EDENet: Echo Direction Encoding Network for Place Recognition Based on Ground Penetrating Radar

Pengyu Zhang[1,2], Xieyuanli Chen[1], Yuwei Chen[1], Beizhen Bi[1], Zhuo Xu[1], Tian Jin[1], Xiaotao Huang[1], Liang Shen[1] *

[1]National University of Defense Technology, Changsha, 410073, China
[2]National University of Singapore, 117576, Singapore
{zhangpengyu19, xieyuanli.chen, yuweichen15, beizhenbi20, xuzhuo19, tianjin, xthuang, liangshen16}@nudt.edu.cn

Abstract

Ground penetrating radar (GPR) based localization has gained significant recognition in robotics due to its ability to detect stable subsurface features, offering advantages in environments where traditional sensors like cameras and LiDAR may struggle. However, existing methods are primarily focused on small-scale place recognition (PR), leaving the challenges of PR in large-scale maps unaddressed. These challenges include the inherent sparsity of underground features and the variability in underground dielectric constants, which complicate robust localization. In this work, we investigate the geometric relationship between GPR echo sequences and underground scenes, leveraging the robustness of directional features to inform our network design. We introduce learnable Gabor filters for the precise extraction of directional responses, coupled with a direction-aware attention mechanism for effective geometric encoding. To further enhance performance, we incorporate a shift-invariant unit and a multi-scale aggregation strategy to better accommodate variations in dielectric constants. Experiments conducted on public datasets demonstrate that our proposed EDENet not only surpasses existing solutions in terms of PR performance but also offers advantages in model size and computational efficiency.

Introduction

Ground Penetrating Radar (GPR) for vertical detection of robust and invariant underground features has garnered significant attention in recent years in vehicle/robot localization (Kouros et al. 2018; Ort, Gilitschenski, and Rus 2023; Zhang et al. 2022). The straightforward idea for this task is to use the GPR scans to perform scan matching in the prior map like LiDAR (Cornick et al. 2016; Ort, Gilitschenski, and Rus 2020). Due to the limited observation range of a single GPR, some work attempts to combine IMU to improve the performance (Baikovitz et al. 2021b). Multiple weather scenarios can cause underground dielectric constants to change, which in turn changes echo characteristics. Some recent work has attempted to overcome this limitation through deep networks (Ort, Gilitschenski, and Rus 2023; Zhang et al. 2024), but it is not robust to interference noise. In order to meet the needs of real-time application, DEC was developed to compress the original echo sequence, and achieve better results than deep networks in local area place recognition (PR), but the DEC features lack uniqueness and have a high mismatch rate on large-scale maps (Li, Guo, and Song 2024).

Using GPR for PR presents several critical limitations. First, unlike the abundant features available for cameras and LiDAR, underground features are limited. Directly adapting existing methods to GPR often results in redundant parameters and convergence difficulties. Second, radar echoes are highly prone to interference, yet most deep learning models rely on CNNs with fixed weights. This uniform convolutional encoding amplifies the network's sensitivity to noise and interference. Third, variability in Dielectric Properties: Multi-temporal GPR data is influenced by fluctuations in underground dielectric constants, causing significant variability in extracted features. Current methods fail to effectively address the challenge of robust feature extraction under these changing conditions.

This paper proposes EDENet, a novel approach for generating compact sequential descriptors through synchronized feature encoding and filtering representation design. Specifically, we propose the incorporation of learnable Gabor filters to integrate geometric priors from GPR echoes, thereby constraining the feature encoding process of the network. Furthermore, the implementation of a direction-aware attention mechanism enables adaptive geometric encoding and feature filtration. Finally, by employing a multi-scale design and cascading existing feature aggregation layers, our approach achieves state-of-the-art performance.

In sum, the contribution of this article is summarized as follows:

- In the scenario of underground dielectric constant changes, the robust directional features in underground echoes are analyzed for the first time to guide the design of the network.
- A method based on learnable Gabor filter layers and direction-aware attention is proposed to adaptively encode robust directional features and filter out useless information. At the same time, EDEBlock is constructed by integrating offset-invariant units.
- Based on multi-scale EDEblock, a new network EDENet is proposed for GPR-based place recognition. Integration with the existing simple feature aggregation layer

*Corresponding authors

can achieve impressive performance. Extensive quantitative and qualitative experiments not only verify the performance, but also show that the EDENet has obvious advantages in model size and computing speed. Our codes are avaliable[1] https://github.com/Pyxel0524/EDE-LGPR.

Related Works

Underground Scene Representing

Handicraft Descriptors often combine near-field electromagnetic reflection characteristics to capture ripple features in GPR images. For example, the Dense Energy Curve (DEC) has been applied to GPR-based localization, but it lacks unique discriminative ability, leading to high false alarm rates in PR tasks (Li, Guo, and Song 2024). Some methods emphasize hyperbolic features of significant underground targets based on geometric relationships (Reichman, Collins, and Malof 2018; Wilson et al. 2007), but these are limited by GPR signal attenuation(Song and Jin 2024; Song et al. 2024). Log-Gabor filtering in the frequency domain has been explored for target detection (Zhang et al. 2021, 2023), yet handcrafted features generally require manual adjustment of filtering parameters, which is time-consuming and dataset-dependent.

Learning-based representation leverage deep networks for specific tasks. In underground target detection, CRNet (Sun, Cheng, and Fan 2022) uses a UNet-based structure to filter clutter and reconstruct features, while M2FNet employs multi-view learning for feature identification (Li et al. 2023). CMU-Net, based on ResNet, overcomes clutter by estimating transformations between GPR slices (Baikovitz et al. 2021b), and LGPRNet uses Inception modules to extract multi-scale features for robot position regression (Ort, Gilitschenski, and Rus 2023). In comparison, MWSNet developed based on Siamese-UNet shows better performance in overcoming location recognition tasks under different weather conditions (Zhang et al. 2024). However, the above CNN-based approaches excel in handling large datasets but are sensitive to redundant details from dielectric constant variations, leading to overfitting. Our approach aims to automatically filter unnecessary details and model directional responses at multiple scales for more robust global representation.

Place Recognition (PR)

Previous research in place recognition has predominantly relied on surface-based sensors, such as cameras and LiDAR. The strategy of compressing raw sensor observations into compact descriptors, as exemplified by methods like NetVLAD (Garg, Vankadari, and Milford 2022; Arandjelovic et al. 2016) for vision and OverlapNet (Chen et al. 2021) for LiDAR, has been widely adopted to accelerate localization. To better account for the continuity of robotic motion, recent advancements, including SeqVLAD and SeqOT (Mereu et al. 2022; Ma et al. 2022), have focused on incorporating sequential information (Yin et al. 2021), thereby improving upon their respective single-observation counterparts.

However, while these methods have demonstrated success in feature-rich surface environments, they face significant challenges in the feature-sparse underground settings. This highlights the need for a targeted analysis to identify effective features for underground environments.

Motivation

In this section, we analyze the key characteristics of GPR echoes, which motivate the design of our proposed method. GPR detects changes in the dielectric constant of subsurface materials by emitting electromagnetic waves. Due to the continuity of the underground medium, significant subsurface features often display directional geometric properties in the resulting echo sequences(Song and Jin 2024). Assuming a homogeneous dielectric constant ϵ_r and considering GPR movement along a trajectory distance x, beamwidth θ, and electromagnetic wave speed c, the GPR echoes can be approximated by a hyperbolic form, as shown in Eq.1.

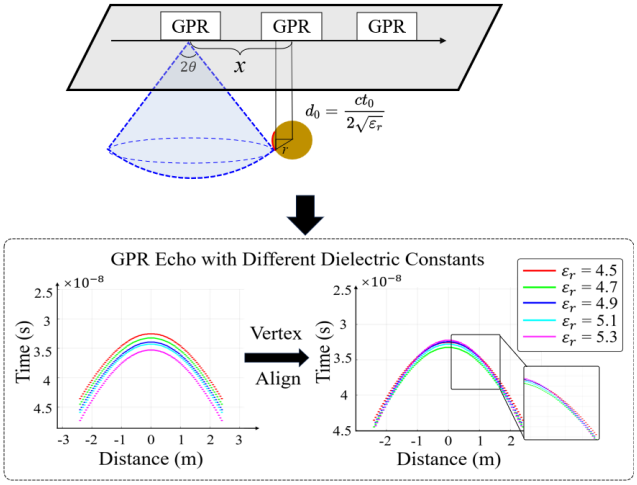

Figure 1: GPR echoes with different dielectric constants.

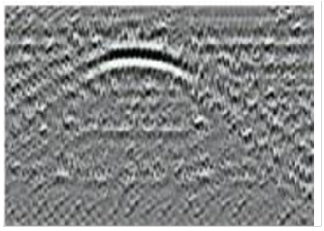

Figure 2: Curve features obscured by interference.

$$\begin{cases} \left[x - \left(\frac{c}{2\sqrt{\varepsilon_r}}\right)\sin\theta\right]^2 + \left[\left(\frac{c}{2\sqrt{\varepsilon_r}}\right)\cos\theta - (r+d_0)\right]^2 & \\ = r^2 & \frac{(d+r)}{cos\theta} < |x| < \frac{(d+2r)}{cos\theta} \\ \frac{\left(\frac{ct}{2\sqrt{\varepsilon_r}}+1\right)^2}{\left(\frac{d_0}{r}+1\right)^2} - \frac{x^2}{(d_0+r)^2} = 1 & others, \end{cases} \tag{1}$$

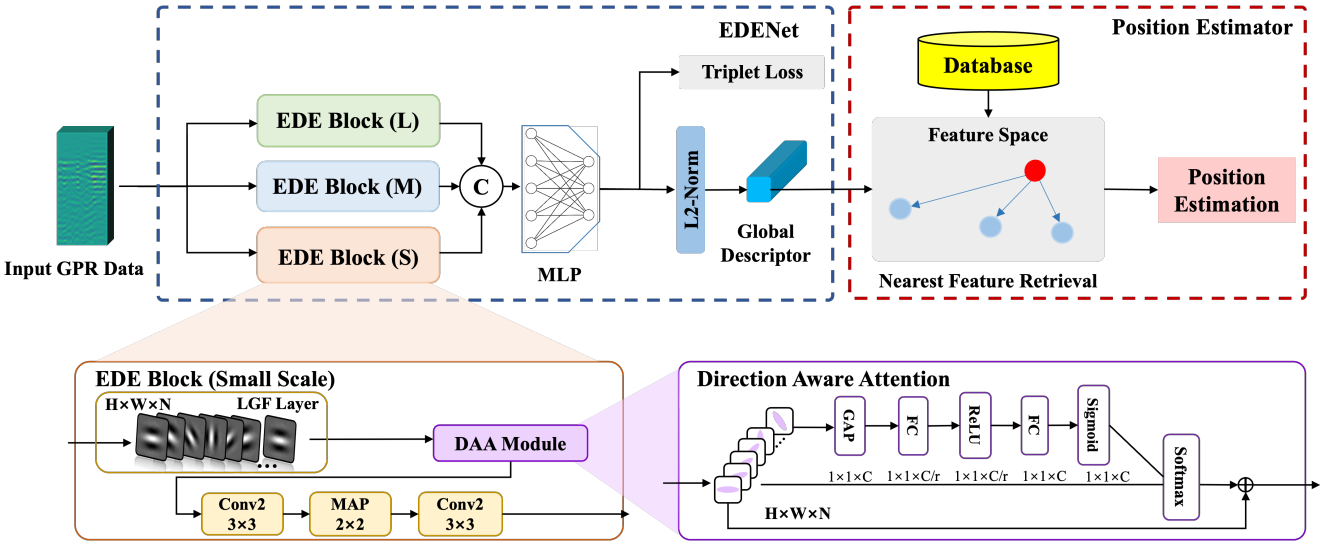

Figure 3: The proposed framework consists of two modules. In EDENet, the sequential GPR data is compressed to generate a compact descriptor. In the position estimator module, we use a vector retrieval scheme to find the best matching.

For radar echoes collected by the same GPR, ϵ_r is the main factor affecting the radar image. Variations in ϵ_r alter the propagation speed of the waves, leading to feature distortions and shifts along the depth axis. However, as shown in Fig. 1, when aligning echoes across different dielectric constants, the geometric features remain relatively consistent, while shifts vary. Therefore, capturing the directional information in the echoes is crucial for encoding the GPR sequence's features.

On the other hand, as illustrated in Fig. 2, electromagnetic interference and noise in extreme conditions introduce random perturbations that can obscure these directional features. Thus, a feature selection model that integrates feature encoding with noise filtering is essential for generating a compact and robust descriptor.

Methodology

Problem Definition

In this study, we address the challenge of PR based on underground sequence echoes. To achieve this goal, we define a sequence descriptor retrieval problem for prior map data $\mathcal{M} = \{(\mathbf{m}_i, \mathbf{T}_i), i = 1, 2, 3, ..., n\}$. The database formed by n GPR echo slices m_i and their associate UTM pose $\mathbf{T}_i = (\mathbf{utm}_x^i, \mathbf{utm}_y^i)$, where each slices $m \in \mathbb{R}^{D \times C}$, with D representing the dimension of depth bins and C representing the dimension of radar array channels, corresponding to the range and azimuth information of the radar echo, respectively. Specifically, $\mathbf{m}_{i,D,C}$ represents the intensity of the radar return in the ith frame at a specific depth and channel. Additionally, each frame m is associated with a pose that contains the pose of the RTK GPS. Next, we have a sequence of query frames $\{\mathbf{q}_i, \mathbf{q}_{i+1}, ..., \mathbf{q}_{i+n}\}$, measured when the vehicle revisits the map. Our goal is to find the optimal correspondence of $\{\mathbf{m}_i, \mathbf{m}_{i+1}, ..., \mathbf{m}_{i+n}\}$in $\mathcal{M}$ which provides the position at the time $\mathbf{m}$ was measured.

Architecture Overview

EDENet encodes directional information from radar echo sequences using multi-scale EDEblocks to construct a geometric representation. Each EDEblock consists of a series of learnable Gabor filters for capturing directional responses and a Directional Attention Module for adaptive reweighting, enhanced with a shift-invariant unit for dimension reduction. Multi-scale features finally are aggregated using an MLP at the output layer to generate compact global descriptor. This design allows EDENet to geometrically encode features while filtering out noise and irrelevant information effectively.

Echo Direction Encoding (EDE) Block

Learnable Gabor Filters (LGF) Layer To extract robust underground features, several convolution operators have been mentioned in the past in underground target detection and image processing tasks, including Gaussian Filter (Deng and Cahill 1993), Gabor filter (Mehrotra, Namuduri, and Ranganathan 1992), and Laden transform (Jong-sen Lee and I. Jurkevich 1990). Inspired by the application of Gabor features in geology and GPR-based tasks (Wang and Alkhalifah 2024; Cheng et al. 2024), this paper proposes a learnable Gabor ($\mathcal{L}G$) filter and constructs a $\mathcal{L}G$-based layer (LGF). The LGF is defined as follows:

$$\mathcal{L}G\left(i, j; \lambda, \gamma, \phi, \theta, \sigma\right) = \\ -exp\left\{-\frac{i'^2\gamma^2 + j'^2}{2\sigma^2}\right\} cos\left(2\pi\frac{i'}{\lambda} + \phi\right), \quad (2)$$

the pixel indices are denoted by i and j, and $i' = icos\theta + jsin\theta$, $j' = -isin\theta + jcos\theta$. λ represents the wavelength of the sinusoidal wave component, γ defines the ellipticity of the Gaussian support of the Gabor filter, ϕ is the phase shift, and θ and σ determines the direction and the standard

deviation of the basic wave. All these parameters control the shape of filter.

We further remove the zero DC component by subtracting the mean, which allows the Gabor filter to approximate the compact support property (Harris, Ho, and Zare 2016):

$$\begin{aligned}\mathcal{F}(i,j,K) = \mathcal{L}G\left(i-\frac{K}{2}, j-\frac{K}{2}\right) \\ -\frac{1}{K^2}\sum_{x=0}^{K}\sum_{y=0}^{K}\mathcal{L}G\left(i-\frac{K}{2}, j-\frac{K}{2}\right),\end{aligned} \tag{3}$$

As illustrated in Eq. 2, when λ tends to ∞ and ϕ approaches 0, the learnable Gabor filter transforms into a Gaussian filter kernel. When λ tends to ∞ and γ and ϕ approach 0, the learnable kernel becomes a edge fitting filter, which is used for edge direction fitting in radar images (Jong-sen Lee and I. Jurkevich 1990). Thus, the learnable Gabor filter can represent various types of filters. It is capable of extracting valuable directional responses and can also act as a filter to remove irrelevant information and clutter from the image.

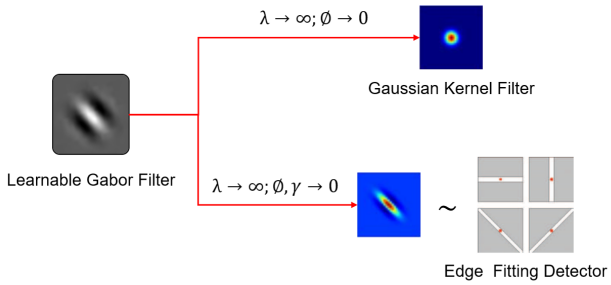

Figure 4: The learnable Gabor filter.

Analyze: The LGF module is designed to extract multi-direction responses. Experiments have found that overly freedom parameter learning will overfitting. Such as setting θ as a learnable parameter will cause the network to focus only on specific main directions. Therefore, in this paper, the LGF module is defined as composed of a set of LG $\{\mathcal{F}(i,j,K;\theta_k)|k=1,2,3,...,n\}$, the k-th θ is defined as $\theta_k = k\cdot\pi/n$, each LGF module shares the same σ and receptive field K. This ensures the diversity and flexibility of the LG modality.

Direction Aware Attention (DAA) Our goal is to ensure that the network can improve sensitivity to specific directional filtering and suppress less useful filtering kernels. To achieve this goal, we suggest modeling the interdependence of kernels $\mathcal{F}$, the weights of filters in different directions are recalibrated through two steps, first by modeling the interdependence of direction kernels and then by incentivizing. In the phase of modeling, the output of LGF module $\mathbf{V} = \sum_{k=1}^{n}\mathcal{F}_k * \mathbf{X}$ as a local spatial representation of GPR sequential data $\mathbf{X}$, is squeezed to a direction-wise descriptor $\mathbf{d}$ by global average pooling (GAP). This is prevalent in feature engineering to model multiple output features, which can be replaced by more complex aggregation operations (Hu, Shen, and Sun 2018). To be specific, the n-th element of $\mathbf{d}\in\mathbb{R}^n$, is defined by:

$$d_n = \frac{1}{K^2}\sum_{i=1}^{K}\sum_{j=1}^{K}V_n(i,j), \tag{4}$$

furthermore, to utilize the local descriptors of group features, we use a simple gating mechanism with sigmoid activation and two parameterized linear layers for dimensionality reduction and enhancement. This is an effective and commonly used operation in feature engineering to model inter-channel correlations (Hu, Shen, and Sun 2018)

$$\mathbf{T} = sigmoid\left(\mathbf{F}_2\delta\left(\mathbf{F}_1\mathbf{d}\right)\right), \tag{5}$$

where, the σ is ReLU function, $\mathbf{F}_1\in\mathbb{R}^{\frac{N}{r}\times N}$, $\mathbf{F}_2\in\mathbb{R}^{N\times\frac{N}{r}}$, the specific selection of r according to (Hu, Shen, and Sun 2018)

To ensure that multiple directional responses are allowed to be emphasized instead of being one-hot activated. In this paper, we add the input of DAA back to the output through skip connections. This provides a shortcut and further improves performance (Huang et al. 2017). The output of the DAA module is obtained by recalibrating the $\mathbf{V}_k$:

$$\widetilde{\mathbf{V}} = \mathbf{T}\left(\mathbf{V},\{\mathbf{F}_1,\mathbf{F}_2\}\right) + \mathbf{V}. \tag{6}$$

Analyze: The DAA module acts as weights adapted to the filter bank, which is different from the idea used for underground object detection (Zhang et al. 2023), so that the design is not just about preserving the index with the maximum response. The DAA module intrinsically emphasizes the main directional features and encourages other kernels to generate other types of filters to remove noise/clutter.

Shift Invariant Unit The above analysis indicates that changes in the underground dielectric constant can cause slight changes in the characteristic positions of significant targets underground. This module designs a simple CNN cascaded max pooling scheme to further boost the specificity of global features $\widetilde{\mathbf{V}}$ to $\eta = \mathbf{W}_2 * pool\left(\mathbf{W}_1 * \widetilde{\mathbf{V}} + b_1\right) + b_2$, while maintaining a certain degree of feature shift invariance, demonstrated in previous studies (Leterme et al. 2022).

Feature Aggregation Layer

In the feature aggregation layer, we convert the output of parallel multi-scale EDEBlocks into a high-level compact descriptor η using an MLP. The underlying idea of feature aggregation is to shift the focus from local features and cluster center extraction to leveraging the fully connected layers' capacity to holistically combine features (Ali-Bey, Chaib-Draa, and Giguere 2023). In this setup, each neuron within the feature aggregation layer can perceive image inputs at different scales, thereby possessing a comprehensive receptive field.

$$\eta = concat\left(\eta_1,\eta_2,...,\eta_i\right), \tag{7}$$

$$\mathbf{O} = \sigma\left(\mathbf{W}*\eta+\mathbf{b}\right), \tag{8}$$

where $\mathbf{W}\in\mathbb{R}^{n\times d}$, n is the dimension of high-level desciptor η, and d is a manually selected dimension. The final

Method/Chanllenge	GROUNDED			CMU-GPR	
	Sunny/Sunny	Sunny/Rainy	Sunny/Snowy	Nsh_b	Gate_g
3D ResNet18	0.94/0.95/0.96	0.49/0.61/0.67	0.46/0.55/0.60	0.46/0.61/0.69	0.70/0.89/0.93
3D ResNet50	0.93/0.95/0.96	0.50/0.64/0.71	0.47/0.56/0.62	0.53/0.64/0.72	0.72/0.91/0.95
3D ResNet101	0.93/0.94/0.95	0.48/0.62/0.70	0.45/0.56/0.62	0.48/0.64/0.70	0.69/0.87/0.95
CMUNet	0.94/0.95/0.96	0.50/0.61/0.69	0.46/0.55/0.61	0.43/0.69/0.76	0.75/0.93/0.97
DEC	0.40/0.47/0.51	0.21/0.31/0.41	0.27/0.37/0.46	0.22/0.48/0.54	0.23/0.33/0.40
MWSNet	0.92/0.93/0.94	0.48/0.55/0.63	0.43/0.54/0.67	0.57/0.69/0.76	0.81/0.94/0.95
SeqVLAD	0.67/0.86/0.90	0.32/0.43/0.50	0.28/0.40/0.47	0.23/0.27/0.29	0.36/0.55/0.65
SeqOT	**0.93/0.95/0.96**	0.49/0.58/0.64	0.40/0.44/0.50	0.36/0.54/0.65	0.51/0.81/0.86
LGPRNet	0.89/0.92/0.94	0.54/0.60/0.67	0.47/0.57/0.66	0.56/0.65/0.70	0.75/0.87/0.90
EDENet	**0.93/0.95/0.96**	**0.71/0.81/0.85**	**0.57/0.66/0.70**	**0.81/0.86/0.89**	**0.95/0.99/0.99**

Table 1: Comparison to SOTA methods on benchmark datasets. The best is highlighted in bold (recall@1/5/10).

output $\mathbf{O}$ of the network is flattened and processed by L2-normalized, which is consistent with the other PR (Ali-Bey, Chaib-Draa, and Giguere 2023).

Loss Function

During the training phase, we follow the common practice of PR works (Arandjelovic and Zisserman 2013; Cao, Araujo, and Sim 2020) to find the nearest reference image for each query in the embedded space as the final positive sample Samples with actual geographic locations exceeding 3 meters are considered negative.

We represent the global embedding features of queries and positive and negative samples as F_q, F_p and F_n, and the network is trained using max margin triplet loss:

$$\mathcal{L}(F_q, \{F_p\}, \{F_n\}) = N_p \left(\delta + max\left(d\left(F_q, F_p\right)\right)\right) - \sum_{N_n} \left(d\left(F_q, F_n\right)\right). \quad (9)$$

Experiment

In this section, we demonstrate our approach through several sets of experiments as follows:

- Comparison with state-of-the-art techniques for GPR based methods and representative techniques for sequence localization using camera and LiDAR, which is our main purpose.
- Ablation experiments to demonstrate the effectiveness of each module and combination of the network.
- Evaluation of the model's computational efficiency and number of parameters, which is important for whether the model can be used in actual deployment.

All the above experiments were validated on the 3D multi-channel public dataset GROUNDED (Ort, Gilitschenski, and Rus 2023) and the 2D single channel CMU-GPR Datasets (Baikovitz et al. 2021a), demonstrating the performance of our proposed method.

Implement Detail

All experiments were conducted on a system equipped with an Intel i7-11700K CPU and an Nvidia RTX 3090 GPU.

	GROUNDED	CMU-GPR
Replacing Gabor with conv	0.63/0.68	0.71/0.78
DAA module (w/o)	0.62/0.70	0.74/0.80
Learnable Gabor + DAA	**0.71/0.76**	**0.75/0.82**

Table 2: Ablation results of learnable Gabor filter and DAA

Network Training and Testing For the GROUNDED, $run\_0008, 0046, 0047, 0080$ are used for training, while $run\_0054, 0056, 0057, 0090$ served as validation and test sets. In the CMU-GPR, Nsh_b and Gate_g were used for cross-validation to prevent data leakage. The training and validation sets were split with a 7:3 ratio, optimize with Adam(Kingma and Ba 2014) and learning rate to 0.0001.

Parameter Setting In the GROUNDED dataset, model parameters and descriptor dimensions were aligned with the settings of SeqVLAD (Mereu et al. 2022). In the CMU-GPR dataset, all descriptor generation dimensions were set to 400 to maintain consistency with the handcrafted DEC features.

Performance of EDENet w.r.t. State-of-the-art

The first experiment compares the proposed method against existing techniques. Given the scarcity of solutions for GPR-based localization, ResNet series methods are used as baselines, with state-of-the-art (SOTA) methods for cameras and LiDAR also included for comparison. These alternative sensor methods are retrained on the GPR dataset as outlined in original papers (Mereu et al. 2022; Ma et al. 2022).

In the GROUNDED dataset, SeqVLAD and SeqOT are selected as SOTA solutions for camera/LiDAR, with feature descriptor parameters set according to (Mereu et al. 2022; Ma et al. 2022). For the CMU-GPR dataset, CMUNet, MWSNet and DEC (all designed only for single-channel GPR data) are added as comparatives.

Table 1 demonstrates that the proposed method outperforms others across different scenarios in both datasets. The DEC method, which describes energy changes in underground scenes, lacks discriminative power. Camera and LiDAR-based methods like SeqVLAD and SeqOT, which

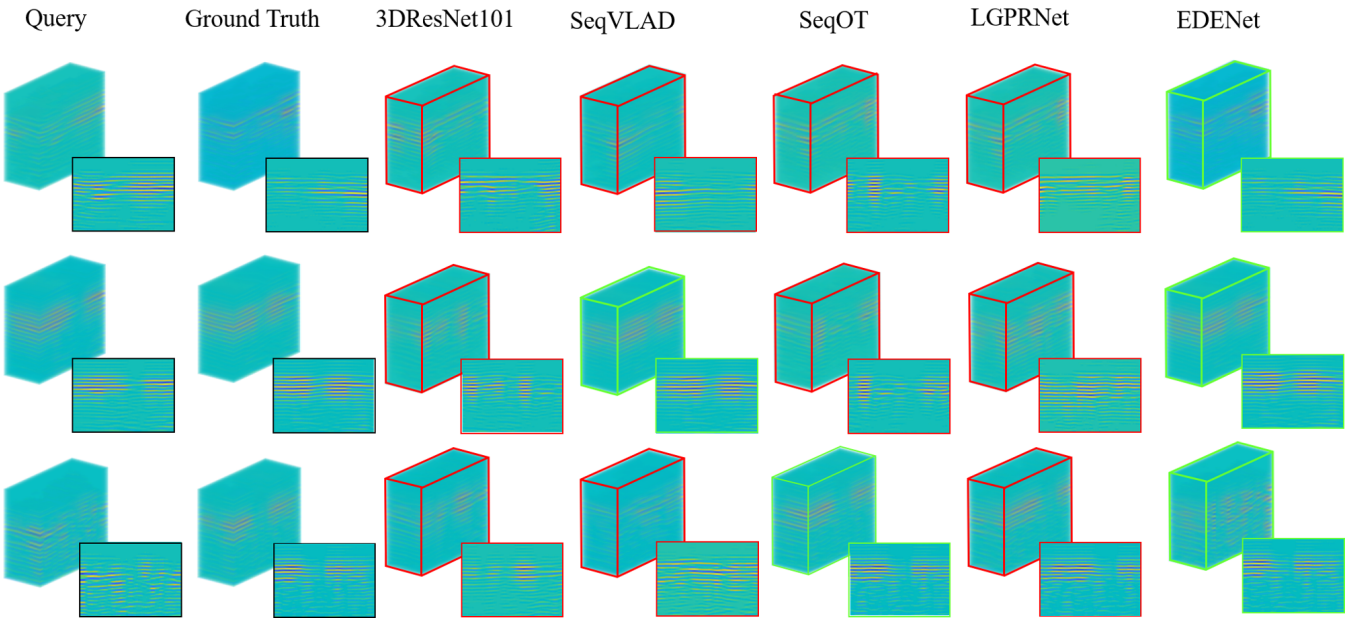

Figure 5: Qualitative results of GROUNDED (Sunny/Rainy). Green and red boxes in indicate correct and wrong match.

depend on rich detail features, tend to overfit and exhibit poor generalization on GPR sequence data due to high feature texture repetition. Additionally, clustering in the feature aggregation stage may overemphasize irrelevant information, limiting performance. CMUNet, MWSNet, and LGPRNet, based on ResNet, U-Net, and InceptionNetwork respectively, excel at capturing multi-scale detail textures but are hindered by an excessive focus on new information introduced by changes in underground dielectric constants, as initially analyzed, leading to model bottlenecks.

Ablation experiment

EDEBlock The primary innovation presented in this paper is the EDEBlock. We conducted ablation experiments on the LGF and DAA modules within EDEBlock using the GROUNDED and CMU-GPR datasets, measuring performance in terms of recall@1 and recall@5. As shown in Table 3, substituting the learnable Gabor filter with a standard convolutional kernel resulted in an average performance drop of 6%, highlighting the importance of the LGF in capturing directional features. Similarly, the DAA module, which is designed as a plug-and-play component, demonstrated its critical role in selectively enhancing the output of the learnable Gabor filter. Its removal led to a 4% drop in performance, underscoring the necessity of adaptive attention in maintaining the integrity of the learned features. The combination of LGF and DAA achieved the best results, validating the proposed method for robust underground feature representation through the main directional description.

The qualitative results and dataset comparisons in Fig. 5 and Fig. 6 further demonstrate the robustness of our method against existing approaches. This is particularly evident in the GROUNDED dataset, where varying weather conditions cause fluctuations in the underground dielectric constant, leading to changes and shifts in characteristic amplitude. Despite these challenges, our proposed scheme effectively captures echo texture variations and accurately identifies corresponding locations

Combination and Parameter of EDENet The scale of EDEBlocks is crucial for capturing the diverse spatial features in GPR data. By employing kernels of different sizes—35×35, 23×23, 11×11, and 5×5—EDENet effectively captures both broad and fine details. Larger kernels, like 35×35, are proficient at capturing low-frequency, global information essential for understanding the overall underground structure. In contrast, smaller kernels, such as 5×5, focus on high-frequency, local details, which are vital for detecting fine-grained features.

The combination of these scales allows EDENet to comprehensively encode GPR data by leveraging the strengths of both large and small kernels. However, introducing additional scales can lead to performance saturation, where the benefit of capturing more information diminishes due to increased redundancy. The learnable Gabor filters, which form the core of EDEBlocks, are not orthogonal and can overlap in the feature space, leading to redundant information that may degrade performance through overfitting.

The parameter k directly impacts the angular resolution of the Gabor filters within each EDEBlock. A larger k increases angular resolution, enabling the Gabor filters to model a broader range of orientations, which is advantageous for GPR-based place recognition as it enhances the model's ability to differentiate between locations. However, a higher k also adds complexity to the filter bank, increasing the risk of redundancy by producing filters that capture similar features without adding new information.

