remote sensing

MDPI

Article

Two-Stage GPR Image Inversion Method Based on Multi-Scale Dilated Convolution and Hybrid Attention Gate

Mingze Wu, Qinghua Liu * and Shan Ouyang

School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China; wumingze@mails.guet.edu.cn (M.W.); hmoysh@guet.edu.cn (S.O.)
* Correspondence: qhliu@guet.edu.cn

Abstract: Ground penetrating radar (GPR) image inversion is of great significance for interpreting GPR data. In practical applications, the complexity and nonuniformity of underground structures bring noise and clutter interference, making GPR inversion problems more challenging. To address these issues, this study proposes a two-stage GPR image inversion network called MHInvNet based on multi-scale dilated convolution (MSDC) and hybrid attention gate (HAG). This method first denoises the B-scan through the first network MHInvNet1, then combines the denoised B-scan from MHInvNet1 with the undenoised B-scan as input to the second network MHInvNet2 for inversion to reconstruct the distribution of the permittivity of underground targets. To further enhance network performance, the MSDC and HAG modules are simultaneously introduced to both networks. Experimental results from simulated and actual measurement data show that MHInvNet can accurately invert the position, shape, size, and permittivity of underground targets. A comparison with existing methods demonstrates the superior inversion performance of MHInvNet.

Keywords: ground penetrating radar; image inversion; dilated convolution; attention gate

check for updates

Academic Editors: Xuan Feng, Haoqiu Zhou and Zejun Dong

Received: 27 November 2024
Revised: 8 January 2025
Accepted: 14 January 2025
Published: 17 January 2025

Citation: Wu, M.; Liu, Q.; Ouyang, S. Two-Stage GPR Image Inversion Method Based on Multi-Scale Dilated Convolution and Hybrid Attention Gate. *Remote Sens.* **2025**, *17*, 322. https://doi.org/10.3390/rs17020322

Copyright: © 2025 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

Ground-penetrating radar is a subsurface structure detection tool that offers portable and fast data acquisition technology, capable of providing dense and accurate high-resolution data [1]. It has been widely used in various fields such as rock strata detection [2], urban road disease detection [3], and building inspection [4]. GPR inversion establishes a mapping relationship between B-scans and electrical characteristics of underground media to obtain information on material structures and parameters in underground spaces.

Currently, GPR inversion methods can be divided into traditional and deep learning-based inversion methods. Traditional inversion methods include reverse time migration imaging, full waveform inversion (FWI), tomography, and others. Chen et al. [5] proposed a method that applies the normalized cross-correlation imaging condition to pre-stack migration imaging of ground-penetrating radar (GPR). Spatial high-pass filtering was used to suppress low-frequency noise generated during the cross-correlation process and reconstruct the spatial morphology and internal structural characteristics of the target. However, reverse time migration algorithms cannot provide electrical parameter information such as permittivity and conductivity of underground targets, which are crucial for identifying and detecting subsurface objects. Feng et al. [6] proposed a dual-parameter FWI method based on total variation regularization to achieve multi-scale inversion of GPR data. The dual-parameter inversion method provides reliable constraints, exhibits good noise adaptability, and can accurately invert the permittivity of underground targets. However, FWI

requires multiple forward simulations to obtain the final inversion results, leading to high computational costs and long processing times. Irving et al. [7] proposed a causal linear model-based inverse Q-filtering algorithm derived from seismic tomography technology to eliminate wavelet dispersion and improve imaging resolution. However, this method relies on precise modeling and strict detection environments, limiting its practicality in real-world engineering applications. In summary, traditional methods suffer from slow computation speeds, low accuracy, and insufficient reliability.

In recent years, deep learning has gradually been applied to GPR inversion problems due to its rapid development and efficiency. Alvarez et al. [8] used three different deep neural networks (DNNs) to reconstruct underground permittivity distribution maps, demonstrating the feasibility of deep learning methods for GPR inversion tasks. Xie et al. [9] proposed a network called Ü-Net, which incorporates instance normalization layers into U-Net for inverting permittivity images. Liu et al. [10] proposed a network called GPRInvNet for tunnel lining defect inversion, which addresses the spatial alignment issue between B-scan images and permittivity images by designing a trace-to-trace encoder. Ji et al. [11] designed an inversion network named PINet, which further extracts global information from B-scan images using a global feature encoder comprising fully connected layers to improve inversion accuracy. Wang et al. [12] enhanced the reliability of reinforced concrete defect detection by introducing a multi-path encoder in the inversion network and using three types of GPR data simultaneously. Dai et al. [13] proposed a network called DMRF-UNet for permittivity inversion of underground targets under nonuniform soil conditions, cascading two U-Net-structured networks to eliminate clutter effects caused by soil heterogeneity during the inversion process.

Existing deep learning-based ground-penetrating radar (GPR) image inversion methods are mostly applied in ideal environments with homogeneous underground media. However, when inverting actually detected B-scan images, they are often influenced by noise and clutter interference in the actual environment, leading to a decrease in the accuracy and stability of the inversion results. To address the above issues, this study proposes a two-stage GPR image inversion network MHInvNet based on the U-Net structure [14]. The main contributions are as follows:

(1) A multi-scale dilated convolution (MSDC) module is developed. To better capture the correlation between adjacent A-scan signals in B-scan images, it is introduced into the network to extract and fuse features of different scales in B-scan images;

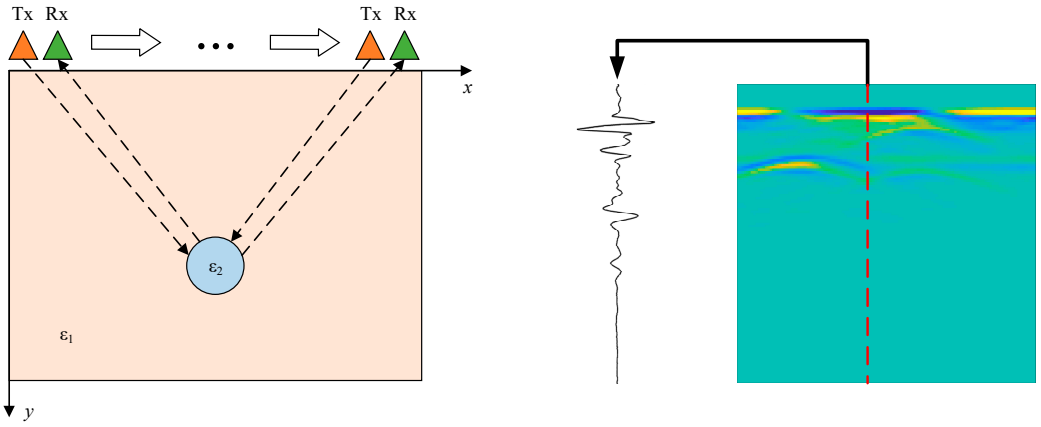
(2) A hybrid attention gate (HAG) module is proposed. Due to the skip connection operation in U-Net where encoder features are directly concatenated with decoder features, unnecessary interference information may be introduced to the decoder network. In order to enhance the network's denoising and inversion performance, it is introduced at the skip connection to highlight important features related to the target signal and suppress unnecessary noise features;

(3) Experiments with simulated data and real data obtained from actual environmental measurements demonstrate the effectiveness and superiority of the proposed network compared to existing methods.

2. Construction and Analysis of GPR Detection Model

Construct the GPR detection model as shown in Figure 1a, where the permittivities of the background medium and the target object in the model region are represented by ε_1 and ε_2, respectively. Move the GPR along the x-axis survey line direction, and the transmitting antenna Tx and receiving antenna Rx will measure multiple A-scan signals at a certain sampling interval. The A-scan reflects the temporal variation of the amplitude intensity of the reflected waves and indirectly reveals the reflection characteristics of electromagnetic

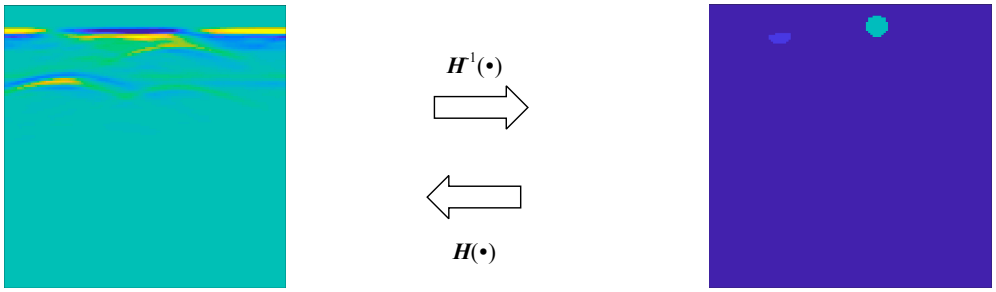
waves propagating in the depth direction. By integrating and visualizing the obtained A-scan signals, B-scan image data can be produced. The B-scan illustrates the changes in the reflected waves in both the survey line direction and the depth direction, with the relationship between the A- and B-scans depicted in Figure 1b. The GPR model detection process described above can be simplified as a mathematical model $\boldsymbol{Y} = \boldsymbol{H}(\boldsymbol{X})$, where $\boldsymbol{X}$ represents the distribution of underground target media, $\boldsymbol{H}$ represents the physical propagation process of GPR electromagnetic signals, and $\boldsymbol{Y}$ represents the B-scans data obtained from the detection.



