Objective, Motivation, and Challenges

Objective
- mmBox aims to design a mmWave-based object detection system, which can effectively predict bounding box for vehicle and pedestrian under extreme outdoor environment.

Motivation
- The high resolution, low-cost, and ability to penetrate small obstacles of mmWave enable the feasibility to be applied in adverse environment.
- Existing works have limited performance, e.g., RODNet[1] only predicts a likelihood cluster on the heatmap, and Radatron[2] lacks distance and height details of objects.

Challenges
- Noise and Sparsity. Only a few parts of transmitted signals are correctly reflected to receivers due to specularity, and the strong reflectors will generate noise.
- Complex Outdoor Environment. More complicated objects and surroundings further increase the difficulty of extracting enough features of targets from the sparse and noisy mmWave reflections.

Data Processing

Static and Dynamic RAMap Generator
- Range FFT is applied to convert time domain signals to the frequency domain, capturing distance details.
- Doppler FFT is applied on varying chirps in a frame to differentiate between stationary and moving entities.
- Angle FFT is applied on signals from non-overlapping virtual antennas to derive the azimuth angle from Range-Doppler data.

Model Architecture

Multi-Scale Bounding Box Generator
- Feature Extractor. This module fuses features from both static and dynamic RAMaps across multiple scales.
- Three-Level Bounding Box Predictor. This module outputs 3 scale predictions. The small size predictions mainly focus on the large bounding boxes, while large-scale predictions consider large ones more.
- Predefined Anchors. These 3x3 anchors are matched with 3-level prediction in 3 different sizes to improve the performance.
- Loss Function:

\[ L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(w^r)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \]

Dataset
- We collected mmWave reflections, gray-scale images, and depth images in outdoor street scenes.
- In total, we got 10,440 samples for training and 2,280 samples for testing.

Results
- mmBox showcases remarkable precision on various metrics including Average Precision (AP) and Classification Accuracy (CA), Average Center Distance (ACD), Average Height/Width Ratio (AWR/AHR), and Average Depth Difference (ADD).

<table>
<thead>
<tr>
<th></th>
<th>CA</th>
<th>AP50</th>
<th>ACD</th>
<th>AHR</th>
<th>AWR</th>
<th>ADD (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle</td>
<td>100%</td>
<td>42%</td>
<td>20 pix.</td>
<td>0.998</td>
<td>1.009</td>
<td>0.80</td>
</tr>
<tr>
<td>Pedestrian</td>
<td>100%</td>
<td>24%</td>
<td>11 pix.</td>
<td>0.995</td>
<td>1.007</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Acknowledgements: