


Article

Super-Resolution Reconstruction of Terahertz Images Based on Residual Generative Adversarial Network with Enhanced Attention

Zhongwei Hou, Xingzeng Cha, Hongyu An, Aiyang Zhang and Dakun Lai * 

School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: dklai@uestc.edu.cn

Abstract: Terahertz (THz) waves are widely used in the field of non-destructive testing (NDT). However, terahertz images have issues with limited spatial resolution and fuzzy features because of the constraints of the imaging equipment and imaging algorithms. To solve these problems, we propose a residual generative adversarial network based on enhanced attention (EA), which aims to pay more attention to the reconstruction of textures and details while not influencing the image outlines. Our method successfully recovers detailed texture information from low-resolution images, as demonstrated by experiments on the benchmark datasets Set5 and Set14. To use the network to improve the resolution of terahertz images, we create an image degradation algorithm and a database of terahertz degradation images. Finally, the real reconstruction of terahertz images confirms the effectiveness of our method.

Keywords: terahertz image; image super-resolution; generative adversarial network; attention mechanism



Citation: Hou, Z.; Cha, X.; An, H.; Zhang, A.; Lai, D. Super-Resolution Reconstruction of Terahertz Images Based on Residual Generative Adversarial Network with Enhanced Attention. *Entropy* **2023**, *25*, 440. <https://doi.org/10.3390/e25030440>

Academic Editor: Amelia Carolina Sparavigna

Received: 30 December 2022

Revised: 13 February 2023

Accepted: 28 February 2023

Published: 2 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Terahertz refers to electromagnetic waves with frequencies between 100 GHz and 10,000 GHz. The waveband is between microwaves and infrared light. It has some unique properties such as transience, low energy, and penetrability. Therefore, terahertz imaging [1] can be applied in fields such as biomedical diagnosis [2,3], non-destructive testing [4], industrial safety testing [5,6], etc.

A crucial technique for terahertz imaging is terahertz tomography [7]. Terahertz tomography makes use of the penetrability of terahertz waves through dielectrics to capture the changes in amplitude and phase of the electromagnetic waves as they pass through the object. Then, tomographic reconstruction algorithms are used to take these changes and figure out how the object is made inside.

However, terahertz tomography images have certain issues that result in poor resolution. Firstly, the reconstructed image is distorted due to the diffraction and scattering of the terahertz waves inside the object. Secondly, because the data acquisition process is discrete, the data in the reconstructed image are also discrete, and the missing data lead to ray-like stripe artifacts in the image.

In addition, due to the strong energy in the center of the THz Gaussian beam and weak energy in the periphery, and the fact that the beam has a certain width, the edges of the reconstructed images are presented as bands with significant widths rather than sharp lines. In other words, the picture quality is poor, and the edge's contour is hazy. Therefore, improving the terahertz image's quality after reconstruction is a crucial challenge.

There are mainly two ways to improve the resolution of terahertz images. One is to improve the imaging system, and the other is to adopt a super-resolution reconstruction method. At present, some research teams focus on improving the terahertz imaging system to improve the spatial resolution [8–10], while other teams propose to obtain a better image

effect via a super-resolution algorithm. Li et al. [11] used the Lucy–Richardson algorithm to improve the resolution of coherent terahertz imaging systems. Ding et al. [12] applied the Lucy–Richardson algorithm to a reflective terahertz imaging system.

In recent years, Dong et al. [13] have proposed a super-resolution reconstruction method based on a convolutional neural network (CNN). They applied a three-layer super-resolution network and achieved state-of-the-art results in a super-resolution reconstruction at that time. Since then, many deeper and better-performing convolutional network models for super-resolution reconstruction have been proposed [14–17]. Meanwhile, there were research teams that applied convolutional super-resolution networks to terahertz image reconstruction [18–20]. However, these methods constructed the network model by increasing the number of convolutional layers, causing the parameters to expand exponentially and they did not introduce specific structures to enhance the reconstruction of image details. Deeper networks can extract more information for picture reconstruction, but network overfitting makes training difficult. After that, Fan et al. [21] proposed a lightweight SR-CNN method to improve the image quality and to meet the demand of small datasets. In 2021, Su et al. [22] suggested a novel subspace-and-attention-guided restoration network (SARNet), which adopted attention guidance to fuse spatio-spectral features of amplitude and phase. Yang et al. [23] introduced the remaining channel mechanism and the residual channel attention mechanism to restore the high-frequency information.

This article proposes a super-resolution reconstruction model for terahertz images based on the residual generative adversarial network with EA. The model has a multi-branch residual block convolutional structure that obtains feature information from each layer during the feature extraction process and conducts feature fusion. Additionally, an enhanced attention mechanism that combines both spatial and channel attention has been added to the residual block. In addition to extracting information from the channels, it also incorporates direction-aware and position-sensitive information, forcing the network to focus more on texture and picture details while maintaining the integrity of the image's shape and lowering network parameters.

The main contributions of this paper are summarized as follows:

1. We design a super-resolution generative adversarial network with attention and residuals that are suitable for multiple super-resolution tasks.
2. We employ an enhanced attention mechanism and make the network pay more attention to the reconstruction of image details and texture information.
3. We use the cosine annealing algorithm to improve the network training process, speed up the training process, and effectively improve the network's performance.
4. We build a terahertz degradation model and image database, and apply the network to terahertz tomography image super-resolution reconstruction.

2. Related Work

2.1. Deep CNN Super-Resolution Based on Residual Block

Due to the development of integrated circuits and the increasing GPU computing power, deep learning has been gradually applied in every field. Dong et al. [13] applied the deep learning method to SR and acquired a far more effective image SR compared with the traditional methods. In 2016, the introduction of residual learning alleviated the vanishing gradient problem [24]. The author of very deep convolutional networks super-resolution (VDSR) [14] applied a residual network at image super-resolution. The low-resolution image carries low-frequency information that is similar to the low-frequency information in the high-resolution image. As a result, the network only needed to learn the residual high-frequency difference between a high-resolution image and a low-resolution image. This method increased the receptive field of the network, improved its performance, and simplified the network training. In 2017, the emergence of Densenet [25] further increased the connectivity of network features. Every layer's feature result became the input once more, allowing the network to learn more detailed feature information. To produce reconstructions that are more accurate than the actual data, Ledig et al. [26] suggested the

super-resolution generative adversarial network (SRGAN). The generative adversarial network (GAN) [27] divided the network into two parts: a generative model and a discriminative model. The generative model was used for generating super-resolution images. The discriminative model was used for discriminating the gap between generated images and ground truth images. The network could train more thoroughly if the loss functions of different models were competing with one another. On the basis of SRGAN, Wang suggested an enhanced super-resolution generative adversarial network (ESRGAN) [28] and a real-world enhanced super-resolution generative adversarial network (Real-ESRGAN) [29]. In order to improve the visual quality and model performance, the ESRGAN introduced residual dense blocks and the Real-ESRGAN proposed a set of degradation models for the degradation process of the real-world.

