Impact Factor:					
	ISRA (India) = 6.317	SIS (USA) = 0.912	ICV (Poland) = 6.630
	ISI (Dubai, UAE) = 1.582	РИНЦ (Russia) = 0.126	PIF (India) = 1.940
	GIF (Australia) = 0.564	ESJI (KZ) = 9.035	IBI (India) = 4.260
	JIF = 1.500	SJIF (Morocco) = 7.184	OAJI (USA) = 0.350

SOI: 1.1/TAS DOI: 10.15863/TAS

International Scientific Journal

Theoretical & Applied Science

p-ISSN: 2308-4944 (print) **e-ISSN:** 2409-0085 (online)

Year: 2021 **Issue:** 03 **Volume:** 95

Published: 26.03.2021 **http://T-Science.org**

QR – Issue QR – Article

Samuel Akwasi Danso
Southwest University of Science & Technology, Information &Communication Engineering
PhD Candiadate, School of Information Engineering
Mianyang-Sichuan, China.

Shang Liping
Southwest University of Science & Technology, Information &Communication Engineering
Professor, School of Information Engineering
Mianyang-Sichuan, China.

Hu Deng
Southwest University of Science & Technology, Information &Communication Engineering
Professor, School of Information Engineering
Mianyang-Sichuan, China.

Patrick Bobbie
Kennesaw State University
Professor, Computer Science
Marietta Campus1100 S. Marietta Parkway Marietta,
GA 30060, U.S.A

Michael Asante
Computer Science Dept, KNUST-Ghana
Professor, Computer Science Dept.

Justice Odoom
Southwest University of Science & Technology, Information &Communication Engineering
PhD Student, School of Computer Science
Mianyang-Sichuan, China.

Daniel Appiah
Nanjing University of Information Science and Technology
PhD Student,School of Information Science

Soglo Richlove Samuel
Southwest University of Science & Technology, Information &Communication Engineering
MSc Student, School of Computer Science
Mianyang-Sichuan, China.

GENERATIVE ADVERSARIAL NETWORK AIDED SECURITY CHECK WITH TERAHERTZ IMAGES USING IMPROVED YOLOv3

Abstract: *Terahertz imaging technology has the advantages of rapid imaging, strong penetration, harmless to the human body hence widely used in a variety of security environments and has become an alternative technology for X-ray imaging. In this paper, a GAN network-assisted deep learning method is proposed to detect whether there are illegal objects in terahertz images. Most importantly, our GAN model is used to improve the low resolutions of*

terahertz image and video. First, the GAN network reconstructs the blurred terahertz image, and then we use the optimized YOLOv3 object detection network to detect. The experimental results show that with the aid of GAN to blur image reconstruction, the accuracy of object detection is improved by 7.49%. On the YOLOv3 detection network, we added additional YOLO heads, which help to improve the ability of the network to detect objects of different sizes. Compared with the original YOLOv3 model, the improved model improves the detection performance by 10.96%. On the quantitative enhancement indexes (PSNR and SSIM), we attain a significant percentage increase.

***Key words**: Terahertz image, Terahertz technology, object detection, GAN model, optimized YOLOv3.*

***Language**: English*

***Citation**: Danso, S. A., et al. (2021). Generative Adversarial Network Aided Security Check with Terahertz Images Using Improved YOLOv3. *ISJ Theoretical & Applied Science, 03 (95),* 301-309.*

***Soi**: http://s-o-i.org/1.1/TAS-03-95-49 **Doi**: https://dx.doi.org/10.15863/TAS.2021.03.95.49*

***Scopus ASCC**: 1700.*

Introduction

The terahertz region is also called T-rays. It ranges from frequencies of about 0.1 to 10 THz. The area compasses between the electronics field and the photonics field in the electromagnetic spectrum band. Terahertz has a wider variety of dielectric materials penetration such as fabric, paper, plastic, leather, and wood. It is non-ionizing and has minimal effects on the human body [1]. It has a very large absorption rate in water as well as high radiation reflection in metals. Nevertheless, Terahertz images have low resolutions because of relatively low power, consequently making the images deblurred and denoised.

