MOSU: Autonomous Long-range Robot Navigation with Multi-modal Scene Understanding

Jing Liang†, Kasun Weerakoon†, Daeun Song‡, Senthurbavan Kirubaharan‡,
Xuesu Xiao‡, and Dinesh Manocha†

†University of Maryland, College Park MD, 20740, USA
‡Goerge Mason University, Fairfax, VA, 22030, USA

Abstract. We present MOSU, a novel autonomous long-range navigation system that enhances global navigation for mobile robots through multimodal perception and on-road scene understanding. MOSU addresses the outdoor robot navigation challenge by integrating geometric, semantic, and contextual information to ensure comprehensive scene understanding. The system combines GPS and QGIS map-based routing for high-level global path planning and multi-modal trajectory generation for local navigation refinement. For trajectory generation, MOSU leverages multi-modalities: LiDAR-based geometric data for precise obstacle avoidance, image-based semantic segmentation for traversability assessment, and Vision-Language Models (VLMs) to capture social context and enable the robot to adhere to social norms in complex environments. This multi-modal integration improves scene understanding and enhances traversability, allowing the robot to adapt to diverse outdoor conditions. We evaluate our system in real-world on-road environments and benchmark it on the GND dataset, achieving a 10% improvement in traversability on navigable terrains while maintaining a comparable navigation distance to existing global navigation methods.

Keywords: Global Navigation, Traversability Analysis, Multiple Modalities

1 Introduction

Global navigation has witnessed significant advancements in recent years [6], playing a crucial role in applications such as autonomous driving [27], logistics [9], and search and rescue [4]. However, several key challenges remain. Many existing approaches depend on highly accurate global maps and precise localization [17], which are costly and difficult to maintain at scale. Additionally, scene understanding, particularly traversability analysis and socially compliant navigation, poses a significant challenge for generating safe and context-appropriate trajectories. Ensuring reliable long-range navigation across diverse and dynamic real-world environments further complicates the problem, as the system must adapt to varying terrain, obstacles, and social contexts. Addressing these challenges requires an approach that integrates global planning with multimodal perception and adaptive local planning to enable robust and scalable autonomous navigation.

Maintaining an accurate map is time- and labor-intensive, and various temporal conditions can degrade planning performance [7,17]. However, humans do not require highly accurate maps for global navigation. Given Google Maps, we can complete most long-range navigation tasks, such as commuting from home to work, traveling in a new city, or walking on trails. Inspired by this observation, instead of maintaining a highly sophisticated global map, we propose separating this task into two easily accessible and generalizable components: routing and trajectory generation. Routing provides raw latitude and longitude directions, while trajectory generation determines the traversability of the environment to guide the robot to the next GPS location.

Scene understanding is essential for trajectory generation in outdoor robot navigation. It is a complex process that requires the integration of geometric perception [11], semantic comprehension [1, 10], and contextual awareness [18]. Geometric perception enables obstacle avoidance, semantic understanding distinguishes traversable paths, and contextual awareness ensures socially compliant behavior. Foundation models have recently demonstrated strong capabilities in capturing social contexts [22, 23]. However, many trajectory generation approaches prioritize only one or two aspects due to the high computational costs on resource-constrained onboard systems [14, 15], leading to incomplete environmental understanding and limiting navigation reliability. To address this, our approach incorporates multiple modalities, including LiDAR-derived geometric confidence [14], image-based color semantics [3], VLM-based context awareness [23], and robot odometry that achieves comprehensive on-road scene understanding for robot navigation.

Problem Statement: Designing a long-range navigation system with autonomous routing and trajectory generation by leveraging multimodal perception and VLMs to achieve comprehensive scene understanding, enhancing both traversability and social awareness.

2 Our Approach

We propose a novel global navigation system, MOSU, with **M**ulti-modal perception and **O**n-road **S**cene **U**nderstanding for mobile robots. While traditional outdoor navigation systems rely on detailed global maps and integrate global path planning with local motion planning, MOSU instead decomposes global path planning into two separate components: routing and trajectory generation. Specifically, we leverage QGIS and GPS data for high-level routing and use multi-modal sensor inputs for low-level trajectory generation and scene understanding. As shown in Fig. 1, the system consists of three stages: routing, trajectory generation, and motion planning.

2.1 Routing

The routing stage computes a high-level path from the robot's current GPS location to the target by generating a sequence of intermediate GPS subgoals. These subgoals are computed by public satellite routing service, such as Google or Openroute. Due to the limited precision of GPS sensors (5m) [26], the subgoals can only serve as directional guides towards the target rather than exact positions, and if the sub-goals are too close the noise will lead the robot to move into non-traversable areas. To prevent this short-range effect of the GPS

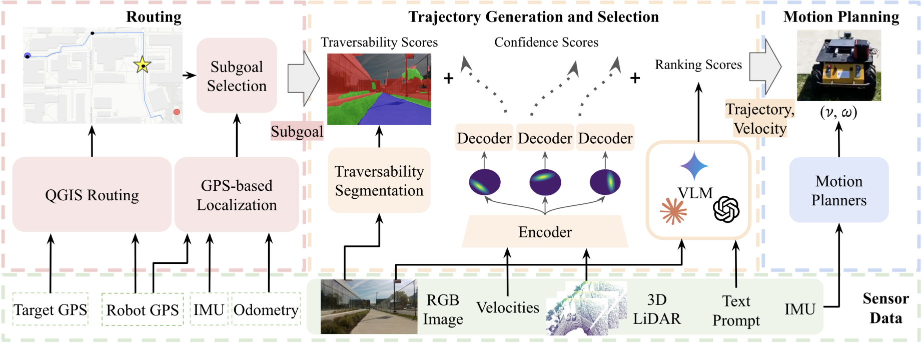


Fig. 1: Overall System Architecture. MOSU leverages QGIS and GPS to generate long-range waypoints, serving as high-level guidance for its trajectory generation system. The trajectory generation system provides local trajectories by integrating multimodal perception cues for traversability assessment and leveraging social cues from Vision-Language Models (VLMs) to ensure social compliance in navigation. This hierarchical approach enables robust, adaptive, and context-aware trajectory generation for long-range autonomous navigation.

sub-goals, we space the sub-goals approximately 50m apart, corresponding to the robot's perceptual range with 3D LiDAR and RGB sensors. As the robot navigates, we convert the current GPS sub-goal into the robot's coordinates and continuously monitor its approximate distance to the current sub-goal and updates to the next one upon reaching a predefined proximity threshold, 10m. This strategy enables scalable long-range navigation while allowing the local trajectory generation module to handle fine-grained motion decisions based on real-time sensor observations.

2.2 Trajectory Generation

Given the current GPS sub-goal, the robot needs a trajectory to navigate toward it. In complex outdoor environments, the robot must handle obstacles, varying terrain traversability, social norms, and traffic rules. We decompose this task into two subproblems: traversability analysis and social navigation. During trajectory generation, we apply different methods to address these challenges and generate trajectories that are both traversable and socially compliant, using multi-modal sensor inputs such as RGB images, robot velocities, 3D LiDAR point clouds, and text prompts (Fig. 1).

For traversability analysis, relying solely on geometric or color information is insufficient to fully understand the environment's traversability (i.e., which areas are safe for the robot to traverse), as shown in Fig. 4. Our system integrates both geometric and color information for more accurate analysis. Social navigation is highly intuitive and depends on the common sense and the culture of the country. Therefore, we utilize generalized vision-language models (VLMs) to address the social navigation problem.

Our trajectory generation method consists of four key components: (1) CVAE-based multiple trajectories generation, which models the distribution of feasible trajectories based on LiDAR and velocity inputs, (2) semantic segmentation for traversability analysis, which leverages RGB images to identify navigable surfaces (3) VLM-based trajectory ranking, which ranks candidate trajectories based on social compliance using overlaid image and natural language prompts.

CVAE-based Multiple Trajectories Generation: This model is to understand the geometric information of the environment for traversability analysis through 3D Lidar point clouds. However, directly modeling the environment using point cloud is memory costly and the segmentation of a large point cloud is also heavy [8, 20]. To address the issue, we adopt MTG [14], which is a very light-weighted learning-based approach to generate multiple candidate trajectories to cover traversable regions in front of the robot. MTG employs a Conditional Variational Autoencoder (CVAE) [21] to model the distribution of feasible trajectories. Given sensor input $\mathbf{x} = \{\mathbf{l}, \mathbf{v}\}$, where $\mathbf{l} \in \mathcal{L}$ represents a sequence of LiDAR observations and $\mathbf{v} \in \mathcal{V}$ denotes the robot's historical velocities, the model generates a set of N candidate trajectories $\mathcal{T} = \{\tau_1, ..., \tau_N\}$. For each trajectory $\tau_n \in \mathcal{T}$, we have the following formulation:

$$p_\theta(\tau_n|\mathbf{x}) = p_\theta(\tau_n|\mathbf{z}_n, \mathbf{c}), \quad \mathbf{z}_n \sim \mathcal{N}(\mu_n, \nu_n), \tag{1}$$

$$\mathbf{c} = f_\theta(\mathbf{x}), \quad \mu, \nu = g_\theta(\mathbf{x}), \quad \mathcal{N}(\mu_n, \nu_n) = h_\theta(\mathcal{N}(\mu, \nu), \mathbf{c}) \tag{2}$$

We slightly abuse the notation of θ to represent the parameters of all neural networks. $\mathbf{c}$ is the conditional vector obtained from an encoder network $f_\theta(\cdot)$. $g_\theta(\cdot)$ is the latent encoder network that takes the sensor inputs $\mathbf{x}$ and predicts the parameters, mean μ and variance ν, of a Gaussian distribution $\mathcal{N}(\mu, \nu)$. In our approach, the LiDAR data is processed by PointCNN [12], and the velocities are processed by a sequence of linear layers, as shown in [14]. Then we linearly transform the distribution to N Gaussian distributions $\mathcal{N}(\mu_n, \nu_n)$ by the learnable linear transformation (neural network) $h_\theta(\cdot)$. We then sample the latent vector $\mathbf{z}_n$ from the distribution $\mathcal{N}(\mu_n, \nu_n)$ and use it as the input of the trajectory decoder $p_\theta(\cdot)$ to generate trajectory τ_n. The predicted variance ν represents the uncertainty in the latent space and is used to compute a confidence score c_τ for each trajectory. This formulation takes the LiDAR geometric information as input and generates trajectories to cover geometrically traversable areas in front of the robot.

Semantic Segmentation for Traversability Analysis: However, this model struggles to accurately detect the boundaries between adjacent traversable and non-traversable areas when they have similar geometric structures (e.g., off-road mud vs. sidewalks), as shown in Fig.4 (c). To address this limitation, we incorporate semantic information by using Mask2Former [3], an RGB-based semantic segmentation model. The model segments the image into multiple regions with different semantic categories by predicting the class label of each pixel. We define five traversability categories [13]: road, sidewalk, vegetation, building, and others. Different types of robots have different traversable areas; for wheeled robots, we constrain them to operate only within the sidewalk and road traversability categories.

Given the segmented image, we overlay the set of candidate trajectories $\mathcal{T}$, generated by the CVAE-based trajectory generator, onto the image and evaluate their traversability scores. First, we use the Bresenham algorithm [2] to convert trajectory waypoints into connected pixels in the image. Then, the semantic traversability score t_τ is computed for each trajectory as the ratio of pixels falling in traversable areas to the total number of pixels along the trajectory.

VLM-based Trajectory Ranking: To ensure social compliance, we incorporate VLMs [22,23] to understand social cues from the robot's observation. As in VL-TGS [23], we project the trajectories onto the image and, from right to left, we assign numbers to the trajectories in sequence, according to the pixel positions of the last waypoint of each trajectory. Then, VLMs take the overlayed RGB image and a text prompt as input. The following is the prompt input to the model:

> The N trajectories are labeled with numbers [0-N] from right to left in sequence. rank trajectories for social navigation.
>
> 1. keep away from the goups of pedestrians. The robot has two mode, Normal and Slow. If the people are approaching, the robot need to Slow.
> 2. follow the traffic rules, and if going across the street, the robot should keep in crosswalks.
> 3. recognize the traffic signs and behave accordingly.
> 4. avoid off-road terrain for small wheeled robots.
>
> Given the picture, the target is at Front Left. Rank the trajectories by the criteria. output the format: [robot mode], [ranked numbers], reason

where the orange text indicates variables based on the current sub-goal and candidate trajectories. The model output consists of three parts: (1) the robot's current velocity mode—either slow or normal; (2) the ranking of trajectories based on social and traffic compliance; and (3) the reasoning behind the ranking, enabling chain-of-thought understanding of the environment. From the ranking of trajectories, we calculate the ranking score, $r_\tau = \frac{1}{N}(N - p_n)$, where p_n is the ranking of the trajectory τ_n. Leveraging chain-of-thought reasoning guided by the prompt, the VLM selects trajectories that are both socially compliant and contextually appropriate.

While each component contributes complementary information, they also come with individual limitations in scene understanding. For example, RGB-based segmentation lacks geometric precision and often fails to capture elevation changes such as curbs. Learning-based models may also misinterpret out-of-distribution regions under unfamiliar conditions, as illustrated in Fig. 4 (d). To mitigate these limitations, we aggregate the outputs from all components to compute a final score for each candidate trajectory. Since trajectories are generated in real time, we further incorporate multiple consecutive frames, transforming them into the current robot frame to ensure consistency. The optimal trajectory τ is selected by maximizing a weighted sum of scores:

$$i^* = \arg\max_{i \in [1,N]} \left\{ \beta_1 c_\tau^i + \beta_2 t_\tau^i + \beta_3 r_\tau^i + \beta_4 g_\tau^i \right\}, \quad \tau = \tau^{i^*}, \tag{3}$$

where $\beta_{1,2,3,4}$ are weights of the components, and $i \in [1, N]$ represents the index of the N generated trajectories. c_τ^i denotes the confidence score from geometric information, t_τ^i represents a semantic traversability score, and r_τ^i corresponds to the ranking score from VLMs, where higher-ranked trajectories receive greater weights. g_τ^i is the distance score to the nearest GPS subgoal, the closer, the higher.

2.3 Motion Planning

This stage generates executable robot actions to follow the selected trajectory τ from Equation 3. We integrate the Dynamic Window Approach (DWA) [5], a widely used reactive local planner that generates safe and feasible motion commands in real time. In addition to standard trajectory following, we apply the velocity mode (normal or slow) predicted by the VLM to constrain the robot's maximum velocity to social-compliantly move the robot. The normal mode allows a maximum velocity of $1, \mathrm{m/s}$, while the slow mode limits the velocity to below $0.5, \mathrm{m/s}$.

3 Experiments

The experiment is designed to evaluate the efficacy of the system in two aspects: (1) Trajectory generation, focusing on traversability analysis and social understanding. (2) Overall system performance in real-world long-range navigation compared with other approaches. Experiments are conducted on a computer equipped with an Intel i9 CPU (31 GB RAM) and an Nvidia RTX 3060 GPU (6.4 GB VRAM). We compare our approach against MTG [14], DTG [15], and VL-TGS [23], NoMaD [24], ViNT [19], and PIVOT [16] with the dataset. We use Gemini [25] for VLM.

We evaluate trajectory generation using the GND dataset [13], which covers 10 campuses with diverse scenarios, including urban and rural environments. For traversability analysis and social routine understanding, as shown in the Fig. 2, we evaluate the approaches a large-scale dataset (GND) with various challenging scenarios.

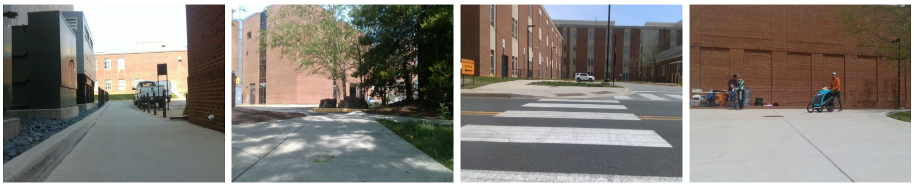
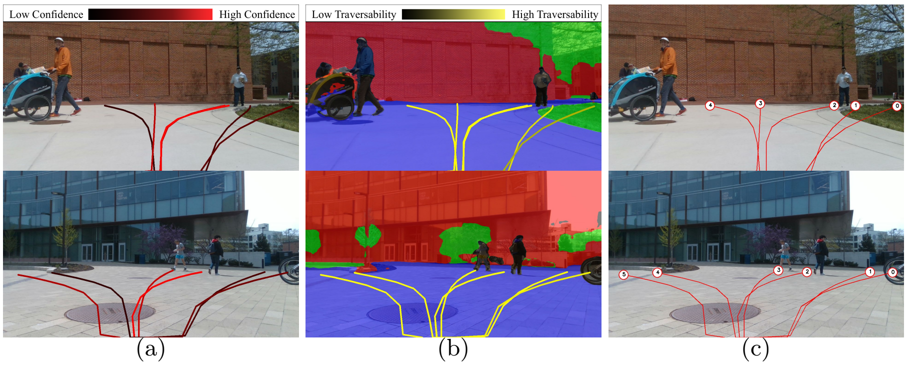

(a) Narrow Space (b) Off-road Terrain (c) Crosswalks (d) Social Scenarios

Fig. 2: **Complex Outdoor Scenarios:** We evaluate the approach in large-scale, complex outdoor environments with various challenging scenarios, such as (a) narrow spaces, (b) areas with dense off-road vegetation, (c) traffic components (e.g., crosswalks), and (d) social situations involving pedestrians.

The metrics in Table 1 are calculated as following:

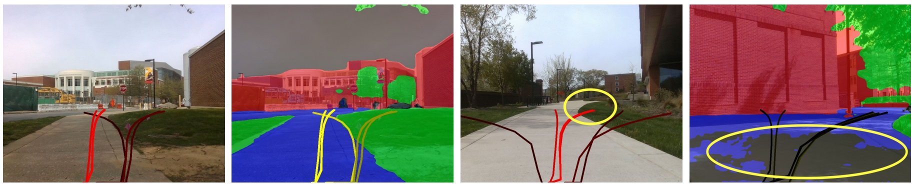
Traversability: We overlay the generated trajectory onto the traversability map from the GND dataset [13]. The traversability score is then calculated as the percentage of fully traversable waypoints over the entire trajectory length.

Distance to Target: $1 - \frac{d_\tau - d_o}{|\tau|}$, where d_τ and d_o represent the distances between the target and the current trajectory and between the target and the optimal ground-truth trajectory, respectively. $|\tau|$ denotes the length of the generated trajectory.



(a) (b) (c)

Fig. 3: **Qualitative Evaluation:** Each column illustrates one of the three components in our trajectory generation pipeline, (a) CVAE-based multiple trajectory generation, (b) semantic segmentation for traversability analysis, and (c) VLM-based trajectory ranking in the outdoor scenarios. In (a) and (b), lighter colors indicate higher scores, confidence scores in (a) and traversability scores in (b). In (c), the VLM ranks trajectories based on social cues. The top example shows a ranking of [3, 0, 1, 2, 4]. The bottom example shows a ranking of [4, 5, 0, 1, 2, 3].



(a) (b) (c) (d)

Fig. 4: **Geometric and Semantic Analysis:** Both geometric trajectory generation and semantic segmentation perform well in most cases, but there are also failure cases, as marked in yellow circles in (c) and (d). When the elevations of the lawn and sidewalk are similar, the geometric model struggles to perform accurately. Additionally, segmentation fails in out-of-distribution scenarios.

3.1 Experimental Insights

As shown in Fig. 4 (a), geometry-based trajectory generation [14] performs well in scenarios where geometric structures are easily detected. However, when the elevations of off-road areas and sidewalks are similar, it often fails, as shown in (c). In Fig. 4 (a) and (b), lighter colors indicate higher scores in both geometric confidence and color-based traversability. As shown in (d), semantic segmentation also encounters out-of-distribution scenarios, where large ground areas cannot be properly segmented. Therefore, we use Equation 3 to integrate all components for optimal performance in traversability analysis. As shown in Table 1, our approach achieves the best traversability among all methods, and the generated trajectories lead to the target closely.

As shown in Fig. 3, beyond traversability analysis, VLMs also process labeled trajectories and analyze social cues from the images. In the first row, the VLM ranks the trajectories as [3, 0, 1, 2, 4], given the target at the right front. It

Method	Modality	Traversability (%) ↑	Distance to Target (%) ↑	Inference Time (s) ↓
PIVOT	RGB + Language	70	69	2.30
ViNT	RGB	57	62	0.69
NoMaD	RGB	59	61	0.24
MTG [14]	Points	61	64	0.01
DTG [15]	Points	67	66	0.13
VL-TGS [23]	Points, RGB, Language	65	70	2.31
MOSU (ours)	Points, RGB, Language	**77**	**73**	2.30

Table 1: **Quantitative Evaluation:** Our approach achieves the best Traversability and comparable Distance to Target.

suggests following trajectory 3 at normal speed, while other trajectories should be taken at a slower speed when encountering humans. The ranking for the second row is [4, 5, 0, 1, 2, 3] with normal speed. In the experiment, we observed that VLMs take a significant amount of time to process, as shown in Table 1, but they demonstrate high accuracy in understanding social cues, particularly in detecting movement directions.

Besides social and traffic understanding, as shown in Tab. 1, our approach achieves a comparable distance-to-target while attaining at least 10% higher traversability than other methods.

4 Conclusion

We propose a system for long-range navigation that considers traversability, as well as social and traffic constraints. The system integrates routing, trajectory generation, and motion planning, leveraging the benefits of geometric, semantic, and language information to enhance scene understanding and trajectory generation. Compared with other state-of-the-art (SOTA) approaches, our method achieves a comparable distance-to-target and improves traversability by at least 10%.

While the system demonstrates strong overall performance, some limitations remain. It has difficulty in detecting small cliffs, such as the vertical surfaces of ramps. When the robot is moving on ramps, it is challenging to detect very low vertical surfaces. Additionally, as with many learning-based approaches, the trajectory generation model lacks generalizability to out-of-distribution scenarios, which can lead to noisy trajectories and poor integration of geometric information. Future improvements may involve incorporating more robust geometric analysis methods, such as foundation geometric understanding models, to better evaluate geometric constraints.

References

1. Agishev, R., Petricek, T., Zimmermann, K.: Trajectory optimization using learned robot-terrain interaction model in exploration of large subterranean environments. IEEE Robotics and Automation Letters **7**(2) (2022) 3365–3371
2. Bresenham, J.E.: Algorithm for computer control of a digital plotter. In: Seminal graphics: pioneering efforts that shaped the field. (1998) 1–6
3. Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R.: Masked-attention mask transformer for universal image segmentation. (2022)

