CODE

High-res 3D Occupancy

nuCraft: Crafting High Resolution 3D Semantic Occupancy for Unified 3D Scene Understanding

Poster Session #1
Board ID #322
Tue 1 Oct 10:30 a.m. CEST — 12:30 p.m. CEST

Project page: https://poodarchu.github.io/publication/eccv2024_nucraft/

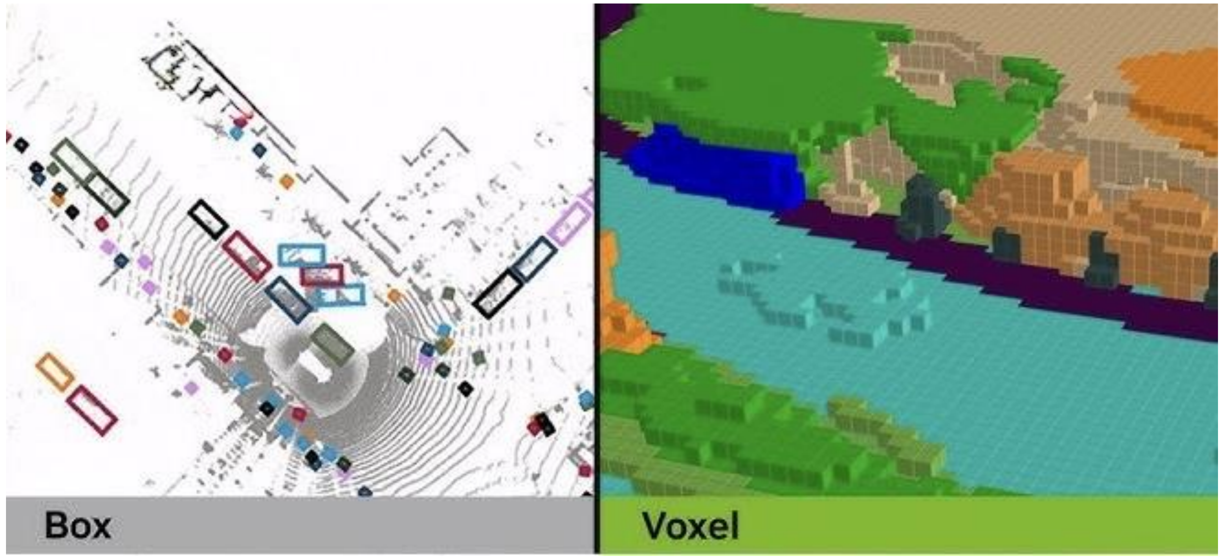
***Benjin ZHU*[1], *Zhe WANG*[2], *Hongsheng LI*[1,3]**

[1]CUHK MMLab [2]SenseTime Research [3]CPII under InnoHK

3D Semantic Occupancy Prediction

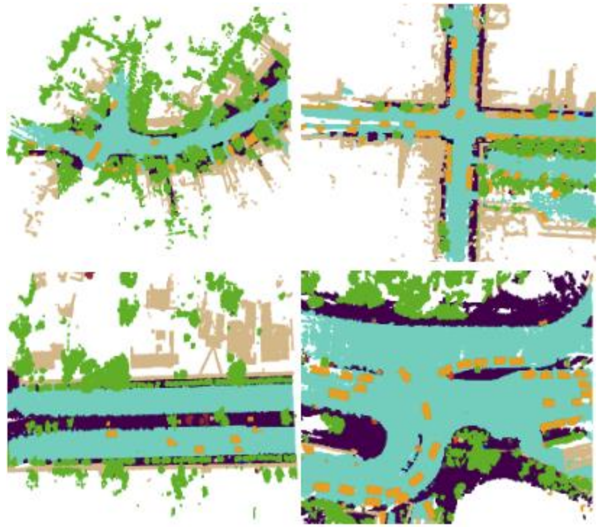
Predicting the complete 3D scene inside a certain fixed volume from camera, LiDAR or other inputs.

