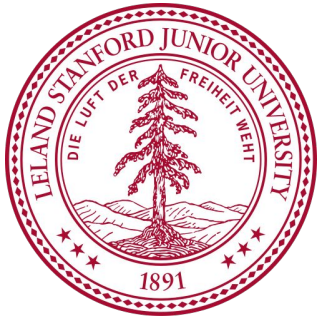
CS231A: Computer Vision, From 3D Reconstruction to Recognition
Winter 2021

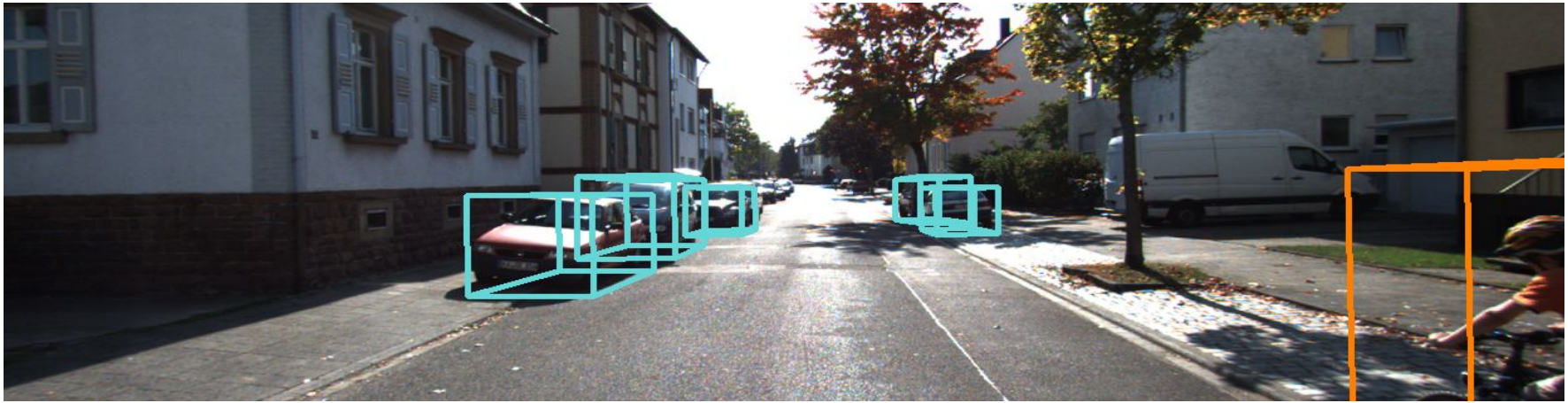
3D Object Detection for Self-Driving Cars



Manoj Rajagopalan | Martin Freeman | Shoaib Lari

March 15th, 2021

- Introduction
- Dataset
- Method
- Results
- Analysis
- Future Work



Background

- Data: Stereo RGB images and Lidar point-clouds

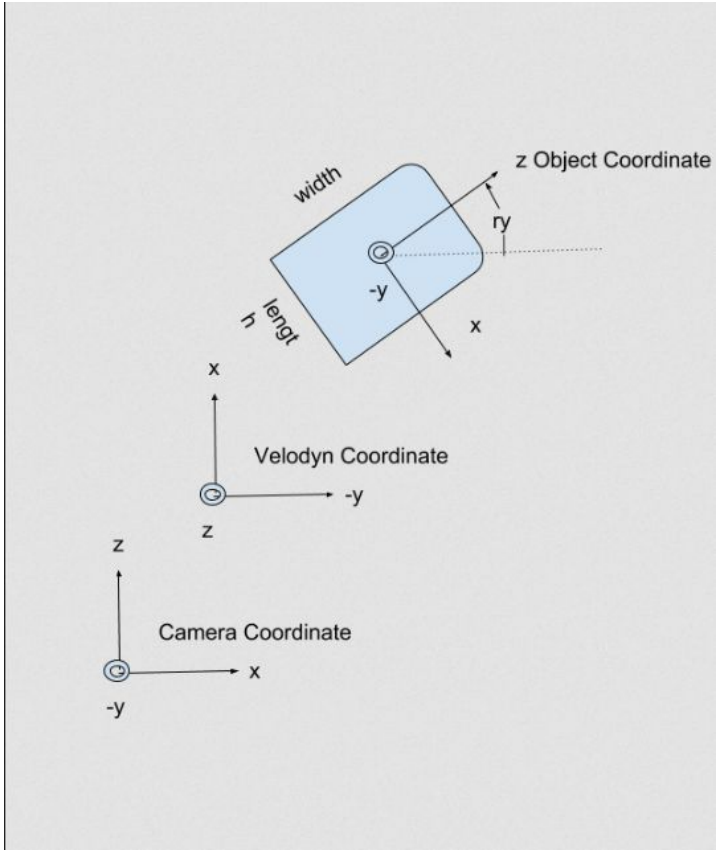
- Grid-based methods
  - Transform irregular point clouds to 3D voxels or 2D BEV (Bird's Eye View) or RV (Range View)
  - Process by 3D or 2D CNNs

- Point-based methods
  - Directly extract discriminative features from raw point clouds

- Comparison
  - Grid-based: more computationally efficient, but less accurate
  - Point-based: higher computation cost, but larger receptive field

Dataset: Kitti Dataset

Dataset: Kitti Dataset



camera_2 image (.png),
camera_2 label (.txt),
calibration (.txt),
velodyne point cloud (.bin),

7481 training images
7518 test images

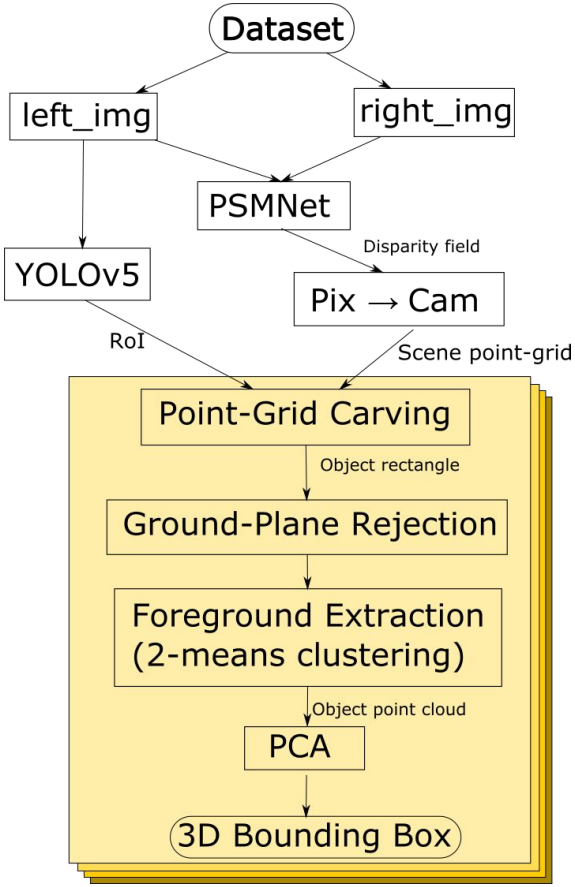
Categories:

1. Car
2. Pedestrian
3. Bike

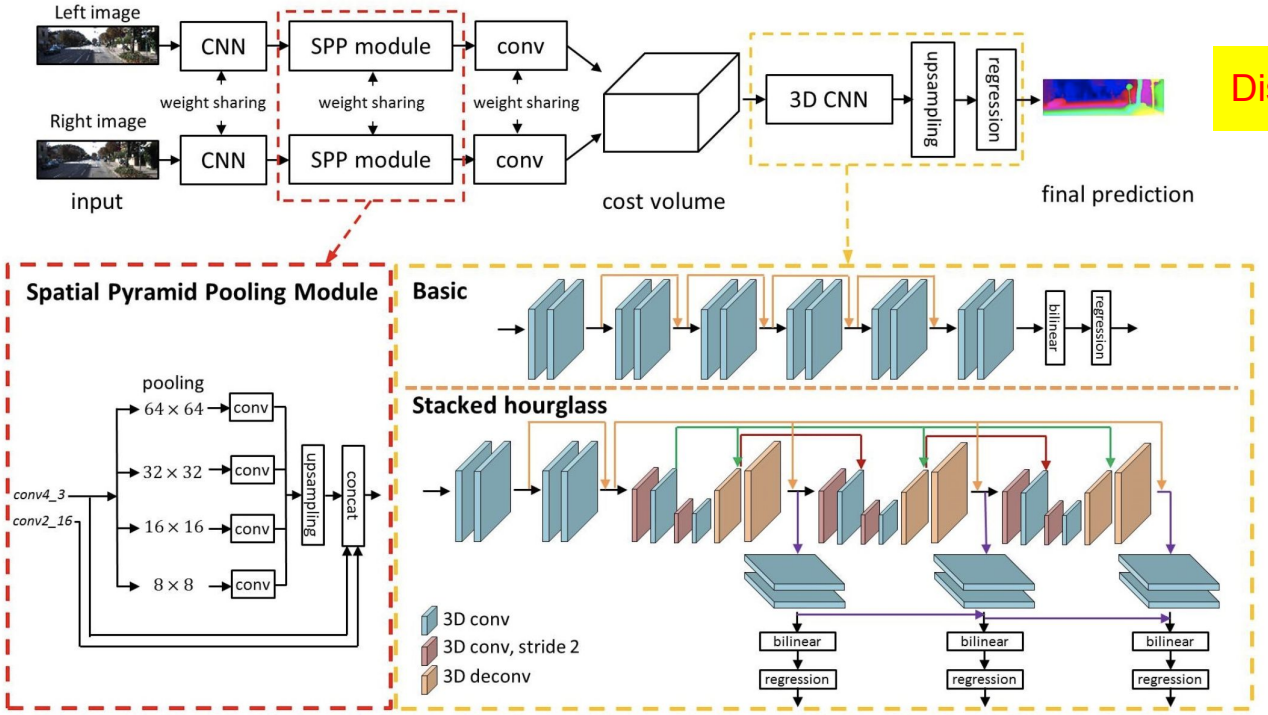
Method

- Dataset: KITTI 2015 stereo images
- PSMNet: Disparity (d) estimation
- Pix → Cam:
  - Depth calculation: $z_{cam} = b \times f / d$
  - $[x_{cam}, y_{cam}, z_{cam}]^T = z_{cam} \ K^{-1} \ [x, y, 1]^T$

