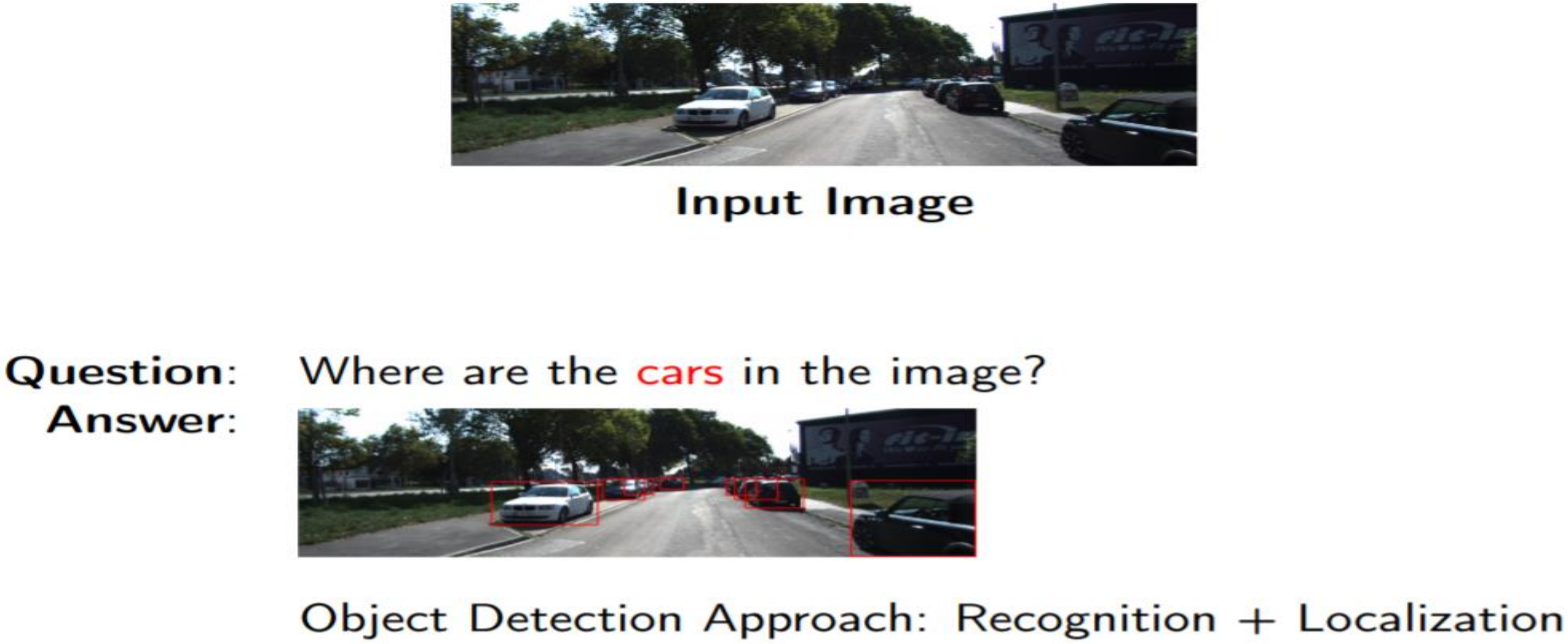
Paper Discussion

Joint 3D Proposal Generation and Object Detection from View Aggregation (AVOD)

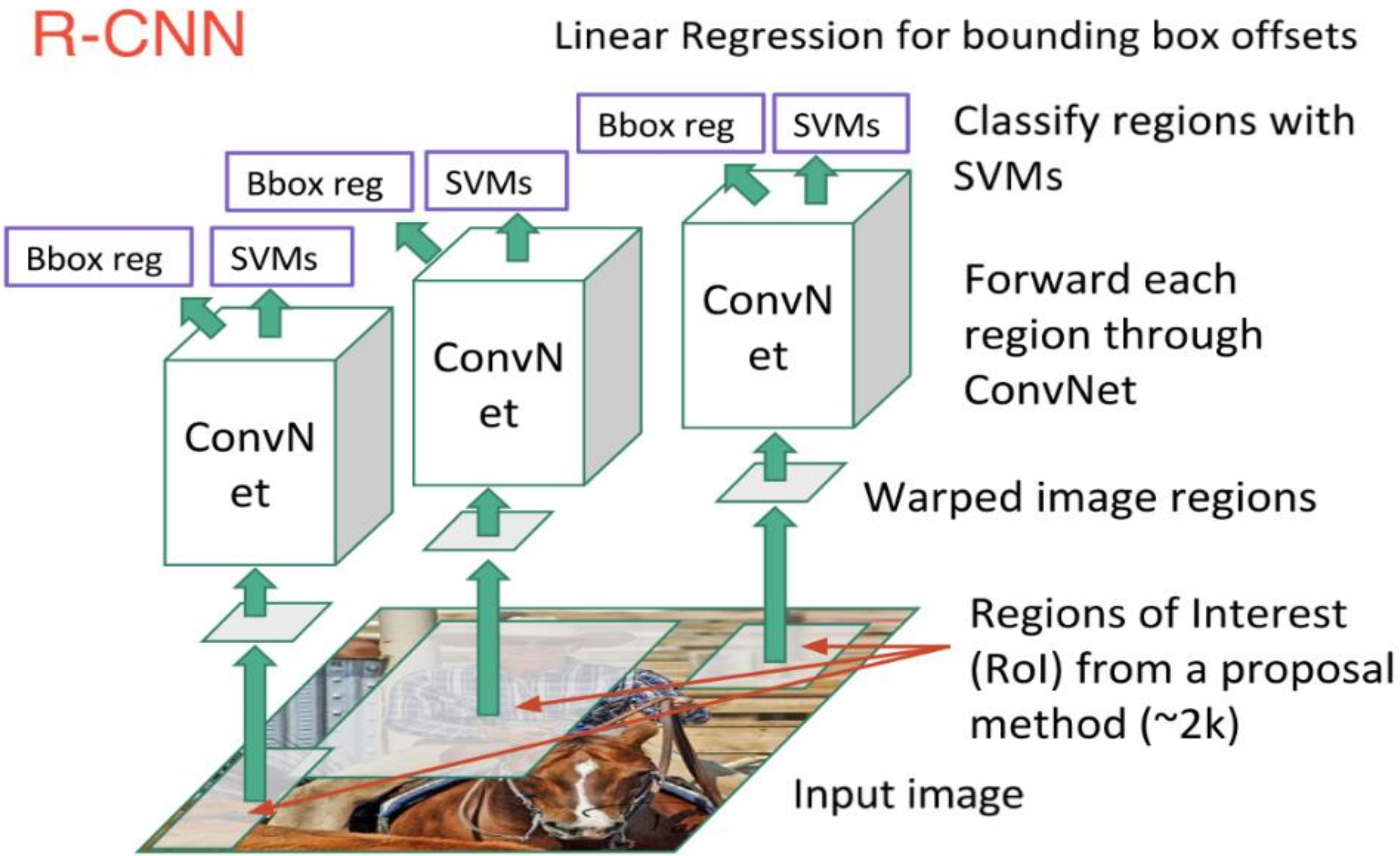
2018-03-29

Background: 2D Object Detection on CNN

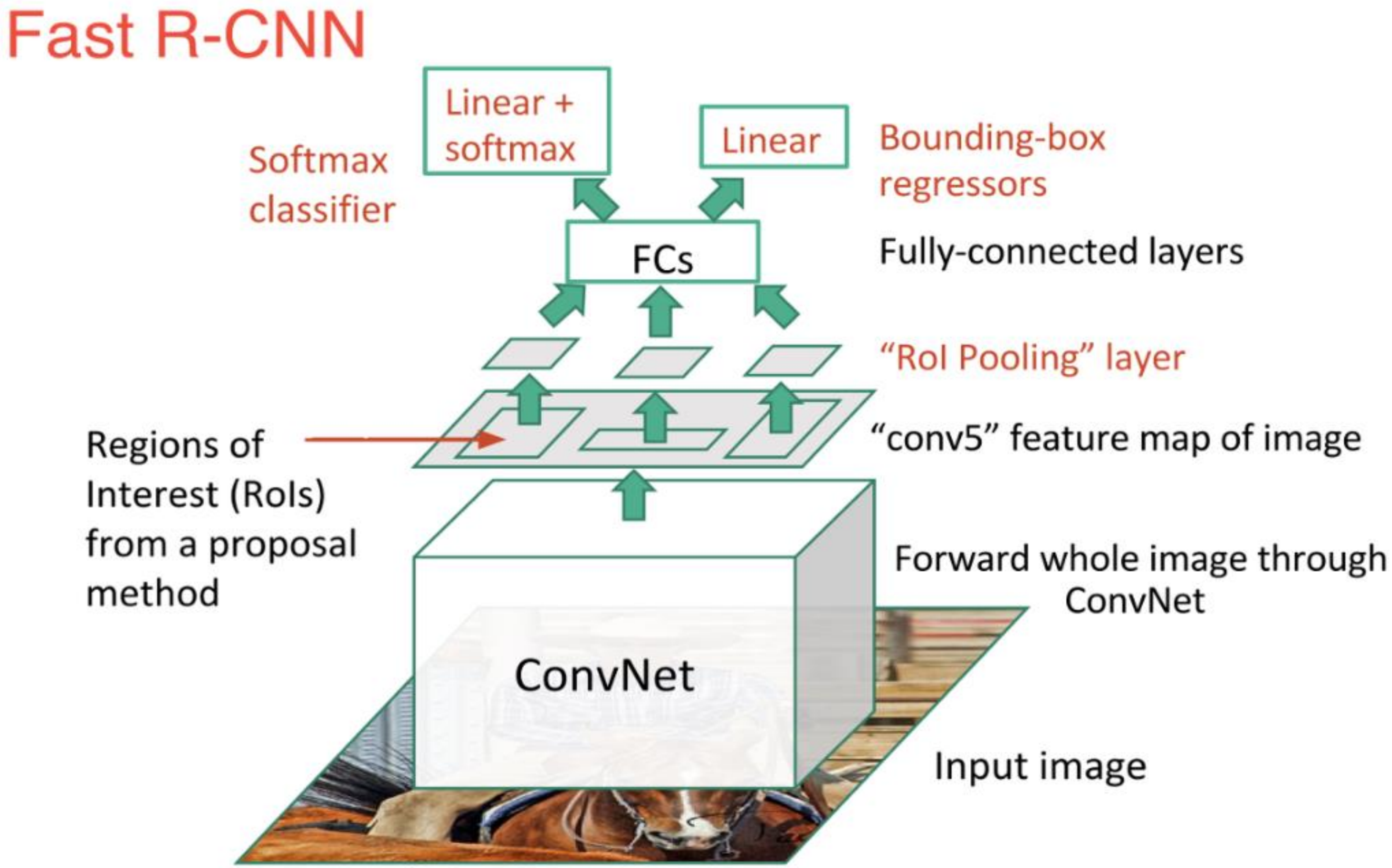


- Candidate Box Selection
- Feature Extraction
- Classification+ Bounding Box Regression

Background: 2D Object Detection on CNN

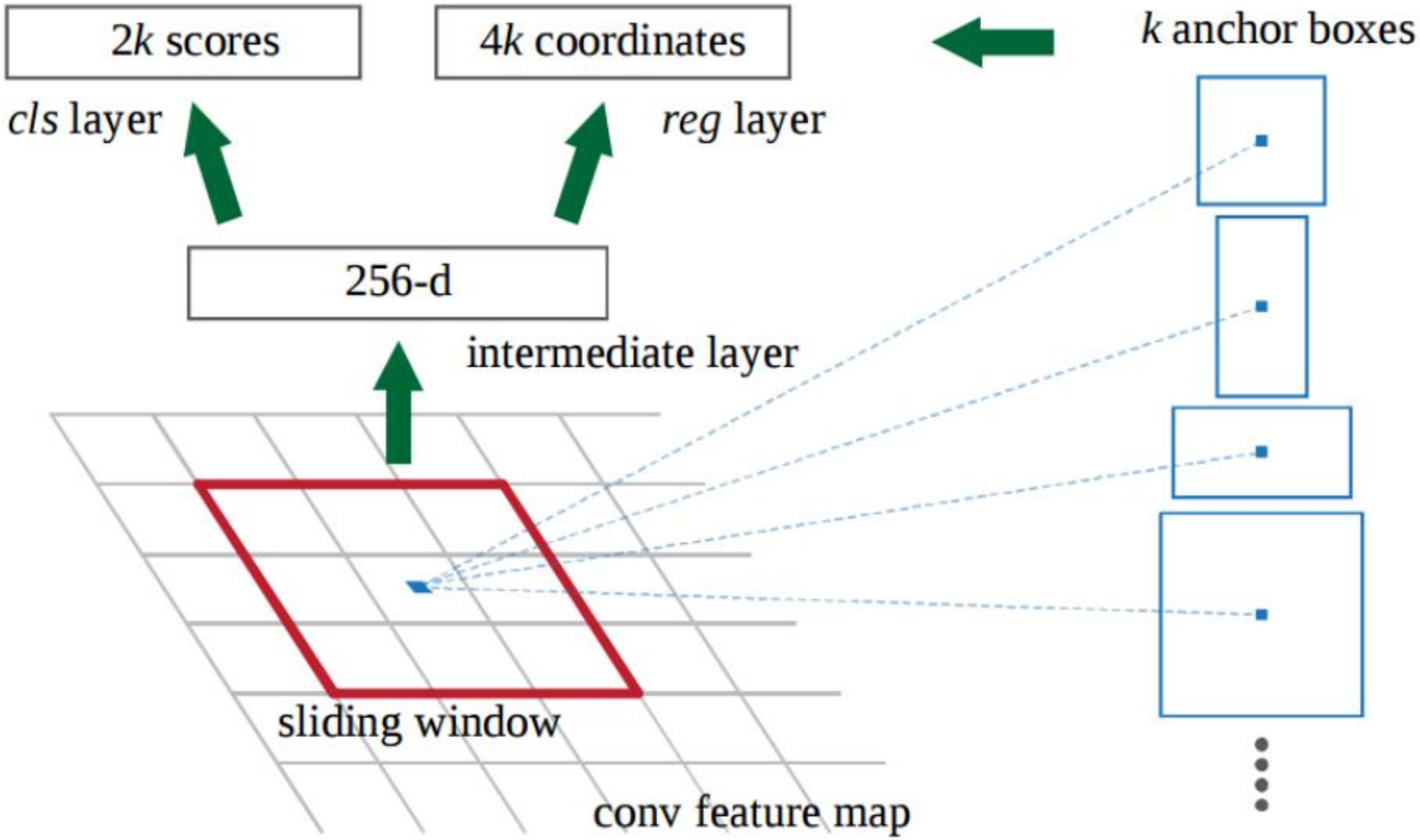


Background: 2D Object Detection on CNN

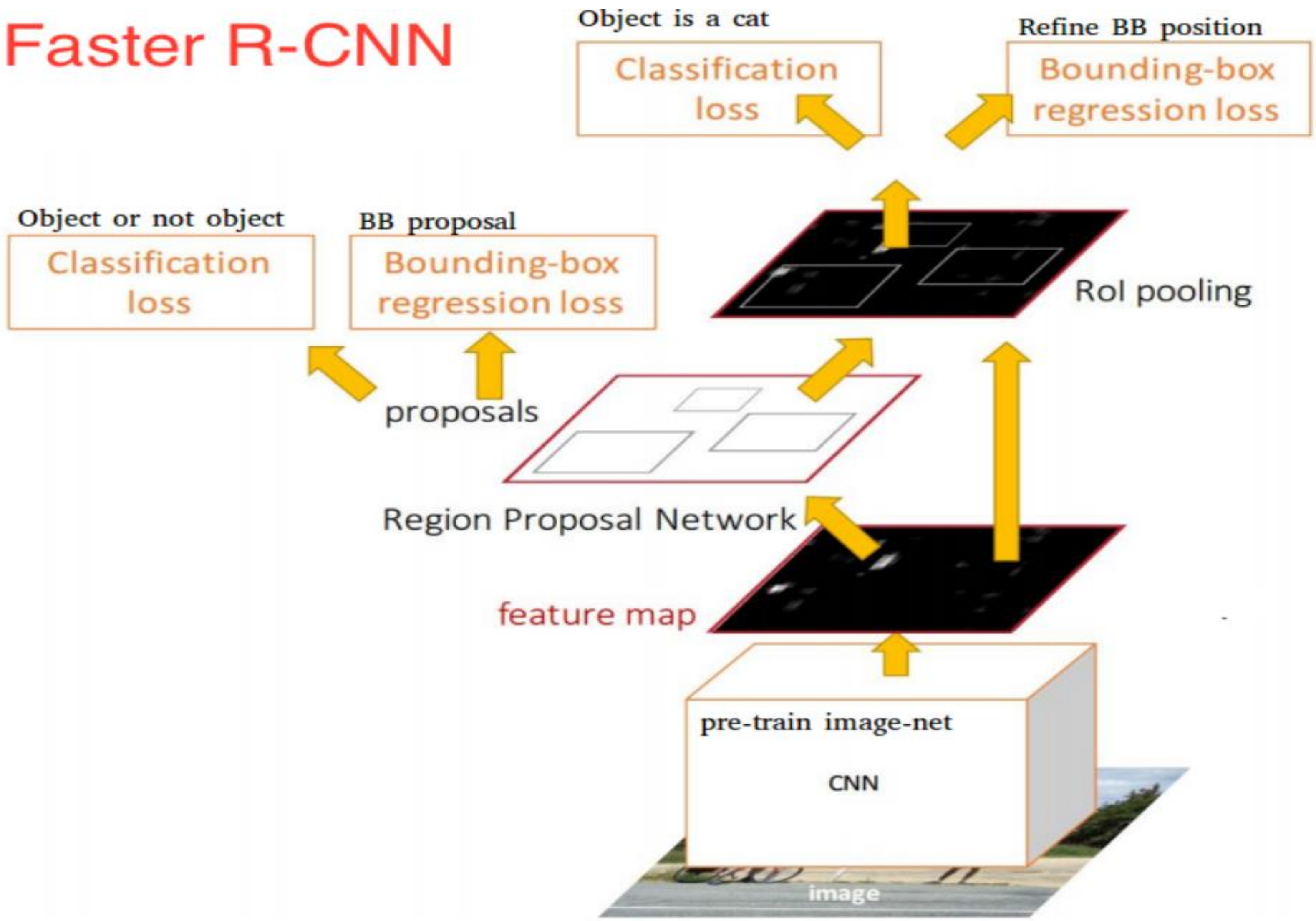


Background: 2D Object Detection on CNN

RPN: Regional Proposal Network

