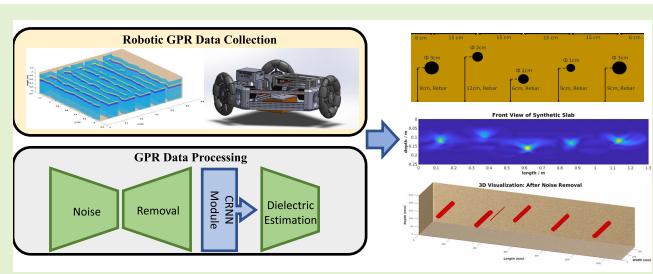


# Improving 3D Metric GPR Imaging Using Automated Data Collection and Learning-Based Processing

Jinglun Feng<sup>1</sup>, Liang Yang, Ejup Hoxha, and Jizhong Xiao, *Senior Member, IEEE*

**Abstract**—Ground Penetrating Radar (GPR) is one of the most important non-destructive evaluation (NDE) devices to detect subsurface objects (i.e., rebars, utility pipes) and reconstruct the underground scene. There are two challenges for GPR-based inspection, which are GPR data collection and 3D subsurface object imaging. To address these challenges, we first propose a robotic solution that automates the GPR data collection process with a free motion pattern. It facilitates the 3D metric GPR imaging by tagging the pose with GPR measurement in real-time. Moreover, to improve the 3D GPR imaging, we introduce a learning-based GPR data analysis method, which includes a noise removal module to clear the background noise in raw GPR data and a Convolutional Recurrent Neural Network (CRNN) to estimate the dielectric value of subsurface medium in each GPR B-scan data. We use both field and synthetic data to verify the proposed methods. Experimental results demonstrate that our proposed methods can achieve higher performance and faster processing speed in 3D GPR imaging than baseline methods.

**Index Terms**—Back-projection (BP) algorithm, ground penetrating radar (GPR), deep neural network (DNN), non-destructive evaluation (NDE), robotics.



## I. INTRODUCTION

GROUND Penetrating Radar (GPR) is widely used in non-destructive evaluation/testing (NDE/NDT), field archaeology investigation, infrastructure inspection, and measurements. GPR works by sending a pulse of polarized high-frequency electromagnetic (EM) wave into the subsurface medium. EM wave attenuates as it travels in medium and reflects when it encounters a material change. GPR antenna would thus record the strength and traveled time of each reflected pulse. The received signal is called an *A-scan*. When GPR antenna surveys over a subsurface object, it produces a series of A-scans at different positions, and the ensemble of A-scans forms a *B-scan*. The B-scan contains hyperbolic features which would indicate the location of the target.

Manuscript received 4 February 2022; revised 29 March 2022; accepted 31 March 2022. Date of publication 4 April 2022; date of current version 28 February 2023. This work was supported in part by the National Science Foundation (NSF) under Grant IIP-1915721 and in part by the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) through the INSPIRE University Transportation Center (<http://inspire-utc.mst.edu>) at the Missouri University of Science and Technology under Grant 69A3551747126. The associate editor coordinating the review of this article and approving it for publication was Prof. Rui Fan. (*Corresponding author: Jizhong Xiao.*)

The authors are with the CCNY Robotics Laboratory, Electrical Engineering Department, The City College of New York, New York, NY 10031 USA (e-mail: [jfeng1@ccny.cuny.edu](mailto:jfeng1@ccny.cuny.edu); [lyang1@ccny.cuny.edu](mailto:lyang1@ccny.cuny.edu); [ehoxha000@citymail.cuny.edu](mailto:ehoxha000@citymail.cuny.edu); [jxiao@ccny.cuny.edu](mailto:jxiao@ccny.cuny.edu)).

Digital Object Identifier 10.1109/JSEN.2022.3164707

However, there are two major challenges to reveal underground objects using GPR. First, in encoder-triggered manual GPR data collections, human inspectors need to pre-mark a grid map on the ground, and push the GPR cart along the straight lines [1], or counting on differential global positioning system (GPS) [2]. It is time-consuming and tedious to manually scan a large area for detailed mapping since current commercial GPR devices can only move forward to trigger survey wheel encoder but cannot make turns. On the other hand, GPS won't work in indoor environments, limiting such GPR applications only to be conducted outdoors. In addition, high-precision GPS is expensive while the accuracy of positioning is around the sub-meter level, which is not sufficient. GPS accuracy is even worse in dense urban environments. Second, the current GPR imaging techniques are easily affected by several constraints, i.e., *GPR background noise*, *permittivity of the surrounding medium*, etc. [3], which makes the images low-quality and noisy. Thus, it is very important to automate the GPR data collection procedure, tag accurate pose with each GPR sample, and propose an efficient metric GPR imaging method to allow non-professional people to understand the result of reconstructed subsurface objects.

Many researches have devoted to the GPR imaging techniques. According to [4], Kirchhoff migration method is firstly introduced in the 1970s, which is able to focus on the target position. However, its processing speed is slower than the rest of the migration approaches. Phase-shift migration is



Fig. 1. Robotic GPR data collection module.

a mathematical method proposed in 1978 [5] that utilizes the exploding source model (ESM) concept to focus the targets from GPR B-scan data. For comparable accuracy, this approach is computationally more efficient than the finite-difference [6] approach. Nevertheless, back-projection (BP) is the most significant and commonly used algorithm among the migration techniques in GPR industry [7]. Mast and Johansson [8], [9] made a breakthrough in 3D GPR imaging reconstruction in 1994. They demonstrate, compare and discuss the advantages and disadvantages of the time-domain algorithm and frequency-domain algorithm of 3D GPR imaging techniques. Based on these pioneering works, Pereira *et al.* [10] employed Hessian-based enhancement filtering to facilitate diffraction-based BP algorithm, thus creating 3-D models of subsurface tubular structures toward advanced underground infrastructure visualization.

Besides the migration methods, it is crucial to automate the GPR data collection and fuse the GPR data with accurate positioning information, rather than counting on encoder-triggered or GPS-based positioning. Pereira *et al.* [11] introduced a new GPR system that used Augmented Reality (AR) based positioning available on Google Tango smartphones. However, Google Tango is obsolete and this project is out of the service now. Mihailescu and Negut [12], [13] proposed a tracking system to localize the real-time position of the GPR sensor. This system depends on the AprilTag [14] to provide accurate positioning information. The limitation of this method is that the user has to print and stick the AprilTag markers before the system starts to work. Li *et al.* [15] and Chou *et al.* [16], [17] proposed a vision-facilitated underground pipeline mapping model to conduct GPR-based 3D reconstruction. Their major contribution is that they used the visual Simultaneous Localization and Mapping (V-SLAM), J-linkage and maximum likelihood method to estimate the radii and locations of all pipelines. However, the authors don't have the privileges to access the GPR source code and synchronize the GPR measurements with positioning, and still need to pre-mark the grid map on the surface and manually push the GPR cart along the grid lines to take GPR measurements.

In this paper, we propose a novel 3D GPR imaging method that consists of three modules. A *robotic data collection* module to provide a free motion pattern for GPR scanning and tag metric positioning with GPR measurements on-the-fly; a *background noise removal* module to clean the noisy data in

the GPR B-scan images; and a *CRNN* module to estimate the dielectric property of subsurface medium.

More specifically, in *robotic data collection* module, an omni-directional robot carries a GPR antenna at the bottom of the chassis to detect and map underground objects and an RGB-D camera to obtain the accurate 6 degree of freedom (DOF) pose in real time. By tagging the GPR measurements with accurate pose on-the-fly in a synchronized way, it enables the robot to scan the ground surface in free-motion pattern and facilitates high-resolution 3D GPR imaging. The learning-based GPR data processing includes two modules, GPR noise removal module and dielectric prediction module. For each input B-scan image, the noise removal module directly decodes the input image into hyperbolic features. The dielectric prediction module takes the segmentation masks from the noise removal module and pools the dielectric property from a Convolutional Recurrent Neural Network (CRNN).

## II. 3D METRIC GPR DATA COLLECTION

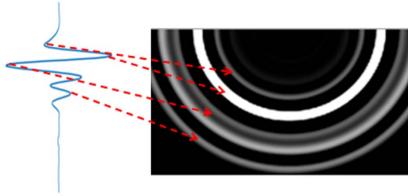
In this section, we first formulate the GPR imaging technique, then introduce a robotic system that automates the GPR data collection with a free motion pattern to facilitate the 3D metric GPR imaging.

### A. GPR 3D Metric Imaging

To create a 3D GPR imaging, the BP algorithm is implemented in this section. The essence of the A-scan represents the amplitude of EM energy, while the BP is a process of aggregation that converts different amplitude of energy into a *semi-sphere* format at different time. As illustrated in Figure 2, the brighter semi-sphere indicates the higher amplitude part in A-scan. Furthermore, the radius of each semi-sphere in BP image indicates the depth between the ground and the object, which is depicted by Equ.1 [15]:

$$\forall A_q^k \in B_k, (x - x_0)^2 + (y - y_0)^2 - (a_l * t)^2, y < 0 \quad (1)$$

where  $x_0, y_0$  represents the specific position of each A-scan measurement in a concrete slab.  $A_q^k = \{a_t | t = 1, \dots, n_q\}$  represents the  $q$ -th A-scan measurement in  $k$ -th B-scan data, while  $t$  and  $a_t$  indicate the traveling time and amplitude of A-scan signal respectively,  $n_q$  means the total number of samples in a A-scan measurement. Meanwhile, we also have  $B_k = \{A_q^k | q = 1, \dots, n_k\}$  that represents the  $k$ -th B-scan



**Fig. 2.** BP algorithm converts the A-scan raw data into a set of semi-spheres.



**Fig. 3.** The Omni-directional robot towards automated GPR data collection, where a GPR antenna is installed at the bottom of the robot chassis.

consisting of  $n_k$  A-scans. By implementing the BP algorithm on a B-scan data, the intersection of multiple back-projected A-scan signals with the highest energy is the possible target location.

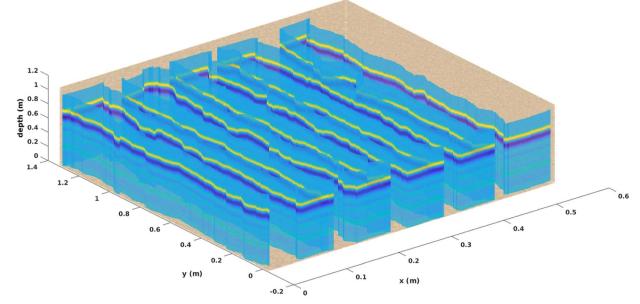
### B. Free Motion Based Data Collection

The current practice for GPR data collection requires a human inspector to mark the grid map, push the GPR device to precisely follow the straight lines in X and Y directions, use the wheel encoder to trigger GPR sampling, and take notes to record these linear motion trajectories for 3D GPR imaging.

To solve this issue, we propose an automatic GPR data collection method that combines a robot and vision-based positioning with GPR signal processing to locate and map subsurface targets. Instead of relying on a wheel encoder to record linear motion, or utilizing the marker to facilitate the localization and mapping [12], [18], [19], we use an Intel RGB-D tracking camera, D435i, to estimate the pose in real time, and record the continuous non-linear trajectory automatically. Note that D435i camera could support indoor and outdoor working environment, which boosts the robustness for our vision positioning system. Meanwhile, a time-domain GPR antenna (PaveScan RDM 1.0) from Geophysical Survey System Inc. (GSSI) with a maximum center frequency of 2GHz, is fixed at the bottom of the robot chassis to conduct GPR data collection.

More specifically, as illustrated in Figure 3, the Omni-wheels allow the robot to move in any directions without spinning and thus provide the GPR with the free motion capacity to scan the area of interest in a fast and swift way. The robot motion satisfies the following relation:

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -d \\ \cos \frac{2\pi}{3} & \sin \frac{2\pi}{3} & -d \\ \cos \frac{-2\pi}{3} & \sin \frac{-2\pi}{3} & -d \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ \omega \end{bmatrix} \quad (2)$$



**Fig. 4.** The B-scan profile tagged with metric positioning information when the robot scan the concrete surface in Zig-Zag pattern.

where  $v_1, v_2, v_3$  represents the linear velocity of each omni-wheel,  $d$  indicates the distance between the center of a wheel and the center of the robot body.  $v_x, v_y$ , and  $\omega$  denote the linear velocity and angular velocity of the robot body respectively.

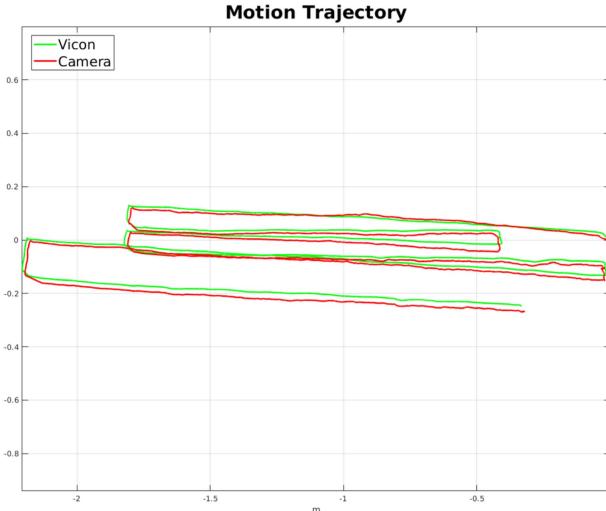
By taking advantage of this non-linear free-motion pattern, the GPR data is automatically and continuously collected and tagged with pose  $[X, Y, \theta]$  at each measurement. Specifically, D435i is embedded with an IMU sensor to collect RGB and depth images of the surrounding environment, together with the corresponding IMU data (e.g., quaternion, angular velocity, and linear velocity). Then, we use ORB-SLAM3 [20] algorithm to fuse the RGB images and depth images with IMU measurements to perform real-time visual-inertial positioning and localization. Then, with a time synchronizer function in Robot Operation System (ROS), we synchronize the GPR sampling with positioning data so that the GPR data collection would not be constrained to the straight line.

The frame rate of the RGB-D camera is 30Hz, and the IMU update rate is 200MHz. Through the interpolation, we achieved 200Hz for position updates. Since the PaveScan GPR sampling rate is 100Hz, we synchronized the vision-based positioning and GPR updating at 100Hz in the experiments. That is to say, our robotic data collection system could collect 100 scan data tagged with pose data per second, and the spacing between the consecutive measurements would be 5 mm when robot moves with a 0.5 m/s linear velocity. It demonstrates that our vision-based positioning solution has met the “low latency” and “high accuracy” requirements by the GPR industry.

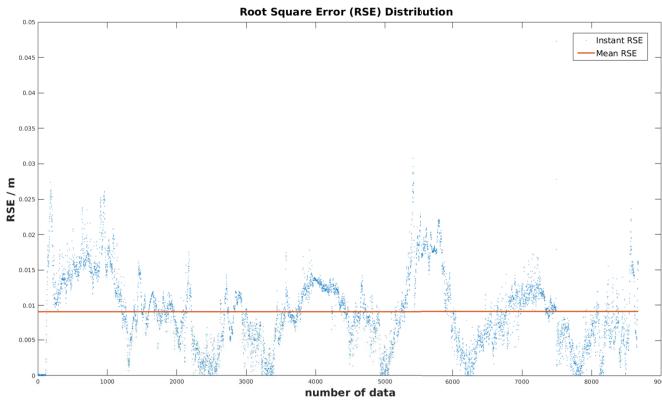
As shown in Equ.3, we update the positioning  $[X, Y]$  according to the orientation  $\theta$ , while  $[X', Y']$  and  $[\delta X, \delta Y]$  represent the previous position state and incremental visual odometry readings respectively. Figure 4 illustrates an example when a robot scans over a surface with a zig-zag pattern, and the B-scan data is tagged with its corresponding positioning information.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \times \begin{bmatrix} \delta X \\ \delta Y \end{bmatrix} + \begin{bmatrix} X' \\ Y' \end{bmatrix} \quad (3)$$

Compared with GPS-based positioning system which could only achieve sub-meter level accuracy, our visual-based positioning system could reach a centimeter-level positioning accuracy. We validate the accuracy using VICON system [21]. VICON system is well-known for the millimeter-level high accuracy of motion capture and thus can provide the ground



**Fig. 5.** The motion trajectory detected by VICON motion capture system as ground truth, and the estimated path estimated by the camera on the robot. Note that this motion is in a zig-zag pattern.



**Fig. 6.** The root square error distribution between the ground truth motion and camera's motion which is around 1cm, with RMSE equals to 0.91cm.

truth of our robot motion. As shown in Figure 5, we command the robot to move along a zig-zag trajectory. The green line indicates the ground truth of the motion trajectory provided by the VICON system, and the red line presents the motion trajectory recorded by our vision-based positioning with RGB-D camera. Then, Equ. 4 denotes the Root Square Error (RSE) between the ground truth and vision-based positioning. Figure 6 illustrates that the Root Mean Square Error (RMSE) is only 0.91 centimeter, which indicates that our vision-based positioning system achieves a high accuracy.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|y(i) - \hat{y}(i)\|^2}{N}} \quad (4)$$

Note that  $N$  indicates the number of position samples while  $i$  means the  $i$ -th position sample.  $y(i)$  and  $\hat{y}(i)$  represent the ground truth position and predicted position respectively.

