

Article

A Multi-Target Detection Method Based on Improved U-Net for UWB MIMO Through-Wall Radar

Jun Pan ^{1,2}, Zhijie Zheng ³ , Di Zhao ^{1,2} , Kun Yan ^{1,2} , Jinliang Nie ^{1,2,*}, Bin Zhou ^{1,2} and Guangyou Fang ^{1,2}

¹ GBA Branch of Aerospace Information Research Institute, Chinese Academy of Sciences, Guangzhou 510700, China; panjun@aircas.ac.cn (J.P.); zhaodi01@aircas.ac.cn (D.Z.); a@casysk.cn (K.Y.); zhb@mail.ie.ac.cn (B.Z.); gyfang@mail.ie.ac.cn (G.F.)

² Guangdong Provincial Key Laboratory of Terahertz Quantum Electromagnetics, Guangzhou 510700, China

³ Key Laboratory of Electromagnetic Radiation and Sensing Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China; zhengzhijie19@mails.ucas.ac.cn

* Correspondence: niejl@aircas.ac.cn; Tel.: +86-(020)-82035388 (ext. 8001)

Abstract: Ultra-wideband (UWB) multiple-input multiple-output (MIMO) through-wall radar is widely used in through-wall human target detection for its good penetration characteristics and resolution. However, in actual detection scenarios, weak target masking and adjacent target unresolving will occur in through-wall imaging due to factors such as resolution limitations and differences in human reflectance, which will reduce the probability of target detection. An improved U-Net model is proposed in this paper to improve the detection probability of through-wall targets. In the proposed detection method, a ResNet module and a squeeze-and-excitation (SE) module are integrated in the traditional U-Net model. The ResNet module can reduce the difficulty of feature learning and improve the accuracy of detection. The SE module allows the network to perform feature recalibration and learn to use global information to emphasize useful features selectively and suppress less useful features. The effectiveness of the proposed method is verified via simulations and experiments. Compared with the order statistics constant false alarm rate (OS-CFAR), the fully convolutional networks (FCN) and the traditional U-Net, the proposed method can detect through-wall weak targets and adjacent unresolving targets effectively. The detection precision of the through-wall target is improved, and the missed detection rate is minimized.

Keywords: through-wall radar; multiple targets; target detection; U-Net; ultra-wideband (UWB); multiple-input multiple-output (MIMO)



Citation: Pan, J.; Zheng, Z.; Zhao, D.; Yan, K.; Nie, J.; Zhou, B.; Fang, G. A Multi-Target Detection Method Based on Improved U-Net for UWB MIMO Through-Wall Radar. *Remote Sens.* **2023**, *15*, 3434. <https://doi.org/10.3390/rs15133434>

Academic Editor: Roberto Orosei

Received: 25 April 2023

Revised: 30 June 2023

Accepted: 5 July 2023

Published: 6 July 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

Ultra-wideband (UWB) radar is widely used in emergency rescue [1–3], geological exploration [4,5], archaeological detection [6], ice detection [7,8], etc. The working frequency of UWB radar is generally 400 MHz~10 GHz [9]. The definition of UWB is that the absolute bandwidth is higher than 1 GHz or the relative bandwidth is larger than 25% [10]. Therefore, UWB radar has good penetration characteristics and high-range resolution [11–14]. The UWB multiple-input multiple-output (MIMO) through-wall radar composed of a UWB radar and a MIMO array can penetrate obstacles for real-time imaging [15–18], which is an important technology in the detection of through-wall targets.

The target detection results of through-wall imaging can be used to determine the location and number of people in a closed room [19,20]. However, in actual detection scenarios, the imaging of strongly reflective targets will mask weakly reflective targets due to the difference in reflectivity of different human targets [21,22]. In addition, unresolving adjacent targets will appear in the imaging results when the distance between targets is relatively close due to the limited array aperture of the through-wall radar. The above factors will reduce the detection probability of through-wall targets.

Constant false alarm rate (CFAR) detectors are often used to detect through-wall targets [23,24], mainly including cell average CFAR (CA-CFAR) detectors and ordered statistical CFAR (OS-CFAR) detectors [25]. In CA-CFAR, the background noise and clutter power are estimated via the sliding window and mean method [26]. In the OS-CFAR, the signal samples in the reference window are sorted by size, and then the single rank of ordered statistics is used instead of the arithmetic mean to estimate the background noise and clutter power [26]. Therefore, the performance of the OS-CFAR detector is better than that of the CA-CFAR detector in the through-wall target detection. In addition, some classical segmentation methods can also be used for target detection. Classical segmentation methods include threshold-based segmentation methods, region-based segmentation methods, and edge detection-based segmentation methods [27]. The threshold-based segmentation method has been introduced into the field of through-wall target detection for its good segmentation effect and high efficiency, such as the OTSU threshold segmentation algorithm [28]. The maximum inter-class variance is used in OTSU to determine the threshold automatically. The threshold is the best when the inter-class variance is the largest, and the image is segmented with this threshold. However, the above-mentioned through-wall target detection methods only consider the intensity characteristics of the target, and it is difficult to detect through-wall weak targets and adjacent unresolving targets.

In recent years, methods of optical image processing have also been used for through-wall target detection. They can learn high-level and robust features from the training data automatically compared to traditional target detection methods. Zheng et al. used the convolutional neural network (CNN) to detect the position and pose of through-wall human targets [29]. The denoising self-encoder is used in [30] to detect the posture of the moving human behind the wall. Liu et al. used robust principal component analysis (RPCA) to detect through-wall targets [31], which can eliminate well the influence of wall clutter. Support vector machine (SVM) and K-Means techniques are also used in through-wall target detection, but they are only suitable for high-SNR environments [32]. U-Net is used for through-wall super-resolution imaging to complete target detection in [33], but it has not been used for through-wall weak target detection. Fully convolutional networks (FCN) are used in [34] for through-wall multi-target detection, which can detect through-wall weak targets, but the recovery of adjacent targets is not perfect. In summary, it is difficult for existing methods to extract through-wall weak targets and adjacent unresolving targets simultaneously.

An improved U-Net model is proposed in this paper to detect through-wall multiple human targets accurately and eliminate weak target masking and adjacent target unresolving. In the proposed detection method, a ResNet module and a squeeze-and-excitation (SE) module are integrated in the traditional U-Net model. The ResNet module can reduce the difficulty of feature learning and improve the accuracy of detection. The SE module allows the network to perform feature recalibration and learn to use global information to emphasize useful features selectively and suppress less useful features. The ability to detect through-wall weak targets and adjacent targets can be improved by introducing these two modules into the U-Net model.

This paper is organized as follows. Section 2 gives the principle of through-wall imaging and analyzes the phenomena of through-wall weak target masking and adjacent target unresolving. Section 3 describes the model of the proposed through-wall target detection method based on the improved U-Net. The establishment of the dataset, the evaluation metrics of the model and the simulation results are given in Section 4. In Section 5, the proposed method is verified via experiments. Finally, Section 6 concludes the paper.

2. Problem Analysis

2.1. Through-Wall Imaging

The scene of through-wall imaging is shown in Figure 1. The through-wall radar is close to the wall, and the target is located on the other side of the wall. Assuming that the

2.2. Unresolving and Masking Phenomena in Through-Wall Multi-Target Imaging

The range resolution, Δ_R , and azimuth resolution, Δ_C , in through-wall imaging are as follows [36]:

$$\Delta_R = \frac{c}{2B}, \quad (4)$$

$$\Delta_C = \frac{\lambda R}{(L_T + L_R) \cos^2 \theta}, \quad (5)$$

where B is the bandwidth of the radar signal, λ is the wavelength corresponding to the center frequency of the radar signal, and L_T and L_R are the aperture lengths of the transmitting and receiving arrays, respectively. R is the range between the target and the center of the radar, and θ is the angle at which the target deviates from the radar's normal angle. It can be seen that the range resolution is related to the bandwidth. Azimuth resolution is limited by the range of the target, the wavelength of the signal and the aperture of the radar.

According to Formula (5), the azimuth resolution increases with the distance of the target when the radar aperture is fixed, which will cause unresolving between two adjacent targets at a long distance in the imaging results. In addition, the amplitude of human echoes will be affected by factors such as the body shape of human, reflectivity, micro-motion and distance when there are multiple human targets. The distance factor can be compensated for via the method of time gain in the distance direction [37]. However, other influencing factors cannot be predicted and eliminated in advance, which will cause masking among multiple targets.

Figure 2 is the simulation schematic diagram of through-wall multi-target imaging. The MIMO through-wall radar used in the simulation model includes two transmitting antennas and six receiving antennas, which are arranged in a sparse and non-redundant line array. The center frequency and bandwidth of the radar are both 1 GHz. There are five point targets in the simulation experiment, with coordinates of (0.3 m, 1 m), (−2 m, 2.5 m), (0.2 m, 4.8 m), (0.8 m, 4.98 m) and (−3.2 m, 8 m). In order to represent the difference in reflection intensity of different targets, the reflectance of each target is different, and the value ranges from 0.2 to 1. The signal-to-noise ratio (SNR) of the simulated echo signal is generated between 3 dB and 10 dB randomly.

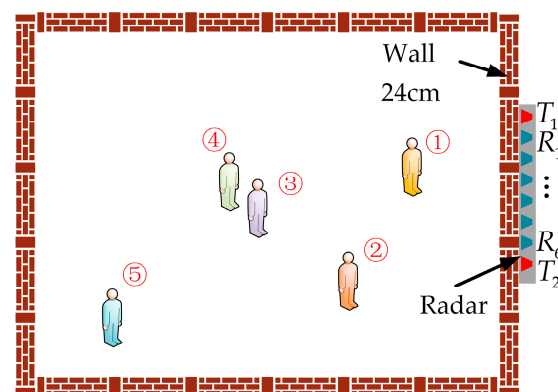


Figure 2. Imaging of through-wall multiple targets.

The time gain method is used to enhance the response of the target at a long distance, and then the improved Kirchhoff algorithm introduced in Section 2.1 is used for imaging. Figure 3 shows the through-wall 2D imaging results. It can be seen that the image of target 2 is weak and is almost masked by several other strong targets. The distance from target 3 to the radar is 4.8 m, and the azimuth resolution of this distance can be calculated as 0.65 m according to the azimuth resolution formula. The azimuth interval between target 3 and target 4 is 0.6 m, which is smaller than the theoretical resolution. It will cause an unresolving phenomenon in the imaging of target 3 and target 4, which will bring difficulties to target detection.

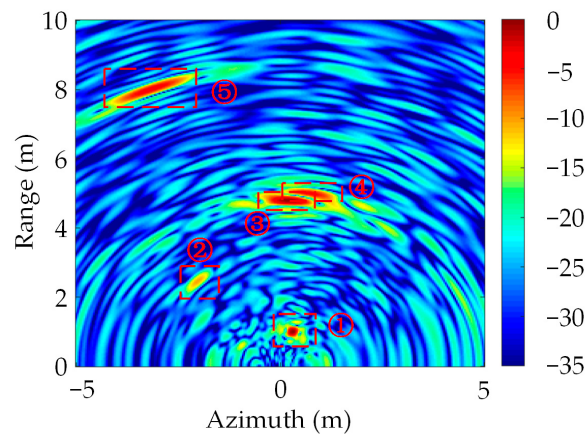


Figure 3. Two-dimensional imaging result.

3. Target Detection Method Based on Improved U-Net

In order to detect through-wall multiple human targets accurately and eliminate the problem of weak target masking and adjacent target unresolving, this paper proposes a multi-target detection method based on the improved U-Net. The designed model integrates the ResNet module and the SE module in the traditional U-Net model to improve its ability to detect through-wall weak targets and adjacent unresolving targets.

3.1. U-Net Model

U-Net is a network proposed for image segmentation based on FCN [38]. On the basis of FCN, the decoding structure of U-Net is modified, and a U-shaped structure network with a symmetrical contraction path and expansion path is formed through multi-channel convolution [39]. Compared to the FCN, skip connections are added to the expansion path in U-Net to incorporate more image features, and a good segmentation effect can also be obtained in the case of a small number of samples. It can solve well the problem of few data samples in the field of through-wall multi-target detection. The U-Net model can reconstruct the contours of adjacent and weak targets at the pixel level and detect the remaining targets accurately.

The U-Net model consists of contraction modules, expansion modules and skip connections [39]. The classic U-Net model generally includes four-layer contraction modules, four-layer expansion modules, and a one-layer output layer. The single-layer contraction module consists of two convolutional layers, one rectified linear unit (ReLU) and a maximum pooling layer. The convolution kernel of the convolutional layer is 3×3 , the convolution kernel of the pooling layer is 2×2 and the stride is 2. The number of feature maps is doubled after each contraction module, and the maximum number is 1024. The single-layer expansion module consists of a 2×2 upconvolution layer (which can halve the number of channels), a skip connection (with the same size feature map as the contraction path), two convolutions (the convolution kernel is 3×3) and the ReLU. Finally, the multi-channel feature map is mapped to the segmentation result via 1×1 convolution at the output of the network.

Compared with the feature map of the shallow network, the target details in the feature map of the deep network are largely lost. In the U-Net model, this problem is well-solved by splicing the shallow and deep feature maps through the skip connection step. Therefore, the U-Net model can achieve high-resolution image segmentation.

3.2. ResNet Module

The designed model needs to have the ability to identify adjacent unresolving targets and weak targets in through-wall imaging. In target detection tasks, deeper networks can provide more accurate detection results. However, in network training based on gradient descent, the problems of gradient explosion and vanishing gradients occur in deep

networks. Gradient clipping can solve the problem of gradient explosion. The problem of gradient disappearance can be initially alleviated via a weight initialization strategy and the batch normalization method, but it cannot guarantee that the training error of the model will decrease with the increase in network depth. The proposal of the ResNet module can solve well the problem of deep network training.

The structure diagram of ResNet is shown in Figure 4a, which contains 2 convolutional layers and 1 skip connection [40]. The relationship between the input and output of the ResNet module can be expressed as follows:

$$F(\mathbf{X}) = \tilde{\mathbf{X}} - \mathbf{X}, \quad (6)$$

where $\mathbf{X}$ is the input of the network, $\tilde{\mathbf{X}}$ is the output of the network and $F(\mathbf{X})$ is the residual map. The output $\tilde{\mathbf{X}}$ of the network can be re-expressed as $F(\mathbf{X}) + \mathbf{X}$ through $\mathbf{X}$ and $F(\mathbf{X})$. The feature map in ResNet is not learned directly, but the residual of the feature map is learned, which is called the learning residual map. It has been proven that optimizing the residual map is easier than learning the original feature map directly [40]. A skip connection is added to ResNet. It is assumed that if only the output of the shallow layer is used for identity mapping (i.e., $F(\mathbf{X}) = 0$) and input to the deep layer, then network degradation will not occur when the network is deepened. The problem of vanishing gradients is solved. Therefore, in through-wall weak target and adjacent target detection, the ResNet module can reduce the difficulty of feature learning and improve the accuracy of detection.

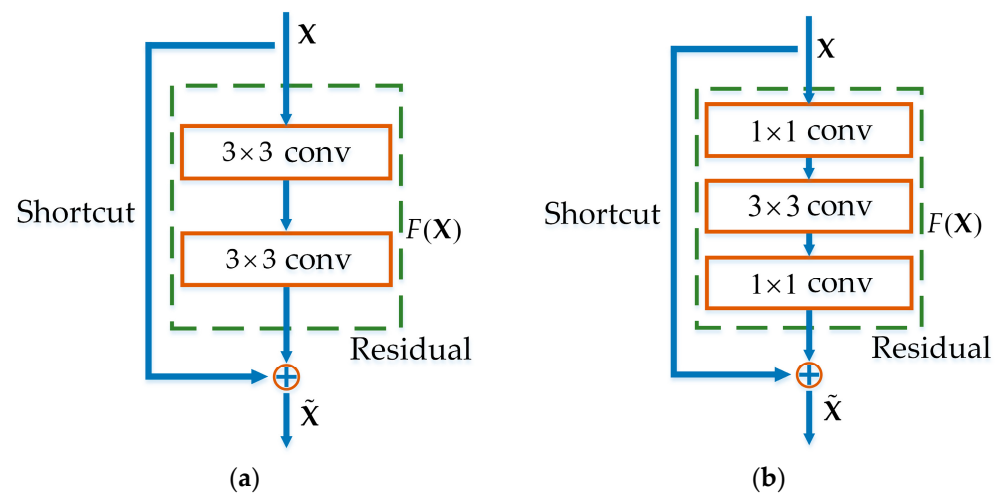


Figure 4. Structure of ResNet. (a) Regular ResNet module; (b) Bottleneck ResNet module.

A regular ResNet module generally consists of two 3×3 convolutional layers, which is accompanied by a large amount of calculation. In this paper, the bottleneck ResNet network is selected to reduce the calculations. The network consists of 1×1 , 3×3 and 1×1 convolutional layers, and the structure is shown in Figure 4b. When the input and output are both 256-dimensional, the parameters of the regular ResNet network and the bottleneck ResNet network are 1,179,648 and 69,632, respectively.

3.3. SE Module

In this paper, the SE module is used to capture the relationship of feature channels and recalibrate the response of channel features adaptively by modeling the interdependence between feature channels, which can enhance useful features while suppressing useless features effectively.

SE contains squeeze and excitation [41]. Taking the SE module embedded in Inception as an example, the module structure is shown in Figure 5, where $\mathbf{X}$ is the input of the

network, the size of the image is $H \times W$, and the number of channels is C . The SE module is mainly implemented through the following three steps:

(1) Squeeze: Squeeze compresses the features of each channel, which is achieved by averaging the feature maps of each channel. The formula is as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \mathbf{x}_c(i, j), \quad (7)$$

where z_c is the compressed feature distribution, $\mathbf{x}_c$ is the feature map of the c -th channel, and i and j are the two-dimensional indices of the pixel, respectively. The number of channels remains unchanged after being processed via Squeeze. In this paper, Squeeze is implemented through global pooling, so as to compress the feature map and obtain the global receptive field.

(2) Excitation: Excitation is mainly realized through the fully connected layer and the activation layer. First, the number of channels becomes $1/r$ of the input after the fully connected layer; in this paper, $r = 64$. The number of channels is expanded to C after the second fully connected layer. This step makes the channels more uncorrelated while reducing the computations. Finally, the 0~1 weight is generated through the activation layer sigmoid. Excitation generates corresponding weights for each channel, and can be adjusted through learning.

(3) Reweight: Reweight means multiplying the weights generated in the previous steps to the corresponding channel features. The weight corresponding to a channel represents the importance of the channel.

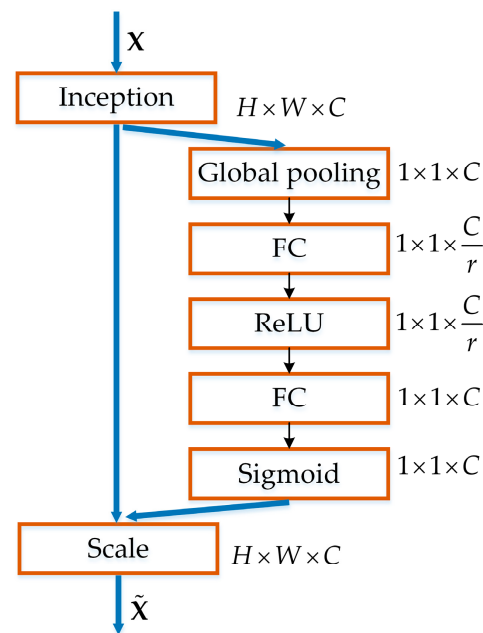


Figure 5. SE module.

$\tilde{\mathbf{X}}$ is the output of the network, and “Scale” is the multiplication operation between the input feature map and the weight. The SE module allows the network to perform feature recalibration and learn to use global information to emphasize useful features selectively and suppress less useful features. Therefore, the ability to extract features is improved. The SE module can extract the key features of the target, which can be well-used for the accurate detection of weak targets and adjacent unresolving targets in through-wall multi-target imaging.

3.4. Improved U-Net Model

The purpose of the improved U-Net model is to deal with the problems of weak target masking and adjacent target unresolving in through-wall multi-target imaging, so as to detect the position and number of targets accurately and reduce the probability of missed detection. The ability of the network to extract useful features and suppress useless features is very important. The SE module is embedded into the ResNet module and combined into a SE-ResNet module, which can extract target features effectively, as shown in Figure 6.

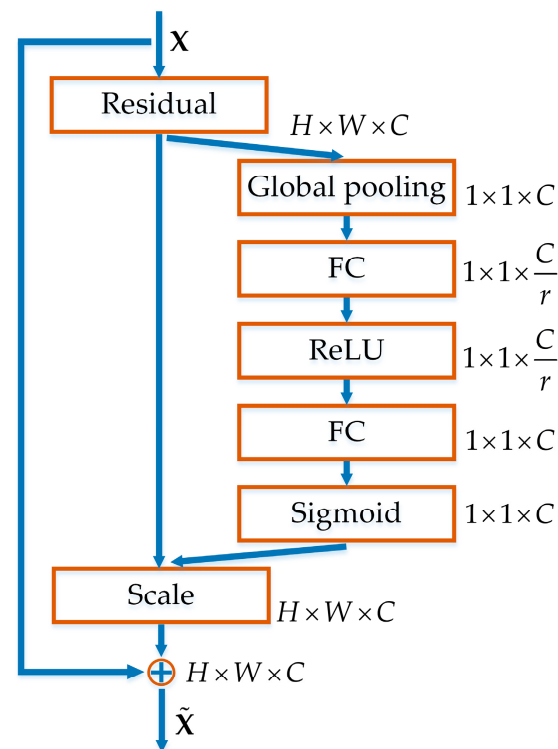


Figure 6. SE-ResNet module.

Then, the SE-ResNet module is further embedded into the traditional U-Net model to form an improved model that can identify through-wall multiple targets at the pixel level. The specific embedding process is to use the SE-ResNet module designed in Figure 6 to replace the convolution module in each layer of the traditional U-Net network. The improved U-Net model contains 36 layers, and the detailed specifications are shown in Table 1, which gives the specifications of the convolution kernel, the step size, and the size of the output feature map specifically.

Table 1. Detailed specifications of the improved U-Net.

SE-ResNet Block	Type	Filter	Stride	Output Size
SE-ResNet block 1	conv	1×1	1	$256 \times 256 \times 64$
	conv	3×3	1	$256 \times 256 \times 64$
	conv	1×1	1	$256 \times 256 \times 64$
SE-ResNet block 2	max pool	2×2	2	$128 \times 128 \times 64$
	conv	1×1	1	$128 \times 128 \times 128$
	conv	3×3	1	$128 \times 128 \times 128$
	conv	1×1	1	$128 \times 128 \times 128$
	max pool	2×2	2	$64 \times 64 \times 128$

Table 1. *Cont.*

SE-ResNet Block	Type	Filter	Stride	Output Size
SE-ResNet block 3	conv	1×1	1	$64 \times 64 \times 256$
	conv	3×3	1	$64 \times 64 \times 256$
	conv	1×1	1	$64 \times 64 \times 256$
	max pool	2×2	2	$32 \times 32 \times 256$
SE-ResNet block 4	conv	1×1	1	$32 \times 32 \times 512$
	conv	3×3	1	$32 \times 32 \times 512$
	conv	1×1	1	$32 \times 32 \times 512$
	max pool	2×2	2	$16 \times 16 \times 512$
SE-ResNet block 5	conv	1×1	1	$16 \times 16 \times 1024$
	conv	3×3	1	$16 \times 16 \times 1024$
	conv	1×1	1	$16 \times 16 \times 1024$
	up-conv	2×2	1	$32 \times 32 \times 512$
SE-ResNet block 6	conv	1×1	1	$32 \times 32 \times 512$
	conv	3×3	1	$32 \times 32 \times 512$
	conv	1×1	1	$32 \times 32 \times 512$
	up-conv	2×2	1	$64 \times 64 \times 256$
SE-ResNet block 7	conv	1×1	1	$64 \times 64 \times 256$
	conv	3×3	1	$64 \times 64 \times 256$
	conv	1×1	1	$64 \times 64 \times 256$
	up-conv	2×2	1	$128 \times 128 \times 128$
SE-ResNet block 8	conv	1×1	1	$128 \times 128 \times 128$
	conv	3×3	1	$128 \times 128 \times 128$
	conv	1×1	1	$128 \times 128 \times 128$
	up-conv	2×2	1	$256 \times 256 \times 64$
SE-ResNet block 9	conv	1×1	1	$256 \times 256 \times 64$
	conv	3×3	1	$256 \times 256 \times 64$
	conv	1×1	1	$256 \times 256 \times 64$
	conv	1×1	1	$256 \times 256 \times 1$

The model structure of the improved U-Net is shown in Figure 7. The improved U-Net used for through-wall target detection has the following two advantages:

- (1) The skip connection in U-Net stitches the shallow target details into the deep feature map, which solves the problem of the lack of details of weak targets or adjacent targets in the deep network;
- (2) The combined SE-ResNet module can make up for the vanishing gradients in the deep network during gradient descent network training through the ResNet module. In addition, it can also increase the weight of useful features and suppress the weight of useless features through the SE module, so as to describe weak targets and adjacent targets in through-wall multi-target imaging accurately.

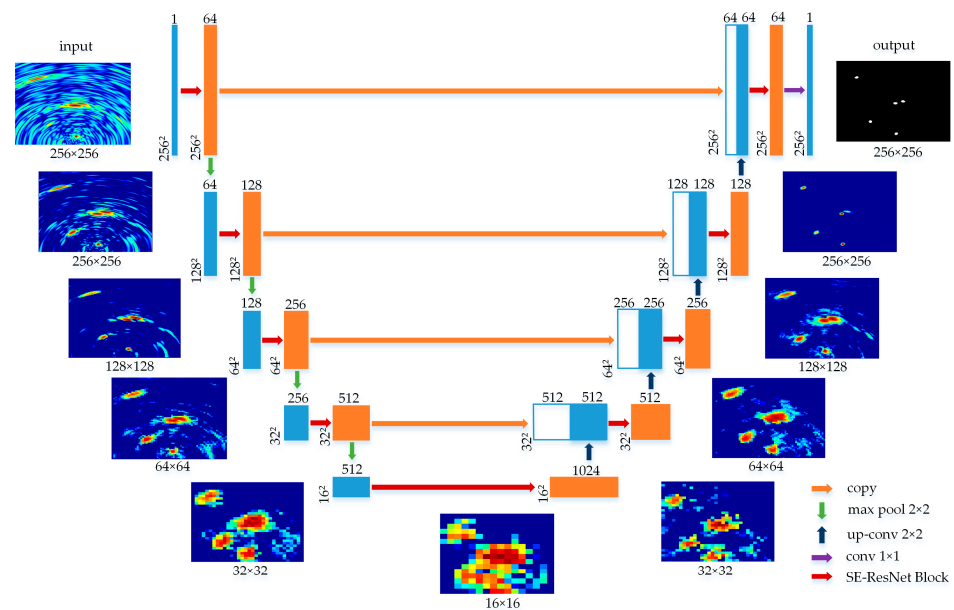


Figure 7. The improved U-Net model.

4. Simulation

4.1. Dataset Generation and Model Training

The imaging result of the through-wall human target is circular or elliptical due to the limitations of resolution, as shown in Figure 3. According to many experiments, the results of through-wall imaging obtained through simulations is similar to the experimental results from the actual scene, so the dataset required in this paper can be established through simulation. Matlab 2018b is used to simulate and obtain the required dataset in this paper. The simulation model constructed is a 2D propagation problem.

The simulation model is shown in Figure 2, and the parameters of the radar have been given in Section 2.2. The thickness of the wall is 24 cm, and the relative permittivity $\epsilon_1 = 7.0$. The imaging area in azimuth is $-5 \sim 5$ m, and the imaging area in range is $0 \sim 10$ m. Gaussian white noise is randomly added to the through-wall echo in the simulation, and the SNR ranges from 2 dB to 10 dB. The number of targets in each simulation is between one and five, and all targets are distributed in the imaging area randomly. In addition, the reflectance of each target is set to $0.2 \sim 1$. The imaging area is divided into 256×256 grid points, and the through-wall images are normalized as the input of the network, as shown in Figure 8a.

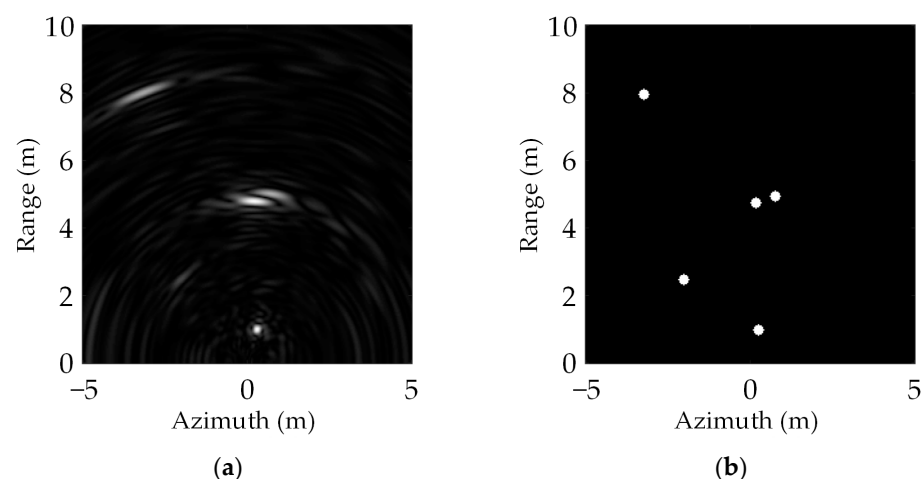


Figure 8. Imaging result and ground truth; (a) 2D imaging result, and (b) labeling result.

It can be seen that the human target is presented as a circle or an ellipse according to the imaging characteristics of the through-wall human target. Therefore, a circular area can be used to represent the labeling result of the imaging target in the simulation; the inner value of the labeling circle is set to 1, and the rest is the background, which is set to 0. In this paper, the radius of the labeling circle of the target is set to six pixels, as shown in Figure 8b. The dataset generated by Matlab software contains 2000 images, 1600 of which are used for training, 200 are used for validation and 200 are used for testing. The loss function is cross entropy in model training, which can be defined by the following formula:

$$L = -\frac{1}{N}(\sum_{p=1}^P (O_p \log \widehat{O}_p + (1 - O_p) \log(1 - \widehat{O}_p))), \quad (8)$$

where N is the batch size, $p = 1, 2, \dots, P$ and P is the total number of pixels in a batch of data. O_p and $\widehat{O}_p$ are the labeling result and prediction result of the p th pixel, respectively.

The designed network is implemented through PyTorch, and the initialization weight of the model is implemented through normal distribution. The Adam optimizer is introduced, and the learning rate is set to 0.001. The batch size is set to 1. The training process consists of 20 epochs, and the final model is obtained after convergence.

4.2. Metric Evaluation of the Model

The Dice coefficient and intersection over union (IoU) parameter are two commonly used evaluation indicators in target segmentation tasks. Dice represents the similarity between the predicted target area and the target labeling area in through-wall imaging. The larger the value, the better the segmentation effect. It can be defined by the following equation:

$$Dice = \frac{2C}{A + B}. \quad (9)$$

The definitions of “ A ” and “ B ” are given in Figure 9. “ A ” indicates the target labeling area, “ B ” indicates the predicted target area of through-wall imaging and “ C ” is the overlapping part of the target labeling area and the predicted target area in through-wall imaging.

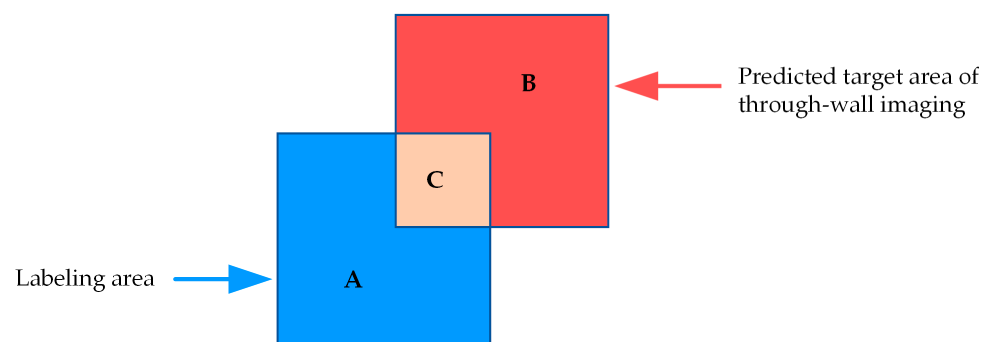


Figure 9. Representation of the target area.

IoU represents the ratio between the intersection and union of the predicted target area and the target labeling area in through-wall imaging. It is used to indicate the degree of coincidence between the predicted target area and the target labeling area in through-wall imaging. It can be defined as follows:

$$IoU = \frac{C}{A \cup B}. \quad (10)$$

The accuracy of target detection and the ability to detect weak targets are the key to through-wall multi-target detection. Therefore, two other important indicators for evaluating the performance of target detection are proposed in this paper: detection precision P_r and the miss detection rate (MDR). Detection precision represents the correct

probability of detected targets compared to the total true targets, which can be defined as follows:

$$P_r = \frac{N_{Correct}}{N_{All}}, \quad (11)$$

where N_{All} represents the total number of real targets, and $N_{Correct}$ is the number of correct targets in the detection results.

The missed detection rate can be defined as follows:

$$MDR = \frac{N_{Miss}}{N_{All}}, \quad (12)$$

where N_{Miss} is the number of undetected targets; most of them are through-wall weak targets or adjacent unresolving targets in this paper.

4.3. Simulation Result

The effectiveness of the proposed improved U-Net model in the detection of through-wall multiple human targets is verified via simulation and experiments. All data processing was performed on a computer with an Intel Core i3-8100 CPU running at 3.60 GHz, which was equipped with 32 GB RAM and GTX 1070 Ti GPU. The version of Python was 3.7 and the version of CUDA was 11.1.

The simulation results of through-wall target detection can be analyzed according to visual effects and specific parameters. Visually, the results are mainly analyzed from the ability to distinguish adjacent targets, the ability to detect weak targets and the degree of restoration of target detection. In terms of parameters, the results of Dice, IoU, P_r and MDR are mainly analyzed.