(a) GPR Detection Model (b) A−scan(left) B−scan(right)

Figure 1. GPR detection model and data representation.

The GPR inversion problem involves the reverse reconstruction of physical information such as location, morphology, and permittivity of underground targets based on the acquired B-scan images. It can be simplified as the mathematical model $\boldsymbol{X} = \boldsymbol{H}^{-1}(\boldsymbol{Y})$, where $\boldsymbol{H}^{-1}$ is the inverse transformation of $\boldsymbol{H}$, representing the inversion operator. The inversion aims to solve the inversion operator $\boldsymbol{H}^{-1}$, which is a nonlinear problem. Deep learning provides a data-driven approach to learning the optimal $\boldsymbol{H}^{-1}$ by minimizing the differences between the predicted permittivity map and the actual permittivity map, thereby achieving a nonlinear mapping from B-scan images to permittivity maps, as shown in Figure 2.



(a) B−scan (b) Permittivity map

Figure 2. Inversion Process.

3. Network Structure

3.1. Overall Framework

The proposed network, MHInvNet, consists of two cascaded U-Nets, namely MHInvNet1 and MHInvNet2. The inversion process can be divided into two stages: first, the noisy B-scan images are denoised to obtain B-scan images that only contain the target signals. Subsequently, the inversion is performed on the dual-channel image obtained by concatenating the original noisy B-scan images with the denoised B-scan images to

reconstruct the permittivity distribution of the underground targets. The overall framework of the network is shown in Figure 3.

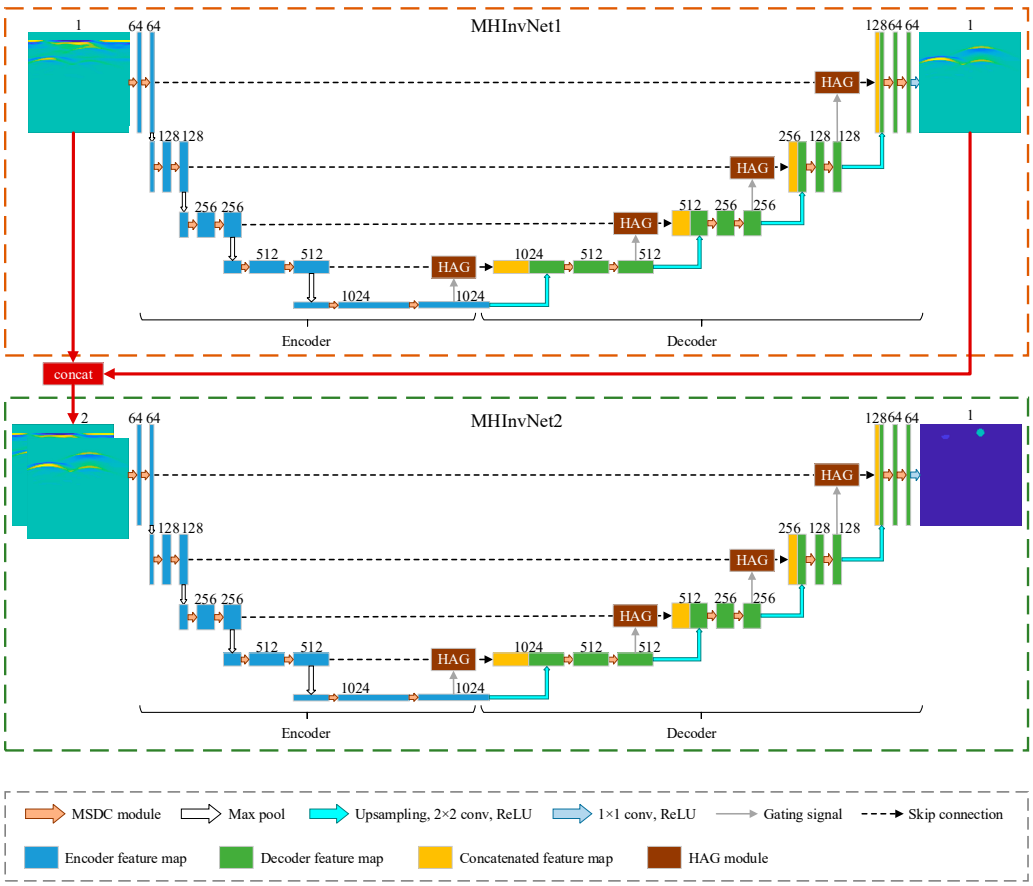


Figure 3. MHInvNet Structure.

MHInvNet1 is used for denoising the noisy input B-scans in the first stage to extract the target signal. MHInvNet1 consists of an encoder, a decoder, and skip connections with the introduction of the HAG module. The encoder consists of five feature extraction blocks, including four feature extraction blocks with downsampling operations comprising two MSDC modules and one max pooling layer, and one feature extraction block without downsampling operation consisting of two MSDC modules. The decoder consists of four upsampling blocks and a 1×1 convolutional layer with a rectified linear unit (ReLU) activation function, where each upsampling block consists of an upsampling layer and two MSDC modules. To compensate for information loss during encoding downsampling, skip connections are established between each layer of the encoder and decoder, with the introduction of the HAG module at the skip connections to help the network focus more on important feature information and suppress other interference.

MHInvNet2 is used for inverse processing of denoised B-scans in the second stage, converting B-scans into permittivity maps of underground targets. The network structure of MHInvNet2 is the same as MHInvNet1 overall, consisting of an encoder, decoder, and skip connections with the introduction of the HAG module. However, only the input part of the network is different. MHInvNet2 concatenates the denoised B-scans obtained from MHInvNet1 with the undenoised B-scans to obtain an image with a channel number of 2, which is used as the input image of the network. In turn, the network can focus more on features related to the target signal during the training process and avoid information loss in the B-scans caused by MHInvNet1's denoising process. This approach allows the

network to learn and acquire more comprehensive feature information, enhancing the reconstruction capability of the permittivity for underground targets.

3.2. Multi-Scale Dilated Convolution Model

Scattering phenomena occur when electromagnetic waves propagate underground and encounter objects with different dielectric properties. The resulting scattered echoes reach the receiving antenna at different times. Therefore, in the B-scan data obtained from GPR detection, there is a certain correlation between adjacent A-scan signals. Additionally, due to the variety of underground targets, correlations between adjacent A-scans are not identical. To better capture the relationships between different A-scans, this paper proposes a Multi-Scale Dense Correlation (MSDC) module, whose structure is shown in Figure 4.

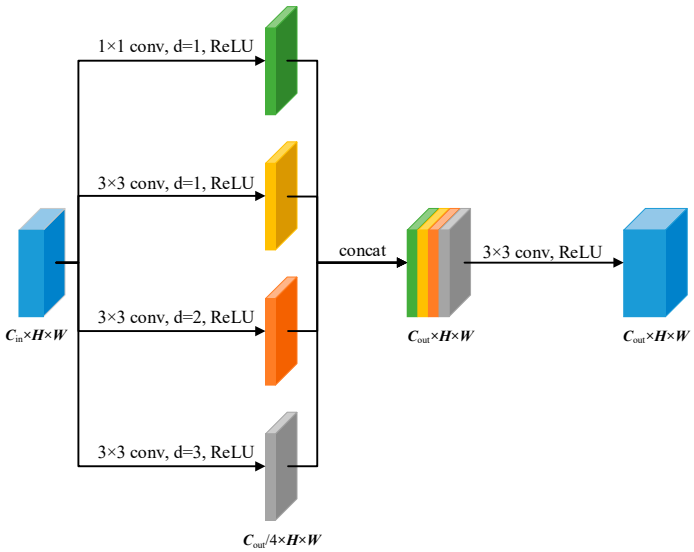


Figure 4. MSDC model.

