American Journal of Optics and Photonics
2021; 9(3): 32-38
http://www.sciencepublishinggroup.com/j/ajop
doi: 10.11648/j.ajop.20210903.11
ISSN: 2330-8486 (Print); ISSN: 2330-8494 (Online)

A Method of Infrared Image Pedestrian Detection with Improved YOLOv3 Algorithm

Yue Sun[1], Yifeng Shao[2], Guanglin Yang[1, *], Haiyan Xie[3]

[1]Laboratory of Signal and Information Processing, Department of Electronics, Peking University, Beijing, China

[2]Optics Research Group, Department of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, Netherlands

[3]Country China Science Patent and Trademark Agent, Beijing, China

Email address:

ygl@pku.edu.cn (Guanglin Yang)

[*]Corresponding author

To cite this article:

Yue Sun, Yifeng Shao, Guanglin Yang, Haiyan Xie. A Method of Infrared Image Pedestrian Detection with Improved YOLOv3 Algorithm. *American Journal of Optics and Photonics*. Vol. 9, No. 3, 2021, pp. 32-38. doi: 10.11648/j.ajop.20210903.11

Received: July 25, 2021; **Accepted**: August 9, 2021; **Published**: August 26, 2021

Abstract: The principle of infrared image is thermal imaging technology. Infrared pedestrian detection technology can be applied to the safety monitoring of the elderly, which can not only protect personal privacy, but also realize pedestrian identification at night, which has strong application value and social significance. A method of infrared image pedestrian detection with improved YOLOv3 algorithm is proposed to increase the detection accuracy and solve the problem of low detection accuracy caused by infrared pedestrian target edge blurring. And according to the characteristics of infrared pedestrian, a complex sample data set is established which is applied to infrared pedestrian detection. The infrared image enhancement method with WDSR-B is adopted to improve the clarity of the data set. In addition, based on YOLOv3 algorithm, the output of the 4-time down-sampling layer is added to obtain richer context information for small targets and improve the detection performance of the network for small-target pedestrians. And the improved YOLOv3 network is trained by the enhanced infrared data set. Experimental results show that the scheme precision of pedestrian detection is higher than that of YOLOv3 algorithm. Therefore, this method can be applied to the detection of pedestrians at night and the safety monitoring of the elderly.

Keywords: Infrared Image, Pedestrian Detection, Neural Network

1. Introduction

The principle of infrared imaging is thermal imaging technology, which can identify pedestrians in a timely manner under unusual environments such as nights and cloudy days. Infrared image pedestrian detection can be applied to the elderly safety monitoring, which has strong social significance. However, there are some problems such as fuzzy edges and indistinct features, which limit the accuracy of infrared image pedestrian detection [16].

Aiming at the problem that the low resolution of the target affects the detection accuracy, scholars have proposed some methods. For examples, the super-resolution algorithm is applied to Fast R CNN network to improve the detection performance of low-resolution targets [1], the positive effect of super-resolution is analyzed on the target detection performance of satellite images [2]. However, the effect of super-resolution algorithm on the performance of the infrared pedestrian target detection algorithm has yet to be verified. In recent years, the super-resolution algorithm has been deeply researched, and there are many approaches, among which WDSR (i.e., Wide Activation Deep Super-Resolution) is one of the better ones [3-5]. In Reference [5], two kinds of networks are proposed, namely WDSR-A and WDSR-B, in which WDSR-B further increases the width of feature graph on the premise of the same computational cost. Experiments show that WDSR-B has better performance. Therefore, we use WDSR-B network to enhance the infrared image and verify the influence of the super-resolution algorithm on the performance of the infrared pedestrian target detection algorithm.

The infrared pedestrian detection method based on deep learning does not need to define the characteristics of pedestrians manually. The algorithm is relatively simple, with strong generalization ability and high detection accuracy

[6-8]. Among the target detection algorithms based on deep learning, YOLOv3 (i.e., You Only Look Once V3) is one of the methods with better detection accuracy and higher real-time performance [9-10]. However, YOLOv3 algorithm performs well in large-scale target application scenarios. Pedestrian targets that occupy fewer pixels in the infrared image have poorer detection accuracy [11-12].

