Learning from Clutter: An Unsupervised Learning-Based Clutter Removal Scheme for GPR B-Scans

Qiqi Dai, Yee Hui Lee, *Senior Member, IEEE*, Hai-Han Sun, Jiwei Qian, Mohamed Lokman Mohd Yusof, Daryl Lee, and Abdulkadir C. Yucel, *Senior Member, IEEE*

***Abstract*—Ground-penetrating radar (GPR) data are often contaminated by hardware and environmental clutter, which significantly affects the accuracy and reliability of target response identification. Existing supervised deep learning techniques for removing clutter in GPR data require generating a large set of clutter-free B-scans as labels for training, which are computationally expensive in simulation and unfeasible in real-world experiments. To tackle this issue, we propose a two-stage unsupervised learning-based clutter removal scheme, called ULCR-Net, to obtain clutter-free GPR B-scans. In the first stage of the proposed scheme, a diffusion model tailored for GPR data augmentation is employed to generate a diverse set of raw B-scans from the input random noise. With the augmented dataset, the second stage of the proposed scheme uses a contrastive learning-based generative adversarial network to learn and estimate clutter patterns in the raw B-scan. The clutter-free B-scan is then obtained by subtracting the clutter pattern from the raw B-scan. The training of the two-stage network only requires a small set of raw B-scans and clutter-only B-scans that are readily available in real-world applications. Extensive experiments have been conducted to validate the effectiveness of the proposed method. Results on simulation and measurement data demonstrate that the proposed method has superior clutter removal accuracy and generalizability and outperforms existing algebraic techniques and supervised learning-based methods with limited training data by a large margin. With its high clutter suppression capability and low training data requirements, the proposed method is well-suited to remove clutter and restore target responses in real-world GPR applications. Upon paper acceptance, the proposed method's code and dataset will be released at *https://github.com/Qiqi-Dai* .**

***Index Terms*— Clutter removal, contrastive learning, data augmentation, diffusion model, generative adversarial network, ground-penetrating radar, unsupervised learning.**

I. Introduction

GROUND-PENETRATING radar (GPR) has been widely used as a non-destructive technique for detecting and imaging subsurface objects. However, system and environmental clutter arise from direct coupling between the transmitter and receiver, strong reflections from the ground surface, and the heterogeneity of the soil environment, which hinders the recognition of the target responses in the obtained B-scan. Therefore, suppressing clutter while retaining the signatures in the GPR B-scan is of great importance for target detection and interpretation [1]-[2].

Over the years, many conventional algebraic algorithms have been investigated to remove clutter in GPR B-scans. The mean subtraction (MS) methods [3]-[5] estimate the clutter as the average of all the traces and subtract it from the raw GPR B-scan. Subspace projection methods, including the singular value decomposition (SVD) [6]-[7], independent component analysis (ICA) [8]-[10], and principal component analysis (PCA) [10]-[11], decompose the raw B-scan into clutter subspace and target subspace based on the strength difference. Low-rank sparse matrix decomposition-based methods, such as the robust PCA (RPCA) [12]-[13], non-negative matrix factorization (NMF) [14], robust NMF (RNMF) [15], and robust autoencoder (RAE) [16], divide the raw B-scan into a low-rank matrix containing the clutter component and a sparse matrix capturing the target component. Space representation-induced methods, such as morphological component analysis (MCA) [17] and dictionary learning [18]-[19], sparsely differentiate the target and clutter components using the dictionaries. However, these conventional clutter removal algorithms suffer from several limitations. 1) Although they are effective in ideal homogeneous subsurface settings, their clutter removal accuracy decreases in diverse real-world heterogeneous environments, and sometimes they even yield distorted object responses. 2) The precise subspace separation is necessary in the subspace projection methods, which can be challenging when the clutter and target responses have similar strengths. 3) The decluttering performance of some of these algorithms highly depends on the hyperparameter selection for different cases. 4) The iterative computations in the optimization process can be time-consuming.

This work was supported by the Ministry of National Development Research Fund, National Parks Board, Singapore, with contract number of COT-V4-2020-6 (*Corresponding authors: Yee Hui Lee; Abdulkadir C. Yucel*).

Q. Dai, Y. H. Lee, Jiwei Qian, and A. C. Yucel are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mails: daiq0004@e.ntu.edu.sg; eyhlee@ntu.edu.sg; qian0069@e.ntu.edu.sg; acyucel@ntu.edu.sg).

H. -H. Sun is with the Department of Electrical and Computer Engineering, University of Wisconsin-Madison, 1415 Engineering Dr, Madison, WI 53706, United States (e-mail: haihan.sun@wisc.edu).

M. L. M. Yusof and D. Lee are with the National Parks Board, Singapore 259569 (e-mails: mohamed_lokman_mohd_yusof@nparks.gov.sg; daryl_lee@nparks.gov.sg).

Recently, due to the impressive capability of learning and extracting data features, deep learning technology has been increasingly applied to process GPR images [20]-[24]. To solve the limitations of conventional algebraic techniques for removing clutter in GPR B-scans, supervised deep learning methods have been developed with enhanced accuracy and efficiency [25]-[31]. These methods employ convolutional encoder-decoder networks [25]-[27], improved U-Net structures [28]-[30], and conditional generative adversarial networks (cGAN) [31] to learn feature representations of object reflections from raw B-scans and generate clutter-free B-scans. To improve the decluttering performance on real measurement data, a hybrid data generation approach is further proposed, which combines measured clutter-only B-scans in real scenarios with simulated clutter-free B-scans to create synthetic raw B-scans for network training [29]-[31]. However, these methods are limited in two respects. 1) As supervised deep learning models, they require pairs of raw and clutter-free B-scans for training, yet obtaining clutter-free B-scans in real GPR measurements is unfeasible. Relying solely on simulated data, which may not fully capture the complexity of clutter, often leads to suboptimal clutter removal performance in real scenarios. 2) While the hybrid dataset improves the network's clutter removal performance in real-world scenarios, it demands thousands of samples in data sets, resulting in a time-consuming and labor-intensive data collection and generation process. Therefore, it is imperative to explore data augmentation methods and unsupervised learning approaches to reduce the demands on GPR datasets while achieving accurate clutter removal performance in real-world GPR applications.

To address the limitations of existing clutter removal methods, in this work, we present a two-stage Unsupervised Learning-based Clutter Removal scheme, called ULCR-Net, which first augments GPR dataset and then removes clutter in GPR B-scans. Extended from our previous study [32], the main contributions of this work are listed as follows.

1) Given the primary challenge in deep learning-based methods is the scarcity of datasets in real-world scenarios, the first stage employs a diffusion model to generate a diverse set of raw B-scans from a very small amount of real B-scan samples. The newly generated dataset, which is highly diverse and of high quality, is then used to train the clutter removal network in the second stage.
2) In the second stage, rather than directly involving clutter-free B-scans that are unfeasible to measure in real-world scenarios for training supervised learning models in the paired setting, we employ a contrastive learning-based GAN trained in the unpaired setting to learn the features of clutter and output the clutter-only B-scan from the input raw B-scan. The clutter-free B-scan is subsequently obtained by subtracting the clutter-only B-scan from the raw B-scan. Throughout the training of the proposed two-stage scheme, only a small set of raw and clutter-only B-scans is required, greatly easing the burden of dataset collection in real applications.
3) The performance of the proposed framework has been tested using both simulation and measurement data on the clutter removal accuracy and generalizability in different scenarios. Comparative studies demonstrate that our method outperforms traditional algebraic techniques and existing supervised learning-based methods, while also avoiding the burden of large labeled dataset generation required by supervised learning methods.

The application of the proposed ULCR-Net to the simulation and measurement data showed that the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), mean absolute error (MAE), mean relative error (MRE), and image quality score (IQS) reached 50.26 dB, 0.9948, 0.0021, 0.61%, and 0.2225, respectively. Compared to state-of-the-art clutter removal network based on supervised learning [30], these metrics have been improved by 10.63%, 0.59%, 30.00%, 43.52%, and 89.36%, respectively, indicating the superiority of the proposed clutter removal scheme for GPR B-scans.

The rest of this paper is organized as follows. Section II introduces the methodology, including the first-stage data augmentation and the second-stage clutter removal networks. Section III presents experimental results based on numerical simulated data to validate the effectiveness of the proposed scheme. Section IV presents the network performance on real measurement data. Conclusions are provided in Section V.

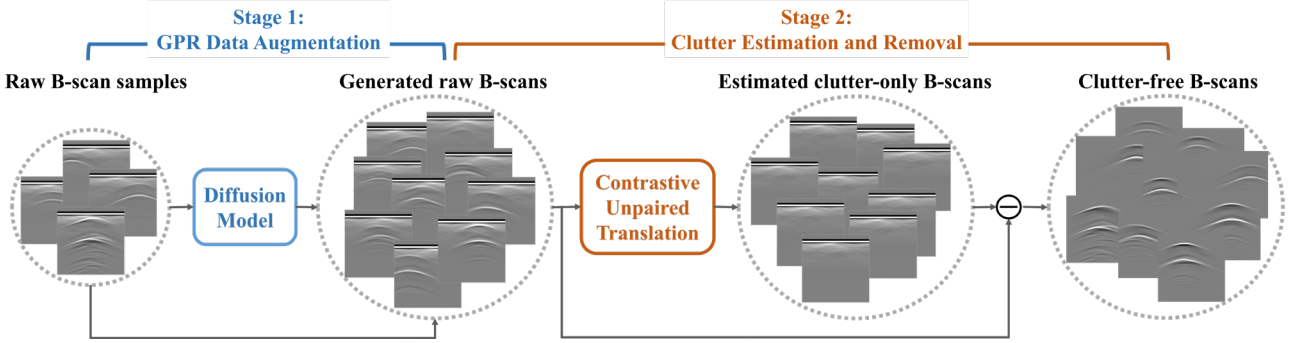

Fig. 1. Overview of the proposed two-stage framework ULCR-Net for removing clutter in GPR images based on unsupervised learning.

