2025 IEEE International Conference on Robotics and Automation (ICRA)
May 19-23, 2025. Atlanta, USA

2025 IEEE International Conference on Robotics and Automation (ICRA) | 979-8-3315-4139-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICRA55743.2025.11127604

Excavating in the Wild: The GOOSE-Ex Dataset for Semantic Segmentation

Raphael Hagmanns[1,3], Peter Mortimer[2], Miguel Granero[1], Thorsten Luettel[2] and Janko Petereit[1]

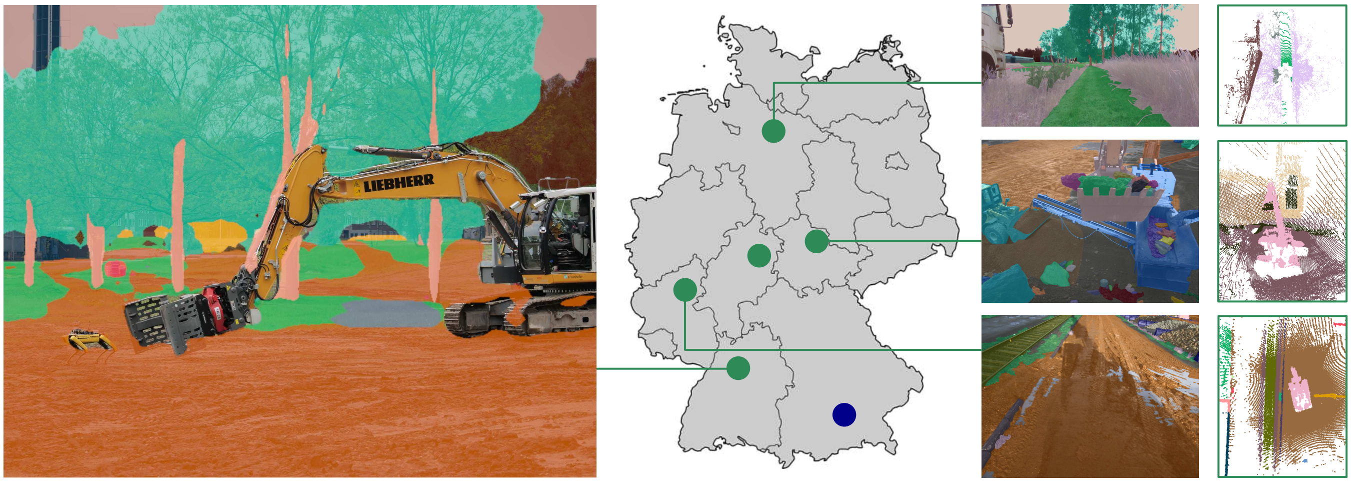


Fig 1. The GOOSE-Ex dataset was recorded over the course of a year in various locations across Germany, covering a wide range of environmental conditions. The left image shows a smartphone image of the two recording platforms, which was semantically segmented using a model trained on a 64-class version of the GOOSE-Ex dataset to demonstrate the platform generalizability of the dataset. The masks of the classes *heavy_machinery* and *obstacle* were removed to better highlight the platforms. The right part of the figure shows exemplary ground truth annotations from different recording locations. This includes both pixel annotated images and annotated point clouds.

***Abstract*— The successful deployment of deep learning-based techniques for autonomous systems is highly dependent on the data availability for the respective system in its deployment environment. Especially for unstructured outdoor environments, very few datasets exist for even fewer robotic platforms and scenarios. In an earlier work, we presented the German Outdoor and Offroad Dataset (GOOSE) framework along with 10 000 multimodal frames from an offroad vehicle to enhance the perception capabilities in unstructured environments. In this work, we address the generalizability of the GOOSE framework. To accomplish this, we open-source the GOOSE-Ex dataset, which contains additional 5 000 labeled multimodal frames from various completely different environments, recorded on a robotic excavator and a quadruped platform. We perform a comprehensive analysis of the semantic segmentation performance on different platforms and sensor modalities in unseen environments. In addition, we demonstrate how the combined datasets can be utilized for different downstream applications or competitions such as offroad navigation, object manipulation or scene completion. The dataset, its platform documentation and pre-trained state-of-the-art models for offroad perception will be made available on https://goose-dataset.de/.**

The perception of unstructured outdoor environments presents significant challenges for autonomous systems, particularly in the domains of free space detection and obstacle avoidance, as well as for manipulation tasks. Attaining complete autonomy in these settings is challenging due to the inherent variability of environmental conditions and terrain types. In recent years, some effort has been made to adapt advanced semantic segmentation models from structured to unstructured environments. However, the adaptation of these models to previously unseen environments and new platforms is a challenging task due to the limited data availability. A particular challenge arises from the platform gap resulting from the platform-specific mounting of cameras and LiDAR sensors, which complicates the transfer-learning to different systems. The GOOSE framework presented in [1] offers a robust basis for segmentation tasks, yet it is constrained to a single platform and a relatively small region. Compared to the original GOOSE dataset, the main objective of the proposed GOOSE-Ex dataset is to facilitate adaptation to heterogeneous platform settings in specialized environments. Platform variations include an excavator setup as well as two quadruped robot setups with varying sensors. The unique excavator setup and its target domains allow for the exploitation and fine-tuning of perception algorithms for unusual and large-scale robotic platforms.

This paper presents a series of contributions designed to enhance the perception of various robots in diverse unstructured environments.

This work was supported by the Federal Office of Bundeswehr Equipment, Information Technology and In-Service Support (BAAINBw).

