

UNRealNet: Learning Uncertainty-Aware Navigation Features from High-Fidelity Scans of Real Environments

Samuel Triest^{1*}, David D. Fan², Sebastian Scherer¹, and Ali-Akbar Agha-Mohammadi²

Abstract— Traversability estimation in rugged, unstructured environments remains a challenging problem in field robotics. Often, the need for precise, accurate traversability estimation is in direct opposition to the limited sensing and compute capability present on affordable, small-scale mobile robots. To address this issue, we present a novel method to learn [u]ncertainty-aware [n]avigation features from high-fidelity scans of [real]-world environments (UNRealNet). This network can be deployed on-robot to predict these high-fidelity features using input from lower-quality sensors. UNRealNet predicts dense, metric-space features directly from single-frame lidar scans, thus reducing the effects of occlusion and odometry error. Our approach is label-free, and is able to produce traversability estimates that are robot-agnostic. Additionally, we can leverage UNRealNet’s predictive uncertainty to both produce risk-aware traversability estimates, and refine our feature predictions over time. We find that our method outperforms traditional local mapping and inpainting baselines by up to 40%, and demonstrate its efficacy on multiple legged platforms.

I. INTRODUCTION

More and more robots are being deployed in rough-terrain applications such as construction and disaster response [1–5]. While these robots (which are often legged) have the promise of performing work that is potentially time-consuming and dangerous, these environments often present mobility challenges due to clutter, dangerous terrain, occlusions, and change in the environment over time. As such, there is need for fast, precise and uncertainty-aware traversability analysis.

At a high level, traversability analysis is the process of constructing cost functions that quantify how desirable it is to reside in a particular region. This process is performed using on-board sensing (such as lidar or camera) and on-board compute. Classical approaches to traversability estimation typically rely on accurate perception to reconstruct the local environment, and apply hand-crafted rules to determine the cost function sent to planning and control modules. Unfortunately, this means that traversability estimation modules are subject to limitations and errors in perception such as odometry drift and occlusion, and often propagate these errors to planning and control, with potentially dangerous consequences for the robot and its operators.

Learning presents a promising solution to this problem, as deep neural networks are able to generate complex predictions that capture abstract environmental priors (e.g. most rooms have flat floors, boxes have a particular shape, etc.). However, generating precise ground-truth labels in high



Fig. 1: We propose UNRealNet (b), a network that produces high-quality, robot-agnostic navigation features and traversability (c) directly from pointclouds from an on-board sensor (a). Colormap for risk maps included for clarity.

quantities to train such a neural network is challenging. Prior work [6,7] often leverages aggregation of future sensing (i.e. predict the map at time $t+k$ from the sensing at time t), but such approaches are limited by the quality of the SLAM algorithm, the accuracy of the sensor, and the mobility limitations of the robot. The end result is that training labels may be noisy and incomplete. While simulation may provide a means to address these issues [8], simulated environments are often overly simplistic and models trained purely in simulation may not successfully bridge the sim-to-real gap.

Additionally, uncertainty is a critical component of traversability estimation. In practice, there will be environmental ambiguities (such as occlusion and out-of-distribution objects) that cannot be resolved by on-board sensing. It is thus critical that traversability estimation methods be able to both recognize and address said uncertainty. Learning based-methods have shown promise in uncertainty estimation [9,10] for their ability to learn environmental priors, producing uncertainty estimates that go beyond common heuristics.

In this work, we leverage a high-precision laser scanner to generate very accurate and dense reconstructions of several unstructured environments. This allows us to generate high-quality traversability features without the need to address common robotic perception limitations such as occlusion or drift in state estimation. In order to deploy on-robot, we use a label-free learning pipeline where we generate large amounts of simulated lidar data, and the corresponding traversability features. We then train a neural network to predict these features directly from raw pointclouds and demonstrate its

* Work done as an intern at Field AI.

¹ Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. {striest, basti}@andrew.cmu.edu

² Field AI, Mission Viejo, CA, USA.

efficacy in producing dense, metric traversability estimates directly from sparse lidar data.

II. RELATED WORK

Given its importance in mobile robot autonomy, traversability estimation has been the focus of much research [11–14]. Of particular importance to this work is the approach described by Fankhauser et al. [15] and Chilian et al. [16], which is a commonly-used baseline for legged traversability. This method relies on computing several terrain features from a digital elevation map (DEM), which are combined via Equation 1 into a cost function for downstream planning. For each feature f , the user is required to specify two parameters α_f, f_{crit} , which specify the importance of that feature, and a critical value over which the robot should not traverse, respectively.

$$c^{i,j} = \sum_f \left(\alpha_f \frac{f^{i,j}}{f_{crit}} \right) \quad (1)$$

While the above methods are well-suited to estimating traversability on fully-observed, accurately-measured terrain, they are brittle with respect to sensing errors and limitations, which are omnipresent in practice. Errors in odometry can lead to terrain surfaces and obstacles being mis-estimated. Dynamic objects such as people and vehicles can cause false positives. Occlusions and negative obstacles can result in no sensing data in areas dangerous to the robot. While methods exist to combat these issues [17–19], they are often heuristic, and time-consuming to compute.

