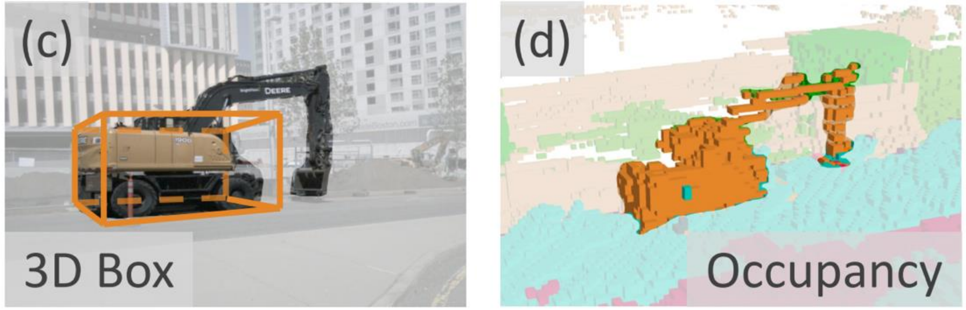
RadarOcc: Robust 3D Occupancy Prediction with 4D Imaging Radar

NeurIPS 2024 Presentation

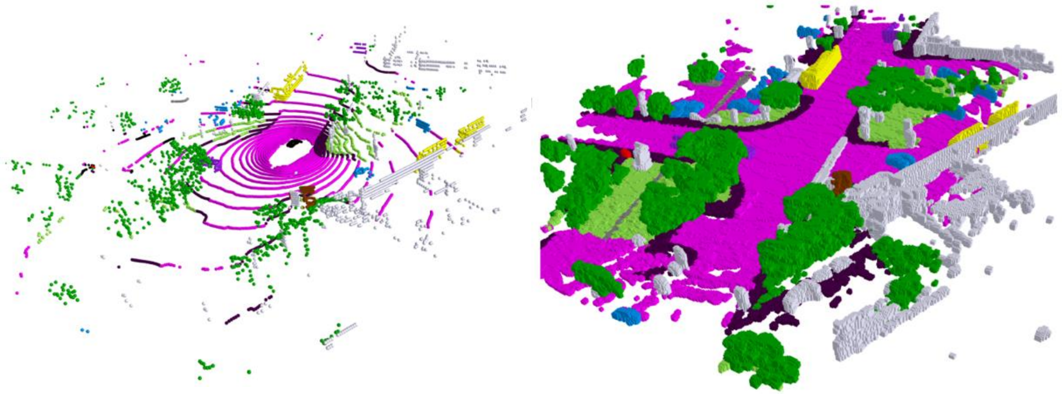
Fangqiang Ding[1], Xiangyu Wen[1], Yunzhou Zhu[2], Yiming Li[3], Chris Xiaoxuan Lu[4]
[1]University of Edinburgh, [2]Georgia Institute of Technology, [3]New York University,
[4]AI Centre, Department of Computer Science, UCL

3D occupancy – a unified scene representation

- Depict scene in both geometric and semantic aspects
- Not limited to foreground-only representation (vs. 3D object detection) and sparse data formats (vs. point cloud segmentation)
- Open-set depiction of scene geometry: out-of-vocabulary items (e.g., animals) and irregular shape (e.g., cranes)



OccNet (ICCV'23)

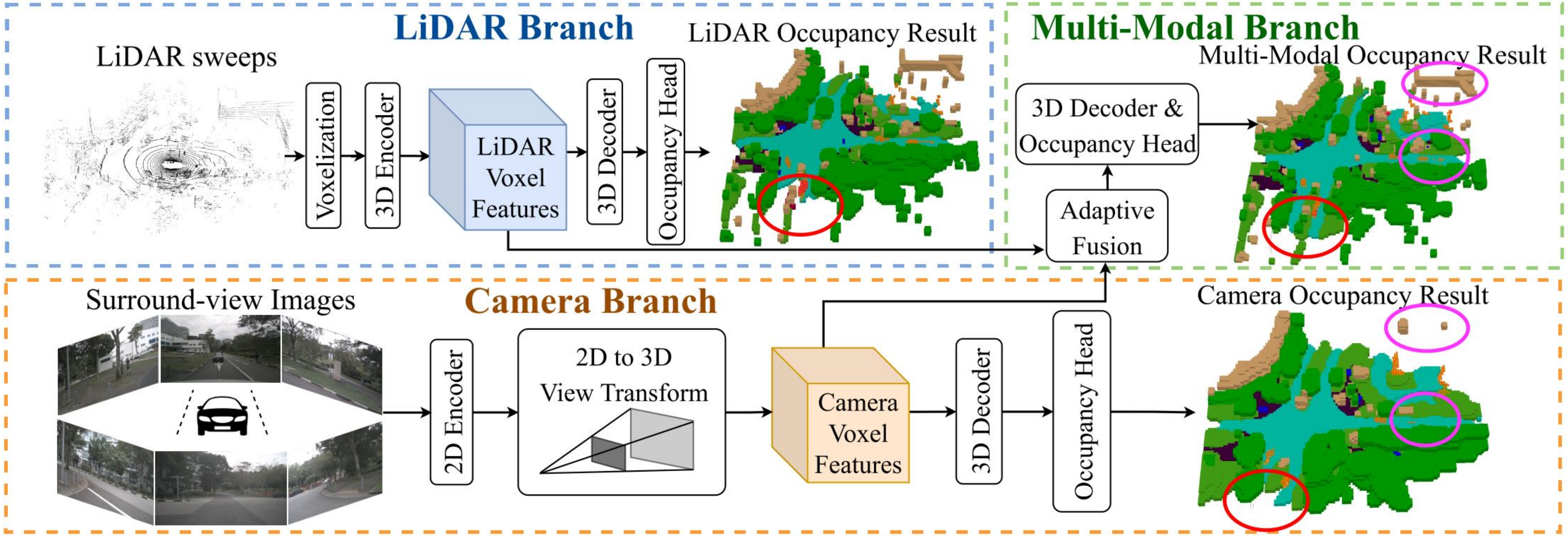


SurroundOcc (ICCV'23)

THE UNIVERSITY of EDINBURGH

Current research gap

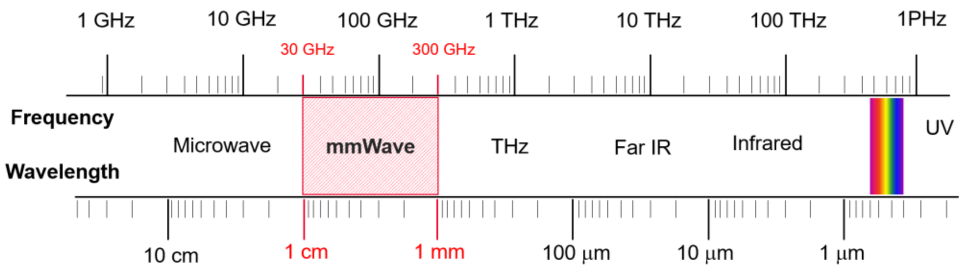
- Current works on 3D occupancy prediction predominantly utilize either **LiDAR point clouds or RGB images**, or a combination of both, overlooking the 4D imaging radar data.



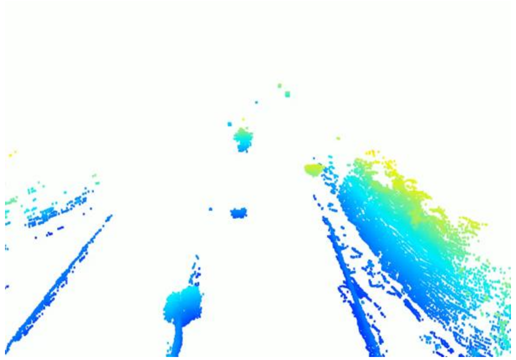
OpenOccupancy (ICCV'23)

Why single-chip mmWave radar

- Robust to adverse weather (e.g., fog, dust, snow) and illumination (e.g., darkness and sun glare)



RGB camera



LiDAR point cloud

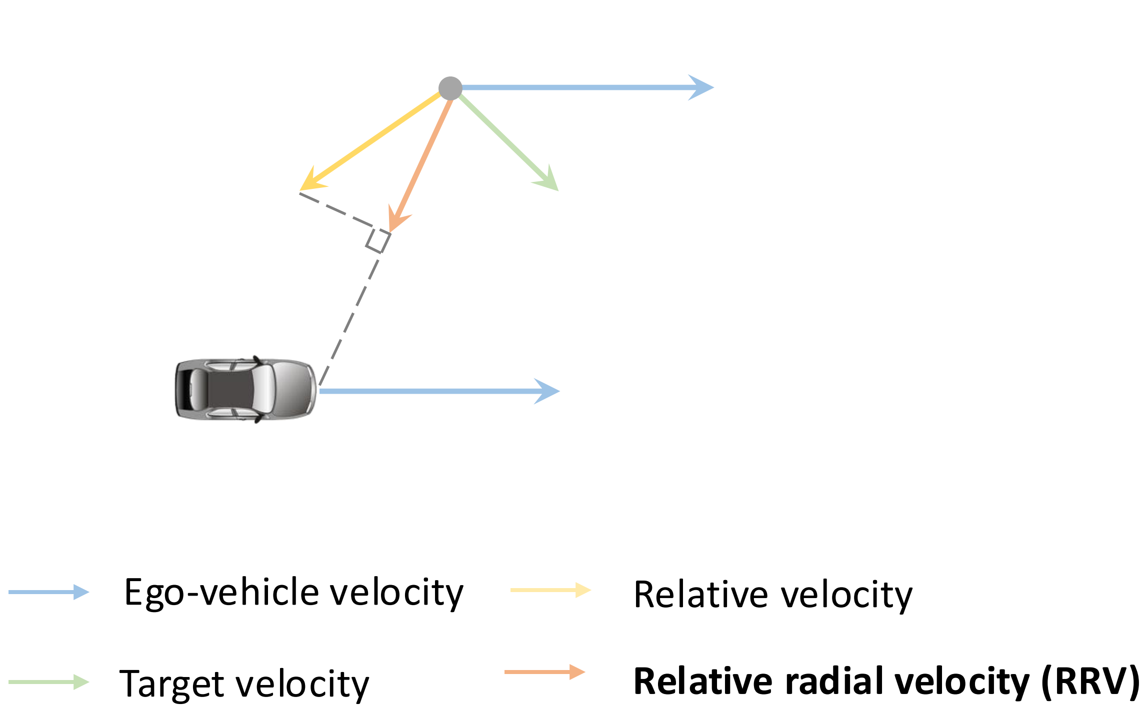
K-RADAR DATASET

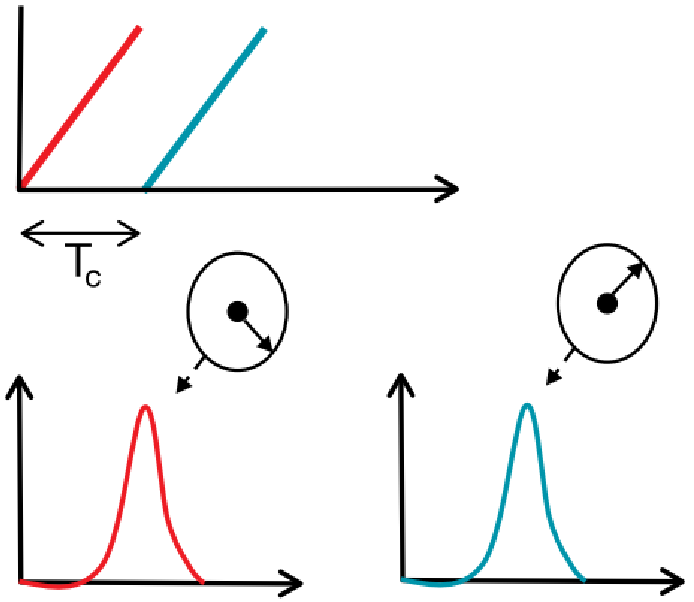
Optical sensors (i.e., camera, LiDAR) can not see through airborne particles.

THE UNIVERSITY of EDINBURGH

Why single-chip mmWave radar

- Radar (Doppler) velocity measurement - relative radial velocity (RRV)

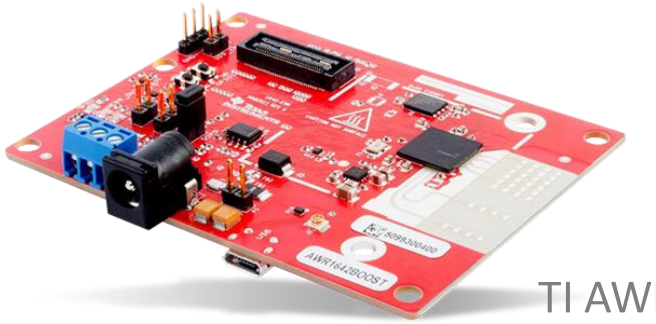




Phase variance across difference chirps contain velocity information

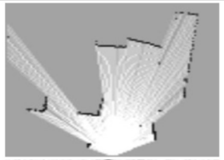
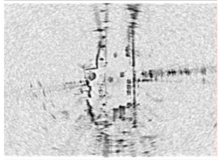
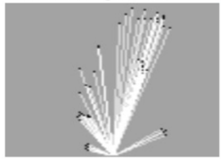
Why single-chip mmWave radar

- Radar-on-a-chip: low cost (vs. LiDAR) and light weight



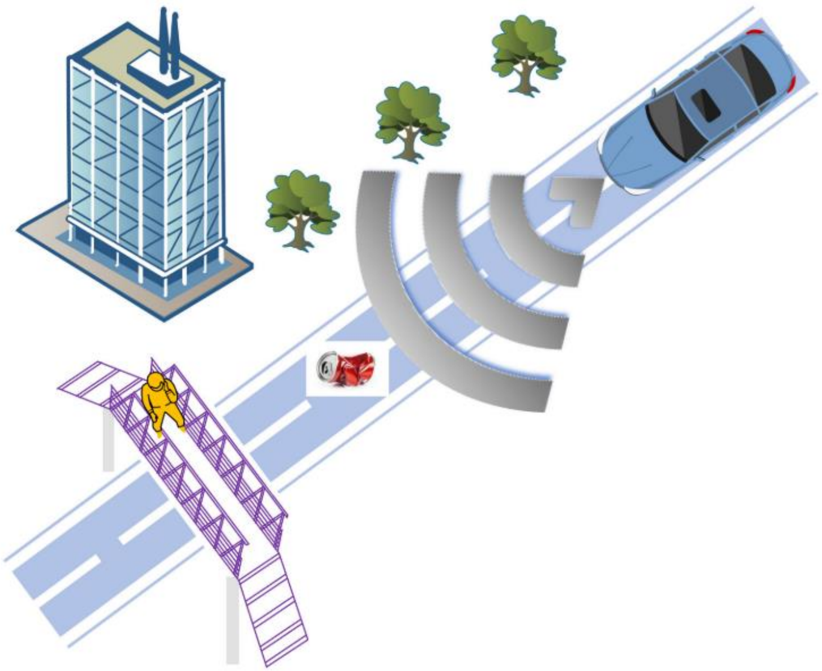
TI AWR1642 RADAR

ARBE 4D RADAR

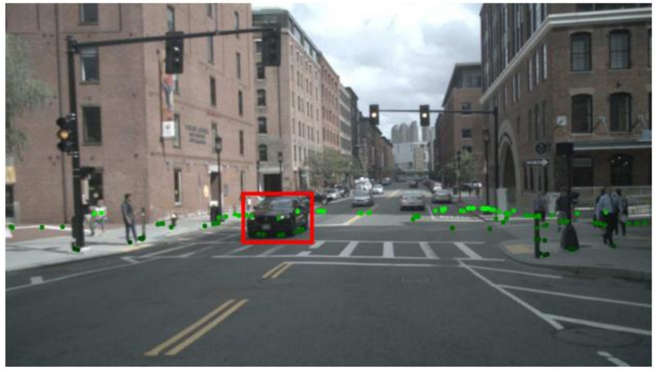
		Cost ($)	Weight (kg)	Power (W)	Scan Points
	Lidar (VLP-16)	8,000	0.83	8	
	Mechnical Radar (CTS-350)	Custo-mized Only	6	24	
	Single-chip Radar (AWR1443)	299	<0.03	2	

- Limited payload or budget

Automotive mmWave radar – toward 4D imaging



- Need to measure elevation information to enable drive-over and drive-under functions



Traditional automotive radar (e.g. nuScenes)

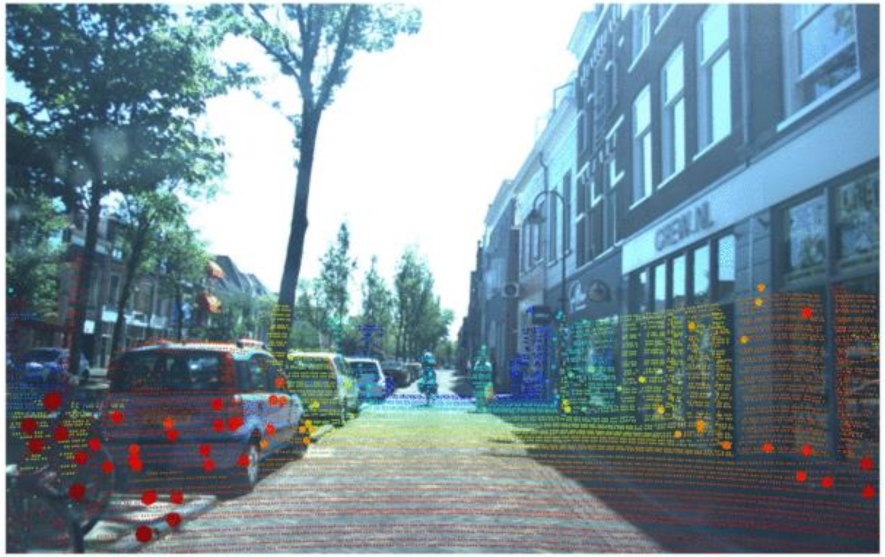
- ✓ MIMO antenna technology
- ✓ **Elevation** measurement
- ✓ Higher angular/range resolution

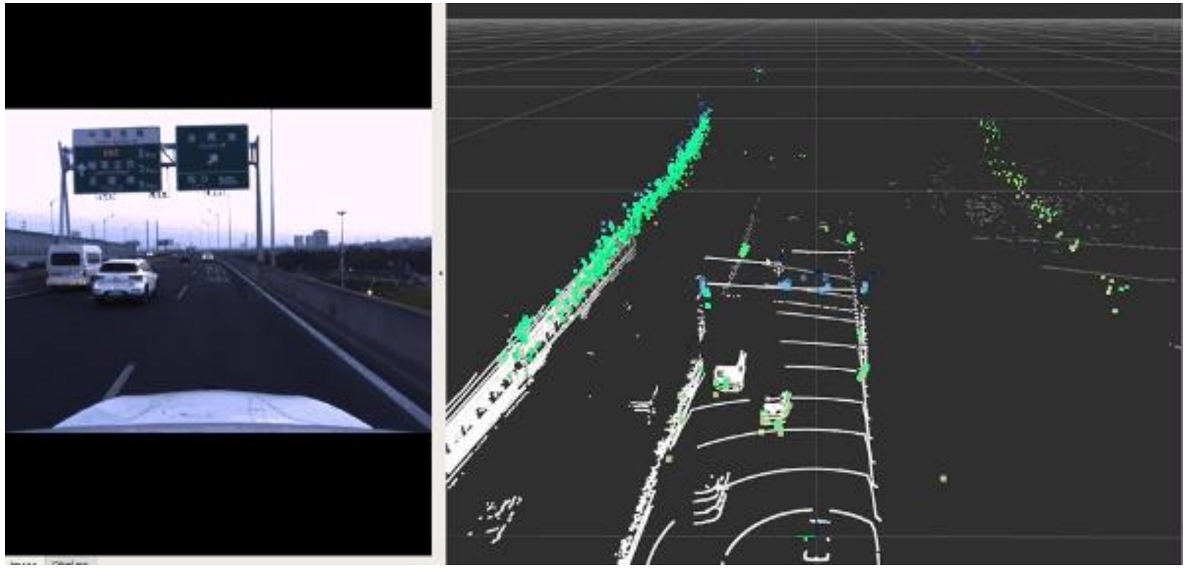
4D imaging automotive radar (e.g. view-of-delft)

THE UNIVERSITY of EDINBURGH

3D occupancy prediction with 4D mmWave radar

- Motivation
  - 'LiDAR-inspired' framework, i.e., relying on **4D radar point clouds**, suffers from the loss of **critical environmental signal** during point cloud generation.
  - For example, the surface of highway, made of low-reflective materials yields **weak signals back,** resulting in very few points being detected.





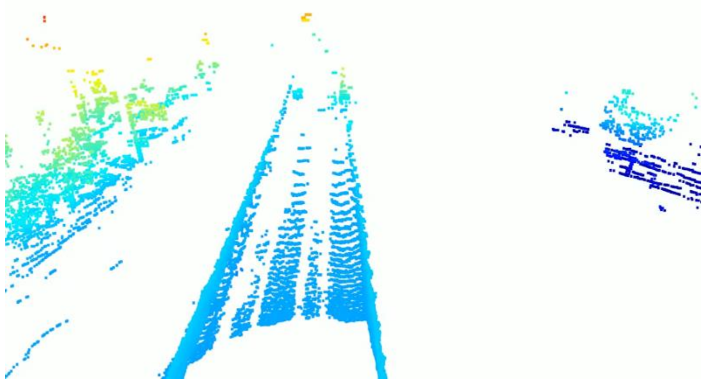
The traditional reliance on sparse radar point clouds, is not optimal for 3D occupancy prediction

Research insights

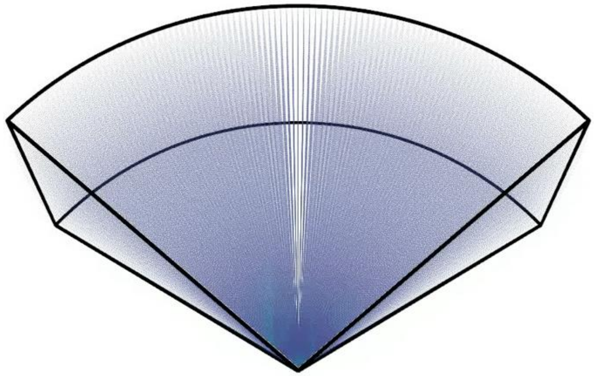
- 4D radar tensor (4DRT), as kind of raw data, **preservers the entirety** of radar measurements. It provides direct 3D measurements in a continuous data format.

- The **volumetric structure** of 4DRTs aligns well with 3D occupancy grids, making them ideally suited for advancing 3D occupancy prediction techniques.

RGB Image



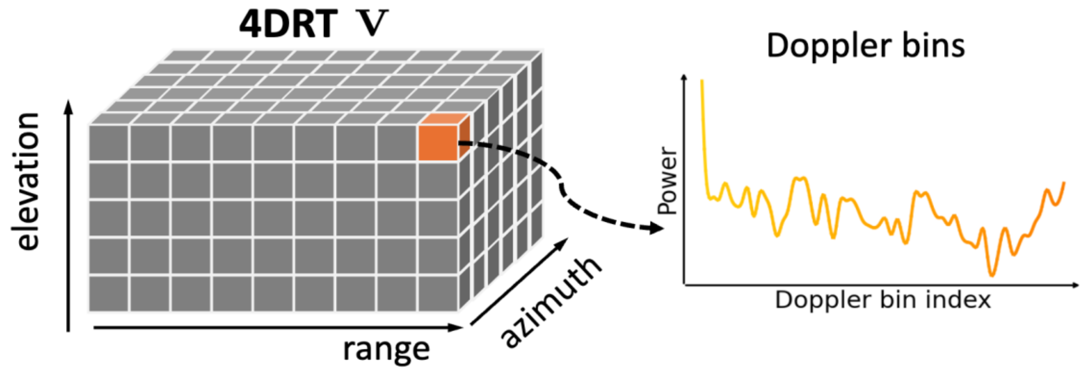
LiDAR Point Cloud

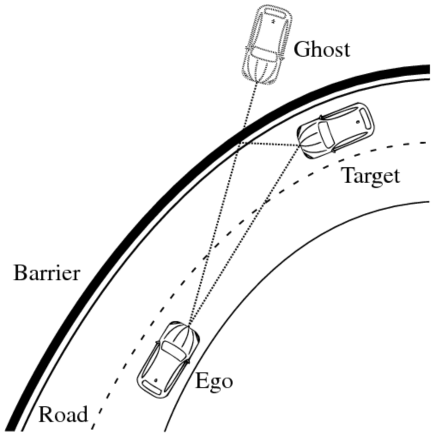


4DRT (reducing Doppler)

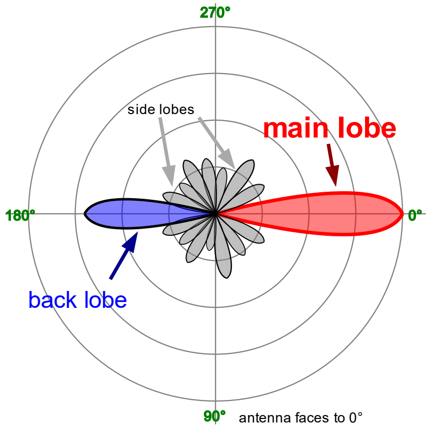
Challenges

- **Substantial size** – up to 500MB per frame, compromise real-time onboard processing
- **Inherently noisy** due to the multi-path effect and sidelobes, threating prediction accuracy
- Stored in **spherical coordinates**, diverges from the preferred 3D Cartesian occupancy grid





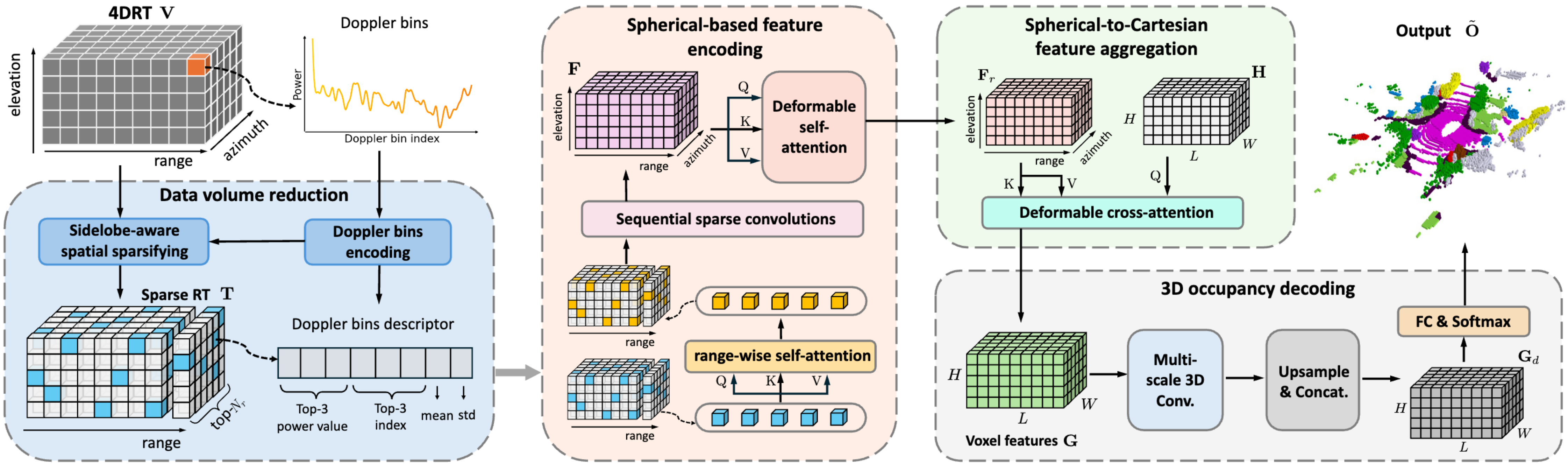
MULTI-PATH EFFECT



Antenna Radiation pattern

THE UNIVERSITY of EDINBURGH

Overall pipeline

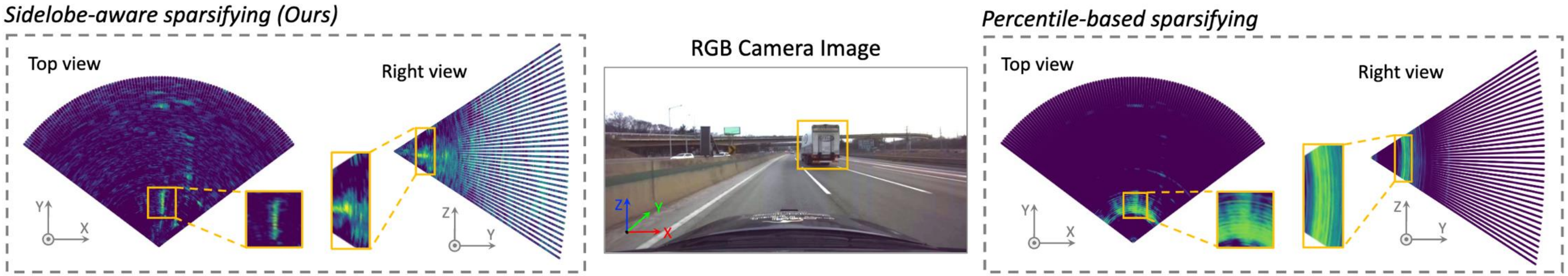


- **Data volume reduction**: reduce the Doppler bins into light-weight, transfer the dense RT into a sparse format
- **Spherical-based feature encoding**: direct encoding of RT features in the spherical coordinates
- **Spherical-to-Cartesian feature aggregation**: learnable voxel queries, aggregate features with deformable attention

THE UNIVERSITY *of* EDINBURGH

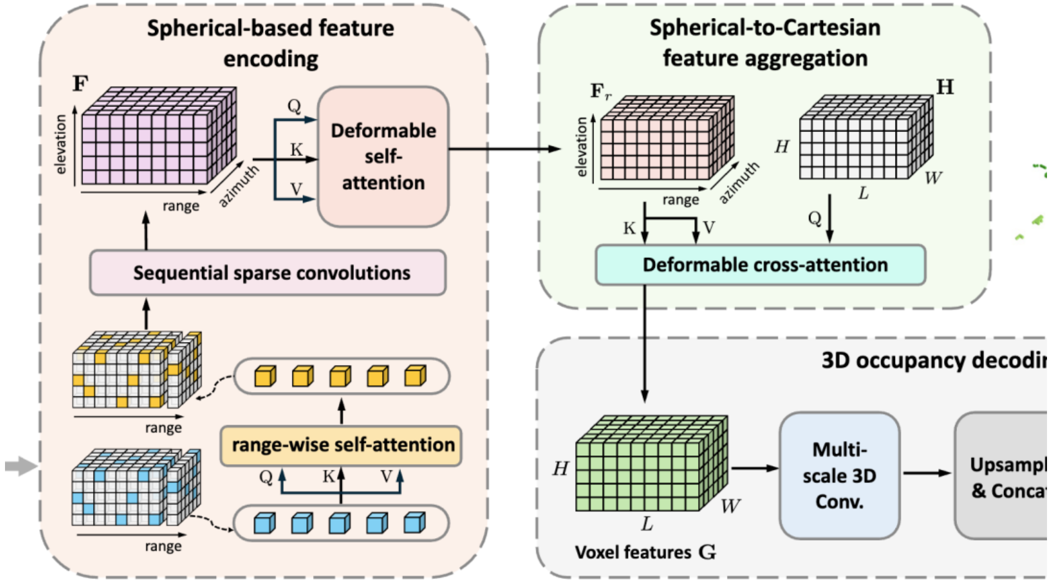
Notable details

- Sidelobe-aware sparsifying: mitigate the concentration of reserved elements at certain ranges



- Interpolation-free transform: from spherical tensor data to Cartesian occupancy prediction

$$\mathrm{DeformAttn}(z, p, \mathbf{X}) = \sum_{m=1}^{M} \mathbf{W}_m \left[\sum_{k=1}^{K} \mathbf{A}_{mk} \cdot \mathbf{W}'_m \mathbf{X}(p + \Delta p_{mk}) \right]$$

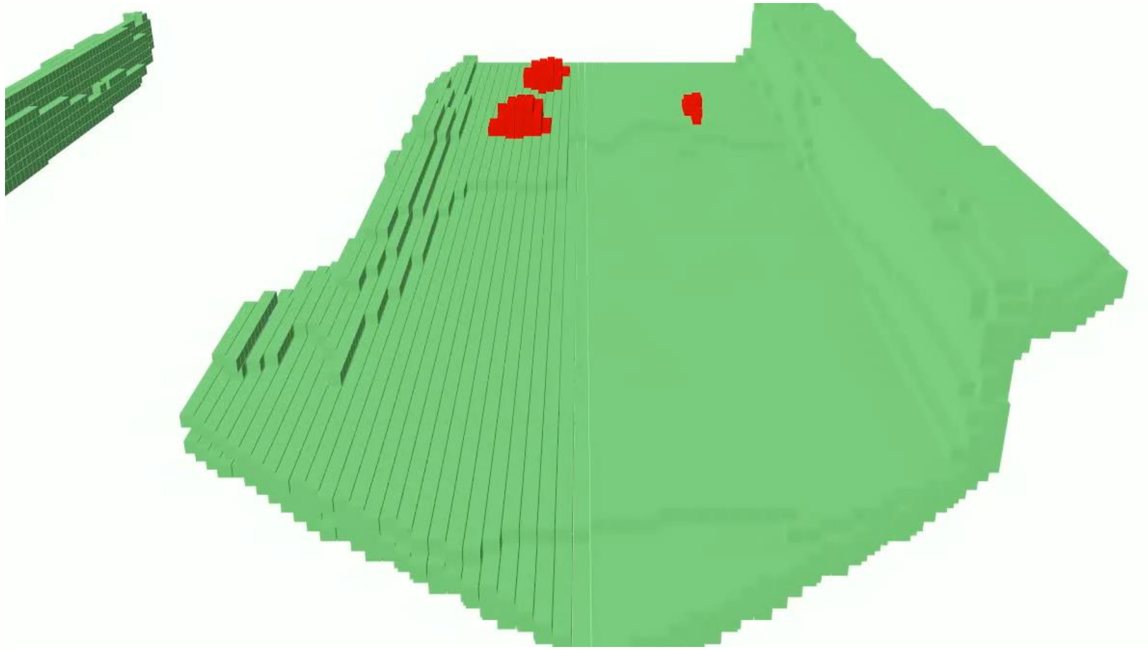


THE UNIVERSITY of EDINBURGH

Demo – 3D occupancy prediction at night

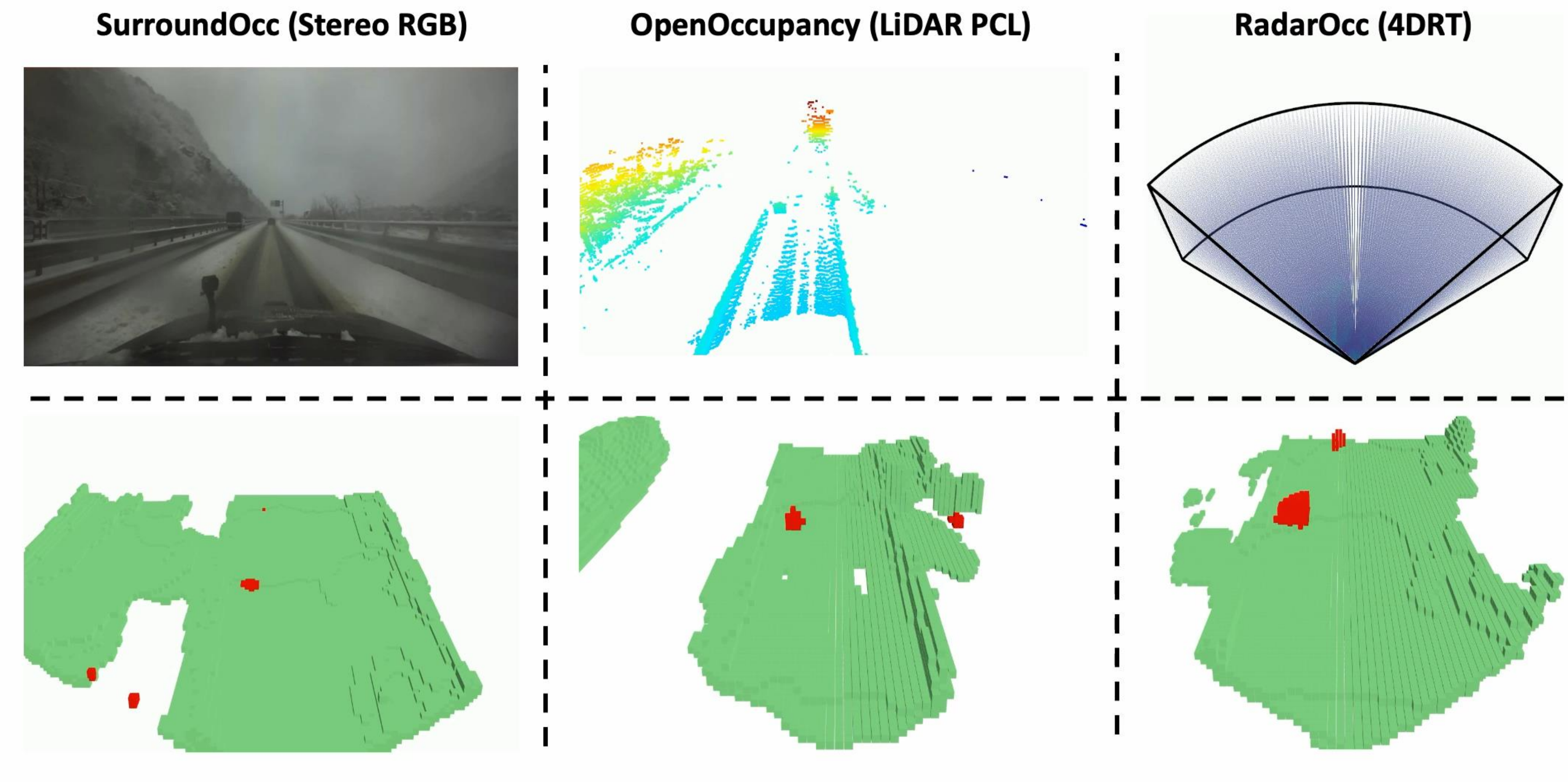
✓ In the right video, green denotes background while red denotes foreground

