LiDAR-based Traversability Estimation for Ground Robots on Construction Sites using Self-Supervised Learning

Jinhee Yu[1], Swayamjit Saha[1], Monika Jayakumar[1], Mikias Gugssa[2], Jingdao Chen[1], and Jun Wang[2]

[1]Computer Science and Engineering, Mississippi State University, Mississippi State, United States, Email: jy603@msstate.edu, ss4706@msstate.edu, mj1599@msstate.edu, chenjingdao@cse.msstate.edu

[2]Civil and Environmental Engineering, Mississippi State University, Mississippi State, United States, Email: mgw285@msstate.edu, jwang@cee.msstate.edu

ABSTRACT

Navigating autonomous robots through the intricate landscapes of construction sites necessitates an accurate evaluation of traversability, a challenge that our research addresses with a self-supervised learning strategy utilizing 3D LiDAR point cloud data. This approach enables ground robots to autonomously identify traversable areas based on the spatial relationships between their trajectories and the surrounding environmental features. By employing self-supervised trajectory area labeling and Positive-Unlabeled (PU) learning, our method classifies points as traversable or non-traversable without labor-intensive manual annotations. Utilizing the RELLIS-3D and Hilti datasets, coupled with Simultaneous Localization and Mapping (SLAM) for trajectory mapping, we conducted qualitative and quantitative assessments with vehicle/robot-mounted LiDAR systems. Results show that the proposed method can reasonably identify traversable areas from LiDAR data for robots in off-road environments as well as robots on construction sites.

INTRODUCTION

The construction industry, a cornerstone of modern infrastructure development, is undergoing a transformative shift with the infusion of robotic technologies such as autonomous excavators (Guan et al. 2021b) and autonomous dump trucks (Komatsu et al. 2021). These advancements highlight the growing role of robotics in shaping the future of work in construction environments. Building on these advancements, the incorporation of sophisticated sensors and a diverse array of learning algorithms into robots equips them to handle the unpredictable nature of construction sites, supporting a variety of tasks including navigation (Jeong et al. 2021) and terrain traversability (Sevastopoulos and Konstantopoulos 2022), alongside material transport and excavation with Autonomous Excavator Systems (AES) (Zhang et al. 2021), quality inspection (Halder and Afsari 2023), surveying, 3D printing (Hossain et al. 2020), logistics, and safety oversight.

The challenge in deploying robots in construction sites lies in the unpredictable and dynamic environments typical of construction sites due to changing terrain, moving equipment, materials, and workers. In order to advance robotic technology, much research is needed to realize robot navigation in construction sites and allow robots to traverse challenging terrain safely. The concept of traversability (Sevastopoulos and Konstantopoulos 2022), refers to a robot's ability to successfully navigate through an area, considering terrain features such as inclines, obstacles, states of the surface such as dirt or concrete, and overall conditions that affect the robot's ability to move efficiently and safely in the environment. Despite progress and highly impactful applications, the dynamic and unpredictable nature of construction sites presents unique challenges for robotic navigation, which is largely influenced by factors such as the steepness of slopes, ground cover, and the presence of stairs or irregularities. The terrain's texture, firmness, and the presence of obstacles also significantly affect a robot's traversability.

In this research, we utilize robots equipped with LiDAR sensors that can capture the geometry of the surrounding environment in the form of a 3D point cloud and use the point cloud to assess terrain traversability. The tasks for the robot can be defined as follows: **Point cloud Segmentation:** Automatically classifying each point in a 3D point cloud acquired by the LiDAR scanner of the robot. **Traversability Estimation:** Performing binary point cloud segmentation where the target classes are (i) traversable and (ii) non-traversable. *Traversable points* should be areas which the robot can navigate through, such as ground or a platform, whereas *non-traversable points* should be areas which the robot cannot navigate through, such as walls, temporary structures, or workers. Supervised learning is a popular machine learning technique where a model can be trained to recognize different types of terrain from an input point cloud. However, the issue with using supervised learning for traversability estimation is that it requires a lot of training data and is extremely time-consuming to annotate.