Recent research in learning for traversability focuses on learning environment priors to improve traversability estimation in occluded or mis-estimated regions of the environment. Ramakrishnan et al. [20] train a neural network to predict occupancy maps from a ego-centric RGB-D images. They demonstrate this network’s ability to leverage environment priors from indoor environments [21,22] to improve navigation and exploration. Additional work by Georgakis et al. [10,23] extends this idea to semantic mapping and uncertainty-aware exploration. Work done by Fan et al. [24,25] estimate traversability using geometric features of a terrain surface. In contrast to prior work, they define a mapping from terrain features to *distributions* of cost, allowing their navigation behaviors to reason about uncertainty. Stolze et al. [26] train a UNet to inpaint elevation maps, given partial elevation maps. They demonstrate that their network is capable of predicting occluded elevation values on real-world data. Gan et al. [7] train a neural network to predict semantics and traversability by building a 3D reconstruction of the environment using the robot’s trajectory and running traversability analysis on it. Meng et al. [6] train a network to predict BEV-space semantics and terrain information from stereo images. Similarly to Gan et al. [7], ground truth is generated by aggregating pointclouds over time and hand-labeling the resulting output. They then demonstrate this network’s ability to produce cost maps for planning by hand-designing a cost function that uses the predicted semantic and terrain information.

We differentiate ourselves from prior work by leveraging high-precision laser scanners to generate high-quality ground truth environment models. This gives us a more accurate and complete pointcloud of the environment to leverage for learning. In fact, the quality of our environment representation is sufficient for us to simulate realistic lidar scans for training. Additionally, we differentiate ourselves from prior work in that we predict a number of geometric features important to traversability with uncertainty estimates, as well as a principled way of leveraging this uncertainty in traversability estimation. This allows our method to be aware of uncertainty and robot-agnostic, while remaining label-free.

III. METHODOLOGY

Our goal is to train a neural network to predict dense map of traversability features ($\mathcal{M} \in \mathbb{R}^{C \times W \times H}$, where C is the number of features, W is the map width, and H is the map height) from a sparse, instantaneous pointcloud ($\mathcal{P} \in \mathbb{R}^{P \times 3}$, where P is the number of points). Doing so will result in a local map that is quick to compute and is robust to occlusions as well as odometry drift. A high-level summary of our method is presented in Figure 2.

A. Creating a High-Quality, Real-World Dataset

1) *Mapping*: To begin, we leverage a high-quality laser scanner (a FARO Focus and Leica BLK360) and their pointcloud registration software to generate a high-quality pointcloud of six locations, over four unique environments. This approach is preferred to relying on offline SLAM algorithms for the following reasons:

- 1) With high precision and accuracy (up to 1 million points per scan, at 1.0mm precision) [27], the precision of these laser scanners exceeds that of common sensors in robotics such as lidar and camera.
- 2) The high density of points in these scans makes it possible to generate high quality data from synthetic viewpoints. This allows us to create training pairs from places the robot did not (or could not) traverse.
- 3) Near full coverage of the environment can be achieved by not being reliant on the traversability capabilities of the robot using the sensor.

In order to generate high quality, metric-space traversability labels from this pointcloud, we run a simple elevation mapping and freespace detection algorithm. The resulting freespace was then manually refined. This step can easily be replaced with a more sophisticated algorithm [28,29] if needed. Given this elevation map, we can then construct high-quality traversability features. As is common in prior work [15,16], we use the step, slope and roughness of a local neighborhood of elevation cells (the same as Chilian et al. [16]). We find that the high precision of the pointcloud allows for relatively nuanced costing of subtle terrain features such as loose wires and smaller changes in elevation (see Figure 3). The end result of this local mapping step is that we now have a dense, accurate pointcloud \mathcal{P}_{gt} and corresponding map features \mathcal{M}_{gt} which can be leveraged for training.

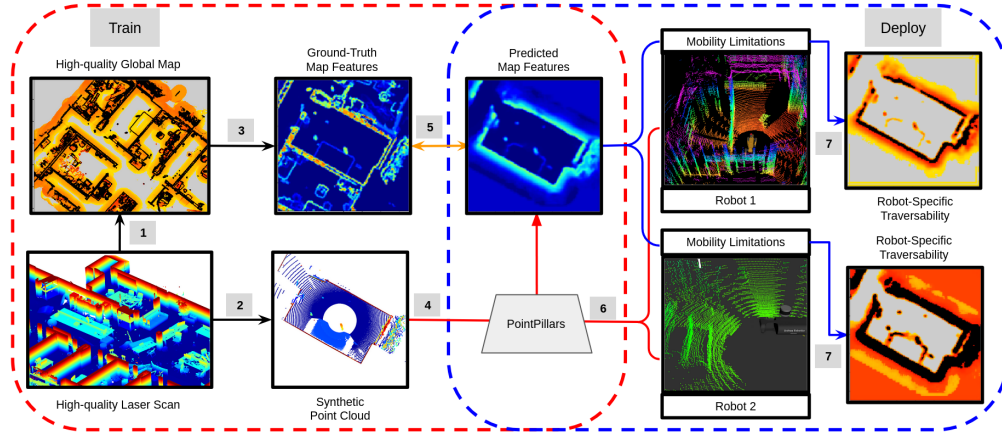


Fig. 2: An overview of our algorithm. We first create a high-quality map (1) from a pointcloud from a laser scanner. We then generate synthetic pointclouds (2) and their corresponding ground-truth map features (3). We train UNRealNet to predict the high-quality, robot-agnostic map features from the noisy, synthetic pointclouds (4, 5). We can then deploy this network on multiple robots (6) by leveraging a robot-specific traversability function (7). Red arrows denote network inputs, blue networks denote network outputs, and the orange arrow denotes the training objective (Equation 3).

