Radar-Camera Sensor Fusion for Joint Object Detection and Distance Estimation in Autonomous Vehicles

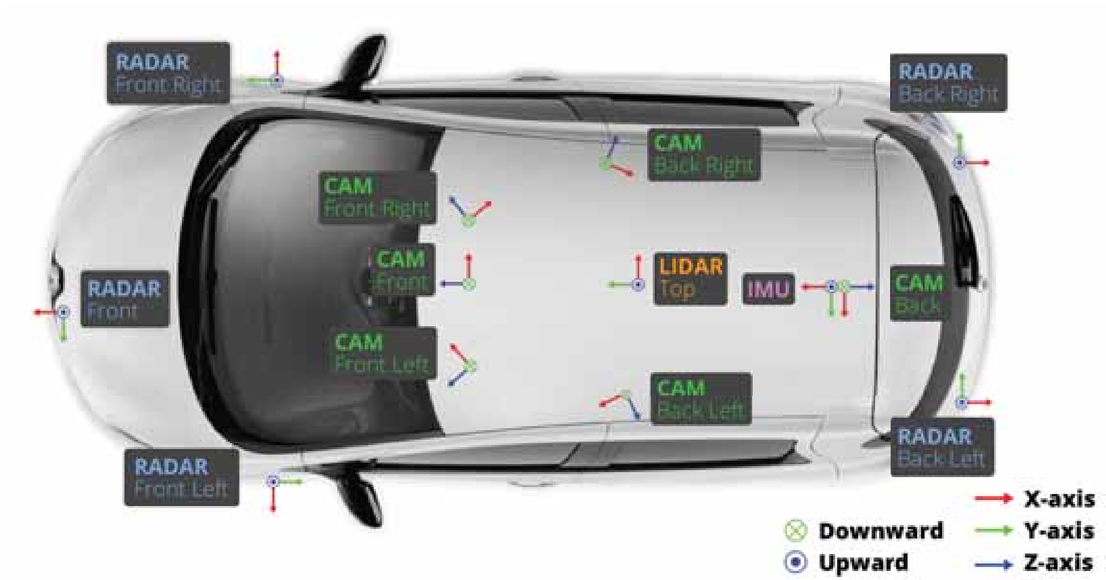
Ramin Nabati   Hairong Qi

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

THE UNIVERSITY OF TENNESSEE KNOXVILLE

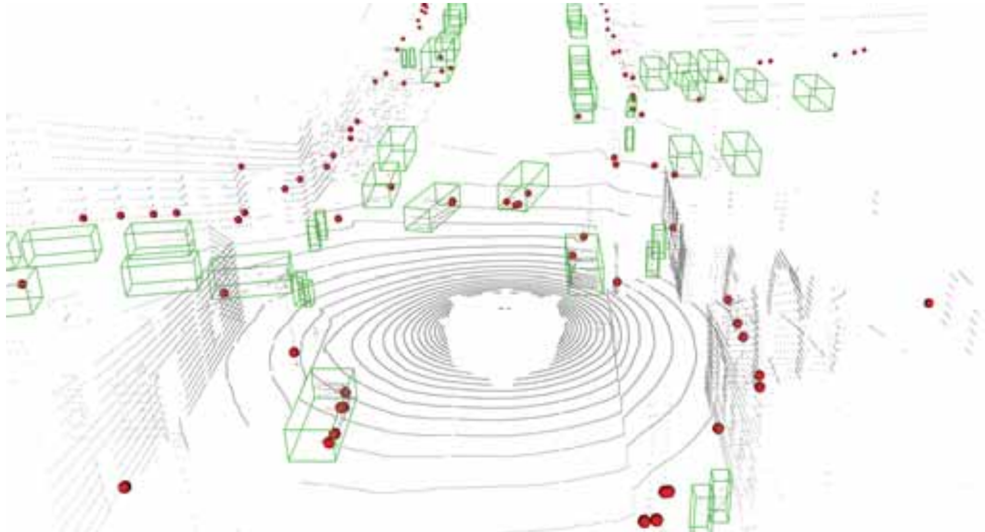
Object Detection for Autonomous Vehicles

- Modern autonomous vehicles are equipped with multiple sensors



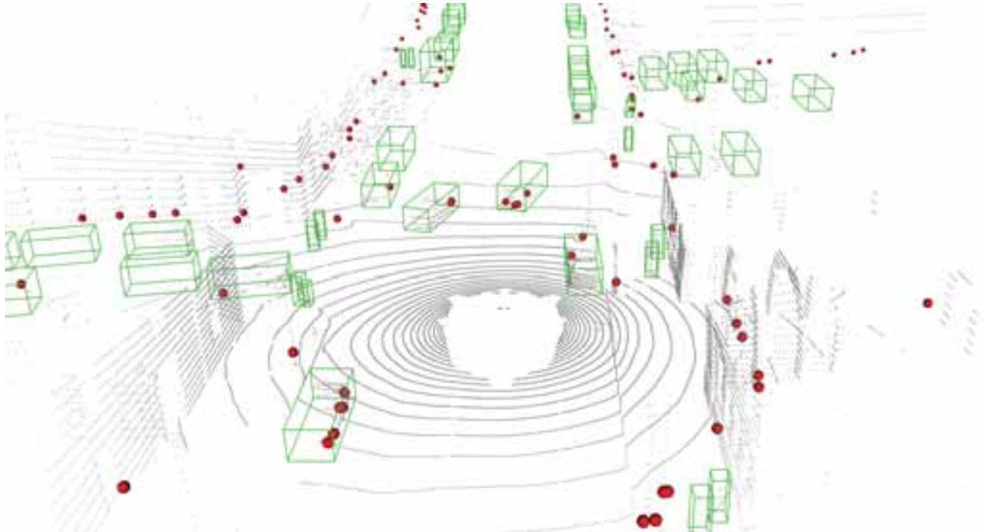
Object Detection for Autonomous Vehicles

- Modern autonomous vehicles are equipped with multiple sensors
- Sensor fusion is a crucial part of the perception system



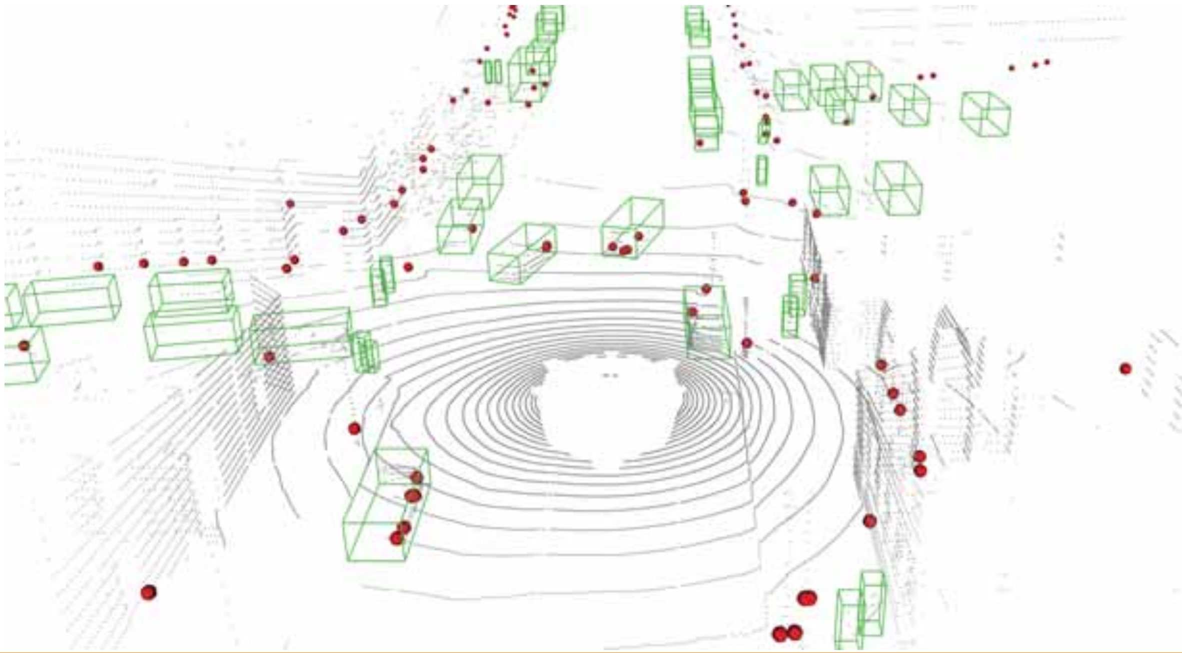
Object Detection for Autonomous Vehicles

- Modern autonomous vehicles are equipped with multiple sensors
- Sensor fusion is a crucial part of the perception system
- The focus of this work is on radar-camera sensor fusion



