

TTM: Terrain Traversability Mapping for Autonomous Excavator Navigation in Unstructured Environments

Tianrui Guan^{1*} Zhenpeng He¹ Dinesh Manocha² Liangjun Zhang¹

¹ Robotics and Auto-Driving Laboratory, Baidu Research

² University of Maryland, College Park

(Full report, video and dataset at <https://gamma.umd.edu/ttm>)

Abstract—We present Terrain Traversability Mapping (TTM), a real-time mapping approach for terrain traversability estimation and path planning for autonomous excavators in an unstructured environment. We propose an efficient learning-based geometric method to extract terrain features from RGB images and 3D pointclouds and incorporate them into a global map for planning and navigation for autonomous excavation. Our method used the physical characteristics of the excavator, including maximum climbing degree and other machine specifications, to determine the traversable area. Our method can adapt to changing environments and update the terrain information in real-time. Moreover, we prepare a novel dataset, Autonomous Excavator Terrain (AET) dataset, consisting of RGB images from construction sites with seven categories according to navigability. We integrate our mapping approach with planning and control modules in an autonomous excavator navigation system, which outperforms previous method by 49.3% in terms of success rate based on existing planning schemes. With our mapping the excavator can navigate through unstructured environments consisting of deep pits, steep hills, rock piles, and other complex terrain features.

I. INTRODUCTION

Excavators are one of the most common types of heavy-duty machinery used for many earth-moving activities, including mining, construction, environmental restoration, etc. According to [1], the size of the global market share for excavators had reached \$44.12 billion in 2018 and is expected to grow to \$63.14 billion by 2026. As the demand for excavators increases, many autonomous excavator systems [1], [2], [3] have been proposed for material loading tasks, which involve use of perception and motion planning techniques. Since excavation is considered one of the most hazardous operations, robustness and safety are two of the most important factors in designing automated systems.

Excavators are usually operated in unstructured and dangerous environments consisting of rock piles, cliffs, deep pits, steep hills, etc. An unstructured environment usually refers to a terrain that lacks structure and has unpredictable and potentially hazardous conditions. Such an environment lacks any lane markings and the arrangement of obstacles tends to be non-uniform. Due to tasks like digging and dumping, the working conditions for excavators are constantly changing. In addition, landfalls and cave-ins might happen, which might cause the excavator to tip over and injure the operator. Therefore, it is crucial to identify different terrains and predict safe regions for navigation. In contrast, the

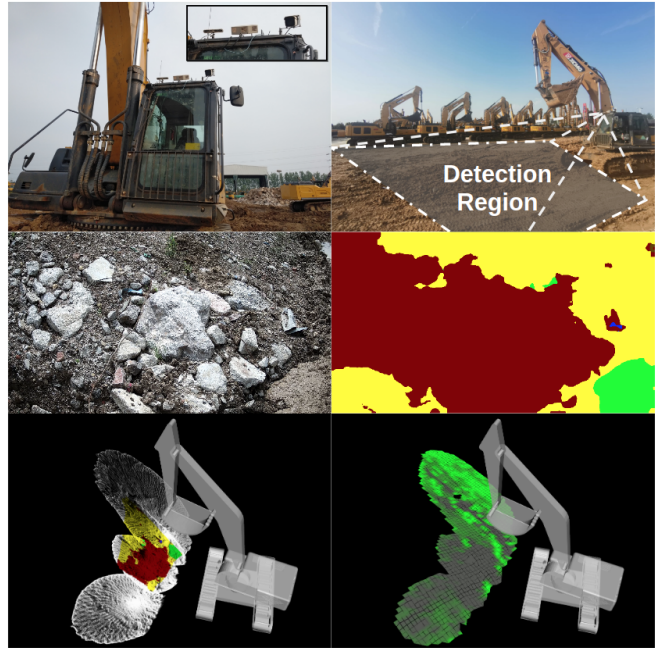


Fig. 1: We highlight our system TTM, the operating environment, input and output of our approach. To the best of our knowledge, TTM is the first approach for excavator navigation in complex, unstructured environments. **Top left:** Sensors on the excavator, including RGB cameras and Livox LiDAR. **Top right:** Detection region from a third-person perspective. **Middle left:** Frontal view captured by the camera. **Middle right:** Semantic segmentation output, where green, yellow, and maroon correspond to flat region, bumpy region, and rock, respectively. **Bottom left:** Colored pointcloud with semantic labels. **Bottom right:** Terrain traversability output, where traversability value decreases from green to grey.

conditions of the roads in structured environments such as highways are usually navigation-friendly, so the core problem during navigation in structured environments is avoiding obstacles rather than determining which part of the surface is easier and safer to navigate. In addition, the boundaries of different objects in structured, urban environments can be clearly captured by perception methods, which reduces the likelihood of mis-classifications.

There have been many works related to unstructured environments, including perception and terrain classification [4], [5], [6], as well as navigation [7], [8], [9], [10]. Perception is a difficult task in unstructured environments because there could be many unexpected or unknown objects. As mentioned in [4], in some cases, the navigable region

* Work done during an internship at Baidu RAL.

and obstacles tend to have similar visual features. Without robust perception results, navigation for excavators would be difficult and could cause damage. Moreover, it requires extra caution to apply navigation methods to excavators due to their enormous size and weight. Therefore, using a better algorithm to obtain a traversable region is an essential first step for safe and reliable navigation for an excavator.

