Progressive Gaussian Transformer with Anisotropy-aware Sampling for Open-vocabulary Occupancy Prediction

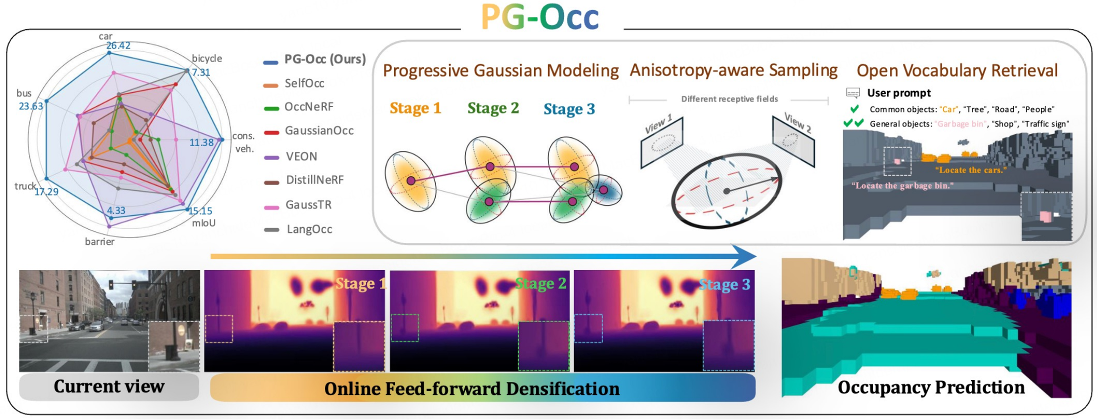
香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Chi Yan[1,2] Dan Xu[1*]

[1]The Hong Kong University of Science and Technology [2]ZEEKR Automobile R&D Co., Ltd

Scan for code & demo

Overview



TL; DR: Progressive densification makes sparse Gaussians effective for 3D occupancy prediction task.

Key Contribution

- **Progressive Gaussian densification for detailed scene modeling**
- **Improves feature learning with anisotropy-aware sampling**
- **Achieves SOTA performance (+14.3% mIoU)**

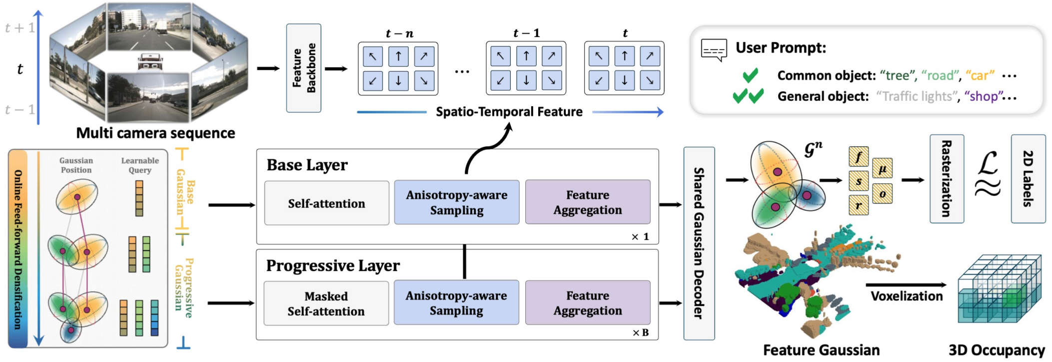
Quantitative Performance

Achieves state-of-the-art performance on Occ3D-nuScenes, especially on small and geometrically complex objects.

Method	Mod.	mIoU	barrier	bicycle	bus	car	cons. veh.	motorcycle	pedestrian	traffic cone	trailer	truck	drive. surf.	sidewalk	terrain	manmade	vegetation
SelfOcc (Huang et al., 2024a)	C	10.54	0.15	0.66	5.46	12.54	0.00	0.80	2.10	0.00	0.00	8.25	55.49	26.30	26.54	14.22	5.60
OccNeRF (Zhang et al., 2023a)	C	10.81	0.83	0.82	5.13	12.49	3.50	0.23	3.10	1.84	0.52	3.90	52.62	20.81	24.75	18.45	13.19
GaussianOcc (Gan et al., 2024)	C	11.26	1.79	5.82	14.58	13.55	1.30	2.82	7.95	9.76	0.56	9.61	44.59	20.10	17.58	8.61	10.29
GaussianFlowOcc (Boeder et al., 2025)	C	14.07	6.27	8.54	13.36	12.38	4.92	10.05	6.84	8.75	1.12	10.43	54.40	26.44	28.89	10.39	9.33
VEON (Zheng et al., 2024a)	C+L+T	13.95	4.80	2.70	14.70	10.90	11.00	3.80	4.70	4.00	5.30	9.60	46.50	21.10	22.10	24.80	23.70
DistillNeRF (Wang et al., 2024)	C+L+T	10.05	1.35	2.08	10.21	10.09	2.56	1.98	5.54	4.62	1.43	7.90	43.02	16.86	15.02	14.06	15.06
LangOcc (Boeder et al., 2024)	C+T	12.04	2.70	7.20	5.80	13.90	0.50	**10.8**	**6.40**	**8.70**	**3.20**	11.00	**42.10**	12.50	**27.20**	14.10	14.50
GaussTR (Jiang et al., 2024)	C+T	13.25	2.09	5.22	14.07	20.43	5.70	7.08	5.12	3.93	0.92	13.36	39.44	15.68	22.89	**21.17**	21.87
PG-Occ (Ours)	C+T	**15.15**	**4.33**	**7.31**	**23.63**	**26.42**	**11.38**	6.33	2.74	5.79	3.07	**17.29**	37.81	**19.29**	20.85	19.02	**21.92**

Architecture

Start with sparse Gaussians → progressively densify under-represented regions → refine features via anisotropic sampling



a) Gaussian Representation.

$$G_i = (\mu_i, s_i, r_i, \sigma_i, f_i)$$

Each Gaussian encodes geometry and text-aligned feature.

b) Progressive Online Densification.

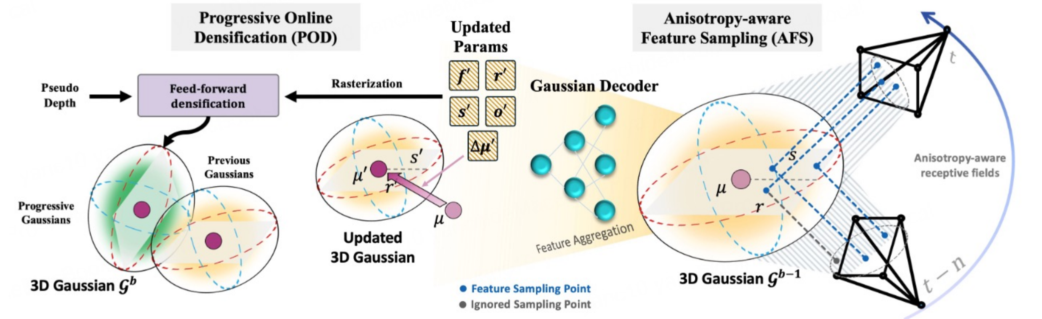
$$(\mu^b, q^b) = (\mu^{b-1}, q^{b-1}) \oplus (\mu^b_{add}, q^b_{add})$$

We iteratively refine the scene by adding Gaussians in under-represented regions.

c) Anisotropy-aware Sampling.

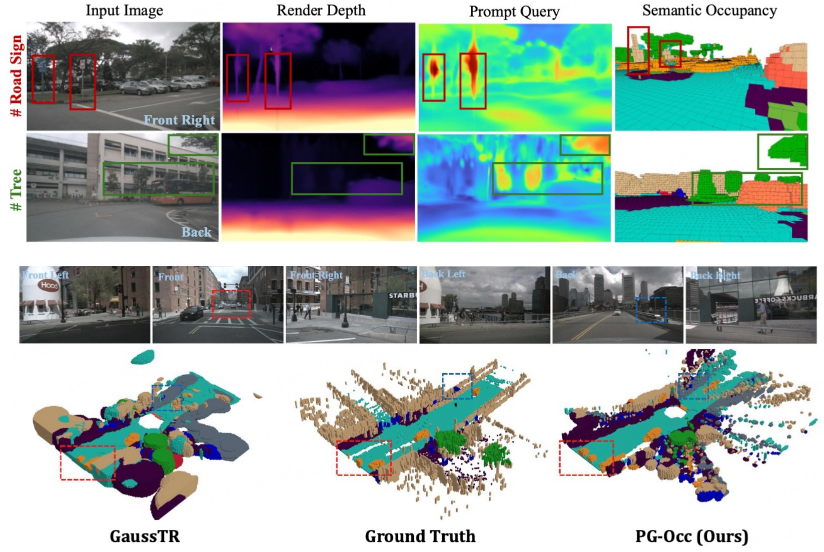
$$\mu_{i,j} = \mu_i + R(r_i)(s_i \odot \delta_{i,j})$$

Sampling within Gaussian ellipsoids improves feature aggregation.



Qualitative Performance

Accurate geometry and open-vocabulary semantics across diverse scenes



Zero-shot generalization

Generalizes across datasets without retraining.