Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving

Yan Wang, Wei-Lun (Harry) Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, Kilian Q. Weinberger

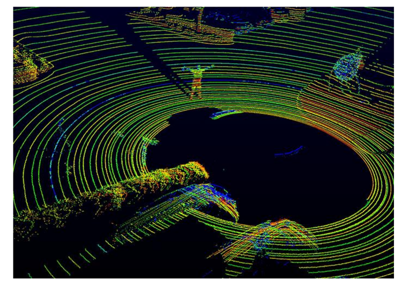
CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Highlights

- ***Propose an image-based 3D detection framework**: converting **image-based depth maps** to **pseudo-LiDAR representation** enables existing LiDAR-based 3D object detectors*
- ***Achieve a 45% AP_{3D} on the KITTI benchmark, almost a 350% improvement over the previous SOTA***

Introduction

- 3D object detection is essential for autonomous driving.
- Most approaches rely on LiDAR for precise depths, but:
  - ➢ Expensive (64-line = $75K USD)
  - ➢ Over-reliance is risky.
  - ➢ Alternatives are needed.



- Image-based approaches fall far behind (10% vs. 74% AP_{3D}), commonly attributed to *poor image-base depth estimation*.

Is image-based depth accurate?

- Image-based depth maps Z can be transformed to 3D points

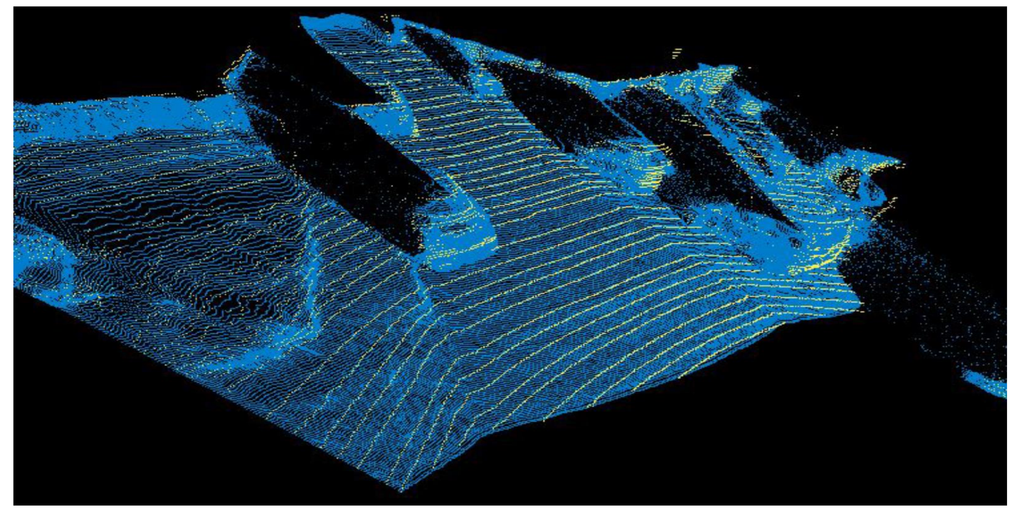
(depth) $z = Z(u, v)$

(width) $x = \frac{(u - c_U) \times z}{f_U}$

(height) $y = \frac{(v - c_V) \times z}{f_V}$

c_U, c_V: image center

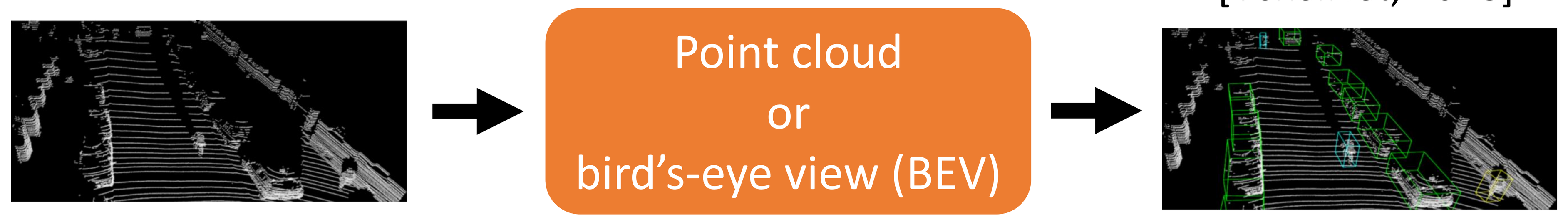
f_U, f_V: focal lengths



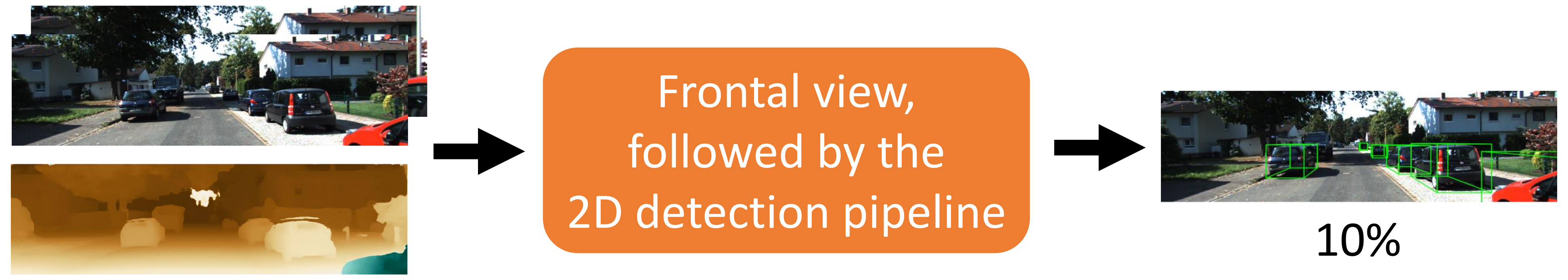
- ***Stereo depth** vs. **LiDAR**: points are surprisingly consistent!*

Data representation matters!

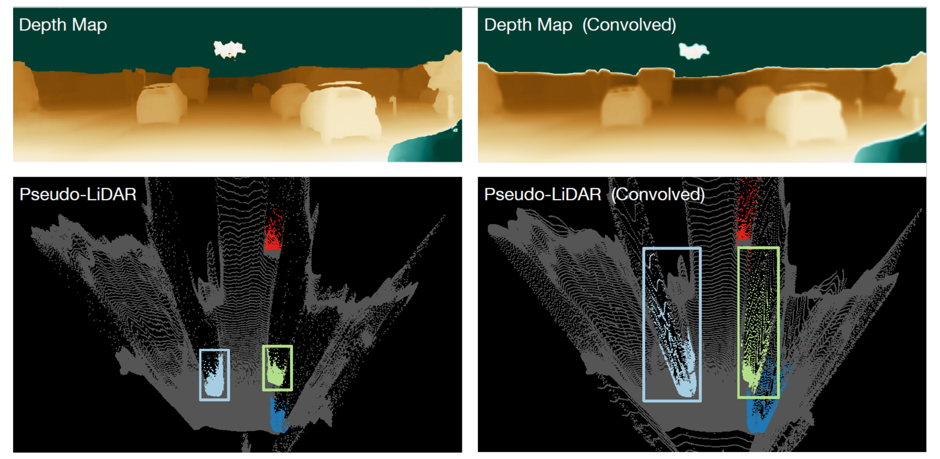
- LiDAR-based 3D detectors



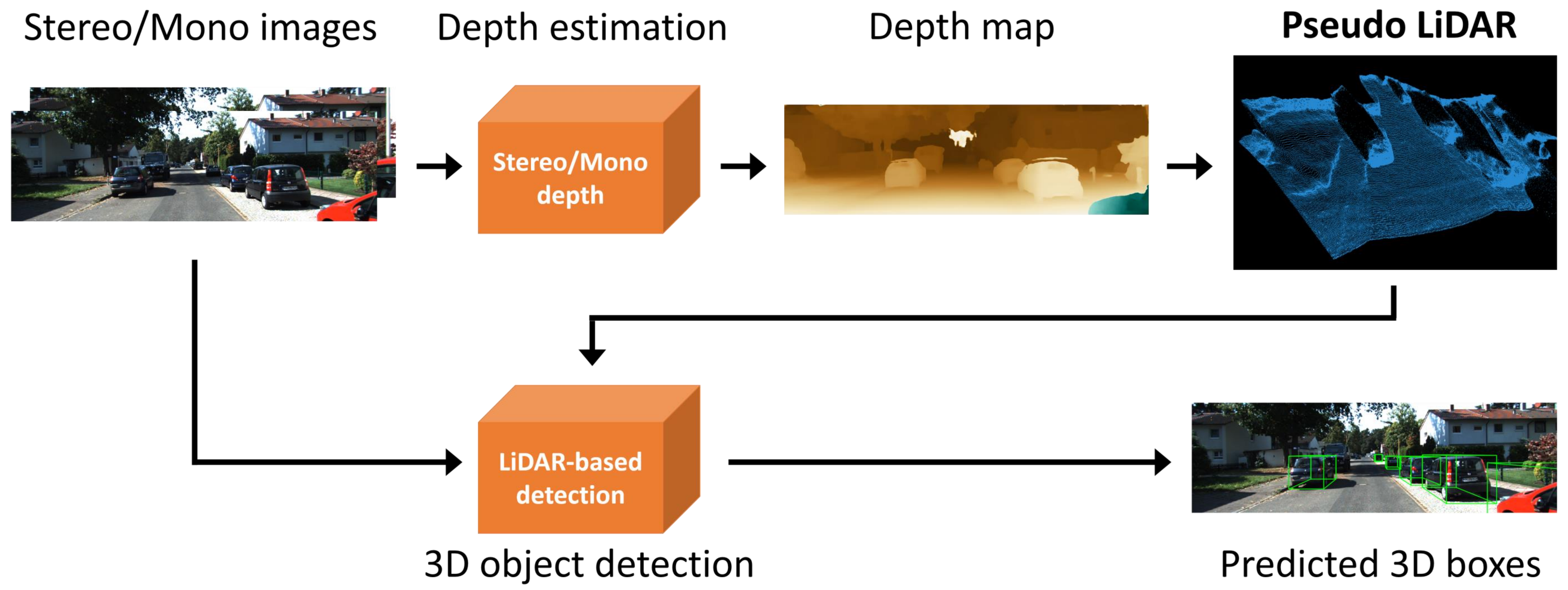
- Image-based 3D detectors



- Issues with convolution from the frontal view:
  - ➢ Object sizes vary with depth.
  - ➢ Neighboring pixels may be far-away in 3D, making it hard for convolutional networks to precisely localize objects in 3D.

