535 pictures Verification sets: 883 pictures →	Different scenarios, altitudes, times and light rays
↓		
Data Labelling →	LabelImag open-source labelling software →	Data type Tarmac and h
↓		
Data Training →	learning rate Gradient number →	Simulate training loss Number of training rounds
↓		
Precision Evaluation →	mAP mAP and training times →	Actual recognition effect of picture

Figure 1. Recognition realization process of Yolo v3 model

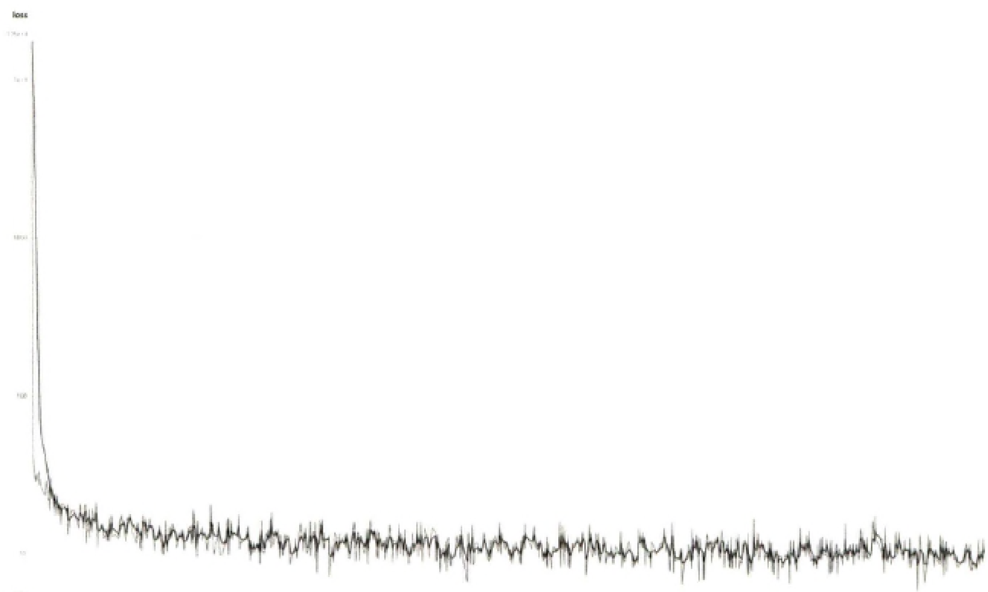

Figure 2. Relational diagram of model training loss and number of training rounds

mAP is the average value index measuring goodness of in-depth learning model in multi-class object detection. Since YOLO V3 can perform detection much faster than Faster R-CNN and SSD on the same mAP, while UAV requires real-time detection positioning during landing process, though YOLO V3 is inferior to the other two means in terms of detection precision, it is not so bad taking precision and speed into comprehensive consideration, therefore, YOLO V3 was used the model of tarmac object detection and recognition in this project.

2.2. Realization process

As shown in Fig. 1, the recognition process of YOLO V3 model consists of data acquisition, data labelling, data training and precision evaluation. Where, the acquired data includes the training sets and verification sets; data labelling includes labelling of both tarmac and h types, of which trarmac refers to the whole tarmac, and h refers to h character in the center of tarmac; data training process takes learning rate, gradient descent and loss conditions into comprehensive consideration; precision evaluation takes actual recognition effect of mAP and pictures into comprehensive consideration. The details will be introduced in the text below.

2.2.1. Data acquisition

During the data acquisition process, tarmac picture acquisition plan has been designed in different day and night scenarios, different times and UAV view from different altitudes, and the acquired data has been amplified technologically by means of image rotation, color switching, adding-noise and interception, etc., to enhance robustness of the data sets. The acquired data sets are divided into test sets and verification sets, where, the test data sets consist of 3535 pictures, and the verification data sets consist of 883 pictures.

2.2.2. Data labelling

Upon completion of data acquisition, LabelImg [4] open-source software was used to make labelling of the objects within the pictures. Data labelling mainly performs labelling of two classes, i.e. tarmac (most peripheral circle of tarmac) and h (H marking on tarmac). The purpose of labelling tarmac and h is to enhance robustness of tarmac recognition, so as to recognize the tarmac in a better way under different conditions (e.g. shield). When both tarmac and h are recognized at the same time, central average value of both can be taken as the reliable center; when only one of them is recognized, they may become backup for each other mutually; during flying process, UAV may make multiple recognitions in time sequence, and then make repeated and multiple judgements of the recognition results in time sequence, to obtain comprehensive results.