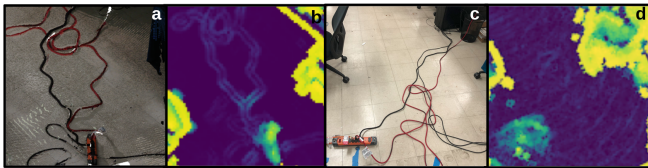


Fig. 3: An example that highlights the value of the high-resolution scanner. (a) the pointcloud from the scanner. (b) the slope feature of the corresponding local map. We are able to identify the edges of each wire. (c) A photo of the environment and (d) the slope feature from the registered pointcloud. The noise from the lidar and SLAM is too high to see the wires.

2) *Synthetic pointclouds*: A key advantage of using a dense pointcloud of the environment is that it is possible to generate high-quality synthetic data from many viewpoints that were not directly experienced by the robot. Thus, in order to generate training pairs $(\mathcal{P}, \mathcal{M})$ from $\mathcal{M}_{gt}, \mathcal{P}_{gt}$ to train our network, we can sample a pose in freespace p and project \mathcal{P}_{gt} into the frame defined by p (equivalent to the CP method described by Langer et al. [30]). This gives us a pointcloud \mathcal{P} that resembles what the sensor would have observed from pose p . In order to generate the corresponding set of map features, we can transform and crop the global map based on p . Finally, we can apply a noising pipeline (including range noise, robot self-hit masks, etc.) to make \mathcal{P} resemble the output of a robot sensor [31].

B. Traversability Prediction with UNRealNet

UNRealNet is a PointPillars-based [32] network designed to predict traversability features directly from the above synthetic pointclouds. Similarly to PointPillars, we first define a metric region around the ego-pose (consisting of a resolution, width and height). All points in this region are passed through a PointPillars network (consisting of

a small PointNet [33], cell-wise max-pooling, and then a UNet [34], which produces a cell-wise, factorized Gaussian distribution for several traversability features (Equation 2). This network is supervised by minimizing the negative log-likelihood of the ground-truth map crop under the predicted map distribution, where unobserved cells in the ground truth map crop are masked out via a mask M_{gt} (Equation 3). Note that the network receives a training signal for cells observed in the label that are not observed in the input, thus requiring the network to inpaint map features.

$$f_{\theta}(\mathcal{P}) = \mathcal{N}(\tilde{\mu}, \tilde{\sigma}), \quad \tilde{\mu} \in \mathbb{R}^{C \times W \times H}, \quad \tilde{\sigma} \in \mathbb{R}^{C \times W \times H} \quad (2)$$

$$\mathcal{L}(\hat{\mathcal{M}}, f_{\theta}(\mathcal{P})) = \frac{M_{gt}}{\sum M_{gt}} \sum_{i=0}^W \sum_{j=0}^H \sum_{f=0}^C \left[-\log(p(\hat{\mathcal{M}}^{f,i,j}, \mathcal{N}(\tilde{\mu}^{f,i,j}, \tilde{\sigma}^{f,i,j}))) \right] \quad (3)$$

C. Probabilistic Traversability Function

While it is possible to compute a single traversability function at train-time, this approach has a significant downside in that traversability estimates cannot be changed at deployment time. This limitation would require re-training of the network for any additional robot platform, or even when tuning the cost function to specific user needs or risk tolerances. As such, it is desirable to be able to re-compute traversability using the predicted features from the network.

It is common practice to use Equation 1 as a cost function for planning, where the features in question are geometric terrain properties such as local slope and roughness. However, given our uncertainty estimates, it is more intuitive to treat traversability as the probability that all map features are below their critical values (Equation 4).

$$p(T^{i,j}) = p(T^{i,j} < f_{crit}, \forall f) \quad (4)$$

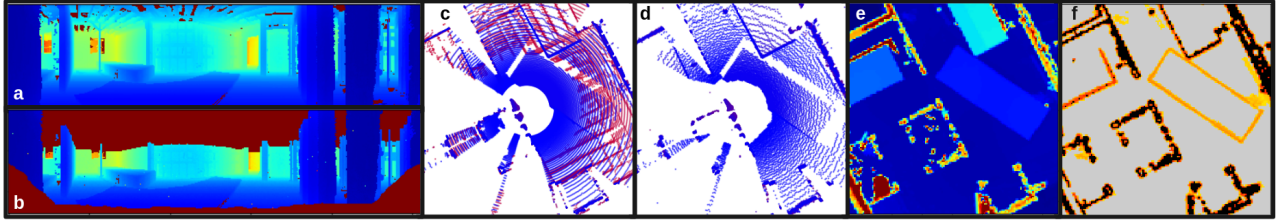


Fig. 4: A visualization of a datapoint in our dataset, showing (a) the depth image obtained from simulating lidar in the ground-truth pointcloud (blue=close, red=far), (b) the depth image obtained by running the original depth image through the noising pipeline, (c) the BEV projection of the original pointcloud, (d) the BEV projection of noised pointcloud, and corresponding crops from the ground-truth elevation map (e) and a sample traversability map (f).