Currently, it is applied in the field of Medicine, pharmaceutical, security checkpoints, industrial & currency detection, postal mails, broadband communications among others. Deep learning models have great effects on x-rays images as well the terahertz images.

Deep learning was the first to shine in the image field and made some great break-through in image classification, target or localization of detection, semantic segmentation, and many more [2].

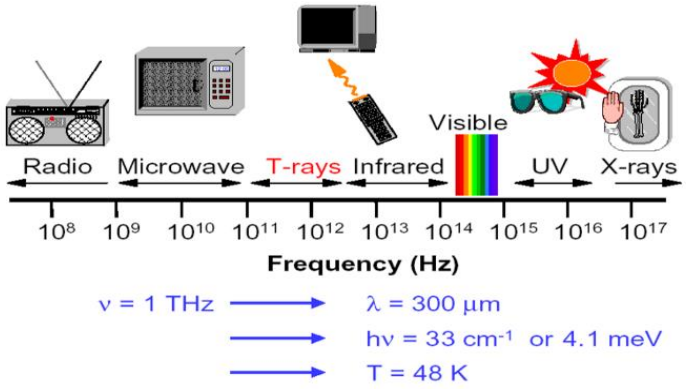

Figure 1 - EM frequency bands

This research belongs to the task of target detection, which is used to detect which objects are contained in terahertz images. This kind of image task includes the location and classification of objects. Classic algorithms include two-stage detetion models: Faster RCNN [3–6], Mask RCNN [7, 8] ,and one-stage detection models :SSD [9], YOLO series [10–12] Redmon et al further improved YOLO called "YOLOv3" [13], RetinaNet [14].

In order to consider both the detection accuracy and the detection speed, this study chooses YOLOv3 as the base detection model and improves our terahertz image target detection task.

Data Acquisition

Due to acquisition rate up to 5000 lines per second, teraFAST-256 device can accommodate scan speed up to 15m/s. The sensor has single sensitivity band at 100±10GHz but experimental power source is between 100GHz. The conveyer belt speed of 10.1m/s is for image capture in Figure 2

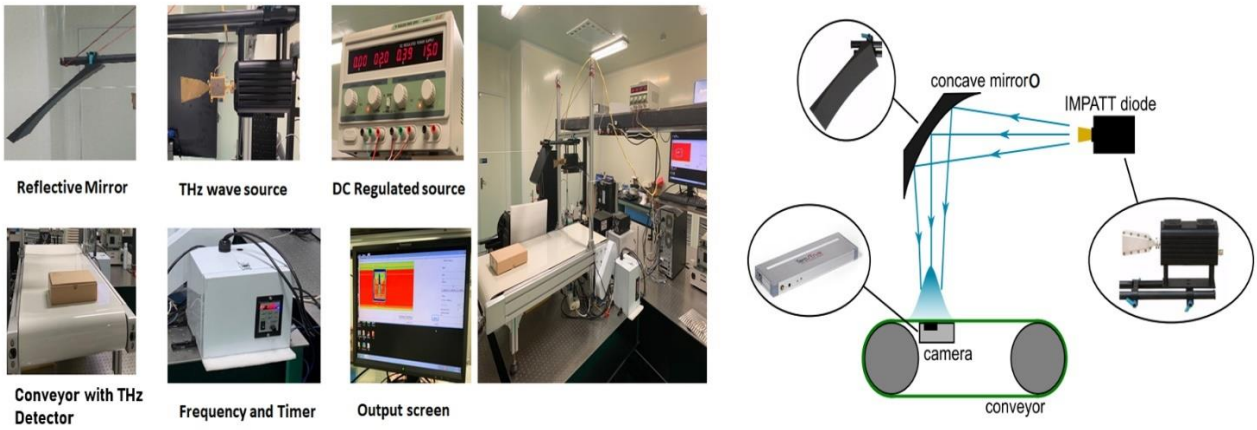

Figure 2 - Terahertz image acquisition

Data augmentation

Terahertz is a novel technology henceforth the image database is scanty. Therefore the number of terahertz image database sets that individually can collect would not be much as the regular database set for object detection model objectives to be established. Consequently, the target detection algorithm may be under-fitted in the case of so little data since terahertz image database is uncommon, and the performance of the model cannot meet the actual detection accuracy requirements. For this reason, it is necessary to expand the original image data. The methods of data argumentation used in this study are shown in figure 3.

