Research on Algorithm of fighter landing gear in bad video image

Miao Tang¹, Wei Fang²*, Wenjun Yan³ and Qing Ling⁴

¹Institute of Information Fusion, Naval Aeronautial University, Yantai, Shandong, 264001, China
²Institute of Information Fusion, Naval Aeronautial University, Yantai, Shandong, 264001, China
³Institute of Information Fusion, Naval Aeronautial University, Yantai, Shandong, 264001, China
⁴Aeronautical Communications Teaching and Research Office, Naval Aeronautial University, Yantai, Shandong, 264001, China

*Corresponding author’s email: bluehawkfw@163.com

Abstract. The retracted and retracted state of landing gear is a problem that needs to be paid close attention to in the landing process of fighter. The traditional visual observation is easy to be affected by the climate environment. Aiming at the problems of small target, low definition and difficult detection of fighter landing gear in bad video images, a YOLOv4 target detection algorithm based on image enhancement is proposed to detect fighter landing gear. In order to improve the accuracy of landing gear detection, ACE algorithm is first applied to the video image to enhance the video image, then the main feature extraction network of YOLOv4 is used to extract the feature information, and K-means clustering algorithm is used to cluster and analyze the number and aspect ratio dimension of target candidate frames. Experiments on the dataset show that the detection accuracy after image enhancement reaches 88.12%, which is 1.19% higher than the original image; The recall rate reached 73.44%, an increase of 4.17%; The value of mAP reached 82.65%, an increase of 1.08%.

1. Introduction

The safety of fighter landing is one of the issues of great concern to the world military powers. It is vulnerable to many factors, such as weather, equipment integrity, landing platform and so on. Among them, the retracted and retracted state of the landing gear is a problem that needs to be paid close attention to in the landing process. The traditional visual observation is easy to be affected by the climate environment, and occasional errors will cause incalculable losses.

In recent years, the target detection algorithm based on deep learning has been widely used because it can automatically extract the characteristics of the target from the image, has strong learning ability, fast recognition speed and high accuracy. At present, the main popular target detection algorithms are Faster R-CNN [1], SSD [2] and YOLO [3]. Faster R-CNN has the highest accuracy, but the speed of detection is very slow, which can not meet the requirements of real-time; SSD algorithm has the lowest detection performance; YOLOv1 performs well in speed, but its generalization ability for small objects is weak. The detection accuracy and speed of YOLOv2 are higher than those of the previous generation, but the improvement of detection accuracy is not obvious. YOLOv3 [4] achieves a balance...
between detection accuracy and speed. Compared with YOLOv3, YOLOv4 [5] has higher accuracy and no reduction in detection speed.

Although these algorithms have achieved good results in many specific datasets, they still face many challenges in fighter landing gear detection. First, the video data itself is not very clear, resulting in low definition of the image after frame cutting; Second, the movement of the fighter will lead to a certain degree of deformation of the landing gear in the intercepted picture; Third, the landing time of fighter is short, and the detection of landing gear needs high real-time and accuracy.

Therefore, this paper combines automatic color equalization [6] (ACE) in image enhancement with YOLOv4 for fighter landing gear detection, so as to reduce the probability of human error and improve the detection accuracy.

2. Image enhancement algorithm

At present, several typical algorithms to change the image quality by adjusting the pixel value of the image mainly include histogram equalization, multi-scale retinal enhancement with color recovery (MSRCR), automatic color equalization (ACE) and so on. This paper mainly uses ACE to enhance the video image of fighter landing.

ACE corrects the pixel value by calculating the difference between the target point and other pixel points to obtain the brightness difference information between pixel points. Firstly, the color difference correction of the image is completed by adjusting the color domain / spatial domain of the image, and the spatial domain reconstructed image is obtained. The calculation formula is

\[
R_c = \sum_{j \neq p} \frac{r(I_c(p) - I_c(j))}{d(p, j)}
\]

(1)

Where \( R_c(p) \) is the intermediate result, \( I_c(p) - I_c(j) \) is the brightness difference of 2 points, \( d(p, j) \) represents the distance measurement function, \( r(*) \) is the brightness expression function, it needs to be an odd function, this step can adapt to the local image contrast, \( r(*) \) can enlarge small differences, enrich large differences, and expand or compress the dynamic range according to local content. In general, \( r(*) \) is:

\[
r(x) = \begin{cases} 
1, & x < -T \\
\frac{x}{T}, & -T \leq x \leq T \\
-1, & x > T 
\end{cases}
\]

(2)

Then the corrected image is dynamically expanded. Finally, the three RGB channels are linearly expanded and dynamically stretched to obtain the final image. ACE can better process image details, realize color correction and improve image brightness and contrast, and has a good image enhancement effect.