2.2. Image Super-Resolution Based on Attention Mechanism

Attention mechanisms can be employed for a variety of deep learning models across many different domains and tasks [30]. In computer vision, attention mechanisms are designed to locate the areas of a picture that capture human attention with a greater priority. In 1998, Itti [31] introduced a technique that employs the remarkable information of various picture elements, locates the image's attention points, and dynamically changes the image's attention points to replicate the shifting process of human visual attention. The Spatial Transformer Network (STN) [32], developed by Google DeepMind in 2015, allows the network to preprocess images by learning the deformation characteristic of the picture using the affine transformation theory. This is a kind of attention model based on space. Hu et al. [33] proposed a novel architectural unit called Squeeze-and-Excitation (SE), which adaptively recalibrates channel-wise feature responses by explicitly modeling interdependencies between channels. This mechanism might cause the network to prioritize the most useful information in the input. Then, using the residual in residual (RIR) structure and SE architectural, a very deep residual channel attention network (RCAN) [34] was deployed. Through adaptive modification of the weight on the feature channel, they could control the influence of the channel on the network feature. In 2018, Woo [35] integrated spatial and channel attention and proposed the convolutional block attention module (CBAM). It can make space information condense into channel information and provide a more precise attention mechanism. Coordinate attention (CA) [36] obtains the horizontal and vertical feature information of each channel, encodes the spatial information using batch normalization (BN) to normalize the data in each batch, and stabilizes its distribution. It then fuses the spatial information through the channel attention mechanism to achieve a composite attention structure. This enhances the relationships between the deep features of pixels.

3. Methodology

In this paper, a generative adversarial network based on an attention mechanism and a residual module was proposed, which consists of a generation network and a discrimination network.

The generation network is used to map low-resolution images to super-resolution images. The discriminator network is used to examine the difference between the generated super-resolution image and the original image, and the discrimination loss is added into the training of the generator network, enabling the network to better recover the true image features.

3.1. Generation Network

In the generation network, the network is divided into four parts: pixel matching, shallow feature extraction, deep feature mapping, and image mapping reconstruction. As shown in Figure 1, I_{lr} and I_{sr} represent the input and output of the generation network, respectively.

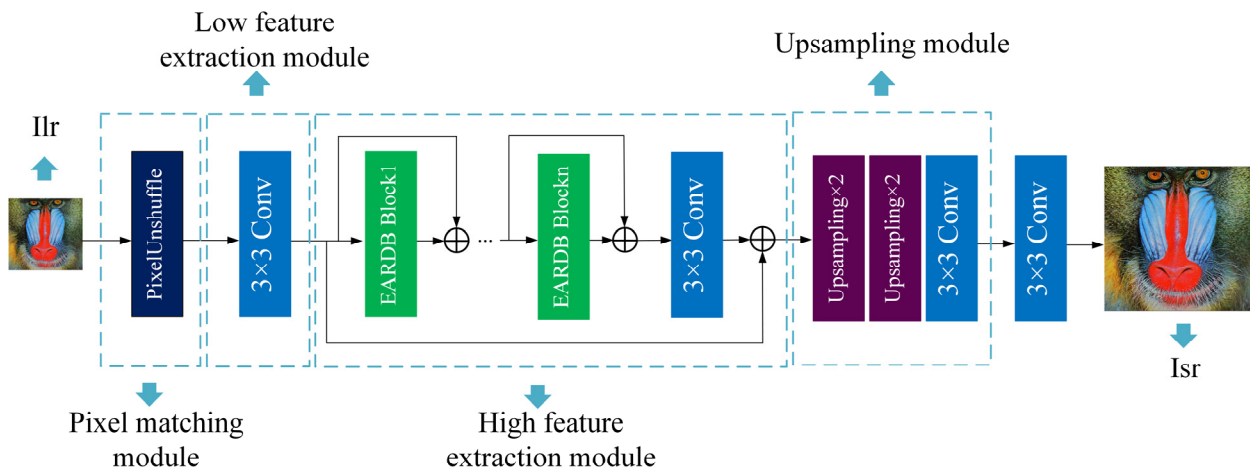


Figure 1. An overview of our generation network.

A layer of PixelUnshuffle is included in the pixel matching module to help with pixel separation. This layer realizes down-sampling by changing the four-dimensional tensor of size (B, C, H, W) to $(B, C \times r^2, H/r, W/r)$. By adjusting the parameter r , it allows the training process for $1\times$, $2\times$, and $4\times$ super-resolution tasks to share the same network. The procedure for pixel matching is depicted as

$$T_{input} = H_{pus}(I_{LR}) \quad (1)$$

where H_{pus} stands for the PixelUnshuffle layer and T_{input} for the tensor output. The $4\times$ super-resolution task networks are used as the fundamental network to share a group of networks. In the $2\times$ super-resolution task, PixelUnshuffle splits the pixels, reduces the image size by 2 times, and increases the number of channels to 4 times. Similarly, for the $1\times$ super-resolution task, the image size is reduced by 4 times, and the number of channels is increased to 16 times. Finally, the reconstruction process realizes the inverse process of PixelUnshuffle.

For shallow feature extraction modules, single-layer convolution is used for simple linear mapping. The shallow feature extraction process is expressed as

$$F_0 = H_{sf}(Input) \quad (2)$$

where H_{sf} represents the mapping process with 3×3 convolution.

The EARDB (enhanced attention residual dense block) structure block is used as the basic skeleton in the deep feature mapping module. The residual structure between EARDBs is shown in Figure 2, which can be expressed as

$$F_n = H_{EARDBn}^n(F_{n-1}) + F_{n-1} = H_{EARDBn}^n(H_{EARDBn-1}^{n-1}(\dots F_0 \dots)) + F_{n-1} \quad (3)$$

where $H_{EARDBn-1}^{n-1}$ and H_{EARDBn}^n represent the $(n-1)$ th and n th EARDB feature extraction structures, respectively. Three attention residual dense block (ARB) structures are connected by residual structures inside the EARDB module.

In order to achieve multiscale feature fusion and reduce network parameters, the dense structure inside the ARB is used to execute feature fusion. The dense process in ARB can be expressed as

$$ARB_{output} = Conv5(Cat(X_1, Conv4_{output}, \dots, Conv1_{output})) \quad (4)$$

where $Conv5$ is the last convolution in the ARB block. Cat is a contact structure that combines 32-dimension growth channel output from each convolution in the ARB.