This analysis underscores the importance of carefully bal-

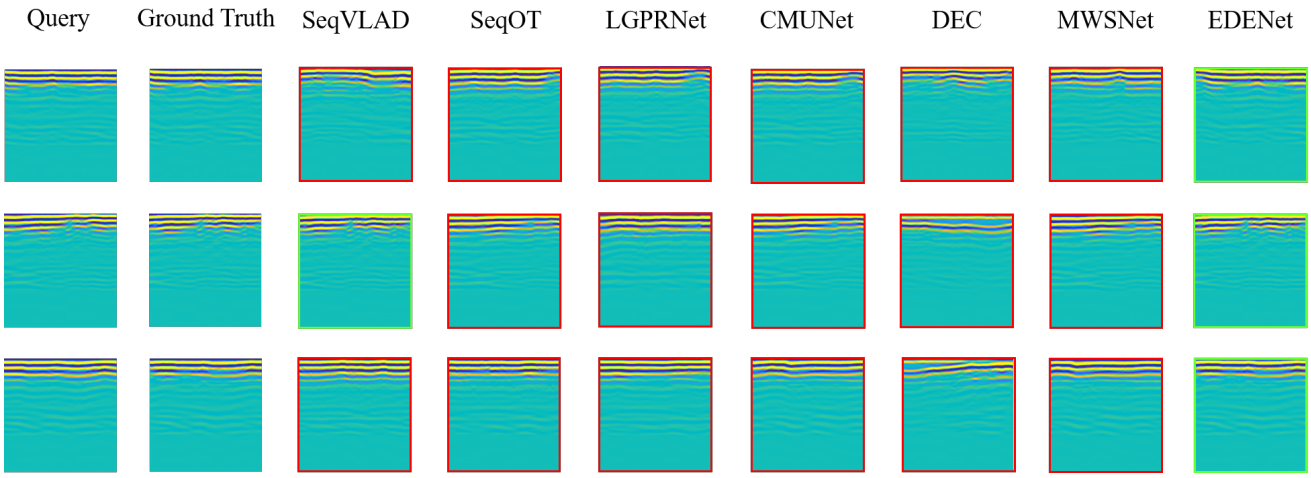

Figure 6: Qualitative results of CMU-GPR. Green and red boxes in indicate correct and wrong match.

EDE-35	EDE-23	EDE-11	EDE-5	$k = 8$	$k = 16$	$k = 32$	$k = 64$	$k = 96$
✓	✓	✓	✓	69.3	72.4	71.5	**76.0**	73.2
✓		✓	✓	65.6	72.1	72.2	74.6	72.3
✓	✓	✓		58.7	62.0	72.6	73.2	70.4
	✓	✓	✓	69.1	71.0	72.1	68.6	72.0
✓			✓	67.2	67.0	70.1	72.6	72.0
✓		✓		66.3	65.1	68.7	70.1	71.2

Table 3: Recall @5 Recognition Rate (%) with Different Scale Combinations and hyperparameters k

ancing the configuration of EDEBlocks in EDENet. The multi-scale kernel combination is essential for capturing the full spectrum of GPR features, but must be chosen to avoid performance saturation. Likewise, the parameter k should be optimized to enhance angular resolution while minimizing redundancy. The selected configuration of 35×35, 11×11, and 5×5 EDEBlocks with $k = 64$ was empirically validated as the optimal setup, offering a balanced and effective approach to GPR-based place recognition.

Performance and Runtime Analysis As shown in Table 4, the fourth experiment reports the average runtime and model size of all methods across ten trials. Our method is the most compact in model size while maintaining competitive performance in parameter count and runtime. Notably, it does not require specialized functions for sequence descriptor similarity evaluation, enabling acceleration with the Faiss library (Douze et al. 2024). This results in a runtime of 188 Hz with a 100-frame GPR sequence, surpassing the 126 Hz GPR data acquisition rate and demonstrating the potential for real-time deployment. In contrast, other methods with the same descriptor size require more processing time due to the excessive parameters in backbones designed for Camera/LiDAR data, reducing efficiency.

Conclusion

This paper proposes a novel neural network, EDENet, designed to generate compact descriptors based on GPR sequence echoes for PR tasks. EDENet introduces a learnable Gabor filter layer and a direction-aware attention mechanism to effectively integrate echo geometric feature encoding while removing redundant information. The current work primarily focuses on the exploration of the backbone network, with a significant concentration of parameters in the final feature aggregation layer. Future work will continue to investigate the design of the feature aggregation layer for positioning based on GPR sequence echoes, aiming to achieve a more efficient and lightweight solution.

	Parameters	Model Size	Runtimes
3DResNet18	1.6×10^10	846Mb	0.006s
3DResNet50	1.71*10^10	1.24G	0.008s
3DResNet101	3.54*10^10	1.63G	0.014s
SeqVLAD	2.06*10^9	**786Mb**	**0.004s**
SeqOT	6.21*10^8	831Mb	0.016s
LGPRNet	3.01*10^9	1.5G	0.011s
EDENet	**5.0*10^8**	828Mb	0.005s

Table 4: Runtime and Model Size

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62303475 and No. 62401576) and the Hunan Province Innovation and Entrepreneurship Project (No. XJZH2024009 and No. XJZH2024010)

References

Ali-Bey, A.; Chaib-Draa, B.; and Giguere, P. 2023. Mixvpr: Feature mixing for visual place recognition. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2998–3007.

Arandjelovic, R.; Gronat, P.; Torii, A.; Pajdla, T.; and Sivic, J. 2016. NetVLAD: CNN architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5297–5307.

Arandjelovic, R.; and Zisserman, A. 2013. All about VLAD. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1578–1585.

Baikovitz, A.; Sodhi, P.; Dille, M.; and Kaess, M. 2021a. Cmu-gpr dataset: Ground penetrating radar dataset for robot localization and mapping. *arXiv preprint arXiv:2107.07606*.

Baikovitz, A.; Sodhi, P.; Dille, M.; and Kaess, M. 2021b. Ground encoding: Learned factor graph-based models for localizing ground penetrating radar. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5476–5483. ISBN 1-66541-714-5.

Cao, B.; Araujo, A.; and Sim, J. 2020. Unifying deep local and global features for image search. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16*, 726–743. ISBN 3-030-58564-6.

Chen, X.; Läbe, T.; Milioto, A.; Röhling, T.; Vysotska, O.; Haag, A.; Behley, J.; and Stachniss, C. 2021. OverlapNet: Loop closing for LiDAR-based SLAM. *arXiv preprint arXiv:2105.11344*.

Cheng, Q.; Cui, F.; Dong, G.; Wang, R.; and Li, S. 2024. A Method of Reconstructing Ground Penetrating Radar Bscan for Advanced Detection Based on High-Order Synchrosqueezing Transform.