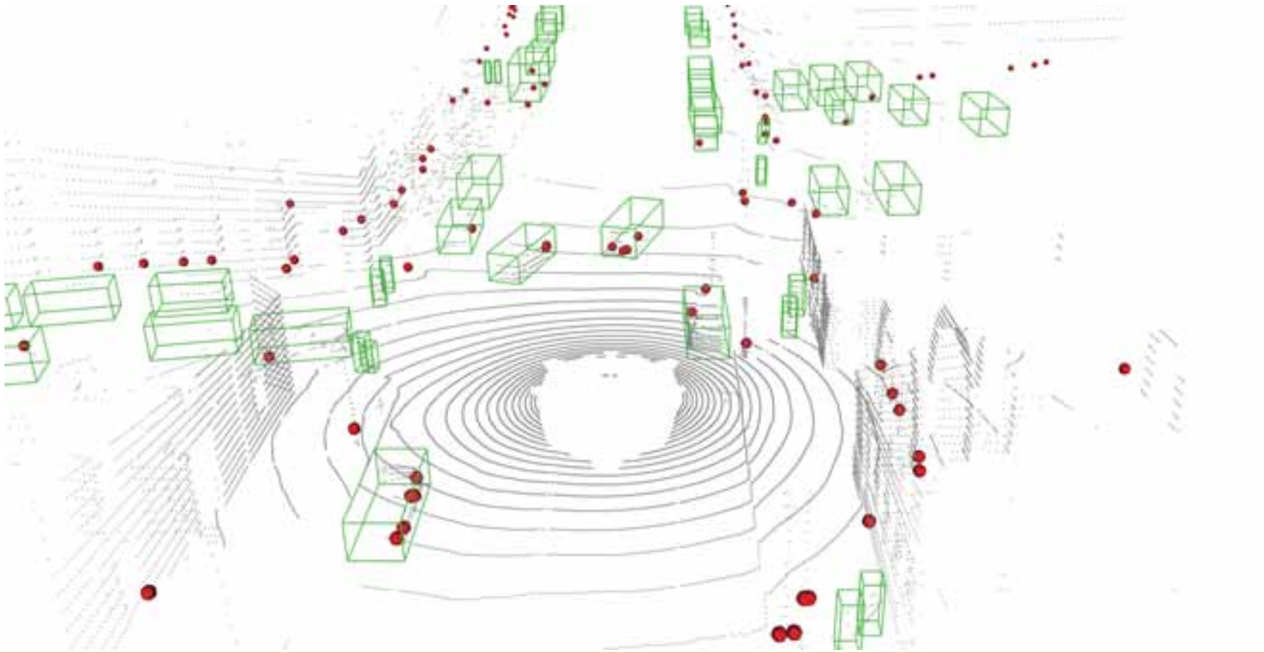
Challenges

- Radar point cloud is extremely sparse compared to LIDARs



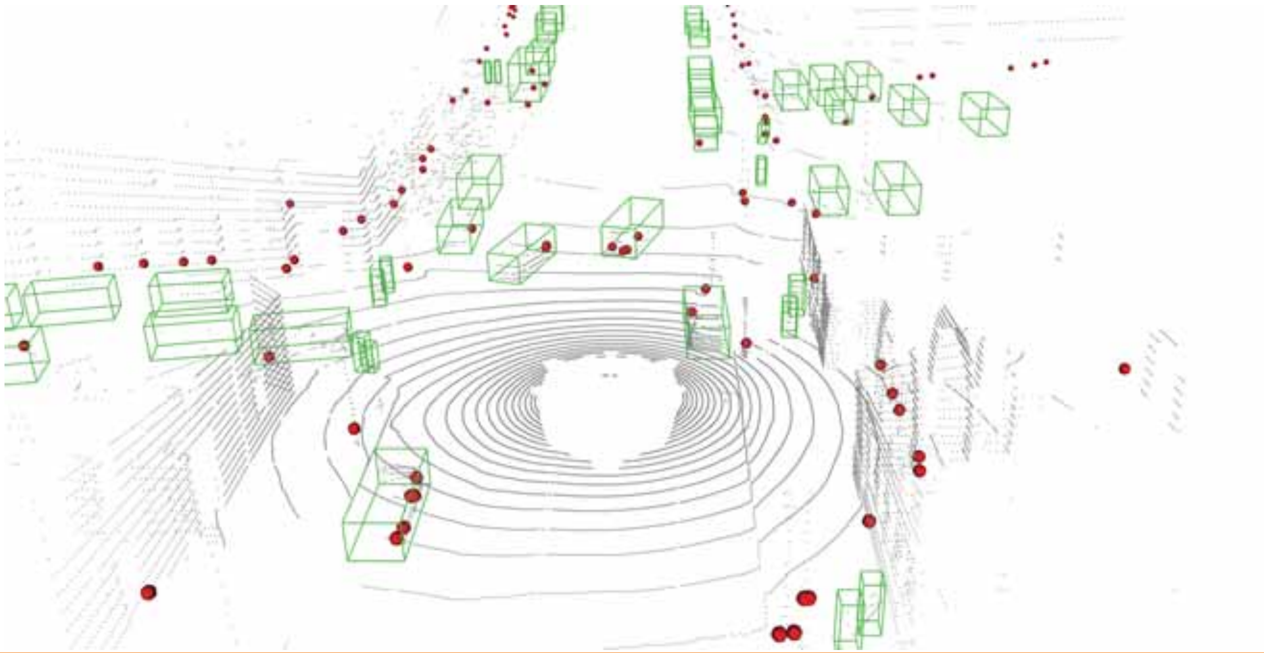
Challenges

- Radar point cloud is extremely sparse compared to LIDARs
- Radars are not good at classifying objects



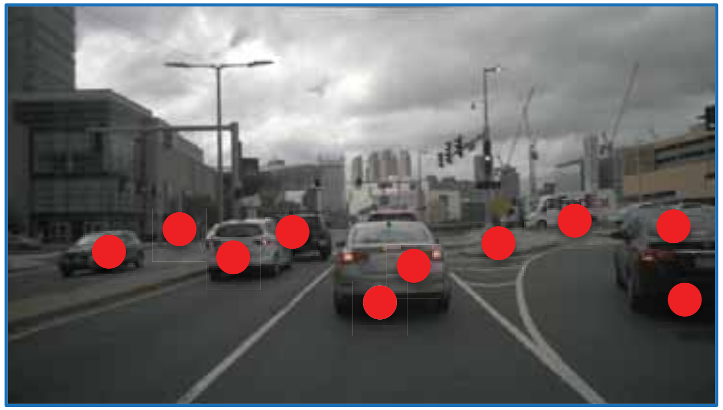
Challenges

- Radar point cloud is extremely sparse compared to LIDARs
- Radars are not good at classifying objects
- Most automotive radars do not provide height information for detections



