Check for updates

Received: 7 March 2024 | Revised: 2 June 2024 | Accepted: 27 June 2024

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12380

ORIGINAL RESEARCH

Terahertz image denoising via multiscale hybrid-convolution residual network

Heng Wu[1,2] | Zijie Guo[1,2] | Chunhua He[1,2] | Shaojuan Luo[3] | Bofang Song[4]

[1]Guangdong Provincial Key Laboratory of Cyber-Physical System, School of Automation, Guangdong University of Technology, Guangzhou, China

[2]School of Computer, Guangdong University of Technology, Guangzhou, China

[3]School of Chemical Engineering and Light Industry, Guangdong University of Technology, Guangzhou, China

[4]Wyant College of Optical Sciences, The University of Arizona, Tucson, Arizona, USA

Correspondence

Chunhua He and Shaojuan Luo.
Email: hechunhua@pku.edu.cn and kesjluo@gdut.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 62173098, 62104047; Guangdong Provincial Key Laboratory of Cyber-Physical System, Grant/Award Number: 2020B1212060069

Abstract

Terahertz imaging technology has great potential applications in areas, such as remote sensing, navigation, security checks, and so on. However, terahertz images usually have the problems of heavy noises and low resolution. Previous terahertz image denoising methods are mainly based on traditional image processing methods, which have limited denoising effects on the terahertz noise. Existing deep learning-based image denoising methods are mostly used in natural images and easily cause a large amount of detail loss when denoising terahertz images. Here, a residual-learning-based multiscale hybrid-convolution residual network (MHRNet) is proposed for terahertz image denoising, which can remove noises while preserving detail features in terahertz images. Specifically, a multiscale hybrid-convolution residual block (MHRB) is designed to extract rich detail features and local prediction residual noise from terahertz images. Specifically, MHRB is a residual structure composed of a multiscale dilated convolution block, a bottleneck layer, and a multiscale convolution block. MHRNet uses the MHRB and global residual learning to achieve terahertz image denoising. Ablation studies are performed to validate the effectiveness of MHRB. A series of experiments are conducted on the public terahertz image datasets. The experimental results demonstrate that MHRNet has an excellent denoising effect on synthetic and real noisy terahertz images. Compared with existing methods, MHRNet achieves comprehensive competitive results.

KEYWORDS

image processing, multiscale hybrid-convolution, residual learning, terahertz image denoising

1 | INTRODUCTION

Terahertz (THz) imaging technology uses a THz radiation source to irradiate an object and capture the transmitted or reflected rays from the object for imaging [1]. Due to the frequency and wavelength characteristics of the THz radiation, combined with the advantages of being harmless to the human body [2], THz imaging technology has been widely used in biomedicine [3] and security fields [4], such as the detection of flammables and explosives, drugs, prohibited guns, and other dangerous goods [5–7]. However, due to factors such as the hardware of the THz imaging system and external environmental interference, the quality of THz images is poor [8, 9], and many problems exist in THz images, such as serious noise interference, low signal-to-noise ratio and contrast, and image blur [10, 11]. These problems have a significant impact on the accuracy of security detection and have become the main obstacle to the application of THz imaging technology. Therefore, the THz image denoising is of great significance and has received extensive attention in recent years [12–14].

Generally, the image denoising methods can be divided into two categories, the traditional and deep learning denoising methods. Traditional image-denoising methods mainly remove image noises directly in the spatial and transform domains. For instance, in the spatial domain, Buades et al. [15] proposed the Non-Local Means method to replace the noise area with a similar area in the image to achieve a good denoising effect. In the transform domain, the Wavelet Transform (WT) method

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Author(s). *CAAI Transactions on Intelligence Technology* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology and Chongqing University of Technology.

[16] homogenises the noise signal in the frequency domain and converts it to the spatial domain to realise image denoising. Better than [15, 16], the Block-Matching and 3D filtering (BM3D) method [17] combines the spatial and transform domains to achieve denoising. The above methods usually have many problems, such as the high model complexity, poor denoising effect for complex noises, and long image processing time.

Recently, deep learning has been widely used in the image-denoising area due to its powerful ability to learn and fit noise distributions. For example, Zhang et al. [18] proposed a denoising convolutional neural network (CNN) (DnCNN) to eliminate image noises by using residual learning. Subsequently, Zhang et al. improved DnCNN and proposed FFDNet [19], which denoises different levels of noise by adding an adjustable noise level map. Gurrola-Ramos et al. [20] developed a residual dense neural network (RDUNet) to predict residual noises using local and global residual learning. Fan et al. [21] presented a blind real image denoising network (SRMNet) by improving M-Net and using a selective residual block for denoising. Zhang et al. [22] designed a deep neural network (DRNet) with DC-ResBlock to extract richer details by introducing dilated convolutions in standard residual blocks. Tian et al. [23] proposed an attention-guided CNN (ADNet) for processing blind noise images. Huang et al. [24] developed the Neighbour2Neighbour method to generate training image pairs by using a random neighbour sub-sampler, which can realise self-supervised denoising. Lequyer et al. [25] developed a Noise2fast method, which uses a small dataset to train a network for the large image denoising. With the development of deep learning, transformer-based models have shown great application potential in zero-shot learning [26], image super-resolution [27] and multivariate time series analysis [28]. However, for THz image denoising, due to the lack of high-quality image data and the difficulty of extracting noise features in THz images, it is difficult to apply these technologies to achieve THz image denoising. Currently, deep learning-based image-denoising methods are mostly used in the field of visible light image denoising. Natural images benefit from the advancement of imaging equipment, and the noise in the image is fine-grained. Different from natural images, the noise in THz images has the characteristics of high complexity and high noise intensity. Moreover, the real THz noise is often irregularly distributed in the target and its edge area. The biggest challenge of THz denoising is to remove the noise from the target and its edges while retaining the original details. Existing methods usually cannot distinguish detailed features from terahertz noise well, and noise reduction often results in a large loss of detailed features. Therefore, developing a model that can remove real THz noise and retain detailed features as much as possible is essential for the applications of THz imaging technology.

To solve the problem of removing noises while preserving target detail features, we propose a multiscale hybrid-convolution residual network (MHRNet) with residual learning for THz image denoising. The multiscale hybrid convolution and residual learning are applied to reuse feature maps and perform local and global residual learning. Specifically, we design a multiscale hybrid-convolution residual block (MHRB) to achieve the extraction of detail features and local prediction residual noises from THz images. In MHRB, we construct a multiscale dilated convolutional block (MDCB) to increase the relative overall receptive field. As the dilated convolution usually causes the continuous loss of feature information due to the discontinuity of the convolution kernel, we develop a multiscale convolution block (MCB) to retain more image details. Many experiments are implemented to verify the denoising performance of MHRNet on THz images.

Although there have been some works using U-Net, residual learning, and dilated convolution for image denoising [29–31], MHRNet has lots of advantages compared to them. Firstly, in ref. [29], Wang et al. proposed incorporating dilated convolutions with different dilation rates into a multi-branch module. The network structure is formed by stacking multiple branches of dilated convolution modules. Compared to ref. [29], the proposed MHRNet also employs a multi-branch structure. However, unlike [29], where the dilated convolution kernel sizes are the same across different branches with distinct dilation rates, we utilise different dilated convolution kernel sizes combined with the same dilation rates (MDCB). Additionally, we integrate an MCB to form a residual structure, addressing the issue of discontinuous feature information caused by the non-uniformity of dilated convolution kernels. Secondly, in ref. [30], Wang et al. introduced a simple dilated convolution module for residual learning of noise. The dilated convolution module consists of a single dilated convolution with a dilation rate of 2, along with the Batch Normalisation (BN) layer and Rectified Linear Unit (ReLU). The network architecture is constructed by stacking these dilated convolution modules to create a single-branch model. Compared to ref. [30], we similarly use dilated convolution for residual learning of noise. However, the proposed MHRNet is structured as a U-shaped network with MHRB to enhance the representative capacity of the model. Thirdly, in ref. [31], Liu et al. presented a hybrid dilated convolution residual module and multi-scale convolution. The hybrid dilated convolution involves combining three 3×3 dilated convolutions with different dilation rates to form a residual structure. The multi-scale convolution module is applied only in the first layer of the model, and its network architecture is similar to that of DnCNN. Compared to ref. [31], the proposed MHRNet also incorporates dilated convolutions with residual structure. However, in contrast to ref. [31], MHRNet integrates MDCB and MCB in the residual structure. The MHRNet architecture follows a U-shaped structure, facilitating a better fusion of shallow and deep features.

The main contributions of this paper are described as follows.

1) We propose a MHRNet for THz image denoising based on residual learning. MHRNet is a U-shaped structure that

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

combines multiscale hybrid-convolution and residual structure. Compared with the existing denoising methods, MHRNet can remove more complex noises and retain more detailed features through global and local residual learning in the condition of low model complexity and small samples. MHRNet is helpful for applying THz imaging technology in the area of detecting hidden dangerous goods.

2) We propose a MHRB, which is a crucial part of the MHRNet. MHRB combines multi-scale dilation convolution, multi-scale convolution and residual structure to integrate the advantages of each module. The MHRB helps to learn richer detail features to improve the quality of the final image denoising results.
3) We design an MDCB, which is a multi-branch structure. In MDCB, the same expansion rate and different convolution kernel sizes are used in combination to increase the receptive field during the image segmentation process, thereby improving the noise segmenting ability of MHRNet.
4) We develop an MCB, which is also a multi-scale structure. We use multi-scale convolution in MCB to extract more feature to make up for the discontinuity of the feature information caused by the dilated convolution. Moreover, MCB can improve the detail feature extraction ability of MHRNet.

2 | RELATED WORK

2.1 | Terahertz image denoising

In the past decades, many denoising schemes have been proposed for THz images [32–34]. Zhu et al. [35] introduced a passive THz image denoising algorithm, which applies the threshold concept to the mean filtering algorithm to remove noises in THz images. Zhang et al. [36] developed a THz image inpainting method that uses wavelet denoising technology to process THz time-frequency signals and restores image details through the point spread function. Ning et al. [37] proposed a THz image enhancement method. This method combines classical deconvolution and complex-valued data processing to restore THz images. Xin et al. [38] presented a method based on adaptive template matching filtering. This method can eliminate the background noise of the image and preserve the image details of the subject. Recently, some scholars have tried to reduce noise in THz images using deep learning. For instance, Guo et al. [39] proposed a deep reinforcement learning-based pixel-wise cross-modal approach to handle noise in THz images. Li et al. [40] applied deep learning to the THz imaging area and proposed a symmetric deep CNN to eliminate noises in the THz images. The above-mentioned methods based on traditional image processing often have poor results when the real THz noise distribution is more complex. Although recent deep learning-based methods can better handle THz noise, the detailed feature loss problem still exists.

2.2 | U-shaped structure

The U-shaped structure can plays an important role in the field of image processing. It essentially reduces the computational complexity of the algorithm and can effectively use memory to deepen the number of neural network layers. The U-shaped structure is usually an encoder-decoder structure. The encoder part first reduces the feature map size, and then the decoder part restores the feature map size in the decoder part. In addition, the model fuses shallow and deep features through feature splicing, which enables the U-shaped structure to can reduce feature loss while adding more learnable parameters. Many U-shaped structure-based networks have been reported over the past few years. Ronneberger et al. [41] combined the U-shaped structure with the CNN and proposed the U-Net model. Liu et al. [42] developed a multi-level wavelet CNN (MWCNN) model based on U-shaped, which introduced the WT and convolutional layers in the U-shaped structure to better balance the receptive field size and computational efficiency. Park et al. [43] designed the U-shaped-based model (DHDN), which replaced the simple convolution in U-Net with a convolutional block composed of dense connectivity and residual learning.

2.3 | Residual structure

The residual structure in deep neural networks can avoid network degradation and gradient explosion caused by deep networks. As a deeper model means that more features can be extracted more features, so the residual structure can stimulate more potential of the model. He et al. [44] proposed a residual learning network (ResNet) framework, which introduces residual connections to achieve identity mapping, ensuring that the network is deepened and higher accuracy is obtained. Zhang et al. [22] presented a residual structure-based network (DRNet). The core of DRNet is a residual structure (DC-ResBlock) composed of common and dilated convolutions, which achieve a better denoising effect.

