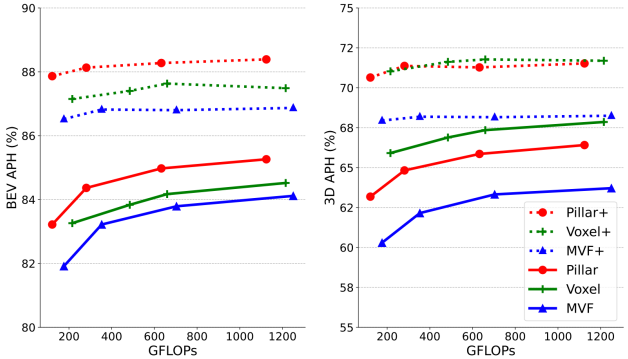
QCRAFT

PillarNeXt: Rethinking Network Designs for 3D Object Detection in LiDAR Point Clouds

Jinyu Li Chenxu Luo Xiaodong Yang

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

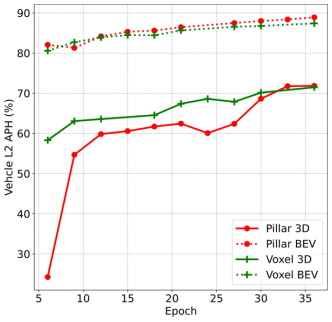
Do We Need Sophisticated Local Point Aggregators?



Overview of pillar, voxel and multi-view fusion (MVF) based 3D object detection networks under different GFLOPs. The dash lines denote the enhanced versions of corresponding models (+). We report the L2 BEV and 3D APH of vehicle on the validation set of Waymo Open Dataset (WOD).

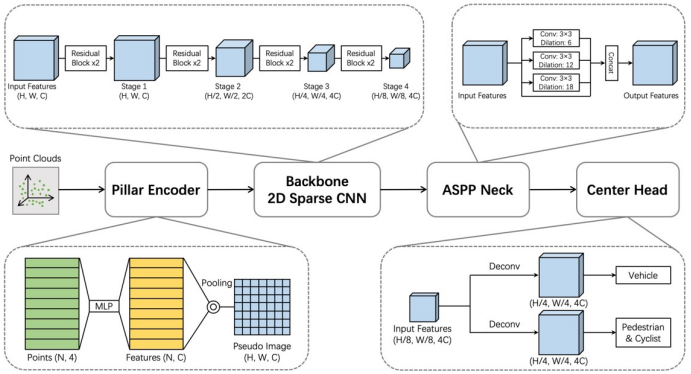
Training Matters

Learning behaviors of the pillar based and the voxel based models. We report the L2 3D and BEV APH of vehicle on the validation set on WOD.



Code Available

Network Architecture



A schematic overview of the network architecture of the proposed PillarNeXt.

Study of Neck Modules

Method	Vehicle L1		Vehicle L2		Pedestrian L1		Pedestrian L2	
	AP	APH	AP	APH	AP	APH	AP	APH
Neck of PillarNet [31]	91.39	90.58	84.54	83.72	**87.90**	**83.02**	81.93	77.20
FPN [17]	92.17	91.35	85.96	85.13	87.88	82.91	82.05	77.23
BiFPN [39]	92.71	91.90	86.92	86.09	87.86	82.88	82.05	77.23
Plain	91.01	90.19	83.86	83.04	87.59	82.61	81.52	76.71
Dilated Block [7]	92.70	91.90	86.61	85.79	87.84	82.91	**82.09**	**77.29**
ASPP [5]	**92.77**	**91.94**	**86.99**	**86.14**	87.74	82.85	82.00	77.26

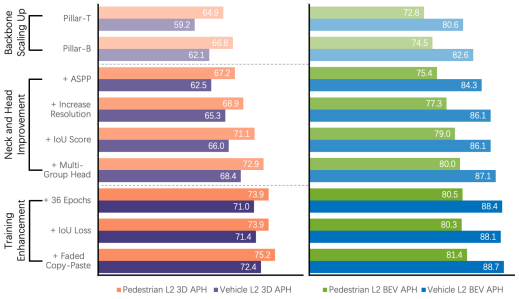
Comparison of different neck modules integrated in our networks. Groups 1 and 2 correspond to the multi-scale and sing-scale necks under the BEV metrics.

Study of Resolutions

Comparison of different resolutions by pillar size (m) of input grids and output features to head.

In Size	Backbone ↓	Head ↑	Out Size	Veh	Ped	Latency
0.3	1	1	0.3	65.0	67.2	255
0.075	8	1	0.6	62.8	66.6	131
0.075	8	2	0.3	64.8	69.0	173

Boosting Roadmap



Experimental Results

Method	Frames	Vehicle L1		Vehicle L2		Pedestrian L1		Pedestrian L2		Cyclist L1		Cyclist L2	
		AP	APH	AP	APH	AP	APH	AP	APH	AP	APH	AP	APH
PillarNet-18 [31]	2	79.59	79.06	71.56	71.08	82.11	78.82	74.49	71.35	70.41	69.57	68.27	67.46
PillarNet-34 [31]	2	79.98	79.47	72.00	71.53	82.52	79.33	75.00	71.95	70.51	69.69	68.38	67.58
PV-RCNN++* [32]	2	80.17	79.70	72.14	71.70	83.48	80.42	75.54	72.61	74.63	73.75	72.35	71.50
RSN* [37]	3	78.4	78.1	69.5	69.1	79.4	76.2	69.9	67.0	-	-	-	-
SST-TS* [11]	3	78.66	78.21	69.98	69.57	83.81	80.14	75.94	72.37				
SWFormer [36]	3	79.4	78.9	71.1	70.6	82.9	79.0	74.8	71.1	-	-	-	-
PillarNeXt-B	3	**80.58**	**80.08**	**72.89**	**72.42**	**85.04**	**82.11**	**78.04**	**75.19**	**78.92**	**77.94**	**76.71**	**75.74**
CenterFormer [49]	8	78.8	78.3	74.3	73.8	82.1	79.3	77.8	75.0	75.2	74.4	73.2	72.3
MPPNet [8]	16	82.74	82.28	75.41	74.96	84.69	82.25	77.43	75.06	77.28	76.66	75.13	74.52
3DAL† [29]	ALL	84.50	-	-	-	82.88	-	-	-	-	-	-	-

Method	Frames	Vehicle L1		Vehicle L2		Pedestrian L1		Pedestrian L2		Cyclist L1		Cyclist L2	
		AP	APH	AP	APH	AP	APH	AP	APH	AP	APH	AP	APH
PV-RCNN++* [32]	1	91.57	-	-	-	85.43	-	-	-	75.94	-	-	-
PillarNeXt-B	1	93.30	92.60	87.26	86.53	88.19	82.13	81.77	75.82	75.67	74.61	72.97	71.95
SWFormer [36]	3	92.60	-	-	-	87.50	-	-	-	-	-	-	-
PillarNeXt-B	3	**94.41**	93.73	89.36	88.66	**90.20**	86.94	84.66	81.36	81.35	80.32	79.23	78.22
3DAL† [29]	ALL	93.30	-	-	-	86.32	-	-	-	-	-	-	-

Comparisons under the 3D (top) and BEV (bottom) metrics on the validation set of WOD.

Method	Encoder	Grid Size	NDS	mAP	mATE↓	mASE↓	mAOE↓	mAVE↓	mAAE↓
CenterPoint [45]	V	0.075	66.8	59.6	0.292	0.255	0.302	0.259	0.193
OHS [6]	V	0.1	66.0	59.5	-	-	-	-	-
PillarNet-18 [31]	P	0.075	67.4	59.9	-	-	-	-	-
Transfusion-L [1]	V	0.075	66.8	60.0	-	-	-	-	-
UVTR-L [15]	V	0.075	67.7	60.9	0.334	0.257	0.300	0.204	0.182
VISTA [9]	V+R	0.1	68.1	60.8	-	-	-	-	-
PillarNeXt-B	P	0.075	**68.8**	**62.5**	0.278	0.251	0.269	0.248	0.201
Our Voxel-B	V	0.075	68.2	62.4	0.278	0.250	0.308	0.263	0.198

Comparisons on the validation set of nuScenes.