PointDrop: Improving Object Detection from Sparse Point Clouds via Adversarial Data Augmentationn

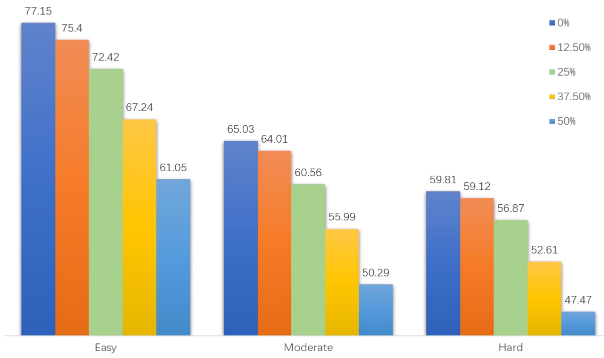
Wenxin Ma[1] , Jian Chen[1], Qing Du[1], Wei Jia[2]

[1] South China University of Technology [2] CVTE

Problem Definition

3D object detection aims to classify the object categories and estimate the oriented 3D bounding boxes of physical objects from 3D sensor data, such as point clouds.

Background

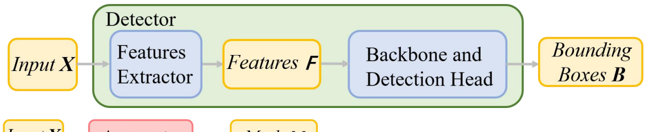


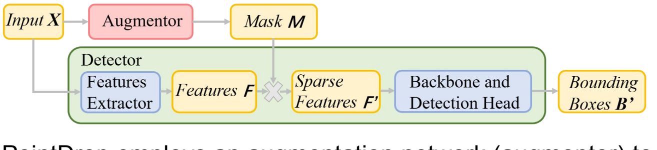
Existing 3D object detection methods have shown good performance on standard 3D object detection datasets. However, in real-world applications, due to various reasons (such as occlusion, low reflectivity of objects and fewer laser beams), the point cloud samples obtained in real-time running may be sparser. Therefore, a well-trained model may perform poorly in these situation

Contributions

- We analyze the role of critical points in 3D object detection and propose to generate point cloud samples with less critical points for data augmentation.
- We propose PointDrop, an adversarial data augmentation method in 3D object detection, which actively generates challenging sparse samples to improve the robustness of the model.
- Experimental results on two sparse point clouds datasets, which are manually created from the KITTI dataset, demonstrate the superiority of our proposed PointDrop.

Overview of PointDrop



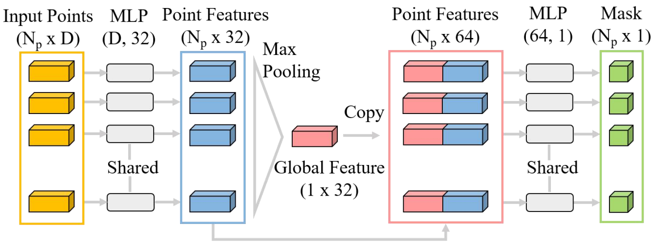


PointDrop employs an augmentation network (augmentor) to provide sparse samples and optimizes the augmentor and the detector in an adversarial way.

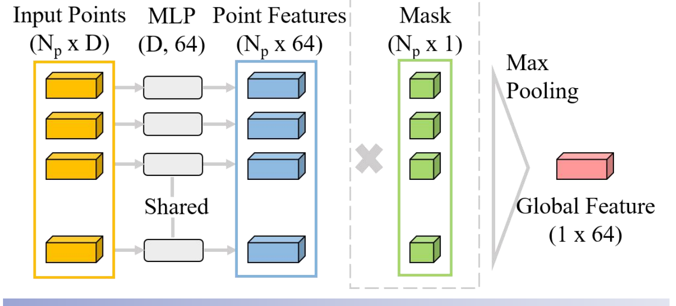
- The augmentor learns to produce hard sparse samples by dropping the features of some critical points in the original samples.
- The detector learns to handle sparse samples robustly by competing against the augmentor.
- The augmentor can adjust the difficulty of the generated sparse samples by taking the detector's loss as feedback.

Network Architectures

An illustration of how the augmentor generates a sparse mask for apillar :



An illustration of how the detector exploits the sparse mask to generatea sparse global feature for a pillar :



Loss Function

The loss for the augmentor:

$$L_A = L(X') + \lambda \,|\, 1.0 - \exp(L(X') - L(X)) \,|$$

The loss for the detector:

$$L_D = L(X) + L(X') + \gamma \, \| F_p - F_p' \|_2$$

Experiments

Result on the KITTI validation 3D detection benchmark

Category	Method	Easy		
		Sparse-0%	Sparse-25%	Sparse-50%
Car	PointPillars [9]	85.44	81.41	78.57
	PointPillars + RandomDrop	85.16	82.15	80.61
	PointPillars + PointDrop	**86.42**	**85.05**	**81.35**
Pedestrians	PointPillars	67.01	64.27	56.55
	PointPillars + RandomDrop	64.97	63.04	61.33
	PointPillars + PointDrop	**67.16**	**65.40**	**61.86**
Cyclists	PointPillars	79.00	71.58	48.04
	PointPillars + RandomDrop	79.17	78.89	71.14
	PointPillars + PointDrop	**80.83**	**80.02**	**72.03**

Result on the KITTI validation BEV detection benchmark

Category	Method	Easy		
		Sparse-0%	Sparse-25%	Sparse-50%
Car	PointPillars [9]	89.87	89.93	89.50
	PointPillars + RandomDrop	89.98	89.96	89.72
	PointPillars + PointDrop	**90.02**	**90.06**	**90.05**
Pedestrians	PointPillars	**72.53**	70.05	66.72
	PointPillars + RandomDrop	70.85	71.14	67.51
	PointPillars + PointDrop	71.41	**71.29**	**70.69**
Cyclists	PointPillars	81.88	75.40	50.97
	PointPillars + RandomDrop	82.02	81.05	73.38
	PointPillars + PointDrop	**82.59**	**81.74**	**74.20**

Alabtion Study

Augmentor	Perceptual Loss	Random Dropping	Easy		
			Sparse-0%	Sparse-25%	Sparse-50%
			77.15	72.42	61.05
✓			76.09	74.12	66.20
✓	✓		76.30	74.51	67.73
✓	✓	✓	**78.14**	**76.82**	**71.75**