3 | METHOD

3.1 | Terahertz image noise model

Generally, the mathematical model of a THz image with noises can be defined as follows:

$$\mathbf{Y} = \mathbf{X} + \mathbf{G}_{\sigma} \tag{1}$$

where $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{G}_{\sigma}$ are the noise image, clear image, and noise signal with a standard deviation of σ, respectively.

The problem of recovering a clear THz image $\mathbf{X}$ from a noisy THz image $\mathbf{Y}$ through supervised learning can be described as follows:

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\hat{\mathbf{X}} = A(\mathbf{Y}, \Theta) \tag{2}$$

where $\hat{\mathbf{X}}$, $A(\cdot)$ and Θ denote the predicted THz image, implicit function of the neural network model, and model parameters, respectively.

Since the detail features in the noisy THz images are very difficult to predict, it is reasonable to estimate the noise signal in the noisy images, which is relatively easy to predict [20]. Then, the noise signal is subtracted from the original image to obtain a clear image. So, the THz denoising model can be defined as follows:

$$\hat{\mathbf{X}} = A(\mathbf{Y}, \Theta) = \mathbf{Y} - \Gamma(\mathbf{Y}, \Theta) \tag{3}$$

where $\Gamma(\cdot)$ is the parameter map of the noise signal, $\Gamma(\mathbf{Y},\Theta) \approx \mathbf{G}_\sigma$.

3.2 | Network structure design

Inspired by U-Net and ResNet, we design a MHRNet that is used to remove noises from the THz images. Figure 1 shows the architecture of MHRNet. The input of the network is a noisy THz image, and the noise map is predicted through the feature extraction module, multi-scale hierarchical cascade structure (MHCS) and feature fusion module. The output clean THz image is obtained by subtracting the predicted noise map from the input THz image.

Given a noisy input THz image $\mathbf{I} \in R^{H\times W\times C}$, where H, W and C are respectively the height, width and channel of the input image, we first perform the feature extraction by using a 3 × 3 convolution (Conv), and this process is expressed as follows:

$$\mathbf{E}_1 = \mathrm{R}(BN(Conv_{3\times3}(\mathbf{I}))) \tag{4}$$

where $\mathbf{E}_1 \in R^{H\times W\times f}$ is the output feature map, $f = 32$ is the input feature map number. $Conv_{3\times3}(\cdot)$, $BN(\cdot)$ and $R(\cdot)$ represent a Conv operation with a kernel size of 3, Batch Normalisation (BN) and Rectified Linear Unit (ReLU), respectively.

Then, the feature map $\mathbf{E}_1$ is fed into the MHCS. The backbone of MHCS has four scales, including three transition-down/up layers, three down/up sampling layers and a middle transition layer. The shortcut layer is used to connect the transition-down layer and transition-up layer of each scale. Two MHRB modules are combined to form the transition layer in four scales, and the final output of each scale can be expressed as follows:

$$\mathbf{L}_l = \xi_{MHRB}(\xi_{MHRB}(\mathbf{E}_l)) \tag{5}$$

where $\xi_{MHRB}(\cdot)$ is a function representing the MHRB, $\mathbf{E}_l$ and $\mathbf{L}_l$ are the input and output feature maps of each scale, respectively. Here, $\mathbf{L}_l \in R^{H\times W\times N}$, $l \in \{1,2,3,4\}$ and $N = 2^{L-1} \times f$ are the lth scale and the feature map numbers, respectively. Note that the input and output feature map numbers of each scale are the same.

After the transition-down layer outputs the feature map $\mathbf{L}_l$, we adopt a Conv (kernel size 2 and stride 2) to reduce the feature map size by half for downsampling. The down-sampling output feature map $\mathbf{D}_k$ process is expressed as follows:

$$\mathbf{E}_{l+1} = \mathbf{D}_k = R\left(Conv_{2\times2,s=2}(\mathbf{L}_l)\right) \tag{6}$$

where $Conv_{2\times2,s=2}(\cdot)$ denotes the Conv operation with a kernel size of 2 and a stride of 2, followed by ReLU. k is the kth sampling operation between two scales, $k = l$, $l \in \{1,2,3\}$. The feature map $\mathbf{D}_k$ is used as the input feature map for the smaller scale.

In the upward transition, we use the transposed Conv to upsample the feature map $\mathbf{L}_l$ that is output by the transition-up

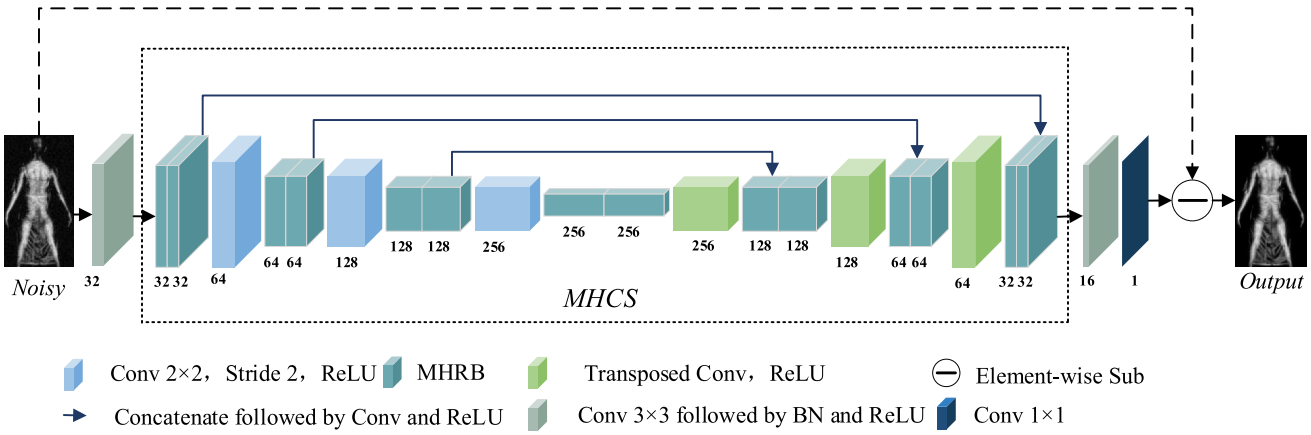

FIGURE 1 Sketch map of MHRNet for THz image denoising.

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

layer. The upsampling and feature map $\mathbf{U}_k$ outputting process is presented as follows:

$$\mathbf{E}_{l-1} = \mathbf{U}_k = R\left(TConv_{2\times2,s=2}(\mathbf{L}_l)\right) \tag{7}$$

where $TConv_{2\times2,s=2}(\cdot)$ represents the transposed Conv, followed by ReLU. Here $k = l - 1$, $l \in \{2,3,4\}$. The feature map $\mathbf{U}_k$ is used as the input feature map for the larger scale.

To reduce the loss of detail features, the numbers of feature maps $\mathbf{D}_k$ and $\mathbf{U}_k$ are respectively increased and decreased,

$$\begin{aligned} N_{D_k} &= 2N_{L_l} \\ N_{U_k} &= \frac{1}{2}N_{L_{l+1}} \end{aligned} \tag{8}$$

where N_{L_l} and N_{D_k} are respectively the output feature map numbers of the lth scale and downsampling, $N_{L_{l+1}}$ and N_{U_k} are respectively the output feature map numbers of the $(l + 1)$-th scale and corresponding upsampling. Here $l = k$, $k \in \{1,2,3\}$.

After fusing the shallow features with the deep features through the Concat operation, we employ a Conv (kernel size 3) to halve the number of feature maps and obtain the feature map $\mathbf{Z}_l$. This process can be written by the following:

$$\mathbf{Z}_l = BN(Conv_{3\times3}(Concat(\mathbf{U}_k, \mathbf{L}_l))), l = k \tag{9}$$

where $Concat(\mathbf{U}_k,\mathbf{L}_l)$ denotes the splicing of feature maps $\mathbf{U}_k$ and $\mathbf{L}_l$. Here, the Conv is followed by BN.

Therefore, given a feature map $\mathbf{E}_1$, the MHCS can output a feature map $\mathbf{T}$. The above process is expressed by the following:

$$\mathbf{T} = \psi_{MHCS}(\mathbf{E}_1) = \xi_{MHRB}(\xi_{MHRB}(\mathbf{Z}_1)) \tag{10}$$

where $\psi_{MHCS}(\cdot)$ represents the implicit function of the MHCS.

Finally, the feature map $\mathbf{T}$ is sent to the output feature module, and a 3 × 3 Conv is used for the feature fusion. Then, a 1 × 1 Conv is used to reduce the feature map numbers. This Conv outputs the residual noise through the global residual learning. After subtracting the residual noise from the input THz image $\mathbf{I}$, the denoised THz image $\mathbf{O} \in R^{H\times W\times C}$ is obtained. This process is expressed as follows:

$$\mathbf{O} = I - Conv_{1\times1}(R(BN(Conv_{3\times3}(\mathbf{T})))) \tag{11}$$

3.3 | multiscale hybrid-convolution residual block

The architecture of MHRB is shown in Figure 2, which is a residual structure and mainly consists of the MDCB, Conv layer and MCB. In the MDCB, the multi-scale dilated structure is realised by two dilated convolutions with different kernel sizes and the same dilation rate in the two branches. Similarly, in the MCB, two convolutions with different kernel sizes and padding in the two branches form a multi-scale convolution structure.

For a feature map $\mathbf{E}_l$ that inputs to the MHRB, we first halve its channel numbers through a 3 × 3 Conv to get the feature map $\mathbf{H}_1$. Then the local receptive field is enlarged, and the feature map $\mathbf{H}_{MDB}$ is generated by MDCB. There are two branches in MDCB, which use 3 × 3 and 5 × 5 dilated convolutions (Convs) to generate feature maps $\mathbf{H}_2$ and $\mathbf{H}_3$, respectively.

The expansion rate of the dilated Conv is 2, followed by BN and ReLU. At the end, the concatenation (Concat) and 3 × 3 Conv are used to fuse the feature maps. This process can be expressed as follows:

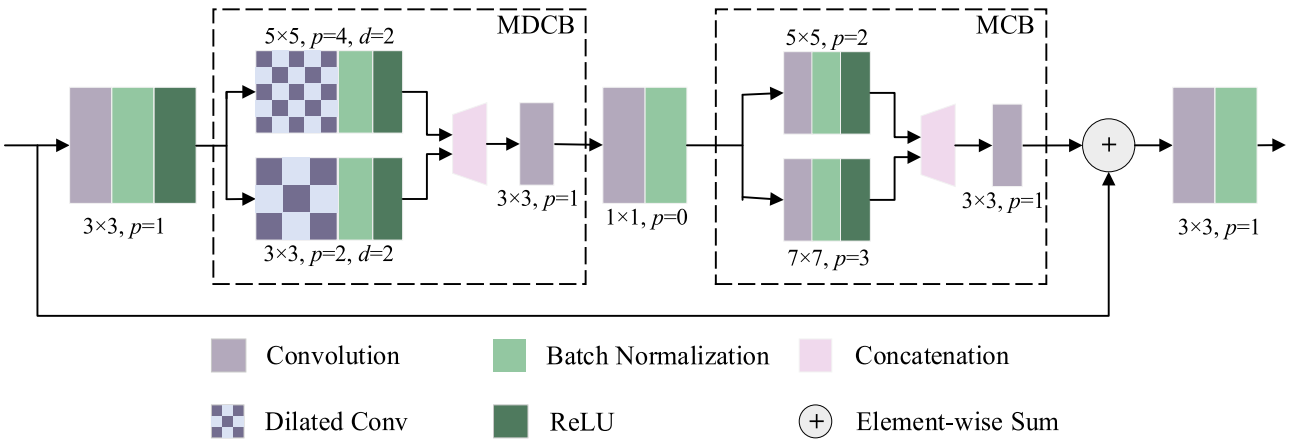

FIGURE 2 The structure of multiscale hybrid-convolution residual block (MHRB). Here, p and d are the padding size and dilation rate, respectively.