Dilated convolution was originally applied to wavelet decomposition tasks [15], but now it is widely used in image processing fields such as semantic segmentation [16] and object detection [17]. Dilated convolution can change the size of the convolutional kernel receptive field by setting different dilation rates. This mechanism can expand the receptive field of the convolutional kernel without increasing additional parameters. The calculation formula for the equivalent standard convolutional kernel size is as follows:

$$K = (d-1) \times (k-1) + k \tag{1}$$

In the formula, K represents the size of the equivalent convolutional kernel, d is the dilation rate, and k is the size of the original convolutional kernel.

The MSDC module has four convolutional layer branches, including one branch with a convolution kernel size of 1×1 and three branches with convolution kernel sizes of 3×3. The dilation rate of the 1×1 convolution kernel is set to 1, and the dilation rates of the 3×3 convolution kernels are set to 1, 2, and 3, respectively. Using Formula (1), it can be seen that the equivalent convolution kernel sizes of the four convolution layer branches from top to bottom in Figure 4 are 1×1, 3×3, 5×5, and 7×7, respectively. The input feature map $\boldsymbol{F}_{\text{in}} \in \boldsymbol{R}^{C_{\text{in}} \times H \times W}$ first goes through the four convolution layer branches to capture feature information at different scales, while compressing the channel dimension to 1/4 of the output feature map size, obtaining the feature map $\boldsymbol{F}_{i \times i} \in \boldsymbol{R}^{C_{\text{out}}/4 \times H \times W}$, where i = 1, 3, 5, 7, represents the sizes of the equivalent convolution kernels corresponding to the four convolution layer branches, C_{in} represents the number of channels in the input

feature map, C_{out} represents the number of channels in the output feature map, and H and W represent the height and width of the feature map. Then, concatenating $\boldsymbol{F}_{i\times i}$ results in the fused feature map $\boldsymbol{F}_c \in \boldsymbol{R}^{C_{out}\times H\times W}$. Finally, after passing through a 3×3 convolutional layer and ReLU activation function, the output feature map $\boldsymbol{F}_{out} \in \boldsymbol{R}^{C_{out}\times H\times W}$ is obtained. The specific implementation process can be described using mathematical formulas:

$$\boldsymbol{F}_{i\times i} = f_D^{i\times i}(\boldsymbol{F}_{in}),\ i = 1,3,5,7 \tag{2}$$

$$\boldsymbol{F}_{out} = \mathrm{ReLU}(f^{3\times 3}(\mathrm{concat}(\boldsymbol{F}_{1\times 1}, \boldsymbol{F}_{3\times 3}, \boldsymbol{F}_{5\times 5}, \boldsymbol{F}_{7\times 7}))) \tag{3}$$

In the formula, $f_D^{i\times i}$ represents the dilated convolution operation with an equivalent kernel size of $i \times i$, $f^{3\times 3}$ represents the convolution operation with a kernel size of 3×3, and concat represents the concatenation operation. In this study, both MHInvNet1 and MHInvNet2 utilize multi-scale dilated convolution modules for denoising and inversion, respectively.

3.3. Hybrid Attention Gate Model

When the encoder of the network downsamples to a deeper level, it extracts higher-level and more abstract feature information while losing some spatial information. By connecting the feature maps of the encoder and decoder through skip connections, it can compensate for lost spatial information. However, not all skip connections are useful. Since the lower-level features of the encoder may contain unwanted interference, this interference will negatively impact the effect of the skip connection fusion feature map. To highlight important features during the skip connection process and suppress unwanted interference, inspired by Attention U-Net [18] and ECANet [19], this study proposes a HAG module consisting of two parts: Channel attention gate (CGA) and Spatial attention gate (SGA), as shown in Figure 5.

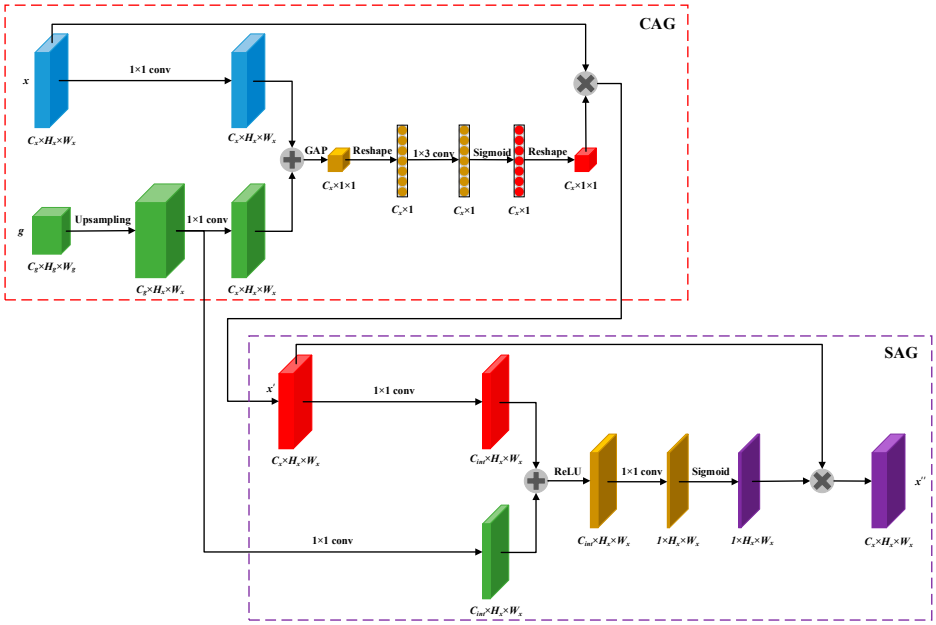


Figure 5. HAG model.

CAG is used to generate channel attention feature maps. Channel attention can capture the importance between different channels of feature maps to highlight those with important feature information. It has two inputs, the feature map $\boldsymbol{x} \in \boldsymbol{R}^{C_x\times H_x\times W_x}$ from the encoder skip connection and the gate signal $\boldsymbol{g} \in \boldsymbol{R}^{C_g\times H_g\times W_g}$ from the lower-level decoder, where C_x, H_x, W_x represent the number of channels, height, and width of the input feature map $\boldsymbol{x}$, respectively, and C_g, H_g, W_g represent the number of channels, height,

and width of the gate signal $\boldsymbol{g}$. CAG first upsamples the gate signal $\boldsymbol{g}$ from the lower-level decoder to match the feature map sizes of $\boldsymbol{x}$ and $\boldsymbol{g}$. Then, the upsampled $\boldsymbol{x}$ and $\boldsymbol{g}$ undergo convolution operations, and after being added, followed by global average pooling (GAP) to obtain the feature map $\boldsymbol{F}_{\mathrm{GAP}} \in \boldsymbol{R}^{C_x \times 1 \times 1}$. In this way, the context information in $\boldsymbol{g}$ adjusts $\boldsymbol{x}$, enhancing the alignment weights of the two signals. Next, $\boldsymbol{F}_{\mathrm{GAP}}$ is reshaped to obtain a one-dimensional feature $\boldsymbol{F}_{\mathrm{1D}} \in \boldsymbol{R}^{C_x \times 1}$, which then undergoes one-dimensional convolution, sigmoid, and reshape operations to produce the channel attention $\boldsymbol{F}_{\mathrm{CA}} \in \boldsymbol{R}^{C_x \times 1 \times 1}$. This local cross-channel interaction strategy avoids dimension reduction and can significantly reduce model complexity while maintaining performance. Finally, the output feature map $\boldsymbol{x}' \in \boldsymbol{R}^{C_x \times H_x \times W_x}$ weighted by the channel attention is obtained by multiplying $\boldsymbol{F}_{\mathrm{CA}}$ by the input $\boldsymbol{x}$. The specific implementation process can be described by the following mathematical formulae:

$$F_{\mathrm{GAP}} = f^{\mathrm{GAP}}(f^{1\times1}(\boldsymbol{x} + f^{\mathrm{UP}}(\boldsymbol{g}))) \tag{4}$$

$$F_{1D} = \mathrm{Reshape}(F_{\mathrm{GAP}}) \tag{5}$$

$$F_{\mathrm{CA}} = \mathrm{Reshape}(\mathrm{Sigmoid}(f^{1\times3}(F_{\mathrm{1D}}))) \tag{6}$$

$$\mathbf{x}' = \mathbf{x} \cdot F_{\mathrm{CA}} \tag{7}$$

In the equation, f^{GAP} represents global average pooling operation, f^{UP} represents upsampling operation, and $f^{1\times1}$, $f^{1\times3}$ represent convolution operations with kernel sizes of 1×1 and 3×3, respectively.

SAG is used to generate spatial attention feature maps. Spatial attention can capture the importance between different positions in the feature map to highlight positions with important feature information. It also has two inputs: the feature map $\boldsymbol{x}' \in \boldsymbol{R}^{C_x \times H_x \times W_x}$ weighted by channel attention from CAG and the gating signal $\boldsymbol{g} \in \boldsymbol{R}^{C_g \times H_g \times W_g}$ from the lower decoder. SAG first performs convolution on $\boldsymbol{x}'$ and upsampled $\boldsymbol{g}$, adds them, and applies the ReLU activation function to obtain the intermediate feature map $\boldsymbol{F}_{\mathrm{int}} \in \boldsymbol{R}^{C_{\mathrm{int}} \times H_x \times W_x}$. This step also aims to adjust $\boldsymbol{x}'$ according to $\boldsymbol{g}$, where C_{int} is the channel number of the input feature map $\boldsymbol{F}_{\mathrm{int}}$, usually set to half of C_x to reduce computation. $\boldsymbol{F}_{\mathrm{int}}$ then undergoes another convolution and sigmoid activation function to generate the spatial attention $\boldsymbol{F}_{\mathrm{SA}} \in \boldsymbol{R}^{1 \times H_x \times W_x}$. Finally, $\boldsymbol{F}_{\mathrm{SA}}$ is multiplied by the feature map $\boldsymbol{x}'$ to obtain the output feature map $\boldsymbol{x}'' \in \boldsymbol{R}^{C_x \times H_x \times W_x}$ after spatial attention weighting. The specific implementation process can be described by the following mathematical formulae:

$$F_{\mathrm{int}} = \mathrm{ReLU}(f^{1\times1}(\boldsymbol{x}' + f^{\mathrm{UP}}(\boldsymbol{g}))) \tag{8}$$

$$F_{\mathrm{SA}} = \mathrm{Sigmoid}(f^{1\times1}(F_{\mathrm{int}})) \tag{9}$$

$$\mathbf{x}'' = \mathbf{x}' \cdot F_{\mathrm{SA}} \tag{10}$$

HAG employs a combination of CAG and SAG in a series configuration. It first applies channel attention weighting to the input feature map, followed by spatial attention weighting to the channel-attended feature map. This approach highlights the important features in both the channel and spatial dimensions, thereby enhancing the overall feature extraction ability of the network. In this study, the HAG module is introduced into each layer's skip connection of both MHInvNet1 for denoising and MHInvNet2 for inversion.

3.4. Loss Function

The structural similarity index measure (SSIM) is a common image evaluation metric that comprehensively evaluates the similarity of two images in terms of brightness, contrast, and structural similarity. In deep learning, a variant of SSIM known as structural dissimilarity (DSSIM) is often used as a loss function, with the following expression:

$$\mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{x,y} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{11}$$

$$\mathrm{DSSIM}(\boldsymbol{x},\boldsymbol{y}) = 1 - \mathrm{SSIM}(\boldsymbol{x},\boldsymbol{y}) \tag{12}$$

In the equation, μ_x and μ_y represent the average values of images x and y, respectively; σ_x^2 and σ_y^2 represent the variances of images x and y, respectively; c_1 and c_2 are stability coefficients.

The total loss function of this study combines the losses of two networks, MHInvNet1 and MHInvNet2, to achieve simultaneous backpropagation of the two networks during the training process. Both MHInvNet1 and MHInvNet2 use DSSIM as the loss function. The definition of the total loss function is as follows:

$$l = l_1 + l_2 \tag{13}$$

In the equation, l represents the total loss function, and l_1 and l_2 represent the loss functions of MHInvNet1 and MHInvNet2, respectively.

4. Experimentation

4.1. Dataset and Evaluation Criteria

To validate the inversion performance and feasibility of the proposed network MHInvNet, this study used the open-source dataset provided by the authors of [13] to train and evaluate the network. The dataset used was simulated using the open-source software gprMax, with each dataset containing three images: the input noisy B-scan, the denoised B-scan, and the corresponding underground target permittivity map. A Peplinski mixed model was employed in the simulation process to simulate a real nonuniform soil environment for obtaining noisy B-scans with realistic dielectric and geometric properties. The specific simulation model size was 1.5 m wide and 0.5 m high. It had a background soil permittivity ranging from 3.82 to 9.99, and a target permittivity ranging from 2 to 32. The targets were randomly selected from circular, semi-circular, triangular, and rectangular shapes, with the radius of the circular and semi-circular shapes and distances from the three vertices of the triangles to the center ranging from 0.05 m to 0.08 m, and the widths and lengths of the rectangles ranging from 0.04 m to 0.06 m and 0.12 m to 0.16 m, respectively. The orientations of the semi-circular, triangular, and rectangular shapes were randomly chosen between 0° and 360°. The dataset consisted of 18,000 data sets, including 8000 single-target scene data and 10,000 double-target scene data, divided into training and testing sets in a ratio of 9:1. All images were normalized to a range of 0 to 1 and adjusted to a size of 128 × 128.

In order to evaluate the denoising performance and inversion performance of the network, three metrics including SSIM, mean absolute error (MAE), and mean square error (MSE) were adopted as evaluation criteria for denoising and inversion quality. SSIM can measure the structural similarity between two images, with a larger value indicating a higher similarity between the two images. MAE and MSE can measure the pixel-level errors between two images, with smaller values indicating minor errors between the two images.

4.2. Experimental Environment and Parameter Settings

The experimental environment was a Windows platform equipped with an NVIDIA TITAN RTX graphics card with 24 GB of graphics memory. The proposed MHInvNet network framework was implemented based on TensorFlow. An Adam optimizer was used for network training, with an initial learning rate set to 0.0001, and a batch_size set to

20. The network was trained for a total of 100 epochs, and the model with the lowest test loss was saved to obtain the best model.

4.3. Simulated Data Experiment

To verify the superior performance of the proposed network, it was compared with the network DMRF-UNet, which follows the same denoising and inversion process. All networks are configured with hyperparameters consistent with those described above. As shown in Table 1, #1 and #2 represent the denoising and inversion results of the networks. The denoising results of MHInvNet achieve SSIM, MAE, and MSE values of 0.99884, 0.16332, and 0.39140, respectively. The inversion results achieve SSIM, MAE, and MSE values of 0.99297, 0.21191, and 21.31944, respectively. Whether in denoising or inversion results, MHInvNet obtains the best scores in all three evaluation metrics. In terms of network model complexity, MHInvNet has 0.87 M fewer parameters than DMRF-UNet, which is attributed to the MSDC module employing dilated convolutions that reduce the computational burden associated with large-sized convolution kernels. Additionally, the HAG module utilizes strategies of local cross-channel interaction and compression of feature map channels to lower the module's complexity and parameter count. Regarding the time consumed to invert a B-scan image, MHInvNet has fewer parameters and requires slightly less time than DMRF-UNet. Overall, MHInvNet improves network performance while reducing model complexity.

Table 1. Comparison of denoising and inversion Performance of different networks.

Network		SSIM	MAE	MSE	Parameter/M	Time/ms
DMRF-UNet	#1	0.99773	0.79116	0.86150	12.41	23.9
	#2	0.99152	0.25348	23.26477		
MHInvNet	#1	0.99884	0.16332	0.39140	11.54	22.1
	#2	0.99297	0.21191	21.31944		

The denoised B-scan and permittivity map outputs are normalized to between −50 V/m and 75 V/m, and 0 to 32, respectively. Figure 6 shows the denoising and inversion results of different networks after normalization. Regarding denoising, as shown in Figure 6i–l, although DMRF-UNet can effectively remove the echoes of direct waves and clutter caused by soil inhomogeneity, there are still some pixel-level noise spots distributed laterally in the background of the B-scan after denoising. Figure 6q–t indicates that MHInvNet can effectively eliminate noise and unrelated target signals in the noisy B-scan, with almost no noise spots remaining in the background of the B-scan after denoising. In terms of inversion, as shown in Figure 6m–p, the target contours obtained by DMRF-UNet are blurry, with artifacts present, and the inversion effect for stacked targets is poor, sometimes leading to the inversion of false targets. In Figure 6u–x, the MHInvNet results are more accurate, showing clear boundaries between the inverted targets, closely resembling the shapes of the actual underground targets, and performing well when inverting stacked targets. Overall, MHInvNet demonstrates better performance in denoising and inversion. This improvement is attributed to the MSDC module, which extracts and integrates feature information at multiple scales, enabling the network to learn richer global features. Additionally, the HAG module provides attention constraints to the skip-connected feature maps, further enhancing important information in the feature maps and collectively improving the network's denoising and inversion performance.

