

Clutter Suppression in Ground-Penetrating Radar Images Using an Improved Pix2Pix Model

Zewen Zi, Ying Liu *, Yuxin Ji, Yanyan Zang

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, 300222, China

* Corresponding author: Ying Liu (Email: liuying@tute.edu.cn)

Abstract: To address complex clutter in measured ground-penetrating radar (GPR) B-scan images and the difficulty of obtaining paired clutter-free target-response ground-truth data, this paper proposes an improved Pix2Pix clutter suppression method based on a paired dataset fused by simulated target responses and measured backgrounds. In terms of network structure, residual dense blocks (RDB) are introduced to strengthen feature transmission and reuse. A Convolutional Block Attention Module (CBAM) is adopted to adaptively recalibrate channel-wise and spatial features. A multi-scale structural similarity (MS-SSIM) constraint is added to the generator objective function to improve the structural integrity and continuity of target hyperbolic responses. To control confounding effects caused by weight differences of the L1 loss, comparative experiments are carried out under $\lambda_{L1} = 100$ and $\lambda_{L1} = 50$ respectively. The experimental results show that under $\lambda_{L1} = 100$, the CBAM-MS-SSIM model, Exp3 achieves the best performance within the current data partition and parameter range. Its test-set PSNR, SSIM and MS-SSIM reach 52.4344 dB, 0.998056 and 0.998774 respectively. Further ablation experiments indicate that introducing either CBAM or MS-SSIM alone does not improve performance. When the two modules are combined under a strong L1 reconstruction constraint, they improve target reconstruction, structure preservation, and clutter suppression.

Keywords: Ablation Experiment; CBAM; Clutter Suppression; Ground Penetrating Radar; MS-SSIM; Pix2Pix.

1. Introduction

Internal cracks in road structures are important hidden defects that affect the service performance and durability of pavements. Ground Penetrating Radar (GPR) enables rapid, continuous, and non-contact inspection. However, crack targets are typically small and produce weak reflections. Their hyperbolic responses are often obscured by background clutter and noise. For GPR clutter suppression, Zhao et al. introduced total-variation regularization into low-rank-sparse decomposition to separate background clutter from target responses [1], yet their method depends on specific low-rank or sparse priors. In recent years, deep learning has provided an alternative data-driven approach to GPR image processing. The Pix2Pix model proposed by Isola et al. learns mapping relations between paired images through joint conditional adversarial loss and L1 loss [2]. Still, its U-Net generator has room for improvement in representing complex hierarchical features and weak fine-scale structures. The Residual Dense Block (RDB) proposed by Zhang et al. enhances feature transmission and reuse through dense connections and local residual learning [3]. In addition, Huang et al. applied the Convolutional Block Attention Module (CBAM) and multi-scale structural similarity (MS-SSIM) constraint to GPR image restoration [4].

Because clean paired ground-truth data for measured GPR images are difficult to obtain, this study constructs a paired simulation-measurement dataset by combining simulated target responses with measured target-free backgrounds. On this basis, improvements are made to the Pix2Pix generator. RDB is used to enhance feature reuse capability. CBAM strengthens features of crack regions. The MS-SSIM loss constrains multi-scale target structures. Thus, the capacity of clutter suppression and preservation of crack responses under complex backgrounds is improved.

2. Dataset Construction

The training of Pix2Pix relies on pixel-wise aligned paired images. Because high-quality paired measured GPR data—namely, cluttered input images and their corresponding clutter-free target-response ground-truth images—are extremely scarce, this study constructs a paired dataset by fusing simulated target responses with measured backgrounds [5].

2.1. Construction of Simulated Targets

This paper builds a four-layer model: “air-asphalt-base layer-subgrade”. The crack target is a vertical rectangular air-filled medium anomaly with random position, width and height. Figure 1 shows the model with cracks fixed in the base layer. Table 1 lists material parameters of the model.

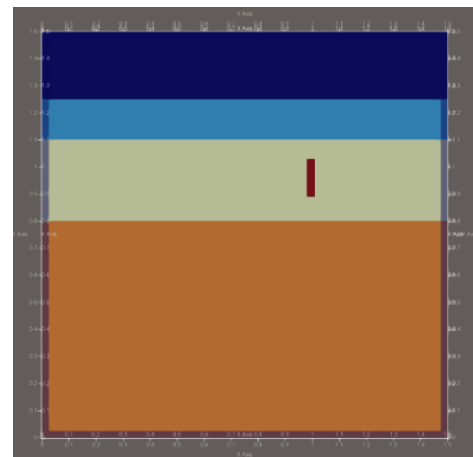


Figure 1. Forward-simulation GPR model of cracks in underground four-layer structure

Table 1. Forward-simulation GPR model parameters for cracks in four-layer underground structure

Layer No.	Name	Thickness	Relative Permittivity	Electrical Conductivity
1	Air	0.25	1	0
2	Asphalt	0.15	4	0.005
3	Base Course	0.3	8	0.004
4	Subgrade	0.8	10	0.003

2.2. Extraction of Simulated Target Responses

This paper uses gprMax to generate paired simulated data (with-target and without-target) sharing identical model parameters and acquisition geometry. Element-wise subtraction is applied to paired with-target and target-free B-scans. Common background components are weakened so that images containing only target responses can be obtained.

Subtraction is completed in the original numerical domain. Regional interpolation is adopted to resize results uniformly to 320×320 and generate target image T:

$$T = X_{\text{with-target}} - X_{\text{without-target}} \quad (1)$$

2.3. Fusion with Measured Backgrounds

Cleaned simulated target T is fused randomly with measured clutter-free background B. Input image A containing real-world background statistical characteristics is generated. The corresponding clean target image is used as the target label Y:

$$A = B + \alpha T \quad (2)$$

$$Y = \mathcal{N}(T) \quad (3)$$

Here, α controls the relative amplitude between targets and measured backgrounds. Input A and label Y are concatenated in left-to-right order to produce AB images required by the aligned mode. Figure 2 shows fused AB images. Figure 3 illustrates the workflow for dataset construction.

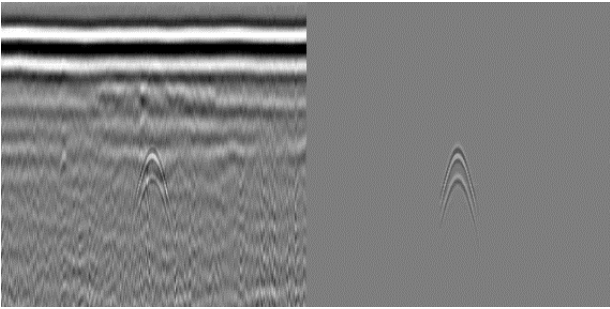


Figure 2. Measured backgrounds fused with simulated targets

3. Improved Pix2Pix Framework

Based on the dataset described above, this paper improves the generator structure and loss function of Pix2Pix. As shown in Figure 4, RDB and CBAM are embedded in U-Net. Adversarial loss, L1 loss and MS-SSIM loss are used jointly. Each improved module is explained below.

3.1. Residual Dense Block (RDB)

To mitigate gradient vanishing in deep networks and maximize feature utilization, residual dense blocks are integrated into both encoder and decoder stages of U-Net. Each RDB consists of densely connected convolution layers. The input of each layer comes from feature-map concatenation of all preceding layers.

Assume the input of one RDB is F_0 . It contains L convolution layers inside:

$$F_l = H_l([F_0, F_1, \dots, F_{l-1}]) \quad (4)$$

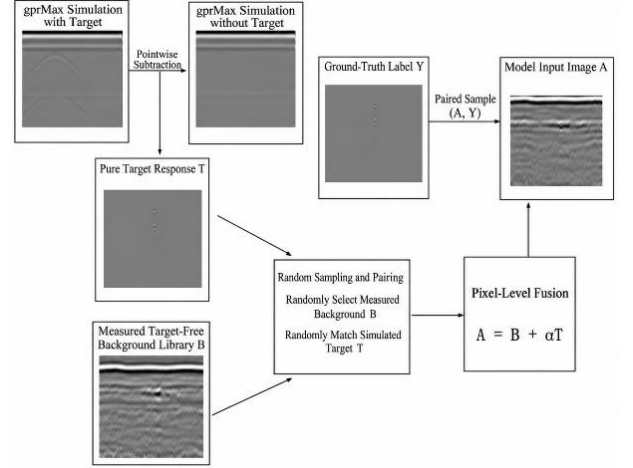


Figure 3. Workflow of dataset construction

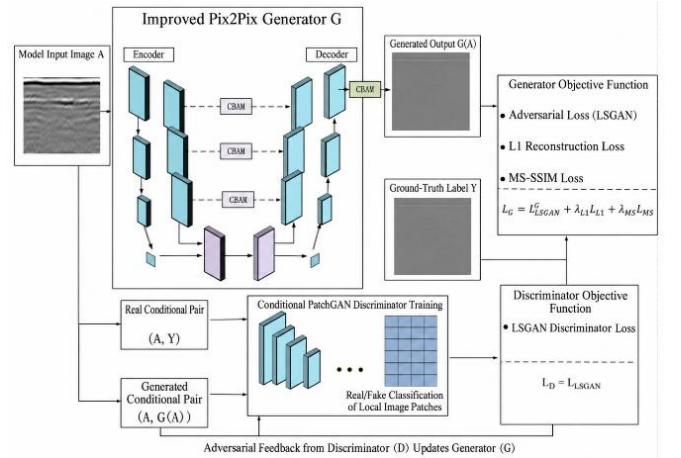


Figure 4. Overall architecture of improved Pix2Pix

Finally, a 1×1 convolution compresses features. Residual connection adds compressed features to the original input F_0 to get RDB output:

$$F_{\text{out}} = F_0 + H_{1 \times 1}([F_0, F_1, \dots, F_L]) \quad (5)$$

3.2. Convolutional Block Attention Module (CBAM)

Crack features in GPR images are often covered by strong background clutter. CBAM is introduced so that networks can automatically focus on these weak target regions [6].

For input feature map $F \in \mathbb{R}^{C \times H \times W}$, channel attention aggregates spatial information via global average pooling and global max pooling. Two channel descriptors are obtained and fed into shared multi-layer perceptrons (MLP) to produce channel-attention weight map $M_c \in \mathbb{R}^{C \times 1 \times 1}$:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (6)$$

Multiply M_c by F channel-wise to obtain the channel-enhanced feature $F' = M_c(F) \otimes F$. Afterwards, spatial attention conducts average pooling and max pooling over F' along the channel dimension. Pooling outputs are concatenated and processed by a 7×7 convolution kernel to produce spatial-attention weight map $M_s \in \mathbb{R}^{1 \times H \times W}$:

$$M_s(F') = \sigma(f^{7 \times 7}([\text{AvgPool}(F'); \text{MaxPool}(F')])) \quad (7)$$

The final output is $F'' = M_s(F') \otimes F'$. Figure 5 presents the architecture of the CBAM module.

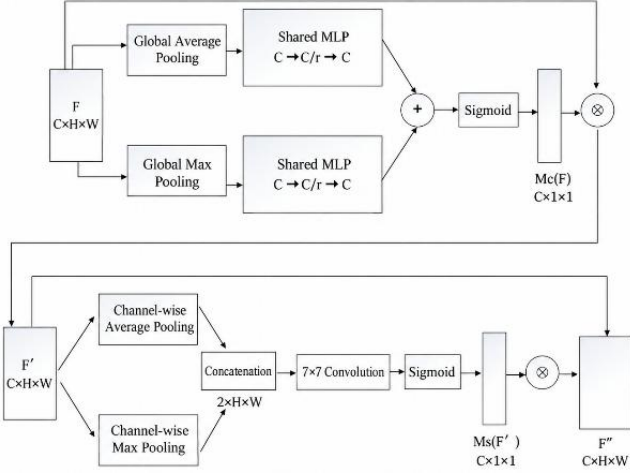


Figure 5. Structure of CBAM module

3.3. MS-SSIM Structural Loss

Cracks in road-structure GPR B-scan images usually appear as faint hyperbolic responses. Traditional pixel-wise losses cannot fully constrain structural features. They may

cause blurred edges or broken arcs. MS-SSIM loss is introduced to improve structural fidelity of generated outputs [7]. Its form is defined as $1 - \text{MS} - \text{SSIM}(G(A), Y)$. It evaluates luminance, contrast and structural consistency between generated images and ground-truth labels at five different scales. In training: $L_{\text{MS}} = 1 - \text{MS} - \text{SSIM}(G(A), Y)$. The weight coefficient λ_{MS} controls its contribution.

To account for pixel reconstruction accuracy, multiscale structural information, and discriminator regularization, different objective functions are employed to optimize the generator and discriminator. The generator loss is expressed as:

$$L_G = L_{\text{adv}}^G + \lambda_{L1} L_{L1} + \lambda_{\text{MS}} L_{\text{MS}} \quad (8)$$

The discriminator is optimized using the following loss function:

$$L_D = L_{\text{adv}}^D + \lambda_{\text{GP}} L_{\text{GP}} \quad (9)$$

The L1 loss constrains the absolute pixel-wise differences between the generated and ground-truth images, thereby helping preserve intensity fidelity and spatial alignment.

4. Ablation Experiments

4.1. Training Parameters

All key experiments adopt aligned data mode, batch size 4 and batch normalization. Candidate check-points for validation include epoch 40, epoch 50, epoch 80 and epoch 100.

For each candidate checkpoint, PSNR, SSIM and MS-SSIM are calculated on the validation set. The best epoch is selected following the priority rule: PSNR first, then SSIM, then MS-SSIM.

4.2. Test-Set Results

Table 2. Test-set results of six groups of experiments under $L1 = 100$

Experiment	Configuration	PSNR	SSIM	MS-SSIM
Exp0	Baseline	44.4596±0.5040	0.988214±0.001131	0.995387±0.001047
Exp1	RDB	49.8374±1.1778	0.996333±0.001209	0.998064±0.000622
Exp2	RDB+CBAM	46.6599±0.9133	0.992166±0.001755	0.997014±0.000927
Exp3a	RDB+MS-SSIM	49.9296±0.9829	0.995754±0.000990	0.997988±0.000503
Exp3	RDB+CBAM+MS-SSIM	52.4344±1.0052	0.998056±0.000729	0.998774±0.000324
Exp4	RDB+CBAM+MS-SSIM+GP	49.5184±0.4668	0.994913±0.000361	0.998909±0.000280

Table 3. Test-set results of six groups of experiments under $L1 = 50$

Experiment	Configuration	PSNR	SSIM	MS-SSIM
Exp0	Baseline	43.9712±1.0527	0.984934±0.004996	0.994849±0.002052
Exp1	RDB	49.5349±2.0177	0.996744±0.001325	0.998124±0.000749
Exp2	RDB+CBAM	47.8285±1.0420	0.993063±0.001070	0.997340±0.000687
Exp3a	RDB+MS-SSIM	48.4757±1.2054	0.996463±0.001269	0.997630±0.000749
Exp3	RDB+CBAM+MS-SSIM	48.2080±0.8866	0.993512±0.000872	0.997642±0.000607
Exp4	RDB+CBAM+MS-SSIM+GP	48.5948±0.4820	0.994325±0.000606	0.998366±0.000405

From Table 2, under $\lambda_{L1} = 100$, Exp1 with RDB achieves a substantial performance improvement over the original Pix2Pix baseline, Exp0. This result indicates that RDB effectively enhances feature extraction and reconstruction capacity. Ablation results show that using CBAM or MS-SSIM alone does not improve the metrics and even degrades performance. When both modules are adopted simultaneously (Exp3), Exp3 achieves the highest PSNR and SSIM, whereas Exp4 achieves the highest MS-SSIM. According to the predefined priority rule of PSNR, SSIM, and

MS-SSIM, Exp3 is selected as the best-performing configuration. This finding indicates that the combined use of CBAM and MS-SSIM improves target reconstruction and structure preservation.

From Table 3, after lowering the L1 weight, Exp3 loses its performance advantage over Exp1. This result indicates that the joint improvement depends on the strength of the L1 reconstruction constraint. Under current dataset settings, Exp3 ($L1 = 100$) is the best-performing configuration under the predefined selection criterion. Compared with Exp0

($L1 = 100$), its PSNR rises by 7.9748 dB. SSIM and MS-SSIM increase by 0.009842 and 0.003387 respectively.

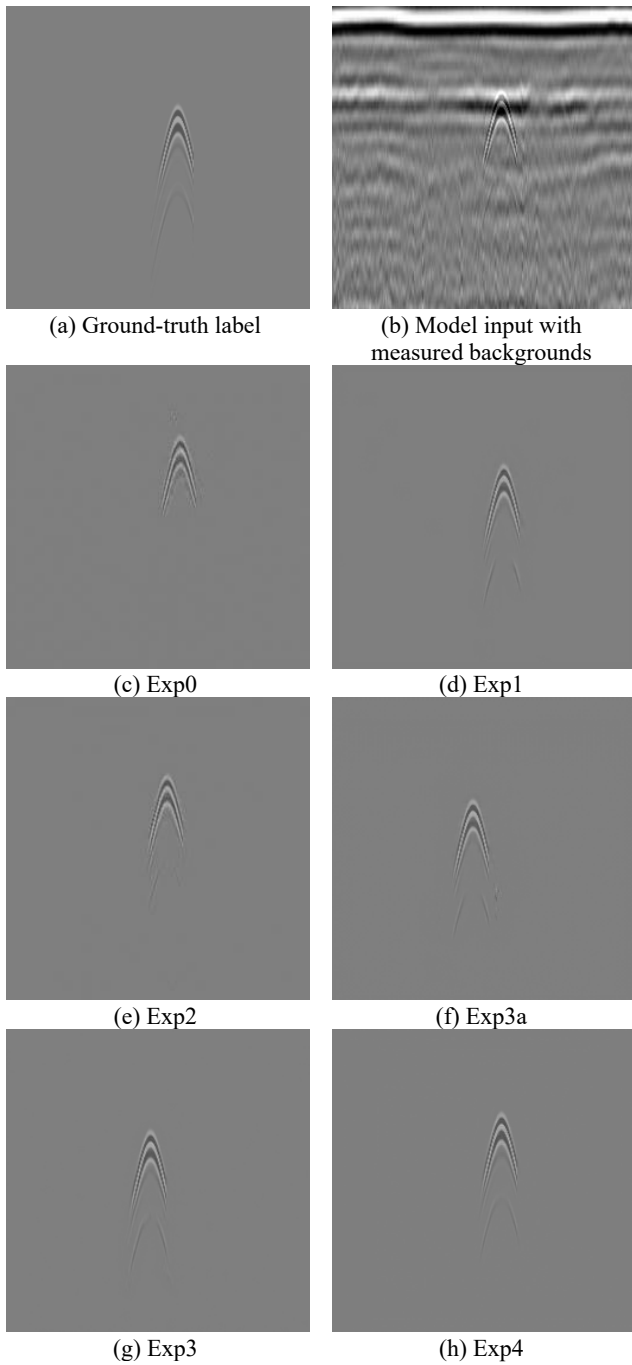


Figure 6. Visual comparison of clutter-suppression outputs from different models under $L1 = 100$

As shown in Figure 6, Exp0 can recover basic outlines of target hyperbolas but leaves local artifacts. After introducing RDB in Exp1, target structures and weak-response layers become more complete. Exp2 with CBAM produces smooth backgrounds yet brings some distortions to target shapes. Exp3a and Exp4 keep major target structures but have defects in fine local details. In contrast, Exp3 combining CBAM and MS-SSIM achieves a better overall balance among target positioning, arc continuity, multilayer-response restoration, and background clutter suppression. Its outputs are closest to the constructed simulated target labels.

5. Conclusion

To enhance crack responses in road-structure GPR data, this study uses gprMax to construct a four-layer forward model containing vertical rectangular cracks and generate simulated data. Target responses are extracted from simulation outputs and fused with randomly selected measured target-free backgrounds. Input images carrying real-world clutter characteristics are generated, while pure target responses are used as target labels to produce paired data for supervised training. Based on the Pix2Pix framework, residual dense blocks (RDBs), a convolutional block attention module (CBAM), a multiscale structural similarity loss (MS-SSIM), and a gradient penalty are introduced to improve the reconstruction of faint hyperbolic structures and stabilize training. The ablation results show that adding CBAM alone does not improve performance at either $L1$ weight. Applying MS-SSIM alone provides only limited benefits. When $\lambda_{L1} = 100$, the joint use of CBAM and MS-SSIM produces clear improvements, whereas this advantage is not observed when $\lambda_{L1} = 50$. Within the constructed paired dataset, the proposed model improves the reconstruction of weak crack responses; however, its applicability to engineering practice requires further validation using measured GPR data containing real crack targets and downstream detection results.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62241109), and Scientific Research Fund of Tianjin Education Commission (Grant No. 2021KJ005).

References

- [1] Zhao, Y., Yang, X., Qu, X., Lan, T., & Gong, J. (2023). Clutter removal method for GPR based on low-rank and sparse decomposition with total variation regularization. *IEEE Geoscience and Remote Sensing Letters*, 20, 3502605, 1–5. <https://doi.org/10.1109/LGRS.2023.3250717>.
- [2] Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004*. <https://doi.org/10.48550/arXiv.1611.07004>.
- [3] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [4] Huang, X., Liu, J., Yang, F., Qiao, X., Gao, L., Fu, T., & Zhao, J. (2025). Study on GPR image restoration for urban complex road surfaces using an improved CycleGAN. *Remote Sensing*, 17(5), 823. <https://doi.org/10.3390/rs17050823>.
- [5] Ni, Z.-K., Shi, C., Pan, J., Zheng, Z., Ye, S., & Fang, G. (2022). Declutter-GAN: GPR B-scan data clutter removal using conditional generative adversarial nets. *IEEE Geoscience and Remote Sensing Letters*, 19, 4023105, 1–5. <https://doi.org/10.1109/LGRS.2022.3159788>.
- [6] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- [7] Wang, Z., Simoncelli, E. P., & Bovik, A. C. (2003). Multiscale structural similarity for image quality assessment. In *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers (Vol. 2, pp. 1398–1402)*. <https://doi.org/10.1109/ACSSC.2003.1292216>.