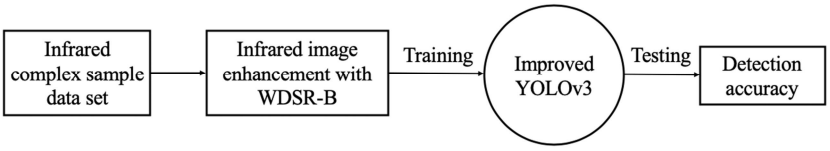

Figure 1. *Infrared image pedestrian detection scheme.*

Therefore, to improve the detection accuracy of pedestrian targets with fuzzy edge and small size, we establish a complex sample data set specifically for infrared pedestrian detection. And WDSR-B network is adopted to enhance infrared image and improve the clarity of pedestrian target. Based on YOLOv3 algorithm, the output of the 4-time down-sampling layer is added to obtain richer context information for small targets. In addition, the improved YOLOv3 network is trained by using the enhanced data set with WDSR-B. The infrared image pedestrian detection scheme proposed in this paper is shown in Figure 1.

2. Infrared Image Enhancement Method with WDSR-B

There are many super-resolution technologies based on the single image, such as the interpolation-based method and the learning-based method. The image super-resolution method based on machine learning enables the model to learn the mapping relationship between the existing low-resolution image and the high-resolution image by establishing an end-to-end neural network. This method increases the information on the basis of the original low-resolution image, instead of simply interpolating the image, which boosts the details of the image, so the effect is obviously better than the traditional interpolation algorithm [13].

WDSR [5] (i.e., Wide Activation Deep Super-Resolution) is one of the most excellent Image super-resolution algorithms based on convolutional neural network. The basic principle of this method is to update the weights of the network through the error back propagation of the output and the true value.

WDSR network improves the width of the feature map by increasing the number of filters before activation. Experimental results prove that WDSR network can further improve the accuracy of super resolution. In addition, WDSR algorithm uses a weighted normalization operation to increase the learning rate of the network. Assume the output y is with the form:

$$y = w \cdot x + b \tag{1}$$

Where w is a k-dimensional weight vector, b is a scalar bias term, x is a k-dimensional vector of input features. WN re-parameterizes the weight vectors in terms of the new parameters using:

$$w = \frac{g}{||v||} \cdot v \tag{2}$$

Where v is a k-dimensional vector, g is a scalar, and ||v|| denotes the Euclidean norm of v. And we will have $||w|| = g$, independent of parameters v. Thereby achieving faster convergence speed and better system performance.

In Reference [5], WDSR-A and WDSR-B networks are proposed, in which WDSR-B further increases the width of feature graph under the premise of the same computational cost. It has been proved in the literature that WDSR-B has better super resolution performance.

WDSR algorithm has been applied to the experiment of visible image super resolution, but its enhancement effect on infrared image remains to be verified. In addition, the effect of super-resolution enhancement on the performance of infrared target detection algorithm is unknown. Therefore, through experiments, we use the original infrared image and the enhanced infrared image with WDSR-B to train the target detection algorithm and analyze the impact of WDSR-B on the performance of the pedestrian detection algorithm based on the infrared image.

3. Improved YOLOv3 Model

YOLO is a special algorithm for target detection that adopts the idea of convolutional neural network. It can scan the entire image at one time and apply predictive filters to identify image classes. YOLOv3 has made some optimizations based on YOLOv1 and YOLOv2 [9, 14, 15], so that the system performance has been significantly improved.

3.1. Cluster Analysis of Infrared Pedestrian Anchors

The YOLOv3 network predicts 5 coordinates for each bounding box, t_x, t_y, t_w, t_h, t_o. The predictions correspond to:

$$b_x = \sigma(t_x) + c_x \tag{3}$$

$$b_y = \sigma(t_y) + c_y \tag{4}$$

$$b_w = p_w e^{t_w} \tag{5}$$

$$b_h = p_h e^{t_h} \tag{6}$$

$$Pr(object) * IOU(b, object) = \sigma(t_o) \tag{7}$$

Where (c_x, c_y) is the offset of the cell to the top left corner

of the image. And p_w, p_h are the width and height of the bounding box prior.

YOLOv3 uses k-means clustering to determine bounding box priors. The calculation formula is as follows：

$$d(box; centroid) = 1 - IOU(box; centroid) \quad (8)$$

On the COCO data set the 9 clusters were: (10*13), (16*30), (33*23), (30*61), (62*45), (59*119), (116*90), (156*198), (373*326). The matching rules of feature maps and bounding boxes of YOLOv3 are shown in the table 1. The result of boundary box clustering of YOLOv3 network is mostly square, which is inconsistent with the shape of pedestrian boundary box.

Table 1. *The matching rules of feature maps and bounding boxes of YOLOv3.*

Feature maps	Receptive field	Bounding box
13*13	Maximum	(373*326), (156*198), (116*90)
26*26	Middle	(59*119), (62*45), (30*61)
52*52	Minimum	(33*23), (16*30), (10*13)

Due to the characteristics of infrared targets such as edge blurring, poor definition, and occlusion, the definition of the complex sample in this paper are pedestrian target with small size, incomplete shape, and fuzzy features. We establish a complex sample data set specifically for infrared pedestrian detection based on the Computer Vision Center-14 (CVC-14) infrared data set. The clustering of sample data is improved for the complex samples proposed in this paper. K=1, 2..., 15. Conduct k-means clustering on the above samples. When K=4, the average IoU tends to be stable.

According to the clustering results, the sizes of the four boundary boxes are: (4*13), (23*29), (15*42), and (55*135). Compared with YOLOv3 algorithm, the clustering results of the complex sample data set in this paper meet the size characteristics of pedestrians in the infrared image basically and have certain robustness for the blocked pedestrians.

3.2. The Improvement to the Output Structure of YOLOv3 Network

YOLOv3 algorithm uses three down-sampling to detect three sizes of objects of the input image. Feature maps with smaller sizes can provide deep semantic information, while feature maps with larger sizes contain rich location information. The YOLOv3 algorithm fuses the adjacent shallow feature maps and deep feature maps, and then outputs the prediction results. However, YOLOv3 uses 8x down-sampling for small target prediction, so feature extraction and target detection cannot be performed accurately when the target size is less than 8∗8 pixels.

To further improve the YOLOv3 algorithm's ability to recognize small targets, the 4 times down-sampling layer in the Darknet-53 network structure is used for target detection to obtain more information on a small target. We use 2 times up-sampling on the output results of the 8 times down-sampling layer in the YOLOv3 network and fuse with the 4 times down-sampling feature map before model output. The improved YOLOv3 algorithm is more suitable for the detection of small targets. The structure of improved network is shown in Figure 2. The part in the dotted box is the output feature graph we added according to the output structure of YOLOv3 algorithm.

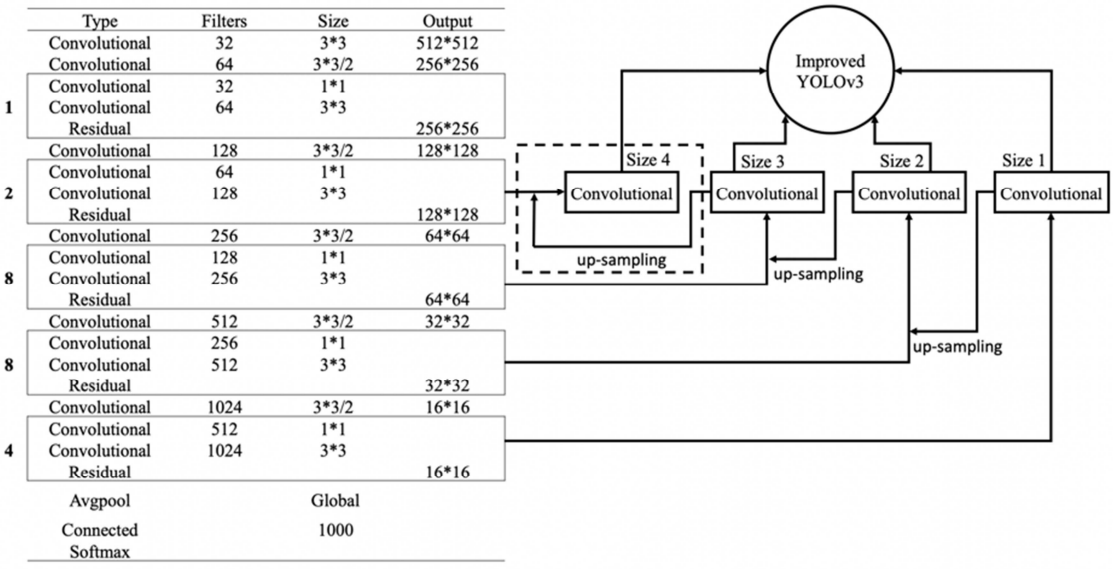

Figure 2. *The output structure of improved YOLOv3.*

4. Experimental Results and Analysis

4.1. Experiment Results of Infrared Image Enhancement with WDSR-B

The evaluation index of super-resolution generally uses the Peak signal-to-noise ratio (PSNR) to measure the performance of the model through the difference between the high-resolution image output by the model and the high-resolution reference image.

Assuming the image resolution is m*n, the high-resolution image output of the model is $P(m,n)$ and the high-resolution

image for reference is $Q(m,n)$, then the mean-square error (MSE) and PSNR of $P(m,n)$ and $Q(m,n)$ can be expressed as:

$$MSE = \frac{1}{mn}\sum_{m=0}^{m-1}\sum_{n=0}^{n-1}[P(m,n) - Q(m,n)^2] \quad (9)$$

$$PSNR = 10 * log(\frac{max_P{}^2}{MSE}) \quad (10)$$

The design of infrared image enhancement scheme based on super-resolution is shown in Figure 3, 500 clear infrared images were selected from CVC-14 data set, and the high-resolution images in the infrared image data set were processed into low-resolution images by bicubic image compression algorithm, which constituted the input and output image pairs of WDSR-B network, and the infrared image enhancement network based on WDSR-B was trained.

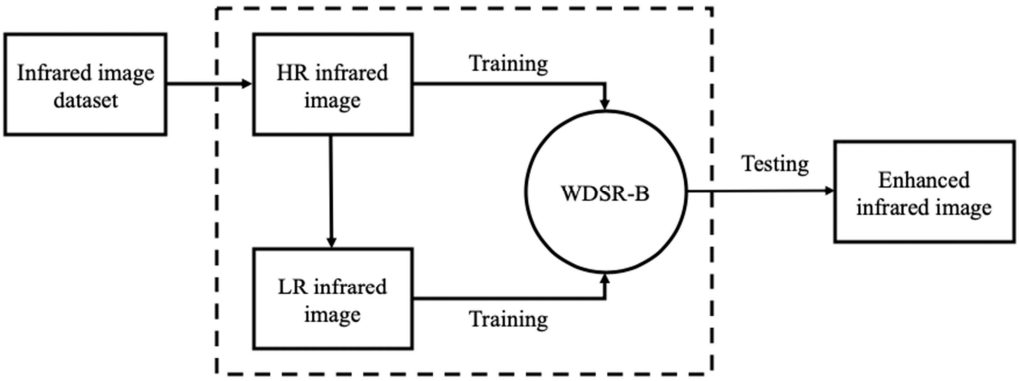

Figure 3. *Infrared image enhancement scheme based on super resolution.*

Test process, we use double sampling under three interpolation methods to reduce to the original image size of 1/2. Then we use nearest neighbor domain interpolation method, bilinear interpolation method, double three interpolation methods, and WDSR-B network respectively to carry out image enhancement experiments and enlarge the input image to the original size.

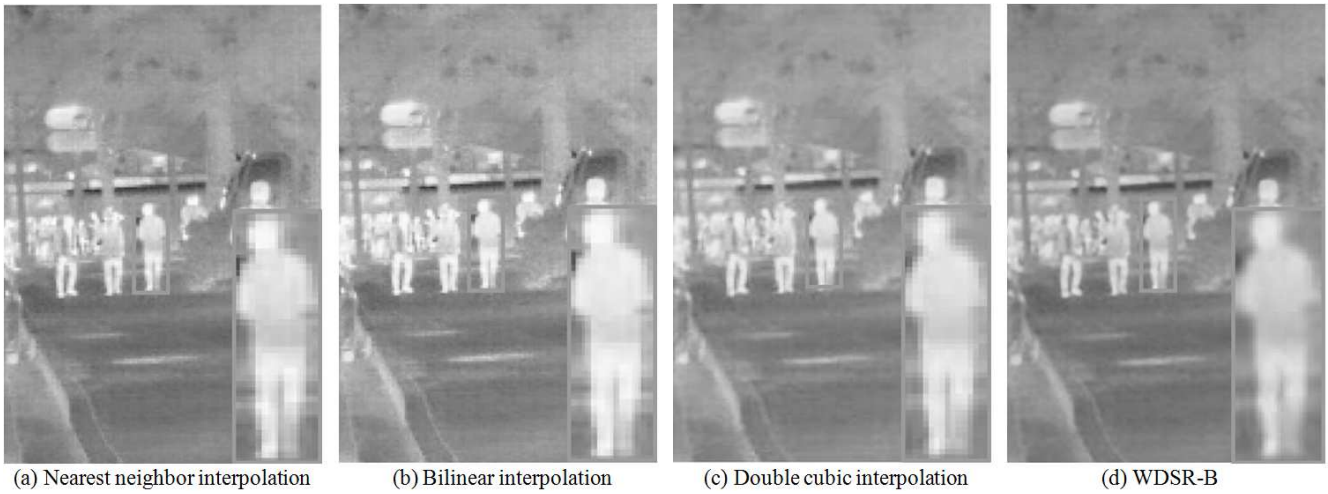

Figure 4. *The experimental results of the above four methods are used for the infrared image.*

Figure 4 shows the experimental results with super resolution using the above four methods. Table 2 shows the experimental PSNR values of infrared images using the above four methods.

Table 2. *Comparison of PSNR values of four super resolution algorithms.*

Algorithm	Infrared
Nearest neighbor	21.9101 dB
Bilinear	21.9703 dB
Bicubic	23.1384 dB
WDSR-B	24.2489 dB

Experimental results show that WDSR-B network performs best among the four super resolution algorithms, and the conclusion is applicable to infrared image. Therefore, the infrared image enhancement method based on WDSR-B can improve the clarity of infrared images.

4.2. Experiment Results of Infrared Image Pedestrian Detection

In this paper, we use precision, recall, and average precision (AP) to evaluate the performance of the target detection algorithm. The calculation formulas are as follow:

$$precision = N_{TP}/(N_{TP} + N_{FP}) \quad (11)$$

$$recall = N_{TP}/(N_{TP} + N_{FN}) \quad (12)$$

Where N_{TP} represents the number of targets correctly identified, N_{FP} represents the number of targets incorrectly identified, and N_{FN} is the number of targets not identified. The area of the graph formed by the accuracy-recall curve and the coordinate axis is the result of average precision.

Three different sets of experiments were used to compare the effectiveness of the proposed scheme: the complex sample data set was used to train the traditional YOLOv3 network and the improved YOLOv3 network respectively. In addition, the complex sample data set was input into the infrared image enhancement network with WDSR-B to obtain the enhanced infrared data set with improved image resolution and clearer pedestrian targets. Then the improved YOLOv3 network was trained with this data set. Finally, the accuracy of three target detection algorithms is tested. The scheme design of infrared image pedestrian detection experiment is shown in Figure 5.

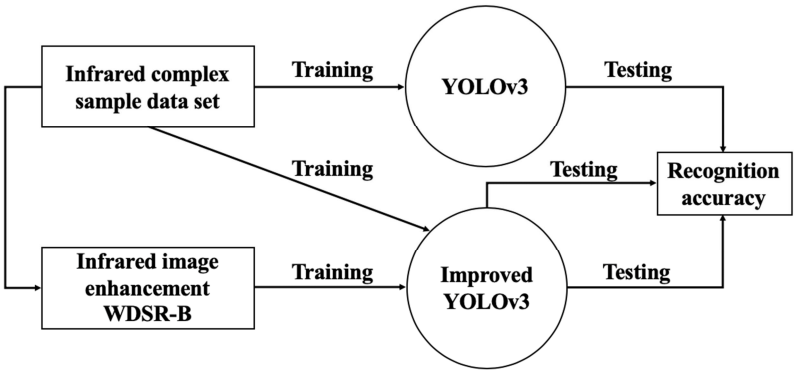

Figure 5. *Infrared image pedestrian detection experimental scheme.*

According to the definition of complex samples proposed in this paper, 500 infrared images were selected, including 2518 pedestrian targets. The experiment selected 400 infrared images as the training set, including 1961 infrared pedestrian targets. The remaining 100 infrared images are the test set, including 557 infrared pedestrian targets.

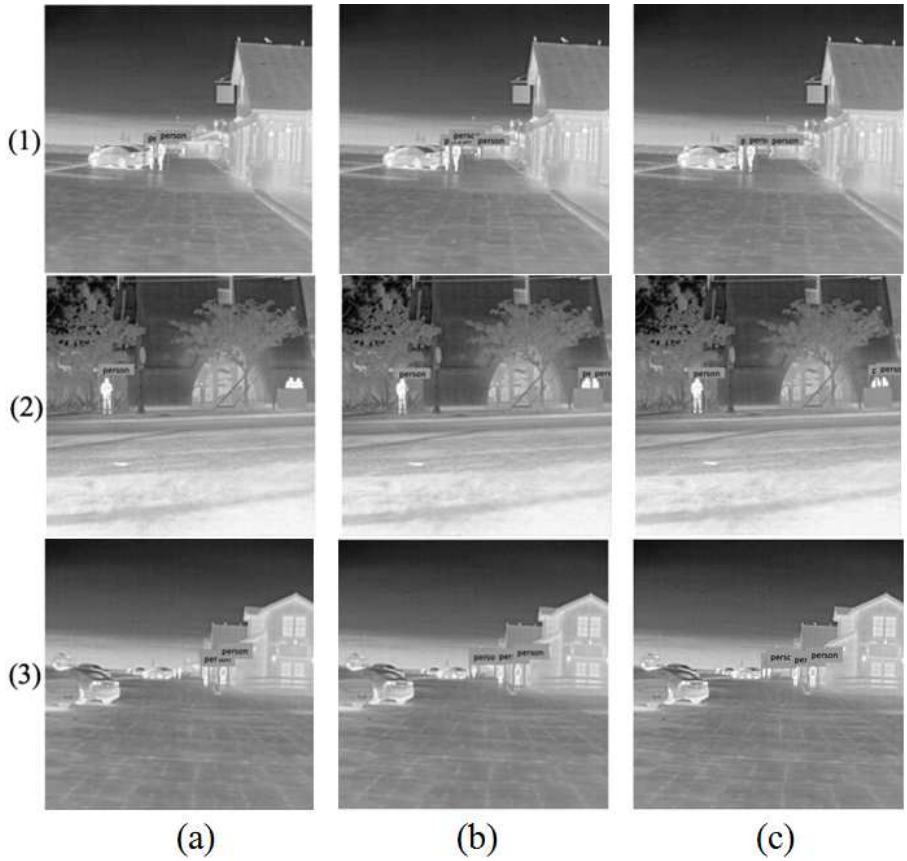

Figure 6. *Comparison of experimental results of three algorithms for pedestrian detection.*

The results of the qualitative assessment of the proposed scheme are shown in Figure 6 (a) are the results of YOLOv3 algorithm. (b) are the results of improved YOLOv3 algorithm. And (c) are the results of improved YOLOv3+ WDSR-B algorithm. (1) are the experimental results when the pedestrian is blocked. And (2) are the detection results of the small target. It can be intuitively seen that the YOLOv3 algorithm has a phenomenon of missed detection. However, the improved YOLOv3 algorithm and the improved YOLOv3+ WDSR-B algorithm can identify targets with occluded and small size. The scheme proposed in this paper has better robustness in a complex environment.

The results of the quantitative evaluation of the proposed scheme are shown in table 3. Precision of traditional YOLOv3 algorithm in infrared image pedestrian detection is only 66.34%, recall is 60.86%. Precision of improved YOLOv3 algorithm is 79.91%, and the recall is 86.71%. The improved YOLOv3+ WDSR-B algorithm can achieve detection precision at 84.55%, recall is 91.38%. Experimental results show that the infrared image pedestrian detection method proposed in this paper can effectively improve the clarity of infrared pedestrian targets. Precision of pedestrian detection is increased by 18.21% and recall is increased by 30.52%.

Table 3. *Experimental results of three algorithms for pedestrian detection.*

Algorithm name	N_{TP}	N_{FP}	N_{FN}	Precision	Recall
YOLOv3	339	172	218	66.34%	60.86%
Improved YOLOv3	483	145	74	76.91%	86.71%
Improved YOLOv3+ WDSR-B	509	93	48	84.55%	91.38%

Figure 7 plots the precision-recall curves of three methods mentioned above. The AP values of the three algorithms in infrared image pedestrian detection are 62.89%, 77.32% and 85.59% respectively. The AP of improved YOLOv3 + WDSR-B algorithm is 22.7% higher than the traditional YOLOv3 algorithm.

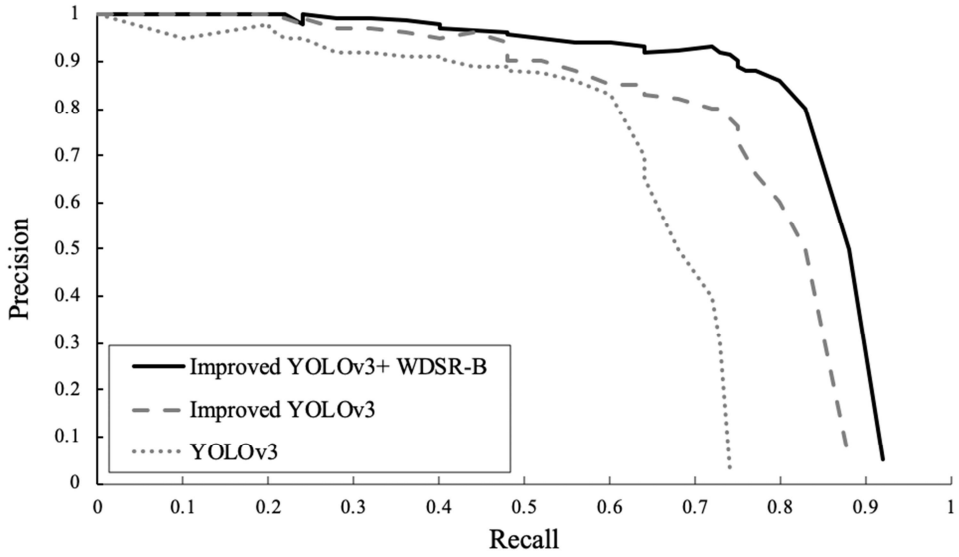

Figure 7. *Accuracy-recall curve of infrared image pedestrian detection.*

5. Conclusion

A method of infrared image pedestrian detection with improved YOLOv3 algorithm is proposed, which can solve the problem of low detection accuracy caused by infrared pedestrian target edge blurring to a certain extent. According to the characteristics of infrared pedestrian, a complex sample data set is established which is specially applied to infrared pedestrian detection. Then the infrared image enhancement method with WDSR-B is used to improve the clarity of the data set. Based on YOLOv3 algorithm, the output of the 4-time down-sampling layer is added to obtain richer context information for small targets and improve the detection performance of the network for small-target pedestrians. And the improved YOLOv3 network is trained by the enhanced infrared data set. Experimental results show that the precision of pedestrian detection of the scheme is 22.7% higher than that of YOLOv3 algorithm. Therefore, this method can be applied to the detection of pedestrians at night and the safety monitoring of the elderly. In the future, we will discuss the influence of super-resolution multiples on infrared pedestrian detection accuracy.

Acknowledgements

The authors would like to thank the Spatial Image Processing Laboratory for their support. This work was supported by the National Science Foundation of China (No. 62071009).

References

[1] H. Krishna, and C. V. Jawahar, "Improving Small Object Detection," 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), Nanjing, pp. 340-345 (2017).

[2] J. Shermeyer, and A. Van Etten, "The Effects of Super-Resolution on Object Detection Performance in Satellite Imagery," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, pp. 1432-1441 (2019).

[3] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence. 38 (2), 295-307, (2016).

[4] B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, pp. 1132-1140, (2017).

[5] Yu J, Fan Y, and Yang J, "Wide Activation for Efficient and Accurate Image Super-Resolution." arXiv: 1808.08718v2 (2018).

[6] Z. Xie, S. Zhang, X. Yu, and G. Liu, "Infrared and visible face fusion recognition based on extended sparse representation classification and local binary patterns for the single sample problem," J. Opt. Technol. 86 (13): 408-413, (2019).

[7] Axel-Christian Guei, and Moulay Akhloufi, "Deep learning enhancement of infrared face images using generative adversarial networks," Appl. Opt. 57 (18), D98-D107, (2018).

[8] Wang C, and Qin S, "Approach for moving small target detection in infrared image sequence based on reinforcement learning." Journal of Electronic Imaging, 25 (5): 053032, (2016).

[9] Redmon J, and Farhadi, "YOLOv3: An Incremental Improvement," arXiv: 1804.02767, (2018).

[10] Xiangfu Zhang, Zhangsong Shi, Zhonghong Wu, and Jian Liu, "Sea surface ships detection method of UAV based on improved YOLOv3," Proc. SPIE 11373, Eleventh International Conference on Graphics and Image Processing (ICGIP 2019), 113730T (2020).

[11] Tian, Wei, et al. "3D Pedestrian Detection in Farmland by Monocular RGB Image and Far-Infrared Sensing." Remote Sensing [J] 13. 15 (2021): 2896.

[12] Zhang C, Li D, Qi J, et al. Infrared Small Target Detection Method with Trajectory Correction Fuze Based on Infrared Image Sensor [J]. Sensors, 2021, 21 (13): 4522.

[13] Shi W, Caballero J, and Ferenc Huszár, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1874-1883 (2016).

[14] Redmon J, Divvala S, and Girshick R, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788 (2016).

[15] Redmon J, and Farhadi A, "YOLO9000: Better, Faster, Stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE: 6517-6525 (2017).

[16] Guanglin Yang, "Output characteristics of pre-amplifying circuit signal for human body infrared detecting," The 17th national academic annual meeting of measuring and controlling instruments (MCMI'2007), 87-90 (2007).