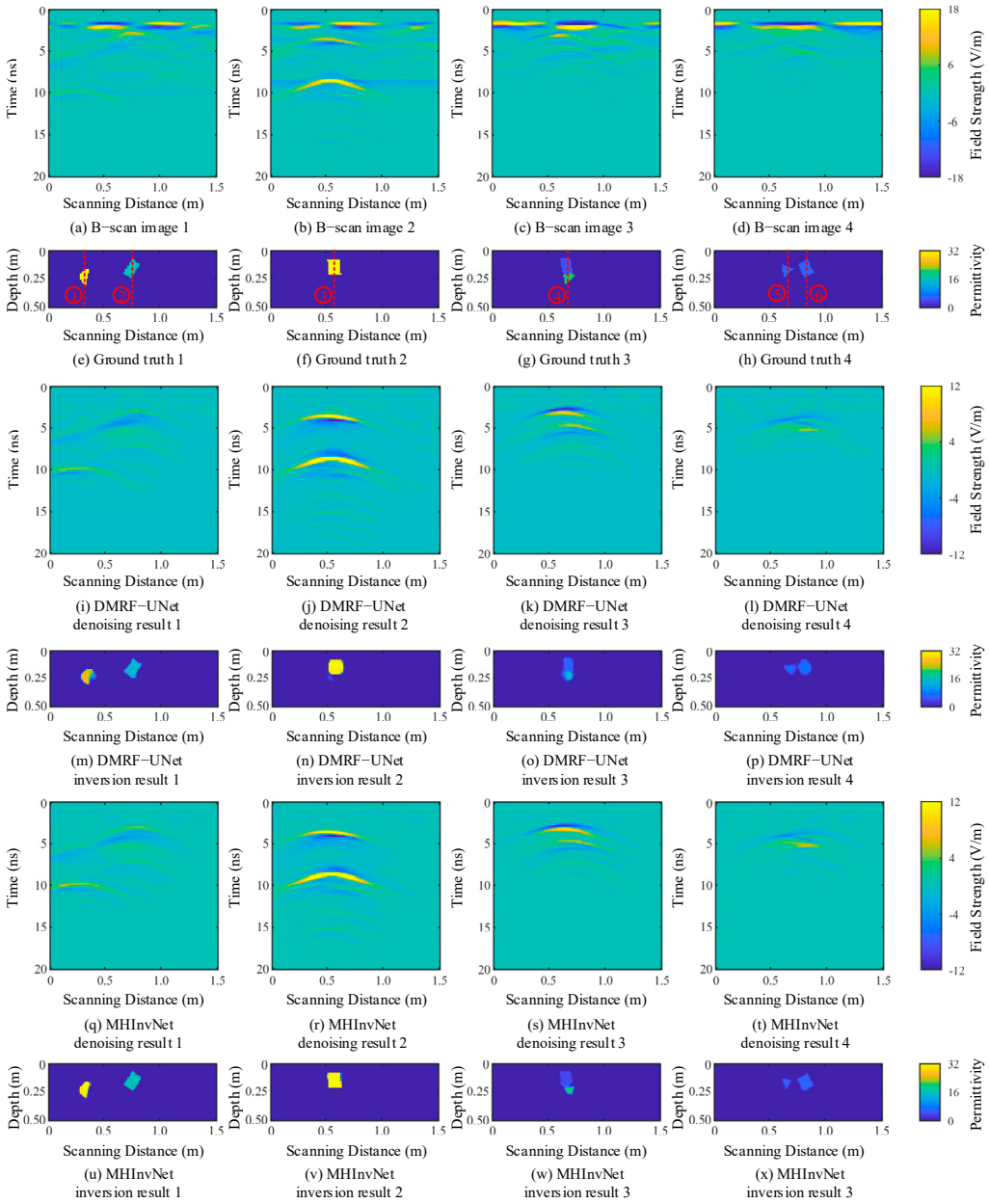


Figure 6. Denoising and inversion results of different networks.

To further validate the accuracy of the proposed network in the specific value inversion of permittivity, this study extracted the permittivity values along the six cutting lines marked from ① to ⑥ in Figure 6b for comparative analysis of the differences between the results inverted by DMRF-UNet and MHInvNet under the same cutting lines and the true permittivity model, as shown in Figure 7. It can be seen from Figure 7a–f that MHInvNet performs better in the specific value inversion of permittivity than DMRF-UNet. A further examination of Figure 7b,d,e,f reveals that the sizes and positions of the targets inverted by MHInvNet exhibit smaller deviations in the depth direction compared to the true model. Overall, the permittivity curve inverted by MHInvNet is closer to the true underground model curve, with only slight differences in the curve amplitudes.

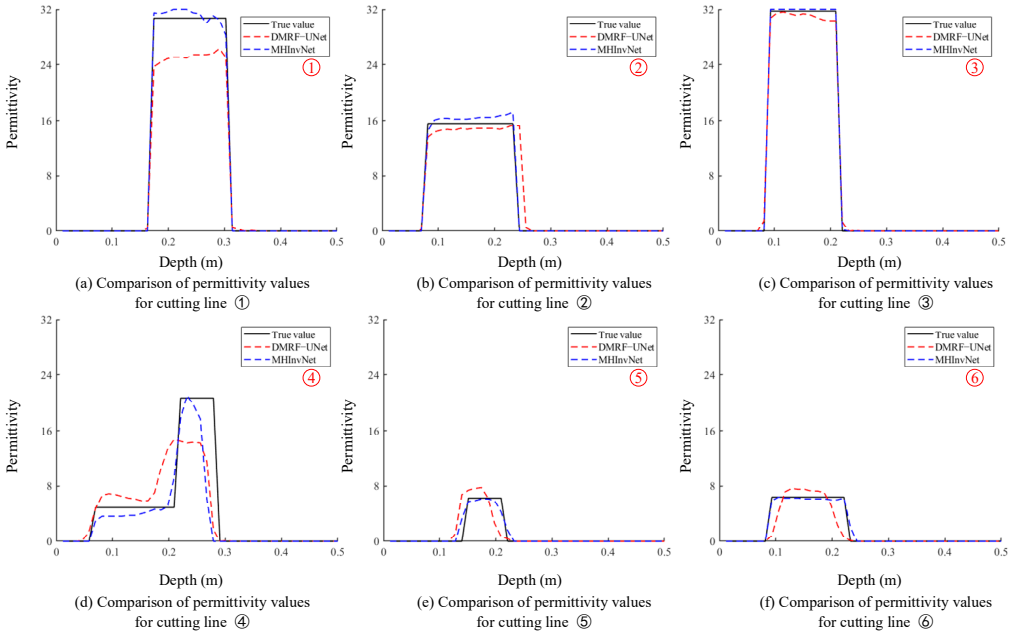


Figure 7. Inversion results of permittivity values at cutting lines ① to ⑥ using different networks.

4.4. Ablation Study

In order to demonstrate the effectiveness of the introduced MSDC module and HAG module, this paper compares and verifies four different network structures. They are MHInvNet_v1 without introducing any modules, MHInvNet_v2 with only the MSDC module, MHInvNet_v3 with only the HAG module, and the proposed network MHInvNet. The performance comparison of the four networks is shown in Table 2. MHInvNet_v1, without introducing any modules, obtained the worst performance indicators. The introduction of the MSDC module in MHInvNet_v2 enhances the network's feature extraction capability, resulting in improved performance indicators. The introduction of the HAG module in MHInvNet_v3 highlights the more important information in the feature map while improving the network's performance indicators. The simultaneous introduction of the MSDC and HAG modules in MHInvNet significantly enhances network performance, achieving the best performance indicators. It also demonstrates the effectiveness of the two modules in the denoising and inversion tasks of B-scan images.

Table 2. Performance comparison of ablation research.