- YOLOv5: 2D object identification
  - RoI (region of interest) rects
- Point-Grid Carving
  - Cut out YOLO rects from scene point-grid
- Ground plane rejection: drop y > 1.6m
  - KITTI cameras at 1.65m above ground
- Foreground extraction:
  - K-means clustering (K=2) over distances to each obj point (pick closer cluster)
- Principal Component Analysis (PCA)
  - Identify orientation of oblong object



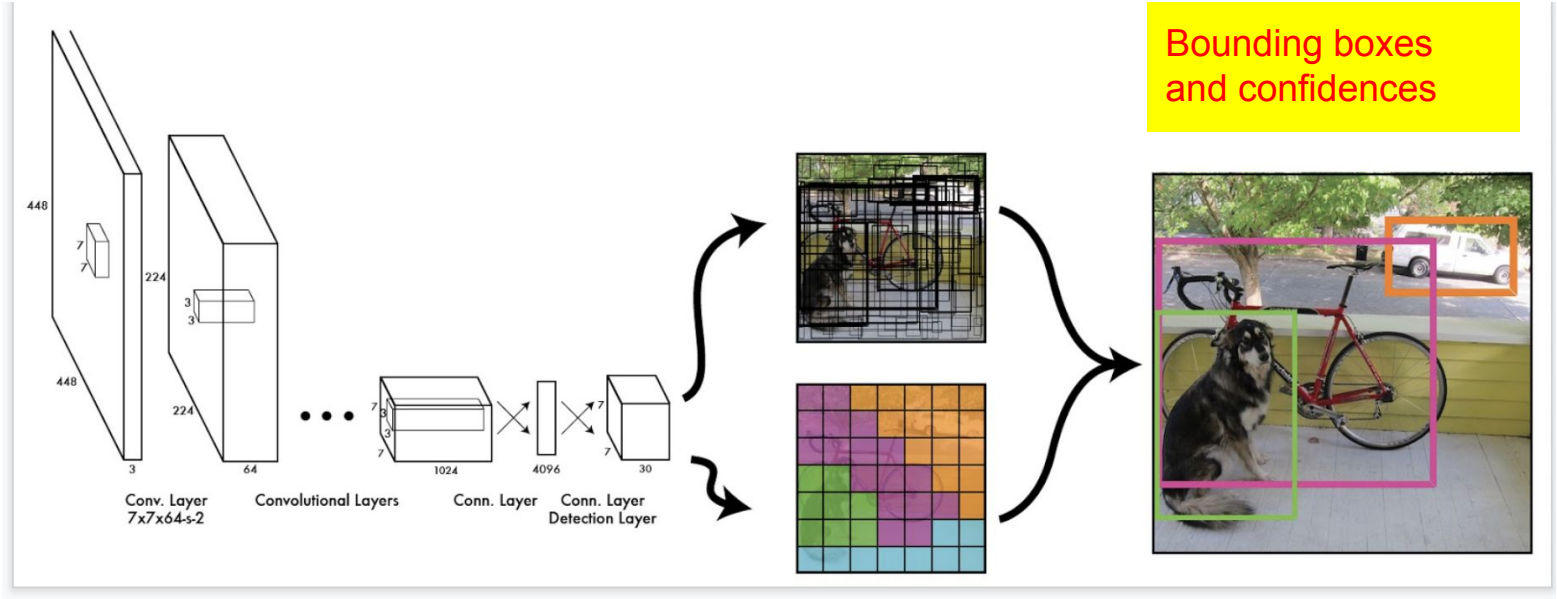
Depth Estimation: PSMNet



Disparity field

J-R. Chang et. al., "**Pyramidal Stereo Matching** Network", CVPR 2018

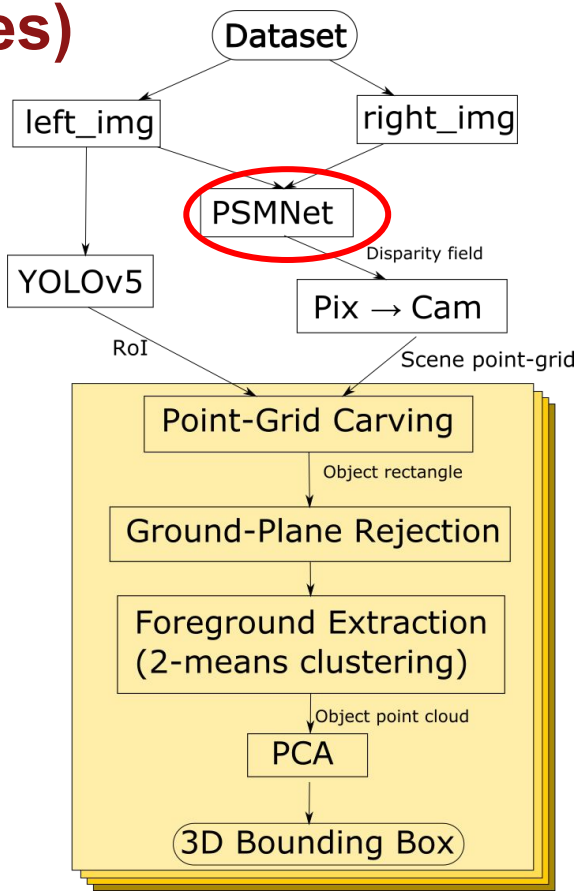
Object (Region) Identification: YOLOv5



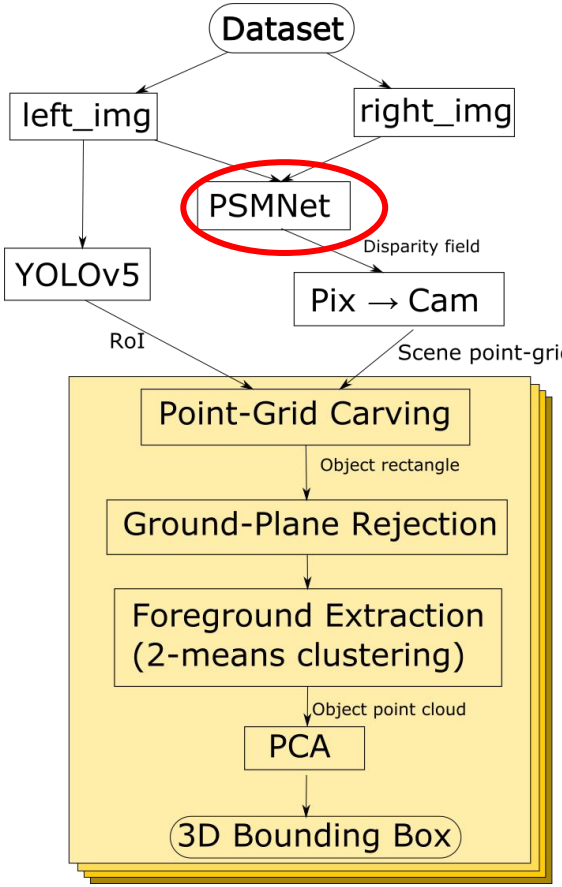
Redmon et. al, “You only look once”, CVPR 2016

Results: Preview

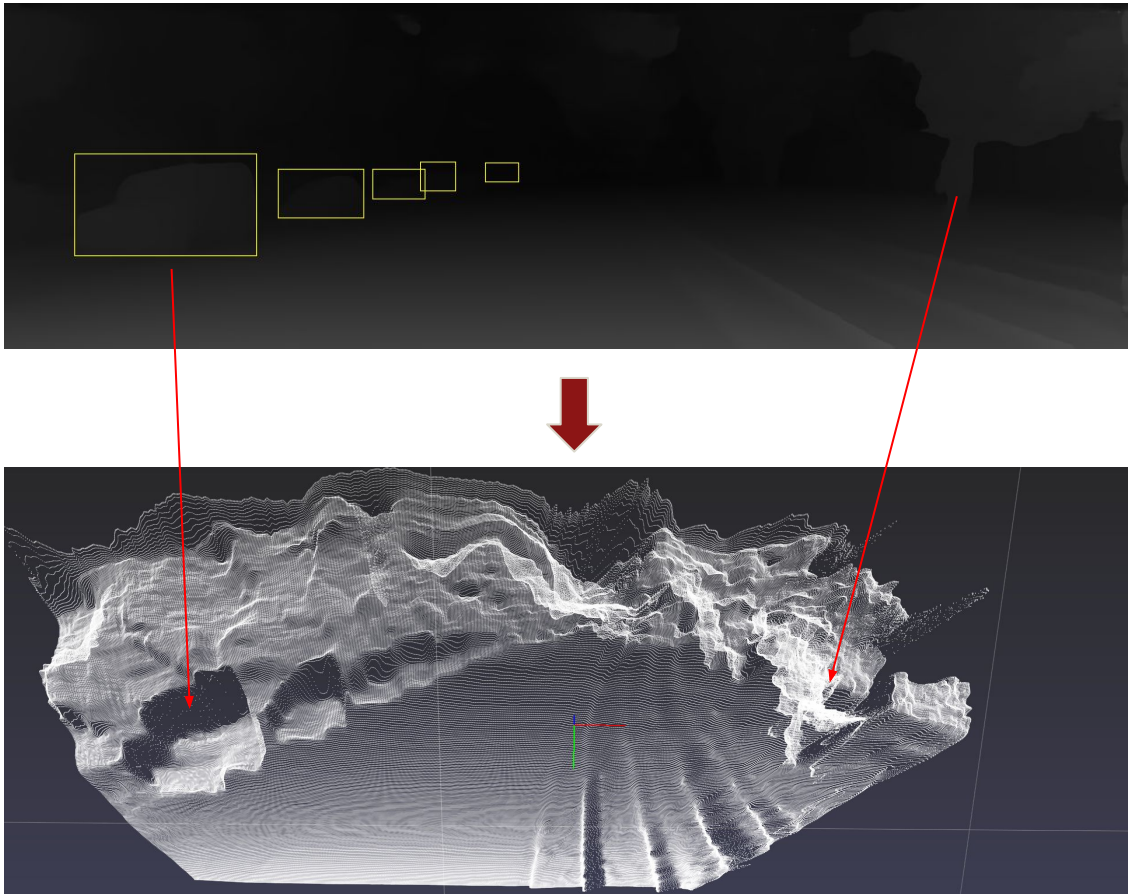
Depth Estimation: PSMNet (Disparities)

