RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles

Ramin Nabati Hairong Qi

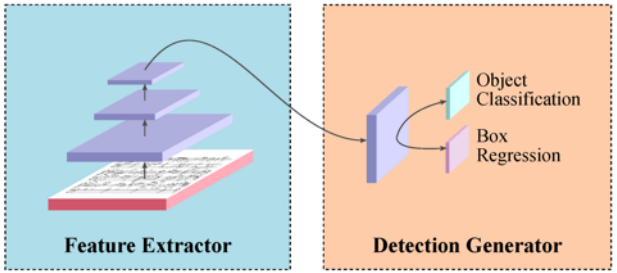
Department of Electrical Engineering and Computer Science
The University of Tennessee Knoxville, USA

THE UNIVERSITY OF TENNESSEE KNOXVILLE

ICIP 2019

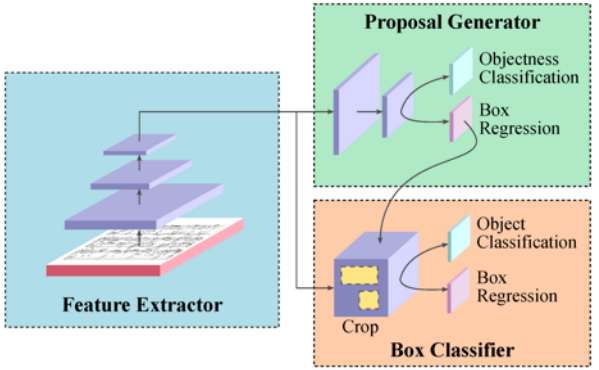
2D Object Detection

- One-Stage Detectors



(a) Basic architecture of a one-stage detector.

- Two-Stage Detectors

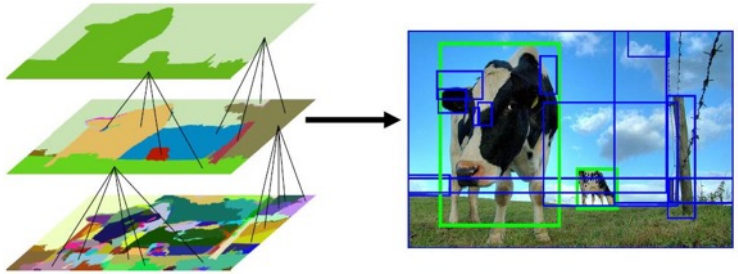


(b) Basic architecture of a two-stage detector.

Images from: A Baseline for General Music Object Detection with Deep Learning, Alexander Pacha et. al

Proposal Generation

- Task: Find candidate regions containing object
  - **Selective Search**[1]: hierarchical grouping of similar regions based on color, texture, size and shape compatibility.



1. Segmentation as Selective Search for Object Recognition, K. E. A. van de Sande et. al

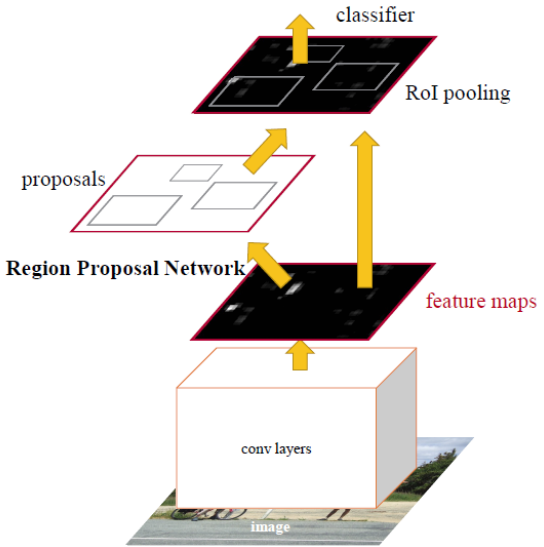
Proposal Generation

- Task: Find candidate regions containing object
  - **Selective Search**[1]: hierarchical grouping of similar regions based on color, texture, size and shape compatibility.
  - **Edge Boxes**[2]: bounding box proposals directly from edges

1. Segmentation as Selective Search for Object Recognition, K. E. A. van de Sande et. al
2. Edge Boxes: Locating Object Proposals from Edges, C. Lawrence et. al

