MCD: Diverse Large-Scale Multi-Campus Dataset for Robot Perception

Supplementary Material

1. Related Works

Tab. 1 provides a detailed comparison in terms of perception modality, ground truth approach and diversity of existing datasets and shows the contributions of MCD. Most well-annotated multi-modal datasets are biased towards autonomous driving scenarios, while environment-diversified datasets primarily focus on SLAM or reconstruction studies with limited annotations. Moreover, autonomous driving datasets often prioritize well-structured highways, overlooking diverse settings where vehicles share unmarked surfaces with pedestrians. This limitation hinders the global adoption of autonomous car solutions. Simulated datasets and methods offer unbiased setups but suffer from a significant domain gap with real-world conditions, making them inadequate for simulating extreme scenarios and corner cases, which pose challenges for robotics research. MCD aims to bridge these gaps with diverse challenges, modalities, and scenarios.

Many existing datasets face challenges related to the accuracy of ground truth estimates. GNSS/INS, the most commonly used localization method for automotive vehicles, often exhibits errors at the decimeter level. Terrestrial Laser Scanner (TLS) is typically employed in conjunction with a limited number of control points to estimate continuous or sparse ground truth coordinates. On the other hand, the Motion Capture system (MoCap) is primarily effective indoors or in low-light conditions. Both TLS and MoCap offer accuracy at the centimeter to millimeter level with clear line-of-sight. Two commonly used methods for trajectory estimation through point cloud matching are Normal Distributions Transform (NDT) and Iterative Closest Point (ICP). These methods often introduce errors in the range of a few decimeters. Simultaneous Localization and Mapping (SLAM) uses onboard LIDAR sensors but may suffer from long-term drift and is generally considered less accurate, with measurement noise ranging from a few decimeters to several meters. The proposed survey-grade prior map continuous time registration (SMCTR) is estimated to provide centimeter-level accuracy without any line-of-sight requirements, making it a promising choice for creating large datasets with centimeter accuracy.

2. Sensor Suites

Two sensor suites are built for data collection. The first is mounted on an All-Terrain Vehicle (ATV) that can run up to 35 km/h, and the second is a portable Handheld Setup (HHS) that allows us to walk through more complex terrains like cobbled paths or staircases. Fig. 1 presents the mobility mode of the two sensor suites.

Figure 1. Mobile platform of each sensor suite: ATV (left) and HHS (right).

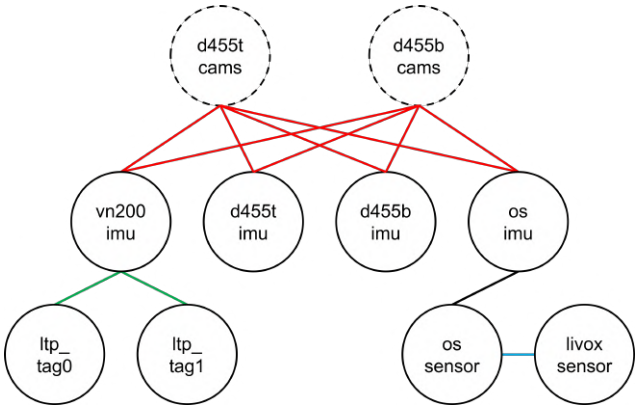

Figure 2. The graph of calibrated extrinsics for the HHS. Each edge is color-coded for different approaches to find their relative extrinsics. More details are given in Sec. 3.

Data from each hardware unit is grouped into separate rosbags so that users can download the modalities of their interest. Tab. 2 reports the topic available for each sensor suite and the nominal rate.

3. Calibration

The calibration process involves finding the camera intrinsics and the extrinsics of all sensors.

First, the Kalibr toolbox [37] is used to find the intrinsic parameters of the D455 cameras. These intrinsic parame-

Year-Dataset	Perception Modalities						Groundtruth			GT Method	MC
	C	L_C	L_N	I	U	O	Pose	Map	Labels		
2012-KITTI [1]	✓	✓	×	×	×	✓	✓	×	✓	GNSS/INS	×
2016-NCLT [2]	✓	✓	×	✓	×	✓	✓	×	×	GNSS/INS	×
2017-Oxford RoboCar [3]	✓	✓	×	✓	×	✓	✓	×	×	GNSS/INS	×
2018-ApolloScape [4]	✓	✓	×	✓	×	×	✓	3D	✓	GNSS/INS	✓
2019-KAIST Urban [5]	✓	✓	×	✓	×	✓	✓	×	✓	SLAM	×
2019-Argoverse [6, 7]	✓	✓	×	×	×	✓	✓	2D	✓	Unspecified	✓
2019-ECP [8]	✓	✓	×	×	×	×	×	×	✓	Unspecified	✓
2019-H3D [9]	✓	✓	×	✓	×	✓	✓	×	✓	GPS/NDT	×
2020-NuScene [10]	✓	✓	×	✓	×	✓	✓	2D	✓	GNSS/INS/ICP	✓
2020-UrbanLoco [11]	✓	✓	×	×	×	✓	✓	×	✓	GNSS/INS	✓
2020-MulRan [12]	✓	✓	×	✓	×	✓	✓	×	×	SLAM	×
2020-Waymo [13]	✓	✓	×	×	×	×	✓	2D	✓	Unspecified	✓
2020-A2D2 [14]	✓	✓	×	✓	×	✓	✓	3D	✓	ICP	✓
2020-A*3D [15]	✓	✓	×	×	×	✓	✓	×	✓	GNSS/INS	×
2020-Ford Multi-AV [16]	✓	✓	×	✓	×	×	✓	3D	×	SLAM	×
2021-ONCE [17]	✓	✓	×	×	×	×	×	3D	✓	Unspecified	✓
2021-Lyft [18]	✓	✓	×	×	×	✓	✓	2D	✓	Unspecified	×
2021-CADC [19]	✓	✓	×	✓	×	✓	✓	×	✓	GNSS/INS	✓
2021-PandaSet [20]	✓	✓	×	✓	×	✓	✓	×	✓	GNSS/INS	✓
2023-Boreas [21]	✓	✓	×	✓	×	×	✓	×	✓	GNSS/INS	×
2014-ICL-NUIM [22]	✓	✓	×	×	×	×	✓	3D	×	Simulator	×
2017-CARLA [23]	✓	✓	×	✓	×	✓	✓	3D	✓	Simulator	×
2020-TartanAir [24]	✓	✓	×	✓	×	×	✓	3D	✓	Simulator	×
2023-AIODrive [25]	✓	✓	×	✓	×	✓	✓	3D	✓	Simulator	×
2012-TUM RGB-D [26]	✓	×	×	×	×	×	✓	×	×	MoCap	×
2016-EuRoC [27]	✓	×	×	✓	×	×	✓	3D	×	TLS/MoCap	×
2016-TorontoCity [28]	✓	✓	×	×	×	×	✓	3D	✓	Photogrammerty	×
2020-Newer College [29]	✓	✓	×	✓	×	×	✓	3D	×	TLS/ICP	×
2021-Hilti [30, 31]	✓	✓	×	✓	×	×	✓	×	×	Control Points	✓
2021-M2DGR [32]	✓	✓	×	✓	×	×	✓	×	×	TLS/MoCap	×
2022-PolyU-BPCoMa [33]	✓	✓	×	✓	×	✓	✓	3D	×	TLS	×
2022-NTU VIRAL [34]	✓	✓	×	✓	✓	✓	✓	×	×	TLS	×
2023-GRACO [35]	✓	✓	×	✓	×	✓	✓	×	×	GNSS/INS	×
2023-FLICAR [36]	✓	✓	✓	✓	×	×	✓	×	×	TLS	×
MCD (proposed) [34]	✓	✓	✓	✓	✓	×	✓	3D	✓	SMCTR	✓

Table 1. C denotes camera(s), L_C represents classical mechanical LIDAR, L_N signifies the newer type of non-repetitive scanning LIDAR, I indicates a raw, high-frequency inertial measurement unit, and U stands for UWB. O refers to other popular sensing modalities, such as thermal imaging, wheel odometry, low-frequency GPS/INS, or RADAR. The 'map' in the ground truth refers to the survey-grade map result in either 2D or 3D, and 'Tools' denotes the methodology employed to generate ground truth besides manual annotations.

ters are then used to find the extrinsic of the 3-camera setup on each D455 w.r.t. each IMU available. Hence, eight 3-camera-1-IMU calibrations are carried out, which are represented by the red edges in Fig. 2.

