



Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences



Master Thesis Proposal

Out of distribution detection in 3D semantic segmentation

Lokesh Veeramacheneni

Supervised by

Prof. Dr. Paul G Plöger
Dr. Matias Valdenegro Toro
Prof. Dr. Sebastian Houben

October 2021

1 Introduction

Many robotic [38] [31], autonomous driving [25] systems deployed in the real world, dynamic environments now use LiDAR as the primary sensor. The 3D LiDAR data acquired offers a true-size replica of rich 3D geometry and can be represented as a 3D point cloud format or using 2D grids. eg: range image representation [28]. Semantic scene understanding is one of the key components of autonomous driving systems. Semantic segmentation is an important task in semantic scene understanding. Semantic segmentation requires the labelling of each data point (3D point in LiDAR/pixel in camera) with its corresponding class.

The existing models for 3D semantic segmentation are insanely complex and uncertain about their detections [6]. This uncertainty in addition to input as out of distribution (OOD) objects questions the safety and performance of the models. One such real-world example is Tesla autopilot stops after misdetecting a billboard with "STOP" written on it as a traffic stop sign [24]. Another such failure is the same autopilot detecting the moon as a yellow sign and slows the car [24]. In all the above scenarios, the model is unable to detect the object as OOD and resulting in undesired outcomes. [27], [5], [9], and [2] define OOD as the input classes which are not included in the training set. This is because most of the training sets of datasets were assumed as a closed world that is the predictions are made on the predefined set of classes. When deployed in the real world the autonomous agent will encounter a new class which leads to dangerous situations as specified in the above example which is still an unsolved problem. Figure 1 displays an example semantic space with decision boundaries between four classes in the training set and one OOD class represented in black. The training classes were vegetation, buildings, trunk and car, whereas the OOD class is a pole. During inference, the pole class is classified as trunk which is a wrong prediction. The ideal behaviour of the classifier is to specify this is an unknown class object and thus probably request for human input. In this thesis, we will try to extend the classifier to detect these unknown objects by exploring OOD detection methods and also quantified uncertainty.

A decade long study has been performed on the task of semantic segmentation, [48] refers to all the traditional approaches based on handcrafted features. With

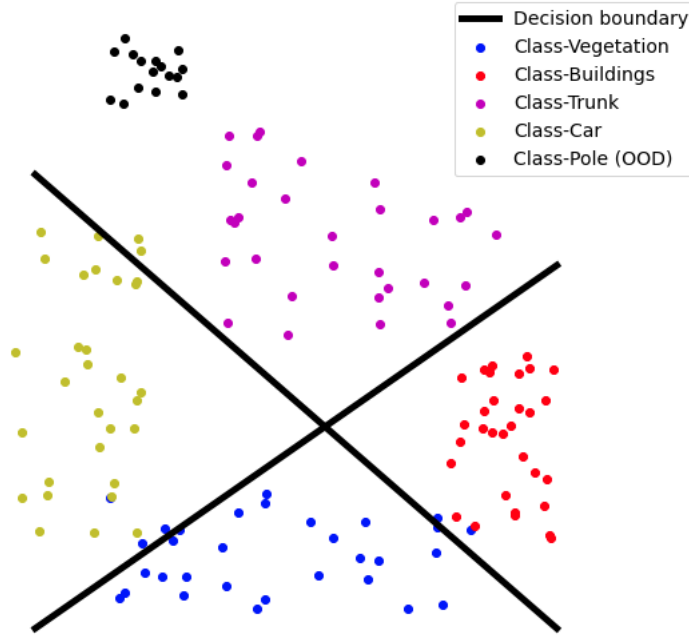


Figure 1: Representation of the semantic space with the decision boundaries of four classes in training set in colour and the points in black are of OOD

the advent of deep learning, a richer feature representation led to mapping input to semantic labels as an end to end procedure. The most popular 3D semantic segmentation datasets in the context of autonomous driving are Semantic KITTI [4], nuScenes-lidarSeg [8] and Sydney Urban [12] collected by various sensors are publicly available. 3D semantic segmentation datasets are limited in size and diversity because labelling is intensive and requires special skills to handle.

There are a diverse range of methods for the task of 3D semantic segmentation. They include point based methods such as PointNet++ [33], RandLA-Net [20], and many more. There also exists image based methods which employ various projection algorithms and popular methods in this segment include SqueezeSegV3 [43], RangeNet++ [28], and SalsaNext [11]. Graph-based and Voxel-based methods are some of their kind to solve the task of 3D semantic segmentation. In the later sections, we will discuss the problem statement, corresponding related work and project plan.

1.1 Problem Statement

In this thesis, we study the application of out of distribution (OOD) detection over the 3D semantic segmentation problem in the context of autonomous driving. Notably, we study the 3D semantic segmentation datasets available and create a benchmark for in and out distribution for the OOD setting.

The other major issue, we address in this thesis is the OOD detection methods themselves. Existing OOD detection methods are developed on 2D classification tasks and the study of these methods on 3D semantic segmentation tasks is not yet studied. This study is also challenging because the existing OOD methods are not easily adaptable to the 3D segmentation models because segmentation involves multi class classification and moreover high dimensionality of the 3D data.

The research questions answered by this thesis are:

- R1** How to create a benchmark over 3D segmentation datasets for the OOD setting? i.e., create the in and out distribution datasets.
- R2** How to extend current OOD detection methods from 2D classification task to 3D semantic segmentation?
- R3** Is uncertainty quantification an effective approach to classify the OOD detection in 3D semantic segmentation models?
- R4** How to evaluate the OOD detections over the 3D semantic segmentation task?

2 Related Work

In this related work, we discuss the methods involving 3D semantic segmentation, out of distribution (OOD) detection methods and performance metrics available.

2.1 3D semantic segmentation

The existing deep learning methods for 3D semantic segmentation is be divided into 4 types based on their mode of input. The first one is point-based methods, these

type of methods perform operations directly on the raw point clouds. These type of methods are more popular and vastly used. Some of the well-known methods in this category are PointNet [32], PointNet++ [33], RandLA-Net [20], SPLATNet [36], TangentConvolution [37], ShellNet [46], PointCNN [26].

The second most kind of deep learning method for 3D semantic segmentation is Image-based methods. These methods require the projection of the 3D point clouds into range images or multi-view images. This projection allows these methods to employ well developed and advanced 2D semantic segmentation methods like (cite). Popular methods in this category include SqueeseSegV3 [43], RangeNet++ [28], 3D-MiniNet [1], SalsaNext [11], LiSeg [45], KPRNet [21] use range image representation.

The other two kinds of methods are voxel-based methods and graph-based methods. Voxel-based methods require a small preprocessing of converting raw point clouds to voxels. Some methods using voxels are 3D U-Net [1], Cylinder3D [47]. Dynamic graph CNN [42], GAC Net [41] are few graph-based methods available.

There exists a few open 3D LiDAR semantic segmentation datasets. Oakland [29], TUM city campus [13], Semantic3D [17], Paris-lille-3D [35], S3DIS [3] are a few datasets with static LiDAR and has point annotations. The other type of datasets includes sequential data. These type of datasets include Semantic KITTI [4], A2D2 [15], nuScenes-lidarSeg [8], KITTI-360 [44], Sydney Urban [12] and they have point annotations. There also exist 3D box annotations with sequential data and they are KITTI [14], H3D [30], and DENSE [7]. GTA-V [34] and SynthCity [16] are only a few synthetic datasets available for 3D semantic segmentation.

2.2 OOD detection

The OOD detection methods are broadly divided into the threshold, non-threshold, uncertainty and generative based methods. Generative methods will not be studied in this thesis because these methods are out of the scope of the project. All the OOD methods discussed below are evaluated on the 2D classification task and no approach was deduced for the 3D semantic segmentation task.

2.2.1 Threshold based

These methods employ a threshold-based detector. The well-known OOD detector is Maximum Softmax Probability (MSP) proposed in [18]. This method considers the max of the softmax output and decides an OOD based on the confidence threshold. Since the softmax scores can be overconfident [27] proposed ODIN which uses calibrated confidence score which uses temperature scaling. Apart from calibrated confidence score, ODIN also makes use of noise perturbations in input. Combining these two, a threshold-based detector is applied at the end. [23] proposed Mahalanobis distance as a confidence score. In addition to this score, input perturbations were also added. The score is calculated at all layers then ensembled and given to threshold based detector.

2.2.2 Non-threshold based

[19] proposed Outlier Exposure which makes use of a large, unlabeled auxiliary dataset to improve the OOD detection performance. Then a new classifier is utilized with a distinct objective function. [10] proposes a novel method for improvement of robustness of OOD detection called Adversarial Learning with inlier and Outlier Exposure (ALOE) which served as an extension to Outlier Exposure proposed by [19].

2.2.3 Uncertainty based

[22] uses ensembled models to evaluate epistemic uncertainty and utilizes predictive entropy to classify the in distribution and out of distribution samples. [39] uses a novel radial basis function approach to calculate the epistemic uncertainty and argues OOD can be detected using epistemic uncertainty. The above methods evaluate the uncertainty on 2D models, there exists only [6] which apply the deep ensemble technique, Dropout and Dropconnet to the 3D semantic segmentation models to evaluate the uncertainty.

2.3 Metrics for OOD detection

There are a few metrics that deal with OOD detection and they include

2.3.1 Area Under Receiver Operating characteristic Curve (AUROC)

AUROC portrays the discrimination ability of the model between positive and negative cases. It is calculated as the area under the ROC curve which in turn is calculated from true positive rate (TPR) and false-positive rate (FPR) across various decision thresholds. The value of AUROC ranges between 0.0 and 1.0 with 0.5 being worst (random classifier) and 1.0 being best.

2.3.2 Area Under Precision-Recall Curve (AUPRC)

The area under the PR curve is calculated as AUPRC. PR curve shows the tradeoff between various thresholded precision and recall. The interpretation of AUPRC is trickier than AUROC. In AUROC the baseline is 0.5 as its random classifier but for AUPRC its fraction of positives ($\frac{\text{positive examples}}{\text{total examples}}$). This implies that various classes have various baseline AUPRC values.

2.3.3 FPR at 95% TPR

This metric calculates the FPR on OOD samples when the TPR is at 95% in distribution samples.

2.3.4 mean intersection-over-union (mIOU)

The metrics discussed till now evaluate the performance on OOD detection, but this metric is used to evaluate the performance of the 3D semantic segmentation model. It is calculated as IOU over all the classes and it is given by

$$mIOU = \frac{1}{C} \sum_{x=1}^C \frac{TP_x}{TP_x + FP_x + FN_x}$$

where C is the number of classes and TP_x, FP_x, FN_x are the true positives, false positives and false negatives of class x.

3 Project Plan

In this section, the project plan including work packages, tasks and timeline for each task is proposed.

3.1 Work Packages

The bare minimum will include the following packages:

WP1 Literature Search

- Literature search on the existing 3D datasets on semantic segmentation for benchmarking.
- Literature search on the existing 3D models for semantic segmentation.
- Literature search over the out of distribution methods.

WP2 Dataset OOD benchmark proposal

- 3D semantic segmentation dataset collection for benchmarking.
- Benchmark datasets for in-distribution and out-distribution.

WP3 Experimentation

- Extend an existing 3D semantic segmentation model as baseline model for out of distribution detection problem.
- Extend the state of the art (SOTA) 3D semantic segmentation model for out of distribution detection problem.

WP4 Evaluation

- Evaluate the baseline model and SOTA model on the proposed evaluation methods.
- Compare both the model on OOD detection

WP5 Project Report

- Write the report

3.2 Milestones

- M1 Literature search
- M2 Dataset benchmarking
- M3 Models implementation
- M4 Evaluation and comparison
- M5 Report submission

3.3 Deliverables

Minimum Viable

- Systematic literature survey of methods over
 - Datasets in 3D semantic segmentation
 - Existing out of distribution methods
 - Semantic segmentation models on 3D datasets
- Proposal of benchmarking datasets for out of distribution detection in 3D semantic segmentation
- Study of uncertainty quantification in 3D semantic segmentation models
- Implementation of method which uses quantified uncertainty for OOD detection

Expected

- Systematic evaluation of the implemented OOD detection method over the proposed dataset
- Extension of the state of the art model for OOD detection
- Evaluation and comparison of the extended state of the art model to baseline algorithm

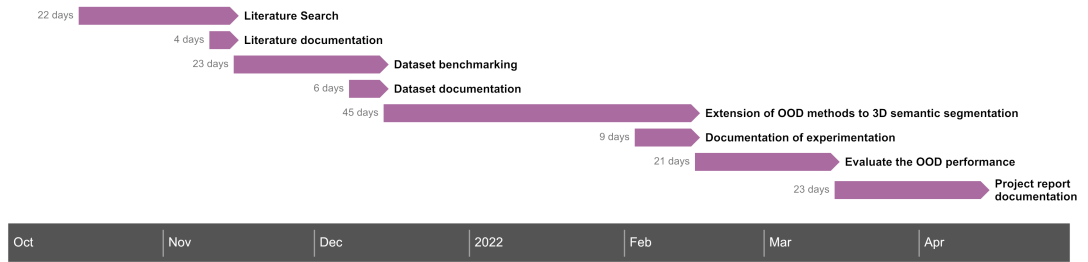
Desired

- Proposal of a refinement over the current OOD model for higher performance

3.4 Project Schedule

The detailed plan of tasks in duration of 6 months of thesis is given in figure 1.

Figure 2: Illustration of project timeline over the period of six months



References

- [1] I. Alonso, L. Riazuelo, L. Montesano, and A. C. Murillo. 3d-mininet: Learning a 2d representation from point clouds for fast and efficient 3d lidar semantic segmentation. *IEEE Robotics and Automation Letters*, 5(4):5432–5439, 2020.
- [2] M. Angus, K. Czarnecki, and R. Salay. Efficacy of pixel-level ood detection for semantic segmentation. *arXiv preprint arXiv:1911.02897*, 2019.

- [3] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [4] 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 the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.
- [5] P. Bevandić, I. Krešo, M. Oršić, and S. Šegvić. Discriminative out-of-distribution detection for semantic segmentation. *arXiv preprint arXiv:1808.07703*, 2018.
- [6] S. Bhandary K, N. Hochgeschwender, P. Plöger, F. Kirchner, and M. Valdenegro-Toro. Evaluating uncertainty estimation methods on 3d semantic segmentation of point clouds. *arXiv e-prints*, pages arXiv–2007, 2020.
- [7] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [8] 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*, pages 11621–11631, 2020.
- [9] R. Chan, M. Rottmann, and H. Gottschalk. Entropy maximization and meta classification for out-of-distribution detection in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5128–5137, October 2021.
- [10] J. Chen, Y. Li, X. Wu, Y. Liang, and S. Jha. Robust out-of-distribution detection for neural networks. *arXiv preprint arXiv:2003.09711*, 2020.

- [11] T. Cortinhal, G. Tzelepis, and E. Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In G. Bebis, Z. Yin, E. Kim, J. Bender, K. Subr, B. C. Kwon, J. Zhao, D. Kalkofen, and G. Baci, editors, *Advances in Visual Computing*, pages 207–222, Cham, 2020. Springer International Publishing.
- [12] M. De Deuge, A. Quadros, C. Hung, and B. Douillard. Unsupervised feature learning for classification of outdoor 3d scans. In *Australasian Conference on Robotics and Automation*, volume 2, page 1, 2013.
- [13] J. Gehring, M. Hebel, M. Arens, and U. Stilla. An approach to extract moving objects from mls data using a volumetric background representation. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4, 2017.
- [14] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.
- [15] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.
- [16] D. Griffiths and J. Boehm. Synthcity: A large scale synthetic point cloud. *arXiv preprint arXiv:1907.04758*, 2019.
- [17] T. Hackel, N. Savinov, L. Ladicky, J. D. Wegner, K. Schindler, and M. Pollefeys. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, 2017.
- [18] D. Hendrycks and K. Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.
- [19] D. Hendrycks, M. Mazeika, and T. Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018.

- [20] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [21] D. Kochanov, F. K. Nejadasl, and O. Booij. Kprnet: Improving projection-based lidar semantic segmentation. *arXiv preprint arXiv:2007.12668*, 2020.
- [22] B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*, 2016.
- [23] K. Lee, K. Lee, H. Lee, and J. Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.
- [24] T. Levin. Tesla’s full self-driving tech keeps getting fooled by the moon, billboards, and burger king signs, 2021.
- [25] B. Li, T. Zhang, and T. Xia. Vehicle detection from 3d lidar using fully convolutional network. *arXiv preprint arXiv:1608.07916*, 2016.
- [26] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31:820–830, 2018.
- [27] S. Liang, Y. Li, and R. Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.
- [28] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss. Rangenet ++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019.
- [29] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert. Contextual classification with functional max-margin markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 975–982, 2009.

- [30] A. Patil, S. Malla, H. Gang, and Y.-T. 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, 2019.
- [31] B. J. Patz, Y. Papelis, R. Pillat, G. Stein, and D. Harper. A practical approach to robotic design for the darpa urban challenge. *Journal of Field Robotics*, 25(8):528–566, 2008.
- [32] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [33] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.
- [34] S. R. Richter, V. Vineet, S. Roth, and V. Koltun. Playing for data: Ground truth from computer games. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, pages 102–118, Cham, 2016. Springer International Publishing.
- [35] X. Roynard, J.-E. Deschaud, and F. Goulette. Paris-lille-3d: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. *The International Journal of Robotics Research*, 37(6):545–557, 2018.
- [36] H. Su, V. Jampani, D. Sun, S. Maji, E. Kalogerakis, M.-H. Yang, and J. Kautz. Splatnet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [37] M. Tatarchenko, J. Park, V. Koltun, and Q.-Y. Zhou. Tangent convolutions for dense prediction in 3d. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

- [38] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, et al. Stanley: The robot that won the darpa grand challenge. *Journal of field Robotics*, 23(9):661–692, 2006.
- [39] J. Van Amersfoort, L. Smith, Y. W. Teh, and Y. Gal. Uncertainty estimation using a single deep deterministic neural network. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9690–9700. PMLR, 13–18 Jul 2020.
- [40] C. Wang, T. MacGillivray, G. Macnaught, G. Yang, and D. Newby. A two-stage 3d unet framework for multi-class segmentation on full resolution image. *arXiv preprint arXiv:1804.04341*, 2018.
- [41] L. Wang, Y. Huang, Y. Hou, S. Zhang, and J. Shan. Graph attention convolution for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [42] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 38(5), Oct. 2019.
- [43] B. Wu, A. Wan, X. Yue, and K. Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1887–1893, 2018.
- [44] J. Xie, M. Kiefel, M.-T. Sun, and A. Geiger. Semantic instance annotation of street scenes by 3d to 2d label transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [45] W. Zhang, C. Zhou, J. Yang, and K. Huang. Liseg: Lightweight road-object semantic segmentation in 3d lidar scans for autonomous driving. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1021–1026, 2018.
- [46] Z. Zhang, B.-S. Hua, and S.-K. Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *Proceedings of the*

IEEE/CVF International Conference on Computer Vision (ICCV), October 2019.

- [47] H. Zhou, X. Zhu, X. Song, Y. Ma, Z. Wang, H. Li, and D. Lin. Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation. *arXiv preprint arXiv:2008.01550*, 2020.
- [48] H. Zhu, F. Meng, J. Cai, and S. Lu. Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation. *Journal of Visual Communication and Image Representation*, 34:12–27, 2016.