Network		SSIM	MAE	MSE
MHInvNet_v1	#1	0.99838	0.79703	0.95108
	#2	0.99146	0.26318	27.65724
MHInvNet_v2	#1	0.99837	0.33972	0.39919
	#2	0.99282	0.22192	21.43576
MHInvNet_v3	#1	0.99871	0.27688	0.59640
	#2	0.99145	0.25849	27.74265
MHInvNet	#1	0.99884	0.16332	0.39140
	#2	0.99297	0.21191	21.31944

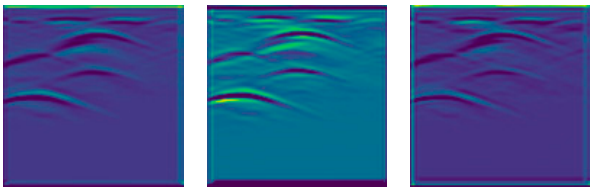
4.5. Visualization Study

In order to further understand the underlying working principles of the MSDC module and the HAG module, this study extracts the feature maps obtained from the convolutions of different downsampling layers in the inversion network for visual analysis, enabling

a more intuitive observation of the effects of these two modules on enhancing network performance. For the MSDC module, feature maps are extracted from the convolution of the shallowest feature extraction layer of the inversion network. These shallow-layer feature maps are relatively close to the input images and contain information such as textures and edges, demonstrating the network's ability to capture image details. As shown in Figure 8, when compared to a network that does not incorporate the MSDC module, the extraction of feature maps from the same position and channel reveals that the MSDC module provides better performance in the feature extraction of target signals. It can clearly delineate the edges and contours of the target signals, preserve details related to them, and capture a greater amount of feature information.



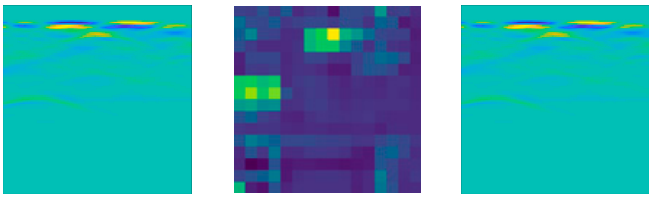
(a)Input B−scan (b) The feature map extracted by conventional convolution



(c) The feature map extracted by the MSDC module

Figure 8. MSDC module visualization.

For the HAG module, since the feature maps obtained from the deepest convolution contain higher-level, more abstract semantic information, the attention matrix extracted by HAG can more distinctly highlight important semantic features. Therefore, the attention matrix output at the deepest skip connection of the inversion network is visualized. Figure 9b shows that the attention matrix effectively represents information about the important parts in the B-scan image. The brighter the color in the matrix, the greater the attention weight at that position, indicating that the corresponding features are more important. By adjusting the extracted attention matrix to the size of the input image and superimposing the two, an attention heatmap is generated, as shown in Figure 9c. The HAG module better highlights the features of the target signal in the B-scan, allowing the network to focus on learning important features.



(a) Input B−scan (b) Attention matrix (c) HAG heat map

Figure 9. HAG module visualization.

4.6. Actual Measurement Data Experiment

In order to validate the effectiveness of the proposed network in denoising and inversion on actual measurement data, this study also used the actual open-source dataset provided by the authors of [13]. The data set was collected using commercial GSSI's Utility

Scan Pro GPR system on outdoor non-flat, nonuniform sand, and the center frequency of the GPR antenna was 400 MHz. Five different shapes, sizes, and relative permittivities of wooden objects were selected as buried objects in the sandy soil. The positions, depths, horizontal angles, and vertical angles of the objects buried underground were selected from ranges of 20 cm to 80 cm, 9 cm to 25 cm, 0° to 60°, and 0° to 60°, respectively. The collected data was normalized to between 0 and 1 and adjusted to a size of 128 × 128, with a total of 196 groups of data collected. Among them, 180 groups of data were selected as the training set for further training the network to fine-tune the weights obtained from training on simulated data. The remaining 16 groups of data were used as a test set to test the network. The network was trained for an additional 150 epochs on top of the previous training, with the remaining training parameters consistent with those described in Section 3.2.

Figure 10 shows the denoising and inversion results of different networks on actual measurement data. From Figures 10i–l and 10q–t, it can be seen that MHInvNet and DMRF-UNet perform similarly in denoising tasks. Although the clutter signals in the background are not completely removed, both can still effectively extract the target signal to reduce the adverse impact of clutter on the accuracy of target inversion. Figure 10m–p,u–x indicate that MHInvNet performs better in terms of the shape of the inverted target contour than DMRF-UNet, as the latter may produce irregular target shapes and artifacts in some cases. Overall, MHInvNet demonstrates better overall performance on actual measurement data.

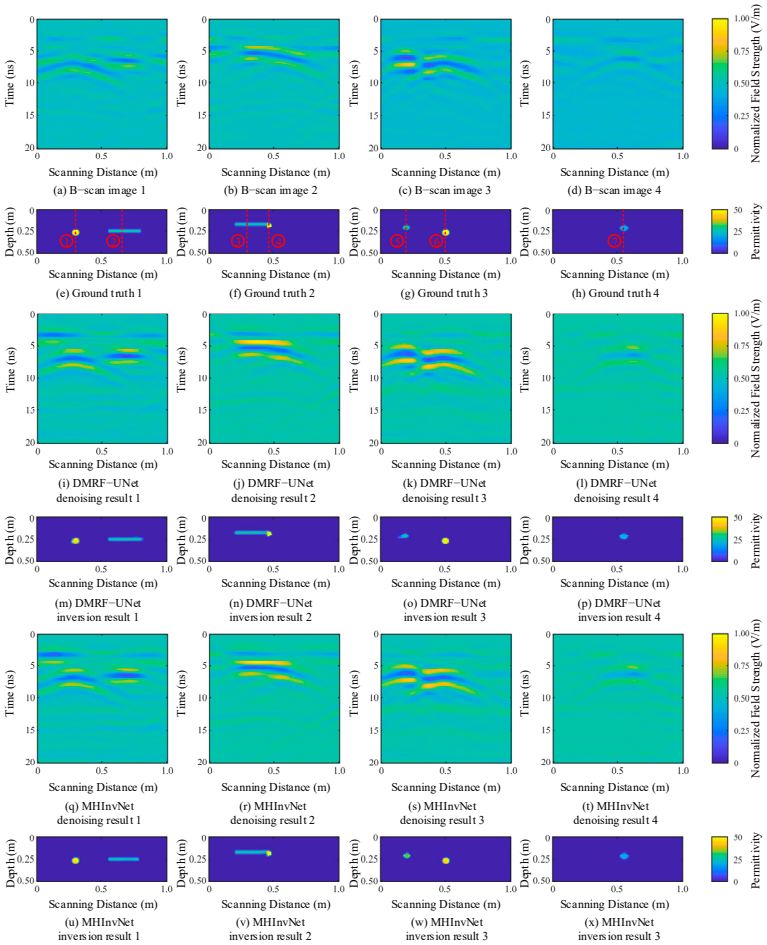


Figure 10. Denoising and inversion results of actual measurement data using different networks.

Similarly, to validate the accuracy of the network in the specific value inversion of permittivity in actual measurement data, a comparative analysis of permittivity values was conducted along the seven cutting lines marked from ① to ⑦ in Figure 10b, as shown

in Figure 11. It can be observed that the permittivity values inverted by MHInvNet are more accurate than DMRF-UNet, with only a relatively large error in the inverted target permittivity along the cutting line ⑤. Further examination of Figure 11a,f,g reveals that the distribution of target permittivity inverted by DMRF-UNet is not uniform, exhibiting slight oscillations in the permittivity curve. By contrast, the permittivity curve of MHInvNet is relatively smoother and closer to the true underground model curve. Overall, MHInvNet achieves better permittivity inversion results along the seven cutting lines.