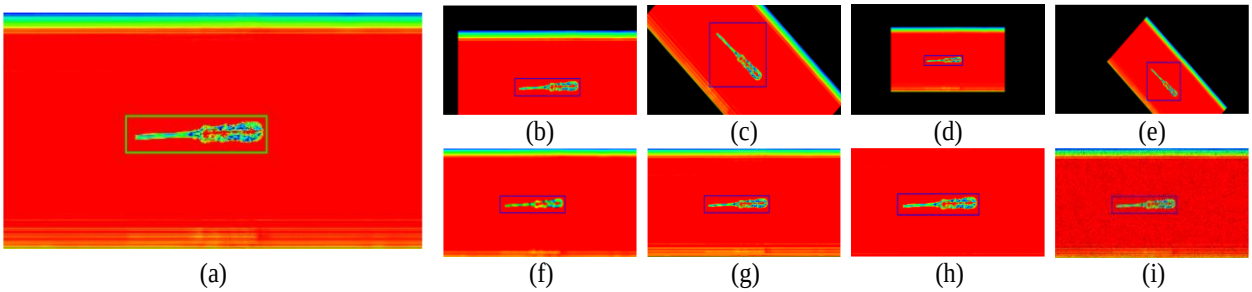

Figure 3 - Methods for augmenting terahertz images. (a) Original image. (b) Translation. (c) Rotation. (d) Scaling. (e) Affine. (f) Blurring. (g) Sharpen. (h) Cropping. (i) Dropout.

These image augmentation methods can be used alone or combined with a variety of transformations that solve the problem of little training data to a certain extent. After augmentation, we got 1884 images, and mark the location of the object. The next section will make a statistical analysis of the augmented data set.

Dataset

This paper introduces the acquisition system of terahertz images. In this study, we selected 314 clear terahertz images, first expanded to 2204 by image argumentation methods, and marked objects. A total of 8 types of objects to be precise.

Table 1. Dataset Statistics Analysis

Class	Number of instances	Average bounding box size
Screw drive	390	108px*84px
Blade	200	36px*35px
Knife	396	89px*75px
Scissors	354	104px*91px
Board marker	240	78px*68px
Mobile phone	240	110px*87px
Wireless mouse	240	70px*75px
Water bottle	240	118px*91px

Structure of YOLOv3

YOLOv3 proposed by Joseph Redmon and his team in 2018 [13] absorbs a lot of excellent target detection ideas, and its main improvements are as follows:

1. In the backbone design, the darknet53 network with residual connection is used. The darknet53 network is mainly composed of DBL (Darknet+Batch Normalization+Leaky Relu) module and Res (Residual) module. The introduction of batch normalization can enhance the generalization ability of the network and reduce the over-fitting phenomenon. The introduction of the residual module alleviates the problems of gradient disappearance and gradient explosion, which makes the model easy to converge.

2. Through the integration of different layers of feature maps, through concatenate operation, the final output of three sizes of feature maps, (13*13, 26*26, and 52*52), that is, YOLO heads, these three feature images are used for multi-scale object location and classification, which enhances the detection effect of small-size objects.

3. The pooling layer is replaced by the convolution operation with a step size mentioned in step two(2) to reduce the loss of the feature graph.

In the design of the Darknet network, the convolution operation with the step size of No.2 is used to reduce the size of feature maps, which reduces the loss of feature information caused by direct pooling and retains more feature information for the YOLO head**.**

Improvement of YOLOv3

There are only three YOLO heads in the original yolov3 model. In order to improve the ability of the model to detect objects of different sizes, we have improved it. On the basis of the original three YOLO heads, we have added two large-size YOLO heads, (104*104, 208*208). The schematic diagram of the improved model is depicted in Figure 4.

The improved YOLOv3 predicts the boxes at 5 scales as we have 5 YOLO heads. At each scale we predict 3 boxes, so the output tensor is N*N*[3 * (4 + 1 + 8)] for N by N size feature map, 4 bonding box offsets, 1 object entity, and 8 classes. In order to adapt to the structure of the 5 YOLO head, two residual modules are added in series to the original backbone foundation. The size of the feature map used by the newly added two YOLOheads is relatively large (104*104, 208*208), and contains more semantic information. In Figure 4 we can see that the later feature map was concatenated with the previous one, which makes full use of shallow and deep features of the convolution neural network.

