The GOOSE Dataset for Perception in Unstructured Environments

Peter Mortimer[1], Raphael Hagmanns[2,3], Miguel Granero[2],
Thorsten Luettel[1], Janko Petereit[2] and Hans-Joachim Wuensche[1]
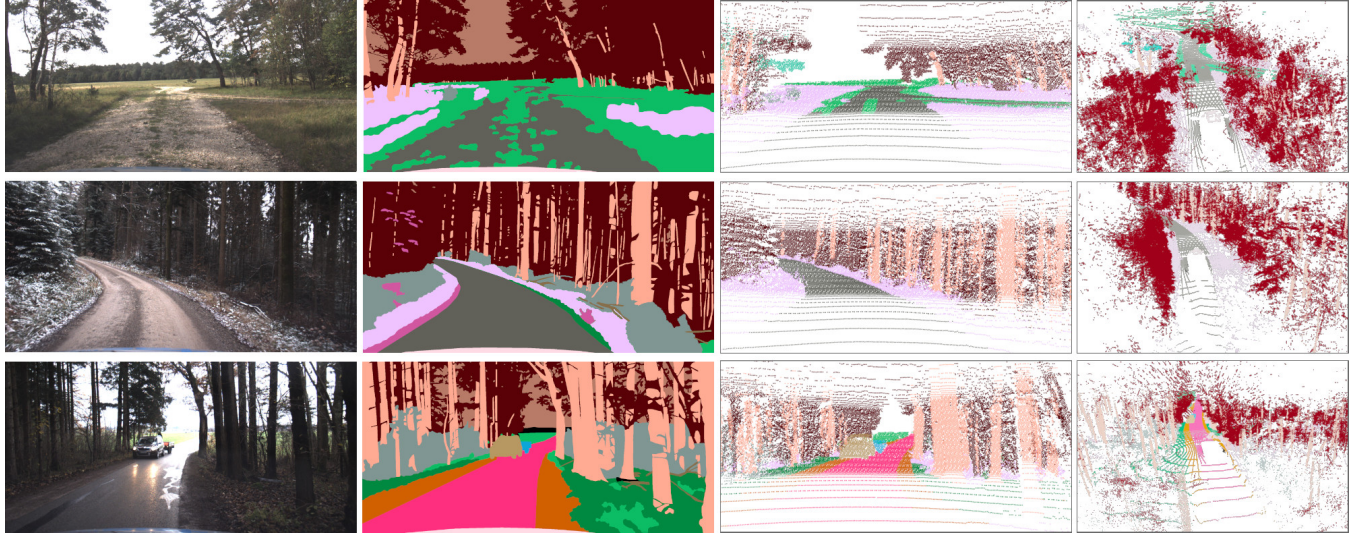
arXiv:2310.16788v2 [cs.CV] 8 Sep 2025

Fig. 1. The GOOSE dataset was recorded over the course of a year and covers all seasons and a wide range of weather conditions. The first column shows color images recorded with the RGB+NIR (near-infrared) camera. The second column is the respective pixel-wise annotated image. The third column shows the annotated point cloud from the 128-channel roof LiDAR system from the same perspective as the camera. The fourth column shows the same point cloud from a perspective outside of the survey vehicle.

Abstract— **The potential for deploying autonomous systems can be significantly increased by improving the perception and interpretation of the environment. However, the development of deep learning-based techniques for autonomous systems in unstructured outdoor environments poses challenges due to limited data availability for training and testing. To address this gap, we present the German Outdoor and Offroad Dataset (GOOSE), a comprehensive dataset specifically designed for unstructured outdoor environments. The GOOSE dataset incorporates 10 000 labeled pairs of images and point clouds, which are utilized to train a range of state-of-the-art segmentation models on both image and point cloud data. We open source the dataset, along with an ontology for unstructured terrain, as well as dataset standards and guidelines. This initiative aims to establish a common framework, enabling the seamless inclusion of existing datasets and a fast way to enhance the perception capabilities of various robots operating in unstructured environments. This framework also makes it possible to query data for specific weather conditions or sensor setups from a database in future. The dataset, pre-trained models for offroad perception, and additional documentation can be found at https://goose-dataset.de/.**

I. INTRODUCTION

Perception in unstructured outdoor environments poses significant challenges for autonomous robots. In offroad environments, mobile robots encounter complex decision-making scenarios that require a fine-grained understanding of their surroundings. One crucial aspect for safe and efficient navigation is the accurate categorization of free space as safely traversable, which requires distinguishing between various obstacle types. Depending on the robotic platform, some regions filled with high grass or bushes might be traversable while this might not be the case for other platforms. In addition, environments can be very different depending on weather conditions or other outer influences. Therefore, most navigation solutions perform poorly in those environments. Many of the current solutions only provide semi-autonomous navigation, as they often rely on teach-and-repeat or similar strategies.

We propose the GOOSE dataset as a novel dataset focused on the perception in unstructured outdoor environments. The size of the GOOSE dataset sets a new standard in annotated 3D point clouds and camera images for unstructured outdoor environments. This will allow the dataset to be a reliable measure of progress by quantitatively evaluating and comparing new methods while simultaneously forming the basis for learning-based methods from other research groups that can use the GOOSE dataset as source for transfer learning on their own robotics platforms.

TABLE I. Comparison of sizes and sensor modalities between existing offroad datasets. In terms of size, the RELLIS-3D provides more annotated laser scans, but fewer annotated images. The individual scans in the GOOSE dataset are denser and were recorded with a 128-channel LiDAR scanner, which is a novelty among driving datasets.

Dataset	Platform	Sensors	Annotated Sensor Modalities	# Annotations	# Classes
OFFSED [1]	ZED	stereo camera	RGB+Depth	203	19
Freiburg Forest [2]	Viona	multi-spectral camera, stereo camera	RGB+NIR+Depth	366	7
TAS500 [3]	MuCAR-3	camera	RGB	640	24
YCOR [4]	ATV	camera	RGB	1 076	8
RUGD [5]	Husky	camera	RGB	7 546	24
RELLIS-3D [6]	Warthog	stereo camera / LiDAR / INS	RGB+Depth / Point Cloud	6 235 / 13 556	20
GOOSE (**ours**)	MuCAR-3	prism camera / LiDAR / INS	RGB+NIR / Point Cloud	10 000 / 10 000	64

This paper presents several key contributions aimed at advancing the perception for unstructured environments:

- We introduce the GOOSE dataset, which comprises of 10 000 pixel-wise annotated pairs of RGB images and LiDAR point clouds, offering a diverse and representative collection of data. The GOOSE dataset encompasses a wide range of environments and includes occurrences of 64 classes that are commonly encountered in unstructured outdoor settings.
- We provide open-source access to the data, including ROS bags and accompanying tools, making it readily available for robotic research. For that reason, additional sensor data, such as near-infrared (NIR) channels of the cameras, surround views, and a high-precision localization based on RTK-GNSS and IMU measurements, are provided to extend the research scope beyond semantic segmentation in 2D and 3D.
- We evaluate the performance of different state-of-the-art models for both 2D semantic segmentation on the RGB camera images and 3D semantic segmentation on the 3D point clouds from the LiDAR scanner.

II. RELATED WORK

Several datasets have been published for urban environments [7], [8], [9], [10], [11]. Modern object detection and segmentation models have had great success in using these large datasets to improve perception capabilities for the classical autonomous driving application case. Typical sensor modalities of these datasets are RGB camera images or LiDAR point clouds, depending on the concrete application. However, there have been very few attempts to replicate the results in urban environments for navigation in unstructured outdoor environments [1], [2], [3], [4], [5], [6]. There are only a few comparatively small datasets with limited sensor modalities, and there is no established ontology for relevant objects in such environments. Table I gives an overview of publicly available datasets and their main characteristics.

Not considered in the comparison in Table I are datasets that were annotated in unstructured environments for other specific tasks such as offroad freespace detection [12], place recognition [13], learning offroad dynamic models [14] or end-to-end driving [15].

Currently, the RUGD dataset [5] includes the largest number of annotations primarily focused on offroad scenes. It has established an ontology that encompasses a variety of typical offroad environments, but the dataset exclusively consists of RGB images. In contrast, datasets like OFFSED [1] and Freiburg Forest [2] include additional multi-spectral and stereo data, although both have a limited number of annotated frames.

Similar limitations apply to the TAS500 dataset [3], which was predominantly created for evaluation purposes but shares the same MuCAR-3 (Munich Cognitive Autonomous Robot Car) platform as GOOSE. On the other hand, the YCOR dataset [4] offers scenes from four distinct locations, showcasing diverse lighting and weather conditions.

In this context, the RELLIS-3D dataset [6] emerges as the largest contribution, incorporating multiple annotated modalities in a substantial number of frames, suitable for modern deep-learning pipelines. By providing both images and point clouds across a diverse range of scenes, it facilitates the training of advanced semantic segmentation models. Nevertheless, the RELLIS-3D dataset has a number of limitations. First, its ontology doesn't allow for detailed classification of trafficability. For instance, the label *grass* constitutes a significant portion of the dataset in terms of pixels count, yet displays considerable visual variability. Furthermore, point clouds are labeled based solely on RGB camera annotations. Multiple views are fused using a SLAM approach for a higher labeling coverage, but the amount of labeled points is still limited. A RELLIS-3D scan comprises up to 13 056 points, which is not much considering the sensor setup. Lastly, the RELLIS-3D dataset only encompasses scenes from a single environment, resulting in relatively consistent lighting conditions and general class appearances across settings. This is also reflected by the comparatively low number of 20 classes in the RELLIS-3D dataset, considering the high amount of annotations.

