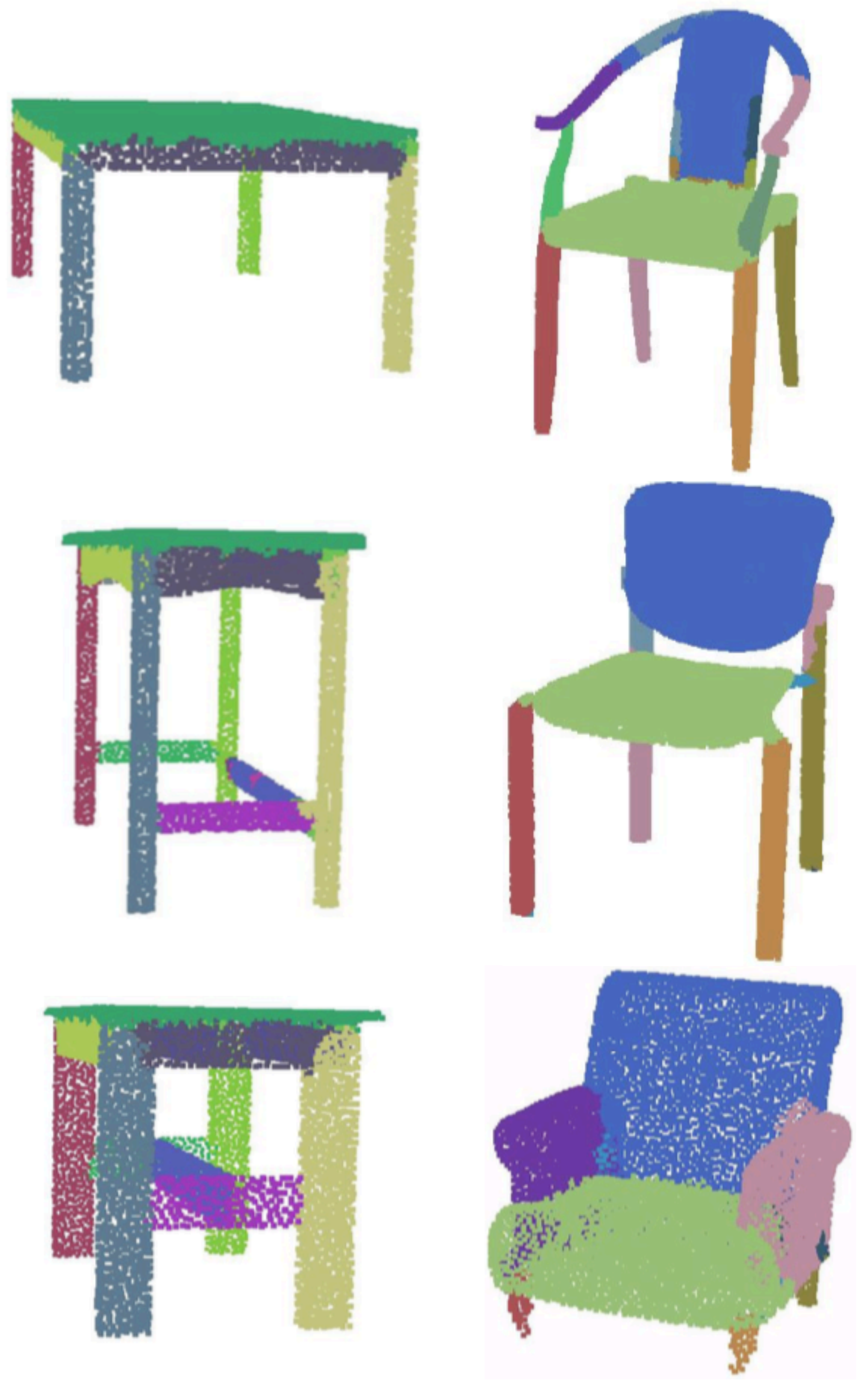
Hardware, AI and Neural-nets

Point-Voxel CNN for Efficient 3D Deep Learning

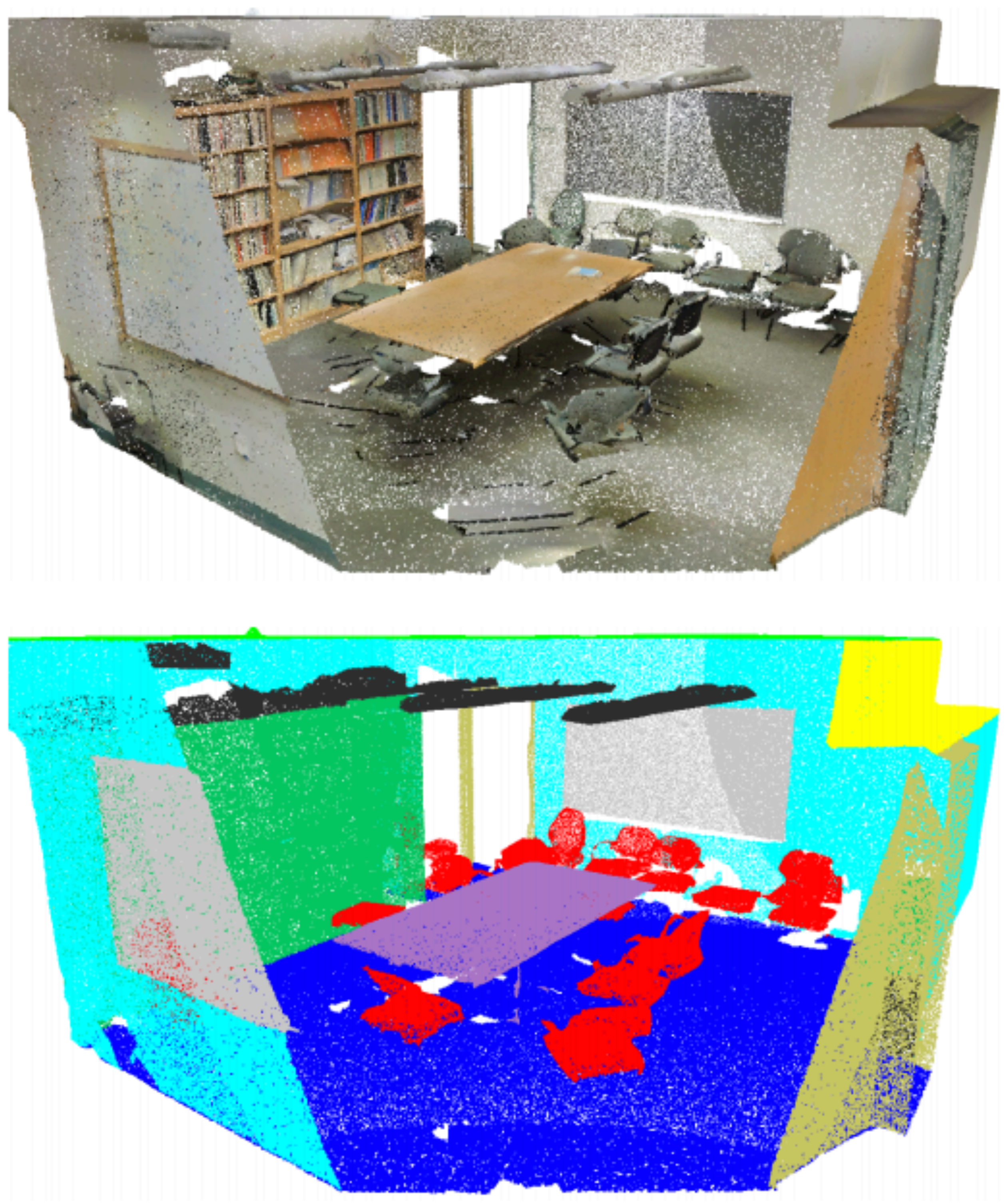
Zhijian Liu*, **Haotian Tang***, **Yujun Lin**, and **Song Han**