3. Landing gear target detection

3.1. The network structure of YOLOv4

YOLOv4 adopts CSPDarknet53 as the backbone feature extraction network, and the network structure is shown in figure 1.
For the input images of 416×416, after the extraction of the backbone feature network, the sizes of three feature images are 52×52, 26×26 and 13×13 respectively. The SPP [7] structure of the spatial feature pyramid is doped in the last feature convolution layer of CSPDarknet53. After the last feature layer is convoluted three times, it is processed by three different scales of maximum pooling. The sizes of the maximum pooling nuclei are 13×13, 9×9 and 5×5 respectively. It can greatly increase the receptive field and separate the most significant contextual features, and the network processing speed has not decreased significantly. In addition, yolov4 also uses the path aggregation network PANet [8] on the three effective feature layers. After the top-down feature extraction, the PANet network structure also realizes the feature extraction from top to bottom, shortens the distance between the lowest layer and the top layer, avoids the loss of a large amount of information and improves the accuracy of prediction.

3.2. K-means clustering adjusts a priori box size

The proposal of a priori frame transforms target detection into two problems at the same time, that is, whether there is a target in the fixed grid and judging the distance between the prediction frame and the real frame. The size of the a priori frame is very important for the YOLOv4 target detection network. The original YOLOv4 a priori frame is clustered on the VOC dataset. However, the type of target to be detected on the video image data set of fighter landing is only the landing gear. The direct use of the original a priori frame can not meet the recognition requirements. In order to improve the detection accuracy of landing gear, The K-means clustering algorithm is re used to cluster the width and height of the data set in this paper, and a new a priori frame size is obtained. Firstly, cluster the labeled data sets, continuously increase the number of cluster centers, and calculate the distance from each data to each cluster center through iteration. The distance calculation formula is

\[ d = 1 - IoU \]  

(3)
The distance function is used as the optimization criterion function for clustering until the criterion function reaches the minimum, and the clustering ends. This will improve the coincidence rate between the prediction boundary box and the marking box of the real target to a certain extent.

For the dataset in this paper, using the above methods for cluster analysis, the dimensions of the a priori frames are (4,7), (4,9), (5,7), (5,10), (6,13), (7,16), (9,12), (10,21), (13,31) respectively.

3.3. Box regression function

YOLOv4 uses the Complete-IoU (CIoU) [9] as the box regression function. CIoU can obtain better convergence speed and accuracy on the rectangular box regression problem. CIoU takes into account the distance between the target and the center point, overlapping area, aspect ratio and penalty term, so that the regression of the target frame becomes more stable, and the function of the penalty term is to control the width and height of the prediction frame to be close to the width and height of the real frame as quickly as possible. CIoU’s penalty items are

\[
R_{\text{CIoU}} = \frac{\rho^2(b, b^\text{gt})}{c^2} + \alpha v
\]

among

\[
\alpha = \frac{v}{1 - \text{IoU} + v}
\]

\[
v = \frac{4}{\pi} (\arctan \frac{w^\text{gt}}{h^\text{gt}} - \arctan \frac{w}{h})^2
\]

The CIoU loss function \(L_{\text{CIoU}}\) formula is

\[
L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^\text{gt})}{c^2} + \alpha v
\]

Among \(\rho^2(b, b^\text{gt})\) represents the Euclidean distance between the center point of the prediction frame and the real frame, \(c\) represents the diagonal distance of the smallest rectangle containing both the prediction box and the real box, \(\alpha\) represents the trade-off parameter, \(v\) represents the consistency of aspect ratio, \(\text{IoU}\) represents the intersection and union ratio of the prediction frame and the real frame, \(w^\text{gt}\) represents the true frame width, \(h^\text{gt}\) represents the real frame height, \(w\) represents the width of the prediction frame, \(h\) represents the height of the prediction frame.

4. Experimental process and result analysis

4.1. Dataset

The dataset used in this paper is based on the video data of a fighter landing process. The video data is processed by OpenCV to obtain a single frame picture to form a data set, a total of 844 pictures, which are divided into training set and test set according to the ratio of 9:1. In this study, the data set is transformed into VOC dataset format. Labelling tool is used to manually label the rectangular frame of the landing gear in the picture, and the real frame ground truth is obtained for training. Among them, 573 pictures are not marked because the front landing gear can not be correctly recognized by human vision.

4.2. Experimental environment setting

In this paper, the CPU is Inter core i9 processor, 32GB internal memory, and the GPU processor is NVIDIA RTX 3080Ti; The experimental platform is windows10; The software environment is Python 3.7, Anaconda 3, CUDA11.0.

4.3. Model training

The input image size is 416\(\times\)416 pixels, and the cosine annealing attenuation method is adopted. Cosine annealing can reduce the learning rate by cosine function. The training parameters are set to 2
training samples per iteration, the minimum learning rate is 0.00001, the maximum learning rate is 0.001, the number of iterations is 100, and the threshold is set to 0.5.