To address these challenges, this study proposes the application of self-supervised learning (SSL), a methodology where an AI robotic system autonomously interprets and labels data in a self-supervised manner, thus minimizing the dependency on extensive human-annotated datasets (Sevastopoulos and Konstantopoulos 2022), which is required for purely supervised learning. Reinforcement learning (RL) is a related concept in which robotic agents learn optimal behaviors through direct interactions with their environment. The disadvantage of reinforcement learning is that it requires long training times to converge and design a custom reward function (Sevastopoulos and Konstantopoulos 2022); whereas our self-supervised learning approach directly uses demonstration trajectories to provide training data. Self-supervised learning techniques for robot traversability estimation have been previously explored in works such as BADGR (Kahn et al. 2021) and ScaTE (Seo et al. 2023). Our study extends previous works in this area by using a more advanced neural network for point cloud processing, proposing an improved traversability labeling process, and validating the method on construction site data.

LITERATURE REVIEW

Traversability estimation for robots

In the pursuit of safe autonomous robot navigation, traditional methods like those relying on thermal inertia for slip measurements (Cunningham et al. 2019) may be effective for open, sandy terrain but impractical for more complex construction terrain with obstacles. Similarly, RADAR systems, though useful, can be too bulky, which may exceed the load capacities for smaller robotic platforms (Wellhausen et al. 2019). More recent research has increasingly relied on camera (Hirose et al. 2018) and LiDAR (Seo et al. 2023) to obtain high-resolution sensor data that can be used to assess traversability. These methods are often paired with Neural Networks (Pomerleau 1992) or Gaussian Process Regression (Ho et al. 2013) to process the high-dimensional sensor data.

Traversability estimation for construction sites

In the field of construction site traversability, the Terrain Traversability Mapping (Guan et al. 2021a) (TTM) and Terrain Traversability Mapping and Navigation System (Guan et al. 2021b) (TNS) were proposed to guide autonomous excavators through complex terrains like construction sites using cameras and LiDAR data. TTM focuses on detailed mapping using a learning-based geometric fusion method and the Autonomous Excavator Terrain (AET) dataset, while TNS employs additional tools like GPS, other navigation aids, and the Complex Worksite Terrain (CWT) dataset for dynamic terrain analysis and generate real-time terrain information. The disadvantage of these methods is that they still rely on training on pre-labeled datasets, which makes them less adaptable and efficient for general-purpose construction site navigation.

3D neural networks for point cloud data

Advancements in 3D deep learning have given rise to neural network architectures that are specifically tailored for processing point clouds. PointNet (Qi et al. 2017) is a pioneering work that directly processes point clouds without the need for voxelization or rendering, thus preserving the fidelity of the original data. Later work includes the Cylindrical and Asymmetrical 3D Convolution Network for LiDAR segmentation (Zhu et al. 2021), which addresses the sparsity of LiDAR data by exploiting cylindrical 3D convolutions, effectively capturing the unique distributions of points in outdoor scenes. Similarly, Spherical Transformer (Lai et al. 2023) was introduced to harness spherical data representations to enlarge the receptive field and improve the recognition performance for sparse distant points. These methods, while effective when training data is available, still operate within the supervised learning paradigm and require extensive manual annotations.

Semi-supervised and self-supervised learning

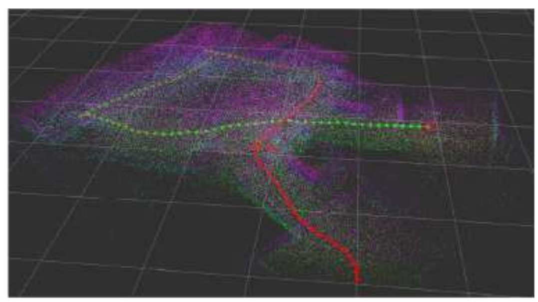
Semi-supervised models like GONet, enhanced by temporal consistency (GONet+T) and stereo imaging (GONet+TS) (Hirose et al. 2018), have made strides in the field, using a DCGAN framework to assess traversability from fisheye camera images. However, they still partially require hand-labeled image data, which can be a limiting factor. ScaTE (Scalable Framework for Self-Supervised Traversability Estimation) (Seo et al.