Assuming that each map feature is an independent Gaussian (i.e. $f_{i,j} = (\mu_{i,j}, \sigma_{i,j})$) allows us to simplify Equation 4 into Equation 5. This formulation only requires the evaluation and multiplication of several Gaussian cdfs, which are easily computable at the speeds required for multi-Hz traversability estimation over maps with tens of thousands of cells. Additionally, this simplifies the burden on the robot operator through the elimination of the α parameter.

$$p(T^{i,j}) = \prod_f \text{cdf}(\mathcal{N}(\mu^{f,i,j}, \sigma^{f,i,j}), f_{crit}) \quad (5)$$

IV. EXPERIMENTS, RESULTS AND ANALYSIS

In total, we collected high-quality laser scans of five environments at three active construction sites. Each scan consists of a single floor (two floors were collected at two sites). The above data generation process was run on each environment, yielding a total of around 30,000 train samples and 12,000 test samples (roughly a 70 – 30 test-train split). Additionally, a sixth environment (at a different site) was collected and used solely for testing. This site yielded an additional 7,000 test samples. UNRealNet was trained to predict $7m \times 7m$ local maps at a resolution of $5cm$, though maps of arbitrary size can be predicted using this pipeline. Our traversability features consisted of the following:

- 1) **terrain**: 1st percentile of heights in a cell.
- 2) **elevation**: 99th percentile of heights in a cell, after removing points over a certain height.
- 3) **step**: difference between elevation and terrain.
- 4) **local slope**: slope in a foothold-sized radius.
- 5) **local rough**: height variance in a foothold-sized radius.
- 6) **slope**: slope in a robot footprint-sized radius.
- 7) **rough**: height variance in a robot footprint radius.

Slope, step and roughness were computed in the same way as Fankhauser et al. [15]. In total, this resulted in a $7 \times 140 \times 140$ tensor of labels per datapoint.

A. Network Performance

In order to evaluate UNRealNet we ask the following:

- 1) Does UNRealNet outperform standard baselines?
- 2) Does the probabilistic traversability function better capture ground-truth traversability?

1) Does UNRealNet Outperform Standard Baselines?:

We first examine UNRealNet’s ability to produce traversability features from simulated lidar scans. In order to quantify this, we use the ground-truth map features for datapoints in our test set. Our primary baseline is the mapping algorithm used to generate features from the aggregated laser scans, but given the simulated pointcloud used as input. This baseline is chosen because it a perfectly observed pointcloud run through this baseline would generate zero error. Thus it serves as a good measure of how well each method can mitigate the effect of partial sensing.

Since the baseline local mapping method does not produce uncertainty estimates, we report RMSE between the mean of the predicted distribution and the ground-truth features for networks. We compare the following ablations:

- 1) (baseline): The local mapping method used to generate ground-truth features, applied to the synthetic pointcloud. Empty cells are filled with a value of zero.
- 2) (Telea): baseline, but leverages the inpainting method proposed by Telea et al. [35] for unknown cells.
- 3) (NS): baseline, but leverages the inpainting method based on Navier-Stokes proposed by Bertalmio et al. [36] for unknown cells.
- 4) (oracle): Equivalent to baseline, except the placeholder is determined by computing the RMSE-minimizing value for each channel (using the ground-truth).
- 5) (no aug): An ablation with the omission of the pointcloud augmentation pipeline. This ablation is provided to quantify the importance of making the simulated lidar reflect real sensing.

We then evaluate each method via the following metrics:

- 1) (both) avg. RMSE for cells with ground-truth values.
- 2) (observed) avg. RMSE for cells observed in the input pointcloud. This captures each method’s ability to correct for occlusion-induced mapping errors.
- 3) (inpaint) avg. RMSE for cells which are not observed in the pointcloud, but have ground-truth values. This captures the ability of each method to interpolate and extrapolate known information to unknown areas.

Results for synthetic scans on a spatially held-out region of each environment are reported in Table I. We also report results for the sixth environment (env6, in red to denote that it