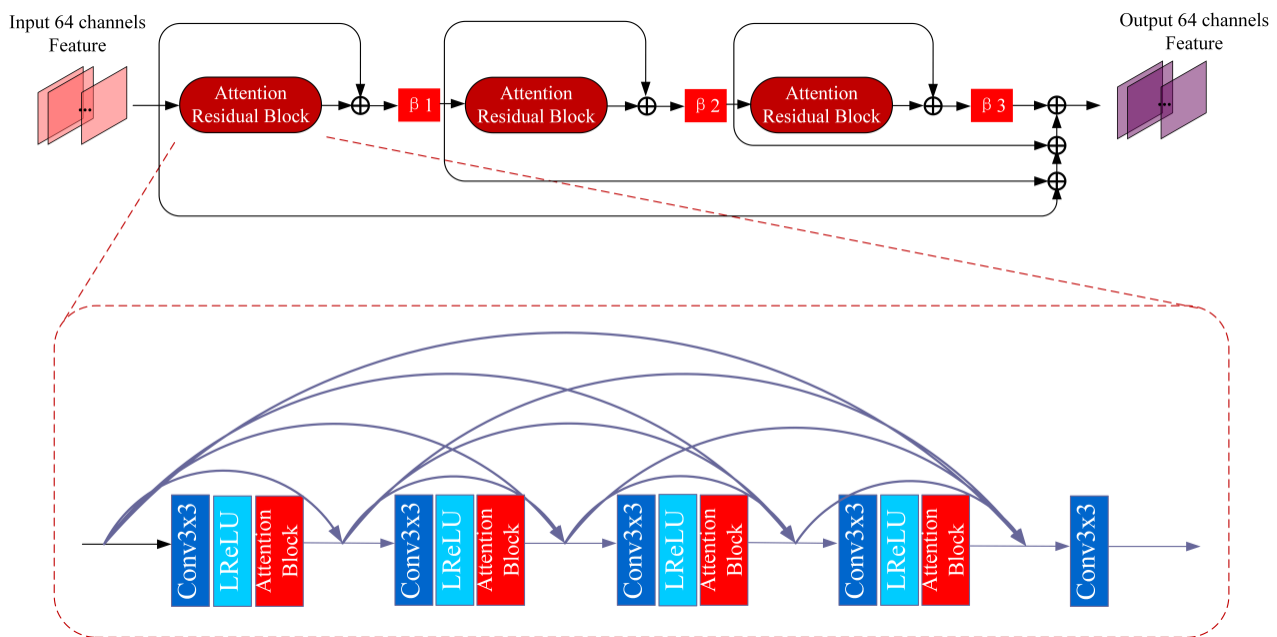


Figure 2. The architecture of EARDDB block.

For the image mapping reconstruction process, we use two up-sampling modules to interpolate the extracted features and make the feature pixels increase 4 times. The upscaling module is made up of the nearest neighbor (NN) layer. Then, the image pixel is combined using two convolution layers. It is shown as follows:

$$I_{SR} = \text{Conv}_{3 \times 3}^2(H_{up}(H_{rec}(I_{LR}) + I_{LR})) \quad (5)$$

where H_{rec} represents the reconstruction module and H_{up} represents the interpolation operation.

3.2. Enhanced Attention

The purpose of EA is to enhance the ability of the network to find key features. The input is $T_{B \times C \times X \times Y}$, a four-dimensional tensor. B represents the number of images input into one iteration of the network batch and C represents the number of characteristic channels of the image in the network. X and Y represent the size of the channels in the X and Y directions, respectively.

The structure of RCAN [34] proves that global mean pooling can build the dependency between channels, increase the sensitivity information of the model to channels, and affect the channel weights in the image reconstruction process. Additionally, inspired by coordinate attention (CA) [36], we can effectively combine channel attention and spatial attention by associating image location information. Therefore, the EA mainly consists of two processes, coordinate information generation and coordinate attention embedding. EA is shown in Figure 3.

In the process of coordinate information generation, two mean pooling kernels are used in the X and Y directions to extract the features of position information. In X direction, it outputs a tensor of $H \times 1$ dimension. The characteristics of row m in the X direction are as follows:

$$Z_{k,m} = \frac{1}{W} \sum_{0 \leq i \leq N} x_k(i, m) \quad (6)$$

where $Z_{k,m}$ represents the characteristics of the k th channel in m line, and x_k represents the characteristics of the k th channel.

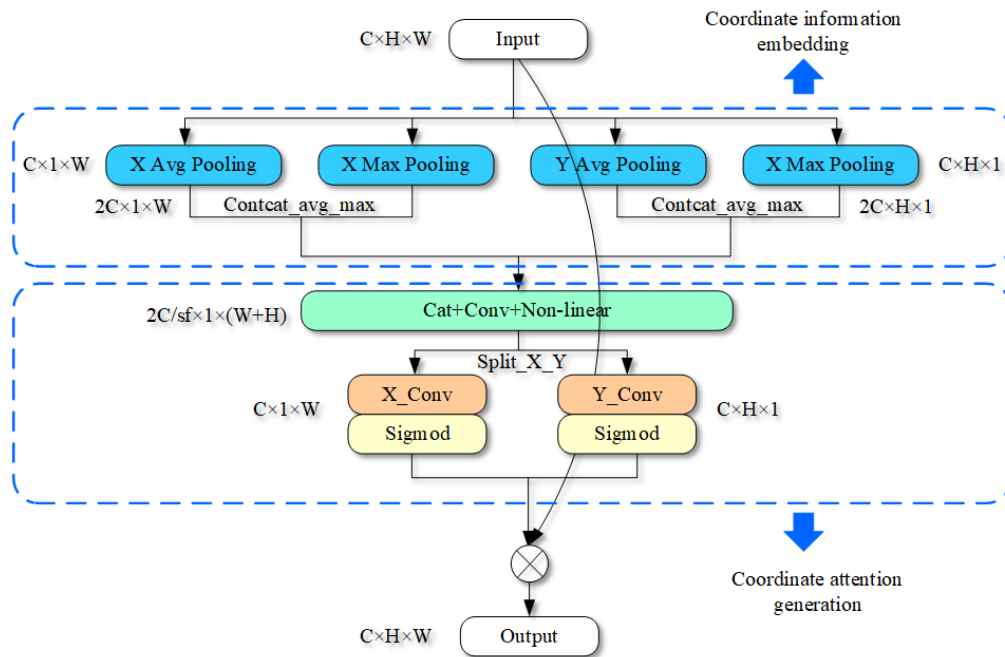


Figure 3. The architecture of enhanced attention.

Similarly, we can achieve the characteristics of Y direction for a tensor of $1 \times W$. The characteristics of the n th row and the k th channel in Y direction are shown as follows:

$$Z_{k,n} = \frac{1}{W} \sum_{0 \leq j \leq M} x_k(j, n) \quad (7)$$

Finally, $M \times N$ dimension tensor is compressed into two low dimensional tensors $M \times 1$ and $1 \times N$.

To preserve the key points in the channel, we use two maximum poolings to record the maximum values of the rows and columns. The X-direction feature tensor of dimensions $H \times 1$ and the Y-direction feature of dimensions $1 \times W$ are finally obtained. The pooling process in X direction is shown as

$$E_{k,m} = \text{Max}_{0 \leq i \leq N} (x_k(i, m)) \quad (8)$$

The pooling process in Y direction is expressed as

$$E_{k,m} = \text{Max}_{0 \leq j \leq M} (x_k(j, n)) \quad (9)$$

The result in Equations (6)–(9) transform into a tensor in $(m + n) \times 2$ through dimension change and contact operation.

For the coordinate attention-embedding process, it needs to encode all the position features and generate attention parameters. These parameters serve to emphasize the area of interest within the picture. In addition, the coding process should also consider the relationship between channels based on location information.

Firstly, coding map characteristics are obtained by 1×1 convolution and non-linear mapping module. In CA, the author adds BN to facilitate network training. However, it has been confirmed in several models such as ESRGAN [28] that BN leads to the loss of image information and the smoothing of strong changes between pixels in the super-resolution task. It is not conducive to the reconstruction of image details. After coding and non-linear mapping, we decode the feature into two tensors T_x and T_y of X and Y dimensions. Then, the two tensors are transformed by convolution, respectively. The decoding process is shown as

$$g_x = \sigma(\text{Conv}_{1 \times 1}(T_x)) \quad (10)$$

and

$$g_y = \sigma(\text{Conv}_{1 \times 1}(T_y)) \quad (11)$$

where g_x, g_y are the attention features of x direction and y direction obtained by the EA. Here, $\text{Conv}_{1 \times 1}$ is the convolutional decoding process. T_x, T_y are the tensors after encoding and non-linear mapping. Finally, the input feature is multiplied by the attention feature result, which is output for EA.

3.3. Discriminator and Loss Function

The discrimination network is shown in Figure 4. Input is the super-resolution image generated by the generation network, and output is the probability that the super-resolution image is close to the real image. The network structure mainly refers to the design idea of the VGGnet [37], which consists of convolution, preLU, and BN. It contains 8 convolutional layers; the convolution kernel size is 3×3 , and the convolution dimension gradually increases from 64 dimensions to 512 dimensions. After obtaining the deep features in the convolutional layer, the final probability value is obtained by two fully connected layers, one preLU layer, and one sigmoid layer.

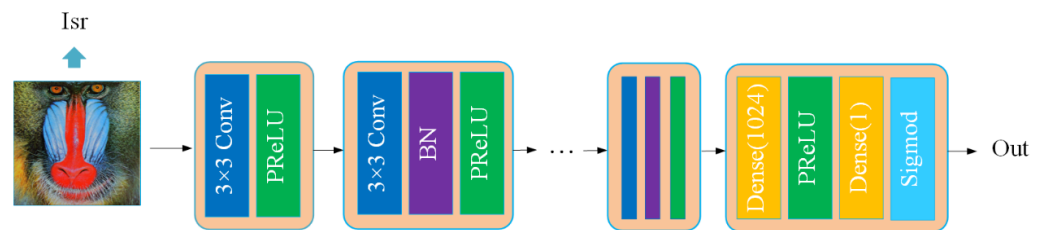


Figure 4. An overview of our discriminator network.

The generation network loss function is denoted by L_G and is shown as follows: it includes content loss function, the perceptual loss function and the adversarial loss function.

$$L_G = L_{per} + \lambda_1 L_{L1} + \lambda_2 L_G^{Ra} \quad (12)$$

In Equation (12), λ_1 and λ_2 are the weighting coefficients used to balance the two loss functions.

The content loss function is used to evaluate the L1 distance of the image I_{sr} generated by the generation network from the original image I_{hr} .

The perceptual loss L_{per} is defined by a pretrained VGG16 network. The perceptual loss function is defined as the Euclidean distance between the features of the reconstructed image I_{sr} and the real image I_{hr} . It is expressed as

$$L_{per} = E \left\{ \|\varphi[G(x)] - \varphi(y)\|_2^2 \right\} \quad (13)$$

where $\varphi[G(x)]$ represents the feature map of the generated super-resolution image through vgg16, and B is the feature map of the original high-resolution image through vgg16.

The adversarial loss L_G^{Ra} is used to judge the image generated by the network, and the adversarial loss function of the discriminator seeks to maximize the proportion of accurate evaluations. The loss function is expressed as

$$L_D = -E \{ \log D_{Ra}[y, G(x)] \} - E \{ \log \{ 1 - D_{Ra}[G(x), y] \} \} \quad (14)$$

where D_{Ra} represents the output of adversarial network.

The purpose of the generator's adversarial loss function is to minimize the probability of the correct judgment, which is expressed as

$$L_D = -E\{\log\{1 - D_{Ra}[y, G(x)]\}\} - E\{\log\{D_{Ra}[G(x), y]\}\} \quad (15)$$

where X is the original image and $G(x)$ is the image generated by the generated network.

4. Experiments

4.1. Discriminator and Loss Function

In order to compare with other SR algorithms, we use the most common training datasets DIV2K [38] and Flickr2K [39] for the training dataset DF2K. Among them, DIV2K includes 1000 2K resolution images, and Flickr2K includes 1450 2K resolution images. These images were cropped into 48,115 pieces of 400×400 pixel images, and the low-resolution images are obtained via the bicubic down-sampling.

In the training process, Set5 [40] is adopted for validation after every 500 iterators. Additionally, for the final result verification process, public benchmark datasets Set5 and Set14 [41] are employed to evaluate our proposed network.

During the imaging process, the scattering and refraction of electromagnetic waves will produce periodic stripes in addition to the noise. In the Fourier frequency spectrum, these periodic stripes have characteristic frequency points with high amplitude [42]. In order to apply the network to super-resolution reconstruction of terahertz images, we design a terahertz image degradation model to simulate real terahertz images. The degenerate expression is as follows:

$$o_1(x, y) = IFFT\{FFT[i(x, y) * PSF(x, y)] * Mask\} \quad (16)$$

$O_1(x, y)$ is a simulated terahertz image that has been degraded, while $i(x, y)$ is the original image. Firstly, the picture is blurred through $PSF(x, y)$, which is a Gaussian blur kernel. Afterwards, we use the fast Fourier transform (FFT) to convert the image to the frequency domain and multiply it with a multiplicative $Mask$. $Mask$ is a matrix used to increase the amplitude value of spectral feature points. The position of the $Mask$ is the characteristic frequency points with high amplitude positions mentioned above and it is usually $1/4$ height up and down from the vertical position of the image center. Finally, degraded terahertz-simulated images are obtained by the inverse Fourier transform ($IFFT$).

In addition, we build a dataset of tomography results, and apply the degradation algorithm to this dataset. It includes 352 pictures, 340 for network training and 12 for testing and verification.

Based on the original image I_{HR} and the reconstructed image I_{SR} , the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) are calculated to evaluate the network effect. The PSNR is expressed as

$$PSNR = 10 \times \lg\left(\frac{255^2}{MSE}\right) \quad (17)$$

where

$$MSE = \frac{1}{M \times N} \left[\sum_{i=1}^M \sum_{j=1}^N (f(i, j) - \hat{f}(i, j))^2 \right] \quad (18)$$

M and N are the height and width of an image, respectively. $f(i, j)$ represents the grayscale values of all pixels in the original image and $\hat{f}(i, j)$ represents the grayscale values of all pixels in the reconstructed image. The reconstructed image looks most like the original image when the PSNR value is high.

The SSIM is formulated as

$$SSIM(F, f) = \frac{(2\mu_F\mu_f + c_1)(2\sigma_{Ff} + c_2)}{(\mu_F^2 + \mu_f^2 + c_1)(\sigma_F^2 + \sigma_f^2 + c_2)} \quad (19)$$

F is the original reference image, f is the image to be evaluated, μ is the image gray level mean, and σ is the image gray level variance. $C1 = k1 \times L$ and $C2 = k2 \times L$. L is the image gray level, where L is the image gray level and $k1$ and $k2$ are equal to 0.01 and 0.03, respectively. SSIM obtains quantitative values by comparing the luminance, contrast, and structure of the original image and the reconstructed image. The larger the SSIM value, the closer the reconstructed image is to the original image.

Texture features in pictures with a high PSNR or SSIM may not match to the visual habits of the human eye. NIQE is a non-parametric evaluation index that measures the impact of image super-resolution by comparing the Gaussian distributions of the original picture and the super-resolution image. NIQE evaluates image quality by equation as

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{((v_1 - v_2)^T (\frac{\Sigma_1 + \Sigma_2}{2})^{-1} (v_1 - v_2))} \quad (20)$$

v_1, Σ_1 is the mean and variance of the original image Gaussian distribution, and v_2, Σ_2 is the mean and variance of the reconstructed image Gaussian distribution.

4.2. Training Details

Our model is trained using the PyTorch framework with an NVIDIA RTX 1660Ti GPU. In the pretraining, the L1 loss function is used to train the generation network and a model with a high PSNR is obtained. The optimizer is set to Adam optimizer, and the initial learning rate is 2×10^{-4} . The optimizer parameters are $\beta_1 = 0.9$, parameter $\beta_2 = 0.99$, and batch size = 16.

In training process, we employ two sets of learning rate adjustment strategies. Firstly, we employ the multistep learning rate (MultiStepLR), a technique that gradually decreases the learning rate. It can reduce the learning rate by fifty percent for every 25,000 iterations.

Secondly, the cosine annealing learning rate algorithm (CosineAnnealingLR) is used to adjust the learning rate. The characteristic of this algorithm is that the learning rate initializes at a small value, and then the rate can rise when the model becomes stable. After that, the learning rate declines gradually. In addition, the training process includes multiple CosineAnnealingLR cycles, and the learning rate of each cycle is initialized.

In the experiment, we set the CosineAnnealingLR algorithm's learning cycle to 30,000, 30,000, and 40,000 iterations. Additionally, the learning rate for the MultiStepLR method is modified every 25,000 iterations. Both of them have a total training time of 10,000 iterations.

The training curves of the two algorithms in training 9block-x4-EARDB are depicted in Figure 5, with both algorithms using 100,000 iterations to obtain the final model. It can be observed that the PSNR experiences a sudden drop after 30,000 and 60,000 iterations when using the CosineAnnealingLR algorithm. However, the curve rapidly rises again after the learning rate restarts. Compared to the MultiStepLR algorithm, the CosineAnnealingLR algorithm results in a 0.13 dB improvement in performance.

After the pretraining, this model is used as the initial model of the generated network and trained with the discriminant network. The initial learning rate is set to 10^{-4} ; Adam optimizer and MultiStepLR algorithm are used to train the generative adversarial network.

4.3. Ablation Study

Under the same training settings, to demonstrate the effectiveness of our proposed architecture, we test different attention networks on the original network. The original network structure is shown in Figure 1.

The quantitative comparisons of different attention networks for $\times 2$ SR task and $\times 4$ SR task are depicted in Table 1 on the datasets Set5 and Set14.

In order to compare the CA mechanism with other advanced attention mechanisms, we have added different attention mechanisms to the same network structure, the residual dense block (RDB) network. In Table 1, the bicubic adopts the linear interpolation method. RDB is a network without the attention structure, and the rest of the networks add SE [33],

CBAM [35], CA [36], and EA, respectively. In addition, the result of the EARDB network is used as a pretraining model to train generative adversarial network with enhanced attention residual dense block (EARDB-GAN).

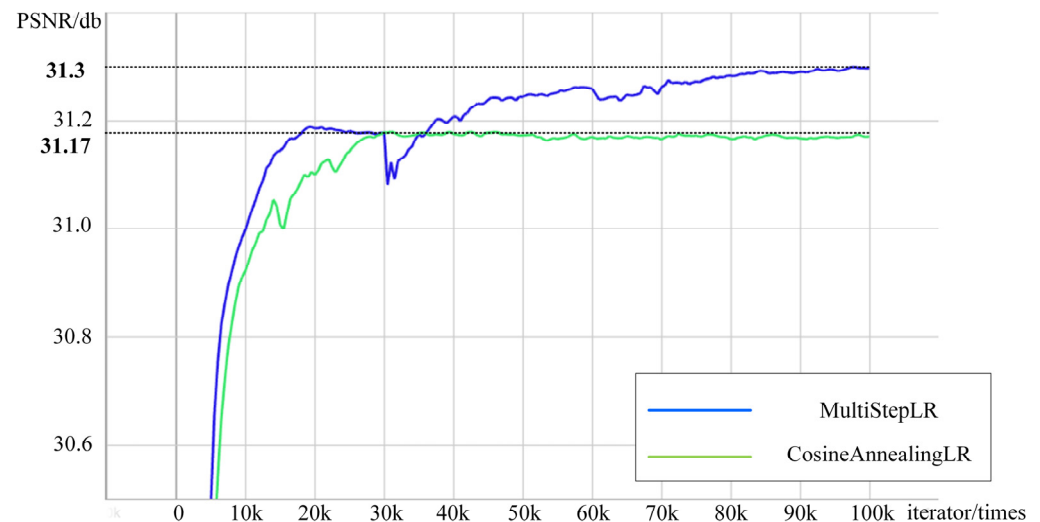


Figure 5. The training curves of two learning rate adjustment schemes.

The attention structure is added after the first four convolutions in the block. The number of blocks is set to 9 and the number of iterations to 100k iterators.

Through comparison, we find that EA has the best effect on $\times 2$ task and $\times 4$ task. On $\times 4$ task, PSNR and SSIM have a greater improvement effect, which indicates that the more pixels the more obvious the EA attention mechanism is on the feature. We also compared the results of the generated network and the discriminant network after each iteration, and found that PSNR and SSIM decreased to some extent, but NIQE reached the maximum.

The reconstruction results of EARDB, EARDB-GAN network, and bicubic algorithm with $\times 4$ image super-resolution are shown in Figure 6. It can be found that through our proposed network, the image details have been better reconstructed. Despite the fact that the EARDB network can achieve a higher PSNR, the images obtained by EARDB-GAN are more similar to real images.

In order to prove the improvement effect of the EARDB-GAN, we compare it with SRGAN [26] and ESRGAN [28] under the same training conditions. The results in Table 2 show that our network is a lightweight model, which reduces the parameters by one time and achieves the same effect as ESRGAN.

In this section, we test different generation networks in Set14 with varying numbers of blocks. The experimental results with a scaling factor of $\times 2$ in five models are shown in Table 3. It can be seen that the network has the best effect in EARDB $\times 4$ with 9 blocks, where PSNR and SSIM achieve the biggest value. In a deeper situation, the model continues to increase the number of network layers, which does not significantly improve the objective evaluation indicators, but increases the number of model parameters.

In order to verify the effect of the algorithm on terahertz images, we use the terahertz image database to resume training with the network EARDB-GAN, which is trained by DF2K. Additionally, the terahertz image database contains 352 computer-generated geometric images. The dataset includes images of various combinations of geometric shapes, such as triangles, circles, pentagons, etc. Figure 7 shows some images of this database, degraded images, and images after EARDB-GAN network reconstruction. It can be seen that the degradation algorithm has successfully simulated some problems of terahertz images, such as their low resolution, blurry edges, and fringe artifacts. Using such a dataset for training, the network can learn the degradation process of terahertz images. From the reconstructed images, the network in this paper can accurately restore the terahertz image details and retain the object contour.

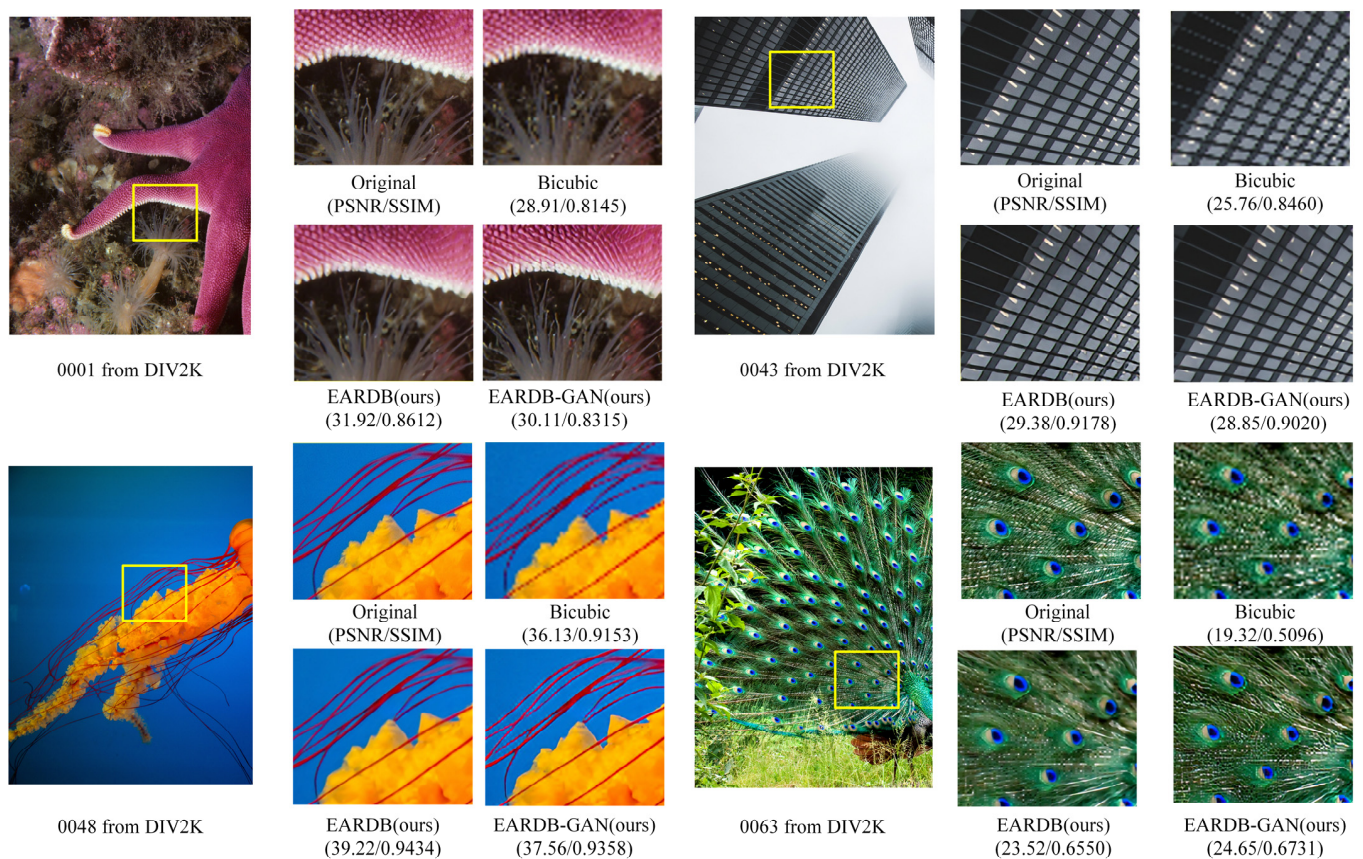


Figure 6. Visual comparison of $\times 4$ super-resolution images on the DIV2K datasets.

Table 1. Quantitative results of several SR models with attention architecture at scaling factors of $\times 2$ and $\times 4$ (average PSNR/SSIM/NIQE). The best performance is highlighted in red.

Method	Scale	Params	Set5	Set14
		(K)	PSNR/SSIM/NIQE	PSNR/SSIM/NIQE
Bicubic	$\times 2$	0	33.52/0.9230/6.2358	30.21/0.8683/5.5834
RDB	$\times 2$	27,715	36.72/0.9435/6.7594	32.69/0.8988/5.8812
SERDB	$\times 2$	27,935	36.89/0.9483/6.7312	33.02/0.9075/5.8619
CBAMRDB	$\times 2$	28,150	36.93/0.9501/6.7248	33.08/0.9083/5.8405
CARD	$\times 2$	28,375	37.01/0.9510/6.7101	33.23/0.9108/5.8322
EARDB	$\times 2$	30,470	37.14/0.9527/6.7262	33.45/0.9113/5.8326
EARDB-GAN	$\times 2$	30,470	36.95/0.9491/6.2162	33.17/0.9091/5.5138
Bicubic	$\times 4$	0	28.41/0.8091/7.2812	25.97/0.7023/6.4523
RDB	$\times 4$	27,695	30.61/0.8813/7.4811	27.53/0.7729/6.6979
SERDB	$\times 4$	27,915	30.85/0.8852/7.4631	27.75/0.7782/6.6810
CBAMRDB	$\times 4$	28,130	30.92/0.8891/7.4566	27.91/0.7829/6.6731
CARD	$\times 4$	28,355	31.11/0.8908/7.4392	28.03/0.7853/6.6607
EARDB	$\times 4$	30,450	31.30/0.8913/7.4401	28.20/0.7890/6.6725
EARDB-GAN	$\times 4$	30,450	31.03/0.8901/7.2293	27.86/0.7851/6.4281

Table 2. Several SR models with GAN at scaling factors of 2 and 4 yielded quantitative results (average PSNR, SSIM, and NIQE). The best performance is highlighted in red.

Method	Scale	Params	Set5	Set14
		(K)	PSNR/SSIM/NIQE	PSNR/SSIM/NIQE
SRGAN	$\times 2$	110,870	36.83/0.9428/6.2257	32.81/0.9041/5.5174
ESRGAN	$\times 2$	130,950	36.91/0.9483/6.2203	32.95/0.9067/5.5153
EARDB-GAN	$\times 2$	30,470	36.95/0.9491/6.2162	33.17/0.9091/5.5138

Table 2. *Cont.*

Method	Scale	Params	Set5	Set14
		(K)	PSNR/SSIM/NIQE	PSNR/SSIM/NIQE
SRGAN	$\times 4$	110,830	30.95/0.8879/7.2317	27.79/0.7831/6.4297
ESRGAN	$\times 4$	130,910	31.01/0.8892/7.2303	27.86/0.7845/6.4282
EARDB-GAN	$\times 4$	30,450	31.03/0.8901/7.2293	27.86/0.7851/6.4281

Table 3. Comparison network with the number of RDB. All the other settings are strictly the same. The best performance is highlighted in red.

Network	EARDBx4 7 Blocks	EARDBx4 8 Blocks	EARDBx4 9 Blocks	EARDBx4 10 Blocks	EARDBx4 11 Blocks
Params	23,752 K	27,101 K	30,450 K	33,799 K	37,148 K
PSNR/SSIM	28.13/0.7881	28.17/0.7887	28.20/0.7890	28.19/0.7890	28.14/0.7883

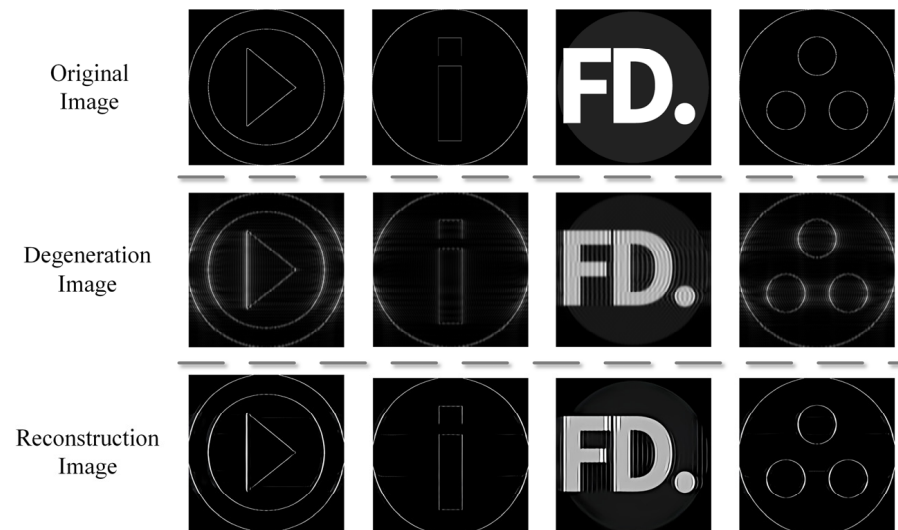
**Figure 7.** Application of degeneration algorithm on terahertz dataset and reconstruction effect of our EARDB-GAN network.

Figure 8 shows a group of real terahertz images. These images were preprocessed using the wavelet adaptive threshold denoising algorithm [43], which employs wavelet decomposition and adaptively adjusts the denoising threshold and wavelet reconstruction to obtain images that are denoising with smoother edges. Finally, these preprocessing images are reconstructed by EARDB-GAN. It can be found that the method in this paper has a good effect on the super-resolution task of terahertz images.

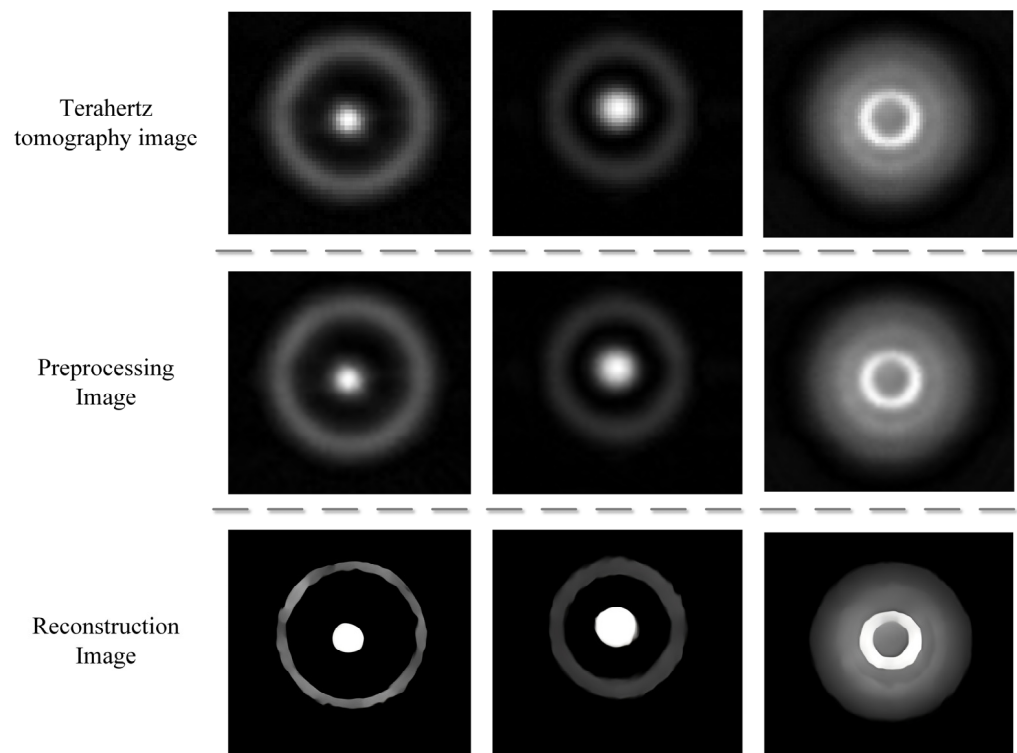


Figure 8. Reconstruction of real terahertz image after wavelet preprocessing.

5. Conclusions

In this paper, we propose a super-resolution reconstruction method for terahertz images based on a residual generative adaptive network with an enhanced attention mechanism. The network's key parameters can be adaptively updated using the attention module, and pixel coordinate information can be incorporated into the attention mechanism. Efficient residual dense connection blocks are used to realize the multiscale information fusion of the image. Extensive quantitative and qualitative experiments demonstrate that our method outperforms most state-of-the-art attention mechanisms.

The network's training effect has been improved through the periodic simulated annealing training method.

To apply the network to terahertz image super-resolution reconstruction, a terahertz image training dataset and image degradation algorithm have been established. The experiments show that our algorithm has a significant impact on terahertz image reconstruction.

Author Contributions: Conceptualization, Z.H. and D.L.; methodology, Z.H.; data curation, X.C.; formal analysis, Z.H. and H.A.; writing—original draft preparation, Z.H. and A.Z.; writing—review and editing, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 61771100 and Sichuan Science and Technology Program under Grant No. 2021YFH0093.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Valušis, G.; Lisauskas, A.; Yuan, H.; Knap, W.; Roskos, H.G. Roadmap of Terahertz Imaging 2021. *Sensors* **2021**, *21*, 4092. [[CrossRef](#)] [[PubMed](#)]
2. Fan, S.; Ung, B.S.; Parrott, E.P.; Wallace, V.P.; Pickwell-MacPherson, E. In vivo terahertz reflection imaging of human scars during and after the healing process. *J. Biophotonics* **2017**, *10*, 1143–1151. [[CrossRef](#)] [[PubMed](#)]

3. Joseph, C.S.; Patel, R.; Neel, V.A.; Giles, R.H.; Yaroslavsky, A.N. Imaging of ex vivo nonmelanoma skin cancers in the optical and terahertz spectral regions optical and terahertz skin cancers imaging. *J. Biophotonics* **2014**, *7*, 295–303. [[CrossRef](#)] [[PubMed](#)]
4. Ahi, K.; Shahbazmohamadi, S.; Asadizanjani, N. Quality control and authentication of packaged integrated circuits using enhanced-spatial-resolution terahertz time-domain spectroscopy and imaging. *Opt. Lasers Eng.* **2018**, *104*, 274–284. [[CrossRef](#)]
5. Friederich, F.; May, K.H.; Baccouche, B.; Matheis, C.; Bauer, M.; Jonuscheit, J.; Moor, M.; Denman, D.; Bramble, J.; Savage, N. Terahertz Radome Inspection. *Photonics* **2018**, *5*, 1. [[CrossRef](#)]
6. Quast, H.; Keil, A.; Löffler, T. Investigation of foam and glass fiber structures used in aerospace applications by all-electronic 3D Terahertz imaging. In Proceedings of the 35th International Conference on Infrared, Millimeter, and Terahertz Waves, Rome, Italy, 5–10 September 2010; pp. 1–2.
7. Guillet, J.P.; Recur, B.; Frederique, L.; Bousquet, B.; Canioni, L.; Manek-Hönninger, I.; Desbarats, P.; Mounaix, P. Review of terahertz tomography techniques. *J. Infrared Millim. Terahertz Waves* **2014**, *35*, 382–411. [[CrossRef](#)]
8. Bitzer, A.; Ortner, A.; Walther, M. Terahertz near-field microscopy with subwavelength spatial resolution based on photoconductive antennas. *Appl. Opt.* **2010**, *49*, E1–E6. [[CrossRef](#)]
9. Guillet, J.P.; Chusseau, L.; Adam, R.; Grosjean, T.; Penarier, A.; Baida, F.; Charraut, D. Continuous-wave scanning terahertz near-field microscope. *Microw. Opt. Technol. Lett.* **2011**, *53*, 580–582. [[CrossRef](#)]
10. Ruan, D.; Li, Z.; Du, L.; Zhou, X.; Zhu, L.; Lin, C.; Yang, M.; Chen, G.; Yuan, W.; Liang, G. Realizing a terahertz far-field sub-diffraction optical needle with sub-wavelength concentric ring structure array. *Appl. Opt.* **2018**, *57*, 7905–7909. [[CrossRef](#)] [[PubMed](#)]
11. Li, Y.; Li, L.; Hellicar, A.; Guo, Y.J. Super-Resolution Reconstruction of Terahertz Images. In Proceedings of the SPIE Defense and Security Symposium, Orlando, FL, USA, 16–20 March 2008.
12. Ding, S.-H.; Li, Q.; Yao, R.; Wang, Q. High-resolution terahertz reflective imaging and image restoration. *Appl. Opt.* **2010**, *49*, 6834–6839. [[CrossRef](#)]
13. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 14 August 2014; pp. 184–199.
14. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
15. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
16. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3147–3155.
17. Wang, Z.; Chen, J.; Hoi, S.C.; Intelligence, M. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal.* **2020**, *43*, 3365–3387. [[CrossRef](#)] [[PubMed](#)]
18. Li, Y.; Hu, W.; Zhang, X.; Xu, Z.; Ni, J.; Ligthart, L.P. Adaptive terahertz image super-resolution with adjustable convolutional neural network. *Opt. Express* **2020**, *28*, 22200–22217. [[CrossRef](#)] [[PubMed](#)]
19. Long, Z.; Wang, T.; You, C.; Yang, Z.; Wang, K.; Liu, J. Terahertz image super-resolution based on a deep convolutional neural network. *Appl. Opt.* **2019**, *58*, 2731–2735. [[CrossRef](#)] [[PubMed](#)]
20. Wang, Y.; Qi, F.; Wang, J. Terahertz image super-resolution based on a complex convolutional neural network. *Opt. Lett.* **2021**, *46*, 3123–3126. [[CrossRef](#)] [[PubMed](#)]
21. Fan, L.; Zeng, Y.; Yang, Q.; Wang, H.; Deng, B. Fast and high-quality 3-D terahertz super-resolution imaging using lightweight SR-CNN. *Remote Sens.* **2021**, *13*, 3800. [[CrossRef](#)]
22. Su, W.-T.; Hung, Y.-C.; Yu, P.-J.; Yang, S.-H.; Lin, C.-W. Seeing Through a Black Box: Toward High-Quality Terahertz Imaging via Subspace-and-Attention Guided Restoration. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–27 October 2022; pp. 453–469.
23. Yang, X.; Zhang, D.; Wang, Z.; Zhang, Y.; Wu, J.; Wu, B.; Wu, X. Super-resolution reconstruction of terahertz images based on a deep-learning network with a residual channel attention mechanism. *Appl. Opt.* **2022**, *61*, 3363–3370. [[CrossRef](#)]
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
25. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
26. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
27. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]
28. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
29. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 1905–1914.

30. Brauwiers, G.; Frasincar, F.; Engineering, D. A general survey on attention mechanisms in deep learning. *IEEE Trans. Knowl.* **2021**. [\[CrossRef\]](#)
31. Itti, L.; Koch, C.; Niebur, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 1254–1259. [\[CrossRef\]](#)
32. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [\[CrossRef\]](#)
33. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ICCV), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
34. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer vision (ECCV), Glasgow, UK, 8–14 September 2018; pp. 286–301.
35. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 8–14 September 2018; pp. 3–19.
36. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 13713–13722.
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
38. Agustsson, E.; Timofte, R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 126–135.
39. Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; Zhang, L. Ntire 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 114–125.
40. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the British Machine Vision Conference, Surrey, UK, 3–7 September 2012.
41. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2010; pp. 711–730.
42. Zou, Y.; Ge, Q.; Han, Y.; Shen, L.; Zhan, C. Stripe noise of THz image processing based on frequency filtering. *Comput. Eng. Appl.* **2009**, *45*, 241–243.
43. Hou, Z.; An, H.; He, L.; Li, E.; Lai, D. Super-Resolution Reconstruction Algorithm for Terahertz Images. In Proceedings of the 2022 3rd International Conference on Pattern Recognition and Machine Learning (PRML), Chengdu, China, 15–17 July 2022; pp. 180–185.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.