

Advancing Off-Road Autonomous Driving: The Large-Scale ORAD-3D Dataset and Comprehensive Benchmarks

Chen Min¹, Jilin Mei^{1,2}, Heng Zhai³, Shuai Wang^{1,2}, Tong Sun^{1,2}, Fanjie Kong⁴, Haoyang Li⁵, Fangyuan Mao^{1,2}, Fuyang Liu^{1,2}, Shuo Wang^{1,2}, Liang Chen¹, Fangzhou Zhao¹, Zhipeng Xiao⁶, Hanzhang Xue⁷, Hao Fu⁸, Yiming Nie⁶, Qi Zhu⁶, Liang Xiao⁶, Dawei Zhao⁶, Yu Hu^{1,2}

Abstract—A major bottleneck in off-road autonomous driving research lies in the scarcity of large-scale, high-quality datasets and benchmarks. To bridge this gap, we present ORAD-3D, which, to the best of our knowledge, is the largest dataset specifically curated for off-road autonomous driving. ORAD-3D covers a wide spectrum of terrains—including woodlands, farmlands, grasslands, riversides, gravel roads, cement roads, and rural areas—while capturing diverse environmental variations across weather conditions (sunny, rainy, foggy, and snowy) and illumination levels (bright daylight, daytime, twilight, and nighttime). Building upon this dataset, we establish a comprehensive suite of benchmark evaluations spanning five fundamental tasks: 2D free-space detection, 3D occupancy prediction, rough GPS-guided path planning, vision-language model-driven autonomous driving, and world model for off-road environments. Together, the dataset and benchmarks provide a unified and robust resource for advancing perception and planning in challenging off-road scenarios. The dataset is publicly available at <https://github.com/chaytonmin/ORAD-3D-Dataset-For-Off-Road-AD>.

I. INTRODUCTION

Off-road autonomous driving has attracted increasing attention in recent years due to its potential in enabling intelligent transportation and robotic systems to operate in unstructured and complex environments [1]. As illustrated in Fig. 1, unlike urban or highway settings—where well-marked lanes, standardized traffic rules, and high-quality maps provide strong priors—off-road scenarios are characterized by irregular terrains, sparse or absent road markings, unpredictable obstacles, and rapidly changing environmental conditions [2]. These factors pose significant challenges for perception, planning, and control systems, necessitating

This work was supported by National Natural Science Foundation of China under Grant No.U23B2034 and No.62176250, Beijing Natural Science Foundation (L259015 and L259049), and the Innovation Program of Institute of Computing Technology, Chinese Academy of Sciences under Grant No. 2024000112.

¹Research Center for Intelligent Computing Systems, SKLP, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 100190.

²University of Chinese Academy of Sciences, Beijing, China, 100190.

³Tongji University, Shanghai, China, 200092.

⁴Xi'an Jiaotong University, Shaanxi, China, 710049.

⁵Nanchang University, Jiangxi, China, 330047.

⁶Defense Innovation Institute, Beijing, China, 100073.

⁷Test Center, National University of Defense Technology, Xi'an, China, 710106.

⁸National University of Defense Technology, Xi'an, China, 410073.

* Corresponding author Dawei Zhao and Yu Hu. Email: adamzdw@163.com and huyu@ict.ac.cn



(a) On-road environment



(b) Off-road environment

Fig. 1. On-road environments are characterized by well-structured infrastructure—such as lane markings, traffic signs, and clearly delineated roadways—that provide consistent and reliable navigational cues. In stark contrast, off-road environments lack paved surfaces and structured guidance, instead featuring irregular, heterogeneous, and often ambiguous terrain. This absence of formal infrastructure greatly amplifies the complexity of perception and navigation tasks.

robust algorithms capable of handling extreme variability in both appearance and geometry [3].

Despite the growing interest in this domain, progress in off-road autonomous driving research has been hindered by the scarcity of large-scale, high-quality datasets. Existing datasets for autonomous driving predominantly focus on structured, on-road environments, limiting the generalization and robustness of models when deployed off-road. Moreover, the few publicly available off-road datasets are often constrained in size, diversity, or sensing modalities, leaving a gap in supporting comprehensive research on perception and planning under diverse off-road conditions.

To address this gap, we present ORAD-3D—to the best of our knowledge, the largest and most diverse dataset dedicated to off-road autonomous driving. The dataset was collected across a broad spectrum of terrains, including woodlands, farmlands, grasslands, riversides, gravel roads, cement roads, and rural areas. It encompasses a rich set of environmental conditions, covering multiple weather scenarios (sunny, rainy, foggy, and snowy) and illumination states (bright daylight, daytime, twilight, and nighttime). The data are captured using multi-sensor configurations, enabling both 2D and 3D perception tasks.

In addition to releasing ORAD-3D, we establish a comprehensive set of benchmark evaluations to enable fair and reproducible comparisons in off-road autonomous driving research. Specifically, we evaluate five core tasks: (1) 2D free-space detection, (2) 3D occupancy prediction, (3) rough GPS-guided path planning, (4) VLM-based autonomous driving, and (5) world model for off-road autonomous driving. Determining traversable areas is essential for safe off-road navigation; thus, we first construct a benchmark for 2D free-space detection based on large-scale off-road imagery. Recognizing the critical role of 3D terrain geometry, we further provide 3D occupancy annotations for predicting terrain structure in 3D space. Given that GPS signals in off-road environments are often noisy or inaccurate—making precise localization challenging—we design a benchmark for rough GPS-guided path planning to evaluate navigation under such uncertainty. Inspired by the rapid progress and strong generalization capability of Vision–Language Models (VLMs), we introduce a benchmark for VLM-based scene understanding and path planning in diverse off-road scenarios. Finally, we construct an off-road world model capable of controllably generating diverse future off-road scenarios. Together, these benchmarks span a representative spectrum of perception and decision-making challenges encountered in real-world off-road environments.