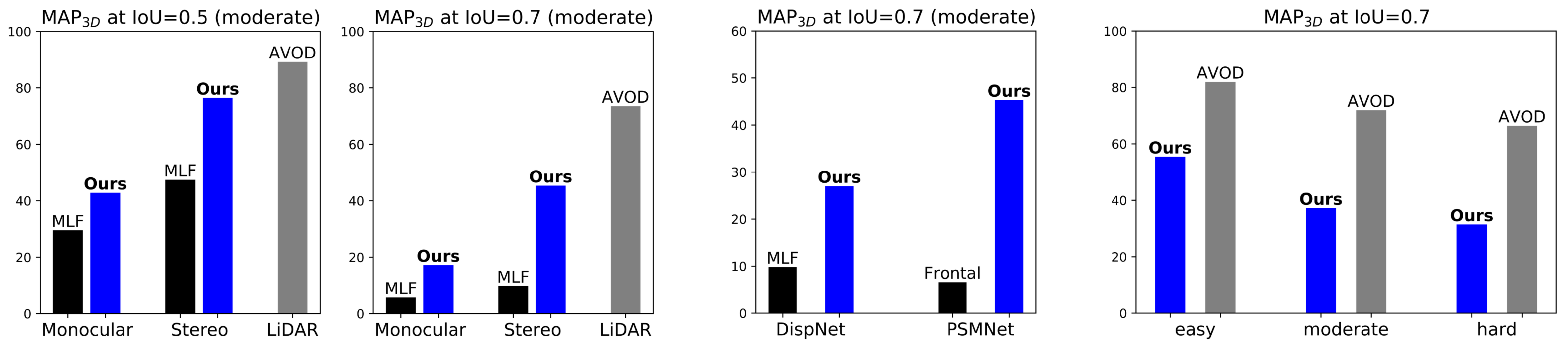


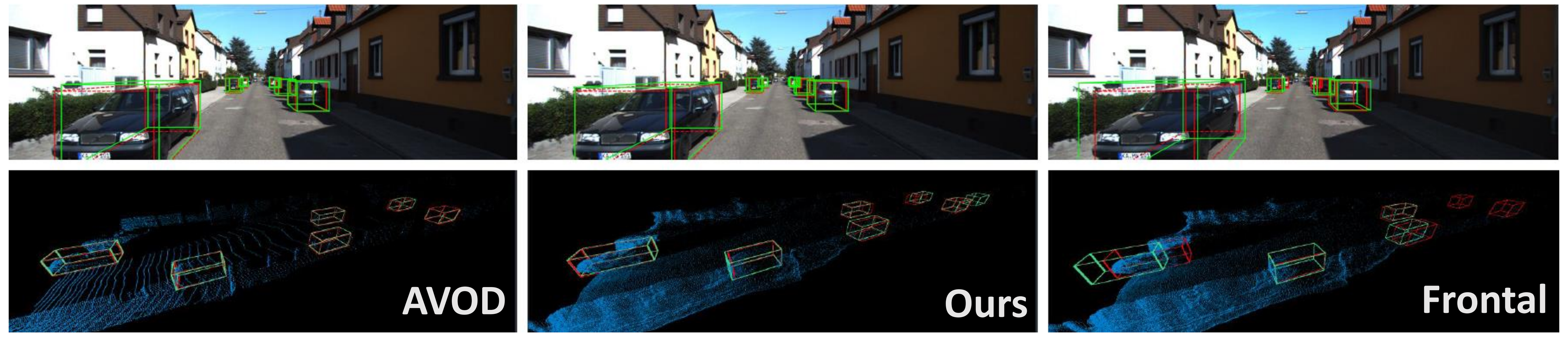
Proposed pseudo-LiDAR framework



Experiments

- **Dataset:** KITTI object detection (4K/4K/8K images for train/val/test), focusing on "car"
- **Our approach:** PSMNet [1]/Dorn [2] for Stereo/monocular depth + AVOD detector [3]





Discussion, conclusion, and future work

- The historic performance gap between image- and LiDAR-based approaches may be more due to differences in processing rather than data quality.
- Pseudo-LiDAR largely improves image-based 3D detection, and may be a promising alternative (or complimentary) to LiDAR.
- **Future directions:** improve stereo depth for far-away objects and computational efficiency
- **Current progress:**
  - ➢ Novel stereo depth network: 45.3% → **50.4%**
  - ➢ Fuse stereo with 4-line LiDAR: 50.4% → **63.4%**
- **Code:** https://github.com/mileyan/pseudo_lidar

[1] Pyramid stereo matching network. In CVPR, 2018.
[2] Deep ordinal regression network for monocular depth estimation. In CVPR, 2018.
[3] Joint 3d proposal generation and object detection from view aggregation. In IROS, 2018.
[4] Multi-level fusion based 3d object detection from monocular images. In CVPR, 2018.
[5] A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In CVPR, 2016.

Acknowledgement: This research is supported in part by grants from the NSF (III-1618134, III-1526012, IIS-1149882, IIS-1724282, TRIPODS-1740822), the Office of Naval Research DOD (N00014-17-1-2175), and the Bill and Melinda Gates Foundation. We are thankful for generous support by Zillow and SAP America Inc. We thank Gao Huang (Tsinghua University) for helpful discussion.