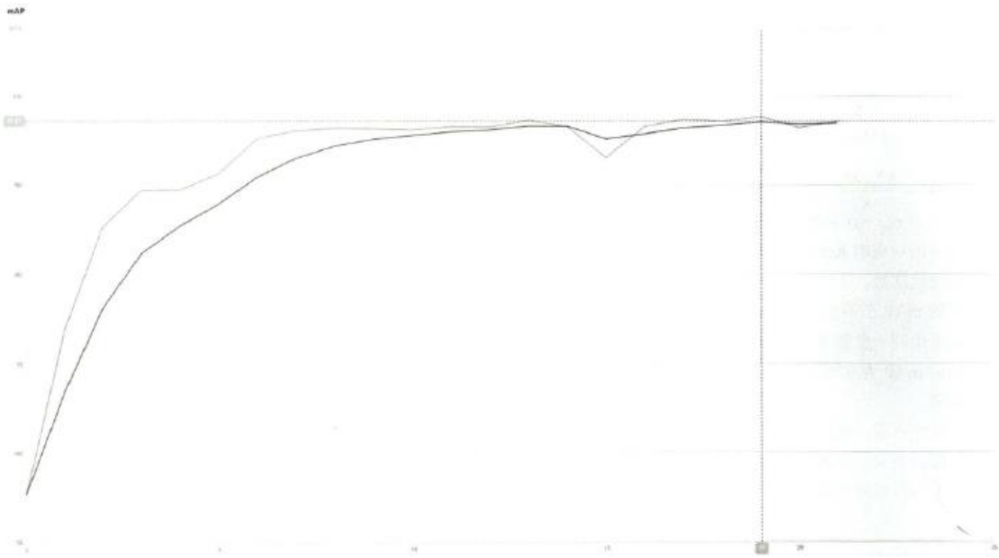

Figure 3. Change in relation between mAP and iteration times after a round of training

Figure 4. Tarmac recognition results from UAV view

2.2.3. Data training

PaddlePaddle is the first industrial-level in-depth learning independently researched and developed in China with complete functions, open source and openness. To support domestic information technology application innovation ("information innovation") business, PaddlePaddle framework was used in this paper to realize YOLO V3 [5], and

Darknet-53 network structure was also applied. To reduce over-fitting or under-fitting, initial value of the learning rate was set at 0.001 in this paper, and piecewise constant decay learning rate strategy i.e. PiecewiseDecay and learning rate optimization strategy – linear learning rate warmup i.e. LinearWarmup were used to make preliminary adjustment of the learning rate. In the meantime, Momentum optimization method was used to optimize the gradient descent direction during the training process, so as to accelerate convergence and reduce turbulence. PiecewiseDecay is refers to a learning rate of stepped descent, which can enable every N iterations of learning rate to decline to 1/K of the original one, where, N and K can be set independently. LinearWarmup can increase the learning rate gradually by the designed formula before normal adjustment of the learning rate, and then restore to the originally set learning rate in case the warmup step value is reached.

As shown in Fig. 2, the X axis refers to the number of training rounds, while the Y axis refers to the training loss, and the smaller the loss the better the training effect. After the loss becomes stable within a certain small scope, it means that the training has been basically converged. After 2,000 training rounds, loss of this model would change around the small range of 10 basically. In this experiment, a good result can be obtained upon convergence via 2,000 training rounds, for the reason that the tarmac pictures taken from UAV view are clear, stable with relatively good data sample; in the meantime, the model and training parameter setting selected conformed to characteristics of the sample data sets.

2.2.4. Precision evaluation

mAP may be used to measure precision of well-trained model, and calculation method of mAP is as shown in (1)-(4), where, TP refers to number of detection boxes whose intersection-over-union (IoU) > 0.5, FP refers to number of detection boxes whose IoU <= 0.5, Precision refers to the precision ratio, Recall refers to the recall ratio, AP refers to prevision evaluation of recognition results of a certain class (i.e. the area under the Precision-Recall curve), and then mAP refers to prevision evaluation of recognition results of all classes.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{FP}{TP+FP} \tag{2}$$

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{interp}(r_i + 1) \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{4}$$

Where, r1, r2, ---, ri refers to the Recall value corresponding to the first interpolating point of Precision interpolation period by ascending order. As shown in Fig. 3, the *X* axis is the number of iterations in a certain round of training, and it can be seen in this round of training that, with increase in number of iterations, mAP of verification sets has been increased constantly to around 97%. During the flying process, UAV can make comprehensive decision-making and judgement using results of the same position in multiple time points, therefore, the current mAP of 97% can be used in practical application and guarantee accurate comprehensive decision-making.

