RadarNet: Exploiting Radar for Robust Perception of Dynamic Objects

Bin Yang*, Runsheng Guo*, Ming Liang,
Sergio Casas, Raquel Urtasun

UberATG University of Toronto University of Waterloo

Sensors for Self-Driving

Camera

- Rich texture information
- Cheap and high-resolution
- No explicit depth information
- Sensitive to lighting conditions

LiDAR

- Accurate geometry
- Invariant to ambient light
- Limited resolution
- Sensitive to weather

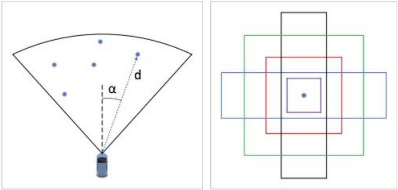
Radar

- Measures radial distance & velocity
- Operates at longer range
- More robust to weather
- Lower resolution than LiDAR
- Noisy returns from clutter & multipaths

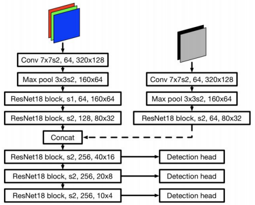
Related Work: Radar as 3D Points

Radar + Camera

- Cascade fusion [1]
- Feature fusion [2,3]



Radar points as anchors



Feature fusion

Strengths

- Radar provides sparse but reliable 3D depth information for images

Weaknesses

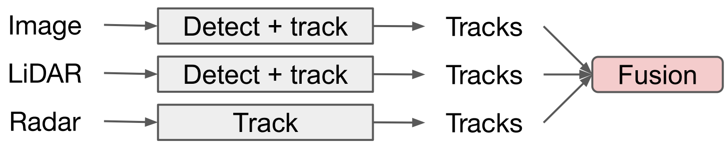
- The performance cannot match LiDAR based systems

[1] RRPN: Radar Region Proposal Network for Object Detection in Autonomous Vehicles. [R. Nabati, et al. ICIP 2019]
[2] RVNet: Deep Sensor Fusion of Monocular Camera and Radar for Image-based Obstacle Detection in Challenging Environments. [V. John, et al. PSIVT 2019]
[3] Distant Vehicle Detection Using Radar and Vision. [S. Chadwick, et al. ICRA 2019]

Related Work: Radar as Objects

Radar tracks + LiDAR tracks [1]

- Track-level sensor fusion with simple object association



Strengths

- Higher object recall by multi-sensor fusion

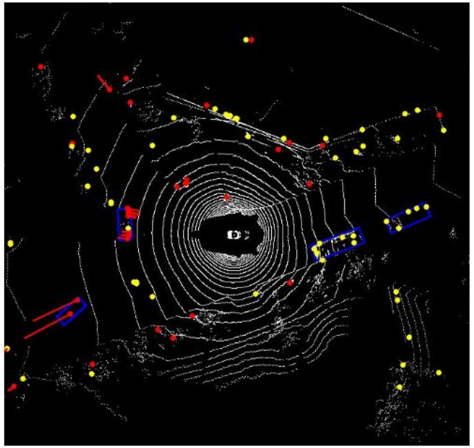
Weaknesses

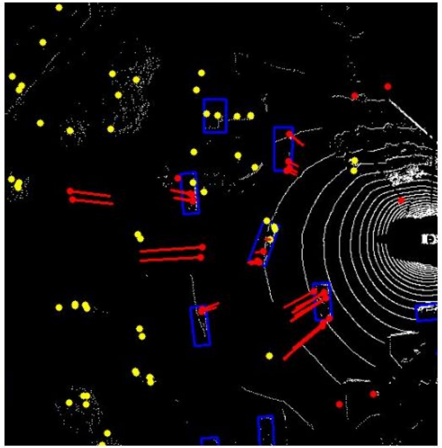
- Limited exploitation of complementary information between sensors

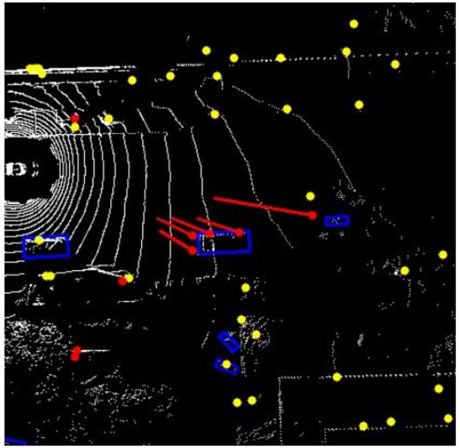
[1] A Multi-Sensor Fusion System for Moving Object Detection and Tracking in Urban Driving Environments. [H. Cho, et al. ICRA 2014]

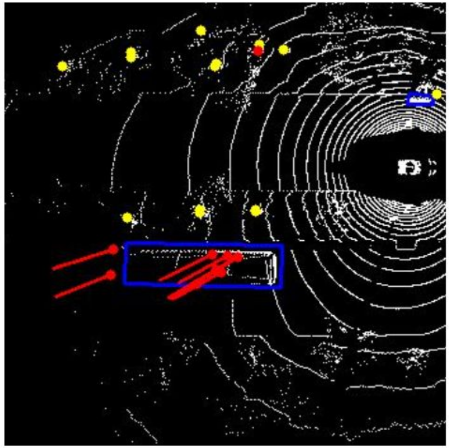
LiDAR v.s. Radar

Sensor Modality	Detection Range	Range Accuracy	Azimuth Resolution	Velocity Accuracy
LiDAR	100 m	2 cm	$0.1^\circ \sim 0.4^\circ$	-
Radar	250 m	10 cm near range 40 cm far range	$3.2^\circ \sim 12.3^\circ$ near range 1.6° far range	0.1 km/h

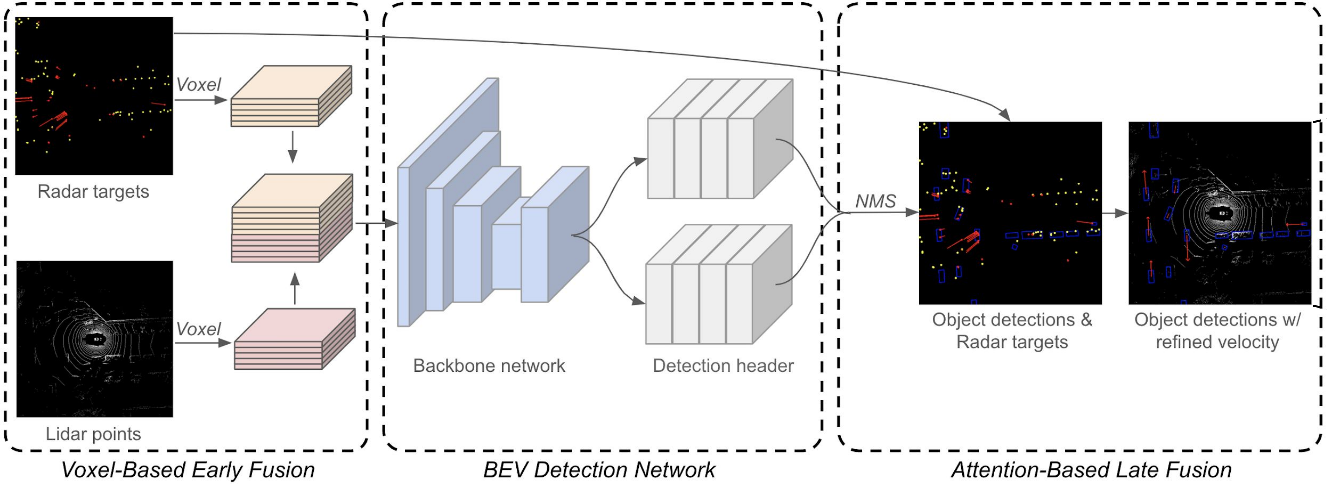

Sparsity

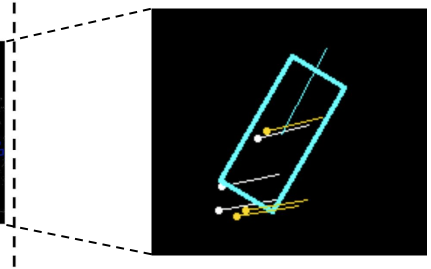

False detections


Inaccurate position


Inaccurate position

RadarNet: Multi-Level Radar Fusion





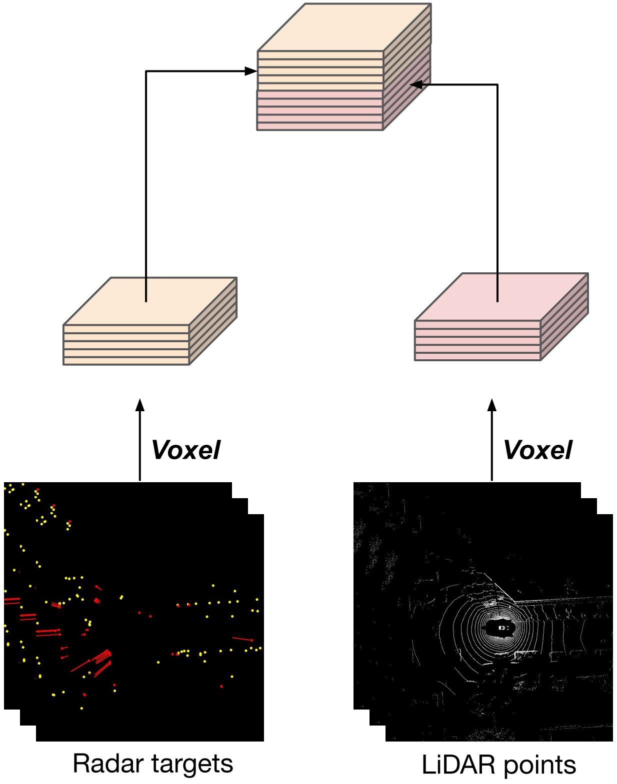
Learned soft association

- **Early fusion:** supplements sparse LiDAR points at long range with Radar returns
- **Late fusion:**
  - takes into account uncertainties in object detections and Radar returns