2023) was later introduced for self-supervised learning of vehicle-specific traversability in dynamic, unpredictable, and unstructured environments where labeling data is challenging or impractical. Utilizing Positive-Unlabeled (PU) learning, ScaTE enhances the capability to accurately identify non-traversable areas, crucial for construction site navigation.

METHODOLOGY

This study proposes an approach for predicting the traversability of autonomous vehicles and robots in construction sites. This approach is based on self-supervised learning methods using past trajectory paths and the corresponding LiDAR point cloud data. The primary objective of this research is to accurately predict traversable areas from LiDAR points by comprehensively considering the terrain characteristics and the driving capabilities and characteristics of the vehicle. To achieve this, Simultaneous Localization and Mapping (SLAM) is used to reconstruct past trajectories of the robot, and areas where the robot has actually traversed are labeled as traversable. These labels are then used to train a deep neural network to generalize and infer the feature characteristics of traversable points and finally to predict the traversability for new, unseen LiDAR data.

(a) Camera Image



(b) Point Cloud Map and Robot Trajectory

Fig. 1. SLAM results for Site 2 of the Hilti 2023 dataset

Data Collection

To benchmark traversability estimation, this study first considers the RELLIS-3D dataset (Jiang et al. 2020), a multimodal off-road dataset that contains sensor data from off-road environments with varied environmental topography. The dataset contains 13,556 LiDAR scans and 6,235 images collected using a Clearpath Warthog robot from the Texas A&M Rellis campus. The Clearpath Warthog unmanned ground vehicle robot is a large-sized robotic platform used to monitor and record site conditions from varied topographical terrains (Clearpath 2022). The sensor setup includes a 64-channel Ouster OS1 LiDAR, a 32-channel Velodyne Ultra Puck LiDAR, a 3D stereo Camera, an RGB Camera, and an Inertial Navigation System (GPS/IMU). In this study, we only used the LiDAR scans from the 64-channel Ouster OS1 LiDAR because the dataset only contained sensor pose information for the Ouster LiDAR. The RELLIS-3D dataset is annotated through crowdsourcing to obtain pixel-wise image labels, then the image labels are projected to point cloud labels through camera-LiDAR calibration and further refined. The labels include 20 semantic classes which are sky, grass, tree, bush, concrete, mud, person, puddle, rubble, barrier, log, fence, vehicle, object, pole, water, asphalt, building and dirt and deep water.

In addition, this study uses the Hilti dataset (Zhang et al. 2023) to evaluate traversability estimation from multiple active construction site environments. The data was collected from a tracked robot platform, the Hilti Jaibot. We used the Site 2 dataset, which was collected from a large parking lot under construction. The robot sensor suite comprises a Robosense BPearl hemisphere LiDAR, an Xsens MTi-670 IMU, and 4x OAK-D cameras (Zhang et al. 2023). The number of LiDAR scans from the robot sensor suite was 3059 scans and the time for data collection was 305 seconds. Note that the Hilti dataset was originally used as a benchmark dataset for SLAM algorithms, but in this study, we use the dataset to evaluate traversability of construction environments. The Hilti dataset does not contain any semantic labels; in this study, we will use self-supervised learning to predict the traversable areas.

Simultaneous Localization and Mapping

Simultaneous Localization and Mapping (SLAM) is the process of simultaneously building a map and localizing the robot's position in it. Our study uses the *hdl_graph_slam* ROS package (Koide et al. 2019), a popular open-source ROS package, to obtain the 6-DOF SLAM poses from the robot sensor data for the Hilti dataset. *hdl_graph_slam* is built upon 3D Graph SLAM using NDT scan matching for odometry estimation and loop detection. Even though the SLAM framework supports GPS and IMU as graph constraints, in this study, we only used the 3D LiDAR scans to estimate robot pose. Figure 1 shows the trajectory map generated by SLAM along with a corresponding sample image captured directly by the camera for the ground robot in the Hilti dataset.

