

RayOcc: Occlusion-Aware Ray Occupancy Estimation via Gaussian Mixture Intensity



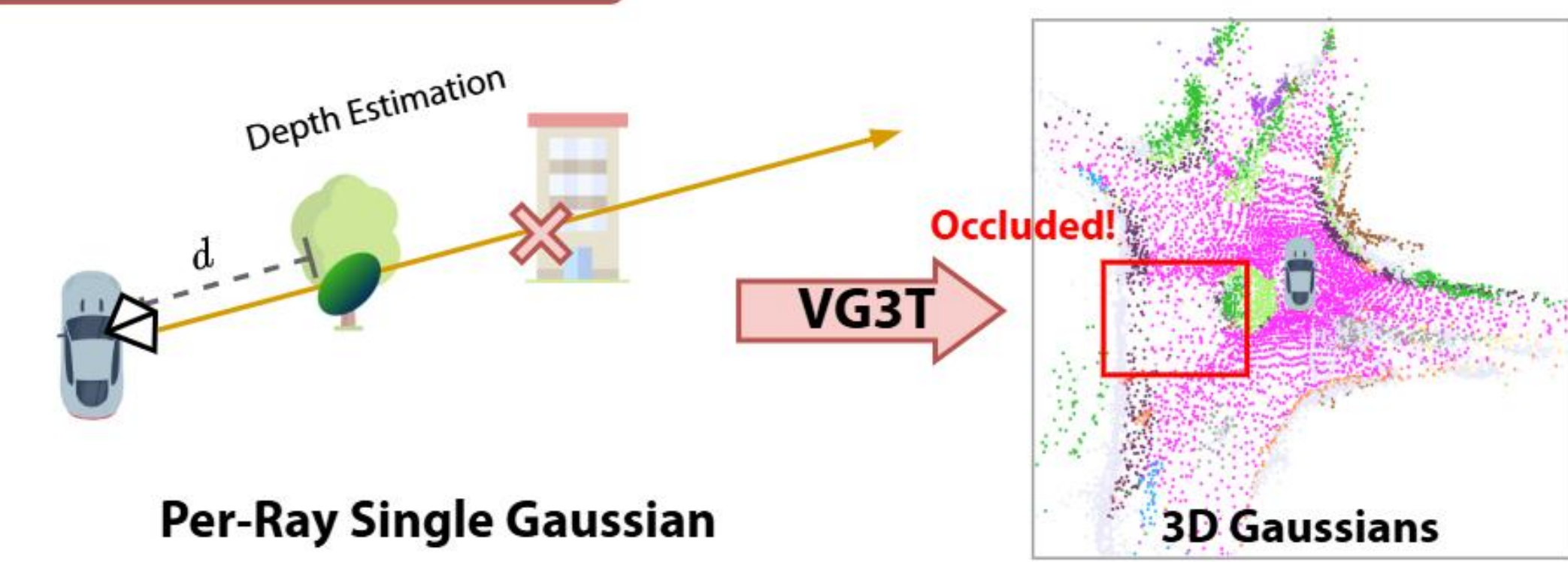
Junho Kim
Kookmin University

Seongwon Lee
Kookmin University

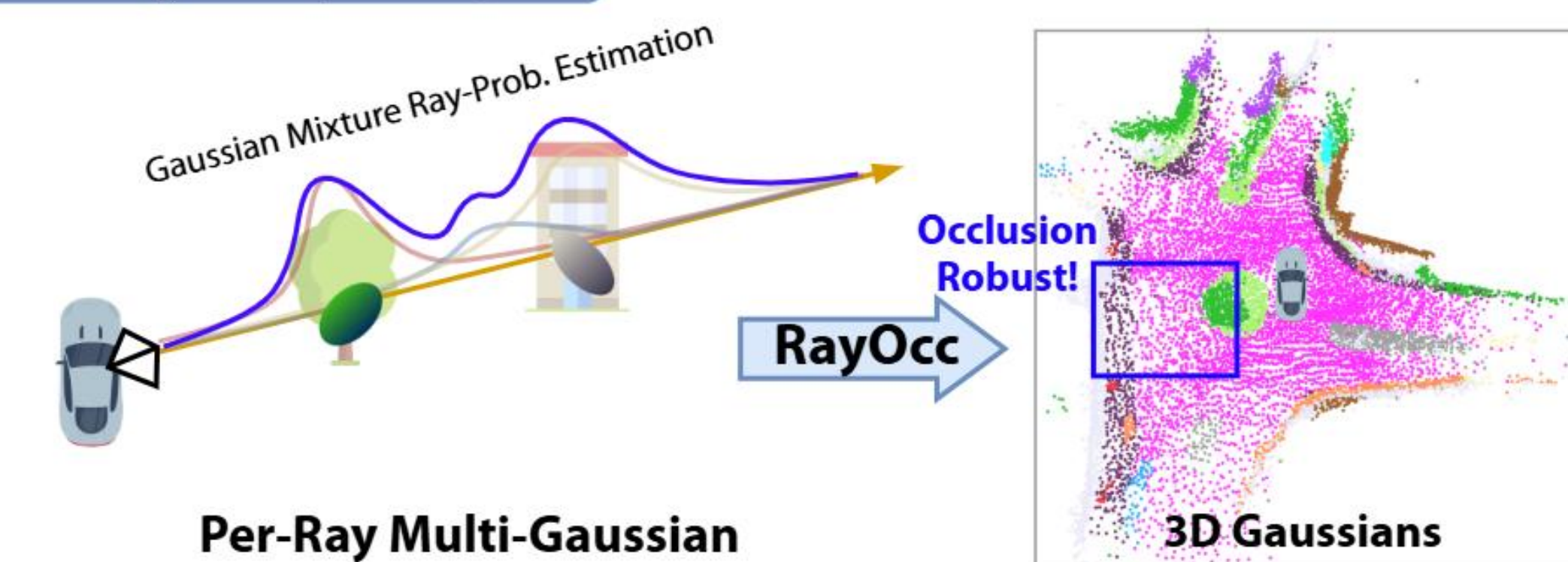


Motivation

Previous Methods



RayOcc (Ours)

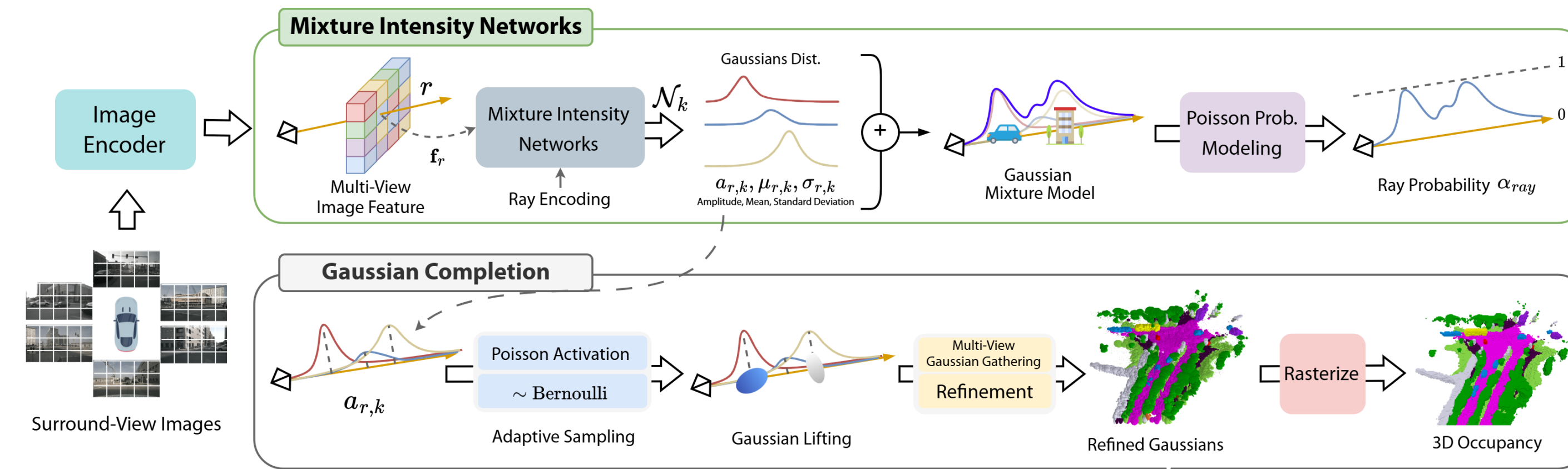


- Existing ray-based methods typically favor a **single dominant depth**, although a camera ray may intersect **multiple occupied surfaces**.
- Softmax-normalization** couples depth hypotheses through **competition**, while **multiple occupied regions should coexist** independently along a ray.

Ideas and Contribution

- Reformulate ray reasoning as **multi-label occupancy existence**, rather than single-depth estimation.
- Predict a **non-normalized Gaussian mixture intensity** and convert it into bin-wise occupancy probabilities via a **Poisson event model**.
- Introduce **adaptive component-wise sampling** to initialize a variable number of 3D Gaussian per ray.
- State-of-the-art nuScenes performance with **34.84 IoU / 22.03 mIoU**.

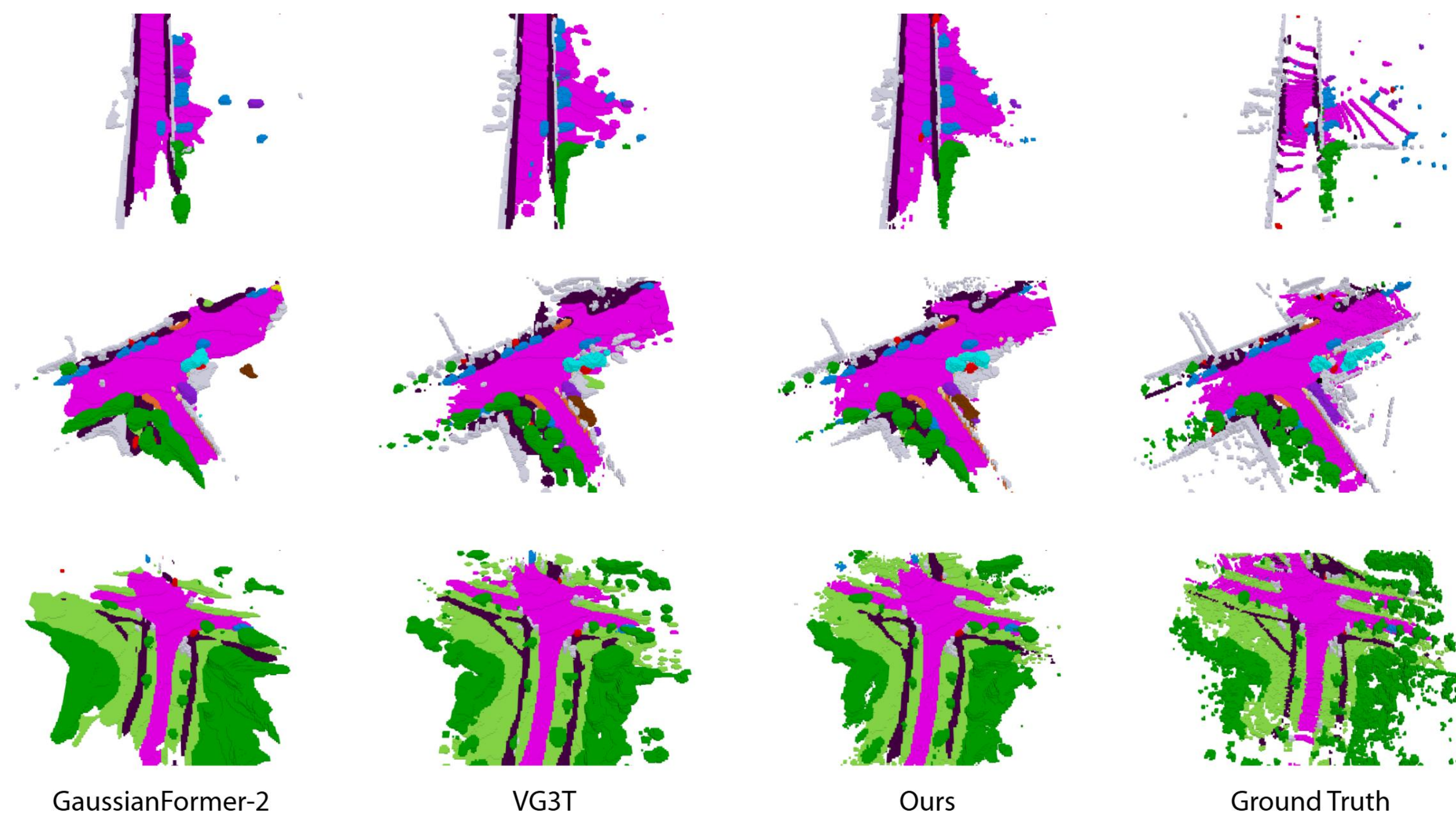
Network Overview



- Multi-view feature with ray geometry** provide a geometry-aware representation for each camera ray.
- Gaussian mixture intensity** is mapped to multi-label occupancy through a **Poisson existence model**.
- Component-wise sampling** initializes multiple Gaussians per ray, followed by refinement and voxel rasterization.

Gaussian Initialization Analysis

- Qualitative Comparison with GF-2 and VG3T



- Adaptive Sampling Ablation

TABLE IV: Ablation Study on RayOcc's Adaptive Sampling

Sampling	mIoU	IoU	Average Number of Gaussians ↓	Position Perc. (%) ↑	Dist. (m) ↓
Bernoulli	21.92	35.08	23136	53.55	0.85
$\tau = 0.5$	22.01	34.90	19822	59.83	0.72
$\tau = 0.7$	22.02	34.87	19681	60.41	0.71
$\tau = 0.8$	22.03	34.84	19251	61.07	0.70
$\tau = 0.9$	22.03	34.76	18679	62.38	0.68

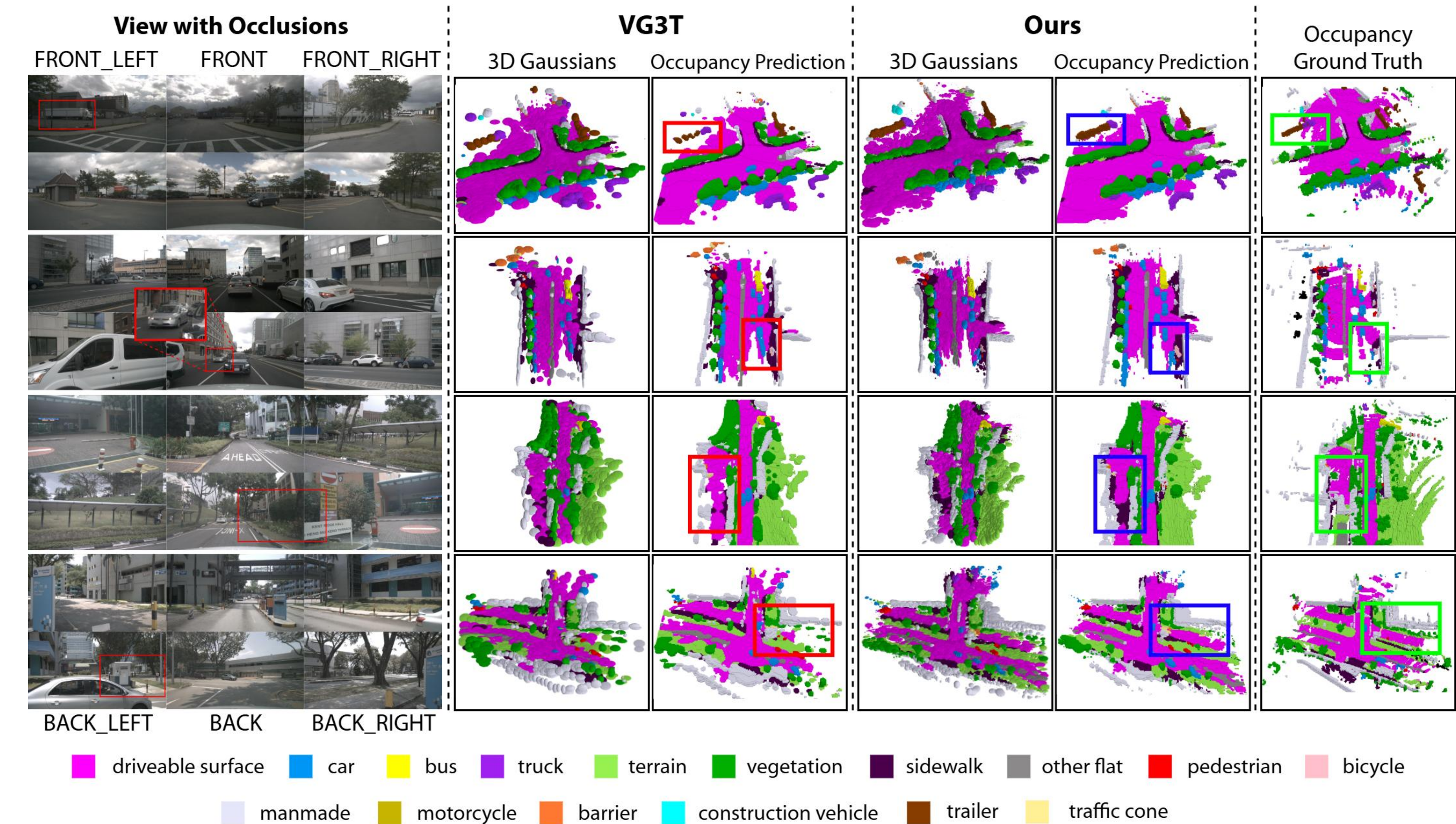
Experimental Results

- Surround-View 3D Semantic Occupancy Prediction

Method	IoU	mIoU	barrier	bicycle	bus	car	const. veh.	motorcycle	pedestrian	traffic cone	trailer	truck	drive. suf.	other flat	sidewalk	terrain	manmade	vegetation
MonoScene [5]	23.96	7.31	4.03	0.35	8.00	8.04	2.90	0.28	1.16	0.67	4.01	4.35	27.72	5.20	15.13	11.29	9.03	14.86
Atlas [42]	28.66	15.00	10.64	5.68	19.66	24.94	8.90	8.84	6.47	3.28	10.42	16.21	34.86	15.46	21.89	20.95	11.21	20.54
BEVFormer [27]	30.50	16.75	14.22	6.58	23.46	28.28	8.66	10.77	6.64	4.05	11.20	17.78	37.28	18.00	22.88	22.17	13.80	22.21
TPVFormer [28]	11.51	11.66	16.14	7.17	22.63	17.13	8.83	11.39	10.46	8.23	9.43	17.02	8.07	13.64	13.85	10.34	4.90	7.37
TPVFormer* [28]	30.86	17.10	15.96	5.31	23.86	27.32	9.79	8.74	7.09	5.20	10.97	19.22	38.87	21.25	24.26	23.15	11.73	20.81
OccFormer [12]	31.39	19.03	18.65	10.41	23.92	<u>30.29</u>	10.31	14.19	13.59	10.13	12.49	20.77	38.78	19.79	24.19	22.21	13.48	21.35
SurroundOcc [11]	31.49	20.30	20.59	11.68	28.06	30.86	10.70	15.14	14.09	12.06	14.38	22.26	37.29	23.70	24.49	22.77	14.89	21.86
GaussianFormer [33]	29.83	19.10	19.52	11.26	26.11	29.78	10.47	13.83	12.58	8.67	12.74	21.57	39.63	23.28	24.46	22.99	9.59	19.12
