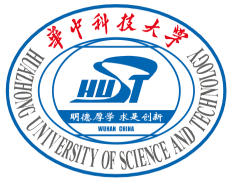
SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection

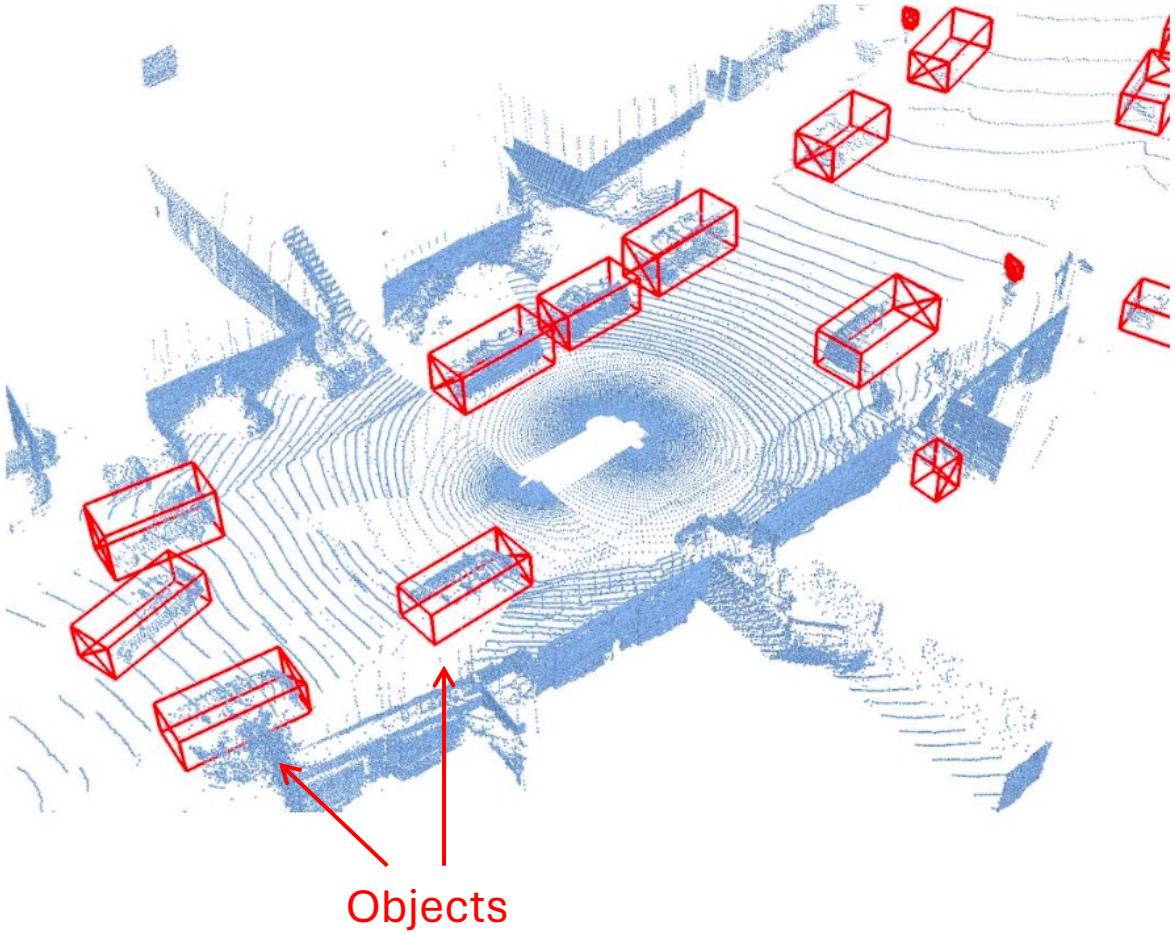
Gang Zhang[1] Junnan Chen[2] Guohuan Gao[3] Jianmin Li[1] Si Liu[4] Xiaolin Hu[1*]