Self-supervised Trajectory Area Labeling

The Trajectory Area Labeling process designed in this study aims to automatically identify the path traversed by autonomous vehicles or wheeled robots from reference trajectories or past trajectories. This process utilizes two Oriented Bounding Boxes (OBBs) to accurately label areas that vehicles or robots can traverse. The first bounding box covers the area from the ground to the top of the robot, while the second bounding box includes the area from the clearance point to the top of the robot. By evaluating the traversability within the areas defined by these two bounding boxes, the traversability status of points within the point cloud scan for each trajectory path point is determined. Points not covered by the path of these bounding boxes (areas not included in the path) are considered non-labeled, ensuring that only areas directly relevant to the vehicle's or robot's potential path are assessed for traversability.

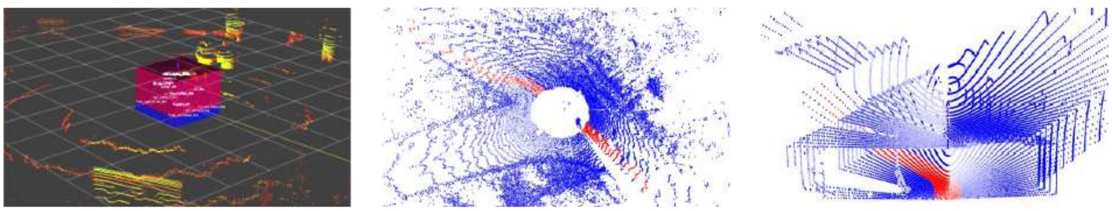
To enhance the precision of the traversability assessment, this study further subdivides the bounding boxes into smaller units. This subdivision is necessary to more accurately distinguish between obstacles of various sizes (e.g., pebbles, grass, people, pillars) that may coexist within the bounding box. Using only large bounding boxes could lead to inaccuracies in distinguishing between obstacles, whereas evaluating traversability independently within each subdivided bounding box achieves more accurate labeling.

For points $\mathbf{p}$ within the subdivided bounding box, the following condition determines if the point is labeled as traversable:

$$\mathbf{p} \in \mathrm{OBB}_{\text{Ground to Top}} \land \mathbf{p} \notin \mathrm{OBB}_{\text{Chassis to Top}}$$

where $\mathbf{p}$ represents points within the subdivided bounding box, $\mathrm{OBB}_{\text{Ground to Top}}$ represents the bounding box covering the area from the ground to the top, and $\mathrm{OBB}_{\text{Chassis to Top}}$ represents the bounding box covering the area from the clearance point to the top.

This condition implies that if all points $\mathbf{p}$ within a subdivided bounding box are included in the Ground to Top bounding box and simultaneously not included in the Chassis to Top bounding box, then all points within that bounding box are considered traversable. This provides a criterion for judging the traversable areas in the presence of static and transient obstacles along the vehicle's trajectory. Figure 2's (b) and (c) illustrate the labeled traversable points in point cloud data.



(a) (b) (c)

Fig. 2. Self-supervised traversability labeling process: (a) Oriented Bounding Boxes covering the space around the robot (b) Traversable points in the RELLIS-3D dataset (c) Traversable points in the Hilti dataset. Red points indicate traversable points and blue points indicate unlabeled points

Traversability Classification

Based on the data labeled through the aforementioned method, this study employs the SphereFormer model (Lai et al. 2023), a LiDAR semantic segmentation model, to classify traversability. During this process, the model's final linear layer is fine-tuned for the binary classification task. In the case of traditional supervised learning for traversability estimation, PN (Positive-Negative) Learning is used where points are labeled as either traversable (Positive) or non-traversable (Negative). In contrast, for self-supervised learning, we have to apply PU (Positive-Unlabeled) Learning, where labeled points are considered positive (traversable), and the remaining points are treated as non-labeled and can be either traversable or non-traversable. In this study, Non-negative PU (nnPU) loss is applied to enhance the stability of the learning process, with the nnPU loss's prior set to 0.3. PU learning is potentially advantageous compared to PN learning, because in the process of self-supervised trajectory area labeling, areas that were not previously traversed by the robot are not strictly "non-traversable", and may be better considered as "unlabeled" areas. Our implementation for nnPU loss is adapted from (Kiryo et al. 2017).