Proposal Generation

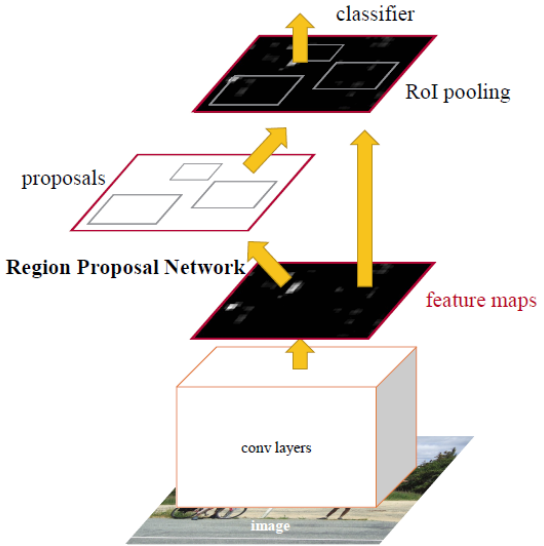
- Task: Find candidate regions containing object
  - **Selective Search**[1]: hierarchical grouping of similar regions based on color, texture, size and shape compatibility.
  - **Edge Boxes**[2]: bounding box proposals directly from edges
  - **RPN**[3]: fully convolutional network that simultaneously predicts object bounds and objectness scores



1. Segmentation as Selective Search for Object Recognition, K. E. A. van de Sande et. al
2. Edge Boxes: Locating Object Proposals from Edges, C. Lawrence et. al
3. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Shaoqing Ren et. al

Proposal Generation

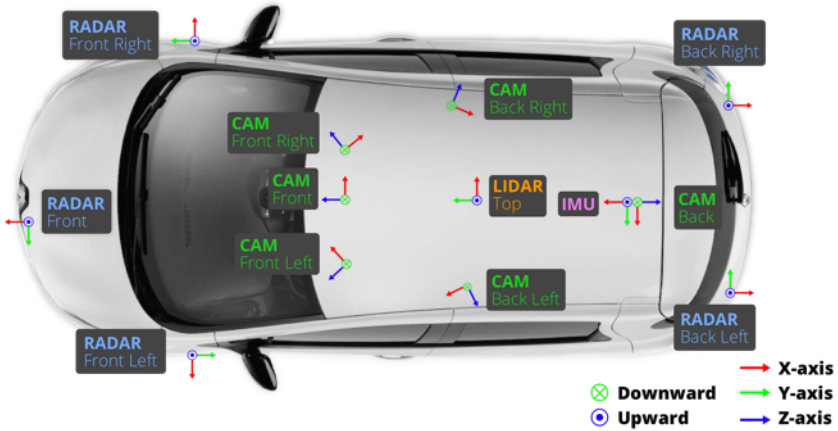
- Task: Find candidate regions containing object
  - **Selective Search**[1]: hierarchical grouping of similar regions based on color, texture, size and shape compatibility.
  - **Edge Boxes**[2]: bounding box proposals directly from edges
  - **RPN**[3]: fully convolutional network that simultaneously predicts object bounds and objectness scores
- Proposal generation is usually the bottleneck in two-stage algorithms

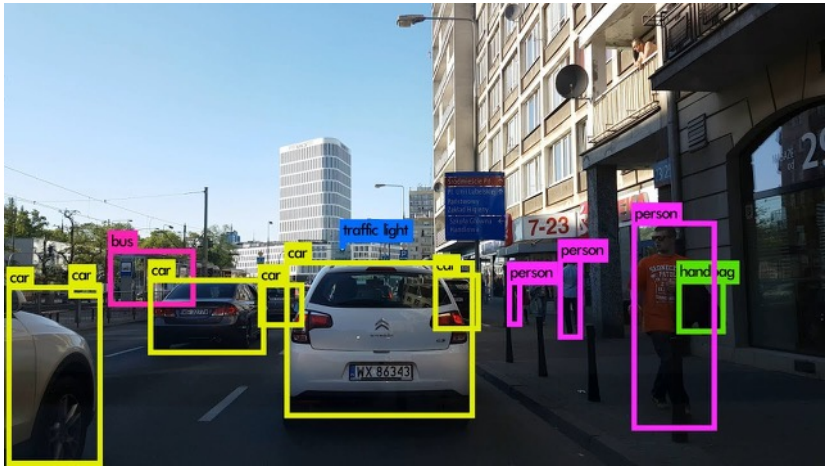


1. Segmentation as Selective Search for Object Recognition, K. E. A. van de Sande et. al
2. Edge Boxes: Locating Object Proposals from Edges, C. Lawrence et. al
3. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Shaoqing Ren et. al