Prior Work

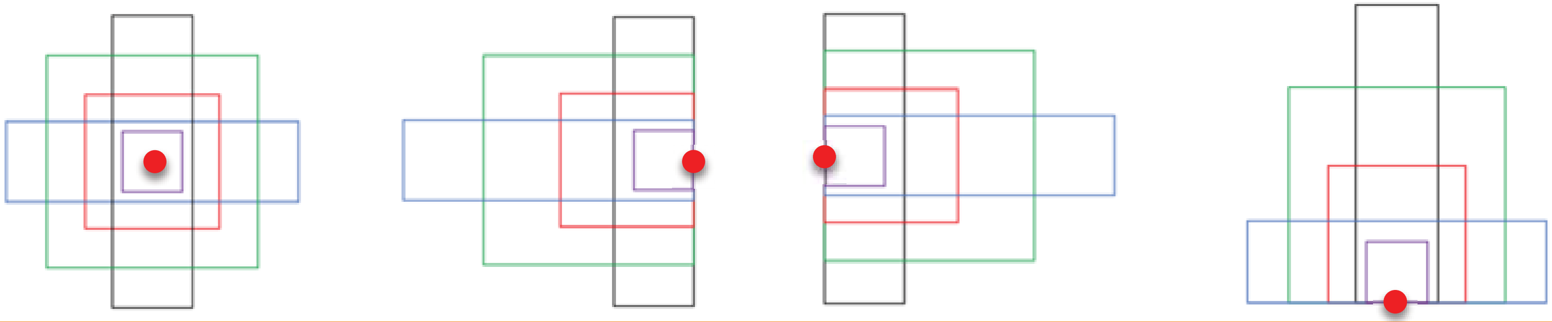
- RRPN: Radar Region Proposal Network



R. Nabati and H. Qi, "RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3093-3097, doi: 10.1109/ICIP.2019.8803392.

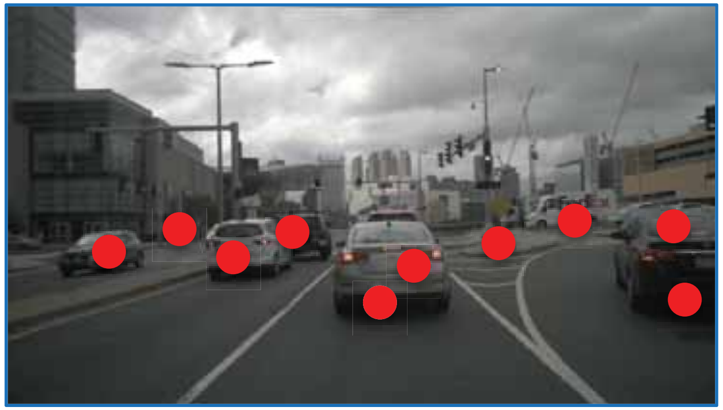
Prior Work

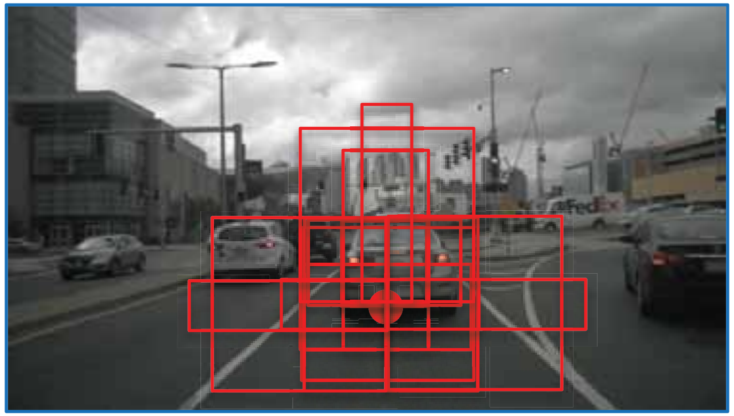
- RRPN: Radar Region Proposal Network

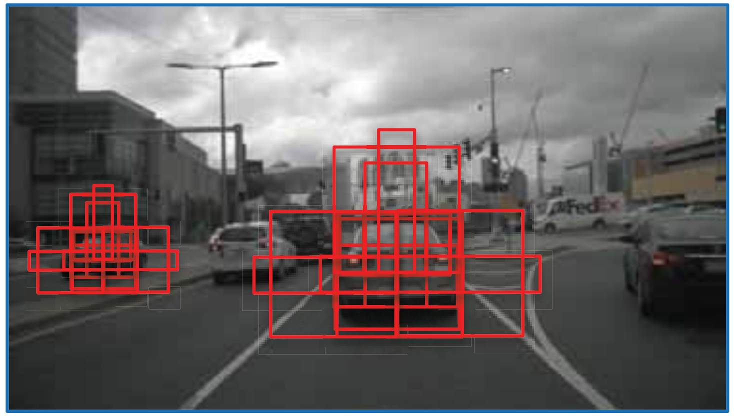


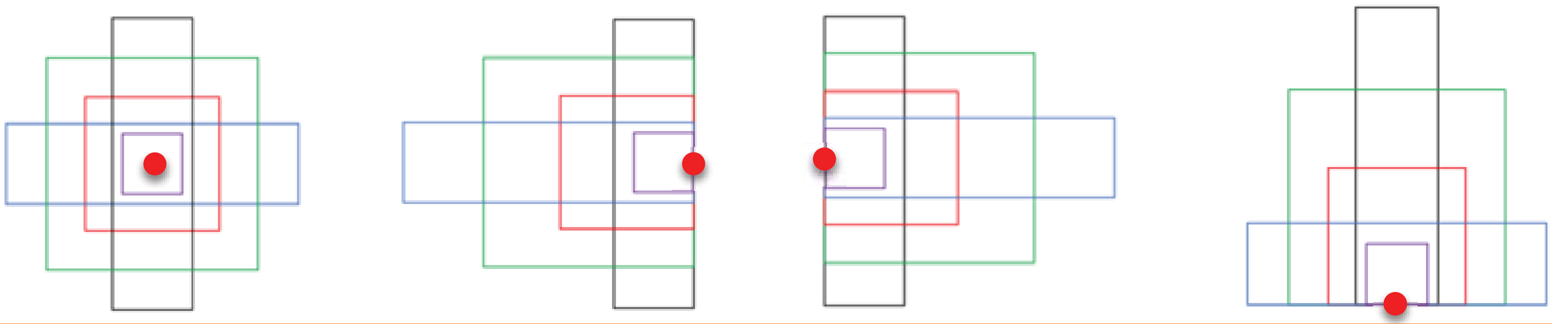
Prior Work

- RRPN: Radar Region Proposal Network





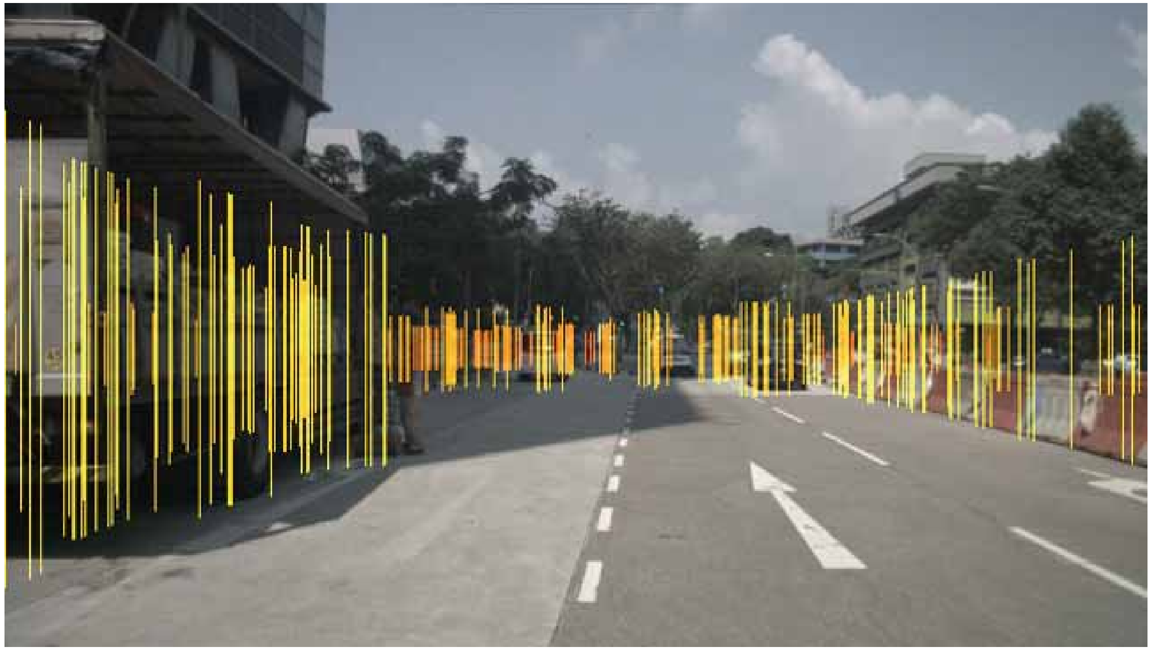




R. Nabati and H. Qi, "RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3093-3097, doi: 10.1109/ICIP.2019.8803392.

Prior Work

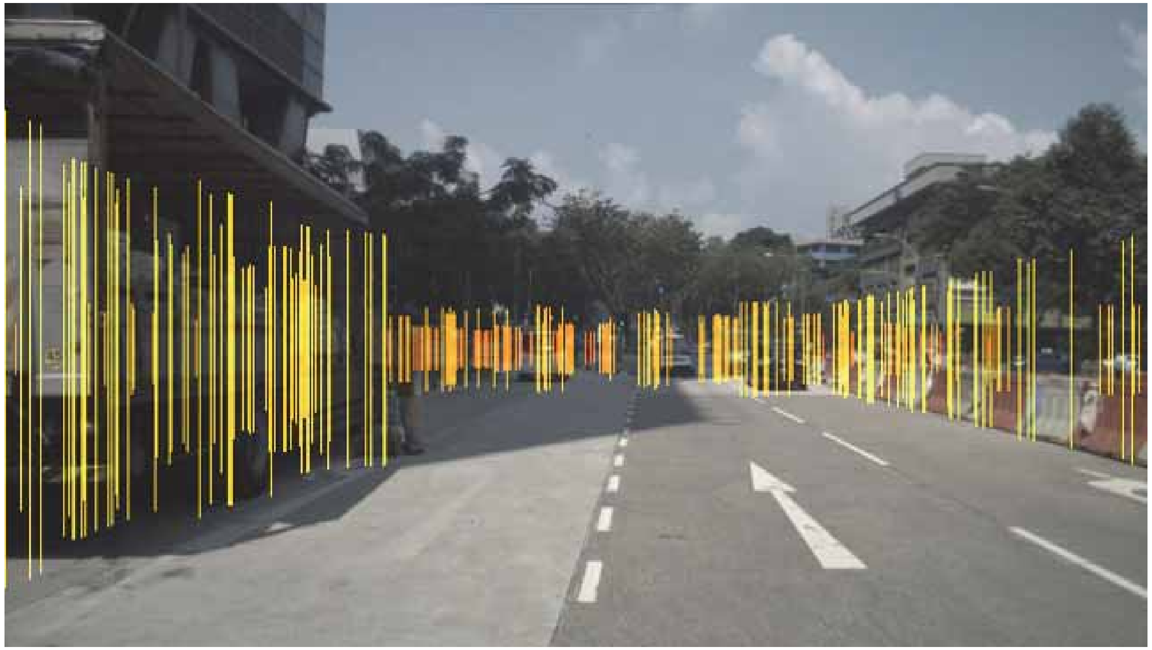
- CRF-Net

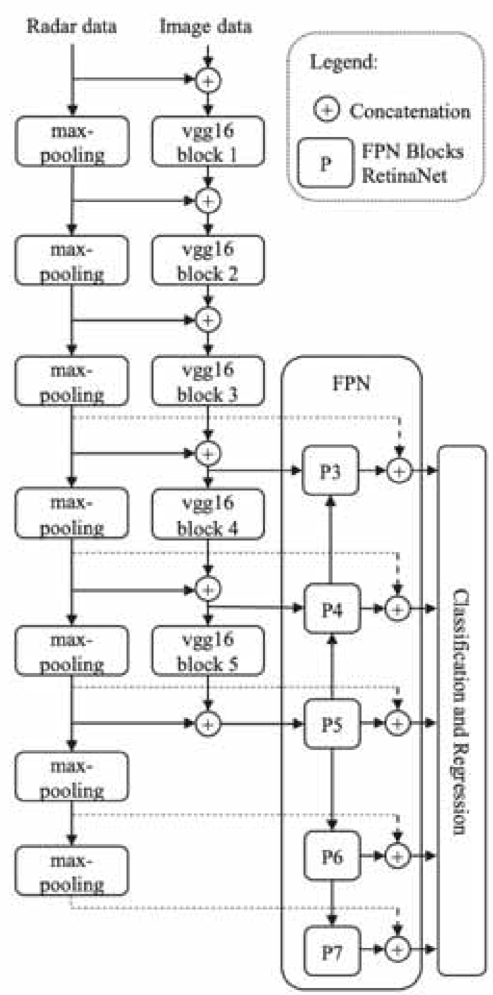


Nobis, Felix, et al. "A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection." *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2019.

Prior Work

- CRF-Net

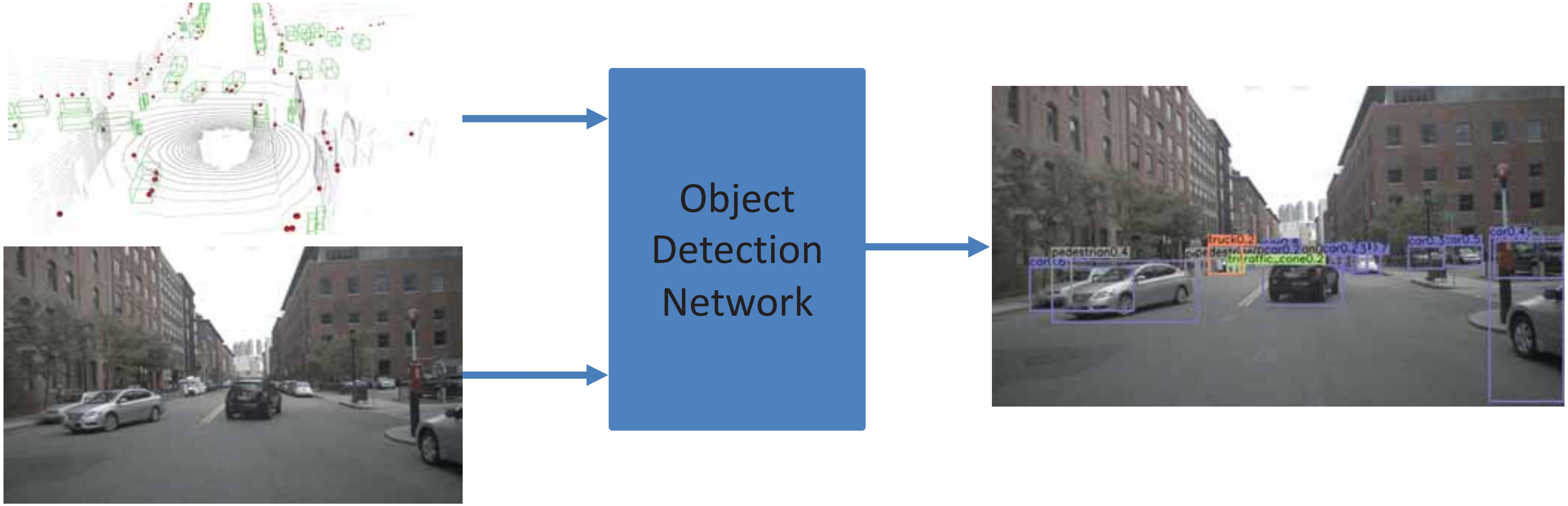




Nobis, Felix, et al. "A Deep Learning-based Radar and Camera Sensor Fusion Architecture for Object Detection." *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*. IEEE, 2019.

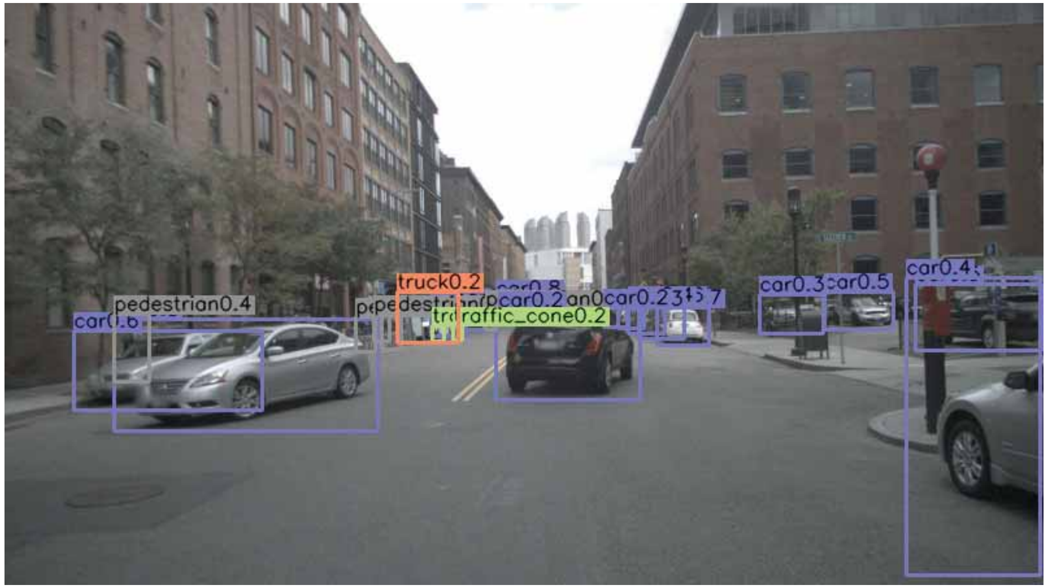
Motivation

- Depth information lost in the network



Motivation

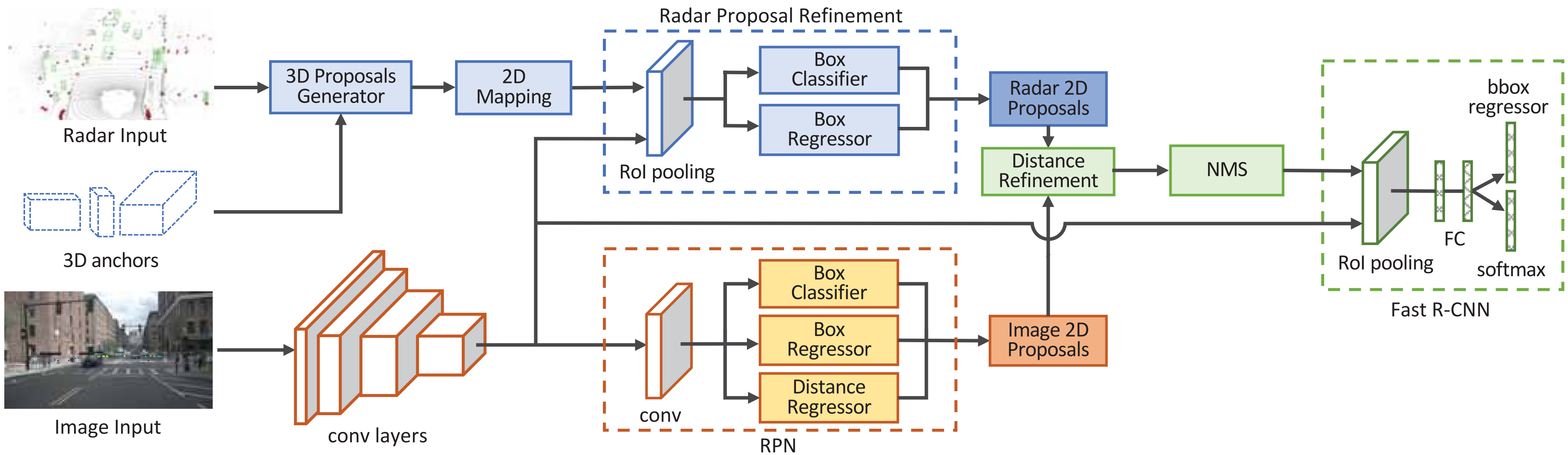
- Depth information lost in the network
- 3D object detection is costly

