

CLEMSON OFF-ROAD DATASET (CORD): A MULTIMODAL BENCHMARK FOR AUTONOMY BEYOND THE FLAT-WORLD

Ashish Patil¹, Prakhar Gupta¹, Mayuresh Bhosale¹,
Arthur Mukwaya³, Akinbobola Jegede³, Dariusz Mikulski²,
Judith Mwakalonge³, and Yunyi Jia¹

¹Clemson University, Greenville, SC, USA.

²US Army DEVCOM GVSC, Warren, MI, USA

³South Carolina State University, Orangeburg, SC, USA

ABSTRACT

While autonomous perception has matured within the structured confines of urban roadways, it remains brittle when confronting the chaotic, non-rigid terrain of the natural world. This paper introduces the Clemson Off-Road Dataset, a high-fidelity, multimodal dataset engineered to bridge this gap by challenging standard "flat-world" assumptions. Featuring 2.90 TB of sensor data, the dataset captures a diverse spectrum of unstructured environments, ranging from the transitional trails of CU-ICAR and the day/night lighting dynamics of TN3 to the unstructured wilderness of Camp Daniels and the novel coastal scenery of Edisto Island. Distinguishing itself from existing forest-centric benchmarks, the Clemson Dataset provides a first-of-its-kind focus on coastal data, featuring unique adversarial conditions such as extreme solar glare, loose sand, and shifting tide lines. The data is collected aboard a Polaris RZR Pro R 4, a high-performance platform integrated with a sensor suite designed to perceive physics beyond geometry. Alongside 360° HD camera coverage, 3D LiDAR, and Radar, we integrate Cubert Ultris Hyperspectral imaging and Prophesee EVK4 Event-based vision to enable material classification and high-dynamic-range motion tracking. To overcome the bottleneck in ground truth generation, we used our "AI LabelMate," a context-aware semi-automated annotation agent that fuses Vision-Language Models (Florence-2) with SAM2 to generate 6331 pixel-perfect annotated frames using a specialized off-road ontology and a human-in-the-loop pipeline. We establish performance baselines using Oneformer for semantic segmentation and used SalsaNext for lidar point clouds labelling. Available in both raw ROS2 bag and extracted standard formats, this Dataset serves as a pivotal testing ground for the next generation of robust autonomous systems. The dataset of this paper is available upon request to the Virtual Prototyping of Autonomy-Enabled Ground Systems (VIPR-GS) Center.

Citation: A. Patil, P. Gupta, M. Bhosale, A. Mukwaya, D. Mikulski, J. Mwakalonge, and Y. Jia, "Clemson Off-Road Dataset (CORD): A Multimodal Benchmark for Autonomy Beyond the Flat-World" In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*, NDIA, Novi, MI, Aug. 11-13, 2026.

1. INTRODUCTION

Autonomous Ground Vehicles (AGVs) have achieved significant levels of maturity in structured urban environments, largely due to the availability

of large, high-quality datasets such as KITTI [1], Cityscapes [2], and nuScenes [3], and BDD100K [4], driven by foundational advances in deep

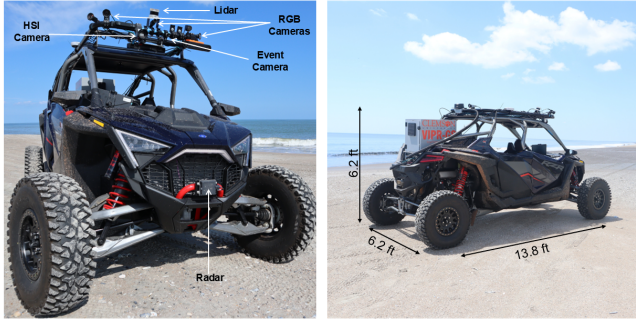


Figure 1: The Clemson Off-Road Dataset (CORD) platform: A Polaris RZR Pro R 4 Ultimate equipped with a multimodal sensor suite including Hyperspectral (HSI), Event, and RGB cameras, LiDAR, and Radar.

learning [13]. These benchmarks operate under a "flat-world" assumption, where drivable surfaces are planar, boundaries are geometric (curbs, lanes), and semantic actors are rigid. However, for applications in search and rescue, military intelligence, and agricultural automation, vehicles must operate in unstructured, off-road environments where these assumptions collapse. In the wild, the terrain is not merely a surface to be traversed but a dynamic medium that interacts with the vehicle's physics. A patch of green may represent solid grass or a non-traversable mud pit; a sandy slope may offer traction or result in a devastating slip depending on moisture content.

Despite the critical need for off-road autonomy, the data structure remains very limited. Datasets such as RUGD [5] and RELIS-3D [6] have advanced semantic understanding in unstructured environments but are limited in particular aspects. RUGD depends on a small mobile robot platform with a low vantage point, failing to capture the suspension dynamics of full-scale passenger vehicles. While RELIS-3D introduces a robust Warthog platform and multimodal LiDAR-camera fusion, it lacks the spectral diversity required to distinguish material properties and is geographically confined to land-based scrub and forest-dominated environments. Similarly, while TartanDrive [7] provides massive interaction data for dynamics modeling, it does not focus on high-fidelity semantic

segmentation of complex environments like coastal zones.

To bridge this gap, we introduce the Clemson Off-Road Dataset, a high-fidelity, multimodal repository specifically engineered to advance perception in environments where standard assumptions fail. Spanning 2.90 TB of off-road data, our dataset captures the dynamics of a high-performance Polaris RZR Pro R 4 traversing diverse venues in South Carolina, ranging from dense forests to coastal dunes.

The primary contributions of this paper are:

1. **First-of-its-kind Coastal Data:** We provide the first major autonomy dataset featuring Edisto Island, capturing coastal topography, shifting tide lines, and extreme solar glare on wet sand scenarios absent in current benchmarks.
2. **Hyper-Heterogeneous Sensor Suite:** Beyond standard Camera-LiDAR setups, we integrate Cubert Hyperspectral imaging for material classification and Prophesee Event-based vision for high-dynamic-range motion tracking, enabling research into "physics-aware" perception.
3. **Semi-Automated Ground Truth:** We utilized our recently designed context-aware annotation agent "AI LabelMate" [25] which fuses Vision-Language Models (Florence-2 [23]) with Segmentation Foundation Models (SAM2) [17] to generate 6,331 pixel-perfect semantic labels efficiently and 1,591 lidar labels using SalsaNext [14].

2. RELATED WORK

The development of the Clemson Off-Road Dataset builds upon a family of foundational efforts in robotic perception. Table 1 provides a comparative overview of the Clemson dataset against prominent existing benchmarks.

Table 1: Comparison of Off-Road Autonomy Datasets

Dataset	Platform	Environment	Sensor Modalities	Annotation Type	# Annotations
RUGD [5]	Husky UGV	Creek, Park, Trail	Camera	Pixel-wise	7,546 (Images)
RELLIS-3D [6]	Warthog UGV	Scrub, Forest	Camera, LiDAR	Pixel/Point-wise	6,235 (Img) / 13,556 (Scan)
TartanDrive [7]	Yamaha ATV	Forest, Mud	Cam, IMU, Shocks	Interaction	~200k Interactions
GO Dataset [8]	Warthog UGV	Forest, Trail	Cam, LiDAR, Radar, Thermal	Pixel-wise	N/A
Clemson (Ours)	Polaris RZR	Coastal, Forest	Cam, LiDAR, Radar, Event, HSI	Pixel-wise (AI-Assisted)/ Point-wise	6,331 (Images)/ 1,591 (Scan)

2.1. Urban and Structured Datasets

The foundation of modern autonomous driving lies in urban datasets. KITTI [1] established the standard for sensor suites (Camera+LiDAR) and benchmarks in odometry and object detection [16]. Cityscapes [2] advanced the field of semantic segmentation [15, 18] with high-quality pixel-level annotations. However, models trained on these datasets fail catastrophically in off-road settings due to domain shifts: “road” changes from asphalt to mud, and “obstacles” change from cars to hanging tree branches or negative obstacles like ditches.

2.2. Off-Road and Unstructured Datasets

To address the domain gap, several off-road datasets and traversability models [11, 19, 20] have emerged:

RUGD [5]: Following early efforts like the Yamaha-CMU dataset [10], the Robot Unstructured Ground Driving dataset was among the first to categorize off-road terrain systematically. Collecting data from a small Husky UGV, RUGD introduced an ontology of 24 classes. However, its small form factor restricts the sensor viewpoint to near-ground levels, limiting its applicability to full-scale vehicle dynamics.

RELLIS-3D [6]: As a direct successor to RUGD, RELLIS-3D utilizes a Clearpath Warthog to capture synchronized LiDAR and camera data in Texas scrubland. It serves as the current state-of-the-art for multimodal off-road segmentation. However, RELLIS-3D lacks the sensors necessary for material property inference [9] (hyperspectral/thermal) and does not cover the coastal domains critical for amphibious autonomy.

TartanDrive [7]: Focusing on vehicle-terrain interaction, TartanDrive collected ~200,000 interactions using a Yamaha ATV. It is the gold

standard for learning dynamics models (e.g., slip prediction). Unlike our dataset, TartanDrive focuses less on dense semantic perception and lacks the high-fidelity sensor suite (e.g., 64-beam LiDAR, Event Cameras) required for detailed 3D mapping.

GO (Great Outdoors) [8]: Recently, the GO dataset introduced diverse modalities including thermal and radar to handle degraded environments. While comprehensive, it remains forest-centric, similar to prior work in under-canopy agricultural navigation [12]. Our dataset differentiates itself by explicitly targeting the Coastal/Beach domain and integrating Neuromorphic (Event) and Hyperspectral vision to solve specific edge cases like sun glare and soil moisture classification.

2.3. Comparison and Analysis

As shown in Table 1, the Clemson Dataset is unique in its integration of specific “physics-aware” sensors (Hyperspectral, Event) and its environmental focus on coastal topography. While RUGD and RELLIS-3D provide excellent baselines for forest trails, they do not capture the non-rigid surface physics of sand dunes or the optical challenges of tide lines found in our venue 4, Edisto Island.

3. SENSOR SUITE & DATA ACQUISITION

To capture the phenomenological complexity of these environments, the research team built a Ros2 based data collection platform that serves as a mobile remote sensing laboratory. The hardware architecture prioritizes high-bandwidth, synchronized sensor data.

3.1. Hardware & Platform: The Autonomous Polaris RZR

The vehicle used for research and data collection is a Polaris RZR Pro R 4 Ultimate, a high-performance

Table 2: CORD Sensor Specifications and Off-Road Utility

Sensor Type	Model	Specification	Primary Utility in Off-Road
LiDAR	Ouster OS1-64	64 Channel, 10-20 Hz	3D geometry, foliage penetration, and night navigation. The 45° V-FOV is critical for overhanging branches.
RGB Camera	Lucid Vision (x4)	Triton Series (IMX490)	High-resolution texture and semantic segmentation; HDR handles contrast of forest vs. sunlit dunes.
Event Camera	Prophesee EVK4	1280x720, μ s latency	High-speed motion estimation and operation in extreme dynamic range or low-light conditions.
Hyperspectral	Cubert Ultris X20P	350-1000 nm, 164 bands	Material classification (mud vs. dry dirt), vegetation health, and water detection via spectral absorption.
Radar	Continental ARS	77GHz (ARS 408-21)	Detection of dynamic agents through dust, fog, and rain; long-range tracking up to 250m.
GNSS/INS	Novatel PwrPak7	RTK-enabled, coupled	Ground truth localization and georeferencing of point clouds.

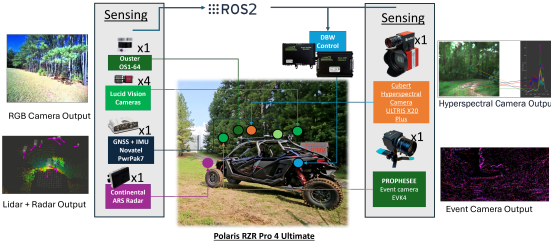


Figure 2: The CORD Research Platform: An autonomy-enabled Polaris RZR Pro R 4 Ultimate serving as a high-performance mobile laboratory. The integrated sensor suite provides 360° coverage and redundant modalities, including 64-channel Ouster LiDAR, Lucid Vision RGB cameras, Continental Radar, Prophesee Event and Cubert Hyperspectral cameras.

vehicle chosen for its mechanical robustness. Unlike smaller platforms used in related research like the Husky (RUGD) or Warthog (RELLIS), the RZR features a long-travel active suspension system capable of navigating dunes and rock crawls at speeds relevant to tactical driving.

Drive-by-Wire (DBW) System: The vehicle is equipped with a Dataspeed DBW kit, enabling precise electronic control over steering, throttle, braking, and shifting. This allows for:

- **Repeatability:** Conducting multiple runs over the same terrain at controlled velocities to test sensor consistency.

- **Safety:** Integrated E-Stop functionality and programmatic intervention layers.

Compute and Network: Data logging is handled by a ruggedized Nuvo on-board computer, selected for its resilience to extreme vibration and thermal loads. High-throughput sensors (LiDAR, 4x Cameras) are routed through a 10GbE-capable network switch to ensure zero frame drops during recording.

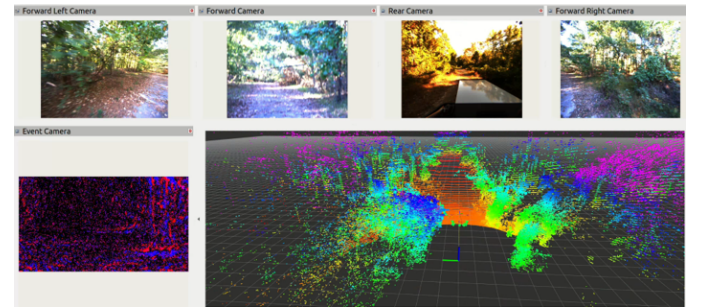


Figure 3: Multi-modal sensor visualization from the CORD platform. The top row displays the 360° coverage provided by the four HDR RGB cameras (Forward Left, Forward, Rear, and Forward Right). The bottom row highlights the specialized off-road perception modalities: the bottom left shows the asynchronous, high-dynamic-range output from the Event Camera, while the bottom right illustrates the dense 3D geometric environment captured by the 64-channel LiDAR point cloud.

3.2. The Sensor Suite

The sensor configuration is designed to provide overlapping, redundant modalities that cover the failure modes of individual sensors (see Table 2).

3.2.1. HDR RGB Cameras

CORD features four Lucid Vision Triton cameras to provide high-resolution texture and semantic context. Equipped with IMX490 sensors, these cameras offer a 120dB High Dynamic Range (HDR) and LED Flicker Mitigation, which are essential for maintaining visibility when transitioning between the deep shadows of forest canopies and the extreme solar glare of coastal dunes. These cameras provide the base data for the semantic segmentation masks generated via our context-aware AI-assisted pipeline, AILabelMate.

3.2.2. High-Resolution 3D LiDAR

The Ouster OS1-64 provides the core geometric foundation, capturing 64 channels of high-density point clouds. Its large 45° vertical field-of-view (V-FOV) is a critical design choice for off-road navigation, enabling the detection of negative obstacles like ditches and high-altitude hazards such as overhanging tree branches. Furthermore, the LiDAR's active sensing capability ensures geometric reliability in the TN3 "Night Scenario," where passive RGB sensors fail due to lack of ambient illumination.

3.2.3. Long-Range Radar

A Continental ARS 77GHz radar is integrated to provide robust object detection in "degraded visual environments" where optical sensors are limited. It maintains the ability to track dynamic agents at distances up to 250m through heavy dust, fog, and rain. In off-road scenarios, this modality serves as a redundant safety layer for identifying other vehicles or pedestrians in low-visibility conditions.

3.2.4. GNSS

The Novatel PwrPak7 provides high-precision ground truth localization and georeferencing. The system employs RTK-enabled GNSS with a tactical-grade IMU to provide accurate pose estimation during aggressive pitch and roll maneuvers encountered while traversing rough terrain or sand dunes.

3.2.5. The Cubert Ultris Hyperspectral Camera

The inclusion of the Cubert Ultris X20P [21] is a distinctive feature. Most datasets rely on 3-band RGB cameras, whereas the Ultris captures 164 bands between 350 nm and 1000 nm. In RGB, wet mud and dark shadows often appear identical. However, in the Near-Infrared (NIR) spectrum, water absorbs light strongly (appearing dark), while vegetation reflects NIR strongly. This allows algorithms to distinguish between a shadow on safe grass and a non-traversable mud pit.

3.2.6. The Prophesee Event Camera

The Prophesee EVK4 [22] utilizes neuromorphic sensing principles where each pixel operates independently, firing an "event" only upon intensity changes. The EVK4 maintains temporal fidelity in the microsecond range, allowing the vehicle to track features while vibrating heavily over off-road terrain.

3.3. Synchronization and Data Architecture

Synchronized multi-modal data requires precise temporal alignment. CORD utilizes ROS 2 bag files as the primary data storage format.

- **PTP Synchronization:** Sensors are synchronized using Precision Time Protocol (PTP) over Ethernet, ensuring clock agreement between the Ouster LiDAR, Lucid Cameras, and the compute unit.
- **Spatial Calibration:** Rigid transforms (extrinsics) are calibrated and provided as

static TF trees, allowing users to project 3D LiDAR points directly onto 2D camera images.

4. VENUE DIVERSITY AND ENVIRONMENTAL BREADTH

To validate autonomy across the full spectrum of operational domains, data collection was executed across four distinct venues. Each venue was selected to represent a specific class of "unstructured" challenge, ranging from visibility-limited forests to traction-limited coastal zones.

4.1. Venue 1: CU-ICAR

Environment: A controlled off-road trail within a research campus featuring patches of mud, manicured grass, and dense localized vegetation.

Specific Challenge: Hanging obstacles and wood logs. This venue explicitly tests the detection of overhanging tree branches. This is a notorious failure case for 2D planar navigation stacks, which often assume obstacles are connected to the ground. CORD captures branches that clear the vehicle hood but impact the sensor tower, necessitating full 3D voxel occupancy checking for safe navigation.

4.2. Venue 2: Technology Neighborhood 3

Environment: A transitional zone combining open fields with dense tree lines.

Specific Challenge: Edge Cases in Lighting TN3 was recorded in two distinct modes to stress-test sensor robustness:

- **Day Scenario:** Features extreme sunny conditions with direct sun-glare. This creates significant lens flare and washes out RGB sensors, forcing reliance on LiDAR reflectivity and the High Dynamic Range (HDR) of the Event Camera.
- **Night Scenario:** As shown in Figure 5 under zero ambient light; standard RGB cameras degrade due to high noise and motion blur, whereas active LiDAR preserves geometric boundaries and the event camera reliably tracks high-speed contrast features.

4.3. Venue 3: Camp Daniels, Elloree, SC

Environment: A complex mixture of gravel roads, sandy leaf-covered trails, and lake-side paths.

Specific Challenge: Water Hazards. Trails in this venue run adjacent to lakes and ponds, where detecting the water's edge is critical. The dataset captures the transition from high-friction gravel to medium-friction sand and finally to water boundaries. It further includes maneuverability obstacles, such as fallen wooden logs, to force complex path planning and swerving behaviors rather than simple reactive stopping.

4.4. Venue 4: Edisto Island, SC

Environment: Public beach, dunes, tidal zones, and camp-to-beach transition trails.

Specific Challenge: Surface Physics and Dynamic Semantics.

- **Traction & Soft Sand:** Several sequences capture the vehicle's loss of traction in soft, dry sand compared to firm, packed wet sand. This variation is geometrically invisible (both surfaces are flat) but critical for vehicle control.
- **High-Speed Runs:** CORD includes high-speed beach maneuvers to capture slip dynamics, essential for training Model Predictive Control (MPC) algorithms that must account for sideslip.
- **Dynamic Agents:** Unlike static forests, the beach environment is highly dynamic. The data includes migrating bird flocks (aerial obstacles), pedestrians near tide lines, and semi-static obstacles like umbrellas and tents.

5. DATASET ONTOLOGY

The CORD ontology is specifically designed to address the semantic complexity of unstructured coastal and forest environments, moving beyond the standard urban classes found in traditional datasets. We categorize most distinct semantic labels into six primary functional groups to facilitate high-fidelity traversability analysis and scene understanding.

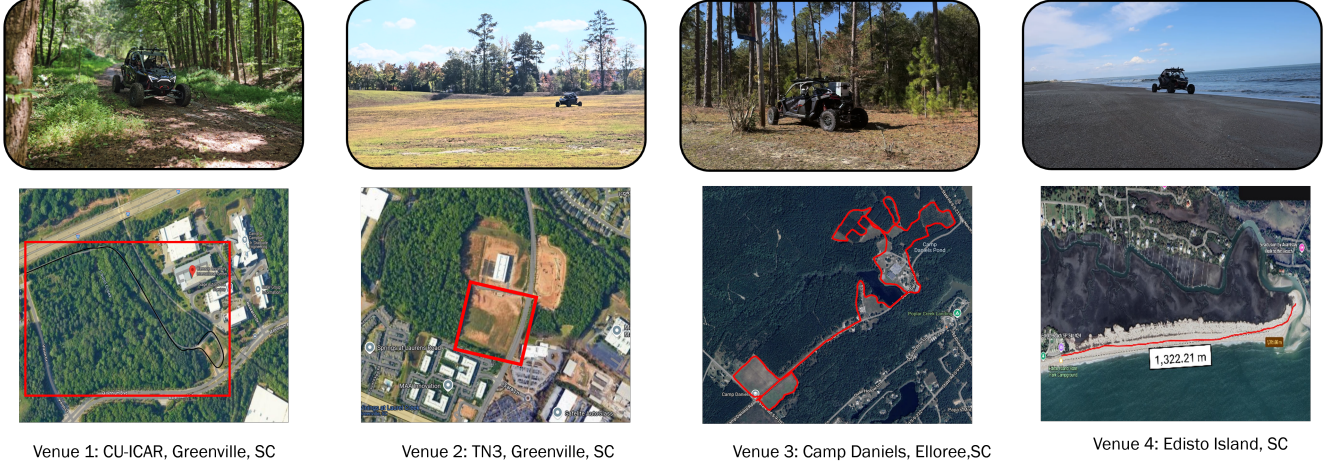


Figure 4: Geographic Layout of CORD Venues: Satellite imagery and GPS trajectories for the primary collection sites: (1) CU-ICAR in Greenville, SC; (2) Technology Neighborhood 3 (TN3) grassland; and (3) Camp Daniels dense forest with water bodies (4) Edisto Island coastal beach and tidal zones. The red boundaries indicate the primary operational domains where high-fidelity multimodal data was recorded.