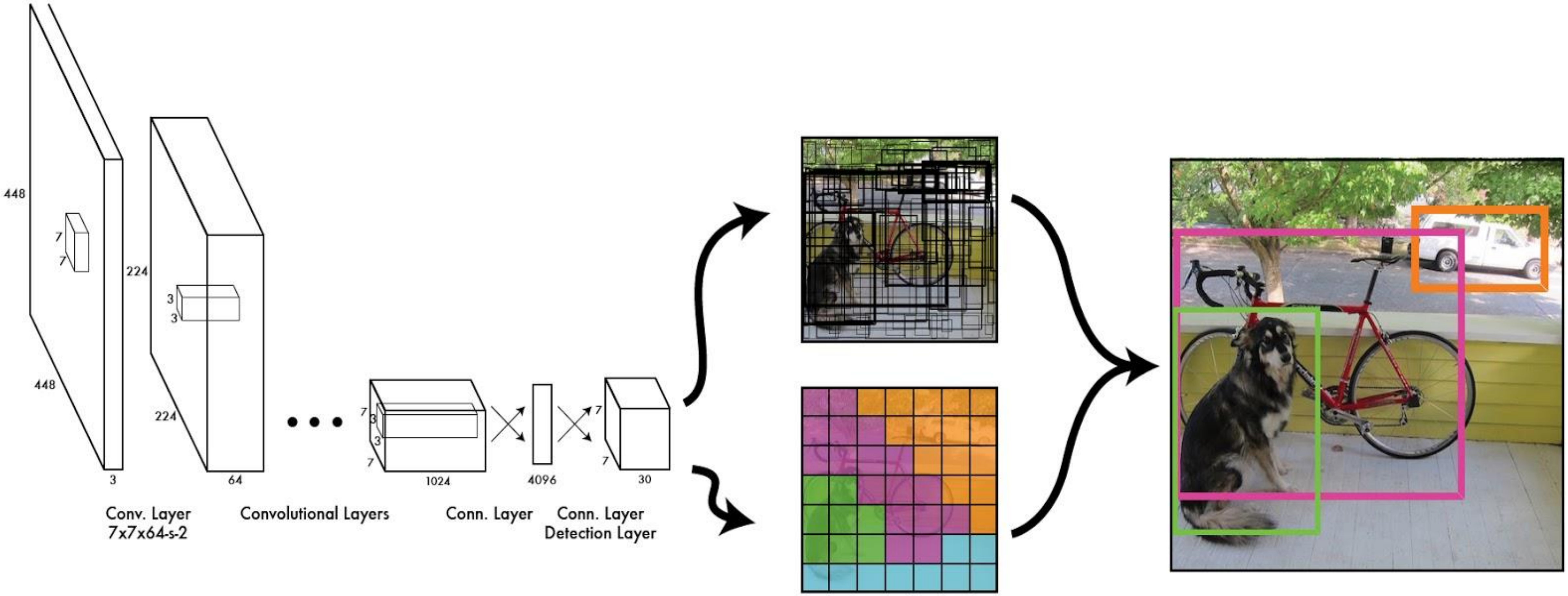


Background: 2D Object Detection on CNN



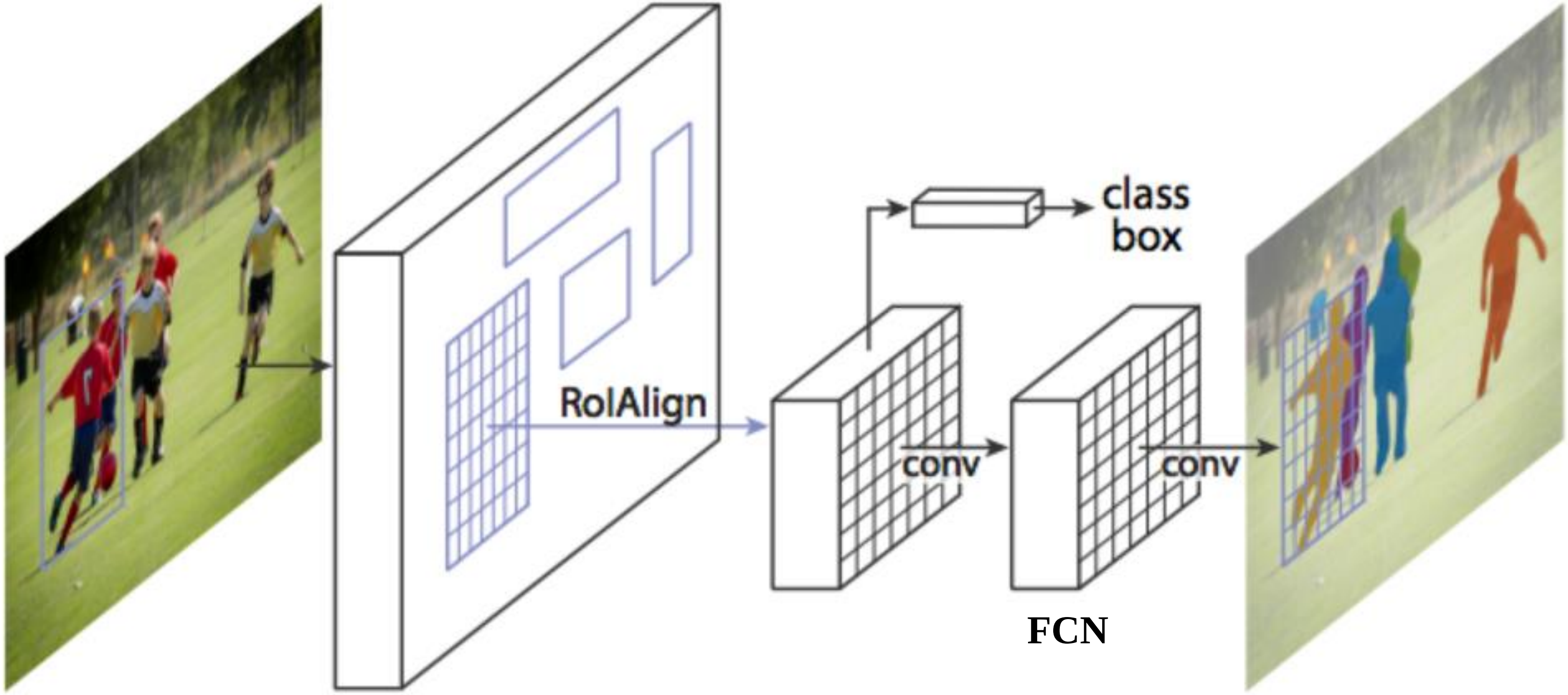
Background: 2D Object Detection on CNN

YOLO



Background: 2D Object Detection on CNN

Mask R-CNN



