

TERAHERTZ IMAGE RESTORATION BENCHMARKING DATASET

Yixiong Zhang¹ Zhipeng Su¹ Feng Qi² Jianyang Zhou³ Xiao-Ping Zhang⁴

¹School of Informatics Xiamen University, Xiamen 361005, China

²Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

³School of Electronic Science and Engineering, Xiamen University, Xiamen, China

⁴Department of Electrical, Computer and Biomedical Engineering, Ryerson University, Canada

ABSTRACT

The paper introduces a new terahertz (THz) image benchmarking dataset for THz imaging. The degradation of THz image quality is one of the main problems caused by system noise, intrinsic long-wavelength, and diffraction phenomena. In this paper, the point spread function of the THz imaging process is reconstructed firstly. The THz datasets with ground-truth and degraded images are then synthesized using the point spread function (PSF). We propose a Dense Instantiation Normalization Block (DIN Block) to reconstruct clean THz images. Based on the DIN block, a powerful multi-stage network is designed, named as DINet. DINet achieves the state-of-the-art (SOTA) restoration performance on image rain removal datasets and the proposed THz datasets. To the best of our knowledge, the THz image benchmarking dataset is the first public dataset, which is available at <https://github.com/hellogry/THzDatasets>

Index Terms— terahertz (THz) image benchmarking dataset, point spread function (PSF), terahertz imaging systems, Dense Instantiation Normalization Block (DIN Block), THz image restoration

1. INTRODUCTION

Terahertz (THz) imaging is considered as one of the most promising technologies because of its penetration characteristics, ultra-wide spectrum and low photon energy. THz radiation is able to penetrate most substances that the visual light and infrared light cannot reach. THz imaging is widely used in military, medical, radar and industrial detection applications. However, the light diffraction and system noise in the imaging system causes degradation of THz images, which is

Corresponding author: Yixiong Zhang and Zhipeng Su. This work was supported in part by the Science and Technology Key Project of Fujian Province (2019H6001, 2019HZ020009, 2020HZ020005, 2021HZ021004 and 2021H61010115), the National Natural Science Foundation of China (Grant No.U1705263), the President's Fund of Xiamen University for Undergraduate (No. 20720212006), and the Open Project of Key Laboratory of Wireless Sensor Network & Communication, Shanghai Institute of Micosystem and Information Technology, CAS.

still a pressing problem to be solved. To address the problem, many algorithms have been designed to solve the task of THz image recovery. The PSF diffraction degradation is reduced by an iterative Bayesian-based method that uses deconvolution to recover clean images [1]. Based on the quantized maximum entropy theory, a technique is proposed to obtain clean images and distributions by obtaining them from noisy and incomplete data [2]. With the development of convolutional neural networks in recent years, there has been a growing number of methods to use deep learning for THz image restoration [3, 4, 5]. For the terahertz images in this paper, the PSF is a terahertz mapping function that is the main factor of image blurring. This paper propose to use AI methods to solve the problem of terahertz image reconstruction.

The restoration methods of THz image can be classified into two types: optimization-based and data-based. Optimization based methods are used to recover images by extracting a priori information from the THz raw data or by modeling the optical physical properties. For instance, the differential spectrum is defined as the difference between the emitted and the received spectra of the THz signals [6]. Then the overall absorption effect is modeled as a linear combination of the differential spectra of individual substances. In [7], the point spread function (PSF) for quantitatively evaluating the out-of-focus effect is derived based on the distance deviation. Then an adaptive focusing algorithm based on the minimum-entropy method was proposed for single-frequency holography to improve the recovered images of targets at different distances.

Recently, the rise of data-based methods plays an important role in driving THz image restoration. In [4], an autoencoder is developed, the encoder network predicts suitable parameters and the decoder is fixed on a physically meaningful model function to train the encoding network in an unsupervised manner. In [3], convolutional neural network (CNN) is used for the first time to reduce THz ringing effect. The removal of the ringing effect is treated as a deblurring problem by denoising the hyperspectral cube and then using a CNN for a band deblurring, since it is difficult to recover the target just by intensity measurements.

2. METHODOLOGY

In this section we first introduce the construction of the THz imaging system and the acquisition of the real PSF. Then the THz dataset is synthesized by this function. Following, we present the implementation details and the composition of the DINet.

2.1. Setup of Terahertz Imaging System

The imaging system is shown in Fig 1, which consists of a THz transmitter and a pair of lenses. A vector network analyzer (VNA) is used to record S11s. However, the phase information is ignored and only amplitude values are recorded. Stepper motors are moved along the x, y and z axes in a space of 70mm × 70mm × 14mm. The lens in Fig 1 is an independently developed lens in the THz lab that can focus a 100 GHz spot to 0.56 wavelengths. The THz transceiver operates at the frequency from 0.17 to 0.22 THz and uses symmetrical Teflon biconvex lenses (for the pupil side) in order to focus the THz beam. Both lenses are 100 mm in diameter and have a focal length of 25 mm (left) and a focal length of 50 mm (right). By moving the stepper motor, the amplitude is received from emitted waves.

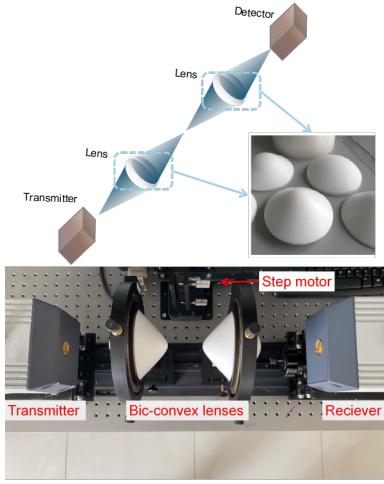


Fig. 1. The picture above shows the principle diagram of THz imaging, and the picture below is the physical imaging system.

Based on the proposed THz imaging system, the PSF is obtained and the center part with resolution of 15*15 pixels is used to synthesize the dataset. As shown in Fig 2, the spot in the center region is enlarged to show more information.

2.2. Terahertz Image Dataset

The imaging system is assumed to be a linear observed system. The synthesized THz image O consists of the convolution result between the PSF and the clean image I , and the

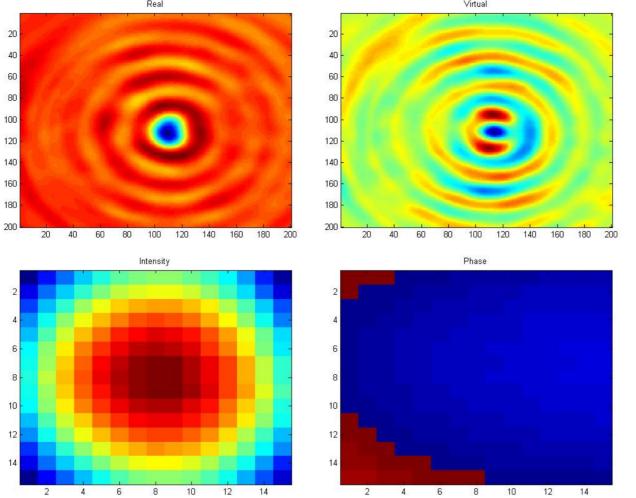


Fig. 2. Visualization of the acquired PSF. The first row shows the real and imaginary parts of the PSF with resolution of 200*200. The second row shows the intensity and phase of the central part extracted from the first row, with resolution of 15*15.

additive noise N generated by the observation system. Thus the image synthesis process can be expressed as follows:

$$O = I * |\mathbf{P}| + N, \quad (1)$$

where $|\cdot|$ is noted as the magnitude of the array elements and $*$ is noted as the convolution. \mathbf{P} is the point spread function. And \mathbf{P} can be expressed as follows:

$$\mathbf{P} = a + bi = Ae^{j2f\pi}, \quad (2)$$

Two THz datasets are constructed. One contains 1800 training images and 200 test images [8]. Another contains 900 training images and 100 test images. Both of the datasets include the ground-truth label I and output O in pairs. Fig 5 shows the synthesized THz images.

2.3. Overall Network Architecture

We design a multi-scale fusion network based on Unet, which is composed of the proposed DIN Block for the up and down sampling. The proposed DINet consists of three sub-networks. The fusion idea of [9] is employed in each sub-network. The first sub-network is used to obtain the coarse features. The other two sub-networks uses DIN Block for fine features extraction. The SAM module [10] is used to suppress the useless information and help pass the useful information to the next sub-network. The entire network structure is shown in Fig 3. In the two Unet-based refinement sub-networks, we first initialize the features using a 1×1 convolutional kernel. Then the DIN block classifies the features into four sub-tensors, each containing 16 channels of information. As shown in

Fig 4, for one of these tensors the half instance normalization [11] is used for information extraction, and the other two are densely connected for feature extraction. The densely connected convolutional layers use a 3×3 convolutional kernel to keep the spatial resolution constant. Finally, the split tensors are merged with the input image.

In each sub-network the upsampling stages are bridged with multi-scale inputs after downsampling [12]. First, for each layer of the downsampled output, it is resized and then concatted for each layer of the corresponding upsampling. Then the four multiscale features are concatenated and the channels are corrected by a 3×3 convolution kernel. Finally, a convolutional kernel of 2×2 with a step size of 2 is used for the upsampling extraction by a deconvolutional layer.

$$\begin{aligned} \text{Dec}_1 &= \{\text{Enc}_1, \text{Enc}_2 \downarrow, \text{Enc}_3 \downarrow, \text{Enc}_4 \downarrow\} \\ \text{Dec}_2 &= \{\text{Enc}_1 \uparrow, \text{Enc}_2, \text{Enc}_3 \downarrow, \text{Enc}_4 \downarrow\} \\ \text{Dec}_3 &= \{\text{Enc}_1 \uparrow, \text{Enc}_2 \uparrow, \text{Enc}_3, \text{Enc}_4 \downarrow\} \\ \text{Dec}_4 &= \{\text{Enc}_1 \uparrow, \text{Enc}_2 \uparrow, \text{Enc}_3 \uparrow, \text{Enc}_4\} \end{aligned} \quad (3)$$

where \uparrow is denoted as upsampling, \downarrow is denoted as downsampling, and $\{\cdot\}$ is the concat connection. Enc_n and Dec_n are noted as the encoding and decoding tensor of the nth layer.

In addition, the DesFus' fusion module is used between different sub-networks. As shown in Fig 4, a 3×3 convolution kernel is used to concatenate the output with consistent resolution. Then, the channel attentions generated by maximum pooling and average pooling are element-wise multiplied with the concatenated output, respectively. Finally, the fusion result is summed up with the corresponding position of the next sub-network. The DesFus module combines residual connectivity with channel attention to fuse different sub-network features as a way to enhance the extraction of multi-scale features.

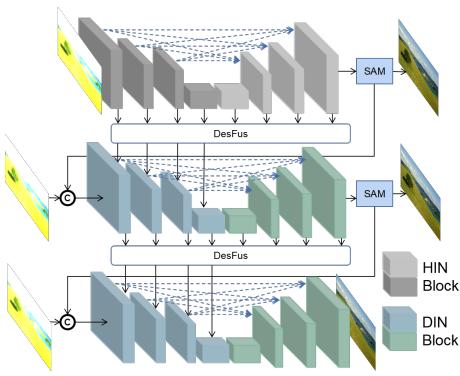


Fig. 3. The overall architecture of our Dense Instantiation Normalization Network for THz image reconstruction.

With the synthetic THz images, the proposed DINet is used for supervised training. A loss function based on the combination of PSNR is adopted. The input dimension is denoted as X , the output dimension of each layer is denoted as

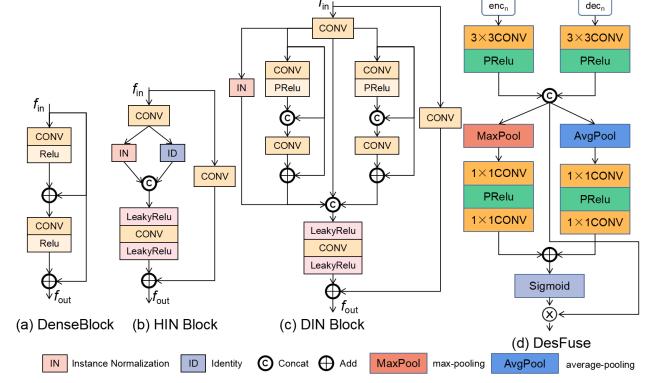


Fig. 4. The DIN Block incorporates the dense connection and half instance properties of (a) and (b). The proposed DIN block and DesFus structure are shown in (c) and (d).

R, and the THz-labeled image is denoted as Y. Then the total loss function is defined as follows:

$$L = - \sum_{i=1}^3 PSNR((R_i + X_i), Y). \quad (4)$$

3. EXPERIMENTS

The proposed DINet model is trained using a 2080Ti GPU. The experiments utilize a runtime environment built with pytorch 1.8.1, python 3.6.9, cuda 10.2 and cudnn 7.6.5. Adam is used to adjust the learning rate from $2e-4$ to $1e-7$ and the iteration number is set as 500000 for training. In addition, the recovery metrics PSNR and SSIM of the images are reported quantitatively for some other existing models. In order to demonstrate the generalization of the proposed model, the proposed DINet is also validated on the deraining task. The results show that the proposed DINet achieves the state-of-the-art (SOTA) metrics under different rain map datasets.

The performances of the proposed DINet on the constructed THz dataset are compared with some existing methods including Unet [13], PReNet [14], MPRNet [10] and HINet [11]. Moreover, the ablation study is carried out on the design of the DINet.

Table 1 shows the quantitative comparison on PSNR and SSIM of the proposed DINet and other methods. Compared to HINet, the PSNR of the proposed DINet is improved by 1.3% and 1.5% for the THz1800 and THz 900 datasets, respectively. Compared to PReNet [14] without multi-scale information, the proposed DINet can better recovered the images by extracting the features at different resolutions. Fig 6 shows the recovered images by different methods. The top two rows show the results of the THz1800 dataset, and the bottom row shows the results of the THz900 dataset. Although the synthesized THz images are degraded by the PSF, the proposed DINet is still able to recover the background details well.

Table 1. Quantitative comparison with PSNR and SSIM of the proposed DINet and other methods.

Model	THz1800		THz900	
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Unet[13]	19.80	0.648	18.369	0.562
PReNet[14]	20.05	0.646	18.477	0.575
MPRNet[10]	20.67	0.679	19.359	0.595
HINet[11]	22.28	0.719	20.084	0.582
DINet	22.56	0.727	20.385	0.588

Table 2. Quantitative comparison with PSNR and SSIM among the different DIN connection.

Connection	SLC	AHLC	ALLC	DHLC	FCC
PSNR/SSIM	19.77/0.646	21.58/0.717	21.94/0.717	22.31/0.723	22.56/0.727

The ablation study on the Unet and the proposed DINet with different number of layers are shown in Table 4. The subscript n in U_{DIN_n} and U_{net_n} denotes the number of layers in the Unet or the proposed DINet network. The structures with different DIN connection are shown in [12], and the results are compared in Table 2.

To verify the generalization performance of the DINet model. Table 3 shows the proposed DINet method on the rain removal dataset. The average time costs are obtained using the testing of the images with size of 1000×1000 . Compared to GCN [19], the proposed DINet improves the PSNR by 1.73%, 1.0%, 2.2% and 3.3% on the four datasets respectively. From Table 3, it is seen that the proposed DINet achieves the best performance among the compared methods, with tolerable resource consumption. Fig 7 shows the recovery result figures of Rain200H on different methods. At the skin area of the zebra, the proposed DINet is able to distinguish well between the white skin and the white rain patterns.



Fig. 5. The THz image synthesized using the PSF. The background image suffers from severe degradation due to the THz optical processing.

4. CONCLUSIONS

The THz images are generally degraded due to the blurring effect of the PSF. It is necessary to restore the THz images in

Table 3. Quantitative comparison with PSNR and SSIM among the SOTA methods and the proposed DINet on Rain200H, Rain200L, DID-Data and DDN-Data datasets.

Model	Time (s) CPU/GPU	Rain200H		Rain200L		DID-Data		DDN-Data	
		PSNR/SSIM							
DDN[15]	3.21/0.34	26.05	0.806	34.68	0.967	30.97	0.912	30.00	0.904
DID-MDN[16]	77.24/0.77	26.61	0.824	35.40	0.962	31.30	0.921	31.49	0.915
RESCAN[17]	70.43/1.53	26.75	0.835	36.09	0.970	33.38	0.942	31.94	0.935
JORDER-E[8]	207.0/1.74	29.35	0.891	37.25	0.975	33.98	0.950	32.01	0.932
PReNet[14]	95.66/0.69	29.04	0.899	37.80	0.981	33.17	0.948	32.60	0.946
RCDNet[18]	34.55/0.57	30.24	0.905	39.17	0.989	34.08	0.953	33.04	0.947
GCN[19]	18.11/0.31	31.15	0.913	40.73	0.989	34.37	0.962	33.01	0.949
DINet	20.31/0.48	31.69	0.922	41.14	0.989	35.11	0.941	34.11	0.942

Table 4. Quantitative comparison with PSNR and SSIM among the number of layers.

Model	<i>Unet</i> ₁	<i>Unet</i> ₂	<i>Unet</i> ₃
	PSNR/SSIM	19.80/0.648	19.83/0.648
Model	<i>U</i> _{DIN₁}	<i>U</i> _{DIN₂}	<i>U</i> _{DIN₃}
PSNR/SSIM	21.97/0.683	22.56/0.727	22.54/0.727

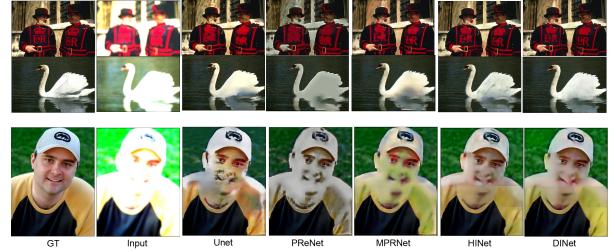


Fig. 6. Visual comparisons on different methods. The THz1800 and THz900 have the same distribution. Please zoom in for better visualization.

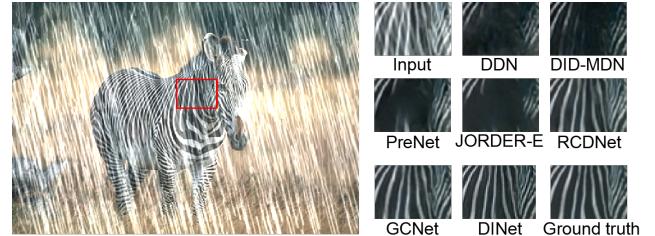


Fig. 7. One visual comparison on Rain200H data sets. Please zoom in for better visualization.

industrial and medical applications. This work presents the construction of a THz imaging system and the corresponding THz datasets. The THz images are generated by the PSF acquired by a THz generator, then recovered by using DINet. Compared with the existing models, DINet achieves the SOTA metrics on our datasets. To demonstrate the generalization of our model, DINet is also validated on the deraining task and achieves the SOTA metrics. The drawback of the system is that it fails to make use of the a priori knowledge of the physical properties of terahertz PSF for image restoration. We will continue to research in this direction.

5. REFERENCES

- [1] William Hadley Richardson, “Bayesian-based iterative method of image restoration,” *JoSA*, vol. 62, no. 1, pp. 55–59, 1972.
- [2] John Skilling, “Quantified maximum entropy,” in *Maximum Entropy and Bayesian Methods*, pp. 341–350. Springer, 1990.
- [3] Marina Ljubenovic, Shabab Bazrafkan, Jan De Beenhouwer, and Jan Sijbers, “Cnn-based deblurring of terahertz images.,” in *VISIGRAPP (4: VISAPP)*, 2020, pp. 323–330.
- [4] Tak Ming Wong, Matthias Kahl, Peter Haring-Bolívar, Andreas Kolb, and Michael Möller, “Training auto-encoder-based optimizers for terahertz image reconstruction,” in *German Conference on Pattern Recognition*. Springer, 2019, pp. 93–106.
- [5] Jinsong Zhang, Wenjie Xing, Mengdao Xing, and Guangcai Sun, “Terahertz image detection with the improved faster region-based convolutional neural network,” *Sensors*, vol. 18, no. 7, pp. 2327, 2018.
- [6] Yang Bai and Hairong Qi, “A new perspective on terahertz image reconstruction based on linear spectral unmixing,” in *Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA*. IEEE, 2008, pp. 2996–2999.
- [7] Sun Zhaoyang, Chao Li, Xiang Gao, and Fang Guangyou, “Minimum-entropy-based adaptive focusing algorithm for image reconstruction of terahertz single-frequency holography with improved depth of focus,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, 01 2015.
- [8] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan, “Deep joint rain detection and removal from a single image,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1357–1366.
- [9] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko, “Rethinking coarse-to-fine approach in single image deblurring,” 2021.
- [10] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, “Multi-stage progressive image restoration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14821–14831.
- [11] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen, “Hinet: Half instance normalization network for image restoration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 182–192.
- [12] Tingting Liang, Xiaojie Chu, Yudong Liu, Yongtao Wang, Zhi Tang, Wei Chu, Jingdong Chen, and Haibin Ling, “Cbnnetv2: A composite backbone network architecture for object detection,” 2021.
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [14] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng, “Progressive image deraining networks: A better and simpler baseline,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
- [15] Xueyang Fu, Jiaxin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley, “Removing rain from single images via a deep detail network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3855–3863.
- [16] He Zhang and Vishal M Patel, “Density-aware single image de-raining using a multi-stream dense network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 695–704.
- [17] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha, “Recurrent squeeze-and-excitation context aggregation net for single image deraining,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 254–269.
- [18] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng, “A model-driven deep neural network for single image rain removal,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [19] Xueyang Fu, Qi Qi, Zheng-Jun Zha, Yurui Zhu, and Xinghao Ding, “Rain streak removal via dual graph convolutional network,” in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 1–9.