The proposed improved U-Net method is compared to the OS-CFAR, FCN and traditional U-Net; the target detection results are shown in Figure 10. Figure 10a is the detection result of the OS-CFAR. The OS-CFAR can detect the weak target 2, but adjacent targets 3 and 4 are aliased together and cannot be distinguished. Only the intensity features of the target are considered in traditional detection methods, so it is difficult to separate adjacent targets. Figure 10b,c shows the detection results of FCN and traditional U-Net, respectively. The detection results of these two methods are similar, as both can detect the weak target 2 and the adjacent targets 3 and 4. However, the area range of the measured weak target 2 and adjacent target 4 is small compared to the real labeling value of the target, i.e., serious distortion. Target 2 will be missed, and targets 3 and 4 will also be aliased in FCN and U-Net when the reflection intensity of weak target 2 is further reduced and the distance between target 3 and target 4 is closer. Figure 10d is the detection result of the proposed improved U-Net. The weak target 2 and adjacent targets 3 and 4 can be detected effectively, returning results similar to the results of the real labeling value of the target.

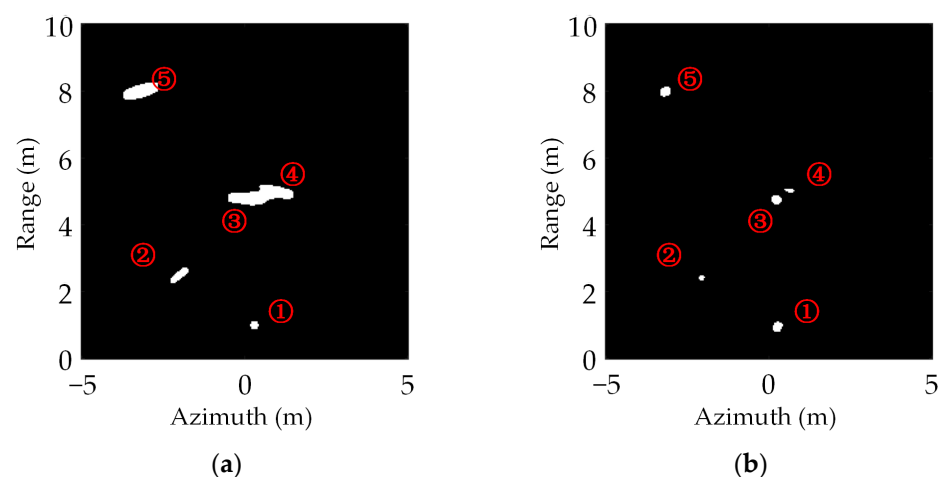


Figure 10. Cont.

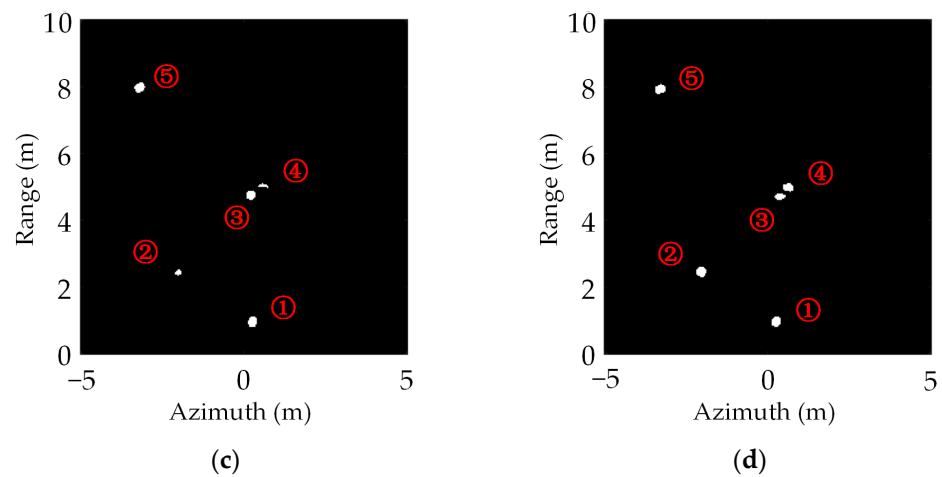


Figure 10. Detection results of simulated targets. (a) OS-CFAR, (b) FCN, (c) U-Net and (d) the proposed improved U-Net.

In terms of positioning error, the errors between the positions of the five target detected via the proposed U-Net method and the actual labeling positions are 0.04 m, 0.07 m, 0.18 m, 0.17 m and 0.09 m, respectively, which meet the practical application requirements.

The results of the training loss function of the FCN, traditional U-Net and the proposed improved U-Net are shown in Figure 11. The loss function of the improved U-Net model converges fastest and has the smallest value after training. It has been explained in Formula (8) that the loss function is cross entropy in this paper. The smaller the loss function is, the closer the detection result is to the real target value, which is consistent with the conclusion in Figure 10.

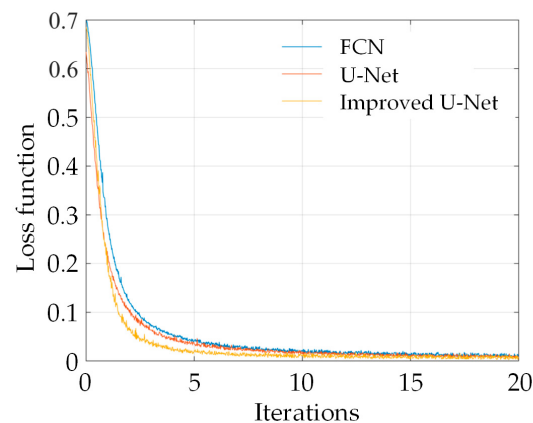


Figure 11. Loss function.

The FLOPs and Params in the traditional U-Net network training are 40.08 G and 17.27 M, respectively. The FLOPs and Params in the improved U-Net network training are 27.88 G and 11.81 M, respectively. The training time of the U-Net network before and after improvement is 1.6 h and 2.8 h, respectively.

In order to further compare the detection performance of the OS-CFAR, FCN and traditional U-Net, 631 targets were detected in 200 through-wall radar images. The detection results of different methods are given in Table 2. The Dice and IoU of the proposed improved U-Net can reach 89.0% and 86.4%, respectively, which are higher than those that the FCN and traditional U-Net can reach. This means that the detection results of the proposed improved U-Net have a higher coincidence with the real result map of through-wall imaging. In terms of the detection precision, the target detection precision of FCN and U-Net is at least 19% higher than that of the traditional OS-CFAR. The detection precision of

the proposed improved U-Net is highest, reaching 91.6%. The parameter of MDR is mainly used to describe through-wall weak targets and adjacent targets. Compared with FCN and the traditional U-Net method, the proposed improved U-Net method can recognize more details, so the MDR is the lowest, which is 3.0%.

Table 2. Simulation results of different methods.

	Dice	IoU	P_r	MDR
OS-CFAR	-	-	70.1%	18.1%
FCN	88.0%	85.2%	89.7%	4.1%
U-Net	88.2%	85.4%	89.9%	4.1%
Improved U-Net	89.0%	86.4%	91.6%	3.0%

4.4. Ablation Study

To evaluate the impact of the SE-Net and ResNet modules in the improved U-Net, we conducted an ablation study by comparing these with two degenerate variants: (1) U-SENet, which incorporates the SE-Net module in the convolution layers of the traditional U-Net, and (2) U-ResNet, which replaces the convolution layers of the traditional U-Net with ResNet blocks.

The results of the ablation study are presented in Table 3. As expected, both variants outperform the traditional U-Net but fall short of the performance achieved using the full model, the improved U-Net. This further validates the benefits of using SE-Net and ResNet modules. Moreover, U-ResNet generally outperforms U-SENet, indicating that the ResNet module, with its stronger feature extraction ability, contributes more to performance improvement.

Table 3. Ablation study on the improved U-Net.

	Dice	IoU	P_r	MDR
U-Net	88.2%	85.4%	89.9%	4.1%
U-SENet	88.6%	85.8%	90.2%	3.4%
U-ResNet	88.8%	86.0%	90.3%	3.4%
Improved U-Net	89.0%	86.4%	91.6%	3.0%

5. Experiment

5.1. Detection of Stationary Targets

The designed 2D MIMO through-wall radar was used to detect through-wall multiple human targets. The MIMO radar used was developed by the Institute of Aerospace Information Research Institute, Chinese Academy of Sciences. The center frequency and bandwidth were both 1 GHz. The length of the array was 1 m, including 2 transmitting antennas at both ends and 6 receiving antennas in the middle. The experimental scene is shown in Figure 12a. The relative permittivity of the wall measured by the vector network analyzer is about 7.0, and the measured thickness of the wall is 37 cm. Taking the detection of 5 human targets as an example, the 5 human targets stood on the other side of the wall and remained stationary, and the position coordinates were (1.0 m, 1.3 m), (−1.5 m, 2.8 m), (0.2 m, 4.5 m), (0.6 m, 4.8 m) and (−1.7 m, 6.6 m). The imaging results of through-wall multiple targets are shown in Figure 12b; the imaging of target 2 is weak, and the energy intensity is much lower than that of other targets. In addition, the imaging results of targets 3 and 4 are almost mixed together due to their close range. Comparing the imaging results of the actual human target with the simulation results, it can be found that the shape of the target in the through-wall experimental results is very similar to that in the simulation results, both of which are circular or elliptical. Therefore, the network trained on the simulation dataset can be used to process the actual experimental data directly to verify its generalization ability.

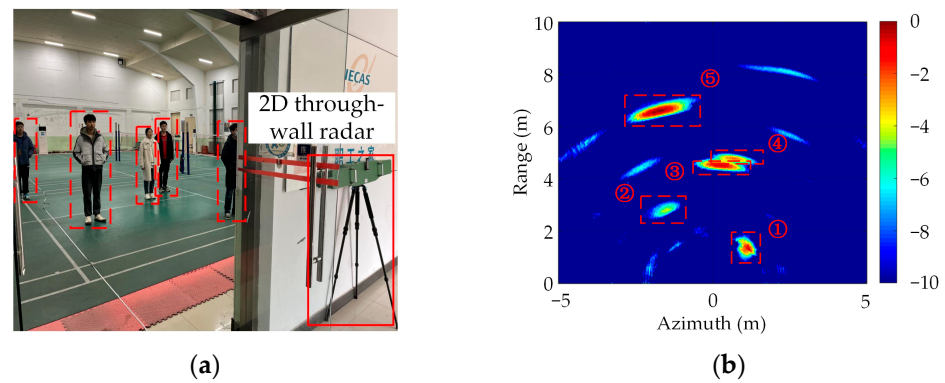


Figure 12. Experimental scene and imaging. (a) Experimental scene; (b) imaging.

The detection results processed using the OS-CFAR, FCN, traditional U-Net and the proposed improved U-Net are given in Figure 13. Figure 13a is the detection result of the OS-CFAR, the weak target 2 is missed, and the adjacent targets 3 and 4 are aliased. Figure 13b,c are the detection results of FCN and traditional U-Net, respectively. In FCN, weak target 2 is missed, and the detection result of adjacent target 4 is weak. In the U-Net result, the weak target 2 is not missed, but the detection result is extremely weak. In addition, the detection result of adjacent target 4 is also weak, which is similar to the case with the FCN. Figure 13d is the detection result of the proposed improved U-Net method. Weak target 2 and adjacent targets 3 and 4 can be detected effectively, resulting in a value similar to the true value map of the target.

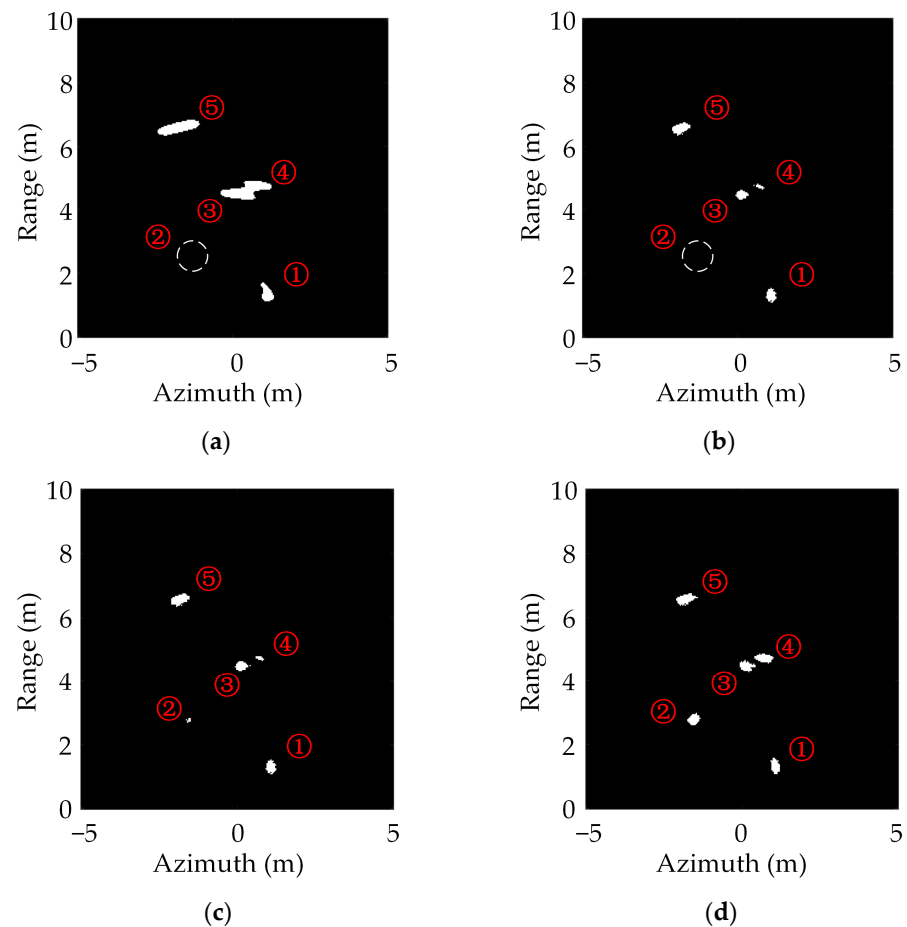


Figure 13. Experimental detection results. (a) OS-CFAR, (b) FCN, (c) U-Net and (d) the proposed improved U-Net.

Briefly, 272 targets in 80 experimental images of through-wall multi targets were detected, the detection precision and MDR of different detection methods are given in Table 4. The detection precision of the OS-CFAR is only 66.5%, and the detection precision of FCN and U-Net is at least 22% higher than that of the traditional OS-CFAR. The precision of the proposed improved U-Net is the highest, at 89.3%. The missed detection rates of FCN and the traditional U-Net are much lower than those of the OS-CFAR, which are 5.5% and 5.2%, respectively. The missed detection rate of the proposed improved U-Net is the lowest, which is 4.4%. Experimental results show that the proposed improved U-Net can better detect through-wall weak targets and adjacent targets. The experimental conclusions are basically consistent with those of the theoretical simulation.

Table 4. Experimental results of different methods.

	P_r	MDR
OS-CFAR	66.5%	22.8%
FCN	87.1%	5.5%
U-Net	87.5%	5.2%
Improved U-Net	89.3%	4.4%

5.2. Detection of Moving Targets

The proposed U-Net method is used to detect through-wall multiple moving targets and then to draw the trajectory of the targets. The experimental scene of through-wall multiple moving targets detection is shown in Figure 14.

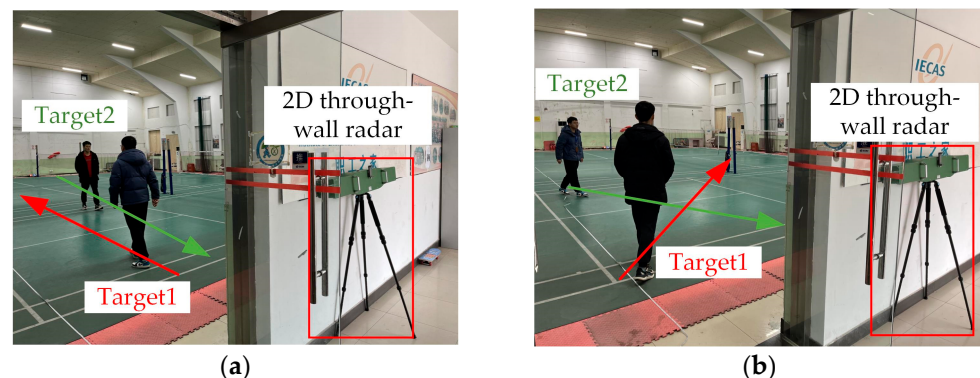


Figure 14. Experimental scene of through-wall multiple moving target detection. (a) Parallel opposing motion, and (b) cross-opposing motion.

Two sets of detection experiments of moving targets were carried out. The test scenario of experiment 1 is shown in Figure 14a; two human targets move towards each other along two trajectories parallel to the normal line of the radar. The interval between the two parallel trajectories is 0.6 m. The starting point and end point of target 1 are (−0.3 m, 1 m) and (−0.3 m, 7 m), respectively. The starting point and end point of target 2 are (0.3 m, 7 m) and (0.3 m, 1 m), respectively. The test scenario of experiment 2 is shown in Figure 14b; two human targets move towards each other along two cross trajectories. The starting point and end point of target 1 are (−2 m, 2 m) and (2 m, 7 m), respectively. The starting point and end point of target 2 are (−2 m, 7 m) and (2 m, 2 m), respectively. The imaging frame rate of the designed 2D MIMO through-wall radar is 5 frames/s. The positions of multiple targets can be extracted after the imaging results are fed into the trained improved U-Net network.