RESULTS

Experimental Setup

The validity of the proposed methodology is verified using the RELLIS-3D and Hilti datasets. For the RELLIS-3D dataset, scenes 00, 01, and 02 are used as the training set, and scenes 03 and 04 are used as the test set. In the Hilti dataset, robot LiDAR data from Site 2 Floor 1 Large room serves as the training set, and Site 2 Robot Floor 2 Large room -dark is used for testing. Test-Time Augmentation (TTA) techniques such as flip and rotation were applied during the testing process. To assess the model's prediction sensitivity and the confidence of its predictions, the final logits output of the neural network was evaluated at confidence thresholds ranging from 0.5 to 0.9.

Evaluation Metrics

Performance evaluation is conducted through both qualitative and quantitative analyses. For the RELLIS-3D dataset, since ground truth labels for semantic segmentation are provided, points belonging to classes such as grass, mud, puddle, asphalt, concrete, rubble, and dirt are considered traversable. Metrics such as accuracy, precision, recall, and F1 score are used to evaluate the model's performance quantitatively, and qualitative evaluations are also conducted by visually inspecting the predicted traversable areas. For the Hilti dataset, only qualitative analysis is performed as ground truth is not provided.

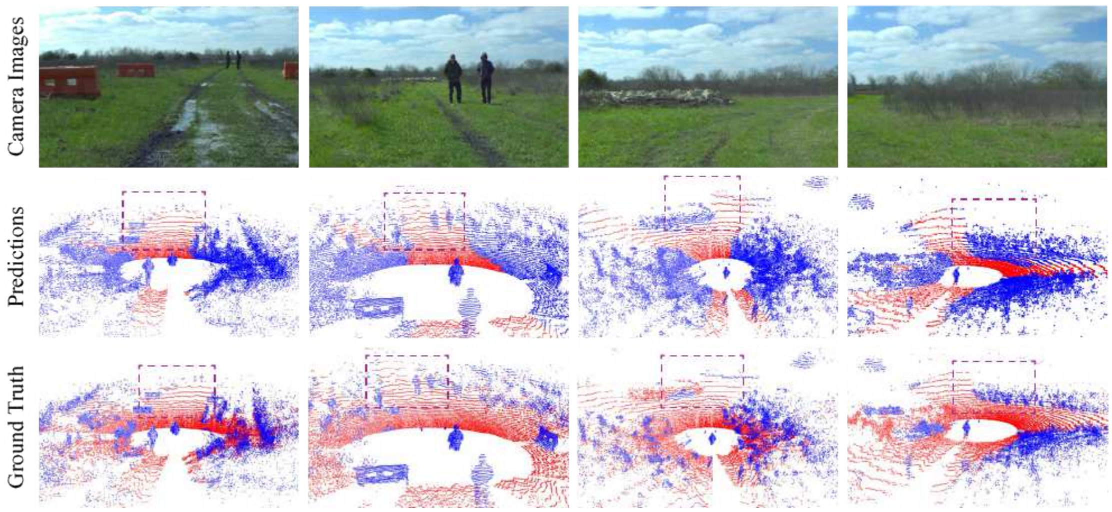
Traversability estimation results on RELLIS 3D dataset

The qualitative results exhibit relatively high precision contrasted with low recall, and the quantitative results also show that the model predicts traversable areas more conservatively. This could be attributed to the nnPU loss function, which assumes that the distributions of positive (traversable) and negative (non-traversable) data do not overlap when calculating positive and negative risks. In reality, the distributions of traversable and non-traversable areas overlap to some extent. This overlap can lead to the model predicting traversable areas more conservatively. Moreover, the task's inherently safety-critical nature necessitates a conservative prior setting for traversability assessment, set to 0.3. This conservative prior encourages the model to be more cautious in predicting traversable areas, increasing precision but reducing recall. The potential inaccuracy of the point cloud segmentation labels provided by the RELLIS-3D dataset must also be considered. However, as seen in Table 1 and Figure 3, the model accurately predicts traversable areas, indicating reliable performance in identifying genuinely traversable regions despite conservative predictions.