RGB Camera



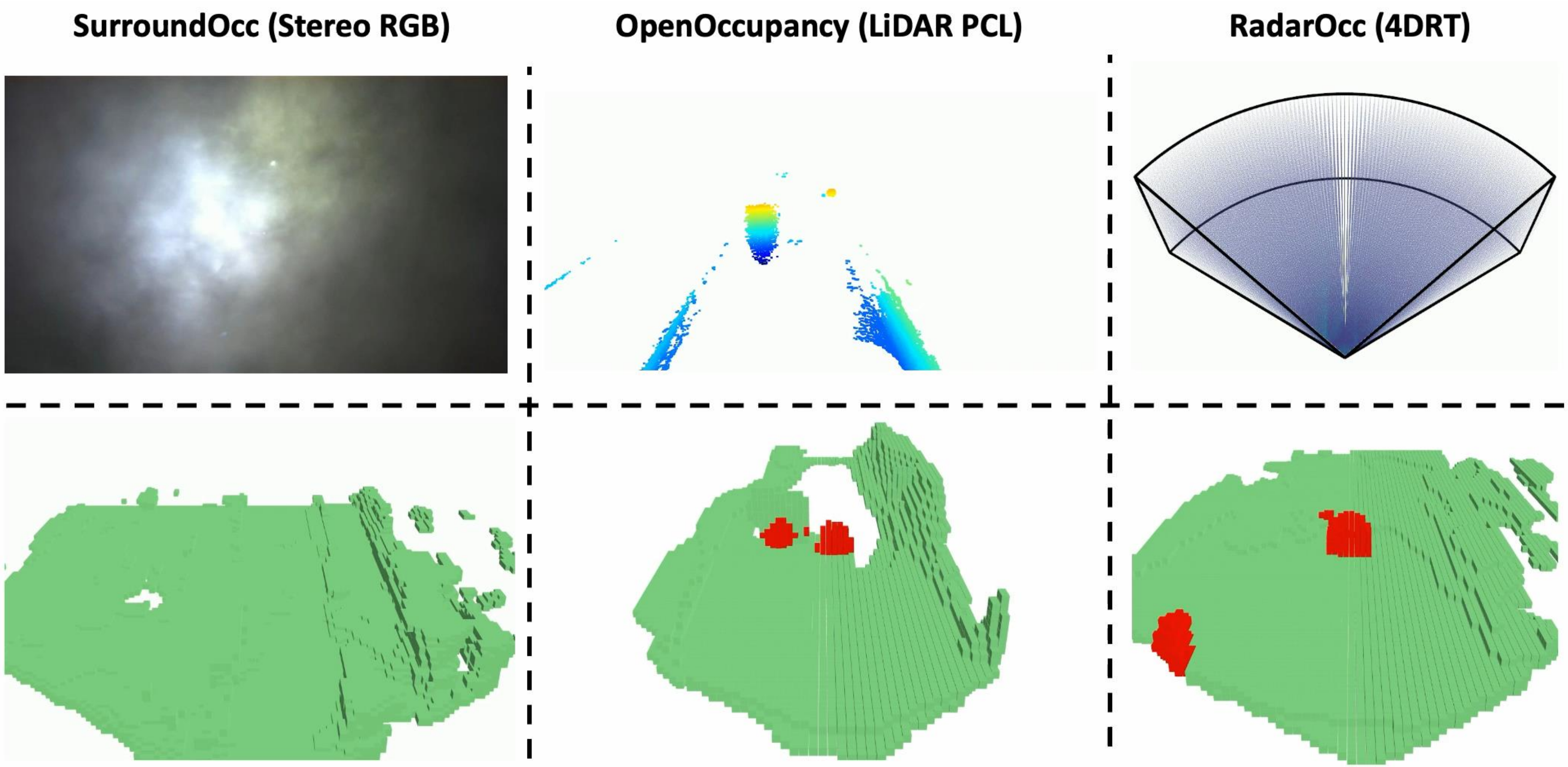
Prediction (RadarOcc)

Demo – 3D occupancy prediction in the snow



THE UNIVERSITY of EDINBURGH

Demo – 3D occupancy prediction in the rain



Thank you!