Env	Method	RMSE (observed)	RMSE (inpaint)	RMSE (both)
env1	net (ours)	0.1562 ± 0.0001	0.1987 ± 0.0009	0.1849 ± 0.0006
	no aug	0.2117 ± 0.0101	0.2618 ± 0.0121	0.2453 ± 0.0113
	oracle	0.2817 ± 0.0003	0.2736 ± 0.0003	0.2779 ± 0.0003
	telea	0.2817 ± 0.0003	2.0402 ± 0.0017	1.6424 ± 0.0029
	ns	0.2817 ± 0.0003	0.3502 ± 0.0003	0.3298 ± 0.0003
baseline	0.2817 ± 0.0003	0.4097 ± 0.0007	0.3740 ± 0.0005	
env2	net (ours)	0.2718 ± 0.0030	0.3608 ± 0.0036	0.3284 ± 0.0035
	no aug	0.3313 ± 0.0141	0.4208 ± 0.0142	0.3877 ± 0.0137
	oracle	0.3731 ± 0.0016	0.4210 ± 0.0007	0.4080 ± 0.0011
	telea	0.3731 ± 0.0016	2.0930 ± 0.0154	1.6450 ± 0.0153
	ns	0.3731 ± 0.0016	0.4638 ± 0.0015	0.4327 ± 0.0017
baseline	0.3731 ± 0.0016	0.4659 ± 0.0007	0.4350 ± 0.0012	
env3	net (ours)	0.1705 ± 0.0021	0.2370 ± 0.0017	0.2172 ± 0.0017
	no aug	0.2319 ± 0.0127	0.2902 ± 0.0114	0.2724 ± 0.0111
	oracle	0.2669 ± 0.0005	0.2937 ± 0.0004	0.2883 ± 0.0004
	telea	0.2669 ± 0.0005	2.3610 ± 0.0058	1.9318 ± 0.0064
	ns	0.2669 ± 0.0005	0.3769 ± 0.0003	0.3474 ± 0.0003
baseline	0.2669 ± 0.0005	0.4370 ± 0.0002	0.3940 ± 0.0002	
env4	net (ours)	0.1413 ± 0.0008	0.2252 ± 0.0014	0.2071 ± 0.0012
	no aug	0.2033 ± 0.0099	0.2845 ± 0.0077	0.2663 ± 0.0078
	oracle	0.2629 ± 0.0001	0.2739 ± 0.0003	0.2728 ± 0.0002
	telea	0.2629 ± 0.0001	2.8353 ± 0.0028	2.4492 ± 0.0017
	ns	0.2629 ± 0.0001	0.4132 ± 0.0005	0.3830 ± 0.0005
baseline	0.2629 ± 0.0001	0.4165 ± 0.0009	0.3868 ± 0.0007	
env5	net (ours)	0.1748 ± 0.0016	0.2792 ± 0.0019	0.2483 ± 0.0016
	no aug	0.2313 ± 0.0104	0.3260 ± 0.0135	0.2969 ± 0.0114
	oracle	0.2439 ± 0.0005	0.3345 ± 0.0005	0.3084 ± 0.0002
	telea	0.2439 ± 0.0005	2.4452 ± 0.0238	1.9733 ± 0.0227
	ns	0.2439 ± 0.0005	0.3635 ± 0.0008	0.3297 ± 0.0008
baseline	0.2439 ± 0.0005	0.4263 ± 0.0005	0.3761 ± 0.0008	
env6	net (ours)	0.1920 ± 0.0012	0.2708 ± 0.0016	0.2537 ± 0.0015
	no aug	0.2466 ± 0.0087	0.3156 ± 0.0086	0.3001 ± 0.0079
	oracle	0.2985 ± 0.0004	0.2998 ± 0.0003	0.3023 ± 0.0003
	telea	0.2985 ± 0.0004	2.8835 ± 0.0098	2.4849 ± 0.0099
	ns	0.2985 ± 0.0004	0.4364 ± 0.0005	0.4092 ± 0.0005
baseline	0.2985 ± 0.0004	0.4598 ± 0.0008	0.4260 ± 0.0008	
env6, ouster	net (ours)	0.2200 ± 0.0014	0.2666 ± 0.0006	0.2567 ± 0.0007
	no aug	0.2217 ± 0.0018	0.2744 ± 0.0014	0.2628 ± 0.0014
	oracle	0.3229 ± 0.0007	0.2941 ± 0.0000	0.3046 ± 0.0002
	telea	0.3229 ± 0.0007	2.4206 ± 0.0213	2.1006 ± 0.0209
	ns	0.3229 ± 0.0007	0.4322 ± 0.0020	0.4118 ± 0.0017
baseline	0.3229 ± 0.0007	0.4658 ± 0.0003	0.4374 ± 0.0004	

TABLE I: Results for local mapping accuracy (env6 colored red to signify that it is a held-out location)

was not in the train set) to demonstrate UNRealNet’s ability to generalize to novel environments. We also report results from a robot traversal of env6 with the on-board lidar (env6, ouster). To get proper correspondences with the ground-truth map, we first align ground-truth cloud and the aggregated pointcloud via ground-plane matching and ICP. Since there is some drift in the SLAM solution, we further refine each individual scan by performing another ICP step for each individual scan. In total, this yields around 800 scans.

Standard deviations are reported over three seeds for *net* and *no aug*. Note that the augmentation pipeline introduces a small amount of variance in the baselines, which are otherwise deterministic. Also note that the RMSE on observed cells for all non-learned methods is nearly identical as they copy the values of observed cells.

Overall, we observe that UNRealNet outperforms all baselines, for all metrics, on all evaluation datasets. Interestingly, we observe that this performance gain is fairly similar for both observed and unobserved cells. This is likely due to factors such as occlusion and sparse sensing resulting in misestimation of terrain features (e.g. only observing the top part of an object and mistaking it for the ground). Furthermore, we observe that both inpainting baselines perform quite poorly, especially when compared to the oracular baseline. This is likely due to the relative sparsity of the input data (we only have returns for a relatively low fraction of cells, compared to traditional CV datasets for inpainting). Furthermore, data is sparser towards the edges of the map, meaning that extrapolation, in addition to interpolation, is necessary to achieve low RMSE. Finally, we observe that

Method	MAE
Det. [16,18]	0.3000 ± 0.0546
Prob. (ours)	0.2326 ± 0.0694

TABLE II: Probabilistic traversability experiment results.