At the end of training, inference code can be performed to make recognition of real images by calling the execution code, as shown in Fig. 4, the tarmac (green box) and H in tarmac (red box) from UAV view can be labeled and recognized in a good way.

2.3. Lightweight package

UAV control software is generally loaded in arm or Android device, of which Android device is the most common. Such device requires miniaturized and lightweight AI algorithms for valid operation due to its own limited computing and storage resources. Therefore, the original model shall be converted to lightweight and then be packaged into SDK available to be called by Android device. PaddlePaddle's model conversion tool i.e. Paddle Lite [6] is an in-depth learning inference framework with high performance, lightweight, strong flexibility and easiness to extend. The inference framework is positioned to support multi-hardware platforms including mobile, embedded and server terminals.

PaddlePaddle model conversion tool i.e. Paddle Lite can convert well-trained model into model document applicable to Android CPU architecture; C++ code was used to call model document to predict and output results of the images to be processed; Java code was used to call well-packaged C++ code via NDK and receive results, and then packages this project into Android SDK in aar format available for to be called. After acquiring images from UAV view, UAV Android control terminal (normal Android tablet) can call such SDK to send the images, thus obtaining recognition results of tarmac. Upon test on normal Android tablet, it would take 200ms-600ms to process a picture taken from UAV view to acquire recognition results of the tarmac, which is close to real-time processing, thus meeting requirements for practical application scenarios completely.

3. Summary

In this paper, a series of operations were carried out on the UAV to take actual pictures of the tarmac, and data of tarmac was acquired in different time periods, places and distances, etc., thus forming two types of data sets i.e. training sets and test sets. Before training, the pictures were pretreated, and Labelling was used to perform labelling of the pictures. YOLO V3 model was established, and the data was passed into the model for training. When both the training loss and precision meet requirements, recognition box results of tarmac and H character in tarmac H in the image from UAV view can be acquired by calling and executing the inference code. At last, it is available for direct call and use by UAV Android control terminal via lightweight package, thus realizing precise recognition of tarmac.

The purpose to recognize tarmac in this paper is to provide ground information for precise landing of UAV. The tarmac pattern recognized is the general tarmac pattern both at home and abroad, and is available for most of the scenarios, without additional QR code and other patterns. In combination with angle of view of UAV camera and the flying altitude acquired by the UAV's own sensor, a three-dimensional reference system with UAV as the center can be established, and then relative positional relation between UAV and the tarmac can be calculated via geometrical relationship; when positional relation between the center of UAV and the center of tarmac meets the landing conditions, vertical landing is available, otherwise, the movement will be automatically controlled in accordance with the positional relation until the conditions are met. Since the UAV's own lighting system at night can provide a clear fill light for the object below its viewing angle, therefore, the tarmac recognition algorithm function realized in this paper can adapt to application scenarios under different lights very well. In addition, processing time required by lightweight algorithm is near real-time, so sampling or continuous recognition can be performed in accordance with practical application demands. To sum up, the contents realized in this paper have laid a solid technical foundation for precise landing of UAV.

References

[1] FAN Qin-min. SSD Object Detection Based on Multi-layer Feature Fusion [D]. Sichuan: Southwest Jiaotong University, 2018.

[2] REN Shao-qing, HE Kai-ming, Ross Girshick, SUN Jian, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, arXiv: 1506. 01497

[3] RAN Xue. Algorithm Design and Realization of YOLO-based Object Detection [D], Chongqing: Chongqing University, 2020.

[4] LabelImg Open-source Software. [2021-11-09]. https://github.com/tzutalin/labellmg.

[5] PaddlePaddle YOLO V3. [2021-11-09]. https://www.paddlepaddle.org.cn/modelbasedetail/yolov3.

[6] PaddlePaddle Paddle Lite. [2021-11-09]. paddlepaddle.org.cn/tutorials/projectdetai1/1978748.