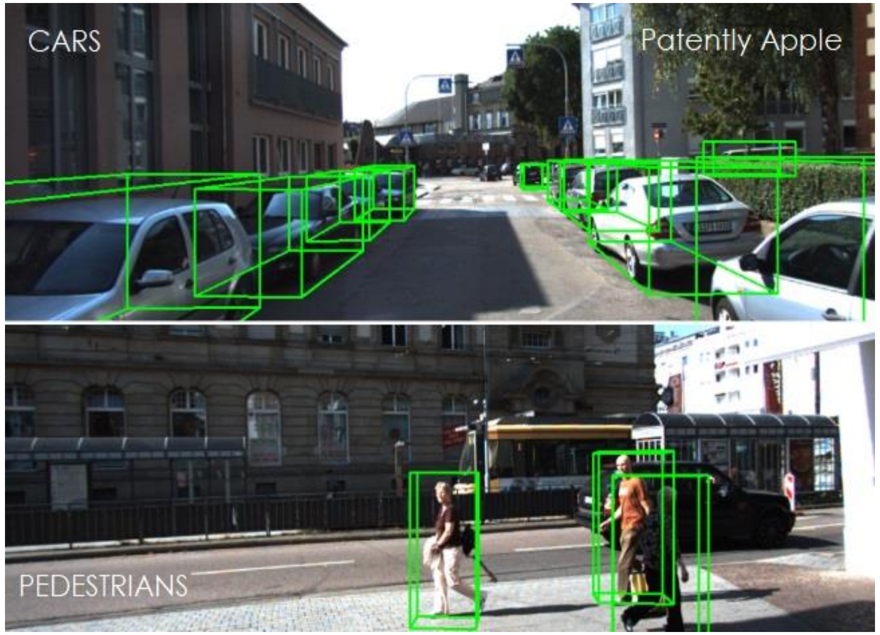
FCN

Problem Domain

- 3D object detection from images and point cloud for autonomous driving

- Deep neural networks

- KITTI 3D object detection benchmark



Challenges

2D $\longrightarrow$ 3D:

- Low resolution of the input data

- Missed instances during region proposal generation cannot be recovered

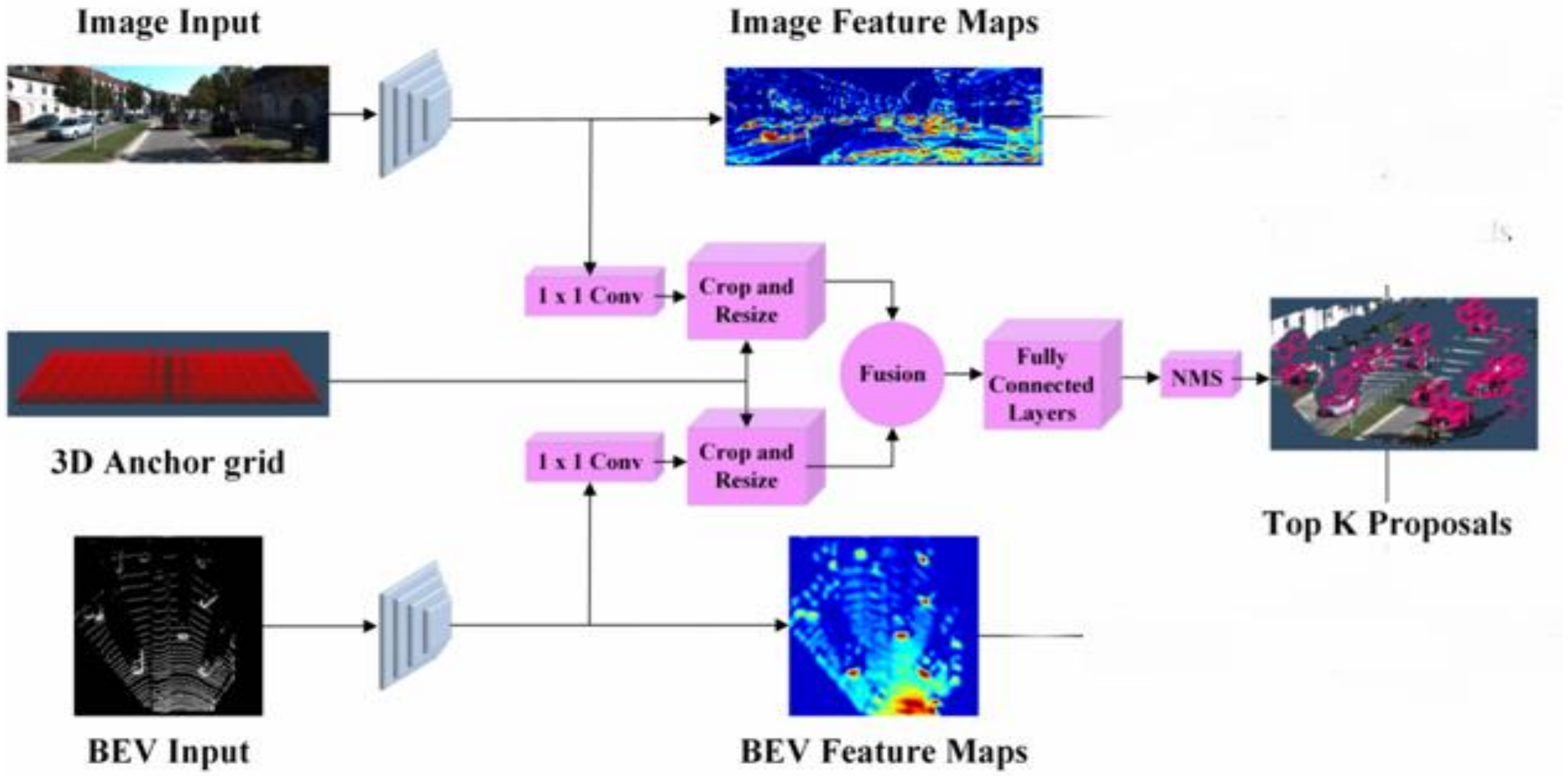
- Oriented bounding box estimation

Related Work

- Hand Crafted Feature For Proposal Generation
- Proposal Free Single Shot Detectors
- Monocular-Based Proposal Generation
- Monocular-Based 3D Object Detections
- 3D Region Proposal Networks (RPN)

Solutions

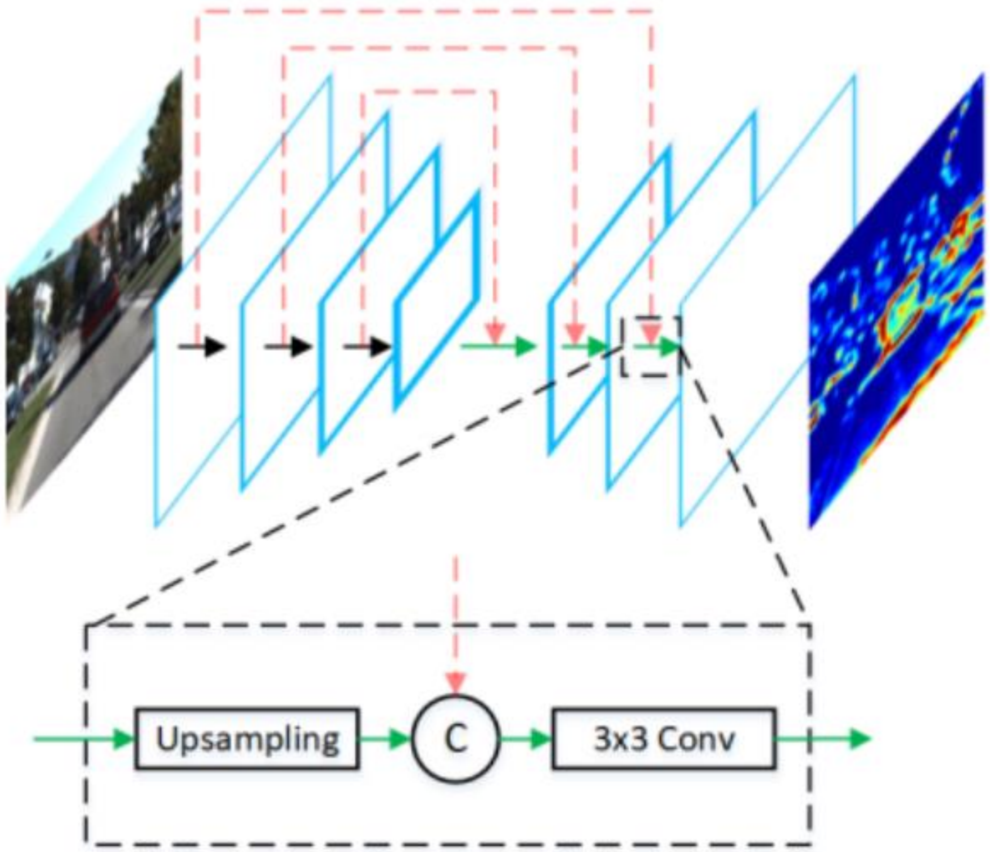
First Stage Detection



Solutions

Feature Map Generating

- BEV images and RGB images
- Feature Pyramid Network (FPN)
- Deconder (Upsampling)



Solutions

3D Anchor Generation (t_x, t_y, t_z) , (d_x, d_y, d_z)

- 3D anchor box grid (80-100K anchors)
- Crop and resize from the feature maps (projection on BEV and RGB feature maps)
- Dimensionality Reduction Via 1*1 Convolutional Layers

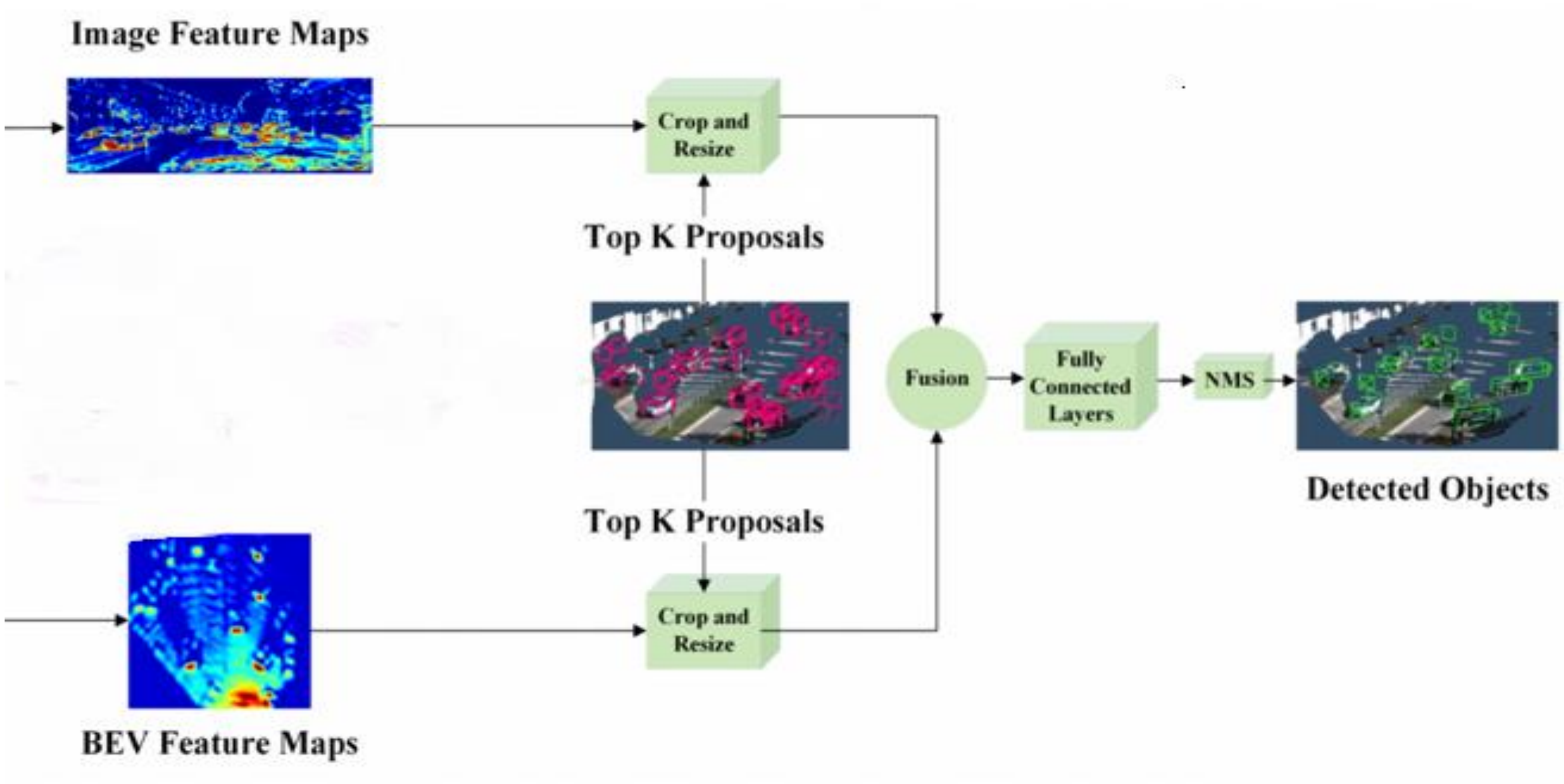
Solutions

3D Proposal Generation

- Two sets of feature crops are fused (element-wise mean)
- Fed into fully collected layers to calculate object score and 3D box regression