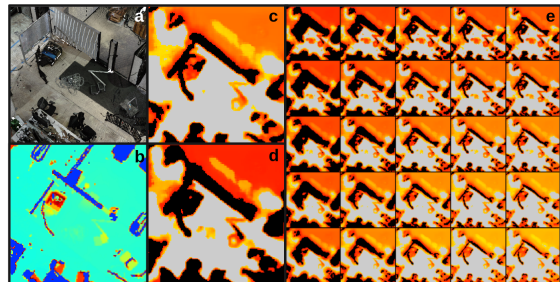


Fig. 5: Traversability predictions using the same network output on a sample in env6. (a) A crop from the FARO-scanned environment. There is a pile of wires in the middle. (b) The ground-truth elevation map. (c) The traversability map generated from parameters for a Spot. (d) The traversability map generated from parameters for an AlienGo. (e) A mosaic of traversability maps by varying the local slope (y axis) and robot slope (x axis) thresholds.

adding augmentations similar to the actual robot payload improves overall performance.

2) *Does the Probabilistic Traversability Function Better Capture Ground-Truth?:* In order to evaluate the efficacy of the probabilistic traversability function, we compare the accuracy to a ground-truth traversability derived from crops of the global map. For each data point, a random set of traversability thresholds was generated. These traversability thresholds were used to generate a traversability map by checking if any ground-truth features exceeded their corresponding threshold, per cell. We then evaluated the proposed traversability function (Prob.) and the one used by Chilian and Fankhauser [15,16] (Det.), on the predicted traversability features from the network. We report the MAE of the predicted traversability value to the ground truth map. Results are presented in Table II for the test set, over a random sampling of traversability thresholds.

Overall, we observe that the uncertainty-aware traversability function is closer to the ground-truth traversability. This is likely because incorporating uncertainty allows the traversability estimation to be less sensitive to mispredictions of features (due to assigning high uncertainty to said areas).

B. Hardware Experiments

We also ran a number of experiments on multiple legged platforms to demonstrate the efficacy of our method in practice. Our traversability pipeline was run on both a Boston Dynamics Spot and Unitree Aliengo, each with a custom payload similar to Bouman et al. [37].

1) *Robot-Agnosticism:* Presented in Figure 5 are multiple traversability maps generated from the same UNRealNet prediction from env 6. Our traversability function is able to take into account robot-specific mobility limitations and produce a corresponding cost. Of note is that UNRealNet

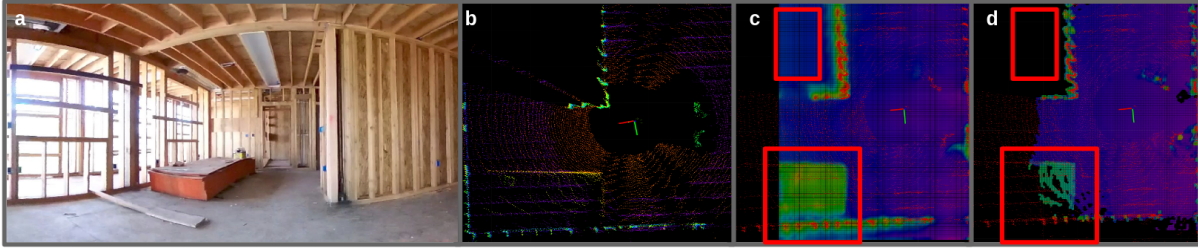


Fig. 6: Case study for elevation prediction. (a) An FPV image of the scene. (b) The input pointcloud (red=low, purple=high) (c) The learned, Kalman-filtered elevation map from UNRealNet (purple=low, red=high) has inpainted the floor behind the wall, as well as the construction material and pillars. (d) Elevation map from STEP [24] is incomplete due to occlusions.

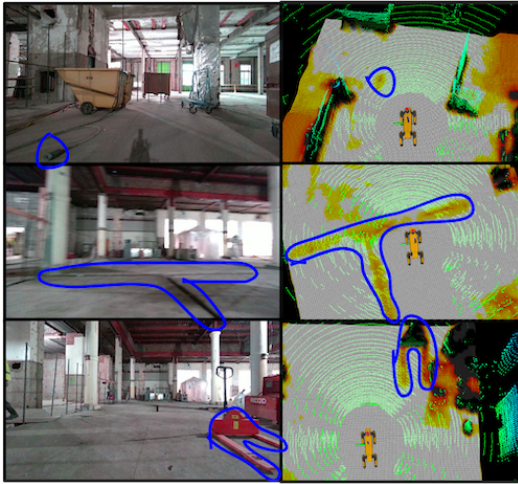


Fig. 7: Additional qualitative results from a seventh environment. We are able to capture detailed traversability estimates including small plugs (top), shallow negative obstacles (middle) and complex geometry (bottom). Correspondences between FPV and BEV are circled in blue. FPV viewpoints were taken at different times than BEV to keep the object of interest in the field of view.

is able to detect and evaluate traversability of a small pile of wires, which becomes more costly the more stringent the traversability thresholds become.

