R-PCNN method to rapidly detect objects on THz images in human body security checks

Hong Xiao
Computer science
Guangdong University of Technology
Guangzhou, China;
wh_red@163.com

Rongyue Zhang
Computer science
Guangdong University of Technology
Guangzhou, China;
flowet@163.com

Hao Wang
Computer science
Norwegian University of Science and Technology
Ålesund, Norway
iswanghao@acm.org

Feng Zhu
Computer science
Guangdong University of Technology
Guangzhou, China;
302761359@qq.com

Cheng Zhang
Computer science
Guangdong University ofTechnology
Guangzhou,China;
25290901204@qq.com

Hongning Dai
Macau University of Science and Technology
Macau, China;

YuBing Zhou
Computer science
Guangdong University of Technology
Guangzhou, China

Abstract **— Terahertz human body security images have low resolution and a low signal-to-noise ratio. In the traditional method, image segmentation, positioning, and identification are applied to detect objects carried by humans in the THz images. However, it is difficult to satisfy the requirements of detection accuracy and speed with this approach. The current research presents a faster detection framework (R-PCNN)[1] combining the preprocessing and structure optimization of Faster R-CNN. The experiment results show that this method can effectively improve the accuracy and speed of object detection in human body THz images. A detection accuracy of 84.5% can be achieved in dense flow scenes, with an average detection time of less than 20 milliseconds for each image.**

Keywords — Human body security check, Terahertz image, Image enhancement, Object detection, Faster R-CNN

Applied science and technology research and development in Guangdong province special fund project: development of urban rail transit security inspection equipment based on terahertz imaging technology(NO. 2015B090923004)National natural science foundation project: research on self-organizing reinforcement strategy of IOT network in intelligent manufacturing(NO. 61672170)

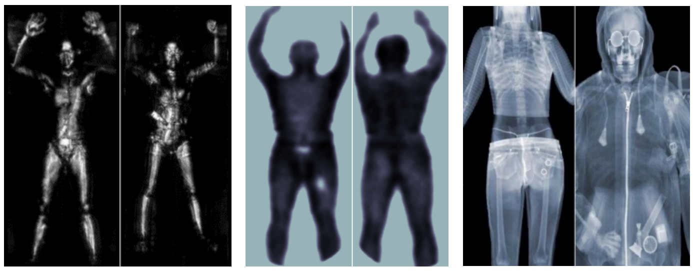
I. INTRODUCTION

Terahertz (THz) imaging technology uses the terahertz wave (which usually refers to far-infrared electromagnetic radiation with a wavelength of 30 μm or 3 mm) as a signal source. The terahertz wave is between the infrared and microwave and has unique properties that X-rays, light/infrared waves, and microwaves do not have. This type of radiation has very low photon energy compared with X-rays and causes no harm to the human body. In contrast to light/infrared waves, the terahertz wave can penetrate materials such as paper, plastic, and cloth. Thus, it can be used to detect objects hidden under human clothing. Compared with microwaves, the terahertz wave has a shorter wavelength and achieves higher imaging accuracy. Terahertz imaging can detect not only metal objects but also nonmetallic contraband items (such as explosives, ceramic knives, glass knives, and drugs). These properties make terahertz imaging highly suitable for human security inspection.

However, the terahertz wave has a much longer wavelength compared with X-rays and visible light. Therefore, the resolution and signal-to-noise ratio of human security imaging based on terahertz technology are different from those of X-rays (as shown in Fig. 1). This makes it difficult to detect automatic objects with the use of THz imaging. Because of its reduced cost, terahertz technology has gradually begun to be used for security checks in subways, public venues, and other crowded pedestrian areas. Passive terahertz imaging currently has a speed of up to 10 frames per second and will be even faster in the future. This puts forward higher requirements for the speed and accuracy of human security check.

Human body terahertz imaging can be classified into active and passive types. The present study concentrates on active THz security images of the human body, which provide more details. At present, the research on object detection in THz images is mainly focused on the segmentation and location of active image objects. First, image segmentation, contour tracking, region growing, fuzzy clustering, and other methods are used to separate the objects, which are then classified and recognized. In recent years, with the rapid development and application of deep learning technology, the accuracy and speed of image object detection and recognition have been greatly improved. However, the application of deep learning algorithms in terahertz image object detection is still in its infancy. The direct use of deep learning algorithms gives rise to two problems: (1) the accuracy of object detection is not high, and (2) the speed of object detection is not fast enough.

This study proposes a kind of Faster detection framework (R-PCNN) combining the preprocessing and structure optimization of Faster R-CNN. Compared with deep learning algorithms without preprocessing, the proposed method can improve the accuracy of detection. Compared with unoptimized deep learning algorithms, this approach can improve the detection speed and satisfy the detection speed requirements in the scene of characters simultaneous check.



(a) Active Terahertz image (b) Passive Terahertz image (c)X-rays image

Figure 1: Difference among active/passive human body terahertz image and X-ray image

II. RELATED STUDIES

To achieve THz images with low resolution and a low signal-to-noise ratio (THR), previous researches have mainly focused on improving two aspects: the image quality and the location detection algorithm.

In [1], to improve the quality of terahertz images, the point spread function of the terahertz scanning imaging system was applied by using the pinhole image analysis, and then the Lucy-Richardson algorithm was used to reconstruct the image. The image quality could be improved by applying gray level transformation and edge detection to enhance the image contrast and target resolution. Reference [2] proposed an improved mean filter algorithm, adding a threshold concept based on the original mean filter algorithm, which improved the image quality and the ability to distinguish hidden objects. Reference [3] presented a terahertz image denoising algorithm based on adaptive manifolds and high-dimensional filtering. Median filtering, adaptive manifold high-dimensional filtering, and the Laplace of Gaussian-based edge enhancement method were carried out to enhance the denoising capability. In [4], an improved threshold gray transform algorithm was used to remove serious noise in the terahertz image. Then, the image was sharpened and enhanced based on the Laplace operator. Reference [5] proposed the norm optimization constraint and enhanced filtering of terahertz images and then reconstructed the image with a higher resolution. However, these methods mainly focused on improving the image quality without considering the detection speed.

In [6], to improve the image object location detection, threshold segmentation was applied to segment objects from the background; the LOG operator was used to extract edge features, and gun detection in the terahertz image was achieved. In [7], by analyzing the gray histogram and fitting the curve, the seed point and growth threshold were used to grow the image region, and object segmentation and detection in the human terahertz image were achieved. In [8], according to the characteristics of the terahertz image, the Otsu threshold was used to segment the threshold, the contour tracking method was applied to extract the contours, and object detection in the human terahertz image was achieved. In [10], a clustering algorithm based on fuzzy local information C-Means was used to detect objects in human terahertz images. In [11], to address the difficulty of detecting edge objects in human terahertz images, an algorithm for automatic recognition of human body edge objects was proposed. In these methods, the objects were first separated from the image, after which they were classified and recognized in a process that was complex and slow. The analysis results on the speed and accuracy of object detection have not been reported in the literature.

In recent years, deep learning has made a breakthrough in target vision detection. Deep learning target detection methods based on region recommendation, especially the R-CNN series (R-CNN, SPP-net, Fast R-CNN, and Faster R-CNN), achieve a very high detection accuracy and have been widely applied; for example, Faster R-CNN has been used to detect human faces, and X-ray diagrams to identify specific bone makers [14][15]. However, R-CNN series methods cannot satisfy the real-time requirements for the detection speed. The deep learning method without region suggestion (for example, the SSD-YOLO series algorithm) achieves a considerably improved detection speed; however, its detection accuracy is still far from that of Faster R-CNN. Thus far, no detailed research has been done on deep learning methods for object detection in terahertz images.

Human body security inspection in dense populations has higher requirements for detection accuracy and speed, thus the need to achieve real-time detection without any loss of precision. In this research, according to the characteristics of terahertz human body security images, a method of object detection that applies Faster R-CNN based on fusion preprocessing and structure optimization is proposed, which can greatly improve the detection speed, if improving the detection accuracy.

III. PREPROCESSING OF THZ HUMAN BODY SECURITY IMAGES

The main task of THz image preprocessing is to improve the signal-to-noise ratio (SNR) and resolution so that the image can be used for the next steps, that is, feature extraction and image recognition. Image preprocessing can improve the image quality but may result in feature change or loss. For example, objects of different materials show different grayscale levels on terahertz images. Some preprocessing algorithms can enhance the image by changing the pixel gray value, which may lead to inaccurate object detection results. The preprocessing of THz human body security images should meet the following requirements: (1) the noise contained in the image should be removed as much as possible; (2) the loss of important features contained in the image, such as edges and textures, should be avoided; and (3) additional noise should not be introduced in the process. The image contrast, image visual effect, and image information entropy should be balanced by using a preprocessing algorithm.

A. Denoising THz human body security images

In the existing image denoising method, the images in the spatial domain are directly denoised in the original airspace; that is, the gray value of the original image is processed directly through linear and nonlinear methods. In transform domain denoising, the image is changed into a transform domain signal, and then the corresponding coefficients in the transform domain are processed to remove the noise. Finally, the inverse transform is returned to the original space domain, and the denoised image is obtained. The frequency domain denoising algorithm has a good denoising effect on the noise with global statistical law; however, it can neither deal with a single pixel nor describe the local detail structure of the image very well. Nevertheless, the features of the THz human body security images are reflected in the local features. The correlation denoising algorithm based on the wavelet transform is another transform domain method. This approach has some advantages in image edge analysis but requires a large amount of computation. TABLE I shows the processing time of common denoising algorithms (experimental hardware configuration: i7-6700K *8 CPU; 4.0GHz main frequency, 16GB memory, and GTX750Ti/GTX1080 GPU)

TABLE I. Processing time of common denoising algorithms for THz human body security images(unit: ms)

Terahertz image resolution	Median filtering	Mean filtering	Gaussian filtering	Bilateral filtering
420*160	8.3	9.1	11	11.7

Considering the requirement of real-time processing, the spatial denoising method is chosen. Median filtering has an excellent denoising capacity for a certain type of random noise. As indicated in the processing results shown in Fig. 2, the median filter achieves better noise removal and reduction while maintaining well the target object (control tool) in the image.

(a) Original figure (b) Man filter (3×3) (c) Median filter (3×3)

Figure 2: Diagram of the mean and median filter effects

B. Enhancing THz Human body Security Images

Image enhancement algorithms can be divided into four classes: histogram equalization, wavelet transform, Retinex, and partial differential equation. Regarding effectiveness, the histogram equalization algorithm can improve the contrast of the original image. However, due to the merging of adjacent gray levels, loss of image information occurs, and the enhancement of information entropy is not obvious. The wavelet transform algorithm can also obviously improve the contrast of the original image; however, the noise is amplified in the process of enhancing the high-frequency image, resulting in a lower signal-to-noise ratio. The enhanced Reitnex algorithm has the best visual effect and the highest signal-to-noise ratio and information entropy. The partial differential equation algorithm achieves a moderate enhancement effect; the contrast, signal-to-noise ratio, and information entropy of the original image are also improved to some extent.

Regarding the processing speed, the histogram equalization algorithm has the shortest computation time and the best real-time performance. The wavelet transform algorithm requires wavelet decomposition and wavelet reconstruction of image, the partial differential equation algorithm requires finding the optimal solution by iterative calculation, and the Retinex algorithm requires computation of the Gaussian filtering . These three algorithms have a relatively long computation time.

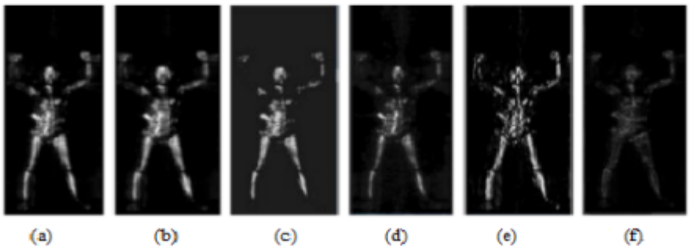
Considering the requirement of real-time processing, the histogram class of enhancement algorithms was selected.

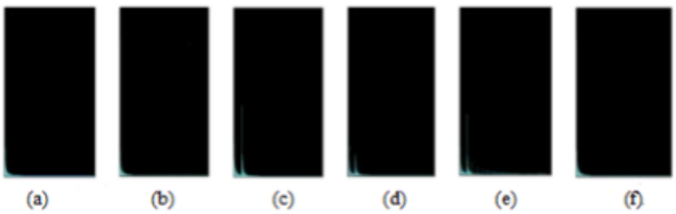
TABLE II shows the processing time of the different algorithms under this class (experimental hardware configuration: i7-6700K *8 CPU, 4.0GHz main frequency, 16GB memory, GTX750Ti/GTX1080 GPU).

According to the diagram of the enhanced effects of processing (shown in Fig. 3), both linear and nonlinear transformation have better enhancement effects on terahertz human images, and the contour of the object (cutter) is obvious. The pixel histogram (shown in Fig. 4) indicates that the pixel distributions in the linear and nonlinear transform are basically the same around the first peak value but are different. Exponential nonlinear transformation achieves better results in enhancing THz human body security images.

TABLE II. Processing time of different enhancement algorithms for terahertz human body security images (unit: ms)

Terahertz image resolution	linear transformation	non-linear transformation	Sobel operator	Gradient operator
420*160	10.8	11	11.8	12





(a) Original figure (b) Mean filter (3×3) (c) Linear transformation
(d) Exponential nonlinear transformation (e) Sobel operator
(f) Gradient operator

Figure 3: Diagram of the effects of different enhancement algorithm

Figure 4: Histogram of different enhancement algorithms

IV. HUMAN BODY TERAHERTZ IMAGE OBJECT DETECTION

A. *Deep learning detection algorithm*

The deep learning methods in the field of target vision detection typically include: (1) R-CNN techniques based on region proposals, such as R-CNN, SPP-net, Fast R-CNN, and Faster R-CNN; and (2) methods without region proposals, such as SSD and YOLO. The methods without region proposals are applied in real time and considerably improve the speed of target detection. TABLE III shows a comparison of the accuracy and speed of object detection in terahertz human body images between various algorithms (experimental hardware configuration: i7-6700K *8 CPU, 4.0GHz main frequency, 16GB memory, GTX750Ti/GTX1080 GPU). Software configuration: Faster R-CNN R-FCNN YOLOv3 neural network model based on Caffe framework. Training set: The objects are divided into two categories: cutting tools and mobile phones, 39 knives, 41 mobile phones, 38 terahertz images of human body security, and the objects also divided into two categories: cutting tools and mobile phones. Of which 15 are knives and 23 are mobile phones).

Regarding terahertz image object detection, the experimental results show that the methods based on region proposals, such as VGG16 and R-FCN ResNet-50, have better accuracy. The methods without region proposals, such as YOLOv3, achieve a considerable improvement in detection speed, but their detection accuracy is still far from that of methods based on region proposals. Therefore, the network structure of terahertz image object detection was optimized based on the Faster R-CNN method, and the detection speed was improved.

B. *Optimizing the Faster R-CNN Network structure*

The Faster R-CNN algorithm uses multilayer convolution to obtain a large number of image features and improve the accuracy of object detection. At present, the data sets used in object detection studies consist of colored pictures, including several object types. For example, the ImageNet data set comprises 1000 classes of objects. The number of PASCAL VOC data sets is relatively small

TABLE III. Comparison of the detection accuracy and speed in terahertz human body security images between deep learning algorithms

Network model	CNN		Training time（**min**）	Detection accuracy	Detection time（**ms**）
	Time complexity	Space complexity			
Faster R-CNN (VGG-16)	3.51×10^{10}	1.10×10^{7}	374	77.3%	37
R-FCN (ResNet-50)	1.78×10^{10}	0.85×10^{7}	300	78.4%	27
YOLOv3	2.25×10^{10}	1.19×10^{7}	330	67.4%	22

because the objects in the image vary greatly; however, each image may contain more than one object of different types, and the target scale also varies widely. The SUN data set covers large scenarios, locations, and object changes and contains 397 class scenarios. The MS COCO data set has 6 or 7 targets per image. Compared with these common data set images, terahertz images are grayscale images, which have a single scene, few kinds of objects, and less features in the object detection task.

In the application of the Faster R-CNN algorithm to object detection in terahertz human body security images, a too-deep structure may, on the one hand, lead to over-fitting and reduce the accuracy of object detection. On the other hand, such structure may lead to too much calculation and a lower detection speed. The network parameters and computational complexity can be effectively reduced by clipping the network. In the application of object detection in THz human body images, the present work applies hierarchical clipping to reduce the number of convolution and pool layers. Figure 5 shows the optimized network structure.

The experiment results show not only that the time complexity and space complexity of the model were greatly reduced but also that the training time was considerably decreased and the detection accuracy was improved. At the same time, compared with R-FCN, the

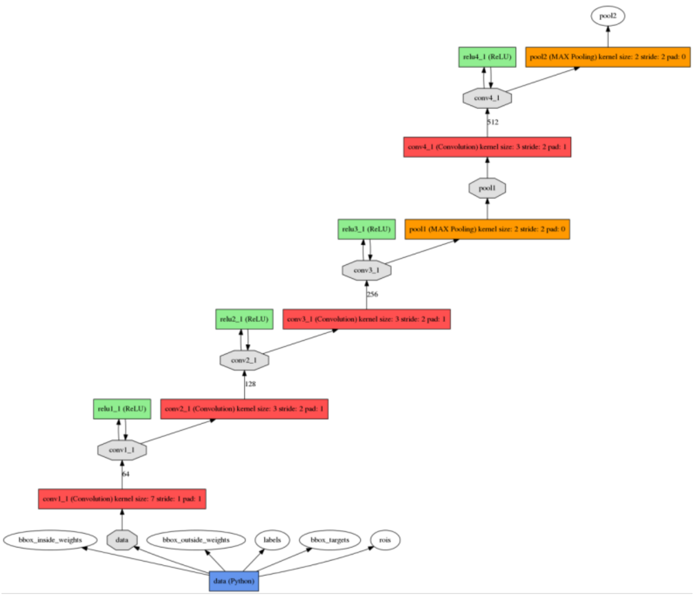


Figure 5: The optimized network structure

detection time was reduced from 27ms to 16ms, and the detection speed increased about 1.7-fold.

V. R-PCNN TO DETECT OBJECT OF HUMAN BODY THZ IMAGE

A R-PCNN detection framework

Figure 6 shows the detection framework (R-PCNN) applied after the image preprocessing and network structure optimization. During object detection: (1) the

TABLE IV. Comparison of the object detection performance in terahertz human body security images between optimized networks

Network model	CNN		Training time (min)	Detection accuracy	Detection time (ms)
	Time complexity	Space complexity			
Faster R-CNN (VGG-16)	3.51×10^{10}	1.10×10^{7}	374	77.3%	37
R-FCN (ResNet-50)	1.78×10^{10}	0.85×10^{7}	300	78.4%	27
Optimized Faster R-CNN	1.31×10^{10}	0.14×10^{7}	150	79.4%	16

original image is entered into the image preprocessing module for denoising and enhancement, and (2) the preprocessed images are detected by the trained network. Denoised and enhanced images are used to train the model in order to avoid the image preprocessing effect of grayscale stretching on the classification results.

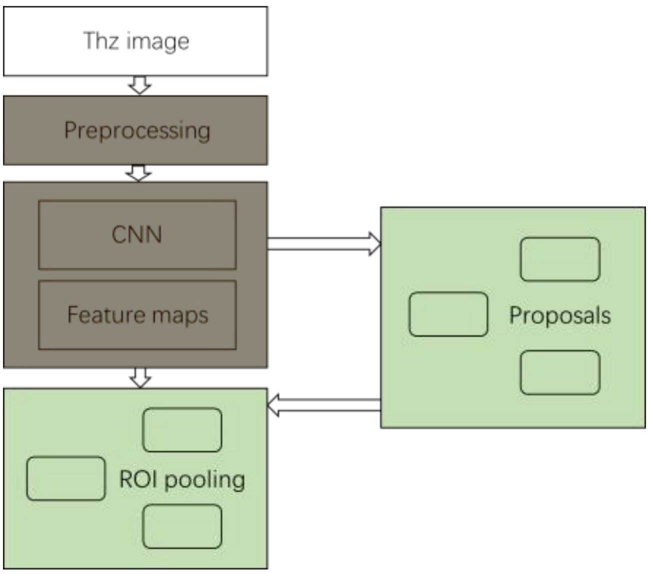


Figure 6: Object detection framework (R-PCNN) with preprocessing of THz human body security image

B Experiments

In this study, the algorithm experiments applied the same system environment. The hardware configuration consisted of: an i7-6700K *8 CPU, 4.0GHz main frequency, 16GB memory, and a GTX750Ti/GTX1080 GPU. The software configuration was Open CV-2.4.11 + Python 2.7 + Caffe +Faster R-CNN. The training set and test set settings used were the same as those described in Section 4.1. For several typical networks, the optimized Faster R-CNN network/R-PCNN framework carries out 60,000 iterative training and object detection tests,

respectively.TABLE V shows the performance results regarding the detection accuracy and speed.

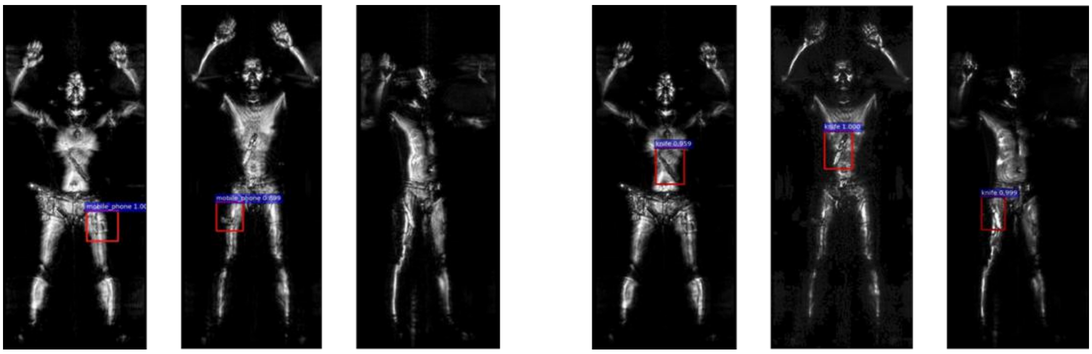
The experimental results indicate that, compared with VGG-16 and ResNet-50, the R-PCNN algorithm greatly reduces the complexity of the model, decreases the training time to less than half, and improves the detection accuracy by more than 6%. Compared with the optimized Faster R-CNN method, the detection time of part of the network is reduced from 16ms to 10 ms, and the detection time (30 ms) after preprocessing is between that of VGG-16 and ResNet-50.

1) Effects of preprocessing on detection accuracy

The human body terahertz image without preprocessing is directly detected by using a neural network; Fig.7(a) shows some errors and missed images. By using the median filter and exponential nonlinear transform for preprocessing of the human terahertz image and by

TABLE V. Comparison of the accuracy and speed of object detection in terahertz human body images between various networks

Network model	CNN Time complexity	CNN Space complexity	Training time（min）	Detection accuracy	Detection time（ms）
Faster R-CNN (VGG-16)	3.51×10^{10}	1.10×10^{7}	374	77.3%	37
R-FCN network (ResNet-50)	1.78×10^{10}	0.85×10^{7}	300	78.4%	27
Optimized Faster R-CNN	1.31×10^{10}	0.14×10^{7}	150	79.4%	16
R-PCNN	1.31×10^{10}	0.14×10^{7}	160	84.5%	10（network time）+20（preprocessing time）



(a) Detecting errors and missed images without preprocessing （b）Correct effect images after preprocessing

Figure 7：Effect of preprocessing on detection accuracy Effects of preprocessing on detection speed

applying an input neural network for object detection, the errors and missed images in Fig. 7(a) can be detected accurately, as shown in Fig. 7(b).

2) Effects of preprocessing on detection speed

Based on the experimental results, the detection speed is further improved after preprocessing. For the same image, the preprocessing and network detection can only be serial; the detection time is around 30 ms. In the dense crowd scene, the preprocessing and network detection are carried out in two stages, which can improve the throughput of the system. In the actual security scene, based on an imaging speed of 25 frames/s in a single channel, the image processing time should be less than 40ms. By applying the algorithm presented in this work, an image processing time of about 20ms after pipeline processing is achieved, which can satisfy the demand of security checks in high-density crowds.

VI. CONCLUSION AND FUTURE WORK

In this research, a Faster R-CNN detection framework (R-PCNN) that combines preprocessing and structural optimization is proposed. The experimental results show that this method can improve the accuracy and speed of object detection in THz human body security images. In a dense crowd scene, the detection accuracy reached 84.5 percent, and the detection time per image was less than 20 milliseconds. However, although this work reports some progress in terahertz human body security imaging, some issues remain in this field. The first is the low resolution and signal-to-noise ratio of terahertz images. Despite preprocessing for image denoising and enhancement, terahertz images still have a low resolution and signal-to-noise ratio, which are the main reasons for the difficulty in improving the accuracy of object detection. The second issue is the lack of diversity in terahertz image data sets. There are few terahertz images, including dangerous items, in actual security checks. Moreover, there are few changes in the types, sizes, and shapes of objects in the data sets currently available. The accuracy of model detection can be improved by increasing the scale and diversity of the training data. To address these two issues, the following two steps are suggested. The first is to reconstruct the super-resolution of the terahertz image by using deep learning methods to enhance the image resolution, which can in turn improve the accuracy of object detection. The second is to enrich the THz image data sets for human body security checks by using artificial scene and data enhancement methods.

REFERENCES

[1]. Rang Zhao, Yuan Meng Zhao, Cun Ling Zhang. Target aided identification in passive human THz image [J]. High Power Laser and Particle Beam, 2014, 26(07):132-136.

[2]. Wei Wen Zhu, Yue Jin Zhao, etc. Improved mean filtering algorithm in THz imaging [J]. Infrared and Laser Engineering, 2013, 42(05):1241-1246.

[3]. Shu Jin Zhu, Yue Hua Li. A terahertz image denoising algorithm based on adaptive manifolds and high-dimensional filtering [J]. Information and Electronic Engineering, 2016, 14(04):497-501+507.

[4]. [J]. Information and Electronic Engineering, 2016, 14(04):497-501+507.

[5]. Jian Zhong Shu, Yue Zhang, Fei Gao, Jiu Sheng Zhu. Quality Improvement and Noise Reduction of Terahertz Continuous Wave Imaging [J]. Spectroscopy and Spectral Analysis, 2017, 37(11):3343-3346.

[6]. Zhi Zhang, Yu Lin Lin, Jian Bin Zhang, Yan Hong He. Improvement method for terahertz imaging quality [J]. Infrared and Laser Engineering, 2017, 46(11):298-302.

[7]. Ping Yan Feng, Hui Wang. Feature Recognition of Pistol in Terahertz Image Based on Threshold Segmentation and Edge Detection [J]. Infrared, 2011, 32(12):23-26.

[8]. Xin Zhang, Yuan Meng Zhao, Cun Lin Zhang. Passive THz Image Segmenta algorithm [J]. High Power Laser and Particle Beam, 2013, 25(06):1597-1600.

[9]. Xin Zhang, Yuan Meng Zhao, Cun Chao Deng, Lin Zhang. Study on the pas Terahertz Image Target Detection

[10]. Wei Wen Zhu, Yue Jing Zhao, etc. Image Segmentation of THz Image [J]. T Beijing Inst Technol, 2013, 33(04):399-402.

[11]. Tian Jiang, Hui Liang Shen, etc. Target Detection of THz image Based on C-m of fuzzy local information [J]. Laser Technology, 2015, 39(03):289-294.

[12]. Lin Hua Wang, Ming Hui Yuan, etc. Recognition of edge object of human boc THz security inspection system [J]. Infrared and Laser Engineering, 2 46(11):149-154.

[13]. Shaoqing Ren, Kaiming He, Ross Grishick, Jian Sun, "Faster R-CNN: Tow Real-Time Object Detection with Region Proposal Networks," IEEE Transact on Pattern Analysis and Machine Intelligence, pp.1137-1149, 2016.

[14]. T. Hoang, Ngan Le, Yutong, Zheng, Chenchen Zhu, Khoa Luu, Marios Savvi "Multiple Scale Faster-RCNN Approach to Driver's Cell-Phone Usage and H on Steering Wheel Detection," 2016 IEEE Conference on Computer Vision Pattern Recognition Workshops (CVPRW), pp. 46-53,2016.

[15]. Huaizu Jiang, Erik Learned-Millter, "Face Detection with the Faster R-CNN," 2 12th IEEE International Conference on Automatic Face & Gesture Recognition 2017), pp. 650-657, 2017.

[16]. Ruhan Sa, William Owens, Raymond Wiegand, Mark Studin, Donald Capol Kenneth Barooha, Alexander Greaux, Robert Rattray, Adam Hutton, Jhon Cinti Vipin Chaudhary. "Intervertebral disc detection in X-ray images using faste CNN," 2017 39th Annual International Conference of the IEEE Engineerin Medicine and Biology Society (EMBC), pp. 564-567,2017.