World Scientific Research Journal Volume 9 Issue 12, 2023
ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

Deep Learning for The Recognition of Terahertz Security Image: A Review

Yu Jiang[1], Yizhang Li[1], Fenggui Wang[1,*], Zhongmin Wang[1], Xiangdong Li[1] and Guangjun Cao[2]

[1]Institute of Automation, Qilu University of Technology (Shandong Academy of Sciences), Key Laboratory of UWB & THz of Shandong Academy of Sciences, Shandong, China

[2]Fangyuan Building Materials Co., Ltd.

* Corresponding author: wfgranger@163.com

Abstract

In recent years, terahertz imaging has been widely used in security screening because terahertz waves can penetrate nonpolar materials without the risk of detonating bombs or harming living organisms. However, terahertz images are low-resolution and difficult to obtain, and terahertz image recognition presents challenges in certain areas, such as accuracy, speed and robustness. To address this challenge, the researchers chose a deep learning-based recognition method to identify terahertz images. This method mainly uses methods such as optimizing the existing model structure and introducing key techniques to make the model better adapt to the characteristics of terahertz images, and it has become the main method in the field of terahertz image recognition at present. In this paper, we provide a systematic review of terahertz image recognition based on deep learning, including the general process of recognition, preprocessing, backbone network, recognition network, model structure, defects and improvements. In addition, we discuss the feasibility of building a practical intelligent terahertz security system based on deep learning algorithms that considers the hardware performance contributions in this field. This work may be of guidance to terahertz security screening developers seeking to upgrade their products without massive hardware configuration changes.

Keywords

Deep learning; Image recognition; Terahertz; Security inspection.

1. INTRODUCTION

The significance of security inspection has been increasing nowadays to cope with challenges from terrorism and extremism. Security inspection by imaging is conventionally achieved at the X-ray band were phonons with high-energy pass-through package materials. Restriction on checking organism is gradually realized and overcome by imaging at the terahertz band. Terahertz wave (THz wave or T-ray) whose frequency bridges microwave and optic wave (1011Hz<f<1013Hz) is known penetrative to nonpolar molecules and harmless to humans. Researchers had realized that recognizing objects with human eyes is labor-consuming and not reliable. Deploying an efficient program to support automatic identification is desired for THz security inspection.

A number of THz imaging techniques have been developed that vary in configuration while not all of them are suitable for security check [1-10]. Terahertz time-domain spectroscopy (THz TDS) is a typical pulse system and it is reported that the advanced THz TDS can perform 3D

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

imaging with bandwidth as wide as 5 THz and dynamic range close to 100 dB [11]. Nevertheless, THz TDS acquires image by scanning multiple spectra with the movement of spot, therefore not competitive in video-rate spectral imaging. As a contrast, ultra-fast sub-Terahertz linear scanner that works at a single frequency for actual security screening has been developed [12]. New terahertz sensor technology makes for producing scalable detector arrays that allow ultrafast operation at room temperature [13, 14]. The device accommodates scanning speeds up to 15 m/s whereas the dynamic range is estimated 40 dB, resolution close to 5 mm. Algorithms are required to deal with trade-off in imaging quality.

At present, compared with light and infrared waves, terahertz images obtained by using terahertz waves have problems such as low resolution, visible fringe, low contrast and blurred edges. These problems are mainly due to the effects of factors such as detection sensitivity and electrical noise, as well as the effect of diffraction interference of optical waves, and also the absorption of terahertz waves by moisture in the air and atmospheric attenuation. Terahertz images present diverse content in security screening tasks, and their unexpected features increase the difficulty of feature engineering. Therefore, problems such as low resolution and blurred edges need to be addressed in practical applications, and methods that can extract complex features are needed for effective automatic identification of potentially hazardous materials.

Deep learning is an influential approach to achieve precise image recognition nowadays, which now has been employed in multiple machine vision tasks [15, 16]. As a data-driven method, precise end-to-end classification is not achieved until abundant images are trained to modify parameters of networks. The history of deep learning can be traced back to 2006 when Geoffrey Hinton torches passion of investigators to build deep networks by solving the problem of vanishing gradient [17]. Nowadays, a model with multiple hidden layers can be effectively trained with help of GPU. Thus, deep learning models are expected to extract complex features of THz images, which may promote efficient object recognition for security inspection.

Excellent algorithms and hardware levels are the main factors that enable successful applications of deep learning in image recognition. Only with mature hardware development can period required to train large neural networks on large data sets be kept within reasonable limits, especially with the innovation of GPUs. Prior to the advent of GPUs in 1999, models have to be trained using a single CPU. Due to low computational power, only models with a few layers could be trained, such as LeNet.5. From the advent of AlexNet in 2012 to the introduction of ResNet in 2015, the training arithmetic consumption of neural networks (in petaflop/s-day) increased by nearly two orders of magnitude. In parallel with the increase in model depth, the demand for computing power has also grown explosively. From K20X in 2012 to A100 in 2020, the inference performance of GPUs has increased to 317 times of the original [18]. It is suggested that the evolution of learning technology depends on the evolution of hardware

Our goal is to showcase state-of-the-art techniques in deep learning-based terahertz image detection. To achieve this, we conduct a comprehensive review of existing literature on deep learning-based terahertz image recognition. Our review covers various aspects such as the general process of recognition, preprocessing, backbone network, recognition network, model structure, flaws and improvements. We also analyze algorithms and evaluate the feasibility of current models in terahertz security systems. This work can serve as a valuable resource for developers working on terahertz security projects.