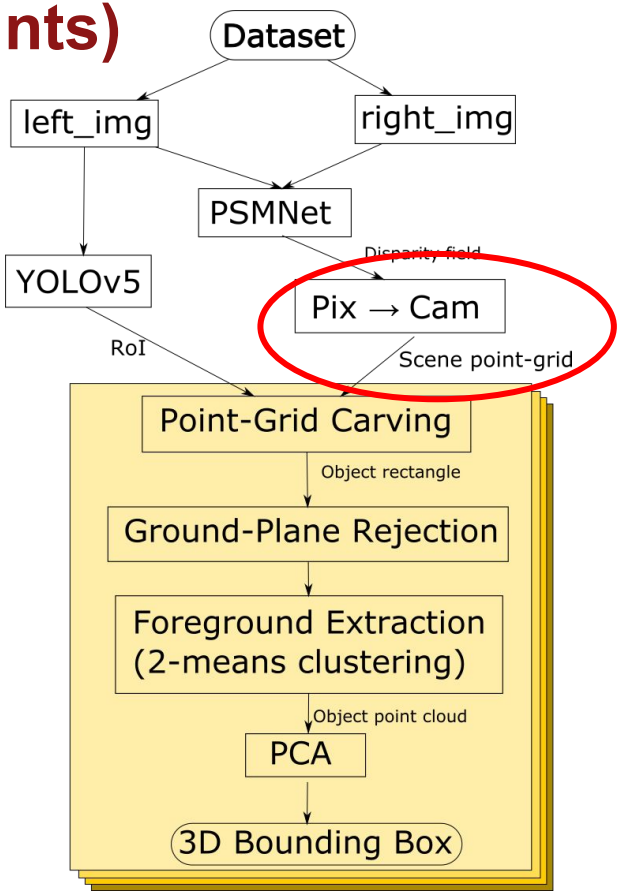


PSMNet: Better Demo

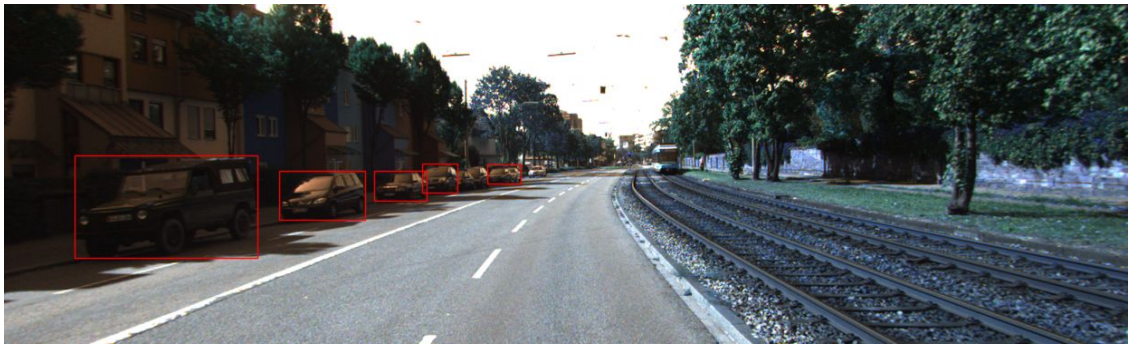


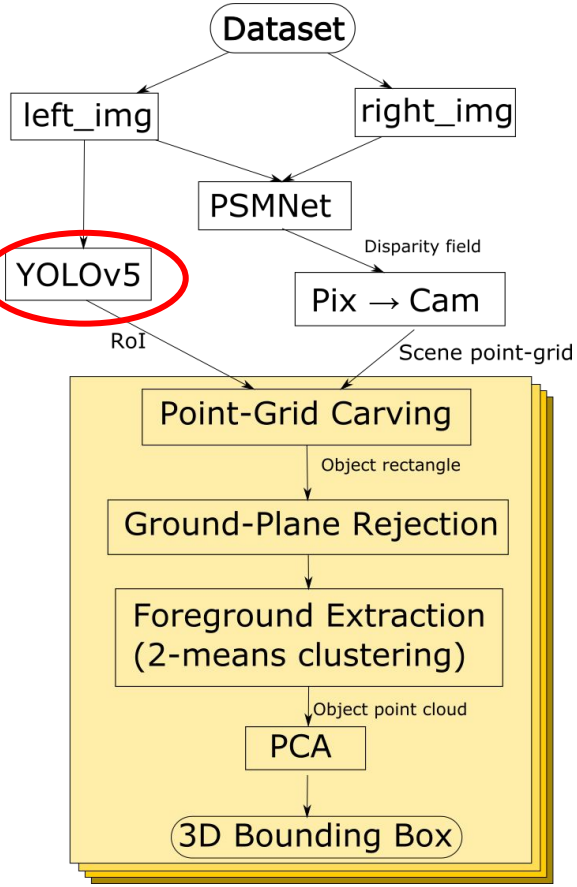
Pix → Cam (Disparity → Depth → Points)