Project Page: **http://pvcnn.mit.edu/**

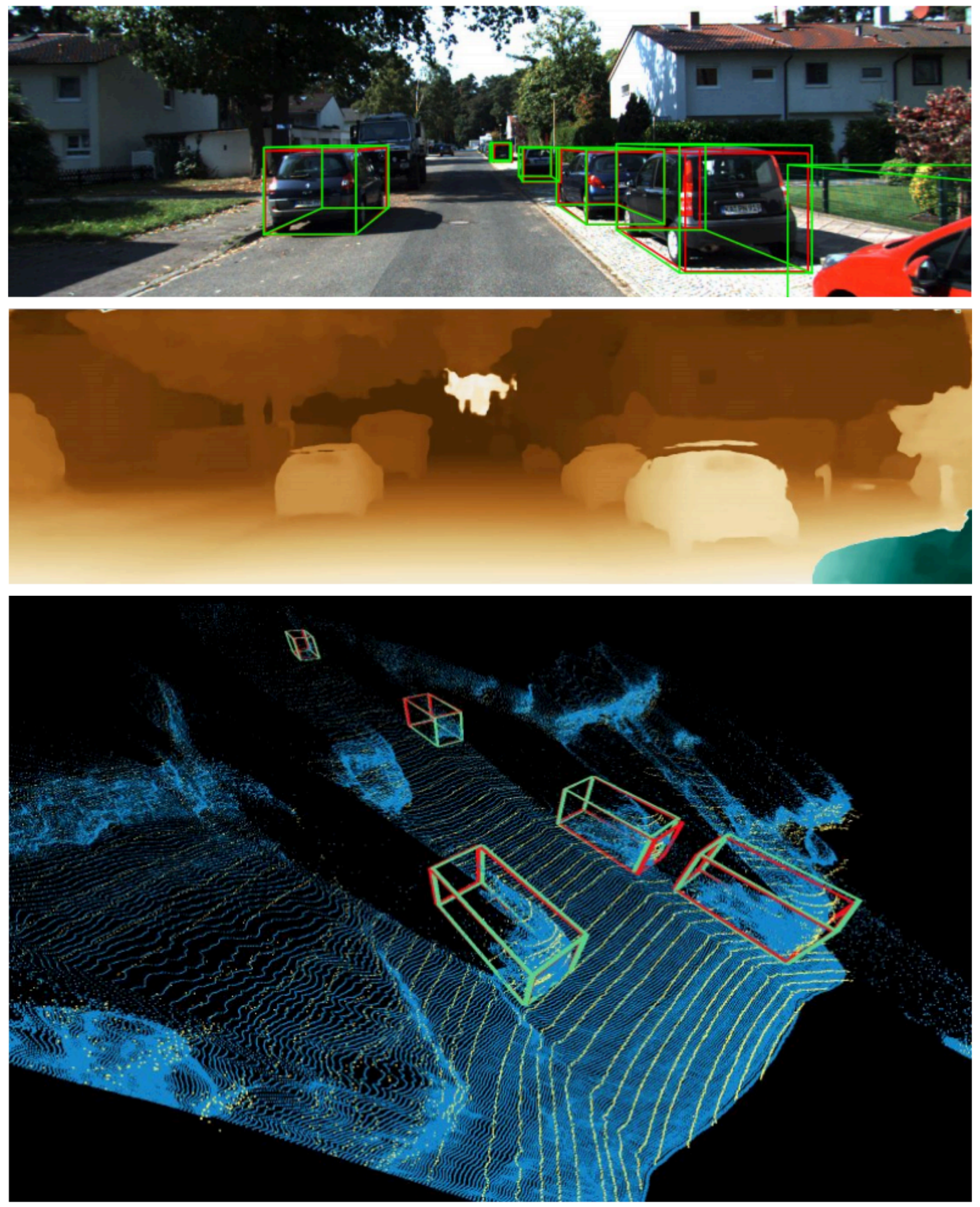
3D Deep Learning



3D Part Segmentation
(for Robotic Systems)

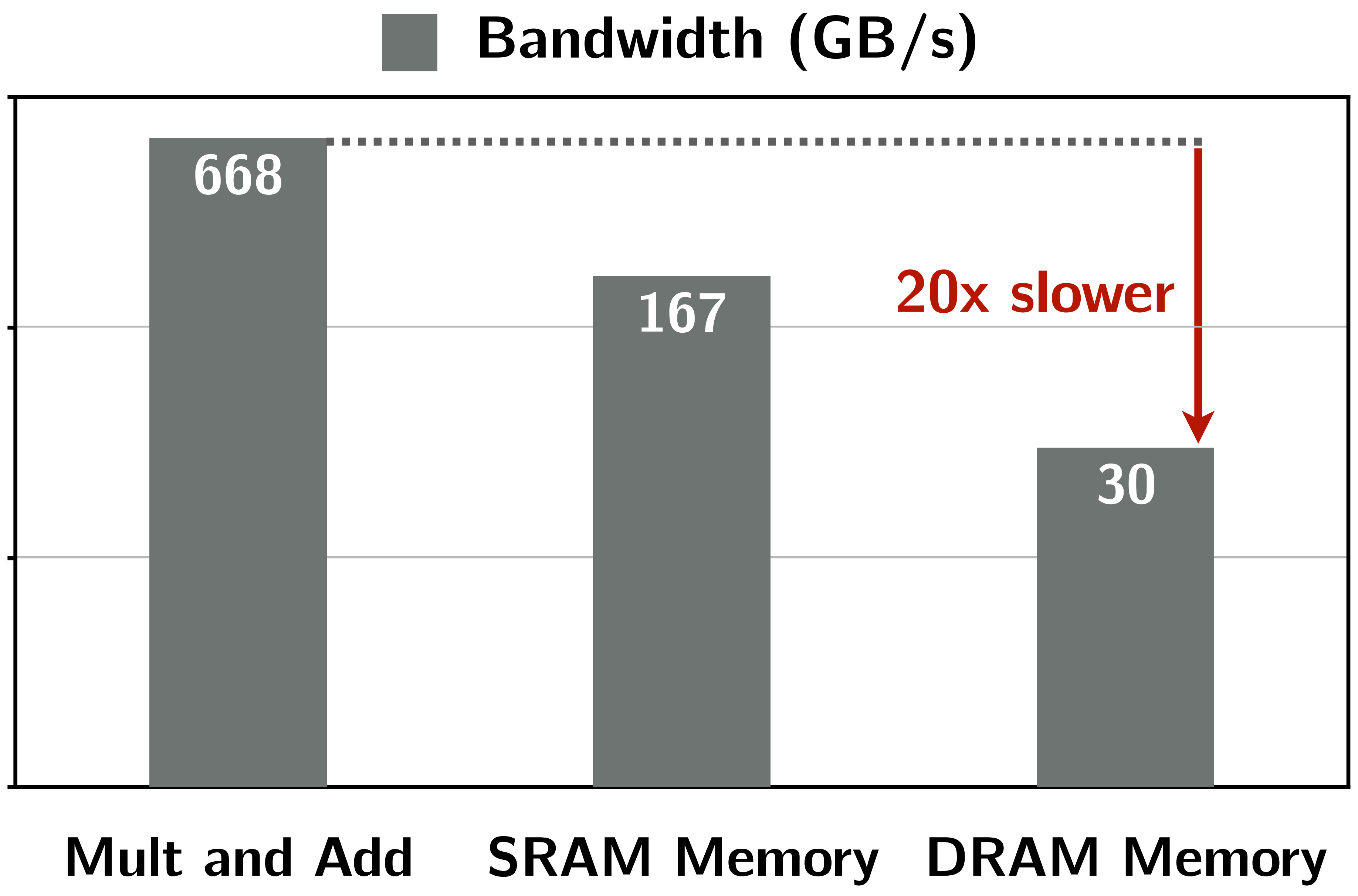


3D Semantic Segmentation
(for VR/AR Headsets)



3D Object Detection
(for Self-Driving Cars)

Efficient 3D Deep Learning



Sequential Memory Access

1	2	3	4	5	6	7	8

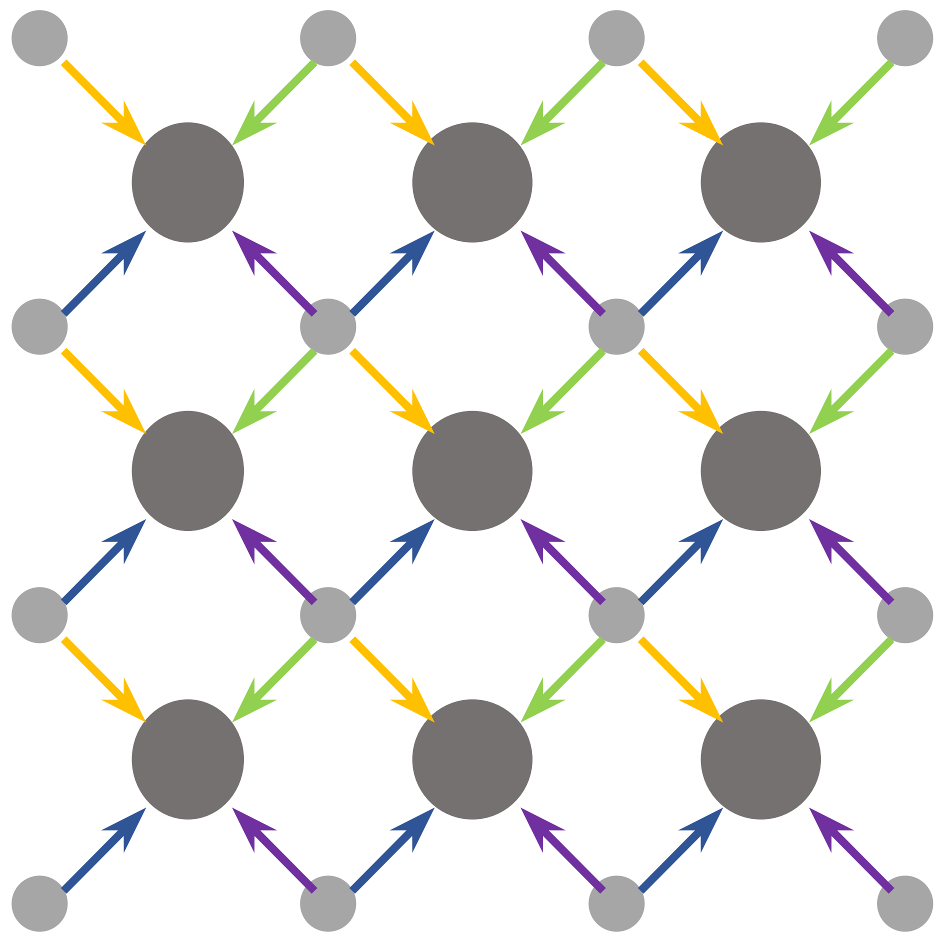
Random Memory Access

7	5	2	4	6	1	8	3

Off-chip DRAM access is much more expensive than arithmetic operation!

Random memory access is inefficient due to the potential bank conflicts!

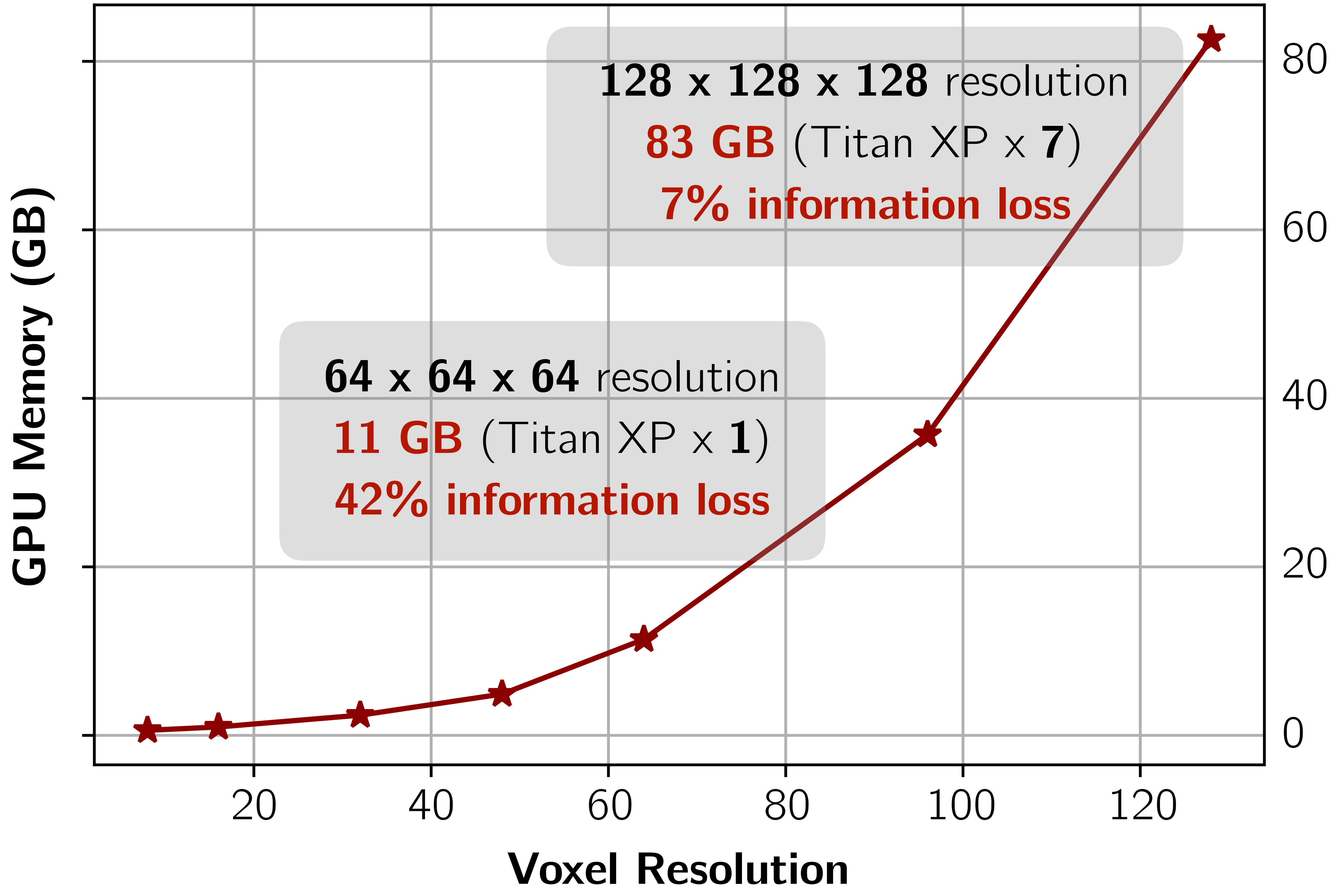
Voxel-Based Models: Cubically-Growing Memory



3D ShapeNets [CVPR'15]

VoxNet [IROS'15]

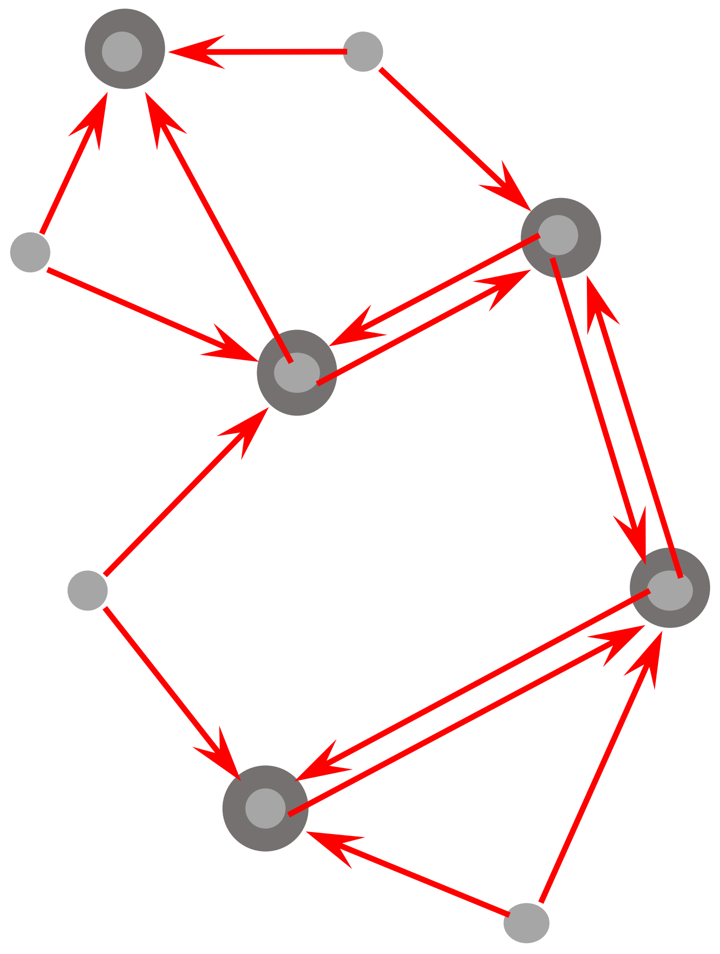
3D U-Net [MICCAI'16]



