Robust 3D Object Detection in Adverse Weather via Rain Noise Semantic

Xun Huang 31520221154206, Pufan Zou 31520221154244
Bochun Yang 23020221154134, Maoji Zheng 23020221154150
Class of School of information

Abstract

The unmanned loading system has the core capability of complex 3D scene perception, and its automatic loading of unmanned equipment can provide key technical support for complex tasks such as logistics supply, improving survivability, and improving equipment application efficiency. However, under the severe weather environment, the unmanned loading system still faces the challenges of missing point cloud scene data, strong data degradation and high noise, and sharp changes in lighting, which affect the detection accuracy. Based on the requirements of complex task demands and existing challenges, this project, based on the 3D scene data acquired by multiple lidars, plans to build a multi-task learning framework for severe weather sensing 3D object detection. Combined with multi-task learning to achieve object candidate generation and weather noise detection, we finally achieve the refinement of 3D object candidates for bad weather noise point perception. Through the research of this project, we enhance the semantic perception ability of the unmanned loading system to the harsh climate noise and realize the 3D object detection task in the harsh climate.

1 Introduction

3D object detection is the basic for many unmanned loading platforms such as autonomous driving, intelligent robots. These unmanned loading platforms are usually required to deal with the disturbance of various uncertain weather in the applications with high reliability requirements like national defense and disaster relief. However, severe weather usually affects the quality of data to varying degrees, resulting in a large amount of noise, which poses a challenge to the classical 3D object detection methods. Therefore, it is meaningful to construct a robust 3D object detection method to deal with the impact of uncertain data quality reduction under harsh conditions.

Some scholars [1] try to simulation the point cloud data under rain and fog conditions through physics-based methods, but it has a big gap between simulated data and real data. While others try to weaken the influence of rain and fog, such as some work [2, 3] fusion multi-modal view, but they fail when every individual view data are disturbed by noise. Some work [4,5] try to filter out the rain and fog noise generated in the point cloud data through semantic segmentation or filtering algorithms, however, this method is not directly learn to deal with the problem but depend on the quantity of denoising model.

Following the experience that introducing the semantic information to point clouds can enhance the 3D object detection model's ability to perceive object instances [6]. Similarly, we hope to introduce the weather semantic information of noise points to help increasing the model's ability to overcome the influence of noise under harsh weather. So, we introduce our PV-RCNN-RA, a two-stage 3D detection framework aiming at more accurate 3D object detection in adverse weather from point clouds.Unlike some works [4,5] try to filter out the rain and fog noise generated in the point cloud data through semantic segmentation or filtering algorithms, however, those methods are not directly learn to deal with the problem but depend on the quantity of denoising model. In state-1, we design a Multi-Task Learning Module which deal with the Weather Noise Detection and Proposal Generation by association and they share the same feature decoder. The two tasks reinforce each other. In state-2, we design a 3D Proposal Refine Module, the method combine the patch features which extract by the original 3D Box Proposal and Weather Semantic Probability Vector to get a more accuracy proposal. And, to get fully use of the Weather Semantic Probability Vector, we also introduce a new kind of loss called Whether Aware Loss, which eliminates error detection and supervises the proposal near to the outline of the vehicle.

Our contributions can be summarized into four-fold.

(1) we design a Multi-Task Learning Module to predict point-wise noise probability and get the 3D proposal box together, experiments show that those two tasks can improve each other.

(2) we propose a novel weather-aware 3D Proposal Refine Module which use the Weather Semantic Probability Vector to get more accuracy result.

(3) we introduce a new kind of loss base on the Weather Semantic Probability Vector, which eliminates error detection and supervises the proposal near to the outline of the vehicle.

(4) experiments show that PV-RCNN-RA can well improve the object detection accuracy in adverse weather on self-driving dataset Waymo.

