



Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences



Master Thesis Proposal

Out of distribution detection in 3D semantic segmentation models

Lokesh Veeramacheneni

Supervised by

Dr. Paul G Plöger

Dr. Matias Valdenegro Toro

Third Supervisor

Month 2021

1 Introduction

Many robotic [36] [28], autonomous driving [22] 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 [25]. 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 [4]. 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 [21]. Another such failure is the same autopilot detecting the moon as a yellow sign and slows the car [21]. In all the above scenarios, the model is unable to detect the object as OOD and resulting in undesired outcomes.

A decade long study has been performed on the task of semantic segmentation, [46] refers to all the traditional approaches based on handcrafted features. With 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 [3], nuScenes-lidarSeg [6] and Sydney Urban [9] 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 is a diverse range of methods for the task of 3D semantic segmentation. They include point based methods such as PointNet++ [30], RandLA-Net [17], and many more. There also exists image based methods which employ various projection algorithms and popular methods in this segment include SqueezeSegV3 [41], RangeNet++ [25], and SalsaNext [8]. 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 of 3D semantic segmentation?
- R3** Is uncertainty quantification an effective approach to classify the OOD detection in 3D semantic segmentation models?
- R4** What metrics can be applied for the OOD detection task over 3D semantic segmentation models?

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 [29], PointNet++ [30], RandLA-Net [17], SPLATNet [33], TangentConvolution [35], ShellNet [44], PointCNN [23].

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 [41], RangeNet++ [25], 3D-MiniNet [1], SalsaNext [8], LiSeg [43], KPRNet [18] 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 [45]. Dynamic graph CNN [40], GAC Net [39] are few graph-based methods available.

There exists a few open 3D LiDAR semantic segmentation datasets. Oakland [26], TUM city campus [10], Semantic3D [14], Paris-lille-3D [32], S3DIS [2] 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 [3], A2D2 [12], nuScenes-lidarSeg [6], KITTI-360 [42], Sydney Urban [9] and they have point annotations. There also exist 3D box annotations with sequential data and they are KITTI [11], H3D [27], Waymo [34], and DENSE [5]. GTA-V [31] and SynthCity [13] 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 [15]. 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 [24] 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. [20] 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

[16] 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. [7] 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 [16].

2.2.3 Uncertainty based

[19] uses ensembled models to evaluate epistemic uncertainty and later used to detect if the input is from in distribution. [37] 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 [4] 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.
- Documentation.

WP2 Dataset OOD benchmark proposal

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

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.
- Documentation.

WP4 Evaluation

- Evaluate the baseline model and SOTA model on the proposed evaluation methods.

- Compare both the model on OOD detection
- Documentation

WP5 Project Report

- Update and review existing report sections (literature, datasets, experimentation and evaluation)
- Report completion and submission

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 LiDAR semantic segmentation
 - Existing out of distribution methods
 - 3D models for semantic segmentation on LiDAR data
- Proposal of 3D benchmarking datasets for out of distribution detection
- Study of uncertainty estimation over 3D models for OOD detection
- Extension of OOD detection method to a baseline 3D model

Expected

- Systematic evaluation of the implemented baseline model over the benchmarked dataset
- Extension of the state of the art model for OOD detection

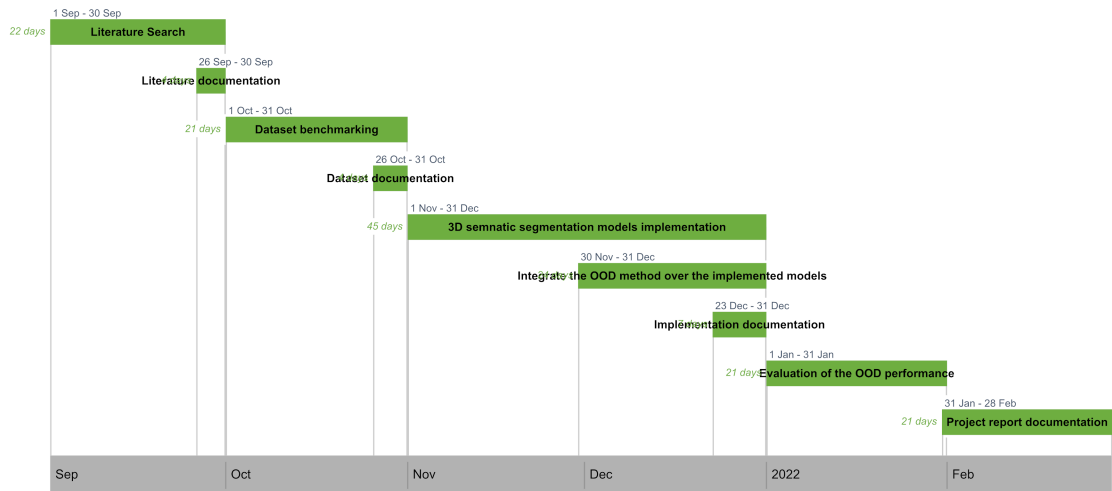
Desired

- Evaluation and comparison of the implemented state of the art model to baseline algorithm
- 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 1: Illustration of project timeline



References

- [1] Iñigo Alonso, Luis Riazuelo, Luis Montesano, and Ana 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. doi: 10.1109/LRA.2020.3007440.
- [2] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [3] 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 (ICCV)*, October 2019.
- [4] Swaroop Bhandary K, Nico Hochgeschwender, Paul Plöger, Frank Kirchner, and Matias Valdenegro-Toro. Evaluating uncertainty estimation methods on 3d semantic segmentation of point clouds. *arXiv e-prints*, pages arXiv–2007, 2020.
- [5] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix 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.
- [6] 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.
- [7] Jiefeng Chen, Yixuan Li, Xi Wu, Yingyu Liang, and Somesh Jha. Robust out-

of-distribution detection for neural networks. *arXiv preprint arXiv:2003.09711*, 2020.

- [8] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In George Bebis, Zhaozheng Yin, Edward Kim, Jan Bender, Kartic Subr, Bum Chul Kwon, Jian Zhao, Denis Kalkofen, and George Baci, editors, *Advances in Visual Computing*, pages 207–222, Cham, 2020. Springer International Publishing. ISBN 978-3-030-64559-5.
- [9] Mark De Deuge, Alastair Quadros, Calvin Hung, and Bertrand Douillard. Unsupervised feature learning for classification of outdoor 3d scans. In *Australasian Conference on Robotics and Automation*, volume 2, page 1, 2013.
- [10] Joachim Gehring, Marcus Hebel, Michael Arens, and Uwe 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.
- [11] Andreas Geiger, Philip Lenz, and Raquel 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. doi: 10.1109/CVPR.2012.6248074.
- [12] 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.
- [13] David Griffiths and Jan Boehm. Synthcity: A large scale synthetic point cloud. *arXiv preprint arXiv:1907.04758*, 2019.
- [14] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D Wegner, Konrad Schindler, and Marc Pollefeys. Semantic3d. net: A new large-scale point cloud classification benchmark. *arXiv preprint arXiv:1704.03847*, 2017.

- [15] Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136*, 2016.
- [16] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018.
- [17] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew 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.
- [18] Deyvid Kochanov, Fatemeh Karimi Nejadasl, and Olaf Booij. Kprnet: Improving projection-based lidar semantic segmentation. *arXiv preprint arXiv:2007.12668*, 2020.
- [19] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*, 2016.
- [20] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in neural information processing systems*, 31, 2018.
- [21] Tim Levin. Tesla’s full self-driving tech keeps getting fooled by the moon, billboards, and burger king signs, 2021. URL <https://www.businessinsider.com/tesla-fsd-full-self-driving-traffic-light-fooled-moon-video-2021-7>.
- [22] Bo Li, Tianlei Zhang, and Tian Xia. Vehicle detection from 3d lidar using fully convolutional network. *arXiv preprint arXiv:1608.07916*, 2016.
- [23] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31:820–830, 2018.

- [24] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017.
- [25] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet ++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019. doi: 10.1109/IROS40897.2019.8967762.
- [26] Daniel Munoz, J. Andrew Bagnell, Nicolas Vandapel, and Martial Hebert. Contextual classification with functional max-margin markov networks. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 975–982, 2009. doi: 10.1109/CVPR.2009.5206590.
- [27] 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, 2019. doi: 10.1109/ICRA.2019.8793925.
- [28] Benjamin J Patz, Yiannis Papelis, Remo Pillat, Gary Stein, and Don Harper. A practical approach to robotic design for the darpa urban challenge. *Journal of Field Robotics*, 25(8):528–566, 2008.
- [29] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas 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.
- [30] Charles R Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017.
- [31] Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 102–118, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46475-6.

- [32] Xavier Roynard, Jean-Emmanuel Deschaud, and François 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.
- [33] Hang Su, Varun Jampani, Deqing Sun, Subhransu Maji, Evangelos Kalogerakis, Ming-Hsuan Yang, and Jan Kautz. Splatnet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [34] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [35] Maxim Tatarchenko, Jaesik Park, Vladlen Koltun, and Qian-Yi Zhou. Tangent convolutions for dense prediction in 3d. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [36] Sebastian Thrun, Mike Montemerlo, Hendrik Dahlkamp, David Stavens, Andrei Aron, James Diebel, Philip Fong, John Gale, Morgan Halpenny, Gabriel Hoffmann, et al. Stanley: The robot that won the darpa grand challenge. *Journal of field Robotics*, 23(9):661–692, 2006.
- [37] Joost Van Amersfoort, Lewis Smith, Yee Whye Teh, and Yarin Gal. Uncertainty estimation using a single deep deterministic neural network. In Hal Daumé III and Aarti 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.
- [38] Chengjia Wang, Tom MacGillivray, Gillian Macnaught, Guang Yang, and David Newby. A two-stage 3d unet framework for multi-class segmentation on full resolution image. *arXiv preprint arXiv:1804.04341*, 2018.

- [39] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie 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.
- [40] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 38(5), October 2019. ISSN 0730-0301. doi: 10.1145/3326362.
- [41] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt 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. doi: 10.1109/ICRA.2018.8462926.
- [42] Jun Xie, Martin Kiefel, Ming-Ting Sun, and Andreas 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.
- [43] Wenquan Zhang, Chancheng Zhou, Junjie Yang, and Kai 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. doi: 10.1109/IVS.2018.8500701.
- [44] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit 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.
- [45] Hui Zhou, Xinge Zhu, Xiao Song, Yuexin Ma, Zhe Wang, Hongsheng Li, and Dahua Lin. Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation. *arXiv preprint arXiv:2008.01550*, 2020.
- [46] Hongyuan Zhu, Fanman Meng, Jianfei Cai, and Shijian 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. ISSN 1047-3203. doi: <https://doi.org/10.1016/j.jvcir.2015.10.012>.