Research on Infrared Image Target Detection Technology Based on YOLOv3 and Computer Vision

Hao Lv*, Desheng Jing, Wentao Zhan

Xi'an Aeronautics Computing Technique Research Institute, Aviation Industry Corporation of China, Xi’an, China

*Corresponding author: haolv2010@mail.nwpu.edu.cn

Abstract. Target detection technology is one of the basic topics in the field of computer vision, and it is also a hot spot, with a very broad application market. However, most of the current target detection technologies based on deep learning are aimed at visible light imaging technology, and there are very few researches on infrared imaging technology. Target detection based on deep learning implements the learning of more features by abstracting, extracting, processing and integrating the essential features of a large number of samples. Therefore, the use of deep learning target detection algorithms for infrared image pedestrian detection applications can make up for the shortcomings of traditional detection methods. YOLOv3 is currently a relatively balanced recognition algorithm. This article analyzes the principles and characteristics of the YOLOv3 series of algorithms to optimize multi-scale detection, which improves the detection accuracy and achieves a relative balance between detection accuracy and speed to a certain extent. This research hopes to provide efficient and feasible solutions and solutions for infrared target detection and recognition in the air through the application of deep learning technology.

Keywords: CNN; YOLOv3; Infrared Image; Target detection.

1. Introduction

Target detection technology is a very important research branch in the field of computer vision. With the development of information technology, digital image has become an indispensable information transmission medium in social production and people's life, Massive image data will be generated all the time.

Infrared thermal imaging refers to the use of photoelectric technology to capture and detect the thermal radiation signals of objects, and convert these signals into graphics or images that can be perceived and received by human vision. Therefore, the infrared thermal imager can image the surrounding environment and detection targets under the conditions of complete darkness, strong light irradiation, thick smoke, haze and no more than moderate rain. It can greatly make up for the deficiency of visible light imaging restricted by lighting conditions and meet the needs of all-time and all-weather observation. In national defense and security applications, it is one of the core tasks to monitor specific targets in infrared images.
The basic principle of the target detection algorithm based on deep learning is to use a large number of convolution operations to realize the detection algorithm, independently abstract and extract the features of the target in the image, splice and summarize the extracted features to obtain more features and improve the accuracy of target detection [1].

The basic principle of the target detection algorithm based on deep learning is to use a large number of convolution operations to realize the detection algorithm, independently abstract and extract the features of the target in the image, splice and summarize the extracted features to obtain more features, and improve the accuracy of target detection. Thanks to the rapid development of artificial intelligence and related technologies in recent years, Deep learning target detection algorithm has made rapid progress, and has achieved very excellent detection results in visible images. However, compared with visible images, infrared images also have some characteristics that are not conducive to target detection. The main technical difficulty lies in [2]:

(1) It is difficult to detect infrared target under complex background because the image contour and edge of infrared image are fuzzy;

(2) The signal-to-noise ratio of infrared image is low due to the influence of infrared detection noise;

(3) The infrared target is weak, the information after imaging is not obvious, the resolution is poor and the contrast is low.

In the field of infrared target detection, opportunities and challenges coexist. In the continuous efforts of researchers at home and abroad, the theory and technical methods of infrared target detection are constantly improved, and the target detection technology in infrared scene is developing by leaps and bounds.

2. Related Technology Status

2.1. Deep learning neural network algorithm

Deep learning detection algorithms are mainly divided into two categories: two-step detection algorithm and one-step detection algorithm. The two-step detection algorithm divides the detection process into two stages. The first stage task is to divide the image candidate area, and the second stage task is to judge the possible targets in the candidate area. The main representatives of such methods are region based convolution neural networks (R-CNN) algorithm and subsequent improved algorithms such as fast R-CNN and fast R-CNN. Single step detection algorithm combines region selection and detection judgment to form an integrated detection network. The main representatives of such algorithms are SSD single shot multibox detector (SSD) algorithm and you only look once (Yolo) Yolo detection algorithm.

Aiming at the low detection accuracy of the first generation Yolo model, Liu et al. [3] proposed an SSD model with both high average detection accuracy and real-time performance. Compared with the first generation Yolo model, SSD model has a better effect on the detection of small-size images and small targets. When VOC 2007 test is input, the average image detection accuracy can reach more than 75.1%. On the other hand, Redmon et al. [4] constructed Yolo V2 on the basis of the first generation Yolo model, which not only maintains the classification accuracy, but also improves the recall rate and positioning accuracy. In 2017, the third generation model of the Yolo family, Yolo V3, came out. Yolov3 model uses darknet-53 to extract image features, and then uses Yolo layer for multi-scale prediction. It is Yolo Series integrator.
2.2. The Status of Infrared Target Detection Technology

All objects whose temperature is above absolute zero will radiate infrared to the outside world, and the infrared spectrum is between microwave and visible light, its wavelength is 1mm to 760nm, and people can’t observe infrared with the naked eye. Target detection and recognition in infrared scene is an important aspect of infrared imaging application, and infrared target detection technology also has far-reaching significance in regional protection and information warfare. However, compared with visible images, infrared images have lower resolution and contrast, fuzzy target edge feature information, and more noise. These problems set many challenges for the development of infrared target detection technology. Therefore, how to overcome the difficulties in infrared target detection technology and improve the detection accuracy and speed is the current development requirement.

Traditional infrared target detection methods include the following modes:

1. The status quo of infrared target detection technology is based on the method of searching regions of interest (ROIs) by target characteristics. This mode applies a selective algorithm to gradually traverse the entire image, but the consequence of traversal is that many areas are repeatedly scanned.
detected, which wastes calculation time and the overall detection efficiency is relatively low. In order to improve detection performance, researchers have proposed a target detection method for segmenting infrared images, the most basic of which is background subtraction. If the pixel difference between the input image and the background in the figure is greater than the initially set threshold, it is determined that there is a target object.

2. Threshold segmentation method using physical characteristics of infrared imaging. Due to the different radiation intensities of different objects, there are brightness differences in different areas in the infrared image. Therefore, the ideal segmentation threshold can be set according to the mean and standard deviation of the pixel gray values to distinguish the background from the target object, and to determine the target object category. Because infrared images have the characteristics of low image contrast and blurred target edge features, Liu [5] proposed an improved fast generalized fuzzy c-means (Improved Fast Generalized Fuzzy C-Means, IFGFCM) algorithm, by applying clustering to the histogram of the infrared image calculation to set the ideal segmentation threshold. After that, Davis [6] made improvements on this basis, applied Gaussian model to reflect the distribution of pixel gray value, and set the threshold.

3. Background predictive modeling method [7]. The core of this method is to apply low-pass filtering, median filtering, morphological filtering and other algorithms to suppress the background, highlight the potential targets in the infrared image, and then detect and recognize these key targets. Honeywell of the United States applies wavelet transform to extract the characteristic information of the target to be measured. Based on this, an infrared radar fusion detection method is proposed, which combines two measuring devices to perform multi-sensor fusion stereo modeling and recognition.

3. The Key Points of YOLOv3 Infrared Detection Technology

3.1. Infrared Target Edge Detection
Infrared targets in the air have the problem of fuzzy edge feature information. The trained network sometimes incorrectly recognizes the background area with no target at the edge of the image as a target, which will cause some interference in identifying the complete target, making the false alarm rate too high and detecting the accuracy rate is low. Darknet53 can output feature maps of three scales for multi-scale prediction. The side length s of the predicted feature maps are 13, 26, 52 respectively. And a grid unit will predict three bounding boxes, so a total of (13×13+26×26+52×52)×3=10647. Each bounding box is set with a corresponding threshold. If the detection score is greater than the threshold, it indicates that there may be a target within the bounding box. The higher the score, the greater the possibility of the existence of the target. After applying the YOLO v3 model to preliminary testing of infrared targets in the air, it is found that if the threshold value of the boundary box of the edge area of the feature map is set too high, the detection effect of some small-sized infrared targets is not good, and the model will detect some false targets. Therefore, the author considers to improve the error function of the edge area, increase the bounding box threshold penalty, and reduce the false alarm rate. According to the characteristics of the aerial infrared target in this data set, the annular area of the output feature map is defined as the edge area of the aerial infrared target.

In order to meet the needs of aerial infrared target detection tasks, a target detection image data set was constructed in accordance with the VOC data set format standard. The data set contains 38 color video images with approximately 25,000 frames. In terms of single frame detection, in order to ensure the diversity of images, frame skipping selection is adopted, with a total of 8000 pictures as samples. Aiming at the improved YOLOv3 model, using a self-made infrared image data set for testing, the detection results of each target are shown in the following figure.
3.2. Infrared Small Target Sensing
The selection and setting of the network layer is an important part of the convolutional neural network design. Considering the amount of data of the perception task and the category of the perception target, this model designs two one-dimensional convolutional layers, and uses multiple filters to convolve with objects to come up with the potential characteristics of the data; design a pooling layer to improve the network's Generalization and abstraction capabilities; two fully connected layers are designed to realize data conversion and state recognition. In order to meet the needs of airborne infrared dim and small target state sensing tasks, a CNN model dedicated to air target state sensing tasks is designed on the basis of the traditional CNN model.

3.3. Comparative analysis
The constructed target detection data set has been trained and tested on the classic model, and the preliminary detection results are shown in Table 1. After the improvement, the accuracy rate has been improved.

| Model algorithm | UAV  | bird | Remark                      |
|-----------------|------|------|-----------------------------|
| SSD             | 76.4%| 75.8%|                             |
| YOLOv3          | 82.4%| 76.7%|                             |
| Improved YOLOv3 | 83.9%| 82.5%| Multi-scale detection optimizatio|

4. Conclusions
The technical status and difficulties of infrared target detection are introduced, and optimization research is carried out in combination with the YOLOv3 algorithm. Through the study and research on the relevant theories of aerial infrared target detection technology based on deep learning, a deep learning network architecture is constructed on this basis, and its key technical issues are explored. According to the low pixel resolution of aerial infrared target images, the image...
blurred edges and poor contrast improve the YOLOv3 model. In order to improve the detection and positioning accuracy, the edge loss function is improved, and the YOLOv3 model's ability to characterize infrared targets in the air is greatly improved, and the target frame dimension clustering is reasonably selected to improve the performance of the model algorithm. Aiming at the state perception of weak and small targets, a CNN model dedicated to the task of aerial target state perception is designed on the basis of the traditional CNN model. Good results have been obtained.

References
[1] LIU Zhi-jia, et al. An improved method of infrared image target detection based on YOLO algorithm. LASER & INFRARED, China, 2020.50(12):1513-1519
[2] Huang Lehong. Research on Aerial Infrared Target Detection Method Based on Deep Learning. Master’s Thesis, University of Chinese Academy of Sciences, China, 2020.
[3] R. Girshick. Fast R-CNN. IEEE International Conference on Computer Vision, Santiago, Chile, December, China, 2015: 1440-1448
[4] S. Q. Ren, K. M. He, R. Girshick, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, China, 2017, 39(6): 1137-1149
[5] J. F. Dai, Y. Li, K. M. He, et al. R-FCN: Object Detection via Region-based Fully Convolutional Networks. Conference on Neural Information Processing Systems (NIPS), China, 2016
[6] J. Redmon, S. Divvala, R. Girshick, et al. You only look once: unified, real-time detection. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, June, China, 2016: 779-788
[7] S Ren, K He, R Girshick, et al. Faster r-cnn:towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, China, 2017, 39(6):1137—1149.