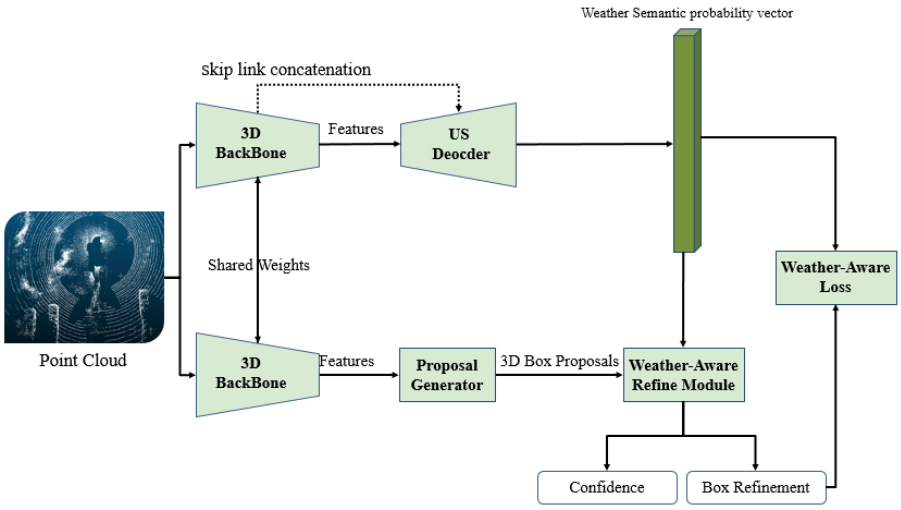

Figure 1: Overview of the proposed method. On the left is the Multi-Task Learning Module, which consists of two branches: Weather Noise Detection(top) and Proposal Generation(bottom), the two tasks share the same feature extractor. On the right is the Proposal Refine Module, the method combine the patch features which extract by the original 3D Box Proposal and Weather Semantic Probability Vector to get a more accuracy proposal. And the method also introduce a new kind of loss called Whether Aware Loss, which eliminates error detection and supervises the proposal near to the outline of the vehicle.

2 Related Work

LiDAR-based 3D object detection. Most current 3D object detection methods are two-stage strategies due to their natural advantages in accuracy. The first stage of the two-stage algorithm generates a candidate frame and the second stage optimizes the candidate frame to obtain an accurate prediction frame. PointRCNN [6] uses PointNet [7] as the backbone network to fuse semantic features and the original point cloud for the second stage of regional candidate network optimization. part-A2 Net [8] uses a voxel-based network to extract regional candidate features to optimize the ambiguity problem in PointRCNN and further improve the feature representation. Similarly, STD [9] converts semantic features obtained from region candidate networks into a representation with voxelization and reduces the number of candidate frames to improve performance. PV-RCNN [10] combines the advantages of point- and voxel-based networks, combining voxels in regions and original points to extract features. 3D-CVF [11] obtains semantic features from multi-view images and fuses them with point cloud features in two stages Frustum-PointNet [12] applies PointNetto a 3D object detection task by combining image-based 2D detection with bbox regression corresponding to point cloud data.Detection Transformer [13] takes the smash-hit Transformer [14] and cleverly adapts it to the target detection task, which improves the detection of large targets.

3D object detection in adverse weather. However, in general, classical 3D object detection methods have poor detection performance in extreme weather. Some work [2, 3] has also mitigated the problem of poor weather data scarcity to some extent by using multimodal view fusion to attenuate the effects of rain and fog noise while simulating bad weather point cloud data for data enhancement. However, these methods do not work on regions where the individual view data are disturbed by noise, and the detection performance is not stable. Some work [4, 5] has also attempted to screen out rain and fog noise generated in point cloud data by methods such as semantic segmentation or filtering algorithms to reduce the impact of noisy data on the detection model, but this method is greatly limited by the performance of the denoised model itself.

3 PV-RCNN-RA for Point Cloud 3D Detection

PV-RCNN-RA is a two-stage 3D detection framework aiming at more accurate 3D object detection in adverse weather from point clouds. In state-1, we design a Multi-Task Learning Module which deal with the Weather Noise Detection and Proposal Generation by association and they share the same feature decoder. In state-2, we design a Proposal Refine Module, the method combine the patch features which extract by the original 3D Box Proposal and Weather Semantic Probability Vector to get a more accuracy proposal. Next sections are details about our methods.

3.1.Multi-Task learning module

3D feature extracter. Voxel CNN with 3D sparse convolution [15–17]is a popular choice by state-of-the-art 3D detectors for effi-ciently converting the point clouds into sparse 3D feature volumes. Because of its high efficiency and accuracy, we adopt it as the backbone of our framework for feature encoding and 3D proposal generation.

The input points P are first divided into small voxels with spatial resolution of L × W × H, where the features of the non-empty voxels are directly calculated as the mean

Method	Vehicle(L1)		Vehicle(L2)	
	mAP	mAPH	mAP	mAPH
PV-RCNN(Baseline)	73.22	72.41	69.50	68.60
PV-RCNN-RA(Ours)	**74.57**	**73.98**	**70.92**	**70.36**
Improvement	*+1.34*	*+1.57*	*+1.42*	*+1.76*

Table 1: Performance comparison on the Waymo Open Dataset with 1000 frames rain weather point clouds for the vehicle detection.

of point-wise features (i.e., 3D coordinates, reflectance intensities) of all inside points. The network utilizes a series of 3×3×3 3D sparse convolution to gradually convert the point clouds into feature volumes with 1×, 2×, 4×, 8× down-sampled sizes. Such sparse feature volumes could be viewed as a set of voxel-wise feature vectors. And, our network has four levels with feature dimensions 16, 32, 64, 64, respectively. So the output of 3D feature extractor is the tensor with shape (N,64) where N is the count of the voxel after down-sample.

Weather noise detection. Follow the feature extractor, we design the decoder as U-Net architecture with skip link concatenation like most semantics segment task. The decoder up-sample the feature map to get a point-wise weather semantic probability vector which indicate the probability a point belong to noise introduced by adverse weather. In order to make the network correctly distinguish the noise points, we designed the following losses:

$$Loss_{seg} = -\sum_{i=1}^{m}(y_i * ln(\sigma_i) + (1 - y_i) * ln(1 - \sigma_i)) \quad (1)$$

which y_i denotes the label of the each point and σ_i denotes the output of the sigmoid function.

3D Proposal generation. By converting the encoded 8×down-sampled 3D feature volumes into 2D bird-view feature maps, high-quality 3D proposals are generated following the anchor-based approaches [18, 19]. Specifically, we stack the 3D feature volume along the Z axis to obtain the $\frac{L}{8} \times \frac{W}{8}$ bird-view feature maps. Each class has $2 \times \frac{L}{8} \times \frac{W}{8}$ 3D anchor boxes which adopt the average 3D object sizes of this class, and two anchors of 0°,90° orientations are evaluated for each pixel of the bird-view feature maps.

3.2.Weather-aware proposal refine module

3D Feature fusion. Based on the point-wise weather noise detection, we fuse the original features of the input point cloud with the noise vector detected at the point level [20].The size of the weather semantic probability vector is N × 1, and each element in the vector represents the probability that the point is predicted as noise. We fuse this with the original feature to obtain the sensed weather feature of size N × 4 for subsequent refinement of the proposal box.

Point-wise feature pooling. After obtaining the perception characteristics of the noise points at the point level, we pool the regions of interest to get the characteristics of the perceived areas of interest.The purpose of RoI pooling is for producing a fixed-size feature from an arbitrary box.In the RoI layer [21, 22],we take the 3D proposal box and the weather noise point-wise features as inputs to refine the object proposal box, and the output is the refined information of the proposal box, including the center, size, Angle, and confidence of the 3D bounding box.

3.3.Final loss design

We define three loss functions to evaluate our Multi-task learning module and Weather-aware proposal refine module. In multi-task learning, we first conducted the generation of 3D object proposal detection, and we used $L_{det}$to evaluate and optimize the learning efficiency of this task. Secondly, in the noise detection task, we separate the noise data from the original data and use L_{seg} to evaluate the de-noising rate of the data. Finally, we propose $L_{consistency}$ to evaluate the proportion of noise points in the final refined proposal.

$$Loss = L_{det} + L_{seg} + L_{consistency} \quad (2)$$

4 Expriments

Datasets. Waymo Open Dataset is a recently released and currently the largest dataset of 3D detection for autonomous driving. There are totally 798 training sequences with around 158, 361 LiDAR samples, and 202 validation sequences with 40, 077 LiDAR samples. It annotated the objects in the full 360◦ field instead of 90◦ in KITTI dataset. We evaluated our model in 1000 frames of rainy days of this large-scale dataset to verify the effectiveness of our proposed method.

Network Architecture. Same as our baseline model PV-RCNN, the 3D voxel CNN in Our PV-RCNN-RA has four levels with feature dimensions 16, 32, 64, 64, respectively. Their two neighboring radii rk of each level in the VSA module are set as (0.4m, 0.8m), (0.8m, 1.2m), (1.2m, 2.4m), (2.4m, 4.8m), and the neighborhood radii of set abstraction for raw points are (0.4m, 0.8m). For the proposed RoI-grid pooling operation, we uniformly sample 6 × 6 × 6 grid points in each 3D proposal and the two neighboring radii r˜ of each grid point are (0.8m, 1.6m).

Training and Inference Details. Our PV-RCNN-RA framework is trained from scratch in an end-to-end manner with the ADAM optimizer. For the Waymo Open Dataset, we train the entire network with batch size 2, learning rate 0.01 for 100 epochs on 1 GTX 3090 Ti GPU, which takes around 72 hours. The cosine annealing learning rate strategy

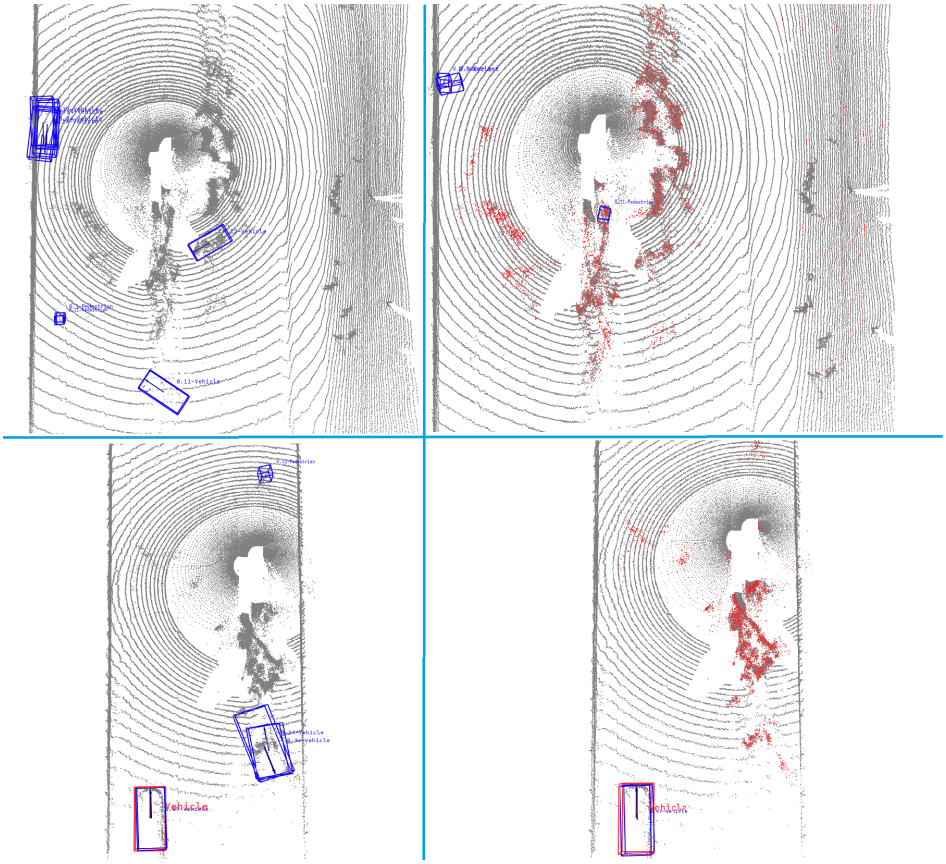

Figure 2: Visualization of PV-RCNN-RA's detection results of rainy weather point cloud data in Waymo Open Dataset. The left column shows the results of the PV-RCNN model, and the right column shows the results of our PV-RCNN-RA model.

is adopted for the learning rate decay. For the proposal refinement stage, we randomly sample 128 proposals with 1:1 ratio for positive and negative proposals, where a proposal is considered as a positive proposal for box refinement branch if it has at least 0.55 3D IoU with the ground-truth boxes, otherwise it is treated as a negative proposal.

For inference, we keep the top-100 proposals generated from the 3D voxel CNN with a 3D IoU threshold of 0.7 for non-maximum-suppression (NMS). These proposals are further refined in the proposal refinement stage with aggregated keypoint features. We finally use an NMS threshold of 0.01 to remove the redundant boxes.

4.1.Quantitative experiment of 3D detection on centralized rain weather data of Waymo Open

Evaluation Metric. We adopt the official released evaluation tools for evaluating our method, where the mean average precision (mAP) and the mean average precision weighted by heading (mAPH) are used for evaluation. The rotated IoU threshold is set as 0.7 for vehicle detection. we split the data into two difficulty levels, where the LEVEL 1 denotes the groundtruth objects with at least 5 inside points while the LEVEL 2 denotes the ground-truth objects with at least 1 inside points.

Comparison with baseline method. Table 1 shows that our method outperforms previous baseline significantly with a 1.34% mAP gain for the Vehicle(L1) object detection. The results show that our method achieves remarkably better mAP under rain weather.

4.2.Qualitative visualization experiment of 3D detection on rain weather data of Waymo Open

In this section, we will visually demonstrate how our PV-RCNN-RA works by visualizing the results. In addition, we show the comparison with the baseline model PV-RCNN in some rainy scenarios, further highlighting the effectiveness of our method.

Visualization of PV-RCNN-RA test results. As shown in Figure 3, we visualize the 3D detection box results and the rain noise semantic segmentation results of the model. For 3D detection box results, the blue rectangle box is the boundary of the detection box, and the top information is confidence and category. And for the rain noise semantic segmentation results, the points predicted as noise are displayed in red, and vice versa. Through observation, we found that although there are some misjudgments, the model can basically distinguish the noise and non noise caused by water spray. And benefiting from the predicted noise information, the model can effectively avoid interference from noise in detection: for example, misjudge the noise as the

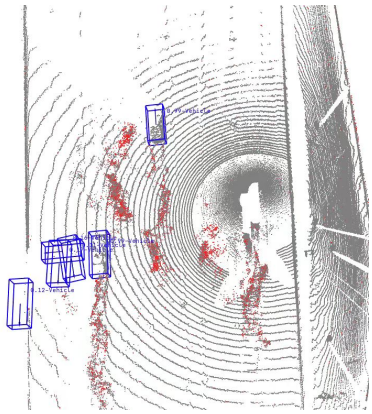

Figure 3: Visualization of PV-RCNN-RA's detection results of rainy weather point cloud data in Waymo Open Dataset

object.This can preliminarily prove the feasibility of introducing rain noise semantic segmentation into the network.
Visualization of comparison between baseline model PV-RCNN and our model PV-RCNN-RA. As in the previous part, we used the same visualization method to show the detection results of PV-RCNN model and PV-RCNN-RA model in the same rain scene. And connect the screenshots of their results together, as shown in Figure 2, to more intuitively compare the detection gap between the two models in rainy days.Through observation, compare the test results of the same line. It can be found that by using the method proposed by us and introducing rain noise semantic information and corresponding noise models into the detection model, the model can be more robust to noise interference. Specifically, the model can greatly avoid misjudging dense noise as the object, and the model can even improve the confidence level of the same object.

5 Conclusion

The simulation of severe weather data can obtain rich point cloud data of severe weather, which can alleviate the bottleneck of data driven 3D target detection research limited by the scarcity of severe weather data. At the same time, the severe weather simulation data has both the target tag and weather semantic tag of the original point cloud data, which greatly reduces the labor cost of data annotation. The introduction of weather semantic tags and the design of weather aware 3D object detection method are expected to solve the problem of interference of a large number of noises to model detection in bad weather. The project plans to generate point cloud data of severe weather 3D scene based on physical simulation, build a multi task learning framework for 3D target detection guided by weather noise information, and further improve the performance of the model in severe weather detection.