Generative adversarial network for debluring

In order to adapt to the structure of the five-YOLO head, two residual modules are added in series to the original backbone foundation.

The size of the feature map used by the newly added two YOLO heads is relatively large, which makes use of more shallow features of the convolution neural network, and the shallow features mainly carry the spatial information of the object, so the improved design is helpful to the location of the object.

In this study, inspired by the DeblurGANv1 [15] DeblurGANv2 model [16, 17], we designed the GAN model for deblurring terahertz images, as shown in Figure5.

Generator of GAN

The function of the generator is to give a fuzzy image and get the output result of the deblurring target. Under the action of the optimizer, the network constantly updates the parameters of the generator, and finally achieves our expected effect as shown in figure 5

In order to fully mine the image information, our generator structure draws lessons from the design of DeblurGANv2, even if the connection feature pyramid network, function after using ResNet50 as the backbone, of the generators is to improve the semantic information of shallow features.

The output feature maps of each layer of the feature pyramid network are stacked first, and then the results of the same size as the input image are obtained by up-sampling, and finally, the final output (output of generator) is obtained by element-wise addition with the original image. For the generator, the output should be as close as possible to the real value.

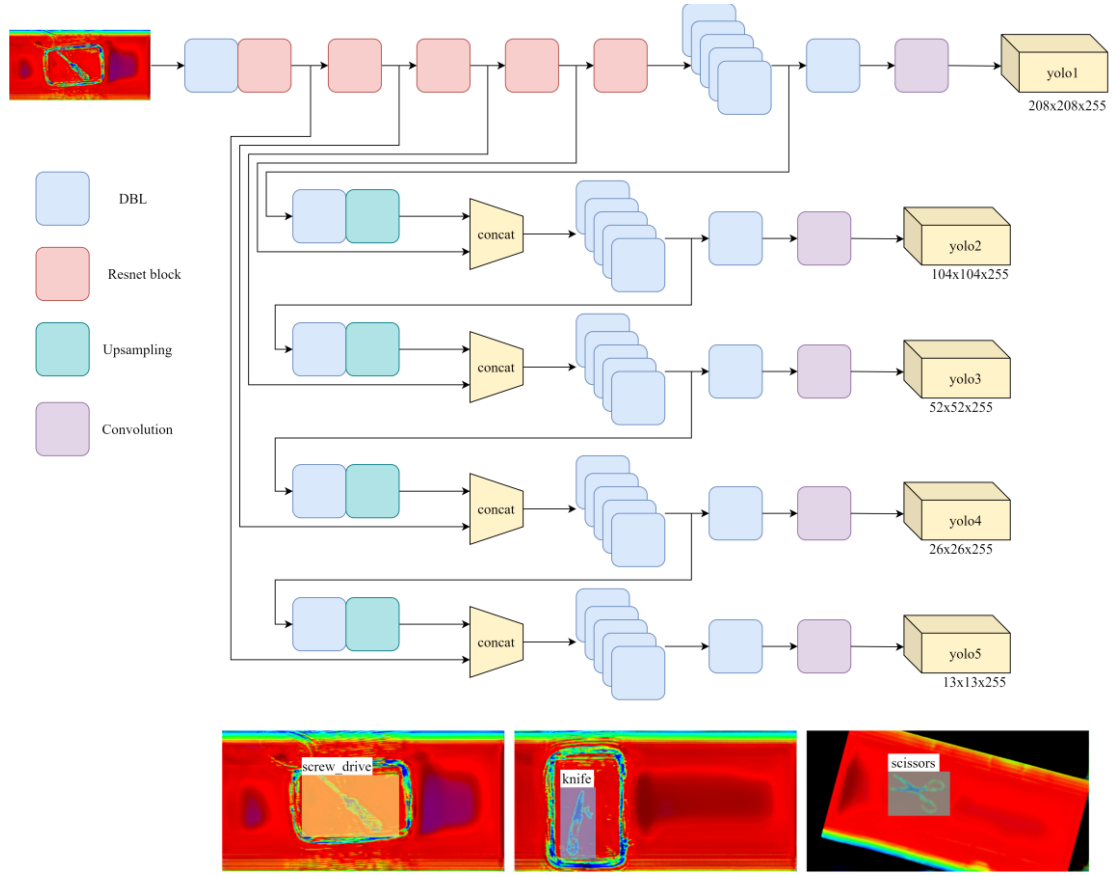

Figure 4 - Improved headsof YOLOv3

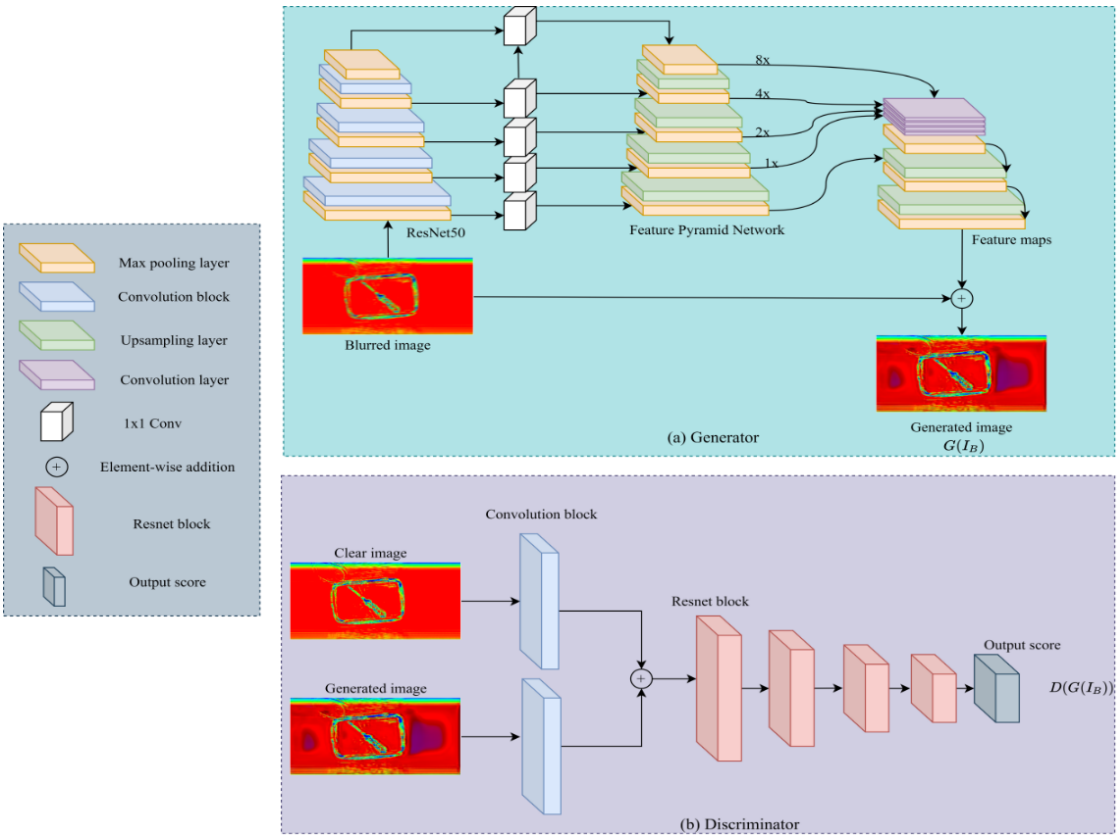

Figure 5 - Our GAN Structure for debluring

Discriminator of GAN

The function of the discriminator is to distinguish the difference between the result generated by the generator and the real input image.

In this paper, the generator needs to generate as clear an image as possible, and the discriminator is used to distinguish the difference between the reconstructed image and the real clear image.

The discriminator of this paper first convolutes the two input images, then concatenates, the feature images, then uses four residual blocks (resnet block) to extract features, and finally outputs through the full connection layer to get output score.

The optimization objective function of the whole GAN is:

$$\min_G \max_D = \mathrm{E}_{x \square p_{data}(x)}[\log D(x)] + \mathrm{E}_{I_B \square p_{I_B}(I_B)}[\log[1 - D(G(I_B))]$$

That is, the discriminator tries at best to distinguish whether the input image is real or generated by the generator, and the generator generates the same value as the real value as far as possible. In effect, the two networks play games against each other. When it is difficult for the discriminator to distinguish, it shows that the image reconstructed by the generator is close to the real clear image.

GAN-YOLOv3 COMBINED METHOD

This section mainly introduces the target detection algorithm for detecting objects contained in terahertz images and the GAN model for deblurring terahertz-blurred images as in figure 4.

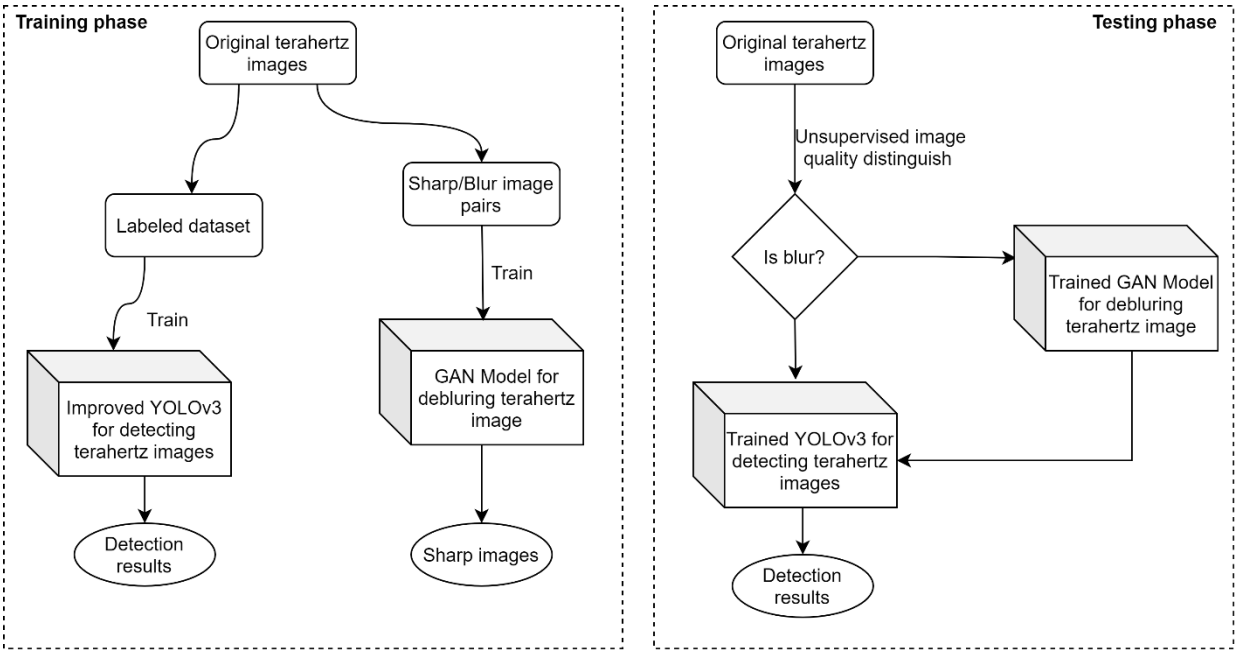

Figure 6 - OUR GAN-YOLOv3 combined method to improve detecting terahertz images.

In order to explore the GAN deblurring method and the improved YOLOv3 detection method proposed in this paper, we first artificially generate sharp-blurred image pairs. The mathematical model of blur image can be expressed as follows:

$$I_S = I_S * PSF + \mathrm{T}$$

PSF for Point Spread Function, * represents convolution operation, T for noise, for dynamic fuzzy (motion blur).

$$PSF(x_1, x_2) = \frac{1}{K}, \quad \text{if } x_1 = x_2 tan\beta, \sqrt{x_1^2 + x_2^2} \le \frac{K}{2}$$

$$PSF(x_1, x_2) = 0, \quad \text{otherwise}$$

In the above formula, K represents the length of fuzzy kernel and β is the direction of fuzzy motion. According to this method, we randomly generate images with different fuzzy kernel sizes, different blurred angle image and form sharp-blur image pairs.