128 x 128 x 128 resolution
83 GB (Titan XP x **7**)
7% information loss

64 x 64 x 64 resolution
11 GB (Titan XP x **1**)
42% information loss

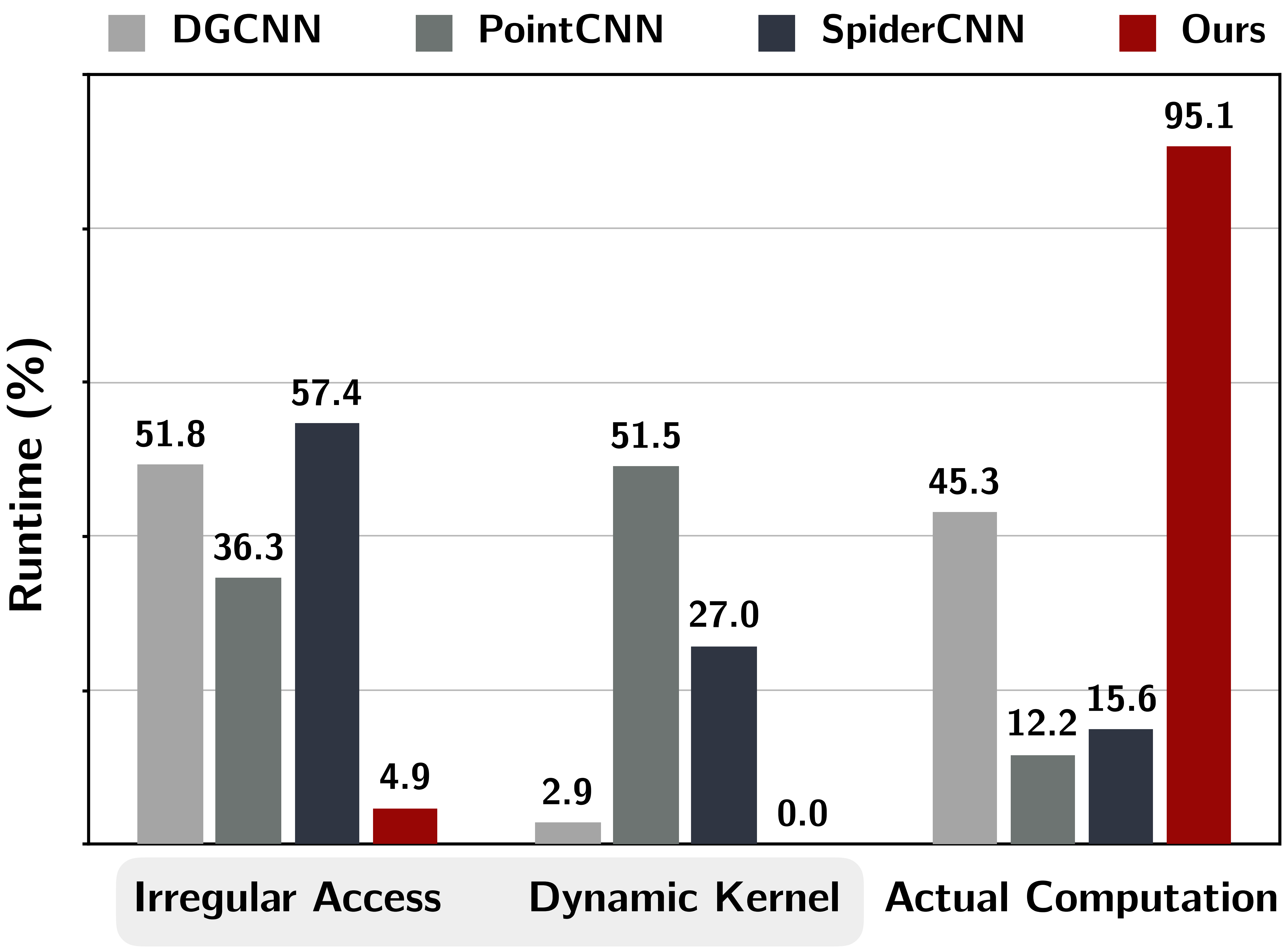
Point-Based Models: Sparsity Overheads



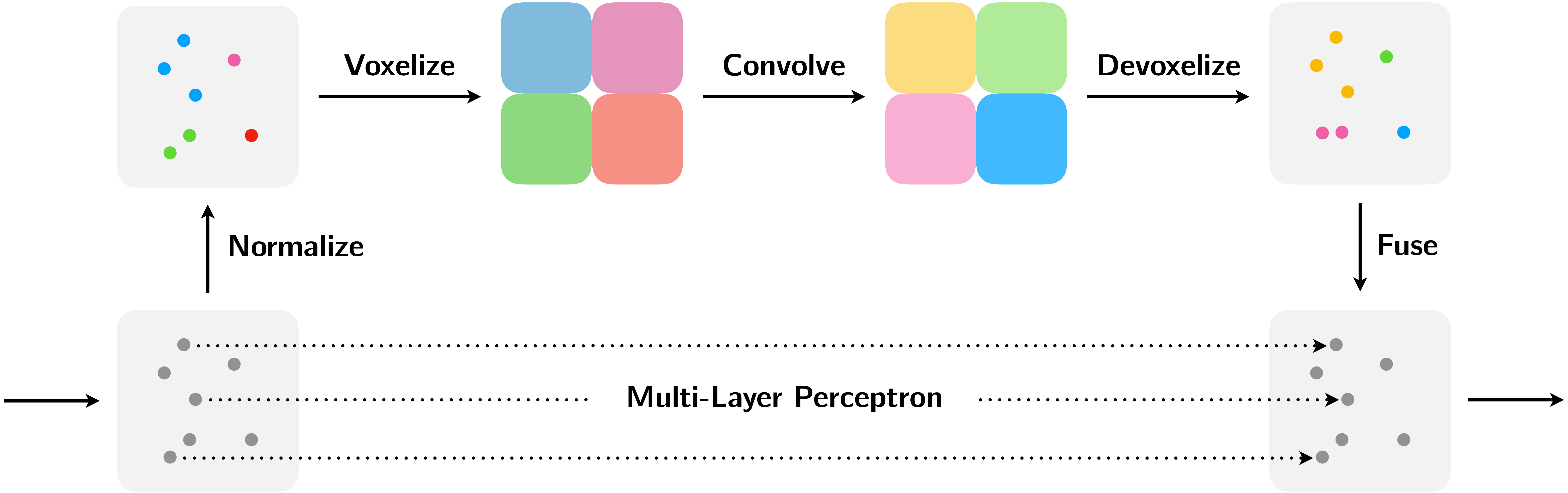
PointNet [CVPR'17]

PointCNN [NeurIPS'18]

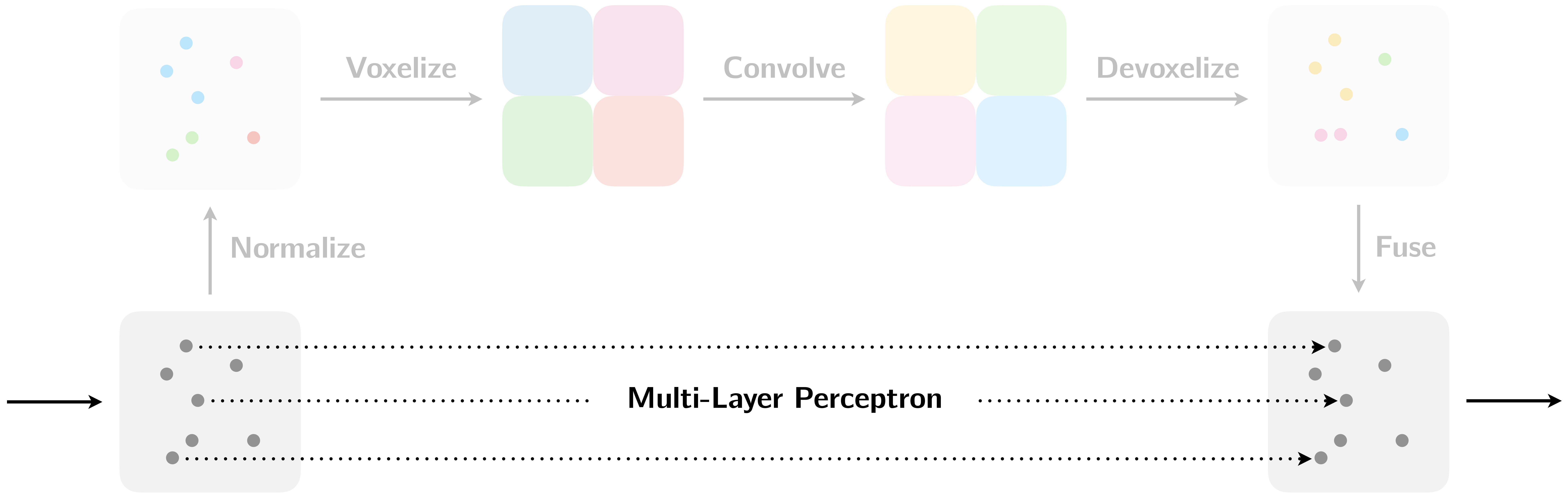
DGCNN [SIGGRAPH'19]



Point-Voxel Convolution (PVConv)

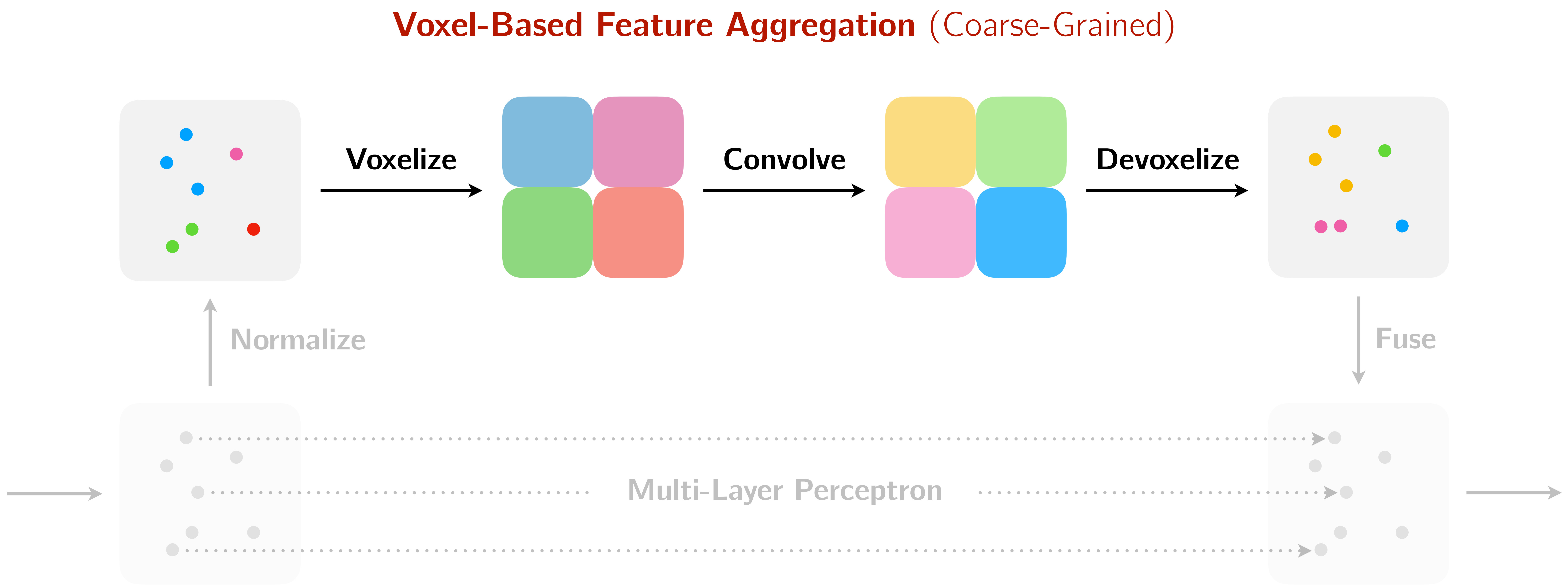


Point-Voxel Convolution (PVConv)

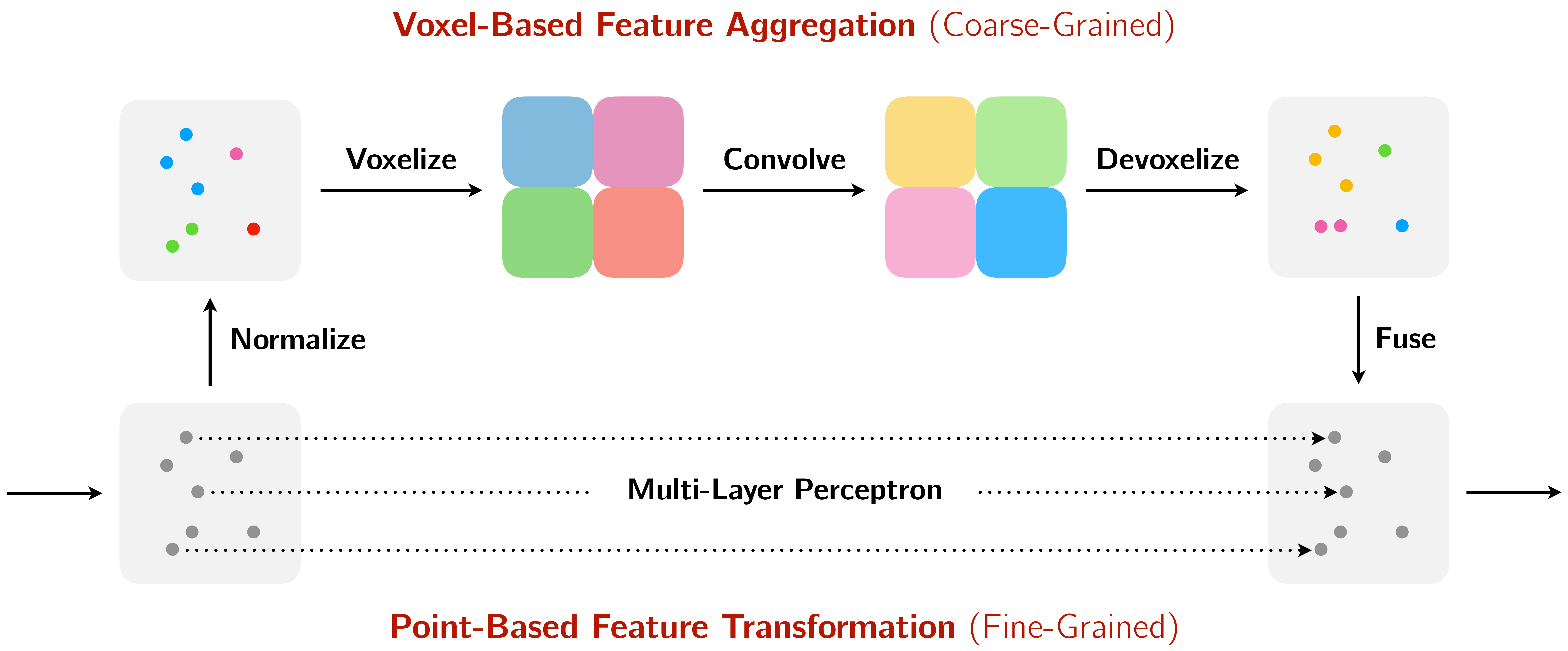


Point-Based Feature Transformation (Fine-Grained)