II. Methodology

A. Overall Framework

The framework of the proposed ULCR-Net is shown in Fig. 1. The first stage, designed for GPR data augmentation, employs a diffusion model to generate a large set of raw B-scans from a small set of real raw B-scans. These generated B-scans, combined with the original B-scans, form a diverse and high-quality training dataset for the second-stage clutter removal network. In the second stage, an unsupervised learning-based GAN using contrastive unpaired translation is designed to estimate the clutter-only B-scan from the input raw B-scan. Finally, the clutter-free B-scan containing only object signatures is obtained by subtracting the estimated clutter-only B-scan from the raw B-scan. The details of the two stages are introduced as follows. It should be noted that the italic symbols/letters refer to scalars, bold symbols/letters refer to two-dimensional metrices, italic bold symbols/letters refer to multi-dimensional arrays, script font set refers to data distributions, and double-struck font set refers to operators.

B. Stage 1: Diffusion Model-Based GPR Data Augmentation

In the first stage, a diffusion model-based data augmentation method is designed to generate a new set of raw B-scans from the provided small set of raw B-scans, enlarging the training dataset size for the second stage. Compared to classical deep generative models, the diffusion model synthesizes images of higher diversity and quality with better stability in the training process [33]. As illustrated in Fig. 2, the diffusion model includes a forward diffusion process and a reverse diffusion process. In the forward diffusion process, the provided raw B-scan is perturbed by successively adding Gaussian noise. In the reverse diffusion process, a deep neural network is trained to predict and remove the added Gaussian noise step by step and finally recover the input B-scan. Using the trained model, new B-scans resembling the provided ones are generated by feeding random noise images into the trained reverse diffusion process.

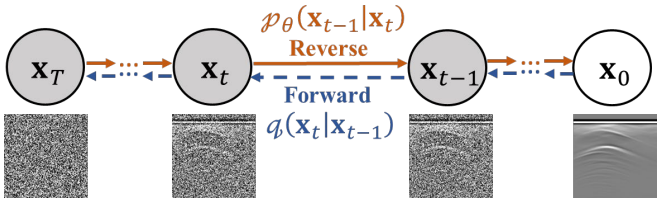

Fig. 2. The chain transitions of the forward and reverse diffusion process for generating the GPR B-scan.

Forward process. In the forward diffusion process, as shown in Fig. 2, given a B-scan image from the training data distribution $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, we add small Gaussian noise to the input B-scan in T steps and obtain a sequence of noised B-scan images $\mathbf{x}_1, \cdots, \mathbf{x}_T$. The input B-scan $\mathbf{x}_0$ gradually loses its distinguishable features as the step t increases. Combined with Markov assumption, the forward process can be expressed as [34]-[35]

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t \mathbf{I}\right), \quad (1)$$

where $\{\beta_t \in (0,1)\}_{t=1}^T$ is a variance schedule to ensure that the final image $\mathbf{x}_T$ is nearly an isotropic Gaussian distribution when T is sufficiently large, $\mathcal{N}(\mathbf{x}; \mu, \sigma)$ represents a normal distribution with mean μ and covariance σ generating the sample $\mathbf{x}$, and $\mathbf{I}$ is an identity matrix with the same dimension as the input B-scan.

Reverse process. Reversely, as shown in Fig. 2, we can generate a new B-scan from an input Gaussian noise image $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$ following the T reverse step $p(\mathbf{x}_{t-1}|\mathbf{x}_t)$. A deep neural network with the trainable parameters θ is trained to approximate each reverse step $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to restore the B-scan $\mathbf{x}_0$ after T steps. The reverse step is expressed as [34]-[35]

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(t, \mathbf{x}_t), \boldsymbol{\Sigma}_\theta(t, \mathbf{x}_t)), \quad (2)$$

where $\boldsymbol{\mu}_\theta(t, \mathbf{x}_t)$ and $\boldsymbol{\Sigma}_\theta(t, \mathbf{x}_t)$ represent the mean and covariance in the step t, respectively. Applying the reverse formula for all time steps, the data distribution becomes

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t). \quad (3)$$

With the known noise image $\mathbf{x}_T$ and the embedding at time step t, the network learns to reversely predict the noise and restore the B-scan.

Learning strategy. The training of a diffusion model is to find the reverse transitions that maximize the likelihood of the training data. Using the variational lower bound (VLB) combining with Kullback-Leibler (KL) divergence [34]-[37], the loss function at each time step τ_t can be expressed as

$$\tau_t = \mathbb{E}_{t\in[1,T], \mathbf{x}_0, \boldsymbol{\epsilon}_t}\left[\left\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta\left(\sqrt{\hat{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\hat{\alpha}_t}\boldsymbol{\epsilon}_t, t\right)\right\|^2\right], \quad (4)$$

where $\alpha_t = 1 - \beta_t$, $\hat{\alpha}_t = \prod_{i=1}^t \alpha_i$, $\mathbb{E}$ represents the expected value, $\boldsymbol{\epsilon}_t$ is the real noise at time step t, and $\boldsymbol{\epsilon}_\theta$ is the predicted noise by the neural network at time step t. This effectively measures the distance between the real noise and the predicted noise. Therefore, the network is trained to estimate the noise image at each time step and finally reconstruct the B-scan.

Network structure. A neural network is required to train a model for predicting $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$. As the input and output dimensions are identical, the U-Net architecture [38] is adopted to predict the noise at each time step. Specifically, the network consists of four encoding modules and four symmetrical decoding modules with skip connections. Each encoding module has two Wide ResNet blocks [39] and one down-sampling layer. Each decoding module has one up-sampling layer and two Wide ResNet blocks. The group normalization [40] is used in each block. Self-attention blocks are introduced between the two residual blocks at the 16×16 resolution [41]. The sinusoidal position embedding is incorporated into each residual block to specify the diffusion time step t [34]. Using the well-trained U-Net structure for T steps, B-scans are generated from input random noise images.

C. Stage 2: Unsupervised GAN-Based Clutter Estimation and Removal

Let $\mathbf{x}$ represent a raw B-scan measured by a GPR, which generally consists of object signatures $\mathbf{x}_s$ and clutter component $\mathbf{x}_c$, i.e., $\mathbf{x} = \mathbf{x}_s + \mathbf{x}_c$. The existing supervised learning schemes in [25]-[31] use a paired raw and clutter-free dataset $[\boldsymbol{X}, \boldsymbol{X}_s]$ to train a network to establish a mapping relationship between the input raw B-scan $\mathbf{x}$ and the output clutter-free B-scan $\mathbf{x}_s$. However, generating a large set of ground-truth clutter-free B-scans is time-consuming in

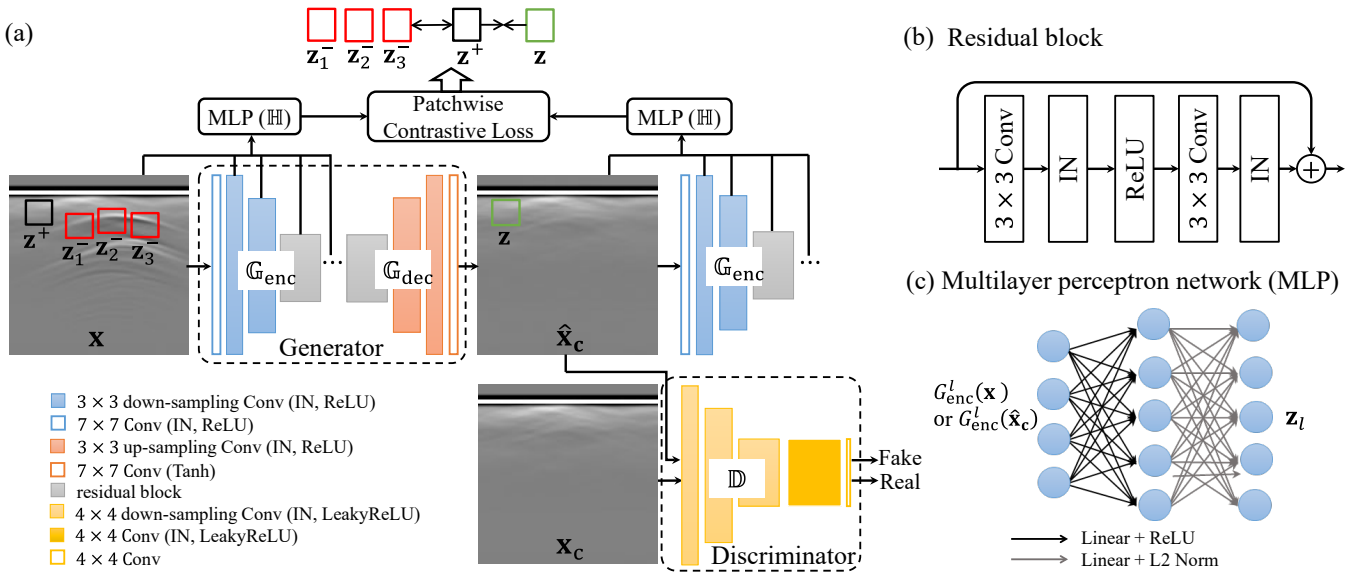

Fig. 3. (a) Architecture of the contrastive learning-based GAN for estimating the clutter-only B-scan $\hat{\mathbf{x}}_c$ from the input noisy B-scan $\mathbf{x}$. (b) Structure of the residual block. (c) Structure of the multilayer perceptron (MLP) network to obtain the projected features via $\mathbf{z}_l = \mathbb{H}_l(\mathbb{G}^l_{\text{enc}}(\mathbf{x} \text{ or } \hat{\mathbf{x}}_c))$. Note that 'Conv', 'IN', 'ReLU', 'Tanh', 'LeakyReLU', 'Linear', and 'L2 Norm' represent the convolution, instance normalization, ReLU activation, Tanh activation, Leaky ReLU activation, fully connected layer, and L2 normalization, respectively.

simulations and impractical in real measurements [29]-[31]. On the contrary, obtaining unpaired clutter-only B-scans is relatively achievable in real measurements and computationally inexpensive in simulations. Therefore, our approach is to train a contrastive learning-based GAN [42] to predict clutter-only B-scans $\hat{\mathbf{x}}_c$ based on an unpaired raw and clutter-only dataset $[\boldsymbol{X}, \boldsymbol{X}_c]$. Subsequently, the clutter-free B-scan can be obtained by subtracting $\hat{\mathbf{x}}_c$ from $\mathbf{x}$.

The architecture of the contrastive learning-based GAN is shown in Fig. 3. The GAN consists of an image generator $\mathbb{G}$ for constructing the clutter-only B-scan and a discriminator $\mathbb{D}$ for distinguishing the generated clutter-only B-scan $\hat{\mathbf{x}}_c$ and the real clutter-only B-scan $\mathbf{x}_c$. The adversarial learning between the generator and the discriminator enables the predicted clutter-only B-scan to be close to the ones in the provided training data. However, as the data sets used for training the network are not paired, there is no guarantee that the clutter distribution of the generated clutter-only B-scan $\hat{\mathbf{x}}_c$ corresponds to that of the input raw B-scan $\mathbf{x}$. Therefore, the contrastive learning between the features extracted from the generated clutter-only B-scan $\hat{\mathbf{x}}_c$ and the input raw B-scan $\mathbf{x}$ is introduced to constrain that $\hat{\mathbf{x}}_c$ retains the clutter information of $\mathbf{x}$.

Generator. As shown in Fig. 3(a), the generator $\mathbb{G}$ has an encoder-decoder structure. It consists of one 7×7 convolutional layer followed by instance normalization (IN) [43] and rectified linear unit (ReLU) activation [44], two 3×3 down-sampling convolutional layers also with IN and ReLU, nine residual blocks, two 3×3 up-sampling convolutional layers with IN and ReLU, and one 7×7 convolutional layer followed by Tanh activation. The structure of the residual block is shown in Fig. 3(b). The input feature map is forwarded to one 3×3 convolutional layer followed by IN and ReLU and one 3×3 convolutional layer followed by IN. Then, a skip connection is applied by adding the input feature map to the processed one. By defining the first half of the generator as the encoder $\mathbb{G}_{\text{enc}}$ and the remaining half as the decoder $\mathbb{G}_{\text{dec}}$, the output B-scan is generated as $\hat{\mathbf{x}}_c = \mathbb{G}(\mathbf{x}) = \mathbb{G}_{\text{dec}}(\mathbb{G}_{\text{enc}}(\mathbf{x}))$.

Discriminator. The structure of the discriminator $\mathbb{D}$ is also shown in Fig. 3(a). Based on PatchGAN [45], the discriminator includes three 4×4 down-sampling convolutional layers, each followed by IN and Leaky ReLU activation, one 4×4 convolutional layer with IN and Leaky ReLU activation, and one 4×4 convolutional layer without any activation function. The final convolutional layer outputs a one-channel prediction map to distinguish the generated B-scan and the real B-scan.

Contrastive learning-augmented loss. To enable the unpaired translation from the input raw B-scans to the output clutter-only ones, the contrastive learning-augmented loss is introduced to train the network [42]. First, the adversarial learning of the generator and the discriminator is employed to constrain the generated clutter-only B-scans to be similar to the clutter-only B-scans in the dataset via the GAN loss:

$$\tau_{GAN}(\mathbb{G}, \mathbb{D}, \boldsymbol{X}, \boldsymbol{X}_c) = \mathbb{E}_{\mathbf{x}_c \in \boldsymbol{X}_c} \log \mathbb{D}(\mathbf{x}_c) + \mathbb{E}_{\mathbf{x} \in \boldsymbol{X}} \log\left(1 - \mathbb{D}(\mathbb{G}(\mathbf{x}))\right). \quad (5)$$

Second, multilayer and patchwise contrastive learning is used to maximize the mutual information regarding the clutter distribution between the input and output B-scans. This ensures that the predicted clutter-only B-scan corresponds accurately to the input raw B-scan. As shown in Fig. 3(a), when considering a clutter patch $\mathbf{z}$ from the output clutter-only B-scan as the query, we aim to associate $\mathbf{z}$ with its corresponding positive patch $\mathbf{z}^+$ at the same spatial location in the input raw B-scan and disassociate $\mathbf{z}$ from other noncorresponding negative patches $\mathbf{z}^-$ via a cross-entropy (CE) loss [46]:

$$\tau(\mathbf{z}, \mathbf{z}^+, \mathbf{z}^-) = -\log\left[\frac{\exp\left(\mathbf{z} \cdot \frac{\mathbf{z}^+}{\varepsilon}\right)}{\exp\left(\mathbf{z} \cdot \frac{\mathbf{z}^+}{\varepsilon}\right) + \sum_{n=1}^{N} \exp\left(\mathbf{z} \cdot \frac{\mathbf{z}_n^-}{\varepsilon}\right)}\right], \quad (6)$$

where ε, n, and N represent the temperature scaler, index, and total number of the negative patches, respectively. After

separately inputting $\mathbf{x}$ and $\hat{\mathbf{x}}_c$ into $\mathbb{G}_{enc}$ and obtaining the feature maps, we select L layers of interest and pass them through a multilayer perceptron (MLP) network, which is defined as $\mathbb{H}$, to further extract features from the selected L layers.

The MLP network consists of two fully connected layers, as shown in Fig. 3(c). For each spatial position $s \in \{1, 2, \cdots, S\}$ at each layer $l \in \{1, 2, \cdots, L\}$, the query patch encoded from the output clutter-only B-scan is represented as $\mathbf{z}_l^s = \left[\mathbb{H}_l\left(\mathbb{G}_{enc}^l(\mathbb{G}(\mathbf{x}))\right)\right]^s$. Then the positive and negative patches encoded from the input raw B-scan are represented as $\mathbf{z}^{+s}_l = \left[\mathbb{H}_l(\mathbb{G}_{enc}^l(\mathbf{x}))\right]^s$ and $\mathbf{z}^{-S\backslash s}_l = \left[\mathbb{H}_l(\mathbb{G}_{enc}^l(\mathbf{x}))\right]^{S\backslash s}$, respectively, where $S\backslash s$ denotes the indices of the noncorresponding positions. The contrastive loss involving Eq. (6) can be expressed as

$$\tau_{PatchNCE}(\mathbb{G}, \mathbb{H}, \boldsymbol{X}) = \mathbb{E}_{\mathbf{x}\in\boldsymbol{X}} \sum_{l=1}^{L} \sum_{s=1}^{S} \tau\left(\mathbf{z}_l^s, \mathbf{z}^{+s}_l, \mathbf{z}^{-S\backslash s}_l\right). \quad (7)$$

The final loss function combining the adversarial loss and the contrastive loss for training the whole network is expressed as

$$\tau_{Total}(\mathbb{G}, \mathbb{D}, \mathbb{H}, \boldsymbol{X}, \boldsymbol{X}_c) = \tau_{GAN}(\mathbb{G}, \mathbb{D}, \boldsymbol{X}, \boldsymbol{X}_c) + \tau_{PatchNCE}(\mathbb{G}, \mathbb{H}, \boldsymbol{X}). \quad (8)$$

Once the network is well-trained, it can generate the corresponding clutter-only B-scan based on a given raw B-scan. Subsequently, the clutter-free B-scan is obtained by subtracting this clutter-only B-scan from the raw one.

III. Numerical Experiments

A. Dataset Preparation and Implementation Details

To evaluate the clutter removal performance of the proposed scheme on numerical simulation data, a set of raw B-scans under heterogeneous soil conditions are generated using the open-source simulation software gprMax [47]. A typical simulation scenario is shown in Fig. 4. The 2D soil domain occupies a size of 1.5×0.5 m^2, and the spatial discretization is 0.0025×0.0025 m^2. To build a heterogeneous soil environment with realistic dielectric and geometric properties, the Peplinski mixing model [48] is used and the soil properties are set as sand fraction 0.5, clay fraction 0.5, bulk density 2 g/cm^3, and sand particle density 2.66 g/cm^3. The water volumetric fractions, relative permittivity, and conductivity of the soil vary within [0.1%, 20%], [3.65, 9.91], and [0.01, 0.07], respectively. A Gaussian waveform with a center frequency of 1 GHz is used as the source waveform. A Hertzian dipole (TX) and a probe (RX) transmit and receive the signals, respectively, and are moved along the straight scanning trajectory to collect B-scan data. 110 raw B-scans for the buried cylindrical object with random size, position, and constitutive parameters are generated. Among them, the object's radius is randomly selected from [0.025, 0.05] m, the horizontal position and depth of the object center are randomly chosen from [0.25, 1.25] m and [0.10, 0.25] m, respectively, and the relative permittivity value of the object varies in [2, 32]. The corresponding clutter-only B-scan is obtained by conducting GPR scanning on the heterogeneous soil without any objects buried. In total, 110 raw B-scans and one clutter-only B-scan are obtained as the simulated dataset. All the B-scans are normalized to [0, 1] and resized to 128×128. Among the 110 raw B-scans, 100 B-scans are used to train the network models in Stages 1 and 2. The one clutter-only B-scan is only used as the unpaired ground truth to train the contrastive learning-based GAN in Stage 2. The rest 10 raw B-scans are used as the testing data.

The proposed two-stage networks are implemented on PyTorch [49]. For the diffusion model at Stage 1, the number of time steps is set to 1000. The channel numbers of the U-Net at each time step are [64, 128, 256, 512]. The batch size, learning rate, and number of training steps are set to 16, 2×10^{-5}, and 20,000, respectively. Using the well-trained diffusion model, we generate 100 raw B-scans from random noise and combine them with the simulated 100 raw B-scans to form the training data of Stage 2. Totally, 200 raw B-scans and the one clutter-only B-scan are used for training the second-stage network. For the contrastive learning-based GAN at Stage 2, five layers of the encoder are used to extract the multilayer features. The batch size is set to 1. In the 400-epoch training process, the learning rate is 2×10^{-4} in the first 200 epochs and linearly decays to 0 in the last 200 epochs. In the testing phase, the clutter-only B-scan is estimated from the input raw B-scan using the well-trained generator.

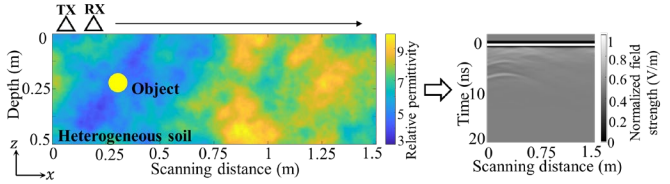

Fig. 4. A typical simulation scenario for generating the training dataset and regular testing dataset with one cylinder object and the obtained B-scan via the GPR scanning.

B. Comparative Result Analysis

The clutter removal performance of the proposed scheme ULCR-Net on the simulated raw B-scans is quantitatively and qualitatively compared with existing GPR image clutter removal methods, including the SVD [6]-[7], RPCA [12]-[13], deep convolutional autoencoder (DCAE) [25], and clutter removal network (CR-Net) [30]. For the SVD, the largest singular value and the corresponding singular vector of the noisy B-scan are removed. For the RPCA, the regularization parameter, the augmented Lagrangian parameter, the reconstruction error tolerance, and the maximum number of iterations are set to 0.025, 0.25, 1×10^{-6}, and 200, respectively. For the DCAE and CR-Net, we follow the network structures as described in [25] and [30], respectively. When training the DCAE using our dataset, the initial learning rate is set to 0.0005 with a decay rate of 99.99% if the training loss does not decrease in the current epoch; the number of epochs and the batch size are set to 300 and 1, respectively. When training the CR-Net using our dataset, the initial learning rate is set to 0.001 with a decay rate of 99% if the training loss does not decrease in the current epoch; the number of epochs and the batch size are set to 300 and 16, respectively. It should be noted that the supervised learning-based DCAE and CR-Net use 100 pairs of raw and clutter-free B-scans for training the networks from scratch, while the proposed ULCR-Net only involves the clutter-only B-scan. Referring to [25]-[31], four evaluation metrics are employed to compare the clutter removal accuracy,

TABLE I
METRICS COMPARISON WITH EXISTING CLUTTER REMOVAL METHODS ON NUMERICAL SIMULATION DATASET

Method	PSNR (dB) (↑)				SSIM ($\times 10^{-2}$) (↑)				MAE ($\times 10^{-2}$) (↓)				MRE (%) (↓)			
	R	G1	G2	G3	R	G1	G2	G3	R	G1	G2	G3	R	G1	G2	G3
SVD	38.15	38.17	37.21	37.98	95.63	95.81	95.60	95.62	0.62	0.62	0.65	0.63	2.47	2.47	2.79	2.52
RPCA	37.74	37.81	37.72	37.61	95.96	96.05	96.30	96.09	0.50	0.49	0.50	0.51	2.58	2.57	2.59	2.63
DCAE	40.91	40.78	38.22	40.17	96.58	96.48	96.22	96.19	0.53	0.55	0.60	0.57	1.81	1.88	2.72	2.02
CR-Net	45.43	44.90	41.03	44.32	98.90	98.71	98.29	98.61	0.30	0.32	0.37	0.33	1.08	1.26	2.47	1.31
ULCR-Net	**50.26**	**49.96**	**48.98**	**49.89**	**99.48**	**99.45**	**99.48**	**99.46**	**0.21**	**0.21**	**0.22**	**0.22**	**0.61**	**0.64**	**0.73**	**0.64**

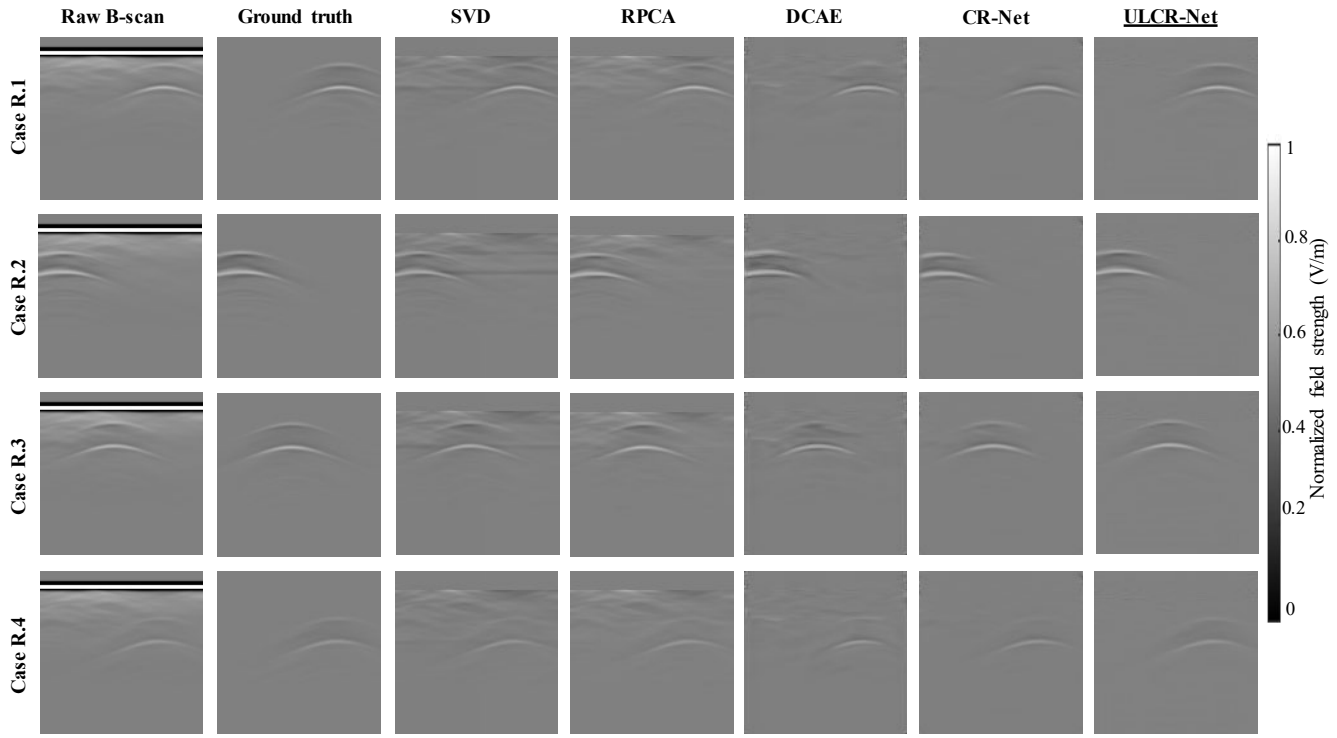

Fig. 5. Clutter removal result comparison on numerical simulation data. Cases R.1-R.4 represent four scenarios in the regular testing dataset, in which a cylindrical object with various properties is buried in the heterogeneous soil environment.

including the PSNR (peak signal-to-noise ratio), SSIM (structural similarity), MAE (mean absolute error), and MRE (mean relative error) between the estimated clutter-free B-scans and the ground-truth clutter-free B-scans. They are defined as:

$$\mathrm{PSNR(dB)} = 10\log_{10}\frac{1}{\frac{1}{H\times W}\sum_{i,j}\left(\mathbf{x}_{\mathrm{s}i,j}-\hat{\mathbf{x}}_{\mathrm{s}i,j}\right)^2}, \tag{9}$$

$$\mathrm{SSIM} = \frac{(2\mu_{\mathbf{x}_\mathrm{s}}\mu_{\hat{\mathbf{x}}_\mathrm{s}}+c_1)(2\sigma_{\mathbf{x}_\mathrm{s}\hat{\mathbf{x}}_\mathrm{s}}+c_2)}{\left(\mu_{\hat{\mathbf{x}}_\mathrm{s}}^2+\mu_{\hat{\mathbf{x}}_\mathrm{s}}^2+c_1\right)\left(\sigma_{\hat{\mathbf{x}}_\mathrm{s}}^2+\sigma_{\hat{\mathbf{x}}_\mathrm{s}}^2+c_2\right)}, \tag{10}$$

$$\mathrm{MAE} = \frac{1}{H\times W}\sum_{i,j}\left|\mathbf{x}_{\mathrm{s}i,j}-\hat{\mathbf{x}}_{\mathrm{s}i,j}\right|, \tag{11}$$

$$\mathrm{MRE} = \frac{\sqrt{\sum_{i,j}\left(\mathbf{x}_{\mathrm{s}i,j}-\hat{\mathbf{x}}_{\mathrm{s}i,j}\right)^2}}{\sqrt{\sum_{i,j}\mathbf{x}_{\mathrm{s}i,j}^2}}\times 100\%, \tag{12}$$

where $\mu_{\hat{\mathbf{x}}_\mathrm{s}}$ and $\mu_{\mathbf{x}_\mathrm{s}}$ are the means of the estimated clutter-free B-scan $\hat{\mathbf{x}}_\mathrm{s}$ and the ground-truth clutter-free B-scan $\mathbf{x}_\mathrm{s}$, respectively. $\sigma_{\mathbf{x}_\mathrm{s}}$, $\sigma_{\hat{\mathbf{x}}_\mathrm{s}}$, and $\sigma_{\mathbf{x}_\mathrm{s}\hat{\mathbf{x}}_\mathrm{s}}$ are the variance of $\mathbf{x}_\mathrm{s}$, variance of $\hat{\mathbf{x}}_\mathrm{s}$, and covariance of $\mathbf{x}_\mathrm{s}$ and $\hat{\mathbf{x}}_\mathrm{s}$, respectively. c_1 and c_2 are two variables. i, j, H, and W are the indices and total dimensions of the B-scan image. Higher PSNR and SSIM and lower MAE and MRE indicate better clutter removal accuracy.

Table I presents the comparative results on the evaluation metrics. "R" column in the table represents the test results on regular 10 testing data, described in subsection A, while the test on the data in other columns, "G1", "G2", and "G3", is explained in subsection C. Their average PSNR, SSIM, MAE, and MRE are calculated and listed. It is observed that the proposed ULCR-Net achieves the highest PSNR and SSIM and the lowest MAE and MRE, demonstrating the superior clutter removal performance of the proposed two-stage network.

To visually compare the clutter removal performance of different methods, Fig. 5 shows the results of four cases within the regular testing dataset. The raw B-scans contain the direct coupling of TX and RX, reflections from the ground, irregular clutter due to the heterogeneous soil environment, and hyperbolic signatures due to reflections from the buried cylindrical object with various properties. Although the conventional SVD and RPCA algorithms effectively remove the direct coupling and reflections from the ground, they cannot remove irregular clutter patterns due to the heterogeneous soil. The supervised deep learning-based DCAE and CR-Net

outperform SVD and RPCA by recognizing the features of the object's reflection patterns and extracting them out, but some clutter remains in their predicted clutter-free B-scans. This is because the small dataset size (only 100 pairs) is insufficient to train these fully supervised learning-based networks well. In contrast, our proposed two-stage scheme ULCR-Net first augments the small dataset and then learns the features of the clutter distribution from the provided clutter-only B-scans. In this way, the clutter patterns are accurately estimated. Subtracting the accurate clutter patterns from raw B-scans yields cleaner and clutter-free B-scans.

C. *Generalizability Test on New Scenarios*

In the training dataset and the regular testing dataset, only the scenarios with a single cylinder object buried in heterogeneous soil [Fig. 4] are considered. To examine the generalizability of the proposed scheme in new scenarios, we generate three generalized testing datasets by simulating new-shape objects, new-material objects, and multiple objects buried in the heterogeneous soil, as shown in Figs. 6(a)-(c), respectively.

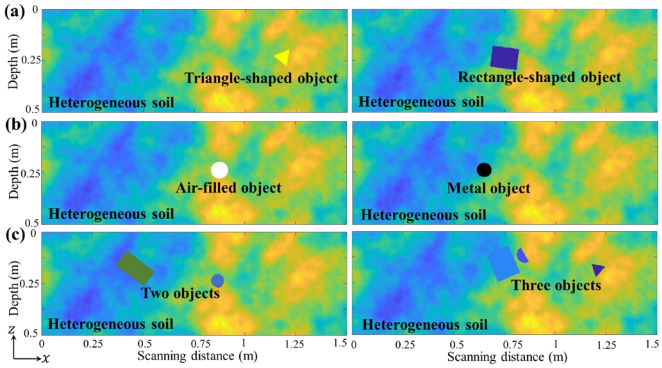

Fig. 6. Simulation scenarios for (a) new-shape objects in dataset G1, (b) new-material objects in dataset G2, and (c) multiple objects in dataset G3, respectively, in the generalizability test.

For these generalized testing datasets, dataset G1 includes 10 scenarios with rectangle-shaped objects and 10 scenarios with triangle-shaped objects, dataset G2 includes 10 scenarios with air-filled objects (or cavities) and 10 scenarios with metal objects, and dataset G3 includes 10 two-object scenarios and 10

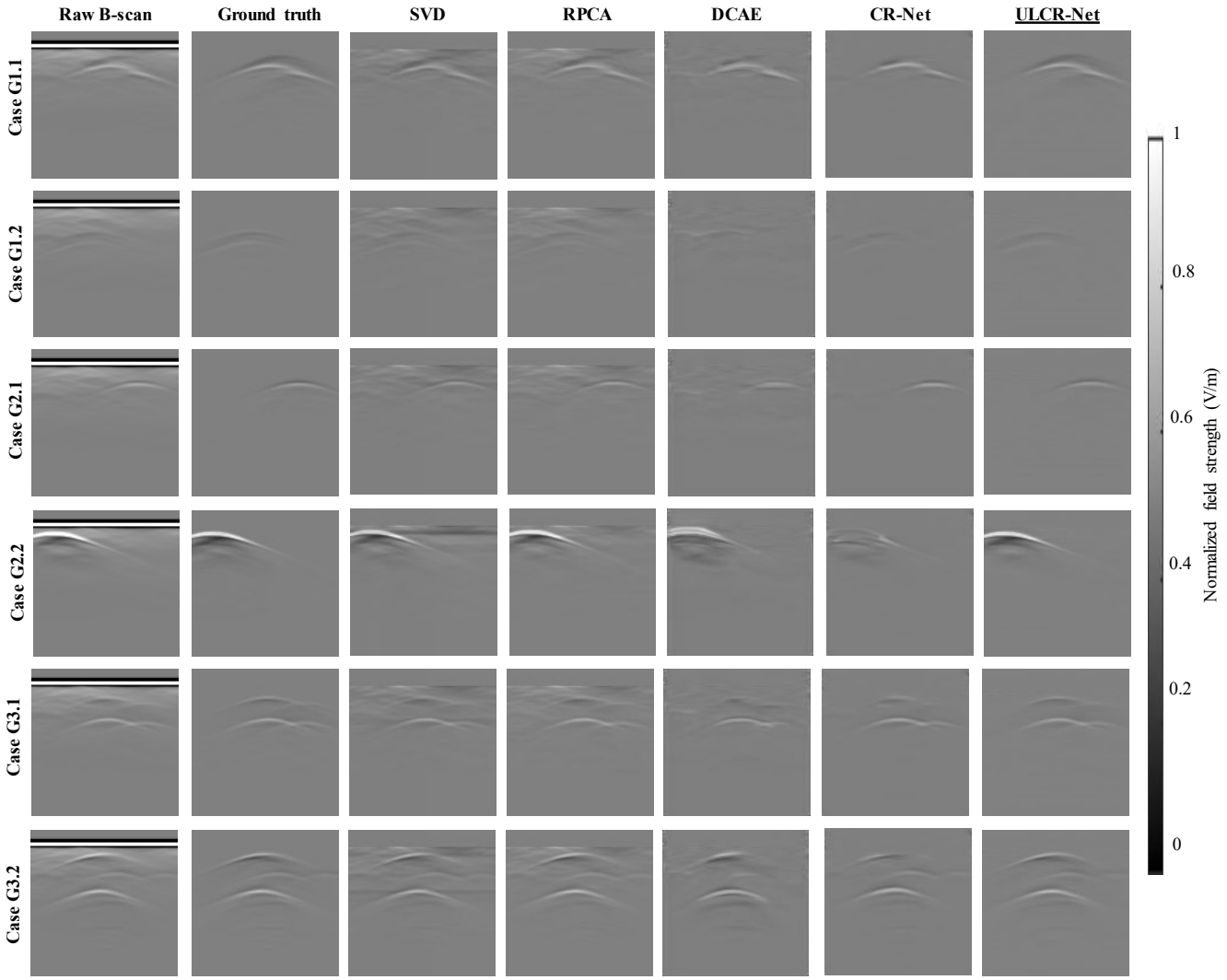

Fig. 7. Imaging result comparison of the generalizability test. Cases G1.1, G1.2, G2.1, G2.2, G3.1, and G3.2 represent the scenarios with one rectangle-shaped object, one triangle-shaped object, one air-filled object, one metal object, two objects, and three objects, respectively.

three-object scenarios. These generalized testing datasets are directly fed into the network without any fine-tuning. The performance of the proposed method, alongside comparison methods, is quantitively compared using PSNR, SSIM, MAE, and MRE, as shown in Table I. Our scheme achieves the highest PSNR and SSIM and the lowest MAE and MRE in all generalized datasets, demonstrating enhanced generalizability compared to existing methods.

The clutter-removal results of the three generalized testing datasets are visualized in Fig. 7. Cases G1.1, G1.2, G2.1, G2.2, G3.1, and G3.2 represent scenarios with one rectangular object, one triangular object, one air-filled object, one metallic object, two objects, and three objects, respectively. For the network trained on only one-cylinder-object scenarios, the reflection signatures in these B-scans are totally new. For example, the reflection patterns of rectangular and triangular objects are non-standard hyperbolic; the field strength values of reflections of the air-filled and metal objects differ significantly from those in the training set; and the multiple-object cases present complex overlapping and interfering reflection patterns. However, our proposed scheme, having learned sufficient information about clutter distribution, can accurately distinguish the clutter patterns from the object signatures in these new B-scans. The estimated clutter-free B-scans using the proposed ULCR-Net are the closest to their ground truths compared to the results of SVD, RPCA, DCAE, and CR-Net. The comparative results shown in Table I and Fig. 7 quantitively and qualitatively verify the superior generalizability of the proposed scheme.

TABLE II
TESTS ON CASES WHEN THE CLUTTER SUPPRESSES THE TARGET

Case No.	Upper Depth (cm)	Diameter (cm)	Horizontal Position (cm)	Relative Permittivity	PSNR (dB)	SSIM ($\times 10^{-2}$)	MAE ($\times 10^{-2}$)	MRE (%)
i	4	2	75	19	49.63	99.43	0.22	0.66
ii	2	3	100	8	52.54	99.55	0.11	0.47
iii	1	3	50	10	50.22	99.46	0.21	0.61

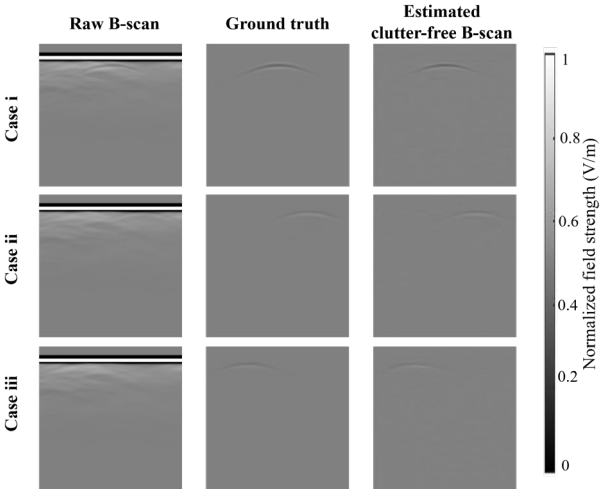

Fig. 8. Imaging results of the tests on scenarios when the clutter suppresses the target. Cases i, ii, and iii represent the scenarios with one cylindrical object buried in the soil, in which the depths of the object's upper surface are 4 cm, 2 cm, and 1 cm, respectively.

Furthermore, to test the cases when the clutter suppresses the target, three new scenarios with the object buried very close to the ground are simulated and then the obtained raw B-scans are directly fed into the trained network. The object parameters and evaluation metrics of the clutter removal results using the proposed scheme are listed in Table II. The high PSNR and low SSIM, MAE, and MRE in each case demonstrate high clutter-removal accuracy of our method. The visual results, displayed in Fig. 8, show that the estimated clutter-free B-scans closely match the ground truth, highlighting the network's capability to effectively remove clutter even when it heavily masks the target signatures in the raw B-scans. These results verify that the proposed scheme effectively learns the features of the clutter and accurately removes it.

D. *Tests on a Complicated Realistic Scenario*

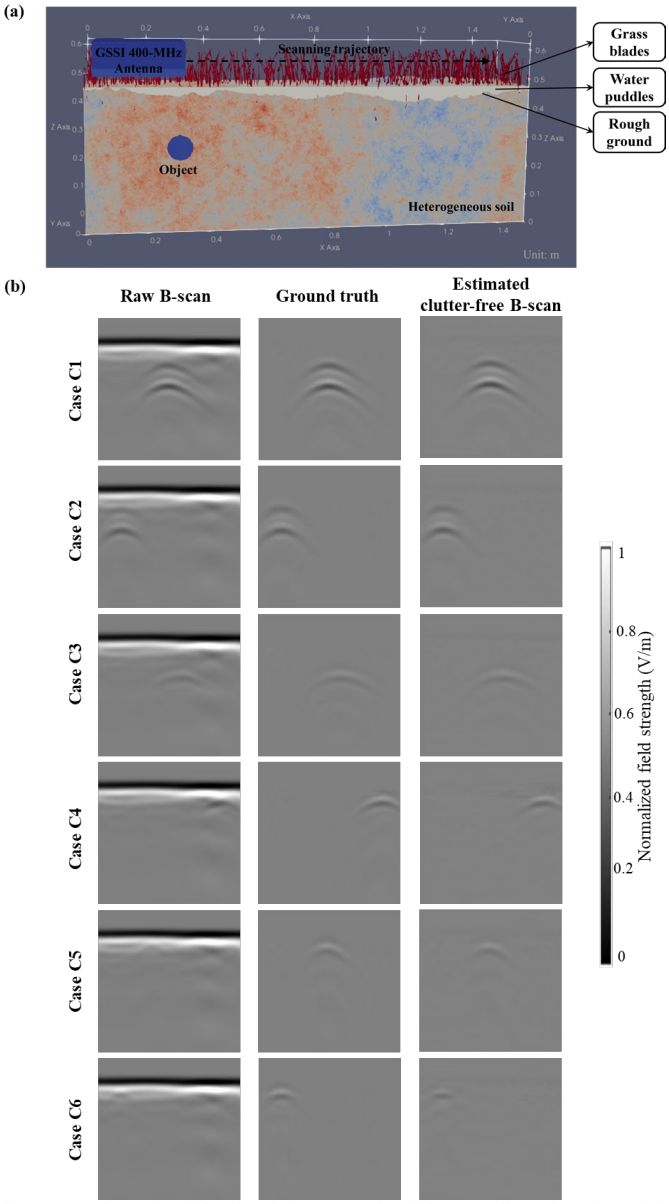

Fig. 9. (a) Schematic of the simulation scenario of the complicated surface with rough ground, water puddles, and grass blades. (b) Imaging results of the clutter removal for the complex scenarios. Cases C1-C3 represent three testing scenarios in which the depth range of the object center is randomly selected from [0.10, 0.25] m. Cases C4-C5 represent three new scenarios in which the depths of the object's upper surface are 4 cm, 2 cm, and 1cm, respectively.

To test more realistic scenarios using the proposed scheme, we add rough ground surface with water puddles and grass blades in the heterogeneous soil model, as shown in Fig. 9 (a). In particular, the surface roughness is 0.1 m from peak to valley. The water puddles on the rough surface have a depth of 0.025 m. The number of blades of the grass with random distributions is set to 500 and the grass height varies from 0 to 0.2 m on the rough surface. The depth of the object center is randomly selected from [0.10, 0.25] m and the radius of the cylindrical object is randomly chosen from [0.025, 0.05] m. The commercial GSSI 400-MHz antenna built in gprMax [50] is adopted as the transceiver for GPR scanning. A total of 100 raw B-scans are simulated for training the proposed networks and 10 for testing. The object parameters, soil properties, and the implementation details of the proposed framework are the same as those described in subsection A. The clutter-only B-scan is obtained via the simulation of the complicated scenario without any buried objects.

The visual results, including the raw B-scans, the ground-truth clutter-free B-scans, and the clutter-free B-scans obtained via the proposed scheme, are shown in Cases C1-C3 of Fig. 9(b). It is observed that the clutter due to the complicated scattering mechanism on the surface and the heterogenous soil environment is accurately removed, and the object signatures are retained in the estimated clutter-free B-scans. The average PSNR, SSIM, MAE, and MRE of the testing data reach 49.88 dB, 99.39×10^{-2}, 0.22×10^{-2}, 0.64%, respectively, which are comparable to those obtained in much simpler scenarios. We further test three new scenarios where the clutter suppresses the target signatures, as shown in Cases C4-C6 of Fig. 9(b), in which the depths of the object's upper surface are 4 cm, 2 cm, and 1 cm, respectively. The raw B-scans of these three scenarios are directly fed into the well-trained network without any additional fine-tuning. As shown, the estimated clutter-free B-scans are close to the ground truths. The average PSNR, SSIM, MAE, and MRE of the three testing samples are 46.96 dB, 99.08×10^{-2}, 0.26×10^{-2}, and 0.90%, respectively, which are degraded compared to those of the scenarios without severe clutter suppression but still acceptable. These results demonstrate the applicability of the proposed scheme in complicated real-world environments.

E. Ablation Study

Stage 1 of the proposed scheme is designed to augment the diversity of the training data using a diffusion model, aiming to enhance clutter removal performance at Stage 2, given a limited amount of original data. To verify the effectiveness of Stage 1 in boosting this performance, we conduct an ablation study that compares the clutter removal results of the proposed network with and without Stage 1 on both the regular testing dataset and the generalized testing datasets. The network without Stage 1 (or only Stage 2) means training the contrastive learning-based GAN at Stage 2 using only the 100 raw B-scans without any data augmentation. Other settings of the second-stage network remain unchanged. The comparative results are listed in Table III. As shown, the proposed two-stage framework ULCR-Net that combines Stages 1 and 2 constantly achieves higher PSNR and SSIM and lower MAE and MRE on all the regular and generalized testing datasets. This improvement suggests that the diverse B-scans generated by the diffusion model at Stage 1 help the network at Stage 2 learn more features of the clutter patterns and object signatures. As a result, the clutter estimation accuracy and generalization capability of the second-stage network are enhanced. The results demonstrate the effectiveness of the diffusion model on GPR data augmentation.

TABLE III
METRICS COMPARISON TO DEMONSTRATE THE EFFECTIVENESS OF DIFFUSION MODEL AT STAGE 1

Metrics	Only Stage 2				Stages 1 and 2 (Proposed ULCR-Net)			
	R	G1	G2	G3	R	G1	G2	G3
PSNR (dB) (↑)	49.60	49.52	48.37	49.18	50.26	49.96	48.98	49.89
SSIM ($\times10^{-2}$) (↑)	99.40	99.40	99.41	99.38	99.48	99.45	99.48	99.46
MAE ($\times10^{-2}$) (↓)	0.23	0.23	0.24	0.24	0.21	0.21	0.22	0.22
MRE (%) (↓)	0.66	0.67	0.78	0.70	0.61	0.64	0.73	0.64

TABLE IV
METRICS COMPARISON USING DIFFERENT AMOUNTS OF TRAINING SAMPLES

Original Sample No.	Generated Sample No.	PSNR (dB) (↑)	SSIM ($\times10^{-2}$) (↑)	MAE ($\times10^{-2}$) (↓)	MRE (%) (↓)
100	100	50.26	99.48	0.21	0.61
80	100	49.66	99.39	0.22	0.66
60	100	49.09	99.33	0.24	0.70
40	100	48.73	99.35	0.25	0.74
20	100	46.93	98.97	0.27	0.90

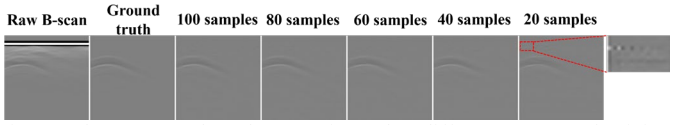

Fig. 10. Clutter-removal result comparison when different amounts of training samples are used for the proposed ULCR-Net.

To further investigate the required amount of the inputted training samples for the proposed scheme, the test results for different amounts of B-scans are compared in Table IV. In the first stage, we use 20, 40, 60, 80, and 100 raw B-scans to train the diffusion model to generate 100 new noisy B-scans, respectively. Then in the second stage, the combined 120, 140, 160, 180, and 200 noisy B-scans are used to train the contrastive learning-based GAN, respectively. The PSNR, SSIM, MAE, and MRE of the same testing data are calculated to evaluate the clutter-removal performance. As shown in Table IV, as the amount of training samples decreases, the clutter-removal accuracy decreases. Especially when the amount of training samples decreases from 40 to 20, the degrading rate of the clutter-removal performance becomes larger. The estimated clutter-free B-scans using various amounts of training samples are visualized in Fig. 10. As shown, when the amount of training samples decreases to 20, the object signatures in the estimated clutter-free B-scan become weaker than those in the ground-truth B-scan, and new clutter patterns appear in the estimated clutter-free B-scan. Based on the comparison, it can be concluded that using more than 40 training samples in the proposed scheme can maintain good performance, and using a

larger number of training samples would yield higher clutter-removal accuracy.

IV. Tests on Real Measurement Data

A. Real Data Collection and Implementation Details

The clutter removal performance of the proposed scheme is also assessed using real measurement data. As shown in Fig. 11, a commercial GSSI Utility Scan Pro GPR system with a 400-MHz antenna is used to collect B-scans in an outdoor uneven sandy field. The scanning trace is 1 m, and the time window is 20 ns. For training data, the raw B-scans were acquired by burying a wooden cylinder object with different diameters and relative permittivity (ε_r) [Fig. 11]. The ranges of the horizontal position, depth, horizontal angle, and vertical angle of the buried object are within [20, 80] cm, [9, 25] cm, [0, 60°], and [0, 60°], respectively. The clutter-only B-scans are collected by conducting GPR scanning in the field without any buried objects. To assess the generalizability of the proposed scheme, box-shaped objects and multiple objects are buried underground to obtain new raw B-scans that are not included in the training dataset. In total, 30 raw B-scans and 30 clutter-only B-scans are collected and used as the training dataset, and four new raw B-scans are obtained for testing purposes. All the real measured B-scans are processed by time-zero correction [5], normalized to [0, 1], and resized to 128×128.

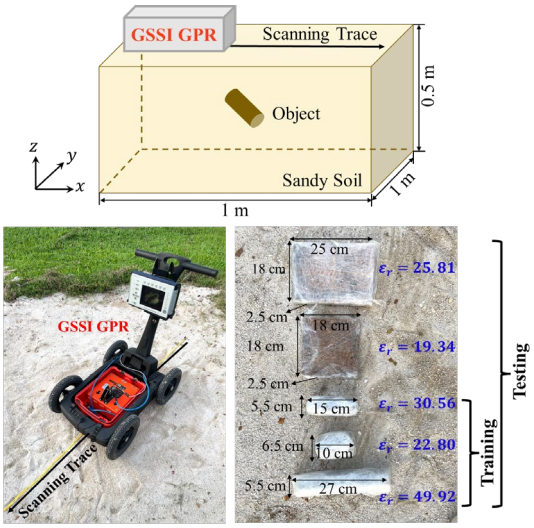

Fig. 11. The experiment setup for collecting real measurement data. Only single-cylinder-object scenarios are used for training. New scenarios with box-shaped objects and multiple objects are used to test network generalizability.

The parameter settings used for training the networks are the same as those described in Section III.A. First, we generate 70 additional raw B-scans using the well-trained diffusion model at Stage 1. These B-scans are combined with the 30 real measured raw B-scans to form the training data for Stage 2. Consequently, the training dataset for the contrastive learning-based GAN at Stage 2 comprises 100 raw B-scans as inputs and 30 clutter-only B-scans as outputs in an unpaired mode. During the testing phase, the four newly acquired raw B-scans are directly fed into the trained generator network at Stage 2 to obtain the predicted clutter-only B-scans. The clutter-free B-scans are then obtained by subtracting the predicted clutter-only B-scans from the raw B-scans.

As the real measured raw B-scans do not have corresponding ground truth, the evaluation metrics in Eqs. (9)-(12) are not applicable. Instead, we employ a non-reference metric, the IQS (image quality score), to evaluate the clutter removal performance based on clutter reduction and structure preservation, as done in [51]-[52]. The IQS is expressed as

$$\mathrm{IQS} = -\mathbb{C}\big(\mathbf{S}(\mathbf{x}, \hat{\mathbf{x}}_c), \mathbf{S}(\mathbf{x}, \hat{\mathbf{x}}_s)\big), \tag{13}$$

where $\mathbf{S}(\mathbf{x}, \hat{\mathbf{x}}_c)$ and $\mathbf{S}(\mathbf{x}, \hat{\mathbf{x}}_s)$ represent the SSIM map of the input raw B-scan $\mathbf{x}$ and the output clutter-only B-scan $\hat{\mathbf{x}}_c$ and the SSIM map of the input raw B-scan $\mathbf{x}$ and the estimated clutter-free B-scan $\hat{\mathbf{x}}_s$, respectively, while $\mathbb{C}$ stands for the Pearson linear correlation coefficient between these two SSIM maps. The SSIM map contains the local SSIM value for each pixel of the B-scan image based on Eq. (10). A larger IQS indicates better clutter removal performance.

B. Result Comparison and Analysis

As performed on simulation data, the clutter removal performance on real measurement data using the proposed two-stage scheme ULCR-Net is compared with the existing clutter removal methods, including SVD [6]-[7], RPCA [12]-[13], DCAE [25], and CR-Net [30]. For training the supervised deep learning-based techniques DCAE and CR-Net, we follow the dataset generation method described in [30]. Specifically, real measured clutter-only B-scans are combined with simulated clutter-free B-scans to form the raw B-scans as input for DCAE and CR-Net, and the corresponding simulated clutter-free B-scans are used as the ground-truth outputs. To maintain consistency between the experimental clutter-only B-scans and simulated clutter-free B-scans, the GSSI 400-MHz GPR antenna model from gprMax [50] is used for transmitting and receiving signals in the one-meter scanning scenarios with a buried cylinder object. In this way, we obtain 100 pairs of data for training the DCAE and CR-Net. Once the networks are well trained, they are used to remove clutter in the four real measured raw B-scans.

The comparative clutter removal results of SVD, RPCA, DCAE, and CR-Net are presented in Fig. 12. The dotted frames in red mark the area of the target signatures and the red arrows indicate the clutter. Cases R, G1, G2, and G3 represent the test scenarios with one cylinder-shaped object, one box-shaped object, two cylinder-shaped objects, and two mixing objects of cylinder and box shapes, respectively. It should be noted that Cases G1, G2, and G3 have never appeared in the training data. We can observe that SVD and RPCA cannot effectively suppress the clutter patterns located above the object signatures due to reflections from the uneven ground as well as those located below the object signatures due to the reflections from the field boundaries; some reflection patterns of the subsurface objects are even distorted. For DCAE, as only a small dataset is employed to train the supervised learning-based network, the network cannot learn adequate informative features of the object signatures, resulting in unsatisfactory clutter removal performance. For the CR-Net using a more advanced clutter removal network, its clutter suppression capability is improved compared to DCAE. However, since it is trained with simulated clutter-free B-scans as ground truths, it tends to estimate field strengths similar to the simulated data, sometimes mistakenly recognizing clutter with strengths close to the objects' reflections as actual signatures, retaining them in the clutter-

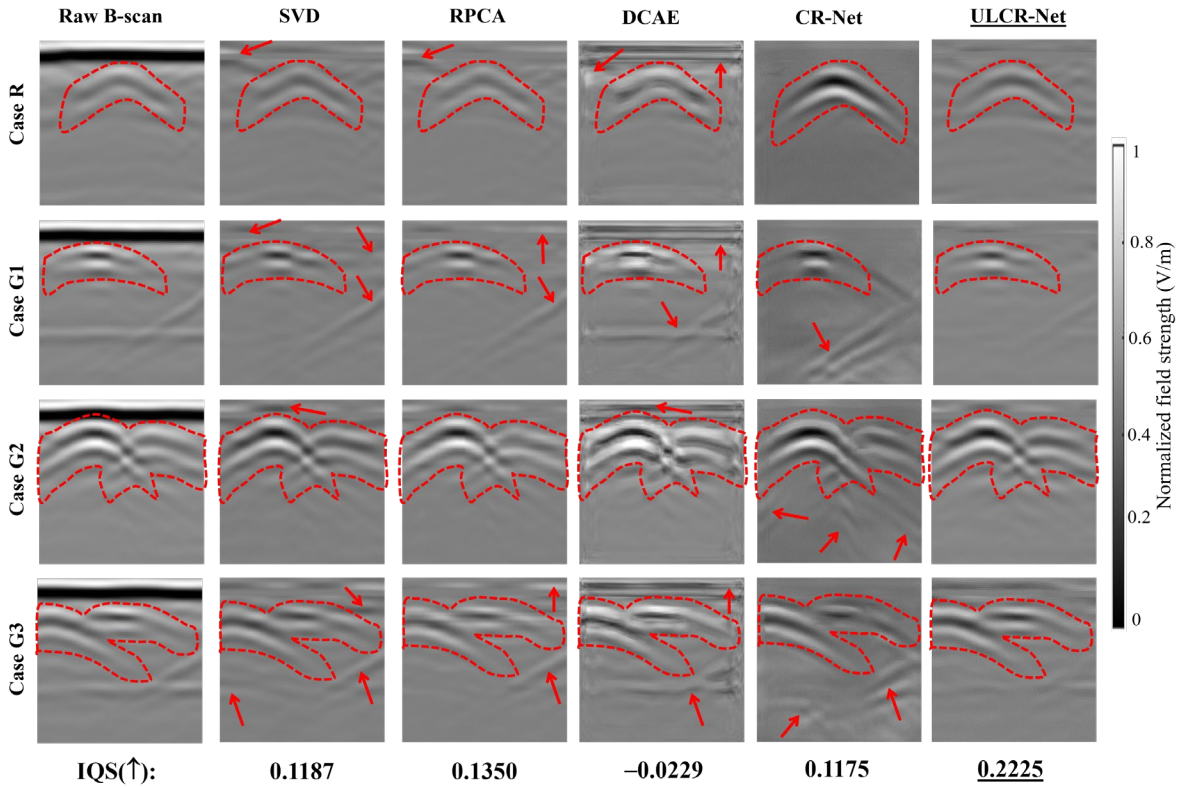

Fig. 12. Imaging result comparison of the real measurement data. Cases R, G1, G2, and G3 represent the test scenarios with one regular cylinder-shaped object, one box-shaped object, two cylinder-shaped objects, and two objects of cylinder shape and box shape, respectively.

free B-scans. On the contrary, our proposed ULCR-Net can suppress the clutter and accurately estimate the object signatures using a very small set of unlabeled real measurement data. Although some irregular clutter still exists, the overall clutter removal performance on the real measurement data is satisfactory. The average IQS of the clutter-free B-scans using the proposed ULCR-Net reaches 0.2225 [Fig. 12], which is the highest compared to those of SVD, RPCA, DCAE, and CR-Net. This demonstrates the superior accuracy and generalizability of the proposed scheme in reducing clutter and effectively retaining the object reflections in diverse real-world B-scans.

V. Conclusions

In this paper, we present a two-stage framework for clutter removal in GPR B-scans based on unsupervised learning. The first stage focuses on data augmentation, where a diffusion model is employed to generate GPR B-scans from random noise. This stage helps build a larger dataset from a small amount of real data, which is crucial for training the learning-based clutter removal network in the second stage. The second stage is dedicated to clutter removal, which involves a contrastive learning-based GAN that learns to predict clutter-only B-scans through unsupervised learning. The final clutter-free B-scans are obtained by subtracting the clutter-only B-scans from the raw B-scans. The entire learning process of the two-stage framework only requires a small set of unpaired raw and clutter-only B-scans, which is relatively easier to acquire in the field. Experimental results on both simulation and measurement data demonstrate the superior clutter removal performance of our method compared to existing algebraic and supervised learning-based techniques in terms of clutter removal accuracy and generalization capability. Besides the effective clutter removal performance, our method successfully addresses the key challenge of acquiring large labelled dataset in real-world scenarios for supervised deep learning-based methods. These advantages make our method highly suitable for clutter removal in practical GPR applications.

References

[1] L. Van Kempen and H. Sahli, “Signal processing techniques for clutter parameters estimation and clutter removal in GPR data for landmine detection,” in *Proc. 11th IEEE Signal Process. Workshop Stat. Signal Process.*, Aug. 2001, pp. 158–161.

[2] R. Solimene, A. Cuccaro, A. Dell’Aversano, I. Catapano, and F. Soldovieri, “Ground clutter removal in GPR surveys,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 3, pp. 792–798, Mar. 2014.

[3] H. Brunzell, “Detection of shallowly buried objects using impulse radar,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 875–886, Mar. 1999.

[4] A. M. Zoubir, I. J. Chant, C. L. Brown, B. Barkat, and C. Abeynayake, “Signal processing techniques for landmine detection using impulse ground penetrating radar,” *IEEE Sensors J.*, vol. 2, no. 1, pp. 41–51, Feb. 2002.

[5] A. Benedetto, F. Tosti, L. B. Ciampoli, and F. D’Amico, “An overview of ground-penetrating radar signal processing techniques for road inspections,” *Signal Process.*, vol. 132, pp. 201–209, Mar. 2016.

[6] F. Abujarad, G. Nadim, and A. Omar, “Clutter reduction and detection of landmine objects in ground penetrating radar data using singular value decomposition (SVD),” in *Proc. 3rd Int. Workshop Adv. Ground Penetrating Radar*, May 2005, pp. 37–42.

[7] C. Liu, C. Song, and Q. Lu, “Random noise de-noising and direct wave eliminating based on SVD method for ground penetrating radar signals,” *J. Appl. Geophys.*, vol. 144, pp. 125–133, Sep. 2017.

[8] B. Karlsen, H. B. Sørensen, J. Larsen, and K. B. Jakobsen, "Independent component analysis for clutter reduction in ground penetrating radar data," in *Proc. SPIE*, vol. 4742, Aug. 2002, pp. 378–389.

[9] A. Zhao, Y. Jiang, W. Wang, and X. Jiaotong, "Exploring independent component analysis for GPR signal processing," in *Proc. Progress Electromagn. Res. Symp.*, Hangzhou, China, Aug. 2005, pp. 750–753.

[10] B. Karlsen, J. Larsen, H. B. D. Sorensen, and K. B. Jakobsen, "Comparison of PCA and ICA based clutter reduction in GPR systems for anti-personal landmine detection," in *Proc. 11th IEEE Signal Process. Workshop Stat. Signal Process.*, Singapore, Aug. 2001, pp. 146–149.

[11] G. Chen, L. Fu, K. Chen, C. D. Boateng, and S. Ge, "Adaptive ground clutter reduction in ground-penetrating radar data based on principal component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3271–3282, Jun. 2019.

[12] D. Kalika, M. T. Knox, L. M. Collins, P. A. Torrione, and K. D. Morton, "Leveraging robust principal component analysis to detect buried explosive threats in handheld ground-penetrating radar data," in *Proc. SPIE*, vol. 9454, May 2015, Art. no. 94541D.

[13] X. Song, D. Xiang, K. Zhou, and Y. Su, "Improving RPCA-based clutter suppression in GPR detection of antipersonnel mines," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1338–1342, Aug. 2017.

[14] D. Kumlu and I. Erer, "Clutter removal in GPR images using nonnegative matrix factorization," *J. Electromagn. Waves Appl.*, vol. 32, no. 16, pp. 2055–2066, Jun. 2018.

[15] D. Kumlu and I. Erer, "Improved clutter removal in GPR by robust nonnegative matrix factorization," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 958–962, Jun. 2020.

[16] Z.-K. Ni, S. Ye, C. Shi, C. Li, and G. Fang, "Clutter suppression in GPR B-scan images using robust autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2020.

[17] E. Temlioglu and I. Erer, "Clutter removal in ground-penetrating radar images using morphological component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1802–1806, Dec. 2016.

[18] F. Giovanneschi, K. V. Mishra, M. A. Gonzalez-Huici, Y. C. Eldar, and J. H. G. Ender, "Dictionary learning for adaptive GPR landmine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10036–10055, Dec. 2019.

[19] Z.-K. Ni, J. Pan, C. Shi, S. Ye, D. Zhao, and G. Fang, "DL-based clutter removal in migrated GPR data for detection of buried target," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Jun. 2022.

[20] Z. Tong, J. Gao, and D. Yuan, "Advances of deep learning applications in ground-penetrating radar: A survey," *Constr. Build Mater.*, vol. 258, p. 120371, Oct. 2020.

[21] X. L. Travassos, S. L. Avila, and N. Ida, "Artificial neural networks and machine learning techniques applied to ground penetrating radar: A review," *Appl. Comput. Inform.*, vol. 17, no. 2, pp. 296–308, Apr. 2021.

[22] Q. Dai, B. Wen, Y. H. Lee, A. C. Yucel, G. Ow, and M. L. M. Yusof, "A deep learning-based methodology for rapidly detecting the defects inside tree trunks via GPR," in *Proc. IEEE USNC-CNC-URSI North Amer. Radio Sci. Meeting (Joint AP-S Symposium)*, Jul. 2020, pp. 139–140.

[23] H. H. Sun, Y. H. Lee, Q. Dai, C. Li, G. Ow, M. L. M. Yusof, and A. C. Yucel, "Estimating parameters of the tree root in heterogeneous soil environments via mask-guided multi-polarimetric integration neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Dec. 2021.

[24] Q. Dai, Y. H. Lee, H.-H. Sun, G. Ow, M. L. M. Yusof, and A. C. Yucel, "3DInvNet: A deep learning-based 3D ground-penetrating radar data inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, May 2023.

[25] E. Temlioglu and I. Erer, "A novel convolutional autoencoder-based clutter removal method for buried threat detection in ground-penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp.1-13, Jul. 2021.

[26] D. Feng, X. Wang, X. Wang, S. Ding, H. Zhang, "Deep convolutional denoising autoencoders with network structure optimization for the high-fidelity attenuation of random GPR noise," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1761.

[27] J. Luo, W. Lei, F. Hou, C. Wang, Q. Ren, S. Zhang, S. Luo, Y. Wang, and L. Xu, "GPR B-scan image denoising via multi-scale convolutional autoencoder with data augmentation," *Electronics*, vol. 10, no. 11, 2021, Art. no. 1269.

[28] Q. Dai, Y. H. Lee, H. H. Sun, G. Ow, M. L. M. Yusof, and A. C. Yucel, "DMRF-UNet: A two-stage deep learning scheme for GPR data inversion under heterogeneous soil conditions," *IEEE Trans. Antennas Propag.*, vol. 70, no. 8, pp. 6313–6328, Aug. 2022.

[29] J. Wang, K. Chen, H. Liu, J. Zhang, W. Kang, S. Li, P. Jiang, Q. Sui, and Z. Wang, "Deep learning-based rebar clutters removal and defect echoes enhancement in GPR images," *IEEE Access*, vol. 9, pp. 87207–87218, 2021.

[30] H. H. Sun, W. Cheng, and Z. Fan, "Learning to remove clutter in real-world GPR images using hybrid data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, May 2022.

[31] Z. -K. Ni, C. Shi, J. Pan, Z. Zheng, S. Ye, and G. Fang, "Declutter-GAN: GPR B-scan data clutter removal using conditional generative adversarial nets," *IEEE Geosci. Remote. Sens. Lett.*, vol. 19, pp. 1-5, Mar. 2022.

[32] Q. Dai, Y. H. Lee, M. L. M. Yusof, D. Lee, and A. C. Yucel, "Learning from noise: An unsupervised GPR data denoising scheme based on generative adversarial networks," in *Proc. CNC-USNC/URSI National Radio Sci. Meet.*, Portland, OR, Jul. 2023.

[33] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," in *Proc. NeurIPS*, vol. 34, 2021, pp. 8780–8794.

[34] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. NeurIPS*, vol. 33, Dec. 2020, pp. 6840–6851.

[35] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using non-equilibrium thermodynamics," in *Proc. ICML*, Jul. 2015, pp. 2256–2265.

[36] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between gaussian mixture models," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, vol. 4, Apr. 2007, pp. IV-317–IV-320.

[37] C. Luo, "Understanding diffusion models: A unified perspective," *arXiv:2208.11970*, Aug. 2022.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Oct. 2015, pp. 234–241.

[39] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2016, pp. 87.1–87.12.

[40] Y. Wu and K. He, "Group normalization," in *Proc. ECCV*, Sep. 2018, pp. 3-19.

[41] X. Chen, N. Mishra, M. Rohaninejad, and P. Abbee, "PixelSNAIL: An improved autoregressive generative model," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2018, pp. 863–871.

[42] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. 16th Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 319–345.

[43] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv:1607.08022*, Nov. 2017.

[44] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," *arXiv:1803.08375*, Feb. 2019.

[45] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jul. 2017, pp. 1125–1134.

[46] A. V. D. Oord, Y. Li, O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv:1807.03748*, Jul. 2018.

[47] C. Warren, A. Giannopoulos, and I. Giannakis, "gprMax: Open source software to simulate electromagnetic wave propagation for Ground Penetrating Radar," *Comput. Phys. Commun.*, vol. 209, pp. 163-170, Dec. 2016.

[48] N. R. Peplinski, F. T. Ulaby, and M. C. Dobson, "Dielectric properties of soils in the 0.3-1.3-GHz range," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 3, pp. 803-807, May 1995.

[49] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, N. Gimelshein, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, "PyTorch: An imperative style, highperformance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, Canada, vol. 32, Dec. 2019, pp. 8026–8037.

[50] I. Giannakis, A. Giannopoulos, and C. Warren, "Realistic FDTD GPR antenna models optimized using a novel linear/nonlinear full-waveform inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1768–1778, Mar. 2019.

[51] X. Kong, K. Li, Q. Yang, L. Wenyin, and M. H. Yang, "A new image quality metric for image auto-denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2888-2895.

[52] L. Si, "No-reference image denoising quality assessment," in *Advances in Computer Vision*, vol. 943, K. Arai and S. Kapoor, Eds., Cham, Switzerland: Springer, 2019, pp. 416–433.