By providing both a large-scale, diverse dataset and well-defined benchmarks, ORAD-3D aims to serve as a foundational resource for the off-road autonomous driving community. We anticipate that it will accelerate the development of algorithms that are robust to terrain variability, adverse weather, and low-visibility conditions—ultimately pushing the frontier of safe and reliable off-road autonomous driving.

The highlights of our work are as follows:

- **ORAD-3D dataset** – We release the largest and most diverse off-road autonomous driving dataset to date, covering multiple terrains, weather conditions, and illumination levels with synchronized multi-sensor data.
- **Environmental diversity** – The dataset includes challenging and underrepresented scenarios such as woodlands, farmlands, grasslands, riversides, and rural roads, with extensive weather and lighting variations.
- **Benchmark suite** – We establish standardized benchmarks for 2D free-space detection, 3D occupancy prediction, rough GPS-guided path planning, VLM-based autonomous driving, and off-road world model.
- **Research resource** – ORAD-3D serves as a comprehensive platform for developing and evaluating robust perception and planning algorithms in unstructured off-road environments.

II. RELATED WORK

A. Off-Road Autonomous Driving

Off-road perception research primarily targets road segmentation and traversability estimation in unstructured, complex terrains. Recent works enhance passable area detection using advanced CNNs and Transformer architectures [4],

TABLE I
OVERVIEW OF OFF-ROAD AUTONOMOUS DRIVING DATASET.

Datasets	Sensors	Frames	Tasks	Extreme Weathers
DeepScene [22]	C	0.3K	Seg	no
RUGD [23]	C	7.5K	Seg	no
Rellis-3D [24]	C+L	13K	3DDet, Seg	no
GOOSE [25]	C+L	10K	Seg	no
ORFD [26]	C+L	12K	Seg	Rain, snow, fog, darkness, grassland, forest, rural area, etc.
ORAD-3D	C+L	57K	Seg, 3D Occ, Path Planning, VLM, World Model	Rain, snow, fog, darkness, grassland, forest, rural area, etc.

[5], while others integrate RGB imagery with point clouds to achieve more robust terrain mapping and navigation [6], [7]. Multi-modal fusion techniques combining RGB, depth, and LiDAR data have been employed to improve free-space detection and obstacle classification [8]–[11]. Noise-robust networks and Transformer-based models further refine segmentation accuracy in challenging conditions [12], [13]. Contrastive and self-supervised learning approaches reduce dependence on dense labels, enabling fine-grained terrain understanding from limited annotations [14]–[16]. Off-road semantic segmentation methods such as OFFSEG [17] and OffRoadTranSeg [18] tackle class imbalance and domain adaptation issues via semi-supervised strategies. Fusion-based frameworks leverage uncertainty modeling and attention mechanisms to enhance mapping and perception in unstructured environments [19], [20]. Xu et al. [21] propose an end-to-end Transformer-based framework for map-less autonomous driving, which takes raw LiDAR data and a noisy topometric map as inputs and generates precise local trajectories for navigation. While these approaches demonstrate promising results, most are trained and evaluated on small-scale datasets. To address this limitation, we introduce a large-scale off-road dataset with comprehensive benchmark results.

B. Datasets for Off-Road Autonomous Driving

Existing off-road autonomous driving datasets predominantly target perception-oriented tasks, such as traversability estimation and semantic segmentation. Identifying traversable areas is particularly critical for autonomous vehicles. TrailNet [27] is among the first datasets to examine road surface types using publicly available camera data. ORFD [26] provides high-resolution imagery and annotations for detecting navigable regions across diverse off-road conditions. Verti-Wheelers [28] addresses wheeled mobility on steep terrain, while M3-GMN [29] advances grid map-based navigation. For semantic segmentation in off-road environments, several datasets have been introduced. YCOR [30] offers multi-season imagery, whereas Botanic-Garden [31] contains annotated images for robot navigation in botanical gardens. RUGD [23] provides multimodal sensor data to support perception in outdoor scenes. More recently, RELIS-3D [24], GOOSE [25], and WildScenes [32] have released multimodal datasets for robust perception



Fig. 2. ORAD-3D dataset contains a variety of off-road scenes.



Fig. 3. Different weather conditions are considered in ORAD-3D.