Point-Voxel Convolution (PVConv)

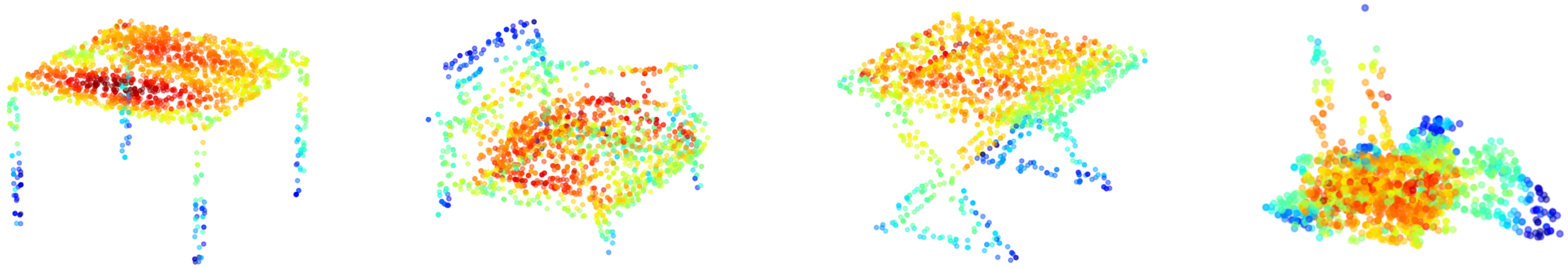


Point-Voxel Convolution (PVConv)

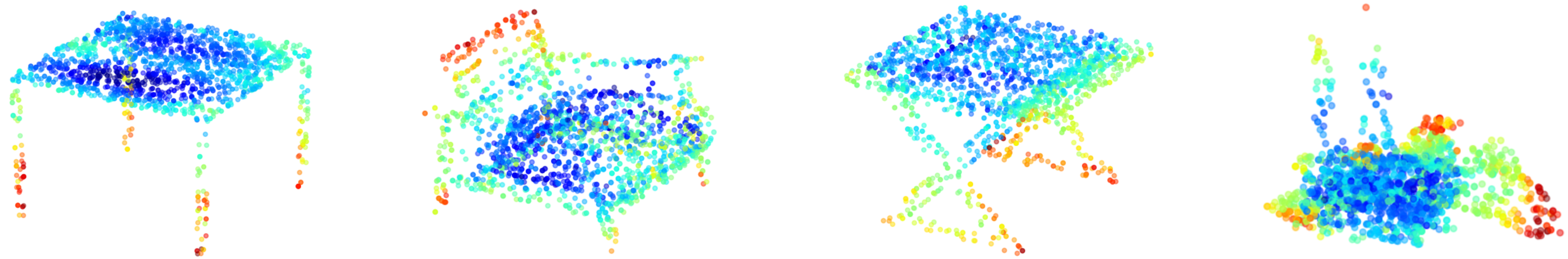

Voxel-Based Feature Aggregation (Coarse-Grained)
Voxelize
Convolve
Devoxelize
Normalize
Fuse
Multi-Layer Perceptron
Point-Based Feature Transformation (Fine-Grained)

Point-Voxel Convolution (PVConv)

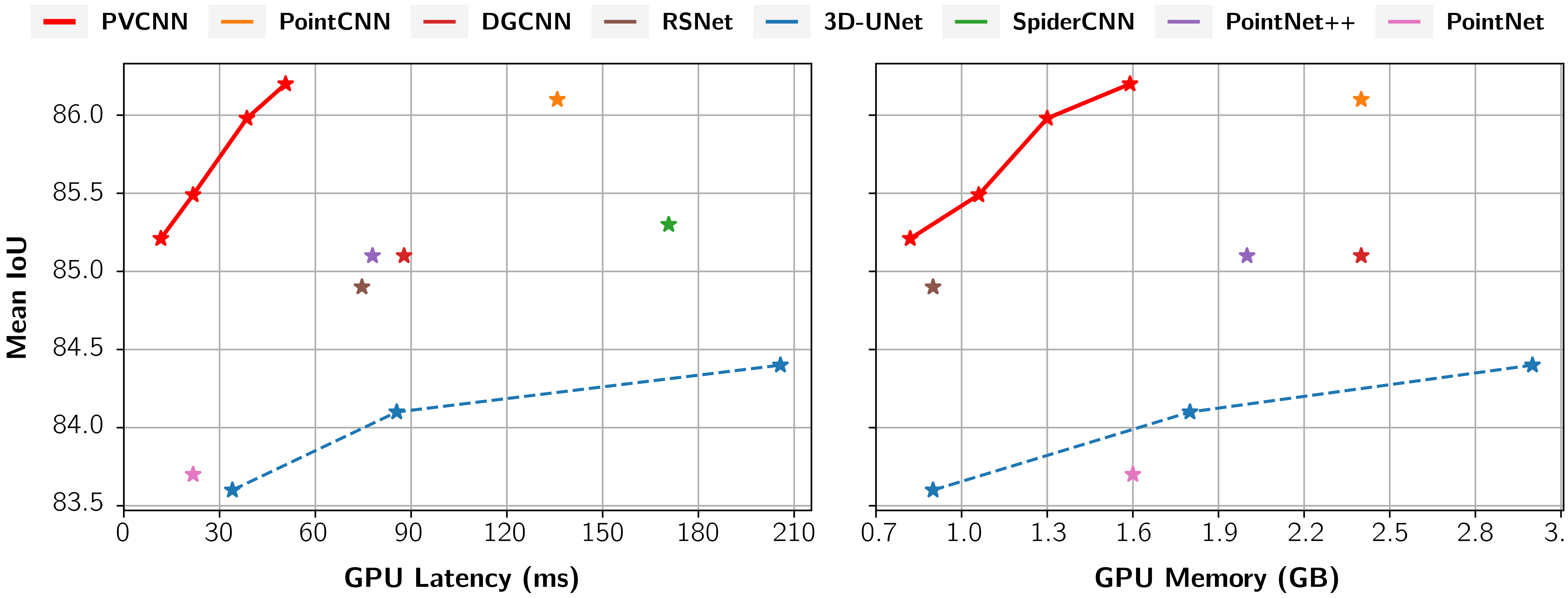
Features from **Voxel-Based Branch**:



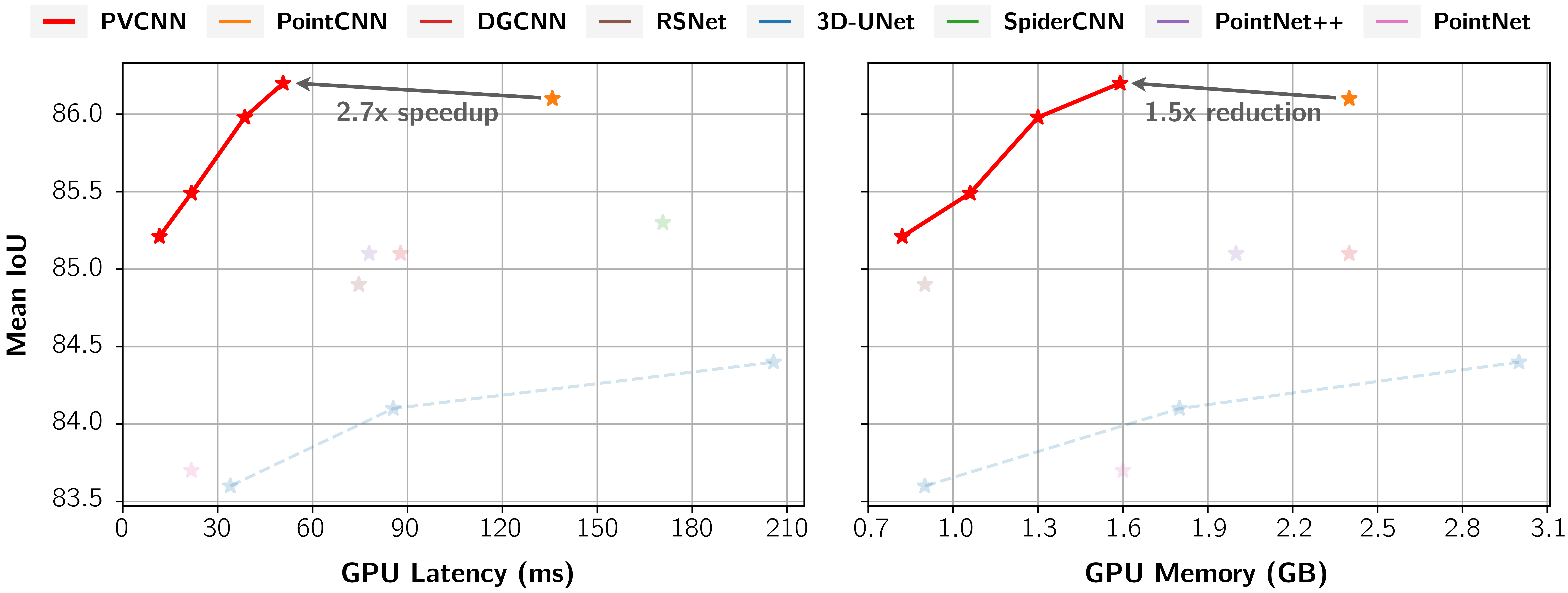
Features from **Point-Based Branch**:



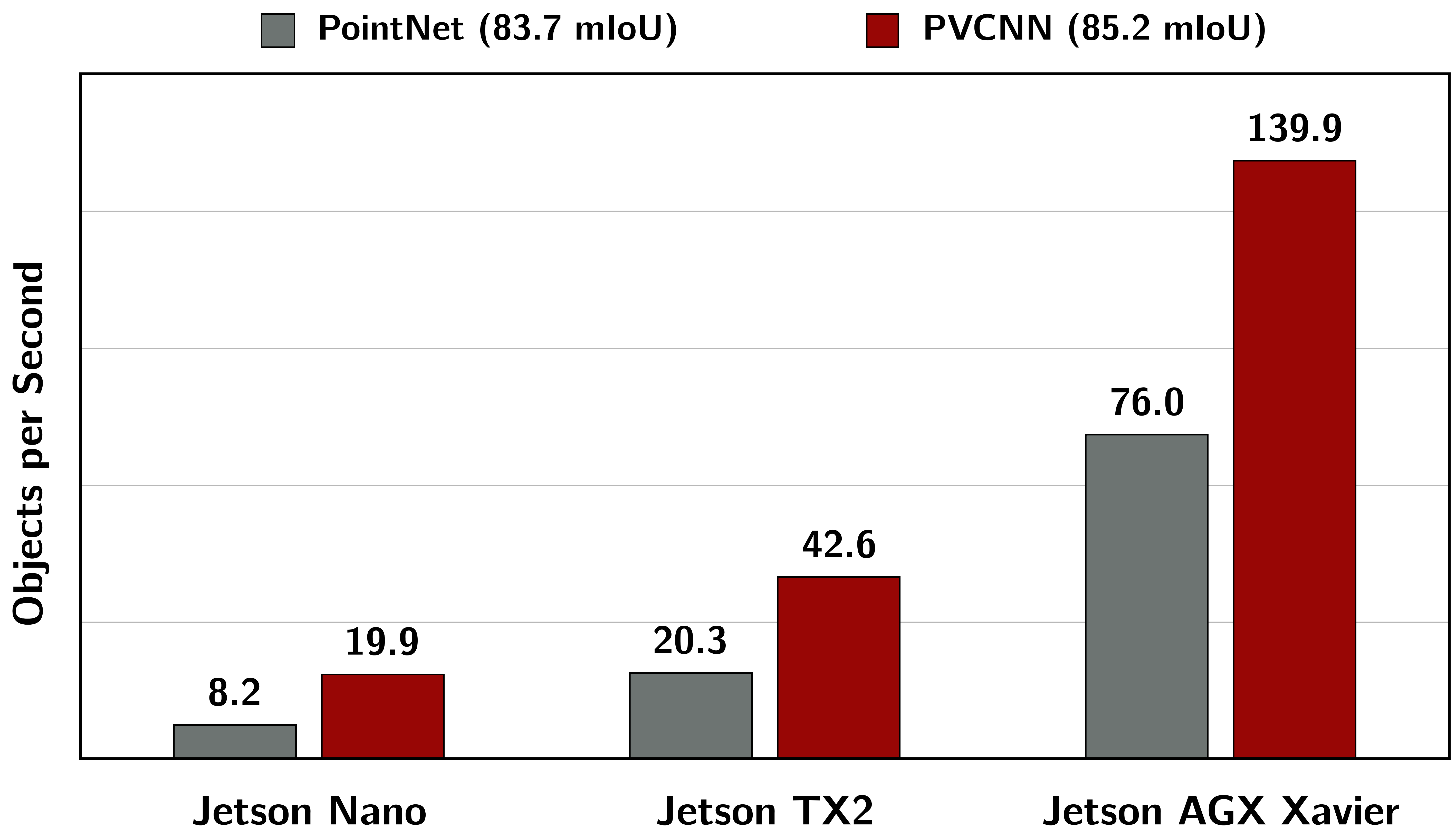
Results: 3D Part Segmentation (ShapeNet)



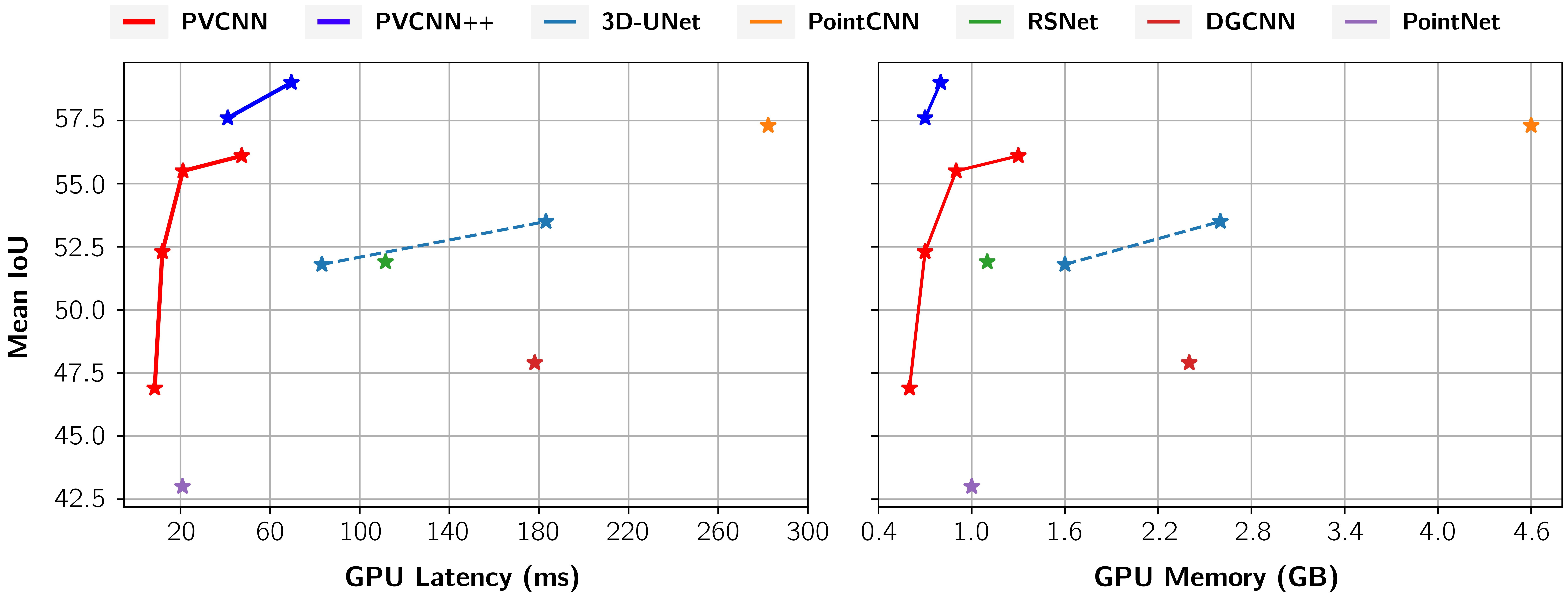
Results: 3D Part Segmentation (ShapeNet)



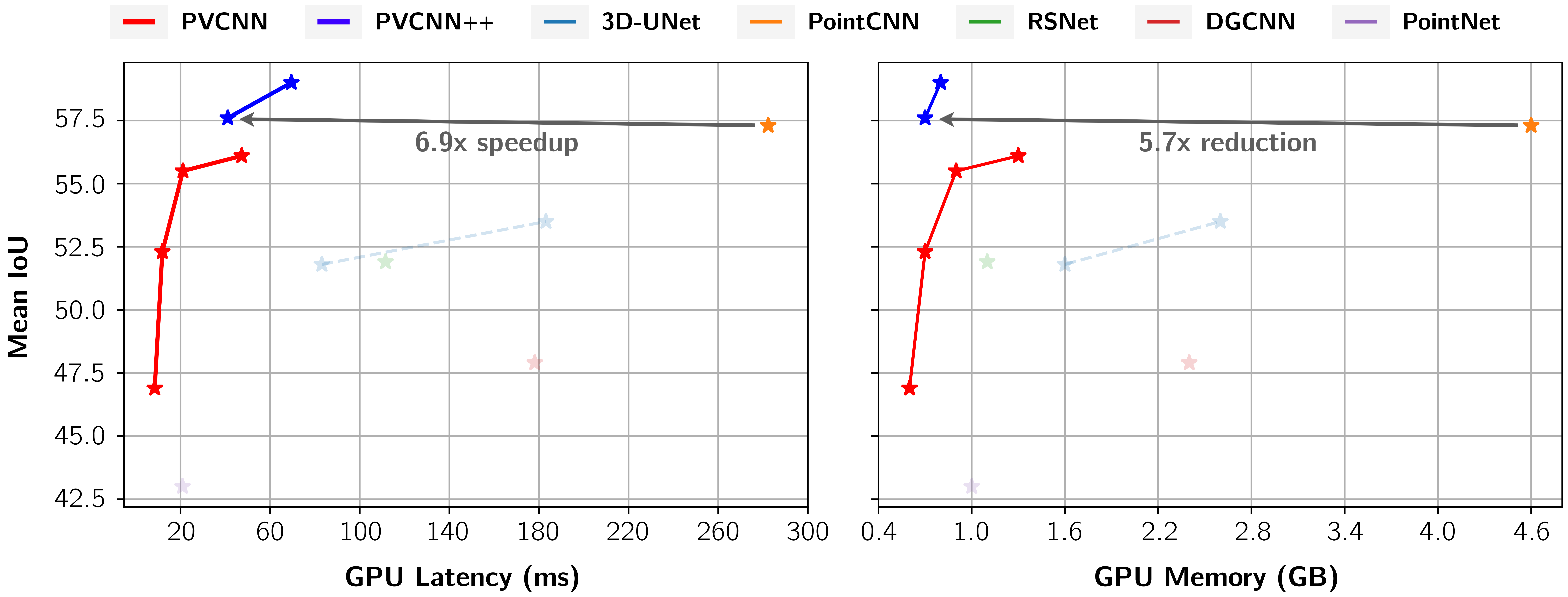
Results: 3D Part Segmentation (ShapeNet)



