RadarNet: Exploiting Radar for Robust Perception of Dynamic Objects

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

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

Weaknesses
- The performance cannot match LiDAR based systems

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

# LiDAR v.s. Radar

<table>
<thead>
<tr>
<th>Sensor Modality</th>
<th>Detection Range</th>
<th>Range Accuracy</th>
<th>Azimuth Resolution</th>
<th>Velocity Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiDAR</td>
<td>100 m</td>
<td>2 cm</td>
<td>0.1° ~ 0.4°</td>
<td>-</td>
</tr>
<tr>
<td>Radar</td>
<td>250 m</td>
<td>10 cm near range, 40 cm far range</td>
<td>3.2° ~ 12.3° near range, 1.6° far range</td>
<td>0.1 km/h</td>
</tr>
</tbody>
</table>

**Images:**
- Sparsity
- False detections
- Inaccurate position
- Inaccurate position
RadarNet: Multi-Level Radar Fusion

- **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
  - $\#\text{channels} = \#\text{height slices} \times \#\text{sweeps}$
  - Voxel feature: distance-weighted density

- Radar BEV voxel
  - Multi-cycle point clouds in current ego coordinates
  - $\#\text{channels} = \#\text{cycles}$ (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

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 $v_\parallel$ alone
  - To address this, we align the radial velocity $v_\parallel$ from Radar with the velocity estimate $\vec{v}$ from detection, and get the back-projected velocity $v_{bp}$
**Step 2:** Soft association between Radar targets & object
- Pairwise features = Detection feature + Radar feature

\[(w, l, \frac{v_x}{\|v\|}, \frac{v_y}{\|v\|}, \cos(\gamma)), \quad (dx, dy, dt, v^{bp})\]
### 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

- Voxel-Based Early Fusion
- BEV Detection Network
- Attention-Based Late Fusion

Learned soft association
### Evaluation Results on nuScenes

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

Model Input: I = image, L = LiDAR, R = Radar
## Ablation Study

### nuScenes (<50m range)

<table>
<thead>
<tr>
<th>Model</th>
<th>LiDAR</th>
<th>Radar Early</th>
<th>Radar Late</th>
<th>Cars AP@2m↑</th>
<th>AVE↓</th>
<th>Motorcycles AP@2m↑</th>
<th>AVE↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiDAR</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>87.6</td>
<td>0.203</td>
<td>53.7</td>
<td>0.316</td>
</tr>
<tr>
<td>Early</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>+0.3</td>
<td>-2%</td>
<td>+1.9</td>
<td>-0%</td>
</tr>
<tr>
<td>Heuristic</td>
<td>✓</td>
<td>✓</td>
<td>heuristic</td>
<td>+0.3</td>
<td>-9%</td>
<td>+1.9</td>
<td>-4%</td>
</tr>
<tr>
<td>RadarNet</td>
<td>✓</td>
<td>✓</td>
<td>attention</td>
<td>+0.3</td>
<td>-14%</td>
<td>+1.9</td>
<td>-15%</td>
</tr>
</tbody>
</table>

### DenseRadar (<100m range)

<table>
<thead>
<tr>
<th>Model</th>
<th>LiDAR</th>
<th>Radar Early</th>
<th>Radar Late</th>
<th>Vehicles AP ↑ 0-40m</th>
<th>40-70m</th>
<th>70-100m</th>
<th>ADVE ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiDAR</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>95.4</td>
<td>88.0</td>
<td>77.5</td>
<td>0.285</td>
</tr>
<tr>
<td>Early</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>+0.3</td>
<td>+0.5</td>
<td>+0.8</td>
<td>-3%</td>
</tr>
<tr>
<td>Heuristic</td>
<td>✓</td>
<td>✓</td>
<td>heuristic</td>
<td>+0.3</td>
<td>+0.5</td>
<td>+0.8</td>
<td>-6%</td>
</tr>
<tr>
<td>RadarNet</td>
<td>✓</td>
<td>✓</td>
<td>attention</td>
<td>+0.3</td>
<td>+0.5</td>
<td>+0.8</td>
<td>-19%</td>
</tr>
</tbody>
</table>
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