

Sparse PointPillars: Exploiting Sparsity in Birds-Eye-View Object Detection

Kyle Vedder and Eric Eaton

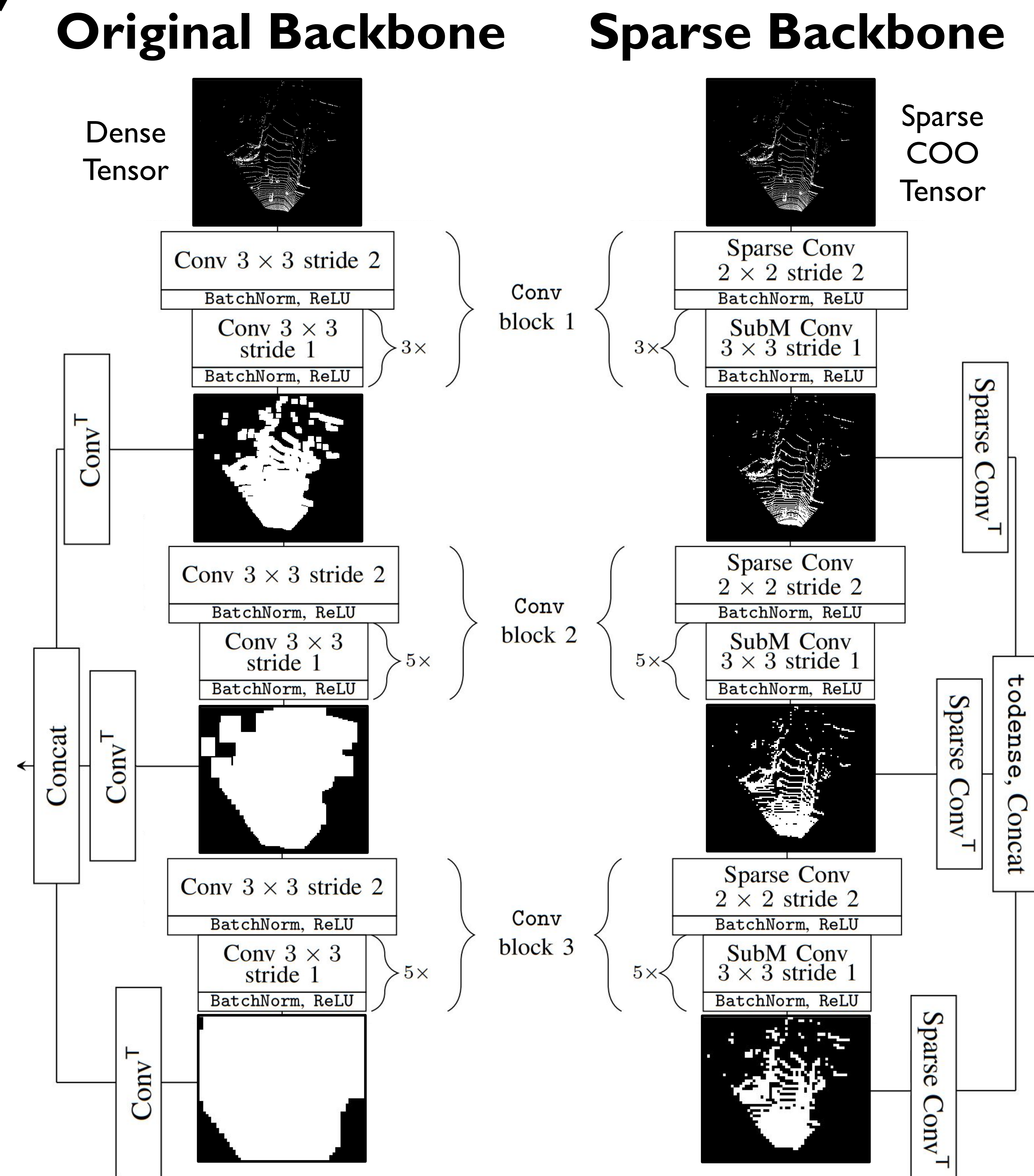
Full Workshop Paper:



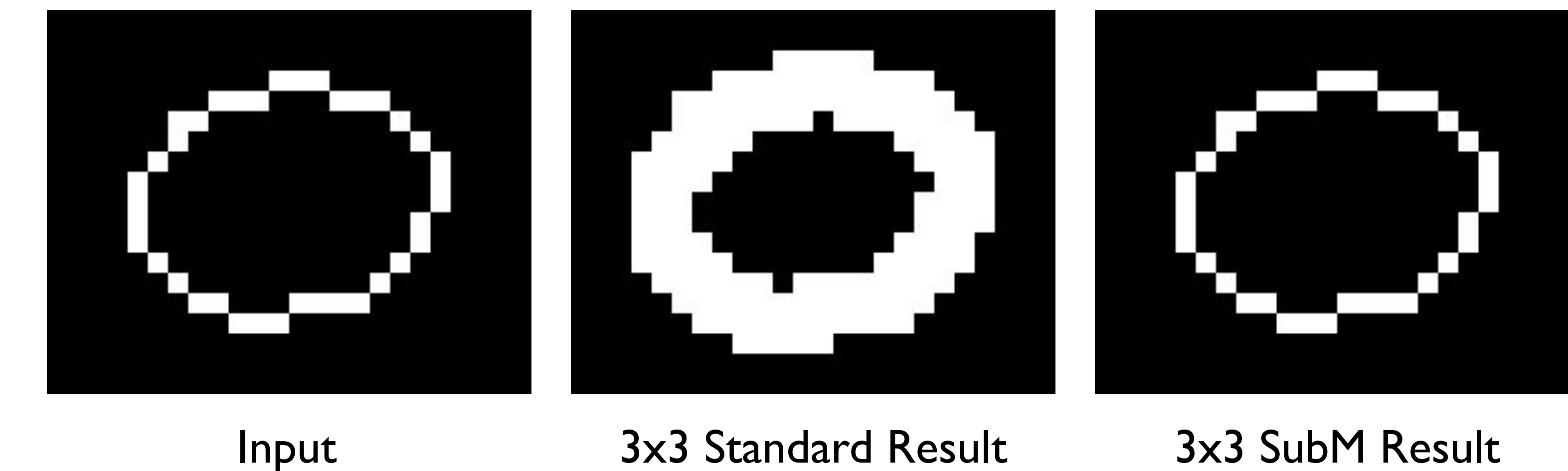
Code:



Key Idea: Modify *Feature Net* and *CNN Backbone* to use **sparse convolutions** and **preserve representational sparsity** in order to **reduce evaluation latency**



SubM vs Standard Conv



Runtimes

Table 1: Model runtime in milliseconds for each network component, averaged over ten trials, run on the KITTI dataset with 16cm×16cm pillars. All models have the same Feature Extractor and Head (runtimes included for completeness), and all non-Original models have the same sparse Feature Net.

	Feat. Extr.	Feat. Net	Backbone	Head	Total vs Original
Original PointPillars	6.904±0.018	1.344±0.043	16.185±0.053	3.638±0.022	-
Sparse PointPillars	6.879±0.016	0.508 ±0.030	14.090 ±0.057	3.778±0.018	-2.817
Sparse1+Dense23	6.898±0.017	0.517 ±0.022	17.321±0.050	3.646±0.021	0.223
Sparse12+Dense3	6.973±0.089	0.498 ±0.021	22.091±0.245	3.578±0.063	5.069
Sparse+WideConv	6.858±0.015	0.480 ±0.022	17.483±0.071	3.684±0.030	0.434

Detection Quality

Table 2: Performance of the original PointPillars as % AP and of our sparse model as the relative % AP difference (Δ) from Original on KITTI with 16cm×16cm pillars. Higher is better.

	Original PointPillars			Sparse PointPillars		
	Easy	Medium	Hard	Easy	Medium	Hard
BBox AP	90.51	88.67	87.06	0.11Δ	-2.68Δ	-4.78Δ
BEV AP	89.93	87.03	84.09	0.25Δ	-5.30Δ	-4.35Δ
3D AP	86.46	76.29	69.73	-1.85Δ	-5.31Δ	-1.39Δ

Table 3: Ablative model % AP difference (Δ) from Original on KITTI with 16cm×16cm pillars.

	Sparse1+Dense23			Sparse12+Dense3			Sparse+WideConv		
	Easy	Med.	Hard	Easy	Med.	Hard	Easy	Med.	Hard
BBox AP	-0.17Δ	-0.35Δ	-0.68Δ	-0.23Δ	-0.79Δ	-0.99Δ	0.00Δ	-2.38Δ	-4.84Δ
BEV AP	-0.03Δ	-0.85Δ	-3.58Δ	-0.24Δ	-1.42Δ	-2.24Δ	-0.06Δ	-5.56Δ	-2.90Δ
3D AP	-5.50Δ	-1.31Δ	-0.75Δ	-2.13Δ	-1.91Δ	-1.32Δ	-5.94Δ	-6.38Δ	-2.18Δ