4.4. Evaluation criterion
This paper evaluates the performance of the model according to the accuracy rate, recall rate and mAP. The formula of accuracy rate and recall rate is as follows:

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{8}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{9}
\]

\(TP\) represents the number of positive samples in the predicted positive samples, \(FP\) represents the number of false samples in the predicted positive samples and \(FN\) represents the number of positive samples in the predicted negative samples. Recall rate represents the probability of being predicted as a positive sample in the actual positive sample, which reflects the ability of the model to identify the target; The accuracy rate represents the prediction accuracy in the positive sample results, which reflects the accuracy of model detection. However, there are limitations in using accuracy and recall as evaluation indexes alone. Therefore, the average accuracy map is introduced as evaluation index, which is one of the most important indexes to evaluate the performance of target detection algorithm.

5. Analysis of experimental results
In this paper, the algorithms of MSRCR and ACE are used to enhance the video images of fighter landing respectively. The image effect is shown in figure 2.
It can be seen from the image effect that the color of the image obtained by using the MSRCR enhancement algorithm changes and the image is dark. The brightness and color of the image obtained by ACE algorithm are improved, and the enhancement effect is better than MSRCR. Therefore, ACE algorithm is selected for image enhancement in this paper.

In order to verify the detection accuracy of several popular target detection algorithms in the data set in this paper, SSD, YOLOv3 and YOLOv4 are selected to detect the landing gear of fighter respectively. The detection results are shown in table 1.

| model  | mAP(%) | time(ms) |
|--------|--------|----------|
| SSD    | 70.12  | 27.17    |
| YOLOv3 | 78.32  | 25.25    |
| YOLOv4 | 81.57  | 24.14    |

It can be seen from the table that based on the data set in this paper, YOLOv4 significantly improves the detection accuracy compared with the other two algorithms, and the detection speed is not much different from the other two algorithms, which is slightly improved, but it can still meet the needs of real-time detection. Overall, the detection effect of YOLOv4 is better. Therefore, this paper selects YOLOv4 algorithm for Fighter landing gear detection.

In order to verify the impact of the enhanced ace algorithm on the detection results, the original data set and the enhanced data set are respectively input into the YOLOv4 network, trained with the same training parameters, and 100 iterations are carried out to calculate the P, R and map of each algorithm respectively. The results are shown in table 2. The detection effect is shown in figure 3.

| algorithm   | Recall(%) | Precision(%) | mAP(%) |
|-------------|-----------|--------------|--------|
| Original drawing | 69.27     | 86.93        | 81.57  |
| ACE         | 73.44     | 88.12        | 82.65  |

As can be seen from table 1, compared with the original image, the accuracy of the image enhanced by ACE is improved by 1.19%, the recall is improved by 4.17%, and the map is improved by 1.08%. It can be seen from the comparison diagram of detection effect in Figure 3 that the front landing gear of the fighter is missed in Figure 3 (a) and (b), which is mainly due to the loss and distortion of the front landing gear caused by the change of flight angle during the movement of the fighter; The rear landing gear can be well detected. The enhanced detection probability is higher and the detection effect is better.
6. Conclusions
Aiming at the problems of small landing gear target and difficult detection in bad video images, this paper proposes to combine ACE and YOLOv4 algorithm to complete landing gear detection. The experimental results show that the accuracy, recall and map of the enhanced image have been improved compared with the original image, which can meet the purpose of real-time detection. This method can reduce the probability of error caused by human fatigue and provide a certain safety guarantee for the landing of fighters in different environments. However, the detection of the front landing gear is still missing. The next step will continue to carry out image enhancement and algorithm optimization to obtain better detection effect.

Acknowledgments
This paper is one of the phased achievements of the project supported by the National Natural Science Foundation of China (91538201), the special fund for Taishan Scholars project (ts201511020) and the Key Laboratory of information system security technology (614211190404).

References
[1] Ren S Q, He K M, Girshick R, et al. (2017) Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1137-1149.
[2] Liu W, Anguelov D, Erhan D, et al. (2016) SSD: single shot multibox detector [M]/Leibe B, Matas J, SebeN, et al. Lecture notes in computer science. Cham: Springer,9905:21-37.
[3] Redmon J, Divvala S, Girshick R, et al. (2016) You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, Las Vegas, NV, USA. New York: IEEE, 779-788.
[4] Redmon J, Farhadi A. (2018) YOLOv3:an incremental improvement [J]. arXiv preprint arXiv: 1804.02767.
[5] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection [J]. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[6] Getreuer P. (2012) Automatic Color Enhancement (ACE) and its Fast Implementation[J]. Image Processing on Line, 2:266-277.
[7] HE K, ZHANG X, REN S, et al. (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916.
[8] LIU S, QI L, QIN H, et al. (2018) Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8759-8768.
[9] ZENG Z, WANG P, LIU W, et al. (2020) Distance-IoU loss: faster and better learning for bounding box regression[C]//AAAI, 12933-13000.