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

Wenxin Ma, Jian Chen, Qing Du, Wei Jia

Presenter:Wenxin Ma
South China University of Technology

Outline

Outline

Background

Existing 3D object detection methods have shown good performance on standard 3D object detection datasets.

Problem

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 situations.

Object Detection Results on Sparse Point Clouds

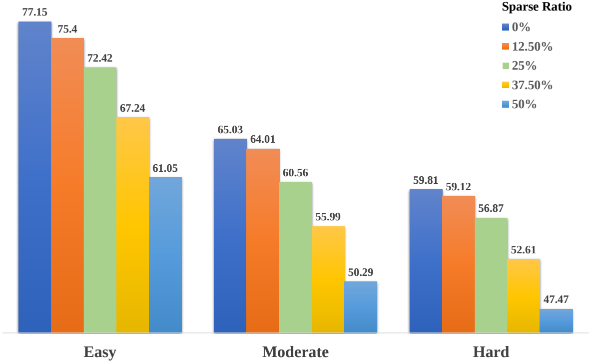


Figure: 3D object detection results (mAP) of PointPillars with different sparse ratios of point clouds on KITTI dataset. As the sparse ratio increases, the accuracy of the model decreases sharply on three difficulty levels.

Challenges

A straightforward method is to generate sparse samples by randomly discarding points in the original samples to augment the training dataset.

Problems of this method

- The point clouds obtained by random sampling may only discard some non-critical points, which is redundant for model training.
- This method separates network training and data augmentation into two independent stages.

Outline

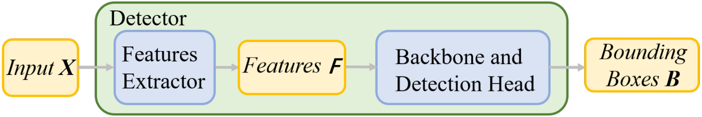
PointDrop

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

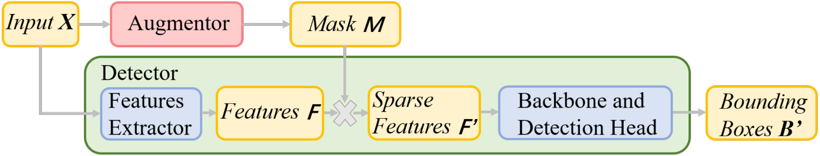
Overview

- 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.

Overview



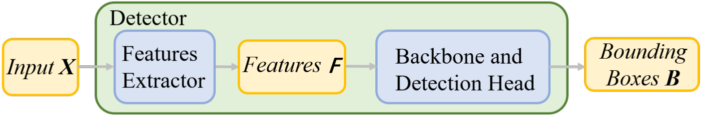
(a)



(b)

Figure: An overview of our method. Given an input sample **X**, we forward it twice to get the loss of the original sample and the sparse sample.

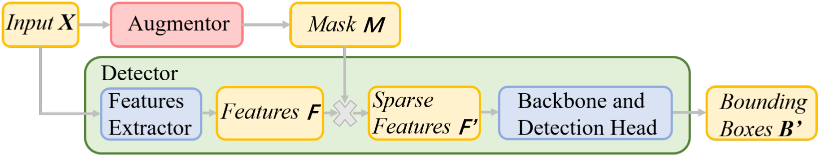
First Forward



The detector:

- extracts the features of the point clouds $\mathbf{F}$,
- further processes the extracted features,
- and finally predict bounding boxes $\mathbf{B}$ and calculate the corresponding loss $\mathbf{L(x)}$.

Second Forward



- The augmentor outputs a mask **M** according to the sample **X**.
- The features **F** are multiplied with the mask to produce sparse features **F'**.
- The detector performs the same operations on the sparse features to predict bounding boxes **B'**, and calculate the corresponding **L (x')**.

The Loss for the Augmentor

$$\mathcal{L}_{\mathcal{A}} = L(\boldsymbol{X}') + \lambda|1.0 - \exp(L(\boldsymbol{X}') - \rho L(\boldsymbol{X}))|$$

- The first term L(X') restricts the augmented sample to be shape distinctive.
- The latter term gets its minimum of zero while L(X') $=$ ρ L(X). Therefore, we can adjust $\rho \geqslant 1$ to control L(X'), thus adjusting the difficulty of the augmented samples.

The Loss for the Detector

$$\mathcal{L}_{\mathcal{D}} = L(\boldsymbol{X}) + L(\boldsymbol{X'}) + \gamma \|\boldsymbol{F}_p - \boldsymbol{F}'_p\|_2$$

- The first two terms L(X) and L(X') encourage the detector should recognize both X and X'.
- The last term is a perceptual loss to encourages the features of the augmented sample and the original one to be similar.

Network Architecture for Augmentor

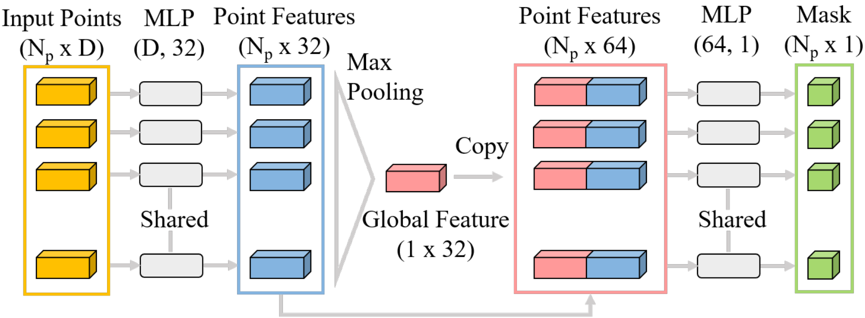


Figure: An illustration of how the augmentor generates a sparse mask for a pillar.

Network Architecture for Detector

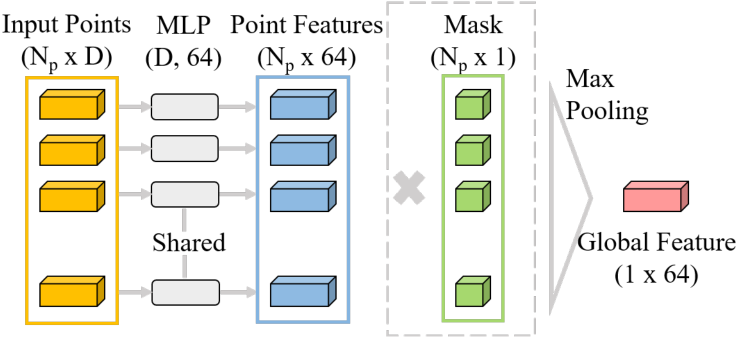


Figure: An illustration of how the detector exploits the sparse mask to generate a sparse global feature for a pillar. The operation inside dotted region is skipped when the detector aims to predict bounding boxes for original point clouds.

Outline

Results

Table: Results on the KITTI validation 3D detection benchmark.

Category	Method	Easy			Moderate			Hard		
		Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%
Car	PointPillars	85.44	81.41	78.57	76.34	70.93	67.50	70.25	68.27	65.04
	PointPillars + RandomDrop	85.16	82.15	80.61	75.65	74.35	69.77	70.33	69.28	67.21
	PointPillars + PointDrop	**86.42**	**85.05**	**81.35**	**76.57**	**75.88**	**70.70**	**70.99**	**70.40**	**68.27**
Pedestrians	PointPillars	67.01	64.27	56.55	60.83	58.77	50.34	54.79	53.37	45.45
	PointPillars + RandomDrop	64.97	63.04	61.33	58.18	58.02	56.08	53.18	53.15	51.49
	PointPillars + PointDrop	**67.16**	**65.40**	**61.86**	**61.13**	**59.67**	**59.48**	**55.63**	**54.43**	**54.36**
Cyclists	PointPillars	79.00	71.58	48.04	57.92	51.99	33.02	54.36	48.96	32.23
	PointPillars + RandomDrop	79.17	78.89	71.14	59.17	**59.13**	50.43	55.93	55.04	47.69
	PointPillars + PointDrop	**80.83**	**80.02**	**72.03**	**62.54**	58.77	**50.44**	**57.92**	**55.69**	**48.34**

Table: Results on the KITTI validation BEV detection benchmark.

Category	Method	Easy			Moderate			Hard		
		Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%
Car	PointPillars[?]	89.87	89.93	89.50	86.85	82.14	79.53	84.68	80.71	78.29
	PointPillars + RandomDrop	89.98	89.96	89.72	86.43	84.70	80.79	**85.88**	80.44	79.96
	PointPillars + PointDrop	**90.02**	**90.06**	**90.05**	**87.45**	**86.63**	**80.96**	85.84	**83.76**	**80.62**
Pedestrians	PointPillars	**72.53**	70.05	66.72	**67.36**	65.47	61.20	**62.39**	60.16	56.35
	PointPillars + RandomDrop	70.85	71.14	67.51	63.80	62.67	61.02	60.01	58.86	55.93
	PointPillars + PointDrop	71.41	**71.29**	**70.69**	66.67	**66.26**	**65.56**	61.82	**60.98**	**60.67**
Cyclists	PointPillars	81.88	75.40	50.97	61.46	55.09	36.21	57.31	52.17	34.93
	PointPillars + RandomDrop	82.02	81.05	73.38	61.76	**61.66**	**53.44**	58.83	56.81	50.19
	PointPillars + PointDrop	**82.59**	**81.74**	**74.20**	**63.96**	61.07	**53.44**	**60.61**	**58.18**	**50.85**

Ablation Study

Augmentor	Perceptual Loss	Random Dropping	Easy			Moderate			Hard		
			Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%	Sparse-0%	Sparse-25%	Sparse-50%
			77.15	72.42	61.05	65.03	60.56	50.29	59.80	56.87	47.57
✓			76.09	74.12	66.20	65.13	61.98	54.52	60.03	58.23	51.60
✓	✓		76.30	74.51	67.73	65.21	62.61	56.25	60.29	58.66	52.16
✓	✓	✓	**78.14**	**76.82**	**71.75**	**66.75**	**64.77**	**60.21**	**61.51**	**60.17**	**56.99**

Outline

Conclusions

Contributions

- 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.

Thank you for your attention!