Camera-LiDAR Fusion based Object Segmentation in Adverse Weather Conditions for Autonomous Driving

Junyi Gu
Dep. of Mechanical and Industrial Engineering
Tallinn University of Technology
Tallinn, Estonia
junyi.gu@taltech.ee
ORCID: 0000-0002-5976-6698

Mauro Bellone
FinEst Centre for Smart Cities
Tallinn University of Technology
Tallinn, Estonia
mauro.bellone@taltech.ee
ORCID: 0000-0003-3692-0688

Artjom Lind
Institute of Computer Science
University of Tartu
Tartu, Estonia
artjom.lind@ut.ee
ORCID: 0000-0002-8498-3547

2024 19th Biennial Baltic Electronics Conference (BEC) | 979-8-3503-6511-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/BEC61458.2024.10737955

***Abstract*— This paper presents the results of a series of machine learning experiments that focus on the domain adaption between different training datasets for autonomous driving. The experiments use a neural network model obtained through transfer learning from the Waymo Open Dataset and then tested with our custom dataset, which was recorded at the TalTech campus in different weather conditions. In this work, we developed a set of tools to extract and process the sensory data from the iseAuto shuttle. The camera and LiDAR sensors of the iseAuto shuttle were calibrated, and the point clouds data of the LiDAR sensor was projected onto the camera plane. We present, here, our publicly available iseAuto dataset which was used for a classification task under adverse weather and low illumination conditions. The iseAuto dataset contains 8000 frames of camera and LiDAR data for traffic object detection and segmentation. We provide ground truth annotations of 2400 frames, in which the object contours were manually labeled out. In addition, we provided the manual-labeled ground-truth segmentations of cars and humans in images, which can be used by the community to test the accuracy of the segmentation of their models. An additional focus of this paper is to demonstrate that with a few custom annotated data, using transfer learning and semi-supervised learning, it is possible to obtain reasonable accuracy on noisy real-world data. The current performance in vehicle segmentation ranges from 65% to 85% in intersection over union (IoU) and between 43% and 60% IoU for pedestrian segmentation in challenging scenarios such as nighttime and rainy weather.**

***Index Terms*—object segmentation, camera-LiDAR fusion, iseAuto Dataset, autonomous driving, semi-supervised learning**

I. Introduction

The popularization of autonomous driving technologies has generated strict requirements to detect and classify objects around vehicles in different visibility conditions. Vehicles rely on sensors to perceive the environment, the most common sensors in current commercial vehicles are radars and cameras, which are used in Advanced Driver Assistance Systems (ADAS) to provide fundamental functions such as collision warning and lane keeping. However, fully autonomous vehicles need to have the capability to recognize the type and assign semantic meaning to the objects, which forces researchers to seek to use other perceptive sensors. In recent years, Light Detection and Ranging (LiDAR) sensors have received the most attention in the research and industry of autonomous driving. Compared with the data acquired from other sensors (e.g., radars and cameras), the awareness of the 3D scenes is one of the most important advantages of LiDAR data. The unique features of LiDAR data, such as accurate distance measurements and rich 3D geometric, provide opportunities for autonomous vehicles to perceive the surrounding environment in three dimensions. However, LiDAR data is usually sparse and unevenly distributed, and the sensor itself is limited by the reflectivity level of the target due to the natural characteristics of all laser-based systems. To mitigate the problems of the LiDAR sensors, cameras are mainly used to complement the objects' texture density and color information. Therefore, it is necessary to effectively fuse two different modalities (LiDAR point clouds and camera images) to fully exploit the advantages of two sensors to generate more reliable and accurate semantic segmentation.

Deep learning technologies have significantly improved the camera-LiDAR fusion algorithms' performance in depth completion, 2D/3D object detection (bounding box), and semantic segmentation. The essence of the camera-LiDAR fusion for depth completion is upsampling sparse and orderless depth data to dense and ordered data. The current studies in this field can be classified by the types of camera images (monocular or stereo) used for guiding depth upsampling.

