A Pillar-based Lightweight 3D Detector with Density-Enhanced Pillar Encoder

Shuyu Ji, Shanzhu Xiao, Huamin Tao, Qiuqun Deng
College of Electronic Science and Technology, National University of Defense Technology

Introduction

In the field of 3D object detection, complex architectures and large parameters hinder edge deployment. We propose a lightweight, real-time neural network with three key components. The Density-Enhanced Pillar Encoding module (DEPE) resolves data sparsity by expanding pillar perception. Our backbone, using reparameterization and depth-wise separable convolutions, avoids sparse operations, easing edge deployment. Evaluated on KITTI dataset, our network achieves 1.87M parameters and 18Hz inference speed, balancing efficiency and accuracy.

2. Density-Enhanced Pillar Encoder

The widely adopted pillar encoding structure originates from VoxelNet and primarily aggregates features through operations such as MLP (Multi-Layer Perceptron), concatenation, and max pooling. Specifically, point-wise MLP is used to extract features from each point, and element-wise max pooling is employed to aggregate features from all points within a pillar. However, a significant number of pillars contain only a single point. In such cases, element-wise max pooling becomes meaningless and fails to effectively aggregate features, leading to computational waste, which is due to the uneven density of point clouds.

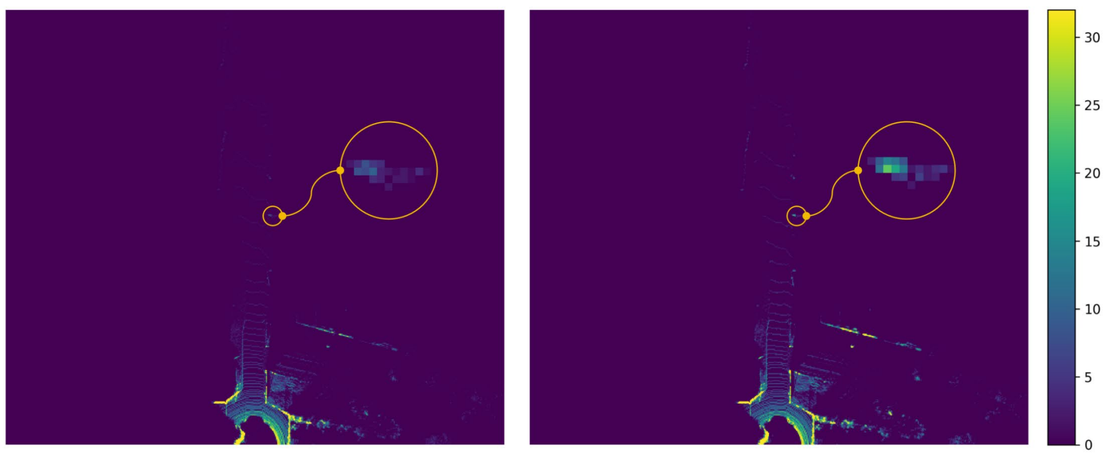

FIG 2. COMPARISON BEFORE AND AFTER DENSITY ENHANCEMENT.

Therefore, we introduce a Density-Enhanced Pillar Encoding module based on density prior information. Specifically, we use average density information to assign different grid sizes to pillars at varying distances, with larger grids allocated to more distant pillars to encompass more potential neighboring points and the grids overlap with each other. By doing so, we expand the perceptual range of distant pillars, reduce the number of single-point pillars, and more fully utilize the feature extraction capabilities of the PFE module.

Network Architecture

1. The overall architecture of network

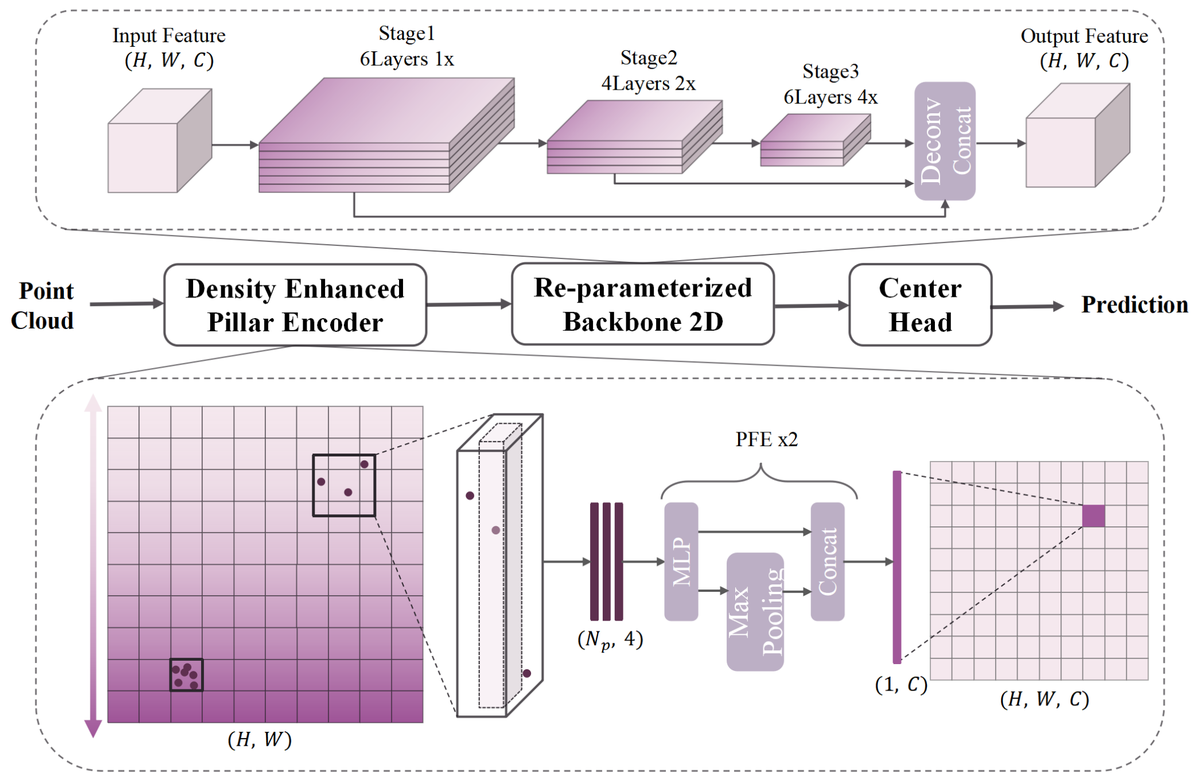

FIG 1. THE OVERALL ARCHITECTURE OF NETWORK

We proposes a lightweight and real-time point cloud object detection neural network based on the pillar structure, which is primarily composed of three components: a density-enhanced encoder, a re-parameterization backbone, and a center-based detection head.

3. Re-parameterized Lightweight Backbone

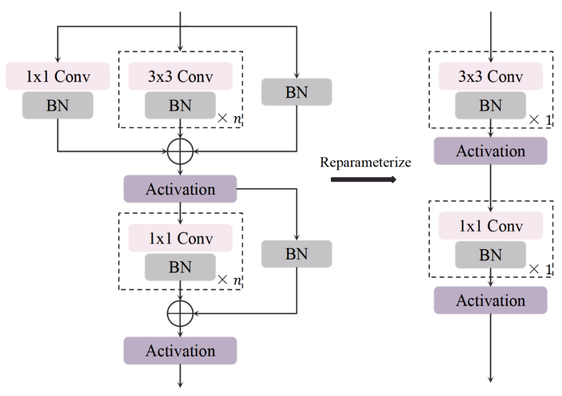

FIG 3. RE-PARAMETERIZED MOBILEONE BLOCK.

Backbone plays a crucial role in extracting inter-pillar features from the pseudo-image generated by the PFE module, constituting the primary component of neural network feature extraction, which is also the part of the network that demands significant computational resources and time.

We construct the backbone using the MobileOne module, which applies the reparameterization method based on the MobileNetV1 structure, employs depth-wise separable convolutions to reduce the network's parameter count, and adds reparameterization branches to enhance the model's feature extraction capabilities.

Experiments

TABLE I. PERFORMANCE COMPARISONS ON KITTI DATASET

Method	mAP	3D Car (IoU=0.7)			3D Ped. (IoU=0.5)			3D Cyc. (IoU=0.5)			Param.
	(Mod.)	Easy	Mod.	Hard	Easy	Mod.	Hard	Easy	Mod.	Hard	(M)
VoxelNet	49.05	77.47	65.11	57.73	39.48	33.69	31.50	61.22	48.36	44.37	6.41
SECOND	57.43	84.65	75.96	68.71	45.31	35.52	33.14	75.83	60.82	53.67	5.33
PointPillars	58.29	82.58	74.31	68.99	51.45	41.92	38.89	77.10	58.65	51.92	4.83
CenterPoint	59.96	88.21	79.80	76.51	46.83	38.97	36.78	76.32	61.11	53.62	7.76
IA-SSD	60.30	88.34	80.13	75.10	46.51	39.03	35.60	78.35	61.94	55.70	2.69
Our work	58.91	84.12	74.94	71.63	45.94	42.99	42.72	78.25	58.81	55.53	1.87

Our work demonstrates a notable advantage in terms of params, and while maintaining a straightforward inference architecture, it incurs an acceptable loss in detection accuracy, achieving a balance between precision and efficiency.

TABLE II. THE ABLATION EXPERIMENTS

Methods	3D Car	3D Ped.	3D Cyc.	Param.	Latency
	Mod.	Mod.	Mod.	(M)	(ms)
baseline	75.58	42.19	59.81	5.22	65.68
+DEPE	75.13	42.53	59.90	5.22	71.22
+backbone	74.42	41.82	54.72	1.87	49.86
+DEPE +backbone	74.94	42.99	58.81	1.87	54.75