[1] The authors are with the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Karlsruhe, Germany firstname.lastname@iosb.fraunhofer.de

[2] The authors are with the Institute for Autonomous Systems Technology, University of the Bundeswehr Munich, Germany firstname.lastname@unibw.de

[3] The author is with the Karlsruhe Institute of Technology (KIT), Germany firstname.lastname@kit.edu

TABLE I: Comparison of sizes and sensor modalities between existing offroad datasets. In terms of size, the RELLIS-3D and GOOSE provides more annotated laser scans, but fewer annotated images. The CWT dataset also contains annotated sensor data from an autonomous excavator, but is notably smaller than GOOSE-Ex. For the parts of GOOSE-Ex recorded with a quadrupedal robot platform, the RUGD and RELLIS-3D are the most similar datasets in terms of the camera and LiDAR sensor height above ground.

Dataset	Platform	Sensors	Annotated Sensor Modalities	# Annotations	# Classes
CWT [2]	Excavator	camera	RGB	669	7
RUGD [3]	Husky	camera	RGB	7 546	24
RELLIS-3D [4]	Warthog	stereo camera / LiDAR / INS	RGB+Depth / Point Cloud	6 235 / 13 556	20
GOOSE [1]	MuCAR-3	prism camera / LiDAR / INS	RGB+NIR / Point Cloud	10 000 / 10 000	64
GOOSE-Ex (**ours**)	Excavator, Spot	prism camera / LiDAR / INS	RGB+(NIR) / Point Cloud	5 000 / 5 000	64

- We present the GOOSE-Ex dataset, which consists of 5 000 calibrated pairs of pixel-wise annotated RGB images and point-wise annotated LiDAR point clouds from a robotic excavator and a quadruped platform. The dataset encompasses over 100 sequences from diverse environments, employing the same format and class hierarchy established in the GOOSE framework [1].
- We open-source the dataset and accompanying tools to enable rapid prototyping. We also provide additional sensor data, such as near-infrared (NIR) channels of many camera frames, surround views, and a high-precision localization.
- We evaluate the performance of various state-of-the-art models for semantic segmentation across different dataset combinations and sensor modalities.
- To the best of our knowledge, GOOSE-Ex is the first multimodal and large-scale semantic segmentation dataset for excavator platforms. Together with the additional excavator-specific sensor data recorded, this can accelerate the progress in a variety of downstream applications, some of which we showcase in Section V.

I. Related Work

The release of datasets with dense semantic and instance-wise annotations of pixels in color images [5–8] and 3D points in LiDAR scans [9–11] have led to ever-improving segmentation models for perception in autonomous driving in urban environments. In recent years, there have been attempts to replicate the results for navigation in unstructured outdoor environments [1, 3, 4, 12–17].

Table I gives an overview of semantic segmentation datasets similar to GOOSE-Ex and their main characteristics. Not included in the comparison in Table I are datasets that have been annotated in unstructured environments for other specific tasks such as offroad free space detection [17–20], place recognition [21], learning offroad dynamic models [22–24] or end-to-end driving [25].

Currently, the recent GOOSE dataset [1] and the RUGD dataset [3] include the largest number of annotated images primarily focused on offroad scenes. The high reflectivity of foliage in the near-infrared (NIR) spectrum [26] motivated the inclusion of this image modailty in datasets like TAS-NIR [27], Freiburg Forest [12] and GOOSE [1].

Many of the early datasets in this domain like the OFFSED dataset [15], the TAS500 dataset [13] and the YCOR dataset [14] lack the size and variety to train deep neural networks that can generalize to a different robot platform.

Among the 3D point cloud datasets, the RELLIS-3D dataset [4] and the GOOSE dataset [1] contain fused LiDAR point cloud data of each annotated scene.

Semantic segmentation datasets have extended on existing datasets before. IDD dataset [28] extended the semantic segmentation schema used in the CityScapes dataset [6] to novel object classes and novel driving scenarios. In a similar vein, datasets were extended with adverse weather and lighting conditions [29–33]. GOOSE-Ex uses the same semantic segmentation scheme as the GOOSE dataset, but extends both the acquisition platforms and domains beyond those in GOOSE by including annotated data from an autonomous excavator and a quadruped robot platform.

A growing number of autonomous excavators [2, 34] exist, but most methods focus on planning excavation tasks [35–37]. The recent CWT dataset [38] consists of 669 images annotated for objects and relevant terrain types (see Table I).

Previous datasets recorded on quadrupedal robots have focused on robot navigation [39, 40], odometry [41], mapping [42] and 3D pose estimation [43]. Most image data from quadrupedal robots is unlabeled [44], making RUGD [3] and RELLIS-3D [4] the most similar semantic segmentation datasets in terms of the sensor height above ground [45]. For both platform types, GOOSE-Ex provides a novel contribution by providing semantically segmented camera and LiDAR data that allows for robustness and fine-tuning of the semantic segmentation models.

II. The GOOSE-Ex Dataset

We summarize the main aspects of the GOOSE framework in Section II-A, which includes the organization of the dataset, its structure, ontology, metadata, labeling policy, and more. For additional details, we refer to our previous work [1], where we published these definitions along with the original GOOSE dataset. In the remaining sections, we discuss the GOOSE-Ex dataset in detail.

A. GOOSE Framework

Annotation of RGB images and LiDAR point clouds allows for 64 classes, enabling fine-grained segmentation tasks, especially for the traversability analysis across different vegetation and terrain types. It also enables fine-grained

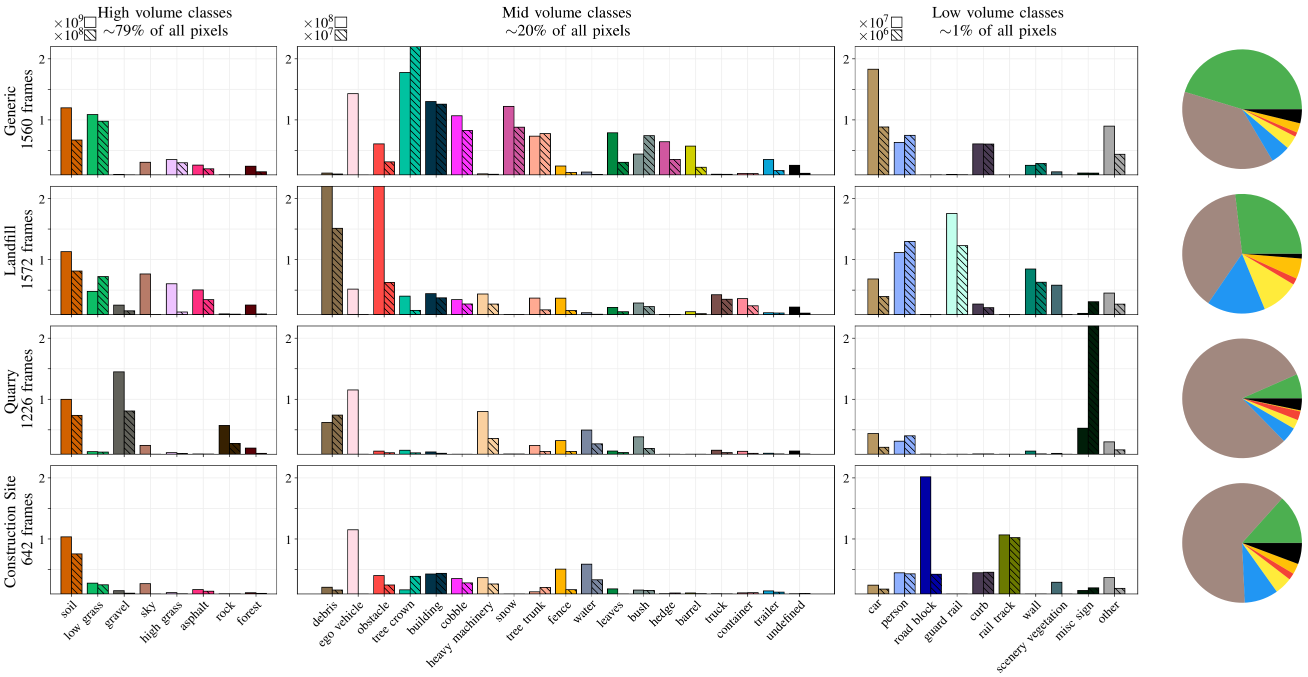


Fig. 2: Best to inspect digitally. Histogram of the annotated pixels □ and points ▨ in the GOOSE-Ex dataset. The classes are split according to their pixel volume into high-, mid- and low-volume classes. We omitted the *ego-vehicle* class in the point clouds and summarized classes with negligible occurrences in the *other* bar. The pie charts show the category distribution from the main categories *Vegetation* 🟩, *Terrain* 🟫, *Sky* 🟦, *Construction* 🟨, *Vehicle* 🟥, *Road* ◼, *Object* 🟧, *Void* ⬛. The remaining categories *Sign*, *Human*, *Water*, *Animal* are summarized as *Other* ◼.

downstream applications such as the barrel detection presented in Section V. For more general applications, we also suggest a rough division into categories, e.g. by aggregating all vegetation classes. The GOOSE ontology is inspired by different datasets and ontologies such as SemanticKITTI [9], TAS500 [13], ATLAS [46], and RELLIS-3D [4] and designed to be as compatible and extendable as possible.

The GOOSE-Ex dataset was manually labeled, merging multiple frames based on platform odometry to enhance the annotation quality. To achieve consistency across all datasets in the GOOSE framework, the same labeling policy was used for both the original GOOSE and GOOSE-Ex datasets. The raw data's hierarchical structure allows easy filtering of environmental conditions or platform configurations. Labeled frames are also provided as a standalone package in a format similar to the SemanticKITTI [9] dataset.

B. Places

Figure 1 shows the different recording areas of the GOOSE (blue) and GOOSE-Ex (green) datasets. We roughly divide the GOOSE-Ex dataset into four different high-level settings:

generic setting, containing a mixture of typical offroad and industrial regions

landfill setting, containing frames from inside and around a landfill as a typical excavator operating environment

quarry setting, as special operating environment for large machines, with complex surface geometries

construction site setting, including an excavator training area with many different heavy machines

This subdivision allows to use specific parts of the dataset to fine-tuning different operational scenarios.

C. Dataset Statistics

The distribution of classes and categories is shown in Figure 2. For unstructured outdoor environments, the *vegetation* and *terrain* categories naturally make up the majority of the dataset. Typical for excavator scenarios, *soil* is a dominant class in the distribution across all environments. The partitioning of the histogram into different settings reveals plausible effects: The *generic* environment contains the most vegetation and many different mid-volume classes. Due to steep slopes in the terrain, the landfill setting contains more *sky* than the others. The piles of trash at the landfill also account for the increase of *debris* and *obstacles*. *Gravel* and *rock* are naturally the most frequent classes in the quarry setting. Other classes are very rare in the quarry environment, except *heavy machinery* due to the large size of those vehicles in quarry environments. Finally, the construction site setting contains an equal appearance of classes with some outliers such as *road blocks* and *fences*.

Of the 64 classes in the GOOSE ontology, only 36 are present in the histogram, all remaining occurrences are summarized in the *other* bars. This illustrates the general problem of class imbalance in natural environments. However, we believe that fine-grained annotations can only be advantageous as they allow to solve fine-grained object recognition and segmentation tasks. If fine granularity is not important for a task, one can refer to the coarser division into categories.

III. Platform Setups

To enhance the transfer learning possibilities onto unique platforms, the GOOSE-Ex dataset was recorded on two robotic platforms as illustrated in Figure 3.

A. Sensor Setup

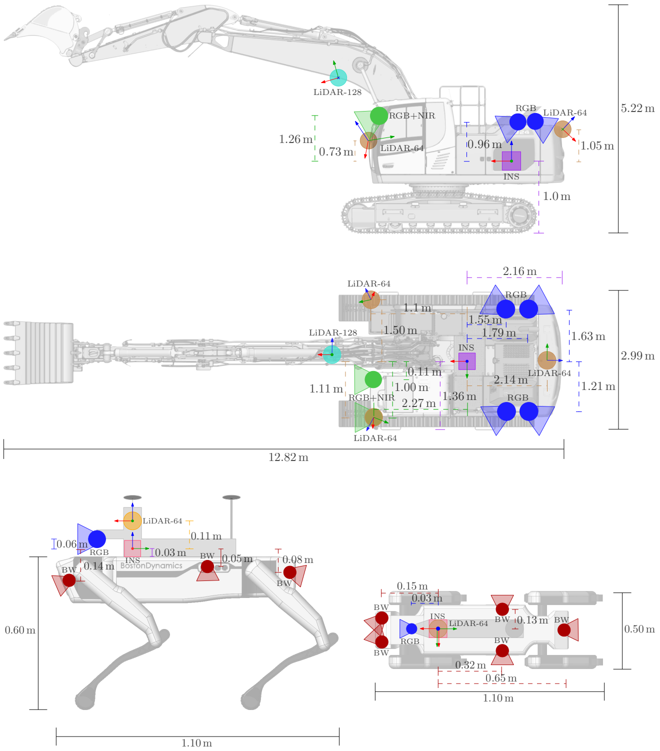


Fig. 3: Schematic of the sensor setups on the Fraunhofer IOSB research excavator *ALICE* and quadruped robot *SpotLow*. All measurements are relative to the IMU frame.

The main recording platform is the Liebherr R924 track excavator *ALICE* [34] with drive-by-wire capabilities and many modifications such as sensors for precise angle, force and track odometry measurements. The main sensors we use for the dataset collection include:

- 3 × Ouster OS0-64 Rev. 6 (●) [47]
- 1 × Ouster OS0-128 Rev. 6 (●), mounted at boom
- 2 × JAI FSFE-3200D (●): global shutter prism camera with two 3.2MP sensors for both RGB and NIR (near-infrared) equipped with a 6 mm 1/1.8" Fujinon TF6MA-1 lens, 5 Hz, 59° hFoV
- 4 × Alvium G1-240C (●): global shutter camera with 2.4MP resolution RGB sensor equipped with a Lensagon B5M3428S123C lens, 5 Hz, 110° hFoV
- SBG Ekinox-D (●): Inertial Navigation System (INS) with differential RTK-GNSS corrections received over a base-station if available or over LTE using NTRIP

We combine point clouds from all LiDARs into a single point cloud that we use for subsequent experiments.

As second platform, we equipped a Boston Dynamics Spot robot with two different sensor setups to further increase the platform generalizability. The setup includes computing hardware as well as a differential SBG Ellipse-D RTK-INS (●) for precise localization. The base platform already includes six low-resolution surround cameras (●), which we utilize for the point cloud labeling and include in the raw data. The remaining setup differs in the following way:

SpotLow equipped with

- Ouster OS0-64 Rev. C (●)
- Intel$^{©}$ RealSense$^{\text{TM}}$ LiDAR Camera L515 (●), 5 Hz, 70° hFoV, rolling shutter

SpotHigh equipped with

- Ouster OS0-128 Rev. 7 (●)
- Alvium G1-240C (●) (see above)

B. Synchronization

We utilize the Precision Time Protocol (PTP, IEEE 1588 [48]) support of the sensors to synchronize the Ouster LiDARs with the cameras and the INS system. Only the Realsense$^{\text{TM}}$ L515 camera of the *SpotLow* does not support PTP, so we use the system clock with some exposure offset instead. During post-processing, we leverage the ROS inbuilt *approximate time synchronizing* mechanism to match point clouds and camera images. The INS system provides the *grandmaster* clock in the sensor network, receiving its time stamps with 200 Hz via GNSS.

C. Calibration

We leverage a custom calibration suite to determine both intrinsic and extrinsic parameters of all cameras and LiDARs. For intrinsic and stereo calibration, we assume the pinhole projection model and use a standard checkerboard calibration procedure within our suite. For extrinsic calibration, we build on [49] and [50], and use a calibration target with Apriltags and three circular holes to allow for target matching in both the point cloud and the camera image. One of our LiDAR scanners on the excavator moves along with the boom, so the extrinsic transformation would change as the excavator moves. We therefore merge all point clouds from different scanners with respect to the INS frame, resulting in a single consistent transformation that can be used for reprojection.