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\begin{aligned}\mathbf{H}_1 &= R(BN(Conv_{3\times3}(\mathbf{E}_l)))\\ \mathbf{H}_2 &= R\big(BN\big(Conv_{3\times3,d=2}(\mathbf{H}_1)\big)\big)\\ \mathbf{H}_3 &= R\big(BN\big(Conv_{5\times5,d=2}(\mathbf{H}_1)\big)\big)\\ \mathbf{H}_{MDB} &= Conv_{3\times3}(Concat(\mathbf{H}_2,\mathbf{H}_3))\end{aligned} \tag{12}$$

Then, the channel numbers of $\mathbf{H}_{MDB}$ is halved by using a Conv to obtain the feature map $\mathbf{H}_4$. To solve the lack of feature information caused by the discontinuous kernel of the expansion Conv, we use the MCB to further extract the feature $\mathbf{H}_{MCB}$ from the THz image. The two branches in the MCB module use 5 × 5 and 7 × 7 Convs to generate the feature maps $\mathbf{H}_5$ and $\mathbf{H}_6$, followed by BN and ReLU. Concat and 3 × 3 Conv are used for feature fusion at the end. Through the combination of MDCB and MCB, richer details can be extracted from the image. This process can be expressed as follows:

$$\begin{aligned}\mathbf{H}_4 &= BN(Conv_{1\times1}(\mathbf{H}_{MDB}))\\ \mathbf{H}_5 &= R(BN(Conv_{5\times5}(\mathbf{H}_4)))\\ \mathbf{H}_6 &= R(BN(Conv_{7\times7}(\mathbf{H}_4)))\\ \mathbf{H}_{MCB} &= Conv_{3\times3}(Concat(\mathbf{H}_5,\mathbf{H}_6))\end{aligned} \tag{13}$$

Finally, after adding the feature map $\mathbf{E}_l$ and the feature map $\mathbf{H}_{MCB}$, a Conv with a kernel size of 3 and BN is used to perform the feature fusion and output the feature map $\mathbf{H}$. The function $\xi_{MHRB}(\cdot)$ of the MHRB module can be expressed as follows:

$$\mathbf{H} = \xi_{MHRB}(\mathbf{E}_l) = BN(Conv_{3\times3}(\mathbf{H}_{MCB} \oplus \mathbf{E}_l)) \tag{14}$$

where $\oplus$ represents the element-wise addition operation.

3.4 | Loss function and training process

The L2 (MSE) loss function is often used to evaluate the difference between the true value, and the model predicted value. Thus, we adopt the L2 loss function to optimise the model parameters of MHRNet. During the training, given a set of clean-noisy THz image.pairs $\{\mathbf{I}_{hr}^n, \mathbf{I}_{lr}^n\}_{n=1}^N$, the loss function of MHRNet is defined as:

$$L(\Theta) = \frac{1}{N}\sum_{n=1}^{N}\left[\mathbf{I}_{hr}^n - A\left(\mathbf{I}_{lr}^n\right)\right]^2 \tag{15}$$

where $\mathbf{I}_{hr}^n$, $\mathbf{I}_{lr}^n$, N, $A(\cdot)$ and Θ denote the high-definition THz image, noisy THz image, number of training data sets, implicit function of the network model, and parameters of the model, respectively. The training of the network model is to find parameters Θ that make $\mathbf{I}_{lr}$ closer to $\mathbf{I}_{hr}$.

Algorithm 1. Training process

```
Input: Clean-noisy THz image pairs.
Output: Recovery image.
Initialise θ from MHRNet.
for n in range maximum epoch do.
    while b < the number of batch number do.
        Select k noisy images {I_lr^1, ..., I_lr^k};
        Select k ground truth {I_hr^1, ..., I_hr^k};
        L(Θ) = 1/N Σ_{n=1}^{N} [I_hr^n − A(I_lr^n)]^2;
        Update paraments by Adam Optimiser;
    ends
ends
```

The training process is described in Algorithm 1.

4 | EXPERIMENT ON SYNTHETIC NOISES

4.1 | Experimental dataset

The THz image dataset used in this study is from the Huaxun Ark Group, which includes 144 real noisy THz images along with the corresponding manually pre-denoised images. Collecting high-quality THz images is currently challenging, and manual denoising is time-consuming and has limited effectiveness. However, self-supervised denoising algorithms can learn denoising network models from noisy images. Thus, we implement the self-supervised Blind2Unblind [45] algorithm and use the manually pre-denoised images from the Huaxun Ark Group as the training dataset, producing 144 further denoised THz images that are used as ground truth. Note that we do not employ Blind2Unblind to denoise real THz noise directly. Instead, we optimise THz images using Blind2Unblind based on the manually pre-denoised images. The optimised images are served as pseudo-labels for the proposed model. Additionally, we conduct experiments with Blind2Unblind to denoise real THz image noise. As shown in Figure 3a and b are a real noisy THz image and a manually pre-denoised image, while (c) and (d) show the denoising results of Blind2Unblind on (b) and (a), respectively. We utilise the no-reference-image quality evaluation indicators NIQE [46] to quantify the image quality. Lower NIQE values indicate better image quality. The result in Figure 3c reflects that using the Blind2Unblind can further reduce noise in Figure 3b, improving the image quality. The result in Figure 3d indicates that directly denoising real noisy THz images using Blind2Unblind has limited effectiveness. The experimental results confirm the limitations of Blind2Unblind in denoising real THz image noise. Since manually denoising THz images is inconvenient and time-consuming, and Blind2Unblind cannot directly denoise real THz noise, thus we develop a model that can remove real THz noise.

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

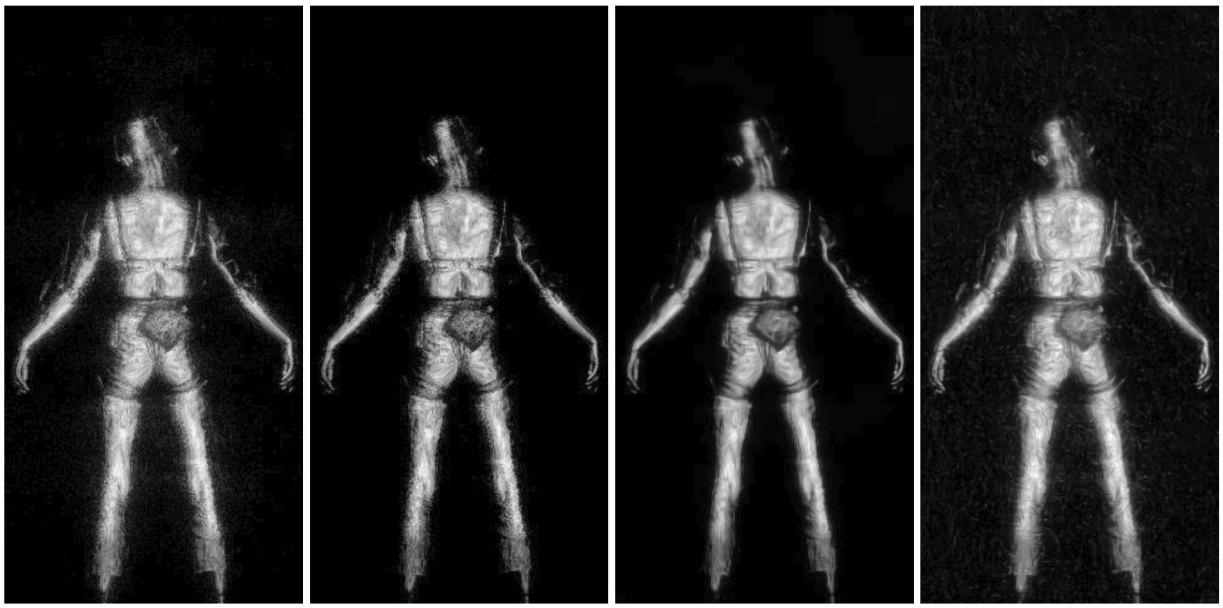

FIGURE 3 Denoising results on different input images using Blind2Unblind.

4.2 | Experimental setting for training

The 144 denoised and real noisy THz images (240 × 480 pixels) are used as the basic training dataset. To improve the training effect, we crop 27 patches of size 128 × 128 with stride 40 for each training image from the 144 denoised images, and 3888 ground truths I_{hr} are generated. When specific prior information about the noise source is not available, adding Gaussian noise is an alternative strategy to simulate the noise distribution. Gaussian noise is a common noise source for electromagnetic electronic devices in the real world. We find that the Gaussian noise added and converted to grayscale images closely resembles real noise. As real THz image noise is often distributed only around the human body and its edges, adding noise to the entire image can increase the noise learning samples of the model. Additionally, adding Gaussian noise can improve the robustness and generalisation of the model. Thus, we add Gaussian noise (standard deviations $\sigma = [10,15,25]$) to the real noisy THz images to generate the noise images $\mathbf{I}_{lr}^{\sigma_1}, \mathbf{I}_{lr}^{\sigma_2}, \mathbf{I}_{lr}^{\sigma_3}$. Finally, 3888 ground truths $\mathbf{I}_{hr}$ combined with 3888 noise images $\mathbf{I}_{lr}^{\sigma_1}, \mathbf{I}_{lr}^{\sigma_2}, \mathbf{I}_{lr}^{\sigma_3}$ are respectively used to construct the three training sets $\left\{\mathbf{I}_{hr}^{n}, \mathbf{I}_{lr}^{n}\right\}_{\sigma_j}^{N=3888}$, $j = 1,2,3$ and $n = 1,2,\ldots,N$. Note that we train the models separately with different training sets σ_j.

During the training, the epoch and batch size are set as 100 and 64, respectively. The initial learning rate is set as 2e-3 and adjusted to 2e-4 at 50 epochs and 2e-5 at 80 epochs until the end of training. The ADAM optimiser is used to optimise the network model, and the parameters of the optimiser are *betas* = [0.9,0.999], *eps* = 10^{-8}. MHRNet is implemented by the Pytorch framework, and all experiments are conducted on the NVIDIA GTX 3070 GPU.

4.3 | Comparisons with the state-of-the-art methods

During the experiments, we select 50 clear THz images from [45] as the ground truth of testset (Set50). The Gaussian noises with standard deviations σ = 10, 15 and 25, are respectively added to the real noise images to generate the noisy images.

To evaluate the performance of MHRNet, nine previously reported state-of-the-art denoising methods are utilised for comparison, including RDUNet [20], SRMNet [21], DRNet [22], ADNet [23], Neighbour2Neighbour [24], Noise2Fast [25], DRANet [47], KBNet [48] and LGBPN [49]. We trained these models from scratch using the same original experimental setup as the authors. Peak signal-to-noise ratio (PSNR) [50, 51] and structural similarity (SSIM) [51, 52] are used as indicators to evaluate the quality of the denoised images.

Table 1 presents the quantitative comparison results between MHRNet (Ours) and the other six methods on Set50. As exhibited in Table 1, when the noise intensity (deviation) increases, the denoising performance of the seven methods is reduced. Additionally, among the seven methods, SRMNet performs the second best on the PSNR and SSIM values. The evaluation metrics of ADNet, RDUNet, Neighbour2Neighbour, Noise2Fast, and DRNet are good when the Gaussian noise standard deviation is 10, but it drops dramatically with increasing noise intensity. The reason may be that these methods cannot identify and segment noises from THz images well, resulting in low denoising performance when the noise intensity is high. The evaluation metrics of MHRNet on Gaussian noises (standard deviations of 10, 15, and 25) are much better than the other six methods. Note that the amount of the training data, parameters, and FLOPs of MHRNet are 3888, 9.97 M, and 84.8 G, respectively. The experimental

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

results confirm that MHRNet can still achieve good denoising in the case of small samples and low model complexity. Especially when the standard deviation is 25, MHRNet can still maintain good denoising performance.