Object Detection for Autonomous Vehicles

- Autonomous vehicles require real-time object detection
- Multiple sensors available on vehicle (camera, Radar, LIDAR, ...)
- Large areas of image usually do not contain any object of interest

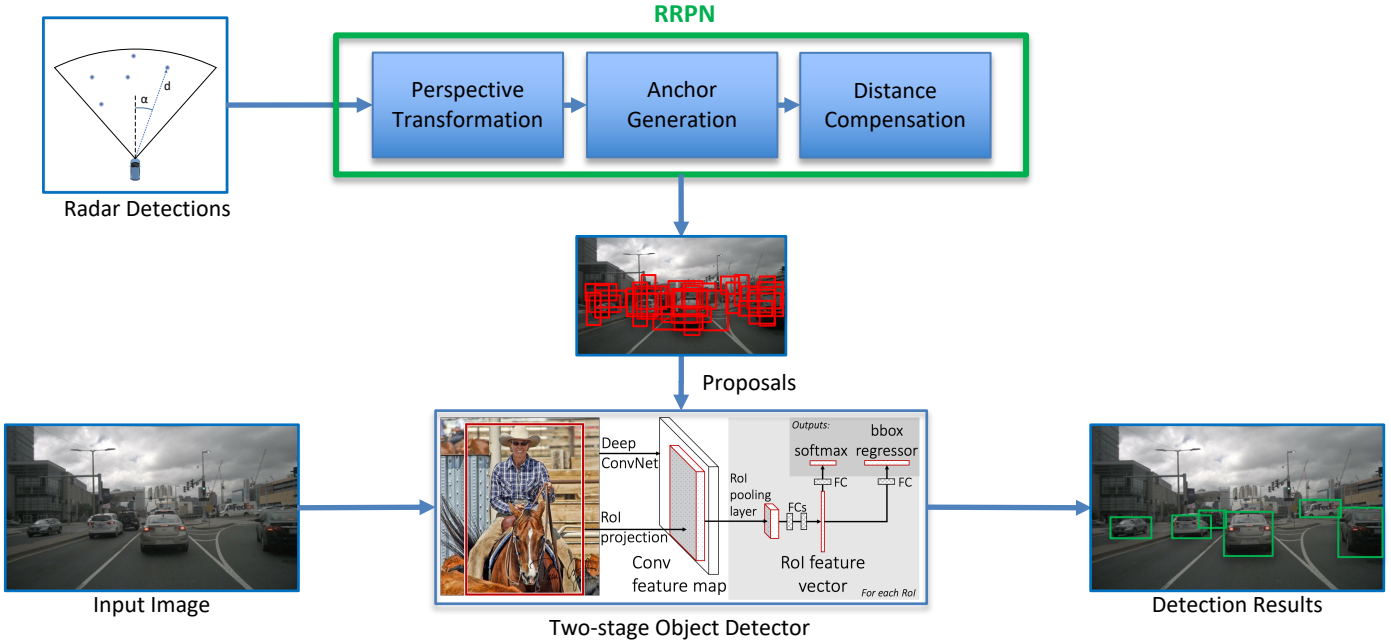




Left image: https://www.nuscenes.org
Right image: https://github.com/karolmajek/darknet

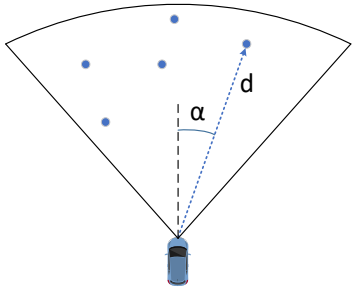
Radar Region Proposal Network (RRPN)

- Real-time algorithm generating object proposals from Radar detections



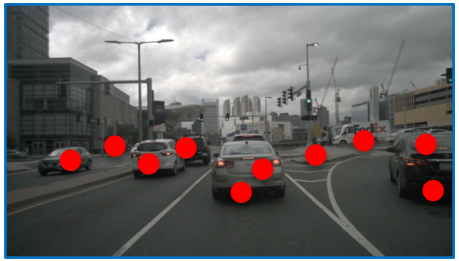
Radar Region Proposal Network (RRPN)

- Perspective Transformation
  - Radar detections are reported in Bird's Eye View (BEV) format
  - Detections need to be projected to the image coordinate system
  - Projection is done using a projection matrix H
  - H is obtained from calibrating the Radar and camera



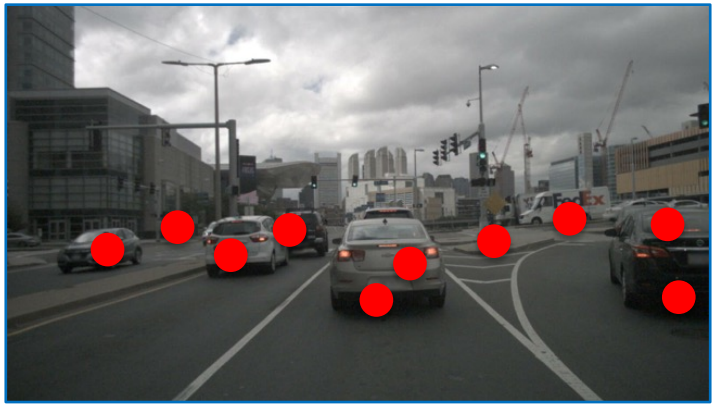
$$p = HP$$

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix}$$



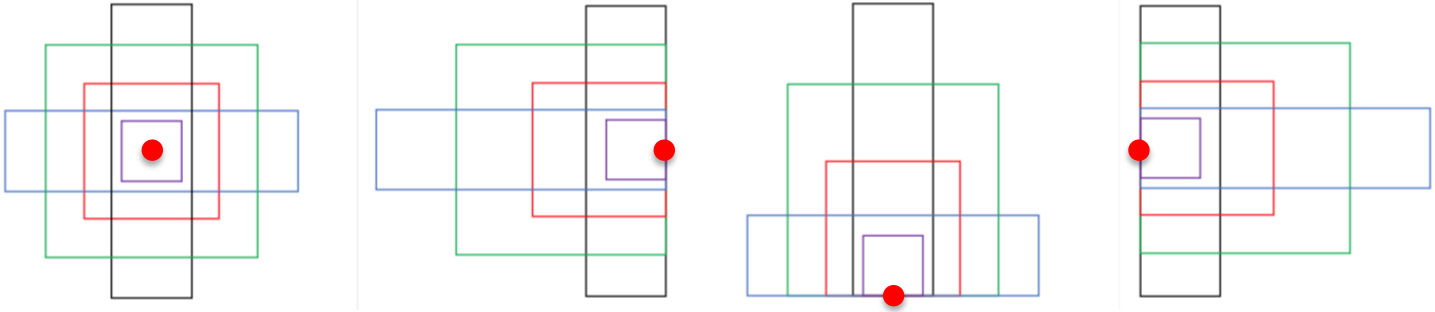
Radar Region Proposal Network (RRPN)

- Anchor Generation
  - Generate object proposals for every mapped Radar detection on image
  - Problems:
    - Radar detections are not always mapped to the center of objects
    - Radars usually do not report the size of the detected objects



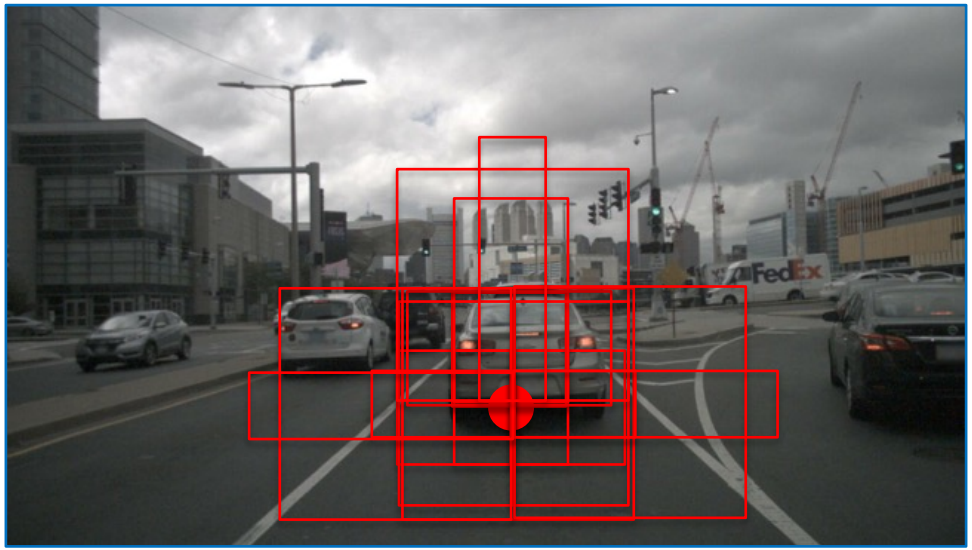
Radar Region Proposal Network (RRPN)

- Anchor Generation
  - Generate multiple bounding boxes with different size and aspect ratios at each point



Radar Region Proposal Network (RRPN)

- Anchor Generation
  - Generate multiple bounding boxes with different size and aspect ratios at each point



Radar Region Proposal Network (RRPN)

- Distance Compensation
  - Size of an object in image has an inverse relationship with distance
  - Radar detection has the range information for the detected object

Radar Region Proposal Network (RRPN)

- Distance Compensation
  - Size of an object in image has an inverse relationship with distance
  - Radar detection has the range information for the detected object
  - This information is used to adjust the size of the proposed bounding boxes

$$S_i = \alpha \frac{1}{d_i} + \beta$$

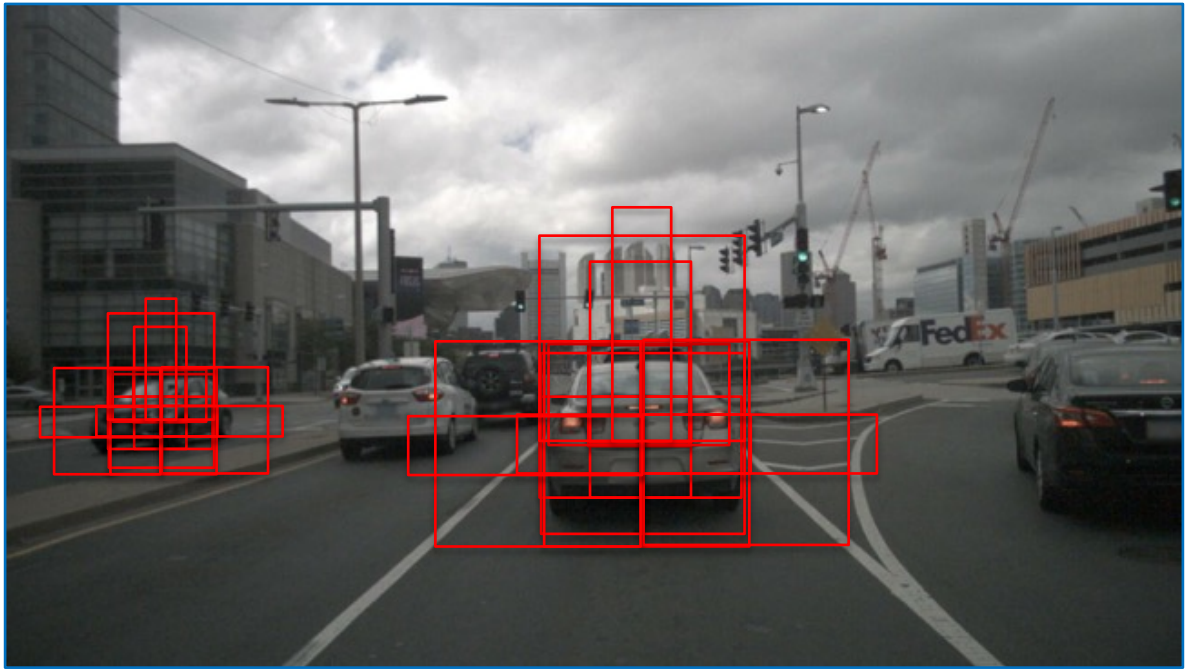
$$\underset{\alpha,\beta}{\operatorname{argmax}} \sum_{i=1}^{N} \sum_{j=1}^{M_i} \max_{1<k<A_i} IOU_{jk}^{i}(\alpha, \beta)$$