IV. Experimental Evaluation

A. Training Split

Similar to [1], we divided the dataset into *scenarios* consisting of multiple *sequences*, one per recorded rosbag. We select sequences from different scenarios to define training (3989 frames), validation (407 frames) and test (604 frames) splits. We withhold the label files for the small test split to include it in a public benchmark of all GOOSE datasets. Of the 5000 total frames, 2800 are from the excavator, the remaining 2200 from the Spot robot.

B. Evaluation Metrics

The standard metric for evaluating the semantic segmentation on both images and point clouds is the mean Intersection over Union (mIoU). For each class i, it compares the prediction region with the ground truth region, resulting in

$$\text{IoU}_i = \frac{\sum_I \sum_{x,y} \mathbb{1}(P(x,y) == i \wedge GT(x,y) == i)}{\sum_I \sum_{x,y} \mathbb{1}(P(x,y) == i \vee GT(x,y) == i)} \quad (1)$$

TABLE II: Comparison of the 2D image segmentation and 3D point cloud segmentation performance on the GOOSE-Ex 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. The models were trained using the GOOSE dataset as a base and fine-tuned on the GOOSE-Ex dataset. *Category* models are trained directly on category labels, whereas *class* IoU values are calculated per class and averaged afterwards.

	network	type	**mIoU↑**	Vegetation	Terrain	Vehicle	Object	Constr.	Road	Sign	Human	Sky
2D	PP-LiteSeg [51]	*category*	**63.60**	85.62	85.84	64.10	44.83	72.08	53.84	1.19	67.21	97.67
		class	**43.83**	26.83	65.50	27.53	18.78	65.63	24.00	3.79	65.47	96.99
	DDRNet [52]	*category*	**62.03**	85.03	85.85	59.81	45.09	67.18	54.02	1.73	61.79	97.76
		class	**47.28**	42.25	70.22	27.81	26.60	66.65	31.74	2.15	61.28	96.77
3D	PTv3 [53, 54]	*category*	**63.83**	73.96	34.12	28.68	59.21	76.10	70.87	85.79	81.94	-
		class	**29.27**	34.00	23.39	32.54	29.15	49.31	23.34	40.99	30.74	-
	MSeg3D [55]	*category*	**36.26**	51.55	80.71	42.53	29.81	19.26	14.11	32.33	19.78	-
		class	**20.87**	27.52	27.78	30.67	1.61	41.99	40.38	17.86	0.00	-

TABLE III: Fine-tuning performance of the GOOSE-Ex dataset. The number in brackets displays the IoU delta between the models trained only on GOOSE and after the fine-tuning.

	network	split	**class mIoU↑**	**category mIoU↑**
2D	PP-LiteSeg [51]	All	43.83 (**+28.71**)	63.60 (**+24.91**)
		Alice	26.89 (**+24.75**)	47.89 (**+22.31**)
		Spot	47.82 (**+30.58**)	64.71 (**+20.58**)
	DDRNet [52]	All	47.28 (**+39.79**)	62.03 (**+25.24**)
		Alice	28.29 (**+25.39**)	42.53 (**+18.72**)
		Spot	49.63 (**+39.56**)	65.08 (**+22.11**)
3D	PTv3 [53, 54]	All	29.27 (**+14.26**)	63.83 (**+32.96**)
		Alice	17.18 (**+ 7.50**)	57.42 (**+29.13**)
		Spot	28.65 (**+11.04**)	70.71 (**+30.07**)
	MSeg3D [55]	All	20.87 (**+12.18**)	36.26 (**+22.18**)
		Alice	14.77 (**+ 8.10**)	34.23 (**+20.14**)
		Spot	27.91 (**+14.10**)	60.85 (**+31.03**)

with I being the image, $P(x,y)$ the predicted label, $GT(x,y)$ the ground truth and $\mathbb{1}$ the indicator function. The IoU is accumulated over the entire test-set and averaged over all classes to yield the mIoU. The IoU metric is known to be biased towards object instances and classes that cover large areas of the image [6], therefore fine-grained datasets with many classes as ours generally produce weaker mIoU results.

C. Semantic Segmentation

In Table II, we provide averaged IoU values for different state-of-the-art 2D and 3D semantic segmentation 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)*. PP-LiteSeg [51] uses an encoder-decoder structure with a lightweight attention-based fusion model in the decoder to enable real-time semantic segmentation. DDRNet [52] is based on BiSeNet [56], which uses a typical two-stream architecture and fuses both branches at different depths in the network. For 3D segmentation, the recent Point Transformer V3 [53] (PTv3) makes use of so-called Point Transformer layers on the point cloud input as its building blocks in a encoder-decoder architecture similar to U-Net [57]. We use the lidar-only variant of MSeg3D [55] that uses a voxel-based feature encoder with sparse convolutions for point-wise feature learning similar to Part-A^2 [58]. The multimodal variant is left for future work as it requires adaptive training across the setups of each platform.

We observe a good performance on labels with a high presence in the data (e.g. *vegetation*, *terrain*, *sky*) and an expected lower performance on poorly represented classes. The results are very similar and comparable to those obtained in [1], with the difference that the scenarios and platforms represented in the GOOSE-Ex data are more diverse than those of the original GOOSE dataset. When the category models are evaluated on the original GOOSE test split, a mIoU of *55.75%* and *49.75%* is obtained for PP-LiteSeg and DDRNet respectively, showing good generalization capabilities on all scenarios. A platform-specific comparison can be seen in Table III. Here we observe a lower performance on the excavator platform, due to an unorthodox field of view that observes mostly ground pixels and points. The appearing vegetation classes are harder to distinguish, especially for the 3D cases, and simple classes like *Sky* appear less often. As stated above, the *class* results are quite low compared to other datasets due to several very rare classes that produce IoU values of zero. The impressive performance of the PTv3 model on categories suggests that 3D segmentation models can provide valuable input for robust navigation solutions even in difficult unstructured environments.