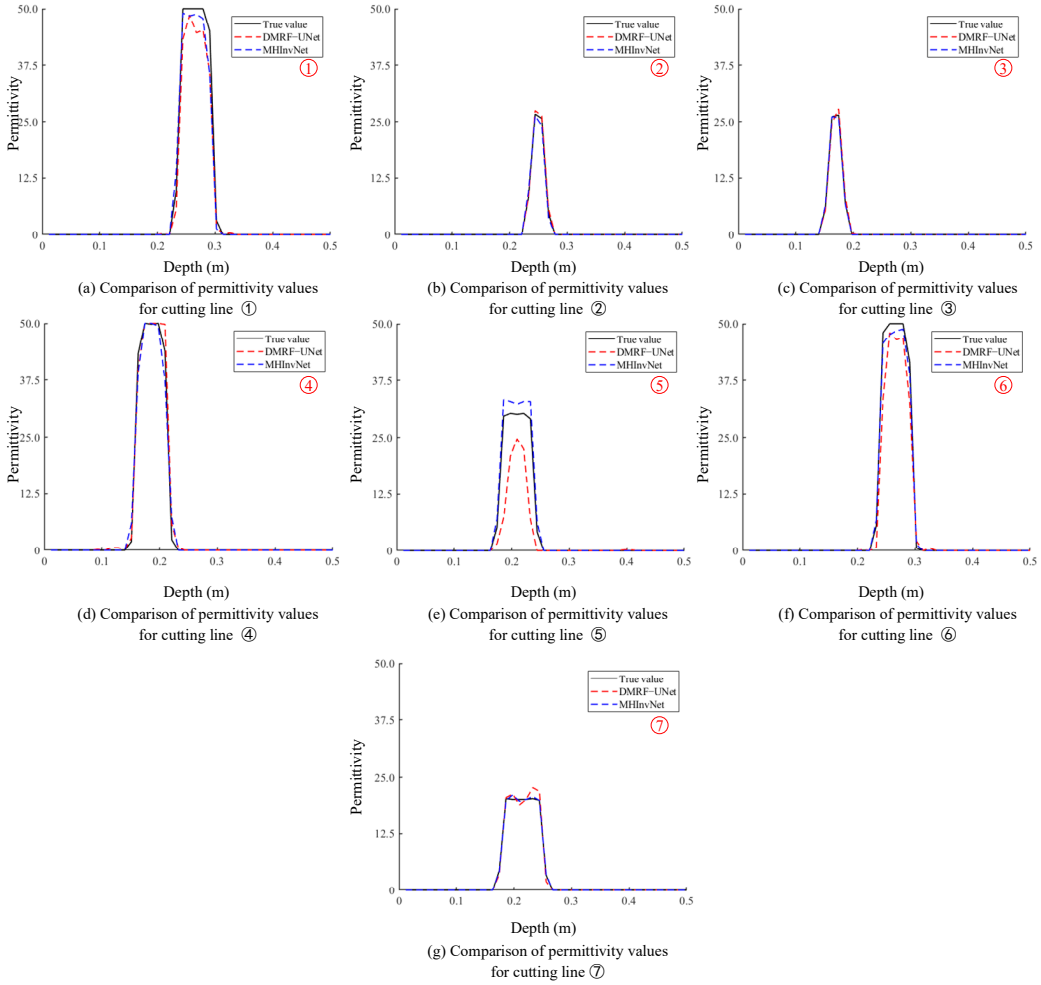


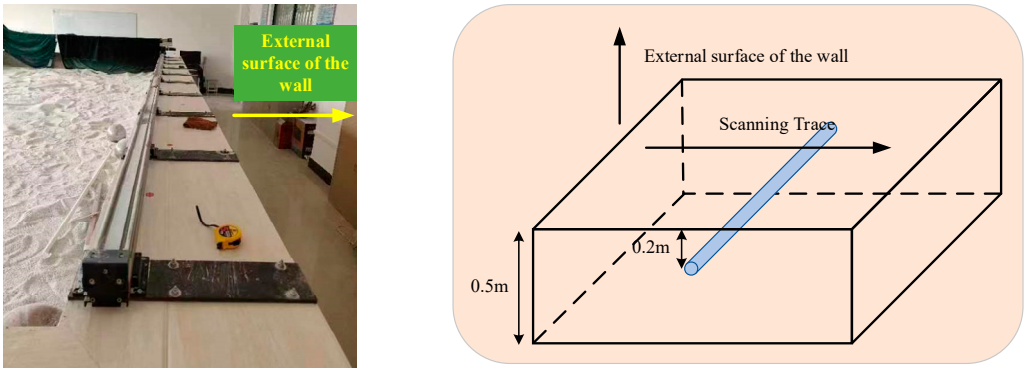
Figure 11. Inversion results of permittivity values at cutting lines ① to ⑦ of actual measurement data using different networks.

4.7. Measured Data Experiments in Different Geographical Environments

To further validate the network's generalization ability in inverting different measured data, we collected measured data from real scenarios that differ from those in Section 4.6 for experimentation. The experiment utilized a GPR from Y-Line, model YL-GPR400M, for data collection, with the antenna center frequency set at 400 MHz. The experimental scenario is depicted in Figure 12a, which shows a concrete wall with a thickness of 0.5 m, within which a solid steel bar with a diameter of 5 cm is embedded, located 0.2 m from the external surface of the wall. Behind the inner side of the wall is dry sand. Data acquisition was conducted along the external surface of the wall using radar, as shown in the experimental schematic in Figure 12b.

The collected data were preprocessed using the same method as the open-source measured datasets (i.e., normalization, resizing, and mean subtraction). Figure 13 presents the denoising and inversion results for the measured steel bar data. In Figure 13b,c, there

are no significant differences in the denoised images obtained by MHInvNet and DMRF-UNet, as both methods effectively suppress the background clutter in the input B-scan. Figure 13e,f indicate that both MHInvNet and DMRF-UNet exhibit some deviation in the inversion at the target location, demonstrating horizontal and depth shifts compared to the actual target. Specifically, the target outline obtained by DMRF-UNet is unclear and irregularly shaped, with an uneven texture. It also produces nonexistent false targets. By contrast, MHInvNet performs target inversion relatively well. Although the dimensions of the inverted target are slightly smaller than those of the actual target, its overall performance still surpasses that of DMRF-UNet, demonstrating better generalization capability.



(a) Experimental scenario (b) Schematic view of the experiment scenario

Figure 12. Measurement experiments of measured data in different geographical environments.

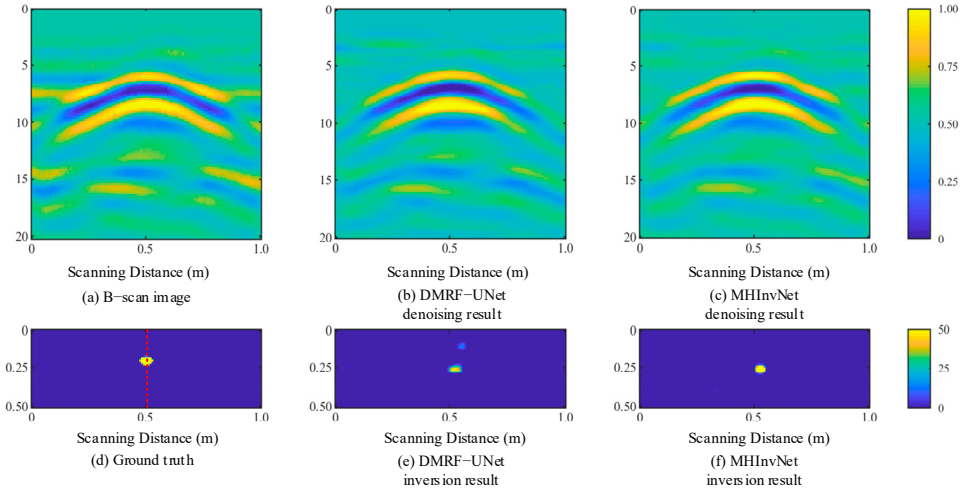


Figure 13. The denoising and inversion results for the measured data of steel bars by different networks.

The accuracy analysis of the permittivity inversion along the cutting lines shown in Figure 13d is presented in Figure 14. Due to the positional inversion deviations of MHInvNet and DMRF-UNet, the permittivity curves of both models do not align well with the true model curve in the horizontal coordinate direction. Regarding the values of permittivity, the maximum permittivity of the targets included in the measured dataset used for weight fine-tuning by both networks is 50. By contrast, the permittivity of the steel bar theoretically approaches infinity. Therefore, the correct value obtained for the permittivity of the steel bar during inversion is theoretically the maximum value that the network can invert, which is 50. It can be concluded that since the peak value of the permittivity curve obtained by MHInvNet reaches the theoretically correct value of 50, while that of DMRF-UNet is less than 50, the accuracy of MHInvNet's permittivity inversion is higher, showing good generalization capability in the specific inversion of permittivity values.

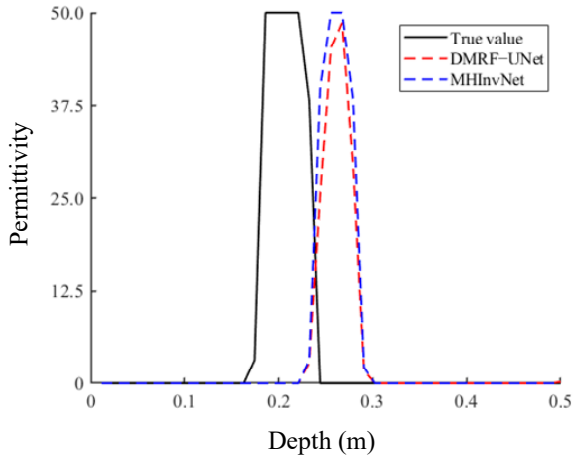


Figure 14. Inversion results of permittivity values for the measured data of steel bars by different networks.