S_i : Scale factor for object i
d_i : Distance to object i
α, β: Parameters to adjust the scale

N : Number of training images
M_i : Number of ground truth bounding boxes in image i
A_i : Number of anchors generated in image i

Radar Region Proposal Network (RRPN)

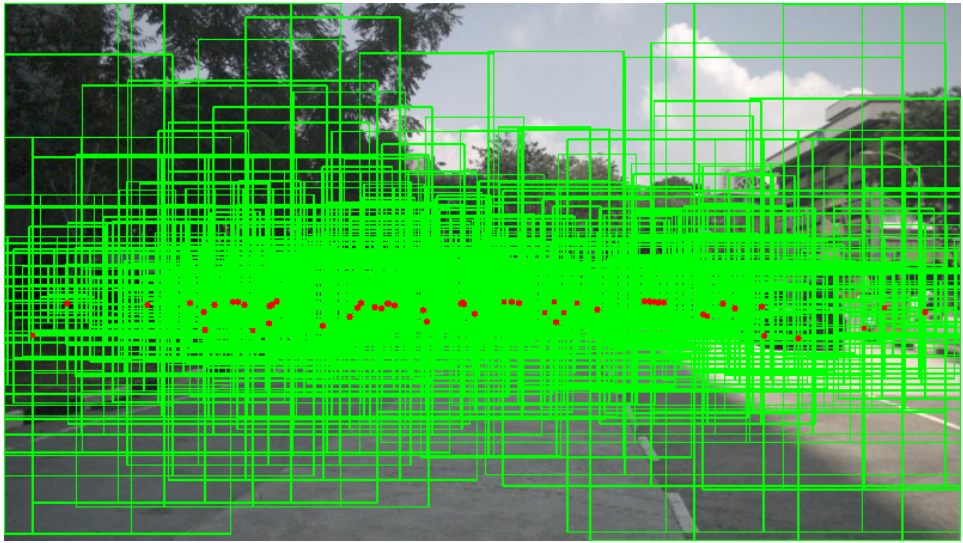
- Distance Compensation



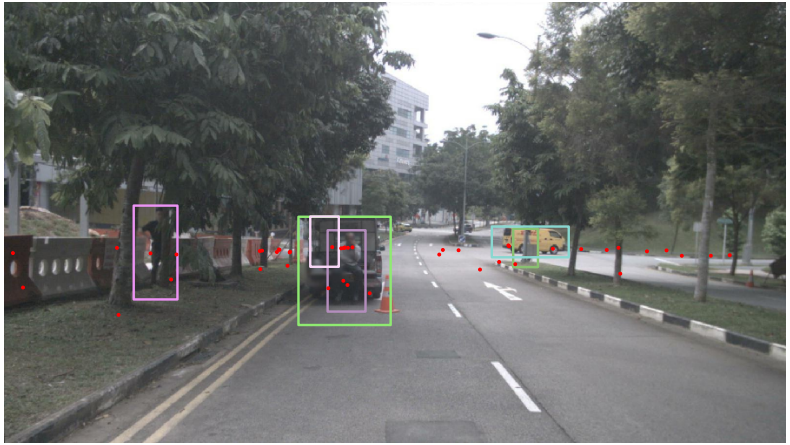
Experiments and Results

- Simulation Setup:
  - Object Detection Network: Fast-RCNN
  - Backbone: ResNet-101 and ResNeXt-101
  - Dataset: NuScenes
  - Compared to: Selective Search
- Running time:
  - Up to 100x faster than Selective Search