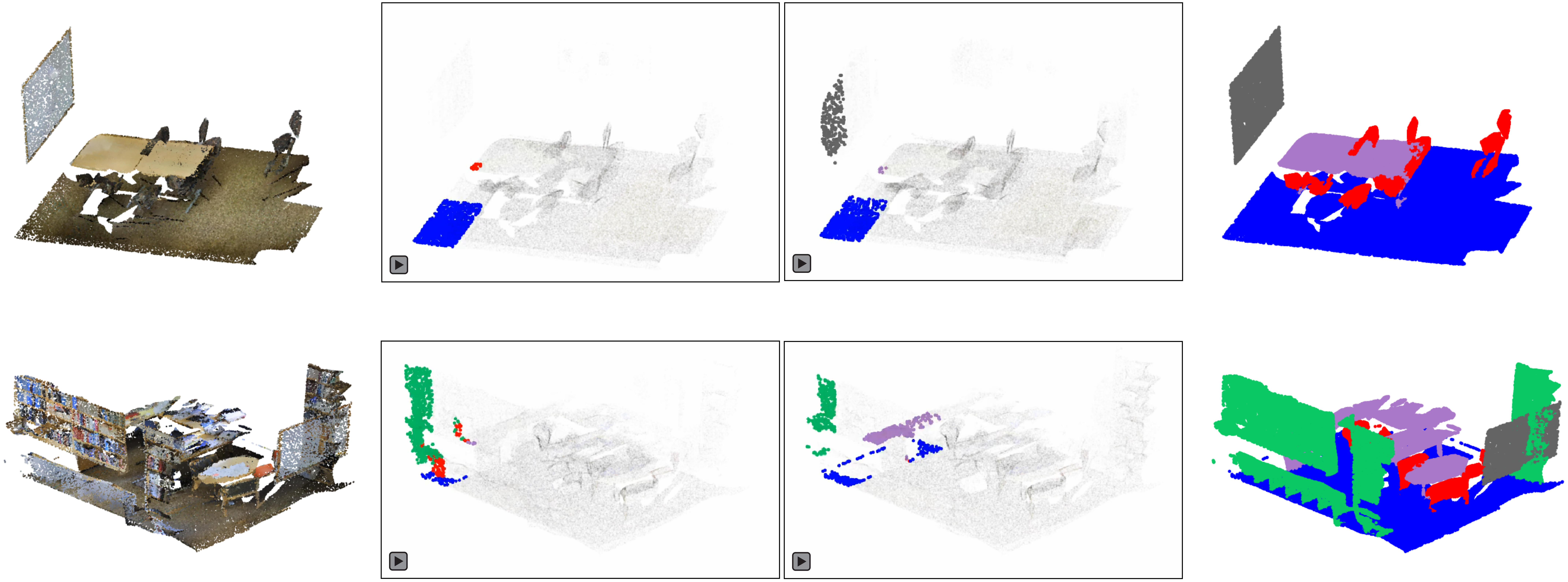
Results: 3D Semantic Segmentation (S3DIS)



Results: 3D Semantic Segmentation (S3DIS)



Results: 3D Semantic Segmentation (S3DIS)



Input Scene

PointNet

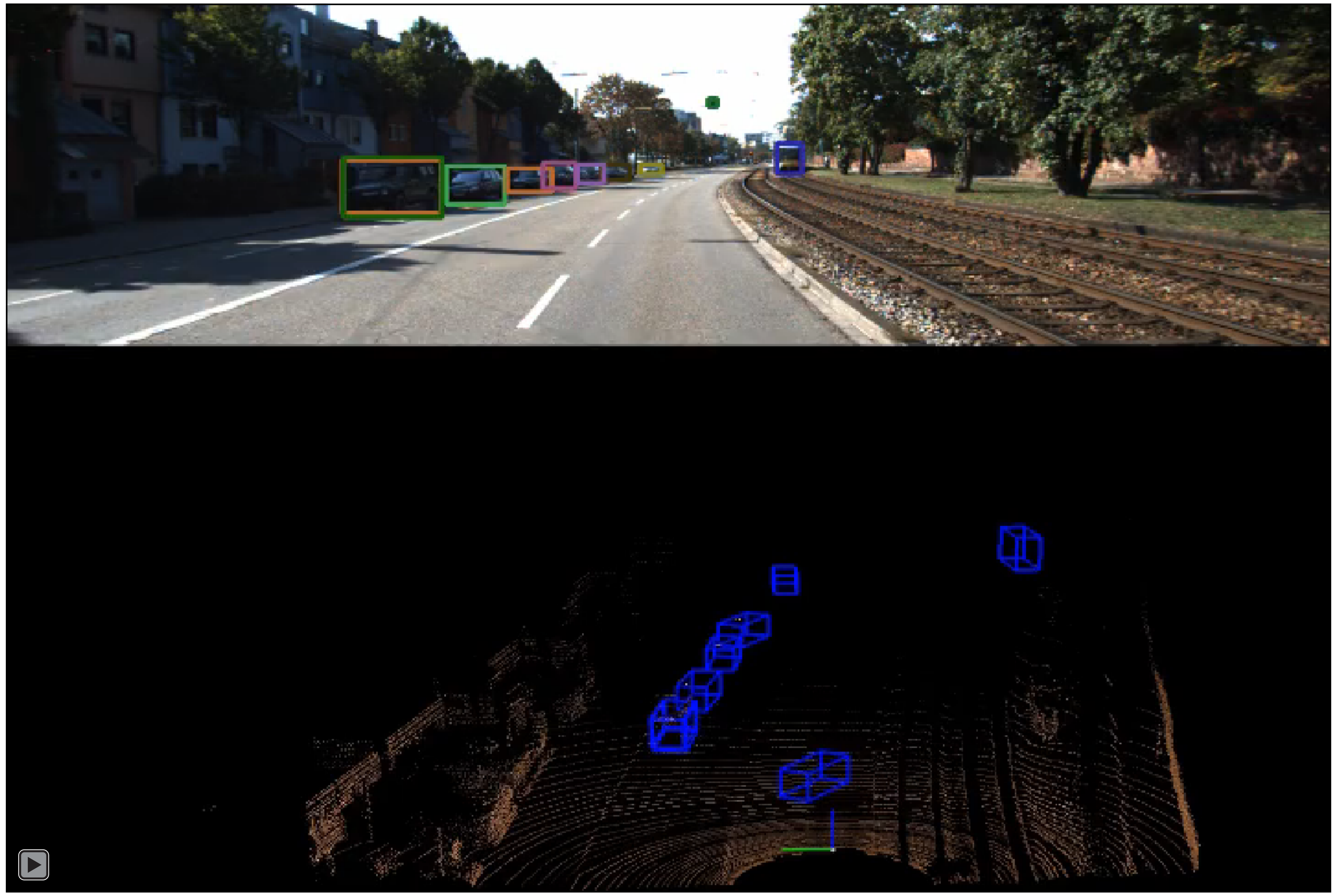
PVCNN
(**1.8x** faster)