With the GOOSE dataset, we have tried to overcome these drawbacks. Thus, our point clouds are pointwise annotated and consist of more than 200 000 points on average. We have also put a focus on diversity in scenes and environments, which is supported by our large 64 class ontology. Furthermore, both image and point clouds are annotated with corresponding instances, which enables a combination of 2D and 3D based perception. We therefore expect that the GOOSE dataset can contribute to deep-learning based navigation in a variety of offroad environments.

III. THE GOOSE DATASET

A. Ontology

The GOOSE dataset uses 64 semantic classes in total to segment the RGB images and LiDAR point clouds. The ontology presented in Table II is inspired by similar autonomous driving datasets such as SemanticKITTI [10], TAS500 [3] and RELLIS-3D [6]. The semantic segmentation of the GOOSE dataset is generally more granular, in particular when segmenting the terrain surface and vegetation types. Many semantic classes can be grouped into broader parent semantic classes, such as {*moss*, *low grass*, *leaves*, *high grass*} → *drivable vegetation*. This composability is motivated by the instance labels of the ATLAS ontology [16], an all-terrain label set for autonomous systems. The fine-granular segmentation also makes the GOOSE dataset compatible for composite datasets for multi-domain semantic segmentation like MSeg [17].

B. Dataset Statistics

The GOOSE dataset consists of pixel-annotated images of a RGB+NIR front-facing camera and an annotated point cloud from the spinning LiDAR system with 128 channels on the roof of MuCAR-3 and two side-facing LiDAR systems with 32 channels. Figure 2 shows a histogram of all annotated pixels in the image data of the GOOSE dataset. Almost 90% of the annotated pixels are in the categories *vegetation*, *terrain* and *sky*. This is consistent with our goal of recording mainly in unstructured outdoor environments. The remaining 10% of the annotated pixels make up different obstacle types like *fence*, *building* and *pole*, as well as dynamic objects like *car*, *person* and *animal*. Figure 4 shows the semantic histogram of the annotated points in the 3D point cloud, which has a very similar relative ordering to the annotated pixels. The GOOSE dataset is very fine-granular with its segmentation into 64 semantic classes. One caveat caused by the fine-granular segmentation in the GOOSE dataset, is that some semantic classes only rarely appear in the dataset like *water* or *rider*.

C. Annotation Process

The RGB image of the RGB+NIR camera was annotated by a human annotator. The annotation consistency was ensured by providing a detailed labeling policy beforehand, which will be co-released on the GOOSE dataset website.

Multiple LiDAR scans from the same sequence are fused in the same coordinate frame. This allows the human annotator to more easily recognize objects in the far distance with a higher point density. The instance IDs of the objects in the annotated RGB image and LiDAR scans of the same timestamp are then matched.

D. Data Format

We provide all raw sensor data from each sequence in the ROS bag format [19]. The annotated sensor data in a given sequence is published as a ROS message. Additionally, the annotated image and point cloud data is available in a format similar to SemanticKITTI [10].

TABLE II. An overview of all 64 classes in the GOOSE dataset. The semantic classes were selected to cover both outdoor and urban driving scenarios. The most common *thing* classes in the dataset are also segmented by instance (✓), allowing the evaluation of panoptic segmentation tasks. The pixelwise and pointwise annotations gain additional context by the annotated *stuff* classes [18].

| category	class	color	instance
vegetation	bush	(128,150,147)	✗
	crops	(0,134,237)	✗
	forest	(90,0,7)	✗
	hedge	(232,211,23)	✗
	high_grass	(238,195,255)	✗
	leaves	(0,137,65)	✗
	low_grass	(12,189,102)	✗
	moss	(180,168,189)	✗
	scenery_vegetation	(69,109,117)	✗
	tree_crown	(0,194,160)	✗
	tree_root	(196,164,132)	✗
	tree_trunk	(255,170,146)	✓
terrain	asphalt	(255,47,128)	✗
	cobble	(255,52,255)	✗
	gravel	(97,97,90)	✗
	snow	(209,87,160)	✗
	soil	(209,97,0)	✗
sky	sky	(183,123,104)	✗
construction	bridge	(160,121,191)	✗
	building	(1,51,73)	✗
	container	(255,138,154)	✓
	debris	(136,111,76)	✗
	fence	(255,181,0)	✗
	guard_rail	(194,255,237)	✗
	tunnel	(204,7,68)	✗
	wall	(0,132,111)	✗
	wire	(255,140,0)	✗
vehicle	bicycle	(0,77,67)	✓
	bus	(153,125,135)	✓
	car	(183,151,98)	✓
	caravan	(48,0,24)	✓
	heavy_machinery	(250,208,159)	✓
	kick_scooter	(111,0,98)	✓
	motorcycle	(79,198,1)	✓
	on_rails	(161,194,153)	✓
	trailer	(10,166,216)	✓
	truck	(123,79,75)	✓
road	bikeway	(163,0,89)	✗
	curb	(74,59,83)	✗
	pedestrian_crossing	(122,73,0)	✗
	rail_track	(107,121,0)	✗
	road_marking	(99,255,172)	✗
	sidewalk	(59,93,255)	✗
object	barrel	(208,208,0)	✓
	obstacle	(255,74,70)	✗
	pipe	(221,0,0)	✗
	pole	(192,185,178)	✓
	rock	(55,33,1)	✓
	street_light	(0,111,166)	✓
void	ego_vehicle	(255,219,229)	✗
	outlier	(120,141,102)	✗
	undefined	(0,0,0)	✗
sign	barrier_tape	(190,196,89)	✗
	boom_barrier	(52,54,45)	✓
	misc_sign	(0,30,9)	✓
	road_block	(0,0,166)	✓
	traffic_cone	(255,255,0)	✗
	traffic_light	(27,68,0)	✓
	traffic_sign	(194,255,153)	✓
human	person	(143,176,255)	✓
	rider	(221,239,255)	✓
water	water	(122,135,161)	✗
animal	animal	(0,0,53)	✓
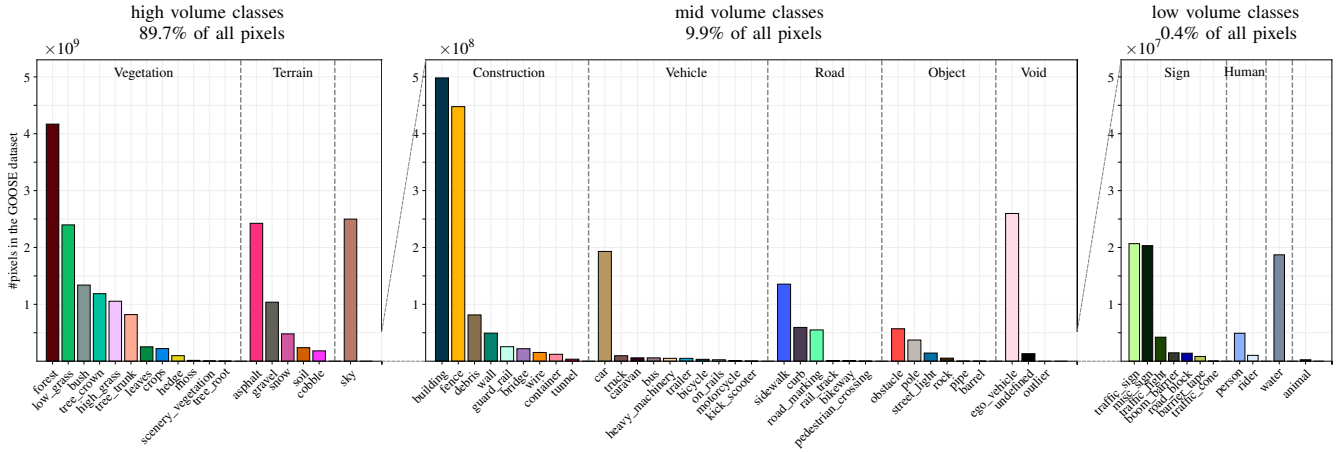
Fig. 2. Best to inspect digitally. Histogram of the annotated pixels in the 2D images of the GOOSE dataset. The 12 categories can be grouped into the high volume classes in the categories {*Vegetation*, *Terrain*, *Sky*}, mid volume classes from the categories {*Construction*, *Vehicle*, *Road*, *Object*, *Void*} and the low volume classes from the categories {*Sign*, *Human*, *Water*, *Animal*}.