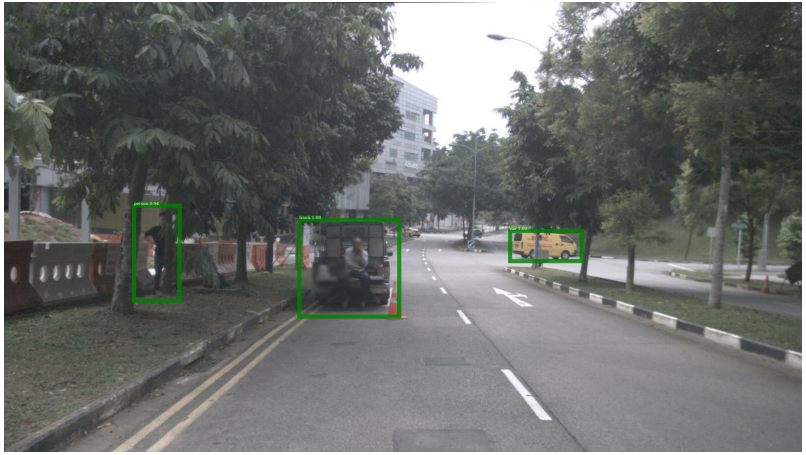
Results



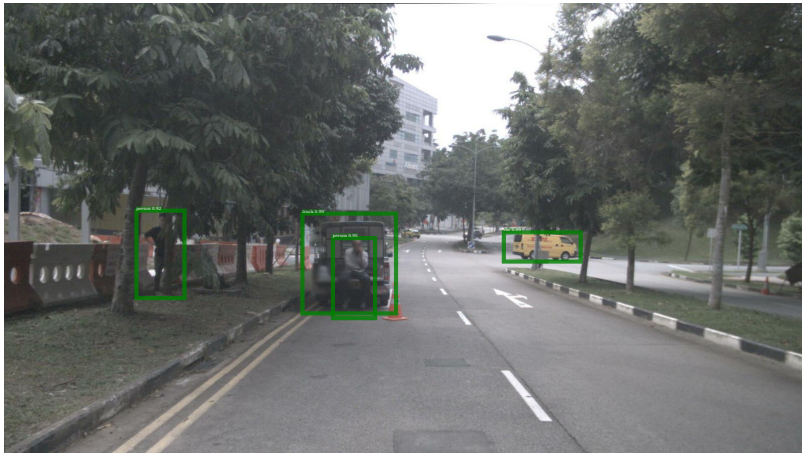
Results



Ground Truth



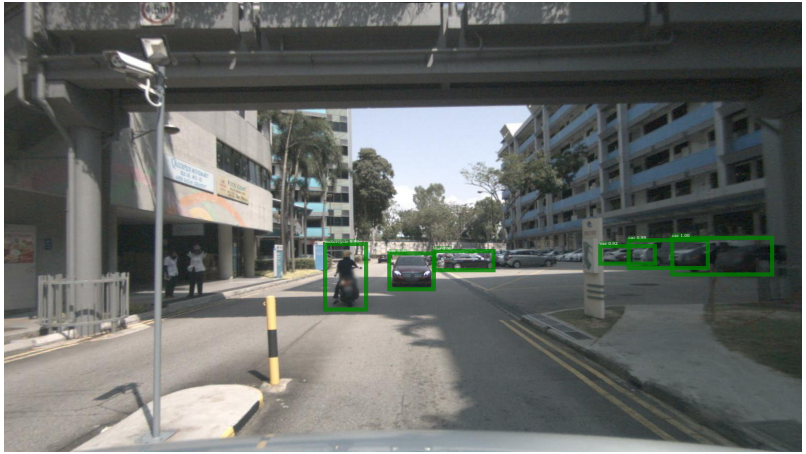
Selective Search



RRPN

Results

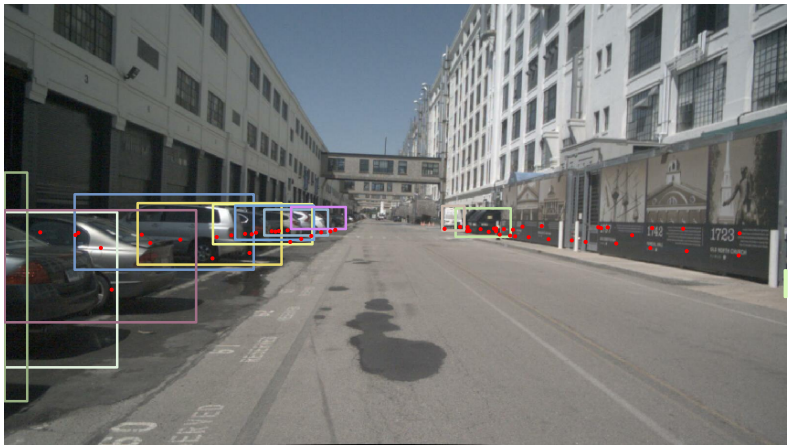
Ground Truth



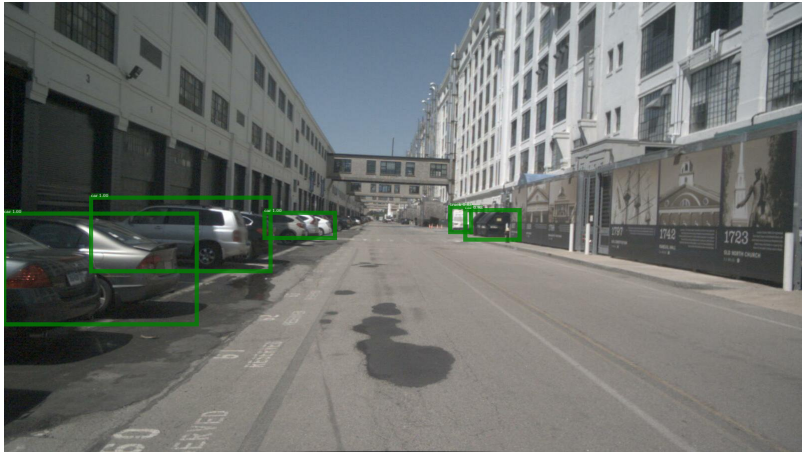
Selective Search

RRPN

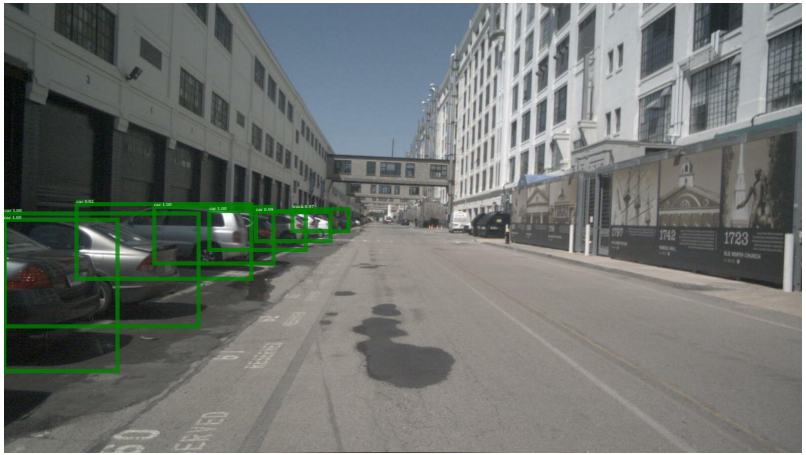
Results



Ground Truth



Selective Search



RRPN

Results

method	AP	AP50	AP75	AR	ARs	ARm	ARl
SS + X101 - F	0.368	0.543	0.406	0.407	0.000	0.277	0.574
SS + R101 - F	0.418	0.628	0.450	0.464	0.001	0.372	0.316
RRPN + X101 - F	0.419	**0.652**	0.463	0.478	**0.041**	0.406	0.573
RRPN + R101 - F	**0.430**	0.649	**0.485**	**0.486**	0.040	**0.412**	**0.582**
SS + X101 - FB	0.332	0.545	0.352	0.382	0.001	0.291	0.585
SS + R101 - FB	0.336	0.548	0.357	0.385	0.001	0.291	**0.591**
RRPN + X101 - FB	0.354	**0.592**	0.369	0.420	0.202	**0.391**	0.510
RRPN + R101 - FB	**0.355**	0.590	**0.370**	**0.421**	**0.211**	**0.391**	0.514

method	Car	Truck	Person	Motorcycle	Bicycle	Bus
SS + X101 - F	0.424	0.509	0.117	0.288	0.190	0.680
SS + R101 - F	**0.472**	**0.545**	0.155	0.354	0.241	**0.722**
RRPN + X101 - F	0.428	0.501	0.212	0.407	0.304	0.660
RRPN + R101 - F	0.442	0.516	**0.220**	**0.434**	**0.306**	0.664
SS + X101 - FB	0.390	0.415	0.122	0.292	0.179	0.592
SS + R101 - FB	0.392	0.420	0.121	0.291	0.191	**0.600**
RRPN + X101 - FB	0.414	**0.449**	**0.174**	0.294	**0.215**	0.579
RRPN + R101 - FB	**0.418**	0.447	0.171	**0.305**	0.214	0.572

F: Front camera only
FB: Front and back camera
SS: Selective Search

Questions?