GaussianFormer-2 [23]	30.56	20.02	20.15	12.99	27.61	30.23	11.19	<u>15.31</u>	12.64	9.63	13.31	22.26	39.68	23.47	25.62	23.20	12.25	20.73
VG3T [34]	<u>34.06</u>	<u>21.74</u>	19.95	13.63	28.69	29.52	<u>12.69</u>	16.02	<u>13.77</u>	<u>10.64</u>	15.75	<u>23.02</u>	<u>41.74</u>	<u>26.26</u>	<u>27.52</u>	<u>26.44</u>	16.51	25.75
Ours	34.84	22.03	<u>20.31</u>	<u>13.32</u>	29.98	29.79	15.31	14.72	13.25	10.55	<u>15.47</u>	23.53	42.67	27.02	28.51	26.96	<u>15.69</u>	<u>25.44</u>

* means supervised by dense occupancy annotations as opposed to the original LiDAR segmentation labels. The best and second-best performances are represented by **bold** and underline respectively.

- Qualitative Results of Our Method and VG3T



RayOcc better preserves partially occluded objects and reduces fragmented prediction by maintaining multiple depth hypotheses along each ray.

- Performance And Efficiency Comparison

TABLE II: Gaussian Sampling Methods Comparison.

Method	Gaussian Sampling	Position Perc. (%) ↑	Dist. (m) ↓
GaussianFormer [33]	Random	16.41	3.07
GaussianFormer-2 [23]	Categorical	28.85	1.24
VG3T [34]	Depth	51.22	0.97
RayOcc (ours)	Component-wise	61.07	0.70

TABLE III: Efficiency Comparison with Gaussian Representation Methods.

Method	Number of Gaussians ↓	Latency (ms) ↓	Memory (MB) ↓	mIoU	IoU
GaussianFormer [33]	144000	372	6229	19.10	29.83
GaussianFormer-2 [23]	25600	513	3063	20.02	30.56
VG3T [34]	13661*	223	1975	<u>21.74</u>	<u>34.06</u>
RayOcc (ours)	<u>19251*</u>	<u>243</u>	<u>2620</u>	22.03	34.84

* denotes average number of Gaussians per scene.