  - learns soft association between them

Voxel-Based Early Fusion

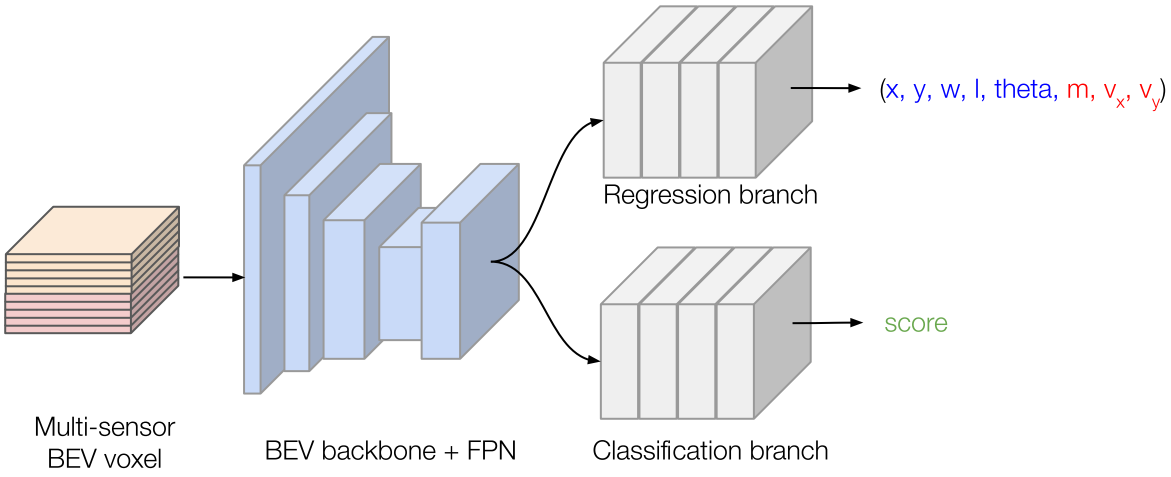
- LiDAR BEV voxel
  - Multi-sweep point clouds in current ego coordinates
  - #channels = #height slices * #sweeps
  - Voxel feature: distance-weighted density

- Radar BEV voxel
  - Multi-cycle point clouds in current ego coordinates
  - #channels = #cycles (ignore height)
  - Voxel feature: motion-aware occupancy



Detection Network

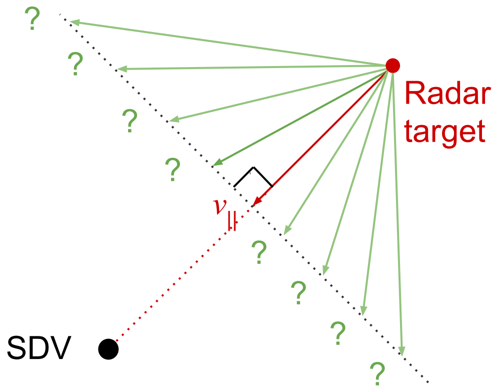
- **Multi-scale BEV Backbone:** same as PnPNet [1]
- Detection Output:
  - BEV bounding box: (x, y, w, l, theta)
  - Velocity estimate: moving probability, 2D velocity (v_x, v_y)
  - Classification score



[1] PnPNet: End-to-End Perception and Prediction with Tracking in the Loop. [M. Liang, et al. CVPR 2020]

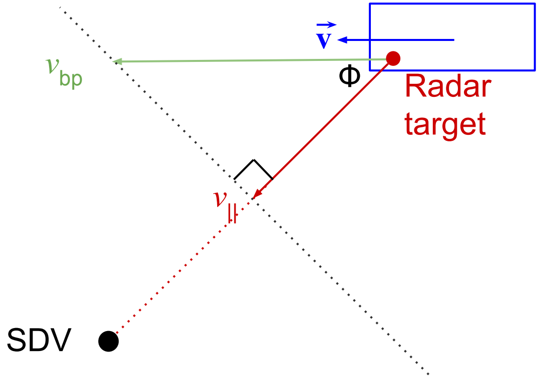
Attention-Based Late Fusion

- **Step 1:** Alignment of Radar velocity to objects
  - It's ambiguous to infer the 2D object velocity given radial velocity $v_{\parallel}$ alone



Attention-Based Late Fusion

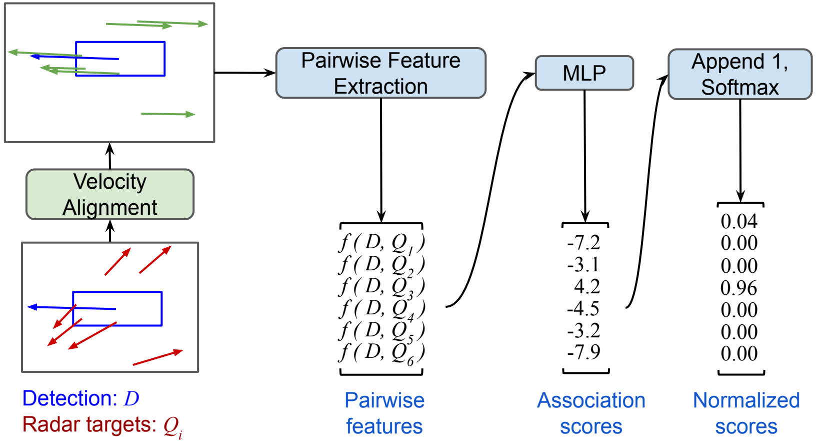
- **Step 1:** Alignment of Radar velocity to objects
  - It's ambiguous to infer the 2D object velocity given radial velocity $v_{||}$ alone
  - To address this, we alignment the radial velocity $v_{||}$ from Radar with the velocity estimate $\vec{\mathbf{v}}$ from detection, and get the back-projected velocity v_{bp}



Attention-Based Late Fusion

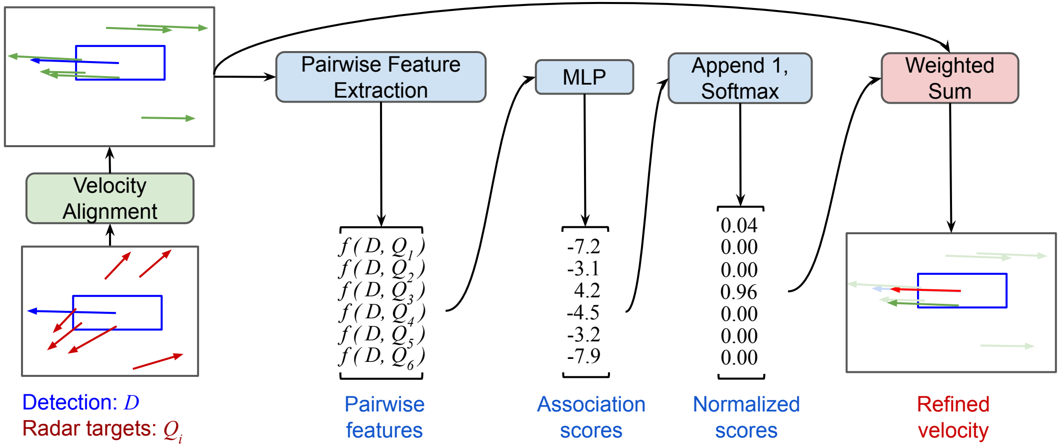
- **Step 2:** Soft association between Radar targets & object
    - Pairwise features = Detection feature + Radar feature

$$\left(w, l, \|\mathbf{v}\|, \frac{v_x}{\|\mathbf{v}\|}, \frac{v_y}{\|\mathbf{v}\|}, \cos(\gamma)\right) \qquad (\mathrm{d}x, \mathrm{d}y, \mathrm{d}t, v^{\mathrm{bp}})$$



Attention-Based Late Fusion

- **Step 3:** Information aggregation
  - The refined velocity is the weighted sum of
    1. back-projected velocities from Radar targets
    2. the initial velocity estimate from detection



Model Training

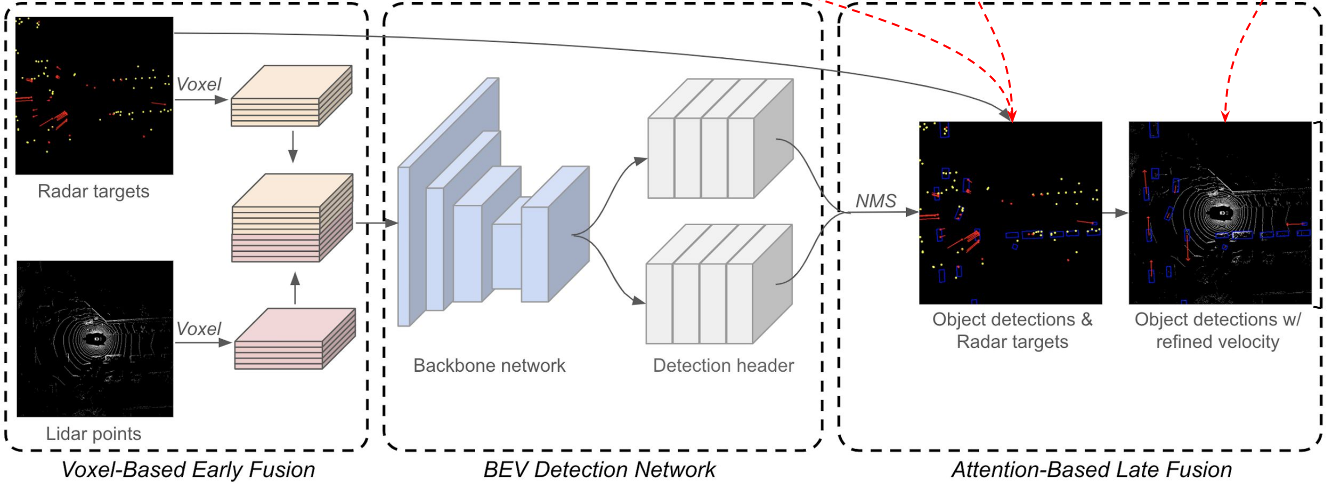
- Multi-task loss function:

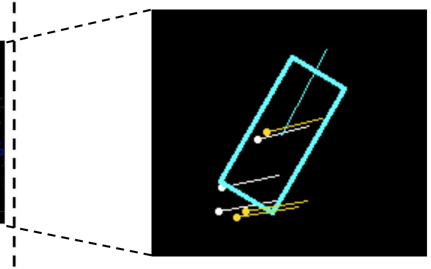
$$\mathcal{L} = (\mathcal{L}_{\text{cls}}^{\text{det}} + \alpha \cdot \mathcal{L}_{\text{reg}}^{\text{det}}) + \beta \cdot (\mathcal{L}_{\text{cls}}^{\text{velo}} + \mathcal{L}_{\text{reg}}^{\text{velo}}) + \delta \cdot \mathcal{L}_{\text{reg}}^{\text{velo\_attn}}$$

Detection loss

Velocity loss on detection outputs

Velocity loss on Radar refined velocity estimates





Learned soft association

Evaluation Results on nuScenes

Method	Input	Cars		Motorcycles	
		AP↑	AVE↓	AP↑	AVE↓
MonoDIS	I	47.8	-	28.1	-
PointPillar	L	70.5	0.269	20.0	0.603
PointPillar+	L	76.7	0.209	35.0	0.371
PointPainting	L+I	78.8	0.206	44.4	0.351
3DSSD	L	81.2	0.188	36.0	0.356
CBGS	L	82.3	0.230	50.6	0.339
RadarNet (LiDAR only)	L	84.2	0.203	51.0	0.316
RadarNet (Full model)	L+R	**84.5**	**0.175**	**52.9**	**0.269**

Model Input: I = image, L = LiDAR, R = Radar

Ablation Study

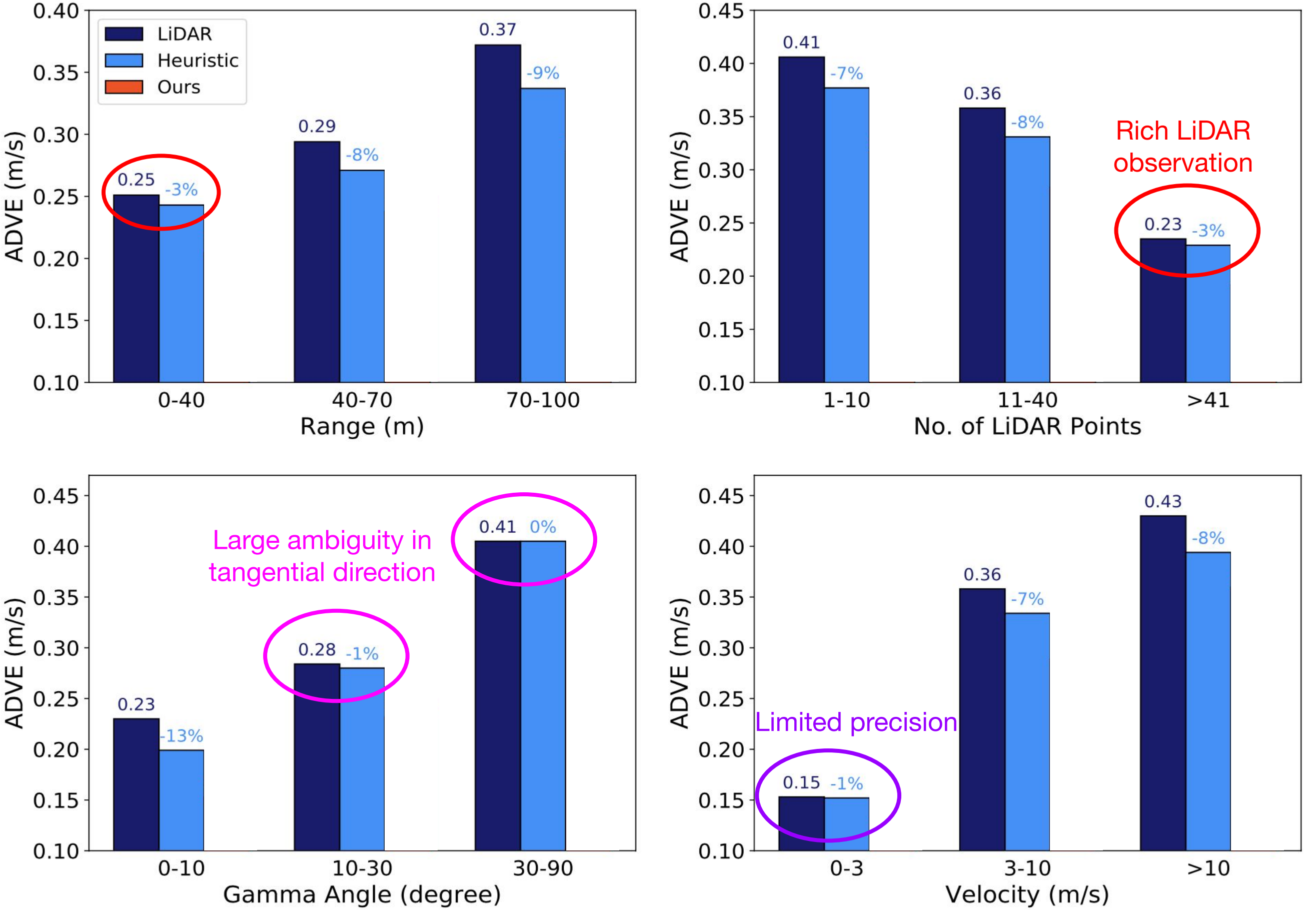
nuScenes (<50m range)

Model	LiDAR	Radar		Cars		Motorcycles	
		Early	Late	AP@2m↑	AVE↓	AP@2m↑	AVE↓
LiDAR	✓	-	-	87.6	0.203	53.7	0.316
Early	✓	✓	-	+0.3	-2%	+1.9	-0%
Heuristic	✓	✓	heuristic	+0.3	-9%	+1.9	-4%
RadarNet	✓	✓	attention	**+0.3**	**-14%**	**+1.9**	**-15%**

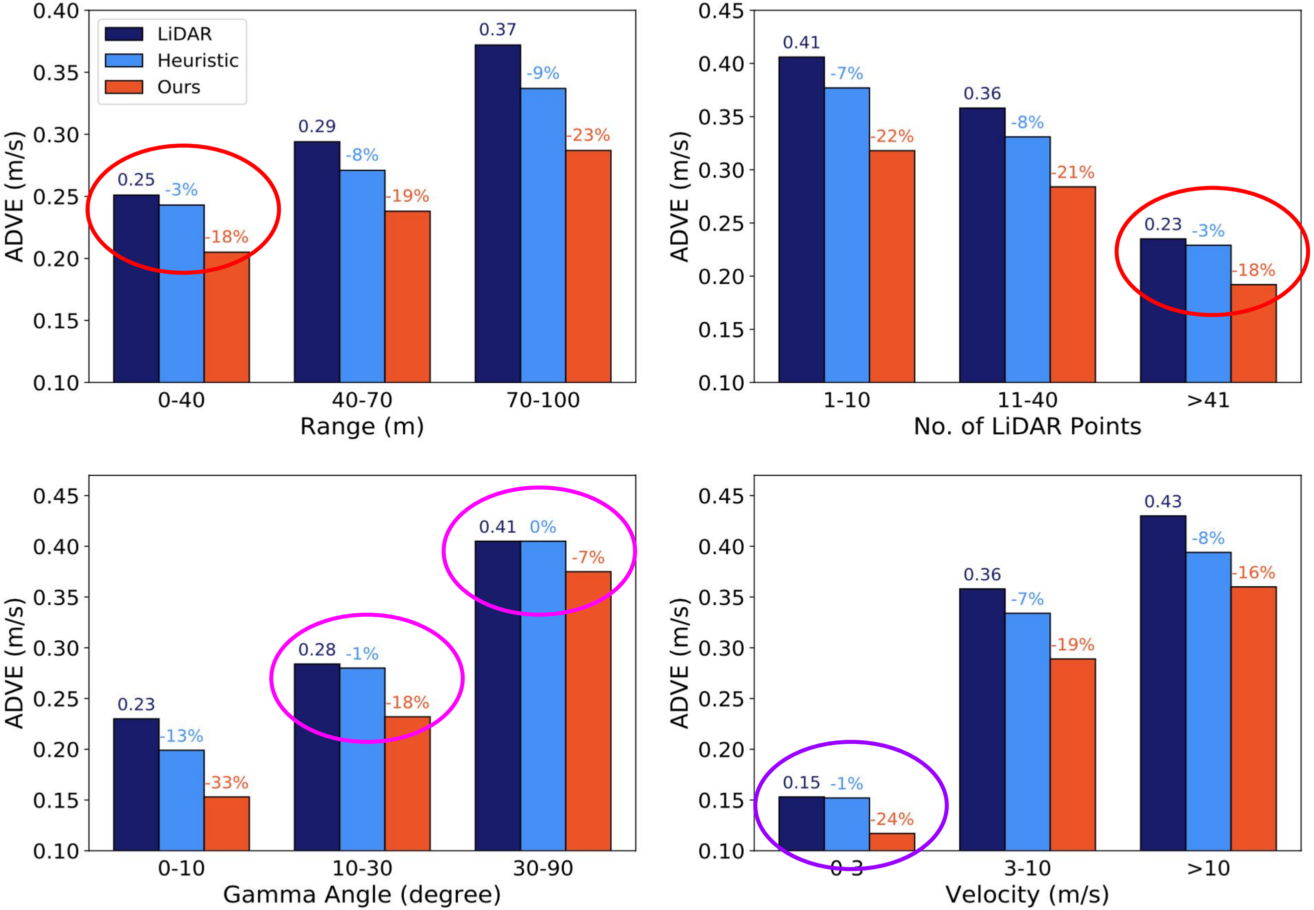
DenseRadar (<100m range)

Model	LiDAR	Radar		Vehicles AP ↑			ADVE ↓
		Early	Late	0-40m	40-70m	70-100m	
LiDAR	✓	-	-	95.4	88.0	77.5	0.285
Early	✓	✓	-	+0.3	+0.5	+0.8	-3%
Heuristic	✓	✓	heuristic	+0.3	+0.5	+0.8	-6%
RadarNet	✓	✓	attention	**+0.3**	**+0.5**	**+0.8**	**-19%**

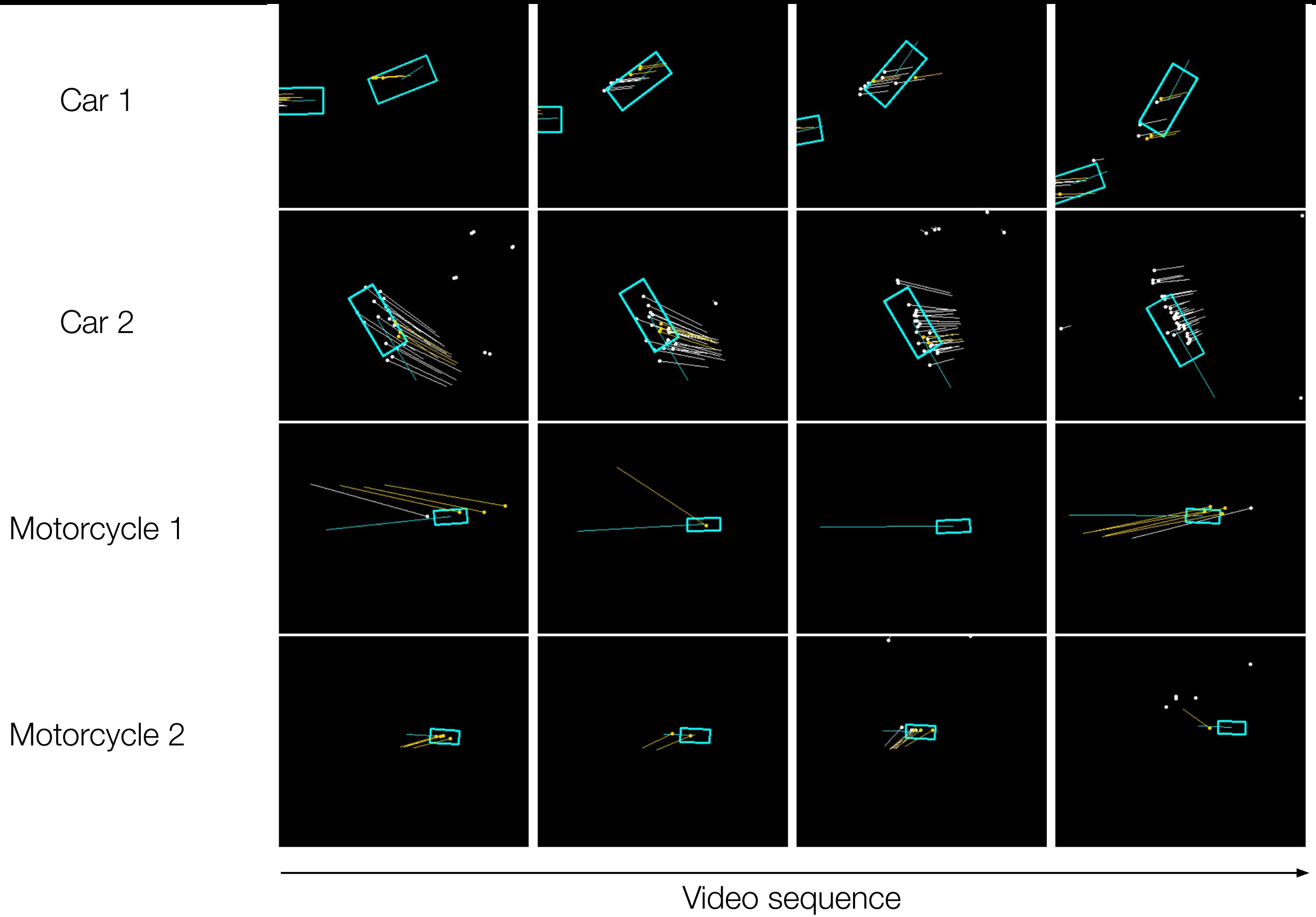
Evaluation on Heuristics (Late Fusion)



Evaluation on Attention (Late Fusion)



Qualitative Results of Object-Radar Association



Conclusion

- Voxel-based early fusion of LiDAR and Radar to exploit long-range evidence of Radar

- Attention-based late fusion of Radar targets and detections to exploit the uncertain Radar velocities

- State-of-the-art results in dynamic object perception

RadarNet: Exploiting Radar for Robust Perception of Dynamic Objects. [B. Yang, et al. ECCV 2020]