Next, the transform from the ouster IMU to the ouster sensor (black edge in Fig. 2) is obtained from the datasheet. The transform from the body frame to the UWB tags (green edge) is measured directly via the motion capture system. Finally, the transformation matrix from the ouster LiDAR to the Livox sensor is obtained by aligning the LiDAR point clouds of a room corner using the ICP method.

Finally, we perform a pose-graph optimization over the graph in Fig. 5 to obtain the final extrinsic of all sensors and summarize them in the calibration report.

Modality	Hardware	Available ROS Topics		Rate
		ATV	HHS	
Camera	D435i	/d435i/infra1/image_rect_raw /d435i/infra1/image_rect_raw	-	30 Hz
	D455t	-	/d455t/infra1/image_rect_raw /d455t/infra2/image_rect_raw /d455t/color/image_raw	30 Hz
	D455b	/d455b/infra1/image_rect_raw /d455b/infra2/image_rect_raw /d455b/color/image_raw	/d455b/infra1/image_rect_raw /d455b/infra2/image_rect_raw /d455b/color/image_raw	30 Hz
IMU	Ouster OS1	/os_cloud_node/imu	/os1_cloud_node/imu	100 Hz
	D435i	/d435i/imu	-	400 Hz
	D455t	-	/d455t/imu	400 Hz
	D455b	/d455b/imu	/d455b/imu	400 Hz
	VN100	/vn100/imu	-	400 Hz
	VN200	/vn200/imu	/vn200/imu	400 Hz
LiDAR	Ouster OS1	/os_cloud_node/points	/os_cloud_node/points	10 Hz
	Livox Mid70	/livox/lidar	/livox/lidar	10 Hz
UWB	Linktrack P	/ltp_tag0/nlnf3 /ltp_tag1/nlnf3	/ltp_tag0/nlnf3 /ltp_tag1/nlnf3	20 Hz

Table 2. Summary of sensing modalities, hardware units, ROS topics, and the nominal rates on each platform.

4. Coordinate System

Fig. 3 illustrates the convention of coordinate frame used in this dataset. In this dataset, we define $\mathtt{W}$ as the coordinate frame of the ground truth map, which is aligned with the GPS coordinate system. The body frame $\mathtt{B}$ is chosen to coincide with the VN100 IMU on the ATV and the VN200 IMU on the HHS. A time index t is attached to $\mathtt{B}$ to differentiate the body frame at different times. It is a common practice that the body frame at initial time $\mathtt{B}_0$ is chosen as the frame of reference for all subsequent pose estimates, i.e., ${}^{\mathtt{B}_t}_{\mathtt{B}_0}\hat{\mathbf{T}}$. Each sensor has a coordinate frame $\mathtt{S}$ whose transform (the extrinsic) to the body frame is provided in a .yaml calibration file. We refer to our online note[1] for more details on the convention and interpretation of the provided data.

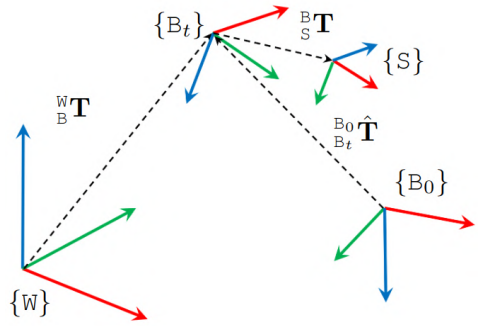

Figure 3. The coordinate frames of ground truth, time-varying body frame, and the body-fixed sensor.

5. Survey-Grade Ground Truth Map

5.1. Survey Mapping Process

3D survey-grade mapping of the campuses was done by using various terrestrial laser scanners (TLS)[2,3]. In essence, TLS acquires point cloud data from discrete, static locations by means of ground-based, high-resolution 3D laser scanners. These individual scans are then combined to produce a complete 3D point cloud via co-visible landmarks (both man-made and natural) and propriety global scan matching software. point cloud data consists of 3D point coordinates and may also include intensity or RGB channels.

Fig. 4 shows the hardware setup for the mapping mission over the most challenging campuses due to its elongated shape. We divided the campus into three sections and

[1] https://mcdviral.github.io/UserManual#coordinate-systems

[2] https://leica-geosystems.com/products/laser-scanners/scanners/leica-rtc360

[3] https://faro.com/en/Products/Hardware/Focus-Laser-Scanners

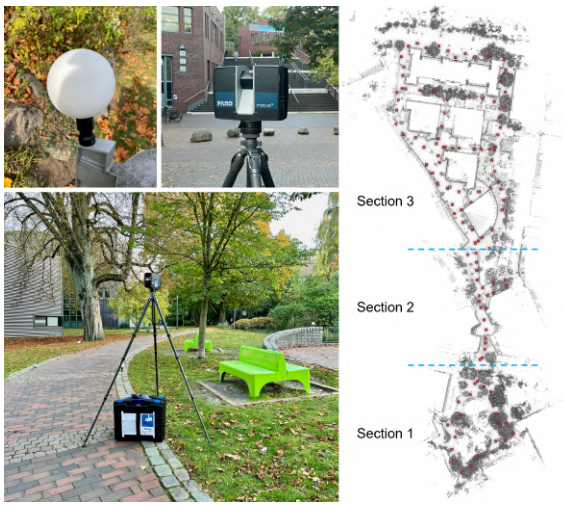

Figure 4. Target sphere, FARO laser scanner, and scan positions distributed across the TUHH campus.

scanned them in succession from north to south. Scanning each section took between 2 to 3 days, requiring at least two field operators.

Upon completion of the scanning process, all individual scans were registered in the manufacturer's software. Registration of individual scans in each of the three sections was performed solely on the basis of the artificial sphere targets recorded in the scans. Cloud-to-cloud registration was then used to align the three sections and create the final point cloud of the entire campus area.

Target-based statistics showed a maximum distance error of less than 10 mm for all sphere pairings that were used in the registration optimization process. Building walls and corners were spot-checked for overlap of scans. A total of 15 reference measurements were made with a total station to verify the global accuracy of the point cloud. The results showed a maximum distance error of less than 5 cm over the entire length of the campus, which is below the resolution of the point cloud.

5.2. Visualization

The survey maps are the key assets of MCD which will be made available on our website. These maps enable the SMCTR algorithm and facilitate the labeling process of thousands of LiDAR scans. In this section, we present an overview of the ground truth maps and their annotated versions. Hence, we delve into visualizing specific details within each asset, see Fig. 6, 7, 8.

The original scanned map for each campus amounts to approximately 200 GB. This data size makes it impossible to visualize the raw point clouds with its annotations. As a result, for the purpose of visualization and comparison, we uniformly downsampled the point clouds to a 5-centimeter

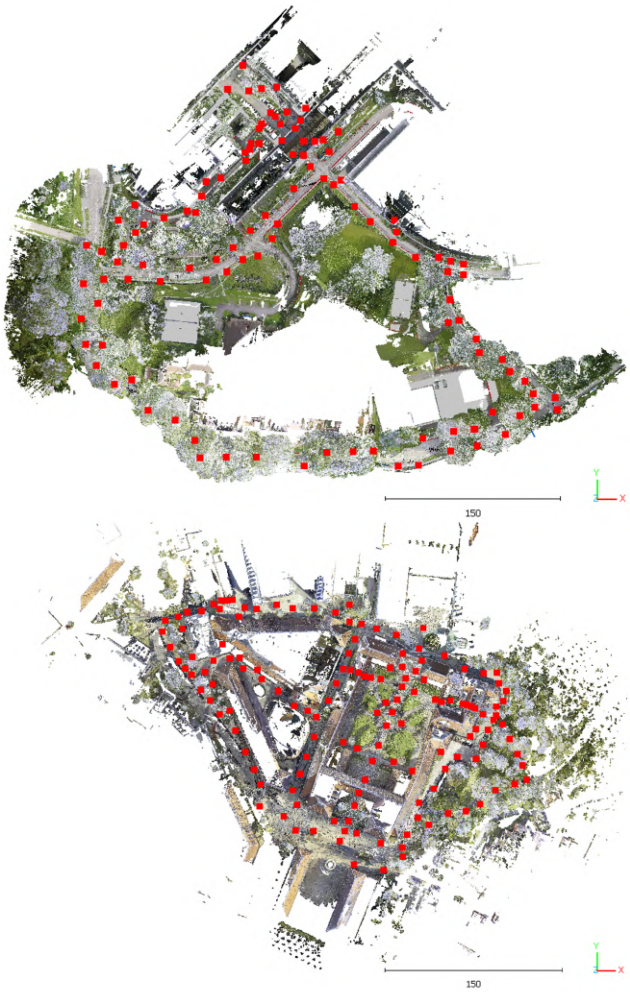

Figure 5. Ground truth scan position distributions across the NTU and KTH campus. TUHH consists of 114 scans and TUHH consists of 145 scans.

voxel size across all campuses and reduced each map to less than 7 GB. Additionally, we applied blurring techniques to conceal road information and human faces. Overall, all three campuses are situated in developed countries and exhibit a balance between vegetation and buildings. However, Campus NTU and KTH stand out as they have less reflectivity, resulting in more comprehensive point cloud coverage. On the other hand, Campus TUHH boasts modern architecture with a significant number of reflective glass surfaces. Consequently, it has the lowest point cloud coverage among the three campuses because of the reflection noise.

6. Continuous-time Trajectory

To make working with continuous-time trajectory ground truth easy, we have created a python wrapper for the basalt

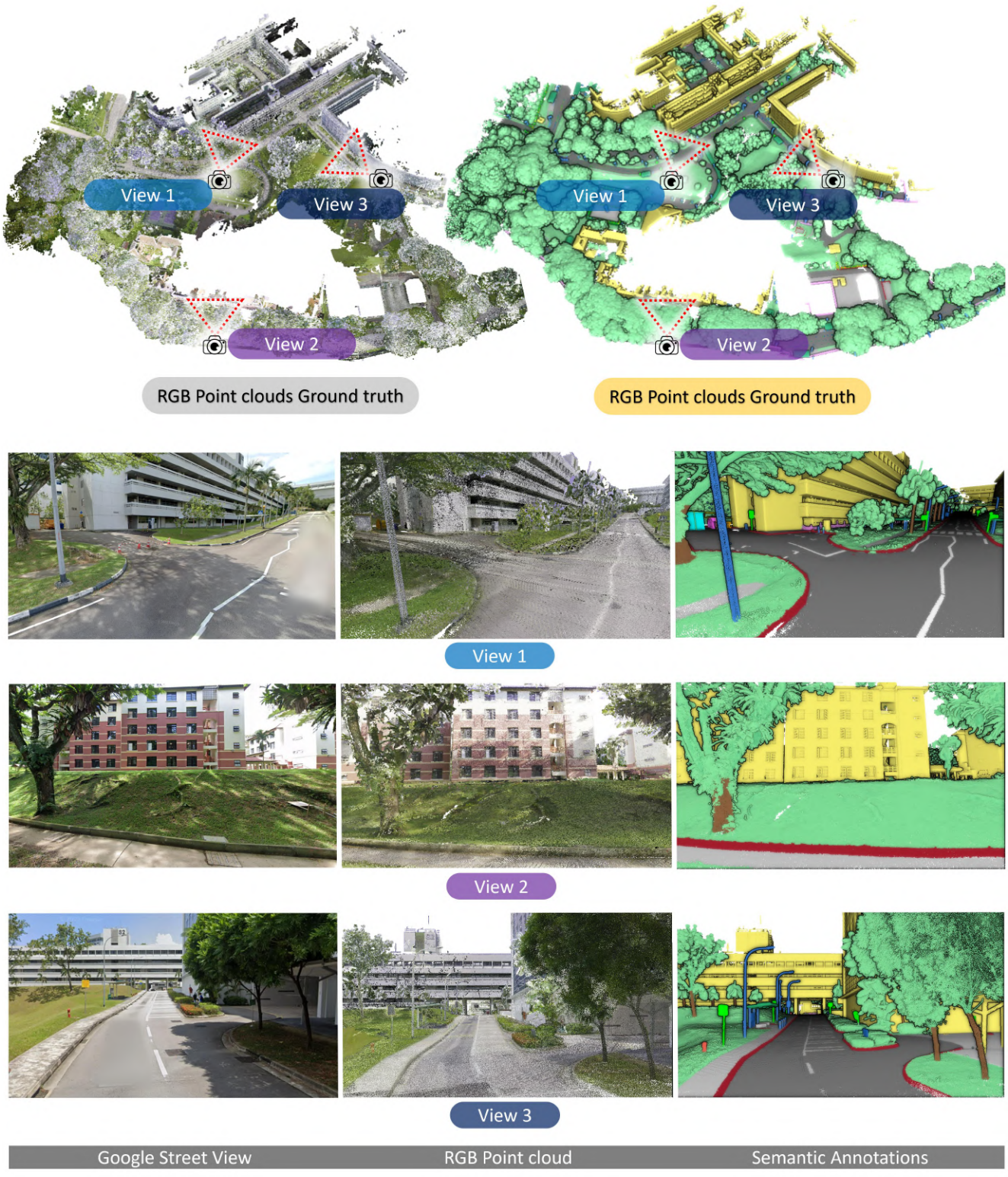

Figure 6. MCD NTU ground truth map highlights.

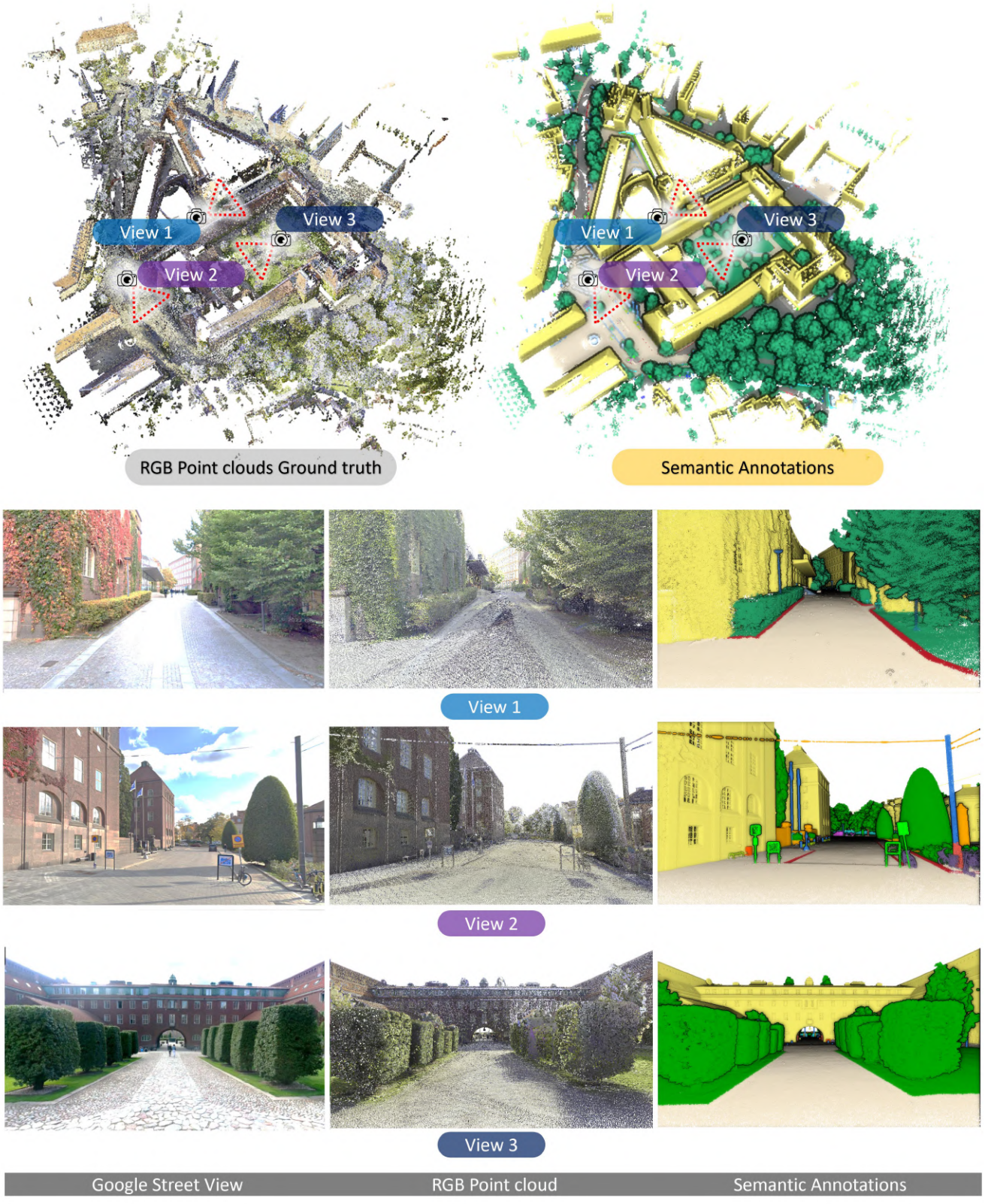

Figure 7. MCD KTH ground truth map highlights.

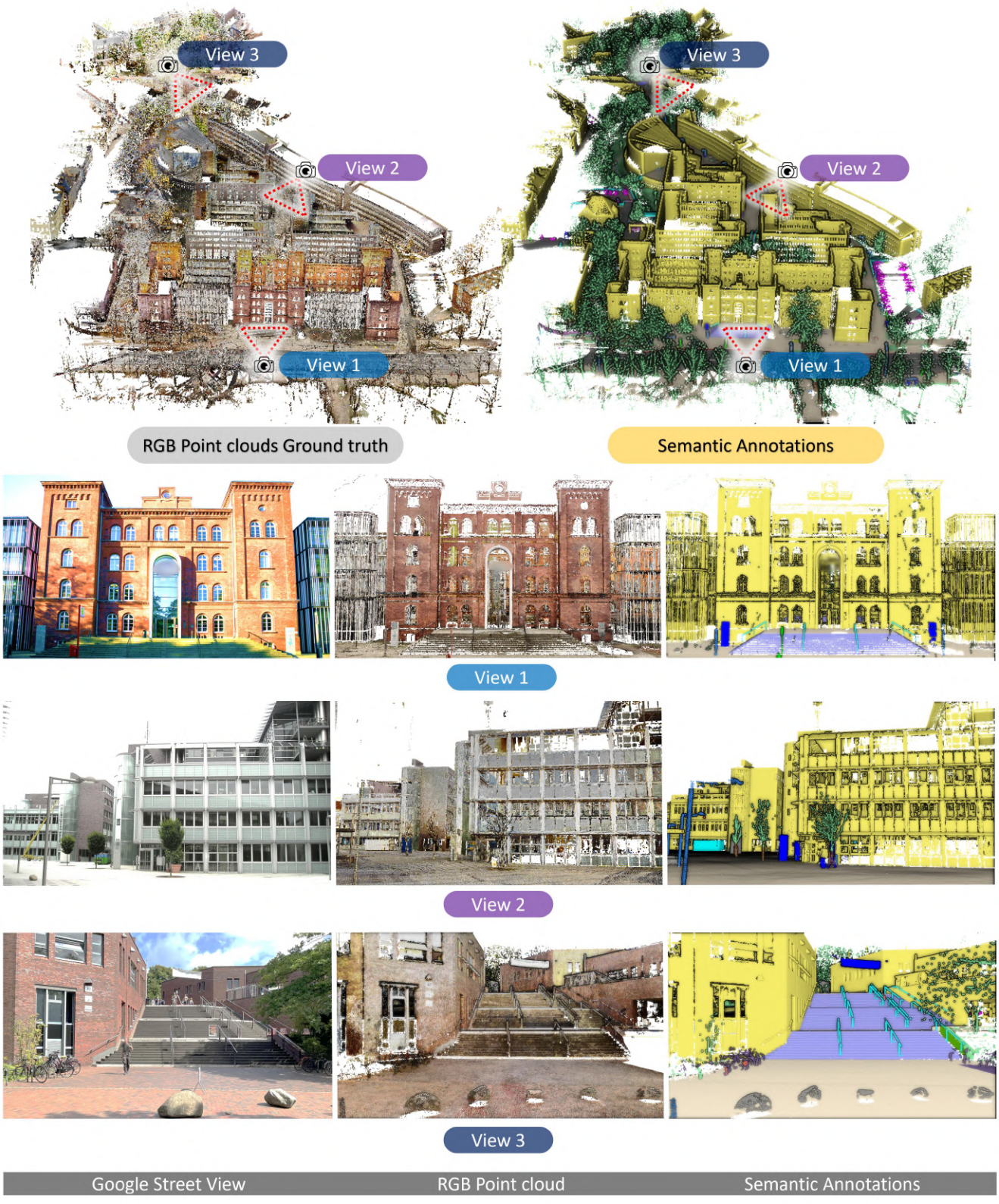

Figure 8. MCD TUHH ground truth map highlights.

library [38] called *CEVA* (**C**ontinuous-time **Eva**luation), which can be installed from `github.com/mcdviral/CEVA` . In many aspects, CEVA exceeds the basalt library in its utilities. Some of the improvements we have made are:

Easy use: CEVA is a Python module, therefore, it can be easily imported into any Python program or Jupyter notebooks instead of having to be included and compiled like the C++ library basalt. Several Jupyter notebooks are provided at scripts to showcase CEVA's utilities.

Run-time selected spline order: the basalt library uses a C++ template for the spline order to allow faster matrix computations. Without compromising the speed, in CEVA, a spline with any spline order from 4 to 9 can be created at run time. This is particularly useful for parameter tuning since when changing the spline order, the program does not have to be recompiled. CEVA achieves this by precompiling the splines of different orders, and after the order is declared, all methods will simply switch to the spline with the corresponding order.

Load/save spline from/to a log: The user can easily load or save a spline module from or to a csv log file containing the spline's intrinsic parameters (order, knot length, min-time, max-time) and the control points.