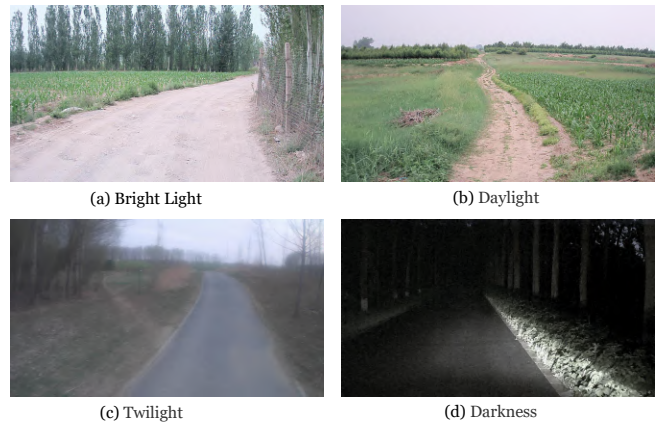


Fig. 4. Different light conditions are considered in ORAD-3D.

in complex natural environments. UnScenes3D [3] focuses on 3D semantic occupancy as a central representation for off-road understanding. While these datasets have significantly contributed to the field, most remain limited in scale and primarily emphasize perception tasks. In contrast, this work introduces a large-scale off-road dataset that not only supports perception tasks but also includes planning tasks and vision–language model–driven autonomy, thereby providing a more comprehensive resource for advancing off-road autonomous driving research.

III. ORAD-3D DATASET

We present ORAD-3D, a comprehensive off-road autonomous driving dataset curated from multi-season recordings spanning winter to summer. It covers diverse terrains, weather, lighting, and road types, providing rich data essential for robust perception and accurate future forecasting. ORAD-3D includes data such as images, LiDAR, pose, and depth. The provided annotations include 2D free-space segmentation, 3D occupancy, driving trajectories, and scene text descriptions.

A. Data Curation and Statistics

1) *Diverse Scenes*: Unlike the uniformity of structured on-road environments, off-road scenarios are characterized by diverse types, such as grasslands, forests, deserts, farmlands, and mountainous terrain, as shown in Fig. 2. Furthermore, off-road environments contain numerous elements that are irregular in shape, with blurred boundaries and semantically ambiguous categories. The heterogeneity and spatial disorder of open off-road scenes significantly exceed the cognitive limits of structured roadways. Structurally, the absence of lane markings and traffic signs removes clear guidance, while natural features like trees and deep ravines contribute to feature sparsity in models. To address these challenges, we have collected off-road data from a variety of environments, including grasslands, forests, deserts, and farmlands, laying the foundation for off-road autonomous driving. The data distribution across various scenes is shown in Table II.

2) *Weather Conditions*: Off-road scenarios often encounter sudden weather changes, such as shifts from sunny to rainy, foggy to snowy conditions, and so on, as shown in Fig. 3. Typically, there is a larger volume of autonomous driving data available for sunny weather. However, obtaining

TABLE II
DISTRIBUTION OF OFF-ROAD SCENES IN THE ORAD-3D DATASET. THE UNIT IS FRAMES.

Split	Grassland	Field	Wilderness	Forest	Riverside	Gravel	Rural	Asphalt	Cement	Mud	Total	%
Train	3,921	5,610	2,146	7,005	1,770	4,424	5,734	1,467	7,481	369	39,927	69.07%
Val	345	960	343	478	334	642	860	336	1,086	333	5,717	9.89%
Test	1,092	1,900	1,047	1,785	393	1,233	1,719	528	1,999	468	12,164	21.04%
Total	9.27%	14.65%	6.12%	16.03%	4.32%	10.90%	14.38%	4.03%	18.28%	2.02%	57,808	100%

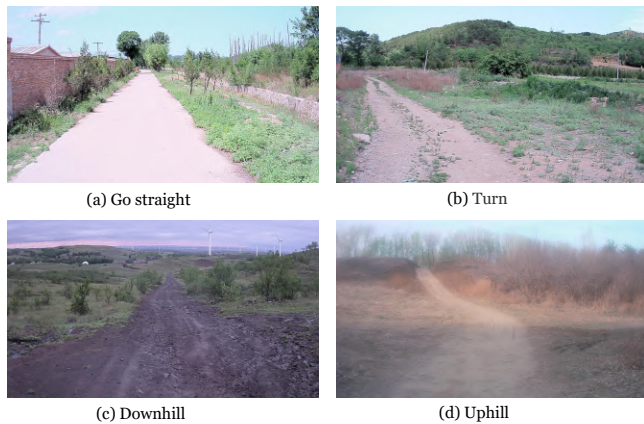


Fig. 5. Different road types are considered in ORAD-3D.

TABLE III
DISTRIBUTION OF WEATHER CONDITIONS IN ORAD-3D.

Split	Sun	Rain	Fog	Snow	Total
Train	15,737	11,741	8,863	3,586	39,927
Val	2,976	679	1,450	612	5,717
Test	5,209	2,654	3,958	343	12,164
Total	41.38%	26.08%	24.69%	7.86%	57,808

corner case data for rare and extreme weather scenarios, such as rain, fog, and snow, is crucial for enhancing the accuracy of off-road autonomous driving models. To address this, we have specifically collected off-road scene data under various extreme weather conditions, including rain, snow, fog, and sunny days. The scenes set on rainy days include moments both in the rain and after it has stopped. The data distribution across weather conditions is shown in Table III.

3) *Lighting Variations*: Lighting variations have a significant impact on autonomous driving performance. To address this, we have collected off-road scene data under various lighting conditions, including bright light, daylight, twilight, and darkness, as shown in Fig. 4. Data collection during the evening and nighttime is particularly challenging due to reduced visibility and the complexity of environmental factors. To overcome these challenges, we specifically focused on gathering data during twilight and nighttime, allowing the model to learn the unique distribution of off-road data under low-light conditions. By incorporating these varying lighting conditions into the training process, we aim to improve the robustness and accuracy of off-road autonomous

TABLE IV
DISTRIBUTION OF LIGHTING VARIATIONS IN ORAD-3D.

Split	Bright light	Daylight	Twilight	Darkness	Total
Train	7,515	15,536	6,840	10,036	39,927
Val	690	2,758	1,599	670	5,717
Test	2,938	5,964	550	2,712	12,164
Total	19.28%	41.96%	15.55%	23.21%	57,808

driving models in challenging lighting situations. Bright-light scenes are those in which local or global illumination exceeds normal levels of natural light or environmental reflectance, affecting scene perception. The data distribution across lighting variations is shown in Table IV.

4) *Road Type*: Autonomous vehicles are required to perform a variety of driving maneuvers, such as traveling in a straight line, turning, and navigating complex terrain. Generating video representations of future scenes based on these driving actions is highly valuable, as it allows the vehicle to anticipate and adapt to its surroundings. In off-road environments, the terrain is often rugged and uneven, with frequent uphill and downhill maneuvers that present significant challenges to both vehicle control and data collection. This makes it difficult to obtain comprehensive and diverse datasets that capture the full range of possible off-road driving scenarios. To address this challenge, we have collected a diverse set of video data corresponding to various driving actions, with a particular focus on off-road scenarios, as illustrated in Fig. 5.

5) *Data collection*: A sensor fusion kit, Pandora, developed by Hesai Technology, is mounted on the vehicle’s roof to capture LiDAR point cloud and RGB image data. The system is equipped with five cameras positioned around the lower part of the vehicle, including one color camera and four wide-angle black-and-white cameras. We collected 145 sequences from various off-road environments across China, spanning from spring to winter. Each sequence covers a distance of approximately 100 meters, and the size of the RGB images is 1280×720 pixels.

IV. ORAD-3D BENCHMARKS

Building on the large-scale off-road autonomous driving dataset ORAD-3D, we evaluate five core tasks: (1) 2D free-space detection, (2) 3D occupancy prediction, (3) rough GPS-guided path planning, (4) VLM-based autonomous driving, and (5) off-road world model.

A. 2D Free-Space Detection



Fig. 6. The annotation 2D off-road free-space detection.

1) *Label Generation*: Unlike the ORFD [26] dataset, which adopts a binary road/non-road labeling scheme, our annotation protocol (Fig. 6) captures the complexity of off-road environments by including fine-grained classes such as safe road, boundary transition zones, puddles, rocks, vehicles, and pedestrians. This detailed semantic representation enables more nuanced terrain understanding and supports safer autonomous navigation.

2) *Benchmark Method*: Building on existing off-road free-space detection methods, we conduct experiments on the ORAD-3D dataset using both multimodal and vision-only approaches. Detailed comparative analyses are presented in the experimental results section.

B. 3D Occupancy Prediction

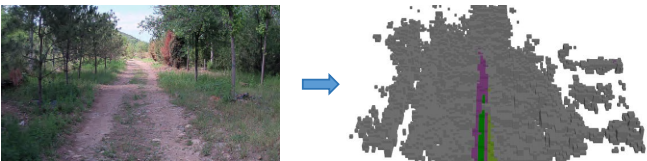


Fig. 7. The Annotation 3D occupancy prediction.

1) *Label Generation*: Relying solely on 2D off-road free-space detection results is insufficient for accurately modeling complex 3D terrains. To address this limitation, we further construct 3D occupancy annotations for off-road environments. Specifically, we first apply the LiDAR odometry method KISS-ICP [33] to register the collected LiDAR point clouds and obtain precise pose estimates. Multiple frames are then accumulated to generate dense point clouds, from which the 3D occupancy labels are subsequently derived (Fig. 7).

Existing off-road 3D occupancy prediction benchmarks, such as Wild-Occ [34], are collected from small-scale platforms with limited coverage and relatively simple terrain.

2) *Benchmark Method*: We conduct extensive experiments on ORAD-3D using a range of 3D occupancy prediction methods, encompassing both general-purpose approaches and algorithms tailored for off-road environments.

C. Rough GPS-guided Path Planning



Fig. 8. The flowchart for rough GPS-guided local path planning.

1) *Label Generation*: In off-road environments, GPS signals are often unreliable, preventing autonomous vehicles from obtaining accurate localization. To address this, we construct a rough GPS-guided path planning benchmark dataset. Specifically, we use the previously estimated poses as ground-truth driving trajectories and apply B-spline interpolation to generate waypoints. We then introduce controlled perturbations to these trajectories to simulate the effects of inaccurate GPS localization.

2) *Benchmark Method*: We adopt the method proposed by Xu et al. [21] as our benchmark for rough GPS-guided path planning (Fig. 8). This approach takes LiDAR BEV maps and rough routes as inputs, employing a Transformer-based backbone to predict waypoints. Building upon this, we further incorporate an uncertainty module that employs an attention mechanism to predict trajectory probabilities for different anchor points on the BEV map, thereby enhancing the accuracy and reliability of path planning.

D. VLM-based Autonomous Driving

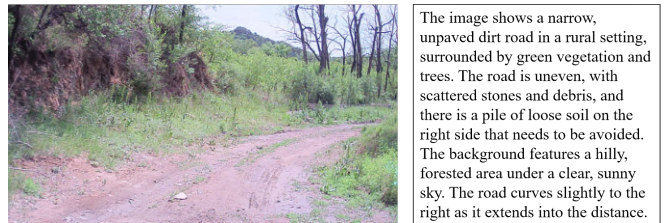


Fig. 9. Textual description annotation of off-road scenes.