- Understand the **3D structure** of a scene
- Determine which spaces are **occupied** or free
- Assign **semantic labels** to occupied areas

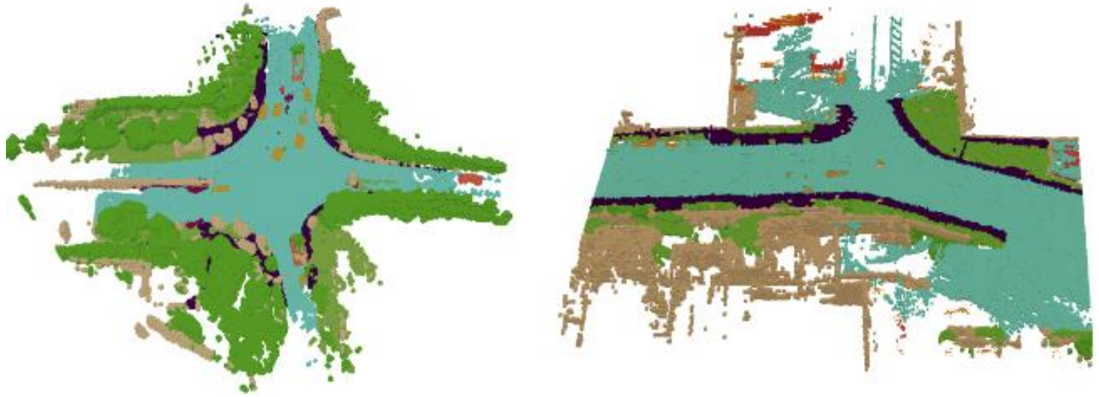


Benchmarks

- Occ3D & its variants



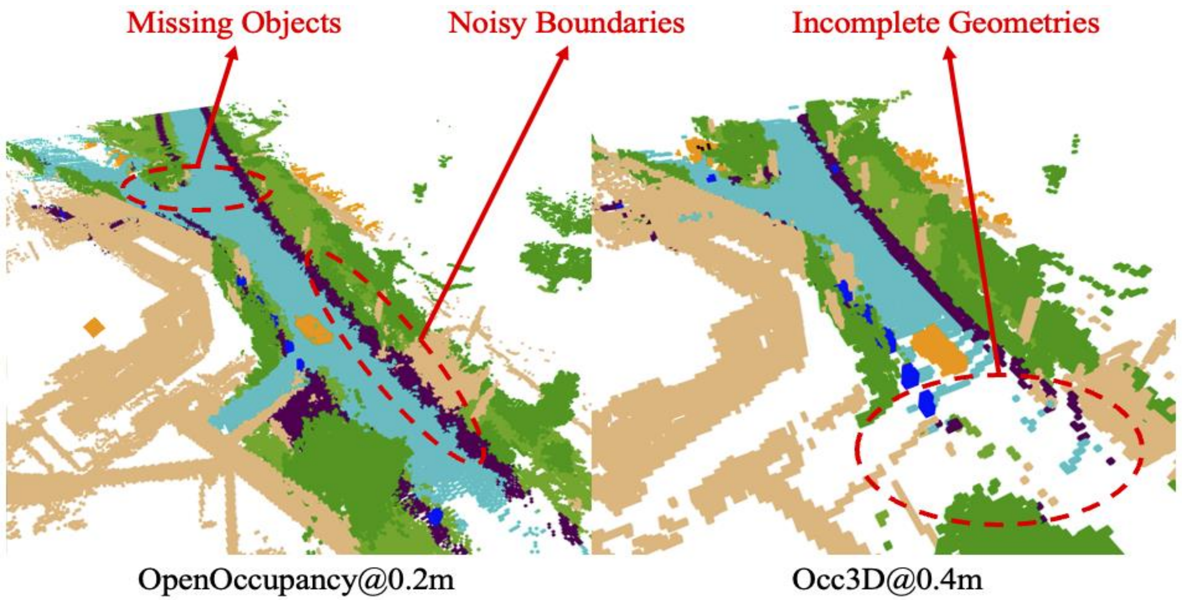
- OpenOccupancy



Limitations

Existing 3D semantic occupancy benchmarks in AD are limited by

- **Low** resolution (up to [512×512×40] with 0.2m voxel size)
- **Inaccurate** annotations inherited from noise and errors in raw data