4. Davids, A.: Urban search and rescue robots: from tragedy to technology. IEEE Intelligent systems **17**(2) (2002) 81–83
5. Fox, D., Burgard, W., Thrun, S.: The dynamic window approach to collision avoidance. IEEE Robotics & Automation Magazine **4**(1) (1997) 23–33
6. Gao, P., Liu, Z., Wu, Z., Wang, D.: A global path planning algorithm for robots using reinforcement learning. In: 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), IEEE (2019) 1693–1698
7. Gasparetto, A., Boscariol, P., Lanzutti, A., Vidoni, R.: Path planning and trajectory planning algorithms: A general overview. Motion and Operation Planning of Robotic Systems: Background and Practical Approaches (2015) 3–27
8. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(12) (2020) 4338–4364
9. Hoffmann, T., Prause, G.: On the regulatory framework for last-mile delivery robots. Machines **6**(3) (2018) 33
10. Kim, Y., Lee, J.H., Lee, C., Mun, J., Youm, D., Park, J., Hwangbo, J.: Learning semantic traversability with egocentric video and automated annotation strategy. IEEE Robotics and Automation Letters (2024)
11. Kong, L., Liu, Y., Li, X., Chen, R., Zhang, W., Ren, J., Pan, L., Chen, K., Liu, Z.: Robo3d: Towards robust and reliable 3d perception against corruptions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). (2023) 19994–20006
12. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: Pointcnn: Convolution on x-transformed points. Advances in neural information processing systems **31** (2018)
13. Liang*, J., Das*, D., Song*, D., Shuvo, M.N.H., Durrani, M., Taranath, K., Penskiy, I., Manocha, D., Xiao, X.: Gnd: Global navigation dataset with multi-modal perception and multi-category traversability in outdoor campus environments. In: 2025 IEEE International Conference on Robotics and Automation (ICRA), IEEE (2025)
14. Liang, J., Gao, P., Xiao, X., Sathyamoorthy, A.J., Elnoor, M., Lin, M.C., Manocha, D.: Mtg: Mapless trajectory generator with traversability coverage for outdoor navigation. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE (2024) 2396–2402
15. Liang, J., Payandeh, A., Song, D., Xiao, X., Manocha, D.: Dtg: Diffusion-based trajectory generation for mapless global navigation. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (2024) 5340–5347
16. Nasiriany, S., Xia, F., Yu, W., Xiao, T., Liang, J., Dasgupta, I., Xie, A., Driess, D., Wahid, A., Xu, Z., et al.: Pivot: Iterative visual prompting elicits actionable knowledge for vlms. arXiv preprint arXiv:2402.07872 (2024)
17. Ozturk, U., Akdaug, M., Ayabakan, T.: A review of path planning algorithms in maritime autonomous surface ships: Navigation safety perspective. Ocean Engineering **251** (2022) 111010
18. Sathyamoorthy, A.J., Weerakoon, K., Elnoor, M., Zore, A., Ichter, B., Xia, F., Tan, J., Yu, W., Manocha, D.: Convoi: Context-aware navigation using vision language models in outdoor and indoor environments. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (2024) 13837–13844
19. Shah, D., Sridhar, A., Dashora, N., Stachowicz, K., Black, K., Hirose, N., Levine, S.: Vint: A foundation model for visual navigation. arXiv preprint arXiv:2306.14846 (2023)
20. Sohail, S.S., Himeur, Y., Kheddar, H., Amira, A., Fadli, F., Atalla, S., Copiaco, A., Mansoor, W.: Advancing 3d point cloud understanding through deep transfer learning: A comprehensive survey. Information Fusion (2024) 102601
21. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. Advances in neural information processing systems **28** (2015)

22. Song, D., Liang, J., Payandeh, A., Raj, A.H., Xiao, X., Manocha, D.: Vlm-socialnav: Socially aware robot navigation through scoring using vision-language models. IEEE Robotics and Automation Letters **10**(1) (2025) 508–515
23. Song, D., Liang, J., Xiao, X., Manocha, D.: Vl-tgs: Trajectory generation and selection using vision language models in mapless outdoor environments. IEEE Robotics and Automation Letters **10**(6) (2025) 5791–5798
24. Sridhar, A., Shah, D., Glossop, C., Levine, S.: Nomad: Goal masked diffusion policies for navigation and exploration. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE (2024) 63–70
25. Team, G.: Gemini: A family of highly capable multimodal models (2024)
26. Van Diggelen, F., Enge, P.: The world's first gps mooc and worldwide laboratory using smartphones. In: Proceedings of the 28th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2015). (2015) 361–369
27. Yurtsever, E., Lambert, J., Carballo, A., Takeda, K.: A survey of autonomous driving: Common practices and emerging technologies. IEEE access **8** (2020) 58443–58469