Information query: CEVA allows users to easily query information related to the spline, such as min-time, max-time, pose, velocity, and acceleration ... by any set of time instances. Moreover, query methods for other parameters, such as the blending matrix, order, knot length, knot pose ... are also provided for easier inspection of the spline object.

Deskew pointcloud: This is one of the new original features of CEVA that basalt does not have. Once a spline has been initialized with all intrinsic parameters and the control points, the user can pass a numpy array of the LiDAR point clouds and their timestamps to a CEVA object and obtain a deskewed point cloud where all LiDAR points are transformed to the body frame at the beginning of the scan. We provide a demo of this deskew process at the script deskew_demo.ipynb. Fig. 9 demonstrates the registration of point clouds by the SLAM deskew process and CEVA's deskew process using the SMCTR trajectory.

Fitting spline: In many applications, one would like to smooth out noisy trajectory measurements for better visualization or use as a filter signal. For this purpose, we implemented a continuous-time optimization scheme to fit a spline to the noisy pose measurements. This also provides a demo of our SMCTR algorithm, which employs LiDAR and IMU factors, besides the pose factor in the spline fitting method of CEVA. The demo script can be found at the fitspline_demo.ipynb script.

7. Details of Segmentation Benchmarks

In this section, we provide experimental details, benchmark arrangement, and more experimental results of the MCD-based 3D semantic segmentation benchmark.

Experimental Details: Most baseline methods are re-implemented utilizing their default configs on SemanticKITTI [46] and their official open-source code. Two exceptions are SPVCNN [41] and MinkowskiNet [39], for which we leverage the re-implementation from the recent 2DPASS [47] since the original code is based on out-of-date software settings. Since our LiDAR data processes a different FOV compared to SemanticKITTI, we confine the range of the input scene from [-51.2, 0, -4] to [51.2, 100, 20] along x, y, z axis, respectively. Class weights are re-computed based on the class-wise point number in the training split for all baselines. During the inference stage, we adopt the single-scan evaluation protocol without test-time augmentations to make a fair comparison. All experiments are conducted on a single or dual RTX 3090, depending on the preference of baseline settings.

Benchmark Arrangement: As mentioned in the main paper, all methods are trained on the NTU campus. As shown in Tab. 4, we utilized the point cloud scan from ntu_day_01, ntu_day_02, and ntu_night_13 as the training split for all benchmarks, which totally includes 10,606 scans as our training samples. For NTU benchmark, ntu_day_10 is chosen as our testing split simply because it shares a larger recording time gap compared to those in the training split. For the KTH and TUHH benchmark, we select kth_night_05 and tuhh_day_02 as their testing split, respectively, since they contain relatively suitable numbers of testing samples.

More Experimental Results: Due to the page limit, the results of MCD KTH benchmark are included here as in Tab. 3. Similar to MCD TUHH, a serious performance drop (relatively more than 63%) can be observed compared to MCD NTU for all baselines. This indicates the existence of cross-campus discrepancy between MCD KTH and MCD NTU and the fact that existing fully supervised baselines are not able to generalize the knowledge in a cross-campus manner. It is worth mentioning that this performance discrepancy exists not only in long-tailed classes but also in some dominant classes, such as "Road" and "Sidewalk".

Additionally, an ablation study about how the existing methods perform against points of various ranges is provided in Fig. 10. Compared to the performance w.r.t distance analysis in [46], recent methods achieve a better performance in segmenting points within 50m. Yet as the point range increases with much sparser points, a similar performance degeneration as in [46] can be observed. Especially for state-of-the-art S-Former [45], its performance significantly drops from about 50% to less than 20% once the point range exceeds 50 meters. On the contrary, SDSeg3D [43] is the least distance-affected method, achieving the best performance in segmentation over 65m. Our results show that segmenting long-range points still remains a critical problem and it is worth further investigation in the

Methods	*Barrier*	*Bike*	*Building*	*Chair*	*Cliff*	*Container*	*Curb*	*Fence*	*Hydrant*	*Sign*	*Lanemark*	*Other*	*Parkinglot*	*Pedestrian*	*Pole*	*Road*	*Shelter*	*Sidewalk*	*Stairs*	*Structure-O*	*Traffic-C*	*Trunk*	*Vegetation*	*Vehicle*	mIoU
MKNet [39]	0.0	10.2	75.2	0.6	0.1	0.0	**9.4**	2.3	0.0	30.5	0.2	0.3	**0.1**	44.8	20.5	10.3	**0.1**	15.8	0.5	0.8	0.7	32.4	66.6	20.4	14.3 (-69.0%)
SalsaNext [40]	0.0	0.4	51.2	0.0	**0.2**	0.0	3.9	0.2	0.0	9.6	0.6	0.0	0.0	8.9	5.9	7.5	0.0	5.0	0.0	0.0	2.5	13.0	42.7	9.0	6.7 (-80.0%)
SPVCNN [41]	0.0	2.6	67.0	0.4	0.1	0.0	7.2	0.4	0.0	32.6	0.4	0.3	0.0	19.4	19.3	8.7	0.0	8.5	4.4	**0.7**	3.2	25.9	68.1	12.7	11.8 (-75.5%)
Cylinder3D [42]	0.0	2.1	73.8	0.1	0.0	0.0	3.0	2.1	0.0	12.7	**0.8**	0.2	0.0	15.3	12.3	10.0	**0.1**	19.4	0.4	0.0	0.7	22.5	64.5	5.7	10.2 (-75.3%)
SDSeg3D [43]	0.1	9.3	**77.8**	1.4	0.0	0.0	6.1	3.0	0.0	**35.6**	0.5	**0.6**	0.0	50.0	19.3	**11.4**	0.0	20.1	4.5	0.0	1.4	42.0	68.5	27.5	16.4 (-64.9%)
WaffleIron [44]	0.1	4.6	73.6	0.4	0.0	0.0	6.7	0.7	0.0	31.9	0.3	0.0	0.0	26.0	24.1	10.7	0.0	11.8	3.4	0.5	0.1	**46.7**	62.6	20.7	13.5 (-73.1%)
S-Former [45]	0.1	**20.8**	73.9	**3.5**	0.0	0.0	8.1	**4.8**	0.0	33.6	0.2	0.1	0.0	**55.6**	**28.3**	9.5	0.0	**30.6**	**11.6**	0.5	**4.7**	39.1	**68.2**	**31.4**	**17.7** (-70.0%)

Table 3. Segmentation performance of existing state-of-the-art methods on the MCD KTH sequences. The intersection over union (IoU) is utilized as our evaluation matrix. The figure in round brackets is the relative mIoU gap on MCD KTH compared to MCD NTU (bottom), respectively. Note that "Traffic-C" and "Structure-O" are short for "Traffic Cone" and "Structure-Other" classes, respectively.

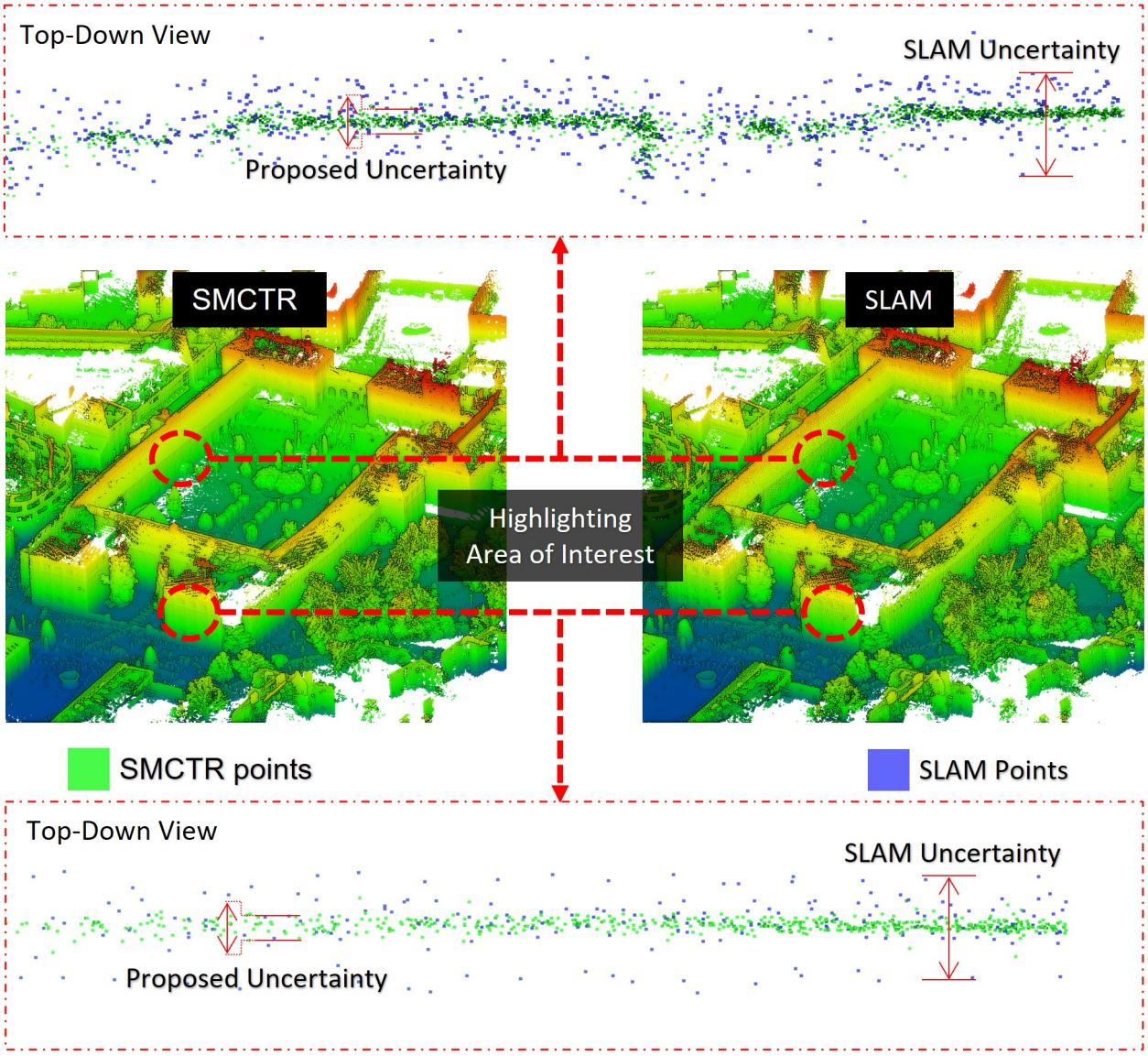

Figure 9. Comparing the SLAM-deskewed point cloud (left) with SMCTR (middle). Notice the Uncertainty of the SLAM-deskewed point cloud on the walls, rooftops and trees. On the other hand, the SMCTR point cloud has fewer registration noises.

Split	Benchmark	Sequence	Seq. No.	Total No.
Train	-	ntu_day_01	6,010	10,606
		ntu_day_02	2,273	
		ntu_night_13	2,323	
Test	NTU	ntu_day_10	3,232	3,232
	KTH	kth_night_05	6,638	6,638
	TUHH	tuhh_day_02	4,986	4,986

Table 4. Split arrangement of semantic segmentation benchmarks. Here "-" means the training split arrangement is utilized for all benchmarks.

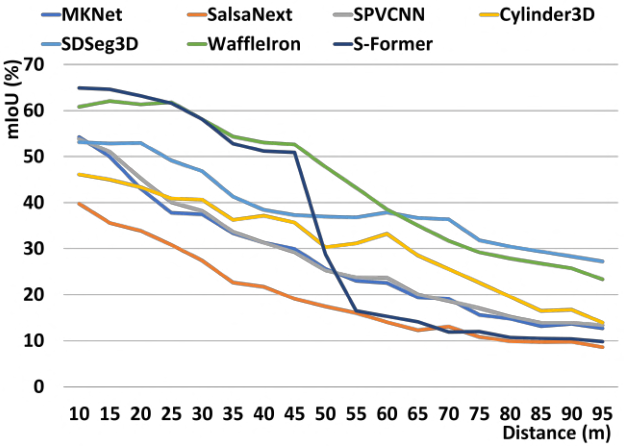

Figure 10. Segmentation performance vs. point range.

future study.

8. The Challenges

LIO Challenge: Fig. 13 report the estimation error over time of the LIO methods upon the three main sequences.

As explained in the SLAM experiment, it is evident that no SLAM method can perform consistently well across all campuses. Noticeably, the LIO-SAM method performs well in the KTH and TUHH environments but has significant drift in the ATV sequences in NTU campus, especially in the z-direction. One reason that we have pointed out is that it is not adept for high-speed scenarios. In addition, since NTU campus has a gradual slope running through a road with dense vegetation, which lacks large horizontal planar features that can constrain its drift on a global scale, especially when it uses a local submap for localization.

Visual Challenge: Fig. 11 illustrates the visual challenges in the MCD. In sequences from campus NTU, solar interference predominantly affects RGB cameras, causing a noticeable purple hue offset. This offset is dynamic, where it appears when the robot is in direct sunlight and disappears when it is in the shade. Additionally, night sequences in campus NTU exhibit significant lighting variations, transitioning from bright regions to darker areas. In general, this scenario closely resembles typical urban areas, making it highly relevant for testing various autonomous driving algorithms.

Campus KTH and TUHH encounter similar issues, with excessive daytime brightness leading to camera overexposure but with lesser hue offset. Furthermore, due to the recent conflict in Ukraine, cities have implemented energy conservation measures, resulting in darker urban environments. This has resulted in consistent low-light conditions, offering a valuable opportunity to evaluate the effectiveness of various night image enhancement algorithms. Moreover, in both of these campuses, the roads are often shared by both vehicles and pedestrians, and many areas lack lane-markings that indicate permissible paths. This presents a significant challenge for both self-driving cars and delivery robots, as they must navigate these complex and unstructured environments.

Cross Modality Noise Challenge: Fig. 12 illustrates a real-world problem where robot cars must contend with corner cases of solar interference when exposed to direct sunlight, along with various reflection issues that can potentially trigger system malfunctions across multiple modalities. While these challenges are common in robotics and autonomous driving applications, there are very few existing datasets that effectively bring them to light and provide annotated data points that enable robots to learn from samples. In our sequence, we not only highlight these challenges but also provide annotations with the aim of offering labels that help machines distinguish between noise and genuine observations.

9. Research Potential

The MCD dataset comprises rich, valuable information encompassing prior maps, ground truth poses, and semantic annotations. Our dataset can be used in numerous applications, including but not limited to: multi-modality sensor fusion, SLAM, Single modality localization, Cross-Modality localization (i.e., Camera localization based on the prior 3D map), lane detection, neural radiance fields reconstructions, traversability analysis, domain adaption, perception fault analysis, image classification, object detection, object localization, semantic segmentation, image captioning, image generation, super-resolution, scene recognition, visual search, depth estimation, 3D object recognition, stereo vision, image denoising, image inpainting, image blurring, image restoration, and image style transfer.

10. Licence and Legal Compliance

Figure 11. Challenging image sequences with various lighting conditions and inferences.

search use. MCD is collected to push the frontier of robotics research. Throughout the project, we have adhered to the Personal Data Protection Act (PDPA) and General Data Protection Regulation (GDPR). To the best of our knowledge, facial blurring has been applied to all sequences, and Data Protection Officers have been appointed to oversee PDPA and GDPR compliance. The data is securely stored on Google Drive to prevent unauthorized access or disclosure. In the event of potential issues or breaches related to anonymization, security, or data access problems, we encourage concerned parties to report or request data corrections using the provided link at `mcdviral.github.io/#licence`. Remixes, transformations, or derivative works of MCD must be distributed under the original license.

References

[1] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013. 2

[2] Nicholas Carlevaris-Bianco, Arash K Ushani, and Ryan M Eustice. University of michigan north campus long-term vision and lidar dataset. *The International Journal of Robotics Research*, 35(9):1023–1035, 2016. 2

[3] Will Maddern, Geoffrey Pascoe, Chris Linegar, and Paul Newman. 1 year, 1000 km: The oxford robotcar dataset. *The International Journal of Robotics Research*, 36(1):3–15, 2017. 2

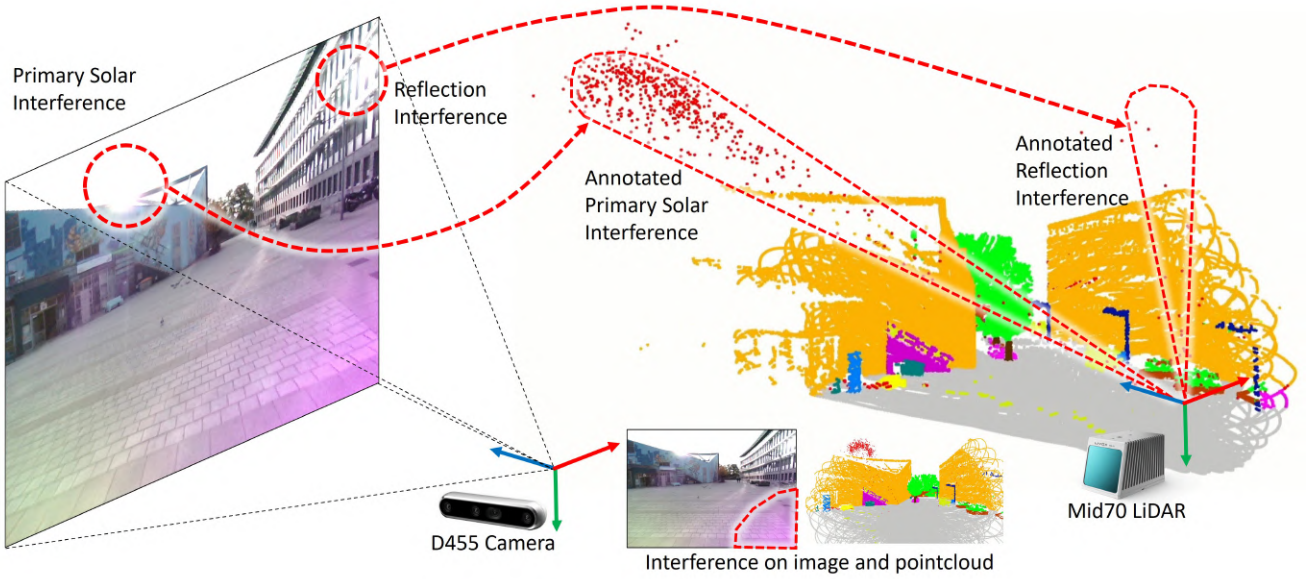

Figure 12. Problem with solar interference and reflection on image and LiDAR.

[4] Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 954–960, 2018. 2

[5] Jinyong Jeong, Younggun Cho, Young-Sik Shin, Hyunchul Roh, and Ayoung Kim. Complex urban dataset with multi-level sensors from highly diverse urban environments. *The International Journal of Robotics Research*, 38(6):642–657, 2019. 2

[6] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, et al. Argoverse: 3d tracking and forecasting with rich maps. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8748–8757, 2019. 2

[7] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, et al. Argoverse 2: Next generation datasets for self-driving perception and forecasting. *arXiv preprint arXiv:2301.00493*, 2023. 2

[8] Markus Braun, Sebastian Krebs, Fabian Flohr, and Dariu M Gavrila. Eurocity persons: A novel benchmark for person detection in traffic scenes. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1844–1861, 2019. 2

[9] Abhishek Patil, Srikanth Malla, Haiming Gang, and Yi-Ting Chen. The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 9552–9557. IEEE, 2019. 2

[10] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020. 2

[11] Weisong Wen, Yiyang Zhou, Guohao Zhang, Saman Fahandezh-Saadi, Xiwei Bai, Wei Zhan, Masayoshi Tomizuka, and Li-Ta Hsu. Urbanloco: A full sensor suite dataset for mapping and localization in urban scenes. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2310–2316. IEEE, 2020. 2

[12] Giseop Kim, Yeong Sang Park, Younghun Cho, Jinyong Jeong, and Ayoung Kim. Mulran: Multimodal range dataset for urban place recognition. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Paris, May 2020. 2

[13] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020. 2

[14] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020. 2

[15] Quang-Hieu Pham, Pierre Sevestre, Ramanpreet Singh Pahwa, Huijing Zhan, Chun Ho Pang, Yuda Chen, Armin Mustafa, Vijay Chandrasekhar, and Jie Lin. A 3d dataset: Towards autonomous driving in challenging environments. In *2020 IEEE International conference on Robotics and Automation (ICRA)*, pages 2267–2273. IEEE, 2020. 2

[16] Siddharth Agarwal, Ankit Vora, Gaurav Pandey, Wayne Williams, Helen Kourous, and James McBride. Ford multi-

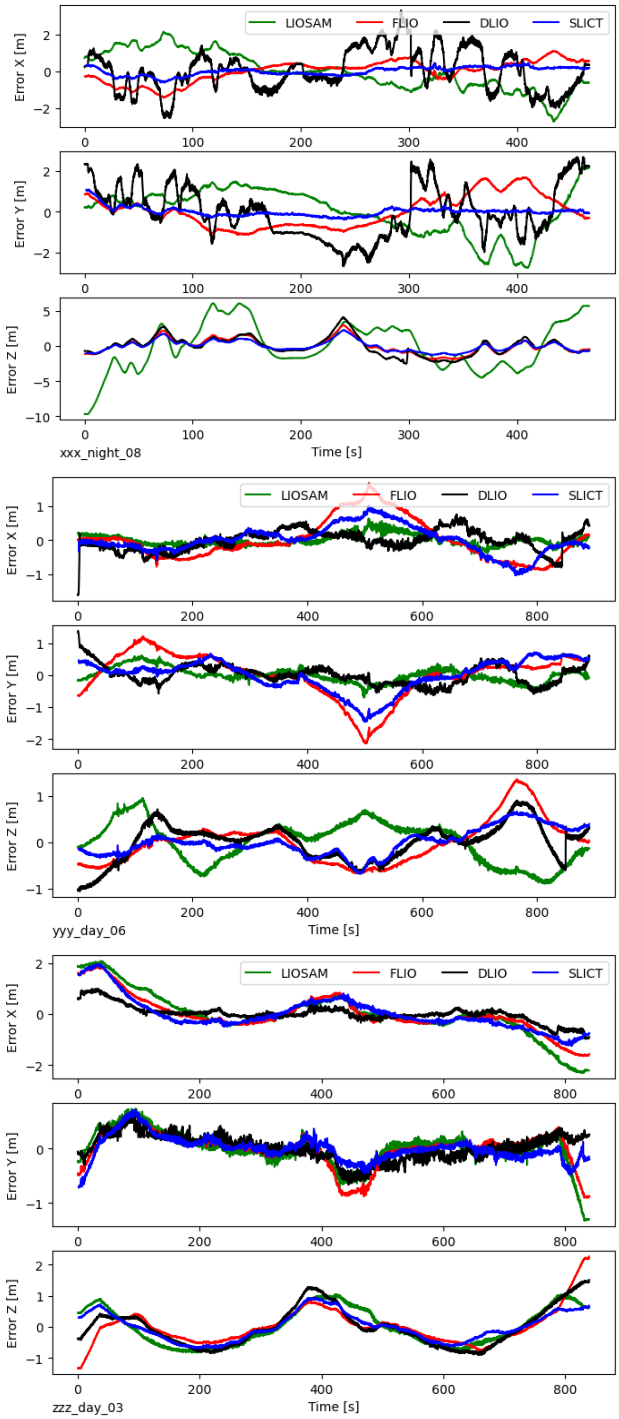

Figure 13. Error of LIO methods on the main sequences.

av seasonal dataset. *The International Journal of Robotics Research*, 39(12):1367–1376, 2020. 2

[17] Jiageng Mao, Minzhe Niu, Chenhan Jiang, Hanxue Liang, Jingheng Chen, Xiaodan Liang, Yamin Li, Chaoqiang Ye, Wei Zhang, Zhenguo Li, et al. One million scenes for autonomous driving: Once dataset. *arXiv preprint arXiv:2106.11037*, 2021. 2

[18] John Houston, Guido Zuidhof, Luca Bergamini, Yawei Ye, Long Chen, Ashesh Jain, Sammy Omari, Vladimir Iglovikov, and Peter Ondruska. One thousand and one hours: Self-driving motion prediction dataset. In *Conference on Robot Learning*, pages 409–418. PMLR, 2021. 2

[19] Matthew Pitropov, Danson Evan Garcia, Jason Rebello, Michael Smart, Carlos Wang, Krzysztof Czarnecki, and Steven Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 40(4-5):681–690, 2021. 2

[20] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, et al. Pandaset: Advanced sensor suite dataset for autonomous driving. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 3095–3101. IEEE, 2021. 2

[21] Keenan Burnett, David J Yoon, Yuchen Wu, Andrew Z Li, Haowei Zhang, Shichen Lu, Jingxing Qian, Wei-Kang Tseng, Andrew Lambert, Keith YK Leung, et al. Boreas: A multi-season autonomous driving dataset. *The International Journal of Robotics Research*, 42(1-2):33–42, 2023. 2

[22] A. Handa, T. Whelan, J.B. McDonald, and A.J. Davison. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In *IEEE Intl. Conf. on Robotics and Automation, ICRA*, Hong Kong, China, May 2014. 2

[23] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2

[24] Wenshan Wang, Delong Zhu, Xiangwei Wang, Yaoyu Hu, Yuheng Qiu, Chen Wang, Yafei Hu, Ashish Kapoor, and Sebastian Scherer. Tartanair: A dataset to push the limits of visual slam. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4909–4916. IEEE, 2020. 2

[25] Xinshuo Weng, Yunze Man, Jinhyung Park, Ye Yuan, Matthew O'Toole, and Kris M Kitani. All-in-one drive: A comprehensive perception dataset with high-density long-range point clouds. 2023. 2

[26] Jürgen Sturm, Wolfram Burgard, and Daniel Cremers. Evaluating egomotion and structure-from-motion approaches using the tum rgb-d benchmark. In *Proc. of the Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RJS International Conference on Intelligent Robot Systems (IROS)*, volume 13, 2012. 2

[27] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016. 2

[28] Shenlong Wang, Min Bai, Gellert Mattyus, Hang Chu, Wenjie Luo, Bin Yang, Justin Liang, Joel Cheverie, Sanja Fidler, and Raquel Urtasun. Torontocity: Seeing the world with a million eyes. *arXiv preprint arXiv:1612.00423*, 2016. 2

[29] Milad Ramezani, Yiduo Wang, Marco Camurri, David Wisth, Matias Mattamala, and Maurice Fallon. The newer

college dataset: Handheld lidar, inertial and vision with ground truth. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4353–4360, 2020. 2

[30] Michael Helmberger, Kristian Morin, Nitish Kumar, Danwei Wang, Yufeng Yue, Giovanni Cioffi, and Davide Scaramuzza. The hilti slam challenge dataset, 2021. 2

[31] Lintong Zhang, Michael Helmberger, Lanke Frank Tarimo Fu, David Wisth, Marco Camurri, Davide Scaramuzza, and Maurice Fallon. Hilti-oxford dataset: A millimeter-accurate benchmark for simultaneous localization and mapping. *IEEE Robotics and Automation Letters*, -1(-1):–1–1, 2022. 2

[32] Jie Yin, Ang Li, Tao Li, Wenxian Yu, and Danping Zou. M2dgr: A multi-sensor and multi-scenario slam dataset for ground robots. *IEEE Robotics and Automation Letters*, 7(2):2266–2273, 2021. 2

[33] Wenzhong Shi, Pengxin Chen, Muyang Wang, Sheng Bao, Haodong Xiang, Yue Yu, and Daping Yang. Polyu-bpcoma: A dataset and benchmark towards mobile colorized mapping using a backpack multisensorial system. *International Journal of Applied Earth Observation and Geoinformation*, 112:102962, 2022. 2

[34] Thien-Minh Nguyen, Shenghai Yuan, Muqing Cao, Yang Lyu, Thien H Nguyen, and Lihua Xie. Ntu viral: A visual-inertial-ranging-lidar dataset, from an aerial vehicle viewpoint. *The International Journal of Robotics Research*, 41(3):270–280, 2022. 2

[35] Yilin Zhu, Yang Kong, Yingrui Jie, Shiyou Xu, and Hui Cheng. Graco: A multimodal dataset for ground and aerial cooperative localization and mapping. *IEEE Robotics and Automation Letters*, 8(2):966–973, 2023. 2

[36] Ziming Wang, Yujiang Liu, Yifan Duan, Xingchen Li, Xinran Zhang, Jianmin Ji, Erbao Dong, and Yanyong Zhang. Ustc flicar: A sensors fusion dataset of lidar-inertial-camera for heavy-duty autonomous aerial work robots. *The International Journal of Robotics Research*, page 02783649231195650, 2023. 2

[37] Paul Furgale, Joern Rehder, and Roland Siegwart. Unified temporal and spatial calibration for multi-sensor systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1280–1286. IEEE, 2013. 1

[38] Christiane Sommer, Vladyslav Usenko, David Schubert, Nikolaus Demmel, and Daniel Cremers. Efficient derivative computation for cumulative b-splines on lie groups. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11148–11156, 2020. 8

[39] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019. 8, 9

[40] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15*, pages 207–222. Springer, 2020. 9

[41] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *European Conference on Computer Vision*, pages 685–702. Springer, 2020. 8, 9

[42] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9939–9948, 2021. 9

[43] Jiale Li, Hang Dai, and Yong Ding. Self-distillation for robust lidar semantic segmentation in autonomous driving. In *European Conference on Computer Vision*, pages 659–676. Springer, 2022. 8, 9

[44] Gilles Puy, Alexandre Boulch, and Renaud Marlet. Using a waffle iron for automotive point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3379–3389, October 2023. 9

[45] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17545–17555, 2023. 8, 9

[46] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9297–9307, 2019. 8

[47] Xu Yan, Jiantao Gao, Chaoda Zheng, Chao Zheng, Ruimao Zhang, Shuguang Cui, and Zhen Li. 2dpass: 2d priors assisted semantic segmentation on lidar point clouds. In *European Conference on Computer Vision*, pages 677–695. Springer, 2022. 8