Learning Surface Terrain Classifications from Ground Penetrating Radar

Anja Sheppard, Jason Brown, Nilton Renno, and Katherine A. Skinner

The University of Michigan

2505 Hayward St, Ann Arbor, MI 48109

anjashep@umich.edu, jaybrow@umich.edu, nrenno@umich.edu, kskin@umich.edu

Abstract

Terrain classification is an important problem for mobile robots operating in extreme environments as it can aid downstream tasks such as autonomous navigation and planning. While RGB cameras are widely used for terrain identification, vision-based methods can suffer due to poor lighting conditions and occlusions. In this paper, we propose the novel use of Ground Penetrating Radar (GPR) for terrain characterization for mobile robot platforms. Our approach leverages machine learning for surface terrain classification from GPR data. We collect a new dataset consisting of four different terrain types, and present qualitative and quantitative results. Our results demonstrate that classification networks can learn terrain categories from GPR signals. Additionally, we integrate our GPR-based classification approach into a multimodal semantic mapping framework to demonstrate a practical use case of GPR for surface terrain classification on mobile robots. Overall, this work extends the usability of GPR sensors deployed on robots to enable terrain classification in addition to GPR's existing scientific use cases.

1. Introduction

The Ground Penetrating Radar (GPR) is an instrument with a long history in non-intrusive surveying of sub-surface features with applications such as transportation infrastructure mapping [22], cemetery grave localization [10], and geologic feature identification [7]. The reflected signal returned from the radar pulse, shown in Fig. 1, encodes information about the velocity of electromagnetic waves through materials below the sensor. Recent literature has shown use cases for GPR to characterize surface properties, such as soil water content [35]. Additionally, GPR has been gaining interest in the robotics community for light-invariant autonomous vehicle localization [5].

Terrain classification is a challenging problem for mobile robots in extreme environments. One challenge is that training RGB segmentation networks generally requires a lot of
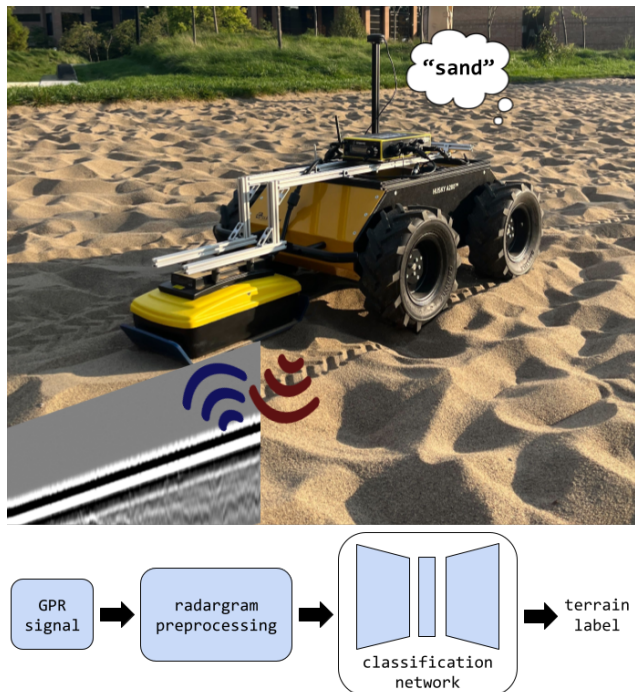

Figure 1. Clearpath Husky robot with a GPR mounted on the back. The radargram, which represents the permittivity of the surface and subsurface materials sensed by the GPR as a time series, is shown in image form in the bottom left.

hand-labeled, diverse data. Additionally, camera-based approaches can suffer when terrain is occluded or the lighting is poor [24]. This has led to an interest in alternative sensor modalities for classifying terrain, such as LiDAR [29], Inertial Measurement Units (IMUs) [11], torque sensors [3], and microphones [38].

As GPR sensors are already being deployed on mobile robots in remote environments to accomplish science tasks – such as the RIMFAX [13] radar on the Perseverance rover – we seek to determine whether this sensor can be leveraged to aid terrain classification. This paper seeks to answer the following question: Can GPR distinguish *surface* terrains?

In our quest to answer this, we present the following contributions:

(i) We investigate the use of GPR direct waves for learning-based surface terrain classification.

(ii) We conduct field experiments with a custom robot platform to collect a real dataset containing overlapping RGBD images, GPR radargrams, and GPS coordinates.

(iii) We develop a framework that takes input GPR data and outputs terrain classifications, and we present qualitative and quantitative results to evaluate this framework.

(iv) We integrate GPR terrain classification into a multimodal framework for terrain mapping to motivate future work.

The remainder of this paper is organized as follows: we first discuss standard uses for the GPR and prior work with other sensing modalities for terrain classification in Sec. 2. In Sec. 3, we present our formulation of GPR-based classification. In Sec. 4, we share details about our robot data collection platform and we demonstrate experimental results on our GPR terrain classification networks. Additionally, we present preliminary results of a multimodal terrain mapping framework that fuses GPR and RGB data. Finally, in Sec. 5, we discuss future work motivated by this novel application of GPR.

2. Background

2.1. Terrain Classification

Terrain classification is an important and well-studied problem for autonomous robots operating in extreme environments, such as the Mars rovers [25]. Prior work on terrain classification has focused primarily on semantic segmentation for RGB and depth images. However, some recent works also demonstrate novel uses of unconventional sensors in multimodal frameworks for learning terrain.

2.1.1 RGB Camera and Depth Sensor Approaches

Exteroceptive sensors, such as cameras and range sensors, provide context about the robot's environment from a distance. The Soil Property and Object Classification (SPOC) network proposed by NASA [1] presents a lightweight, vision-based approach to the terrain classification problem for Mars rovers. Another approach aims to provide more fine-grained labels of RGB rover images that are useful for science-based navigation goals by focusing particularly on terrain texture [27]. In order to extend advances in Martian terrain classification, [33] uses unsupervised contrastive pretraining to improve generalization to sites on Mars that were not present in the image training set.

LiDAR, while typically a sensor used to provide geo-metric information about a scene, has also proven useful for classifying semantics, as demonstrated in [29]. However, LiDARs are resource-intensive, making them difficult to include on planetary rovers or other resource-constrained mobile robots.

Vision-only approaches to terrain classification suffer from several sensor-specific weaknesses: lighting, weather, and the potential for obstacles to be obscured can cause unreliable results in the real world. Additionally, vision-based segmentation models often require tedious pixel-wise labeling. The AI4MARS dataset released by NASA's Jet Propulsion Laboratory [32] relied on crowdsourcing volunteers to label over 35,000 images. Multimodal approaches can help compensate for weaknesses in unimodal RGB-based terrain classification and can also potentially reduce the burden of manual image labeling.

2.1.2 Multimodal Approaches

Some works have leveraged additional sensor modalities to supervise the training of vision-based terrain classification networks. One foundational study uses a traversability metric calculated from wheel torque sensors to supervise the labeling of images [3]. Similarly, [38] presents a multimodal model that learns a deep clustering encoder for audio data of the wheel/terrain interaction and then produces pixel-wise semantic labels for the corresponding forward-facing camera images.

Recent approaches have also focused on sensor fusion within a multimodal network architecture, such as [11], which fuses visual and inertial data via a shared embedding space to detect both known and unknown terrain types. Additionally, [16] proposes the use of Multimodal Variational Autoencoders (VAEs) and Gaussian Mixture Models to extract features from audio and visual data into a shared embedding space. This approach allows for a more robust framework that still performs well when one of the modalities is missing.

In this work, we aim to add GPR to the growing list of sensors that can classify terrain. We also hope to demonstrate its potential in compensating for weaknesses of unimodal RGB approaches through a multimodal terrain mapping framework.

2.2. Ground Penetrating Radar

Over the past few decades, autonomous robots with GPR have been deployed in remote environments for various tasks including subsurface infrastructure reconstruction [9], crevasse mapping for safe traversal [34], and water detection on Mars [13, 15].

GPR is typically used to investigate subsurface features [28]. In recent years, some works have explored radar-based property mapping, such as GPR-mounted drones for map-

ping soil moisture [35], identifying the extent of subway deterioration [6], and reconstructing underground tree roots beneath streets [23]. Due to the generally static nature of subsurface features, GPR has also demonstrated usefulness in aiding loop closure for autonomous vehicle localization [5]. Datasets recently released by Carnegie Mellon University [2] and the Massachusetts Institute of Technology [26] are explicitly aimed at accelerating research in the use of GPR for spatio-temporal mapping.

To the best of our knowledge, no one has extended the usage of GPR to characterizing surface terrain properties for the purpose of terrain classification. A unique property of GPR called the direct wave (discussed in Sec. 3.1) allows this sensor to be used for surface characterization in addition to scientific objectives. The nature of the GPR sensor model makes classification based on this metric light-agnostic, allowing it to be a suitable complement for vision-based terrain identification. In this paper, we begin to unlock a new dimension of GPR use cases for robotic navigation in extreme environments by demonstrating surface terrain classification with GPR data.

3. Technical Approach

Within this section, we detail the key technical components of our approach to GPR terrain classification. Figure 1 shows an overview of our proposed framework. Our framework takes GPR data as input and performs radargram pre-processing. The preprocessed data is passed into a neural network to output a terrain label. Each component of this framework is detailed further below.

3.1. GPR Signal Characterization

GPR uses radio waves to probe subsurface composition in low loss dielectric material areas. The complex permittivity, ε, expresses how well a material responds to electric flux [4]:

$$\varepsilon = \varepsilon' - i \cdot \varepsilon'' \tag{1}$$

The real component, ε', quantifies the stored energy, and the imaginary component, ε'', quantifies the dielectric loss of the material.

The relative permittivity constant, κ, also called the dielectric constant, is often expressed as the ratio between the material's permittivity and that of a vacuum (ε_0) [17]:

$$\kappa = \frac{\varepsilon}{\varepsilon_0} \tag{2}$$

Materials such as wet sand, clay, or pockets of groundwater have high relative permittivity, making signal penetration much more challenging. This is reflected in a high value of the imaginary portion of the dielectric constant, strongly damping the GPR signal.

The amount of the radio signal that is reflected at a boundary between substances is a ratio of the magnitude

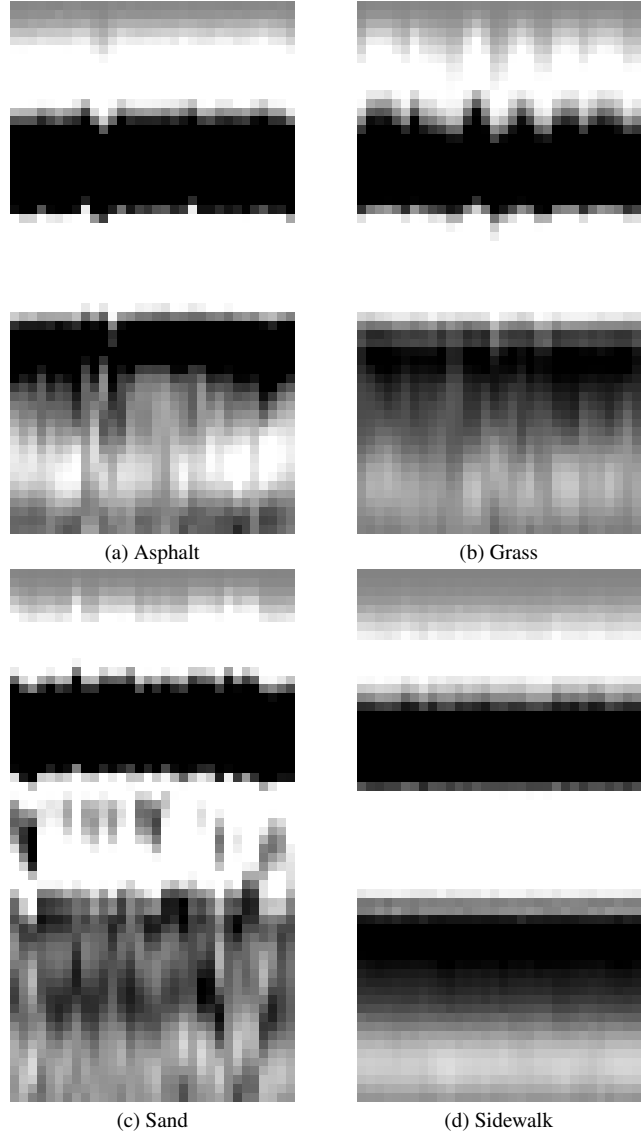(a) Asphalt   (b) Grass

(c) Sand   (d) Sidewalk

Figure 2. 60 x 32 pixel slice of a GPR radargram direct wave for four terrain types: a. asphalt, b. grass, c. sand, and d. sidewalk.

of the dielectric constants of both materials, κ_1 and κ_2, and it is called the reflection coefficient (R) [20]:

$$R = \frac{\sqrt{\kappa_1} - \sqrt{\kappa_2}}{\sqrt{\kappa_1} + \sqrt{\kappa_2}} \tag{3}$$

The higher the reflection coefficient, the stronger the reflection. For example, when the radar signal passes from a high permittivity material, such as water, to a low permittivity material, such as dry sand, there is a positive reflection coefficient. Higher reflectivity at a boundary results in a more visible marker in the radargram (see the contrasting light and dark returns in Fig. 2). Each vertical line of pixels in the radargram corresponds to one radar ping return from the GPR, where the returns are in mV.

The depth d of the features shown in a radargram can be calculated by:

$$d = v \times \frac{t}{2} \tag{4}$$

where v is the wave travel velocity of the ground material (a value dependent on the soil makeup) and t is the two-way travel time [36]. Calculating the depth is very difficult unless wave velocity is well-characterized or if the depth to a feature is manually measured. However, this tells us that the pixels towards the top of the radargram represent radiowave returns that were received the fastest, and thus represent terrain that was closer to the sensor.

An important property of GPR is the direct wave, which is an effect of the radar signals travelling directly through the air and top surface of the ground to the receiver. See Fig. 3 for a visual comparison of direct and reflected waves. In many cases where GPR is being used to characterize sub-surface features such as pipes or geologic formations, the direct waves are removed with background subtraction filters [30]. However, in this work we are especially focused on the surface terrain, making the direct waves of particular interest. Figure 2 shows us the qualitative visual distinctiveness of the direct waves for the four terrains we have focused on in this work. In Sec. 4.4, we directly investigate how the network performs on both the direct wave and the reflected wave in order to support our use of the direct wave phenomena for surface terrain classification.

The frequency of the GPR dictates its penetration depth and resolution [17]. A low frequency GPR can penetrate hundreds of meters underground, but with a lower resolution. As we are interested primarily in surface and shallow subsurface terrain information, we chose a high frequency 500 MHz GPR sensor in order to benefit from higher resolution data.

3.2. Radargram Preprocessing

As discussed in Sec. 3.1, a GPR sensor returns a vector for each sample that represents the electromagnetic wave reflectivity of materials below the sensor. These returns are then concatenated into an image representation called a radargram. In order to create our training image set, we first pad each radargram from $h \times w$ into $h \times w_{\text{pad}}$ according to the following equation, where % represents the modulus operator:

$$w_{\text{pad}} = w + ((w - w_{\text{resize}})\% s) \tag{5}$$

Next, we slice horizontally with resize width, w_{resize}, along the area of the image representing the direct wave at an interval of s, the stride length. For our experiments, we used a stride length of 4 pixels.