Cornick, M.; Koechling, J.; Stanley, B.; and Zhang, B. 2016. Localizing ground penetrating radar: A step toward robust autonomous ground vehicle localization. *Journal of field robotics*, 33(1): 82–102.

Deng, G.; and Cahill, L. 1993. An adaptive Gaussian filter for noise reduction and edge detection. In *1993 IEEE conference record nuclear science symposium and medical imaging conference*, 1615–1619. ISBN 0-7803-1487-5.

Douze, M.; Guzhva, A.; Deng, C.; Johnson, J.; Szilvasy, G.; Mazaré, P.-E.; Lomeli, M.; Hosseini, L.; and Jégou, H. 2024. The faiss library. *arXiv preprint arXiv:2401.08281*.

Garg, S.; Vankadari, M.; and Milford, M. 2022. Seqmatchnet: Contrastive learning with sequence matching for place recognition & relocalization. In *Conference on Robot Learning*, 429–443. ISBN 2640-3498.

Harris, S.; Ho, K.; and Zare, A. 2016. On the use of log-gabor features for subsurface object detection using ground penetrating radar. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXI*, volume 9823, 161–169.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Jong-sen Lee; and I. Jurkevich. 1990. Coastline Detection And Tracing In SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4): 662–668.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kouros, G.; Kotavelis, I.; Skartados, E.; Giakoumis, D.; Tzovaras, D.; Simi, A.; and Manacorda, G. 2018. 3d underground mapping with a mobile robot and a gpr antenna. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3218–3224. ISBN 1-5386-8094-7.

Leterme, H.; Polisano, K.; Perrier, V.; and Alahari, K. 2022. On the shift invariance of max pooling feature maps in convolutional neural networks. *arXiv preprint arXiv:2209.11740*.

Li, H.; Guo, J.; and Song, D. 2024. Subsurface Feature-based Ground Robot/Vehicle Localization Using a Ground Penetrating Radar. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 1716–1722.

Li, N.; Wu, R.; Li, H.; Wang, H.; Gui, Z.; and Song, D. 2023. M 2 FNet: Multi-modal Fusion Network for Airport Runway Subsurface Defect Detection using GPR Data. *IEEE Transactions on Geoscience and Remote Sensing*.

Ma, J.; Chen, X.; Xu, J.; and Xiong, G. 2022. SeqOT: A spatial–temporal transformer network for place recognition using sequential LiDAR data. *IEEE Transactions on Industrial Electronics*, 70(8): 8225–8234.

Mehrotra, R.; Namuduri, K. R.; and Ranganathan, N. 1992. Gabor filter-based edge detection. *Pattern recognition*, 25(12): 1479–1494.

Mereu, R.; Trivigno, G.; Berton, G.; Masone, C.; and Caputo, B. 2022. Learning sequential descriptors for sequence-based visual place recognition. *IEEE Robotics and Automation Letters*, 7(4): 10383–10390.

Ort, T.; Gilitschenski, I.; and Rus, D. 2020. Autonomous navigation in inclement weather based on a localizing ground penetrating radar. *IEEE Robotics and Automation Letters*, 5(2): 3267–3274.

Ort, T.; Gilitschenski, I.; and Rus, D. 2023. GROUNDED: A localizing ground penetrating radar evaluation dataset for learning to localize in inclement weather. *The International Journal of Robotics Research*, 42(10): 901–916.

Reichman, D.; Collins, L. M.; and Malof, J. M. 2018. gprHOG and the popularity of histogram of oriented gradients (HOG) for buried threat detection in ground-penetrating radar. *arXiv preprint arXiv:1806.01349*.

Song, S.; Dai, Y.; Sun, S.; and Jin, T. 2024. Efficient Image Reconstruction Methods Based on Structured Sparsity for Short-Range Radar. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–15.

Song, Y.; and Jin, T. 2024. Efficient Near-Field Radar Microwave Imaging Based on Joint Constraints of Low-Rank

and Structured Sparsity at Low SNR. *IEEE Transactions on Microwave Theory and Techniques*, 1–16.

Sun, H.-H.; Cheng, W.; and Fan, Z. 2022. Learning to remove clutter in real-world GPR images using hybrid data. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–14.

Wang, F.; and Alkhalifah, T. 2024. Learnable Gabor kernels in convolutional neural networks for seismic interpretation tasks.

Wilson, J. N.; Gader, P.; Lee, W.-H.; Frigui, H.; and Ho, K. 2007. A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination. *IEEE Transactions on Geoscience and Remote Sensing*, 45(8): 2560–2572.

Yin, P.; Wang, F.; Egorov, A.; Hou, J.; Jia, Z.; and Han, J. 2021. Fast sequence-matching enhanced viewpoint-invariant 3-d place recognition. *IEEE Transactions on Industrial Electronics*, 69(2): 2127–2135.

Zhang, P.; Shen, L.; Chen, Y.; Huang, X.; and Xin, Q. 2023. A Novel Feature Descriptor for Hyperbola Recognition in GPR Images Based on Symmetry Model. *IEEE Geoscience and Remote Sensing Letters*.

Zhang, P.; Shen, L.; Huang, X.; and Xin, Q. 2022. A new registration method with improved phase congruency for application to GPR images. *Remote Sensing Letters*, 13(7): 726–737.

Zhang, P.; Shen, L.; Wen, T.; Huang, X.; and Xin, Q. 2021. Vector phase symmetry for stable hyperbola detection in ground-penetrating radar images. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–12.

Zhang, P.; Zhi, S.; Yuan, Y.; Bi, B.; Xin, Q.; Huang, X.; and Shen, L. 2024. Looking Beneath More: A Sequence-based Localizing Ground Penetrating Radar Framework. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 8515–8521. ISBN 9798350384574.