Radar+RGB Fusion for Robust Object Detection in Autonomous Vehicles

Authors: Ritu Yadav, Axel Vierling, Karsten Berns
Presenter: Ritu Yadav

Session: Machine Learning in Object Detection

Affiliation
Robotics Research Lab
Department of Computer Science
Technical University of Kaiserslautern, Germany

Object Detection in Autonomous Vehicles

Need to be:

- Highly accurate
- Fast
- Reliable
- Affordable

Object Detection in Autonomous Vehicles

Need to be:

- Highly accurate
- Fast
- Reliable
- Affordable

Commonly Used Sensor: Camera

- High resolution, High accuracy.
- **But** not Reliable.
  - Sensitive towards Noise like reflection, sun glare, bad lighting, dusty or rainy weather conditions etc.

Object Detection in Autonomous Vehicles

Need to be:

- Highly accurate
- Fast
- Reliable
- Affordable

Commonly Used Sensor: Camera

- High resolution, Fast, High accuracy.
- **But** not Reliable.
  - Sensitive towards Noise like reflection, sun glare, bad lighting, dusty or rainy weather conditions etc.

Other Sensors:

- **LIDAR**: High resolution, 3D detection, **but** Sensitive towards Noise & Expensive.
- **Radar**: Low resolution, sparse data, **but** Robust towards noise, Long range & Low cost.

Object Detection in Autonomous Vehicles

Need to be:

- Highly accurate
- Fast
- Reliable
- Affordable

Commonly Used Sensor: Camera

- High resolution, Fast, High accuracy.
- **But** not Reliable.
  - Sensitive towards Noise like reflection, sun glare, bad lighting, dusty or rainy weather conditions etc.

Other Sensors:

- **LIDAR**: High resolution, 3D detection, **but** Sensitive towards Noise & Expensive.
- **Radar**: Low resolution, sparse data, **but** Robust towards noise , Long range & Low cost.

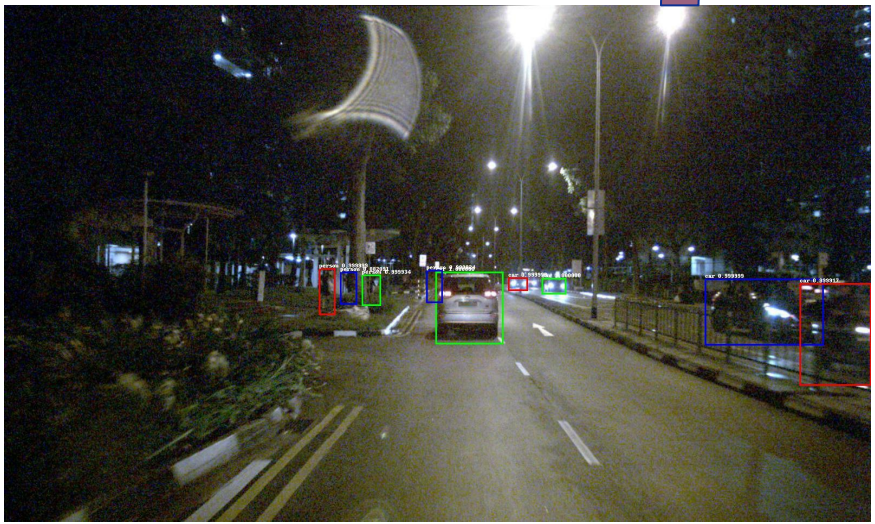
Impact of Noise on Vision based Detection Networks

Detection Results on image with no extra Noise

+Noise

-Performance

Detection Results on image with extra Noise



Proposed Solution

(RGB + Radar) Fusion

Radar Sensor

Radar Sensor Data:

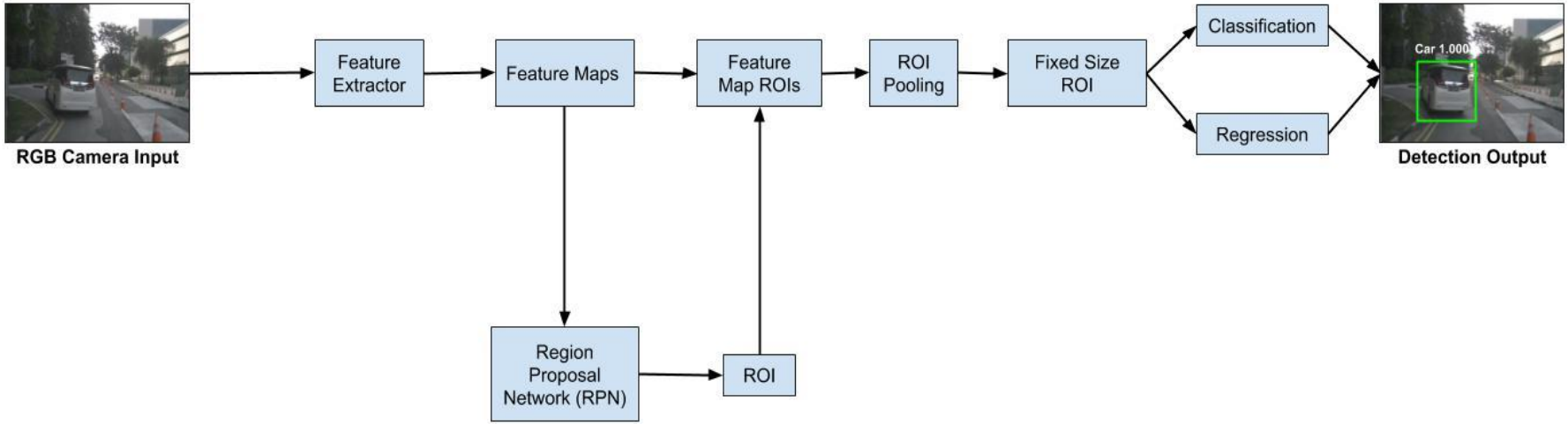
- Radar points(x,y,z)
- Point at interesting objects
- Robust
- Sparse

Radar +RGB Sensor Fusion:

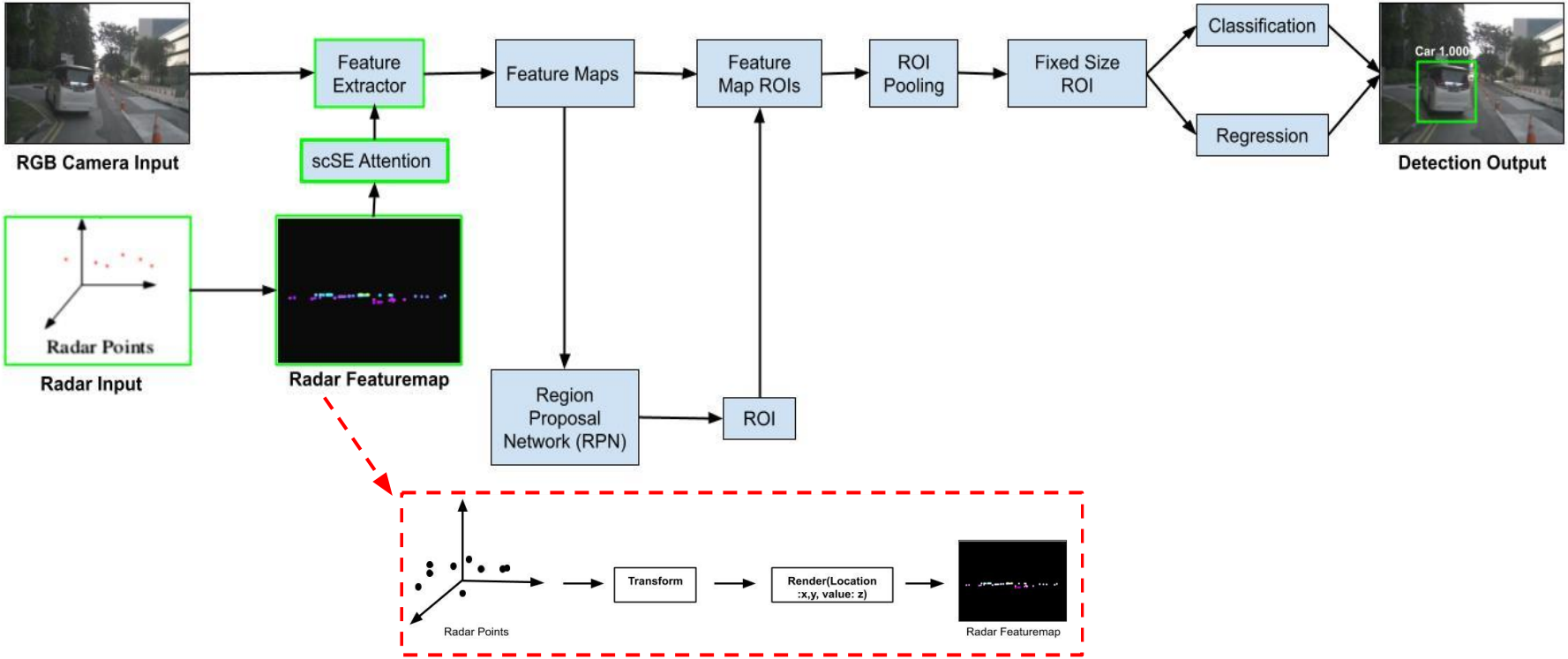
1. Attentive Feature Level Fusion
2. Data Level Fusion

Image with superimposed radar data.

Two-Stage Object Detector with Camera Input



Radar Attentive Feature level Fusion



Radar Data Level Fusion

RGB Camera Input

Feature Extractor

Feature Maps

Feature Map ROIs

ROI Pooling

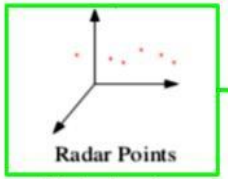
Fixed Size ROI

Classification

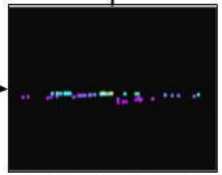
Regression

Detection Output

scSE Attention



Radar Input

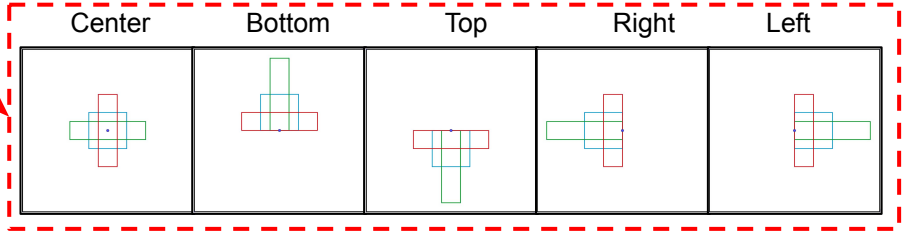


Radar Featuremap

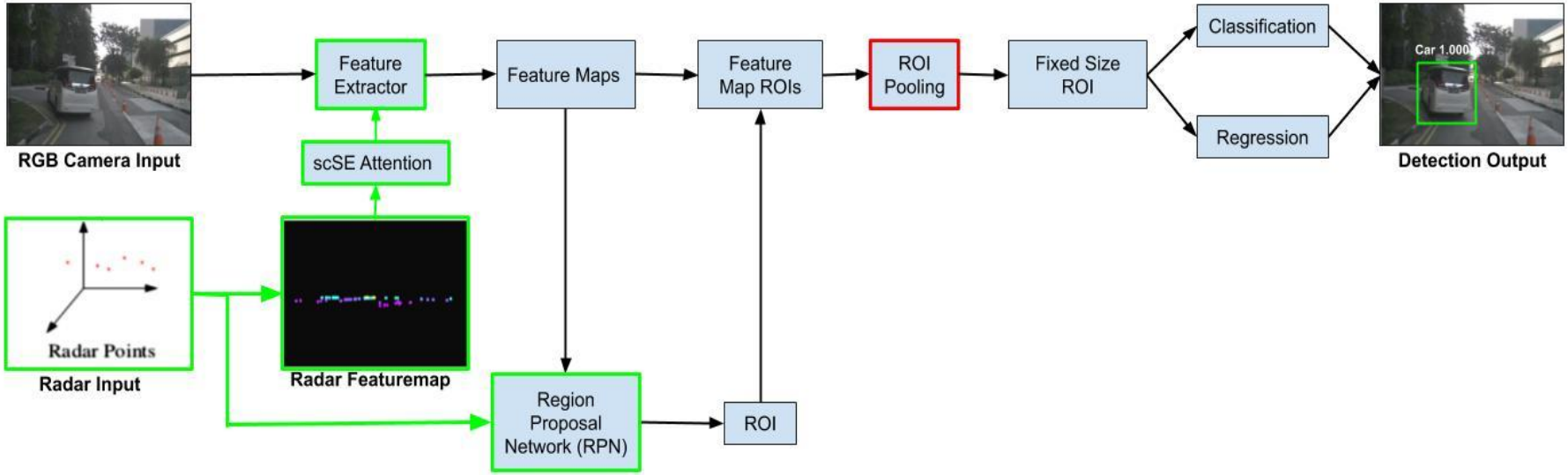
Region Proposal Network (RPN)

ROI

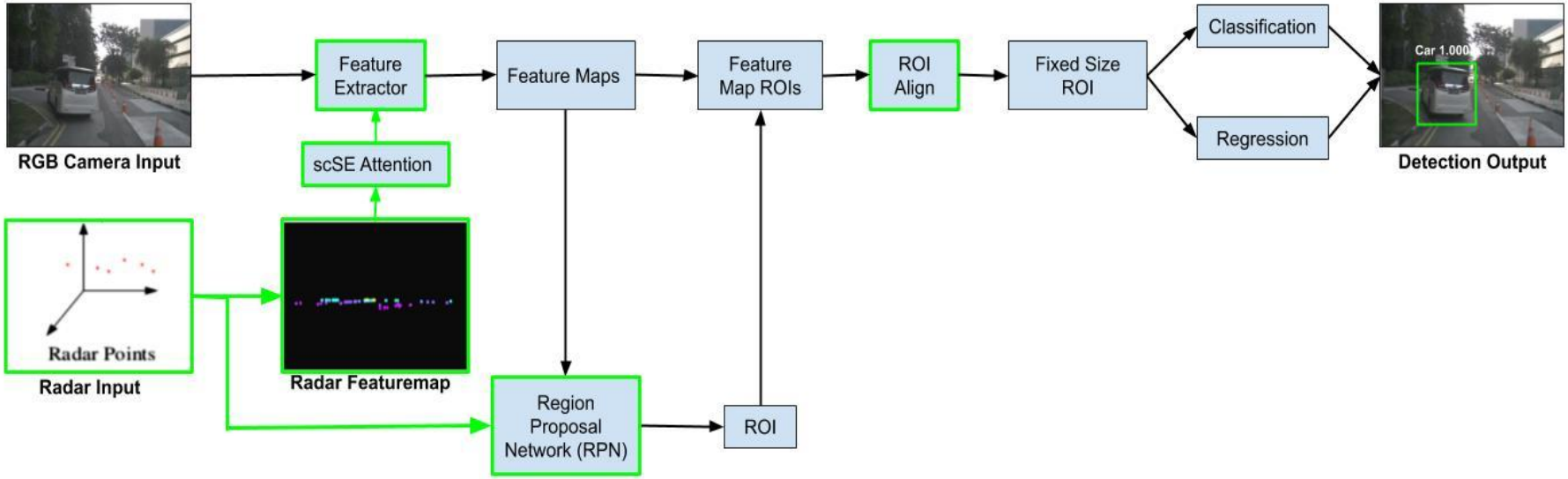
Radar Point Based Anchoring



Other Addition



Other Addition



Two Fusion Networks Proposed:

RANet(RAdar **Net**work**):**

- Radar point based anchors only

BIRANet(Best of RGB **I**mage and **RA**dar **Net**work**):**

- Vision(FFPN) + Radar point based anchors
- **Best of Two Method**:
  - Calculate IOU of both radar and vision based anchors with each ground truth bounding box
  - Select the anchor with highest IOU

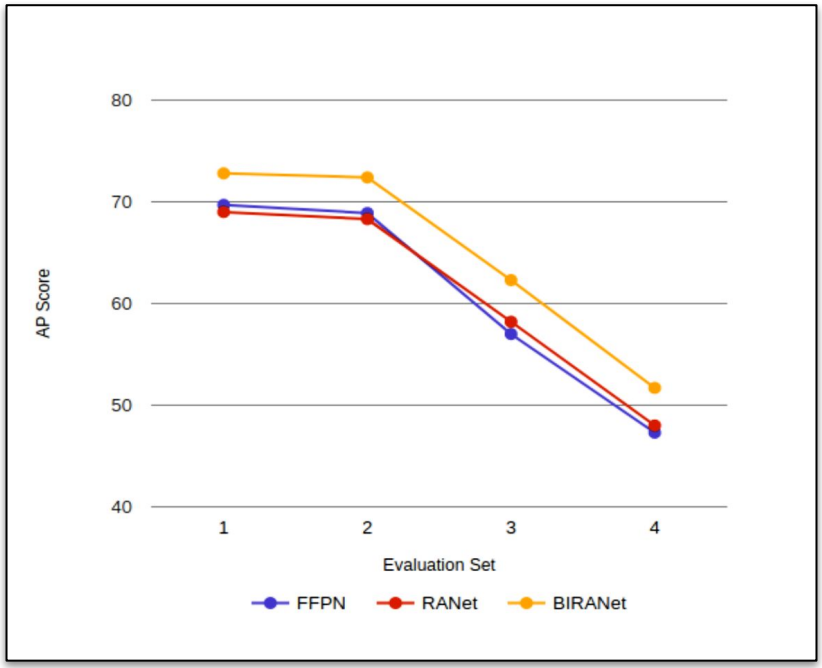
Dataset

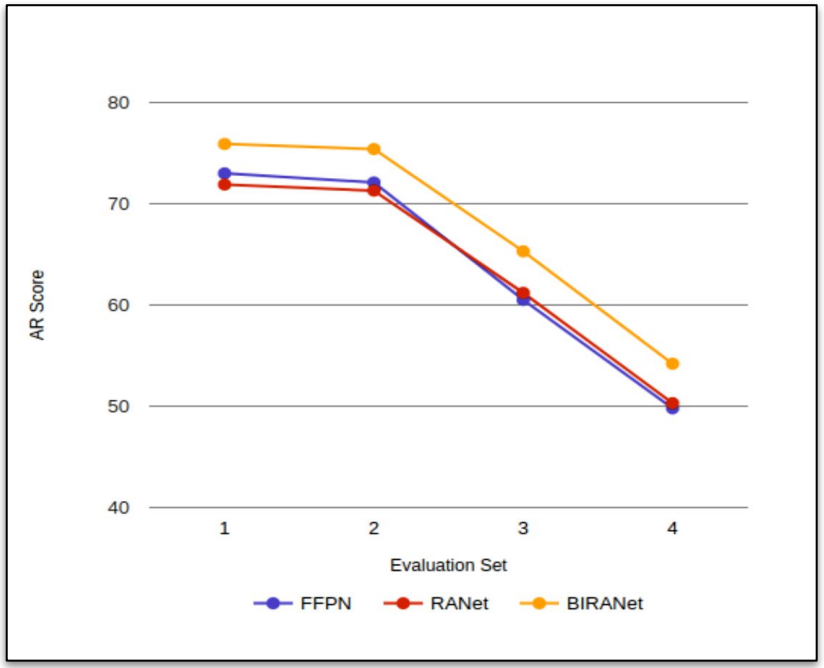
NuScenes(Training & Evaluation):

- 900 x 1600 image resolution
- 32254 training image and 5782 validation images

- **Evaluation Set:**
  - **Eval 1:** Default NuScenes evaluation set
  - **Eval 2, 3, 4:** NuScenes evaluation set with small, medium and large amount of augmented noise

Quantitative Comparison





Note: *Base Network : FFPN(Faster R-CNN with Feature Pyramid Network), Proposed Networks : RANet & BIRANet.*

Result comparison on Image Resolution Scale

Resolution Drop 1024x1024 → 512x512

BIRANet Evaluation on Image Resolution Scale

Resolution Drop 1024x1024 → 512x512

	Eval 1	Eval 4
AP & AR Drop	**~4%**	**~1%**
Inference Time Drop	**55%**	**55%**

If High Noise → Processing low-resolution images is a better option.

Qualitative Results: Detection Comparison on Eval 1.