Solutions

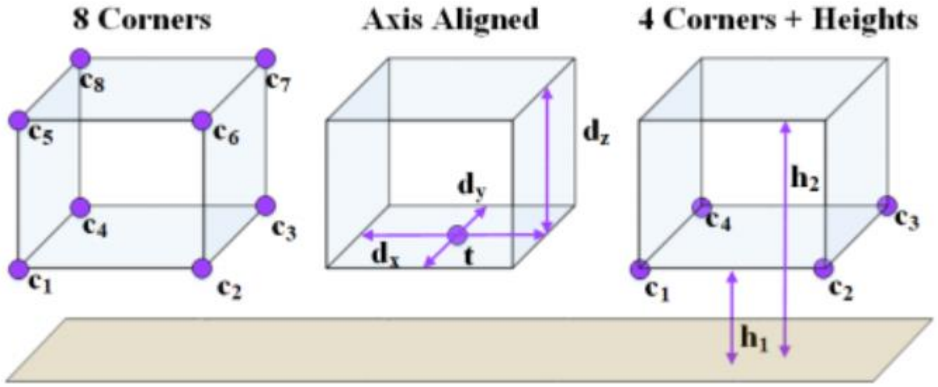
Second Stage Detection



Solutions

Second Stage Detection

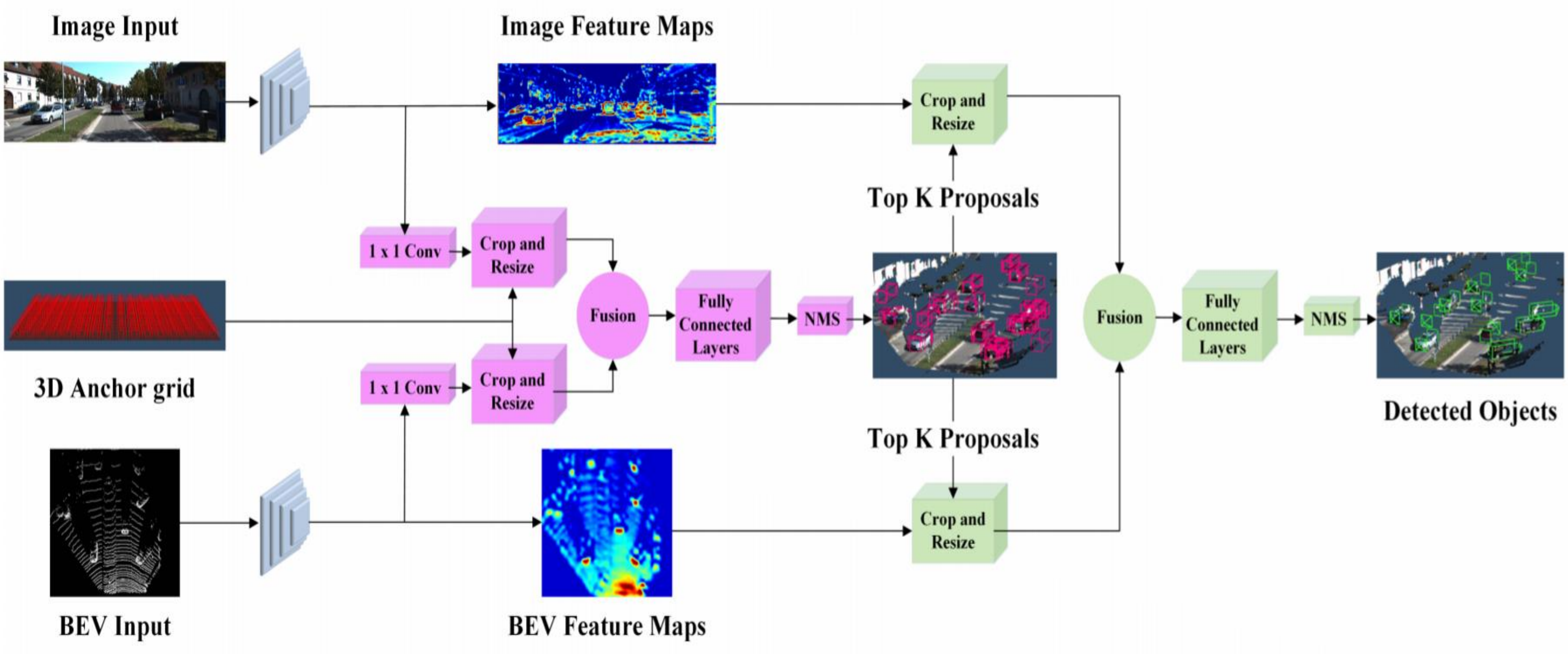
- 3D Bounding Box Encoding (8 corners to 4 corners+ heights)



- Orientation Vector Regression

Solutions

Overview



Result

Car

	Method	Setting	Code	Moderate	Easy	Hard	Runtime
1	AVOD-FPN		code	**71.88 %**	**81.94 %**	**66.38 %**	0.1 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View A							
2	F-PointNet			70.39 %	81.20 %	62.19 %	0.17 s
3	DF-PC_CNN			66.22 %	80.28 %	58.94 %	0.5 s
4	AVOD		code	65.78 %	73.59 %	58.38 %	0.08 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View A							
5	VxNet(LiDAR)			65.11 %	77.47 %	57.73 %	0.03 s
6	MV3D			62.35 %	71.09 %	55.12 %	0.36 s

Pedestrian

	Method	Setting	Code	Moderate	Easy	Hard	Runtime
1	F-PointNet			**44.89 %**	**51.21 %**	**40.23 %**	0.17 s
2	AVOD-FPN		code	39.00 %	46.35 %	36.58 %	0.1 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View							
3	VxNet(LiDAR)			33.69 %	39.48 %	31.51 %	0.03 s
4	AVOD		code	31.51 %	38.28 %	26.98 %	0.08 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View							
5	3dSSD			17.35 %	20.22 %	17.20 %	0.03 s
6	LMNetV2			11.46 %	13.64 %	11.57 %	0.02 s

Cyclist

	Method	Setting	Code	Moderate	Easy	Hard	Runtime
1	F-PointNet			**56.77 %**	**71.96 %**	**50.39 %**	0.17 s
2	VxNet(LiDAR)			48.36 %	61.22 %	44.37 %	0.03 s
3	AVOD-FPN		code	46.12 %	59.97 %	42.36 %	0.1 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View A							
4	AVOD		code	44.90 %	60.11 %	38.80 %	0.08 s
J. Ku, M. Mozifian, J. Lee, A. Harakeh and S. Waslander: Joint 3D Proposal Generation and Object Detection from View A							
5	LMNetV2			3.23 %	2.84 %	3.28 %	0.02 s