Ground Truth

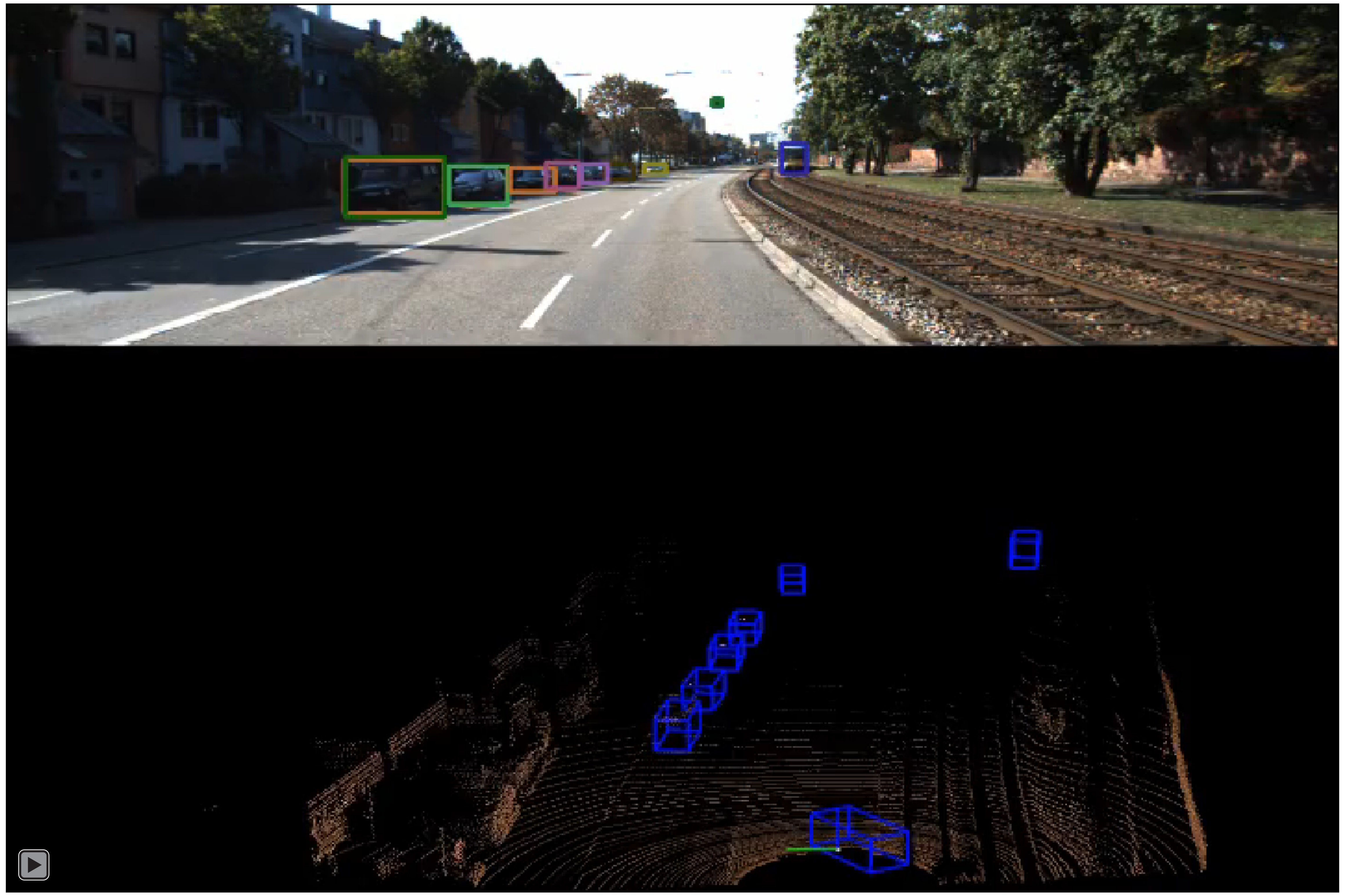
Results: 3D Object Detection (KITTI)

	GPU Latency	GPU Memory	Pedestrian	Cyclist	Car
F-PointNet++	105.2 ms	2.0 GB	61.6	62.4	72.8
PVCNN (efficient)	**58.9 ms** **(1.8x)**	**1.4 GB** **(1.4x)**	60.7 (-0.9)	63.6 (+1.2)	73.0 (+0.2)
PVCNN (complete)	**69.6 ms** **(1.5x)**	**1.4 GB** **(1.4x)**	**64.9** **(+3.3)**	**65.9** **(+3.5)**	**73.1** **(+0.3)**
	Faster	**Lower**	**More Accurate**		

Results: 3D Object Detection (KITTI)



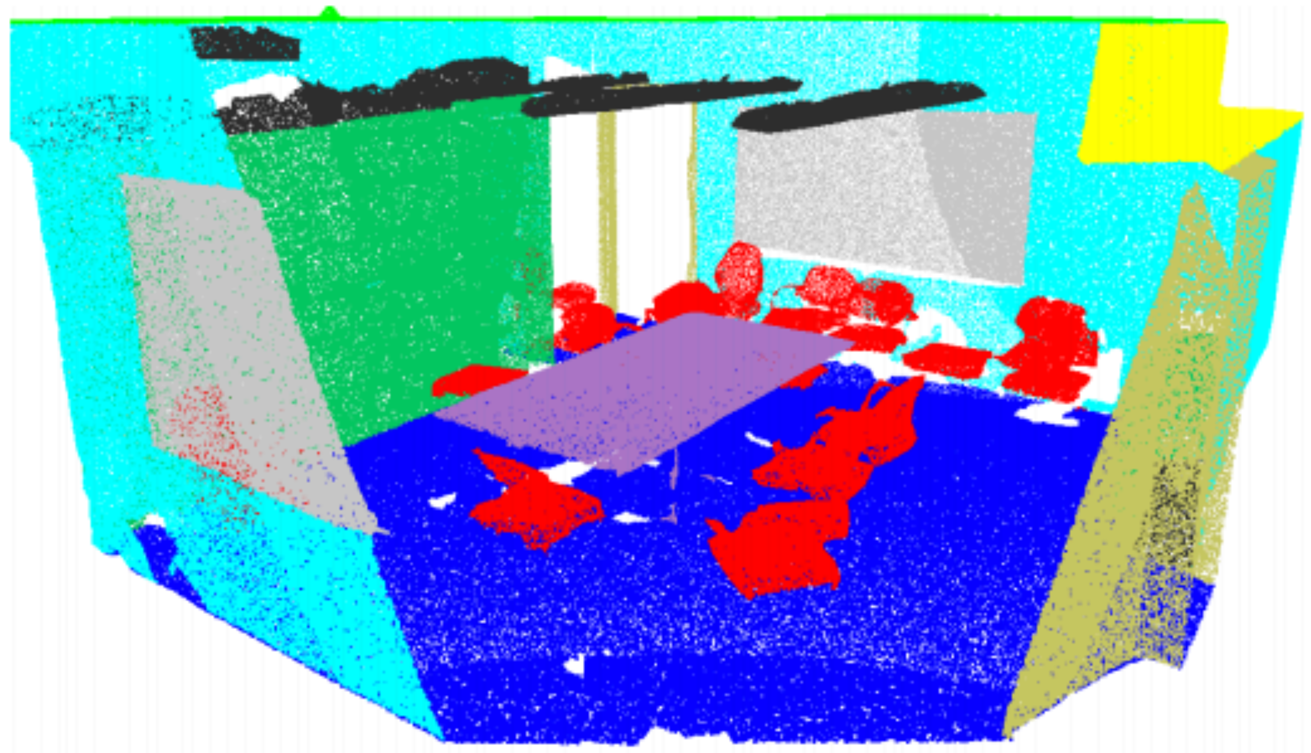
F-PointNet++
(10 FPS)



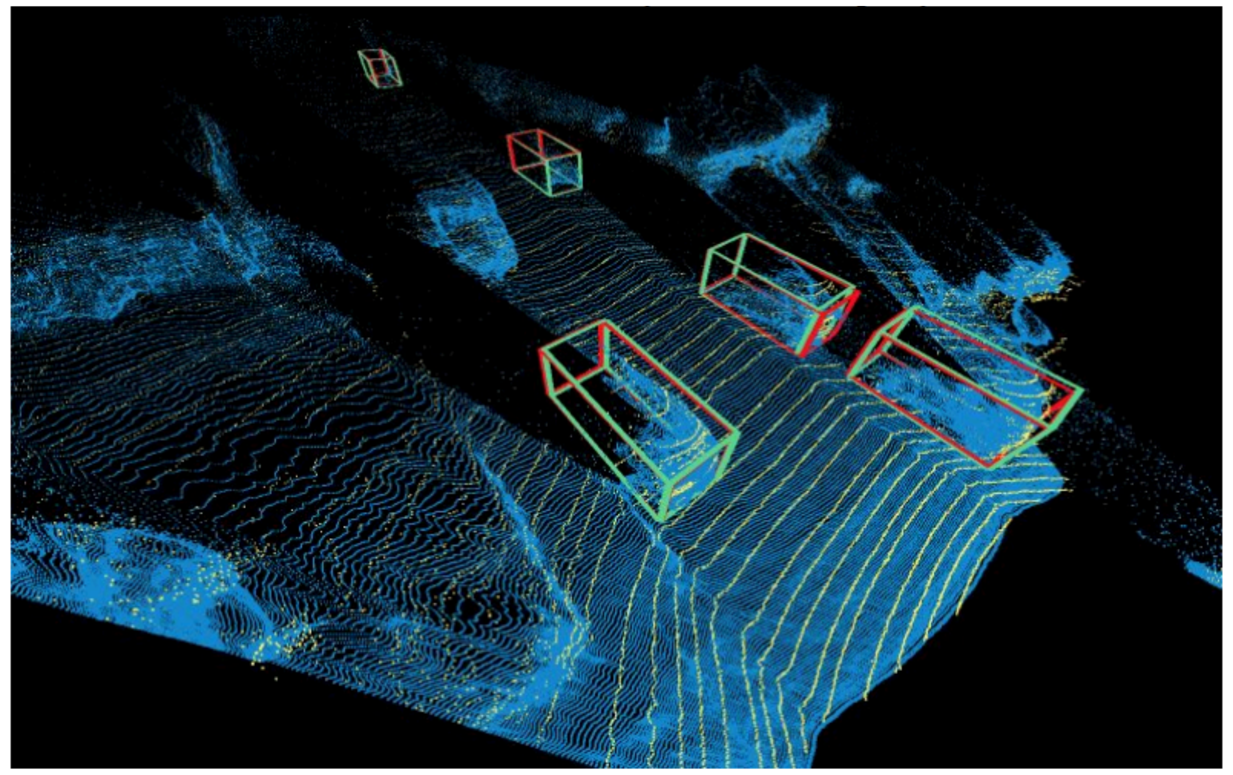
PVCNN
(**17** FPS, **1.8x** faster)

Point-Voxel CNN for Efficient 3D Deep Learning

2.7x measured speedup
1.5x memory reduction



6.9x measured speedup
5.7x memory reduction



1.8x measured speedup
1.4x memory reduction

Gold Medal in Lyft Challenge on 3D Object Detection for Autonomous Vehicles

Poster: 10:45-12:45 PM @ East Exhibition Hall B + C **#112**

GitHub: **https://github.com/mit-han-lab/pvcnn**

Project Page: **http://pvcnn.mit.edu**