1) *Label Generation*: Recent advances in VLMs have demonstrated their ability to analyze complex scenes through chain-of-thought reasoning, enabling end-to-end autonomous driving [35]. However, research on applying VLMs to off-road autonomous driving remains limited. In this work, we first leverage the multimodal large model Qwen2.5-VL [36] to annotate images from the ORAD-3D dataset, generating detailed scene descriptions as illustrated in Fig. 9. Subsequently, using pose data as ground-truth trajectories, we prompt the VLM to predict future paths, enabling end-to-end path planning in off-road scenarios.

2) *Benchmark Method*: We construct an off-road VLM-based autonomous driving benchmark on OpenEMMA [37], integrating a chain-of-thought (CoT) reasoning mechanism to handle the complexity of unstructured environments. As shown in Fig. 10, the model takes RGB images and prompt as inputs, performs step-by-step analysis of terrain, obstacles, and navigable areas, and outputs future trajectory waypoints

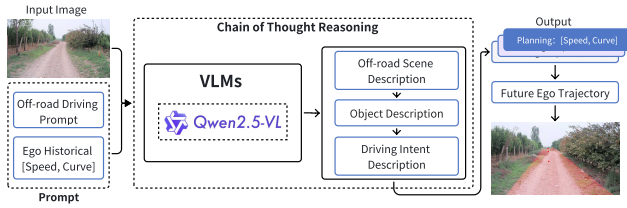


Fig. 10. The flowchart for VLM-based off-road autonomous driving.

for end-to-end path planning, leveraging VLM generalization while adapting reasoning to off-road challenges.

E. Off-road World Model

In contrast to the abundance of large-scale datasets readily available for urban autonomous driving, research on off-road autonomous driving remains significantly constrained by the scarcity of suitable data. This paper seeks to address this critical limitation by investigating off-road scene data generation through the development of a world model capable of producing diverse and controllable off-road scenarios [38]. Specifically, the proposed approach enables the synthesis of off-road data under extreme weather and illumination conditions, from multiple viewpoints, and across heterogeneous road environments, thereby substantially enhancing both the scale and diversity of off-road datasets.

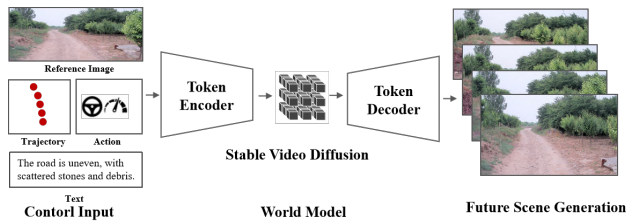


Fig. 11. The flowchart for off-road world model.

1) *Benchmark Method:* We introduce an off-road world model framework (Fig. 11) that redefines the architecture of Stable Video Diffusion (SVD) [39] for off-road autonomous driving applications. Diverging from conventional image-to-video generation paradigms, our approach introduces a dual-stream conditioning mechanism that simultaneously processes visual observations and vehicular control signals [40]. These features are then spatially aligned with the encoded visual features from the initial observation frame using cross-attention layers in the modified U-Net backbone [41]. This synergistic integration enables the diffusion model to generate temporally coherent video predictions that strictly adhere to specified conditions while maintaining visual consistency with the environmental context.

V. EXPERIMENTAL RESULTS

In this section, experiments are conducted to validate the performance of the proposed dataset and baseline method.

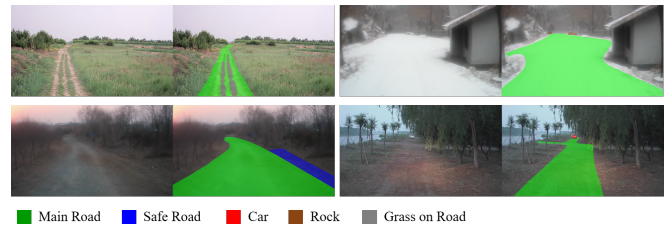


Fig. 12. Visualization of 2D free-space detection.

TABLE V
QUANTITATIVE RESULTS ON THE ORAD-3D TESTING SET

Method	OFF-Net [26]	M2F2-Net [10]	ROD [42]
mIoU \uparrow	62.4%	65.9%	70.3%

A. 2D Free-Space Detection

As shown in Fig. 12, the proposed off-road free-space detection algorithm (ROD [42]) effectively delineates off-road boundaries, median grass strips, and safe regions, thereby providing fine-grained road information that facilitates off-road autonomous driving. Quantitative results are presented in Table V. Compared to multimodal baselines such as M2F2-Net [10] and OFF-Net [26], the vision-only ROD [42], built upon a ViT backbone, demonstrates superior accuracy in predicting detailed road information.

B. 3D Occupancy Prediction

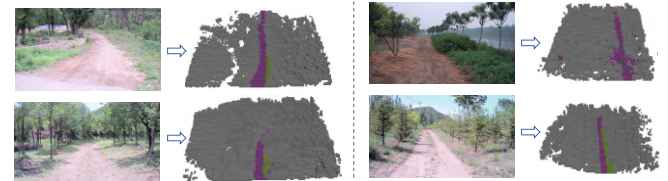


Fig. 13. Visualization of 3D occupancy prediction.