### III. IMPROVING GPR IMAGING WITH LEARNING-BASED DATA ANALYTICS

GPR background noise and the uncertainty of the material dielectric affect the accuracy of GPR imaging. Therefore, we propose two models, which are the GPR object segmentation model and the dielectric prediction model, to facilitate

accurate subsurface object reconstruction and mapping. The learning-based data analytics is illustrated in Figure 7 and Figure 8, where we use a U-Net [22] to perform subsurface object segmentation and a DielectricNet to predict the dielectric value of the subsurface surrounding medium.

#### A. GPR Background Noise Removal

GPR works by sending an EM wave into the ground. EM wave attenuates as it travels in the medium and reflects when it encounters a change in material dielectric. GPR antenna would thus record the strength of each reflection as well as the travel-time it takes back to the receiver. Therefore, when GPR surveys over an underground object, some of the GPR energy pulses reflect when they hit the subsurface object, some energy continues to penetrate through the object until they become too weak to get back. However, those weak and scattered reflections in the B-scan image could affect the GPR imaging result of the targets. We classify those scatter signals as *GPR Background Noise* (see in Figure 7). Hence, it is important for us not only to focus on enhancing and sharpening B-scan features but also removing those background noise.

---

#### Algorithm 1 GPR Background Noise Removal

---

**Require:** The B-scan set,  $\mathcal{S} = \{B_k \mid k = 0, \dots, n\}$   
**Ensure:** The noise filtered B-scan set,  $\mathcal{S}^f = \{B_k^f \mid k = 0, \dots, n\}$   
 Convert  $\mathcal{S}$  to image set  $\mathcal{S}^I = \{B_k^I \mid k = 0, \dots, n\}$   
**for**  $k \leftarrow 1, n$  **do**  
 1. In contraction path, extracting 2D feature map  $F_i$  from  $B_k$   
 2. In symmetric expanding path, up-convolution the extracted 2D feature map  $F_i$  and represent as  $F_i^v$ ;  
 3. Then fuse the up-convolution feature  $F_i^v$  with the correspondingly cropped feature map from the contracting path, reshape the up-sampled feature to satisfy the prediction requirements and get the mask image set  $\mathcal{M} = \{M_k \mid k = 0, \dots, n\}$   
 4. At last, filter the B-scan set  $\mathcal{S}$  through the mask image set  $\mathcal{M}$ , and get the filtered B-scan set  $\mathcal{S}^f$  as the output.  
**end for**  
**return** filtered B-scan set  $\mathcal{S}^f$

---

Algorithm 1 describes this procedure. At the beginning, we convert the B-scan data into B-scan image set  $\mathcal{S}^I$  so that a segmentation model could be implemented, i.e., U-Net [22]. Then, for each B-scan image in  $\mathcal{S}^I$ , we apply regular convolutions and max pooling layers in contraction path. In the expansion path, we apply transposed convolutions along with regular convolutions. Intuitively, we get the hyperbolic masks  $\mathcal{M} = \{M_k \mid k = 0, \dots, n\}$  from the B-scan image set  $\mathcal{S}^I$ . Furthermore, we deploy the masks  $\mathcal{M}$  as the filter to remove the noise in the raw B-scan data set  $\mathcal{S}$ , so that the filtered B-scan set  $\mathcal{S}^f$  would just keep the strongest target reflection signal, which is shown in Figure 7. As proposed in [23], a classification model is implemented as the noise removal model. However, their results still contain much more noise compared to the current method. The reason is that for the

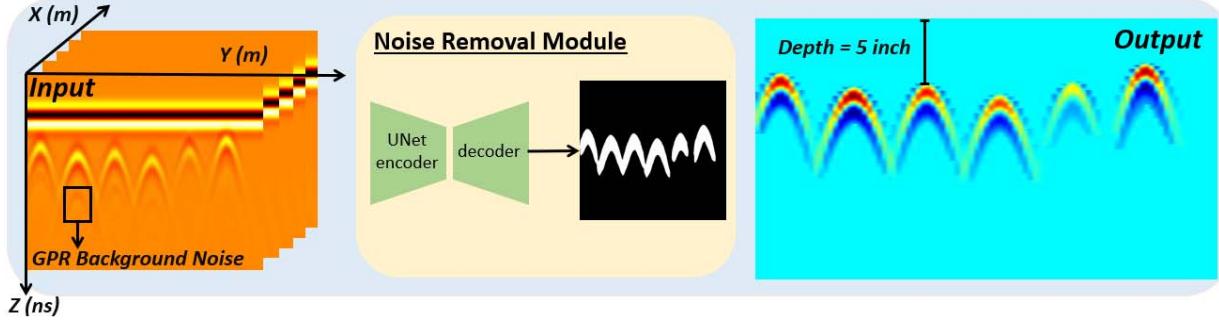


Fig. 7. GPR background noise removal module.

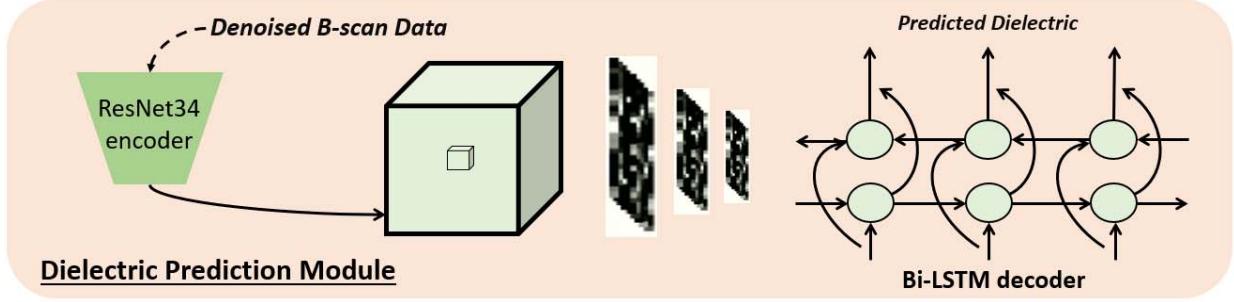


Fig. 8. Target dielectric estimation module.

same region of interest, a classification bounding box occupies more noise data than a segmentation mask. We provide the qualitative and quantitative studies, which can be found in Section IV-B and Section IV-D, to support this claim.

*Loss Design and Training:* U-Net uses a weighted cross-entropy as the loss function, and the per-pixel weights are given by a formula as shown in Equ 5. This design balances the weights between classes and has an extra term to penalize joining two bits of the segmentation [22].

$$w(\mathbf{x}) = w_c(\mathbf{x}) + w_0 \cdot \exp\left(-\frac{(d_1(\mathbf{x}) + d_2(\mathbf{x}))^2}{2\sigma^2}\right) \quad (5)$$

where  $\mathbf{x}$  represents the pixel position  $\mathbf{x} \in \mathcal{M}$ .  $d_1$  and  $d_2$  denote the distance to the border of the nearest and the second cell respectively.  $w_c$  is the weight map to balance the class frequencies while  $w_0$  and  $\sigma$  are constant numbers which are usually set to 10 and 5 respectively [22].

The cross entropy loss function is described as following.

$$\mathcal{L} = \sum_{\mathbf{x} \in \mathcal{M}} w(\mathbf{x}) \log(p(\mathbf{x})) \quad (6)$$

where  $p(\mathbf{x})$  is the probabilistic prediction. Note that the details of our annotated GPR data is introduced in Section IV-A.

We train our model on a server with Intel Core i9-9900K 3.2GHz CPU, GeForce RTX 2080 Ti GPU, and 32GB RAM. Our noise removal model is trained for 100 epochs with stochastic gradient descent (SGD) optimizer whose momentum is equal to 0.9. The initial learning rate is set to  $5e-5$  while the batch size is 16. The weight decay is  $1e-8$  for every 10000 iterations.

### B. Targets Dielectric Estimation

The dielectric constant is crucial for the GPR migration process since it determines the depth of each target shown in a B-scan image. As shown in Equ.7, the EM wave which is emitted from the GPR antenna is supposed to be the speed of light  $c$  in a vacuum. However, the dielectric  $\mathcal{D}_{gt}$  value of the subsurface surrounding medium would make a difference to the traveling speed, which further impacts the depth prediction of the target.

$$v = \frac{C}{\sqrt{\mathcal{D}_{gt}}}, \quad \mathbf{D} = \frac{T_{tr} * v}{2} \quad (7)$$

where  $v$  and  $T_{tr}$  are the velocity and two-way travel time of GPR EM signal,  $\mathbf{D}$  denotes the depth of subsurface targets.

Thus, for any underground object, if it is buried at the same depth underground, when the dielectric value of its surrounding medium is high, the hyperbolic feature in B-scan data would be shown in a deeper position. On account of this property and inspired by [24], we introduce a CRNN model (i.e., DielectricNet) to estimate the dielectric property value  $\mathcal{D}$  for each B-scan. Algorithm 2 describes this procedure.

Given a filtered GPR scan set  $\mathcal{S}^f$  (as shown in Figure 7) in medium with an unknown dielectric, an informative and discriminative feature representation plays a significant role in dielectric estimation. The feature representation should preserve the dielectric property of the underground medium. Moreover, it should consider the spatial distribution of all hyperbolic features to enable better dielectric prediction. We take advantage of Recurrent Neural Network (RNN), which is able to take a sequence of signals, either spatially continuous or temporally continuous, to estimate the dielectric.

**Algorithm 2** Targets Dielectric Estimation

**Require:** The noise filtered B-scan set,  $\mathcal{S}^f = \{B_k^f \mid k = 0, \dots, n\}$

**Ensure:** The predicted dielectric property  $\mathcal{D}$  and its corresponding estimated depth  $\mathbf{D}'$

The predicted dielectric property  $\mathcal{D}$  and its corresponding estimated depth  $\mathbf{D}'$

**for**  $k \leftarrow 1, n$  **do**

1. Extracting a latent feature map  $W_k$  via encoder  $\Phi$
2. Embedding each feature map  $W_k$  through a Bi-LSTM decoder  $\Theta$  and outputs the predicted feature distribution  $W_d$
3. Converting the per-frame predictions made by Bi-LSTM into a label  $\mathcal{D}$ , which represents the dielectric estimation
4. Calculate the corresponding target depth with Equ 7

**end for**

**return** estimated dielectric property  $\mathcal{D}$  and target depth  $\mathbf{D}'$

**1) Model Architecture:** We introduce the technical details of our CRNN model as follows. A CRNN model contains three components [24]: convolutional layers as a encoder, recurrent layer as a decoder and a transaction layer to output the predicted label.

In particular, we illustrate the module structure in **Table I**, the input is a sequence of the filtered B-scan images. The width of each image is flexible since it is determined by the number of A-scans. The height of each B-scan image is equal to 1024, which represents the total sample number of an A-scan.

In the convolutional layers, we employ *ResNet34* [25] as the encoder  $\Phi$ , which contains five residual convolutional blocks, and it automatically extracts a feature sequence from each input B-scan  $B_k^f$  to obtain the feature map  $W_k$  [26]. Meanwhile, we deploy a bidirectional long short-term memory (*Bi-LSTM*) [27] as the decoder  $\Theta$ . It is built for making dielectric prediction distributions  $W_d$  for each frame of the feature map  $W_k$ , outputted by the encoder  $\Phi$ . Specifically, the recurrent layer contains two Bi-LSTM blocks, in which their hidden units are 512 and 256 respectively. Equ. 8 illustrates the definition of a LSTM model,  $c_k$  is the memory states at time step  $k$  and  $f$  represents the forget gate's activation vector. In addition,  $O$  is the output gate's activation vector,  $i$  is the input/update gate's activation vector and  $g$  stands for the cell input activation vector. The LSTM model is able to operate on the input with arbitrary width, traversing from starts to ends. This property benefits us since the input B-scan  $B_k^f$  contains arbitrary A-scans. Besides, in image-based sequences, contexts from both directions are useful and complementary to each other, thus a Bi-LSTM model is implemented according to [28]. At last, in the transaction layer  $\Psi$ , the prediction distributions are converted into a label, which is represented as the dielectric property  $\mathcal{D}$ .

$$\begin{aligned} \text{LSTM: } & (W_k, h_{k-1}, c_{k-1}) \rightarrow (h_k, c_k) \\ & c_k = f \odot c_{k-1} + i \odot g \\ & h_k = O \odot \tanh(c_k) = W_d \end{aligned} \quad (8)$$

**TABLE I**  
DIELECTRIC PREDICTION MODULE STRUCTURE. |\*| STANDS FOR  
RESIDUAL CONVOLUTIONAL BLOCKS

Layer	Configurations
Input	Width $\times$ 1024 B-scan $B_k^f$
conv1	7 $\times$ 7, 64, stride 2
MaxPooling	3 $\times$ 3, max pooling stride 2
conv2_x	$3 \times 3, 64$   $\times 3$ $3 \times 3, 64$
conv3_x	$3 \times 3, 128$   $\times 4$ $3 \times 3, 128$
conv4_x	$3 \times 3, 256$   $\times 6$ $3 \times 3, 256$
conv5_x	$3 \times 3, 512$   $\times 3$ $3 \times 3, 512$
Bidirectional-LSTM	hidden units: 512
Bidirectional-LSTM	hidden units: 256
Transcription	$\mathcal{D}$

As shown in Equ. 9,  $\Phi$  extracts the feature representations  $W_k$  of the input B-scan  $B_k^f$ .  $\Theta$  aims to predict the dielectric distributions  $h_k$  of each B-scan. In the end,  $\Psi$  interprets this learned dielectric property as a single value  $\mathcal{D}$ .

$$\begin{aligned} \Phi: & \mathbb{R}^2 \rightarrow \mathbb{R}^2, B_k^f \mapsto \Phi(B_k^f) = W_k \\ \Theta: & \mathbb{R}^2 \rightarrow \mathbb{R}^2, W_k \mapsto \Theta(W_k) = W_d \\ \Psi: & \mathbb{R}^2 \rightarrow \mathbb{R}^1, W_d \mapsto \Psi(W_d) = \mathcal{D} \end{aligned} \quad (9)$$

**2) Loss Design and Training:** Once the dielectric property is predicted, we optimize the model by using a Connectionist Temporal Categorical (CTC) loss [29].

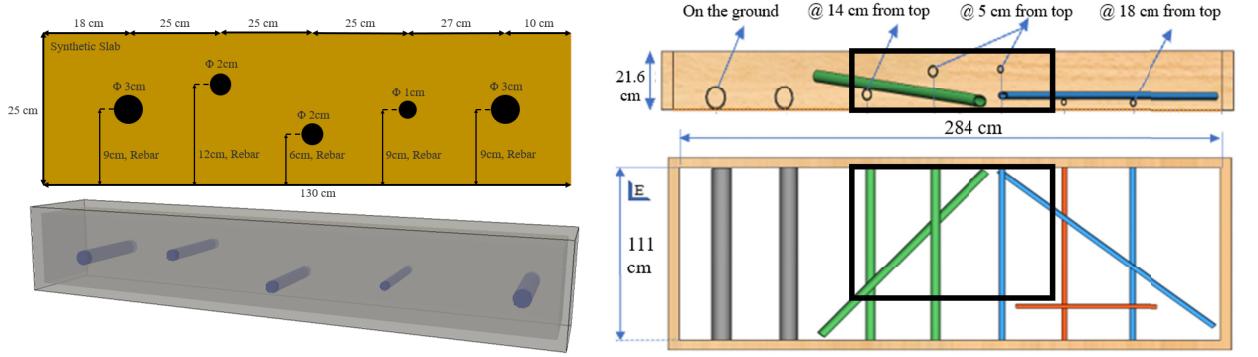
$$\sum_i P(\mathcal{D}_{gt} \mid W) = \sum_i P(\mathcal{D}_{gt} \mid \pi_i) P(\pi_i \mid W) \quad (10)$$

where  $\mathcal{D}_{gt}$  is the ground truth of the dielectric property,  $W$  is the input sequence of the RNN model, and  $\pi_i$  denotes each possible value of the dielectric. The depth of the subsurface target can be further calculated according to Equ. 7.

We train our model on a same server mentioned in the last section. This model is trained for 200 epochs with an 0.01 initial learning rate. We use Root Mean Squared Propagation (RMSprop) optimizer without the momentum to update the learning rate and the weight decay is set to  $1e-08$  for every 10000 iterations.

#### IV. EXPERIMENTAL STUDY

To demonstrate the effectiveness of the proposed 3D metric GPR imaging method, we conducted both simulation and experimental study. The synthetic data were generated by gprMax [30] while the field data were collected on a concrete slab at CCNY Robotics Lab Testing Pit. We also compare the



(a) The front view and top view of the synthetic concrete slab generated by gprMax.

(b) The front view and top view of the field CCNY concrete slab. It contains 7 horizontally buried pipes, 2 diagonally buried pipes, and 1 short copper pipe.

**Fig. 9.** The ground truth of the testing slabs. (a) synthetic concrete slab. (b) field concrete slab. Note that in (b), the black bounding box demonstrates the field data collection area.

effectiveness between our method and other approaches, and discuss the effectiveness of the DielectricNet model.

#### A. Data Preparation

To verify the proposed Deep Neural Network (DNN) models in this paper, we prepare a GPR B-scan dataset for training and testing purposes. The dataset we provide contains both synthetic and field B-scan data.

**1) Field GPR Data Collection:** We firstly collected the field GPR data with our Omni-directional robot on the CCNY concrete slab whose layout is defined in Figure 9(b). This concrete slab have a dimension of  $2.40(m) \times 0.90(m) \times 0.25(m)$  ( $length \times width \times thickness$ ). The dielectric value of this slab is approximately 7. There are 10 pipes buried in the slab with different size, depth and material. Specifically, it contains 7 horizontally buried pipes, 2 diagonally buried pipes, and 1 short copper pipe. As mentioned in Section II, the GPR sensor we used is a GSSI PaveScan RDM 1.0, with 2 GHz frequency and 20cm max depth detection range. We conducted 24 automated GPR tests which contribute 120 raw B-scan data to our dataset.

**2) Synthetic GPR Data Generation:** However, the collected field data is still not enough for the DNN model training purpose. Thus, by taking advantage of gprMax [30], we build a synthetic testing environment that simulates the real NDT condition.

Specifically, to match with the real commercial GPR (i.e., GSSI PaveScan RDM 1.0), we setup the simulated GPR pulse to a Gaussian norm wave, which has a central frequency  $f_c = 2GHz$ . In addition, the distance between transmitter and receiver of the antenna is set to 5cm, with a sensing time window as 5ns. Then, we generate a synthetic slab dataset using gprMax. The surrounding medium of the synthetic slab is set to multiple dielectric values to emulate different environments. In particular, the relative permittivity varies from 4 to 10 with a 0.5 increment as the interval. The conductivity is set to 0.01. We also assume the surrounding medium has a non-magnetic property, thus the relative permeability is set to 1. We then randomly put 1 to 5 cylindrical objects into each slab, where all the objects are designed as the Perfect

Electric Conductor (PEC). When conducting synthetic GPR data generation, we also make the spacing of consecutive GPR measurements to 5 mm to match with the same property in our field data collection.

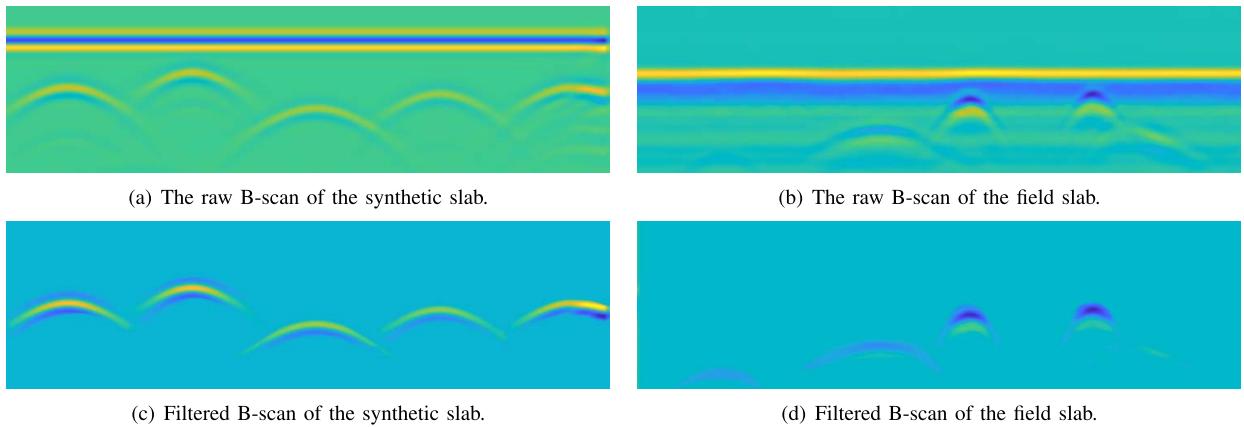
In the end, we generate 160 synthetic models and collect 9 B-scan data over each synthetic model. Combined with 120 field B-scan data, we eventually obtain the 1560 B-scan data in total, and use 1100 for training, 300 for testing, and 160 for evaluation.

**3) Radargram Data Annotation:** To train the learning-based modules and conduct the corresponding experiments, we hence introduce the data annotation for ground truth data generation.

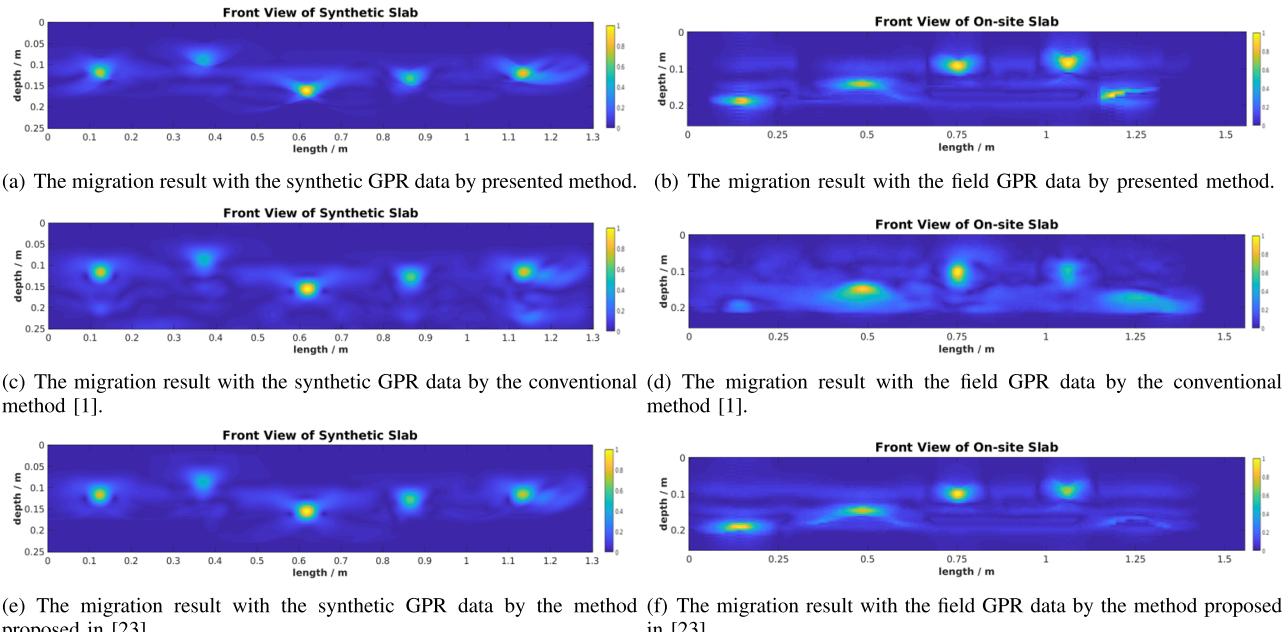
We use LabelMe [31] software to annotate the hyperbolic features for B-scan image set  $\mathcal{S}^I$ . The annotated radargram images are further used as the ground truth in *noise cancellation module*. As for the *dielectric estimation module*, on the one hand we can easily determine the ground truth dielectric value for synthetic GPR data, because we designed it on purpose. On the other hand, since we know the material of the field slab is concrete, we can check the dielectric table for the nominal value and make a calibration to determine its real value. Hence, once a dielectric value of the surrounding medium is predicted, we can calculate the traveling velocity of the signal and thus estimate the target depth according to Equ 7. Note that the estimation precision of the dielectric prediction is discussed in Section IV-C.

#### B. Comparison Study

As illustrated in Figure 9 to Figure 12, both synthetic and field GPR data have been used to validate the effectiveness of our method. Specifically, Figure 9(a) shows a synthetic concrete slab which is 1.3 meters long, 0.25 meters deep and 0.25 meters wide. In addition, Figure 9(b) shows the ground truth of the field concrete slab, where the dielectric value of the concrete is close to 7, and the dielectric value of the PVC pipes is equal to 3. Note that the black bounding box showing in Figure 9(b) demonstrates the field GPR data collection area, which is 1.55 meters long, 0.25 meters deep and 0.25 meters wide. Specifically, Figure 10 shows the raw B-scan data and filtered B-scan data respectively. Figure 11



**Fig. 10.** Qualitative results on B-scan images. The raw B-scan data and denoised B-scan data.



**Fig. 11.** GPR 2D imaging qualitative comparison results in front view. (a), (b) show the GPR imaging result with our method; (c), (d) show the conventional GPR 2D imaging method introduced in [1]; (e), (f) show another learning-based method introduced in [23].

and [Figure 12](#) show the 2D and 3D metric GPR imaging results.

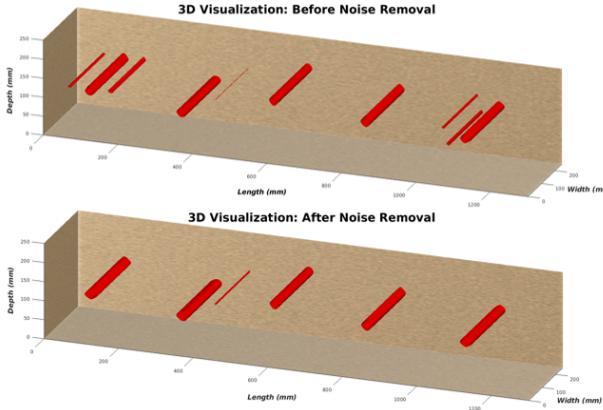
To further validate the proposed method, we compare our approach with the conventional BP method [1] and our previous learning-based GPR imaging method proposed in [23], where different DNN models were implemented for noise removal module and dielectric estimation module.

[Figure 11](#) demonstrate 2D GPR imaging results generated by the above three methods respectively. In particular, [Figure 11\(c\)](#) and [Figure 11\(d\)](#) show that a lot of background noise exist in 2D imaging results using BP method. In [Figure 11\(e\)](#) and [Figure 11\(f\)](#), the noise is partially removed and the result is sharper using the learning-based GPR imaging method in [23]. The [Figure 11\(a\)](#) and [Figure 11\(b\)](#) show the 2D GPR imaging generated by our method. It can be concluded that the results of our method are sharper and have less noise than other methods. [Figure 12](#) illustrates 3D visualization *top* view of the reconstruction results before and after using our noise removal module.

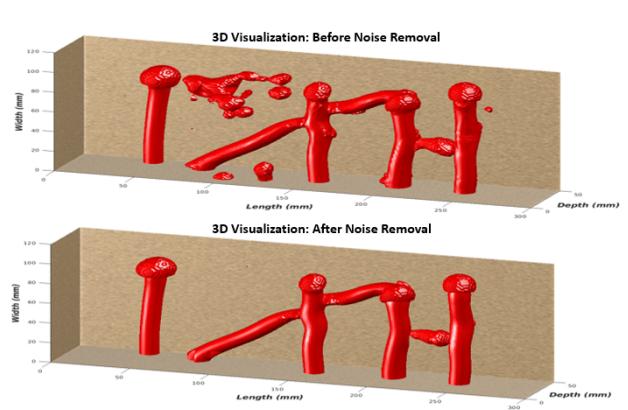
**TABLE II**  
QUANTITATIVE RESULTS ON GPR IMAGING EFFECTIVENESS  
COMPARISON. EFFECTIVENESS COMPARISON BETWEEN THE  
PROPOSED DNN-FACILITATED GPR IMAGING METHOD  
AND OTHERS WITH MULTIPLE METRICS

	<i>MSE</i> ↓	<i>SNR</i> ↑	<i>SSIM</i> ↑
Ours w/ synthetic dataset	<b>284.774</b>	<b>16.982</b>	<b>0.987</b>
[23] w/ synthetic dataset	351.455	16.068	0.976
[1] w/ synthetic dataset	453.541	14.961	0.983
Ours w/ field dataset	<b>397.384</b>	<b>15.810</b>	<b>0.980</b>
[23] w/ field dataset	485.132	14.943	0.971
[1] w/ field dataset	983.156	11.876	0.938

[Table II](#) provides a quantitative evaluation of three baseline methods regarding the image noise and similarity level compared to the ground truth image. The ground truth image indicates the cross section of the pipes in a concrete slab, similar to [Figure 9\(a\)](#). Given the 2D reconstruction image *X* and 2D ground truth image *Y*, we evaluate them with the



(a) The 3D GPR data reconstruction comparison result with the synthetic GPR data from the *front* view.



(b) The 3D GPR data reconstruction comparison result with the field GPR data from the *top* view.

**Fig. 12.** GPR 3D imaging qualitative comparison results in top view. (a) shows the 3D GPR imaging results, which is corresponding to the synthetic data in Figure 9(a). (b) shows the 3D GPR imaging results, which is corresponding to the field data collected in the region of interest (black box) in Figure 9(b).

following different metrics: *Mean Square Error*, *Signal-to-Noise-Ratio* and *Structural Similarity Index* on both synthetic dataset and field dataset as illustrated in Equ.11-Equ.13. For Mean Square Error, the lower the value, the better performance is. For Signal-to-Noise-Ratio and Structural Similarity Index, the larger the value, the better effectiveness it represents.

$$\text{MSE} = \frac{1}{n} \sum_{i=1, j=1}^n (X_{i,j} - Y_{i,j})^2 \quad (11)$$

$$\text{SNR} = 10 * (\log_{10}(X/Y)) \quad (12)$$

$$\text{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (13)$$

where in Equ.11,  $i, j$  denotes pixel index in image  $X$  and image  $Y$ ; in Equ.12, we take  $X$  as noise signal and compare it w.r.t ground truth image  $Y$ . At last, in Equ.13,  $\mu_X$ ,  $\mu_Y$ ,  $\sigma_X$ ,  $\sigma_Y$ , and  $\sigma_{XY}$  are the local means, standard deviations, and cross-covariance for images  $X, Y$ ,  $C_1$  and  $C_2$  are constant values.

We can find that our new learning-based method outperforms baseline methods. This is because our noise removal module uses a segmentation model, and it returns a mask region that covers the hyperbolic feature. In contrast, the noise removal module in [23] uses a detection network, and it would only return a bounding box region on the B-scan image. Besides covering the hyperbolic feature in the B-scan data, the bounding box region usually contains more noise from B-scan data than the mask region.

### C. Why CRNN Module Is Used for Dielectric Prediction?

The CRNN module has widely been used for the *scene text recognition* tasks in computer vision community [24]. The input to the CRNN module is the image that contains text, while the output is the text sequence extracted from the image. This end-to-end method is proven to be effective and inspired us to consider that the hyperbolic features contained in GPR B-scan images could also be analyzed in a similar way by the CRNN module. The hyperbolic features are normally

determined by several factors including the dielectric property of surrounding medium, the dimension, material, and depth of the buried utility pipelines. As long as the dataset contains enough objects with different combinations of the above properties, the CRNN module shall be able to predict the dielectric value of the surrounding medium with the variant of other factors. Since the dielectric value and depth of hyperbolic features are correlated, we can use a CRNN module to encode this property as an abstract sequence, and further decode and estimate this abstract sequence as a particular dielectric value. Specifically, the CNN encoder we implement is *ResNet34* and the RNN decoder is *Bi-LSTM*. In addition, to better illustrate the effectiveness of our CRNN module, we compare it with other architectures of learning models, such as the combinations of *VGG16* encoder and *Bi-LSTM* decoder; a plain CNN encoder with a *Bi-LSTM* decoder; a single *VGG16* model and *ResNet34* model respectively. All baseline model architectures are illustrated in Table III, the module input is a filtered B-scan while the module output is an estimated dielectric value.

In this quantitative experiment, we compare the ground truth of dielectric value with the estimated value predicted by our CRNN module and other baseline models. If the estimated dielectric value equals to the ground truth value, we consider this prediction as a true positive prediction; otherwise, this prediction is a false positive prediction. To keep the consistence, we use the same hyperparameters such as learning rate, batch size and the number of epochs in each baseline training process. The precision score and running time is shown in Table IV. We can conclude that our CRNN model achieves the highest dielectric prediction precision and the second to the fastest in inference running time compared with other baselines. Equ. 14 illustrates the metric for precision calculation. TP and FP indicate the number of true positive predictions and the number of false positive predictions in dielectric value estimation respectively.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (14)$$

TABLE III

STRUCTURE OF BASELINE MODELS FOR DIELECTRIC ESTIMATION. [\*], [|] AND <\*> STAND FOR PLAIN CONVOLUTIONAL BLOCKS, RESIDUAL CONVOLUTIONAL BLOCKS AND FULLY-CONNECTED BLOCKS RESPECTIVELY

VGG16 + Bi-LSTM		Plain CNN + Bi-LSTM		VGG16		ResNet34	
Layer	Configurations	Layer	Configurations	Layer	Configurations	Layer	Configurations
Input						$W \times 1024$ B-scan $B_k^f$	
conv1_x	$\left[ 3 \times 3, 64 \right] \times 2$	conv1	$7 \times 7, 64$ , stride 2	conv1_x	$\left[ 3 \times 3, 64 \right] \times 2$	conv1	$7 \times 7, 64$ , stride 2
MaxPooling	$3 \times 3$ , stride 2	MaxPooling	$3 \times 3$ , stride 2	MaxPooling	$3 \times 3$ , stride 2	MaxPooling	$3 \times 3$ , stride 2
conv2_x	$\left[ 3 \times 3, 128 \right] \times 2$	conv2_x	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	conv2_x	$\left[ 3 \times 3, 128 \right] \times 2$	conv2_x	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$
MaxPooling	$3 \times 3$ , stride 2	—	—	MaxPooling	$3 \times 3$ , stride 2	—	—
conv3_x	$\left[ 3 \times 3, 256 \right] \times 3$	conv3_x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	conv3_x	$\left[ 3 \times 3, 256 \right] \times 3$	conv3_x	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$
MaxPooling	$3 \times 3$ , stride 2	—	—	MaxPooling	$3 \times 3$ , stride 2	—	—
conv4_x	$\left[ 3 \times 3, 512 \right] \times 3$	conv4_x	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	conv4_x	$\left[ 3 \times 3, 512 \right] \times 3$	conv4_x	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$
MaxPooling	$3 \times 3$ , stride 2	—	—	MaxPooling	$3 \times 3$ , stride 2	—	—
conv5_x	$\left[ 3 \times 3, 512 \right] \times 3$	conv5_x	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	conv5_x	$\left[ 3 \times 3, 512 \right] \times 3$	conv5_x	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$
—	—	—	—	MaxPooling	$3 \times 3$ , stride 2	Average Pool	$7 \times 7$ , stride 1
Bidirectional-LSTM	hidden units:512	Bidirectional-LSTM	hidden units:512	FC	$\langle 4096 \rangle \times 2$	FC	1000
Bidirectional-LSTM	hidden units:256	Bidirectional-LSTM	hidden units:256	FC	1000	—	—
Transcription	$\mathcal{D}$	Transcription	$\mathcal{D}$	softmax	$\mathcal{D}$	softmax	$\mathcal{D}$

#### D. Ablation Study

To better analyze the contributions of different DNN modules (e.g., noise removal module and dielectric estimation module) of our GPR imaging system, we further perform ablation studies. This experiment is designed by replacing or dropping off each DNN module. We first replace our noise removal module with a classification-based model, *Faster RCNN*, the same as in our previous work [23]; then, we replace our dielectric estimation module with the other three baselines, such as *VGG16 + Bi-LSTM*, *VGG16*, and *ResNet34*; in the end, we only keep the proposed noise removal module and dielectric estimation module respectively and drop the other module, to make a fair and comprehensive comparison.

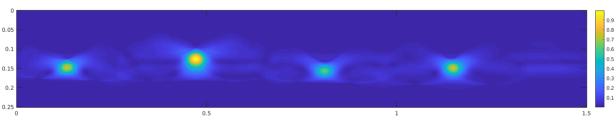
We list different model architectures of the dielectric estimation module, such as *Bi-LSTM*, *VGG16*, and *ResNet34*, in Table III. The input to these models is the same, which is the filtered GPR B-scan data  $B_k^f$ . As for the *UNet* and *Faster RCNN*, we use the original network structure and utilize them to process the input B-scan data  $B_k$ . Note that we need to adjust the input image size according to the model implemented in noise removal module. Specifically,

when we use *Faster RCNN* to remove the noise in B-scan data [23], we need to convert the B-scan data  $B_k$  to an image, and then resize it to  $224 * 224$ . Once we obtain the output image with a bounding box, we up scale the image back to its original size and further filter the B-scan based on the bounding box coordinates. Similarly, when we use a segmentation-based model, *UNet*, as a filter, the input GPR image size is  $1/4$  down-scaled of its original size, which equals to  $w * 256$ , where  $w$  indicates the width of a GPR image. Because in our experiments, the width of GPR image varies while the depth of GPR images equals to 1024, which represents the number of samples in a A-scan. Then, when we obtain the mask image, we up scale the mask image to its original size and filter the input B-scan data based on the mask pixels' coordinate.

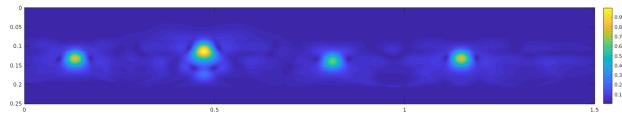
In this evaluation, we first quantitatively evaluate those baselines with the following metrics: MSE, SNR and SSIM. The results are shown in Table V and Table VI. Then, we provide the qualitative comparison between the baselines as shown in Figure 13. It can be concluded that our segmentation-based noise removal module contributes



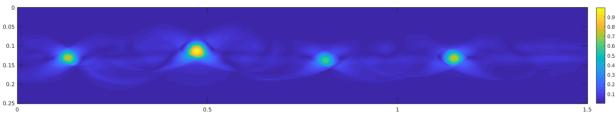
(a) The original B-scan image.



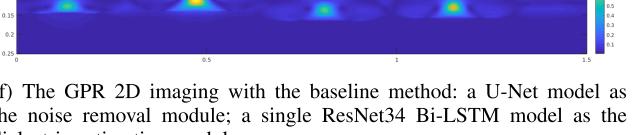
(b) The GPR 2D imaging with our proposed method: a U-Net model as the noise removal module; a ResNet34 with Bi-LSTM model as the dielectric estimation module.



(c) The GPR 2D imaging with the baseline method: a Faster-RCNN model as the noise removal module; a ResNet34 with Bi-LSTM model as the dielectric estimation module.



(e) The GPR 2D imaging with the baseline method: a U-Net model as the noise removal module; a single VGG16 model as the dielectric estimation module.



(g) The GPR 2D imaging with the baseline method, which only contains a U-Net model as the noise removal module.

(h) The GPR 2D imaging with the baseline method, which only contains a CRNN model as the dielectric estimation module.

**Fig. 13.** Qualitative comparison between the baseline methods. Different baselines in noise removal module and dielectric estimation module are replaced and evaluated.**TABLE IV**

QUANTITATIVE RESULTS BETWEEN PROPOSED DIELECTRIC ESTIMATION METHOD AND OTHERS ON PRECISION AND TIME COST

	Precision%↑	Time(sec.)↓
Ours (ResNet34 + Bi-LSTM)	<b>98.3</b>	0.0052
VGG16 + Bi-LSTM	98.1	0.0056
Plain CNN + Bi-LSTM	98.1	0.0051
VGG16	97.6	<b>0.049</b>
ResNet34	98.1	0.0050

**TABLE V**

ABLATION STUDY ON THE PROPOSED DNN-FACILITATED GPR IMAGING METHOD ON SYNTHETIC DATA

	MSE ↓	SNR ↑	SSIM ↑
Ours (U-Net + ResNet34 + Bi-LSTM)	<b>284.774</b>	<b>16.982</b>	<b>0.987</b>
Faster R-CNN + ResNet34 + Bi-LSTM	335.256	16.538	0.979
U-Net + VGG16 + Bi-LSTM	305.432	16.012	0.985
U-Net + VGG16	317.921	15.912	0.974
U-Net + ResNet34	309.310	16.462	0.983
U-Net	364.264	16.488	0.973
ResNet34 + Bi-LSTM	382.194	15.363	0.969

around 3% metric improvements compared with the previous classification-based method [23]. It indicates that a B-scan data filtered by a segmentation model gains less noise than a classification model. When replacing different backbones in dielectric estimation modules, the performance is not significantly reduced, but it increases the inference time as indicated in **Table IV**. At last, when removing any of the two DNN

**TABLE VI**

ABLATION STUDY ON THE PROPOSED DNN-FACILITATED GPR IMAGING METHOD ON FIELD DATA

	MSE ↓	SNR ↑	SSIM ↑
Ours (U-Net + ResNet34 + Bi-LSTM)	<b>397.384</b>	<b>15.810</b>	<b>0.980</b>
Faster R-CNN + ResNet34 + Bi-LSTM	442.737	15.083	0.973
U-Net + VGG16 + Bi-LSTM	422.454	15.209	0.980
U-Net + VGG16	433.144	15.394	0.976
U-Net + ResNet34	424.562	15.093	0.978
U-Net	502.294	13.902	0.966
ResNet34 + Bi-LSTM	521.936	13.109	0.955

modules, the performance is degraded but still outperforms the conventional method [1]. That is to say, the DNN-facilitated modules are essential in terms of improving the effectiveness of the GPR imaging.

## V. CONCLUSION

This paper introduces a robotic GPR data collection platform and a learning-based 3D metric GPR imaging method, which can improve the GPR imaging performance by taking the following steps. 1) Introducing a robotic GPR data collection platform that provides free motion pattern to collect the GPR scans and tag the position with the GPR measurement. 2) Using a learning-based segmentation model to remove the background noise from the raw GPR B-scan images. 3) Using a dielectric estimation module to estimate the dielectric value of the underground environment. Experimental results show that our proposed method outperforms the baseline methods in

terms of noise removal and dielectric estimation. Our method makes the GPR data collection significantly easier by enabling the automatic scan of the flat surface in a free-motion pattern with minimal human intervention. It eliminates the time and cost to laying out grid lines on flat terrain and reducing the hassle to closely follow the grid lines and the note-taken time to record the linear motion trajectory in X-Y directions of current GPR data collection practice.

Our proposed method assumes the dielectric property is stable which is valid only in certain conditions. In some natural situations, the dielectric value may not be a constant and varies with depth, subsurface surrounding medium, water content and other variables. For future work, we would like to address this issue and collect more GPR data for learning-based GPR imaging processing.

### ACKNOWLEDGMENT

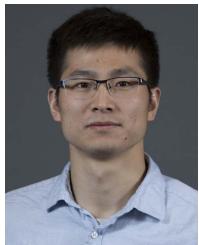
Jizhong Xiao has significant financial interest in InnovBot LLC, a company involved in research and development and commercialization of the technology.

### REFERENCES

- [1] X. Xie, J. Zhai, and B. Zhou, "Back-fill grouting quality evaluation of the shield tunnel using ground penetrating radar with bi-frequency back projection method," *Autom. Construct.*, vol. 121, Jan. 2021, Art. no. 103435.
- [2] W. A. Tanoli, A. Sharafat, J. Park, and J. W. Seo, "Damage prevention for underground utilities using machine guidance," *Autom. Construct.*, vol. 107, Nov. 2019, Art. no. 102893.
- [3] U. Peer and J. G. Dy, "Automated target detection for geophysical applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1563–1572, Mar. 2017.
- [4] N. Smitha, D. R. Ullas Bharadwaj, S. Abilash, S. N. Sridhara, and V. Singh, "Kirchhoff and F-K migration to focus ground penetrating radar images," *Int. J. Geo-Eng.*, vol. 7, no. 1, p. 4, Dec. 2016.
- [5] J. Gazdag, "Wave equation migration with the phase-shift method," *Geophysics*, vol. 43, no. 7, pp. 1342–1351, 1978.
- [6] J. F. Claerbout and S. M. Doherty, "Downward continuation of moveout-corrected seismograms," *Geophysics*, vol. 37, no. 5, pp. 741–768, 1972.
- [7] S. Demirci, E. Yigit, I. H. Eskidemir, and C. Ozdemir, "Ground penetrating radar imaging of water leaks from buried pipes based on back-projection method," *NDT & E Int.*, vol. 47, pp. 35–42, Apr. 2012.
- [8] J. E. Mast and E. M. Johansson, "Three-dimensional ground-penetrating radar imaging using multifrequency diffraction tomography," *Proc. SPIE*, vol. 2275, pp. 196–203, Sep. 1994.
- [9] E. M. Johansson and J. E. Mast, "Three-dimensional ground-penetrating radar imaging using synthetic aperture time-domain focusing," *Proc. SPIE*, vol. 2275, pp. 205–214, Sep. 1994.
- [10] M. Pereira *et al.*, "3-D multistatic ground penetrating radar imaging for augmented reality visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5666–5675, Aug. 2020.
- [11] M. Pereira, D. Burns, D. Orfeo, R. Farrel, D. Hutson, and T. Xia, "New GPR system integration with augmented reality based positioning," in *Proc. Great Lakes Symp. VLSI*, May 2018, pp. 341–346.
- [12] L. Mihailescu and V. A. Negut, "Methods and systems for tracking and guiding sensors and instruments," U.S. Patent 9 561 019, Feb. 7, 2017.
- [13] L. Mihailescu and V. A. Negut, "Methods and systems for tracking and guiding sensors and instruments," U.S. Patent 10 426 350, Oct. 1, 2019.
- [14] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 3400–3407.
- [15] H. Li, C. Chou, L. Fan, B. Li, D. Wang, and D. Song, "Toward automatic subsurface pipeline mapping by fusing a ground-penetrating radar and a camera," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 722–734, Apr. 2020.
- [16] C. Chou, S.-H. Yeh, and D. Song, "Mirror-assisted calibration of a multi-modal sensing array with a ground penetrating radar and a camera," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1457–1463.
- [17] C. Chou, A. Kingery, D. Wang, H. Li, and D. Song, "Encoder-camera-ground penetrating radar tri-sensor mapping for surface and subsurface transportation infrastructure inspection," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1452–1457.
- [18] J. Feng, L. Yang, E. Hoxha, D. Sanakov, S. Sotnikov, and J. Xiao, "GPR-based model reconstruction system for underground utilities using GPRNet," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 845–851.
- [19] L. Yang, Y. Chang, S. Sotnikov, J. Feng, B. Li, and J. Xiao, "Wind-turbine blade inspection using impact-sounding module and acoustic analysis," in *Proc. IEEE 9th Annu. Int. Conf. CYBER Technol. Autom., Control, Intell. Syst. (CYBER)*, Jul. 2019, pp. 253–258.
- [20] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. Montiel, and J. D. Tardós, "ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimodal slam," *IEEE Trans. Robot.*, vol. 37, no. 6, pp. 1874–1890, Dec. 2021.
- [21] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of vicon system positioning performance," *Sensors*, vol. 17, no. 7, p. 1591, Jul. 2017.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.
- [23] J. Feng, L. Yang, H. Wang, Y. Song, and J. Xiao, "GPR-based subsurface object detection and reconstruction using random motion and DepthNet," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 7035–7041.
- [24] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.
- [25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [26] X. Rong, C. Yi, and Y. Tian, "Unambiguous text localization and retrieval for cluttered scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5494–5502.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
- [29] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 369–376.
- [30] 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.
- [31] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2008.



**Jinglun Feng** received the B.S. degree in electrical engineering from Shandong Jianzhu University in 2015 and the M.S. degree in control engineering from Shandong University in 2018. He is currently pursuing the Ph.D. degree in electrical engineering with the City College of New York, CUNY, New York, NY, USA. His research interests are in 3D vision, deep learning, object pose estimation, sensor fusion, and intelligent inspection for robotics applications.



**Liang Yang** received the B.S. degree from Shenyang Aerospace University, Shenyang, China, in 2012, the Ph.D. degree in electronics engineering from the City College of New York (CUNY City College) in 2019, and the Ph.D. degree in pattern recognition and intelligent system from the University of Chinese Academy of Sciences in 2019. In 2019, he joined Apple as a Senior 3D Computer Vision Researcher, where he is currently working on 3D visual perception and understanding. His research interests cover motion and path planning, 3D perception and understanding, visual SLAM and reconstruction, multi-sensor fusion, and robotic control.



**Ejup Hoxha** received the B.S. degree in electrical and computer engineering with focus in control theory and robotics from the University of Prishtina and the M.S. degree in computer engineering from The City College, City University of New York, where he is pursuing the Ph.D. degree. He worked several years in industry as a control engineer, SCADA developer, and software developer. He is currently working as a Researcher with the CCNY Robotics Lab, with a focus on robotics, control, visual SLAM, deep learning, sensor fusion, and acoustic NDT.



**Jizhong Xiao** (Senior Member, IEEE) received the B.S. and M.S. degrees from the East China Institute of Technology, Nanjing, China, in 1990 and 1993, respectively, the M.E. degree from Nanyang Technological University, Singapore, in 1999, and the Ph.D. degree from Michigan State University in 2002. He is a Professor of Electrical Engineering at the City College of New York (CCNYCUNY City College) and a Doctoral Faculty Member of the Ph.D. Program in Computer Science at the CUNY Graduate Center. He started the robotics research program at CCNY in 2002 as the Founding Director of the CCNY Robotics Lab. He has published more than 160 research articles in peer-reviewed journal and conferences. His current research interests include robotics and control, cyber-physical systems, autonomous navigation and 3D simultaneous localization and mapping (SLAM), real-time and embedded computing, assistive technology, multi-agent systems and swarm robotics. He received the U.S. National Science Foundation CAREER Award in 2007, the CCNY Outstanding Mentor Award in 2011, and the Humboldt Research Fellowship for Experienced Researchers from the Alexander von Humboldt Foundation, Germany, from 2013 to 2015.