According to the detection results of imaging through-wall multiple static targets, the detection effect of the FCN and traditional U-Net on through-wall multiple targets is almost the same. Therefore, the FCN will not be repeated in the future experiment on through-wall multiple moving target detection. The improved U-Net, traditional U-Net and OS-CFAR

are used to detect a through-wall moving target, and then to draw the trajectory of the targets. The detection results of parallel opposing motion and cross-opposing motion under different methods are shown in Figure 15.

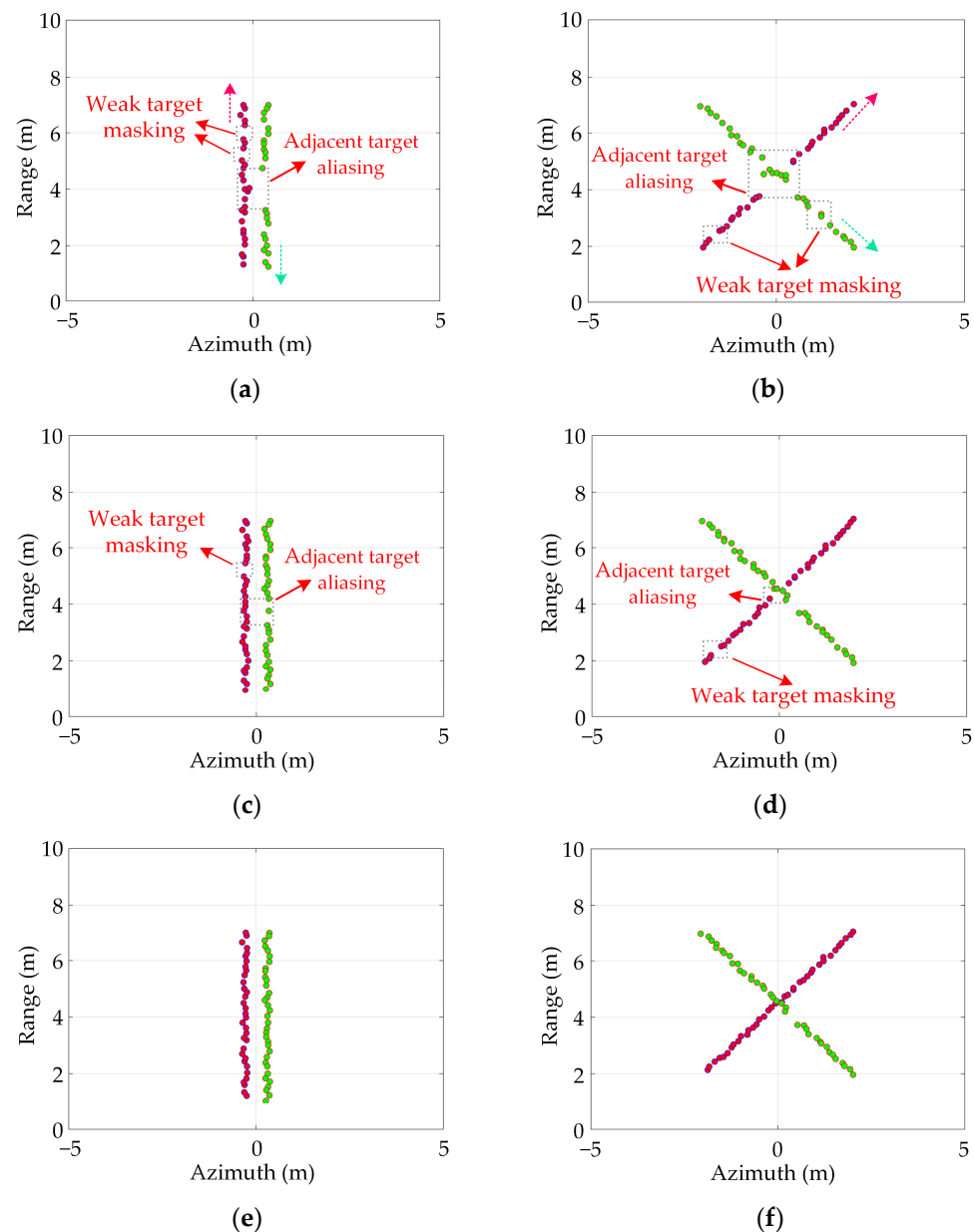


Figure 15. Detection results of moving targets. (a) The detection result of parallel opposing motion after OS-CFAR. (b) The detection result of cross opposing motion after OS-CFAR. (c) The detection result of parallel opposing motion after U-Net. (d) The detection result of cross opposing motion after U-Net. (e) The detection result of parallel opposing motion after the improved U-Net. (f) The detection result of cross opposing motion after the improved U-Net.

Figure 15a,b shows the detection results of parallel opposing motion and cross-opposing motion based on the OS-CFAR, respectively. It can be seen that the unresolving phenomenon of the targets is more obvious when moving targets are close, and only a single target can be detected. In addition, target masking occurs in many places globally during the detection process and leads to the loss of the target. Figure 15c,d shows the detection results of parallel opposing motion and cross-opposing motion based on the traditional U-Net, respectively. The results show that the unresolving phenomenon is alleviated when moving targets are close, and the masking points of the weak target are

also reduced, but the two phenomena still exist. Figure 15e,f includes the detection results of parallel opposing motion and cross-opposing motion based on the proposed improved U-Net, respectively. In both scenarios, the unresolving phenomenon basically disappears when the two targets are close, and the number of masking points is minimized.

6. Conclusions

This paper proposes an improved through-wall multi-target detection method based on U-Net. The ResNet module and SE module are integrated in the traditional U-Net model. The ResNet module can reduce the difficulty of feature learning and improve the accuracy of detection. The SE module allows the network to perform feature recalibration and learn to use global information to emphasize useful features selectively and suppress less useful features. The ability to detect through-wall weak targets and adjacent targets can be improved by introducing these two modules into the U-Net model. The proposed U-Net method is compared with the OS-CFAR, FCN and traditional U-Net methods. Compared to the OS-CFAR, the detection precision, P_r , of the proposed method is increased by 21.5%, and the missed detection rate MDR is reduced by 15.1%. Compared to the FCN and traditional U-Net, the Dice of the proposed method is increased by at least 0.8%, the IoU is increased by at least 1.0%, P_r is increased by at least 1.7%, and MDR is decreased by at least 1.1%. In addition, the proposed method can be used to detect through-wall multiple moving targets. The results show that the proposed method can eliminate weak target masking and target unresolving effectively when crossing during motion. The proposed method is also applicable to the detection of stationary and moving targets in other through-wall scenarios (walls of different thicknesses, different numbers of targets, etc.), which makes the high-precision detection of through-wall targets possible.

Author Contributions: Conceptualization, J.P. and Z.Z.; methodology, J.P.; software, D.Z.; validation, K.Y., J.N. and B.Z.; formal analysis, J.P.; investigation, J.N.; resources, G.F.; data curation, B.Z.; writing—original draft preparation, J.P.; writing—review and editing, J.N. and G.F.; visualization, Z.Z.; supervision, B.Z.; project administration, G.F.; funding acquisition, G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Basic Science Center Project of National Natural Science Foundation of China under grant 61988102, in part by the Key Research and Development Program of Guangdong Province under grant 2019B090917007, and in part by the Science and Technology Planning Project of Guangdong Province under grant 2019B090909011.

Data Availability Statement: Data sharing is not applicable to this article.

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

References

1. Shi, C.; Ni, Z.-K.; Pan, J.; Zheng, Z.; Ye, S.; Fang, G. A Method for Reducing Timing Jitter's Impact in through-Wall Human Detection by Ultra-Wideband Impulse Radar. *Remote Sens.* **2021**, *13*, 3577. [\[CrossRef\]](#)
2. Li, Z.; Jin, T.; Dai, Y.; Song, Y. Through-Wall Multi-Subject Localization and Vital Signs Monitoring Using UWB MIMO Imaging Radar. *Remote Sens.* **2021**, *13*, 2905. [\[CrossRef\]](#)
3. Maiti, S.; Bhattacharya, A. Microwave Detection of Respiration Rate of a Living Human Hidden Behind an Inhomogeneous Optically Opaque Medium. *IEEE Sens. J.* **2021**, *21*, 6133–6144. [\[CrossRef\]](#)
4. Ye, S.; Zhou, B.; Fang, G. Design of a Novel Ultrawideband Digital Receiver for Pulse Ground-Penetrating Radar. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 656–660. [\[CrossRef\]](#)
5. Guangyou, F.; Pipan, M. Instantaneous Parameters Calculation and Analysis of Impulse Ground Penetrating Radar (GPR) Data. In Proceedings of the IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No.01CH37217), Sydney, Australia, 9–13 July 2001; Volume 6, pp. 2695–2697.
6. Fang, G. The Research Activities of Ultrawide-Band (UWB) Radar in China. In Proceedings of the IEEE International Conference on Ultra-Wideband, Singapore, 24–26 September 2007.
7. Zhao, B.; Zhang, Y.; Lang, S.; Liu, Y.; Zhang, F.; Tang, C.; Liu, X.; Fang, G.; Cui, X. Shallow-Layers-Detection Ice Sounding Radar for Mapping of Polar Ice Sheets. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4301010. [\[CrossRef\]](#)