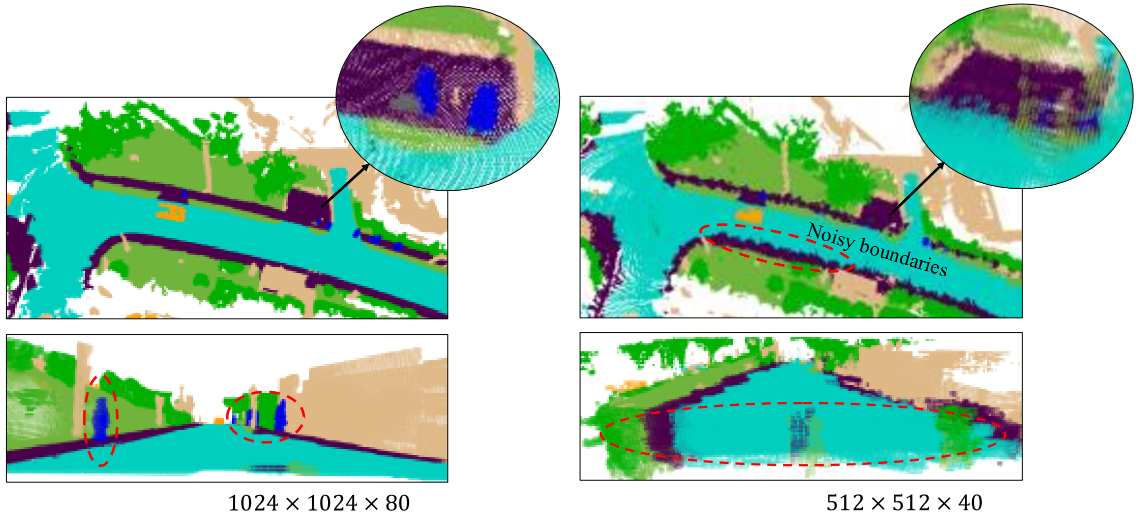


- **Indirect** predictions at **0.4**m or lower, followed by **post upsampling**

The Chinese University of Hong Kong 香港中文大學

Contribution

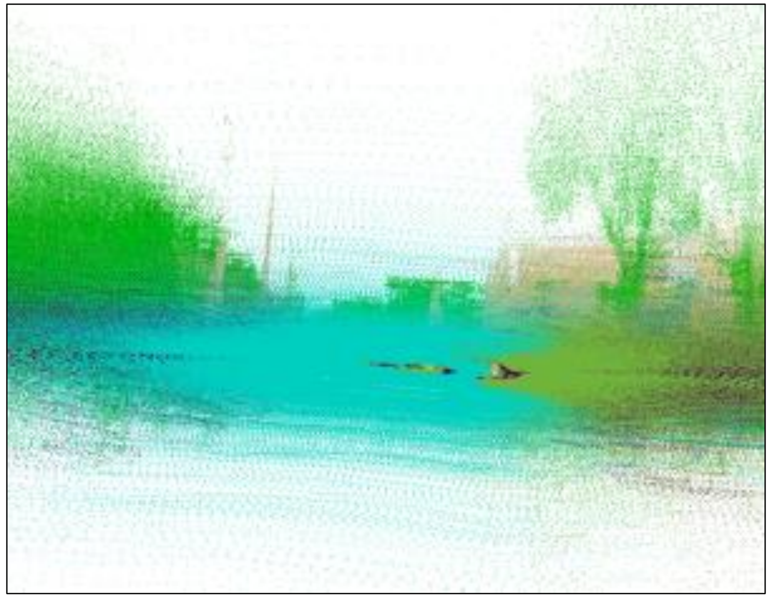
- nuCraft, a **high-resolution** 3D semantic occupancy dataset which offers **8× resolution** and more **precise annotations**
- A **general** and **robust** data generation pipeline for creating high-quality 3D semantic occupancy GT **from noisy data** with no human effort

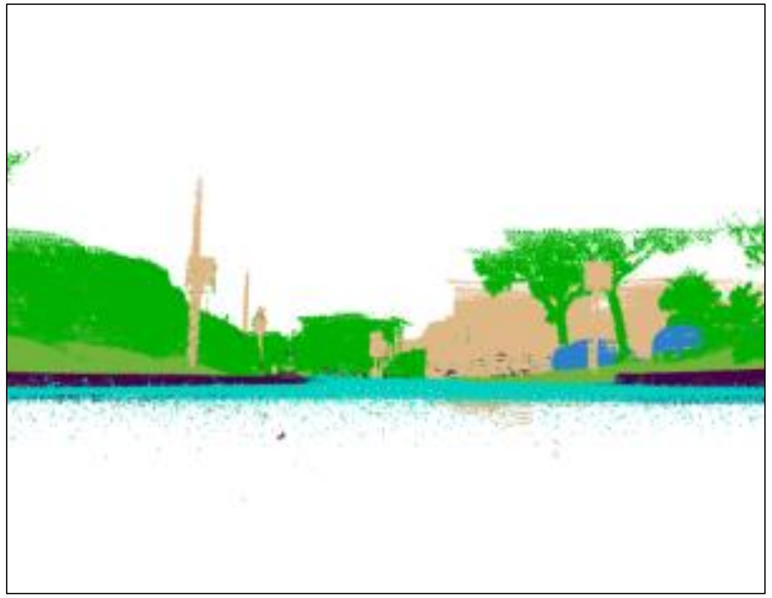


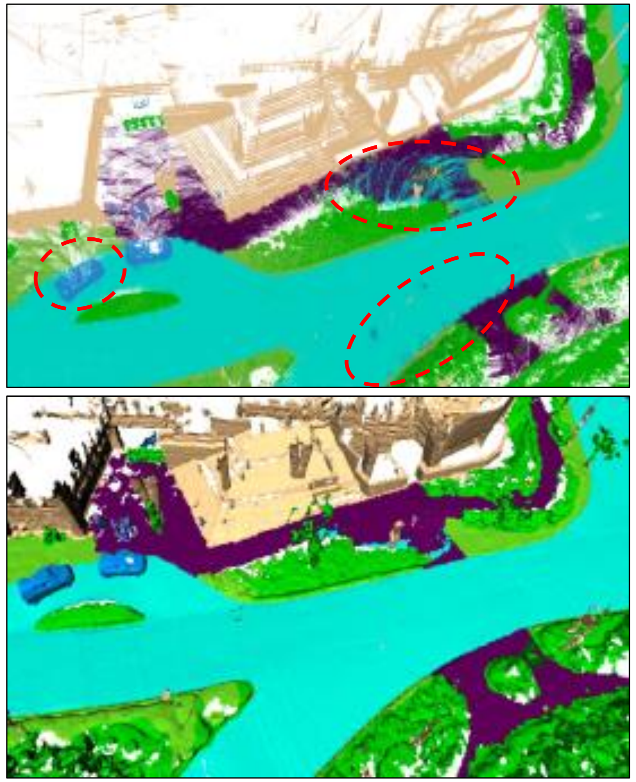
- VQ-Occ, a novel 3D occupancy prediction framework which decouples the encoding of occupancy GT and semantic occupancy prediction, achieving **direct** high-resolution prediction and SOTA performance

Creation of nuCraft

1. Static/Moving Parts **Separation** & Continuous **Scenes Grouping**
2. LiDAR Sequence Aggregation with **Pose Estimation**
3. Voxel Densification with **Semantic Mesh Reconstruction**
4. **Occupancy GT Generation** with necessary post-processing







Dataset Quality Check

- **Foreground**: 3D Object Detection & **Background**: Drivable Area Segmentation

Table 1: Comparison of 3D object detection performance.