Terrain traversability is a binary value, or a probability score, measuring the difficulty of navigating on a region. Terrain traversability estimation is a critical step between perception and navigation. In many cases [9], [11], a method capable of detecting obstacles and distinguishing road and non-road regions tends to be sufficient for navigation. On the other hand, in an unstructured, hazardous environment where off-road navigation is unavoidable, there are many factors that must be considered, including efficiency, adaptability, and safety. In such case, not only a more detailed classification according to terrain features is needed, but also a continuous value for traversability is preferred to describe the complication of terrain and provide the safest option for the navigation module.

Main Results:

We present Terrain Traversability Mapping (TTM), a real-time mapping approach for terrain traversability classification and navigation for excavators. We propose an efficient learning-based geometric fusion method to extract traversability maps for planning and navigation tasks. Our method leverages the physical constraints of the excavator, including maximum climbing degree, width of the excavator tracks, etc., to determine a traversability value for the surface. Our method can process RGB images and point-cloud streams in real-time and the global map is updated with semantic and geometric terrain information at a rate of 10 Hz, which can satisfy the requirements of excavators during operation. In addition, we present Autonomous Excavator Terrain (AET) dataset, which includes 669 RGB images from the working environments of excavators with seven categories according to surface navigability. AET dataset provides us good traversability estimation on similar construction sites where excavators usually operate. We deploy our approach on an excavator on several scenes with rock piles, steep ramps, pits, hills, and puddles, as shown in Figure 1.

Our method TTM is the first traversability estimation and mapping approach based on 3D geometric and 2D semantic information that is used for path planning and navigation of heavy-duty excavation machines in real-world scenes. Our approach has been integrated with a state-of-the-art Autonomous Excavator System (AES) [1] and we highlight the benefits. The perception method in original AES is capable of detecting rocks and target materials for digging and dumping in the RGB images. Our TTM approach strengthens the perception capability and semantic understanding of the surrounding scene, and provides an estimated traversability map for navigation. As a result, the excavator can avoid dangerous regions and safely navigate to the goal in unstructured environments. The novel aspects of our work include:

- 1) We present a real-time terrain traversability estimation and mapping algorithm (TTM) from 3D LiDAR and RGB camera inputs for excavator navigation. We develop a novel learning-based geometric fusion strategy with machine specifications for terrain traversability prediction in unstructured environments. We use learned semantic terrain information as visual guidance for dangerous surface detection and use geometric terrain features as supplemental information in uncertain and unknown regions.
- 2) We implement TTM in both simulated environments and real-world settings on an excavator in various challenging construction scenes. Our approach is the first traversability mapping algorithm integrated with planning and control modules, which are used to safely guide an excavator in unstructured environments.
- 3) We present Autonomous Excavator Terrain (AET) dataset, consisting of RGB images in unstructured environments with seven different classes based on terrain types, traversable regions, and obstacles.

II. RELATED WORK

A. Field Robots

Field robots usually refer to machines that operate in off-road, hazardous environments. They are typically any heavy-duty service robots for industrial usage in mining [12], excavation [1], agriculture [13], construction [14], etc. To satisfy industrial needs and save money and labor for the company, many automated systems [1], [2], [3] have been developed for those service robots in the field. However, it remains a challenge to fully automate many tasks because most of the applications are used in unknown, unstructured environments. The automation must be robust and adaptable to the changing circumstances of the surroundings.

B. Perception in Unknown Environments

There are four types of perception methods used in navigation and terrain classification tasks: proprioceptive-based methods, geometry-based methods, vision-based methods, and hybrid methods. Proprioceptive-based methods like [15] primarily use vibration or accelerometer signals collected by IMU and tactile sensors to classify the terrain. For these methods, the robot must traverse the region to make predictions, so one significant disadvantage of those methods is that their applications are constrained to safe and predictable environments and do not work well in unknown and hazardous environments. Geometry-based methods like [16] utilize stereo cameras or LiDAR to obtain pointcloud data and extract relevant information. Geometric methods can capture useful terrain attributes like shape, slope, and drastic height variation, as well as the location of obstacles.

There are many vision-based methods proposed for various perception tasks. For example, [17] use SIFT features extracted from the RGB images and classify terrains using simple models like MLP. In recent years, as many fast and reliable deep learning solutions for computer vision have been developed, researchers have begun to train a deep model

for perception using a large amount of annotated data. For example, [18] performs semantic segmentation for urban scenario benchmarks, while [4] focuses on coarse-grain semantic segmentation for terrain classification. Computer vision is a powerful tool in detecting textures that geometry-based methods are unable to identify.

Given the complimentary outputs of different perception methods, there are also hybrid methods that take advantage of the strengths of each perception method to make classification decisions based on collective information from multiple sensors and multiple data representations. For example, [19] uses both image input and audio input to cluster road features with a Gaussian mixture model. Our proposed method is also a hybrid method, leveraging the strengths of RGB cameras and 3D LiDAR.

C. Terrain Traversability Analysis

There are plenty of works [20], [21], [22] on classifying different terrains based on either material categories or navigability properties. [20] takes a voxel-based octree representation of the terrain and outputs semantic labels for each voxel using Conditional Random Field. [21] uses semantic mapping to associate points with 2D semantics to obtain a 3D semantic pointcloud. The goal of these classification methods is to correctly predict surface types, and they lack understanding of the navigability of the surface.

Compared to terrain classification, traversability is a more subjective concept and is specific to navigation. For different robots, the definition of traversability could be slightly different. In [16], the notion of traversability is entirely based on geometric attributes of the surface, including slope, height variation, and roughness of the surface. [4] proposes an attention-based deep network for traversability prediction in a natural and unstructured environment. [23] uses pointcloud and RGB images to classify 2D terrains with safe, risky, and obstacle labels.

D. Datasets in Unstructured Environments

Most recent developments in perception tasks like object detection and semantic segmentation focus on urban driving scene datasets like KITTI [24], Waymo [25], etc., which achieve high accuracy in terms of average precision. On the other hand, unstructured scenes like the natural environment, construction sites, and complicated traffic scenarios are less explored, primarily for two reasons. The first reason is that there are fewer datasets with unstructured environments, and the second is that perception and autonomous navigation in unstructured off-road environments is challenging due to its unpredictability and diverse terrain types.

Recent efforts in off-road perception and navigation include RUGD [26] and RELLIS-3D [27], which are semantic segmentation datasets collected from a robot navigating in off-road and natural environments and contain scenes like trails, forests, creeks, etc. [28] is a construction dataset containing annotation of heavy-duty vehicles for detection, tracking, and activity classifications.

III. TERRAIN TRAVERSABILITY MAPPING

We propose an efficient 2D-3D hybrid method for terrain traversability mapping (TTM) to extract terrain features and generate a traversability map as the robot navigates through an unknown environment. TTM takes a 3D pointcloud stream from the LiDAR, an RGB camera stream from the RGB camera, and the corresponding poses of the excavator extracted from the GPS-RTK module. The output of our method is a global map consisting of terrain information, including semantic information, geometric information, and a final traversability score.

The map is represented as an elevation grid map and is updated based on incoming pointclouds and RGB images. Internally, each grid cell in the map stores the average height value of the latest N points within this cell, as well as overall information about those points like update time, slope, step height, and their semantic information. A traversability score is calculated for each grid cell. In Figure 2, we present an overview of our approach. Our code is based on the open-source grid map library [29].

Our method is a hybrid method that combines 2D semantic and 3D geometric information. Most of the existing methods [16], [20] use 3D pointcloud or voxels to detect obstacles and classify them into different terrain types, while some methods like [21], [22] only use 2D semantic segmentation and register terrain labels on the pointcloud. Combining both 2D semantic information and 3D geometric information is essential for a robust prediction. For example, we can use geometric information from pointclouds to detect pits or hills, while it is difficult to get such terrain information from an RGB photo. On the other hand, 2D semantic segmentation can provide terrain semantics through visual features. For example, a terrain with a flat surface indicated through LiDAR detection could still be a water puddle, or a surface with a high slope could be a small pile of dirt, which is traversable, or a rock pile, which needs to be avoided.

There are some reasons why we choose not to use 3D segmentation methods directly on the pointcloud. First, it is difficult to annotate 3D data with semantic labels, while 2D labels are more accessible and easier to annotate. Second, 3D segmentation networks are usually heavy models and require longer inference time. Finally, during navigation in unstructured environments, excavator tracks could cause dust in the air, which introduces a lot of noise for the LiDAR input. 2D RGB images have rich color and texture information and are more immune to dust.

The rest of this section is structured as follows: We start with two components of our method, semantic mapping and geometric computation in Section III-A and Section III-B, respectively. We discuss how to use combined information to estimate terrain traversability in Section III-C. We explain how we process the output traversability map and integrate it with path planning and navigation in Section III-D.

A. Learning-based Semantic Mapping to Pointcloud

To extract semantic information of the terrain, we choose to use 2D semantic segmentation on unstructured terrains.

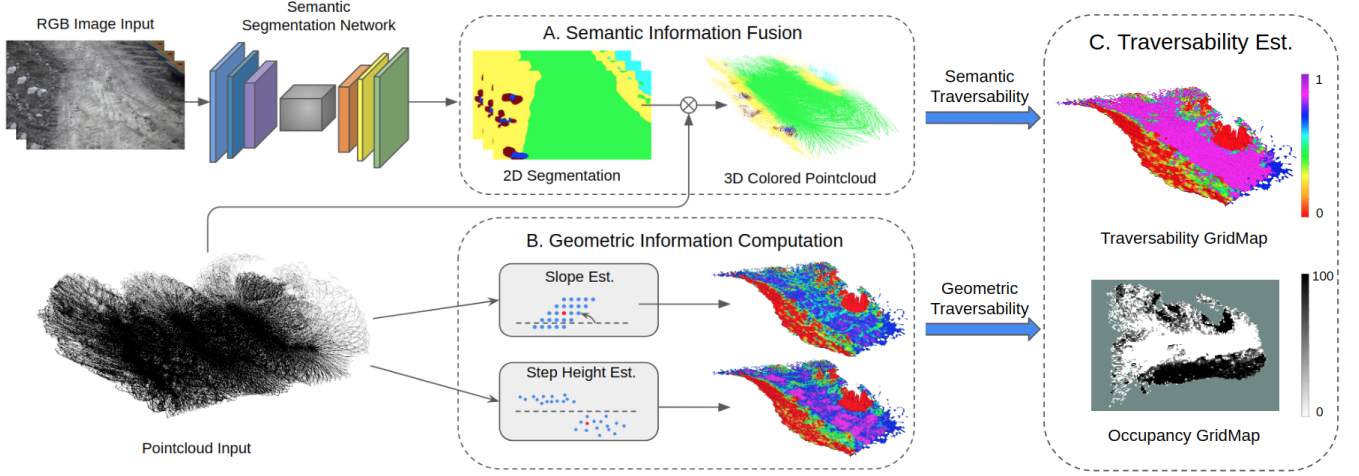


Fig. 2: **Overview of our method (TTM):** Our method takes RGB images and pointclouds as inputs to infer traversability. We extract semantic information using segmentation and associate terrain labels with pointclouds, as shown in A (**top**). We extract geometric information using slope and step height estimation, as shown in B (**bottom**). Finally, we produce a traversability gridmap based on semantic and geometric information and convert it to a 2D occupancy map for path planning, as shown in C (**right**).

Given an input RGB image $I \in \mathbb{R}^{3 \times H \times W}$, the goal is to generate a probabilistic map $M \in \mathbb{R}^{N \times H \times W}$, where N is the number of classes. After obtaining the probabilistic map P , our final output P is equal to $\text{argmax}(M) \in \mathbb{Z}^{H \times W}$.

To better facilitate the learning process, we create a novel Autonomous Excavator Terrain (AET) dataset for terrain classification and obstacle detection. According to the traversability criterion of the navigable regions defined in latest work [4] and the operating environments of excavators, we define seven classes in AET dataset, including flat regions, bumpy regions, mixed water and dirt, water puddles, rock piles, obstacles, and other excavators. The goal of the annotation is to identify different traversable regions and obstacles during navigation. Based on AET dataset, we train a Fast-SCNN [30] model for finding traversable regions with real-time performance. Our training code is based on [31]. For more details about AET dataset, please refer to Section IV-D.

After we get the segmentation prediction P , we use a timestamp to locate the corresponding pointcloud C . Given a point $\vec{p} = [x_p, y_p, z_p] \in C$ in the world coordinate, camera intrinsic matrix $K \in \mathbb{R}^{3 \times 3}$, and extrinsic matrix $E \in \mathbb{R}^{4 \times 4}$, we can calculate its corresponding image coordinates $\vec{p}_i = [u, v]$ based on the following formula:

$$\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix} = E \cdot \begin{pmatrix} x_p \\ y_p \\ z_p \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \cdot \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} / z_c$$

For each point captured by the LiDAR, we use calibration matrices to find its correspondence to the segmentation results and save its terrain label in the gridmap cell. In the end, each gridmap cell is assigned a terrain label based on the majority of the semantic labels. For cells that are not assigned

with a semantic label, we extract geometric information as a replacement for traversability estimation.

B. Geometric Information Computation

In this section, we present details of slope and step height estimation, and highlight how machine specifications are considered for calculation of geometric traversability score as our novel contribution.

Slope Estimation: Each grid cell g is abstracted to a single point $p = \{x, y, z\}$, where x, y is the center of the cell in the global coordinate frame and z is the height value of the grid. The slope s in arbitrary grid cell g is computed by the angle between the surface normal and the z-axis¹ of the global coordinate frame:

$$s = \arccos(n^z), n^z \in [0, 1]$$

where n^z is the component of normal \vec{n} on the z-axis.

We use Principal Component Analysis (PCA) to calculate the normal direction of a grid cell. By analyzing the eigenvectors and eigenvalues of covariance matrix C , created from the nearest neighbors of the query grid cell, we can estimate the direction of the surface normal vector as well as slope. The covariance matrix C is calculated as follows:

$$C = \frac{1}{k} \sum_{i=1}^k (p_i - \bar{p}) \cdot (p_i - \bar{p})^T, \quad C \cdot \vec{v}_j = \lambda_j \cdot \vec{v}_j,$$

$$j \in \{0, 1, 2\}, \lambda_i < \lambda_j \text{ if } i < j,$$

where k is the number of neighbors considered in the neighborhood of g , $p_i = \{x, y, z\}$ is the position of the neighbor grid in the global coordinate frame, \bar{p} is the 3D centroid of the neighbors, λ_j is the j -th eigenvalue of the covariance matrix, and \vec{v}_j is the j -th eigenvector. The surface normal \vec{n} of grid g is the eigenvector \vec{v}_0 that has the smallest absolute value of eigenvalue λ_0 .

The purpose of the slope estimation is to get the shape of the terrain and avoid navigating on a steep surface. Based on

¹Up direction in the real world

the experience of several expert operators, the width of the track is a good indicator of the navigation stability on rough terrain. Usually, when the area of a rough region is less than half of the track width, the excavator can navigate through it without any trouble. In our case, the width of our excavator track is 0.6 m, so we chose the grid resolution $d_{res} = 0.2m$ and use eight neighborhood search methods, which means the search area size equals the track width.

Step Height Estimation: The step height h is computed as the largest height difference between the center point p of the grid and its k' nearest neighbors:

$$h = \max(\text{abs}(p^z - p_i^z)), i \in [1, k']$$

Since slope is a description of variation in the terrain in a relatively small region, we choose to use a neighbor search parameter $k' = 7 * 7 > k$ that spans 1.4 m to measure height change in a larger scope. The step height calculation guarantees that the track does not traverse areas with extreme height differences.

Geometric Traversability Estimation: Based on information about slope and step height of the terrain, we can calculate a geometric traversability score T_{geo} . According to the physical constraints of the robot, we create some critical values s_{cri} , s_{safe} , h_{cri} , h_{safe} , as the thresholds for safety and danger detection. The purpose of those threshold values is to avoid danger when the surface condition exceeds the limits of the robot and to avoid more calculations when the surface is very flat. In other cases, the traversability is calculated as a continuous value. The formula for geometric traversability T_{geo} is:

$$T_{geo} = \begin{cases} 0 & s > s_{cri} \text{ or } h > h_{cri} \\ 1 & s < s_{safe} \text{ and } h < h_{safe} \\ \max(1 - (\alpha_1 \frac{s}{s_{cri}} + \alpha_2 \frac{h}{h_{cri}}), 0) & \text{otherwise} \end{cases}$$

where the weights α_1 and α_2 sum up to 1. In our case, we use 0.5 and 0.5, respectively.

Unlike the method described in [16], we exclude the use of a roughness score in T_{geo} . In our context, the excavator is more sensitive and prone to failure under extreme slopes and step heights, while it can usually handle uneven ground and relatively rough terrain. We also discover that the roughness measurement is not effective in the traversability computation. For more detailed comparisons, please refer to the full report [32]. In addition, the measurement of terrain roughness is accomplished through visual cues and texture information, as bumpy regions are classified from 2D semantic information, and this method produces a more meaningful output than geometric measurement.

Because the value of RTK on the Z-axis is unstable, the value of Z may vary significantly at different times, so we save update time in the grid and only use the grid within a certain range Δt at the current time for geometric information calculation.

Safe and Critical Threshold:

We use the specifications of our excavator to determine the safe and critical values. We use an XCMG XE490D excavator, which has a maximum climbing angle of 35

degrees. We use the typical recommended climbing angles for any vehicle as a safe climbing angle, which is 10 degrees. As a result, we set $s_{cri} = 30 \text{ deg}$ and $s_{safe} = 10 \text{ deg}$.

For step height, we make sure that h_{cri} and h_{safe} comply with slope s_{cri} and s_{safe} . We get an approximation of the maximum height allowed by s_{cri} and s_{safe} , expanding three times the resolution d_{res} along the surface:

$$\begin{aligned} h_{cri} &= 3 \tan(s_{cri}) \times d_{res} = 0.35 \text{ m} \\ h_{safe} &= 3 \tan(s_{safe}) \times d_{res} = 0.10 \text{ m} \end{aligned}$$

Thus, the step height estimation is complementary to slope estimation; it provides a global perspective, whereas slope is local terrain information. Combining these two specifications can help us remove noise in the map, such as bumps caused by dust, and ensure the robustness of the T_{geo} .

C. Traversability with Geometric and Semantic Fusion

In this section, we describe our novel algorithm for geometric-semantic fusion. From the semantic and geometric information, we use a continuous traversability score $T \in [0, 1]$ to measure how easily the surface can be navigated. This is especially relevant to off-road scenarios because we prefer flat regions over bumpy roads to save energy. Moreover, when an excavator is navigating on a construction site, being able to correctly identify different regions is critical to avoid hazardous situations like flipping over.

The overall traversability score T is calculated based on semantic terrain classes C_{sem} and geometric traversability T_{geo} as follows:

$$T = \begin{cases} 0 & C_{sem} = \{\text{rock, excavator, obstacle, water}\} \\ 1 & C_{sem} = \{\text{flat}\} \text{ and } T_{geo} > 0 \\ T_{geo} & \text{otherwise} \end{cases},$$

Based on our formulation, we make sure that the robot avoids any obstacles like rocks or excavators and forbidden regions like water. We also consider flat regions as traversable regions with a score of 1. In other cases like bumpy, mixed, or unassigned regions, we set the score according to the geometric information. With our current scheme, we manage to minimize the use of continuous values in the final traversability map to reduce the number of grey regions to adapt to the effectiveness of existing planners.

D. Planning Based on TTM

We start with a novel post-processing step on the traversability map for planning. We remove some small non-traversable regions that satisfy all of the following criteria:

- The region has an average traversability value less than some occupied threshold t_{occ} .
- The region has a height less than the critical threshold of step height h_{cri} .
- The spans of the region along the x-axis and the y-axis are less than half the distance between two tracks of the excavator d_{track} .

This post-processing step is mostly to accommodate existing planners, since they do not work well and fail to

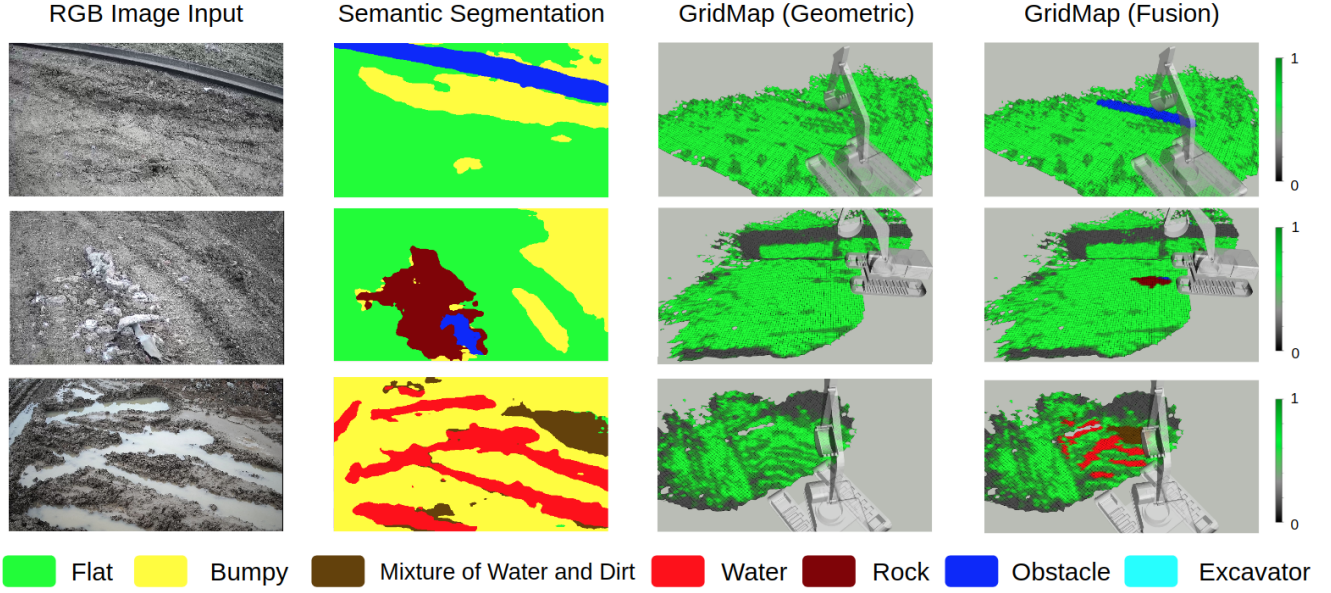


Fig. 3: **Visual results of TTM:** Each row shows the input RGB image, the corresponding semantic segmentation result, the traversability map with only geometric information, and the traversability map after fusion. In the final traversability map, the higher the traversability score is, the easier it is for robots to navigate the corresponding terrain. We use the color scheme from green to grey to highlight some semantic labels like rock, obstacle and water. We see that there are many places that are not detected through their geometric properties, but that can be recognized using their visual features and are reflected in the final traversability map.

plan a feasible trajectory with too many scattered, noisy regions. After post-processing, we transform the traversability gridmap to a 2D occupancy gridmap as an input for path planning. We use the Hybrid A* planner [33] to generate a path and send the trajectory to the motion controller, which guides the excavator to follow this trajectory. Some planning results can be found in IV-E.

IV. IMPLEMENTATION DETAILS AND ANALYSIS

A. Implementation Details

We use an XCMG XE490D excavator to perform our experiments. The excavator is equipped with a Livox-Mid100 LiDAR and a HIK web camera with FOV of 56.8 degrees to detect the environment, and a Huace real-time kinematic (RTK) positioning device to provide the location. We run our code with a laptop with Intel Core i7-10875H CPU, 16 GB RAM, and GeForce RTX 2060 6GB GPU on the excavator.

We train a Fast-SCNN [30] model on two NVIDIA GeForce GTX 1080 Graphics cards for 240K iterations, which takes approximately two days. We use a stochastic gradient descent optimizer with a learning rate of 0.12 and a momentum of 0.9. We also use polynomial learning rate decay with a power of 0.9. The final model achieves a mean IoU of 68.75% and an overall accuracy of 84.49%.

B. Visual Results and Analysis

In this section, we evaluate our method in real world with visual results. In Figure 3, we show some typical scenarios excavators encounter to illustrate the advantages of geometric and semantic fusion. For visualization purposes, we change the color scheme to range from green to grey and paint the detected obstacles and forbidden regions like water to the corresponding color in Figure 3. In the first and second

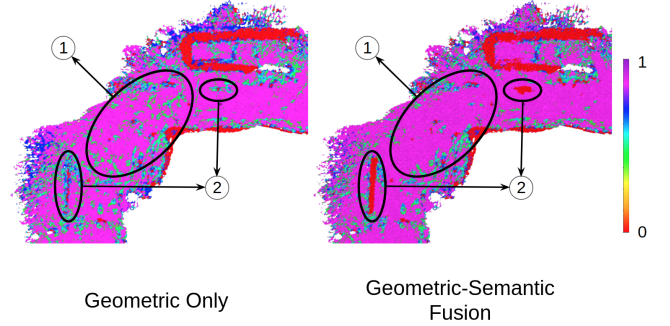


Fig. 4: **Gridmap comparison between the geometric-only scheme [16] and ours:** (1) Our method is less noisy and has more connected regions to plan a feasible trajectory. (2) Our method can detect obstacles that the geometric method could not recognize.

cases, the steel bar and stone were not captured by geometric calculation, while with semantic information, those obstacles can be detected and avoided. In the last case, the rough and bumpy region is in fact water and should not be traversed.

In Figure 4, we compare traversability maps generated by using a geometric-only method [16] and using TTM with geometric-semantic fusion. The output after fusion is less noisy since segmentation results can smooth out safe regions. Our method detects more non-traversable regions based on obstacles and dangerous regions from semantic information.

C. Performance in the Real World

We set up our experiments on an excavator in the real world. Our method consists of the following major parts, which contribute to the overall runtime of the system:

- **Segmentation** generates a pixel-wise semantic classification on each image in the RGB input stream.
- **Projection** casts the 2D segmentation result onto the

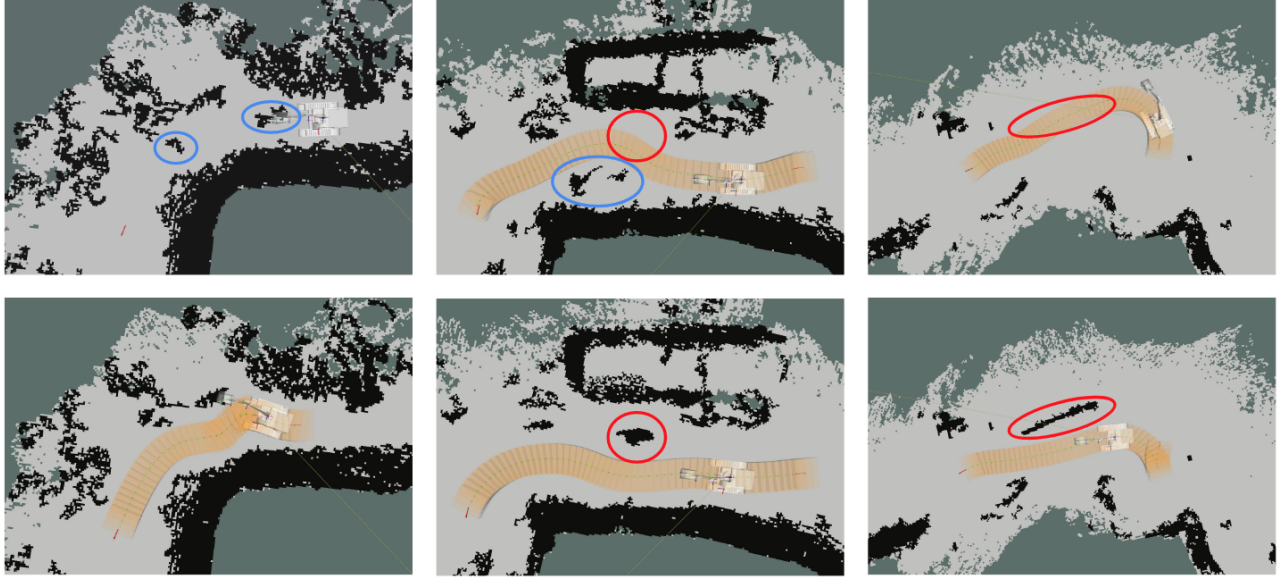


Fig. 5: **Planner output comparisons between geometric-only scheme [16] (top) and TTM (bottom):** We show planned trajectories based on the Hybrid A* [33] planner. The planning is based on a global traversability map. We highlight some obstacles that are not observed by geometric method (red), as well as some traversable regions that are falsely observed by geometric method (blue).

Run-time (ms)	Max	Min	Mean
Segmentation	100.2	53.5	75.4
Projection	54.0	35.0	42.3
T_{geo} Calculation	38.0	9.0	22.1

TABLE I: **Runtimes of different modules.** Our method can be run in real-time and update the traversability map at a rate of 10 Hz. The hardware is specified in Section IV-A.

3D pointcloud and assigns each point a semantic label through the calibration matrix.

- **Geometric traversability calculation** estimates and updates slope and step height based on point cloud data in gridmap representation.

The final fusion step is under 2 ms and do not contribute to the overall runtime of the method. As shown in Table I, the running time per update of each component mentioned above is 75.4 ms, 42.3 ms, and 22.1 ms, respectively.

Our method can handle an RGB image stream of 25 Hz and a pointcloud stream of 10 Hz in the real world without lagging, and the map can update semantic and geometric information at a rate of 10 Hz. Please refer to the video for more visual results of excavator navigation.

D. Dataset Collection

Our dataset is collected at a construction site while an excavator is navigating through the work area. We collect several videos (around 30 minutes in total) under different circumstances and sample 669 images of size 1920×1080 for our dataset. The training and testing sets are split according to a 9:1 ratio. As shown in Figure 6, AET dataset covers most of the situations that might be encountered on a work site, including rock-piles, pits, stagnant water after rain, etc.

Flat surfaces, bumpy surfaces, and mixtures of water and dirt are usually navigable. In most cases, when flat surfaces are detected, they are preferable to other surfaces. Water

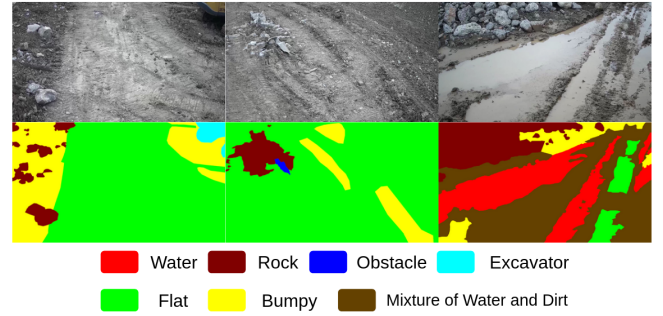


Fig. 6: **Autonomous Excavator Terrain (AET) dataset:** We show a few samples from our AET dataset (top) and corresponding annotations (bottom). All images are collected in unstructured environments with various terrain types.

is a forbidden region, since it is difficult to gauge how deep the water is and the soil near the water could be soft and easily deformed. Even though the excavator can usually traverse through shallow water, we choose to minimize the risk and damage. Obstacles like rocks and other excavators must be avoided for safety. The annotation is decided and implemented based on the opinion of a team of excavator operators.

E. Planning Based on a Traversability Map

Based on the resulting occupancy gridmaps from geometric-only method [16] and the proposed TTM, we randomly choose start and goal position on an unoccupied grid with over 90 trials. Using Hybrid A* [33] planner, the success rates of finding a valid path without collision for our TTM and the other method are 82.6% and 33.3%, respectively. We show some comparisons on planning results in Figure 5. In particular, we use an occupied threshold t_{occ} of 0.6, a critical threshold h_{cri} of 0.35 m for maximum step height, and a distance between two tracks d_{track} of 2.75 m for map post-processing and planner configuration.

For more details of the experimentation, please refer to the full report[32].

V. CONCLUSIONS, LIMITATIONS, AND FUTURE WORKS

In this paper, we present Terrain Traversability Mapping (TTM) that combines 2D semantic terrain information and 3D geometric terrain information to make joint predictions on terrain traversability in unstructured environments. We present a novel LiDAR-camera fusion strategy based on a learning-based geometric method. We prepare Autonomous Excavator Terrain (AET) dataset with difficult real-world scenes in unstructured construction sites for perception in the wild. Finally, we show results on our approach with planning and control modules for excavators in the real world and in simulation on various scenes.

Our work has some limitations. First, due to safety issues, we are not able to extensively test navigation in the real world (especially in some dangerous scenarios where most regions are bumpy or steep) to avoid failure cases like flipping over. Second, we have tested our performance on existing planners, which are not able to exploit the full potential of traversability. For example, sometimes the excavator would have been able to run over small obstacles with the space between two tracks, but there is no existing planner that could handle such a case.

As part of our future work, we need to consider how to design a planner that can take full advantage of the traversability map and utilize the specifications of the excavator like a human operator. Our final goal is using this framework to guide the excavator to navigate on its own in extreme and hazardous environments.

REFERENCES

- [1] L. Zhang, J. Zhao, P. Long, L. Wang, L. Qian, F. Lu, X. Song, and D. Manocha, "An autonomous excavator system for material loading tasks," *Science Robotics*, vol. 6, no. 55, 2021. [Online]. Available: <https://robotics.sciencemag.org/content/6/55/eabc3164>
- [2] S.-K. Kim and J. Russell, "Framework for an intelligent earthwork system: Part i. system architecture," *Automation in Construction*, vol. 12, pp. 1–13, 2003.
- [3] J. Seo, S. Lee, J. Kim, and S.-K. Kim, "Task planner design for an automated excavation system," *Automation in Construction*, vol. 20, no. 7, pp. 954–966, 2011.
- [4] T. Guan, D. Kothandaraman, R. Chandra, and D. Manocha, "Ganav: Group-wise attention network for classifying navigable regions in unstructured outdoor environments," 2021.
- [5] K. Viswanath, K. Singh, P. Jiang, S. P. B., and S. Saripalli, "Offseg: A semantic segmentation framework for off-road driving," 2021.
- [6] A. Singh, K. Singh, and P. B. Sujit, "Offroadtranseg: Semi-supervised segmentation using transformers on offroad environments," 2021.
- [7] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous Robots*, vol. 18, pp. 81–102, 01 2005.
- [8] G. Kahn, P. Abbeel, and S. Levine, "Badgr: An autonomous self-supervised learning-based navigation system," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1312–1319, 2021.
- [9] M. J. Procopio, J. Mulligan, and G. Grudic, "Learning terrain segmentation with classifier ensembles for autonomous robot navigation in unstructured environments," *Journal of Field Robotics*, vol. 26, no. 2, pp. 145–175, 2009.
- [10] A. Kumar, Z. Fu, D. Pathak, and J. Malik, "Rma: Rapid motor adaptation for legged robots," *arXiv:2107.04034*, 2021.
- [11] R. A. Hewitt, A. Ellery, and A. de Ruiter, "Training a terrain traversability classifier for a planetary rover through simulation," *International Journal of Advanced Robotic Systems*, vol. 14, no. 5, p. 1729881417735401, 2017.
- [12] H. Shariati, A. Yeraliyev, B. Terai, S. Tafazoli, and M. Ramezani, "Towards autonomous mining via intelligent excavators," in *CVPR Workshops*, 2019.
- [13] R. R. Shamshiri, C. Weltzien, I. A. Hameed, I. J. Yule, T. E. Grift, S. K. Balasundram, L. Pitonakova, D. Ahmad, and G. Chowdhary, "Research and development in agricultural robotics: A perspective of digital farming," 2018.
- [14] N. D. Nath and A. Behzadan, "Deep convolutional networks for construction object detection under different visual conditions," in *Frontiers in Built Environment*, 2020.
- [15] A. Vicente, Jindong Liu, and Guang-Zhong Yang, "Surface classification based on vibration on omni-wheel mobile base," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 916–921.
- [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*, 2009, pp. 4571–4576.
- [17] S. Matsuzaki, K. Yamazaki, Y. Hara, and T. Tsubouchi, "Traversable region estimation for mobile robots in an outdoor image," *J. Intell. Robotic Syst.*, vol. 92, no. 3–4, pp. 453–463, 2018.
- [18] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "Adapnet: Adaptive semantic segmentation in adverse environmental conditions," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 4644–4651.
- [19] R. Ishikawa, R. Hachiuma, and H. Saito, "Self-supervised audio-visual feature learning for single-modal incremental terrain type clustering," *IEEE Access*, vol. 9, pp. 64 346–64 357, 2021.
- [20] D. Belter, J. Wietrzykowski, and P. Skrzypczyński, "Employing natural terrain semantics in motion planning for a multi-legged robot," *Journal of Intelligent & Robotic Systems*, vol. 93, 03 2019.
- [21] D. Paz, H. Zhang, Q. Li, H. Xiang, and H. I. Christensen, "Probabilistic semantic mapping for urban autonomous driving applications," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 2059–2064.
- [22] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics*, M. Hutter and R. Siegwart, Eds. Cham: Springer International Publishing, 2018, pp. 335–350.
- [23] F. Schilling, X. Chen, J. Folkesson, and P. Jensfelt, "Geometric and visual terrain classification for autonomous mobile navigation," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 2678–2684.
- [24] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [25] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, et al., "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2446–2454.
- [26] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugged dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *International Conference on Intelligent Robots and Systems (IROS)*, 2019.
- [27] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," *arXiv:2011.12954*, 2020.
- [28] D. Roberts and M. Golparvar-Fard, "End-to-end vision-based detection, tracking and activity analysis of earthmoving equipment filmed at ground level," *Automation in Construction*, 2019.
- [29] 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.
- [30] R. P. K. Poudel, S. Liwicki, and R. Cipolla, "Fast-scnn: Fast semantic segmentation network," in *BMVC*, 2019.
- [31] M. Contributors, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark," <https://github.com/open-mmlab/mms Segmentation>, 2020.
- [32] T. Guan, Z. He, D. Manocha, and L. Zhang, "TTM: Terrain traversability mapping for autonomous excavator navigation in unstructured environments," in *arXiv*, 2021.
- [33] K. Kurzer, "Path planning in unstructured environments : A real-time hybrid a* implementation for fast and deterministic path generation for the kth research concept vehicle," Master's thesis, 2016.