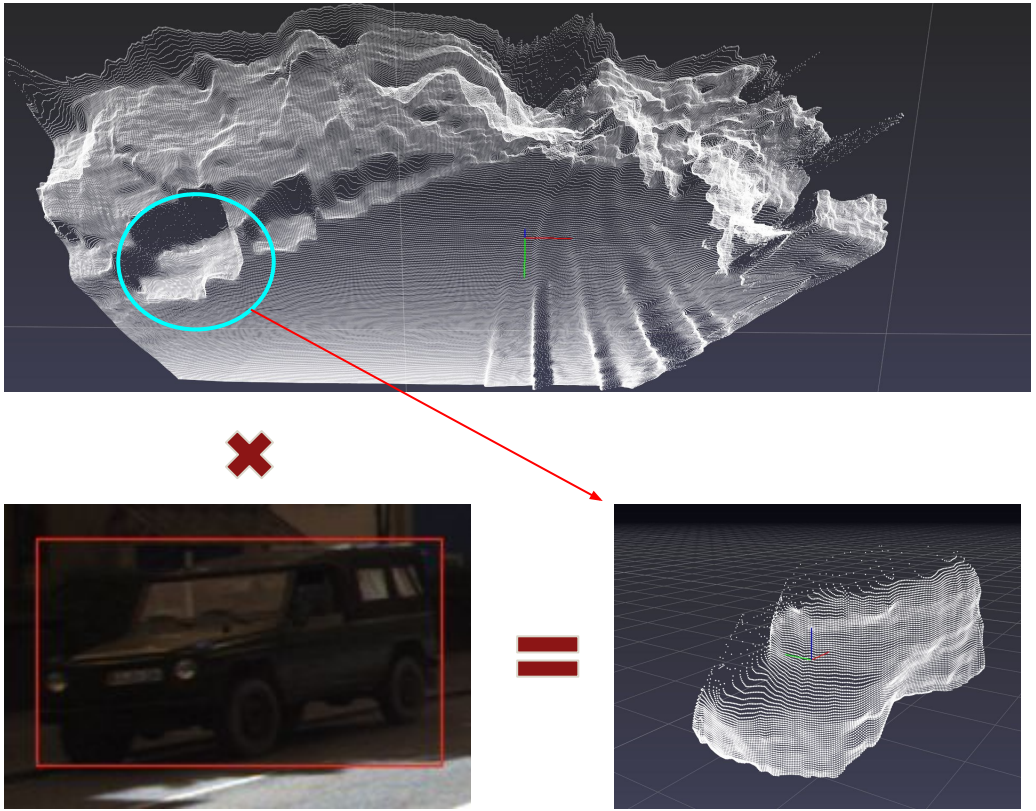


Object Identification: YOLOv5



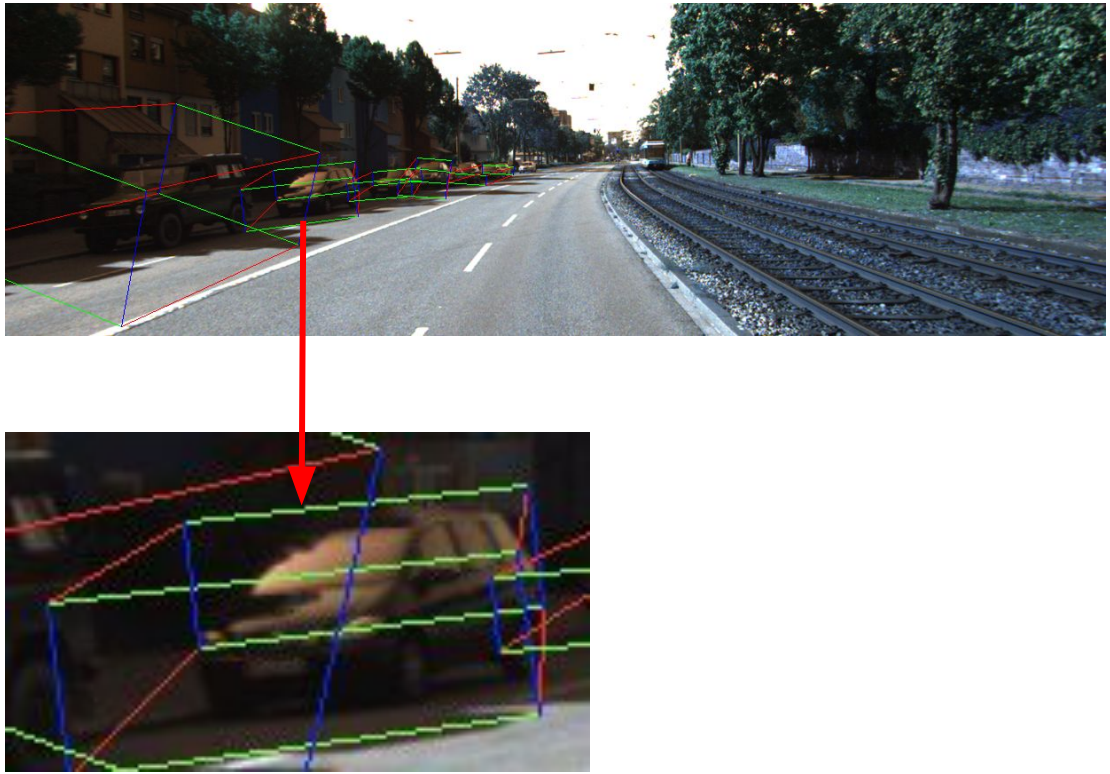


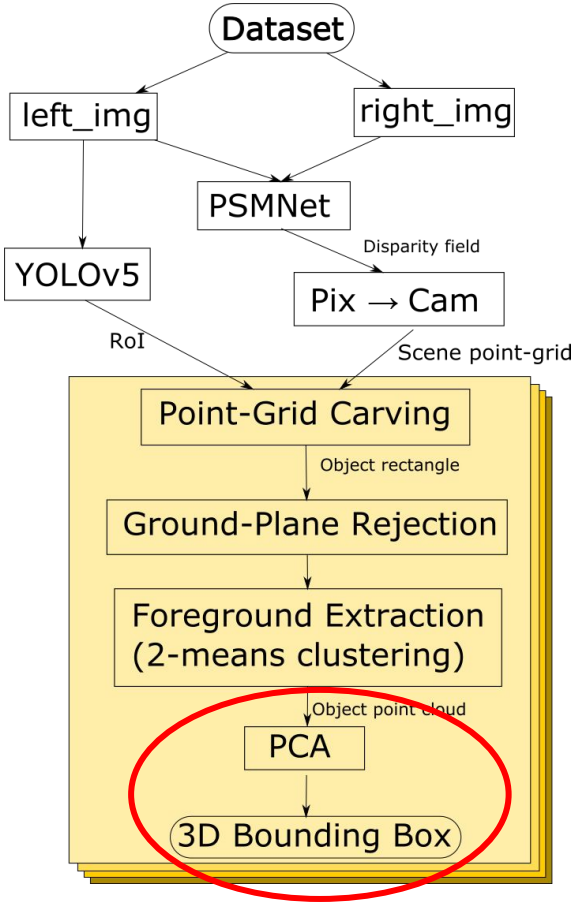
Point-Grid Carving → Cluster



Manoj Rajagopalan, Martin Freeman, Shoaib Lari

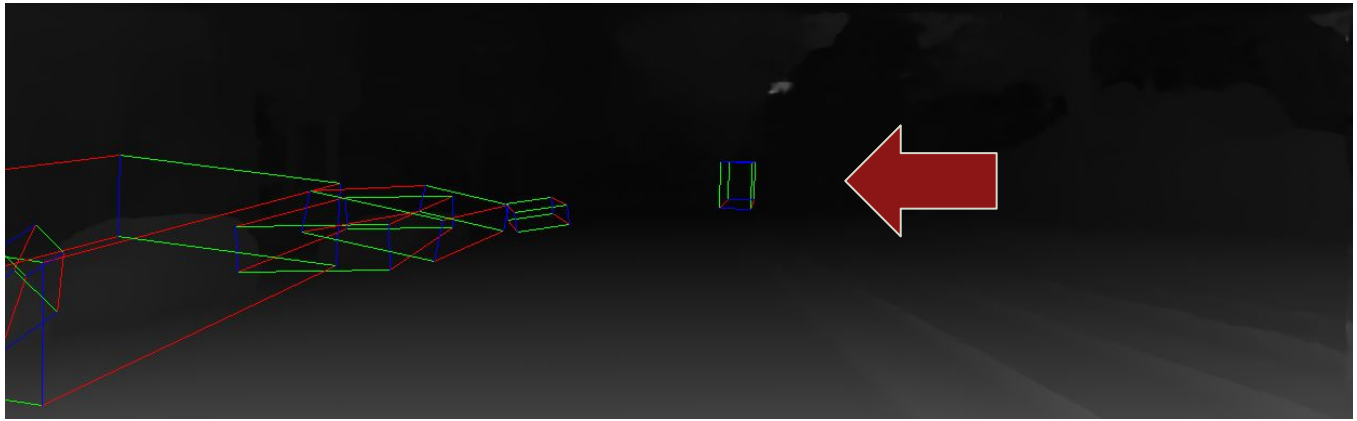
PCA: 3D oriented bounding boxes





Results: Recap

Other classes: Trains



Analysis

- 3D Bounding Boxes are reasonably positioned
  - Not well-fit and oriented
- K(=2) means clustering is important to get this close
  - Background points within RoI cause big distortion
- PCA provides only crude estimate of orientation
  - Fails as object approaches edges