V. Downstream Applications

We verify the practicability of the GOOSE-Ex dataset on multiple downstream applications.

A. Terrain Traversability Estimation

In many cases, coarse-grained semantic segmentation is sufficient to interface a traversability analysis with a path planner. GANav [59] proposes a group-wise attention mechanism to segment RGB images of unstructured environments into navigable regions. The attention mechanism and the corresponding group-wise attention loss help the transformer architecture to efficiently fuse multi-scale image features. The approach has originally been tested on the RUGD [3] and the RELLIS-3D [4] datasets. We train GANav on the GOOSE and GOOSE-Ex datasets and test on all four datasets to verify the generalizability properties. We follow the original categorization into 6 semantic classes: *smooth*, *rough*, *bumby*, *forbidden*, *obstacle*, *background*.

TABLE IV: Results of GANav [59] trained and tested on different datasets to investigate their generalizability.

test	mIoU↑ train on RUGD	train on RELLIS-3D	train on GOOSE	fine-tune on GOOSE-Ex
RUGD [3]	89.08	15.38	21.59	29.35
RELLIS-3D [4]	24.76	74.44	36.86	45.56
GOOSE [1]	17.74	17.87	37.99	41.36
GOOSE-Ex	22.95	19.74	31.45	54.89

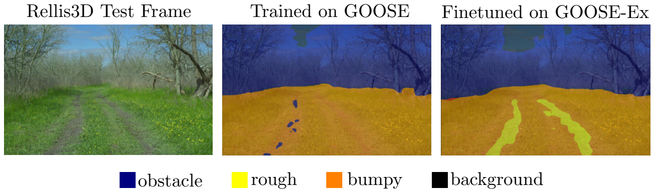


Fig. 4: GANav [59]: Qualitative Segmentation on Unseen Dataset. Fine-tuning on GOOSE-Ex exhibits the best generalization to other datasets like RELLIS-3D [4].

According to the GANav [59] segmentation results in Table IV and Figure 4, the GOOSE-Ex fine-tuning achieves the best generalization results across all four datasets. The high mIoU results on the RELLIS-3D and RUGD test sets trained on their respective train sets indicate a low inter-set variance compared to the GOOSE datasests. Also, the models have not been optimized for a high performance on the GOOSE datasets, as all hyperparameters such as crop sizes and aspect ratios are designed to work well on the RUGD dataset.

B. Object Manipulation

Construction machines are often employed to manipulate heavy or hazardous objects. For these tasks, it is crucial to accurately identify and distinguish the target object from the surrounding environment. We trained Mask2Former [60] as a panoptic segmentation approach on the GOOSE-Ex datasets to exploit the instance labels of many objects in the dataset. Specifically, we trained a model for panoptic segmentation of barrels, some qualitative results are shown in Figure 5. After a fusion with the depth perception, this allows the excavator to perform accurate pose estimation and subsequent grasping of barrels in complex environments.

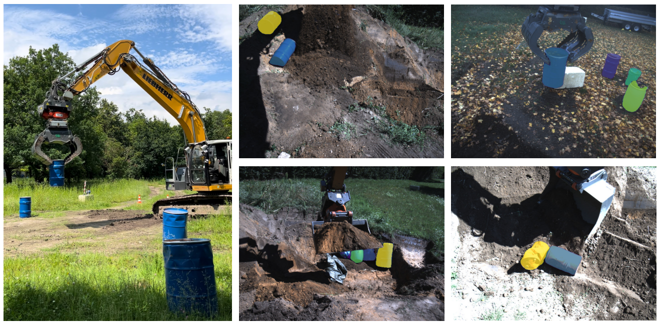
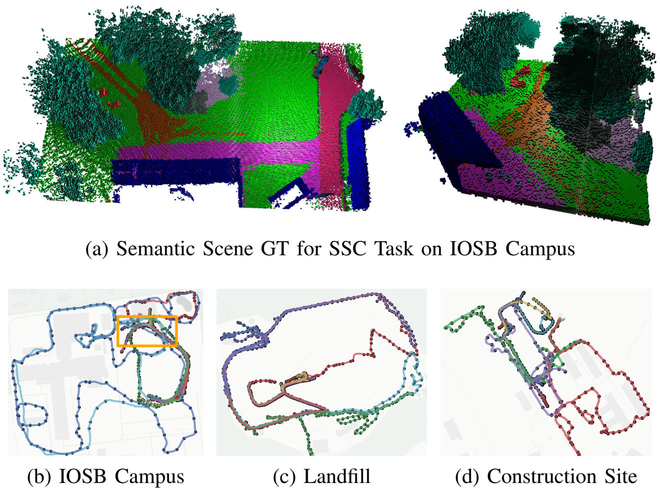


Fig. 5: Excavator *ALICE* grasping barrels (left) using panoptic segmentation masks (right) obtained from a Mask2Former [60] model trained on the GOOSE-Ex dataset.

C. SLAM Benchmark and Semantic Scene Completion

Beyond the scope of semantic segmentation, the GOOSE-Ex dataset could be used as an odometry or SLAM test set for several tasks: Single-robot odometry estimation, loop-closure detection based on images or point clouds, loop-closure detection based on semantics, or even multi-robot SLAM using the overlapping trajectories of different robots. Figures 6b-6d provide an overview of sequences of selected scenarios that may prove useful for these tasks. All sequences come with a GNSS-based ground truth estimate.



(a) Semantic Scene GT for SSC Task on IOSB Campus

(b) IOSB Campus (c) Landfill (d) Construction Site

Fig. 6: A selection of sequences that can be used as testbed for (semantic-)SLAM approaches. Dots ● indicate an annotated frame. The highlighted area □ in (b) corresponds to the reconstructed semantic scene displayed in (a). Background maps: ©OpenStreetMap contributors

We have post-processed several groups of sequences with a SLAM approach [61] based on GTSAM [62] to achieve a high annotation density for different regions. By fusing multiple frames we can generate ground truth semantic maps that can be used to tackle the task of Semantic Scene Completion (SSC) [63] in unstructured environments. The goal of the SSC task is to predict both geometry and semantics for a specified target volume, given a single LiDAR scan as input. We use the processing pipeline of SemanticKITTI [9] to generate voxel maps with a voxel resolution of 0.2 m. Figure 6a shows an example region. We plan to release this as standalone SSC dataset in the future.

VI. Conclusion and Future Work

We present the GOOSE-Ex dataset for semantic segmentation in unstructured environments across domains and platforms. In future work, we want to explore approaches that benefit from additional modalities, such as NIR or LiDAR, and extend the research scope from semantic segmentation to other tasks such as semantic SLAM.

Acknowledgements We thank the Autonomous Robotic Systems group at Fraunhofer IOSB for their support during several data collection trips. We also thank Roman Abayev and Anselm von Gladiß for maintaining the GOOSE DB for querying sequences within the GOOSE datasets.

References

[1] P. Mortimer, R. Hagmanns, M. Granero, T. Luettel, J. Petereit, and H.-J. Wuensche, "The GOOSE Dataset for Perception in Unstructured Environments," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2024. 1, 2, 4, 5, 6

[2] T. Guan, Z. He, R. Song, D. Manocha, and L. Zhang, "TNS: Terrain Traversability Mapping and Navigation System for Autonomous Excavators," in *Proceedings of International Conference on Robotics: Science and Systems (RSS)*, 2022. 2

[3] 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)*, 2019. 2, 5, 6

[4] 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)*, 2021. 2, 3, 5, 6

[5] 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

[6] 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)*, 2016. 2, 5

[7] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[8] 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

[9] 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)*, 2019. 2, 3, 6

[10] 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.

[11] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

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

[13] 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

[14] 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

[15] 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

[16] R. Nunes, J. F. Ferreira, and P. Peixoto, "Procedural Generation of Synthetic Forest Environments to Train Machine Learning Algorithms," in *Proceedings of IEEE International Conference on Robotics and Automation Workshops (ICRAW)*, 2022.

[17] S. Sharma, L. Dabbiru, T. Hannis, G. Mason, D. W. Carruth, M. Doude, C. Goodin, C. Hudson, S. Ozier, J. E. Ball, and B. Tang, "CaT: CAVS Traversability Dataset for Off-Road Autonomous Driving," *IEEE Access*, vol. 10, 2022. 2

[18] M. Hoveidar-Sefid and M. Jenkin, "Autonomous Trail Following using a Pre-trained Deep Neural Network," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

[19] S. Hosseinpoor, J. Torresen, M. Mantelli, D. Pitto, M. Kolberg, R. Maffei, and E. Prestes, "Traversability Analysis by Semantic Terrain Segmentation for Mobile Robots," in *Proceedings of IEEE International Conference on Automation Science and Engineering (CASE)*, 2021.

[20] 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

[21] 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

[22] 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

[23] M. Sivaprakasam, P. Maheshwari, M. G. Castro, S. Triest, M. Nye, S. Willits, A. Saba, W. Wang, and S. Scherer, “TartanDrive 2.0: More Modalities and Better Infrastructure to Further Self-Supervised Learning Research in Off-Road Driving Tasks,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[24] A. Datar, C. Pan, M. Nazeri, and X. Xiao, “Toward Wheeled Mobility on Vertically Challenging Terrain: Platforms, Datasets, and Algorithms,” in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2024. 2

[25] 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

[26] R. W. Wood, “Photography by Invisible Rays,” *Photographic Journal*, pp. 329–338, 1910. 2

[27] P. Mortimer and H.-J. Wuensche, “TAS-NIR: A VIS+NIR Dataset for Fine-grained Semantic Segmentation in Unstructured Outdoor Environments,” in *Proceedings of 12th Workshop On Planning, Perception and Navigation for Intelligent Vehicles (PPNIV), IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2022. 2

[28] G. Varma, A. Subramanian, A. Namboodiri, M. Chandraker, and C. V. Jawahar, “India Driving Dataset (IDD): A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments,” in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019. 2

[29] C. Sakaridis, D. Dai, and L. Van Gool, “Semantic Foggy Scene Understanding with Synthetic Data,” *International Journal of Computer Vision*, 2018. 2

[30] C. Sakaridis, D. Dai, S. Hecker, and L. Van Gool, “Model Adaptation with Synthetic and Real Data for Semantic Dense Foggy Scene Understanding,” in *Proceedings of European Conference on Computer Vision (ECCV)*, 2018, pp. 707–724.