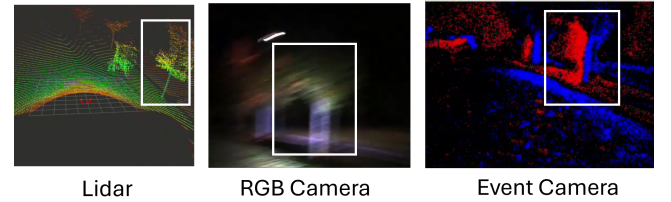


Figure 5: Venue 2 (TN3) Night Data Comparison: A qualitative evaluation of sensor performance in total darkness. While the passive *RGB Camera* (middle) acts as a noise generator with significant motion blur and low signal-to-noise ratio, the active *LiDAR* (left) maintains geometric fidelity. The *Event Camera* (right) demonstrates its microsecond-level temporal resolution and high dynamic range by successfully tracking features that are invisible to standard frame-based sensors in dark conditions.

Table 3: CORD Semantic Ontology: Categories and Included Labels

Category	Included Semantic Labels
Terrain / Ground	Sand, Asphalt, Concrete, Mud, Dirt, Gravel, Grass, Fallen Leaves/Mulch.
Water Features	Ocean, Pond, Puddle.
Vegetation	Trees, Low Trees, Bush, Brush Wood, Tree Log.
Dynamic Agents	Person, Car, Ego Vehicle, Bicycle, Animal, Travel Trailer.
Infrastructure	Building, Wall, Pier, Street Light, Pole, Traffic Sign, Fence, Road Barrier.
Beach Specifics	Dune Fence, Beach Gear, Beach Sign, Turtle Nest.

5.1. Semantic Ontology and Labeling Protocol

To facilitate robust traversability analysis and scene understanding, the CORD dataset employs a comprehensive ontology tailored for unstructured domains. This ontology encompasses 42 distinct semantic classes organized into six high-level functional categories (see Table 3).

6. DATASET STATISTICS

Figure 6 shows the distribution of classes in image annotations in the CORD dataset. The distribution is highly imbalanced, which is a common characteristic for off-road semantic segmentation benchmarks. The

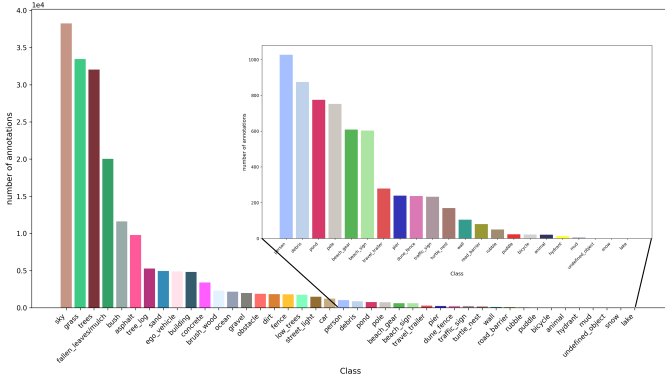


Figure 6: Semantic Annotation Class Distribution. The main bar chart illustrates the frequency of image annotations across the 42 semantic categories of the CORD dataset. Natural terrain elements—*sky*, *grass*, *trees*, and *fallen_leaves/mulch*—dominate the distribution, accounting for 64.79% of the 191,003 total labeled instances. The inset provides a magnified view of the long-tail distribution, highlighting the extreme rarity of critical navigational classes such as *turtle_nest*, *debris*, and dynamic agents like *person*, which underscore the inherent challenges of off-road perception compared to urban environments.

image annotations are dominated by *sky*, *grass*, *trees* and *fallen_leaves/mulch* which add up to 64.79% of the total labeled annotations (191,003 annotations over 42 classes). The top three classes *sky* (20.02%), *grass* (17.52%), and *trees* (16.77%) contribute to 54.31% of all the annotations, showing that natural terrain features are highly concentrated in terms of annotations.

The dataset has a clear long-tail distribution, with many classes containing very few samples. Less than 1% of the total annotations are in the categories of *person* (0.54%), *debris* (0.46%), *pond* (0.41%) and *pole* (0.39%). Rare classes are *mud* (<0.01%), *hydrant* (0.01%), *animal* (0.01%) and *bicycle* (0.01%) with less than 25 annotations per class. This stark imbalance reflects the inherent challenges of off-road perception, where the natural environment dominates the visual field, while man-made objects and specific natural hazards are sparse.

Non-uniform distributions are common in semantic segmentation datasets, but the imbalances in off-road environments are strongly tied to the venue’s topography. This imbalance is more pronounced than is usually seen in urban datasets,

and it poses a substantially greater challenge to segment rare classes such as obstacles, infrastructure, and dynamic agents. These minority classes are crucial for safe autonomous navigation but are significantly less frequent than the dominant natural terrain classes.

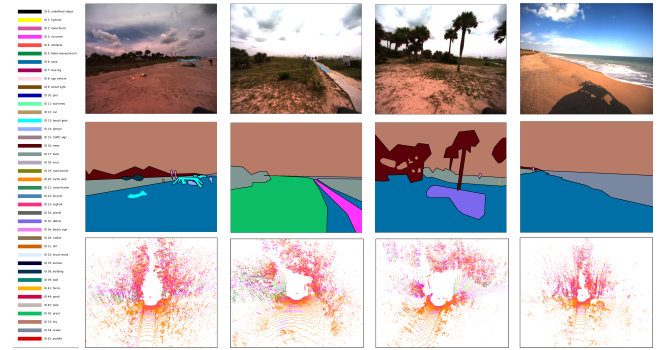


Figure 7: CORD Semantic Segmentation Samples: Multimodal Annotation Overview of the Clemson Off-Road Dataset (CORD). The figure showcases (top row) raw RGB camera images from diverse coastal and trail venues, (middle row) corresponding high-fidelity semantic segmentation masks generated via the AI LabelMate pipeline, and (bottom row) 3D LiDAR point clouds annotated using the same 42-class ontology. The legend (left) illustrates the comprehensive semantic mapping organized into functional categories, ranging from dominant terrain features to rare obstacles like tree logs and coastal infrastructure. Qualitative examples of pixel-perfect ground truth generated across venue 4. To ensure temporal diversity, we have labeled 6,331 RGB images downsampled at 5-second intervals across the total recording duration. Additionally, 1,591 LiDAR point clouds have been annotated to support 3D multimodal perception. The masks demonstrate the specialized off-road ontology, accurately delineating non-rigid terrain boundaries, complex vegetation, and coastal tide lines.

Table 4: CORD Dataset Distribution: Duration and Volume per Venue

Venue Environment	Total Data Volume (GB)
Venue 1: CU-ICAR	219
Venue 2: Technology Neighborhood 3	526
Venue 3: Camp Daniels	905
Venue 4: Edisto Island	1250
Total Aggregate	2900 GB (2.90 TB)

7. BENCHMARKS, EVALUATION METRICS AND EXPERIMENTS

7.1. Experimental Evaluation

We evaluated the OneFormer [24] architecture on the CORD validation set, achieving a mean Intersection over Union (mIoU) of 27.56% and a pixel accuracy (pACC) of 64.99%. As shown in Table 5, the high frequency-weighted IoU (59.59%) compared to the mIoU suggests that the model performs significantly better on dominant classes than on rare ones, confirming the severe class imbalance inherent in off-road environments.

Table 5: Quantitative Performance of OneFormer on CORD. Results represent the baseline performance across all semantic classes.

Model	mIoU (%)	fwIoU (%)	mAcc (%)	pAcc (%)
OneFormer (Swin-L)	27.56	59.59	54.46	64.99

7.2. Class-Specific Performance

The model shows exceptional proficiency in segmenting dominant natural features and static structures, with *Trees* (94.64% IoU) and *Fallen Leaves* (85.11% IoU) reaching high scores. Notably, the model achieved a 76.88% IoU for the *Ocean* class, demonstrating its ability to handle the novel coastal environments of Edisto Island.

However, performance drops sharply for rare or geometrically complex objects. Critical navigation hazards such as *Tree Logs* (1.55% IoU) and *Turtle Nests* (3.59% IoU) remain difficult for the network to distinguish. Furthermore, the low score for *Sand* (5.59% IoU)—despite its high frequency in the dataset—indicates a significant ‘texture-blurring’ challenge where the model struggles to differentiate traversable sand from surrounding dirt or gravel trails.

Table 6: Class-wise Intersection over Union (IoU) Results. Representative classes showing the performance variance between dominant and rare categories.

High Performance Classes	IoU (%)	Challenging/Rare Classes	IoU (%)
Ego-Vehicle	98.30	Dune Fence	44.25
Trees	94.64	Debris	35.78
Fallen Leaves/Mulch	85.11	Obstacle	22.71
Ocean	76.88	Turtle Nest	3.59
Asphalt	66.38	Sand	5.59
Person	66.48	Tree Log	1.55

8. CONCLUSIONS

We introduced the Clemson Off-Road Dataset (CORD), a first-of-its-kind, high-fidelity multimodal benchmark engineered to push autonomous perception beyond standard flat-world assumptions. Spanning over 2.90 TB of sensor data, CORD provides a hyper-heterogeneous sensor suite including Hyperspectral imaging, Event-based vision, 3D LiDAR, Radar, and HDR RGB cameras mounted on a high-performance Polaris RZR platform. By capturing diverse topographies from the dense, visibility-limited forests of Camp Daniels to the novel, traction-limited coastal zones of Edisto Island, this dataset introduces unique adversarial conditions such as extreme solar glare, shifting tide lines, and non-rigid terrain physics.

To overcome the bottleneck of unstructured data annotation, we used AI LabelMate, a semi-automated pipeline fusing Vision-Language and Segmentation Foundation Models to generate 6,331 pixel-perfect semantic masks across a specialized off-road ontology.

In future work, we plan to leverage the unique modalities of CORD to develop perception models. Specifically, we aim to explore the fusion of Hyperspectral and Event camera data to improve material classification and maintain robust tracking during day-to-night transitions. Ultimately, CORD serves as a pivotal testbed for bridging the gap between geometric mapping and true semantic-based scene understanding.

References

- [1] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *Int. J. Robot. Res. (IJRR)*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [2] M. Cordts *et al.*, “The cityscapes dataset for semantic urban scene understanding,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 3213–3223.
- [3] H. Caesar *et al.*, “nuScenes: A multimodal dataset for autonomous driving,” in *Proc.*

- IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11621–11631.
- [4] F. Yu *et al.*, “BDD100K: A diverse driving video database with scalable annotation tooling,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 8620–8629.
- [5] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2019, pp. 5000–5007.
- [6] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, “RELLIS-3D Dataset: Data, benchmarks and analysis,” *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2484–2491, April 2021.
- [7] S. Triest *et al.*, “TartanDrive: A large-scale dataset for learning off-road dynamics models,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2022, pp. 2546–2552.
- [8] P. Jiang *et al.*, “GO: The Great Outdoors Multimodal Dataset,” *arXiv preprint arXiv:2501.19274*, Jan 2025.
- [9] A. Valada, G. L. Oliveira, T. Brox, and W. Burgard, “Deep multispectral semantic scene understanding of forested environments using multimodal fusion,” in *Proc. Int. Symp. Experimental Robotics (ISER)*, Springer, 2017, pp. 465–477.
- [10] J. G. Rogers *et al.*, “Yamaha-CMU off-road dataset,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2011.
- [11] P. Papadakis, “Terrain traversability analysis methods for unmanned ground vehicles: A survey,” *Eng. Appl. Artif. Intell.*, vol. 26, no. 4, pp. 1373–1385, 2013.
- [12] A. N. Sivakumar *et al.*, “Learned visual navigation for under-canopy agricultural robots,” in *Proc. Robotics: Science and Systems (RSS)*, 2021.
- [13] Y. Bengio, A. Lecun, and G. Hinton, “Deep learning for AI,” *Commun. ACM*, vol. 64, no. 7, pp. 58–65, 2021.
- [14] T. Cortinhal, G. Tzelepis, and E. E. Aksoy, “SalsaNext: Fast, uncertainty-aware semantic segmentation of LiDAR point clouds,” in *Int. Symp. Visual Computing (ISVC)*, Springer, 2020, pp. 207–222.
- [15] J. Wang *et al.*, “Deep high-resolution representation learning for visual recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.
- [16] J. Philion and S. Fidler, “Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3D,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Springer, 2020, pp. 194–210.
- [17] A. Kirillov *et al.*, “Segment anything,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 4015–4026.
- [18] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *arXiv preprint arXiv:1706.05587*, 2017.
- [19] M. V. Gasparino *et al.*, “Wayfast: Navigation with predictive traversability in the field,” *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10651–10658, 2022.
- [20] D. D. Fan *et al.*, “STEP: Stochastic traversability evaluation and planning for risk-aware off-road navigation,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2021, pp. 8004–8010.
- [21] P. Ghamisi *et al.*, “Advances in hyperspectral image and signal processing: A comprehensive

overview of the state of the art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, 2017.

Phone: +1 8642011814
Email: patil5@clemson.edu

- [22] G. Gallego *et al.*, “Event-based vision: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.
- [23] B. Xiao *et al.*, “Florence-2: Advancing a unified representation for a variety of vision tasks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024, pp. 4818–4829.
- [24] J. Jain *et al.*, “OneFormer: One transformer to rule universal image segmentation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 2989–2998.
- [25] A. Patil, D. Mikulski, J. Mwakalonge, and Y. Jia, “AI LabelMate: A Context-Aware Annotation Agent for Reducing Semantic Fragmentation,” *WCX SAE World Congress Experience*, Detroit, MI, USA, Technical Paper 2026-01-0261, Apr. 14, 2026. doi: 10.4271/2026-01-0261.

9. ACKNOWLEDGMENTS

This work was supported by Clemson University’s Virtual Prototyping of Autonomy Enabled Ground Systems (VIPR-GS), a US Army Center of Excellence for modeling and simulation of ground vehicles, under Cooperative Agreement W56HZV-21-2-0001 with the US Army DEVCOM Ground Vehicle Systems Center (GVSC).

The authors would also like to express their gratitude to Denis Ruganuza, Nana Gyimah, Debbie Indah, Hannah Musau, Jacob Davis, and Awadh Simba for their extensive and invaluable efforts in data collection and labeling.

10. CONTACT INFORMATION

Ashish Patil
PhD Candidate, Automotive Engineering
Clemson University