As each radargram slice represents multiple samples from the GPR, there is the chance for more than one terrain
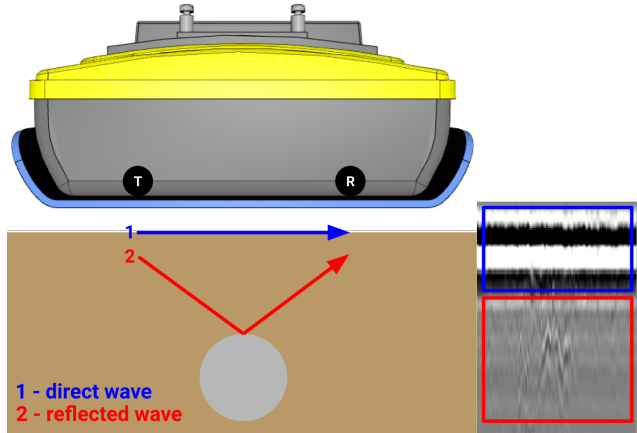

Figure 3. The radar propagation from GPR pulses has a unique property where it travels through the air and the surface of the ground (direct wave, blue), and also through the subsurface material itself (reflected wave, red). The radargram on the right shows how these different received waves appear in a GPR radargram.

to be represented in a sliced image. This will certainly happen in the real world as the sensor traverses terrain boundaries. For the purposes of these experiments, the radargrams were sliced to isolate each terrain into separate images before further processing into the test and train sets.

The radargrams were split with an 80/20 train/test split before being further sliced into widths 1, 8, 16, 24, and 32 pixels for use in our experiments. This was done in order to provide a frozen test set across all experiments.

3.3. Classification Networks

We investigated four neural network architectures for their effectiveness in classifying terrain from GPR data. Each of the networks takes in GPR signals – either concatenated into a 2D image or just each individual 1D sample vector. The output of each network is a single class prediction: asphalt, grass, sand, or sidewalk. Once a terrain class for a segment of GPR samples is predicted, this can be integrated into a terrain mapping framework as shown in Sec. 4.7.

3.3.1 AlexNet

The AlexNet [21], first proposed in 2012, is a common supervised approach to RGB image classification. Although GPR radargrams are not images in the RGB space, each vector return from the sensor can be concatenated into a 2D radargram image for training and inference. The input to the AlexNet is this 2D grayscale radargram image, as shown in Fig. 2.

3.3.2 ResNet101

ResNet [14] is typically viewed as an improvement upon AlexNet in the supervised image classification problem do-

main. We include ResNet101 as a supervised baseline for radargram classification. Similarly to the AlexNet, the input to the ResNet network is a 2D grayscale radargram image.

3.3.3 1D Convolutional Neural Network

One consequence of formulating GPR data as a 2D image is that a real-time terrain classification system must wait for multiple data points to be obtained before being deployed. Additionally, it is challenging to determine the boundary between two terrain types present in one radargram slice, as the labels are per trace sample. This can be alleviated by evaluating signal-based networks that take in a single 1D GPR trace for classification. A common approach to 1D signal classification is the 1D Convolutional Neural Network (CNN) [19], which takes as an input the (200×1) GPR sample return vector and outputs a class prediction.

3.3.4 Variational Autoencoders for Deep Clustering

VAEs are unsupervised networks designed to encode an image into a 1D latent space, and then reconstruct that image from the encoding [18]. VAEs can be useful for learning underlying embedded structures in an image [31]. In our particular approach, modeled after [12], we modify the standard VAE reconstruction loss function to include incentive for better clusters. After performing k-means clustering on the embedding space, we can use the autoencoder to predict the terrain class of the test set.

4. Experiments and Results

4.1. Mobile Robot Platform

A mobile robot platform (shown in Fig. 4) was used for collecting data and deploying the GPR classification networks explored in this work. We used a Clearpath Husky robot running off of a Jetson Nano with Ubuntu 19.04 and ROS Melodic. The Husky robot has built-in wheel encoders for use in filtered odometry data calculation. Our modifications to the platform included custom mounts for a Noggin 500 MHz GPR, a Garmin GPS module, and an Intel Realsense D455 RGBD camera. The 500 MHz GPR can penetrate 5 - 10 meters below the surface (depending on the material) with a high resolution, making it the most suitable for the task of surface terrain classification compared to lower frequency sensors. A custom ROS driver for the GPR was also developed as a part of this work.

4.2. Data Collection

Data collection occurred over several months from July through November 2023, in temperatures ranging from 40° F to 70° F. Table 1 details the specific locations in southeast Michigan and at the University of Michigan where different terrain data was collected. At each site, the radargram was
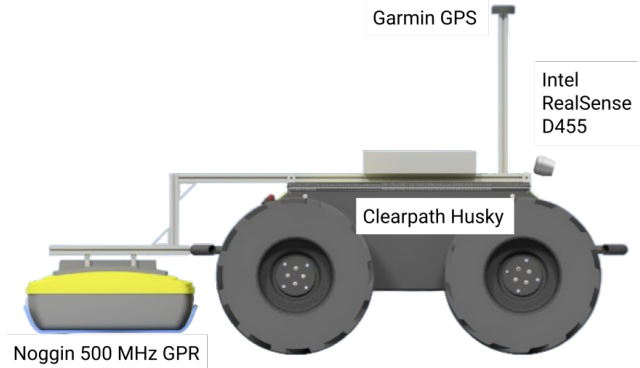

Figure 4. Our custom Clearpath Husky robotic sensor platform equipped with a GPR, a GPS, an RGBD camera, and wheel encoders.

processed for training by slicing along terrain boundaries, if any were present. This was done by cross-referencing camera images of the site. A total of 64 radargram images was collected of varying lengths. This amounts to 7752 individual traces across all radargrams and terrain categories. Some data collection was done following days of precipitation. This is important to note as radio wave propagation is heavily impacted by water.

Field Site	Terrain Type Collected
Independence Lake, MI	asphalt, grass, sand
UM North Campus	asphalt, grass, sand, sidewalk
UM MCity	asphalt

Table 1. Field data collection sites

4.3. Terrain Classification Networks

As described in Sec. 3.3, we investigated four different networks for classifying GPR radargrams: AlexNet [21], ResNet101 [14], a 1D CNN [19], and an autoencoder for deep clustering [12]. All networks were trained on an NVIDIA GeForce RTX 3090 with 24 GB of memory. AlexNet and ResNet101 were pretrained on ImageNet, trained until convergence at 20 epochs, and employed cross entropy loss. The 1D CNN was also trained until convergence at 20 epochs. We stopped training after 20 epochs to avoid overfitting. The VAE was trained for 4000 epochs. Each network was trained five times and we report the average results in Tab. 2. The AlexNet and the ResNet101 perform equally the best, however the AlexNet trained faster so we use it for further experiments.

4.4. Selection of Radargram Subsections

As previously mentioned, we sectioned out the portion of the radargram that represents the direct wave as this isolates the GPR's representation of the surface terrain. In this ex-

	Overall (↑)	Asphalt (↑)	Grass (↑)	Sand (↑)	Sidewalk (↑)
1D CNN [19]	0.795	0.921	0.750	0.721	0.737
Autoencoder [12]	0.873	0.922	0.967	0.653	1.000
ResNet101 [14]	**0.985**	**1.000**	**1.000**	0.947	**1.000**
AlexNet [21]	**0.985**	**1.000**	0.995	**0.952**	**1.000**

Table 2. GPR classification network accuracy, averaged over five trained networks per architecture.

periment, we seek to confirm our hypothesis that the direct wave is more informative for surface terrain classification than the reflected wave section of the radargram. With our frozen train and test radargram sets, we slice the training images into two sections: the direct wave and the reflected wave. Each image is of size 60×32 pixels. The total number of image slices in the train set 1033, and the total number in test is 278. Additionally, we train networks on the full (combined direct wave and reflected wave) 200×32 radargram slices. We use AlexNet classifiers for this experiment and report average accuracy across five trained networks for each approach.

The results are shown in Tab. 3, where the full radargram performs marginally better than the direct wave. The AlexNet trained on the full image may be focusing on learning features in the region of the image that correspond to the direct wave, hence the similar performance between the two networks. The network trained on the direct wave slice significantly outperforms the AlexNet trained only on the reflected wave. These results are intuitive because the reflected wave represents the subsurface materials several feet deep, and it may also capture features such as underground pipes or water deposits. There may be some correlation between the subsurface and surface terrains that is learned by the network, but ultimately the direct wave is a much better representation. Additionally, the validation loss for the reflected wave train set converged poorly during training. Overall, this experiment is a good indicator that the direct wave is the most important part of the radargram for surface terrain classification. Still, using the full slice provides the best performance.

Radargram Input Slice	Overall Accuracy (↑)
Reflected Wave	0.768
Direct Wave	0.985
Full Slice	**0.989**

Table 3. Comparison of AlexNet networks trained on different portions of the GPR radargram. See Fig. 3 for an example of direct wave and reflected wave in a radargram.

4.5. Time Series Length

A radargram is a time series of multiple returns from the GPR sensor stacked horizontally into an image. In this

(a) 1 trace (b) 8 traces (c) 16 traces (d) 24 traces (e) 32 traces
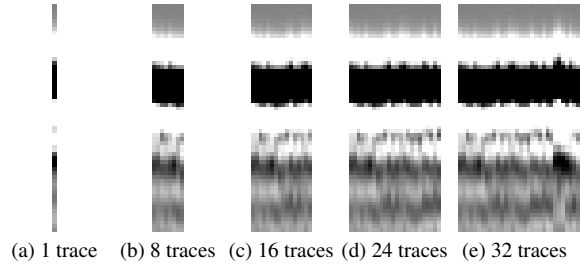
Figure 5. The five time series lengths used for the experiment shown on a single radargram of sandy terrain. Note that the pattern distinct to the terrain type makes itself more apparent as the number of traces increases.
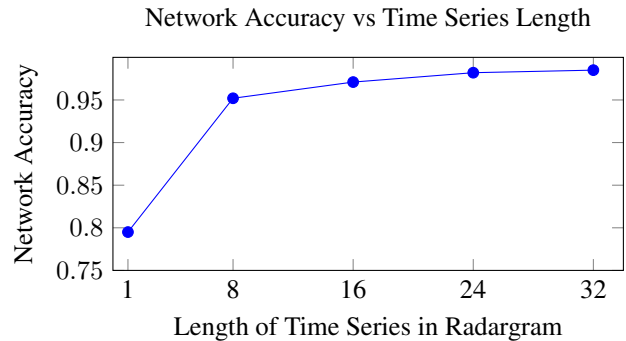

Figure 6. Averaged network accuracy plotted against the input radargram length. Notice that the performance increases minimally after time series length of 8.

experiment, we seek to understand how many consecutive samples must be collected for the terrain to become recognizable to the classification network. In Fig. 5, we see how as the width of the radargram increases, the pattern defined by the terrain becomes more apparent. However, depending on the sampling frequency, we may have to wait upwards of 15 seconds for 32 traces to be recorded – making real-time terrain classification with this size less convenient. Additionally, passing over terrain boundaries will take longer to detect with a longer radargram. The smaller the radargram necessary for accurate classification, the more fine-grained and spatially precise our terrain predictions. We train a 1D CNN on just single traces, and also train four different AlexNets on radargrams with 8, 16, 24, and 32 stacked traces. Each network is trained five times, and the average

results are presented in Fig. 6.

As shown in Fig. 6, the radargram with 32 traces performs the best, as we might expect. However, the radargrams of lengths 8, 16, and 24 perform only marginally worse, making these viable alternatives that can also provide more granular predictions. The 1D CNN, trained only on individual vector signal returns from the GPR, performs much worse – indicating that terrain class in GPR direct wave radargrams is more detectable over multiple traces.

4.6. Deep Clustering

The VAE deep clustering approach to classification is unsupervised, so we may expect less accurate results compared to the supervised AlexNet and ResNet101. In Fig. 7, we see distinct clusters for asphalt and sidewalk, and two more mixed clusters for grass and sand. The individual terrain class accuracies presented in Tab. 2 similarly reflect that the networks have a harder time learning the difference between sand and grass. This may be due to stronger similarities between radargrams of these two types.

4.7. Future Applications to Mapping

We conclude that GPR shows promise as an additional exteroceptive modality for terrain classification and semantic mapping. In this section, we show preliminary results that build off of the Bayesian approach in [8] to fuse GPR classification network outputs and RGB semantic segmentation network outputs into a labeled terrain map. We compare our fused vision and GPR approach against a vision-only terrain segmentation network from [37].

In Fig. 8, preliminary results on a test site where the robot traversed a sidewalk are shown. Of particular note

Figure 8. Bird's eye view terrain mapping comparison between ground truth, the RGB-only baseline, and our method (RGB and GPR) on a test site. Red represents asphalt, blue represents sidewalk, green represents grass, gray represents other, and black represents no data. Notice in our approach that the GPR is able to correct for the RGB camera network's misclassification of the sidewalk.
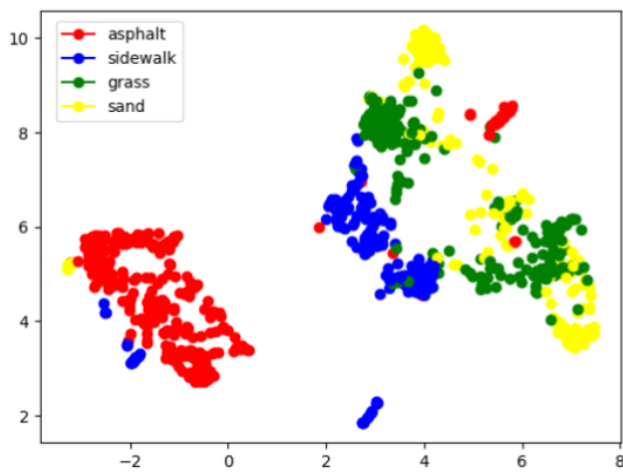

Figure 7. Deep clustering VAE results with combined loss function. Notice the distinct clusters for asphalt and sidewalk, and the more mixed clusters for sand and grass.
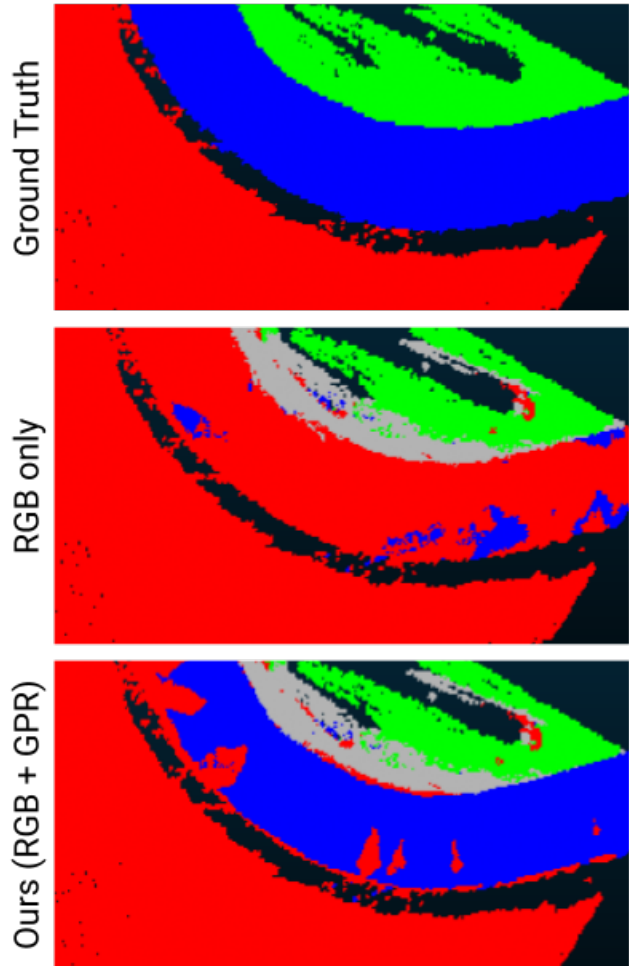
is that the vision-only approach misidentifies the sidewalk (blue) as asphalt (red). In our approach, which fuses vision and GPR data, the GPR is able to successfully identify the sidewalk and correct the map semantics as the robot drives over the terrain.

5. Discussion

Our results present a preliminary case for the use of GPR for surface terrain classification. Through the experiments presented, we show how different radargram subsections and lengths affect classification network performance. With our dataset and choice of networks, the direct wave portion of the sensor return and a time series length of at least 8 pixels

provides optimal prediction results. We believe that the results presented demonstrate great promise in the novel use of GPR for surface terrain classification.

Alternative sensor modalities such as microphones, IMUs, and torque sensors are gaining increasing traction as advances in multimodal learning demonstrate their capability to compensate for the weaknesses of more traditional sensing modalities such as RGB and depth. GPR is typically a downward facing sensor, and in order to collect samples the sensor must pass directly over the terrain. As the sensor is usually not forward-looking, this limitation motivates a multimodal approach to navigation stacks that include GPR and forward-facing cameras. We show preliminary results from an RGB and GPR fusion terrain classification and mapping framework, demonstrating that GPR can help correct for misclassifications from forward-facing camera-based networks. Additionally, it is of note that labeling GPR traces is trivial compared to other modalities, reducing the manual labeling burden. We hope to motivate applications with robotic platforms that already have a GPR to incorporate this sensor as an additional modality for the task of terrain classification.

6. Conclusion

Mobile systems are cost and power constrained, making it more effective to utilize existing sensors rather than include new ones. In this work, we propose the novel use of GPR for surface terrain classification. We explore supervised and unsupervised networks for classification, and we also perform several experiments to learn more about how the radargram size and direct wave impact the classification network performance. Lastly, we demonstrate preliminary results of integrating GPR terrain classifications for a multimodal mapping framework. This motivates future work on leveraging GPR to complement RGB for terrain classification and semantic mapping.

Acknowledgements

References

[1] Deegan Atha, R. Michael Swan, Annie Didier, Zaki Hasnain, and Masahiro Ono. Multi-mission terrain classifier for safe rover navigation and automated science. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–13. IEEE, 2022. 2

[2] Alexander Baikovitz, Paloma Sodhi, Michael Dille, and Michael Kaess. CMU-GPR dataset: Ground penetrating radar dataset for robot localization and mapping. In *Proceedings of ICRA '21 Radar in Robotics Workshop*, 2021. 3

[3] Christopher A. Brooks and Karl Iagnemma. Self-supervised terrain classification for planetary surface exploration rovers. *Journal of Field Robotics*, 29:445–468, 2012. 1, 2

[4] Kenneth S. Cole and Robert H. Cole. Dispersion and Absorption in Dielectrics I. Alternating Current Characteristics. *The Journal of Chemical Physics*, 9(4):341–351, 1941. 3

[5] Matthew Cornick, Jeffrey Koechling, Byron Stanley, and Beijia Zhang. Localizing ground penetrating radar: A step toward robust autonomous ground vehicle localization. *Journal of Field Robotics*, 33(1):82–102, 2016. 1, 3

[6] Thikra Dawood, Zhenhua Zhu, and Tarek Zayed. Deterioration mapping in subway infrastructure using sensory data of GPR. *Tunnelling and Underground Space Technology*, 103: 103487, 2020. 3

[7] K Eder, C Reidler, C Mayer, and M Leopold. Crevasse detection in alpine areas using ground penetrating radar as a component for a mountain guide system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37:837–841, 2008. 1

[8] Parker Ewen, Adam Li, Yuxin Chen, Steven Hong, and Ram Vasudevan. These maps are made for walking: Real-time terrain property estimation for mobile robots. *IEEE Robotics and Automation Letters*, 7(3):7083–7090, 2022. 7

[9] Jinglun Feng, Liang Yang, Ejup Hoxha, Biao Jiang, and Jizhong Xiao. Robotic inspection of underground utilities for construction survey using a ground penetrating radar. *Journal of Computing in Civil Engineering*, 37(1), 2023. 2

[10] Sabine Fiedler, Bernhard Illich, Jochen Berger, and Matthias Graw. The effectiveness of ground-penetrating radar surveys in the location of unmarked burial sites in modern cemeteries. *Journal of Applied Geophysics*, 68(3):380–385, 2009. 1

[11] Tianrui Guan, Ruitao Song, Zhixian Ye, and Liangjun Zhang. Vinet: Visual and inertial-based terrain classification and adaptive navigation over unknown terrain. In *IEEE International Conference on Robotics and Automation (ICRA) Proceedings*, 2022. 1, 2

[12] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1753–1759, 2017. 5, 6

[13] Svein-Erik Hamran, David A Paige, Hans EF Amundsen, Tor Berger, Sverre Brovoll, Lynn Carter, Leif Damsgård, Henning Dypvik, Jo Eide, Sigurd Eide, et al. Radar imager for mars' subsurface experiment—RIMFAX. *Space Science Reviews*, 216:1–39, 2020. 1, 2

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 4, 5, 6

[15] Y. Hervé, V. Ciarletti, A. Le Gall, C. Corbel, R. Hassen-Khodja, W.S. Benedix, D. Plettemeier, O. Humeau, A.J.