IV. ROBOT SETUP

A. Sensor Setup

The GOOSE dataset was recorded on the UniBw Munich research vehicle MuCAR-3 [20]. MuCAR-3 is a modified Volkswagen Touareg with full drive-by-wire capabilities. The sensor setup of MuCAR-3 is illustrated in Figure 3 and includes:

- 1 × Velodyne Alpha Prime (■): 128 channels, 10 Hz, 40° vertical field of view
- 2 × Ouster OS0 (■): 32 channels, 10 Hz, 90° vertical field of view
- 4 × Basler acA2440-20gc (■): RGB camera with a 4 mm 1/1.8" Lensation BM4018S118C lens, 10 Hz, 94° horizontal field of view
- 1 × FLIR A615 (■): thermal infrared camera with a built-in lens, 50 Hz, 45° horizontal field of view
- 1 × JAI FSFE-3200D (■): prism camera with a sensor behind a RGB color filter and a sensor behind a near infrared filter with a 6 mm 1/1.8" Fujinon TF6MA-1 lens, 10 Hz, 59° horizontal field of view
- 1 × Basler acA2440-20gc (■): RGB camera with a 8 mm 1/1.4" Kowa LM8HC mounted on a mobile platform, 10 Hz, 54° horizontal field of view
- 5 × Smartmicro UMRR-96 Type 153 (■): 79 GHz band, set in medium-range mode with a detection range of 0.4 m-55 m
- 1 × Smartmicro UMRR-11 Type 132 (■): 77 GHz band, long-range mode with a detection range of 1 m-175 m
- 1 × Oxford RT3000v3 (■): Inertial Navigation System (INS) with differential RTK-GNSS corrections received over LTE using the NTRIP protocol

The low level vehicle control and mobile camera platform control is handled by two real-time capable dSPACE computers. The main computing unit for sensor data accumulation and recording is equipped with a Titan RTX GPU and an AMD Ryzen Threadripper PRO 3975WX CPU.
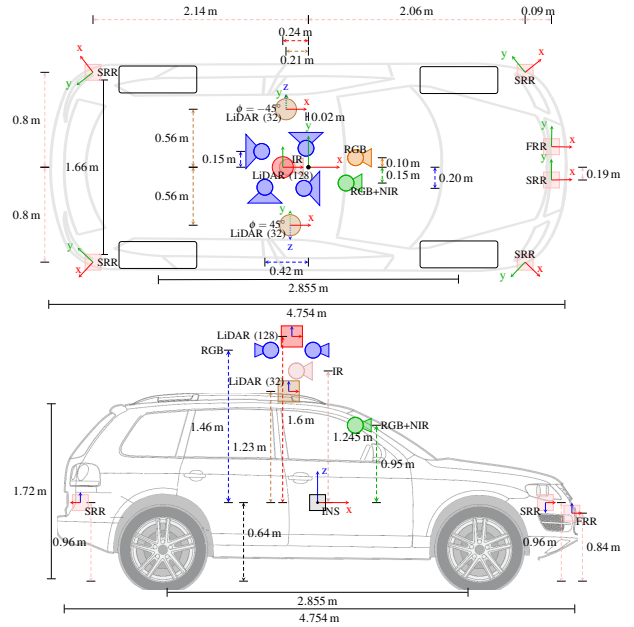

Fig. 3. Schematic of the sensor setup on MuCAR-3. MuCAR-3 is a modified Volkswagen Touareg with full drive-by-wire capabilities. The base_link coordinate frame is situated between both axles at the vehicle's center of gravity and coincides with the **INS** (■) position. The **RGB+NIR** (■) camera image is annotated as part of the GOOSE dataset and is mounted behind the windshield. The point cloud of the **128-beam LiDAR** (■) scanner on the roof of MuCAR-3 is also annotated. The two **32-beam LiDAR** (■) scanners are tilted at an ±45° angle to detect close by obstacles lateral to MuCAR-3. These points are also annotated for the same timestamp in the GOOSE dataset. The four **RGB** (■) color cameras on the roof provide a 360° view of the surrounding. A front-facing **thermal infrared (IR)** (■) camera provides an additional modality for low-light conditions. Six **radar** (■) sensors are mounted around the vehicle to provide 360 degree radar detections with only small blind spots on the sides. An additional **RGB** (■) color camera mounted on a mobile platform [21] is positioned behind the windshield and pans the camera when turning into corners or following objects of interest.
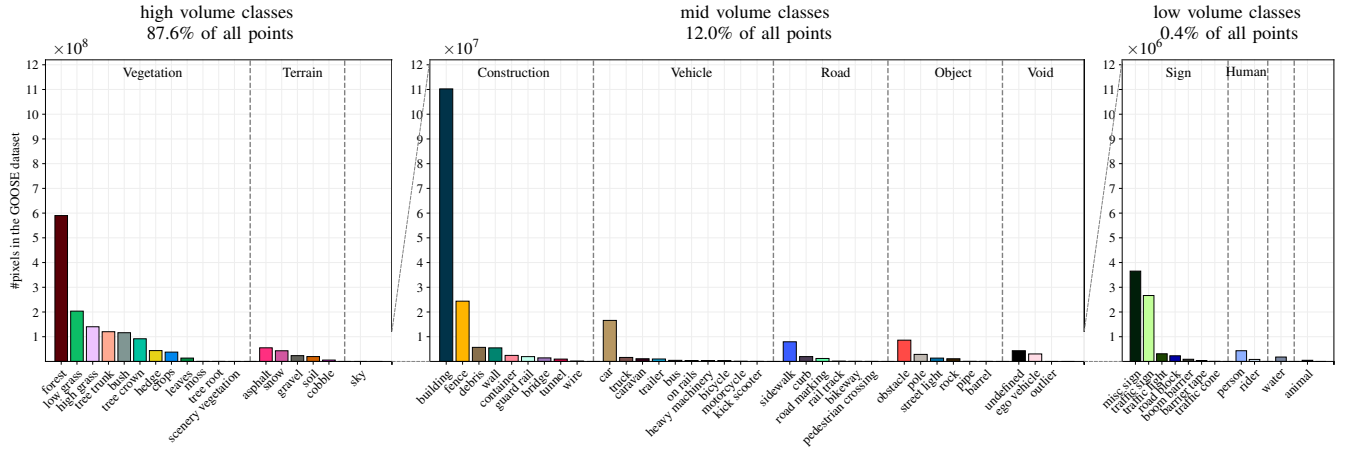
Fig. 4. Best to inspect digitally. Histogram of the annotated points in the 3D point clouds of the GOOSE dataset. We observe a relative order of the semantic classes similar to that of the annotated images in the GOOSE data set (see Figure 2). The elevated position of the 128-channel LiDAR scanner leads to many detections of building roofs, which explains the high occurrence of annotated *building* points among the mid volume classes.

B. Synchronization

The sensor clocks of the cameras, LiDAR scanners and INS system are synchronized using the Precision Time Protocol (PTP, IEEE 1588 [22]). In essence, PTP synchronizes clocks across a sensor network by determining a leader clock in a local network and synchronizing the follower clocks to the leader clock while also taking the message delay into consideration.

We predetermined our INS system as leader clock. The INS obtains its time at 100 Hz over GNSS. The color cameras from Basler and JAI, the main computer's system clock, as well as the side-facing Ouster OS0 LiDAR sensors support the synchronization of their clocks using PTP. This synchronization also allows for triggering the cameras with respect to the current LiDAR orientation

C. Calibration

The intrinsic parameters of each camera are determined by calibrating them on a mosaic of AprilTags [23] with checkerboard corners as described in [24]. All cameras are assumed to fit a pinhole projection model. The extrinsic calibration of the sensors is essential for sensor fusion. The relative transformation between the roof LiDAR scanner and color cameras is determined using a calibration target with four circular holes, which can be automatically detected in the LiDAR point cloud [25]. Similar to [26], the calibration target has four additional AprilTags, which can be detected accurately in the color camera images. Detailed calibration procedures will be made available in the dataset documentation.

V. EXPERIMENTAL EVALUATION

For the initial evaluations, we decided to focus on the 2D image segmentation from the RGB camera image (see Subsection V-C) and 3D point cloud segmentation from a single LiDAR scan (see Subsection V-D).

A. Training Split

The GOOSE dataset was recorded over the course of a year. Each recording day is considered a *scenario* consisting of multiple *sequences*. The raw sensor data of a sequence is stored in ROS bag format for later reprocessing. To use the GOOSE dataset as a benchmark, we define a fixed training, validation and test split for supervised learning algorithms. We ensured that each split contains sequences from each season and a wide range of weather conditions, while still sharing a similar distribution among the semantic classes that appear in each split. The training split holds 7830 annotated images and point clouds, the validation split 960 and the test split 1210. The annotations of the test split are withheld from the publication of the GOOSE dataset to allow the setup of a public benchmarking platform in the future.

B. Evaluation Metrics

The standard metric for evaluating the semantic segmentation on both images and point clouds is the mean Intersection over Union (mIoU), also known as the mean Jaccard index. The IoU is widely used for evaluating the segmentation of a semantic class by comparing the prediction region with the ground truth region. It can be described using the result cases true positive (TP), false negative (FN) and false positive (FP) of a binary classification in the following way:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}}. \qquad (1)$$

The IoU is determined by accumulating TP, FP and FN of the whole test set. The mIoU is then calculated by averaging the IoU over all classes in the dataset. The IoU is known to be biased towards object instances and classes that cover large areas of the image [8]. In Table III, we provide averaged IoU values for models which were trained on the full set of classes *(class)* as well as IoU values for models trained on the broader category labels *(category)*.

TABLE III. Comparison of the 2D image segmentation and 3D point cloud segmentation performance on the GOOSE test set. The IoU scores are specified in percent. For class-based evaluation, classes with occurences less than 20 are omitted. No classes of the category *Sky* exist for the 3D point cloud segmentation.

	network	type	**mIoU**	Vegetation	Terrain	Vehicle	Object	Constr.	Road	Sign	Human	Sky	Water
2D	PP-LiteSeg [27]	*category*	**67.21**	92.43	90.62	87.83	46.86	78.81	51.23	58.19	60.53	96.28	9.29
		class	**45.09**	47.69	70.97	37.69	29.15	43.96	49.71	33.48	34.07	96.11	8.23
	DDRNet [28]	*category*	**70.23**	92.94	91.34	91.05	52.66	81.67	59.27	62.90	64.65	96.55	9.23
		class	**46.53**	49.54	72.20	31.22	35.63	44.52	54.50	37.97	34.38	96.24	9.09
	Mask2Former [29]	*category*	**64.26**	91.76	91.34	91.05	42.57	75.07	44.26	43.22	61.72	96.49	5.10
3D	PVKD [30], [31]	*class*	**34.32**	47.40	28.69	40.82	20.61	37.94	36.21	29.89	57.94	-	9.42
	SPVNAS [32]	*class*	**17.32**	41.08	19.27	14.58	21.29	21.27	6.38	12.94	19.08	-	0.0

C. Image Segmentation

Our selection of image semantic segmentation algorithms was determined by comparing the state-of-the-art methods on other public benchmarks. Since the GOOSE dataset will primarily be used in a outdoor robotics context, inference speed is also an important criterion for method selection. DDRNet [28] uses a typical two-stream architecture originated in BiSeNet [33], but fuses both branches at different depths in the network. PPLiteSeg [27] uses an encoder-decoder structure with a lightweight attention-based fusion model in the decoder to enable realtime semantic segmentation. Mask2former [29] is a recent, universal, transformer-based model that can be conditioned at both semantic and instance segmentation.

From the results in Table III we observe a solid image segmentation quality in general. Most categories with many appearances have IoU values over 90% (*Vegetation*, *Terrain*, *Vehicle*, *Sky*). However, we also encounter some categories with very small IoU values, leading to a reduced value (10-20%) for the overall mIoU compared to the same algorithm applied on the Cityscapes [8] dataset. The main reason for this is the natural imbalance in the class distribution within our dataset. Limited training data availability as well as similar appearances of different classes lead to difficulties in recognizing those classes correctly. The comparatively low class-based IoU values in Table III show that performing fine-graded segmentation is difficult in environments, where the transitions between regions are not always sharp. These effects might be mitigated by applying typical class balancing strategies, such as weighted loss functions or specialized fine-graded architectures, which will be further explored in future research. Although classes with only a few labels may have a negative impact on the mIoU score, their existence is highly relevant for practical navigation solutions since misclassified regions such as *water* may also lead to wrong classification of trafficability.

D. Point Cloud Segmentation

For the evaluation of the annotated 3D point cloud data, we focus on methods that produce an annotated point cloud instead of a voxel grid or a birds-eye-view segmentation of the surroundings. Point-Voxel-KD [30] (PVKD) trains a slim student network by mimicking a larger teacher model both in its voxel and point representation. The student and teacher architectures are based on Cylinder3D [31]. SPVNAS [32] propose the Sparse Point-Voxel Convolution (SPVConv) block that makes use of both a point and a voxel-based representation, while also searching for an efficient network architecture of SPVConv blocks through neural architecture search (NAS).

In the lower part of Table III, we compare the 3D segmentation performance of Point-Voxel-KD and SPVNAS on the GOOSE dataset. As input, both models use the 3D position and the reflection intensity of each point in the point cloud. For both models, we observe low mIoU scores for rare categories like *Human*, *Sky* and *Water*. Categories like *Terrain* contain many classes that are geometrically similar and differ mainly in their surface texture like *gravel*, *soil* and *snow*. In these categories, we also observe that both models perform poorly compared to the image segmentation results for the same category.

VI. CONCLUSION

We introduce the GOOSE dataset, an open-source, multimodal dataset for perception in unstructured outdoor environments. As our benchmarks show, GOOSE dataset provides challenges in terms of imbalanced class distributions across a diverse set of outdoor environments. In our future work, we want to investigate additional semantic segmentation models, particularly those that can utilize different modalities. Additionally, we will expand the GOOSE-dataset by adding data from additional platforms in different environments to further enhance the adaptability of segmentation techniques for mobile robot navigation.

REFERENCES

[1] P. Neigel, J. Rambach, and D. Stricker, "OFFSED: Off-Road Semantic Segmentation Dataset," in *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP)*, 2021. 2

[2] A. Valada, G. Oliveira, T. Brox, and W. Burgard, "Deep Multispectral Semantic Scene Understanding of Forested Environments using Multi-modal Fusion," in *International Symposium on Experimental Robotics (ISER)*, 2016. 2

[3] K. A. Metzger, P. Mortimer, and H.-J. Wuensche, "A Fine-Grained Dataset and its Efficient Semantic Segmentation for Unstructured Driving Scenarios," in *International Conference on Pattern Recognition (ICPR)*, 2021. 2, 3

[4] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-Time Semantic Mapping for Autonomous Off-Road Navigation," in *Field and Service Robotics*. Springer, 2018. 2

[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 *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, China, 2019. 2

[6] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "RELLIS-3D Dataset: Data, Benchmarks and Analysis," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, May 2021. 2, 3

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, 2013. 2

[8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016. 2, 5, 6

[9] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A Multimodal Dataset for Autonomous Driving," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[10] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019. 2, 3

[11] O. Zendel, M. Schörghuber, B. Rainer, M. Murschitz, and C. Beleznai, "Unifying Panoptic Segmentation for Autonomous Driving," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[12] 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 *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2532–2538. 2

[13] J. Knights, K. Vidanapathirana, M. Ramezani, S. Sridharan, C. Fookes, and P. Moghadam, "Wild-Places: A Large-Scale Dataset for Lidar Place Recognition in Unstructured Natural Environments," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2023. 2

[14] S. Triest, M. Sivaprakasam, S. J. Wang, W. Wang, A. M. Johnson, and S. Scherer, "TartanDrive: A Large-Scale Dataset for Learning Off-Road Dynamics Models," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2022. 2

[15] A. Tampuu, R. Aidla, J. A. van Gent, and T. Matiisen, "LiDAR-as-Camera for End-to-End Driving," *Sensors*, vol. 23, no. 5, 2023. 2

[16] W. Smith, D. Grabowsky, and D. Mikulski, "ATLAS, an All-Terrain Lalbelset for Autonomous Systems," in *Ground Vehicle Systems Engineering and Technology Symposium*, 2022. 3

[17] J. Lambert, Z. Liu, O. Sener, J. Hays, and V. Koltun, "MSeg: A Composite Dataset for Multi-domain Semantic Segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[18] H. Caesar, J. Uijlings, and V. Ferrari, "COCO-Stuff: Thing and Stuff Classes in Context," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[19] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS: an open-source Robot Operating System," in *Proceedings of IEEE International Conference on Robotics and Automation Workshops (ICRAW)*, 2009. 3

[20] M. Himmelsbach, T. Luettel, F. Hecker, F. v. Hundelshausen, and H.-J. Wuensche, "Autonomous Off-Road Navigation for MuCAR-3," *KI-Künstliche Intelligenz*, vol. 25, no. 2, 2011. 4

[21] A. Unterholzner and H.-J. Wuensche, "Hybrid Adaptive Control of a Multi-Focal Vision System," in *IEEE Transactions on Intelligent Vehicles*, 2010. 4

[22] "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," *IEEE Std 1588-2019 (Revision of IEEE Std 1588-2008)*, pp. 1–499, 2020. 5

[23] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016. 5

[24] J. Kallwies, B. Forkel, and H.-J. Wuensche, "Determining and Improving the Localization Accuracy of AprilTag Detection," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2020. 5

[25] C. Guindel, J. Beltrán, D. Martín, and F. García, "Automatic Extrinsic Calibration for Lidar-Stereo Vehicle Sensor Setups," in *Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC)*, 2017. 5

[26] J. Beltrán, C. Guindel, A. de la Escalera, and F. García, "Automatic Extrinsic Calibration Method for LiDAR and Camera Sensor Setups," *IEEE Transactions on Intelligent Transportation Systems*, 2022. 5

[27] J. Peng, Y. Liu, S. Tang, Y. Hao, L. Chu, G. Chen, Z. Wu, Z. Chen, Z. Yu, Y. Du, Q. Dang, B. Lai, Q. Liu, X. Hu, D. Yu, and Y. Ma, "PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model," 2022. 6

[28] H. Pan, Y. Hong, W. Sun, and Y. Jia, "Deep Dual-Resolution Networks for Real-Time and Accurate Semantic Segmentation of Traffic Scenes," *IEEE Transactions on Intelligent Transportation Systems*, 2022. 6

[29] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," 2022. 6

[30] Y. Hou, X. Zhu, Y. Ma, C. C. Loy, and Y. Li, "Point-to-Voxel Knowledge Distillation for LiDAR Semantic Segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 6

[31] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, and D. Lin, "Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 6

[32] H. Tang, Z. Liu, S. Zhao, Y. Lin, J. Lin, H. Wang, and S. Han, "Searching Efficient 3D Architectures with Sparse Point-Voxel Convolution," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2020. 6

[33] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018. 6