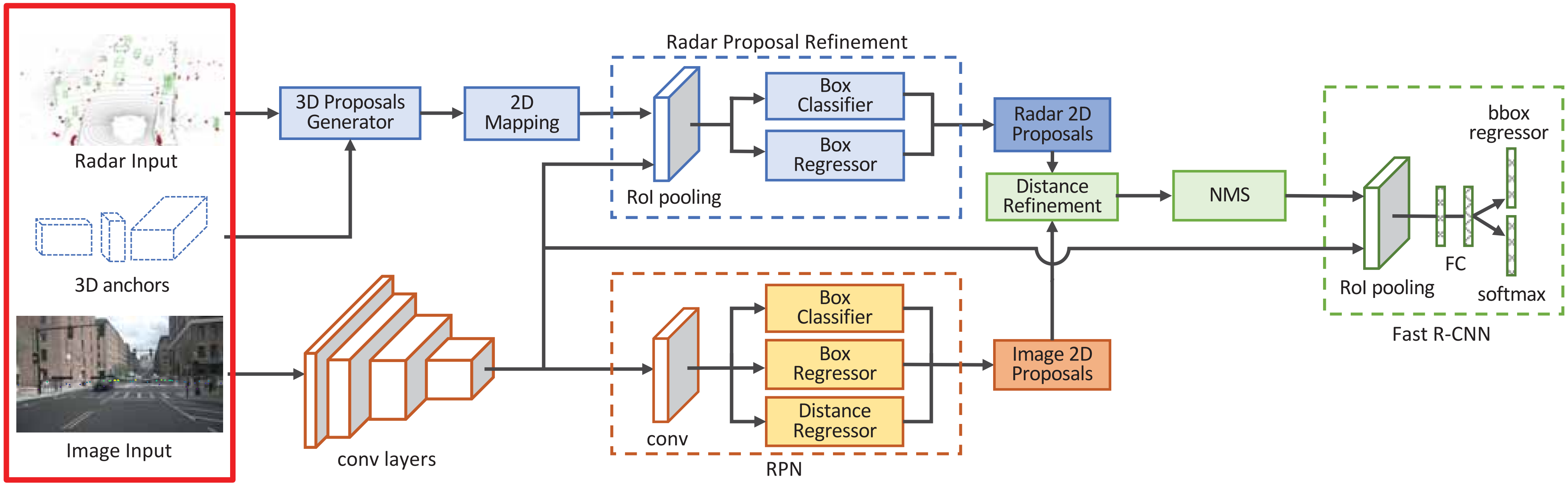


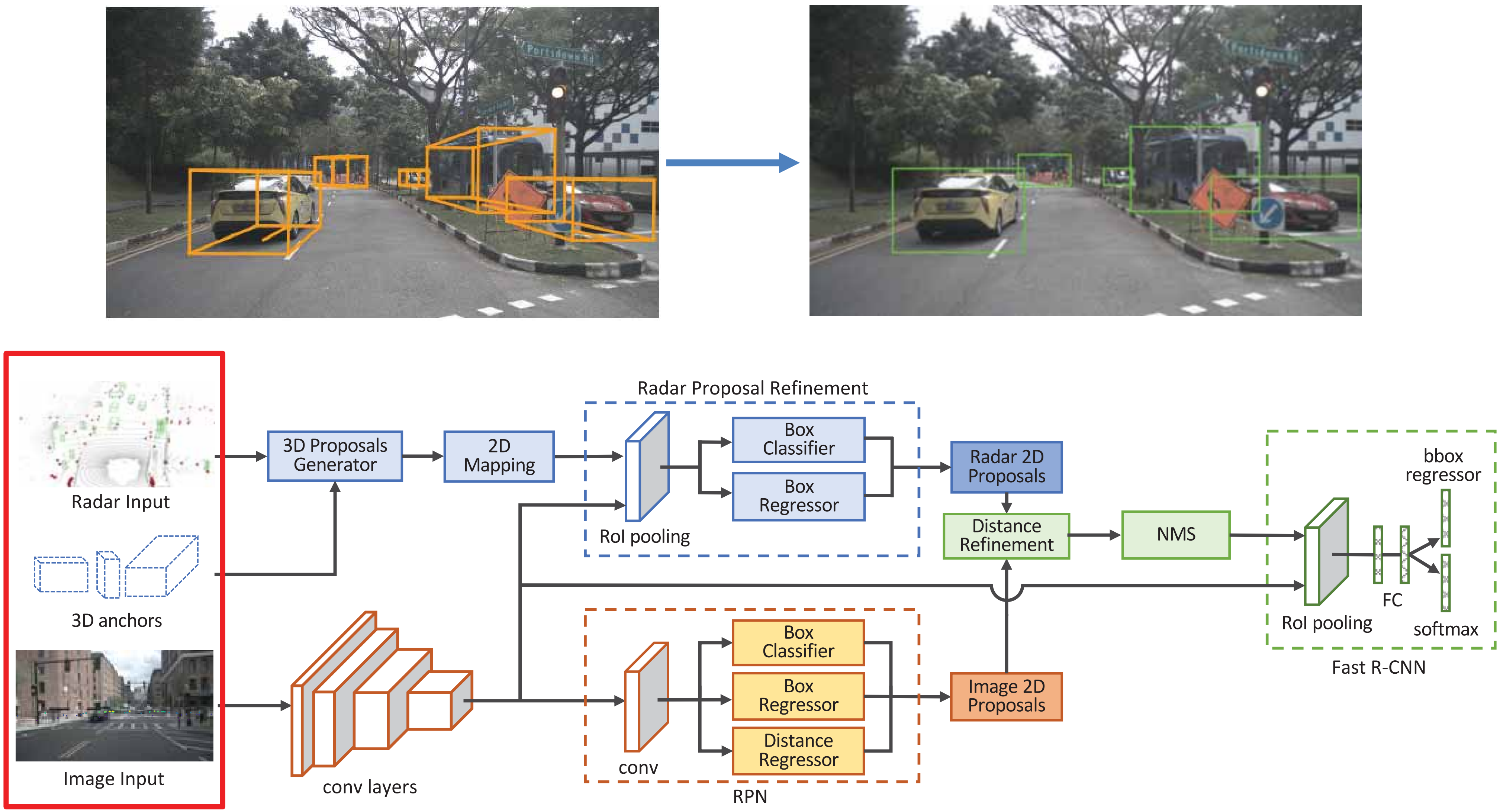
Motivation

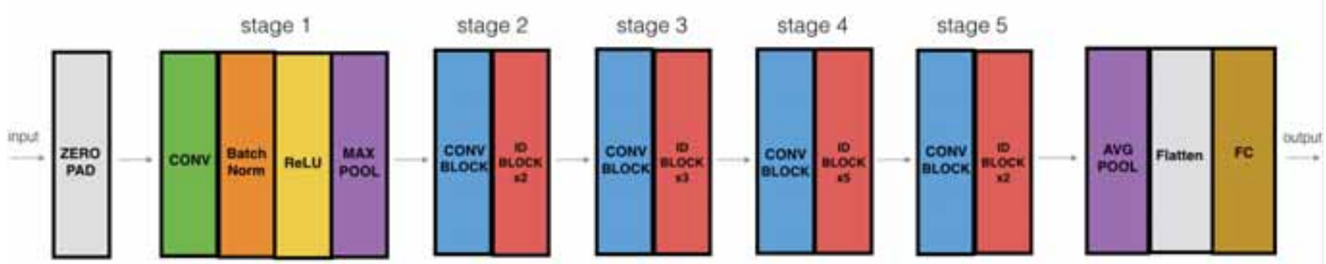
- Depth information lost in the network
- 3D object detection is costly

Network Architecture

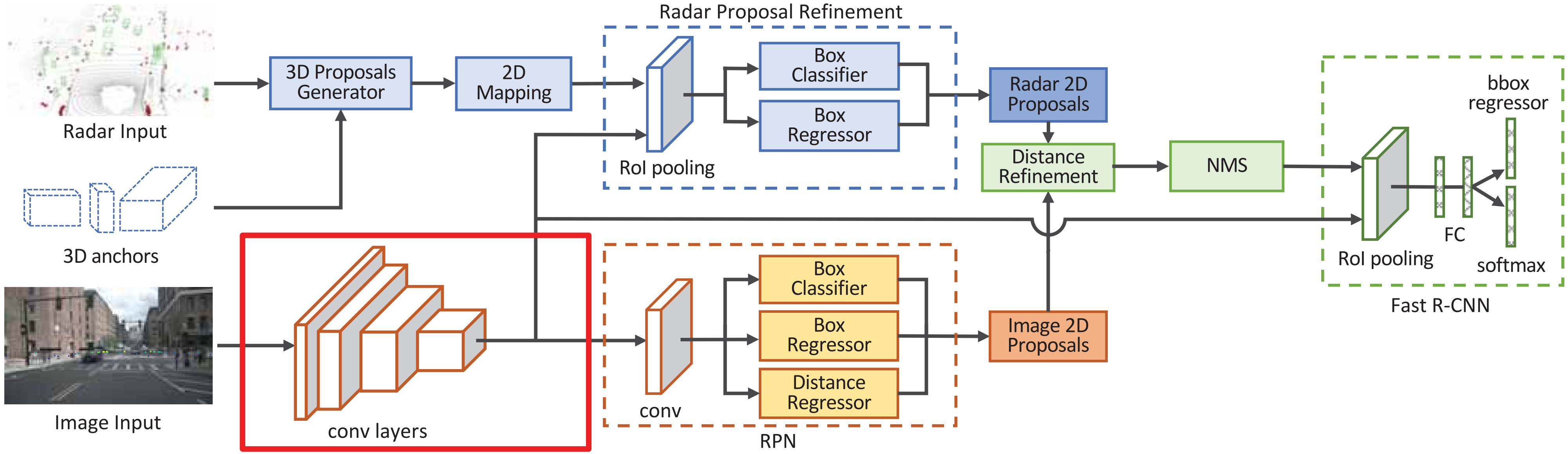


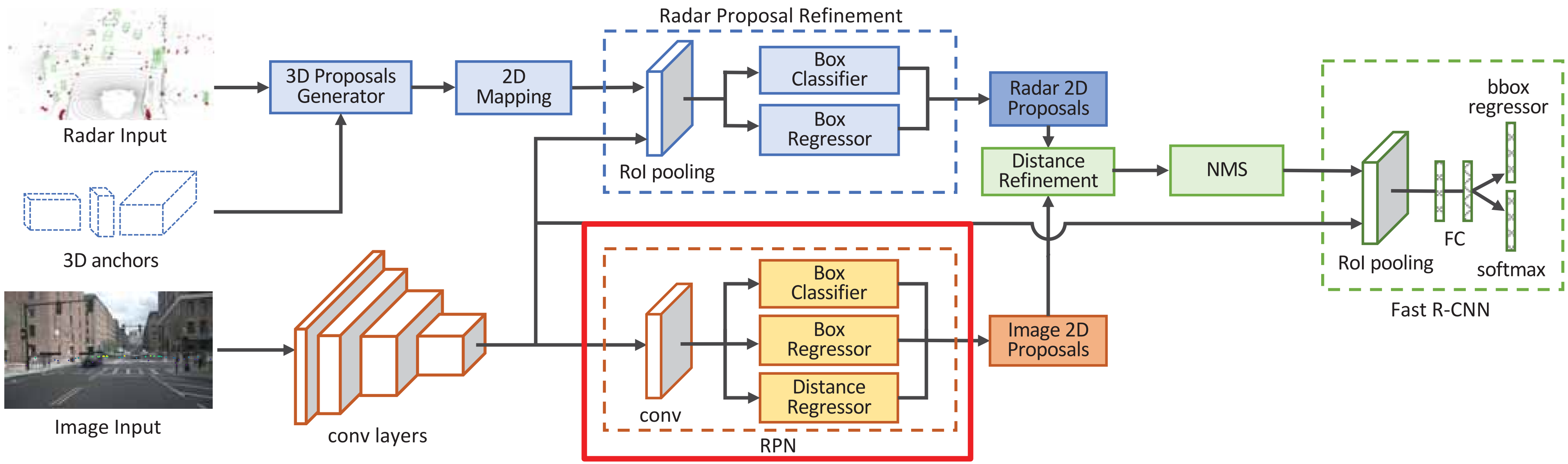

Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals
bbox regressor
RoI pooling
FC
softmax
Fast R-CNN


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals
RoI pooling
FC
bbox regressor
softmax
Fast R-CNN

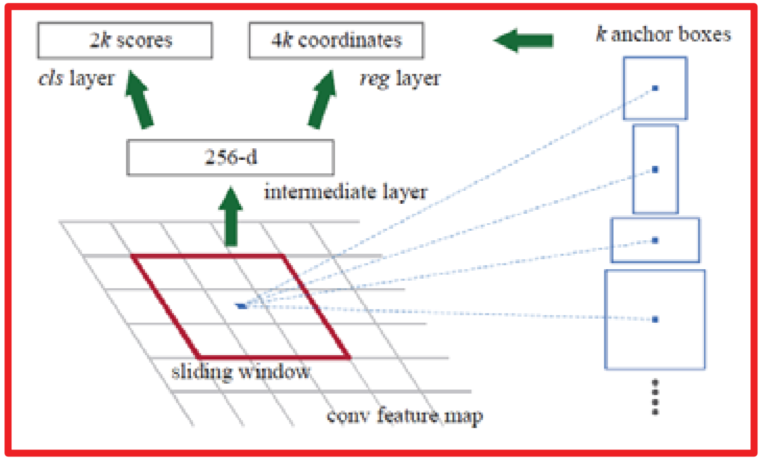


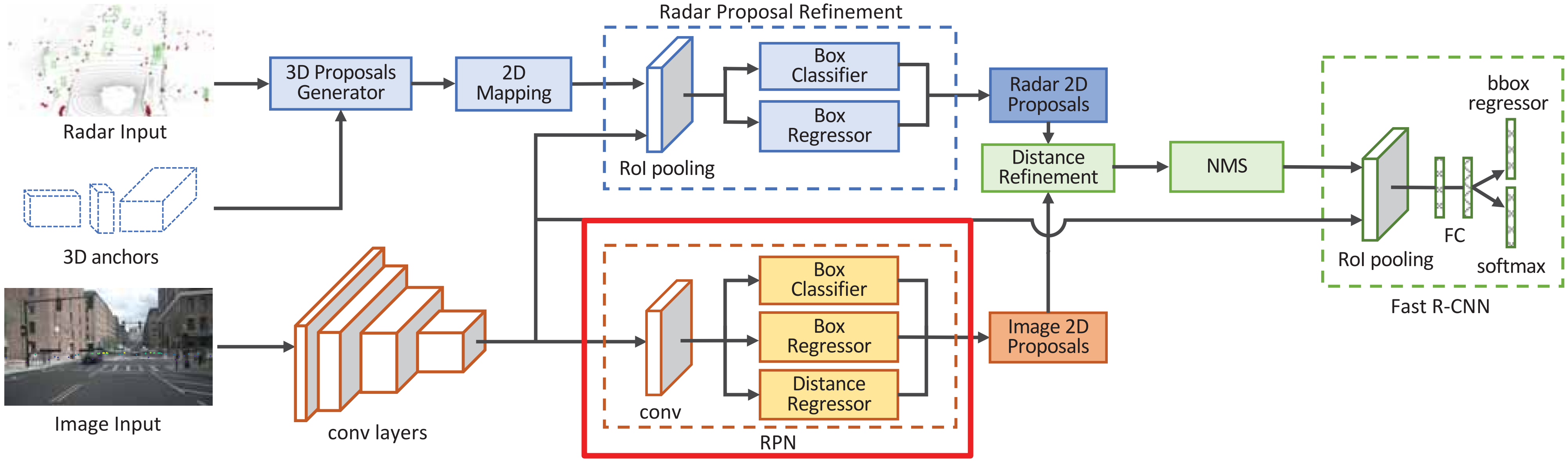
ResNet-50 Model*



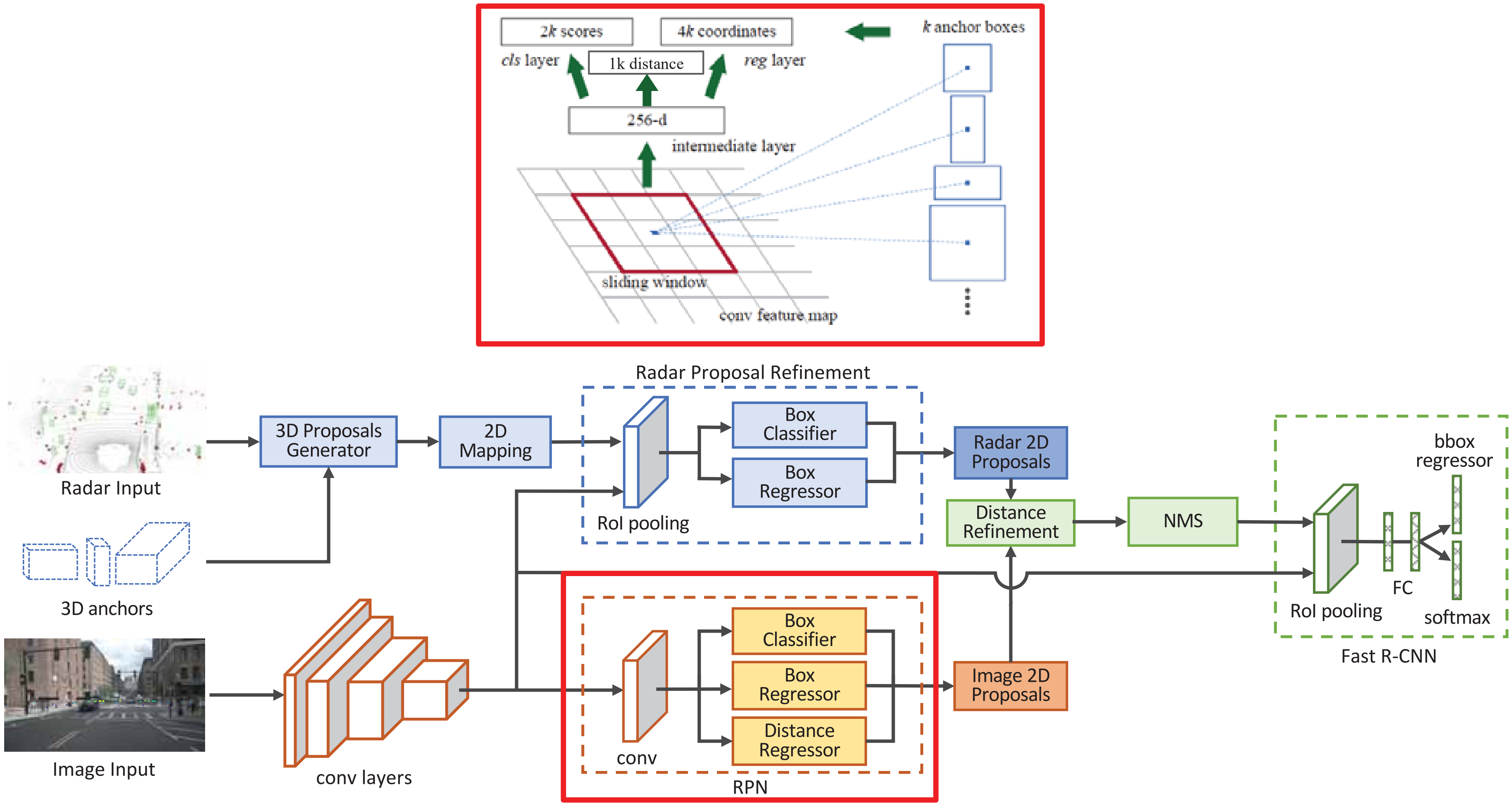


Radar Proposal Refinement
Radar Input
3D Proposals Generator
2D Mapping
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
3D anchors
Distance Refinement
NMS
bbox regressor
RoI pooling
FC
softmax
Fast R-CNN
Image Input
conv layers
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals

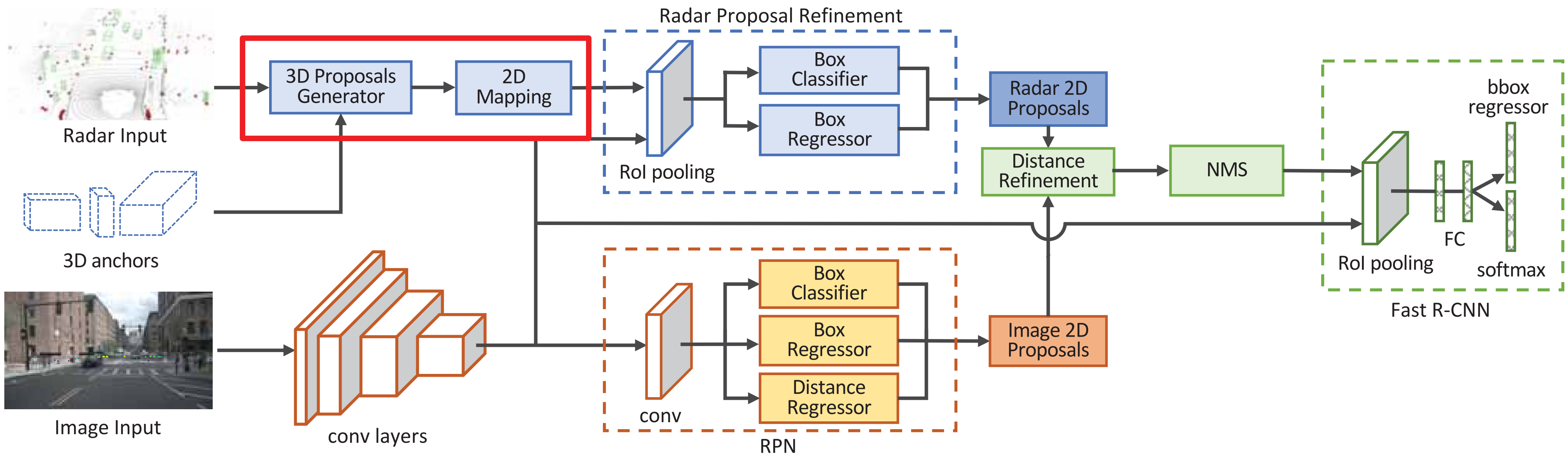


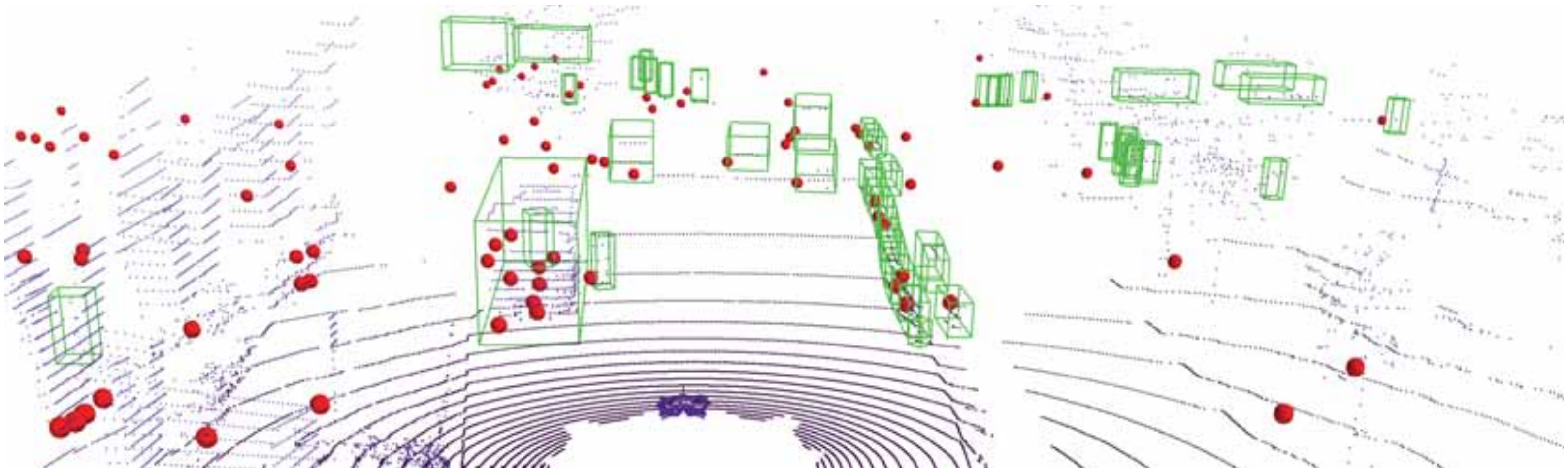


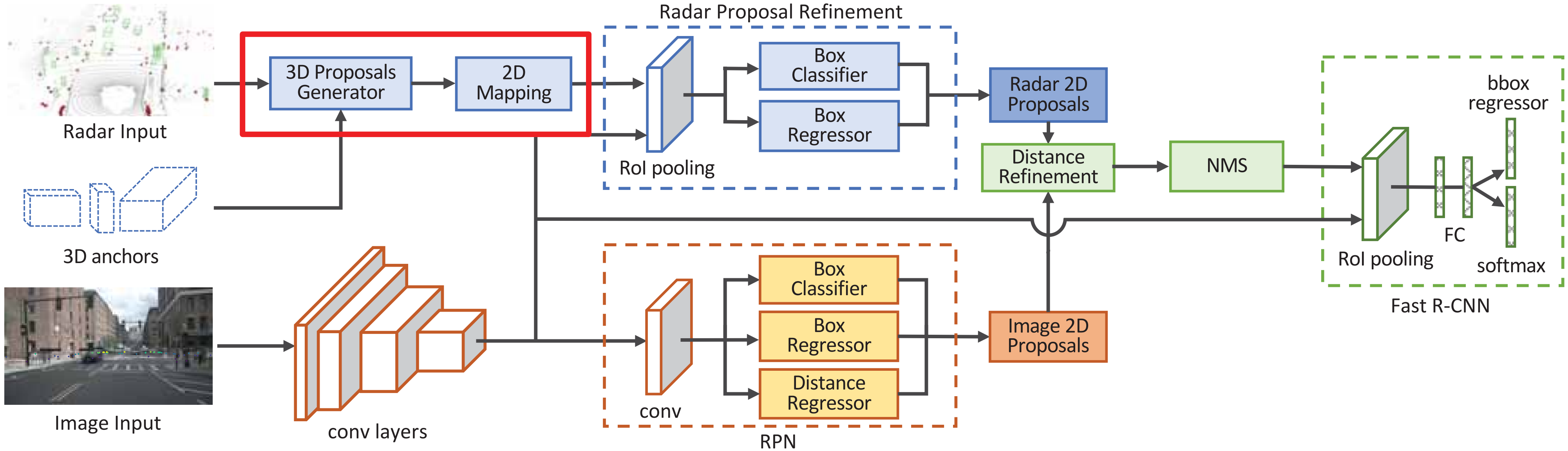
Source: https://arxiv.org/abs/1311.2524

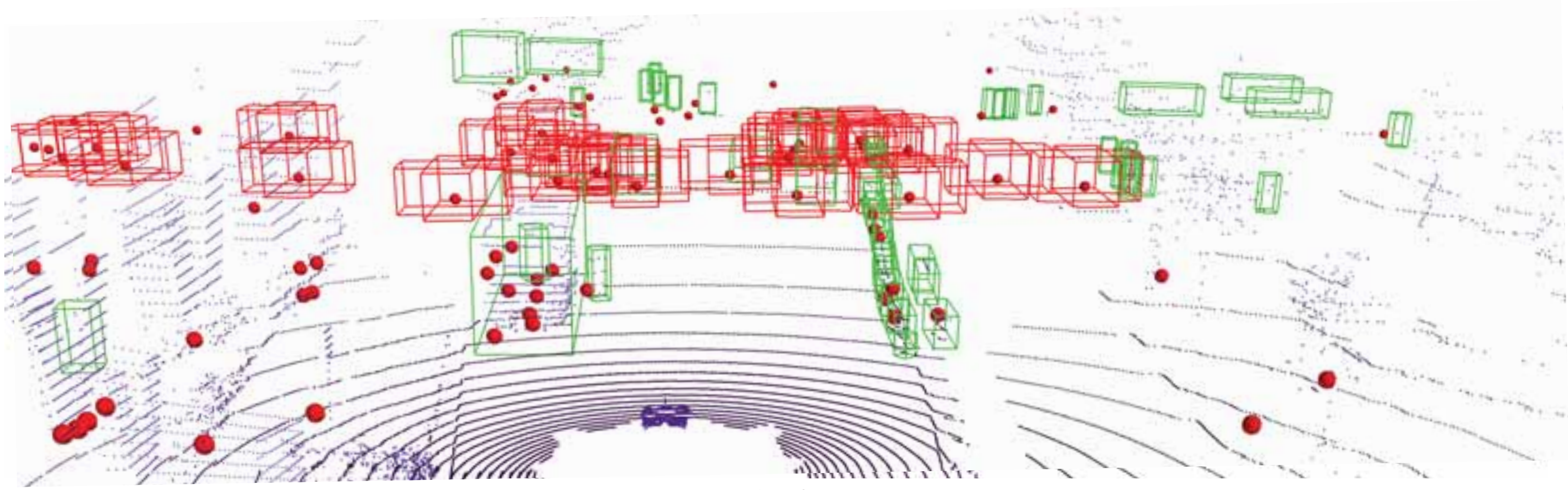


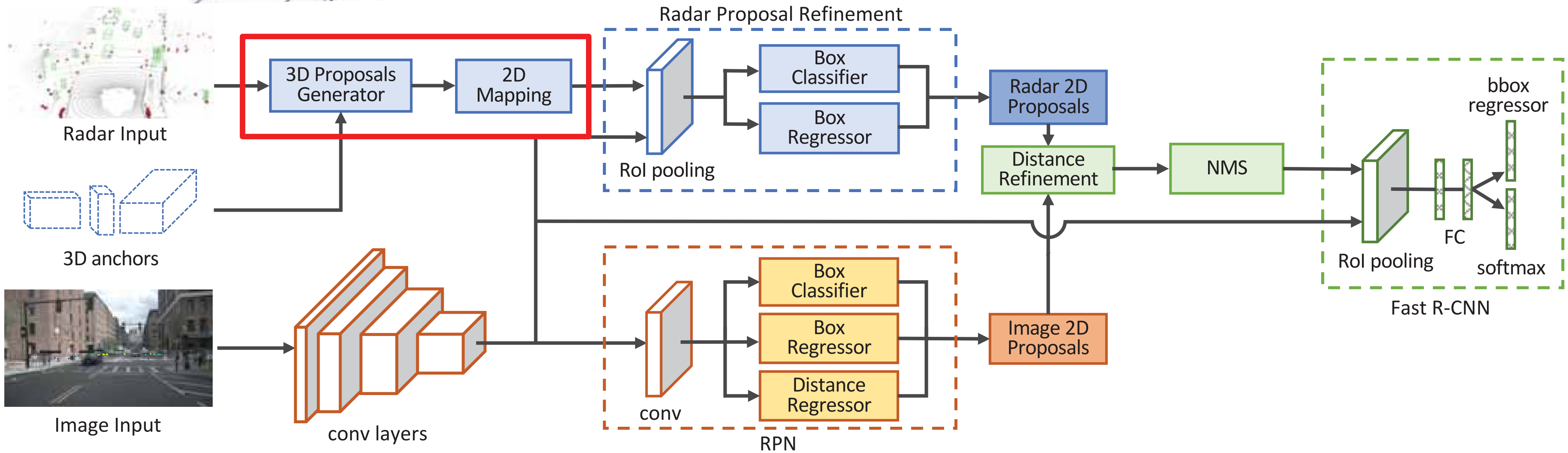
2k scores
4k coordinates
k anchor boxes
cls layer
1k distance
reg layer
256-d
intermediate layer
sliding window
conv feature map
Radar Proposal Refinement
Radar Input
3D Proposals Generator
2D Mapping
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
bbox regressor
FC
softmax
RoI pooling
Fast R-CNN
3D anchors
Image Input
conv layers
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
Image 2D Proposals
NMS
RoI pooling
FC
bbox regressor
softmax
Fast R-CNN



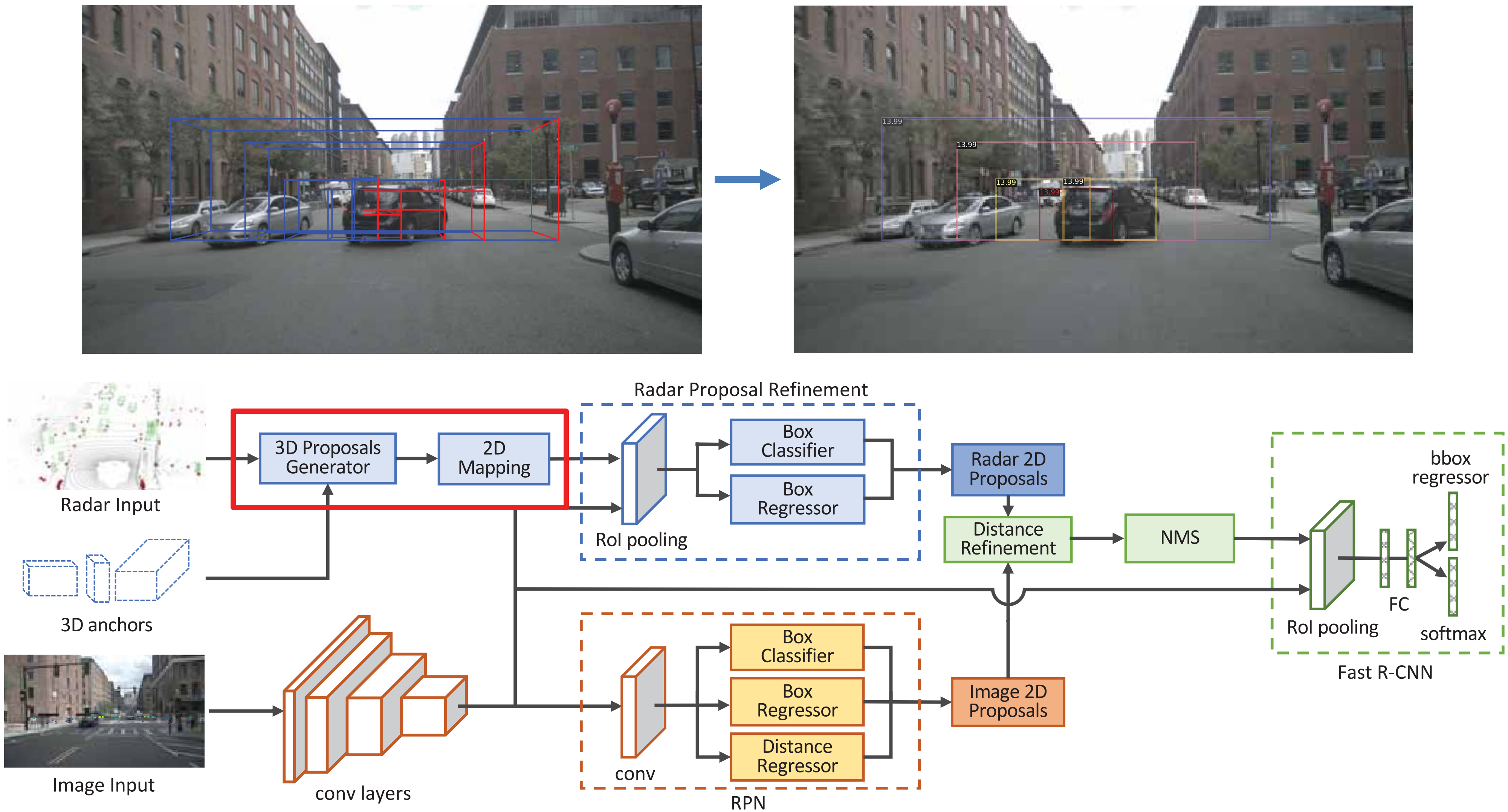

Radar Proposal Refinement
Radar Input
3D Proposals Generator
2D Mapping
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
bbox regressor
FC
softmax
RoI pooling
Fast R-CNN
3D anchors
Image Input
conv layers
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals

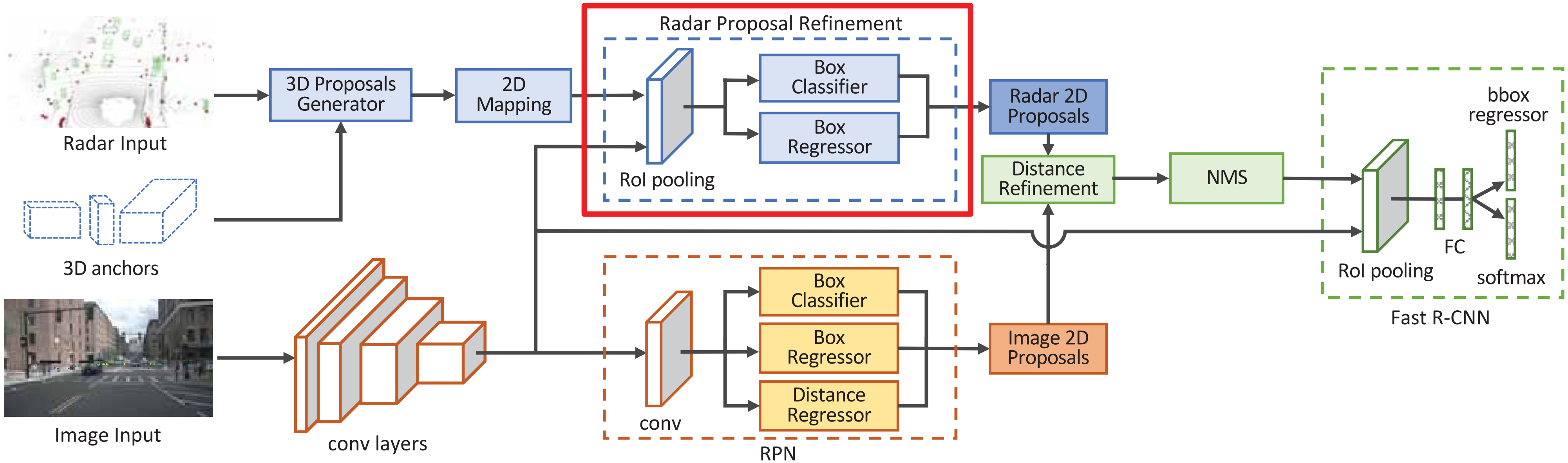


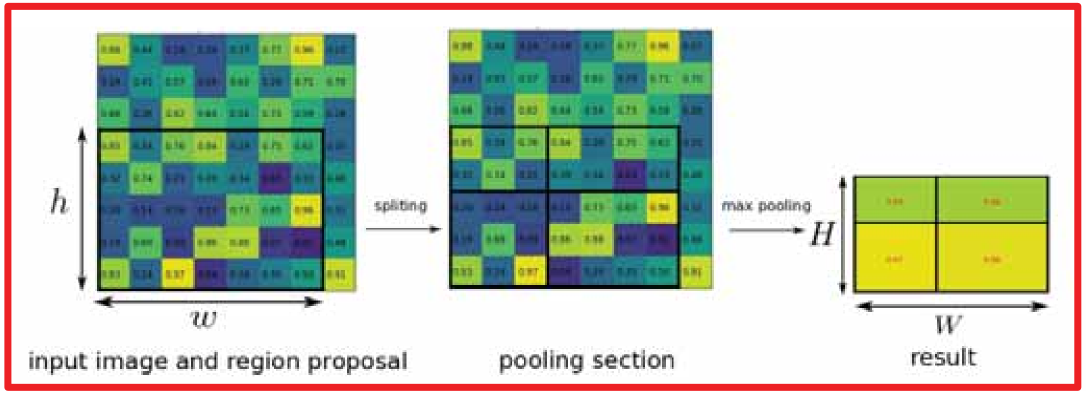


PPNIV'20 iROS

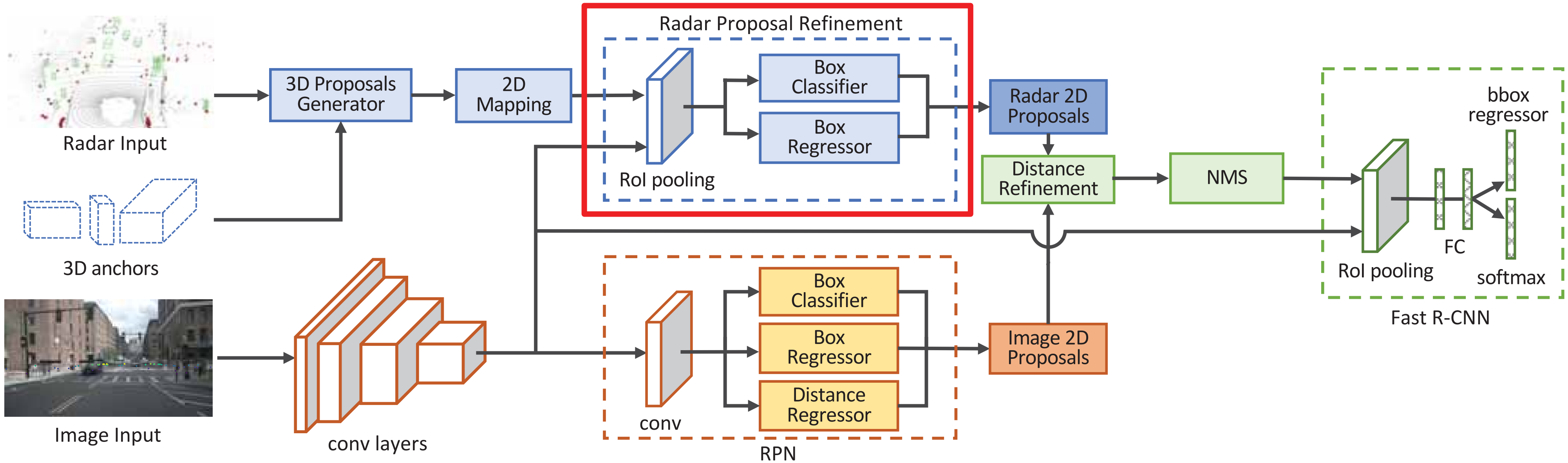
THE UNIVERSITY OF TENNESSEE KNOXVILLE


13.99
Radar Proposal Refinement
Radar Input
3D Proposals Generator
2D Mapping
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
bbox regressor
FC
softmax
RoI pooling
Fast R-CNN
3D anchors
Image Input
conv layers
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals

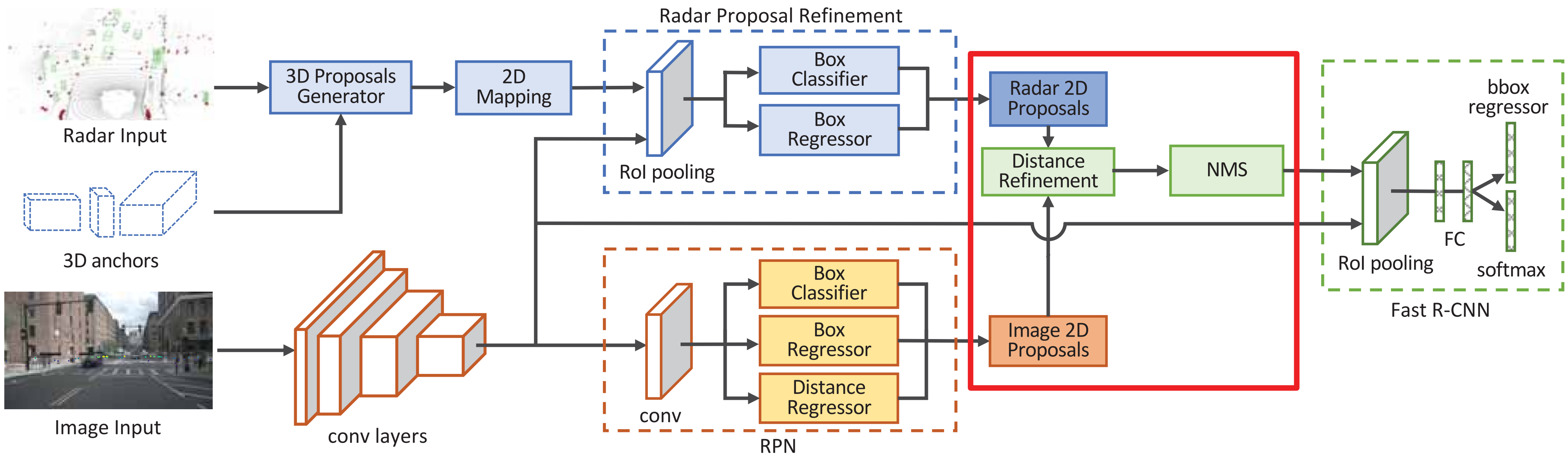

Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
Image 2D Proposals
NMS
RoI pooling
FC
bbox regressor
softmax
Fast R-CNN

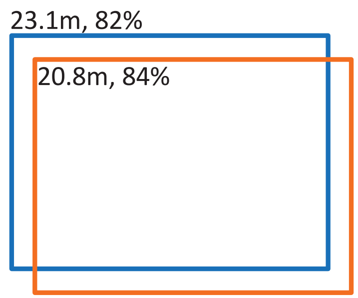


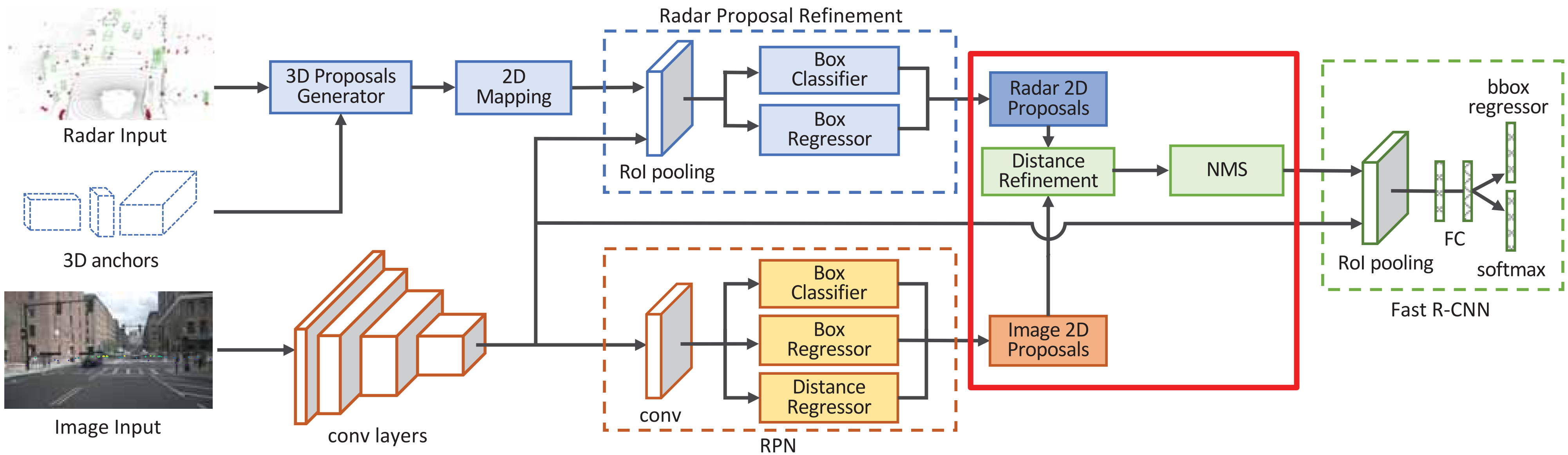
h
w
spliting
max pooling
H
W
input image and region proposal
pooling section
result