Tsinghua University



3D Object Detection in Point Clouds

Tsinghua University



Point clouds

- highly sparse
- distribute on objects' surfaces

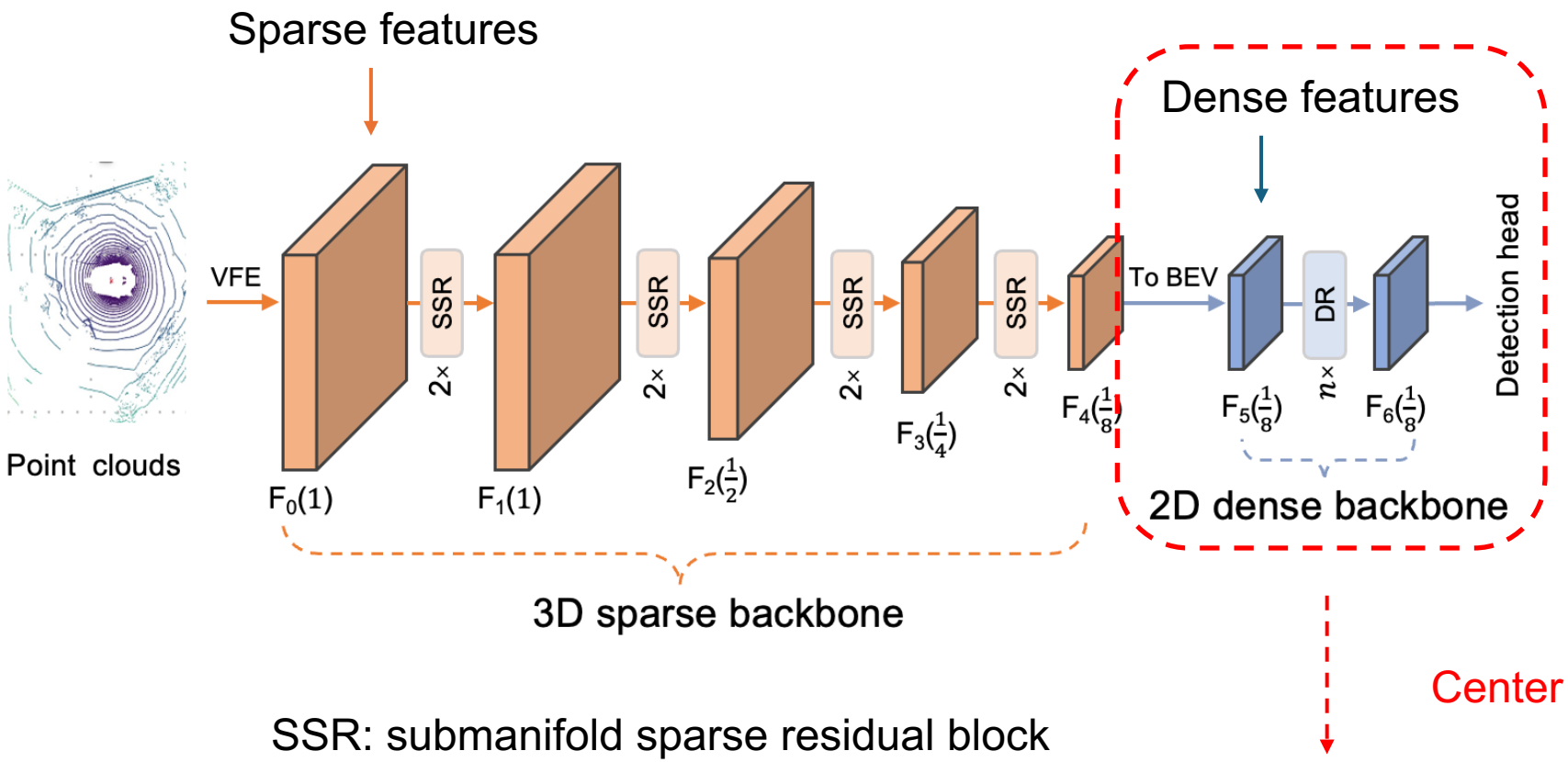
3D Object Detection for Autonomous Driving: A Comprehensive Survey. Mao, et al., 2023.

