



Visibility Extension of 1-D Aperture Synthesis by a Residual CNN for Spatial Resolution Enhancement

Guanghui Zhao , Qingxia Li ^{*}, Zhiwei Chen , Zhenyu Lei, Chengwang Xiao and Yuhang Huang

Science and Technology on Multi-Spectral Information Processing Laboratory, School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

* Correspondence: qingxia_li@hust.edu.cn

Abstract: In order to improve the spatial resolution of a one-dimensional aperture synthesis (1-D AS) radiometer without increasing the size of the antenna array, the method of visibility extension (VE) is proposed in this article. In the VE method, prior information about the visibility distribution of various scenes is learnt by a residual convolutional neural network (ResCNN). Specifically, the relationship between the distribution of low-frequency visibility and that of high-frequency visibility is learnt. Then, the ResCNN is used to estimate the high-frequency visibility samples from the low-frequency visibility samples obtained by the AS system. Furthermore, the low- and high-frequency visibility samples are combined to reconstruct the brightness temperature image of the scene, to enhance the spatial resolution of AS. The simulation and experiment both demonstrate that the VE method can enhance the spatial resolution of 1-D AS.

Keywords: aperture synthesis (AS); visibility extension; resolution enhancement; spatial frequency; residual convolutional neural network (ResCNN)



Citation: Zhao, G.; Li, Q.; Chen, Z.; Lei, Z.; Xiao, C.; Huang, Y. Visibility Extension of 1-D Aperture Synthesis by a Residual CNN for Spatial Resolution Enhancement. *Remote Sens.* **2023**, *15*, 941. <https://doi.org/10.3390/rs15040941>

Academic Editor: Antonio Iodice

Received: 31 December 2022

Revised: 1 February 2023

Accepted: 6 February 2023

Published: 8 February 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

For sea-surface-temperature (SST) remote sensing, the demand for improving SST feature resolution has become urgent in recent years [1]. This demands improvement of the spatial resolution of sensors. Compared to infrared, due to the ability of microwaves to penetrate clouds, passive microwave measurements can provide more SST information during cloudy weather when infrared measurements are disturbed by cloud coverage; but microwave measurements have a lower spatial resolution [2].

To improve the spatial resolution of passive microwave radiometry for remote sensing of the Earth, the aperture synthesis (AS) radiometer was proposed [3]. The larger the size of an antenna array, the higher the spatial resolution the radiometer will achieve. However, limited by the carrying capacity of satellites, the antenna array size in an AS system cannot be too large, which limits the spatial resolution of the AS system. For example, the first AS in orbit, Microwave Interferometric Radiometer with Aperture Synthesis (MIRAS), was first designed to use a larger antenna array with 130 antennas [4], but was launched with a smaller antenna array consisting of 69 antennas [5].

In recent years, researchers have developed effective methods for AS radiometers to reconstruct brightness temperature (BT) images using neural networks to improve image quality [6–8]. However, no public literature can be found on enhancing the spatial resolution of an AS radiometer without increasing the size of the antenna array.

In this article, a method of visibility extension (VE) to enhance spatial resolution without enlarging the size of an antenna array is proposed for 1-D AS. The proposal of the VE method is inspired by two discoveries. First, the visibility (i.e., the spatial spectrum) of a scene is truncated in AS observations, and the cutoff spatial frequency (v_{cutoff}) is determined by the largest spacing of the antenna pairs in an AS system. Second, visibility generally distributes continuously; therefore, high-frequency visibility samples are related

to low-frequency visibility samples. If prior information about the visibility distribution of various scenes is learnt by a neural network, it is possible to extend the visibility.

Several types of neural network could be candidates for extending visibility. Long short-term memory (LSTM) networks have been mainly applied in the field of speech recognition [9,10] and natural language processing [11], and have achieved good performance in time-series forecasting [12,13]. Multi-layer perceptron (MLP) networks have been widely used, and can be designed and trained depending on specific applications [14,15]. Convolutional neural networks (CNNs) are widely used in various fields, such as computer vision [16], speech processing [17], face recognition [18], etc. The residual learning framework has solved the problem of degradation in very deep networks and has made them easier to be trained [19].

To find a satisfactory model, seven neural network models with various configurations are tested, including LSTM network models, MLP network models, CNN models, the residual convolutional neural network (ResCNN) models, and so on. Of all the models tested in this article, a ResCNN model outperforms the others. Because the visibility generally distributes continuously, the visibility samples at local adjacent spatial frequencies have a connection with each other, by which high-frequency visibility samples can be estimated from low-frequency visibility samples. CNN can effectively learn the local features, and with a residual learning framework, the problem of degradation can be suppressed. This is the reason that ResCNN can work effectively in visibility extension. Therefore, the ResCNN model is chosen to extend visibility in the VE method.

Thus, the main idea of the VE method is that visibility samples at spatial frequencies higher than v_{cutoff} are estimated by a ResCNN, and they are combined with low-frequency visibility samples obtained by an AS system to reconstruct the BT image of a scene, in order to enhance spatial resolution.

The main contributions of this article are: (1) visibility extension is proposed to enhance the spatial resolution of AS without enlarging the antenna array size; (2) ResCNN is proposed to perform visibility extension.

The VE method is introduced in Section 2. The simulation and experiment are described in Sections 3 and 4, respectively. Discussion is provided in Section 5, and the conclusion is presented in Section 6.

2. VE Method

2.1. Theory of Spatial Resolution Enhancement by Visibility Extension

The visibility of 1-D AS is defined below [3].

$$V(v) = \int_{-\pi}^{\pi} T_b(\varphi) e^{-j2\pi v \sin\varphi} d\varphi \quad (1)$$

where V is the visibility, v is the spatial frequency, T_b is the scene brightness temperature (BT), and φ is the azimuth angle. The visibility has the property of Hermitian symmetry: $V(-v) = V^*(v)$, where the superscript * denotes the conjugate.

In theory, the scene BT can be reconstructed by (2).

$$\hat{T}_b(\varphi) = \int_{-\infty}^{\infty} V(v) e^{j2\pi v \sin\varphi} dv \quad (2)$$

where \hat{T}_b is the reconstructed BT.

In fact, only finite and discrete samples of visibility can be obtained by an AS system; therefore, the scene BT can instead only be reconstructed by (3).

$$\hat{T}_b(\varphi) = \sum_{n=-L}^L V(v_n) e^{-j2\pi v_n \sin\varphi} \Delta v \quad (3)$$

where Δv is the minimum spacing of the antenna pairs in the AS system, L is the number of visibility samples at $v_n > 0$, and $v_n = n\Delta v$. As can be seen from (3), the visibility samples obtained by an AS system are truncated, and the cutoff frequency $v_{cutoff} = L\Delta v$.

If the visibility is extended, the spatial resolution of \hat{T}_b can be enhanced. An AS system can only obtain visibility samples at spatial frequencies lower than v_{cutoff} (denoted as a complex vector V_{low}). If the visibility samples at spatial frequencies higher than v_{cutoff} (denoted as a complex vector V_{high}) are added, the corresponding BT image reconstructed from the union of V_{low} and V_{high} will have higher spatial resolution and will be closer to the original BT, especially at the position where the BT changes rapidly, as illustrated in Figure 1. Because V_{high} is added and there are more high-frequency visibility samples used in BT image reconstruction, the spatial resolution is higher. This is the theory of enhancing spatial resolution by visibility extension.

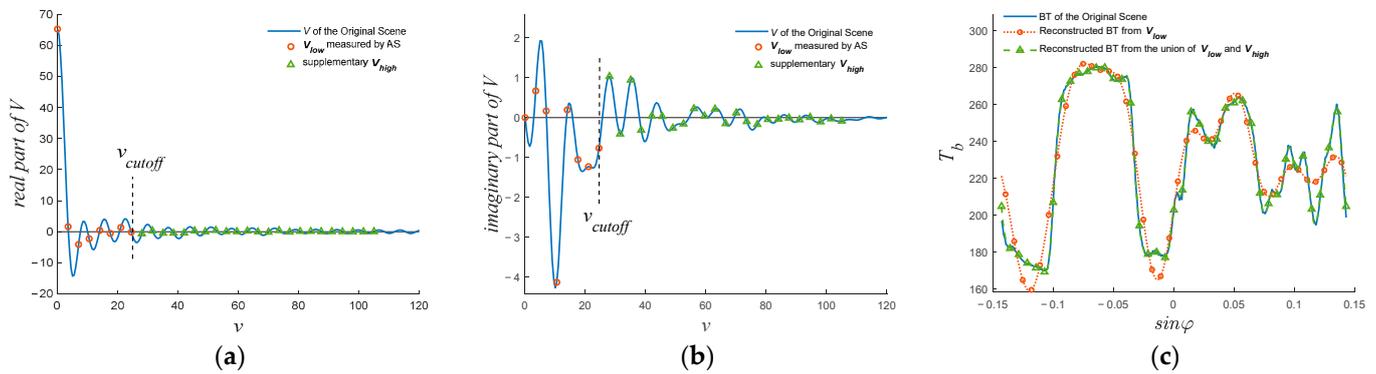


Figure 1. Illustration of visibility extension. Only the visibility at $v \geq 0$ is illustrated because of Hermitian symmetry. (a) The distribution of the real part of V . (b) The distribution of the imaginary part of V . (c) The distribution of the BT.

2.2. Procedure

The visibility of ordinary scenes distributes continuously; in other words, the visibility changes continuously. This implies that there is a relationship between the distribution of low- and high-frequency visibility. If the prior information of visibility distribution is learnt by a neural network, it is possible to estimate V_{high} by the neural network from the V_{low} obtained by an AS system.

In the VE method, a neural network is trained by the dataset generated from the visibility of various scenes to learn prior information about the visibility distribution, specifically to learn the relationship between the distribution of low- and high-frequency visibility. The trained neural network is used to estimate V_{high} according to V_{low} obtained by an AS system, and the estimate of V_{high} is denoted as V'_{high} . Then, V_{low} and V'_{high} are combined. Furthermore, the visibility samples at $v_n < 0$ are added, and the BT image of the scene is reconstructed by the Inverse Discrete Fourier Transform (IDFT). Because high-frequency visibility samples are added in the image reconstruction, the spatial resolution of the reconstructed BT image with the VE method is higher than that of the original reconstructed BT image.

The steps of the VE method are illustrated in Figure 2. There are five steps:

1. V_{low} is obtained through observation by an AS system.
2. V'_{high} is output from the trained neural network with the input complex vector V_{low} .
3. V_{low} and V'_{high} are combined.
4. The visibility samples at $v_n < 0$ are added according to $V(-v_n) = V^*(v_n)$.
5. The BT image of the scene is reconstructed by IDFT.

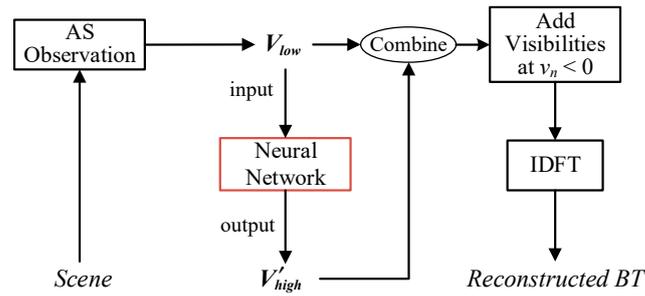


Figure 2. Steps of the VE method.

The most important step of the VE method is step 2. The dataset must be generated before the neural network is designed and trained. Then V'_{high} can be estimated. The dataset generation, neural network design, and network training need to be performed only once for an AS system.

2.3. Dataset

2.3.1. Procedure of Dataset Generation

The neural network needs to be trained by the dataset to determine the weights and biases. The procedure for dataset generation is listed below.

1. The BT image of a scene is generated. The scene can be an ideal scene or a natural scene. The BT image of an ideal scene is generated by simulation, while that of a natural scene is generated from satellite data.
2. The visibility samples $V(v_0) - V(v_{n+p-1})$ (where $v_j = j\Delta v, j = 0, 1, \dots, n + p - 1$) are acquired by a Discrete Fourier Transformation (DFT) from the scene BT according to (1).
3. The first n visibility samples, $V(v_0) - V(v_{n-1})$, are selected as the input complex vector V_{low} . The following p visibility samples, $V(v_n) - V(v_{n+p-1})$, are selected as the target complex vector V_{high} . V_{low} and V_{high} are combined into one sample, as illustrated in Figure 3.
4. Multiple samples are generated by repeating steps 1 to 3, and multiple samples compose the dataset.

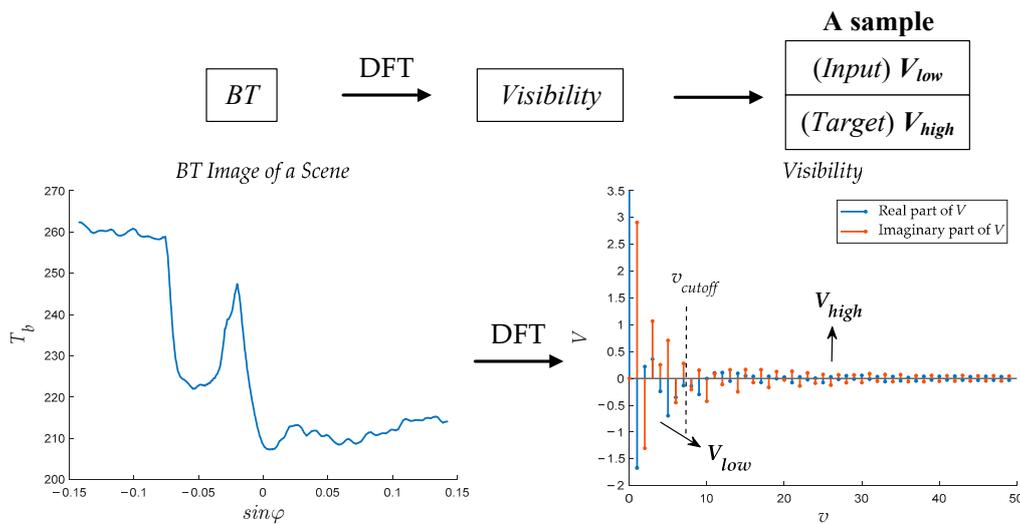


Figure 3. Generation of a sample in the dataset.

2.3.2. Determination of the Parameters in Dataset Generation

The parameter n in steps 2 and 3, i.e., the length of the input complex vector V_{low} , is determined by the AS system, specifically by the number of baselines of the AS system. For the MAS-V experiment system, which has eight baselines, mentioned in Section 4, $n = 8$.

When determining the parameter p in steps 2 and 3, i.e., the length of the output complex vector V'_{high} , the balance of the spatial resolution and the reconstruction error should be considered. When p is greater, the advantage is that there will be more high-frequency visibility samples used in the BT image reconstruction, so the spatial resolution of \hat{T}_b is higher. However, the disadvantage is that \hat{T}_b suffers larger reconstruction error. The reconstruction error is assessed by the root-mean-square error (RMSE) between the reconstructed BT and the original BT of a scene, as expressed in (4).

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^M (\hat{T}_b(m) - T_b(m))^2} \quad (4)$$

where \hat{T}_b is the reconstructed BT, T_b is the original BT, and M is the number of the pixels of the BT image. In this article, an average reconstruction error of approximately 40,000 scenes with different values of p was calculated, as illustrated in Figure 4. With an increase in p , the reconstruction error first reduced, but then increased. The value of p should be as large as possible with an acceptable increase in RMSE. Thus, according to Figure 4, $p = 42$ in this article, which is a trade-off.

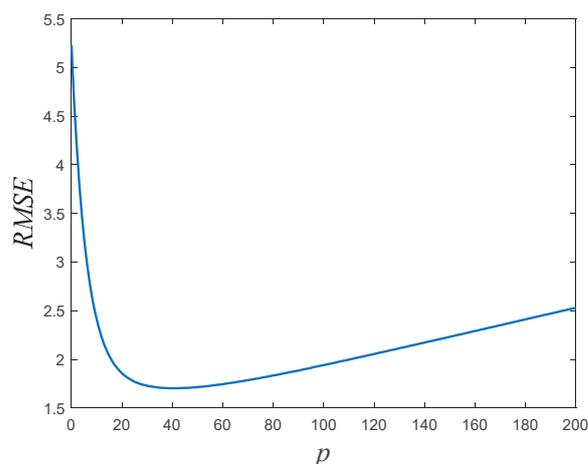


Figure 4. The trend of the reconstruction error with respect to p .

Besides this, in step 2, $\Delta v = 3.5$ according to the experiment system MAS-V mentioned in Section 4.

2.3.3. BT Images for the Dataset

According to the procedure of dataset generation described above, the first step is to generate BT images of the scenes. The BT images of numerous scenes are generated, in which two types of scene are involved: (a) ideal scenes, including scenes of point source and homogeneous scenes; (b) natural scenes, including scenes observed by the microwave radiometer carried by a satellite in orbit.

For the ideal scenes, the scenes of the point source at different locations in the field of view (FOV) are chosen, as well as homogeneous scenes with different widths and located at different positions in the FOV. BT images of ideal scenes are generated through simulation.

For the natural scenes, the scenes observed by the Scanning Microwave Radiometer (SMR) [20] onboard HaiYang-2B (HY-2B) are chosen. The SMR was designed to measure oceanic and atmospheric parameters, such as sea-surface temperature (SST), sea-surface

wind speed, water vapor, and cloud liquid water. It is a linearly polarized passive microwave radiometer that measures microwave radiation with the 6.925, 10.7, 18.7, 23.8, and 37 GHz channels, and it employs a conical scanning mechanism [20]. In each scan, the SMR generates a 150-pixel BT image of the Earth's surface. The SMR observation data obtained from the 37 GHz channels with horizontal and vertical polarization in February 2020 are selected and downloaded from the website (<https://osdds.nsoas.org.cn>) on 12 April 2021.

For the training dataset, 9000 BT images of ideal scenes and 46,685 BT images of natural scenes are selected; and for the testing dataset, 30,589 BT images of natural scenes are selected. Some examples of the selected BT images are shown in Figure 5.

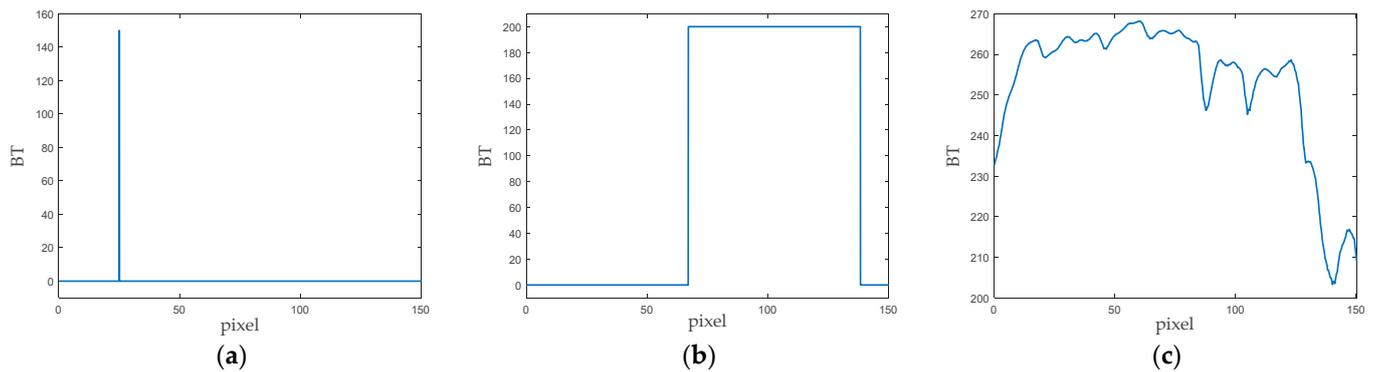


Figure 5. Examples of the BT images selected to generate the dataset. (a) The BT image of a point source. (b) The BT image of a homogeneous scene. (c) The BT image of a natural scene.

Thus, a training dataset containing 55,658 samples and a testing dataset containing 30,589 samples are generated.

2.4. Selection of Neural Network Model

In order to find the appropriate neural network model to learn the prior information of visibility distribution and extend the visibility, the LSTM network models, the MLP network model, the CNN models, and the residual convolutional neural network (ResCNN) models are tested in this article.

In Table 1, the parameter-search range of the architecture configuration for seven types of neural network model are presented. To generate the output with designed size, a layer (denoted as CN1), illustrated in Figure 6, composed of two convolutional layers convolving 42 and 84 filters of 1×1 and a global average pooling layer, is used as the last layer of the “ResCNN + CN1” and “CNN + CN1” models; a layer (denoted as FC1) composed of two fully connected layers with 512 and 84 nodes, respectively, is used as the last layer of the “ResCNN + FC1”, “CNN + FC1”, and “LSTM + FC1” models; and a fully connected layer containing 84 nodes (denoted as FC2) is used as the last layer of the “MLP + FC2” models. Because the last layers for the models are determined, the parameter-search range presented in Table 1 are for the layers excluding the last layer. In the table, the value of the parameter “ResCNN blocks” indicates the number of ResCNN blocks in the models; the value of the parameter “Filters” (“Nodes”) indicates the number of filters (nodes) contained in the first layer, and the number of filters (nodes) in the following layers is a multiple of the value of “Filters” (“Nodes”); the value of the parameter “Kernel size” is the size of the convolution kernel; the value of the parameter “Layers” indicates the number of layers (excluding the last layer) of the models; the value of the parameter “Units” indicates the number of features in the hidden state of LSTM.

Table 1. Parameter-search range of the architecture configuration for the seven types of neural network model.

Models	Parameters	Values
ResCNN + CN1 *	ResCNN blocks *	11, 12, 13
	Filters	128 to 512
	Kernel size	3, 5, 7
ResCNN + FC1	ResCNN blocks *	11, 12, 13
	Filters	128 to 512
	Kernel size	3, 5, 7
CNN + CN1 *	Layers	1 to 9
	Filters	64, 128, 256
	Kernel size	3, 5, 7
CNN + FC1	Layers	3 to 9
	Filters	128 to 512
	Kernel size	3, 5, 7
MLP + FC2	Layers	1 to 7
	Nodes	32, 64, 128
LSTM + FC1	Layers	1, 2, 4
	Units	32, 64, 128
LSTM	Layers	1, 2, 4
	Units	32, 64, 128

* The structures of the ResCNN block and CN1 are illustrated in Figure 6.

Other hyperparameters, such as the batch size, learning rate, momentum, and dropout, are searched simultaneously. The Bayesian hyperparameter-search method is used for hyperparameter searching. The Bayesian hyperparameter-search method uses a Gaussian process to model the relationship between the parameters and the test loss, and chooses parameters to optimize the probability of improvement.

To speed up the hyperparameter search, an early termination strategy called hyperband [21] is set to prevent poorly performing runs of the neural network models. Hyperband stopping evaluates whether a run should be stopped or permitted to continue at one or more pre-set iteration counts, which are set as the 4th, 12th, 36th, and 108th epochs in each run in this article.

For the seven types of neural network model, more than 1300 different configurations are tested in visibility extension. The training results of the most effective models in each type of neural network model are shown in Table 2. It can be seen from Table 2 that the ResCNN model outperforms the CNN model; the CNN model outperforms the MLP model; and the MLP model outperforms the LSTM model. The most effective of the “ResCNN + CN1” models is the most effective for visibility extension of all the tested models.

The distribution of visibility only depends on the distribution of the scene BT, so the high-frequency visibility samples have no memory of low-frequency visibility samples, which is also the reason for the poor performance of LSTM. In addition, for the general scenes, visibility distributes continuously. Therefore, there is a connection between the visibility samples at local adjacent spatial frequencies, which means that the high-frequency visibility samples are related to the low-frequency visibility samples. Compared with MLP, CNN can learn the local features of visibility distribution, which is why CNN outperforms MLP in visibility extension. For a deeper network, residual connections can reduce the impact of the degradation problem, so ResCNN performs better than other types of network model mentioned above.

Thus, the most effective of the “ResCNN + CN1” models (i.e., the VE-ResCNN) is introduced in this article and used to extend the visibility to enhance spatial resolution.

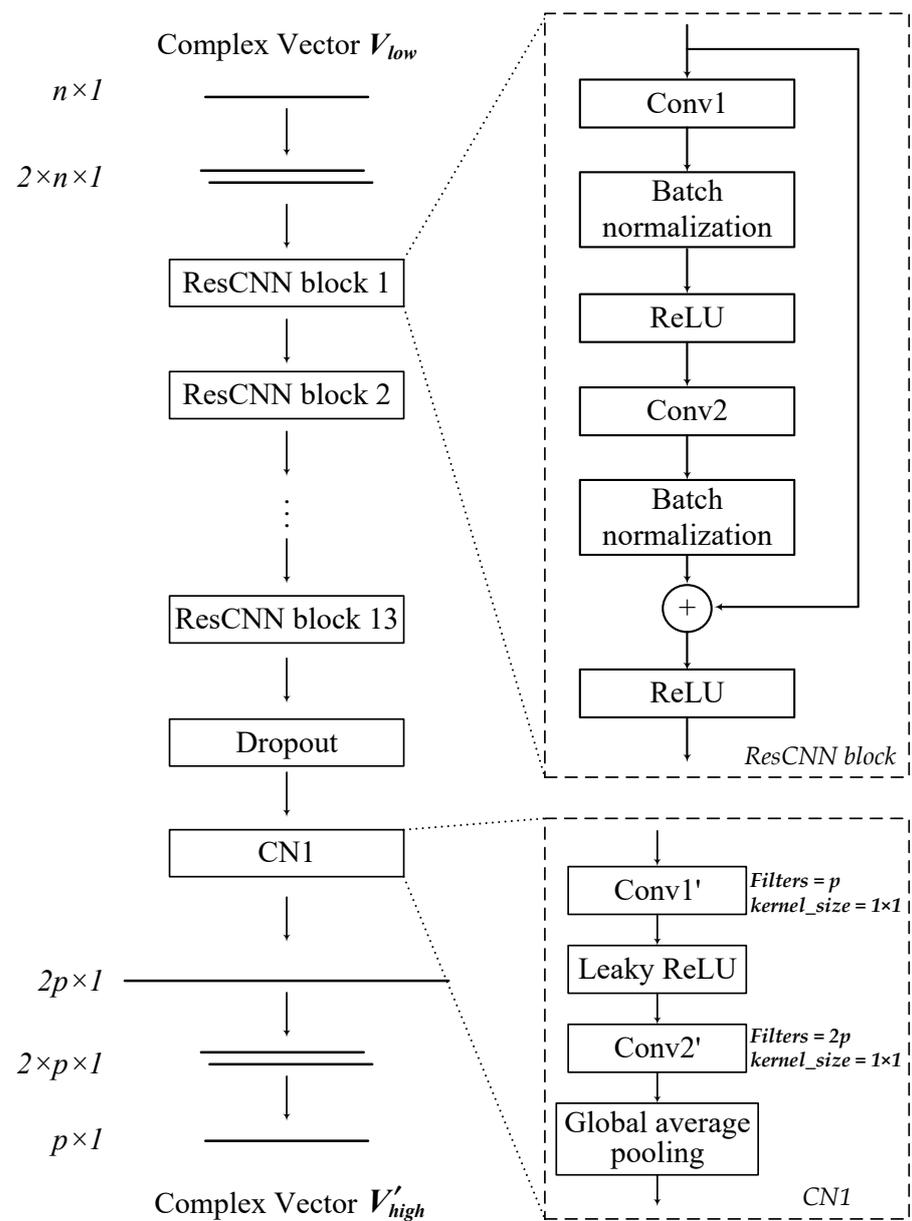


Figure 6. Architecture of the VE-ResCNN.

Table 2. The training results of the most effective models in the seven types of neural network model.

Models	Blocks	Total Layers	Trainable Parameters	Test Loss ($\times 10^{-3}$)	Training Time
* ResCNN + CN1	13	$13 \times 2 + 2$	182,084,166	2.863	4 h 28 min 16 s
ResCNN + FC1	12	$12 \times 2 + 2$	30,884,604	3.116	1 h 42 min 40 s
CNN + CN1	\	$6 + 2$	313,268,550	3.229	7 h 14 min 36 s
CNN + FC1	\	$8 + 2$	14,254,532	3.606	1 h 9 min 37 s
MLP + FC2	\	$7 + 1$	7,516,244	4.102	55 min 10 s
LSTM + FC1	\	$2 + 2$	53,460	9.699	53 min 44 s
LSTM	\	2	81,408	11.10	50 min 4 s

* The most effective model of “ResCNN + CN1” (i.e., the VE-ResCNN) is used as the neural network in the VE method.

2.5. VE-ResCNN

2.5.1. Architecture

The VE-ResCNN proposed in this article is composed of 13 ResCNN blocks, a dropout layer, and a last layer (denoted as CN1), as illustrated in Figure 6. The ResCNN blocks are used to extract the features of the input complex vector. The novelty of the residual network is in the use of the bypass pathway concept, which was employed in Highway Nets to address the problem of training a deeper network [19]. The dropout layer is used to prevent the neural network from overfitting [22]. In addition, the CN1 layer is used for dimensionality reduction to output the vector with the designed size.

As illustrated in Figure 6, the network input is V_{low} , which is a complex vector with n elements. Then, a matrix with two columns formed by the real and imaginary parts of V_{low} is input into the first ResCNN block of the VE-ResCNN.

The ResCNN block is a conventional convolutional neural network plus a residual connection. The block output, denoted as x_i , can be expressed as (5).

$$x_i = \text{ReLU}(F(x_{i-1}) + x_{i-1}) \quad (5)$$

where x_{i-1} is the block input (when $i = 0$, x_0 is the matrix with two columns transformed from the input complex vector V_{low}), $\text{ReLU} = \max(0, x)$ is the rectified linear unit function, and $F(x_{i-1})$ is the output of the feed-forward network in the block, expressed as (6).

$$F(x_{i-1}) = \text{BN}(\mathbf{w}_2 * \text{ReLU}(\text{BN}(\mathbf{w}_1 * x_{i-1} + \mathbf{b}_1))) + \mathbf{b}_2 \quad (6)$$

where BN is the batch-normalization process [23], \mathbf{w}_1 (\mathbf{w}_2) is the convolution weight matrix of the first (second) convolutional layer Conv1 (Conv2) in the ResCNN block, the symbol $*$ represents the convolution operation, and \mathbf{b}_1 and \mathbf{b}_2 are the biases.

The input of the dropout layer is the output of the ResCNN block 13. The dropout layer randomly sets some input elements to zero with a pre-set probability using samples from a Bernoulli distribution. This has been proven to be an effective technique for regularization and to prevent the co-adaptation of neurons [24].

Then, the output matrix of the dropout layer is input into the CN1 layer. The architecture of CN1 is inspired by previous works [25,26]. As illustrated in Figure 6, the kernel size of the convolutional layers (Conv1' and Conv2') in CN1 is 1×1 , and the number of filters (i.e., the convolutional weight matrices) in Conv1' and Conv2' are p and $2p$, respectively. Conv1' is followed by a leaky ReLU activation whose expression is $\text{LeakyReLU}(x) = \max(0, x) + f \times \min(0, x)$, where f is the leak factor. As a substitution for a fully connected layer, the global average pooling layer can significantly reduce the number of model parameters. A vector consisting of $2p$ elements is output by the CN1 layer. The output of the CN1 layer is reshaped into a two-column matrix with size $2 \times p \times 1$. Then, the two columns are combined as the real and imaginary parts of a complex vector. Thus, a complex vector with the size $p \times 1$ (i.e., V'_{high}) is obtained.

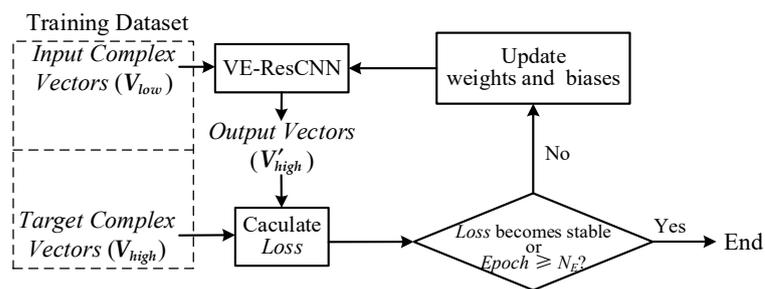
The architecture parameters of the VE-ResCNN are listed in Table 3. The dropout probability of the dropout layer is 0.413. In addition, the leak factor (f) of the leaky ReLU activation in CN1 is 0.01.