Vieau, B. Lustrement, S. Abbaki, E. Bertran, L. Lapauw, V. Tranier, N. Oudart, F. Vivat, C. Statz, Y. Lu, S. Hegler, and A. Hérique. The wisdom radar on board the exomars 2022 rover: Characterization and calibration of the flight model. *Planetary and Space Science*, 189:104939, 2020. 2

[16] Reina Ishikawa, Ryo Hachiuma, and Hideo Saito. Self-supervised audio-visual feature learning for single-modal incremental terrain type clustering. *IEEE Access*, 9:64346–64357, 2021. 2

[17] Harry M. Jol. *Ground Penetrating Radar Theory and Applications*. Elsevier, 2009. 3, 4

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 5

[19] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 2021. 5, 6

[20] Rosemary Knight. Ground penetrating radar for environmental applications. *Annual Review of Earth and Planetary Sciences*, 29(1):229–255, 2001. 3

[21] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. 4, 5, 6

[22] Anne Lalagüe. *Use of ground penetrating radar for transportation infrastructure maintenance*. PhD thesis, Norwegian University of Science and Technology, 2015. 1

[23] Livia Lantini, Fabio Tosti, Iraklis Giannakis, Lilong Zou, Andrea Benedetto, and Amir M. Alani. An enhanced data processing framework for mapping tree root systems using ground penetrating radar. *Remote Sensing*, 12(20), 2020. 3

[24] Michael Milford, Eleonora Vig, Walter Scheirer, and David Cox. Vision-based simultaneous localization and mapping in changing outdoor environments. *Journal of Field Robotics*, 31(5):780–802, 2014. 1

[25] Masahiro Ono, Thoams J. Fuchs, Amanda Steffy, Mark Maimone, and Jeng Yen. Risk-aware planetary rover operation: Autonomous terrain classification and path planning. In *2015 IEEE Aerospace Conference*, pages 1–10, 2015. 2

[26] Teddy Ort, Igor Gilitschenski, and Daniela Rus. GROUNDED: The localizing ground penetrating radar evaluation dataset. In *Robotics: Science and Systems*, 2021. 3

[27] Tejas Panambur, Deep Chakraborty, Melissa Meyer, Ralph Milliken, Erik Learned-Miller, and Mario Parente. Self-supervised learning to guide scientifically relevant categorization of martian terrain images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1321–1331. IEEE, 2022. 2

[28] L.P. Peters, J.J. Daniels, and J.D. Young. Ground penetrating radar as a subsurface environmental sensing tool. *Proceedings of the IEEE*, 82(12):1802–1822, 1994. 2

[29] Amirreza Shaban, Xiangyun Meng, Joonho Lee, Byron Boots, and Dieter Fox. Semantic terrain classification for off-road autonomous driving. In *Proceedings of the 5th Conference on Robot Learning PMLR*, 2022. 1, 2

[30] Prabhat Sharma, Bambam Kumar, Dharmendra Singh, and SP Gaba. Critical analysis of background subtraction techniques on real GPR data. *Defence Science Journal*, 67(5), 2017. 4

[31] Chunfeng Song, Feng Liu, Yongzhen Huang, Liang Wang, and Tieniu Tan. Auto-encoder based data clustering. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 117–124, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 5

[32] R. Michael Swan, Deegan Atha, Henry A. Leopold, Matthew Gildner, Stephanie Oij, Cindy Chiu, and Masahiro Ono. AI4MARS: A dataset for terrain-aware autonomous driving on mars. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1982–1991. IEEE, 2021. 2

[33] Grace Vincent, Alice Yepremyan, Jingdao Chen, and Edwin Goh. Mixed-domain training improves multi-mission terrain segmentation. *ECCV Workshop: AI4SPACE*, 2022. 2

[34] Rebecca M. Williams, Laura E. Ray, and James Lever. An autonomous robotic platform for ground penetrating radar surveys. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 3174–3177, 2012. 2

[35] Kaijun Wu, Gabriela Arambulo Rodriguez, Marjana Zajc, Elodie Jacquemin, Michiels Clément, Albéric De Coster, and Sébastien Lambot. A new drone-borne GPR for soil moisture mapping. *Remote Sensing of Environment*, 235:111456, 2019. 1, 3

[36] Fei Xie, Wallace W.L. Lai, and Xavier Dérobert. GPR-based depth measurement of buried objects based on constrained least-square (CLS) fitting method of reflections. *Measurement*, 168:108330, 2021. 4

[37] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 7

[38] Jannik Zürn, Wolfram Burgard, and Abhinav Valada. Self-supervised visual terrain classification from unsupervised acoustic feature learning. *IEEE Transactions on Robotics*, 2019. 1, 2