- PSMNet noisy! (qualitatively good)
  - Min ~ 1.2 (328 m) , Max ~ 133 (2.9 m)
  - Never zero! (infinite objects)
- YOLOv5 identifies many object classes
  - Person, motorcycle, car, bus, train, truck (in this project)
- Runs at ~ 2-3 seconds / frame
  - Image dims: 1242 x 345 (each)

Object Identification: YOLO vs. Pointcloud Approaches

Stereo Images

↓

Disparity Map

↓

Dense pointcloud

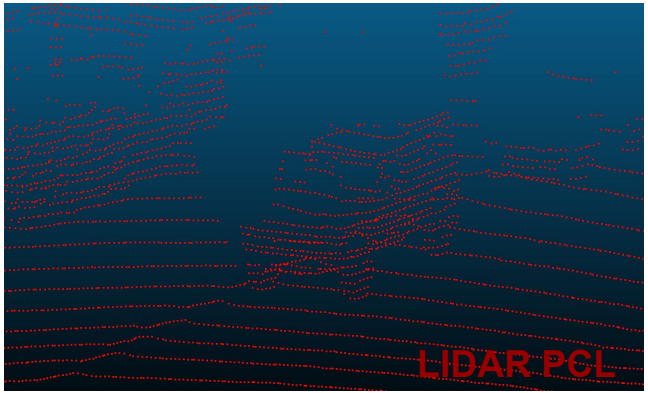
↓

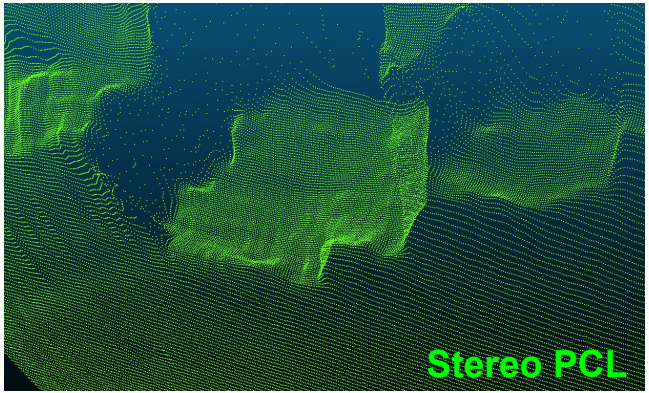
PV-RCNN Detector

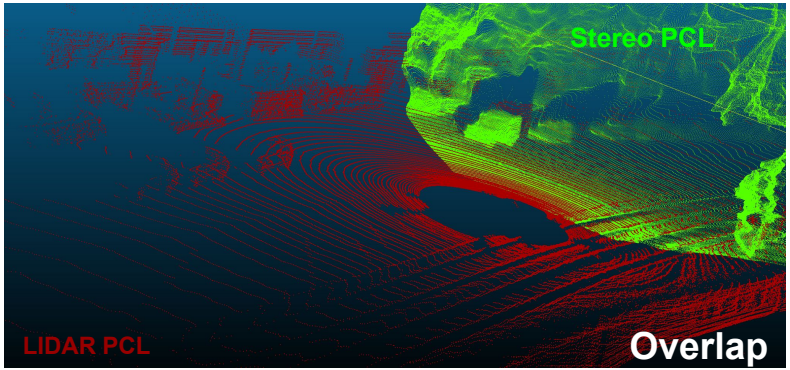
↓

CNN-based detections (bounding boxes)

Object Identification: YOLO vs. Pointcloud Approaches







Future Work

- Better depth estimation
  - Try OpenCV Semi-Global Block Matching (SGBM) for disparities
  - Retrain PSMNet (fine-tune)
  - Try other network types
  - Fuse geometric and NN-based approaches

- Better bounding box fitting
  - Use car shapes/orientation (eg: perpendicular to ground) as prior (current PCA considers each object to be a fuzzy point cloud)

- Track objects using Kalman Filters
  - Multi-frame “smoothing” could help with NN noise

- Metric-based performance of LIDAR vs Stereo pointcloud
  - Must standardize pointclouds by trimming LIDAR scan FOV and range