Table 1. Performance metrics for the SphereFormer model on the RELLIS 3D dataset at different output thresholds

Output Thresholds	Accuracy	Precision	Recall	F1 Score
0.5	0.78	0.89	0.32	0.47
0.7	0.78	0.93	0.32	0.47
0.9	0.78	0.97	0.31	0.47

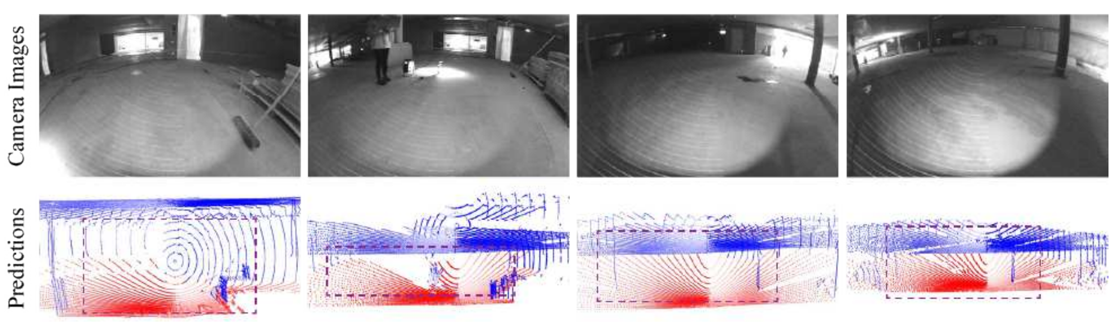
Fig. 3. Traversability estimation results from 3D point clouds in the RELLIS-3D dataset. The purple dashed boxes highlight the regions in the point clouds that correspond to the camera images above.



Qualitative results on Hilti dataset

Figure 4 shows the traversability estimation results for a construction site from the Hilti dataset. Overall, the method is able to predict ground points as traversable and columns, walls, and other obstructions as non-traversable.

Fig. 4. Traversability estimation results from 3D point clouds in the Hilti dataset along with corresponding camera images.



CONCLUSION

In conclusion, this study demonstrated a self-supervised learning method for traversability estimation on construction sites where a ground robot can learn to navigate through reference demonstration trajectories instead of manual annotations, thus saving valuable annotation time. The proposed trajectory area labeling approach accurately labels the traversable parts of the vehicle's path across various environments and transient objects on the vehicle's trajectory. However, the method shows limitations in recognizing flexible objects, such as curtains and tall grass, as traversable, even when they do not impede navigation. Furthermore, it relies predominantly on geometric features to assess traversability, which poses challenges in accurately identifying traversable surfaces like calm water due to LiDAR data's limited ability to capture texture details. To address these challenges and enhance the method's effectiveness, future research will be directed toward developing and integrating strategies to overcome the identified limitations. Moreover, the training and test sets we used primarily represent specific construction environments, which may lead to overfitting and cause the model to perform poorly in unfamiliar settings. A broader range of datasets and testing scenarios is necessary to fully evaluate the method's adaptability. Future work will expand the method to construction sites with more challenging terrain features and obstacles, such as uneven ground, raised floors, and stairs.

Acknowledgements

The work reported herein was supported by the National Science Foundation (NSF) (Award #IIS-2153101). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

Clearpath (2022). "Warthog unmanned ground vehicle, <https://clearpathrobotics.com/warthog-unmanned-ground-vehicle-robot/>. Accessed on 2024-01-26.

Cunningham, C., Nesnas, I. A., and Whittaker, W. L. (2019). "Improving slip prediction on mars using thermal inertia measurements." *Autonomous Robots*, 43, 503–521.

Guan, T., He, Z., Manocha, D., and Zhang, L. (2021a). "Ttm: Terrain traversability mapping for autonomous excavator navigation in unstructured environments." *arXiv preprint arXiv:2109.06250*.

Guan, T., He, Z., Song, R., Manocha, D., and Zhang, L. (2021b). "Tns: Terrain traversability mapping and navigation system for autonomous excavators." *arXiv preprint arXiv:2109.06250*.

Halder, S. and Afsari, K. (2023). "Robots in inspection and monitoring of buildings and infrastructure: A systematic review." *Applied Sciences*, 13(4), 2304.

Hirose, N., Sadeghian, A., Vázquez, M., Goebel, P., and Savarese, S. (2018). "Gonet: A semi-supervised deep learning approach for traversability estimation." *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 3044–3051.

Ho, K., Peynot, T., and Sukkarieh, S. (2013). "Traversability estimation for a planetary rover via experimental kernel learning in a gaussian process framework." *2013 IEEE International Conference on Robotics and Automation*, IEEE, 3475–3482.

Hossain, M. A., Zhumabekova, A., Paul, S. C., and Kim, J. R. (2020). "A review of 3d printing in construction and its impact on the labor market." *Sustainability*, 12(20), 8492.

Jeong, I., Jang, Y., Park, J., and Cho, Y. K. (2021). “Motion planning of mobile robots for autonomous navigation on uneven ground surfaces.” *Journal of Computing in Civil Engineering*, 35(3), 04021001.

Jiang, P., Osteen, P., Wigness, M., and Saripalli, S. (2020). “Rellis-3d dataset: Data, benchmarks and analysis.” *arXiv preprint arXiv:2011.12954*.

Kahn, G., Abbeel, P., and Levine, S. (2021). “Badgr: An autonomous self-supervised learning-based navigation system.” *IEEE Robotics and Automation Letters*, 6(2), 1312–1319.

Kiryo, R., Niu, G., Du Plessis, M. C., and Sugiyama, M. (2017). “Positive-unlabeled learning with non-negative risk estimator.” *Advances in neural information processing systems*, 30.

Koide, K., Miura, J., and Menegatti, E. (2019). “A portable three-dimensional lidar-based system for long-term and wide-area people behavior measurement.” *International Journal of Advanced Robotic Systems*, 16(2), 1729881419841532.

Komatsu, T., Konno, Y., Kiribayashi, S., Nagatani, K., Suzuki, T., Ohno, K., Suzuki, T., Miyamoto, N., Shibata, Y., and Asano, K. (2021). “Autonomous driving of six-wheeled dump truck with a retrofitted robot.” *Field and Service Robotics*, G. Ishigami and K. Yoshida, eds., Singapore, Springer Singapore, 59–72.

Lai, X., Chen, Y., Lu, F., Liu, J., and Jia, J. (2023). “Spherical transformer for lidar-based 3d recognition.” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17545–17555.

Pomerleau, D. A. (1992). “Progress in neural network-based vision for autonomous robot driving.” *Proceedings of the Intelligent Vehicles92 Symposium*, IEEE, 391–396.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). “Pointnet: Deep learning on point sets for 3d classification and segmentation.” *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Seo, J., Kim, T., Kwak, K., Min, J., and Shim, I. (2023). “Scate: A scalable framework for self-supervised traversability estimation in unstructured environments.” *IEEE Robotics and Automation Letters*, 8(2), 888–895.

Sevastopoulos, C. and Konstantopoulos, S. (2022). “A survey of traversability estimation for mobile robots.” *IEEE Access*, 10, 96331–96347.

Wellhausen, L., Dosovitskiy, A., Ranftl, R., Walas, K., Cadena, C., and Hutter, M. (2019). “Where should i walk? predicting terrain properties from images via self-supervised learning.” *IEEE Robotics and Automation Letters*, 4(2), 1509–1516.

Zhang, L., Helmberger, M., Fu, L. F. T., Wisth, D., Camurri, M., Scaramuzza, D., and Fallon, M. (2023). “Hilti-oxford dataset: A millimeter-accurate benchmark for simultaneous localization and mapping.” *IEEE Robotics and Automation Letters*, 8(1), 408–415.

Zhang, L., Zhao, J., Long, P., Wang, L., Qian, L., Lu, F., Song, X., and Manocha, D. (2021). “An autonomous excavator system for material loading tasks.” *Science Robotics*, 6(55), eabc3164.

Zhu, X., Zhou, H., Wang, T., Hong, F., Ma, Y., Li, W., Li, H., and Lin, D. (2021). “Cylindrical and asymmetrical 3d convolution networks for lidar segmentation.” *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9939–9948.