Figures 4–9 show the denoising effects of different methods on the Gaussian noise images with the noise standard deviations $\sigma = 10$, $\sigma = 15$ and $\sigma = 25$. For better visual comparison, some specific regions of sample images are zoomed in, and PSNR values are used for qualitative analysis. The red boxes in the enlarged areas of Figures 4–9 show the differences in denoising details between MHRNet and various comparison algorithms. As shown in Figures 4 and 5, when the Gaussian noise standard deviation is 10, ADNet can eliminate the image noises to a certain extent, but the processed image becomes blurred. Noise2Fast can restore detailed features, but it also preserves some noises. However, many noises still exist in the images of Neighbour2Neighbour and RDUNet, and the denoising effect looks poor. On the contrary, DRNet and SRMNet remove most of the noises and retain more edge features in the denoised images. However, as shown in Figure 5, when reconstructing the texture features on the clothing surface in the back of the human body, only MHRNet can fully restore them, and the other six methods have varying degrees of omission.

When the Gaussian noise standard deviation is 15, the denoised images of ADNet, Neighbour2Neighbour, RDUNet,

TABLE 1 Average PSNR (dB)/SSIM results of different methods on Gaussian noisy images.

	$\sigma = 10$		$\sigma = 15$		$\sigma = 25$		Parameters	FLOPs	
Methods	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**(M)**	**(G)**	**Input size**
ADNet	36.227	0.968	34.802	0.951	33.071	0.935	*0.52*	60.03	240 × 480
Neighbor2Neighbor	37.107	0.963	34.562	0.949	31.975	0.925	1.26	*40.68*	256 × 480
RDUNet	37.236	0.965	34.844	0.942	30.459	0.796	166.37	1419.08	240 × 480
Noise2Fast	37.485	0.958	35.702	0.944	33.226	0.914	**0.26**	**29.80**	240 × 480
DRNet	39.296	0.980	36.689	0.967	33.928	0.935	2.22	255.38	240 × 480
LGBPN	31.282	0.911	29.054	0.886	29.227	0.869	6.20	-	240 × 480
SRMNet	41.600	*0.989*	*40.885*	*0.986*	*39.642*	*0.981*	37.58	500.32	240 × 480
KBNet	-	-	40.298	0.978	37.347	0.963	118.49	120.79	240 × 480
DRANet	*42.547*	0.987	40.063	0.978	37.118	0.959	1.62	1041.20	240 × 480
Ours	**47.952**	**0.996**	**46.078**	**0.994**	**43.158**	**0.990**	9.97	84.82	240 × 480

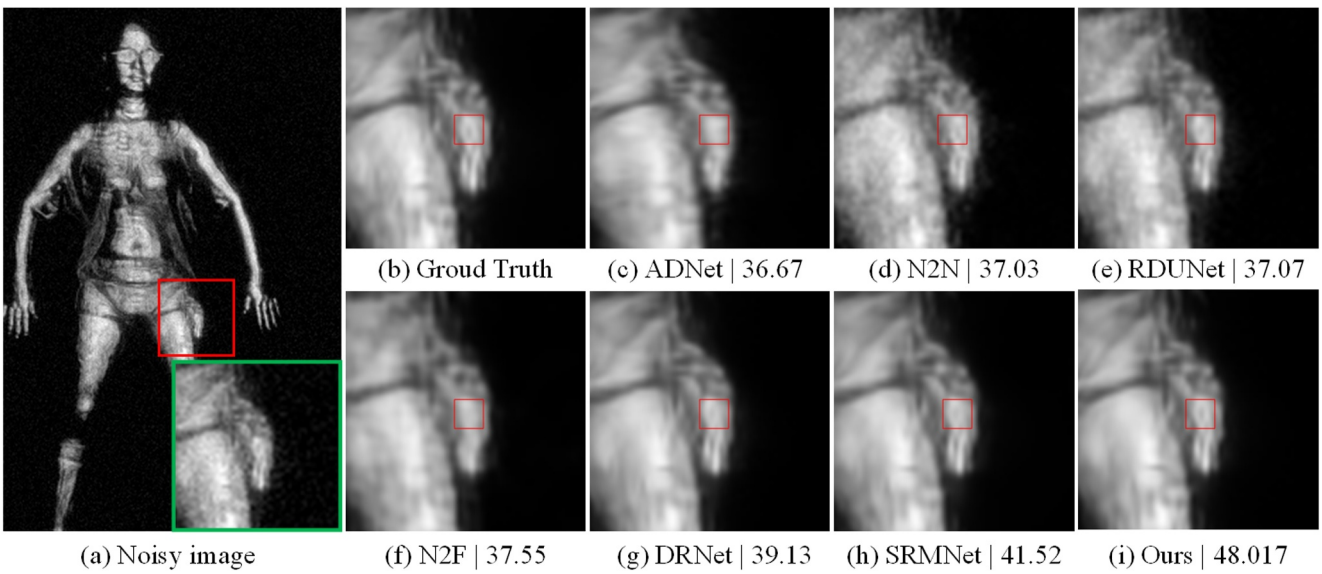

FIGURE 4 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 10$ (front).

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

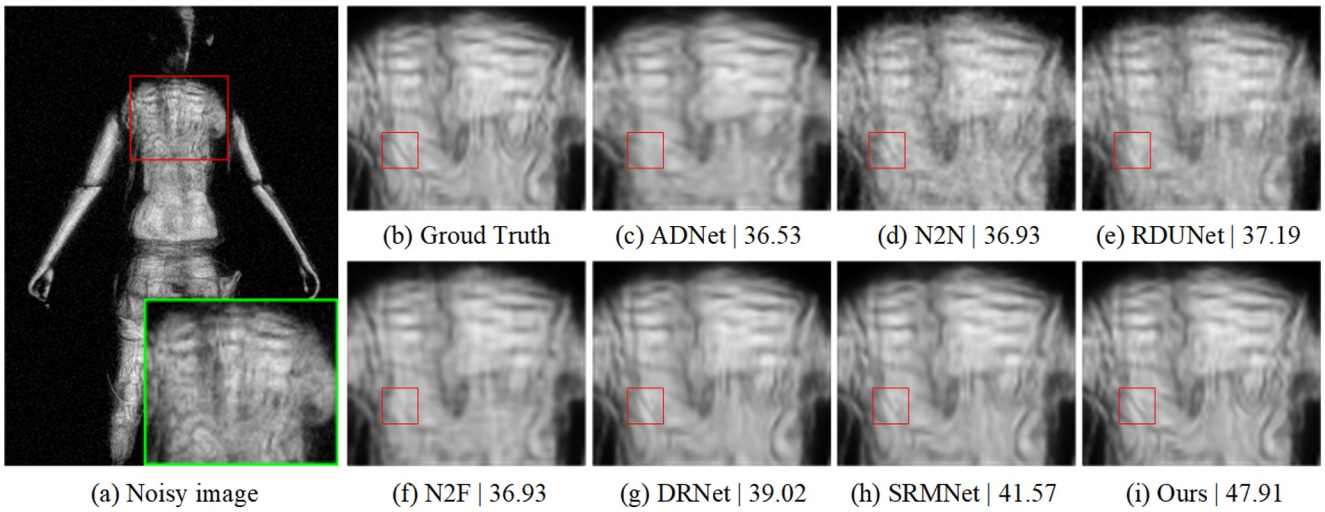

FIGURE 5 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 10$ (back).

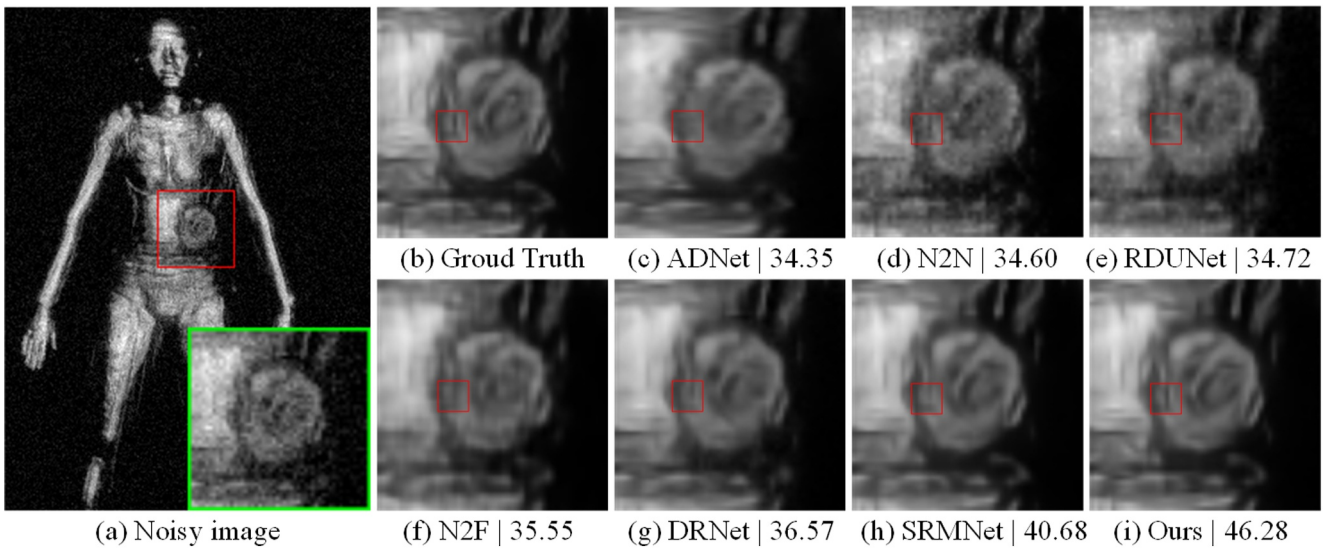

FIGURE 6 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 15$ (front).

and Noise2Fast are blurred as displayed in Figures 6 and 7. Better than the mentioned four methods, the denoising effect of DRNet and SRMNet is relatively good. However, MHRNet can reconstruct detail features in the denoised images, which is significantly better than the other six methods. For example, only MHRNet can clearly restore the texture and contour features of guns in Figure 7.

When the standard deviation of Gaussian noises reaches 25, most of the existing methods perform very poorly in denoising as shown in Figures 8 and 9. The images of ADNet and Noise2fast lose lots of detail features. After processing by Neighbour2Neighbour and RDUNet, the images still have many noises. On the contrary, MHRNet can remove noises while preserving detail features even in very poor image quality. Although the denoising effects of DRNet and SRMNet are similar to MHRNet, MHRNet is much higher than the other denoising methods in terms of PSNR and SSIM values as presented in Table 1. Consequently, the experimental results show that MHRNet outperforms the other six methods in both quantitative and qualitative results.

4.4 | Ablation experiment

The key component of MHRNet is the MHRB module. To validate the denoising effect of the MHRB module on THz images, we conduct multiple ablation experiments. Specifically,

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

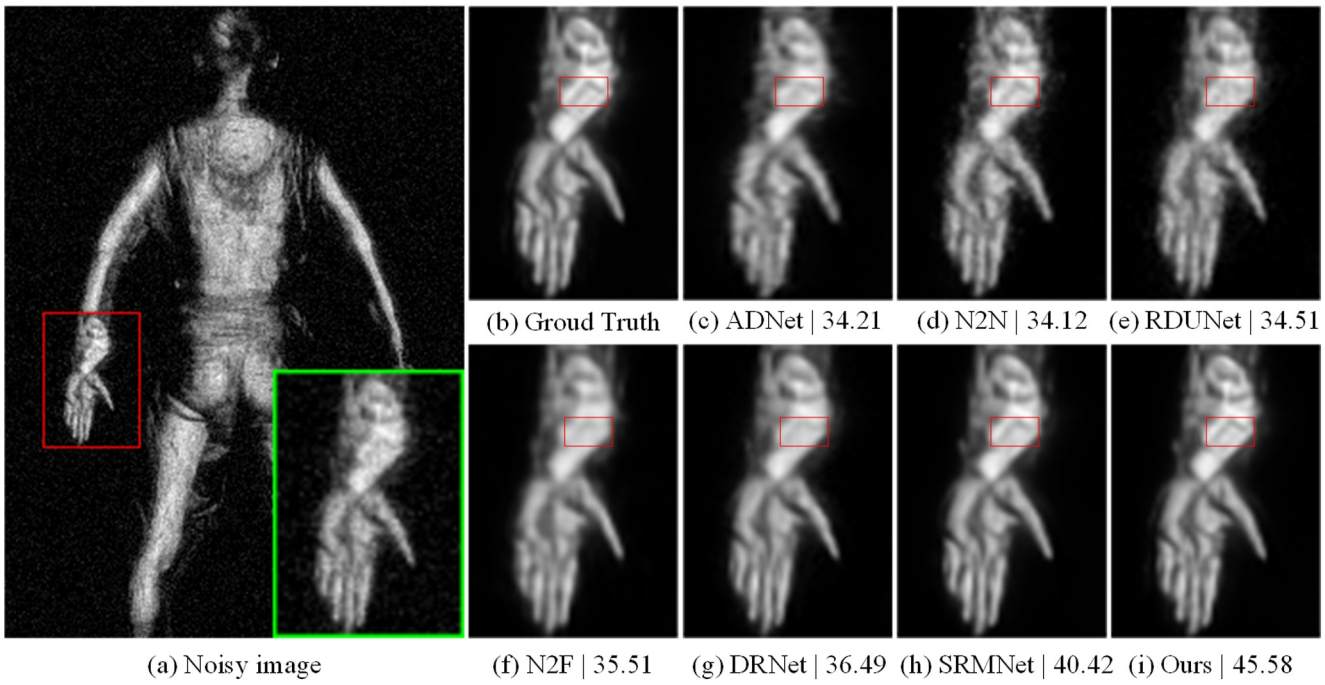

FIGURE 7 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 15$ (back).

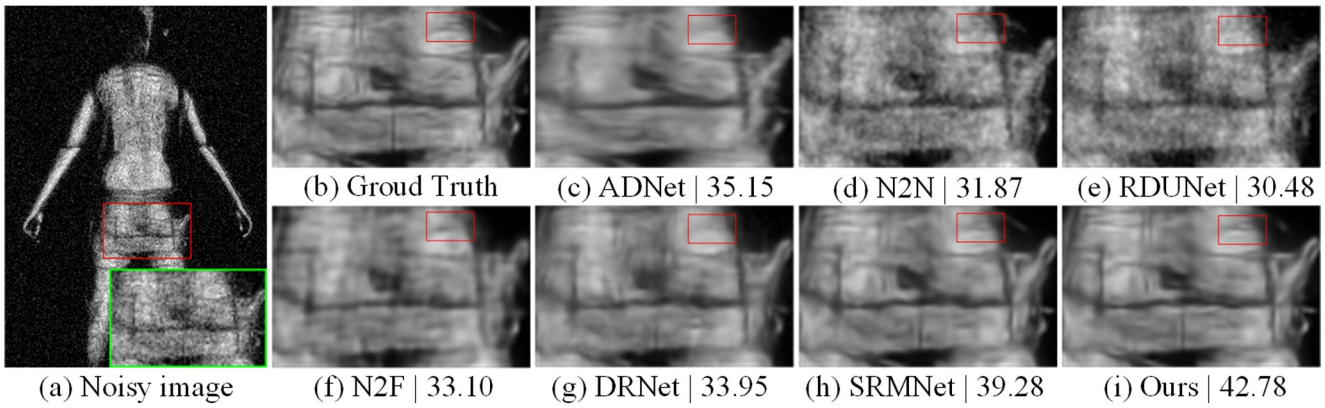

FIGURE 8 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 25$ (back).

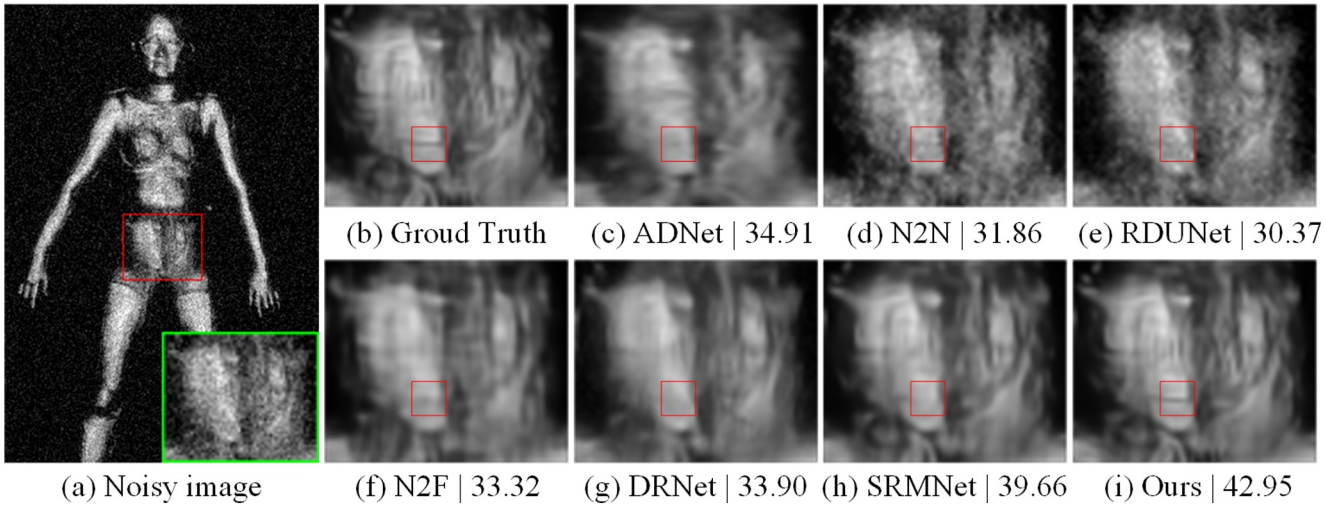

FIGURE 9 Denoising results and PSNR value of different methods on the Gaussian noise with a variance of $\sigma = 25$ (front).

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the U-Net [41] is used as the baseline. We implement comparative MHRB (MDCB and MCB) experiments in different scenarios to assess the effectiveness of the proposed MHRB module. In the ablation experiments, we use the dataset of 3888 pairs of Clean-noise THz images ($\sigma = 25$) from Section 4.2 to train MHRNet.

The test set is Set50, as detailed in Section 4.3. In addition, PSNR and SSIM evaluation indicators are used to evaluate the denoising effect of the MHRB module. The results of the ablation experiments are presented in Table 2, providing the average outcomes of different models on Set50. The methods '+MDCB' and '+MCB' mean that the network has the multiscale dilated convolution and only multiscale convolution, respectively. When individually incorporating the MDCB and MCB modules compared to the baseline model, the PSNR values are increased by 2.69 and 3.15 dB, respectively, and the SSIM values are improved by 0.007 and 0.013, respectively. MHRNet achieves the highest average PSNR, surpassing U-Net by 8.5 dB. This indicates that the U-Net structure composed of the MHRB module effectively removes noise from THz images. MHRNet obtains the highest average SSIM value, demonstrating its ability to retain more detailed features. As dilated convolutions of the proposed model are only included in MDCB, the results of '+MCB' and 'MHRNet' reflect the effect of dilated convolution. When the network has dilated convolutions, the PSNR values are increased by 5.351 dB. As shown in Figure 10, the results of the ablation experiments confirm that MHRB brings significant improvements in THz image denoising and preserving detailed features.

TABLE 2 Denoising performance comparison with different modules by PSNR and SSIM.

Methods	PSNR	SSIM
U-Net	34.662	0.9654
+MDCB	37.356	0.9721
+MCB	37.807	0.9779
MHRNet	**43.158**	**0.9896**

5 | EXPERIMENT ON REAL NOISES

5.1 | Real noise dataset and experimental setup

We select 50 noisy THz images from the Huaxun Ark Group as the test set to verify the denoising effect of MHRNet. Note that the test set contains real-world THz noises. We use four different training sets to assess the denoising effect on two types of noise: (1) only containing real noise and (2) containing real noise and Gaussian noise (with standard deviations of 10, 15, and 25). The experimental results are shown in Table 3. As the test set does not include high-quality images for reference, we utilise the no-reference-image quality evaluation indicators NIQE [46] to evaluate the denoising performance of comparison models.

TABLE 3 Denoising performance comparison with different training set by NIQE. + denotes adding noise.

Training dataset	NIQE↓
Baseline	8.694
+Gaussian-10	*7.447*
+Gaussian-15	**7.380**
+Gaussian-25	7.766
Real noise	7.993

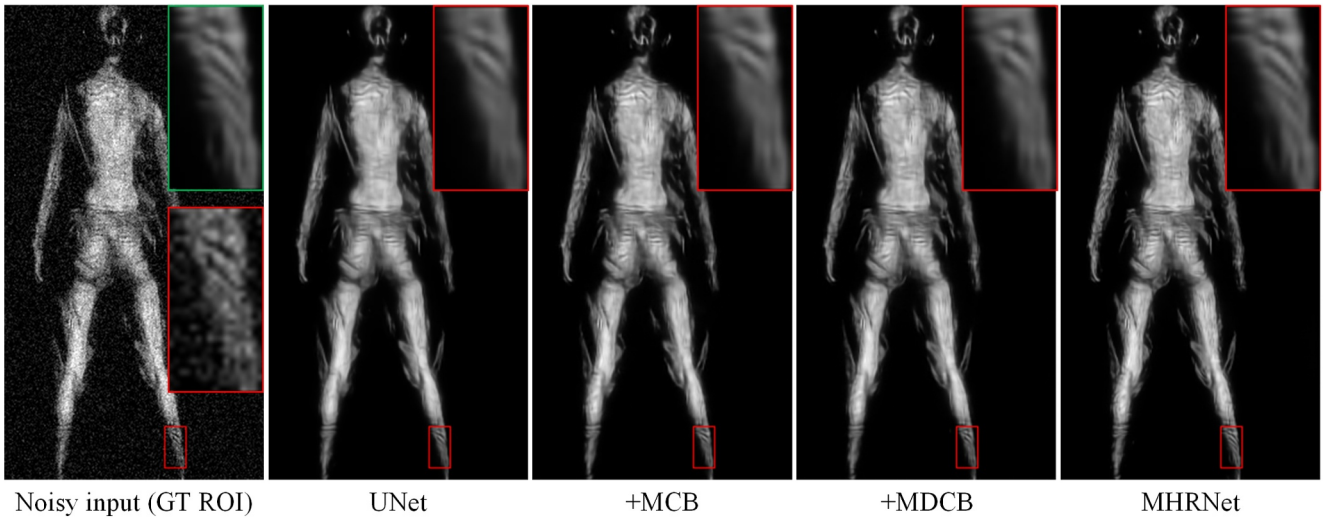

FIGURE 10 The denoising images of the ablation experiments.

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

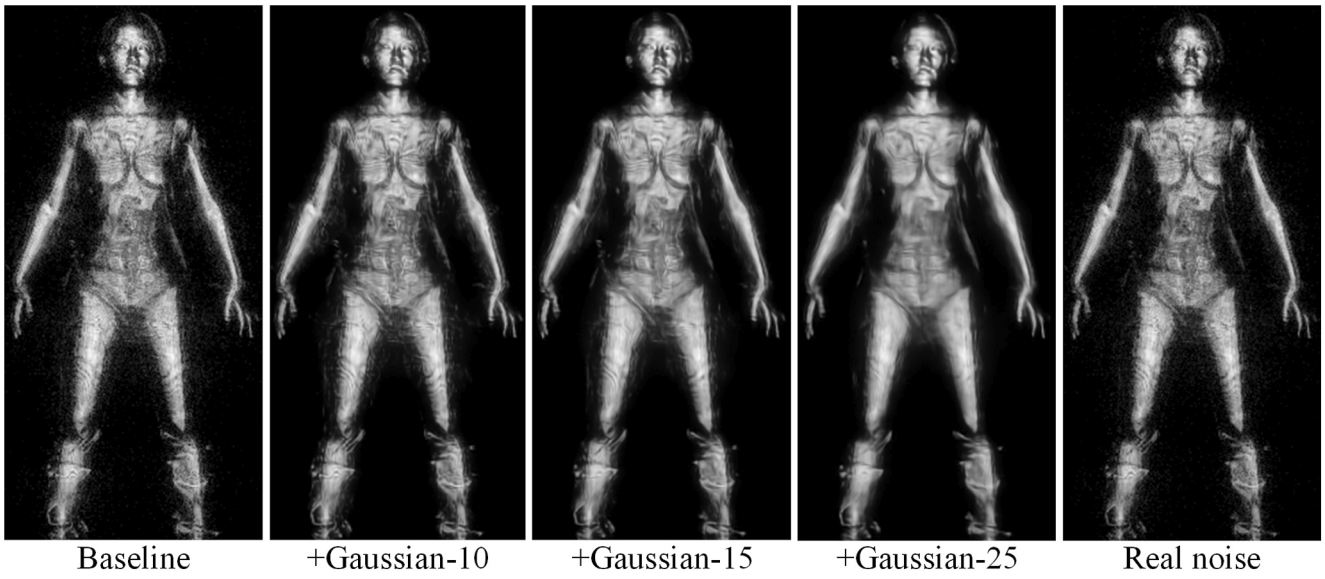

FIGURE 11 Denoising effect of different training models on real noise.

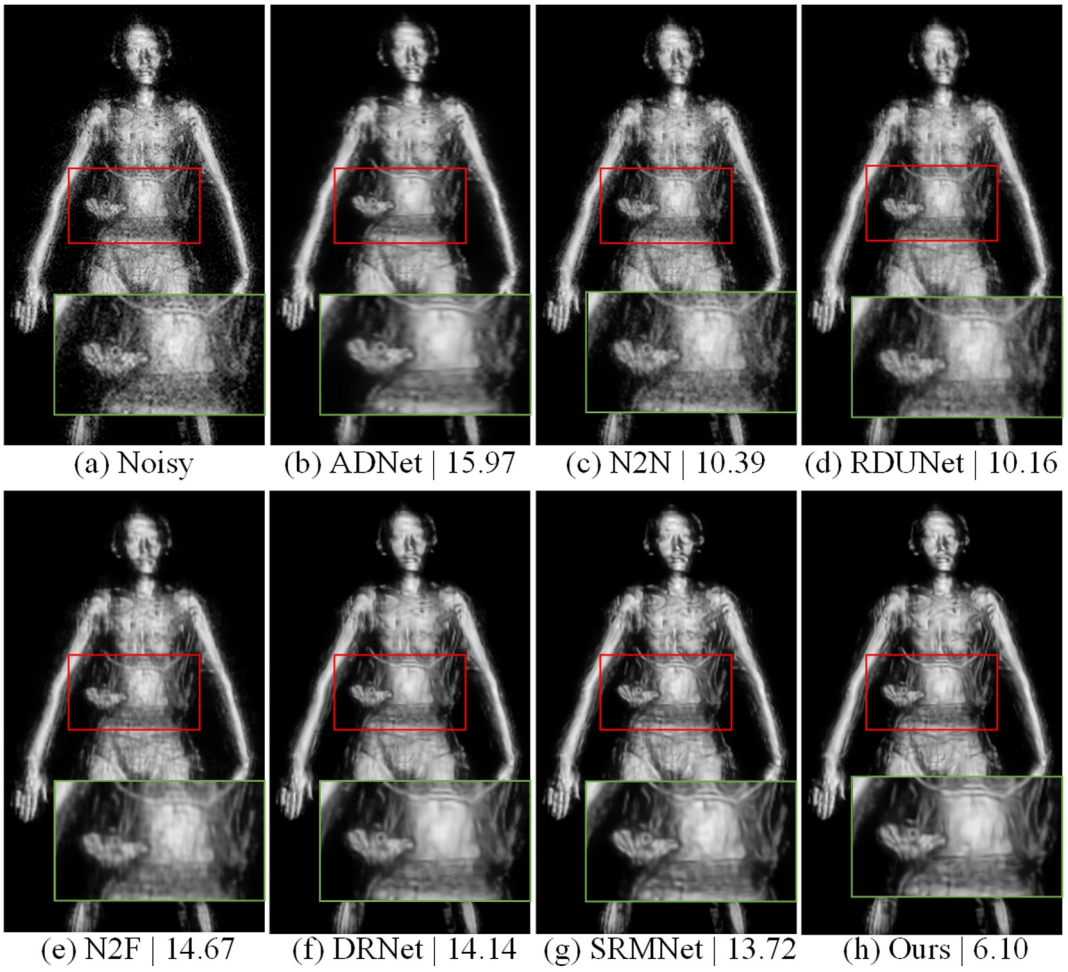

FIGURE 12 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

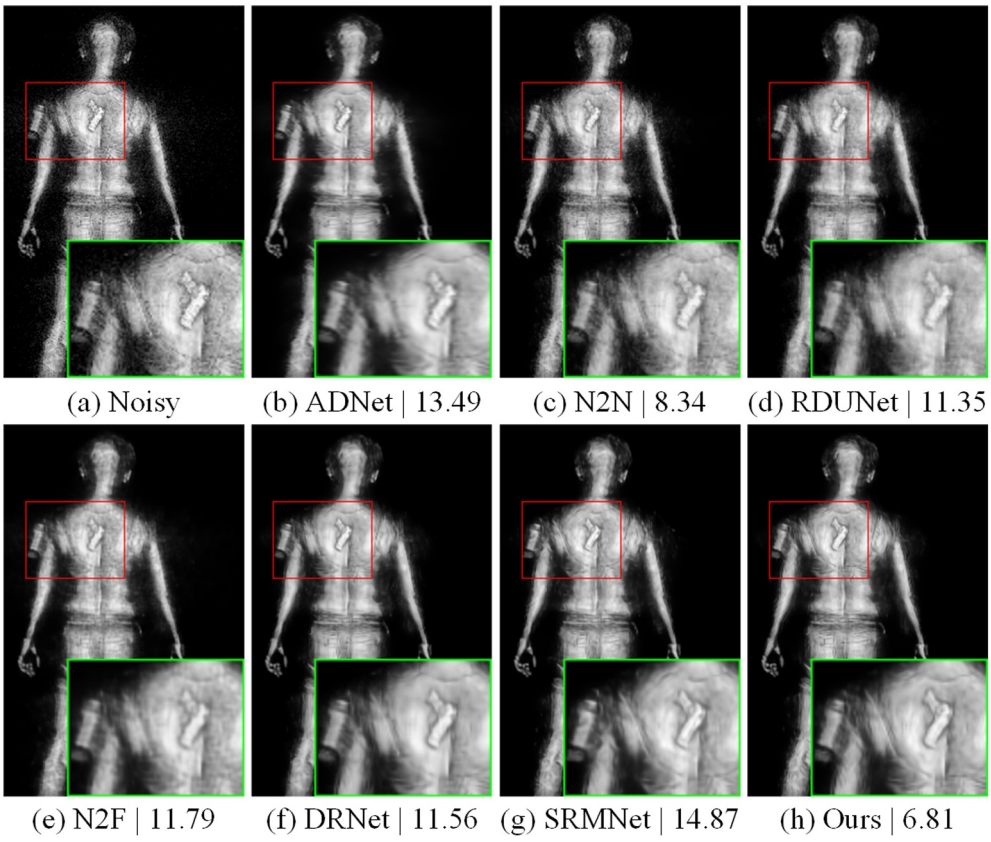

FIGURE 13 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

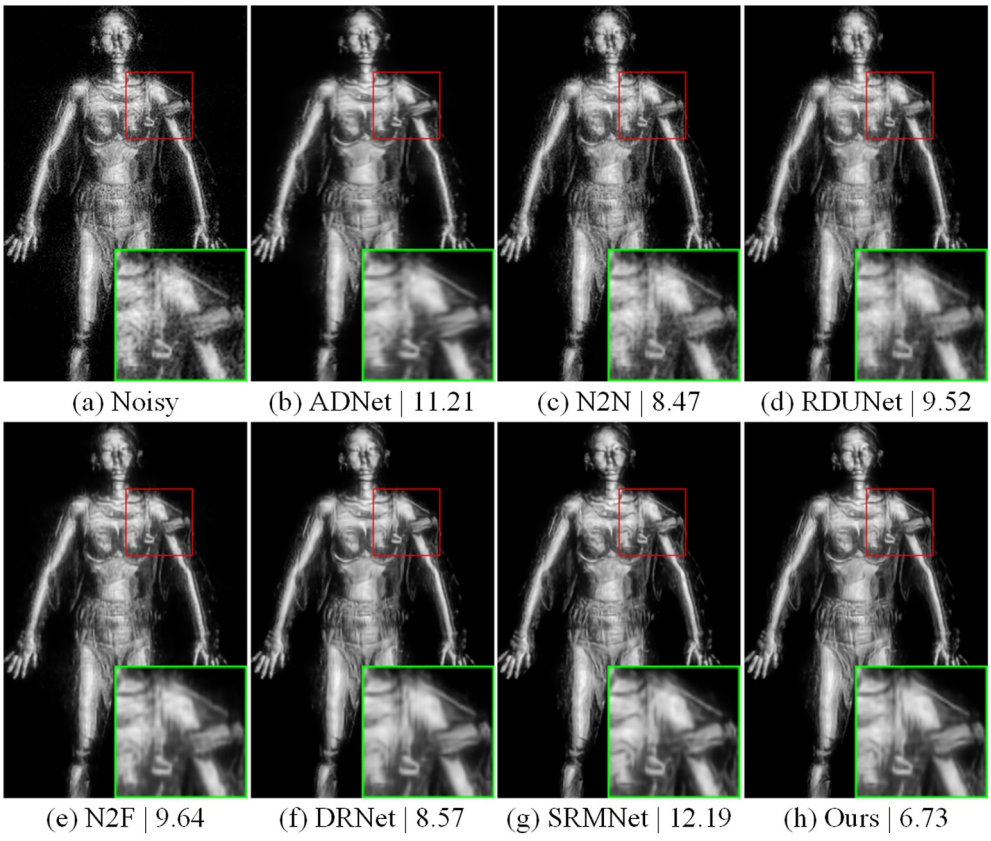

FIGURE 14 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

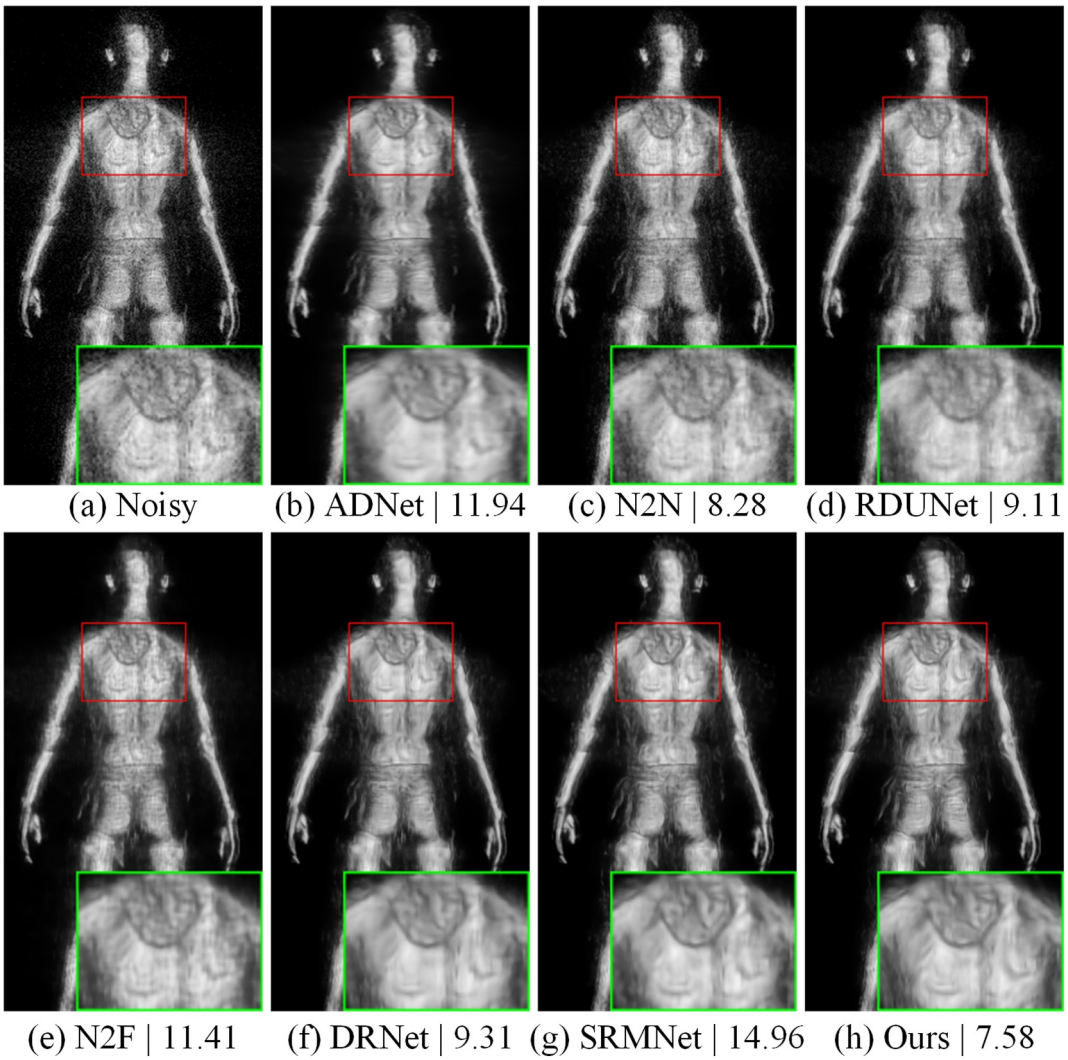

FIGURE 15 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

In Table 3, 'Baseline' denotes the original test images that are not denoised, and lower NIQE values indicate better image quality. Table 3 shows that the model trained on the dataset with a standard deviation of 15 for Gaussian noise achieves the lowest NIQE value when removing real THz image noise. However, using real THz noise as the training set produces the poorest denoising performance on THz images with real noise. The reason is that the real THz noise is often distributed only around the human body and its edges, making it difficult for the model to learn the noise distribution. Conversely, adding Gaussian noise to the entire image increases the noise distribution area, enabling the model to learn noise patterns better. Figure 11 shows the denoising effect of different training models on real noise. In Figure 11, the denoised image generated by using real noise and the Gaussian-10 model still contains a lot of noise, while the image generated by using the Gaussian-25 model causes blurred details. The comparison results indicate that the model using Gaussian-15 has the best denoising effect. According to the results in Table 3 and Figure 11, the MHRNet is trained by using a THz image dataset that adds Gaussian noise with a standard deviation of 15. The training parameter settings are the same as those in Section 4.2.

5.2 | Experimental result

Figures 12–15 show the denoising results of the real noisy THz images. To better show the denoising effect, the selected area in the figure is enlarged and the NIQE value is displayed. As shown in Figures 12–15, the original THz image contains a lot of real noises. It can be observed from the denoised images that ADNet eliminates most of the real noises but causes the image blur. In Figures 12 and 13, the contours of dangerous goods and the surface texture features of the back of the human body look blurred in the processed image. In most denoised images, many noises still

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

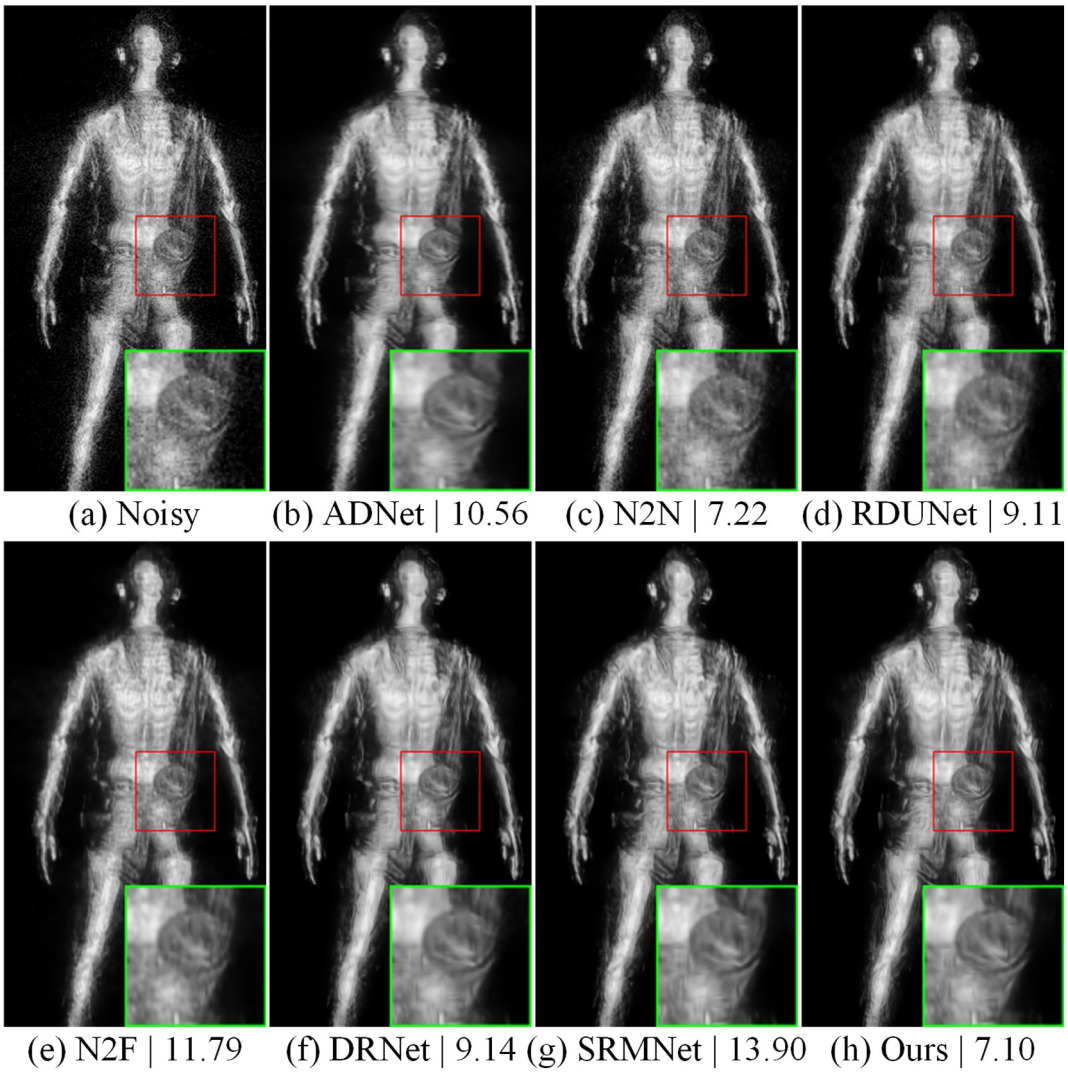

FIGURE 16 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

exist in the images processed by Neighbour2Neighbour and RDUNet.

In Figures 12 and 14, Noise2Fast eliminates part of the real noises but loses many detail features. Additionally, in Figures 13 and 15, the image of Noise2Fast still has some real noises. DRNet and SRMNet can remove most of the real noises, but some texture features are lost as shown in Figures 16 and 17. These indicate that ADNet, RDUNet, Neighbour2Neighbour, Noise2Fast, DRNet, and SRMNet have poor denoising results in the case of the small sample data.

However, with low model complexity and small sample data, MHRNet shows a good denoising effect on complex real THz noises both in selected regions and in the whole image. It can retain a large amount of detailed feature information while denoising. Therefore, MHRNet not only effectively removes synthetic noises but also performs equally well on the real THz noises.

We conduct denoising experiments on THz noise image datasets of other real scenes. The experimental results are shown in Figure 18. As the green box shows in Figure 18, MHRNet has a good generalisation capability for noise in active THz imaging security inspection scene. However, MHRNet does not handle the noise in the red box in the figure well. Since MHRNet uses a supervised training method, and the noise in the red box does not have a similar noise distribution in the previous training set, MHRNet can not perform well when dealing with some novel noise distributions. However, this situation can be improved with the diversity of training data.

6 | CONCLUSION

We propose a residual learning-based multiscale hybrid-convolution residual network MHRNet for the THz image denoising. MHRNet fuses shallow features and deep feature information and uses local and global residuals to improve denoising performance while retaining more image details. We develop an MHRB module composed of an MDCB and

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12390, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

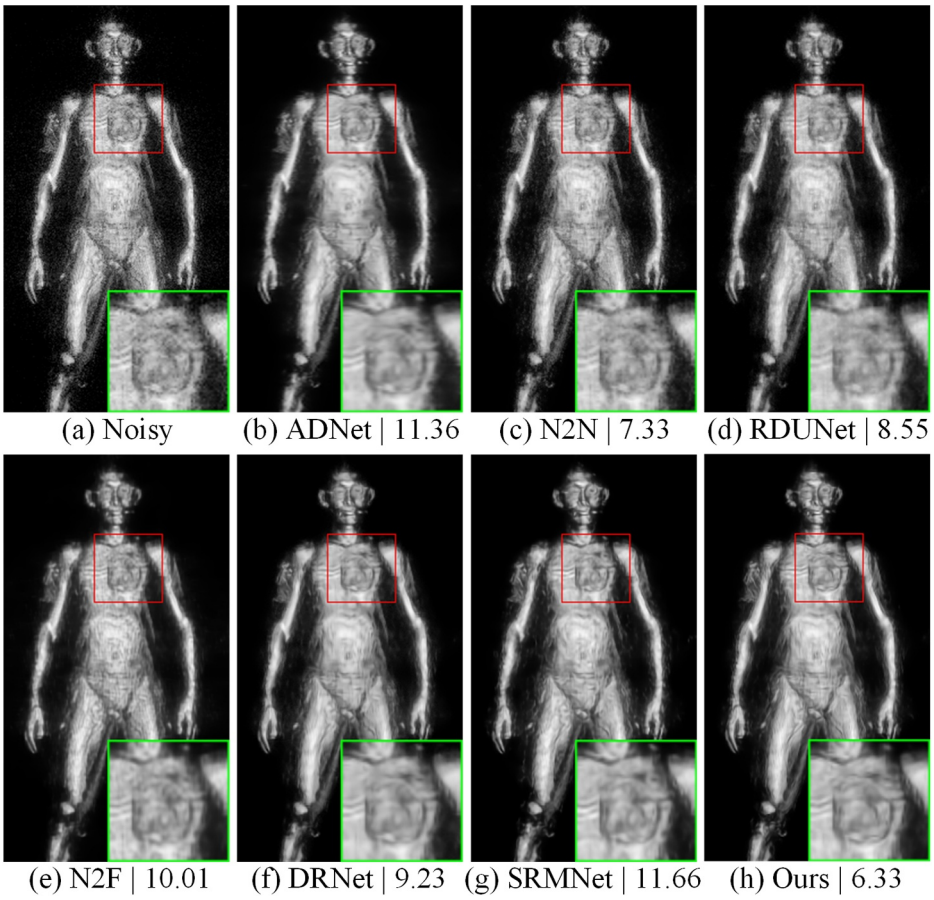

FIGURE 17 Denoising results and NIQE value of different methods on the real noisy terahertz image (female/front).

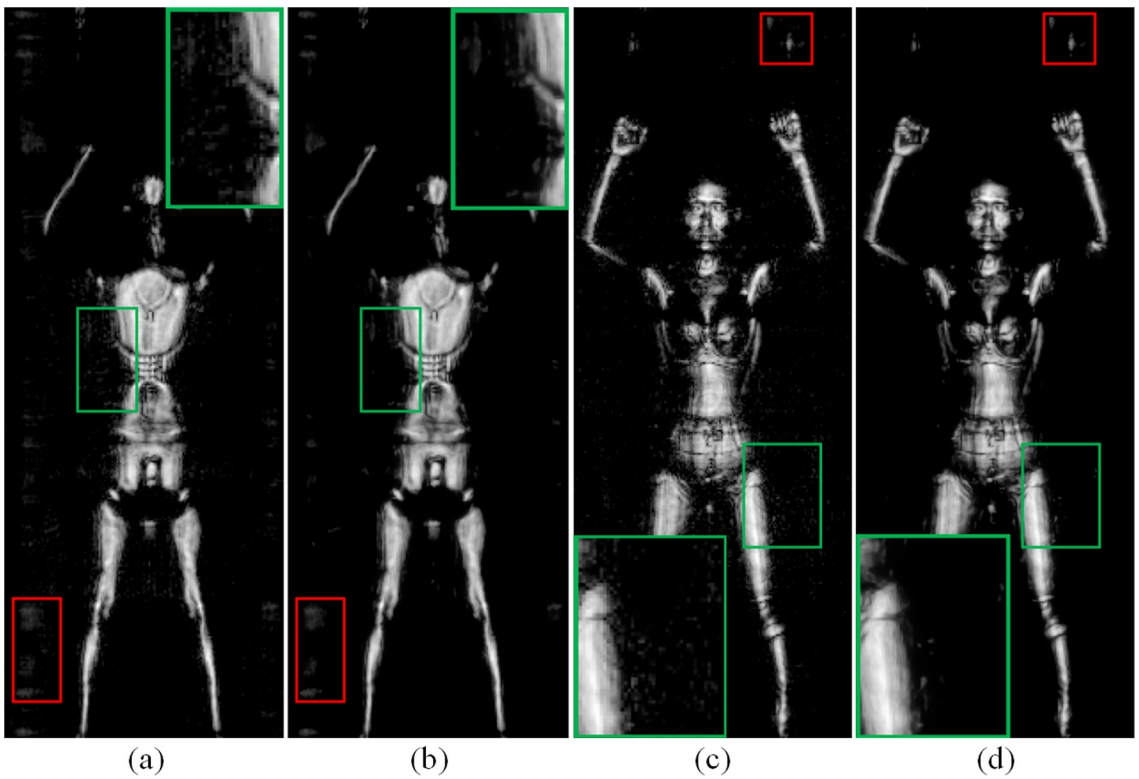

FIGURE 18 Denoising results of MHRNet on the active THz imaging security inspection scene of real noisy terahertz image. (a) and (c) are noise images. (b) and (d) are the denoised images.

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MCB for the local residual learning. The effectiveness and advantages of MHRB are verified by ablation experiments. Additionally, we implement experiments on 3 THz image datasets with different Gaussian standard deviations and one real THz image dataset to verify the denoising effect of MHRNet. Experimental results show that MHRNet has a better denoising effect and retains more image detail features compared with the ADNet, Neighbour2Neighbour, RDUNet, Noise2Fast, DRNet, and SRMNet. In the case of low model complexity and few samples, the denoising performance of MHRNet is still very competitive. This is helpful for identifying targets in THz images and has great application potential in the field of THz hidden dangerous goods detection.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (62173098, 62104047) and Guangdong Provincial Key Laboratory of Cyber-Physical System (2020B1212060069).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

ORCID

Heng Wu https://orcid.org/0000-0003-0832-2218

REFERENCES

1. Tonouchi, M.: Cutting-edge terahertz technology. Nat. Photonics 1(2), 97–105 (2007). https://doi.org/10.1038/nphoton.2007.3
2. Pawar, A.Y., et al.: Terahertz technology and its applications. Drug Invent. Today 5(2), 157–163 (2013). https://doi.org/10.1016/j.dit.2013.03.009
3. Siegel, P.H.: Terahertz technology in biology and medicine. IEEE Trans. Microw. Theor. Tech. 52(10), 2438–2447 (2004). https://doi.org/10.1109/tmtt.2004.835916
4. Tribe, W.R., et al.: Hidden object detection: security applications of terahertz technology. In: Terahertz and Gigahertz Electronics and Photonics III, vol. 5354 (2004)
5. Ergün, S., Sönmez, S.: Terahertz technology for military applications. Journal of Management and Information Science 3(1), 13–16 (2015). https://doi.org/10.17858/jmisci.58124
6. Kemp, M.C.: Detecting hidden objects: security imaging using millimetre-waves and terahertz. In: IEEE Conference on Advanced Video and Signal Based Surveillance. IEEE (2007)
7. Zhang, Xi-C., Xu, J.: THz technology in security checks. In: Introduction to THz Wave Photonics, pp. 201–219. Springer, Boston (2010)
8. Hou, L., et al.: Enhancing terahertz image quality by finite impulse response digital filter. In: 2014 39th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz). IEEE (2014)
9. Wang, Z., et al.: How do detected objects affect the noise distribution of terahertz security images? IEEE Access 6, 41087–41092 (2018). https://doi.org/10.1109/access.2018.2859359
10. Ahi, K., Anwar, M.: Developing terahertz imaging equation and enhancement of the resolution of terahertz images using deconvolution. In: Terahertz Physics, Devices, and Systems X: Advanced Applications in Industry and Defense, vol. 9856 (2016)
11. Yardimci, N.T., Jarrahi, M.: Nanostructure-enhanced Photoconductive terahertz Emission and detection. Small 14.44(44), 1802437 (2018). https://doi.org/10.1002/smll.201802437
12. Cui, S.-shan, Qi, Li, Chen, G.: Terahertz digital holography image denoising using stationary wavelet transform. In: Selected Papers from Conferences of the Photoelectronic Technology Committee of the Chinese Society of Astronautics 2014, Part II, vol. 9522, pp. 952230 (2015). https://doi.org/10.1117/12.2182877
13. Gao, J., Chen, Q., Blasch, E.: Image denoising in the presence of non-Gaussian, power-law noise. In: 2012 IEEE National Aerospace and Electronics Conference (NAECON). IEEE (2012)
14. Li, Qi, et al.: Real-time terahertz scanning imaging by use of a pyroelectric array camera and image denoising. JOSA A 27(11), 2381–2386 (2010). https://doi.org/10.1364/josaa.27.002381
15. Buades, B.C., Morel, J.: A Non-local Algorithm for Image Denoising, pp. 60–65. IEEE Computer Vision and Pattern Recognition (2005)
16. Zhao, R., Cui, H.: Improved Threshold Denoising Method Based on Wavelet Transform, pp. 1–4. IEEE International Conference on Modelling, Identification and Control (2015)
17. Dabov, K., et al.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. 16(8), 2080–2095 (2007). https://doi.org/10.1109/tip.2007.901238
18. Zhang, K., et al.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. IEEE Trans. Image Process. 26(7), 3142–3155 (2017). https://doi.org/10.1109/tip.2017.2662206
19. Zhang, K., Zuo, W., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans. Image Process. 27(9), 4608–4622 (2018). https://doi.org/10.1109/tip.2018.2839891
20. Gurrola-Ramos, J., Dalmau, O., Alarcón, T.E.: A residual dense U-Net neural network for image denoising. IEEE Access 9, 31742–31754 (2021). https://doi.org/10.1109/access.2021.3061062
21. Fan, C.-M., et al.: Selective residual M-net for real image denoising. In: 2022 30th European Signal Processing Conference (EUSIPCO). IEEE (2022)
22. Zhang, J., et al.: DRNet: a deep neural network with multi-layer residual blocks improves image denoising. IEEE Access 9, 79936–79946 (2021). https://doi.org/10.1109/access.2021.3084951
23. Tian, C., et al.: Attention-guided CNN for image denoising. Neural Network. 124, 117–129 (2020). https://doi.org/10.1016/j.neunet.2019.12.024
24. Huang, T., et al.: Neighbor2neighbor: Self-Supervised Denoising from Single Noisy Images, pp. 14781–14790 (2021)
25. Lequyer, J., et al.: Noise2Fast: fast self-supervised single image blind denoising. arXiv preprint arXiv:2108, 10209 (2021)
26. Cheng, D., et al.: Hybrid routing transformer for zero-shot learning. Pattern Recogn., (2023), 137: 109270, https://doi.org/10.1016/j.patcog.2022.109270
27. Fang, C., et al.: Cross-modality high-frequency transformer for MR image super-resolution[C]. In: Proceedings of the 30th ACM International Conference on Multimedia, pp. 1584–1592 (2022)
28. Yao, J., et al.: Contextual Dependency Vision Transformer for spectrogram-based multivariate time series analysis[J]. Neurocomputing, (2024), 572: 127215, https://doi.org/10.1016/j.neucom.2023.127215
29. Wang, Y., et al.: Multi-scale dilated convolution of convolutional neural network for image denoising. Multimed. Tool. Appl. 78(14), 19945–19960 (2019). https://doi.org/10.1007/s11042-019-7377-y
30. Wang, T., Sun, M., Hu, K.: Dilated deep residual network for image denoising. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 1272–1279. IEEE (2017)
31. Liu, C., Shang, Z., Qin, A.: A multiscale image denoising algorithm based on dilated residual convolution network. In: Image and Graphics Technologies and Applications: 14th Conference on Image and Graphics Technologies and Applications, IGTA 2019, Beijing, China, April 19–20, 2019, Revised Selected Papers 14, pp. 193–203. Springer, Singapore (2019)
32. Ljubenovic, M., et al.: CNN-Based Deblurring of terahertz images. VISIGRAPP 4(VISAPP), 323–330 (2020). https://doi.org/10.5220/0008973103230330

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

33. Liang, Y., Fan, W., Xue, B.: Image enhancement techniques used for THz imaging. In: International Symposium on Photoelectronic Detection and Imaging 2011: Terahertz Wave Technologies and Applications, vol. 8195, pp. 819515 (2011). https://doi.org/10.1117/12.900775
34. Zhang, C., et al.: Image denoising algorithm for millimeter wave imaging. In: 2022 47th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz), pp. 1–2 (2022). https://doi.org/10.1109/IRMMW-THz50927.2022.9895555
35. Zhu, W., et al.: Improved mean filtering algorithm in Thz imaging. Infrared Laser Eng. 42, 1241–1246 (2013)
36. Zhang, Z., et al.: Restoration of integrated circuit terahertz image based on wavelet denoising technique and the point spread function model. Opt Laser. Eng. 138, 106413 (2021). https://doi.org/10.1016/j.optlaseng.2020.106413
37. Ning, W., Qi, F., Wang, J.: Image restoration for terahertz image based on complex-valued deconvolution. In: 2019 8th Asia-Pacific Conference on Antennas and Propagation (APCAP). IEEE (2019)
38. Xin, Le, et al.: A template matching background filtering method for millimeter wave human security image. In: 2019 6th Asia-Pacific Conference on Synthetic Aperture Radar (APSAR). IEEE (2019)
39. Guo, Y., et al.: CMID: Crossmodal image denoising via pixel-wise deep reinforcement learning[J]. Sensors, (2023), 24(1): 42, https://doi.org/10.3390/s24010042
40. Li, K., Stantchev, R.I., Pickwell-MacPherson, E.: Convolutional Neural Network based denoising method for rapid THz Imaging. In: 2021 46th International Conference on Infrared, Millimeter and Terahertz Waves (IRMMW-THz). IEEE (2021)
41. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham (2015)
42. Liu, P., et al.: Multi-level wavelet-CNN for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018)
43. Park, B., Yu, S., Jeong, J.: Densely connected hierarchical network for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2019)
44. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
45. Wang, Z., et al.: Blind2Unblind: self-supervised image denoising with visible blind Spots. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)
46. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Process. Lett. 20, 209–212 (2012)
47. Wu, W., et al.: Dual residual attention network for image denoising. Pattern Recogn., (2024), 149: 110291
48. Zhang, Y., et al.: Kbnet: kernel basis network for image restoration. arXiv preprint arXiv:2303.02881, (2023)
49. Wang, Z., et al.: LG-BPN: local and global blind-Patch network for self-supervised real-world denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18156–18165 (2023)
50. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. 44(13), 800–801 (2008)
51. Hore, A., Ziou, D.: Image Quality Metrics: PSNR vs. SSIM, pp. 2366–2369. IEEE (2010)
52. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861

How to cite this article: Wu, H., et al.: Terahertz image denoising via multiscale hybrid-convolution residual network. CAAI Trans. Intell. Technol. 10(1), 235–252 (2025). https://doi.org/10.1049/cit2.12380

24682322, 2025, 1, Downloaded from https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12380, Wiley Online Library on [28/01/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License