In this work, we introduce our PV-RCNN-RA, a two-stage 3D detection framework which is able to still effectively complete 3D object detection task in harsh environmental conditions, such as rain, fog, and scenes with sharp changes in light. We used a new method, which is not limited to how to remove noise directly through semantic segmentation, but to associate the noise generation and detection modules and share their feature encoder. The project plans to build a multi task collaborative learning network with weather noise detection and target candidate generation based on multi task learning. According to the feature extractor, we design the decoder as a U-Net architecture and perform skip link connection. It upsamples the feature map to obtain the point by point weather semantic probability vector, which is used to indicate the probability that the point belongs to the noise introduced by bad weather. On the basis of weather noise detection, weather noise detection at the point level is fused with the original features to obtain weather sensing features. Later, similar to most two-stage 3D target detection work, the point level weather noise point perception features are pooled to obtain weather sensing region of interest features and used to optimize the detection frame to obtain more accurate target detection frame information.

Our experimental results show that PV-RCNN-RA has a good prospect and is promising. With the addition of a large number of research, we can build a 3D object detection dataset with weather semantic tags based on real data and simulated weather data, and build a 3D target detection multi task learning framework for weather perception.

References

[1] Kilic V, Hegde D, Sindagi V, et al. Lidar Light Scattering Augmentation (LISA): Physics-based simulation of adverse weather conditions for 3D object detection. arXiv e-prints, arXiv: 2107.07004.

[2] Bijelic M, Gruber T, Mannan F, et al. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11682-11692.

[3] Mai N A M, Duthon P, Khoudour L, et al. 3D object detection with SLS-fusion network in foggy weather conditions. Sensors. 2021, 21(20): 6711.

[4] Heinzler R, Piewak F, Schindler P, et al. CNN-based LiDAR point cloud de-noising in adverse weather. IEEE Robotics and Automation Letters. 2020, 5(2): 2514-2521.

[5] Charron N, Phillips S, Waslander S L. De-noising of lidar point clouds corrupted by snowfall. Proceedings of the Conference on Computer and Robot Vision. IEEE, 2018: 254-261.

[6] Shi S, Wang X, Li H. Pointrcnn: 3D object proposal generation and detection from point cloud. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 770-779.

[7] Qi C R, Su H, Mo K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 652-660.

[8] Shi S, Wang Z, Shi J, et al. From Points to Parts: 3D object detection from point cloud with part-aware and part-aggregation network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020: 1-1.

[9] Yang Z, Sun Y, Liu S, et al. Std: Sparse-to-dense 3d object detector for point cloud. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1951-1960.

[10] Shi S, Guo C, Jiang L, et al. Pv-rcnn: Point-voxel feature set abstraction for 3D object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10529-10538.

[11] Yoo J H, Kim Y, Kim J, et al. 3D-cvf: Generating joint camera and lidar features using cross-view spatial feature fusion for 3D object detection. Proceedings of the European Conference on Computer Vision European Conference. 2020: 720-736.

[12] Qi C R , Liu W , Wu C , et al. Frustum PointNets for 3D Object Detection from RGB-D Data[J]. 2017.

[13] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), vol 12346.

[14] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.

[15] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. CVPR, 2018

[16] Benjamin Graham and Laurens van der Maaten. Submanifold sparse convolutional networks. CoRR, abs/1706.01307, 2017.

[17] Shaoshuai Shi, Zhe Wang, Jianping Shi, Xiaogang Wang, and Hongsheng Li. From points to parts: 3d object detection from point cloud with part-aware and part-aggregation network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

[18] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. Sensors, 18(10):3337, 2018.

[19] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. CVPR, 2019.

[20] Xin Li, Botian Shi, Yuenan Hou, Xingjiao Wu, Tianlong Ma, Yikang Li, Liang He. Homogeneous Multimodal Feature Fusion and Interaction for 3D Object Detection,2022.

[21] Buyu Li, Wanli Ouyang, Lu Sheng, Zeng Xingyu, Xiaogang Wang. GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving. computer vision and pattern recognition,2019

[22] Jifeng Dai, Kaiming He, Jian Sun. Instance-aware Semantic Segmentation via Multi-task Network Cascades computer vision and pattern recognition,2015.