Ground Truth

FFPN

RANet

BIRANet

Qualitative Results: Detection Comparison on Eval 4.

Ground Truth

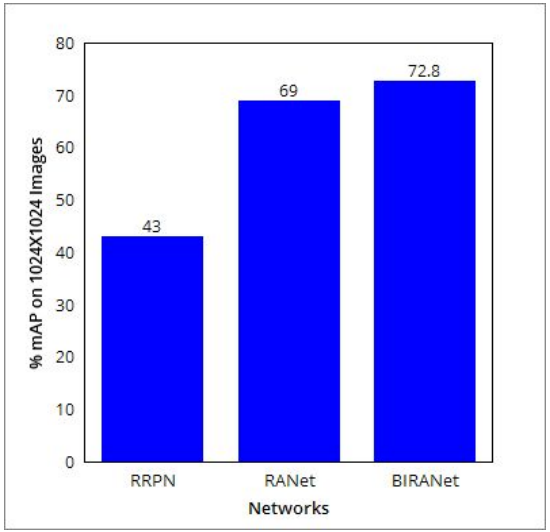
FFPN

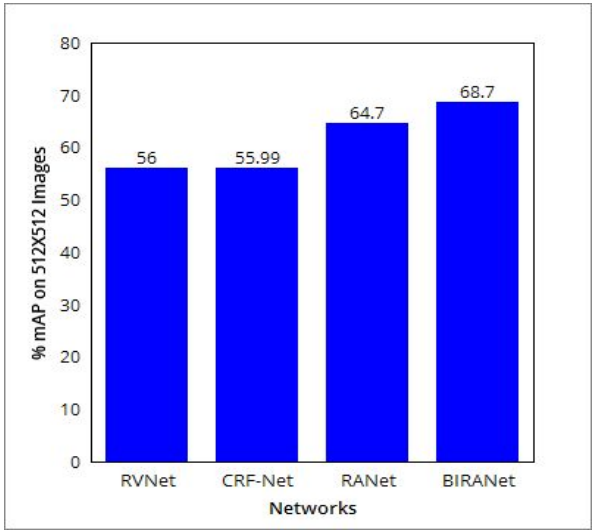
RANet

BIRANet

Comparison with Existing Methods

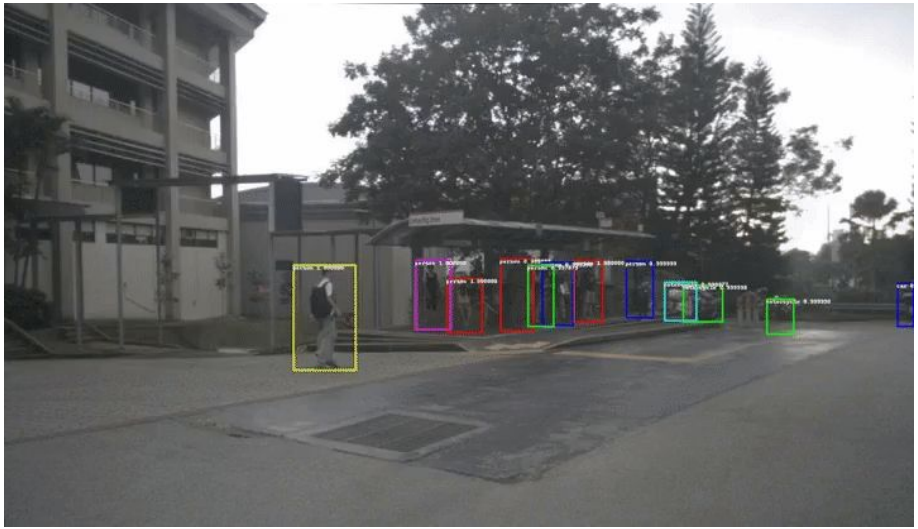
- **RRPN :** RANet is **26%** and BIRANet is **29.8%** better.
- **RVNet, CRF-Net :** RANet is **8.7 %** and BIRANet is **12.7%** better.





RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles, RVNet: Deep Sensor Fusion of Monocular Camera and Radar for Image-Based Obstacle Detection in Challenging Environments, CRF-Net: A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection.

Thank You.



Front - Camera Detection

Back - Camera Detection