Not All Points Are Equal: Learning Highly Efficient Point-based Detectors for 3D LiDAR Point Clouds

Yifan Zhang, Qingyong Hu, Guoquan Xu, Yanxin Ma, Jianwei Wan, Yulan Guo
National University of Defense Technology, University of Oxford

UNIVERSITÀ DEGLI STUDI DI PADOVA

ABSTRACT

- Single-stage point-based 3D detector (**IA-SSD**)
  - Reduce memory and computational costs:
    - ➢ Two-learnable, task-oriented, instance-aware downsampling
    - ➢ Contextual centroid perception
- Compare with task-agnostic sampling
- Outperform in terms of detection accuracy and runtime efficiency

PREVIOUS WORKS

- **Voxel-based** Detectors: first convert raw point clouds into intermediate regular representation

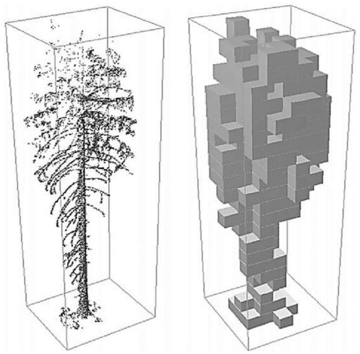
  ↓

  Good detection performance with complex modules (two- stage) to overcome quantization error but greatly increasing computational costs

- **Point-based** Detectors: directly operate on raw point clouds (PointNet or variants)

  ↓

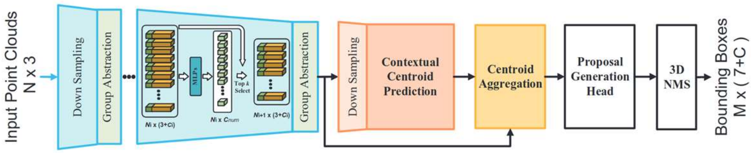
  Limited detection performance and expensive computational/memory costs due to the unstructured and orderless nature of 3d point clouds.



METHODS

The single-stage point-based 3D Detector **IA-SSD** main differences with previous point-based detectors:

1. Instance-aware sampling
2. Contextual instance centroid perception



In this paper, we argue that ***not all points are equally important to the task of object detection***. We only care about the foreground points.

1. **Feature extraction:** Set Abstraction layers
2. **Instance-aware Downsampling Strategy**: preserve foreground points
   - ***Class-aware Sampling:*** 2 MLPs used to learn semantic categories of each point
   - ***Centroid-aware Sampling:*** give higher weights to points closer to the instance centroid

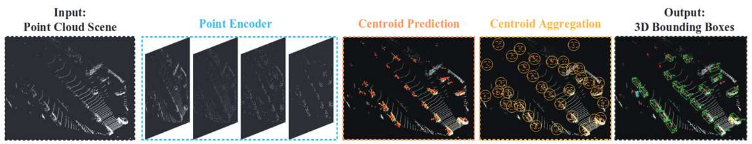
$$L_{cls} = -\sum_{i=1}^{C} (Mask_i * s_i \log(\hat{s_i}) + (1 - s_i)\log(1 - \hat{s_i}))$$

3. **Contextual Instance Centroid Perception**: exploit also surrounding points from a large context around the objects

$$L_{cent} = -\frac{1}{|F_+|}\frac{1}{|S_+|}\sum_i \sum_j (|\Delta\widehat{c_{ij}} - \Delta c_{ij}| + |\widehat{c_{ij}} - \overline{c_i}|) I_s(p_{ij})$$

4. Centroid-based Instance Aggregation: MLPs and symmetric function to learn a feature representation for each instance using the neighboring points.

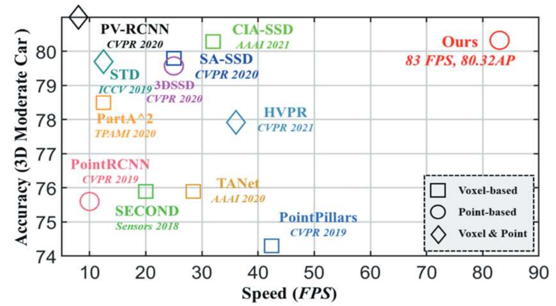
5. Proposal Generation Head: multidimensional representation of the bounding box with location, scale and orientation. Then filtering using 3D-NMS with a IoU threshold.