- Monocular images contain 3D geometric information in their color and gray pixel values, which can be used as

Corresponding author: Junyi Gu
This research received funding from two grants: the European Union's Horizon 2020 Research and Innovation Programme, under the grant agreement No. 856602, and the European Regional Development Fund, co-funded by the Estonian Ministry of Education and Research, under grant agreement No 2014-2020.4.01.20-0289.

a reference for depth upsampling. Moreover, there are more options fusion strategy for monocular images. The examples of early fusion (data-level fusion and feature-level fusion) are [1] [2]. Cheng et al. [1] proposed a convolution spatial propagation network to extract the affinity matrix from the RGB-D images to generate a sharp and dense depth map in real time. In [2], only sequences of RGB and sparse depth images were required for training a regression model to learn from sparse depth input to dense depth prediction. The late fusion (decision-level fusion) application for monocular images is [3], which aligned the LiDAR and camera data with a geometric model and then used Gaussian regression to complete the depth computation.

- Stereo images have richer ground truth 3D geometry in their disparity maps. Although the stereo images are limited by effective distance range and unreliable in high-occlusion and less-texture scenarios, theoretically, they are still more complementary than monocular images with the LiDAR data for depth completion. Recent deep learning work of stereo-LiDAR fusion is [4], which proposed a geometry-aware network to exploit sparse and accurate point clouds for guiding correspondences of stereo images in a unified 3D volume space.

The taxonomy according to the fusion stages (early and late) can be used for object detection methods based on the camera-LiDAR fusion.

- For early fusion methods, one-step fusion work [5] directly fed the depth map and color image into a Convolutional Neural Network (CNN) for training to detect objects. In addition, a two-step feature-level fusion example [6] first obtained Regions of Interest (ROI) by clustering the LiDAR point clouds data, then used CNN to detect objects in the corresponding image areas of the ROI.
- Compared with the early fusion, the late fusion strategy carried out the fusion of point clouds and image data at the last stages of the network [7] [8]. Therefore, late fusion methods are not sensitive to the sensor interference that happens at early stages and also work normally when a sensor fails.

Semantic segmentation has become a popular research topic in recent years because it gives a per-pixel and per-class classification, which is the direction of absolute scene understanding. The practical applications of camera-LiDAR fusion-based semantic segmentation are various, including object tracking [9], lane detection [10], and traffic sign recognition [11]. The deep architectures of most semantic segmentation algorithms for autonomous driving are CNNs, which have surpassed other approaches, such as Recurrent Neural Networks (RNNs) and Feed Forward Neural Networks (FFNs), in terms of accuracy and even efficiency.

In this work, we adopt, but leverage a Fully Convolutional Neural Network (FCN) originally proposed by the work [12] to perform 3D semantic segmentation. Exclusively, we use three submodels for different modalities, namely, camera-only, LiDAR-only, and fusion. One of the merits of such a strategy is the best-performed fusion detector is independent of the whole system, which is relatively easy to be built and implemented in practical applications.

Another contribution of this work is that we present a custom dataset recorded by the iseAuto autonomous shuttle designed and developed in the Autonomous Vehicles lab in TalTech, Estonia [13]. The vehicle was based on design toolkit [14], and early design conceptual methodology OPAS [15]. The latest public datasets, such as Waymo [16] and nuScenes [17], tended to use cutting-edge sensors and cover different weather conditions. The sensors used to produce the iseAuto dataset are Velodyne VLP-32 and FLIR Grasshopper3, which can provide dense point clouds and high-resolution (4k) images. The iseAuto dataset was primarily recorded in the TalTech campus, and covers all-weather scenarios featuring rainy and dark conditions. The dataset is available for download at (https://autolab.taltech.ee/data/). Moreover, we developed a set of tools to post-process the raw data directly from the sensors. The iseAuto shuttle was operated by the Robot Operating System (ROS); thus, LiDAR and camera were produced in ROS formats and stored as bag files. Our tools load the bag files and extract the sensory data to the formats that can be used for the neural network models. Furthermore, there are tools to calibrate the LiDAR and camera sensors, which is critical for all dataset works. With the intrinsic and extrinsic information from the calibration, it is possible to eliminate the distortion of the data and, most important, project the point clouds to the image plane. Finally, it is worth mentioning that with the help of our tools, human annotators can select the contours of the objects in images and save the segmented area with its label. When the illumination is too dark to recognize the objects in images, point clouds projection can be used as assistance for labeling. In particular, these tools were designed for universal usage. All the autonomous vehicles that rely on ROS to handle the data can use them to produce datasets that are applicable to machine learning.

The experiments of our FCN focus on supervised and semi-supervised learning. We use the Waymo Open Dataset as the reference and baseline for learning procedures. The evaluation of all models for the iseAuto dataset is based on the human-annotated frames, which was regarded as the ground truth in our case, although in most cases, human uncertainty is unavoidable in manual labeling work. Our model detects objects as semantic segmentation. Therefore, a per-pixel and per-class counting intersection of union (IoU) metric [18] was used to evaluate the model performance. This paper extracts the essential contents of our early work which was published at Electronics [19]. Exclusively, this paper focuses on our private iseAuto dataset, elaborates more details, and illustrates the dataset from a comprehensive perspective. Please refer to the work [19] for more details of experiments and evaluation criteria related to domain adaptation.

II. iseAuto Dataset

Recently, data is believed to be a valuable property. The rapid development of autonomous driving technologies has required the datasets recorded by the mainstream sensors and contain various scenarios. It is generally agreed that producing a private dataset is a challenging task for every research group, let alone maintaining the consistency of the dataset from the perspectives of sensor configuration and environmental conditions. As a result, public datasets such as Waymo and nuScenes are the first choice for researchers to test their systems. In our case, the iseAuto shuttle was supposed to deploy on the real traffic pilot case, and there are very few open datasets specifically for the low-speed shuttle bus. Furthermore, most of the public datasets are synchronized, denoised, and precisely annotated, which is unsuitable for us to evaluate the performance of the models in practical situations and real driving conditions.

The LiDAR and camera data collection of the iseAuto dataset was conducted by the primary sensors of the iseAuto shuttle, which are Velodyne VLP-32C and FLIR Grasshopper3. Velodyne VLP-32C has 32 channels, and its vertical Field of View (FoV) is 40°. The maximum range and update rate of the Velodyne VLP-32C are 200 meters and 20 Hz. For the FLIR Grasshopper3 camera, the most important specification is a maximum 4240x2824 resolution, which is the main reason we chose it as the primary camera, to ensure that the vision of the small objects is sharp and detailed even in the long range. However, the frequency of the camera is only 7 Hz when running at the maximum resolution. Therefore, synchronization was needed when projecting the point clouds to the images.

In the iseAuto dataset, both LiDAR and camera data were frame-based and presented as a single file. We differentiate the dataset by illumination (day and night) and weather (sunny and rain) conditions as four subsets: day-fair, day-rain, night-fair, and night-rain. It is important to point out that data allocation on different weather conditions is unbalanced in some works [12] [20], in which the objects are better represented in sunny daytimes than rainy nights. However, it is critical for the iseAuto shuttle to detect objects in adverse environments with high confidence, considering the general weather conditions in Estonia. Therefore, the total number of frames for each weather subset is the same in the iseAuto dataset to ensure that the knowledge that can be attained from different weather conditions is the same. To further test the model's object detection capability in poor illumination scenarios, we intentionally turn down the camera's exposure; thus, the night subset of the iseAuto dataset has a relatively lower illumination than the other datasets. Fig. 1 compares the iseAuto and Waymo datasets at a similar night scene.

The process of LiDAR data is prospectively projecting the 3D point clouds into the 2D occupancy grids; thus, the LiDAR data can be used in existing 2D convolution networks. For the iseAuto dataset, the camera was chosen as the reference frame to project LiDAR data into X, Y, and Z three planes, as

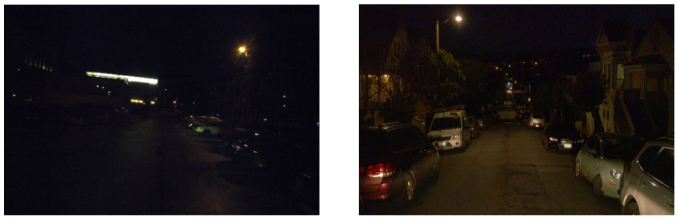


Fig. 1. Comparison of the night scene of iseAuto and Waymo datasets. (a) is the iseAuto dataset, (b) is the Waymo dataset.

shown in Fig. 2. Mathematically, the first step is to transform the point clouds data to the camera reference frame by $p^C = T_C^L p^L$, where $p^C = [x, y, z, 1]$ is a point presented in the camera reference frame, $p^L = [x, y, z]$ is the corresponding point presented in LiDAR frame. The $T_C^L \in \mathbb{R}^{4,4}$ is the transformation matrix from LiDAR to the camera, which can be obtained by using our tools to execute extrinsic camera-LiDAR calibration [21]. There is also a need to consider the camera's intrinsic information, which the rectification R and projection matrices P should be used as $[u, v, 1] = PRp^C$, $(u \in [1, h], v \in [1, w])$ represents the coordinate of each pixel in a 2D image, h and w are the height and width of the image.

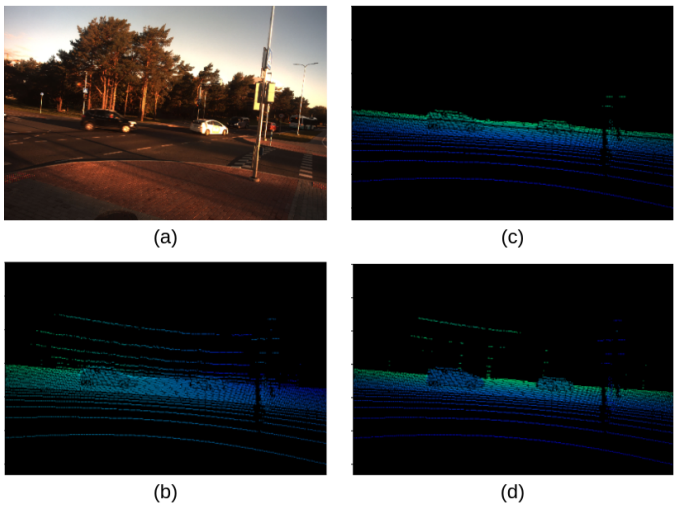
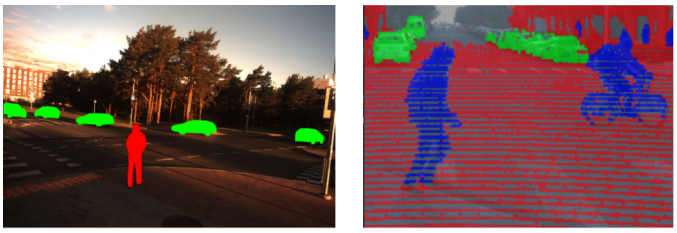


Fig. 2. The projection of the point cloud onto the camera plane in X, Y, and Z channels. (a) is RGB image, (b) is X channel, (c) is Y channel, and (d) is Z channel. The colormap of all three LiDAR projection images is the same. The color of pixels in each channel is proportionally scaled based on the numerical coordinate value of the corresponding LiDAR points. The black region means no LiDAR point falls on this image area.

The annotations of the iseAuto dataset were produced based on high-resolution images because of their rich details for the small and far-away objects. We developed a labeling tool that allows annotators to select the contours of objects in images and save the segmentation with the corresponding label. Compared with other public datasets, the most important characteristic of our annotations is that they are aware of the object contours and the semantic masks are flood-filled, which means all mask pixels belong to a unique class. Fig. 3

shows a comparison between the annotations of the iseAuto and Waymo datasets. The annotations of the Waymo dataset are based on LiDAR data; therefore, there are no-label-zones in annotations because of the sparsity of point clouds. The pixels with no label meanings must be excluded during the evaluation of the models; this would reduce the quantity of the total samples for the experiments, which is not an issue for the iseAuto dataset thanks to our solid-filled annotations.



(a) (b)

Fig. 3. The comparison of the annotations of the iseAuto and Waymo datasets. (a) is iseAuto, (b) is Waymo.

III. Experiments and Results

Our experiments explore the benefits of transfer learning and semi-supervised learning techniques from the perspective of object segmentation accuracy. Here, we use only the iseAuto dataset to train the baseline models for comparison. The external knowledge for transfer learning and semi-supervised learning is from the Waymo Open Dataset. There are 1200 frames of the iseAuto dataset used for training, which is much smaller than the Waymo dataset (16188 frames) used for transfer and semi-supervised learning. Both Waymo and iseAuto dataset splits for training, validation, and testing are 60%, 20%, and 20% of the total number of frames, respectively. Data normalization, augmentation and early stopping are also used in all training. This is to test the performance of our models for domain adaptation, which is an important capability for modern neural network models because it reduces the amount of slow manual work on data processing to review all the samples one by one. In our semi-supervised learning experiments, the unlabeled dataset was machine-annotated by best-performed Waymo-to-iseAuto transfer learning models. Then machine-annotated and human-annotated (ground truth) iseAuto data were mixed to train the transfer learning models continuously; also train the iseAuto baseline model from scratch.

The first column of Table I (iseAuto baseline) and Table II (Waymo2iseAuto TL) reports the performance of the model trained on 1200 iseAuto frames with and without transfer learning (TL) from the Waymo dataset. As expected, in this case, there is an average performance increase between 2-5 percent using transfer learning even in night and rain conditions. The best model obtained in this stage is the fusion one; see the first column of Table II (Waymo2iseAuto TL) was then used to produce the machine-made labels for the unlabeled data. Noted that there is an improvement between the second column of Table I (SSL - iseAuto baseline) and Table II (SSL - Waymo2iseAuto TL), For example, in the night-rain scenario, the vehicle class segmentation accuracy increases to 66.92% from 63.97%, and the human class increases to 48.36% from 43.63%, which proves semi-supervised learning process still uses the knowledge attained from the Waymo dataset.

TABLE I
Pixel-wise performance comparison between iseAuto supervised and semi-supervised baseline models. (in percentage unit)

		iseAuto baseline IoU(%)		SSL-iseAuto baseline IoU(%)	
		Vehicle	Human	Vehicle	Human
Day-Fair	camera	75.97	71.31	79.85	67.06
	LiDAR	71.19	56.87	73.69	58.05
	fusion	80.39	74.56	82.38	68.98
Day-Rain	camera	77.71	39.87	80.27	53.61
	LiDAR	76.00	42.10	80.58	44.09
	fusion	83.2	56.24	83.98	54.28
Night-Fair	camera	68.89	54.98	73.14	55.07
	LiDAR	74.25	47.19	75.75	49.59
	fusion	76.79	62.48	79.28	56.32
Night-Rain	camera	52.17	29.40	60.42	42.06
	LiDAR	59.49	36.76	64.89	41.32
	fusion	64.68	46.09	63.97	43.63

TABLE II
Pixel-wise performance comparison between the transfer learning models with and without semi-supervised learning. (in percentage unit)

		waymo2iseAuto TL IoU(%)		SSL-waymo2iseAuto TL IoU(%)	
		Vehicle	Human	Vehicle	Human
Day-Fair	camera	77.10	75.87	80.32	69.25
	LiDAR	72.14	55.71	76.10	61.81
	fusion	83.27	74.24	82.85	71.09
Day-Rain	camera	80.26	48.11	82.49	57.12
	LiDAR	77.33	40.27	81.00	44.85
	fusion	84.92	57.61	85.04	54.84
Night-Fair	camera	66.07	52.38	75.97	55.46
	LiDAR	74.5	45.38	76.01	51.63
	fusion	80.43	64.03	79.82	60.21
Night-Rain	camera	51.7	41.39	60.79	48.3
	LiDAR	62.51	26.46	64.4	41.15
	fusion	67.89	45.68	66.92	48.36

Inside Table I, there is a comparison of the iseAuto baseline model with and without semi-supervised learning. With the help of semi-supervised learning, there is a noticeable improvement in vehicle segmentation in difficult scenarios. For instance, from 80.39% to 82.38% and from 76.79% to 79.28% for fusion modality in day-fair and night-fair scenarios, respectively. It is still a struggle for our model to segment the human class. This can be explained by too few human samples being recorded in the iseAuto dataset, especially in the labeled split. The models are more uncertain towards the human class in semi-supervised learning after a large amount of unlabeled data was used in training. Table II provides the performance of the transfer learning model from Waymo with and without the support of semi-supervised learning, resulting in an up to 10% apparent performance increase in RGB and LiDAR modalities.

In conclusion, of all scenarios, there is a 2 to 5 percent average increase in vehicle segmentation with the help of

domain adaptation and semi-supervised learning. Notably, the average accuracy of vehicle segmentation increase from 76% (iseAuto baseline) to 79% (SSL-waymo2iseAuto TL) in fusion models for night conditions. However, it is still challenging to segment the less-represented human class, especially in the scenarios that there are not enough of them in dataset and machine-labels are too noisy.

IV. Conclusion

This work presents a custom dataset produced by the real-traffic-deployed shuttle bus. The dataset contains high-resolution RGB images and point clouds projected in the camera plane. The wide variety, in weather and illumination conditions of the iseAuto dataset, provides an opportunity for other researchers in the autonomous driving field to test their algorithms' limits in object detection and segmentation. On the other hand, we present an FCN model for object segmentation and analyze its performance in knowledge transfer from the well-known Waymo dataset to our custom dataset. It proves that domain adaptation is possible, and provides additional performance accuracy, mainly when used in challenging conditions with high uncertainty, such as rain and night conditions. Furthermore, it shows that little annotation and semi-supervised learning leveraged camera and LiDAR-based semantic segmentation performances. Future work includes testing this algorithm in different conditions, such as snow, and improving its capability for domain adaptation, probably learning from other datasets and more challenging small classes.

Acknowledgment

This research received funding from two grants: the European Union's Horizon 2020 Research and Innovation Programme, under the grant agreement No. 856602, and the European Regional Development Fund, co-funded by the Estonian Ministry of Education and Research, under grant agreement No 2014-2020.4.01.20-0289.

References

[1] X. Cheng, P. Wang, and R. Yang, "Depth estimation via affinity learned with convolutional spatial propagation network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 103–119.

[2] F. Ma, G. V. Cavalheiro, and S. Karaman, "Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 3288–3295.

[3] V. De Silva, J. Roche, and A. Kondoz, "Robust fusion of lidar and wide-angle camera data for autonomous mobile robots," *Sensors*, vol. 18, no. 8, p. 2730, 2018.

[4] J. Choe, K. Joo, T. Imtiaz, and I. S. Kweon, "Volumetric propagation network: Stereo-lidar fusion for long-range depth estimation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4672–4679, 2021.

[5] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li, "Object classification using cnn-based fusion of vision and lidar in autonomous vehicle environment," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4224–4231, 2018.

[6] H. Wang, X. Lou, Y. Cai, Y. Li, and L. Chen, "Real-time vehicle detection algorithm based on vision and lidar point cloud fusion," *Journal of Sensors*, vol. 2019, 2019.

[7] S. Pang, D. Morris, and H. Radha, "Clocs: Camera-lidar object candidates fusion for 3d object detection," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 386–10 393.

[8] W. Van Gansbeke, D. Neven, B. De Brabandere, and L. Van Gool, "Sparse and noisy lidar completion with rgb guidance and uncertainty," in *2019 16th international conference on machine vision applications (MVA)*. IEEE, 2019, pp. 1–6.

[9] W. Zhang, H. Zhou, S. Sun, Z. Wang, J. Shi, and C. C. Loy, "Robust multi-modality multi-object tracking," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2365–2374.

[10] L. Caltagirone, M. Bellone, L. Svensson, and M. Wahde, "Lidar–camera fusion for road detection using fully convolutional neural networks," *Robotics and Autonomous Systems*, vol. 111, pp. 125–131, 2019.

[11] M. Soilán, B. Riveiro, J. Martinez-Sanchez, and P. Arias, "Traffic sign detection in mls acquired point clouds for geometric and image-based semantic inventory," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 92–101, 2016.

[12] L. Caltagirone, M. Bellone, L. Svensson, M. Wahde, and R. Sell, "Lidar–camera semi-supervised learning for semantic segmentation," *Sensors*, vol. 21, no. 14, p. 4813, 2021.

[13] R. Sell, R.-M. Soe, R. Wang, and A. Rassõlkin, "Autonomous vehicle shuttle in smart city testbed," in *Intelligent System Solutions for Auto Mobility and Beyond: Advanced Microsystems for Automotive Applications 2020*. Springer, 2021, pp. 143–157.

[14] R. Sell and A. Petritsenko, "Early design and simulation toolkit for mobile robot platforms," *International Journal of Product Development*, vol. 18, no. 2, pp. 168–192, 2013.

[15] F. Christophe, R. Sell, A. Bernard, and E. Coatanéa, "Opas: Ontology processing for assisted synthesis of conceptual design solutions," in *Proceedings of the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2009, pp. 249–260.

[16] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. 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*, 2020, pp. 2446–2454.

[17] 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 the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.

[18] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.

[19] J. Gu, M. Bellone, R. Sell, and A. Lind, "Object segmentation for autonomous driving using iseauto data," *Electronics*, vol. 11, no. 7, 2022. [Online]. Available: https://www.mdpi.com/2079-9292/11/7/1119

[20] J. Zhao, H. Wu, and L. Chen, "Road surface state recognition based on svm optimization and image segmentation processing," *Journal of Advanced Transportation*, vol. 2017, 2017.

[21] J. Gu, A. Lind, T. R. Chhetri, M. Bellone, and R. Sell, "End-to-end multimodal sensor dataset collection framework for autonomous vehicles," *Sensors*, vol. 23, no. 15, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/15/6783