2. TERAHERTZ IMAGE RECOGNITION PROCESS

Due to the characteristics of terahertz images, terahertz images have certain deficiencies in terms of details and information content. This also leads to a certain degree of inapplicability of existing natural image recognition algorithms for terahertz image recognition, which faces great

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

challenges in complex scenarios such as target overlap, target too small, and target rotation. To solve these problems, in recent years, many researchers have started to optimize for the recognition of terahertz images. At present, the optimization of terahertz image recognition mainly focuses on these three processes: (1) Preprocessing process. Operations such as denoising, reconstruction and enhancement are often used to improve the quality of terahertz images; (2) Backbone network. Extraction of feature information in terahertz images by optimizing the backbone network structure; (3) Detection network. In addition to the three processes mentioned above, the incorporation of a number of other key technologies also has a profound impact on the accuracy and speed of terahertz image recognition.

2.1. Preprocessing

Image preprocessing is the first step in terahertz image recognition, and its main purpose is to eliminate irrelevant information from the image, retain useful and true information, and enhance the detectability and reliability of the information in question. The preprocessing enables better feature extraction, image segmentation, matching and recognition. Terahertz image preprocessing includes grayscale, filtering, histogram equalization, corrosion expansion, edge smoothing, noise reduction and reconstruction. Each step has a different role, but their common purpose is to extract image features as efficiently as possible for subsequent image analysis and processing. In the field of deep learning, the main methods for improving terahertz image quality are super-resolution reconstruction, and images denoising techniques. Super-resolution reconstruction solves the problem of low image resolution by providing an additional fraction of image detail. These include methods such as SRCNN [19], SRGAN [20] and EDSR[21]. These methods convert low-resolution images into high-resolution images by extracting features in terahertz images. Meanwhile, image denoising techniques mainly include blind denoising and non-blind denoising methods, and denoising can effectively reduce the noise interference in the terahertz image extraction process, eliminate irrelevant information and accelerate model training.

In 2016, the SRCNN super-resolution reconstruction model was first proposed by C. Dong et al. [19]. They divided the super-resolution reconstruction task into three parts: feature extraction, nonlinear mapping, and reconstruction. The following year, image super-resolution algorithms based on deep convolutional neural networks were applied to terahertz image reconstruction by Marina Ljubenovic et al. [22]. This research explores deep learning methods in the field of terahertz image reconstruction and achieves results that lay the foundation for subsequent related work. Immediately following this, two teams, Zhenyu Long [23] and Li, Yade [24], respectively, incorporated multi-level noise kernels with adjustable interpolation coefficients into the CNN structure for super-resolution reconstruction based on convolutional neural networks. The road to super-resolution reconstruction of terahertz images by convolutional neural networks is opened. In 2019, A generative adversarial network based terahertz image reconstruction method for terahertz image reconstruction has been proposed by W. Yibin et al. [25]. The teams of Z. Zhang [26] and Hou Z [27] have improved the terahertz image reconstruction method based on generative adversarial network respectively. Z. Zhang et al. used multiple degraded images with non-destructive detection to build an experimental dataset by using an improved GAN and introduced a pre-training model, which solved the problem of missing dataset to a certain extent. Hou Z et al. incorporated the enhanced attention (EA) into the network so that it can pay more attention to texture and detail reconstruction without affecting the image contour. In addition to the super-resolution reconstruction mentioned above, there is also the terahertz image deblurring technique investigated by Ljubenović M [28, 29] et al. They proposed for the first time a joint deblurring and denoising method for terahertz time-domain images that takes into account correlation blurring and different noise types. From the development history, the existing image super-resolution

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

techniques and denoising techniques mainly focus on studying how to map high-resolution images to artificially generated low-resolution images. To achieve this goal, researchers have used some key techniques, such as the multi-level noise kernel proposed in the literature [23], the method of adjusting the interpolation coefficients in the literature [24], and the enhanced attention mechanism introduced in the literature [27].

2.2. Backbone network

The role of the backbone network is to extract high level feature representations of the input image that contain abstract concepts that the target has, such as information about shape, texture and color. The backbone network serves as the foundation for many advanced image processing tasks in modern computer vision, such as object detection [30], human pose estimation [31], video classification [32], and image segmentation [33]. Considering the characteristics of terahertz images, we need a backbone network suitable for extracting features from low-resolution images. In the following, we briefly introduce the development status of backbone networks and some backbone networks mainly applied to terahertz images.

2.2.1. Milestones

In 1998, attempts to apply deep learning models to image classification gave rise to some pioneering backbone networks, as illustrated in Figure 1. LeCun et al. were the first to apply a backpropagation algorithm to convolutional neural networks (CNNs), and proposed the LeNet5 [34]. However, research on CNNs remained stagnant for a considerable period due to limitations in hardware performance and sample size. It was not until 2006 that Hinton et al. introduced the concepts of "deep learning" [17] and deep belief networks (DBNs) [35-37]. In 2012, the AlexNet [38] was trained using two GPUs, which partially addressed the issues of vanishing gradients and overfitting. In 2014, deep learning entered a real boom period with the emergence of a number of classic skeleton networks, such as VGGNet [39] and GoogleNet [40-42] that won the top two spots in the ILSVRC Challenge [43] that year for classification event. In 2015, Microsoft proposed a deep residual network (ResNet) [44] that addresses network degradation and improves characterization through unique cross-layer connectivity information mobility and fusion. In the same year, Jonathan Long et al. proposed full Convolutional Networks (FCN) [45]. In 2017, Huang et al. created the DenseNet [46] by "dense connectivity" to alleviate the gradient disappearance problem. Dense connectivity can greatly improve the information flow and feature sharing, and through dense connectivity, DenseNet and its improved algorithms are able to better capture the details and features of terahertz images, thus improving the recognition accuracy and robustness. To date, DenseNet has played an important role in backbone networks for terahertz image recognition. For example, Darknet-19 was used in YOLOv2, Darknet-53 in YOLOv3 [47], and CSPDarknet53 in YOLOv4 [48] and YOLOv5. In 2022, YOLOv6 [49] introduces a new backbone network called EfficientRep.

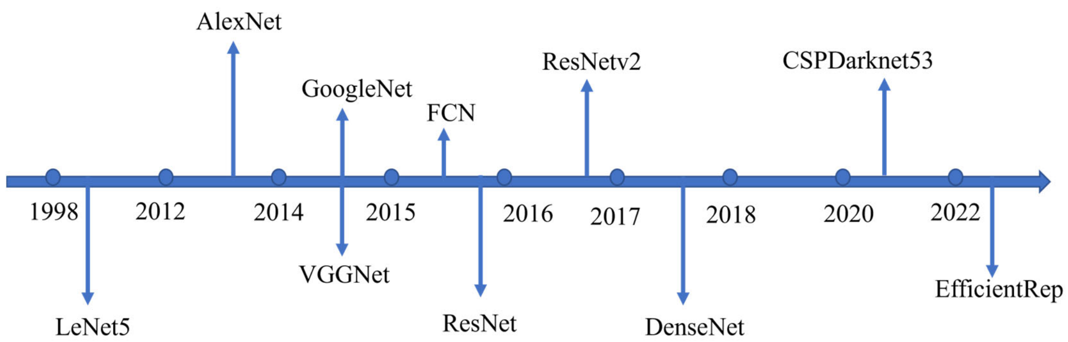

Figure 1. Major milestone in the backbone network

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

2.2.2 Comparison between different backbone

In order to better extract the features of complex data, researchers have made many optimizations to the backbone network. From Table 1, we can see that these optimizations focus on the following aspects: deepening, internal interoperability, and shrinking scale. First, deepening is an important direction. Early backbone networks had only a few layers of network structure and could not handle more complex data, such as LeNet.5 However, with the passage of time and the development of technology, the number of layers of backbone networks has been increasing, and models such as ResNet and Inception can even reach hundreds of layers, which greatly improves the capacity and expressiveness of the models. Second, interoperability within the backbone network is also very important. By connecting some or all layers of the backbone network, feature reuse can be achieved, such as DenseNet. finally, shrinking the network size has also become a trend, especially for large-scale deep neural networks, which are often difficult to run with limited computational resources. Therefore, researchers have explored ways to reduce computational complexity by reducing the network size, such as Inception structures that enable smaller parameters to reach deeper layers. This can greatly reduce the computational cost without significant loss of accuracy. With the popularity of new hardware devices such as mobile and embedded systems and the emergence of large scale data challenges, the subsequent development trend is bound to be more colorful, so let's wait and see.

Table 1. Comparison between different backbone

Backbone	Years	Input image size	Top-5 error rate (%)	Loss function	Params (MB)	Model size (MB)	GFLOPs (Forward pass)	Highlight
AlexNet	2012	224*224*3	20.91	cross-entropy	3MB	233 MB	0.7	ReLU and dropout; Data enhancement
VGG-16	2014	224*224*3	7.7	cross-entropy	58MB	528 MB	15.5	Small size convolution kernel and maximum pooling of 2x2
GoogleNet	2014	224*224*3	10.47	cross-entropy	51MB	27MB	1.6	Average pooling; Inception Module
ResNet-152	2015	224*224*3	5.94	cross-entropy	219MB	230 MB	11.3	Residual structure
DenseNet-121	2017	224*224	7.83	cross-entropy	126MB	31MB	3	The features of all layers are linked to channels

2.3. Detection network

In the field of computer vision, target detection is an important task that automatically identifies and localizes specific objects in images or videos. A deep learning-based target detection network then uses these features to classify and localize targets. Through continuous iterative training, such target detection networks can gradually learn to be more accurate and reliable in target recognition, leading to better results in practical applications.

2.3.1 Deep learning-based terahertz image recognition

The convolutional neural network (CNN) has shown excellent ability in image classification, but we need the target to be precisely located in addition to classifying the image in the process of hazardous object recognition, therefore, the above-mentioned classification model is difficult to meet the recognition needs. To solve the problem of accurate localization, many researchers have proposed two target detection methods based on deep learning, which are the candidate area-based target detection method and the regression-based target detection method.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

