

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

SUPPLEMENTARY MATERIALS

Jinyu Li Chenxu Luo Xiaodong Yang*
QCraft

1. More Implementation Details

We use 0.01 weight decay in the optimizer AdamW, and set the one-cycle scheduled learning rate to 0.001 when the batch size is 16. Our basic augmentations include random flipping, random rotation within the range of $[-\frac{\pi}{4}, \frac{\pi}{4}]$, random scaling between $[0.9, 1.1]$, and random translation with a noise factor of 0.5, which are used for all experiments. Table 9 shows the detailed architecture and training settings used in various experiments. As can be seen in this table, for the experiments on WOD, “Study” denotes the regular setting in Section 4.2, and “Study+” indicates the enhanced version. Pillar/Voxel+ are trained with such enhanced strategy. We use the full settings with additional faded copy-and-paste augmentation when comparing with the single-frame or multi-frame based state-of-the-art methods.

2. Qualitative Results

Figure 5 further provides the qualitative comparison of different neck modules integrated in our pillar based model. In this figure, the first row illustrates the detection results of using the multi-scale neck proposed in PillarNet, and the second row demonstrates the detection results of using the single-scale ASPP neck. It is observed that the first model fails to detect large vehicles, while after enlarging the receptive field, the second model is capable of detecting the two large vehicles accurately. In the two cases, we find that an additional second-stage refinement of a two-stage method also struggles to repair such failures.

*Corresponding author xiaodong@qcrafter.ai

	Waymo Open Dataset (WOD)				nuScenes
	Study	Study+	Full (Single-Frame)	Full (Multi-Frame)	
Flip/Scaling/Rotation/Translation	✓	✓	✓	✓	✓
Random Drop Frames	✓	✓		✓	
Faded Copy-and-Paste			✓	✓	✓
IoU Regression Loss		✓	✓	✓	✓
IoU Score Branch		✓	✓	✓	
Multi-Group Head		✓	✓	✓	✓
Epochs	12	36	36	36	20

Table 9. Details of the experimental settings for different experiments on the two benchmark datasets.

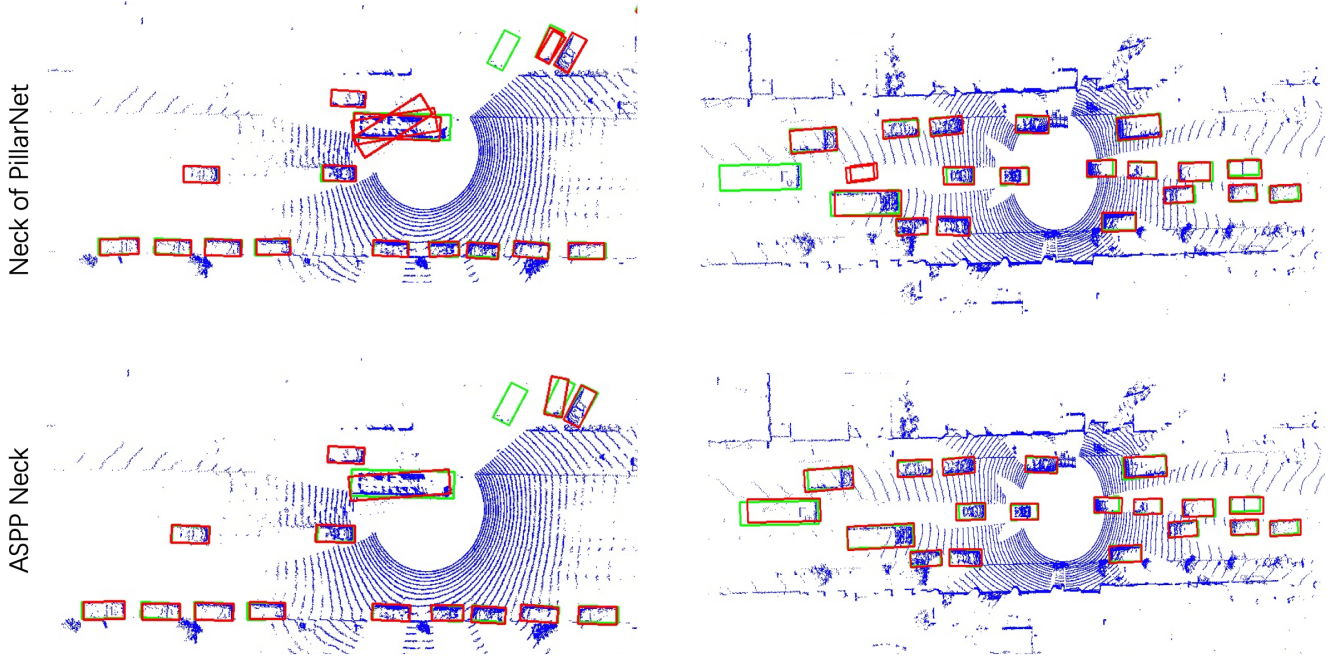


Figure 5. Comparison of the qualitative detection results of using different necks in our model on the validation set of WOD. The first and second rows correspond to the results based on the necks of PillarNet and ASPP, respectively. We use the green and red boxes to denote the ground-truth annotations and model predictions.