ISSN: 2664-9640

Millimeter wave image detection based on Yolo v3-tiny

Zhiling Peng, Guoping Chen, Yuchuan Zhou and Yue Huang

School of Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract

In recent years, the safety of airports, stations and other important public places has become more and more important. Millimeter wave can not only penetrate clothes to detect hidden objects on the human body, but also not produce harmful radiation to the human body. It is suitable for human imaging detection in a safe environment. Taking the security inspection environment as the background, aiming at the problems of large amount of calculation and poor real-time performance of the current millimeter wave detection algorithm, this paper proposes a millimeter wave image target detection algorithm based on Yolo v3-tiny, which mainly improves the loss function of the model. Firstly, in order to make the frame regression of the model more effective, CIoU function is used to replace the original coordinate frame loss function to improve the detection accuracy of the model. Secondly, in order to improve the imbalance between positive and negative samples of millimeter wave image, the Focal loss function is used in the loss function. Finally, a better detection effect is obtained on the millimeter wave data set.

Keywords

Millimeter wave image; object detection; deep learning; loss function.

1. Introduction

The wavelength of millimeter wave is between microwave and light wave, and its frequency range is 30-300GHz. It can penetrate insulating materials and clothing fabrics to display hidden contraband, such as pistols, bombs, tubular products and knives [1]. At the same time, the millimeter wave camera will not produce ionizing radiation harmful to the human body and will not cause discomfort to the human body. It is a safe and healthy safety detection equipment [2]. These characteristics make millimeter wave have great application prospects in the field of human safety detection [3-4], but there are also some deficiencies in the current millimeter wave safety detection results. Due to the large flow of people in the security inspection environment, it is necessary to quickly detect the human body. The scanning and imaging of human body by millimeter wave camera takes a certain time, so the millimeter wave image target detection algorithm has high speed requirements. Human safety inspection involves personal safety, and the detection of hazardous article must be very accurate. Therefore, the target detection algorithm of millimeter wave image has high accuracy requirements. In the field of computer vision, region-based target detection algorithm usually has high detection accuracy and can satisfied the needs of most application scenarios. However, this method requires a lot of computation and consumes a lot of time, such as Fast R-CNN network [5]. Another one-stage algorithm can maintain high speed while maintaining accuracy. For example, Yolo v3 algorithm [6], the category and location of the target can be obtained through one calculation. Yolo v3-tiny algorithm is a simplified version of Yolo v3, with faster speed and lighter volume. This paper presents a millimeter wave image target detection algorithm based on Yolo v3-tiny. According to the characteristics of millimeter wave image, CIoU and Fcoal loss are used in the loss function to improve the detection performance.

ISSN: 2664-9640

2. Millimeter wave image detection model

2.1. Network model structure

The network structure used in this paper is Yolo v3-tiny model. As shown in the Fig.1, Yolo v3-tiny consists of 13 convolution layers, 6 largest pooling layers, 1 upper sampling layer, 1 feature fusion layer and 2 prediction layers. To ensure that the network extracts enough information, the input pictures are interpolated into the network after the size of 608x608. First, the image passes through 10 convolution layers and 6 pooling layers to get the first prediction layer with a size of 19×19. Then, a second prediction layer is obtained by fusing a size of 19×19 feature map with a size of 38×38 feature map after convolution and upsampling. Finally, two prediction layers are used to make regression predictions to get the types and locations of the targets to be measured. Two feature maps of different sizes can predict the targets of different scales.

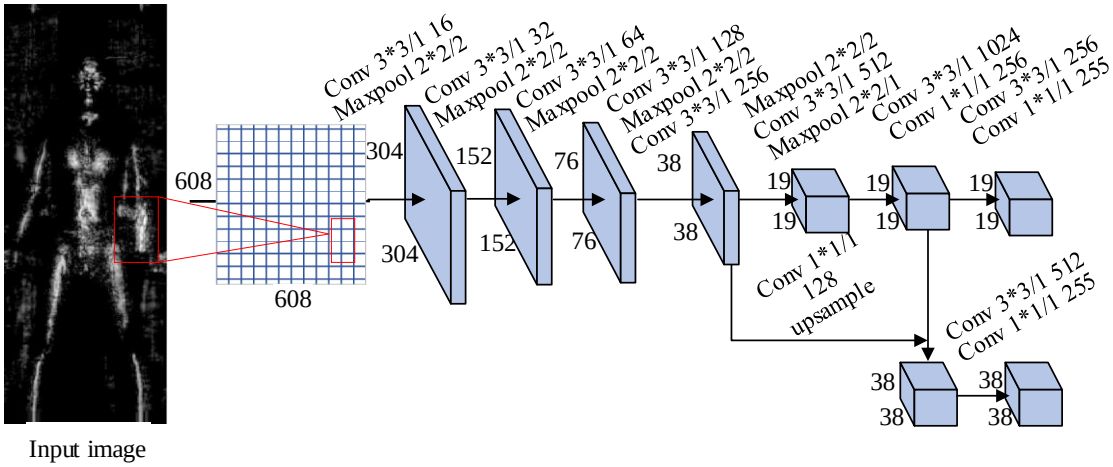

Fig. 1 The structure of Yolo v3-tiny target detection model

2.2. loss function.

2.2.1. Use CIoU loss as bounding box loss

Yolo v3-tiny uses MSE function to calculate the bounding box loss, and regresses through the center coordinate offset, width and height of the prediction box and the real box. The MSE loss function is quite sensitive to the scale of the target and does not always accurately reflect the positioning accuracy. In the calculation, we want to obtain the prediction frame with larger IOU value. One method is to directly use IOU value to measure the target loss. IOU represents the overlap rate between the prediction frame and the real frame, and the calculation formula is:

$$IoU_{(A,B)} = \frac{A \cap B}{A \cup B} \tag{1}$$

In formula (1), A is the area of the prediction frame and B is the area of the real frame, $IoU_{(A,B)}$ indicating the intersection of the area of the prediction frame and the area of the real frame divided by the union.

IOU can directly represent the coincident size of two bounding boxes, but for two objects with the same IOU, IOU may not be able to represent their different positional relationships. When the IOU value is the same, the coincident positions of the two bounding boxes may be different. When IOU is 0, the distance between the two bounding boxes cannot be reflected, and the function cannot be regressed. In order to enhance the sensitivity of the network to location, this paper uses CloU function to replace the original MSE bounding box loss function. CloU takes into account the distance, overlap rate and scale between the target and the bounding box to stabilize the regression of the target box, and adds a penalty factor. The penalty factor considers

the length width ratio of the prediction frame and the length width ratio of the fitting target frame, so that the boundary box under various intersection conditions can be effectively measured.

$$\alpha = \frac{\upsilon}{(1-IOU)+\upsilon} \tag{2}$$

$$\upsilon = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - atc\tan\frac{w}{h})^2 \tag{3}$$

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha\upsilon \tag{4}$$

Where, b and b^{gt} respectively represent the center points of the prediction box and the real box, and ρ represent the Euclidean distance between the two center points. c represent the diagonal distance of the smallest box that can contain the prediction frame and the real frame at the same time. The α parameters used to balance the proportion, υ measure the proportion consistency between the prediction frame and the real frame.

2.2.2. Use Focal loss

In millimeter wave images, the background accounts for a large proportion, and there are few dangerous goods hidden in the human body compared with the background. When generating the bounding box, the negative sample is much larger than the positive sample, and the original loss function is not conducive to optimization. Aiming at the imbalance of positive and negative samples, focal loss is introduced into the category loss of Yolo v3-tiny to improve the imbalance of positive and negative samples. The formula of focal loss is as follows:

$$FL(p,y) = \begin{cases} -\alpha(1-p)^{\gamma}\log a^{p}, y=1 \\ -(1-\alpha)p^{\gamma}\log a^{(1-p)}, y=0 \end{cases} \tag{5}$$

y=1 indicates that the sample is a positive sample, and y=0 indicates that the sample is a negative sample. In focal loss, there are γ and α two parameters, which represent the focusing factor and the inverse scale coefficient. At that time, for positive samples, the predicted result is 0.95. If it is a simple sample, the loss function value will become smaller. When the prediction result is 0.3, it belongs to hard to distinguish samples, which will lead to greater losses. The same is true for negative samples. Through this process, the influence of simple samples can be greatly reduced, the proportion of difficult samples can be increased, and the training tends to difficult samples. The inverse proportion coefficient can correct the proportion of positive and negative samples in the network. If the weight coefficient of a small number of positive samples is large, its contribution to the model will also increase. If the weight coefficient of a large number of negative samples is small, its contribution to the model will be relatively weakened. Therefore, the model will learn more useful information.

2.2.3. Loss function of model

The loss function of Yolo v3-tiny consists of three parts: bounding box loss, confidence loss and classification loss.

CIoU solves the problem that the IOU cannot accurately reflect the degree of overlap and cannot optimize when the IOU=0. CIoU loss function adds more considerations. For the case that two boxes are included in the horizontal and vertical directions, it can make the regression faster. Therefore, the CIoU loss function is applied to the Yolo v3-tiny network model and used $clou\_loss = 1 - CIoU$ as the bounding box regression loss function. Aiming at the problem of imbalance between positive and negative samples caused by less objectives in one stage algorithm, focal loss is used for optimization, and more attention is paid to the samples that are

difficult to distinguish. Therefore, focal loss is used as the class loss. The final improved loss function is as follows:.

$$
\begin{aligned}
loss = & \sum_{i=0}^{S^2}\sum_{j=0}^{B} l_{ij}^{obj}(1-IOU+\frac{\rho^2(b,b^{gt})}{c^2}+\alpha\upsilon) \\
& -\sum_{i=0}^{S^2}\sum_{j=0}^{B} l_{ij}^{obj}(\alpha(1-C_i^j)^{\gamma}C_i^j \lg \hat{C}_i^j+(1-\alpha)(C_i^j)^{\gamma}(1-C_i^j)\lg(1-\hat{C}_i^j)) \\
& -\sum_{i=0}^{S^2} l_{ij}^{obj}\sum_{c\in class}[\hat{P}_i^j \lg P_i^j+(1-\hat{P}_i^j)\lg(1-P_i^j)]
\end{aligned}
\tag{5}
$$

3. Experiments and results

The millimeter wave human body imaging data set is used to detect three kinds of typical dangerous goods (gun, knife and knife) hidden in different positions on the human body. A total of 6000 millimeter wave images (2000 for gun, knife and broadsword) were used in this experiment, of which the ratio of training set to test set is 7:3. The experiment is carried out under the pytorch deep learning framework, and the experimental hardware is NVIDIA 1080ti GPU. It verifies the practical feasibility and effectiveness of the improved Yolo V3 tiny algorithm in millimeter wave image concealment detection under the same data set and experimental equipment.

The experiment compares and analyzes the map and single frame detection speed of Yolo v3-tiny and Yolo v3-tiny model with improved loss function. The results are shown in Table 1.

Table 1 The improved algorithm in this paper is compared with Yolo V3 tiny

Model	mAP	ms/frame
Yolo v3-tiny	85.2	14
Ours	88.1	14

It can be seen from the table that the performance of the improved method in this paper is better than Yolo v3-tiny algorithm on millimeter wave image data set, and the mAP is increased by 5 percentage points. The CIoU loss function used in this model can better reflect the positioning accuracy of the prediction frame, so as to improve the detection accuracy of the model. For the millimeter wave image data set with unbalanced positive and negative samples, the focal loss function can adjust the proportion of positive and negative samples and the proportion of difficult samples through parameters, so as to improve the detection accuracy of the algorithm. After using focal loss as the loss function, the imbalance between positive and negative samples of the model is solved to some extent. At the same time, because focal loss can improve the detection effect of difficult targets, the detection effect of the model is improved as a whole. The improved loss function model is no different from Yolo v3-tiny in detection time, which shows that the improvement of loss function hardly affects the detection speed of the network.

4. Summary

Based on the Yolo v3-tiny network, aiming at the imbalance of positive and negative samples in millimeter wave image and the problem that the loss function of Yolo v3-tiny network can not well reflect the positioning accuracy, this paper uses CIoU instead of MSE as the regression loss function of prediction bounding box and Focal loss function as the classification loss function

to improve the detection accuracy of hidden objects in millimeter wave image. Then, the effectiveness of the improved loss function is verified on the millimeter wave data set, which shows that the improved loss function can improve the detection rate of hidden objects. The results show that the map of the improved model on millimeter wave data set is 88.1%, and the single frame detection speed is 14ms/frame, which can contented the detection requirements of security inspection environment.

References

[1] Oka S, Togo H, Kukutsu N, et al. Latest trends in millimeter-wave imaging technology[J]. Progress In Electromagnetics Research Letters, 2008, 1: 197-204.

[2] Jaeger I, Zhang L, Stiens J, et al. Millimeter wave inspection of concealed objects[J]. Microwave and optical technology letters, 2007, 49(11): 2733-2737.

[3] Yuan M, Zhang Q, Li Y, et al. A Suspicious Multi-Object Detection and Recognition Method for Millimeter Wave SAR Security Inspection Images Based on Multi-Path Extraction Network[J]. Remote Sensing, 2021, 13(24): 4978.

[4] Zhang G F, Lou G W, Wang H. A novel alternating current imaging radiometer for the security inspection[J]. Microwave and Optical Technology Letters, 2008, 50(12): 3259-3261.

[5] Ren S., He K., Girshick R., et al. Faster R-CNN: Towards realtime object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[6] Redmon J., Farhadi A. YOLO V3: An incremental improvement [C]//IEEE conference on computer vision and pattern recognition, Long Beach: IEEE Press, 2019: 1-6.