The candidate area-based target detection method uses candidate area to achieve feature area extraction. Based on this idea, Girshick et al. gave out named regions with CNN features (R-CNN) [50] to bridge the gap between image classification and object detection in 2013. The emergence of R-CNN has made deep learning methods a popular algorithm in the intelligent identification of security check contraband. With the continuous depth of research, the classical representative of the candidate region-based target detection algorithm was gradually developed: the Faster R-CNN [30]. The Faster R-CNN generates regions of interest through the Region of Proposal Network (RPN), which greatly improves the model detection performance and operation speed.The regression-based target detection approach is to view the target detection problem as a regression problem, partitioning the input image into bounding boxes and corresponding classes of probabilities, avoiding the extraction process of the region of interest in two target detection methods, and focusing more on speed. Based on this idea, in 2015, Joseph Redmon of the University of Washington, together with Ross Girshick, a researcher of Faster R-CNN, proposed the classical YOLO [51]. YOLOv1, although faster, has lower accuracy and is less effective in detecting small targets. In their further work, Redmon et al. [52] designed the backbone network DarkNet19 and proposed YOLOv2, while introducing anchor points and multiscale training to further improve detection accuracy. To improve the detection accuracy, Redmon et al. [47] introduced the idea of residuals and the idea of feature pyramids, and further proposed the backbone networks DarkNet53 and YOLOv3. In 2020, YOLOv4 and YOLOv5 were proposed. In 2022, ChienYao et al. [53] proposed YOLOv7, which improves detection by increasing the cost of the training process but does not improve the cost of the inference process.

Many subsequent improvements are based on two classical network models for dangerous object detection, Faster R-CNN and YOLO. Besides that, there are other methods such as Mask R-CNN [54]、SSD [55]、Mask Scoring R-CNN [56], and R-FCN [57] et al.

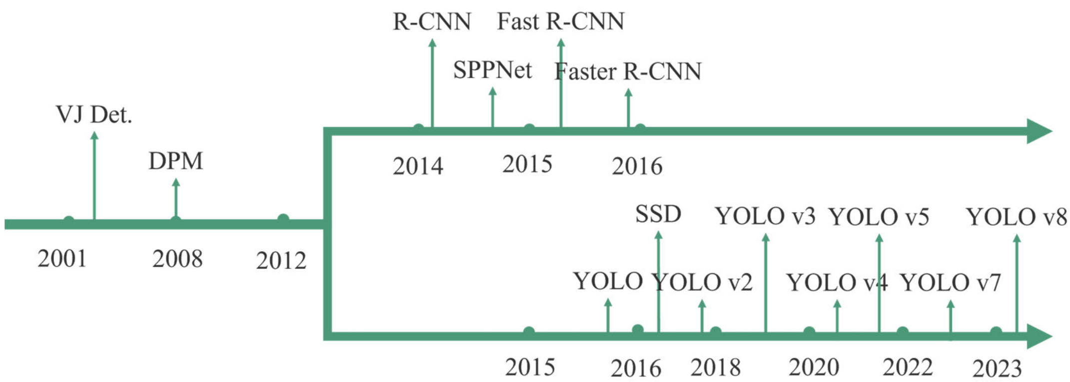

Figure 2. Major milestone in object detection research based on deep convolution neural networks

2.3.2 Application in terahertz image recognition

High-level features extracted through the backbone network can be used for object recognition, localization, and segmentation in target detection tasks, while deep learning-based target detection networks can use these features to detect and recognize different classes of objects more accurately. Therefore, only the combination of both techniques can lead to a higher level of performance and effectiveness of terahertz image recognition models. For example, while the VGG-16-based SSD struggles with small target detection, the ResNet50-based SSD overcomes this limitation (as shown in Table 2).

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

Table 2. Terahertz image recognition model based on SSD

Model	Author	Backbone	Years	Flash Point
SSD	Zhang C et al. [58]	VGG	2020	For small target detection, the k-mean clustering algorithm is used on the prior frame and the network structure of SSD is modified.
SSD	Cheng, L. et al. [59]	ResNet50	2022	The problem of repeated detection and missing detection of small and medium targets in terahertz images is solved

In the field of terahertz images, both VGG and Darknet are public backbone networks for feature extraction, and these networks possess powerful feature extraction capabilities for terahertz images. Faster R-CNN and YOLO based on them also produce powerful results in terahertz image recognition.As shown in Table 3, for the diversity and characteristics of terahertz images, the VGG backbone network-based recognition models are generally used in Faster R-CNNs for security screening tasks, with a few using SSDs (see literature [59] in Table 3 for details) or self-built models. In order to improve the model performance, researchers have introduced some key techniques in Faster R-CNN and modified the related structures. Among them, for the problem of multi-category targets and complex background interference in security screening tasks, the researchers used techniques such as threshold segmentation [60], semantic segmentation [61] and migration learning to continuously optimize the model. In addition, considering the special nature of terahertz images (e.g., low signal-to-noise ratio and small targets), researchers have also employed sparse low-rank decomposition (SLD) [62], preprocessing modules [63], and Inception-ResNet-A asymmetric convolution modules [64] to mine image information, thus further improving recognition accuracy.

Table 3. Terahertz image recognition model based on Faster R-CNN

Model	Author	Years	Flash Point
Faster R-CNN	Jinsong Zhang et al. [60]	2018	Terahertz classification dataset was built; Migration learning and threshold segmentation techniques were introduced to Faster R-CNN for independent detection of human and other objects.
Faster R-CNN	Hou B et al. [65]	2018	Faster R-CNN with deconvolution and shortcut concatenation; Optimizing the loss function of Faster RCNN by applying online hard sample mining to solve the sample imbalance problem.
Faster R-CNN	Xi Yang et al. [62]	2019	The sparse low-rank decomposition (SLD) mining spatio-temporal contextual information capability is applied to terahertz image detection.
Faster R-CNN	Xiuwei Yang et al. [66]	2020	The detection accuracy and efficiency of the network are improved by changing the backbone network, optimizing the training parameters, and improving the a priori box algorithm.
R-PCNN	H. Xiao et al. [63]	2018	The number of convolutional and pooling layers of Faster R-CNN is reduced by layered cropping, and a preprocessing module is introduced.
ResDeepNet	Guo L et al. [61]	2020	A deep convolutional network for semantic segmentation of terahertz images is designed.
——	Wang Q et al. [64]	2023	The Inception-ResNet-A asymmetric convolution module in the Inception-ResNet-V2 network is introduced into the VGG19 network structure.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

According to Table 4, the Darknet-based recognition model is mainly based on YOLO. In the YOLO model, researchers proposed a new loss function, called CIoU Loss [67], which can measure the similarity of bounding boxes more accurately. In addition, for the problems of low contrast and noise in terahertz images, researchers proposed a series of data enhancement strategies, such as MSFG enhancement method [68], BiFP structure [69], and CBAM and ASFF modules [70], to improve the robustness and generalization ability of the model. Also, researchers have modified the relevant structures to adapt to the low resolution of terahertz images, which in turn improves the recognition accuracy and speed. Overall, the continuous improvement and optimization of the model makes YOLO more suitable for application areas in terahertz image recognition and security screening tasks.

Table 4. Terahertz image recognition model based on YOLO

Model	Author	Years	Work
YOLOv3	Danso, S. A. et al. [67]	2021	Improve the accuracy of YOLOv3 by modifying the backbone and neck sections.
YOLOv4	X. Wang et al. [71]	2021	The degradation model is proposed to generate PCB pseudo-terahertz images, improve YOLOv4 to accurately detect four defects, and introduce migration learning to improve detection and classification accuracy.
YOLOv5i	Xu, Fan et al. [68]	2021	Application of MSFG enhancement method to YOLO networks.
YOLOv5	Danso et al. [69]	2022	Application of BiFP structures to the neck of YOLOv5 as a mechanism for improving low resolution.
YOLOv4	JIN, Bo-Yang et al. [72]	2021	A hazardous material carry detection dataset is constructed, and the parameters of the original YOLO are adapted to the hazardous material carry detection task by normalizing the terahertz images and adjusting them.
YOLOX	Tian.N et al.[70]	2022	To reduce the effect of noise in terahertz images, the improved YOLOX network adds CBAM and ASFF modules.
YOLOv4	H. Xiong et al.[73]	2021	Improved accuracy by fusing terahertz imaging and optical imaging.

In the field of deep learning, there are many different models that can be used to solve various tasks. These models have different structures and parameter settings, and therefore can have different effects on different optimization methods. For example, optimizing one model using a gradient descent algorithm may yield better results than using a stochastic gradient descent algorithm, while the opposite is true for another model. Researchers need to keep trying different optimization methods to determine the best choice. They also need to understand the advantages and disadvantages of each optimization method and consider how they can be tuned for a particular model. It is also important to note that different datasets and training objectives can also have an impact on the optimization of the model, and thus also require flexibility to adapt to the situation. In conclusion, researchers in the field of deep learning need

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

to continuously focus on the impact of different models on optimization and actively explore new approaches to improve the performance of deep learning models. A review of the work done by researchers in terahertz image recognition can be summarized in three directions. First, the researchers introduced some key techniques into the backbone network or recognition network, thus improving the training speed, recognition speed, and accuracy. Second, they have further improved the accuracy and speed of terahertz image recognition by modifying and optimizing the structure of existing detection models. Finally, due to the long terahertz wavelength and fast signal decay, there are multiple scales of targets in terahertz images. To better detect these targets, researchers explore the use of multiscale detection techniques, which is a direction worthy of further research.

2.4. Key technology

In 2023, Cheng, L. et al. [74] proposed a passive terahertz human security image intelligent detection method named HE-DETR-DC5.The method combines histogram equalization (HE) processing and defense modeling (DETR) to improve the effectiveness of terahertz image recognition. At present, researchers generally apply some key techniques to improve the accuracy of terahertz image recognition. For example, to overcome the problem of low signal-to-noise ratio in terahertz images, they introduced techniques such as adaptive local thresholding methods, wavelet analysis, and robustness estimation, which further improved the quality of the images. In addition, in terms of the backbone network, the researchers tried different network structures and training methods, such as deep residual network-based and attention mechanism-based approaches (e.g., literature [75]), to better capture the texture features of terahertz images. In terms of detection networks, they have also explored and innovated model design and learning strategies, such as enhanced multi-scale feature graph (MSFG)-based methods and sparse low-rank decomposition (SLD)-based methods. By integrating and optimizing these techniques, the detection accuracy and real-time performance of terahertz images will be further improved, providing more reliable support for applications in nondestructive testing, medical imaging, security screening, and other fields.

3. THE PROSPECT ANALYSIS OF DEEP LEARNING IN THZ IMAGING

Deep learning outperforms other methods in analyzing visible-light images and hyperspectral images, thus it is assumed reasonable to promote the method in tasks based on THz images. The successful implementation of deep learning image is attributed to algorithm and dataset, which determines floor and ceiling of recognition. Classical models are influential to form investigation routine where inputting images, building backbone networks, extracting features, pooling, network head, classification or regression, non-maximum suppression (NMS) are focused and conducted in order. In other words, recognition of THz images does not contribute to kernel innovation of deep learning framework. It is believed that a majority of algorithms would work for recognizing images at various bands although the performance would vary. Data insufficiency is the negative factor for deploying deep learning. It is hard to acquire abundant security inspection images at THz band due to limited performance of hardware. An expedient would work that researchers record a video in which objects pass through the detection area repeatedly while changing their relative location and orientation every time. Afterward, video segmentation at frame scale is needed to prepare dataset with diverse contents. Similar work will be less difficult as humans would cooperate with researchers, posing randomly to extent sample generalization. It is proposed that global investigators build large scale database of THz images, marking hardware performance, environment parameters and labels. Such plan shall be initiated by well-known specialists and influential labs or companies. A probable parameter set may consist of frequency, power, Gamma value (if Gamma Correction is taken), SNR, aperture and etc. Investigators from

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

different affiliations would benefit from unified and inclusive dataset, shifting their attention to algorithm, making it possible to improve performance by deepening models.

The general goal of deep learning in THz image recognition is the same with that in other image recognition tasks. Investigators would pursue higher recognition accuracy and faster response for all possible input; they would also pursue robustness and minimum complexity of model if possible. It is assumed feasible and economical to combine present models to obtain better expectations. As an example, the candidate area-based target detection and the regression-based target detection algorithm are complementary to the other in speed and precision. Thus, a combination of two may worth further study.

Object recognition is convinced crucial in various fields including but not limited to public security and state defense. The studies centering deep learning and their applications in object recognition is believed plentiful and fruitful according to Figure 3. Deep learning may play a prominent role in recognition tasks in THz band.

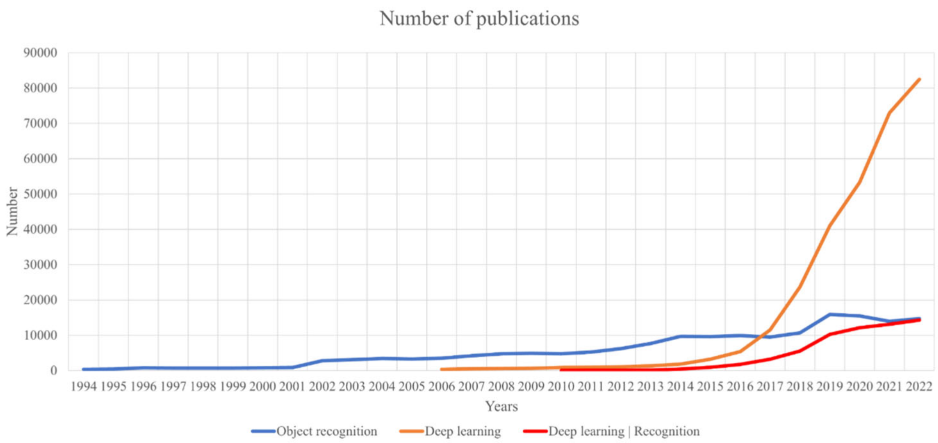

Figure 3. The increasing number of publications in object detection from 1994 to 2022. (Data from Web of Science advanced search: SU= “Object recognition”; “Deep learning”; “Deep learning & Recognition”)

4. SUMMARY

Compared with other methods, image recognition based on deep learning has strong learning ability and data-scale driven performance, which has great potential and advantages in terahertz image recognition. Meanwhile, With the continuous upgrading of computing power, deep learning-based target detection techniques will usher in breakthroughs in complex and large scale tasks. In this paper, we review popular deep learning models for image recognition and sketch overall progress in recent years with representative studies. The defects and improvement of existing methods are also discussed. Furthermore, we discuss the feasibility of building a practical and intelligent terahertz security inspection system based on a deep learning algorithm considering hardware performance and trace pioneer contributions in this field. We hope this paper can facilitate understanding of the field and guide future evolutions of deep learning in THz image recognition for security check.

ACKNOWLEDGMENTS

Funding Information This work is supported by Natural Science Foundation of Shandong Province Youth Foundation (ZR2020QF089), Shandong Province Key research and Development Program (2021JMRH0108), and Bidded Project in Qilu University of Technology (Shandong Academy of Sciences) (2022JBZ02-02, Major Innovation Special Project).

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

Data Availability Statement Data sharing not applicable to this article as no datasets were generated or analysed during the current study

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent As this article does not contain any studies with human participants or animals performed by any of the authors, the informed consent is not applicable.

Authors' contributions Yu Jiang and Yizhang Li: wrote the main manuscripting. Fenggui Wang and Zhongmin Wang: Supervision. Xiangdong Li: Investigation. Guangjun Cao: Methodology.

REFERENCES

[1] Usami M, Iwamoto T, Fukasawa R, Tani M, Watanabe M, Sakai K. Development of a THz spectroscopic imaging system. Physics in Medicine & Biology. 2002;47(21):3749.

[2] Zimdars D, Valdmanis J, White JS, Stuk G, Winfree WP, Madaras EI, editors. Time domain terahertz detection of flaws within space shuttle sprayed on foam insulation. Conference on Lasers and Electro-Optics; 2004: Optica Publishing Group.

[3] Baker C, Tribe WR, Lo T, Cole BE, Chandler S, Kemp MC, editors. People screening using terahertz technology. Terahertz for Military and Security Applications III; 2005: SPIE.

[4] Liu L, Rahman SM, Jiang Z, Li W, Fay P. Advanced terahertz sensing and imaging systems based on integrated III-V interband tunneling devices. Proceedings of the IEEE. 2017;105(6):1020-34.