Review 3D Object Detectors in Point Clouds

Tsinghua University

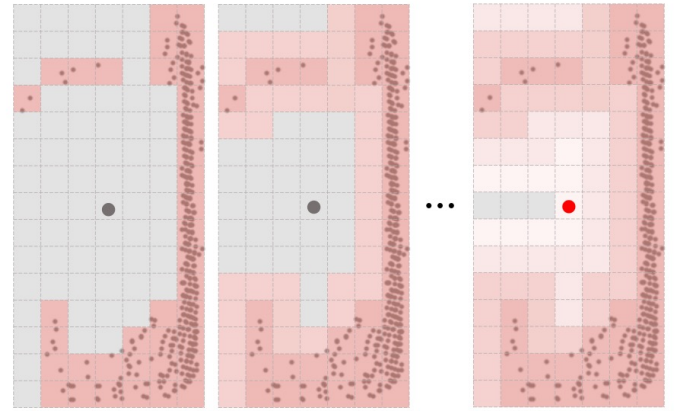
	2D/3D Backbone	2D Backbone	Head	Methods	Drawbacks
Dense Detectors	Dense	Dense	Dense	VoxelNet, PointPillars	Low efficiency, Inferior accuracy
Hybrid Detectors	Sparse	Dense	Dense	CenterPoint, TransFusion-L	Low efficiency in long-range detection
Sparse Detectors	Sparse	Sparse	Sparse	FSD v1/v2, VoxelNeXt	Complex pipeline, Inferior accuracy

Review CenterPoint w/ SECOND (Hybrid Detector)



SSR: submanifold sparse residual block

From FSD (Fan et al., 2021)



Feature Diffusion
(2D dense backbone)

Center Feature Missing (CFM) problem

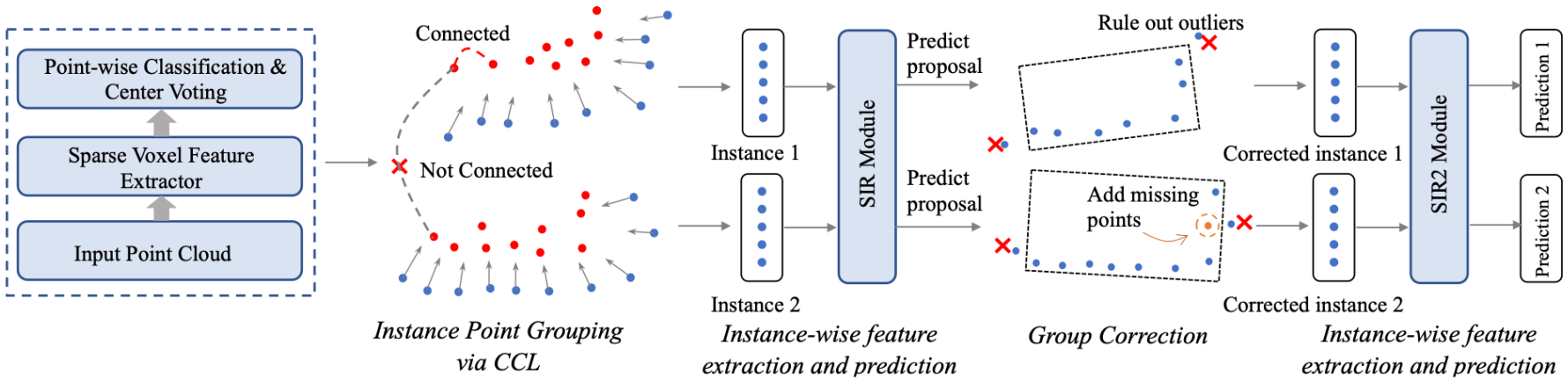
Low efficiency

Non-empty voxels, fewer than 10%

CenterPoint, Yin et al., 2021; SECOND, Yan et al., 2018.

Review FSD and VoxelNeXt (Sparse Detector)

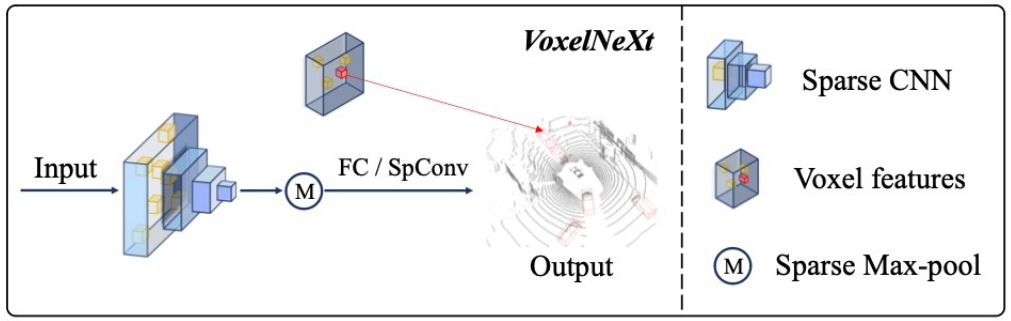
Tsinghua University



()

FSDv1: instance clustering

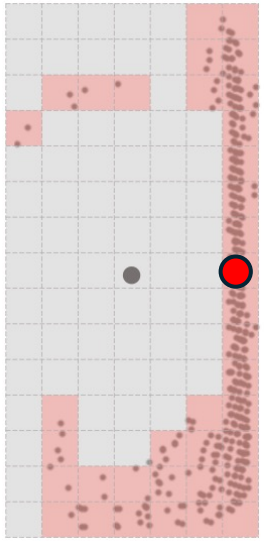
Different from CenterPoint, and complex pipline



VoxelNeXt: nearest voxels

Derived from CenterPoint, but inferior accuracy

From FSD



The desired way to solve the CFM problem should:

- Introduce minimal changes to existing detectors like CenterPoint
- Maintain performance

FSD, Fan et al., 2021; VoxelNeXt, Chen et al., 2021.

Preliminary Experiments on Waymo Open dataset

Tsinghua University

Method	Type	mAPH	Vehicle	Pedestrian	Cyclist
HEDNet	Hybrid	73.2	72.1	72.0	75.6
Nearest	Sparse	71.5	68.9	70.9	74.7

The previous best hybrid detector (Ours, NeurIPS'23)

Backbone of HEDNet + Head of VoxelNeXt

drop dramatically on large objects

HEDNet (Hybrid detector)

Make predictions based on center voxels

VoxelNeXt Style (Sparse detector)

Make predictions based on nearest voxels

Desired solution:

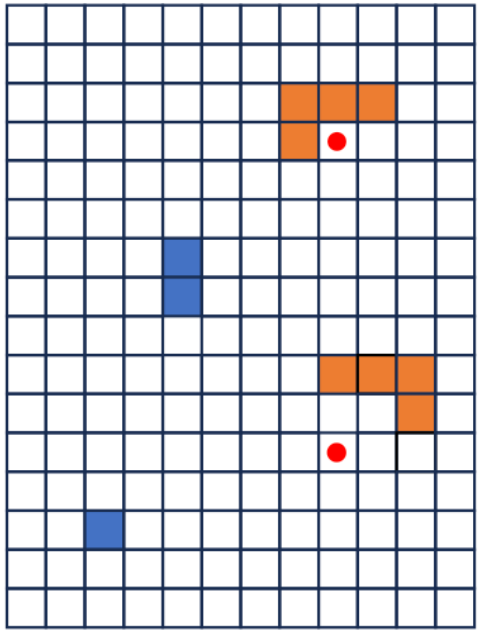
- Extract center voxel features
- Scale-aware
- Maintain feature sparsity as much as possible

HEDNet: A Hierarchical Encoder-Decoder Network for 3D Object Detection in Point Clouds. Zhang et al., NeurIPS 2023.

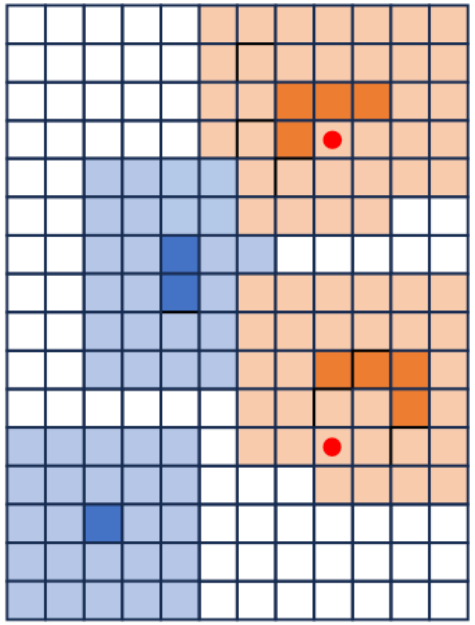
Adaptive Feature Diffusion (AFD)

Tsinghua University

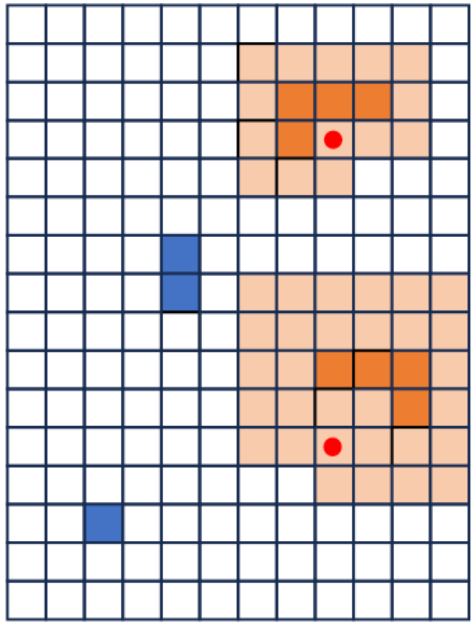
Blue: background; Orange: objects



Input feature map



Uniform feature diffusion

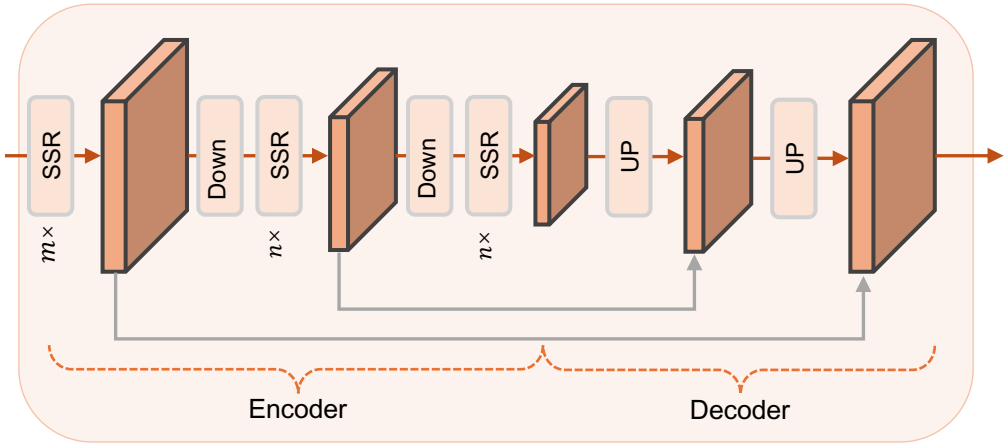


Adaptive feature diffusion

Adaptive feature diffusion: how to perform voxel classification? category group.

Larger objects, larger diffusion ranges.

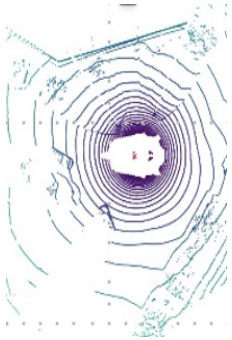
SAFDNet



3D/2D-SED

3D sparse feature map

2D sparse feature map



Point clouds

VFE → 2×SSR → 2×SSR → 4×SSR → 3D-SED → To BEV → AFD → 2D-SED → Sparse detection head

Voxel classification

$L\times W\times H$ | $L\times W\times H$ | $\frac{L}{2}\times\frac{W}{2}\times\frac{H}{2}$ | $\frac{L}{4}\times\frac{W}{4}\times\frac{H}{4}$ | $\frac{L}{8}\times\frac{W}{8}\times\frac{H}{4}$ | $\frac{L}{8}\times\frac{W}{8}$ | $\frac{L}{8}\times\frac{W}{8}$ | $\frac{L}{8}\times\frac{W}{8}$

3D sparse backbone

2D sparse backbone

Tsinghua University

HEDNet: A Hierarchical Encoder-Decoder Network for 3D Object Detection in Point Clouds. Zhang et al., NeurIPS 2023.

Experimental Results

Tsinghua University

Method	Waymo Open			Argoverse2			
	mAPH	FPS	Speedup	mAP	FPS	Speedup	Mem.
HEDNet [7]	73.4	17.2	1.0×	37.1	7.3	1.0×	28.7G
VoxelNeXt [13]†	70.1	15.7	0.9×	30.7	19.6	2.7×	6.2G
FSDv2 [12]†	73.5	10.3	0.6×	37.6	11.5	1.6×	8.6G
SAFDNet†(Ours)	73.9	20.2	1.2×	39.7	15.1	2.1×	7.3G

Hybrid detector

Sparse detector

On Waymo Open (Short-range detection)

- 2x faster than sparse detector FSDv2
- comparable to hybrid detector HEDNet

On Argoverse2 (Long-range detection)

- 2.1% better than sparse detector FSDv2
- 2.1x faster than hybrid detector HEDNet

Tsinghua University

Q & A

Code

Paper

Poster #22, Arch 4A-E, 5 p.m.