RESULTS

Sampling strategies	4096 points			1024 points			512 points			256 points		
	Car	*Ped.*	*Cyc.*	*Car*	*Ped.*	*Cyc.*	*Car*	*Ped.*	*Cyc.*	*Car*	*Ped.*	*Cyc.*
Random [14]	96.6%	99.1%	97.4%	87.5%	92.7%	84.1%	78.8%	84.9%	73.3%	67.4%	72.1%	57.3%
D-FPS [32]	98.3%	100%	97.2%	97.9%	99.3%	97.2%	96.8%	90.6%	90.8%	91.4%	69.1%	71.6%
Feat-FPS [52]	98.3%	100%	97.2%	97.7%	98.0%	97.2%	96.3%	87.6%	94.5%	95.3%	80.1%	91.7%
Cls-aware (Ours)	98.3%	100%	97.2%	97.9%	99.3%	97.2%	97.9%	99.0%	95.4%	97.9%	97.4%	92.7%
Ctr-aware (Ours)	98.3%	100%	97.2%	97.9%	99.3%	97.2%	97.9%	99.0%	97.2%	97.9%	98.4%	97.2%

	Method	Reference	Type	3D Car (IoU=0.7)			3D Ped. (IoU=0.5)			3D Cyc. (IoU=0.5)			Speed
				Easy	*Mod.*	*Hard*	*Easy*	*Mod.*	*Hard*	*Easy*	*Mod.*	*Hard*	
Voxel-based	VoxelNet [61]	CVPR 2018	1-stage	77.47	65.11	57.73	39.48	33.69	31.5	61.22	48.36	44.37	4.5
	SECOND [49]	Sensors 2018	1-stage	84.65	75.96	68.71	45.31	35.52	33.14	75.83	60.82	53.67	20
	PointPillars [18]	CVPR 2019	1-stage	82.58	74.31	68.99	51.45	41.92	38.89	77.10	58.65	51.92	42.4
	3D IoU Loss [59]	3DV 2019	1-stage	86.16	76.50	71.39	-	-	-	-	-	-	12.5
	Associate-3Ddet [7]	CVPR 2020	1-stage	85.99	77.40	70.53	-	-	-	-	-	-	20
	SA-SSD [11]	CVPR 2020	1-stage	88.75	79.79	74.16	-	-	-	-	-	-	25
	CIA-SSD [58]	AAAI 2021	1-stage	89.59	80.28	72.87	-	-	-	-	-	-	32
	TANet [24]	AAAI 2020	2-stage	84.39	75.94	68.82	53.72	<u>44.34</u>	40.49	75.70	59.44	52.53	28.5
	Part-A^2 [39]	TPAMI 2020	2-stage	87.81	78.49	73.51	53.10	43.35	40.06	79.17	63.52	56.93	12.5
Point-Voxel	Fast Point R-CNN [3]	ICCV 2019	2-stage	85.29	77.40	70.24	-	-	-	-	-	-	16.7
	STD [53]	ICCV 2019	2-stage	87.95	79.71	75.09	53.29	42.47	38.35	78.69	61.59	55.30	12.5
	PV-RCNN [36]	CVPR 2020	2-stage	<u>90.25</u>	<u>81.43</u>	<u>76.82</u>	52.17	43.29	40.29	78.60	63.71	<u>57.65</u>	12.5
	VIC-Net [16]	ICRA 2021	1-stage	88.25	80.61	75.83	43.82	37.18	35.35	78.29	63.65	57.27	17
	HVPR [29]	CVPR 2021	1-stage	86.38	77.92	73.04	53.47	43.96	<u>40.64</u>	-	-	-	36.1
Point-based	PointRCNN [38]	CVPR 2019	2-stage	86.96	75.64	70.70	47.98	39.37	36.01	74.96	58.82	52.53	10
	3D IoU-Net [19]	Arxiv 2020	2-stage	87.96	79.03	72.78	-	-	-	-	-	-	10
	Point-GNN [40]	CVPR 2020	1-stage	88.33	79.47	72.29	51.92	43.77	40.14	78.60	63.48	57.08	1.6
	3DSSD [52]	CVPR 2020	1-stage	88.36	79.57	74.55	<u>54.64</u>	44.27	40.23	<u>82.48</u>	64.10	56.90	25
	3DSSD† (Reproduced)	CVPR 2020	1-stage	87.73	78.58	72.01	35.03	27.76	26.08	66.69	59.00	55.62	23
	3DSSD‡ (OpenPCDet)	CVPR 2020	1-stage	87.91	79.55	74.71	3.63	3.18	2.57	27.08	21.38	19.68	28
	IA-SSD (single-class)	-	**1-stage**	**88.87**	**80.32**	**75.10**	**49.01**	**41.20**	**38.03**	**80.78**	<u>**66.01**</u>	<u>**58.12**</u>	<u>**85**</u>
	IA-SSD (multi-class)	-	**1-stage**	**88.34**	**80.13**	**75.04**	**46.51**	**39.03**	**35.60**	**78.35**	**61.94**	**55.70**	<u>**83**</u>



LIMITATIONS

The instance-aware downsampling relies on the semantic prediction of each point, which is susceptible to class imbalances distribution.