2) *Kalman Filter*: When deploying on the robot, we can fuse multiple predictions together to get a higher-quality map (at the cost of becoming reliant on accurate odometry). This is accomplished by leveraging the uncertainty in the neural network predictions (μ, σ , per feature, per cell) to run a one-dimensional Kalman Filter (Equation 6, where $\hat{\mu}, \hat{\sigma}$ are the running estimates of the Kalman filter) [38]. Traversability $\hat{T}^{i,j}$ can then be computed from the Kalman-filtered estimates ($\hat{\mu}, \hat{\sigma}$). We observe that the Kalman filter generally improves the estimation of terrain features.

$$\hat{\mu}' = \frac{\hat{\mu}\sigma^2 + \mu\hat{\sigma}^2}{\hat{\sigma}^2 + \sigma^2}, \quad \hat{\sigma}' = \sqrt{\frac{\hat{\sigma}^2\sigma^2}{\hat{\sigma}^2 + \sigma^2}} \quad (6)$$

3) *Inference Speed*: We first report the speed of our method in comparison to STEP [24]. Both methods were run with an AMD Ryzen 9 5900hx CPU and NVIDIA 3060 Laptop GPU. We observe that although slower, the runtime

of our method is comparable to STEP (around 7hz compared to STEP's 13hz, and 10hz rate of pointclouds from the lidar). The majority of the runtime in our method is consumed by preprocessing the pointcloud for the network, and converting the network's output into a grid map. We argue that much of this runtime can be reduced by re-implementing our Python inference node in C++, and leveraging tools such as TensorRT [39]. Additionally, we note that 7hz inference is fast enough for legged platforms with a top speed of under 2m/s, though this may not be sufficient for faster platforms.

4) *Case Studies*: One particular case study is presented in Figure 6. We compare the learned elevation map to the elevation map produced by STEP. We can observe that UNRealNet has learned to extrapolate using environmental priors, in that it successfully inpaints the crate, pillars, and the floor in between. Additionally, it correctly predicts the floor behind the wall. We also provide additional qualitative results from a seventh environment (for which we didn't have a ground-truth pointcloud) in Figure 7 that highlight additional nuances in our traversability estimation.

V. CONCLUSION AND FUTURE WORK

We presented a method for learning traversability information by generating synthetic sensing from high-quality scans of real-world environments. We demonstrate that this method results in improved local mapping performance and demonstrate its efficacy for traversability estimation on hardware.

Future work includes generating better training labels, such as correspondences between observed terrain features and proprioception [40–42] or demonstrations [43–45]. The work done by Meng et al. [6] provides a good roadmap for how to incorporate semantics into a similar mapping pipeline. Additionally, datasets such as Matterport 3d [21] and TartanAir [8] can be leveraged to obtain joint geometric and semantic labels without any additional human labeling (although fine-tuning on real data may be required). Lastly, combining UNRealNet with object detection and tracking algorithms would allow for increased performance in the presence of dynamic obstacles.

VI. ACKNOWLEDGEMENTS

The authors thank the rest of the Field AI team for their help in collecting data and testing on hardware.