[31] X. Hu, C.-W. Fu, L. Zhu, and P.-A. Heng, “Depth-Attentional Features for Single-Image Rain Removal,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[32] F. A. Shaik, A. Reddy, N. R. Billa, K. Chaudhary, S. Manchanda, and G. Varma, “IDD-AW: A Benchmark for Safe and Robust Segmentation of Drive Scenes in Unstructured Traffic and Adverse Weather,” in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2024.

[33] C. Sakaridis, D. Dai, and L. Van Gool, “Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation,” in *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2019. 2

[34] C. Frese, A. Zube, P. Woock, T. Emter, N. F. Heide, A. Albrecht, and J. Petereit, “An autonomous crawler excavator for hazardous environments,” *at - Automatisierungstechnik*, vol. 70, no. 10, 2022. 2, 4

[35] L. Wang, Z. Ye, and L. Zhang, “Hierarchical Planning for Autonomous Excavator on Material Loading Tasks,” in *Proceedings of International Symposium on Automation and Robotics in Construction (ISARC)*, 2021. 2

[36] Q. Guo, Z. Ye, L. Wang, and L. Zhang, “Imitation Learning and Model Integrated Excavator Trajectory Planning,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[37] Y. Zhu, L. Wang, and L. Zhang, “Excavation of Fragmented Rocks with Multi-modal Model-based Reinforcement Learning,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022. 2

[38] R. Song, S. Ong, L. Kang, S. Jin, Y.-C. Peng, Z. He, L. Qian, and L. Zhang, “Autonomous Excavator System for Construction Automation,” in *Proceedings of International Conference on Robotics: Science and Systems Workshops (RSSW)*, 2023. 2

[39] M. Eder, R. Prinz, F. Schöggl, and G. Steinbauer-Wagner, “Generating Robot-Dependent Cost Maps for Off-Road Environments Using Locomotion Experiments and Earth Observation Data,” in *Proceedings of IEEE International Conference on Robotic Computing (IRC)*, 2022. 2

[40] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone, “Socially CompliAnt Navigation Dataset (SCAND): A Large-Scale Dataset Of Demonstrations For Social Navigation,” *IEEE Robotics and Automation Letters*, 2022. 2

[41] S. Jung, W. Yang, and A. Kim, “Co-RaL: Complementary Radar-Leg Odometry with 4-DoF Optimization and Rolling Contact,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024. 2

[42] K. Chaney, F. Cladera, Z. Wang, A. Bisulco, M. A. Hsieh, C. Korpela, V. Kumar, C. J. Taylor, and K. Daniilidis, “M3ED: Multi-Robot, Multi-Sensor, Multi-Environment Event Dataset,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 2

[43] A. Avogaro, A. Toaiari, F. Cunico, X. Xu, H. Dafas, A. Vinciarelli, E. Li, and M. Cristani, “Exploring 3D Human Pose Estimation and Forecasting from the Robot’s Perspective: The HARPER Dataset,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024. 2

[44] F. Angelini, P. Angelini, C. Angiolini, S. Bagella, F. Bonomo, M. Caccianiga, C. D. Santina, D. Gigante, M. Hutter, T. Nanayakkara, P. Remagnino, D. Torricelli, and M. Garabini, “Robotic Monitoring of Habitats: The Natural Intelligence Approach,” *IEEE Access*, 2023. 2

[45] P. Mortimer and M. Maehlisch, “Survey on Datasets for Perception in Unstructured Outdoor Environments,” in *Proceedings of IEEE International Conference on Robotics and Automation Workshops (ICRAW)*, 2024. 2

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

[47] "Datasheets, CAD files, and guides for our lidar sensor | Ouster — ouster.com," https://ouster.com/downloads, [Accessed 11-09-2024]. 4

[48] "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. 4

[49] 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. 4

[50] 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. 4

[51] 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. 5

[52] 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. 5

[53] X. Wu, L. Jiang, P.-S. Wang, Z. Liu, X. Liu, Y. Qiao, W. Ouyang, T. He, and H. Zhao, "Point Transformer V3: Simpler, Faster, Stronger," in *CVPR*, 2024. 5

[54] Pointcept Contributors, "Pointcept: A Codebase for Point Cloud Perception Research," https://github.com/Pointcept/Pointcept, 2023. 5

[55] J. Li, H. Dai, H. Han, and Y. Ding, "MSeg3D: Multi-Modal 3D Semantic Segmentation for Autonomous Driving," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 5

[56] 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. 5

[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015. 5

[58] S. Shi, Z. Wang, J. Shi, X. Wang, and H. Li, "From Points to Parts: 3D Object Detection From Point Cloud With Part-Aware and Part-Aggregation Network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 5

[59] T. Guan, D. Kothandaraman, R. Chandra, A. J. Sathyamoorthy, K. Weerakoon, and D. Manocha, "GA-Nav: Efficient Terrain Segmentation for Robot Navigation in Unstructured Outdoor Environments," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 8138–8145, 2022. 5, 6

[60] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 6

[61] T. Emter and J. Petereit, "3D SLAM With Scan Matching and Factor Graph Optimization," in *ISR 2018; 50th International Symposium on Robotics*, 2018. 6

[62] F. Dellaert and GTSAM Contributors, "borglab/gtsam," May 2022. [Online]. Available: https://github.com/borglab/gtsam 6

[63] L. Roldão, R. de Charette, and A. Verroust-Blondet, "3D Semantic Scene Completion: A Survey," *International Journal of Computer Vision*, vol. 130, 2022. 6