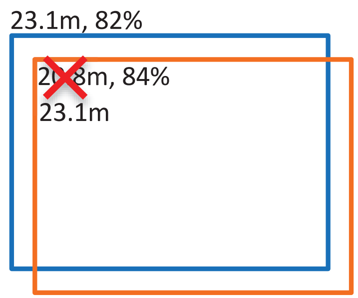


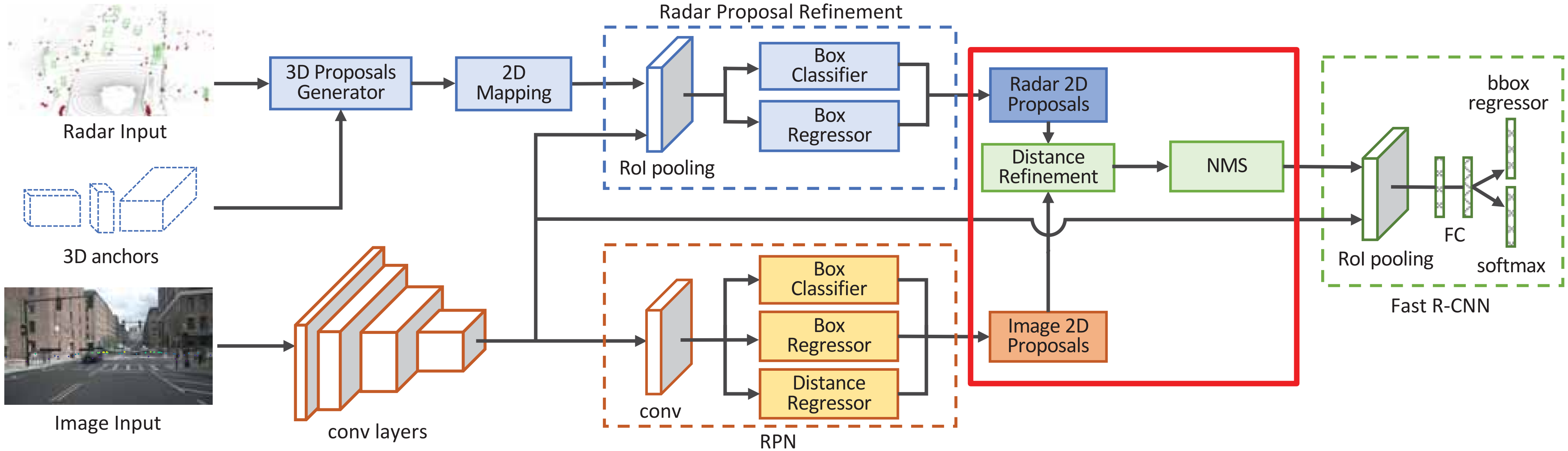
Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
Image 2D Proposals
NMS
bbox regressor
FC
softmax
Fast R-CNN


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
NMS
Image 2D Proposals
RoI pooling
FC
bbox regressor
softmax
Fast R-CNN

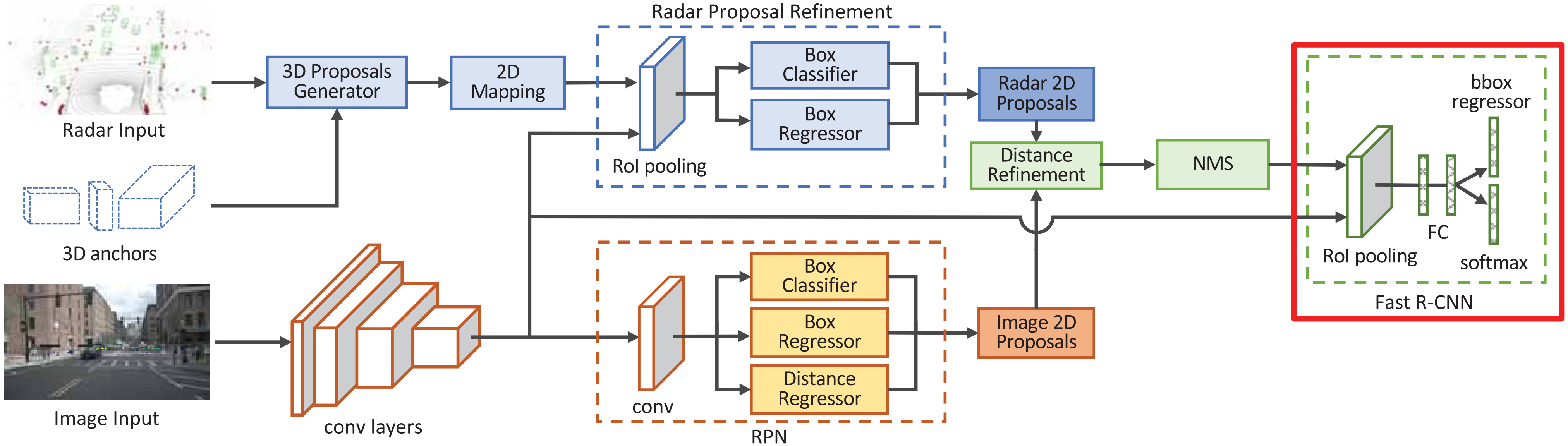

23.1m, 82%
20.8m, 84%


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
NMS
Image 2D Proposals
RoI pooling
FC
bbox regressor
softmax
Fast R-CNN


23.1m, 82%
84%
23.1m


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Radar 2D Proposals
Distance Refinement
NMS
Image 2D Proposals
bbox regressor
FC
softmax
RoI pooling
Fast R-CNN


pedestrian 75% - 19.5m
car 90% - 18.5m
car 87% - 54.9m
car 100% - 16.5m
car 83% - 47.4m
car 70% - 40.8m
car 95% - 42.0m
car 94% - 26.6m


Radar Input
3D anchors
Image Input
3D Proposals Generator
2D Mapping
conv layers
Radar Proposal Refinement
RoI pooling
Box Classifier
Box Regressor
Radar 2D Proposals
Distance Refinement
NMS
conv
Box Classifier
Box Regressor
Distance Regressor
RPN
Image 2D Proposals
bbox regressor
FC
RoI pooling
softmax
Fast R-CNN

Results

TABLE I: Performance on the nuScenes validation set.

	Weighted AP	AP	AP50	AP75	AR	MAE
Faster R-CNN	No	34.95	58.23	36.89	40.21	-
RRPN	No	35.45	59.00	37.00	**42.10**	-
Ours	No	**35.60**	**60.53**	**37.38**	**42.10**	2.65
Faster R-CNN	Yes	43.78	-	-	-	-
CRF-Net	Yes	43.95	-	-	-	-
Ours	Yes	**44.49**	-	-	-	-

TABLE II: Per-class performance

	Car	Truck	Person	Bus	Bicycle	Motorcycle
Faster R-CNN	51.46	33.26	27.06	47.73	24.27	25.93
RRPN	41.80	**44.70**	17.10	**57.20**	21.40	**30.50**
Ours	**52.31**	34.45	**27.59**	48.30	**25.00**	25.97

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}.$$

TABLE III: Per-class Mean Absolute Error (MAE) for distance estimation

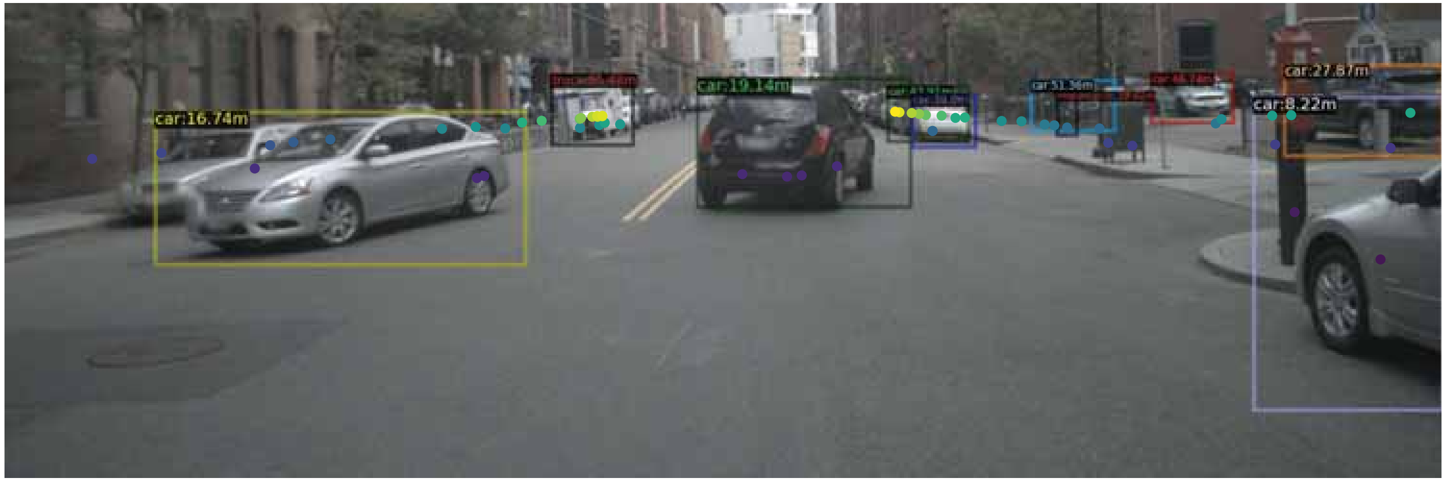
Category	Car	Truck	Person	Bus	Bicycle	Motorcycle
MAE	2.66	3.26	2.99	3.187	1.97	2.81

Results


pedestrian:8.26m
pedestrian:13.16m
pedestrian:17.56m
pedestrian:20.16
pedestrian:21.43m
pedestrian:20.77m
estrian:15.7m


pedestrian 99% - 8.5m
pedestrian 94% - 12.0m
pedestrian 89% - 12.2m
- 10.7m
pedestrian 74% -
pedestrian 87% - 14.6m
pedestrian 87%
pedestrian 81% - 20.5m
pedestrian 78% - 21.6m
pedestrian 63% - 23.2m

Results


car:16.74m
car:19.14m
car:51.36m
car:27.87m
car:8.22m


pedestrian 75% - 19.5m
car 90% - 18.5m
car 87% - 54.9m
car 100% - 16.5m
car 83% - 47.4m
car 70% - 40.8m
car 95% - 42.0m
car 94% - 26.6m

Conclusion

- Contributions
  - A framework for radar-camera sensor fusion
  - Joint object detection and distance estimation

- Future work
  - Improve distance estimation by aggregating radar detections for large objects
  - Perform a thorough ablation study to evaluate the influence of each part of the proposed network

Thanks!

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE