remote sensing

MDPI

Article

CGL-Net and GPCL: A Hybrid Deep Learning Approach for Coarse Root Localization and Multi-Parameter Estimation Using Ground-Penetrating Radar

Jiaying Li [1,2], Guoqiu Fan [1,2], Yili Zheng [1,2,3], Lifeng Wang [1,2,3], Benye Xi [4], Qiaoling Han [1,2,3] and Hao Liang [1,2,3,*]

[1] School of Technology, Beijing Forestry University, Beijing 100083, China; lijiaying@bifu.edu.cn (J.L.)
[2] Key Laboratory of Forestry Equipment and Automation, National Forestry and Grassland Administration, Beijing 100083, China
[3] State Key Laboratory of Efficient Production of Forest Resources, Beijing 100083, China
[4] College of Forestry, Beijing Forestry University, Beijing 100083, China
* Correspondence: lianghao@bjfu.edu.cn

Highlights

What are the main findings?

- A multi-source GPR root dataset of simulated, style-transferred and measured data was built.
- CGL-Net with CBAM, Ghost convolutions, and LSTM reaches an *F1-score* of 0.894.
- GPCL was integrated with CGL-Net for root localization, lifting precision to 0.901.
- Field validation: mean depth error of 3.3 cm and mean absolute diameter error of 0.7 cm in the root-by-root excavation comparison.

What are the implications of the main findings?

- The proposed pipeline offers an accurate tool for non-destructive root mapping.

Abstract

To address the challenges of complex imagery, strong clutter interference, and low localization accuracy in ground-penetrating radar (GPR) detection of tree roots, this study proposes a coarse root localization and multi-parameter estimation method integrating CGL-Net with GPCL. A multi-source dataset combining purely simulated data, CycleGAN style-transferred data, and field-measured data is constructed, and CGL-Net, which fuses CBAM attention, Ghost double convolutions, and an LSTM temporal module, generates root weight localization maps end-to-end. The three-stage localization framework of GPCL is then integrated with the network output, and its parameters are made sample-adaptive so that the weight localization maps can be converted into root locations robustly through multi-directional gradient search, adaptive DBSCAN clustering, and weighted centroid localization. On the simulated test set, CGL-Net reaches an F1-score of 0.894 while using 44.8% fewer parameters than standard U-Net, and its sample-adaptive localization stage raises precision by 0.171 and F1 by 0.073 relative to threshold segmentation, outperforming YOLOv5s, RT-DETR, CMT-Net, GPR-UNet, and GPRInvNet. Across the pooled three-source dataset, the correct rates of depth and diameter estimation are 87.2% and 91.7% at tolerances of 30% and 20%, and on the field-measured data the correct rates of elevation and azimuth angle are 92.1% and 72.4% at a tolerance of 5.0°. In the root-by-root comparison with the excavation of nine coarse roots, the mean absolute errors are 3.3 cm in depth and 0.7 cm in diameter. Field-measured errors are further characterized at the distribution level, with median depth and diameter errors of 22.2% and 22.7% (95% cluster-bootstrap *CIs*: 15.2–28.5% and 15.9–27.4%). The results show that a network emitting continuous weight

Check for updates

Academic Editors: Feifei Hou, Jian Dong and Xianghuan Luo

Received: 13 October 2026
Revised: 20 November 2026
Accepted: 23 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

maps, combined with a parameter-adaptive localization stage, provides a practical route to non-destructive coarse-root detection at the individual-root scale.

Keywords: ground penetrating radar; deep learning; root localization; root morphological parameters; GPCL; DBSCAN; CBAM

1. Introduction

Roots are the core organs of plants for anchorage, water and nutrient uptake, and the synthesis of active substances [1]. Their growth and architecture not only directly regulate aboveground yield and quality, but also profoundly influence the carbon cycle and nutrient balance of terrestrial ecosystems through carbon input and soil structure improvement [2,3]. However, the invisibility of the underground environment and the limitations of traditional destructive sampling methods have long kept root research behind aboveground studies, creating an urgent need for efficient, non-destructive detection and quantitative assessment techniques [4,5]. Traditional root detection methods such as the profile method and soil coring can directly obtain root biological parameters, but they destroy the soil structure of the root zone and the roots themselves, are complicated and time-consuming, and are constrained by the spatial limitations of experimental sites, making it difficult to comprehensively evaluate the spatial distribution of root systems. Ground penetrating radar (GPR), as a non-invasive geophysical technique [6,7], offers the advantages of rapidity and operational simplicity and has been widely used for non-destructive root detection over the past two decades [8]. Because roots and soil differ in water content, their dielectric properties differ accordingly; the high-frequency electromagnetic pulses emitted by GPR into the subsurface are reflected at roots, and by analyzing the echo signals recorded by the receiving antenna, root identification and localization as well as inversion of morphological parameters can be achieved [9].

In recent years, automatic GPR-based root recognition has gradually become a research hotspot, and various machine learning and deep learning methods have been employed to improve recognition accuracy and robustness. Lei et al. proposed a deep learning framework based on Faster R-CNN and data augmentation to recognize hyperbolic features in GPR B-scan images, and combined the DCSE algorithm with the CTFP method to separate target point clusters and extract hyperbola fitting points, validating its effectiveness on both synthetic and measured datasets [10]. Hou et al. proposed a detection method based on an improved Mask Scoring R-CNN that effectively detects and segments hyperbolic targets in GPR images, outperforming several existing methods on field tree-root data [11]. Wang et al. proposed a random Hough transform algorithm based on YOLOv3 for complex environmental backgrounds, achieving automatic recognition of root reflection hyperbolic features and non-destructive estimation of root characteristic parameters, with recognition accuracy and recall of 96.62% and 86.94%, respectively [12]. To address the limited interpretability of current deep learning algorithms for B-scan data, Xu proposed an improved root target recognition and localization algorithm based on DNN back-projection, achieving precise root localization while reducing the average computation time [13]. Chen et al. proposed an automatic localization and prediction method based on convolutional neural networks that improves the detection accuracy of geological features and provides an effective basis for tunnel construction decisions [14]. Wang augmented GPR datasets with GANs and combined Swin Transformer for automatic root target detection, with the detection network achieving a recall of 0.987 and an mAP of 0.871; based on CNN, effective estimation of root radius and depth was also achieved, with a total mean relative error of only 5.83% [15]. Yang et al. proposed an adaptive variational mode decompo-

sition (VMD)-based radar data denoising scheme that adaptively determines the mode parameters using the whale optimization algorithm and removes noise modes, significantly improving the signal-to-noise ratio of radar signals and providing a reliable data basis for target recognition in complex backgrounds [16]. Shi et al. proposed Yolov4-HPV, which combines multi-CycleGAN and wave velocity estimation to reduce the RMSE of diameter estimation to 3.43% [17]. In addition, physics-informed variational networks [18] and GPR inversion methods with an embedded prior velocity field [19] have been used to improve cross-domain generalization and uncertainty quantification. Nevertheless, these studies focus mainly on hyperbolic target detection or individual parameter estimation, whereas joint root localization and multi-parameter estimation systematically validated against field excavation data remains limited.

These gaps define the three research questions addressed in this study: whether a lightweight segmentation network can reliably delineate root hyperbolic responses in GPR B-scans; how root positions can be recovered from the predicted weight maps without fixed-threshold segmentation; and whether the accuracy established on simulated data transfers to field conditions. We hypothesize that coupling channel–spatial attention with temporal feature compensation within a U-Net-style architecture preserves segmentation accuracy at a substantially reduced model size; clustering of gradient points with per-sample adaptive parameters localizes roots more accurately than threshold-based extraction; and a multi-source training set combining simulated, style-transferred, and measured images, validated against excavation ground truth, enables reliable estimation of root depth, diameter, and angles under real conditions.

To answer these questions, we propose a multi-target root analysis pipeline that couples a segmentation network with a geometric localization stage. The CGL-Net network augments a U-Net encoder–decoder [20] with the CBAM attention mechanism (Convolutional Block Attention Module) [21], Ghost convolutions [22], and the LSTM temporal processing module (Long Short-Term Memory) [23,24]; it is trained end-to-end to emit a continuous root weight localization map rather than a binary mask. On this map, the gradient points clustering-based localization (GPCL) framework [25] converts the weight map into root positions through multi-directional gradient search [26], adaptive DBSCAN clustering [27,28], and weighted centroid estimation, and the geometric morphology and physical parameters of each root are then resolved from the clustered responses. The pipeline is trained with supervised learning on annotated simulated data and fine-tuned and validated with measured data, and is finally checked against the excavation ground truth. Accordingly, the contributions of this paper are threefold. First, the CGL-Net architecture, which augments a U-Net-style encoder–decoder with Ghost convolutions [22] and LSTM temporal compensation to segment root hyperbolas and, more importantly, to emit a continuous weight localization map rather than a binary mask. Second, an adaptive, confidence-scored variant of GPCL, which combines multi-directional gradient search [26] with DBSCAN clustering [27,28] and weighted centroid localization on the predicted weight maps, that makes the three-stage localization framework of Hou et al. [25] operable on such weakly supervised weight maps, by replacing its fixed gradient and clustering parameters with per-sample adaptive ones, adding a refinement step that separates adjacent roots, and rejecting low-confidence detections via a multi-dimensional score. Third, a systematic validation protocol built on a multi-source GPR root dataset integrating simulated, CycleGAN style-transferred [29,30], and field-measured data, with distribution-level field uncertainty quantified by root-level cluster-bootstrap confidence intervals. The quantitative results supporting each hypothesis are reported in Section 3.

2. Materials and Methods

2.1. Data Acquisition and Preprocessing

The data in this study consist of simulated data and field-measured data. The simulated data comprise 2600 B-scans, including 1600 purely simulated B-scans generated with gprMax [31] and 1000 B-scans obtained via CycleGAN style transfer; both categories of simulated data have pixel-level root masks and root instance-level labels—the labels of the purely simulated data include root center positions, burial depths, and radii, while the style-transferred data inherit the pixel-level masks and root instance-level labels of the purely simulated data. The measured data are 400 field GPR B-scans with root instance-level labels obtained through root-by-root excavation, including root positions, burial depths, diameters, and growth angles, but without pixel-level masks. Model training follows the two-stage workflow of "supervised training on annotated simulated data—fine-tuning on field data without manual pixel annotation": in the first stage, CGL-Net is trained using supervised learning on the annotated simulated data (purely simulated and style-transferred data), and all ablation and model comparison experiments are completed on this dataset; in the second stage, the field data without manual pixel annotation are added on the basis of the first-stage model to fine-tune the recognition model, in which pixel-level training targets are generated by a threshold-based pseudo-label strategy and the loss function is identical to that of the first stage, followed by testing. The fine-tuned final model is evaluated on the three data categories of simulated, style-transferred, and measured data, where the measured data are used for root localization and depth, diameter, and angle estimation, and are compared root by root with the excavation results. All data are uniformly preprocessed before being fed into the network.

2.1.1. Purely Simulated Data

Based on realistic tree-root architectures, 1600 root-system simulation scenarios were constructed, each containing two to five non-overlapping cylindrical roots randomly placed in a soil domain of 120 cm × 10 cm × 100 cm (x, y, z). The root radii were randomly generated in the range of 8–40 mm, and the root centers were constrained to $20 \leq x \leq 100$ cm and $0 \leq z \leq 80$ cm to limit boundary effects [32] (Figure 1a). The recording window of 10 ns corresponds to a maximal one-way recordable depth of $1.5/\sqrt{\varepsilon_r}$ m; roots whose two-way travel time exceeds the window are recorded with truncated responses, and their pixel-level masks are restricted to the recorded window, so that the supervision remains consistent with the image content. Because the depth bound and the permittivity are drawn independently, a small fraction of deep roots in high-permittivity soils are partially or entirely outside the recording window; these instances are retained, since a truncated or absent response paired with a correspondingly truncated mask also teaches the network the field-relevant case of a root at the edge of detectability. The root axis was first set parallel to the y-axis (perpendicular to the scan direction) and was then randomly rotated, with the yaw, pitch, and roll angles drawn from a uniform distribution over the range of −60° to 60°. With reference to the physicochemical parameters of soils and roots in real scenarios, the soil relative permittivity was randomly set in the range of 3–7 [33], the root relative permittivity in the range of 15–35 [34], and the electrical conductivities of the soil and roots were 0.05 S/m and 0.004 S/m, respectively (Figure 1b). This cylindrical, homogeneous representation follows standard practice in GPR-based root modeling [32–34] and is consistent with the observables available to the radar: at the 1.5 GHz center frequency, the wavelength in soil (7.6–11.6 cm for $\varepsilon r = 3$–7) is comparable to or larger than the root diameters (16–80 mm), so internal root anatomy is not resolvable, and the hyperbola apex, curvature, and amplitude are governed by the bulk root–soil contrast, while intra-root permittivity variation yields only secondary reflections well below the clutter level.

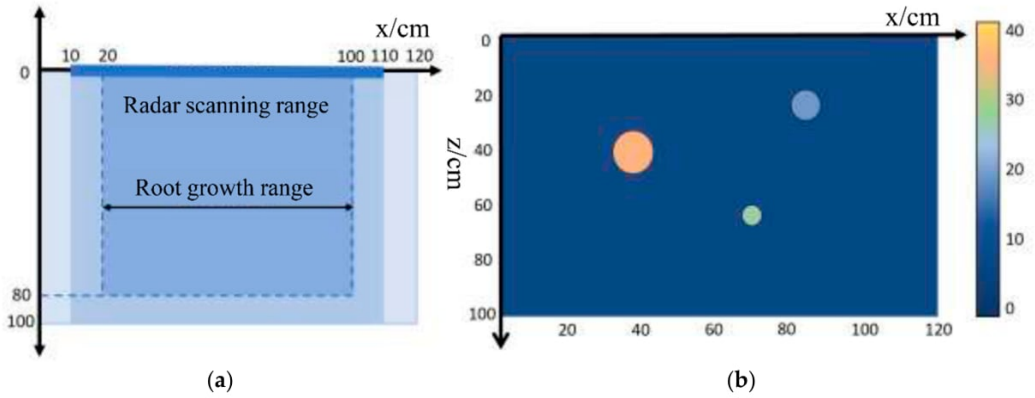


Figure 1. Schematic diagram of forward modeling: (**a**) Space and scope division. (**b**) Root distribution model for one of the samples.

The B-scan dataset was generated through forward simulation. The finite-difference time-domain (FDTD) method has been widely used in GPR forward modeling owing to its simplicity, fully explicit formulation, and robustness. In this study, the open-source software gprMax 3.1.7 [31], which is based on the FDTD method with second-order spatial and temporal differencing, was used to perform the forward simulations of the constructed root-system models. A GPR equipped with the built-in GSSI 1.5 GHz antenna model was moved along the ground surface from x = 10 cm to x = 110 cm at equal intervals of 0.5 cm to acquire the simulated B-scan images, and a total of 1600 B-scan images were obtained. The main parameters of the forward simulations are listed in Table 1.

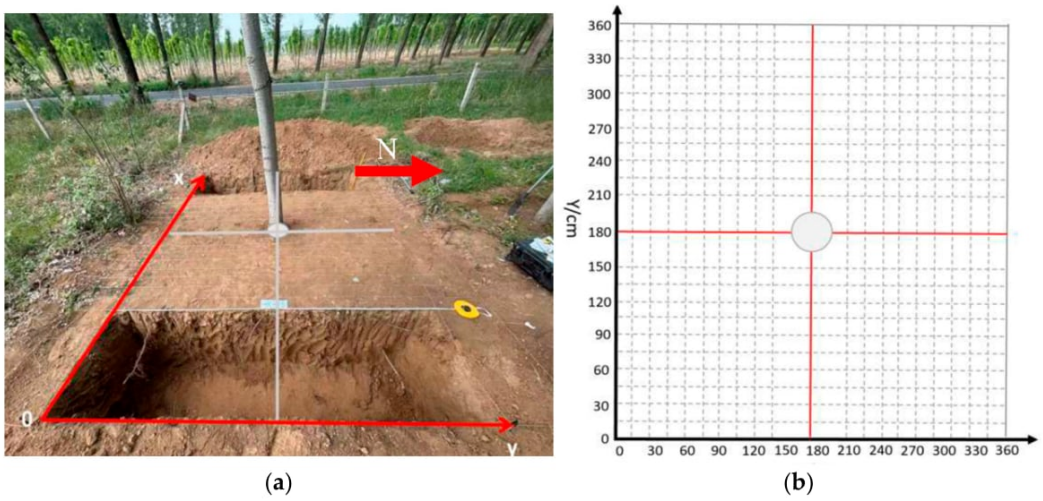
Table 1. Parameter settings of the gprMax FDTD forward simulation.

Parameter	Value
Waveform	Gaussian (built-in excitation at 1.71 GHz)
Antenna model	Built-in GSSI 1.5 GHz antenna model [32]
Resolution	2 mm
Number of A-Scans (traces per B-scan)	201
Trace spacing	5 mm
Time window	10 ns
Boundary condition	CPML (gprMax default)
Added noise	None

2.1.2. Field-Measured Data

Field data were collected at the Qingping Town forest farm, Gaotang County, Shandong Province, China (116.10°E, 36.82°N). The forest farm is located in the southern North China Plain and has a warm temperate semi-humid monsoon climate, with a mean annual temperature of 13.1 °C, a mean annual precipitation of 574.8 mm, and a mean annual sunshine duration of 2663.3 h [35]. Triploid *Populus tomentosa* cultivated in the plantation forest was selected as the research object; this species grows rapidly, has a deep and extensive root system and strong adaptability, and is a typical species in North China [36]. An example of the research object is shown in Figure 2a, with a tree height of 15.1 m and a diameter at breast height of 20.8 cm. Root profile images were collected using the grid method. First, the site was cleaned, including the removal of surface litter and the leveling of the ground. Second, a measurement area of 3.6 m × 3.6 m was delineated centered on the tree root collar, and 11 survey lines spaced 15 cm apart were laid out in each of four directions, covering radial distances of 30–180 cm from the root collar, as shown in Figure 2b. Finally, the GPR was operated along the preset survey lines to scan the roots. repeated acquisitions along

each survey line yielded a total of 400 B-scan images as the measured dataset. Unlike the simulated data, which use a single antenna (the built-in GSSI 1.5 GHz model, Table 1), the field system acquires each line at three antenna frequencies; the simulation antenna is center-frequency matched to the 1500 MHz channel, which is also the channel used exclusively for field testing (Section 2.3), whereas the 400 and 900 MHz acquisitions enter the pipeline only through the fine-tuning stage. The trace interval likewise differs between the domains (5 mm simulated; 0.00518 m raw and 0.036 m after resampling in the field) and is normalized by the per-domain resampling to the 32-column network input (Section 2.1.4). Three P. tomentosa individuals were scanned with the same 44-line grid (11 lines per direction at 15 cm spacing). Each survey line was acquired at all three antenna frequencies, yielding 44 × 3 = 132 B-scans per tree (396 in total), plus four supplementary scans along additional lines; repeated acquisitions along each line raised the total to 400 B-scan images, which constitute the measured dataset. During preprocessing, the 0–16 ns portion of each recording was retained for analysis, and the traces were resampled to an equivalent trace interval of 0.036 m, consistent with the parameters used in the subsequent processing chain (Section 2.5.1). Root excavation and ground-truth measurement. The field campaign was conducted in April 2024. Before scanning, surface litter was removed and the ground within the surveyed area was leveled as far as possible to reduce interference with the GPR measurements. Immediately after GPR acquisition, profile trenches were excavated along each survey line at 15 cm intervals to a depth of 100 cm. For each profile, a local Cartesian coordinate system was established, and the position and diameter of every root intersecting the profile with a diameter greater than 1 cm were measured at each sampling node, with depths referenced to the undisturbed soil surface. Roots were tracked across adjacent profiles by continuity of position and diameter. The ground-truth elevation and azimuth angles used in the angle evaluation (Section 3.1.2) were reconstructed from the root-axis positions measured in adjacent excavation profiles, following the same angular definitions as the GPR-based estimates (Section 2.5.2).



(**a**) (**b**)

Figure 2. Schematic diagram of the field experimental site: (**a**) Field excavation site. (**b**) Measurement lines using the grid method.

2.1.3. Simulation Data Augmentation

Acquiring a large number of high-quality root scan images under field conditions with GPR is difficult. To compensate for the shortage of measured-style images, this study introduced the CycleGAN framework on the basis of the purely simulated data, taking the purely simulated B-scans as the source domain and the measured B-scans as the unlabeled target domain, and transformed the simulated data into radar data with the measured style,

thereby augmenting the data. The effectiveness of this strategy has been verified by Qin et al. [37] by measuring the Bhattacharyya distance between the transferred images and the target-domain images: the mean distance between the transferred and measured images (0.0073) was much smaller than that between the transferred and simulated images (0.3105), indicating a high style consistency with the target domain. On this basis, a CycleGAN was trained on 200 simulated B-scans (source domain) and 200 measured B-scans (unlabeled target domain), both sets being drawn from images subsequently assigned to the training subset (Section 2.3). The generator follows the ResNet architecture with nine residual blocks, and the discriminator is a 70 × 70 PatchGAN; training was performed for 200 epochs with a batch size of 1 at a resolution of 256 × 256, using the Adam optimizer with an initial learning rate of 0.0002 ($\beta_1 = 0.5$), least-squares adversarial losses, and cycle-consistency and identity losses weighted by 10.0 and 5.0, respectively, with the learning rate kept constant for the first 100 epochs and then linearly decayed to zero, and no early stopping was applied. The trained generator was then applied to 1000 simulated B-scans to produce the style-transferred images, which inherit the pixel-level masks and root instance-level labels of the corresponding purely simulated data owing to the spatially aligned image-to-image translation, and were incorporated into the simulated dataset. The test subset was held out of training, fine-tuning and model selection. Examples of the simulated and measured B-scans, together with a style-transferred example, are shown in Figure 3.

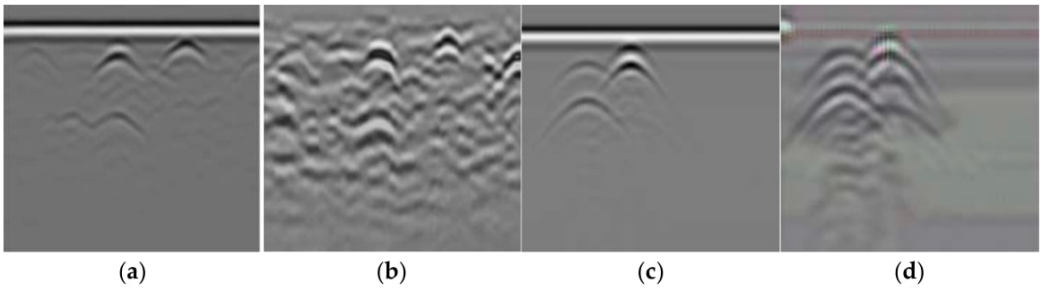


Figure 3. Examples of the datasets: (**a**) A B-scan with the simulated style. (**b**) A B-scan with the measured style. (**c**) The simulated B-scan selected for style transfer. (**d**) The style-transferred image corresponding to (**c**).

2.1.4. Data Preprocessing

To improve the dataset quality, all data were preprocessed in this study. The preprocessing consists of two stages: conventional preprocessing and multi-velocity F-K migration preprocessing, with the effects shown in Figure 4. Conventional preprocessing includes clutter suppression [38], signal enhancement [39], and data normalization. Clutter suppression was implemented by subtracting the mean value of each trace (direct-current removal) followed by adaptive background subtraction, in which the average trace over all traces was attenuated to 10–30% of its original amplitude according to an adaptive threshold and then subtracted, thereby removing the zero-frequency offset in the echo signals and weakening surface reflections and background noise (Figure 4b). Signal enhancement was implemented by a 3 × 3 median filter blended with the original data at a ratio of 7:3, followed by a linear contrast stretch with a gain of 1.2 about the mean amplitude, which suppresses isolated impulse noise and amplifies weak reflections (Figure 4c). Data normalization rescaled all amplitudes to the range of [0, 1] by min–max normalization (Figure 4d). Stolt F-K migration is a constant-velocity wave-equation migration: in the frequency–wavenumber domain, components are mapped via the constant-velocity dispersion relation, collapsing a point reflector's diffraction hyperbola to its apex when migration velocity equals medium velocity [40,41]. It is exact for a homogeneous half-space; in weakly

heterogeneous soil, a site-averaged velocity focuses most diffracted energy, with residual curvature where local velocity deviates. Because field velocity is spatially variable and unknown a priori, a single velocity would propagate bias downstream. Thus, 15 migration velocities uniformly sampled from 0.5–2.0 × 10^8 m/s, ($\varepsilon r \approx$ 2.3–36, covering simulated ε = 3–7, i.e., 1.13–1.73 × 10^8 m/s) are treated as focusing hypotheses: the matching channel yields the sharpest apex, while over/under-migration defocuses it. Stacking all channels preserves every focusing state, so the input contains each root's velocity-focusing signature and localization becomes insensitive to the velocity uncertainty in Section 4.4; the mean channel provides a stabilized aggregate. Practically, each conventionally preprocessed B-scan is Stolt-migrated at every velocity, resized to 128 × 32, stacked in ascending velocity order with the mean image as the final channel, yielding a 16 × 128 × 32 input tensor. The simulated and measured B-scans use recording windows of 10 ns and 16 ns, respectively, so this common 128-row resampling is performed within each domain over its own window, giving different temporal sampling intervals (≈0.078 vs. ≈0.126 ns per row). No absolute-time alignment is imposed: the network operates on per-domain normalized migration images, and the time-to-depth conversion is applied separately in each domain with its own window and wave velocity (Section 2.5.1), so the differing sampling intervals do not bias the depth estimates. Stolt migration was chosen for its single-pass, deterministic, phase-unambiguous transform at the B-scan scale, ensuring reproducibility and physical interpretability.

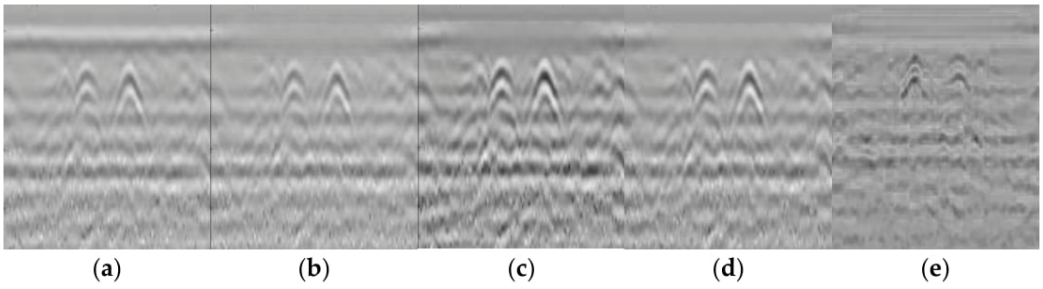


Figure 4. Comparison of the B-scan data at each preprocessing step, all panels displayed with a common amplitude scale of [0, 1]: (**a**) Raw data (horizontal axis: distance along the survey line, m; vertical axis: two-way travel time, ns). (**b**) After clutter suppression (DC removal and background subtraction). (**c**) After signal enhancement (median filtering and contrast stretching). (**d**) After amplitude normalization to [0, 1]. (**e**) channel-wise mean of the 15 single-velocity Stolt F-K migration images over 0.5–2.0 × 10^8 m/s (i.e., the mean channel of the 16-channel output tensor; vertical axis: two-way travel time, ns).

Because a single field B-scan is 3.6 m long whereas the network input covers only a 1.15 m segment, each B-scan was divided along the survey direction into non-overlapping blocks of 32 traces, and each block was fed to the network independently. This decomposition does not change the unit of the dataset. During training the loss is computed block by block, so that the network is never required to produce a prediction spanning a block boundary; during inference the block outputs of one B-scan are returned to their original along-line offsets and reassembled into a single response map covering the whole survey line, from which root positions and diameters are read. Responses of the same root that were separated at a block boundary are then merged when their centroids lie within 0.10 m along the line and 0.05 m in depth, and the depth span used for the diameter regression (Section 2.5.1) is measured over the union of the two clusters.

The mean image of the multi-velocity migration is shown in Figure 4e, where the color represents the normalized amplitude, the horizontal axis denotes distance (m), and the

vertical axis denotes the two-way travel time (ns); all panels share the same color scale for comparability. The parameter settings of each preprocessing step are listed in Table 2.

Table 2. Parameter settings of the preprocessing.

Stage	Step	Parameter	Value
Conventional	DC removal	—	per-trace mean subtraction
	Background subtraction	0.3 (above threshold)/0.1 (below)	attenuation of average trace
	Median filtering	mean + 0.5 σ of the average trace	adaptive threshold
	Contrast enhancement	3 × 3/0.7:0.3	kernel/blending ratio
	Normalization	1.2	linear gain about the mean
		[0, 1] (min–max)	range
F-K migration	Velocity sampling	0.5–2.0 × 10^8 m/s/15 (uniform)	range/number
	Output	15 velocity images (ascending) + 1 mean image, 16 × 128 × 32	channels
Time–depth conversion	Wave velocity/depth offset	0.1272 m/ns/0.2048 ns	-

2.2. Root Detection Based on CGL-Net

This study proposes a root detection algorithm based on CGL-Net (CBAM-Ghost-LSTM Network) to automatically identify the spatial positions of roots in GPR B-scans. The network architecture is shown in Figure 5 and consists of an input module, an encoder–decoder module, and an output module; the layer parameters of each module are listed in Table 3. CBAM and the LSTM temporal module are designed for two challenges of root detection in GPR B-scans: CBAM re-weights features along the channel and spatial dimensions to alleviate target–background feature confusion, whereas the LSTM module models contextual dependencies along the depth direction of each trace to enhance the responses of weak deep targets. The combined benefit of the two modules is quantified by the ablation study in Section 3.3.2, where the full CGL-Net achieves the highest F1 score of 0.894.

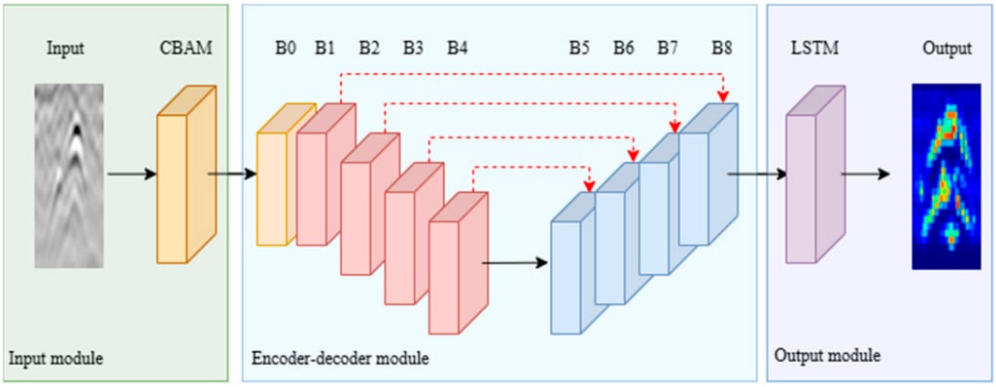


Figure 5. Schematic diagram of CGL-Net model.

Table 3. Parameter settings of CGL-Net.

Module Name	Layer Name	Block Name	Kernel	Stride	Input Size	Output Size
Input Module	INPUT	—	—	—	—	(16, 128, 32)
	CBAM-CA	Channel Attention (AvgPool + MaxPool + MLP)	—	—	(16, 128, 32)	(16, 128, 32)
	CBAM-SA	Spatial Attention (Conv + Sigmoid)	(7, 7)	(1, 1)	(16, 128, 32)	(16, 128, 32)

Table 3. *Cont.*

Module Name	Layer Name	Block Name	Kernel	Stride	Input Size	Output Size
Encoder–decoder Module	B0	GhostDoubleConv	(3, 3)	(1, 1)	(16, 128, 32)	(64, 128, 32)
	B1	MaxPool + GhostDoubleConv	(2, 2) (3, 3)	(2, 2) (1, 1)	(64, 128, 32)	(128, 64, 16)
	B2	MaxPool + GhostDoubleConv	(2, 2) (3, 3)	(2, 2) (1, 1)	(128, 64, 16)	(256, 32, 8)
	B3	MaxPool + GhostDoubleConv	(2, 2) (3, 3)	(2, 2) (1, 1)	(256, 32, 8)	(512, 16, 4)
	B4	MaxPool + GhostDoubleConv	(2, 2) (3, 3)	(2, 2) (1, 1)	(512, 16, 4)	(1024, 8, 2)
	B5	Upsample + GhostDoubleConv	—(3, 3)	—(1, 1)	(1024, 8, 2)	(512,16,4)
	B6	Upsample + GhostDoubleConv	—(3, 3)	—(1, 1)	(512, 16, 4)	(256, 32, 8)
	B7	Upsample + GhostDoubleConv	—(3, 3)	—(1, 1)	(256, 32, 8)	(128, 64, 16)
	B8	Upsample + GhostDoubleConv	—(3, 3)	—(1, 1)	(128, 64, 16)	(64, 128, 32)
Output Module	LSTM Reshape	Sequence Reshape	—	—	(64, 128, 32)	(32, 128, 64)
	LSTM (BiLSTM × 2)	Bidirectional LSTM	—	—	(32, 128, 64)	(32, 128, 64)
	LSTM FC1	Linear + ReLU + Dropout	—	—	(32, 128, 64)	(32, 128, 64)
	LSTM FC2	Linear	—	—	(32, 128, 64)	(32, 128, 64)
	LSTM Reshape^{-1}	Inverse Reshape	—	—	(32, 128, 64)	(64, 128, 32)
	LSTM Residual	Residual Connection	—	—	(64, 128, 32)	(64, 128, 32)
	OutConv	—	(1, 1)	(1, 1)	(64, 128, 32)	(1, 128, 32)

2.2.1. Input Module

To emphasize target-related channels and spatial positions, CBAM is introduced to adaptively re-weight the input features along the channel and spatial dimensions; its contribution to detection accuracy is quantified by the ablation study in Section 3.3.2. The 16-channel focused feature maps output by the multi-velocity F-K migration preprocessing are fed into the network at 128 × 32 pixels. CBAM [21] cascades channel attention and spatial attention to re-weight the input features along the channel and spatial dimensions, alleviating the target–background feature confusion caused by clutter (Figure 6); its contribution to detection accuracy is quantified by the ablation study in Section 3.3.2.

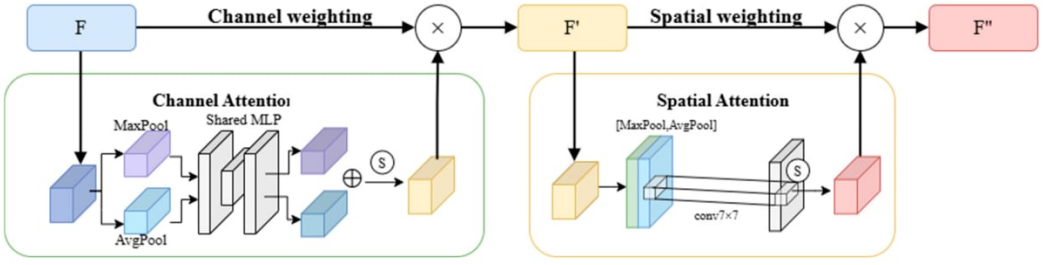


Figure 6. Structure of the CBAM module. Note: "⊗" denotes the channel-wise broadcast multiplication operation; the same notation applies below.

2.2.2. Encoder–Decoder Module

With its encoder–decoder structure and skip connections, UNet can effectively extract the multi-scale spatial features of root targets in GPR B-scan images. To reduce the computational cost of the model while maintaining the feature extraction capability, this study replaces the standard double-convolution blocks in UNet with Ghost double-convolution

blocks (GhostDoubleConv, ratio = 2), whose structure is shown in Figure 7. UNet extracts the multi-scale spatial features of root targets with its encoder–decoder structure and skip connections. To reduce the computational cost while preserving this capability, the standard double-convolution blocks are replaced by Ghost double-convolution blocks (ratio = 2) [22] (Figure 7); the layer-by-layer parameters are listed in Table 3.

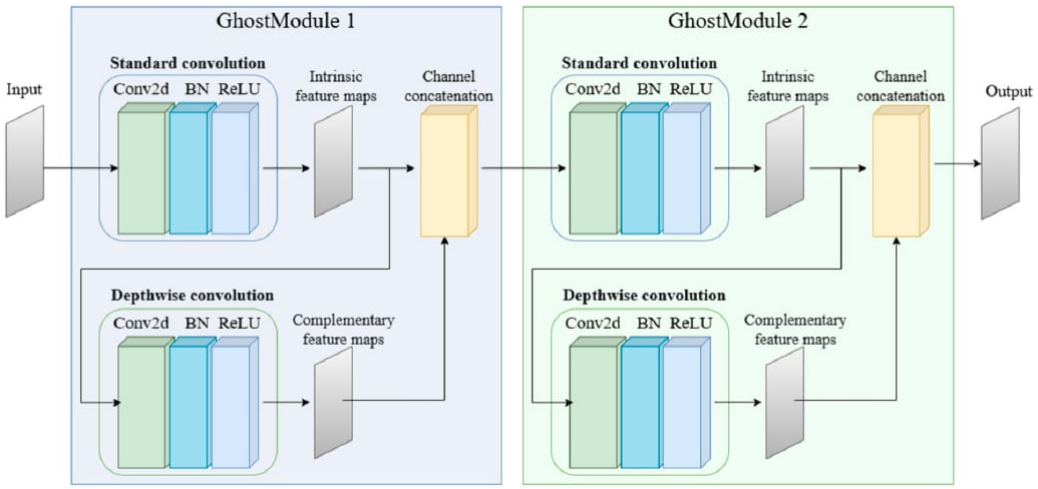


Figure 7. Schematic diagram of the Ghost double-convolution block.

2.2.3. Output Module

After spatial feature extraction, an LSTM temporal module [23] models contextual dependencies along the depth direction of each trace: the decoder features are unfolded, trace by trace, into sequences of depth steps, processed by a two-layer bidirectional LSTM with projection layers, and fused with the module input through a learnable weighted residual connection (Figure 8). A 1 × 1 convolution compresses the result into a single-channel response map of root presence, mapped to [0, 1] by a sigmoid activation, which is supervised against the target response maps and thresholded for root localization in Section 2.3; its contribution to deep weak-target detection is quantified by the ablation study in Section 3.3.2.

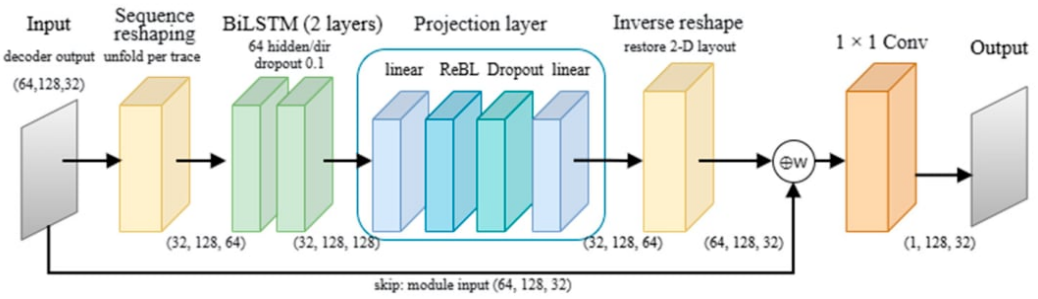


Figure 8. Structure of LSTM Module. Note: "⊕w" denotes a learnable weighted summation, $y = w \cdot x_{LSTM} + (1 - w) \cdot x_{in}$, where w is a learnable scalar initialized to 0.5.

2.3. Model Training

All models were implemented in PyTorch, and the training hardware platform was a workstation equipped with an Intel® Core™ Ultra 9 185H processor, 32 GB of RAM, and an NVIDIA GeForce RTX 4060 laptop GPU (mixed precision enabled). Training used a batch size of 1, the Adam optimizer (initial learning rate 0.001, weight decay 1×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$), and a maximum of 100 epochs; the learning rate was halved when

the validation loss stagnated for 5 epochs. The dataset was split at the B-scan sample level into training, validation and test subsets at a ratio of 70%:15%:15%. The 1000 purely simulated B-scans that served as sources for style transfer were distributed over the three subsets as 700, 150 and 150. Each translated image was assigned to the same subset as its source B-scan: translated images derived from training-subset sources were used for training, those derived from validation-subset sources for validation, and those derived from test-subset sources for testing. A scene and its translated version therefore never fall on opposite sides of the split, and the CycleGAN was trained only on training-subset images (Section 2.1.3). The held-out test subset was excluded from training, validation and model selection, and test scans were processed only by the frozen fine-tuned model during evaluation. For the measured data, the split was blocked by antenna frequency: all 60 field test images (44 B-scans of tree A and 16 of tree B, all at 1500 MHz) are held out, whereas the fine-tuning subset (280 scans) contains only 400 and 900 MHz scans of the same trees, including different-frequency acquisitions of the test lines. No scan of a test line at the test frequency therefore appears in training. Because the trees themselves are seen during fine-tuning at other frequencies, the field metrics do not measure generalization to unseen trees; they measure cross-frequency transfer to the 1500 MHz antenna on previously scanned trees, and the gap between tree-level and site-level generalization is discussed as a limitation in Section 4.4.

The training loss is a weighted combination of the mean squared error (MSE) and a structural-similarity-based loss [42,43]: the MSE term enforces pixel-wise numerical accuracy, while the SSIM-based term, computed following the standard SSIM formulation [42], preserves the structural integrity of the hyperbolic responses. Before loss computation, the network output is mapped to $[0, 1]$ by a sigmoid activation.

The experiments in this study are divided into two stages. In the first stage, the model is trained and tested on the annotated simulated data (purely simulated and CycleGAN style-transferred images), which have pixel-level root mask labels; detection performance is evaluated at the pixel level as defined in Section 2.6. In the second stage, the first-stage model is fine-tuned for 100 epochs on the measured scans of the training subset (280 scans) using thresholded pseudo-labels as supervision, with the same combined loss retained. Since pixel-level annotation of measured data is impractical, the burial depths obtained through root-by-root excavation serve as the ground truth for root-wise evaluation, and detections are matched one-to-one with excavated roots as defined in Section 2.6. The matching procedure was fixed prior to evaluation and was not tuned on the test set.

Pseudo-labels were derived from the post-processed prediction weight maps of the first-stage model using an adaptive, per-sample threshold $T = \mu + 1.5\sigma$, where μ and σ denote the mean and standard deviation of the weight map of that sample, followed by morphological closing and opening with a 3×3 kernel to remove isolated noise and fill small gaps. The same rule was applied to all subsets. Supervision during fine-tuning came exclusively from pseudo-labels on training-subset scans; validation and test subsets were never involved in fine-tuning in any form, and test scans were processed only by the frozen fine-tuned model during evaluation, so no test information entered training, fine-tuning or model selection.

2.4. Root Localization

On the basis of the weight localization maps output by CGL-Net, this study uses root localization as a post-processing step whose overall framework follows the Gradient Points Clustering-based Localization (GPCL) method of Hou et al. [25]: a three-stage pipeline of gradient point search, DBSCAN, and weighted centroid localization (Figure 9). Because the input here is not a binary or energy image but a continuous, weakly supervised weight

map whose amplitude, edge sharpness, and point density vary strongly from sample to sample, the three stages of [25] cannot be transferred as they are: a single global gradient threshold either keeps clutter or loses weak hyperbolic edges, and a fixed neighborhood radius either merges neighbouring roots or fragments a single root. Therefore, building on the GPCL framework and its three basic operators (gradient magnitude screening, density clustering, and weighted centroid estimation), the following four points are introduced: multi-directional gradient fusion with a morphological closing, an adaptive gradient threshold, adaptive DBSCAN parameters with two refinement steps that separate adjacent roots, and a multi-dimensional confidence score with rejection of low-confidence detections.

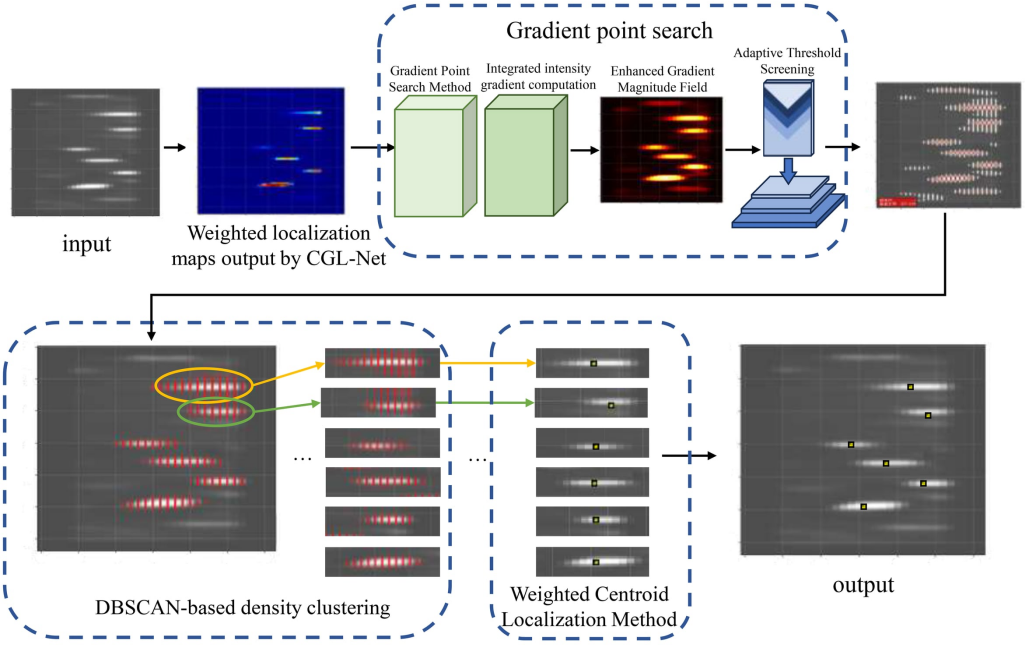


Figure 9. Schematic diagram of the GPCL pipeline. The weighted localization map predicted by CGL-Net is converted into root centres through three successive stages: gradient point search (upper dashed box), DBSCAN-based density clustering (lower left), and weighted centroid localization (lower middle); the resulting root positions are given in the output panel (lower right). The panels showing the input B-scan, the weighted localization maps and the intermediate gradient maps are schematic illustrations of the processing steps rather than quantitative results, and their coordinate tick labels are therefore omitted. Colours are used only to distinguish the different modules and carry no quantitative meaning. The arrows indicate the order in which the modules are processed within the pipeline; in particular, the yellow and green ellipses mark two adjacent clusters belonging to two neighbouring roots, and the matching yellow and green arrows indicate that these two clusters are separated and passed on individually, so that adjacent roots are not merged during clustering.

2.4.1. Gradient Point Extraction

In place of the single-directional gradient magnitude used in [25], this work fuses the gradients of four directions. To capture the hyperbolic edge information in full, the gradient components of the preprocessed weight map in the vertical, horizontal, and two diagonal directions are computed separately, with a higher weight assigned to the vertical direction to highlight the vertex features, constructing a comprehensive gradient intensity map:

$$G = \sqrt{\alpha_y G_y^2 + \alpha_x G_x^2 + \alpha_{d_1} G_d^2 + \alpha_{d_2} G_{d_2}^2}, \alpha_y > \alpha_x > \alpha_{d_1} = \alpha_{d_2} \quad (1)$$

where G denotes the fused comprehensive gradient intensity map; G_y, G_x, G_{d_1}, and G_{d_2} denote the gradient components in the vertical, horizontal, and two diagonal directions, respectively; and α_y, α_X, α_{d_1}, and α_{d_2} denote the weight coefficients of the corresponding directions.

Subsequently, a morphological closing operation with a 5 × 5 elliptical structuring element is applied to the comprehensive gradient intensity map to fill the small breaks along the hyperbolic edges, yielding an enhanced gradient intensity field with a more complete structure. Candidate gradient feature points are then extracted using an adaptive thresholding strategy, in which the gradient threshold is taken as the smallest of several statistics-based candidates and is subject to an absolute lower bound:

$$T_{\text{det}} = \max[\min(T_0, \mu_g + \alpha_1 \cdot \sigma_g, m_g + \alpha_2 \cdot \sigma_g, q_k, \alpha_3 \cdot \sigma_g, T_a), T_{\min}] \quad (2)$$

where T_{det} denotes the preset lower-bound threshold; T_0 denotes the preset base threshold, μ_g, σ_g, and m_g denote the mean, standard deviation, and median of the enhanced gradient intensity field, respectively; q_k is the k-th quantile (k = 70%); $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, and $\alpha_3 = 0.3$ are weighting coefficients; $T_a = 2.0$ is an absolute fallback threshold; and $T_{\min} = 0.2$ is the preset lower bound of the gradient intensity. In addition, a pixel is marked as a candidate gradient feature point only when $G(i,j) \geq T_{\text{det}}$ and the corresponding smoothed weight value exceeds the preset weight lower bound $\omega_{\min} = 0.4$. The feature point set P and the weight ω_k and gradient magnitude g_k associated with each point are thereby constructed.

2.4.2. DBSCAN Spatial Clustering

This study introduces the DBSCAN density clustering algorithm to spatially aggregate the gradient points and automatically partition the root targets, effectively distinguishing the spatially aggregated gradient points of target regions from randomly scattered clutter points. On the basis of the candidate gradient points, and to address the limitation that fixed parameters can hardly adapt to different density distributions, an adaptive parameter selection strategy is proposed. The neighborhood radius ξ is dynamically determined based on the median of the distances between all pairs of feature points:

$$\xi = \max\{\xi_{\min}, \beta \cdot median[d(p_a, p_b), p_a \in P, p_b \in P, a < b]\} \quad (3)$$

where $\beta = 1.5$ is the scaling factor, $d(p_a, p_b)$ denotes the Euclidean distance, the median is computed over all $N(N-1)/2$ pairwise distances of the feature point set P, and $\xi_{\min} = 5$ pixels is the minimum radius.

The minimum number of neighborhood points is set according to the total number of feature points:

$$N_{\min} = \max\left(3, \left\lfloor \frac{|P|}{\delta} \right\rfloor\right) \quad (4)$$

where $|P|$ is the total number of feature points and $\delta = 20$ is the adaptive divisor. After clustering, two refinement steps are applied to separate adjacent roots that may be merged by density-based clustering: clusters are split at horizontal gaps larger than 3 pixels, and a cluster whose mean intra-cluster pairwise distance exceeds ξ is re-partitioned by DBSCAN with a reduced radius of 0.8ξ.

2.4.3. Weighted Centroid Localization

This study adopts the weighted centroid method to estimate the centre of each root response, assigning different contributions according to the weight ω_k of each feature point:

$$x_c = \frac{\sum_{k \in C_c} \omega_k \cdot j_k}{\sum_{k \in C_c} \omega_k}, z_c = \frac{\sum_{k \in C_c} \omega_k \cdot i_k}{\sum_{k \in C_c} \omega_k} \quad (5)$$

where $p_k = (i_k, j_k)$ is the k-th feature point in the cluster, with i_k and j_k being its depth and horizontal coordinates, respectively; ω_k is the weight value of this point; and (x_c, z_c) is the centre coordinate of the root response estimated by the c-th cluster, with x_c the horizontal and z_c the depth position.

To evaluate the reliability of each detection result, a multi-dimensional confidence score is introduced, comprehensively considering the cluster size, the mean and consistency of the weights, the average gradient intensity, and the spatial compactness:

$$\varphi_c = 0.3 \cdot S_{size} + 0.25 \cdot S_{\omega} + 0.15 \cdot S_{cons} + 0.2 \cdot S_{grad} + 0.1 \cdot S_{comp} \quad (6)$$

where $S_{size} = \min(|C_c|/20,1)$ is the cluster size score; $S_{\omega} = \min(\overline{\omega}_c/0.8,1)$ is the mean-weight score, with $\overline{\omega}_c$ the mean weight within C_c; $S_{cons} = 1 - \min(\sigma_{\omega}, \overline{\omega}_c, 1)$ is the weight consistency score based on the coefficient of variation in the weights; $S_{grsd} = \min(\overline{g}_c/20,1)$ is the mean gradient intensity score; and $S_{comp} = \max(0, 1 - \overline{d}_c/30)$ is the spatial compactness score, with $\overline{d}_c$ the mean pairwise distance (in pixels) within C_c. All component scores are normalized to $[0, 1]$, and φ_c is clipped to $[0.1, 1.0]$. Detections with confidence below the threshold $\tau = 0.2$ are discarded as low-confidence detections.

2.5. Root Multi-Parameter Estimation

Based on the weight localization maps generated by CGL-Net, this study develops a procedure for estimating root morphological parameters. The algorithm adopts a multi-level feature extraction strategy to precisely resolve the geometric morphology and physical parameters of roots from the weight distribution patterns.

2.5.1. Diameter Estimation

Owing to the dielectric contrast at the root–soil interface, energy is reflected at both the upper and the lower root boundary, and the resulting echoes are assigned to the same cluster during clustering (Figure 10a). At the 1.5 GHz center frequency, the wavelength in soil is comparable to or larger than the root diameters considered, so the two boundary reflections are generally not resolved as separate pulses but superpose into a composite wave packet whose total duration increases with the root diameter, consistent with Cui et al. [44]. The feature adopted here is therefore the total vertical extent of the reflected wave packet: the vertical span of the gradient feature points within a cluster (Figure 10b), computed for the c-th cluster as:

$$\Delta z_c = \left(\max_{k \in C_c} i_k - \min_{k \in C_c} i_k\right) \cdot \rho_z \quad (7)$$

where i_k is the depth-direction (row) coordinate of the *k-th* feature point in pixels, and ρ_z is the vertical resolution converting the span into physical units. The row index corresponds to the two-way travel time of the recorded time axis, which spans 16 ns over 128 samples ($\Delta t = 16/127 \approx 0.126$ ns per row); a fixed time-zero offset of 0.2048 ns (≈ 0.013 m in depth) is subtracted before conversion, and the depth is obtained by $z = -(t_k - t_0) \cdot v/2$ with the wave velocity v = 0.1272 m/ns. The soil relative permittivity was measured at 12 locations, one sample per quadrant of the x–y grid around each of the three scanned trees: the volumetric water content of each extracted soil core was determined by oven-drying and weighing, and converted to εr with the Topp relation [45]. The values ranged from 5.0 to 6.0 with a mean of 5.56; since the three trees stand close together on the same soil unit, the spatial variability is limited, and the corresponding velocity deviation of approximately ±3–6% about $v = 0.1272$ m/ns propagates proportionally into the depth estimates and is discussed in Section 4.4. Accordingly, for $N_z = 128$ depth samples, and the estimated root depth corresponds to the weighted centroid of the cluster, i.e., the center of the response, which is consistent with the root center recorded in the excavation reference. Because the cluster span is also affected by factors such as hyperbola overlap and the strength of the dielectric contrast, Δz_c is not directly equated with the root diameter; instead, Δz_c (in mm) serves as the regressor of a linear calibration model between Δz_c and the root diameter, which is fitted on the simulated dataset and independently validated on the measured data. The calibration is established

within 16–80 mm, the diameter range of the simulated corpus; diameters outside this range are estimated by extrapolation and are discussed as limitations in Sections 4.2 and 4.4.

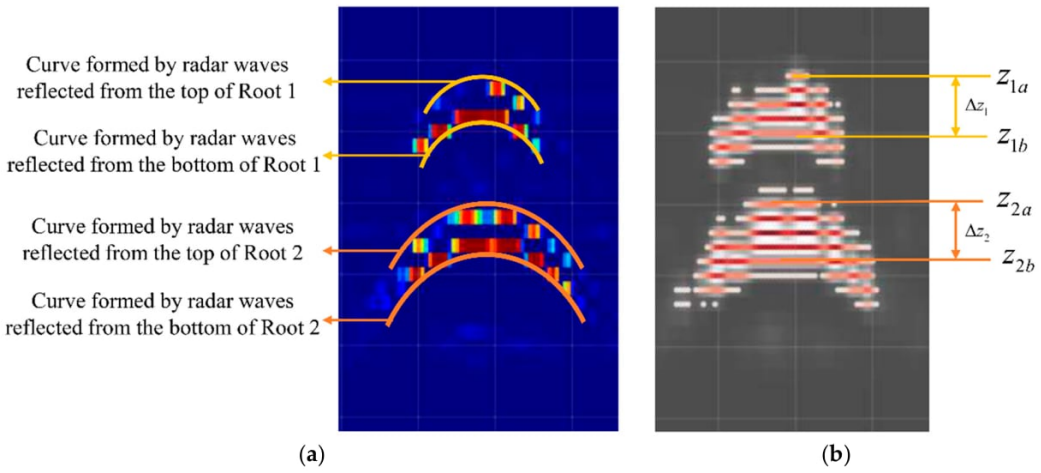


Figure 10. Schematic of the diameter-related feature in GPR B-scans: (**a**) hyperbolic echoes from the upper and lower root boundaries assigned to one cluster; (**b**) vertical span of the gradient feature points within the cluster (the feature Δz_c). The figure is schematic and does not represent measured signal evidence.

2.5.2. Growth Angle Estimation

The orientation of a local root segment is described by the elevation angle (El) and the horizontal azimuth angle (Az). A root segment is observed by pairing the detections of the same root on two adjacent survey lines: for each pair of adjacent lines, the detections in the in-line (x, z) plane are matched by nearest neighbor with a distance tolerance of 0.2 m under one-to-one assignment. Each matched pair yields two points and $p_2 = (x_2,y_2,z_2)$ in metric coordinates, where x is the in-line horizontal position, y is the survey-line position, and z is the depth obtained in Section 2.5.1 (z negative downward). The segment direction is the vector $d = p_2 - p_1$, and the angles are computed as:

$$Az = \arctan2(d_y,d_x) \in [0°,360°) \tag{8}$$

$$El = \arctan2(d_z, \sqrt{d_x^2 + d_y^2}) \in [-90°, 90°] \tag{9}$$

where the z axis points upward, so that a segment descending away from the trunk has $d_z < 0$ and a negative elevation angle. Because the ordering of the two survey lines (from line i to the adjacent line $i + 1$) imposes an arbitrary sign on d, the estimated axis is undirected: the same root axis is equally described by (Az,El) and $(Az + 180°,-El)$. Accordingly, for a predicted segment and its corresponding excavated root segment, with predicted angles (Az_p,El_p) and ground-truth angles, the azimuth and elevation errors used in the angle evaluation (Section 3.1.2) are computed as $e_{Az} = \min(|Az_p - Az_g|,360°-|Az_p - Az_g|)$ and $e_{El} = \min(|El_p - El_g|,|El_p + El_g|)$ where the two(Az_g,El_g) angles are always taken from the same representation of the ground-truth axis, i.e., either (Az_g,El_g) or its equivalent $(Az_g + 180°,-El_g)$, and the representation that minimizes the combined error is used. Each predicted segment is compared with the ground-truth angles of its corresponding excavated root segment.

The geometric construction above relies on two assumptions: the diffraction apex on each survey line marks the point where the root crosses that line, and the root axis is approximately straight over the 15 cm baseline between adjacent lines. The former is largely insensitive to the root–line crossing angle, which alters the amplitude and shape of the recorded diffraction but not the apex position; it is the detectability of the root, rather

than the angle computation, that the crossing angle affects. Root curvature and locally variable growth directions are not captured by this straight-segment approximation; the ground-truth angles, however, are reconstructed from the adjacent excavation profiles with the same geometric definitions (Section 2.1.2), so that the predicted and measured angles refer to the same segment representation. The propagation of soil velocity heterogeneity and surface roughness into the depth estimates—and hence into the angles—is analyzed in Section 4.4.

2.6. Evaluation Metrics

To verify the effectiveness of CGL-Net, the model is evaluated from three dimensions: model complexity and inference efficiency, root detection, and geometric parameter accuracy. Model complexity is characterized by the number of parameters (Params), the *FLOPs* of a single forward pass on a 16 × 128 × 32 input, and the frames per second (*FPS*); the training time is reported as an auxiliary measure of the training overhead. The measurement conditions are given in the note of the corresponding table in Section 3.3.2, and these metrics are used in the ablation experiments to compare the efficiency of standard U-Net, Ghost-UNet, and the various CGL-Net configurations.

Detection performance is reported at two levels, consistent with the two-stage data protocol of Section 2.3: pixel-level P, R and F1-score computed on the root masks (Sections 3.3.2, 3.3.3 and 4.1), and instance-level P, R and F1-score computed on matched root instances for all three datasets (Section 3.1.1; parameter errors are reported in Section 3.1.2), where a detection counts as a true positive only when it is matched to exactly one excavated root by one-to-one nearest-depth matching within the same survey profile with a depth tolerance of 0.1 m, with $P = TP/(TP + FP)$, $R = TP/(TP + FN)$, and $F_1 = 2TP/(2TP + FP + FN)$. The notes of all tables state the counting basis (root-mask pixels, root instances, or matched root instances) and the decision rule (per-sample adaptive thresholding in GPCL for all reported results, except the loss-function ablation in Section 4.3, which is evaluated by image-similarity metrics rather than by detection scores). All detection results in Section 3 follow this common rule and are not re-explained thereafter.

Geometric accuracy is evaluated by the absolute error (AE), relative error (RE), and correct rate (CR) for depth, diameter, and angles, computed only over successfully matched roots: $AE_i = |\hat{y}_i - y_i|$, $RE_i = |\hat{y}_i - y_i|/|y_i|$, and $CR = \frac{1}{N}\sum_{i=1}^{N} I_A(e_i < T) \times 100\%$, where $\hat{y}_i$ and y_i are the estimated and ground-truth values, e_i denotes the relative error for depth and diameter and the absolute error for angles, $I(\cdot)$ is the indicator function, and T is the pre-set tolerance; the coefficient of determination $R^2 = 1 - \frac{\sum_{i=1}^{N}(d_i - \hat{d}_i)^2}{\sum_{i=1}^{N}(d_i - \bar{d})^2}$ and $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \hat{d}_i)^2}$ are used for the diameter hypothesis validation, with $\bar{y}$ the mean of the true diameters. The sample size N differs among the tables: for depth and diameter (Sections 3.1.2 and 3.2) it is the number of matched detection–excavation pairs; for the angle metrics (Section 3.1.2) it is the number of root-segment observations formed by adjacent survey-line pairs, which are not independent roots. The tolerances were set with reference to the accuracy reported for GPR-based root parameter estimation—diameter errors of 13–16% [46] and an overall mean relative error within 10.57% [12]—slightly above these levels so that CR reflects the estimation quality rather than the measurement uncertainty of the excavation reference; the angular tolerance was set to 5.0°.

Besides the tolerance-based CR, distribution-level statistics are reported for burial depth, diameter, and orientation angles: *Median* and interquartile range (IQR) of absolute relative errors, with 95% bootstrap confidence intervals (CIs) of the median. For simulated and style-transferred data (statistically independent instances), CIs use an ordinary bootstrap over instances ($B = 10{,}000$). For measured data, where paired observations may

share an excavated root and are thus non-independent, a root-level cluster bootstrap is used ($B = 10{,}000$). *CR* remains a tolerance-based usability metric and is reported alongside, not instead of, these distribution-level statistics.

2.7. Ablation Experiment Design

To isolate the contribution of the input representation, the first ablation fixes CGL-Net, GPCL, data splits, optimization settings, and input tensor shape ($16 \times 128 \times 32$), and compares three configurations: P0 (15-velocity Stolt F-K stack + mean channel), P2 (single site velocity 0.1272 m/ns replicated across 16 channels), and P1 (pre-processed B-scan without migration replicated across 16 channels). Since the channel dimension is fixed at 16, differences reflect channel content rather than network capacity. Results are reported in Section 3.3.1.

To clarify the independent contributions of each component of CGL-Net and of the GPCL algorithm, a systematic ablation study was designed, and all ablation and comparison experiments were conducted on the same annotated simulated dataset (purely simulated and style-transferred data). The second ablation configures five models—standard U-Net as the baseline, Ghost-UNet, Ghost-UNet + CBAM, Ghost-UNet + LSTM, and the complete CGL-Net—with the same post-processing configuration applied to all five (Section 3.3.2).

The third ablation quantifies the contribution of each post-processing step by applying three strategies to the same weight localization maps produced by the complete CGL-Net, sharing the identical input, network weights, test set and evaluation protocol and differing only in the operations that convert the weight map into root centre coordinates: threshold segmentation only (baseline), gradient-point search only (control), and the complete GPCL (gradient-point search with adaptive DBSCAN clustering and confidence filtering, Equations (1)–(6)) (Section 3.3.3). The full operating details of the three strategies are given in the note of the corresponding table.

3. Results

3.1. Root Detection Results and Multi-Parameter Estimation

The complete CGL-Net + GPCL pipeline was applied to the purely simulated, style-transferred, and field-measured B-scans. Its detection performance on the three datasets is reported in Section 3.1.1 (Table 4), where representative detection results of the field data are shown together with the number of roots identified in each B-scan (Figure 11). The accuracy of the estimated depth, diameter, and angles of the matched roots is then validated against the excavation measurements in Section 3.1.2, the root-by-root agreement is examined in Section 3.2, the individual contributions of the network modules and the GPCL strategy are quantified in Section 3.3, and the comparison with existing methods is given in Section 4.1.

3.1.1. Root Detection Results

All three datasets were evaluated with the same matching rule and the same post-processing setting, so that the differences among rows reflect the domain gap between training and evaluation rather than a change in the evaluation protocol. Table 4 lists the instance-level precision, recall, and F1-score obtained on each dataset, and Figure 11 shows representative detection results of the three datasets together with the roots identified in each B-scan. Precision, recall, and F1-score decrease from the purely simulated data to the style-transferred data and further to the measured data, which is consistent with the progressively larger mismatch between the simulated source domain and the field acquisition conditions. The style-transferred data occupy an intermediate position, as expected, since they retain the geometric labels of the simulated data while inheriting

the background texture, clutter, and signal attenuation of the measured B-scans. On the measured data, the pipeline reaches a precision of 0.816 and a recall of 0.792 within the detectable depth range. Compared with the simulated results, the gap of 0.102 in F1-score on the measured set (0.906 versus 0.804) quantifies the residual cost of transferring the pipeline from simulation to the field, and this gap is the quantity that the style transfer in Section 2.1.3 is designed to reduce. In terms of the speed–accuracy trade-off, the attention and temporal modules of CGL-Net are sequential along the depth axis and therefore reduce the inference throughput: 126.3 *FPS* for CGL-Net against 190.9 *FPS* for standard U-Net and 210.5 *FPS* for the Ghost-UNet backbone, i.e., a 33.8% reduction relative to U-Net. This reduction is the deliberate price paid for the 0.081 improvement in *F1-score*, the 44.8% reduction in parameters, and the 22.6% reduction in *FLOPs*.

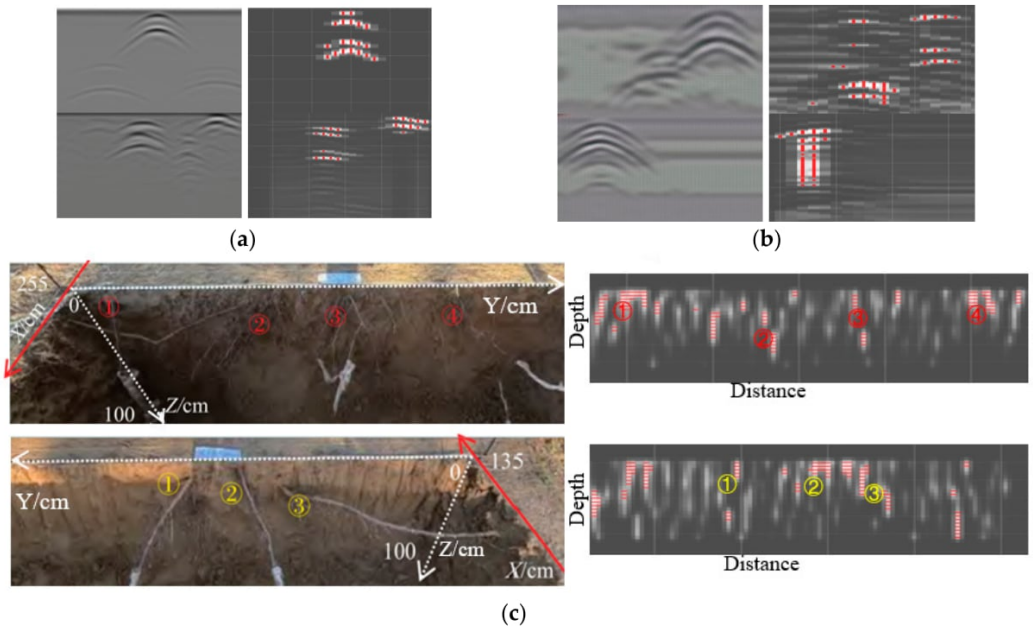


Figure 11. Detection results of the complete CGL-Net + GPCL pipeline on representative B-scans of the three datasets: (**a**) purely simulated data; (**b**) style-transferred data; (**c**) measured data. For (**a**) and (**b**), the left panel is the input B-scan and the right panel is the localization result of GPCL. For (**c**), the panel shows the acquisition results of survey lines X = 255 and X = 135 in the field measurements.

Table 4. Instance-level detection performance of CGL-Net + GPCL on the three datasets.

Dataset	B-Scans	GT Instances	*TP*	*FP*	*FN*	*P*	*R*	*F1-Score*
Purely simulated data	240	782	705	69	77	0.911	0.902	0.906
Style-transferred data	150	468	405	57	63	0.877	0.865	0.871
Measured data	60	335	265	60	70	0.816	0.792	0.804
All data (fine-tuned model)	450	1585	1375	186	210	0.881	0.868	0.874

The residual errors in Table 4 are further examined by source. On the measured data, the model produces 60 false positives and 70 false negatives over 60 B-scans (Table 4), i.e., close to one false positive per B-scan. Most false positives correspond to isolated non-root reflectors—stones and possible fragments of buried pipes—whose diffraction hyperbolas resemble those of coarse roots in a single B-scan; the spatial-context rules of GPCL (Section 2.4) remove part of them, and the remainder sets the practical upper bound of about one false positive per B-scan implied by the field precision (P = 0.816). Most false negatives are roots thinner than the vertical resolution of the 900 MHz antenna (below approximately 5 mm) or buried near the lower boundary of the excavated range, for which

the hyperbola amplitude is close to the background level. These two sources define the current operating envelope of the method and motivate the discussion in Section 4.4.

3.1.2. Accuracy of Root Depth, Diameter, and Angle Estimation

The statistical results of the root burial depth and diameter estimation errors and accuracies are shown in Table 5 and Figure 12. The depth and diameter relative errors are reported together with the corresponding accuracies for each dataset, evaluated with the pre-set tolerances of 30% for depth and 20% for diameter defined in Section 2.6 (*CR*, the proportion of matched root instances whose non-strict relative error does not exceed the tolerance). The relative errors of depth and diameter increase from the simulated data to the style-transferred data and to the measured data, with the corresponding accuracies decreasing. The metrics of the style-transferred data lie between those of the simulated and measured data. The last row reports the fine-tuned model evaluated on the three data sources pooled together namely 450 evaluated B-scans and the matched root instances detected in them, for which the depth and diameter accuracies are 87.2% and 91.7%; these values are computed on the pooled instance set.

Table 5. Root Depth and Diameter Estimation Accuracy Across Different Datasets.

Dataset	Depth			Diameter		
	***AE* (cm)**	***RE* (%)**	***CR* (%)**	***AE* (cm)**	***RE* (%)**	***CR* (%)**
Purely simulated data [1]	3.8	8.8	91.2	0.3	4.5	95.5
Style-transferred data [1]	2.7	18.7	81.4	0.7	9.8	86.2
Measured data [2]	3.0	24.0	76.0	1.5	18.8	78.5
All data (fine-tuned model) [3]	3.4	13.1	87.2	0.5	7.4	91.7

[1] The model was supervised on the simulated dataset and evaluated separately on the simulated test set (240 B-scans) and on the style-transferred test subset (150 B-scans). [2] The model was fine-tuned with the measured data and evaluated on the measured set, which comprises 60 B-scans, namely the 44 survey lines of the first tree and the 16 survey lines of the second tree, all of which were excluded from fine-tuning. All metrics in this table are computed over matched detection–excavation instances, that is, one value per matched root instance. [3] The fine-tuned model was evaluated over the three sources pooled together, that is, 450 evaluated B-scans (240 purely simulated, 150 style-transferred, and 60 measured) together with the matched root instances detected in them; errors of all matched root instances were pooled and metrics were computed once on the pooled set.

Figure 12 compares the estimated and ground-truth depth and diameter of the 1375 matched root instances. For depth (Figure 12a), the markers of all three datasets follow the 1:1 line closely over the whole excavated range, with no systematic drift toward over- or under-estimation, which shows that the residual errors summarized in Table 5 are random rather than depth-dependent. The diameter estimates (Figure 12b) are similarly centred on the 1:1 line but are visibly more dispersed, and the scatter widens towards larger diameters; this is consistent with the diameter being inferred from the depth span of the clustered response rather than measured directly, so that the absolute uncertainty grows with the root size while the relative error stays within the stated tolerance. Among the three datasets, the purely simulated group (grey) shows a trend of smaller dispersion in both panels, while the style-transferred group (blue) is closest to the measured group (orange) in both panels, with performance slightly worse than that of the purely simulated group.

The dispersion pattern in Figure 12b is examined against root size. The relative error of the diameter estimate decreases markedly with increasing true diameter: for roots above 60 mm the diameter REs are concentrated below 15%, whereas roots below 30 mm dominate the upper tail of the distribution, consistent with the peak-to-peak separation of the upper and lower boundary reflections approaching the resolution limit of the antenna. The depth

errors, in contrast, show no comparable size dependence (Figure 12a), supporting the interpretation that depth accuracy is governed by the velocity calibration rather than by root geometry. This size-dependent behavior is quantified for nine excavated roots in Section 3.2.

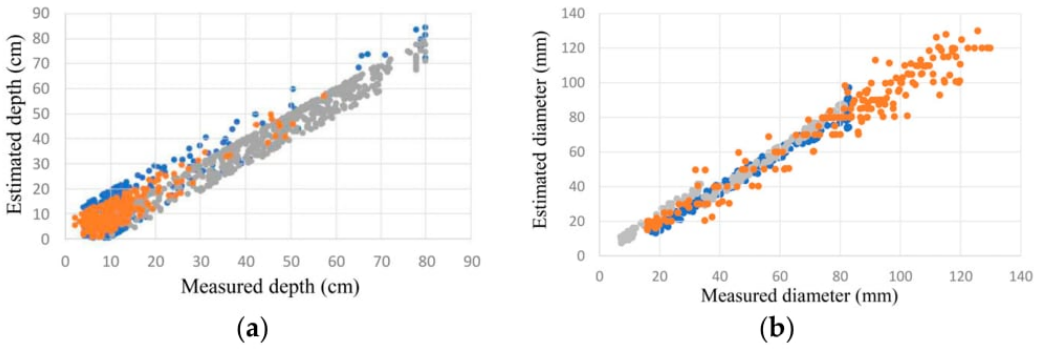


Figure 12. Scatter plots of the estimated versus ground-truth root depth (**a**) and root diameter (**b**) for the matched instances. Orange circles denote the measured roots ($n = 265$), grey circles the purely simulated roots ($n = 705$), and blue circles the style-transferred roots ($n = 405$); the number of markers in each group equals the number of true positives of the corresponding dataset in Table 4. The dashed line is the 1:1 reference, so a marker lying on it indicates exact agreement. Only detections matched one-to-one to a ground-truth root within ±0.1 m in depth (Section 2.3) are retained; unmatched roots and unmatched detections are not shown.

To provide the distribution-level view that the tolerance-based CR cannot capture, the depth and diameter error distributions of the three data sources are summarized separately in Table 6. The median absolute relative error increases monotonically from the purely simulated data (6.1% depth, 4.4% diameter) to the style-transferred data (15.8%, 9.2%) and further to the measured data (22.2%, 22.7%), and the interquartile ranges widen in the same order (14.6% and 18.3% on the measured data), consistent with the progressively larger domain gap quantified in Section 3.1.1. The 95% bootstrap CIs of the medians follow the same pattern: the simulated and style-transferred intervals are narrow (e.g., 4.9–6.3% for simulated depth), whereas the measured-data intervals are substantially wider (15.2–28.5% for depth, 15.9–27.4% for diameter), reflecting both the smaller field sample and the clustering of observations within roots, which is accounted for by the root-level cluster bootstrap described in Section 2.6. The CR values in Table 5 therefore characterize the practical usability of the estimates against the pre-set tolerances, while Table 6 provides the complementary distribution-level uncertainty quantification.

Table 6. Distribution-level error statistics per data source. *Median*, *IQR*, and *95%* CI of the mean relative error (%).

Dataset	Depth			Diameter		
	***Median* (%)**	***IQR* (%)**	***95% CI* (%)**	***Median* (%)**	***IQR* (%)**	***95% CI* (%)**
Purely simulated data [1]	6.1	2.8	4.9–6.3	4.4	3.9	3.1–4.7
Style-transferred data [1]	15.8	5.1	14.4–17.2	9.2	7.5	7.6–11.8
Measured data [2]	22.2	14.6	15.2–28.5	22.7	18.3	15.9–27.4

[1] The simulated and style-transferred instances are statistically independent (one simulated root per instance), and their CIs are computed by an ordinary bootstrap over instances (B = 10,000). [2] The measured instances are clustered within excavated roots, so their CI is computed by a root-level cluster bootstrap: the 22 excavated roots are resampled with replacement, all paired observations of the selected roots are retained, and the median is recomputed in each of B = 10,000 replicates. All errors are absolute relative errors (%), computed with the same matching rule and tolerances as in Table 5.

To further evaluate the estimation of the root spatial orientation, the estimation errors of the elevation angle (El) and the horizontal azimuth (Az) were computed on the measured data, with the results shown in Table 7. For El the mean absolute error is 2.0°, the mean relative error is 27.7%, and the accuracy within the 5.0° tolerance (*CR*) is 92.1%; for Az the corresponding values are 4.4°, 7.3%, and 72.4%. The comparatively large relative error of El stems from observations with small true elevation angles, for which the same absolute deviation corresponds to a large relative error. The angle estimates are derived from the same root observed on adjacent survey lines: the statistics are computed over 76 paired observations obtained from 22 independent roots, and the total number of pairings, the number of independent roots, and the treatment of repeated observations of the same root are specified in the note of Table 6 and Figure 13.

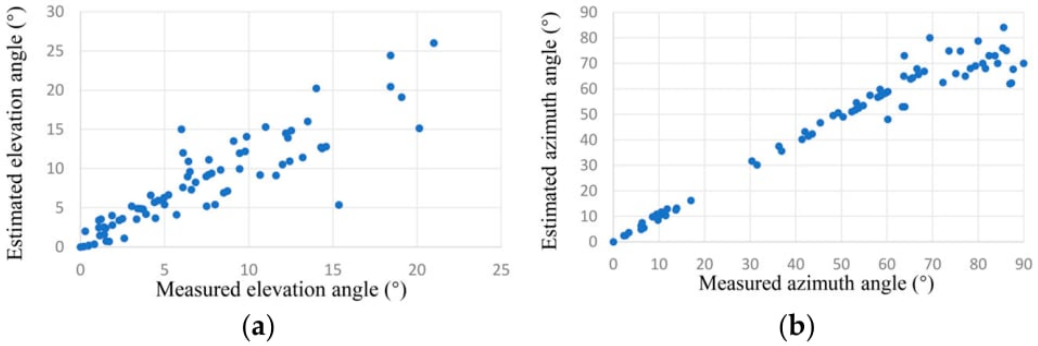


Figure 13. Scatter plots of estimated versus excavation-measured root orientation for the measured roots: (**a**) elevation angle; (**b**) azimuth angle. Only the measured dataset is shown, because the orientation of a root can be recovered only when the same root is observed on adjacent survey lines and its reference value is obtained by excavation. The 76 paired observations obtained from 22 independent roots (Table 7) are plotted, with one marker per observation. The dashed line is the 1:1 reference. Because the elevation and azimuth angles of adjacent survey lines are defined on the closed interval of −90° to 90°, panels (**a**) and (**b**) record the absolute values of the elevation and azimuth angles.

Table 7. Root Angle Estimation Accuracy on Measured Data.

Angle Type [1]	*AE* (°)	*RE* (%)	*CR* (%)
El	2.0	27.7	92.1
Az	4.4	7.3	72.4

[1] The angle estimates are derived from the same root observed on adjacent survey lines, so one root may contribute several paired observations. The statistics are computed over $N = 76$ paired observations obtained from 22 independent roots, and each observation contributes one value. The evaluated set includes the nine roots listed in Section 3.2 as well as additional roots, so the two tables are not independent. Errors are computed on the absolute angular deviation $|\Delta\theta|$, and the relative error of an angle is normalized by the absolute value of the true angle of that observation. The tolerance is 5.0° for both El and Az.

The corresponding distribution-level statistics are reported in Table 8: the median relative error is 23.4% for El (*IQR* 5.8%, 95% cluster-bootstrap *CI* 17.2–27.0%) and 6.4% for Az (*IQR* 7.3%, *CI* 2.5–12.6%). The median of El is larger than that of Az even though its *CR* is higher, because the relative error of El is inflated by observations with small true elevation angles, whereas its absolute deviation remains small; the wide *CI* of Az reflects both the right-skewed distribution of the azimuth errors and the limited number of independent roots, and is consistent with the azimuth being the weaker of the two orientation components.

Table 8. Distribution-level relative error statistics of the estimated root orientation angles on the measured data: *Median*, interquartile range (*IQR*), and 95% bootstrap confidence interval (*CI*) of the median.

Angle Type [1]	*Median* (%)	*IQR* (%)	*95% CI* (%)
El	23.4	5.8	17.2–27.0
Az	6.4	7.3	2.5–12.6

[1] Statistics are computed over the 76 paired observations from 22 independent roots (Table 7), with the relative error of each observation normalized by the true angle of that observation, as in Table 7. Because the observations are clustered within roots, the *CI* is computed by a root-level cluster bootstrap: the 22 roots are resampled with replacement, all paired observations of the selected roots are retained, and the median is recomputed in each of $B = 10{,}000$ replicates. The comparatively large relative errors of El stem from observations with small true elevation angles, for which the same absolute deviation corresponds to a large relative error.

Combining the results across the three dimensions of burial depth, diameter, and angle, CGL-Net maintains reliable estimation accuracy on the measured data and can provide effective support for non-destructive detection of tree roots and synchronous estimation of multiple parameters.

3.2. Case Study: Root-by-Root Comparison with Excavation

To further verify the reliability of the model in estimating individual roots, the GPR inversion results were compared item by item with the true root data obtained after excavation. As a case study, the nine roots of one representative P. tomentosa individual are listed individually in Table 9, which reports both the excavated absolute values and the GPR inversion errors of the burial depth, diameter, and growth angles. Because each root is recorded only where it crosses the 15 cm-spaced excavation profiles, the depth and diameter errors refer to the profile section on which the root was excavated, and the angle errors refer to the root segment spanning two adjacent profiles; the table therefore characterizes the root by the geometric parameters observable on the excavation record, while the corresponding GPR evidence (the migrated B-scans and their GPCL interpretation) is shown for representative lines in Figure 11c and for the full matched sample in Figure 12.

Table 9. Comparison of GPR inversion results and field excavation measurements [1].

Root No.	Depth			Diameter			El			Az		
	Measured (cm)	*AE* (cm)	*RE* (%)	Measured (cm)	*AE* (cm)	*RE* (%)	Measured (°)	*AE* (°)	*RE* (%)	Measured (°)	*AE* (°)	*RE* (%)
1	12.4	3.2	25.7	8.0	2.2	27.5	3.4	1.1	31.7	67.4	7.6	11.3
2	15.7	2.8	18.0	5.3	2.0	37.3	3.2	1.5	46.3	15.9	2.1	13.2
3	17.1	3.6	20.8	1.2	0.1	5.9	8.5	1.0	11.3	20.6	4.0	19.5
4	16.9	3.8	22.2	7.0	0.6	9.1	26.3	5.1	19.3	12.7	1.0	8.1
5	17.9	5.7	32.0	6.3	0.1	1.3	26.7	1.0	3.8	65.3	6.4	9.7
6	12.1	3.2	26.1	2.1	0.5	26.0	9.5	3.6	38.1	20.2	5.4	26.9
7	20.9	3.3	16.0	2.2	0.3	14.4	31.7	0.9	2.7	80.0	3.1	3.9
8	15.1	1.5	10.0	7.9	0.2	1.9	3.8	1.6	42.0	15.0	0.7	4.3
9	10.3	2.7	26.3	2.2	0.3	14.6	8.0	1.6	19.7	19.2	3.4	17.8
Mean	15.4	3.3	21.9	4.7	0.7	15.3	13.5	1.9	23.9	35.1	3.7	12.7

[1] Each row corresponds to one excavated root: the depth and diameter errors are computed from the inversion result on a single B-scan section, and the elevation (El) and azimuth (Az) errors are computed from a root segment observed on adjacent survey lines. The nine roots listed here are a subset of the matched instances reported in Table 5, selected to illustrate root-by-root agreement rather than as an independent sample; the aggregate statistics over all matched instances are given in Table 5. *AE* is the absolute error of that root, *RE* is its relative error with respect to the field-measured value, the mean row gives the mean of the nine values in the corresponding column; the tolerance is 30% for depth, 20% for diameter, and 5.0° for both angles.

The comparison results show that the proposed method achieves centimeter-level accuracy in root depth inversion: the mean absolute error of the burial depth for the nine roots is 3.3 cm. Among them, root 8 has the smallest burial depth error, with an absolute error of 1.5 cm and a relative error of 10.0%, and root 5 the largest, with an absolute error of 5.7 cm and a relative error of 32.0%. In terms of diameter estimation, the nine roots have a mean absolute error of 0.7 cm, a mean relative error of 15.3%; roots 5 and 8 have the smallest relative errors, 1.3% and 1.9%, respectively, root 2 the largest relative error, 37.3%, and root 1 the largest absolute error, 2.2 cm. In terms of angles, the elevation angle is the best for root 7 (an *AE* of 0.9° and an *RE* of 2.7%) and the worst for root 4 (an *AE* of 5.1° and an *RE* of 19.3%); the azimuth is the best for root 8 (an *AE* of 0.7° and an *RE* of 4.3%) and the worst for root 6 (an *AE* of 5.4° and an *RE* of 26.9%), with mean absolute errors of 1.9° (El) and 3.7° (Az). Across the nine roots, the inverted diameters are positively correlated with the true values, with a coefficient of determination of 0.621. The above field validation results indicate that CGL-Net supports the non-destructive detection and quantitative estimation of tree roots under field-measured conditions.

3.3. Ablation Experiments

3.3.1. Ablation of the Input Representation

To further examine the role of the input representation, CGL-Net was retrained and re-evaluated under the three input configurations defined in Section 2.7, with the network, GPCL post-processing, data splits and optimization settings held fixed; the results are listed in Table 10 and the corresponding localization maps are shown in Figure 14. The complete multi-velocity representation (P0) reaches a precision of 0.901, a recall of 0.888 and an F1-score of 0.894. Collapsing the stack to the single site velocity of 0.1272 m/ns (P2) lowers the F1-score to 0.882 ($P = 0.894$, $R = 0.870$), and removing the migration altogether (P1) lowers it further to 0.793 ($P = 0.809$, $R = 0.778$). The degradation is monotonic in both steps, which separates two effects that a channel-count argument alone cannot distinguish. The smaller first step (-0.012 in *F1-score* from P0 to P2) reflects the residual benefit of sampling a range of migration velocities: because the test site is homogeneous and its velocity is well calibrated, every velocity in the stack is close to the correct one, so this step measures the value of having the focusing criterion represented explicitly in the channel dimension rather than the ability to compensate for a mis-set velocity. The second and much larger step (-0.089 from P2 to P1) measures the value of the migration itself, which is adopted in this work as a standard pre-processing step following common practice in GPR-based root detection: without migration the diffraction hyperbolas are not collapsed, their energy is spread along the two-way travel-time axis, and the network must perform the focusing implicitly from a single image. Both components of this degradation are consistent with the error sources identified in Section 3.1.1: the recall rise from 0.778 to 0.888 (P1 to P0) indicates that the collapsed hyperbola makes weak roots easier to recover, whereas the precision rise from 0.809 to 0.901 shows that the migration also reduces the ambiguity between root apices and isolated non-root reflectors. Because the three configurations share an identical 16 × 128 × 32 input tensor and an identical parameter count of 17.13 M, these differences cannot be attributed to model capacity. It should be noted that the multi-velocity input representation is a pre-processing design adopted and validated in this pipeline rather than a methodological contribution of this work; the ablation is included to quantify its role within the pipeline, and the core contributions of this study remain the CGL-Net architecture (Section 3.3.3) and the GPCL strategy (Section 3.3.3).

Table 10. Ablation results of the input representation [1].

Input Representation	Channel Content	*P*	*R*	*F1-Score*
P1: raw B-scan without migration, replicated across 16 channels	single image replicated ×16	0.809	0.778	0.793
P2: single-velocity F-K migration (v = 0.1272 m/ns), replicated across 16 channels	single migrated image replicated ×16	0.894	0.870	0.882
P0: 15-velocity F-K stack + mean channel (proposed)	15 migrated velocities + 1 mean	0.901	0.888	0.894

[1] All three configurations use the same CGL-Net backbone, the same GPCL post-processing, and the same 16 × 128 × 32 input tensor; they differ only in the content of the 16 channels. In P1 and P2 the single (migrated or non-migrated) image is replicated across all 16 channels, so the parameter count (17.13 M) and *FLOPs* (2.36 G) are identical across rows and the comparison isolates the effect of the input representation rather than that of model capacity. P1 and P2 were trained from scratch with the same data split, loss, and training schedule as P0.

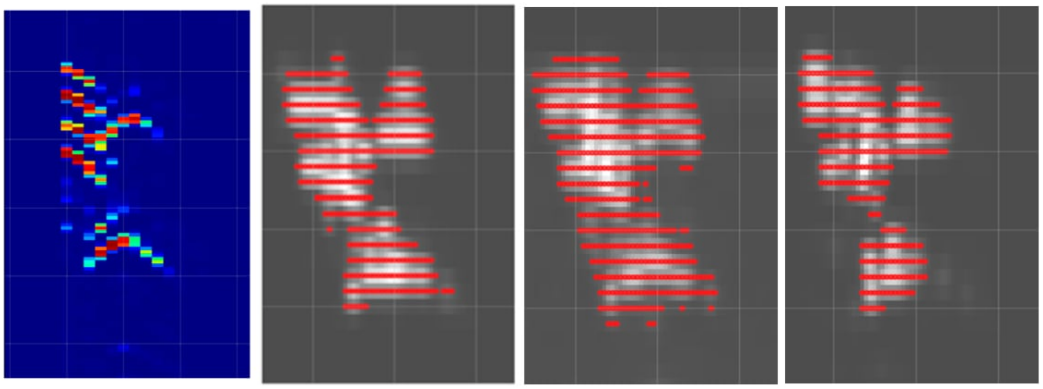


Figure 14. Localization results obtained with the three input representations examined in the input-representation ablation (Section 3.3.1): (**a**) input; (**b**) P1, the conventionally pre-processed B-scan without migration, replicated across the 16 input channels; (**c**) P2, single-velocity F-K migration at the site velocity of 0.1272 m/ns, replicated across the 16 input channels; (**d**) P0, the 15-velocity Stolt F-K stack plus one mean channel (the configuration used throughout this paper). Coloured rectangles denote detection results, light-green rectangles denote ground-truth annotations, and cross marks inside the boxes indicate the estimated centre positions. All four panels show the same B-scan; the same CGL-Net backbone and GPCL post-processing are used in (**b**–**d**), and only the input representation differs.

3.3.2. Ablation of CGL-Net Internal Modules

Ablation experiments were conducted on CGL-Net, and the quantitative evaluation results of the different model configurations on the test set are shown in Table 11 and Figure 15. Ghost-UNet (Figure 15c) achieves a comparable *F1-score* to standard U-Net (Figure 15b), both being 0.813; in terms of model complexity, the number of parameters is reduced by 45.5%, *FLOPs* by 47.9%, the inference speed is improved by 10.3%, and the training time is reduced by 19.3% compared with standard U-Net, indicating that the Ghost module reduces the model parameters and computational cost while preserving the detection accuracy. With the CBAM attention module introduced alone, the detection confidence increases but there are many missed detections (Figure 15d): the precision improves by 0.084 and the recall improves by 0.029. With the LSTM temporal module introduced alone, the missed detections decrease but the localization becomes diffuse (Figure 15e): the recall reaches 0.949 and the *F1-score* reaches 0.858, with detailed detection

results and few missed detections, while false alarms are prominent. The root responses of the complete CGL-Net are precisely concentrated on the root positions (Figure 15f), and the precision, recall, and *F1-score* reach 0.901, 0.888, and 0.894, respectively, achieving the highest *F1-score* among the evaluated configurations; the complete model also reduces the number of parameters from 31.05 M to 17.13 M (about 44.8%) and the *FLOPs* from 3.05 G to 2.36 G (about 22.6%) relative to standard U-Net, so that a higher detection performance is obtained at a smaller model scale.

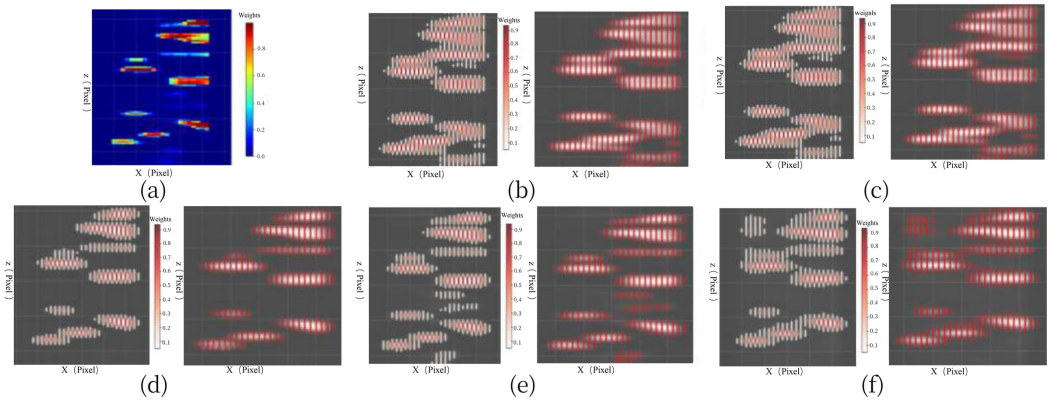


Figure 15. Results and analysis of different improved modules: (**a**) Input. (**b**) U-Net. (**c**) Ghost-Unet. (**d**) Ghost-UNet + CBAM. (**e**) Ghost-UNet + LSTM. (**f**) CGL-Net. Each panel shows the weight localisation map predicted by the corresponding model on the same B-scan, overlaid on the B-scan itself; the red regions indicate the root responses predicted by each model, and the colour bar gives the normalized weight from 0.1 (low response) to 1.0 (high response).

Table 11. Ablation results of internal modules in CGL-Net.

Model Configuration	*Params* (M)	*FLOPs* (G)	*FPS*	Training Time	*P*	*R*	*F1-Score*
Standard U-Net	31.05	3.05	190.9	75 min 44 s	0.830	0.796	0.813
Ghost-UNet	16.93	1.59	210.5	61 min 7 s	0.828	0.798	0.813
Ghost-UNet + CBAM	16.95	1.62	187.0	89 min 58 s	0.912	0.827	0.867
Ghost-UNet + LSTM	17.11	2.33	171.3	127 min 57 s	0.783	0.949	0.858
CGL-Net [1]	17.13	2.36	126.3	167 min 9 s	0.901	0.888	0.894

[1] The three ablation tables (Sections 3.3.2, 3.3.3 and 4.1) differ in which component is varied, not in the evaluation protocol. This table compares backbone networks, each combined with the full GPCL module; the post-processing ablation in Section 3.3.3 ablates the GPCL stage on the outputs of the same CGL-Net; the model comparison in Section 4.1 reports the complete CGL-Net + GPCL system. The last rows of the three tables therefore describe one and the same system evaluated identically, and their identical values are expected rather than an inconsistency.

The pattern across the four configurations—the largest single-module contribution from the LSTM module, followed by CBAM and the Ghost backbone—mirrors the error sources identified in Section 3.1.1: the temporal module addresses depth-direction ambiguity, the attention module addresses clutter confusion, and the backbone affects efficiency more than accuracy.

3.3.3. Ablation of the GPCL Fusion Strategy

To intuitively analyze the effects of the gradient point search method and the DBSCAN clustering algorithm on root localization, quantitative evaluations of the different post-processing strategies were performed on the basis of the CGL-Net output, with the results shown in Table 12 and Figure 16. All strategies are evaluated on the same set of fixed CGL-

Net outputs, on the same test set, under the same pixel-level counting rule; the thresholds are selected before the evaluation and remain unchanged throughout. Compared with the threshold segmentation strategy, the gradient point search method improves the precision by 0.141, indicating that the gradient clustering-based localization method can effectively suppress the interference of unstructured clutter on the detection results; at the same time, its recall decreases by 0.289. The complete GPCL applies the localization stages in sequence, so that the gradient point search supplies the structural evidence of the root apices and the adaptive DBSCAN clustering converts each spatially dense group of gradient points into an independent root detection: the precision is further raised to 0.901 and the recall recovers to 0.888, giving an *F1-score* of 0.894, i.e., 0.073 higher than the threshold segmentation baseline, which is the best balance between precision and recall among the evaluated strategies.

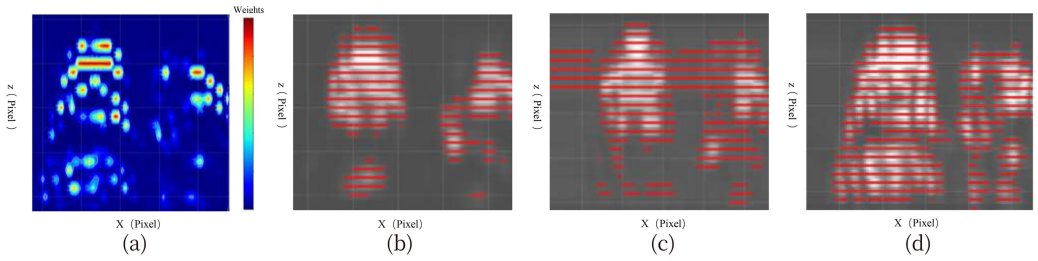


Figure 16. Experimental results and analysis of different post-processing strategies: (**a**) Original weight maps. (**b**) Threshold-segmented localization map. (**c**) Gradient point search-based localization map. (**d**) GPCL map. Note: All strategies are applied to the same set of fixed CGL-Net weight localization maps, and are evaluated on the same set under the same pixel-level counting rule; the threshold used for the threshold segmentation strategy is determined before the evaluation and is not tuned on the test set. The precision, recall and F1-score are pixel-level metrics computed on the root masks, and the F1 improvement of 0.073 refers to the threshold segmentation baseline.

Table 12. Ablation results of the different post-processing strategies (pixel level).

Post-Processing Strategy	*P*	*R*	*F1-Score*
Threshold segmentation	0.730	0.939	0.821
Gradient point search	0.871	0.650	0.744
GPCL	0.901	0.888	0.894

4. Discussion

4.1. Comparison with Existing Root Localization Methods

CGL-Net combined with GPCL was compared with three general-purpose detection models—the YOLOv5s detection module reported by Li et al. [47], the RT-DETR reported by Zhao et al. [48], and the CMT-Net reported by Cui et al. [49], a deep network designed specifically for GPR-based root localization and diameter extraction—and with two specialized deep-learning GPR methods: GPRInvNet [50], a deep network for physics-guided GPR inversion and target imaging, and GPR-UNet [51], a U-Net-based GPR semantic-segmentation architecture. Since neither architecture was originally designed for target localization, their outputs were converted into root positions by direct extraction: for GPR-UNet, the connected components of the thresholded segmentation mask serve as the root positions; for GPRInvNet, the local maxima (peaks) of the inverted image serve as the root positions. Both models consume the same 16-channel focused input as CGL-Net, so that the three specialized GPR methods share the input representation and differ only in the network that produces the output map and in how naturally that map translates into root positions; in particular, no additional post-processing is applied to the baselines beyond

this direct extraction, while the localization stage of the proposed method (GPCL) is an integral part of it, whose individual contribution has been quantified separately by the ablation in Table 12. The three general-purpose detectors were run in their original RGB three-channel input format: the single-channel B-scan was replicated to three channels and resized to 128 × 128 pixels to meet their input requirements. GPR-UNet, GPRInvNet, and CGL-Net + GPCL shared the same input representation—the 16-channel focused feature maps (16 × 128 × 32)—so that the comparison among the three specialized GPR methods is strictly input-controlled: any performance difference between them can be attributed to the network architecture and the localization post-processing rather than to the input representation. The three detectors, whose architectures require image-like inputs, are instead compared under their respective original input formats. All six methods were trained with the Adam optimizer (initial learning rate 0.001), a batch size of 1, a maximum of 100 epochs, the same training/validation/test split (70%:15%:15%), and the same data preprocessing pipeline. The comparison methods adopted the default backbones and hyperparameters recommended in their original publications, while the threshold parameters of CGL-Net and the parameters of GPCL were kept unchanged. The comparison results are shown in Table 13 and Figure 17. YOLOv5s (Figure 17a) produced a considerable number of missed detections in root localization; RT-DETR (Figure 17b) improved over YOLOv5s but its localization remained imprecise; CMT-Net (Figure 17c) showed further improvement, although the target edges were still diffuse. The two specialized GPR methods, which consume the same migration features as the proposed method, clearly outperformed the three general-purpose detectors (F1 of 0.850 for GPRInvNet and 0.830 for GPR-UNet, versus 0.600–0.790), confirming that the shared input representation carries a substantial part of the discriminative information. Nevertheless, the localization response of GPR-UNet (Figure 17d) and GPRInvNet (Figure 17e) remained partially diffuse along the root axis, whereas the response of CGL-Net + GPCL (Figure 17f) was more concentrated on the root regions, with a clear reduction in background clutter response. Among all comparison models, GPRInvNet ranks second by F1-score (0.850), followed by GPR-UNet (0.830); the two are complementary rather than ordered consistently across metrics, since GPR-UNet attains the highest precision of all baselines (0.870) while GPRInvNet attains the highest recall (0.860). CGL-Net + GPCL improves the P, R and F1-score by 0.051, 0.028 and 0.044, respectively, over GPRInvNet, and by 0.061, 0.138 and 0.104 over CMT-Net. Its F1-score is higher than that of every comparison model, and its recall exceeds that of all baselines, whereas its precision is 0.031 lower than that of GPR-UNet—a deliberate consequence of GPCL, which converts densely distributed gradient points into separate root detections instead of merging them into one large connected region, and which therefore trades a small amount of precision for a considerably higher recall (Section 3.3.3). This indicates that the proposed architecture and the GPCL post-processing add a consistent gain beyond the shared input representation in background clutter suppression and root localization accuracy.

Table 13. Statistical results of model comparison experiments. All metrics are computed at the pixel level on the same simulation test set.

Model	Training Time	*P*	*R*	*F1-Score*
YOLOv5s	32 min 47 s	0.61	0.59	0.60
RT-DETR	281 min 14 s	0.78	0.72	0.75
CMT-Net	75 min 26 s	0.84	0.75	0.79
GPR-UNet	139 min 03 s	0.87	0.80	0.83
GPRInvNet	113 min 32 s	0.85	0.86	0.85
CGL-Net + GPCL	167 min 9 s	0.90	0.89	0.89

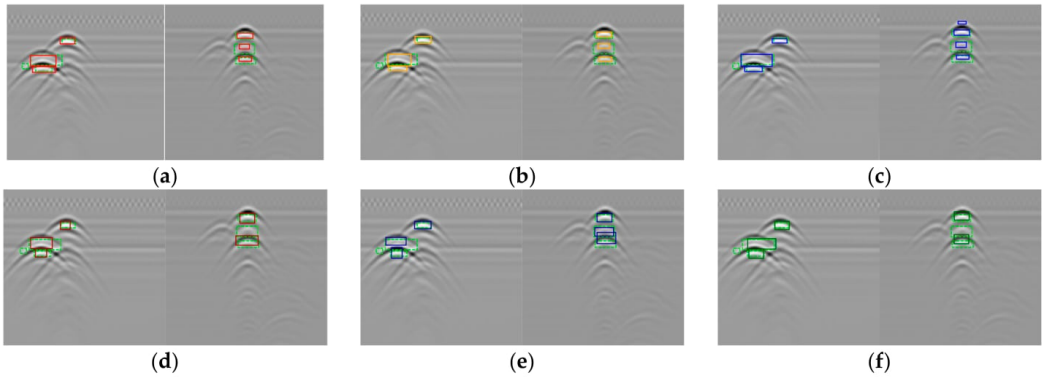


Figure 17. Comparison of results from existing models: (**a**) YOLOv5s. (**b**) RT-DETR. (**c**) CMT-Net. (**d**) GPR-UNet. (**e**) GPRInvNet. (**f**) CGL-Net + GPCL. Colored rectangles denote the detection results of each method, light-green rectangles denote the ground-truth root annotations.

4.2. Root Diameter Estimation and Geometric Hypothesis Validation

The diameter estimator rests on one hypothesis: the vertical span of the root response, Δz_c, grows linearly with the root diameter D. Physically, a near-horizontal cylindrical root appears in the migrated section as two reflectors—the upper and lower soil–root interfaces—whose vertical separation is the projected diameter $D{\cdot}cos\varphi$, where φ is the inclination of the root axis with respect to the horizontal plane; each interface response additionally has a finite vertical extent set by the effective wavelet length, approximately $\lambda/4 = v/(4f_0) \approx 21$ mm at 1.5 GHz. Δz_c is therefore linear in D, with an intercept contributed by the wavelet duration. The remaining factors are absorbed into the empirical coefficients of the fit: the propagation velocity is calibrated on site (Section 2.2), the root–soil dielectric contrast affects echo amplitude and hence detectability rather than echo geometry, the antenna bandwidth sets the intercept, and migration focuses point-like responses so that their residual extent is dominated by the wavelet. The hypothesis was tested on 150 root instances with diameters of 16–80 mm drawn at random from the simulated corpus. The size of this subset is set by the measurement rather than by the size of the corpus: Δz_c is read by hand on the migrated section of each instance, following the definition given in Section 2.5.1, so the calibration was carried out on 150 hand measurements instead of being extended to the whole corpus. The relation between Δz_c and D is therefore established independently of the network response, and the instances that define it are a random, not a selected, subset of the simulated data. As shown in Figure 18, Δz_c is significantly linearly correlated with the root diameter, and the regression equation is:

$$D = 0.9374\Delta z_c + 5.28 \quad (10)$$

where D is the root diameter (mm) and Δz_c is the feature quantity (mm). The regression model yields a coefficient of determination of $R^2 = 0.8045$ and a root mean square error *RMSE* of 6.59 mm, supporting the use of Δz_c as an empirical predictor of root diameter. The definition of Δz_c, the conversion from its pixel value to the physical depth span, and the rule for deriving the root diameter from the fitted equation are given in Section 2.5.1.

The definition of Δz_c throughout this calibration is identical to that of Section 2.5.1: the vertical span of the gradient feature points within the cluster. With this definition, the fitted intercept of 5.28 mm is physically consistent: it reflects the finite duration of the superposed boundary wave packet, which does not cancel when the span of the whole cluster, rather than the peak-to-peak spacing, is measured. The calibration is frequency-specific through the resolvability of the two interface reflections: the separation at which the two reflections can still contribute distinct gradient points is approximately $\lambda/4$—21 mm at 1.5 GHz, 33 mm at 900 MHz and 79 mm at 400 MHz—so that coarse roots are measured by essentially the same slope at every frequency, while for fine roots the two boundaries merge into a single

packet and the span approaches the wavelet duration rather than the projected diameter. The slope can therefore be expected to transfer across antennas for roots well above the resolution floor, whereas the intercept and the small-diameter limit must be recalibrated for each frequency.

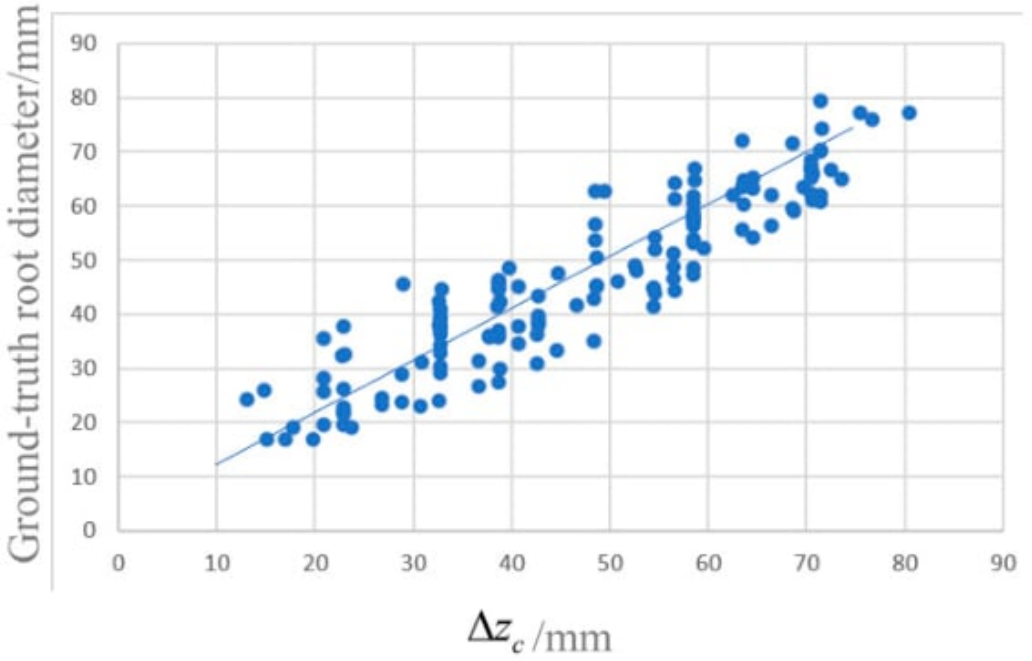


Figure 18. Linear relationship between the clustering depth span Δz_c and the root diameter D (both in mm), fitted as $D = 0.9374\Delta z_c + 5.28$. The calibration was tested on 150 synthetic root instances with diameters of 16–80 mm, for which Δz was measured manually on the migrated sections (Section 2.5.1), independently of the network response; the regression yields $R^2 = 0.8045$ and $RMSE = 6.59$ mm. A leave-one-out cross-validation over the same 150 instances yields $R^2 = 0.8026$ and $RMSE = 6.67$ mm, essentially identical to the fitting statistics.

The calibrated range is bounded by the simulated corpus itself: the simulated root diameters span 16–80 mm (Section 2.1.1), so the linear model is defined, and its coefficients identified, only within this interval. No simulated instances exist outside it, and the excavation ground truth of the field campaign does not extend the calibration beyond the simulated corpus. Outside the range, the behavior of the estimator is bounded by physics rather than by the fit: below 16 mm, the two boundary reflections merge as the separation approaches the resolution floor of approximately $\lambda/4$ (21 mm at 1.5 GHz), so the span saturates and the diameter is progressively underestimated; above 80 mm, the estimate is a linear extrapolation whose uncertainty grows with distance from the calibration interval. Both regimes are therefore stated as limitations (Section 4.4), and 16–80 mm is the validated operating range of the diameter estimator.

4.3. Analysis of the Effect of Loss Functions

Using the CGL-Net model proposed in this study with a learning rate of 0.001, the validation behavior and segmentation quality of different loss functions (MSE, SSIM, and MSE + SSIM) were analyzed under identical training conditions, and the results are shown in Table 14 and Figure 19. L_{MSE} converged stably (Figure 19a). When L_{SSIM} was used alone, the validation loss curve fluctuated continuously during training and no stable convergence was observed (Figure 19b). Combining SSIM with MSE gave a training time of 167 min 9 s, shorter than that of MSE alone (189 min 47 s) and SSIM alone (182 min 55 s). In terms of

segmentation quality, the SSIM and PSNR reached 0.875 and 23.75 dB, respectively, which are 0.300 and 5.28 dB higher than those of L_{MSE}, and 0.031 and 2.68 dB higher than those of L_{SSIM} (Figure 19c).

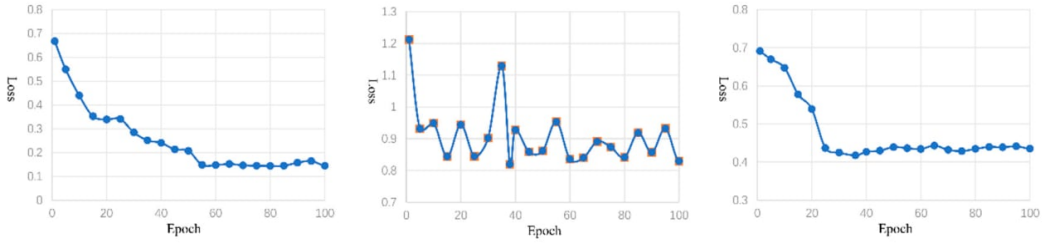


Figure 19. Loss curves of different loss functions: (**a**) L_{MSE}. (**b**) L_{SSIM}. (**c**) $L_{SSIM+MSE}$.

Table 14. Ablation results of loss functions [1].

Loss Function Configuration	SSIM	PSNR (dB)	Epochs to Convergence	Training Time
MSE	0.575	18.47	52	189 min 47 s
SSIM	0.844	21.07	-	182 min 55 s
MSE + SSIM	0.875	23.75	36	167 min 9 s

[1] All configurations were trained on the same simulation training set with identical optimizer settings. *SSIM* and *PSNR* are reported as image-quality metrics, where the loss corresponding to SSIM is $1 - SSIM$. "Convergence epoch" denotes the epoch at which the validation loss stabilizes; "-" indicates that no stable convergence was observed within the 100-epoch budget. *PSNR* was computed with a data peak of 1.0.

4.4. Limitations and Future Work

This study provides a complete technical scheme for coarse-root detection at the individual-root scale, with application prospects in root architecture surveys, health assessment of ancient trees, and belowground biomass carbon estimation. The field validation was conducted on *Populus tomentosa*; the framework itself is tree-species independent and can be transferred directly once annotated data for other species become available.

The principal limitation of the present framework is its two-dimensional scope: the analysis is performed on individual profiles, so the spatial distribution, connectivity, and overall topology of a root system cannot be reconstructed, unlike the three-dimensional architecture that stand-scale carbon-sink estimation requires. This limitation concerns architectural completeness only, not the accuracy of the within-section estimates, all of which were validated against excavation ground truth (Tables 5, 7 and 9).

The applicability of the framework is bounded in three respects: dielectric properties (the training data span soil $\varepsilon r = 3$–7 and root $\varepsilon r = 15$–35, and performance outside this range is unvalidated); species and architecture (cross-species transfer is unverified, as the field validation involved a single species and the simulations used simplified two-to-five-root configurations); and root diameter (the calibration is valid for 16–80 mm, whereas the excavated roots span 10–130 mm, so coarser roots are extrapolated and finer roots approach the resolution floor and are progressively underestimated). 16–80 mm is therefore the validated operating range, site-specific recalibration is required for stands dominated by thicker roots, and a systematic sensitivity analysis across these conditions is left as future work.

The excavation-based ground truth also bounds the absolute accuracy: the 15 cm profile spacing makes the angular reference coarser than the directly measured depths and diameters (consistent with the larger angle errors in Table 7), the angle observations

from adjacent survey-line pairs are not independent (Section 2.6), and the 0.1 m matching tolerance admits a corresponding positional uncertainty (Section 2.3). These factors do not affect the relative comparison among methods.

The field split was also examined under line-level and tree-level grouping strategies. Because the test images comprise all 1500 MHz acquisitions of the two scanned trees while fine-tuning uses only their 400 and 900 MHz scans, the frequency-blocked split already prevents any test-frequency scan from entering training; a tree-level hold-out would additionally remove the cross-frequency fine-tuning signal, but with only two scanned trees it would leave a single tree for fine-tuning and reduce the fine-tuning set from 280 to roughly 130 scans. Tree-level generalization to unseen individuals therefore remains unverified in this study.

The field conditions add two further constraints: the time–depth conversion uses a single site-averaged velocity ($v = 0.1272$ m/ns), so local deviations in water content and bulk density translate proportionally into depth biases—mitigated but not eliminated by the multi-velocity migration input and the 12-location permittivity sampling, with per-B-scan velocity focusing left as a refinement; and real signals contain effects only idealized in the simulations (surface clutter, moisture-dependent attenuation, contrast-dependent visibility), which the CycleGAN style transfer narrows only at the texture level.

Future work will extend the CGL-Net framework to a three-dimensional convolutional neural network, build a three-dimensional simulation dataset of tree roots, and combine it with grid-based multi-survey-line GPR scanning data to achieve an end-to-end mapping from three-dimensional radar data volumes to three-dimensional root structures.

5. Conclusions

To address the challenges in GPR root detection, such as complex hyperbola recognition, strong environmental noise interference, and difficult parameter acquisition, this study proposes and implements a root detection and parameter inversion system integrating prior physical knowledge and deep learning—the CGL-Net and GPCL algorithms. The main conclusions are as follows:

1. The complete CGL-Net configuration delivers the best overall detection performance, with a precision, recall, and *F1-score* of 0.901, 0.888, and 0.894, respectively, on the simulated test set, achieving the highest *F1-score* among the evaluated configurations.
2. Replacing fixed-threshold segmentation with a sample-adaptive realization of the GPCL framework raises the precision and F1-score of root detection: relative to threshold segmentation alone, it increases the precision by 0.171 and the F1-score by 0.073. Because the framework and the three basic operators are those of [25], whereas the adaptive gradient threshold and the adaptive DBSCAN parameters are introduced here to cope with the variable amplitude and point density of the CGL-Net weight maps, these gains should be attributed to the adaptation of the localization step to weakly supervised maps rather than to the localization framework itself.
3. In terms of multi-parameter root estimation, the correct rates (CR) of depth and diameter estimation reach 87.2% and 91.7%, respectively, on the pooled dataset of all three sources, which comprises 450 evaluated B-scans (240 purely simulated, 150 style-transferred, and 60 measured scans) together with the matched root instances detected in them, under the depth and diameter tolerances of 30% and 20%. On the field-measured data, the CRs of the elevation angle and the horizontal azimuth angle are 92.1% and 72.4% under the 5.0° tolerance. In the root-by-root comparison with the field excavation of nine coarse roots, the mean absolute depth error is 3.3 cm and the mean absolute diameter error is 0.7 cm. Distribution-level statistics with 95% cluster-

bootstrap *CIs* (median depth error 22.2%, *CI* 15.2–28.5%; median diameter error 22.7%, *CI* 15.9–27.4%) quantify the field uncertainty of the estimates.

This study demonstrates that the hierarchical integration of physical models, deep learning, and traditional image processing is an effective approach to improving the accuracy and reliability of non-destructive GPR root detection, providing a technical means for the quantitative characterization of coarse roots.

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

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 32672346, 32471858), the National Key R&D Program of China (Grant No. 2023YFC3006800) and 5·5 Engineering Research & Innovation Team Project of Beijing Forestry University.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request.

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

References

1. Zhang, Z.Y.; Fan, B.M.; Song, C.; Zhang, X.X.; Zhao, Q.W.; Ye, B. Advances in root system architecture: Functionality, plasticity, and research methods. *J. Resour. Ecol.* **2023**, *14*, 15–24. [CrossRef]
2. Liu, H.; Fu, X. Phytohormonal networks facilitate plant root developmental adaptations to environmental changes. *Sci. Bull.* **2024**, *69*, 709–713. [CrossRef] [PubMed]
3. Mo, L.; Zohner, C.M.; Reich, P.B.; Liang, J.; de Miguel, S.; Nabuurs, G.-J.; Renner, S.S.; Hoogen, J.v.D.; Araza, A.; Herold, M.; et al. Integrated global assessment of the natural forest carbon potential. *Nature* **2023**, *624*, 92–101. [CrossRef] [PubMed]
4. Li, G.H.; Wang, Z.X.; Xu, H.; Liu, M. Root diameter prediction method of fruit trees based on ground penetrating radar and deep learning. *Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 306–313. (In Chinese) [CrossRef]
5. Wang, B.; Zhang, J.J.; Lai, Z.R. Effects of soil water content on the accuracy of ground penetrating radar for detecting plant root system architecture. *Arid Zone Res.* **2024**, *41*, 456–466. (In Chinese) [CrossRef]
6. Lombardi, F.; Podd, F.; Solla, M. From its core to the niche: Insights from GPR applications. *Remote Sens.* **2022**, *14*, 3033. [CrossRef]
7. Kaur, P.; Dana, K.J.; Romero, F.A.; Gucunski, N. Automated GPR rebar analysis for robotic bridge deck evaluation. *IEEE Trans. Cybern.* **2016**, *46*, 2265–2276. [CrossRef] [PubMed]
8. Hruška, J.; Čermák, J.; Sustek, S. Mapping tree root systems with ground-penetrating radar. *Tree Physiol.* **1999**, *19*, 125–130. [CrossRef] [PubMed]
9. Li, X.B.; Cheng, X.; Zhao, Y.J.; Xiang, B.B.; Zhang, T.H. Deep learning-based ground-penetrating radar inversion for tree roots in heterogeneous soil. *Sensors* **2025**, *25*, 947. [CrossRef] [PubMed]
10. Lei, W.; Hou, F.; Xi, J.; Tan, Q.; Xu, M.; Jiang, X.; Liu, G.; Gu, Q. Automatic hyperbola detection and fitting in GPR B-scan image. *Autom. Constr.* **2019**, *106*, 102839. [CrossRef]
11. Hou, F.; Lei, W.; Li, S.; Xi, J. Deep learning-based subsurface target detection from GPR scans. *IEEE Sens. J.* **2021**, *21*, 8161–8171. [CrossRef]
12. Wang, Z.P.; Zhang, X.W.; Xue, F.X.; Wen, J.; Han, H.Y.; Huang, Y. Estimating the location and diameter of tree roots using ground penetrating radar. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 160–168. (In Chinese) [CrossRef]
13. Xu, H. Tree-Root Localization Algorithm Based on Ground Penetrating Radar. Master's Thesis, Jiangnan University, Wuxi, China, 2022. (In Chinese)
14. Chen, P.S.; Yuan, Q.; Zhang, Z.P.; Yang, L.; Chen, Z.L.; Wu, L. Image interpretation method for geological advance prediction in water-rich fractured zones of tunnels based on convolutional neural network. *J. Basic Sci. Eng.* **2022**, *30*, 196–207. (In Chinese) [CrossRef]
15. Wang, Z.X. Tree-Root Detection Algorithm Based on Ground Penetrating Radar. Master's Thesis, Jiangnan University, Wuxi, China, 2023. (In Chinese)
16. Yang, D.; Guo, C.; Persico, R.; Liu, Y.J.; Liu, H.D.; Bai, C.J.; Lian, C.; Zhao, Q. Adaptive variational mode decomposition and principal component analysis-based denoising scheme for borehole radar data. *Remote Sens.* **2025**, *17*, 525. [CrossRef]

17. Shi, J.F.; Guo, L.; Zhang, L.Y.; Cui, X.H.; Hou, F.F.; Huang, T.B.; Xu, H.; Li, X.J.; Hu, Y. Root point recognition and wave velocity estimation of GPR images based on YOLOv4 model. *Prog. Geophys.* **2025**, *40*, 2211–2226. (In Chinese) [CrossRef]
18. Zhao, X.; Zhang, X.W.; Lv, S.H.; Li, M.; Zhang, J.H.; Lin, C.; Wen, J. Uncertainty-aware tree root morphology and phenotypic parameters reconstructed from ground-penetrating radar signal via physics-informed variational networks. *IEEE Trans. Geosci. Remote Sens.* **2026**, *64*, 4702214. [CrossRef]
19. Zhang, X.W.; Zhao, X.; Lv, S.H.; Xv, L.F.; Lin, C.; Wen, J. Deep learning inversion of ground-penetrating radar with prior physical velocity field for complex subsurface root system. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4407215. [CrossRef]
20. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
21. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [CrossRef]
22. Han, K.; Wang, Y.H.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1577–1586. [CrossRef]
23. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
24. Lei, W.; Luo, J.; Hou, F.; Xu, L.; Wang, R.; Jiang, X. Underground cylindrical objects detection and diameter identification in GPR B-scans via the CNN-LSTM framework. *Electronics* **2020**, *9*, 1804. [CrossRef]
25. Hou, F.; Fang, M.; Luo, T.X.-H.; Fan, X.; Guo, Y. Dual-task GPR method: Improved generative adversarial clutter suppression network and adaptive target localization algorithm in GPR image. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5108313. [CrossRef]
26. Zhang, X.; Xue, F.; Wang, Z.; Wen, J.; Guan, C.; Wang, F.; Han, L.; Ying, N. A novel method of hyperbola recognition in ground penetrating radar (GPR) B-scan image for tree roots detection. *Forests* **2021**, *12*, 1019. [CrossRef]
27. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR, USA, 2–4 August 1996; pp. 226–231.
28. Zhao, X.; Wang, Y.; Gui, R. Accurate automated GPR target detection via SFCW response synthesis and DBSCAN. In Proceedings of the 2026 IEEE International Conference on Computational Electromagnetics (ICCEM), Shanghai, China, 2026; pp. 1–4. [CrossRef]
29. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232. [CrossRef]
30. Wu, J.L.; Yang, H.J. Research on image style transfer technology based on improved CycleGAN network. *Inf. Technol. Informatiz.* **2025**, *2*, 79–82. (In Chinese)
31. Warren, C.; Giannopoulos, A.; Giannakis, I. gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Comput. Phys. Commun.* **2016**, *209*, 163–170. [CrossRef]
32. Feng, D.S.; Tan, Y.; Yang, J.; Zhang, L.J.; Yuan, Z.M.; Liu, J.; Wang, X. Complex frequency shifted perfectly matched layer for symplectic partitioned Runge-Kutta method forward modeling of ground penetrating radar. *Sci. Technol. Eng.* **2022**, *22*, 3473–3484. (In Chinese) [CrossRef]
33. Al Hagrey, S.A. Geophysical imaging of root-zone, trunk, and moisture heterogeneity. *J. Exp. Bot.* **2007**, *58*, 839–854. [CrossRef] [PubMed]
34. Guo, L.; Cui, X.H.; Chen, J. Sensitivity analysis of ground penetrating radar root detection based on GprMax forward modeling. *Prog. Geophys.* **2012**, *27*, 1754–1763. (In Chinese) [CrossRef]
35. Guo, Y.Y.; Li, Z. Analysis of land use status and sustainable land use countermeasures in Gaotang County. *Low. Carbon Estate* **2016**, *19*. (In Chinese) [CrossRef]
36. Mao, X.H.; Yan, S.B.; Ge, L.; Wang, L.; Zhang, F.; Dong, Y.F.; Zang, Z.R.; Li, S.W. Study on construction method of microcore germplasm of *Populus tomentosa*. *J. Beijing For. Univ.* **2023**, *45*, 58–67. (In Chinese) [CrossRef]
37. Qin, R.H.; Fan, G.Q.; Han, Q.L.; Zheng, Y.L.; Xu, J.C.; Liang, H. Inversion of relative dielectric constant of tree root zone based on MWFCNet. *Sci. Silvae Sin.* **2026**, *62*, 109–121. (In Chinese)
38. Lin, Y.; Wang, J.C.; Ma, D.Y.; Wang, Y.P.; Ye, S.B. Improved cycle-consistency generative adversarial network-based clutter suppression methods for ground-penetrating radar pipeline data. *Remote Sens.* **2024**, *16*, 1043. [CrossRef]
39. Lei, W.T.; Tan, X.; Luo, C.P.; Xue, W. Mutual interference suppression and signal enhancement method for ground-penetrating radar based on deep learning. *Electronics* **2024**, *13*, 4722. [CrossRef]
40. Yu, C.; Li, A.L.; Wu, L.L.; Qing, L.M. Study and application of migration technology in ground penetrating radar (GPR) data processing. In *Proceedings of the Third International Conference on Advanced Algorithms and Signal Image Processing (AASIP 2023)*; SPIE: Bellingham, WA, USA, 2023; Volume 12799, p. 127990H. (In Chinese) [CrossRef]

41. Stolt, R.H. Migration by Fourier transform. *Geophysics* **1978**, *43*, 23–48. [CrossRef]
42. Sheng, S.; Yang, X.; Lan, T. Enhanced GPR inversion method in complex underground environment. In Proceedings of the 2024 Photonics and Electromagnetics Research Symposium (PIERS), Chengdu, China, 21–25 April 2024. [CrossRef]
43. Huang, M.; Liang, J.; Yin, P.; Zhu, X.; Jia, Z. SAU-GAN: A shuffle attention U-Net generative adversarial network for GPR inversion. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 7505005. [CrossRef]
44. Zhu, S.; Huang, C.; Su, Y.; Sato, M. 3D GPR to detect tree roots and estimate root biomass in the field. *Remote Sens.* **2014**, *6*, 5754–5773. [CrossRef]
45. Topp, G.C.; Davis, J.L.; Annan, A.P. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.* **1980**, *16*, 574–582. [CrossRef]
46. Cui, X.; Chen, J.; Shen, J.; Cao, X.; Chen, X.; Zhu, X. Modeling tree root diameter and biomass by ground-penetrating radar. *Sci. China Earth Sci.* **2011**, *54*, 711–719. [CrossRef]
47. Li, S.; Zhang, X.W.; Tan, X.; Xu, L.F.; Lyu, S.H.; Wen, J. Deep learning-based inverse identification of multi-target parameters for tree rooting ground-penetrating radar. *J. Beijing For. Univ.* **2024**, *46*, 103–114. (In Chinese) [CrossRef]
48. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. DETRs beat YOLOs on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 16965–16974. [CrossRef]
49. Cui, X.; Li, S.; Zhang, L.; Peng, L.; Guo, L.; Cao, X.; Chen, X.; Yin, H.; Shen, M. Integrated extraction of root diameter and location in ground-penetrating radar images via CycleGAN-guided multi-task neural network. *Forests* **2025**, *16*, 110. [CrossRef]
50. Liu, B.; Ren, Y.; Liu, H.; Xu, H.; Wang, Z.; Cohn, A.G.; Jiang, P. GPRInvNet: Deep Learning-Based Ground-Penetrating Radar Data Inversion for Tunnel Linings. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8305–8325. [CrossRef]
51. Li, G.H.; Ma, J.H.; Wang, Z.X.; Wei, S. Suppressing ground penetrating radar clutter to predict root parameters using deep neural networks. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 171–180. (In Chinese) [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.