Method	mAP
OpenOccupancy	58.4
nuCraft@0.2	62.7
nuCraft@0.1	77.9

Table 2: Comparison of drivable area segmentation performance.

Method	IoU	mIoU
OpenOccupancy	71.2	64.1
nuCraft@0.2	78.3	69.8
nuCraft@0.1	84.2	75.7

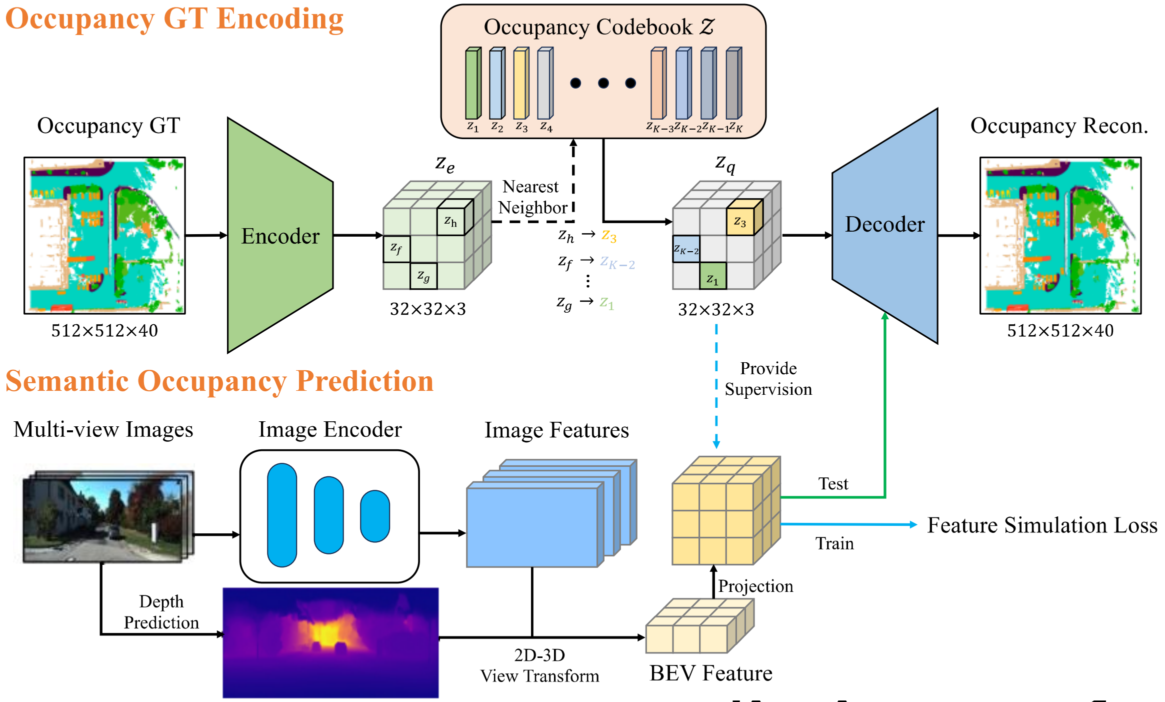
- **Ablation**

Table 3: Component of data generation. The first row denotes the default nuCraft GT at 0.2m resolution. Components are removed one-by-one from the first-row.

Component Removed	mAP	mIoU
None (Full nuCraft@0.2)	**62.7**	**69.8**
S/MPS	62.1	67.2
CSG	62.4	66.5
PE	60.6	66.1
Mesh	58.7	63.2
Clean	62.7	69.7

VQ-Occ: Vector Quantized Occupancy Prediction

- **Occupancy GT Encoding**: VQ-VAE efficiently encodes high-res occupancy GT into a compact latent space, transforming 3D occ prediction into **VQ feature simulation.**



- VQ-Occ achieves better occupancy prediction performance, and achieves direct prediction of 3D semantic occupancy at high resolution (0.2 or finer) without post-upsampling

Results

Table 4: 3D Semantic occupancy prediction results on OpenOccupancy *val* set. VQ-Occ achieves better performance than all previous methods from all input modalities.