TABLE VI
3D SEMANTIC OCCUPANCY PREDICTION RESULTS ON ORAD-3D TEST SET.

Method	IoU \uparrow	mIoU \uparrow
OpenOcc [43]	9.54	5.83
OccFormer [44]	12.24	7.45
SurroundOcc [45]	11.87	7.60

Accurate 3D terrain modeling is crucial for off-road autonomous driving. As shown in Fig. 13, the predicted 3D occupancy results demonstrate that the model can capture the 3D structure of complex off-road environments. Quantitative results are reported in Table VI, where the fusion of LiDAR and vision yields higher prediction accuracy for 3D occupancy estimation.

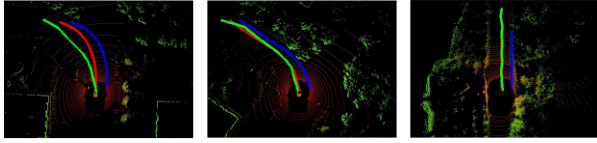


Fig. 14. Visualization of local path planning under the guidance of rough GPS. Blue: local route. Red: ground truth trajectory. Green: predicted trajectory.

TABLE VII
ROUGH GPS-GUIDED PATH PLANNING.

Method	FDE \downarrow	minADE \downarrow	HitRate \uparrow
Xu et al. [21]	2.34	1.07	0.55
Baseline	2.12	0.93	0.63

C. Rough GPS-guided Path Planning

As shown in Fig. 14, although GPS localization is imprecise, the model can effectively plan trajectories along the safe road centerline under the guidance of rough GPS, particularly in turning regions. This capability is crucial for off-road autonomous driving, where localization is often unreliable. Table VII provides the quantitative comparison of trajectory planning performance.

D. VLM-based Autonomous Driving

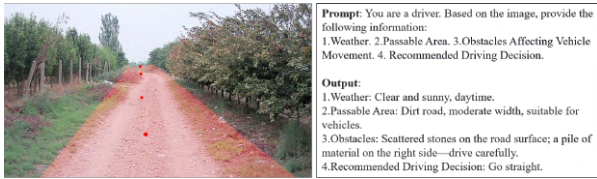


Fig. 15. Visualization of the predicted trajectories via VLM.

TABLE VIII
VLM-BASED PATH PLANNING.

Method	L2(m) \downarrow				FDE (m) \downarrow
	1s	2s	3s	Avg.	
Qwen2.5-VL-3B [36]	0.2123	0.4945	0.8392	0.5154	0.9324
Qwen2.5-VL-7B [36]	0.1867	0.4518	0.8161	0.4849	0.9219
Qwen3-VL-2B [46]	0.1818	0.4332	0.7643	0.4598	0.8575
Qwen3-VL-4B [46]	0.1625	0.3932	0.7278	0.4278	0.8255

As illustrated in Fig. 15, VLMs can perform fine-grained analysis of off-road scenes conditioned on prompts, such as recognizing objects, weather, and road conditions, and subsequently generating a planned trajectory. Quantitative results in Table VIII demonstrate that Qwen3-VL-4B outperform other models in path planning tasks.

E. Off-road World Model

The results in Table IX demonstrate that models trained on our dataset deliver superior visual quality. As shown in Fig. 16, the world model trained on the ORAD-3D dataset

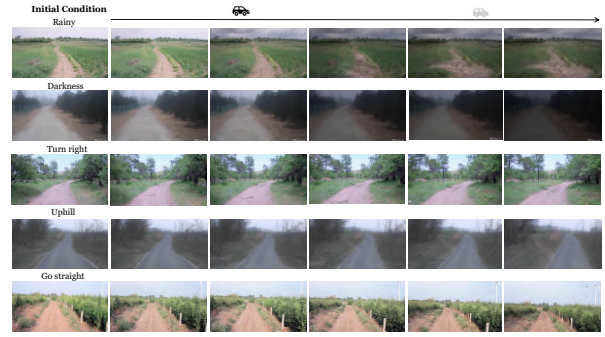


Fig. 16. Future driving video generation will involve the interaction of driving actions with various conditions, enabling the creation of driving videos that correspond to different scenarios and driving behaviors.

TABLE IX
QUANTITATIVE RESULTS OF OFF-ROAD WORLD MODEL.

Method	Dataset	FID \downarrow	FVD \downarrow
SVD [39]	-	79.5	534.2
Baseline	ORAD-3D	70.2	486.6

is capable of accurately generating future scene videos. Notably, the model can generate the required future scenarios based on specific conditions, greatly reducing the cost of collecting off-road scene data and advancing research in off-road autonomous driving.

VI. CONCLUSION

In this work, we introduced ORAD-3D, the largest and most diverse dataset for off-road autonomous driving to date. Collected across a wide variety of terrains, weather conditions, and illumination scenarios, ORAD-3D provides multi-sensor data to support both 2D and 3D perception tasks. Beyond data collection, we established four representative benchmarks—2D free-space detection, 3D occupancy prediction, rough GPS-guided path planning, VLM-based autonomous driving, and off-road world model—covering essential challenges in perception and decision-making for unstructured environments.

Limitations and Future Work. Off-road environments are highly complex, yet current datasets remain limited in scale and obstacle diversity. Future work will expand terrain coverage and capture more challenging scenarios.

REFERENCES

- [1] C. Min, S. Si, X. Wang, H. Xue, L. Luo *et al.*, “Autonomous driving in unstructured environments: How far have we come?” *arXiv preprint arXiv:2410.07701*, 2024.
- [2] Z. Zhao, G. Li, C. Min, and K. Lu, “Ot-drive: Out-of-distribution off-road traversable area segmentation via optimal transport,” *arXiv preprint arXiv:2601.09952*, 2026.
- [3] L. Chen, R. Song, H. Wu, B. Ding, L. Li, and F.-Y. Wang, “Scene as occupancy and reconstruction: A comprehensive dataset for unstructured scene understanding,” *Scientific Data*, vol. 12, no. 1, p. 1232, 2025.
- [4] C. Sun, P.-S. Ge, T. Zhang, and Q.-Y. Xiang, “Passable area detection in off-road environments based on improved pspnet,” in *International Conference on Computer Vision, Image and Deep Learning*. IEEE, 2023, pp. 687–692.

- [5] C. Chung, G. Georgakis, P. Spieler, C. Padgett, A. Agha, and S. Khattak, "Pixel to elevation: Learning to predict elevation maps at long range using images for autonomous off-road navigation," *IEEE Robotics and Automation Letters*, 2024.
- [6] T. Guan, Z. He, R. Song, D. Manocha, and L. Zhang, "Tns: Terrain traversability mapping and navigation system for autonomous excavators," *arXiv preprint arXiv:2109.06250*, 2021.
- [7] T. Guan, R. Song, Z. Ye, and L. Zhang, "Vinet: Visual and inertial-based terrain classification and adaptive navigation over unknown terrain," in *IEEE International Conference on Robotics and Automation*. IEEE, 2023, pp. 4106–4112.
- [8] J. Bae, J. Seo, T. Kim, H.-G. Jeon, K. Kwak, and I. Shim, "Self-supervised 3d traversability estimation with proxy bank guidance," *IEEE Access*, 2023.
- [9] Z. Feng, Y. Feng, Y. Guo, and Y. Sun, "Adaptive-mask fusion network for segmentation of drivable road and negative obstacle with untrustworthy features," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2023, pp. 1–6.
- [10] H. Ye, J. Mei, and Y. Hu, "M2f2-net: Multi-modal feature fusion for unstructured off-road freespace detection," in *IEEE Intelligent Vehicles Symposium*. IEEE, 2023, pp. 1–7.
- [11] O. Kim, J. Seo, S. Ahn, and C. H. Kim, "Ufo: Uncertainty-aware lidar-image fusion for off-road semantic terrain map estimation," *arXiv preprint arXiv:2403.02642*, 2024.
- [12] Y. Lv, Z. Liu, G. Li, and X. Chang, "Noise-aware intermediary fusion network for off-road freespace detection," *IEEE Transactions on Intelligent Vehicles*, 2024.
- [13] T. Yan, Y. Wang, H. Lv, H. Sun, D. Zhang, and Y. Yang, "Fsn-swin: A network for freespace detection in unstructured environments," *IEEE Access*, 2024.
- [14] B. Gao, S. Hu, X. Zhao, and H. Zhao, "Fine-grained off-road semantic segmentation and mapping via contrastive learning," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2021, pp. 5950–5957.
- [15] J. Seo, S. Sim, and I. Shim, "Learning off-road terrain traversability with self-supervisions only," *IEEE Robotics and Automation Letters*, 2023.
- [16] S. Jung, J. Lee, X. Meng, B. Boots, and A. Lambert, "V-strong: Visual self-supervised traversability learning for off-road navigation," *arXiv preprint arXiv:2312.16016*, 2023.
- [17] K. Viswanath, K. Singh, P. Jiang, P. Sujit, and S. Saripalli, "Offseg: A semantic segmentation framework for off-road driving," in *IEEE International Conference on Automation Science and Engineering*. IEEE, 2021, pp. 354–359.
- [18] A. Singh, K. Singh, and P. Sujit, "Offroadtranseg: Semi-supervised segmentation using transformers on offroad environments," *arXiv preprint arXiv:2106.13963*, 2021.
- [19] J. Lian, Y. Wang, L. Li, D. Sui, W. Ren, and T. Wang, "Research on off-road environment perception based on multi-sensor fusion," in *International Conference on Vehicular Control and Intelligence*. IEEE, 2023, pp. 1–6.
- [20] J. Kim and J. Seo, "Uncertainty-aware semantic mapping in off-road environments with dempster-shafer theory of evidence," *arXiv preprint arXiv:2405.06265*, 2024.
- [21] J. Xu, L. Xiao, D. Zhao, Y. Nie, and B. Dai, "Trajectory prediction for autonomous driving with topometric map," in *International Conference on Robotics and Automation*. IEEE, 2022, pp. 8403–8408.
- [22] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard, "Deep multispectral semantic scene understanding of forested environments using multimodal fusion," in *International symposium on experimental robotics*. Springer, 2016, pp. 465–477.
- [23] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2019, pp. 5000–5007.
- [24] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," *arXiv preprint arXiv:2011.12954*, 2020.
- [25] P. Mortimer, R. Hagmanns, M. Granero, T. Luettel, J. Petereit, and H.-J. Wuensche, "The goose dataset for perception in unstructured environments," *arXiv preprint arXiv:2310.16788*, 2023.
- [26] C. Min, W. Jiang, D. Zhao, J. Xu, L. Xiao, Y. Nie, and B. Dai, "Orfd: A dataset and benchmark for off-road freespace detection," in *International Conference on Robotics and Automation*. IEEE, 2022, pp. 2532–2538.
- [27] M. Hoveidar-Sefid and M. Jenkin, "Autonomous trail following using a pre-trained deep neural network," in *ICINCO*, 2018, pp. 113–120.
- [28] A. Datar, C. Pan, M. Nazeri, and X. Xiao, "Toward wheeled mobility on vertically challenging terrain: Platforms, datasets, and algorithms," *arXiv preprint arXiv:2303.00998*, 2023.
- [29] H. X. B. L. X. X. L. Guanglei Xie, Hao Fu and Z. Sun, "M3-gmn: A multi-environment, multi-lidar, multi-task dataset for grid map based navigation," in *International Conference on Intelligent Robots and Systems*. IEEE, 2024.
- [30] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics: Results of the 11th International Conference*. Springer, 2018, pp. 335–350.
- [31] Y. Liu, Y. Fu, M. Qin, Y. Xu, B. Xu, F. Chen, B. Goossens, P. Z. Sun, H. Yu, C. Liu *et al.*, "Botanicgarden: A high-quality dataset for robot navigation in unstructured natural environments," *IEEE Robotics and Automation Letters*, 2024.
- [32] K. Vidanapathirana, J. Knights, S. Hausler, M. Cox, M. Ramezani, J. Jooste, E. Griffiths, S. Mohamed, S. Sridharan, C. Fookes *et al.*, "Wildscenes: A benchmark for 2d and 3d semantic segmentation in large-scale natural environments," *arXiv preprint arXiv:2312.15364*, 2023.
- [33] I. Vizzo, T. Guadagnino, B. Mersch, L. Wiesmann, J. Behley, and C. Stachniss, "KISS-ICP: In Defense of Point-to-Point ICP – Simple, Accurate, and Robust Registration If Done the Right Way," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1029–1036, 2023.
- [34] H. Zhai, J. Mei, C. Min, L. Chen, F. Zhao, and Y. Hu, "Wildocc: A benchmark for off-road 3d semantic occupancy prediction," *arXiv preprint arXiv:2410.15792*, 2024.
- [35] L. Zhang, Y. Nie, H. Li, F. Kong, B. Zhang, S. Huang, K. Fu, C. Min, and L. Xiao, "A vision-language-action model with visual prompt for off-road autonomous driving," *arXiv preprint arXiv:2601.03519*, 2026.
- [36] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge, S. Song, K. Dang, P. Wang, S. Wang, J. Tang, H. Zhong, Y. Zhu, M. Yang, Z. Li, J. Wan, P. Wang, W. Ding, Z. Fu, Y. Xu, J. Ye, X. Zhang, T. Xie, Z. Cheng, H. Zhang, Z. Yang, H. Xu, and J. Lin, "Qwen2.5-vl technical report," *arXiv preprint arXiv:2502.13923*, 2025.
- [37] S. Xing, C. Qian, Y. Wang, H. Hua, K. Tian, Y. Zhou, and Z. Tu, "Openemma: Open-source multimodal model for end-to-end autonomous driving," in *Proceedings of the Winter Conference on Applications of Computer Vision*, 2025, pp. 1001–1009.
- [38] Z. Zhu, X. Wang, W. Zhao, C. Min, N. Deng, M. Dou, Y. Wang, B. Shi, K. Wang, C. Zhang *et al.*, "Is sora a world simulator? a comprehensive survey on general world models and beyond," *arXiv preprint arXiv:2405.03520*, 2024.
- [39] A. Blattmann, T. Dockhorn, S. Kulal, D. Mendelevitch, M. Kilian, D. Lorenz, Y. Levi, Z. English, V. Voleti, A. Letts *et al.*, "Stable video diffusion: Scaling latent video diffusion models to large datasets," *arXiv preprint arXiv:2311.15127*, 2023.
- [40] Y. Wang, K. Cheng, J. He, Q. Wang, H. Dai, Y. Chen, F. Xia, and Z. Zhang, "Drivingdojo dataset: Advancing interactive and knowledge-enriched driving world model," *arXiv preprint arXiv:2410.10738*, 2024.
- [41] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-assisted Intervention*. Springer, 2015, pp. 234–241.
- [42] T. Sun, H. Ye, J. Mei, L. Chen, F. Zhao, L. Zong, and Y. Hu, "Rod: Rgb-only fast and efficient off-road freespace detection," *arXiv preprint arXiv:2508.08697*, 2025.
- [43] H. Jiang, Y. Xu, Y. Zeng, H. Xu, W. Zhang, J. Feng, and L. Zhang, "Openocc: Open vocabulary 3d scene reconstruction via occupancy representation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2024, pp. 1145–1152.
- [44] Y. Zhang, Z. Zhu, and D. Du, "OccFormer: Dual-path Transformer for Vision-based 3D Semantic Occupancy Prediction," *IEEE/CVF International Conference on Computer Vision*, pp. 9399–9409, 2023.
- [45] Y. Wei, L. Zhao, W. Zheng, Z. Zhu, J. Zhou, and J. Lu, "Surroundocc: Multi-camera 3d occupancy prediction for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 21 729–21 740.
- [46] S. Bai, Y. Cai, *et al.*, "Qwen3-vl technical report," *arXiv preprint arXiv:2511.21631*, 2025.