Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/19/20231016

Survey of object detection in terahertz images

Hongshan Li

International College Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

hongshanli2001@163.com

Abstract. In recent years, object detection algorithms have undergone a further development and improvement, resulting in a wider range of application scenarios. As one of the most fundamental and challenging issues in the field of computer vision, the application of object detection in the field of security has also received considerable attention. Terahertz (THz) imaging which is widely used in this area because of the ability to detect hidden objects, as a type of electromagnetic wave imaging with poor imaging performance and low resolution, traditional target detection methods cannot achieve high robustness and effectiveness simultaneously. However, anyway in this possible application scenario, many possible ideas and algorithms have been proposed. This paper analyzes the possibility of applying different object detection methods to terahertz images and analyzes the existing problems in order to give the learner some basic idea and future direction. Several detectors are covered, including the traditional object detection methods and the algorithms based on Convolutional Neural Network(CNN) framework.

Keywords: object detection, THz images, CNN framework.

1. Introduction

Nowadays, the importance of counter-terrorism has gradually been recognized by more people. The safety inspection of public transport such as airports and subway stations has received widespread attention. Terahertz (THz) is an electromagnetic wave (EM) with a frequency range of 0.1-10THz. It has strong penetrability which can image opaque objects, and is harmless to organisms. Due to these characteristics, terahertz imaging technology has been widely used in safety inspection. However, the terahertz safety inspection is still manual so far, which not only leads to low inspection efficiency, but also consumes a lot of labor costs. Therefore, the application of target detection technology in terahertz images is crucial for improving efficiency and reducing costs.

However, for passive terahertz images, the imaging quality depends largely on the hardware system and is vulnerable to external interference. On the one hand, because the EM energy reflected by the living body is relatively weak, the detection of the detector is difficult; On the other hand, THz image has serious noise pollution and low SNR and resolution. Due to the above factors, traditional computer vision processing methods have some difficulties in processing THz images.

Object detection is a basic and important task of computer vision, which has been widely used in real life. Its purpose is to answer two questions: whether the target exists and where it is located. The two important indicators are accuracy (classification accuracy and positioning accuracy, etc.) and detection efficiency [1]. This article aims to provide learners with some possible computer vision processing

DOI: 10.54254/2755-2721/19/20231016

schemes for terahertz images, including traditional target detection algorithms and target detection algorithms based on depth learning.

2. Traditional object detection

2.1. Segmentation-based detection algorithms

The segmentation algorithms is to separate the target from the background in an image. For gray-level image, the pixels in the region generally have gray similarity, while the boundaries of the region generally have gray discontinuity. The following are several common segmentation algorithms:

2.1.1. Maximum entropy segmentation method. The entropy of image can indicate the chaotic degree of image information. The larger entropy is, the more unclear information is which means that it is difficult to get effective information from the message [2]. This method uses the entropy of the image as the criterion to segment the image.

2.1.2. Region growth algorithm. The basic idea of region growing algorithm is to merge pixels with similar properties. For each region, it first specifies a seed point as the starting point for growth, and then compares the pixels in the area around the seed point with the seed point. Then, it is necessary to combine points with similar attributes and continue to grow outwards until non compliant pixels are included[3]. The speed of segmentation-based detection algorithms is relatively fast and the basic idea is is relatively simple which make it easy to complete. However, for THz images, due to the large noise and low image resolution during imaging, the detection effect is not ideal. In addition, people being tested often carry different kinds of items, which increases the number of detection targets and make detection more difficult.

2.2. Feature matching-based detection algorithms

In traditional image processing, image feature matching has the following three basic steps:

feature extraction:

This step is to extract key points (feature points, corner points, etc.) from the picture, generally include the location, scale and direction of key points

feature description:

The description is to describe feature points with a set of mathematical vectors, which mainly ensures that different vectors and different feature points have a corresponding relationship and make the differences between similar key points are as small as possible.

feature matching:

The matching process between feature points is actually to calculate the distance between feature vectors and judge whether the feature points are similar by calculating the distance between different feature description vectors. The common distances include European distance, Hamming distance, cosine distance, etc.

2.2.1. SIFT. The scale space of an image refers to the blurring degree of the image, not the size of the image. The blurring degree of objects from different distances is different. The more blurred the image, the larger the scale space of the image is.

Scale-invariant feature transform (SIFT), a computer vision algorithm used to get and display features in local of an image, searches for extreme points in the spatial scale and shows their location, scale, rotation in-variants, etc. SIFT feature is the image of feature in local, which can confront in a certain degree to rotation, scaling, brightness change, and also contains the stability to angle change, affine transformation, and noise [4]. Besides, the optimized SIFT matching algorithm can achieve the effect of real-time detection.

DOI: 10.54254/2755-2721/19/20231016

2.2.2. HOG. Histograms of Oriented Gradients (HOG) is a description operator based on shape edge features that can detect objects [5].

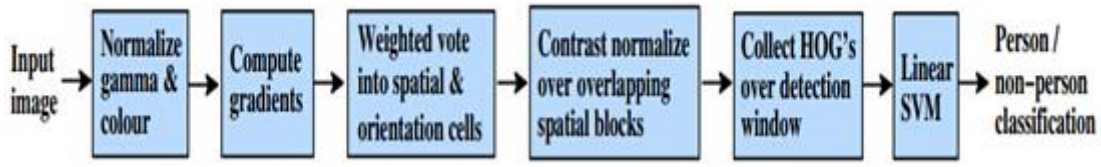

Figure 1. HOG feature extraction process.

In general, the feature extraction of feature matching-based detection algorithms mainly depends on manual feature extraction, as shown in Figure 1, and these algorithms have relatively good resistance to noise. However, the types of artificially constructed features are limited and the number of items carried by people subject to security inspection is very large, which makes the detection effect often unable to meet the requirements.

3. Improved SSD network

The current image object detection algorithms in computer vision are generally depended on R-CNN models and people divided them into two categories according to the number of networks, one is two-stage detector. The common algorithms are R-CNN,Fast R-CNN, Faster R-CNN. The other one is one-stage detector. At present, the mainstream algorithms are Yolo (you only look once) and SSD. Two-stage have higher detection accuracy but the speed is lower than one -stage

3.1. SSD introduction

Single Shot MultiBox Detector (SSD), according to the above, is a one-stage algorithm. The main idea of these detectors is to uniformly conduct dense sampling at different locations of the image, then use CNN to extract features which is used to perform classification and regression directly. Different scales and aspect ratios can be used for sampling. SSD abandons the full connectivity layer and uses CNN for direct analysis, which makes SSD more efficient [6]. However, there are still some problems in SSD algorithm. To begin with, the resolution of the largest feature map which is used for prediction can only reach to 3838. A lower resolution may result in the loss of detailed information on smaller targets contained in the lower layer at the pooling layer. In summary, in order to improve the ability to detect hidden objects in passive terahertz images, an improved network structure of existing SSD has been proposed by Lu Cheng to increase the ability of detector to detect small objects in complex environments.

3.2. Improvement of the netwok

The improved SSD framework can be shown in Figure 2. Overall, the improvement of the algorithm can be divided into four parts. First, VGGNet-16 network with stronger feature extraction ability is used to replace the original ResNet-50 network with poor performance. Adding a new feature layer to the existing basic network which enhance the object detection layer in feature expression. In the next part, in order to increase the correlation of semantic information between the front and rear scale maps, three different scale feature maps will be added to the feature extraction layer [7]. Then, they introduce a hybrid attention mechanism in SSD to enhance the Semantic information of high-level feature maps. This method improves the detector's ability to extract object details and information of position which help reducing the miss rate and false rate of detection [7]. Finally, they introduce the Focal Loss function to increase the negative samples weight and hard samples weight in the loss function which largely improve the robustness of the algorithm.

DOI: 10.54254/2755-2721/19/20231016

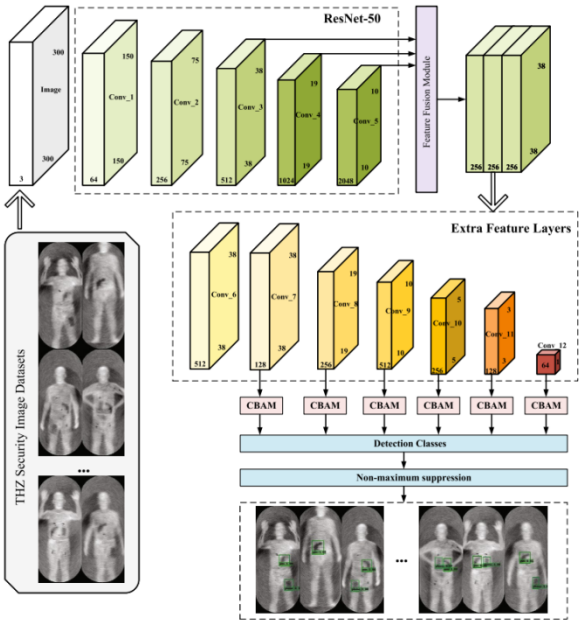

Figure 2. Improved SSD network architecture [7].

The improved algorithm process can be represented by the following block diagram Figure 3, and overall, it is not much different from other 2-stage detector processes.

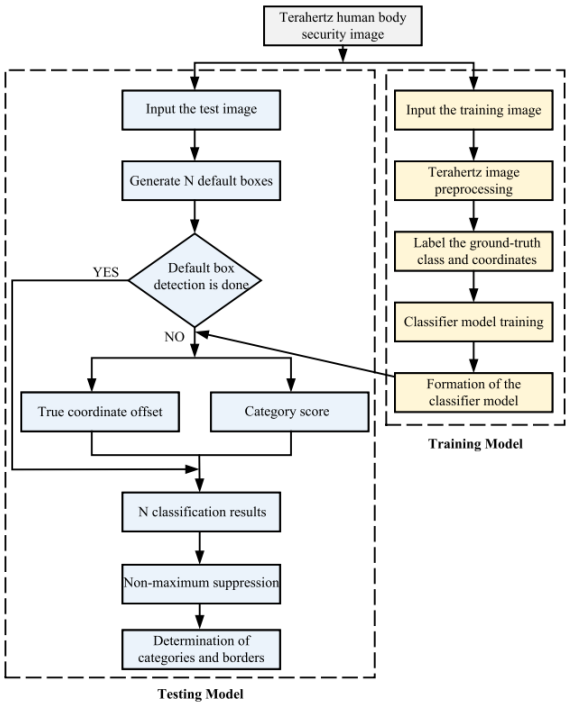

Figure 3. Flowchart of the algorithm [7].

4. CNN with spatial-temporal information

Because of the similarity of the human body in continuous terahertz safety images, the differences in suspicious objects result in different images. Based on this feature, Professor Xi Yang et al. proposed a detection method that utilizes the SLD algorithm [8] to import spatial-temporal information into the

DOI: 10.54254/2755-2721/19/20231016

Faster R-CNN framework to realize fast and accurate detection and recognition of suspicious objects. The algorithm is generally divided into two parts like Figure 4, rough detection and detailed detection.

4.1. Rough detection

The goal of rough detection is to determine the location of suspicious objects, that is, to output a report saying, 'there may be an object'. According to the differences of detected objects, SLD is used to exploit this spatial-temporal information. Here, L represents the static background, and the output is the body of people; S represents the displaced targets which output is the detected objects [9]. Besides, the morphological processing with shape knowledge is used in order to reduce noise influence. Finally, in the last part of the rough detection, the binary image is generated to determine the rough position of suspicious objects which will be marked in the image.

4.2. Detailed recognition

In rough detection, the detector can only obtain the approximate position of the suspicious object but cannot provide more details. Consequently, based on the original algorithm, the results of rough detection are imported into the Faster R-CNN framework to complete further detection. They divided this module into two stages, i.e., offline training and online testing. In the training part, a great number of training images which the detected objects have been marked their corresponding objects labels will be trained in the Faster R-CNN. The model after trained by the Faster R-CNN framework can effectively extract advanced semantic features of all region proposal which is beneficial to recognize the suspicious objects [10]. In the testing stage, the trained Faster R-CNN first create the feature maps. Next, as the locations of detected objects can be described by the result of the rough detection, a narrow-band domain can be achieved by the projection of location. Therefore, the detector only computes region proposals centered at the anchors in the narrow-band domain. As the reduction of the background influence and calculated region proposals amount, the detection efficiency and accuracy are improved compared with the traditional recognition methods.

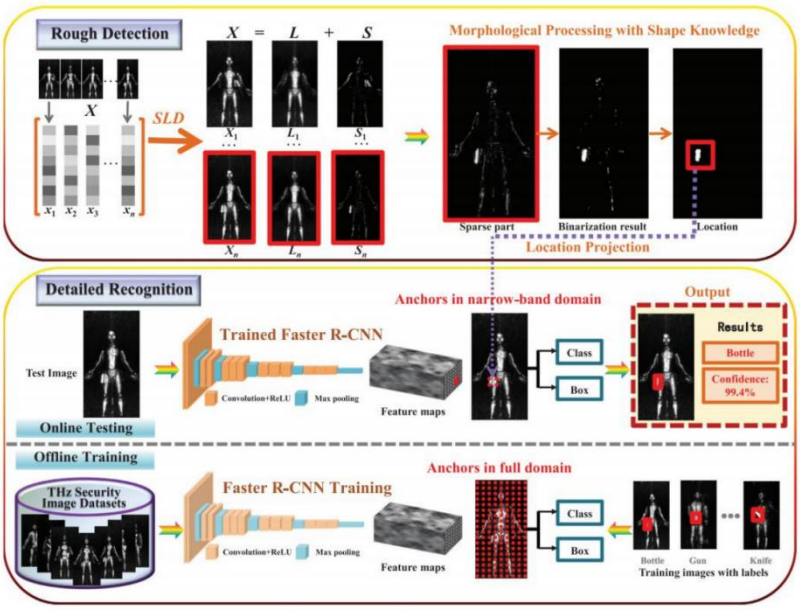

Figure 4. Diagram of the proposed two detection.

5. Conclusion and future directions

In the past two decades, significant breakthroughs have been made in the field of object detection algorithms. With the emergence and improvement of more advanced algorithms, the application scenarios of neural network algorithms have become increasingly widespread. However, object

DOI: 10.54254/2755-2721/19/20231016

detection of passive THz image which is used in the security checking is still a relatively incomplete field in computer vision. This technology still needs more development to mature, and there is currently no established standard evaluation protocol. This article aims to provide learners with some processing ideas or possible solutions for THz images to help them gain more insights.

The advantage of traditional object detection lies in its relatively simple approach and relatively fast speed. However, due to the high noise and low resolution of THZ images, it currently appears that this is not a suitable solution. At present, there are many application scenarios for CNN based frameworks, and the algorithm itself is relatively complete. The author believes that it is the most likely solution in the future. However, there is a common problem with object detection algorithms, which is that it is currently difficult to master the detection ability of unknown category targets. For the field of security, the items people carry are diverse, and there will inevitably be a large number of unknown items. Therefore, achieving the detection of unknown objects is a problem that must be solved in the future application of object detection algorithms in THz images.

References

[1] Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object detection in 20 years: A survey. Proceedings of the IEEE.

[2] Kapur, J. N., Sahoo, P. K., & Wong, A. K. (1985). A new method for gray-level picture thresholding using the entropy of the histogram. Computer vision, graphics, and image processing, 29(3), 273-285.

[3] Adams, R., & Bischof, L. (1994). Seeded region growing. IEEE Transactions on pattern analysis and machine intelligence, 16(6), 641-647.

[4] Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision (Vol. 2, pp. 1150-1157). Ieee.

[5] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). Ieee.

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing.

[7] Cheng, L., Ji, Y., Li, C., Liu, X., & Fang, G. (2022). Improved SSD network for fast concealed object detection and recognition in passive terahertz security images. Scientific Reports, 12(1), 1-16.

[8] Hu, Y., Zhang, D., Ye, J., Li, X., & He, X. (2012). Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE transactions on pattern analysis and machine intelligence, 35(9), 2117-2130.

[9] Yang, X., Wu, T., Zhang, L., Yang, D., Wang, N., Song, B., & Gao, X. (2019). CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images. Signal Processing, 160, 202-214.

[10] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).