5. Discussion

This study proposes an underground target permittivity inversion network MHInvNet for nonuniform soil environments to reduce the negative impact of clutter interference on inversion accuracy through a strategy of denoising followed by inversion. Introducing the MSDC and HAG modules enhanced network performance. Experimental validations were also conducted on both simulated and measured data. The results of the study are discussed as follows:

(1) The MSDC module consists of dilated convolutions with different receptive field sizes. Dilated convolutions allow for an increase in the receptive field of the convolution kernel without introducing extra parameters, thus reducing network complexity to some extent. Convolution kernels with larger receptive fields can capture global features with high-level semantic information in the image. By contrast, smaller receptive fields focus more on local detail features. MSDC learns the correlations of adjacent A-scans more comprehensively by extracting and fusing feature maps of different scales;

(2) The HAG module comprises CAG and SAG, which combine channel attention and spatial attention to weight the feature maps from skip connections. This process highlights important features related to the target signal while suppressing irrelevant features such as clutter and noise. By adjusting the input feature maps using gated signals from lower-level decoders, CAG and SAG further enhance the alignment of weights in feature maps of two different sizes, which guides the feature maps from the skip connection to some extent;

(3) Experimental verification using simulated and measured data shows that the proposed network achieves performance improvements in both denoising and inversion, outperforming existing inversion networks. To validate the generalization ability of the network, measured data from different geographical environments were collected for inversion. The results indicate that the proposed network can achieve high-quality inversions for measured data in various geographical contexts, demonstrating excellent generalization ability. Furthermore, visualization studies provide an in-depth analysis of the impact of the MSDC and HAG modules on network feature extraction, offering a clear understanding of their working principles;

(4) The proposed network can effectively invert the permittivity of targets in underground nonuniform background media, reducing noise and clutter interference to some extent. However, when faced with harsh and complex underground environments,

the collected B-scan images may contain strong interference signals. The noise and clutter intensity exceeds the capacity of the denoising network, rendering it unable to effectively denoise the B-scan and subsequently leading to failure in the inversion network. To overcome this limitation, it is necessary to study how to achieve high-quality inversion of B-scan images under low signal-to-noise ratio and low signal-to-clutter ratio conditions. Additionally, the types of underground targets that the proposed network can invert depend on the sample size of the training dataset; therefore, to achieve accurate inversions for more types and forms of underground targets, it is essential to collect higher-quality and more diverse data to build the dataset and enhance network learning.

6. Conclusions

This study proposes a two-stage GPR image inversion network MHInvNet based on the U-Net structure. Firstly, the first-stage network denoises the input B-scan images and then concatenates the denoised B-scan images with the undenoised B-scan images as inputs for the second-stage network for inversion, obtaining permittivity maps of underground targets. The MSDC module is introduced at both stages of the network to extract features from multiple scales. It merges them to capture the correlation between adjacent A-scans in the B-scan. The HAG module was also introduced to sequentially perform channel-wise attention weighting and spatial attention weighting on the feature maps of skip connections to reduce potential interference and enhance the network's ability to learn important feature information. Comparative experiments with existing methods show that the proposed network achieves good denoising and inversion effects on both simulated and actual measurement data. This study aims to invert the permittivity of targets buried in nonuniform soil environments. In future work, we will further research the inversion tasks of background media permittivity and more complex targets such as underground voids, cracks, and cavities.

Author Contributions: Conceptualization, M.W. and Q.L.; methodology, M.W. and Q.L.; validation, M.W. and Q.L.; writing—original draft, M.W.; writing—review and editing, Q.L. and S.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 62361015 and the Guangxi Special Fund Project for Innovation-Driven Development under Grant AA21077008.

Data Availability Statement: The original data presented in the study is open available in Google Drive at https://drive.google.com/drive/folders/1s_C7Cfp5XlbWF-MjW1z0XIaiq-zpNltN?usp=sharing, access date 9 August 2024. The original author of the data production is Qiqi Dai.

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

References

1. Solla, M.; Pérez-Gracia, V.; Fontul, S. A Review of GPR Application on Transport Infrastructures: Troubleshooting and Best Practices. *Remote Sens.* **2021**, *13*, 672. [CrossRef]
2. Tolooiyan, A.; Dyson, A.P.; Karami, M.; Shaghaghi, T.; Ghadrdan, M. Application of Ground Penetrating Radar (GPR) to Detect Joints in Organic Soft Rock. *Geotech. Test. J.* **2019**, *42*, 257–274. [CrossRef]
3. Wang, Z.; Wan, B.; Han, M. A Three-Dimensional Visualization Framework for Underground Geohazard Recognition on Urban Road-Facing GPRData. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 668. [CrossRef]
4. Urban, T.M.; Leon, J.F.; Manning, S.W.; Fisher, K.D. High resolution GPR mapping of Late Bronze Age architecture at Kalavasos-Ayios Dhimitrios, Cyprus. *J. Appl. Geophys.* **2014**, *107*, 129–136. [CrossRef]
5. Chen, D.P.; Dai, Q.W.; Feng, D.S.; Wang, H.H.; Zhang, B. Reverse time migration of ground penetrating radar based on normalized cross correlation imaging condition. *J. Cent. South Univ. (Sci. Technol.)* **2018**, *49*, 7. (In Chinese)

6. Feng, D.; Cao, C.; Wang, X. Multiscale Full-Waveform Dual-Parameter Inversion Based on Total Variation Regularization to On-Ground GPR Data. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9450–9465. [CrossRef]
7. Irving, J.D.; Knight, R.J. Removal of wavelet dispersion from ground-penetrating radar data. *Geophysics* **2003**, *68*, 960–970. [CrossRef]
8. Alvarez, J.K.; Kodagoda, S. Application of deep learning image-to-image transformation networks to GPR radargrams for sub-surface imaging in infrastructure monitoring. In Proceedings of the 13th IEEE Conference on Industrial Electronics and Applications, Wuhan, China, 31 May–2 June 2018; pp. 611–616.
9. Xie, L.; Zhao, Q.; Ma, C.; Liao, B.; Huo, J. Ü-net: Deep-learning schemes for ground penetrating radar data inversion. *J. Environ. Eng. Geophys.* **2020**, *25*, 287–292. [CrossRef]
10. Liu, B.; Ren, Y.; Liu, H.; Xu, H.; Wang, Z.; Cohn, A.G.; Jiang, P. GPRInvNet: Deep learning-based ground-penetrating radar data inversion for tunnel linings. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8305–8325. [CrossRef]
11. Ji, Y.; Zhang, F.; Wang, J.; Wang, Z.; Jiang, P.; Liu, H.; Sui, Q. Deep neural network-based permittivity inversions for ground penetrating radar data. *IEEE Sens. J.* **2021**, *21*, 8172–8183. [CrossRef]
12. Wang, Y.; Qin, H.; Miao, F. A Multi-Path Encoder Network for GPR Data Inversion to Improve Defect Detection in Reinforced Concrete. *Remote Sens.* **2022**, *14*, 5871. [CrossRef]
13. Dai, Q.; Lee, Y.H.; Sun, H.-H.; Ow, G.; Yusof, M.L.M.; Yucel, A.C. Dmrf-unet: A two-stage deep learning scheme for gpr data inversion under heterogeneous soil conditions. *IEEE Trans. Antennas Propag.* **2022**, *70*, 6313–6328. [CrossRef]
14. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional net-works for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
15. Holschneider, M.; Kronland-Martinet, R.; Morlet, J.; Tchamitchian, P. *A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform*; Springer: Berlin/Heidelberg, Germany, 1989; pp. 286–297.
16. Yu, F.; Koltun, V. Multi-Scale Context Aggregation by Dilated Convolutions. In Proceedings of the ICLR, San Juan, Puerto Rico, 2–4 May 2016.
17. Xu, Q.; Zhu, H.; Fan, H.; Zhou, H.; Yu, G. Study on detection of steel plate surface defects by improved YOLOv3 network. *Comput. Eng. Appl.* **2020**, *56*, 265–272.
18. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
19. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.

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.