8. Xu, B.; Lang, S.; Cui, X.; Li, L.; Liu, X.; Guo, J.; Sun, B. Focused Synthetic Aperture Radar Processing of Ice-Sounding Data Collected Over East Antarctic Ice Sheet via Spatial-Correlation-Based Algorithm Using Fast Back Projection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5233009. [\[CrossRef\]](#)
9. *IEEE Std 1672-2006*; IEEE Standard for Ultrawideband Radar Definitions. IEEE: Piscataway, NJ, USA, 2007; pp. 1–19. [\[CrossRef\]](#)
10. *IEEE 1672-2006Cor 1-2008 Corrigendum IEEE Std 1672-2006*; IEEE Standard for Ultrawideband Radar Definitions—Corrigendum 1. IEEE: Piscataway, NJ, USA, 2008; pp. 1–5. [\[CrossRef\]](#)
11. Ma, Y.; Liang, F.; Wang, P.; Lv, H.; Yu, X.; Zhang, Y.; Wang, J. An Accurate Method to Distinguish Between Stationary Human and Dog Targets under Through-Wall Condition Using UWB Radar. *Remote Sens.* **2019**, *11*, 2571. [\[CrossRef\]](#)
12. Randazzo, A.; Ponti, C.; Fedeli, A.; Estatico, C.; D’Atanasio, P.; Pastorino, M.; Schettini, G. A Through-the-Wall Imaging Approach Based on a TSVD/Variable-Exponent Lebesgue-Space Method. *Remote Sens.* **2021**, *13*, 2028. [\[CrossRef\]](#)
13. Pan, J.; Ni, Z.-K.; Shi, C.; Zheng, Z.; Ye, S.; Fang, G. Motion Compensation Method Based on MFDF of Moving Target for UWB MIMO Through-Wall Radar System. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3509205. [\[CrossRef\]](#)
14. Rohman, B.P.A.; Andra, M.B.; Nishimoto, M. Through-the-Wall Human Respiration Detection Using UWB Impulse Radar on Hovering Drone. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6572–6584. [\[CrossRef\]](#)
15. Hu, Z.; Zeng, Z.; Wang, K.; Feng, W.; Zhang, J.; Lu, Q.; Kang, X. Design and Analysis of a UWB MIMO Radar System with Miniaturized Vivaldi Antenna for Through-Wall Imaging. *Remote Sens.* **2019**, *11*, 1867. [\[CrossRef\]](#)
16. Song, Y.; Jin, T.; Dai, Y.; Song, Y.; Zhou, X. Through-Wall Human Pose Reconstruction via UWB MIMO Radar and 3D CNN. *Remote Sens.* **2021**, *13*, 241. [\[CrossRef\]](#)
17. Li, H.; Cui, G.; Kong, L.; Chen, G.; Wang, M.; Guo, S. Robust Human Targets Tracking for MIMO Through-Wall Radar via Multi-Algorithm Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 1154–1164. [\[CrossRef\]](#)
18. Qi, F.; Li, Z.; Ma, Y.; Liang, F.; Lv, H.; Wang, J.; Fathy, A.E. Generalization of Channel Micro-Doppler Capacity Evaluation for Improved Finer-Grained Human Activity Classification Using MIMO UWB Radar. *IEEE Trans. Microw. Theory Tech.* **2021**, *69*, 4748–4761. [\[CrossRef\]](#)
19. Tivive, F.H.C.; Bouzerdoun, A. Toward Moving Target Detection in Through-the-Wall Radar Imaging. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2028–2040. [\[CrossRef\]](#)
20. Li, H.; Cui, G.; Kong, L.; Guo, S.; Wang, M. Scale-Adaptive Human Target Tracking for Through-Wall Imaging Radar. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1348–1352. [\[CrossRef\]](#)
21. Guo, S.; Chen, J.; Shi, Z.; Li, H.; Wu, P.; Li, S.; Cui, G.; Yang, X. Graph Matching Based Image Registration for Multi-View Through-the-Wall Imaging Radar. *IEEE Sens. J.* **2022**, *22*, 1486–1494. [\[CrossRef\]](#)
22. Tang, V.H.; Bouzerdoun, A.; Phung, S.L. Compressive Radar Imaging of Stationary Indoor Targets with Low-Rank Plus Jointly Sparse and Total Variation Regularizations. *IEEE Trans. Image Process.* **2020**, *29*, 4598–4613. [\[CrossRef\]](#)
23. Song, Y.; Hu, J.; Chu, N.; Jin, T.; Zhang, J.; Zhou, Z. Building Layout Reconstruction in Concealed Human Target Sensing via UWB MIMO Through-Wall Imaging Radar. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1199–1203. [\[CrossRef\]](#)
24. Wankhade, A.; Nandedkar, A.V.; Rathod, S.M.; Amritkar, Y.; Rathod, S.; Sharma, A. Multiple Target Vital Sign Detection Using Ultra-Wideband Radar. In Proceedings of the 2022 International Conference on Signal and Information Processing (ICoNSIP), Pune, India, 26–27 August 2022; pp. 1–4.
25. Urdzík, D.; Kocur, D. CFAR Detectors for through Wall Tracking of Moving Targets by M-Sequence UWB Radar. In Proceedings of the 20th International Conference Radioelektronika 2010, Brno, Czech Republic, 19–21 April 2010.
26. Liu, X.; Xu, S.; Tang, S. CFAR Strategy Formulation and Evaluation Based on Fox’s H-Function in Positive Alpha-Stable Sea Clutter. *Remote Sens.* **2020**, *12*, 1273. [\[CrossRef\]](#)
27. Bhandari, A.K.; Singh, A.; Kumar, I.V. Spatial Context Energy Curve-Based Multilevel 3-D Otsu Algorithm for Image Segmentation. *IEEE Trans. Syst. Man Cybern. Syst.* **2021**, *51*, 2760–2773. [\[CrossRef\]](#)
28. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **2007**, *9*, 62–66. [\[CrossRef\]](#)
29. Zheng, Z.; Pan, J.; Ni, Z.; Shi, C.; Ye, S.; Fang, G. Human Posture Reconstruction for Through-the-Wall Radar Imaging Using Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3505205. [\[CrossRef\]](#)
30. Vishwakarma, S.; Ram, S.S. Mitigation of Through-Wall Distortions of Frontal Radar Images Using Denoising Autoencoders. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 6650–6663. [\[CrossRef\]](#)
31. Zhang, Y.; Xia, T. In-Wall Clutter Suppression Based on Low-Rank and Sparse Representation for Through-the-Wall Radar. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 671–675. [\[CrossRef\]](#)
32. Wang, F.F.; Zhang, Y.R.; Zhang, H.M. Through Wall Detection with SVD and SVM under Unknown Wall Characteristics. In Proceedings of the 2016 IEEE International Workshop on Electromagnetics: Applications and Student Innovation Competition (iWEM), Nanjing, China, 16–18 May 2016; pp. 1–3.
33. Huang, S.; Qian, J.; Wang, Y.; Yang, X.; Yang, L. Through-the-Wall Radar Super-Resolution Imaging Based on Cascade U-Net. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 2933–2936.
34. Li, H.; Cui, G.; Guo, S.; Kong, L.; Yang, X. Human Target Detection Based on FCN for Through-the-Wall Radar Imaging. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1565–1569. [\[CrossRef\]](#)
35. Pan, J.; Ye, S.; Shi, C.; Yan, K.; Liu, X.; Ni, Z.; Yang, G.; Fang, G. 3D Imaging of Moving Targets for Ultra-Wideband MIMO through-Wall Radar System. *IET Radar Sonar Navig.* **2021**, *15*, 261–273. [\[CrossRef\]](#)

36. Tan, K. A Fast Omega-K Algorithm for Near-Field 3-D Imaging of MIMO Synthetic Aperture Radar Data. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1431–1435. [[CrossRef](#)]
37. Pan, J.; Ye, S.; Ni, Z.K.; Shi, C.; Zheng, Z.; Zhao, D.; Fang, G. Enhancement of Vital Signals Based on Low-Rank, Sparse Representation for UWB through-Wall Radar. *Remote Sens. Lett.* **2022**, *13*, 98–106. [[CrossRef](#)]
38. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [[CrossRef](#)]
39. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Berlin/Heidelberg, Germany, 2015.
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
41. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [[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.