In order to ensure the rationality of the algorithm comparison, we uniformly use the parameters shown in table II during the model-training phase.

Table 2. Configuration of models training

Parameter	Value
Training epochs	300
Optimizer	Adam
Learning rate	1e-4
Weight decay	1e-4
Image size	3*608*608
Batch size	4

Quantitative evaluation

From Yolo, our terahertz images are firstly denoised before enhancement to avoid amplification of image noise caused in the captured process. Image visualization quality or improvement is subjective to the human eyes perceived. Quantitatively evaluation scale of images gives its mathematical proofs. This paper seeks to use two quantitative analyze such as PSNR and SSIM [18, 19].

a) Peak Signal–To-Noise Ratio (PSNR): In most cases, the higher the PSNR, the better the visual quality of the image. Table III shows our PSNR mean values indicating that our model is superior therefore its image and video frames quality has been enhanced using the equation below

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\frac{1}{3mn} \sum_{R,G,B} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{color}(i,j) - K_{color}(i,j)]^2} \right)$$

b) Structural Similarity Measure (SSIM): High values indicate good quality of images. Our model proofs again that both image and video of the terahertz dataset are enhanced better using the equation below

$$l(\mathbf{x},\mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha}[c(\mathbf{x},\mathbf{y})]^{\beta}[s(\mathbf{x},\mathbf{y})]^{\gamma}$$

Metrics results for blur and object detection

Table 3. Image quality under different deblurring methods

Metric	PSNR	SSIM
No processing	18.9	0.56
DeblurGANv1[xx]	23.1	0.76
DeblurGANv2[xx]	25.7	0.82
Our model	**26.6**	**0.89**

As can be seen from table 3, after GAN deburring, the image quality has been greatly improved when DeblurGANv1 and DeblurGANv2 are compared with our model. Our deblurring GAN model has achieved higher performance results. Specifically, PSNR and SSIM were 15.15% and 17.10% respectively for DeblurGANv1 whereas 3.50% and 8.53% respectively were attained for DeblurGANv2. More so, for terahertz images at raw state, PSNR and SSIM were 42.32% and 53.57% respectively.

COMPARISON WITH OTHER SOTA DETECTION METHOD

In this section, we introduce the improved YOLOv3 and compare it with other states of the art detection algorithms on the terahertz test set. The results are shown in the table below. + GAN means to use the GAN model to de-blur the dataset, otherwise use the original test set.

Considering the results of each detection model without GAN processing, our improved YOLOv3 model is 10.96% higher than the original YOLOv3 on average precision, and the detection performance of

other detection algorithms is lower than that of the improved YOLOv3 model.

Furthermore, the effect of GAN de-blurring on improving detection accuracy is evident. Concretely, after using GAN to deblur the test set, the seven models used in this paper have been improved in varying degrees in the detection performance indicators, indicating that the GAN model can improve the accuracy of the object detection model after deblurring the image. Our YOLOv3 +GAN shows 7.49% higher detection performance.

Table 4. Map results of difference detection methods

Model	Screw driver	Blade	Knife	Scissors	Board marker	Mobile phone	Wireless mouse	Water bottle	AP
YOLOv3 (ours)	0.742	0.648	0.749	0.785	0.860	0.933	0.920	0.924	0.82
YOLOv3 (ours)+GAN	0.757	0.680	0.794	0.812	0.901	0.942	0.938	0.941	0.846
YOLOv3	0.659	0.539	0.642	0.695	0.78	0.899	0.832	0.87	0.739
YOLOv3+GAN	0.683	0.612	0.707	0.741	0.832	0.911	0.895	0.916	0.787
RetinaNet	0.326	0.41	0.219	0.323	0.706	0.843	0.824	0.802	0.557
RetinaNet+GAN	0.382	0.515	0.288	0.394	0.792	0.903	0.897	0.862	0.629
SSD512	0.63	0.557	0.594	0.608	0.771	0.875	0.826	0.865	0.716
SSD512+GAN	0.683	0.617	0.688	0.691	0.824	0.902	0.878	0.912	0.774
Faster RCNN	0.409	0.26	0.326	0.433	0.663	0.837	0.748	0.816	0.562
Faster RCNN+GAN	0.489	0.356	0.399	0.501	0.711	0.894	0.811	0.885	0.63

CONCLUSION

In this paper, we performed data argumentation and transformational techniques to increase the terahertz image database. In Deep learning, the greater the Dataset, the greater the improved detection method and its efficiency achieved. Furthermore, structural changes were performed on the YOLOv3 model with respect to backbone redesign, integration of different layers of feature maps, and replacement of pooling with convolutional layers. Additionally, we designed the terahertz structure of the GAN model for image deblurring based on DeblurGANv1 and DeblurGANv2 and integrated it into our Yolov3 model. A vigorous experimental work coupled with empirical analysis performed showed the superiority of our model compared with existing models.

References:

1. Zhang, J., Xing, W., Xing, M., & Sun, G. (2018). Terahertz Image Detection with the Improved Faster Region-Based Convolutional Neural Network. *Sensors*. https://doi.org/10.3390/s18072327
2. López-Tapia, S., Molina, R., & Pérez de la Blanca, N. (2018). Using machine learning to detect and localize concealed objects in passive millimeter-wave images. *Engineering Applications of Artificial Intelligence*. https://doi.org/10.1016/j.engappai.2017.09.005
3. Zhang, L., Lin, L., Liang, X., & He, K. (2016). *Is Faster R-CNN Doing Well for Pedestrian Detection?* Springer International Publishing.
4. Chen, X., & Gupta, A. (2017). *An Implementation of Faster RCNN with Study for Region Sampling*. CoRR abs/1702.02138 (2017)
5. Liu, T., & Stathaki, T. (2018). Faster R-CNN for Robust Pedestrian Detection Using Semantic Segmentation Network. Front. *Neurorobot*. https://doi.org/10.3389/fnbot.2018.00064
6. Ren, S., He, K., Girshick, R., & Sun, J. (2017). *Faster R-CNN. Towards Real-Time Object Detection with Region Proposal Networks.* IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/tpami.2016.2577031
7. He, K., Gkioxari, G., Dollár, P., & Girshick, R.B. (2017). *Mask R-CNN.* CoRR abs/1703.06870

8. He, K., Gkioxari, G., Dollár, P., & Girshick, R.B. (2020). *Mask R-CNN*. IEEE Trans. Pattern Anal. Mach. Intell. (2020). https://doi.org/10.1109/TPAMI.2018.2844175
9. (2016). Liu: SSD. *Single Shot MultiBox Detector Volume* 9905 (2016).
10. Redmon, J., & Farhadi, A. (2016). YOLO9000. Better, Faster, Stronger - abs/1612.08242
11. Redmon, J., & Farhadi, A. (2017). YOLO9000. Better, Faster, Stronger, - 6525
12. Wu, X., Sun, S., Chen, N., Fu, M., & Hou, X. (2018). Real-Time Vehicle Color Recognition Based on YOLO9000, - 89
13. Redmon, J., & Farhadi, A. (2018). YOLOv3. An Incremental Improvement. CoRR abs/1804.02767
14. Lin, T., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection. (None), - 3007.
15. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., & Matas, J. (2018). *DeblurGAN. Blind Motion Deblurring Using Conditional Adversarial Networks.* In: CVPR 2018. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition : proceedings : 18-22 June 2018, Salt Lake City, Utah. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018/06. IEEE Computer Society, Los Alamitos, California (2018). https://doi.org/10.1109/cvpr.2018.00854
16. Kupyn, O., Martyniuk, T., Wu, J., & Wang, Z. (n.d.). *DeblurGAN-v2: Deblurring* (Orders-of-Magnitude) Faster and Better
17. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: *DeblurGAN-v2.* Deblurring (Orders-of-Magnitude) Faster and Better. http://arxiv.org/pdf/1908.03826v1
18. Mittal, A., Moorthy, A.K., & Bovik, A.C. (2012). *No-Reference Image Quality Assessment in the Spatial Domain.* IEEE Transactions on Image Processing (2012). https://doi.org/10.1109/tip.2012.2214050
19. Horé, A., & Ziou, D. (2010). Ima*ge quality metrics. PSNR vs. SSIM.* 20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010 (2010). https://doi.org/10.1109/ICPR.2010.579