[5] Hillger P, Grzyb J, Jain R, Pfeiffer UR. Terahertz imaging and sensing applications with silicon-based technologies. IEEE Transactions on Terahertz Science and Technology. 2018;9(1):1-19.

[6] Chan WL, Deibel J, Mittleman DM. Imaging with terahertz radiation. Reports on progress in physics. 2007;70(8):1325.

[7] Jepsen PU, Cooke DG, Koch M. Terahertz spectroscopy and imaging–Modern techniques and applications. Laser & Photonics Reviews. 2011;5(1):124-66.

[8] Knap W, Dyakonov M, Coquillat D, Teppe F, Dyakonova N, Łusakowski J, et al. Field effect transistors for terahertz detection: Physics and first imaging applications. Journal of Infrared, Millimeter, and Terahertz Waves. 2009;30:1319-37.

[9] Ruan S, Yang J, Zhang M, editors. Real-time terahertz imaging using a 1.63-THz optically pumped terahertz laser and a pyroelectric camera. 28th International Congress on High-Speed Imaging and Photonics; 2009: SPIE.

[10] Guillet JP, Recur B, Frederique L, Bousquet B, Canioni L, Manek-Hönninger I, et al. Review of terahertz tomography techniques. Journal of Infrared, Millimeter, and Terahertz Waves. 2014;35:382-411.

[11] Murate K, Kawase K. Perspective: Terahertz wave parametric generator and its applications. Journal of Applied Physics. 2018;124(16):160901.

[12] Shchepetilnikov A, Gusikhin P, Muravev V, Tsydynzhapov G, Nefyodov YA, Dremin A, et al. New ultra-fast sub-terahertz linear scanner for postal security screening. Journal of Infrared, Millimeter, and Terahertz Waves. 2020;41:655-64.

[13] Muravev V, Kukushkin I. Plasmonic detector/spectrometer of subterahertz radiation based on two-dimensional electron system with embedded defect. Applied physics letters. 2012;100(8):082102.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

[14] Muravev VM, Solov'ev VVe, Fortunatov AA, Tsydynzhapov G, Kukushkin IV. On the response time of plasmonic terahertz detectors. JETP Letters. 2016;103:792-4.

[15] Zou Z, Chen K, Shi Z, Guo Y, Ye J. Object Detection in 20 Years: A Survey. Proceedings of the IEEE. 2023:1-20.

[16] Wu X, Sahoo D, Hoi SCH. Recent advances in deep learning for object detection. Neurocomputing. 2020;396:39-64.

[17] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. science. 2006;313(5786):504-7.

[18] Dally B, editor Efficient methods and hardware for deep learning. Proceedings of the Workshop on Trends in Machine-Learning (and impact on computer architecture); 2017.

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

[20] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al., editors. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 21-26 July 2017.

[21] Lim B, Son S, Kim H, Nah S, Lee KM, editors. Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2017 21-26 July 2017.

[22] Ljubenović M, Zhuang L, De Beenhouwer J, Sijbers J. Joint deblurring and denoising of THz time-domain images. IEEE Access. 2020;9:162-76.

[23] Long Z, Wang T, You C, Yang Z, Wang K, Liu J. Terahertz image super-resolution based on a deep convolutional neural network. Applied optics. 2019;58(10):2731-5.

[24] Li Y, Hu W, Zhang X, Xu Z, Ni J, Ligthart LP. Adaptive terahertz image super-resolution with adjustable convolutional neural network. Opt Express. 2020;28(15):22200-17.

[25] Yibin W, Rongyue Z, Hong X, Hao W, Pan Y, Zhou Y, editors. Terahertz Image Super-Resolution Reconstruction of Passive Safety Inspection Based on Generative Adversarial Network. 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData); 2019 14-17 July 2019.

[26] Zhang Z, Zhang L, Chen X, Xu Y, editors. Modified Generative Adversarial Network for Super-Resolution of Terahertz Image. 2020 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD); 2020 15-17 Oct. 2020.

[27] Hou Z, Cha X, An H, Zhang A, Lai D. Super-Resolution Reconstruction of Terahertz Images Based on Residual Generative Adversarial Network with Enhanced Attention. Entropy. 2023;25(3):440.

[28] Ljubenović M, Bazrafkan S, De Beenhouwer J, Sijbers J, editors. CNN-based deblurring of terahertz images. Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2020, Valletta, Malta; 2020.

[29] Ljubenović M, Bazrafkan S, Paramonov P, Beenhouwer JD, Sijbers J, editors. CNN-based deblurring of THz time-domain images. Computer Vision, Imaging and Computer Graphics Theory and Applications: 15th International Joint Conference, VISIGRAPP 2020 Valletta, Malta, February 27–29, 2020, Revised Selected Papers; 2022: Springer.

[30] Girshick R, Donahue J, Darrell T, Malik J, editors. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

[31] Toshev A, Szegedy C, editors. Deeppose: Human pose estimation via deep neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.

[32] Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L, editors. Large-scale video classification with convolutional neural networks. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition; 2014.

[33] Haralick RM, Shapiro LG. Image segmentation techniques. Computer vision, graphics, and image processing. 1985;29(1):100-32.

[34] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-324.

[35] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. Neural computation. 2006;18(7):1527-54.

[36] Lee H, Grosse R, Ranganath R, Ng AY, editors. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. Proceedings of the 26th annual international conference on machine learning; 2009.

[37] Lee H, Grosse R, Ranganath R, Ng AY. Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM. 2011;54(10):95-103.

[38] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Communications of the ACM. 2017;60(6):84-90.

[39] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.

[40] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

[41] Ioffe S, Szegedy C, editors. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning; 2015: pmlr.

[42] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z, editors. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

[43] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. International journal of computer vision. 2015;115:211-52.

[44] He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

[45] Long J, Shelhamer E, Darrell T, editors. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.

[46] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

[47] Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:180402767. 2018.

[48] Bochkovskiy A, Wang C-Y, Liao H-YM. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:200410934. 2020.

[49] Li C, Li L, Jiang H, Weng K, Geng Y, Li L, et al. YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:220902976. 2022.

[50] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence. 2015;37(9):1904-16.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

[51] Redmon J, Divvala S, Girshick R, Farhadi A, editors. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

[52] Redmon J, Farhadi A, editors. YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

[53] Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:220702696. 2022.

[54] He K, Gkioxari G, Dollár P, Girshick R, editors. Mask r-cnn. Proceedings of the IEEE international conference on computer vision; 2017.

[55] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al., editors. Ssd: Single shot multibox detector. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14; 2016: Springer.

[56] Huang Z, Huang L, Gong Y, Huang C, Wang X, editors. Mask scoring r-cnn. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019.

[57] Dai J, Li Y, He K, Sun J. R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems. 2016;29.

[58] Zhang C, Shi X, Huang Z, Fu J. A Method for Small Object Detection of Passive Terahertz Image in Human Body Security Check—Subtitle as Needed. 2020.

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

[60] Zhang J, Xing W, Xing M, Sun G. Terahertz image detection with the improved faster region-based convolutional neural network. Sensors. 2018;18(7):2327.

[61] Guo L, Zhang H-K, editors. ResDeep convolutional networks for terahertz image semantic segmentation. AOPC 2020: Optical Sensing and Imaging Technology; 2020: SPIE.

[62] Yang X, Wu T, Zhang L, Yang D, Wang N, Song B, et al. CNN with spatio-temporal information for fast suspicious object detection and recognition in THz security images. Signal Processing. 2019;160:202-14.

[63] Xiao H, Zhang R, Wang H, Zhu F, Zhang C, Dai H, et al., editors. R-PCNN Method to Rapidly Detect Objects on THz Images in Human Body Security Checks. 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); 2018 8-12 Oct. 2018.

[64] Wang Q, Zhang Y, Ge H, Jiang Y, Qin Y, editors. Evaluation of rice freshness based on a modified VGG network and terahertz imaging. Earth and Space: From Infrared to Terahertz (ESIT 2022); 2023: SPIE.

[65] Hou B, Yang M, Sun X. Real-time object detection for millimeter-wave images based on improved faster regions with convolutional neural networks. Laser & Optoelectronics Progress. 2019; 56 (13): 131009.

[66] Yang X, Liu P, Wang S, Wu B, Zhang K, Yang B, et al. Defect Detection of Composite Material Terahertz Image Based on Faster Region-Convolutional Neural Networks. Materials. 2022;16(1):317.

[67] Danso SA, Liping S, Deng H, Odoom J, Chen L, Xiong Z-g. OPTIMIZING YOLOV3 DETECTION MODEL USING TERAHERTZ ACTIVE SECURITY SCANNED LOW-RESOLUTION IMAGES. Theoretical & Applied Science. 2021(3):235-53.

[68] Xu F, Huang X, Wu Q, Zhang X, Shang Z, Zhang Y. YOLO-MSFG: toward real-time detection of concealed objects in passive terahertz images. IEEE Sensors Journal. 2021;22(1):520-34.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202312_9(12).0012

[69] Danso SA, Shang L, Hu D, Odoom J, Liu Q, Nana Esi Nyarko B. Hidden Dangerous Object Recognition in Terahertz Images Using Deep Learning Methods. Applied Sciences. 2022;12(15):7354.

[70] Tian N, Wang X, Xu Y, Wang R, Lian G, Zhang L, editors. Deep Learning Enabled Hidden Target Detection in Terahertz Images. 2022 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD); 2022 30 Nov.-2 Dec. 2022.

[71] Wang X, Zhang Z, Xu Y, Zhang L, Yan R, Chen X, editors. Real-time minor defect recognition of pseudo-terahertz images via the improved yolo network. 2021 International Conference on Sensing, Measurement & Data Analytics in the era of Artificial Intelligence (ICSMD); 2021: IEEE.

[72] JIN B-Y, YAN F, LIU T-H. Research on Security Inspection Method Based on Terahertz Imaging and Convolution Neural Network.

[73] Xiong H, Cai J, Zhang W, Wu X, editors. Silkworm Egg Development Intelligent Recognition by THz Imaging. 2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz); 2021 29 Aug.-3 Sept. 2021.

[74] Cheng L, Zhang W, Li C, Ji Y, Liu X, Fang G. HE-DETR-DC5 method to rapidly detect objects on passive THz images in human safety inspection. Electronics Letters. 2023;59(2):e12702.

[75] Yang X, Zhang D, Wang Z, Zhang Y, Wu J, Wu B, et al. Super-resolution reconstruction of terahertz images based on a deep-learning network with a residual channel attention mechanism. Applied Optics. 2022;61(12):3363-70.