Method	Input	IoU	mIoU	barrier	bicycle	bus	car	constr. veh.	motorcycle	pedestrian	traffic cone	trailer	truck	driveable	vegetation
MonoScene [3]	C	18.4	6.9	7.1	3.9	9.3	7.2	5.6	3.0	5.9	4.4	4.9	4.2	14.9	6.3
TPVFormer [16]	C	15.3	7.8	9.3	4.1	11.3	10.1	5.2	4.3	5.9	5.3	6.8	6.5	13.6	9.0
C-CONet [31]	C	20.1	12.8	13.2	8.1	15.4	17.2	6.3	11.2	10.0	8.3	4.7	12.1	31.4	18.8
VQ-Occ (Ours)	C	**21.5**	**13.6**	**14.1**	**8.8**	**16.4**	**18.3**	**6.8**	**11.9**	**10.7**	**8.9**	**5.1**	**12.9**	**33.2**	**20.0**
LMSCNet [24]	L	27.3	11.5	12.4	4.2	12.8	12.1	6.2	4.7	6.2	6.3	8.8	7.2	24.2	12.3
JS3C-Net [36]	L	30.2	12.5	14.2	3.4	13.6	12.0	7.2	4.3	7.3	6.8	9.2	9.1	27.9	15.3
L-CONet [31]	L	30.9	15.8	17.5	5.2	13.3	18.1	7.8	5.4	9.6	5.6	13.2	13.6	34.9	21.5
PointOcc [41]	L	34.1	23.9	24.9	19.0	20.9	25.7	13.4	25.6	30.6	17.9	16.7	21.2	36.5	25.6
VQ-Occ (Ours)	L	**35.3**	**24.8**	**25.8**	**19.8**	**21.8**	**26.7**	**13.9**	**26.6**	**31.7**	**18.7**	**17.4**	**22.1**	**37.8**	**26.5**
M-CONet [31]	C&L	29.5	20.1	23.3	13.3	21.2	24.3	15.3	15.9	18.0	13.3	15.3	20.7	33.2	21.0
VQ-Occ (Ours)	C&L	**36.8**	**25.5**	**26.5**	**20.5**	**22.6**	**27.7**	**14.4**	**27.6**	**32.7**	**19.4**	**18.1**	**22.9**	**39.0**	**27.5**

Table 5: Semantic occupancy prediction results on nuCraft *val* set at 0.2m resolution.

Method	Input	IoU	mIoU	barrier	bicycle	bus	car	constr. veh.	motorcycle	pedestrian	traffic cone	trailer	truck	driveable	vegetation
C-CONet [31]	C	20.8	13.4	14.3	9.1	16.5	18.3	7.4	12.3	11.1	9.4	5.8	13.2	32.5	19.9
VQ-Occ (Ours)	C	**21.9**	**14.1**	**15.2**	**9.9**	**17.5**	**19.4**	**8.0**	**13.0**	**11.8**	**10.0**	**6.2**	**14.0**	**34.3**	**21.1**
L-CONet [31]	L	31.3	16.5	18.6	6.3	14.4	19.2	8.9	6.5	10.7	6.7	14.3	14.7	36.0	22.6
PointOcc [41]	L	34.8	24.6	26.0	20.1	22.0	26.8	14.5	27.1	32.1	19.0	17.8	22.3	37.6	26.7
VQ-Occ (Ours)	L	**36.1**	**25.5**	**26.9**	**20.9**	**22.9**	**27.8**	**15.0**	**28.1**	**33.2**	**19.8**	**18.5**	**23.2**	**38.9**	**27.6**
M-CONet [31]	C+L	29.9	20.7	24.4	14.4	22.3	25.4	16.4	17.0	19.1	14.4	16.4	21.8	34.3	22.1
VQ-Occ (Ours)	C+L	**37.5**	**26.2**	**27.6**	**21.6**	**23.7**	**28.8**	**15.5**	**29.2**	**34.3**	**20.5**	**19.2**	**24.0**	**40.1**	**28.6**