Paper ID： MT566

The 3rd International Conference on Mechatronics and Automation Technology (ICMAT 2024)

Threat Target Infrared Image Recognition Method Based on Improved One-stage deep Network

Liming GAO, Chao LI, Xuemei WEI , Yanxin SUN, Wei LIN, Xiangyong ZHANG

Intelligent Technology Center, Norinco Group Test and Measuring Academy, Xi’an, China

1、Introduction

Due to the small size and complex background of threat target images used in military applications, it has become one of the difficulties in infrared image recognition of current threat targets. Deep learning is widely applied in target recognition tasks, but small target infrared image recognition has always been a hot topic in the field of target recognition. The full name of YOLO is You Only Look Once, which belongs to a single-stage object detection network. It takes the entire detected image as the input of the detection network. Unlike two-stage networks, it directly fits the position, size, and image category of the bounding box when outputting. In addition, YOLO uses a feature pyramid structure, which can extract features from feature maps of different scales as the basis for predicting target box parameters. Specifically, YOLO uses feature maps of three scales to predict targets of different sizes, with deep feature maps responsible for detecting large targets and low-level feature maps responsible for detecting small targets

2、YOLOv4

YOLOv4 is an end-to-end object detection algorithm that can be quickly applied to various fields, and can be optimized quickly after pre training, achieving a good balance between detection speed and recognition accuracy. At the same time, the minimum hardware configuration for YOLOv4 is not high, and a regular NVIDIA GTX-1080 computer configuration can train YOLOv4 and achieve good detection results.

Compared with previous YOLO versions, YOLOv4 has many improvements. For the surface defect recognition of insulated bearings in this article, this v4 version is beneficial for achieving good detection results even with fewer defect samples and less obvious defects.

Like other CNN based object detection algorithms, YOLO's feature extraction network also uses multiple convolutional layers and pooling layers to extract target feature information. Finally, a fully connected layer is used to obtain a feature map, which is a condensed version of the original image information and contains the main features of the original image content. The specific implementation steps are as follows:

(1) Uniformly resize the input image.

(2) The scaled image is convolved multiple times to obtain feature maps of (1024, 7, 7).

(3) Finally, two fully connected layers were used to process the feature maps of (30, 7, 7).

3、Improvement Strategy

GIOU

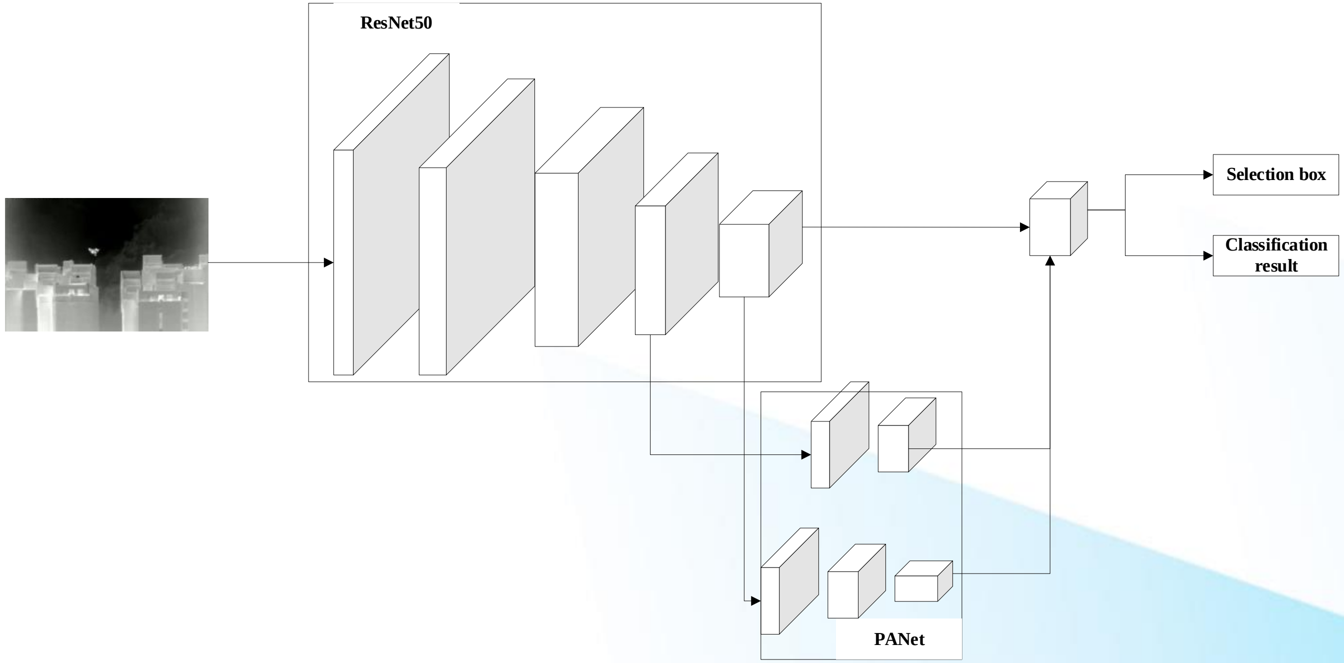
Using GIOU_Loss as the loss function for the prediction box regression, the GIOU_Loss formula is as follows:

$$\text{GIOU\_loss} = 1 - \text{GIOU} = 1 - \left(\text{IOU} - \frac{|A_C - U|}{A_C}\right) \quad (1)$$

A_C is the minimum closed area of two detection boxes, i.e. the area of the red box, U is the area of the green part, $\frac{|A_C-U|}{A_C}$ represents the proportion of the difference set in the red box area in the figure 1, and IOU is the ratio of the intersection and union of the yellow and blue boxes.

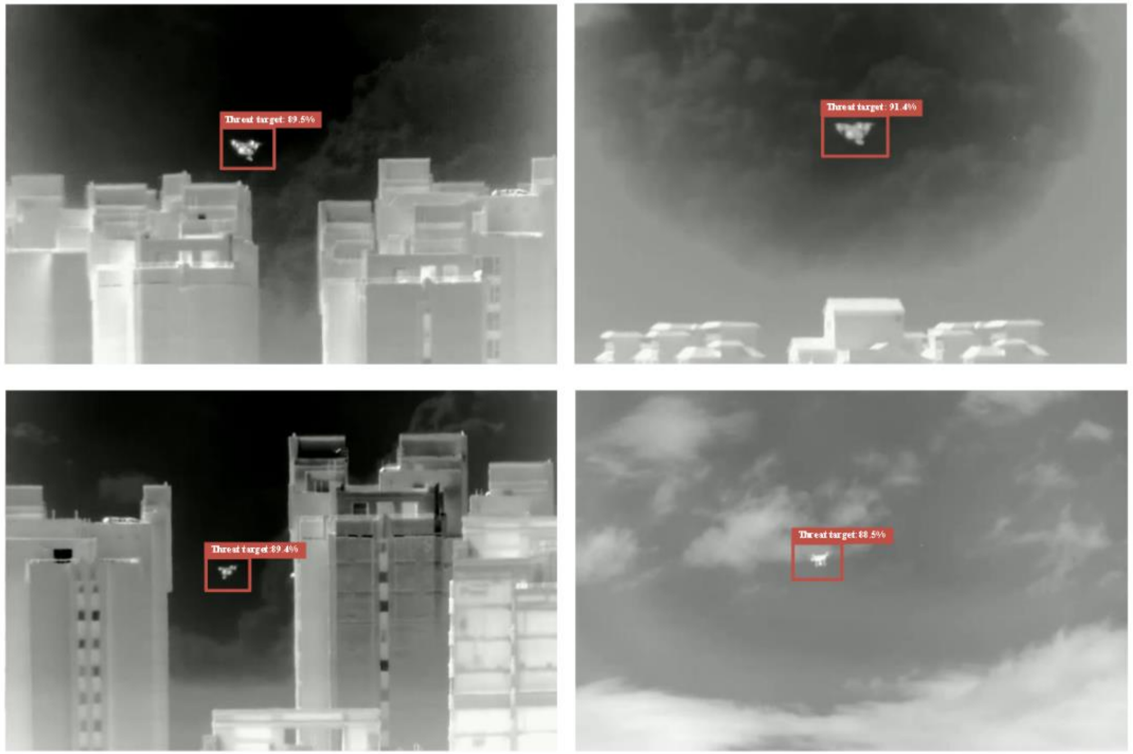
ResNet50

This article chooses the residual network ResNet50 as the basic model. ResNet mainly solves the problem of "degradation" of deep convolutional networks during depth deepening. In general convolutional neural networks, the first problem that comes with increasing the depth of the network is gradient vanishing or exploding. The BN layer can normalize the output of each layer, so that the gradient can remain stable in size even after passing through the layers in reverse, without being too small or too large. However, increasing the depth after BN is still not easy to converge, which is not caused by the disappearance of gradients or overfitting, but rather due to the complexity of the network, making it difficult to achieve the ideal error rate solely through unconstrained free range training



4、Experiment

The network structure of YOLOv4 is set according to default settings; Pre trained models are used on models that have already been trained on the ImageNet dataset, and the dataset format is converted according to the YOLO dataset; The training optimizer uses Adam optimization with a learning rate set to 0.005, and the learning rate reduction method adopts exponential step size reduction; During the training process, the feature extraction network is frozen, meaning that in the specified training generation, the parameters of the feature extraction network are not backpropagated or updated, and only the prediction network is trained, with updated parameters and weights. After training to a specified generation, the entire network is unfrozen and backpropagation is performed on the feature extraction network and prediction network, with all parameters participating in the update. The frozen algebra of the training process in this section is 200, and the total training algebra is 300.



Algorithms	Accuracy	Times
YOLOv3	84.3%	0.098s
YOLOv4	86.1%	0.078s
Proposed algorithm	91.2%	0.067s

5、Conclusion

Conclusion

The infrared image recognition of threat targets applied to military is one of the hotspots in the field of image detection today. Most of the infrared images of threat targets are urban backgrounds, so the backgrounds are complex, the targets are small, and the detection difficulty is high. Therefore, we used an improved YOLO algorithm to complete this task, replacing CIOU with GIOU to improve detection accuracy, and used ResNet50 as a feature extraction network to improve feature extraction efficiency and accuracy. Finally, the effectiveness of the algorithm was verified through experiments, and its detection accuracy reached 91.2%. This algorithm provides a good method for the field of threat target detection.