REFERENCES

- [1] C. D. Bellicoso, M. Bjelonic, L. Wellhausen *et al.*, “Advances in real-world applications for legged robots,” *Journal of Field Robotics*, vol. 35, no. 8, pp. 1311–1326, 2018.
- [2] K. Afsari, S. Halder, M. Ensafi *et al.*, “Fundamentals and prospects of four-legged robot application in construction progress monitoring,” *EPiC Series in Built Environment*, vol. 2, pp. 274–283, 2021.
- [3] C. Gehring, P. Fankhauser, L. Isler *et al.*, “Anymal in the field: Solving industrial inspection of an offshore hvdc platform with a quadrupedal robot,” in *Field and Service Robotics: Results of the 12th International Conference*. Springer, 2021, pp. 247–260.
- [4] A. Agha, K. Otsu, B. Morrell *et al.*, “Nebula: Quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge,” *arXiv preprint arXiv:2103.11470*, 2021.
- [5] S. Scherer, V. Agrawal, G. Best *et al.*, “Resilient and modular subterranean exploration with a team of roving and flying robots,” *Field Robotics Journal*, pp. 678–734, May 2022.
- [6] X. Meng, N. Hatch, A. Lambert *et al.*, “Terrainnet: Visual modeling of complex terrain for high-speed, off-road navigation,” *arXiv preprint arXiv:2303.15771*, 2023.
- [7] L. Gan, Y. Kim, J. W. Grizzle *et al.*, “Multitask learning for scalable and dense multilayer bayesian map inference,” *IEEE Transactions on Robotics*, vol. 39, no. 1, pp. 699–717, 2022.
- [8] W. Wang, D. Zhu, X. Wang *et al.*, “Tartanair: A dataset to push the limits of visual slam,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4909–4916.
- [9] A. G. Kendall, “Geometry and uncertainty in deep learning for computer vision,” Ph.D. dissertation, University of Cambridge, UK, 2019.
- [10] G. Georgakis, B. Bucher, A. Arapin *et al.*, “Uncertainty-driven planner for exploration and navigation,” in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 11 295–11 302.
- [11] S. R. Team, “Stanford racing team’s entry in the 2005 darpa grand challenge,” *Published on DARPA Grand Challenge*, 2005.
- [12] L. D. Jackel, E. Krotkov, M. Perschbacher *et al.*, “The darpa lagr program: Goals, challenges, methodology, and phase i results,” *Journal of Field robotics*, vol. 23, no. 11-12, pp. 945–973, 2006.
- [13] P. Papadakis, “Terrain traversability analysis methods for unmanned ground vehicles: A survey,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1373–1385, 2013.
- [14] “Arl autonomy stack,” 2022, accessed 13-September-2022. [Online]. Available: <https://www.arl.army.mil/business/collaborative-alliances/current-cras/sara-cra/sara-overview/>
- [15] P. Fankhauser, M. Bjelonic, C. D. Bellicoso *et al.*, “Robust rough-terrain locomotion with a quadrupedal robot,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 5761–5768.
- [16] A. Chilian and H. Hirschmüller, “Stereo camera based navigation of mobile robots on rough terrain,” in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 4571–4576.
- [17] J. Larson and M. Trivedi, “Lidar based off-road negative obstacle detection and analysis,” in *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2011, pp. 192–197.
- [18] P. Fankhauser and M. Hutter, “A universal grid map library: Implementation and use case for rough terrain navigation,” in *Robot Operating System (ROS)*. Springer, 2016, pp. 99–120.
- [19] T. Miki, L. Wellhausen, R. Grandia *et al.*, “Elevation mapping for locomotion and navigation using gpu,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2273–2280.
- [20] S. K. Ramakrishnan, Z. Al-Halah, and K. Grauman, “Occupancy anticipation for efficient exploration and navigation,” in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 2020, pp. 400–418.
- [21] A. Chang, A. Dai, T. Funkhouser *et al.*, “Matterport3d: Learning from rgb-d data in indoor environments,” *arXiv preprint arXiv:1709.06158*, 2017.
- [22] F. Xia, A. R. Zamir, Z. He *et al.*, “Gibson env: Real-world perception for embodied agents,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9068–9079.
- [23] G. Georgakis, B. Bucher, K. Schmeckpeper *et al.*, “Learning to map for active semantic goal navigation,” *arXiv preprint arXiv:2106.15648*, 2021.
- [24] D. D. Fan, K. Otsu, Y. Kubo *et al.*, “Step: Stochastic traversability evaluation and planning for risk-aware off-road navigation,” *arXiv preprint arXiv:2103.02828*, 2021.
- [25] D. D. Fan, A.-A. Agha-Mohammadi, and E. A. Theodorou, “Learning risk-aware costmaps for traversability in challenging environments,” *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 279–286, 2021.
- [26] M. Stölzle, T. Miki, L. Gerdes *et al.*, “Reconstructing occluded elevation information in terrain maps with self-supervised learning,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1697–1704, 2022.
- [27] “Faro focus datasheet,” 2023, accessed 04-September-2023. [Online]. Available: https://media.faro.com/-/media/Project/FARO/FARO/RO/2.TECH-SHEET/TechSheet_Focus_Laser_Scanner/TechSheet_Focus_Laser_Scanner_EN.pdf?rev=1291d7f5ef814fe99d471dde60321833
- [28] W. Zhang, J. Qi, P. Wan *et al.*, “An easy-to-use airborne lidar data filtering method based on cloth simulation,” *Remote sensing*, vol. 8, no. 6, p. 501, 2016.
- [29] M. Velas, M. Spanel, M. Hradis, and A. Herout, “Cnn for very fast ground segmentation in velodyne lidar data,” in *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 2018, pp. 97–103.
- [30] F. Langer, A. Milioto, A. Haag *et al.*, “Domain transfer for semantic segmentation of lidar data using deep neural networks,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 8263–8270.
- [31] “Os1 mid-range high-resolution imaging lidar,” 2021, accessed 01-September-2023. [Online]. Available: <https://data.ouster.io/downloads/datasheets/datasheet-revd-v2p0-os1.pdf>
- [32] A. H. Lang, S. Vora, H. Caesar *et al.*, “Pointpillars: Fast encoders for object detection from point clouds,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 697–12 705.
- [33] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [34] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.
- [35] A. Telea, “An image inpainting technique based on the fast marching method,” *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [36] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, “Navier-stokes, fluid dynamics, and image and video inpainting,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.
- [37] A. Bouman, M. F. Ginting, N. Alatur *et al.*, “Autonomous spot: Long-range autonomous exploration of extreme environments with legged locomotion,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 2518–2525.
- [38] P. Fankhauser, M. Bloesch, and M. Hutter, “Probabilistic terrain mapping for mobile robots with uncertain localization,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3019–3026, 2018.
- [39] “Nvidia tensorrt,” 2023, accessed 01-September-2023. [Online]. Available: <https://developer.nvidia.com/tensorrt>
- [40] M. G. Castro, S. Triest, W. Wang *et al.*, “How does it feel? self-supervised costmap learning for off-road vehicle traversability,” *arXiv preprint arXiv:2209.10788*, 2022.
- [41] X. Cai, M. Everett, J. Fink, and J. P. How, “Risk-aware off-road navigation via a learned speed distribution map,” *arXiv preprint arXiv:2203.13429*, 2022.
- [42] J. Frey, M. Mattamala, N. Chebrolu *et al.*, “Fast traversability estimation for wild visual navigation,” *arXiv preprint arXiv:2305.08510*, 2023.
- [43] J. A. Bagnell, D. Bradley, D. Silver *et al.*, “Learning for autonomous navigation,” *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 74–84, 2010.
- [44] M. Wigness, J. G. Rogers, and L. E. Navarro-Serment, “Robot navigation from human demonstration: Learning control behaviors,” in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 1150–1157.

- [45] S. Triest, M. G. Castro, P. Maheshwari *et al.*, “Learning risk-aware costmaps via inverse reinforcement learning for off-road navigation,” *arXiv preprint arXiv:2302.00134*, 2023.