Table 3. Architectural parameters of the VE-ResCNN.

		Parameter	Value			Parameter	Value
ResCNN block 1	CN1	Filters	512	ResCNN block 8	CN1	Filters	1024
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	512		CN2	Filters	1024
		Kernel size	7×1			CN2	Kernel size
ResCNN block 2	CN1	Filters	512	ResCNN block 9	CN1	Filters	1024
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	512		CN2	Filters	1024
		Kernel size	7×1			CN2	Kernel size
ResCNN block 3	CN1	Filters	512	ResCNN block 10	CN1	Filters	1024
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	512		CN2	Filters	1024
		Kernel size	7×1			CN2	Kernel size
ResCNN block 4	CN1	Filters	512	ResCNN block 11	CN1	Filters	1536
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	512		CN2	Filters	1536
		Kernel size	7×1			CN2	Kernel size
ResCNN block 5	CN1	Filters	512	ResCNN block 12	CN1	Filters	1536
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	512		CN2	Filters	1536
		Kernel size	7×1			CN2	Kernel size
ResCNN block 6	CN1	Filters	1024	ResCNN block 13	CN1	Filters	1536
		Kernel size	7×1			CN1	Kernel size
	CN2	Filters	1024		CN2	Filters	1536
		Kernel size	7×1			CN2	Kernel size
ResCNN block 7	CN1	Filters	1024	CN1	CN1'	Filters	42
		Kernel size	7×1			CN1'	Kernel size
	CN2	Filters	1024		CN2'	Filters	84
		Kernel size	7×1			CN2'	Kernel size

2.5.2. Training

The training process of the VE-ResCNN is illustrated in Figure 7. The mini-batch gradient descent algorithm with a momentum is used to train the VE-ResCNN.

**Figure 7.** Training process.

When training the VE-ResCNN, a batch of samples is selected from the training dataset, and the input complex vectors (V_{low}) of the batch of samples are input into the VE-ResCNN. Then, the output complex vectors (V'_{high}) are compared to the target complex vectors (V_{high}) of the batch of samples, and the mean-square error is calculated as the loss function according to (7).

$$Loss = \frac{1}{S} \sum_{s=1}^S \left[\frac{1}{2p} \left(\left| \text{real}(V'_{high,s} - V_{high,s}) \right|^2 + \left| \text{imag}(V'_{high,s} - V_{high,s}) \right|^2 \right) \right] \quad (7)$$

where S is the number of samples in the batch, s is the serial number, p is the length of the complex vector $V'_{high,s}$ (or $V_{high,s}$), $V'_{high,s}$ is the output complex vector corresponding to the s th sample, $V_{high,s}$ is the target complex vector corresponding to the s th sample, $\text{real}(\cdot)$ is the real part of a complex vector, $\text{imag}(\cdot)$ is the imaginary part of a complex vector, and $|\cdot|$ represents the modulo of a vector.

After $Loss$ is calculated, the weights and biases are updated by the gradient-descent algorithm with momentum. The main disadvantage of gradient-descent learning algorithms is that they sometimes become stuck in a local minimum rather than a global minimum. Momentum is used along with the gradient-descent algorithm to solve this issue [27].

Then, another batch of samples from the training dataset are input into the VE-ResCNN to continue the training. When all the samples in the training dataset are used, an epoch is finished. Then, the next epoch begins.

A learning-rate scheduler is employed to adjust the learning rate based on the number of epochs during training. When $Loss$ stops decreasing for ten epochs, the learning rate is reduced by a factor of five. Learning-rate schedulers can often benefit model training.

Epoch by epoch, the weights and biases are updated iteratively to decrease $Loss$. When the epoch number reaches the pre-set number (N_E) or $Loss$ becomes stable, training is stopped. Thus, the training process is finished.

When training the VE-ResCNN, the batch size is set to 512; the initial learning rate is set to 0.015; the momentum is set to 0.356; and the maximum number of epochs (N_E) is set to 300 in this article.

In the training process, with an increase in epoch, both the $Loss$ of the training dataset (denoted as train-loss) and that of the testing dataset (denoted as test-loss) first decrease, and then remain relatively stable with little effect in further training, as illustrated in Figure 8.

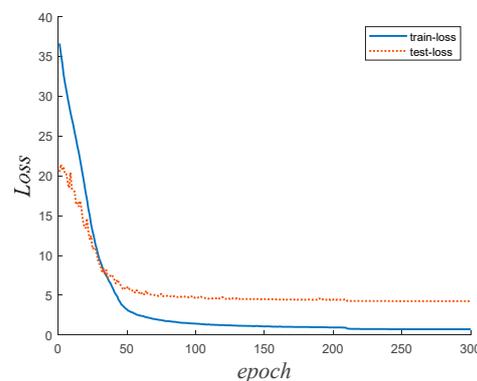


Figure 8. Convergence curve in VE-ResCNN training.

The VE-ResCNN is programmed on the Pytorch framework with Python 3.9, trained on a computer with one CPU (AMD Ryzen 7 2800 H) and one GPU (NVIDIA GeForce RTX 3070). It takes approximately 4.5 h to train the VE-ResCNN for 300 epochs.

The VE-ResCNN only needs to be trained once for an AS system. The trained VE-ResCNN is then used to extend the visibility to reconstruct the BT image, which is rapid, needing less than 0.02 s, as described in a later section.

3. Simulation

3.1. Simulation Procedure

The simulation procedure is illustrated in Figure 9. The following steps are included.

1. V_{low} of the selected scene is obtained by AS simulation according to Formula (1). The parameters of AS simulation are the same as the MAS-V experiment system mentioned in Section 4: the minimum antenna spacing is 3.5λ where λ is the wavelength, and the antenna array arrangement is {1, 2, 3, 4, 5, 6, 7, 8}.

2. The visibility samples at $v_n < 0$ corresponding to V_{low} are obtained by acquiring the conjugate of V_{low} . Moreover, the corresponding reconstructed BT image (denoted as observed BT) is obtained by IDFT.
3. V_{low} is input into the trained VE-ResCNN, and V'_{high} is output. Then, V_{low} and V'_{high} are combined to extend the visibility samples.
4. The visibility samples at $v_n < 0$ corresponding to the extended visibility samples are obtained by acquiring the conjugate of the extended visibility samples. Additionally, the corresponding reconstructed BT image (denoted as VE BT) is obtained by IDFT.

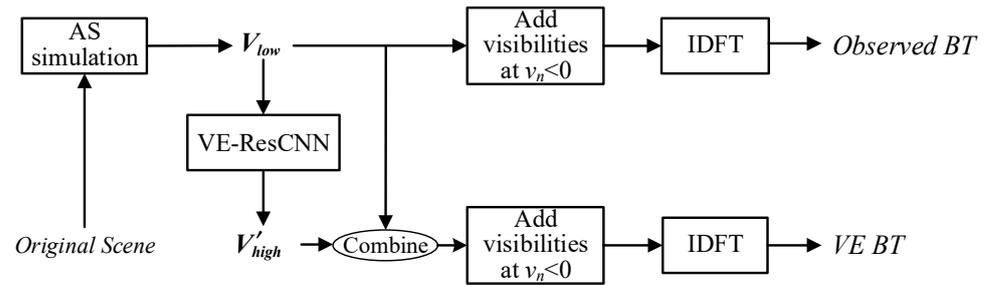


Figure 9. Simulation procedure.

3.2. Simulation Results

In order to show the effect of the spatial resolution enhancement, several scenes are picked out randomly from the scenes for generating the testing dataset, and the corresponding simulation results are illustrated in Figures 10 and 11.

The simulation results of noise sources are illustrated in Figure 10. A narrow pulse signal is used to simulate the noise signal radiated from a horn antenna, i.e., to simulate a noise source (see Figure 10a); and two narrow pulse signals are used to simulate dual noise sources (see Figure 10b,c). In Figure 10a, the half-power beamwidth (HPBW, corresponding to spatial resolution) of the observed BT of the single noise source is approximately 0.0231 (1.32°); however, for the VE BT of the single noise source, the BPBW is approximately 0.0079 (0.45°). In Figure 10b, the observed BT of the dual noise sources cannot separate the two noise sources, although the VE BT can. In Figure 10c, although the two noise sources are separated by both observed BT and VE BT, the beamwidth of VE BT is narrower than that of observed BT. These all indicate that the VE method can enhance spatial resolution.

It also can be seen from Figure 10 that with spatial resolution enhanced, the reconstruction error is also reduced. In Figure 10a–c, the RMSE of observed BT are 0.2171, 0.2271, and 0.3214, respectively; the RMSE of VE BT are 0.0742, 0.1299, and 0.1204, respectively, reduced by 65.82%, 42.80%, and 62.54%, respectively.

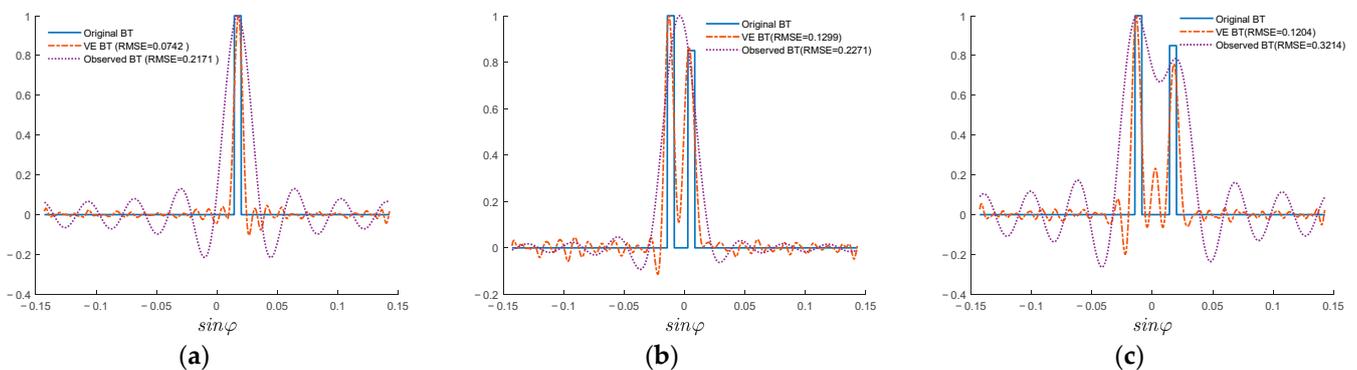


Figure 10. Simulation results of noise sources. (a) Single noise source. (b) Dual noise source 1. (c) Dual noise source 2.

The simulation results of natural scenes are shown in Figure 11. Compared with the lines of observed BT, the lines of VE BT are closer to the lines of original BT, especially where the BT changes rapidly. The results indicate that spatial resolution can be enhanced by the VE method.

With the spatial resolution enhanced, the reconstruction error of the scenes in Figure 11 is reduced, as listed in Table 4.

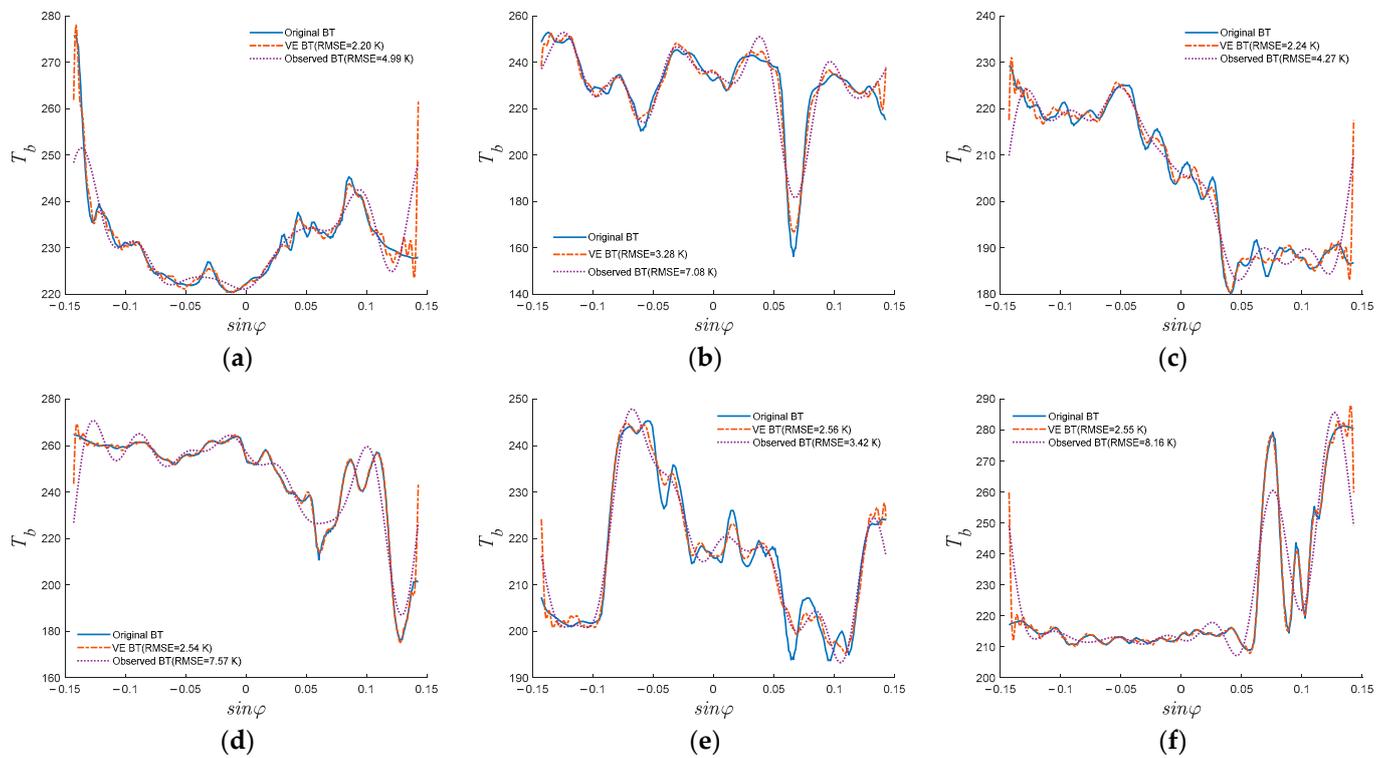


Figure 11. Simulation results of natural scenes. (a) Scene 1. (b) Scene 2. (c) Scene 3. (d) Scene 4. (e) Scene 5. (f) Scene 6.

Table 4. Reconstruction error of the scenes.

	RMSE of observed BT (K)	RMSE of VE BT (K)	Reduced
Scene 1	4.99	2.2	55.91%
Scene 2	7.08	3.28	53.67%
Scene 3	4.27	2.24	47.54%
Scene 4	7.57	2.54	66.45%
Scene 5	3.42	2.56	25.15%
Scene 6	8.16	2.55	68.75%

To further validate the effect of VE-ResCNN, all 30,589 samples in the testing dataset are tested in simulation. Then, the average of both the mean error and the RMSE of the reconstructed BT of all the scenes are calculated, as shown in Table 5. The average RMSE of the observed BT is 5.22 K; however, the average RMSE of the VE BT is 2.73 K (approximately 47.7% less). The reconstruction error of VE BT is smaller because VE BT is closer to the original BT with the spatial resolution enhanced. The average time for reconstructing one observed BT image is approximately 0.003 s, and the average time for reconstructing a VE BT image is approximately 0.019 s, as shown in Table 6.

Table 5. The average reconstruction error of the whole testing dataset.

	Mean Error (K)	RMSE (K)
Observed BT	0.11	5.22
VE BT	0.11	2.73

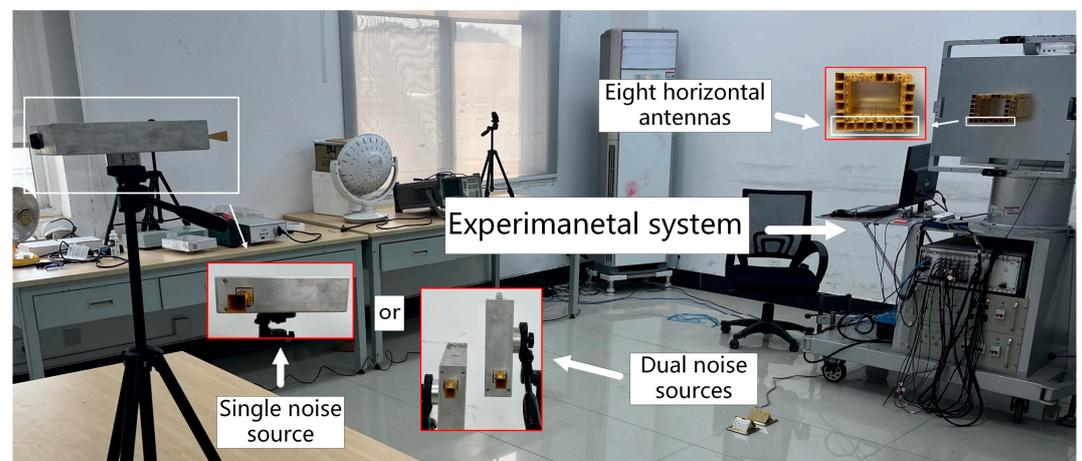
Table 6. The average time for reconstructing a BT image.

	Average Time (s)
Observed BT	0.003
VE BT	0.019

The simulation demonstrates that the spatial resolution of AS observation can be effectively enhanced by the VE method. Additionally, with enhanced spatial resolution, the reconstruction error also decreases.

4. Experiment

The performance of the VE method is also confirmed by an experiment in this section. The experiment system MAS-V [28,29] is used, which was developed by Huazhong University of Science Technology (HUST). MAS-V can be used for conventional AS experiments by removal of its reflectors, as illustrated in Figure 12. The antenna array arrangement of the system in the experiment is {1, 2, 3, 4, 5, 6, 7, 8}, and the minimum antenna spacing is 3.5λ .

**Figure 12.** Experimental scenario.

4.1. Experimental Procedure

The experiment procedure included four steps. Except for step 1, the other steps are the same as those of the simulation procedure described in Section 3.1.

In step 1, the scenes in the experiment are real, which is different from the simulation. There are six scenes used to conduct the experiment: a single noise source and two noise sources 6 cm, 7 cm, 8 cm, 10 cm, and 12 cm apart. When conducting the experiment, MAS-V is used to observe the scene, and V_{low} is obtained after error calibration.

4.2. Experimental Results

The results of the experiment are illustrated in Figure 13. As can be seen from Figure 13a, the HPBW of the observed BT of the single noise source is approximately 0.022 (1.26°); however, the HPBW of the VE BT of the single noise source is approximately 0.010 (0.58°), which indicates an enhancement in spatial resolution. As can be seen from Figure 13b,e,

the observed BT of dual noise sources cannot separate the two noise sources; however, the VE BT is able to, which also indicates an enhancement in spatial resolution. In addition, in Figure 13f, although the two noise sources can be separated by both the observed BT and the VE BT, the beamwidth of the two noise sources of the VE BT is apparently narrower than that of the observed BT. These all indicate an enhancement in spatial resolution. The results are listed in Table 7.

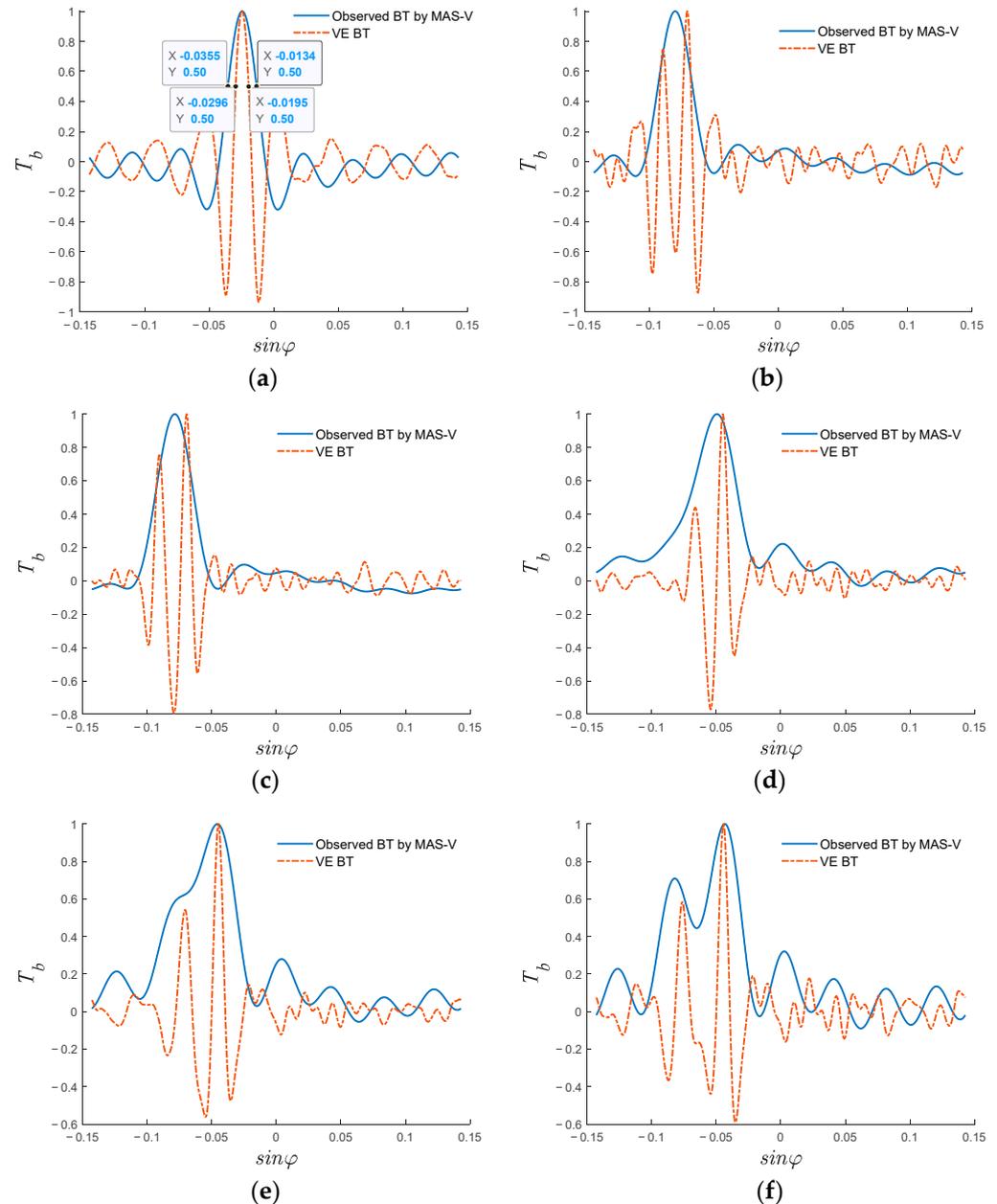


Figure 13. Experiment results. (a) Single noise source. (b) Dual noise sources 6 cm apart. (c) Dual noise sources 7 cm apart. (d) Dual noise sources 8 cm apart. (e) Dual noise sources 10 cm apart. (f) Dual noise sources 12 cm apart.

Table 7. Experimental results.

	Evaluation	Observed BT	VE BT
Single noise source	HPBW	1.26°	0.58°
Dual noise sources 6 cm apart	Separated or not	No	Yes
Dual noise sources 7 cm apart	Separated or not	No	Yes
Dual noise sources 8 cm apart	Separated or not	No	Yes
Dual noise sources 10 cm apart	Separated or not	No	Yes
Dual noise sources 12 cm apart	Beamwidth	Wider	Narrower

The time taken to reconstruct the BT image is the same as that in the simulation: 0.003 s for the observed BT image and 0.019 s for the VE BT image, shown in Table 6.

The experiment indicates that spatial resolution can be effectively enhanced by the VE method, which is consistent with the simulation.

5. Discussion

The simulation and experiment results demonstrate that the VE method can enhance the spatial resolution of 1-D AS. Moreover, the simulation results indicate that the VE method can reduce the reconstruction error by enhancing spatial resolution.

But there are still some disadvantages and limitations of the VE method proposed in this article.

The training dataset must contain a large number of samples in order to make the prior information comprehensive for the VE-ResCNN to learn. This necessitates that there must be enough BT images of scenes to generate the dataset.

The neural network needs to be trained first to learn the prior information of visibility distribution, which means extra computational costs. The training of VE-ResCNN costs nearly 4.5 h in this article, as shown in Table 2. Although it takes approximately 4.5 h to train the VE-ResCNN, the training process needs to be performed only once for an AS system.

However, it takes little time to perform the VE method after the neural network is trained. It takes less than 0.02 s to reconstruct a BT image, as shown in Table 6.

6. Conclusions

The spatial resolution of AS is proportional to the size of an antenna array, which is limited by the carrying capacity of satellites. To enhance the spatial resolution of 1D-AS without increasing the size of an antenna array, the VE method is proposed in this article.

The key idea of the VE method is that the high-frequency visibility samples of a scene are estimated by VE-ResCNN from the low-frequency visibility samples observed by an AS system, and the high- and low-frequency visibility samples are combined to reconstruct the BT image of a scene to enhance the spatial resolution. Only the visibility samples at spatial frequencies lower than v_{cutoff} can be obtained by the AS system in observations. With the additional high-frequency visibility samples estimated from the low-frequency visibility samples, the spatial resolution of the reconstructed BT can be improved.

The VE-ResCNN stands out from seven types of neural network model with over 1300 different configurations. The visibility of scenes generally distributes continuously, so the visibility samples at adjacent spatial frequencies have a connection with each other. ResCNN can learn the local features more effectively than other neural networks mentioned in this article. Therefore, after training by various scenes to learn the prior information of the visibility distribution of scenes, VE-ResCNN achieves good performance in visibility extension.

The simulation indicates that the spatial resolution of 1D-AS can be effectively enhanced by the VE method; and with spatial resolution enhanced, the reconstruction error decreases by approximately 47.7%. The single noise source experiment shows that the HPBW of the single noise source of the original observed BT and the VE BT are approx-

imately 1.26° and 0.58° , respectively. Moreover, the experiment with dual noise sources 6 cm, 7cm, 8cm, and 10 cm apart shows that the original observed BT cannot separate between the dual noise sources, although the VE BT is able to. In addition, in the experiment with dual noise sources 12 cm apart, although the two noise sources can be separated by both observed BT and VE BT, the beamwidth of VE BT is narrower than that of observed BT. These all demonstrate enhanced spatial resolution.

Author Contributions: Conceptualization, G.Z. and Q.L.; methodology and simulation, G.Z.; data curation, G.Z.; writing—original draft preparation, G.Z.; writing—review and editing, G.Z. and Q.L.; experiment, G.Z., Q.L., Z.C., Z.L., C.X. and Y.H.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 62271218 with 530,000 RMB. The project is to study the radiometry theory and method of tilted mirrored aperture synthesis from January 2023 to December 2026, under the leadership of Qingxia Li.

Data Availability Statement: Not applicable.

Acknowledgments: The HY-2B data used for dataset generation were obtained from <https://osdds.nsoas.org.cn>. The authors would like to thank National Satellite Ocean Application Service (NSOAS) for providing the data.

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

References

- O'Carroll, A.G.; Armstrong, E.M.; Beggs, H.M.; Bouali, M.; Casey, K.S.; Corlett, G.K.; Dash, P.; Donlon, C.J.; Gentemann, C.L.; Hoyer, J.L.; et al. Observational Needs of Sea Surface Temperature. *Front. Mar. Sci.* **2019**, *6*, 420.
- Guan, L.; Kawamura, H. Merging satellite infrared and microwave SSTs: Methodology and evaluation of the new SST. *J. Oceanogr.* **2004**, *60*, 905–912. [[CrossRef](#)]
- Ruf, C.; Swift, C.; Tanner, A.; Le Vine, D. Interferometric synthetic aperture microwave radiometry for the remote sensing of the Earth. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 597–611. [[CrossRef](#)]
- Torres, F.; Camps, A.; Bara, J.; Corbella, I.; Ferrero, R. On-board phase and modulus calibration of large aperture synthesis radiometers: Study applied to MIRAS. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 1000–1009. [[CrossRef](#)]
- McMullan, K.D.; Brown, M.A.; Martin-Neira, M.; Rits, W.; Ekholm, S.; Marti, J.; Lemanczyk, J. SMOS: The Payload. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 594–605. [[CrossRef](#)]
- Zhang, Y.; Ren, Y.; Miao, W.; Lin, Z.; Gao, H.; Shi, S. Microwave SAIR Imaging Approach Based on Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10376–10389. [[CrossRef](#)]
- Xiao, C.; Li, Q.; Lei, Z.; Zhao, G.; Chen, Z.; Huang, Y. Image Reconstruction with Deep CNN for Mirrored Aperture Synthesis. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–11. [[CrossRef](#)]
- Xiao, C.; Wang, X.; Dou, H.; Li, H.; Lv, R.; Wu, Y.; Song, G.; Wang, W.; Zhai, R. Non-Uniform Synthetic Aperture Radiometer Image Reconstruction Based on Deep Convolutional Neural Network. *Remote Sens.* **2022**, *14*, 2359. [[CrossRef](#)]
- Graves, A.; Mohamed, A.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649. [[CrossRef](#)]
- Sak, H.; Senior, A.; Beaufays, F. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. In Proceedings of the Interspeech 2014: 15th Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014.
- Gers, F.A.; Schmidhuber, J. LSTM recurrent networks learn simple context free and context sensitive languages. *IEEE Trans. Neural Netw.* **2001**, *12*, 1333–1340. [[CrossRef](#)]
- Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2018**, *323*, 203–213. [[CrossRef](#)]
- Lara-Benitez, P.; Carranza-Garcia, M.; Riquelme, J.C. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *Int. J. Neural Syst.* **2021**, *31*, 2130001. [[CrossRef](#)] [[PubMed](#)]
- Eleuteri, A.; Tagliaferri, R.; Milano, L. A novel information geometric approach to variable selection in MLP networks. *Neural Networks* **2005**, *18*, 1309–1318. [[CrossRef](#)]
- Bhattacharjee, K.; Pant, M. Hybrid particle swarm optimization-genetic algorithm trained multi-layer perceptron for classification of human glioma from molecular brain neoplasia data. *Cogn. Syst. Res.* **2019**, *58*, 173–194. [[CrossRef](#)]
- Fang, W.; Love, P.E.; Luo, H.; Ding, L. Computer vision for behaviour-based safety in construction: A review and future directions. *Adv. Eng. Informatics* **2019**, *43*, 100980. [[CrossRef](#)]

17. Palaz, D.; Magimai-Doss, M.; Collobert, R. End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition. *Speech Commun.* **2019**, *108*, 15–32. [[CrossRef](#)]
18. Li, H.-C.; Deng, Z.-Y.; Chiang, H.-H. Lightweight and Resource-Constrained Learning Network for Face Recognition with Performance Optimization. *Sensors* **2020**, *20*, 6114. [[CrossRef](#)]
19. 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 (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
20. Yu, R.; Lu, H.; Li, S.; Zhu, D.; Zhou, W.; Dang, P.; Wang, C.; Jin, X.; Lv, R.; Li, H. Instrument Design and Early In-Orbit Performance of HY-2B Scanning Microwave Radiometer. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [[CrossRef](#)]
21. Li, L.; Jamieson, K.; Desalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J. Mach. Learn. Res.* **2016**, *18*, 1–52.
22. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
23. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proc. Mach. Learn. Res.* **2015**, *37*, 448–456.
24. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
25. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
26. Li, Z.; Wang, S.H.; Fan, R.R.; Cao, G.; Zhang, Y.D.; Guo, T. Teeth category classification via seven-layer deep convolutional neural network with max pooling and global average pooling. *Int. J. Imaging Syst. Technol.* **2019**, *29*, 577–583.
27. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 1–74. [[CrossRef](#)]
28. Dou, H.; Lang, L.; Guo, W.; Gui, L.; Li, Q.; Chen, L.; Bi, X.; Wu, Y.; Lei, Z.; Li, Y.; et al. Initial Results of Microwave Radiometric Imaging with Mirrored Aperture Synthesis. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8105–8117. [[CrossRef](#)]
29. Li, Q.; Dou, H.; Gui, L.; Chen, L.; Chen, K.; Wu, Y.; Lei, Z.; Li, Y.; Lang, L.; Guo, W. MAS-V: Experimental System of Mirrored Aperture Synthesis at V BAND. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 1013–1016. [[CrossRef](#)]

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.