ODG: Occupancy Prediction Using Dual Gaussians

Yunxiao Shi, Yinhao Zhu, Shizhong Han, Jisoo Jeong, Amin Ansari, Hong Cai, Fatih Porikli

Qualcomm AI Research*

Qualcomm AI research

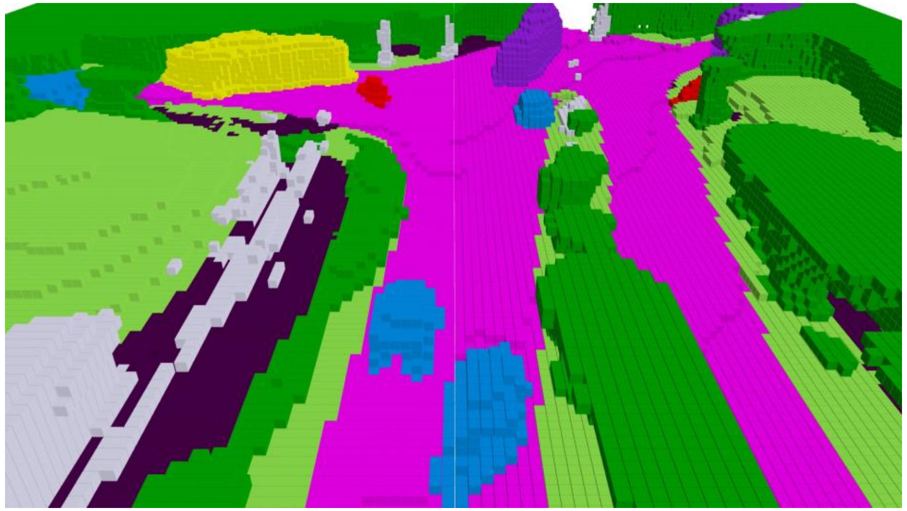
NEURAL INFORMATION PROCESSING SYSTEMS

*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

3D Occupancy Prediction

Goal is to infer 3D geometry + semantics from camera images

Multi-camera images

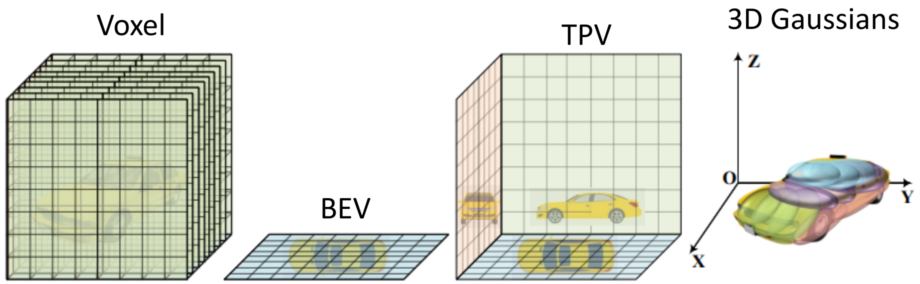


3D occupancy

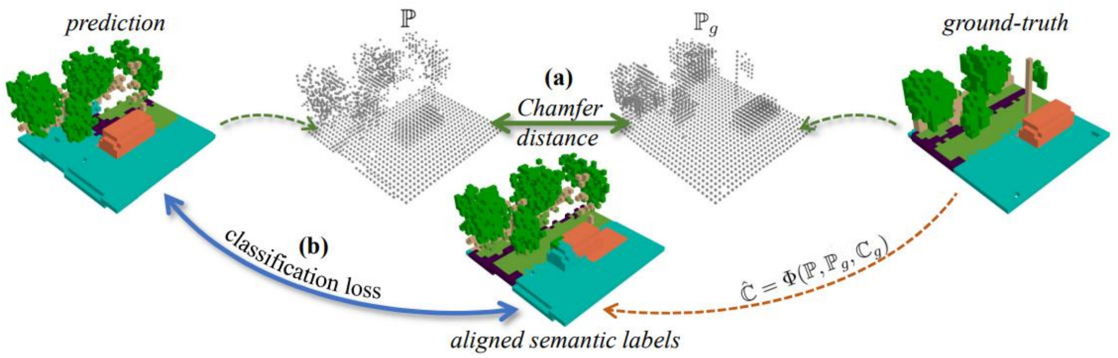
*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

3D Occupancy Prediction

Scene representation is a central problem



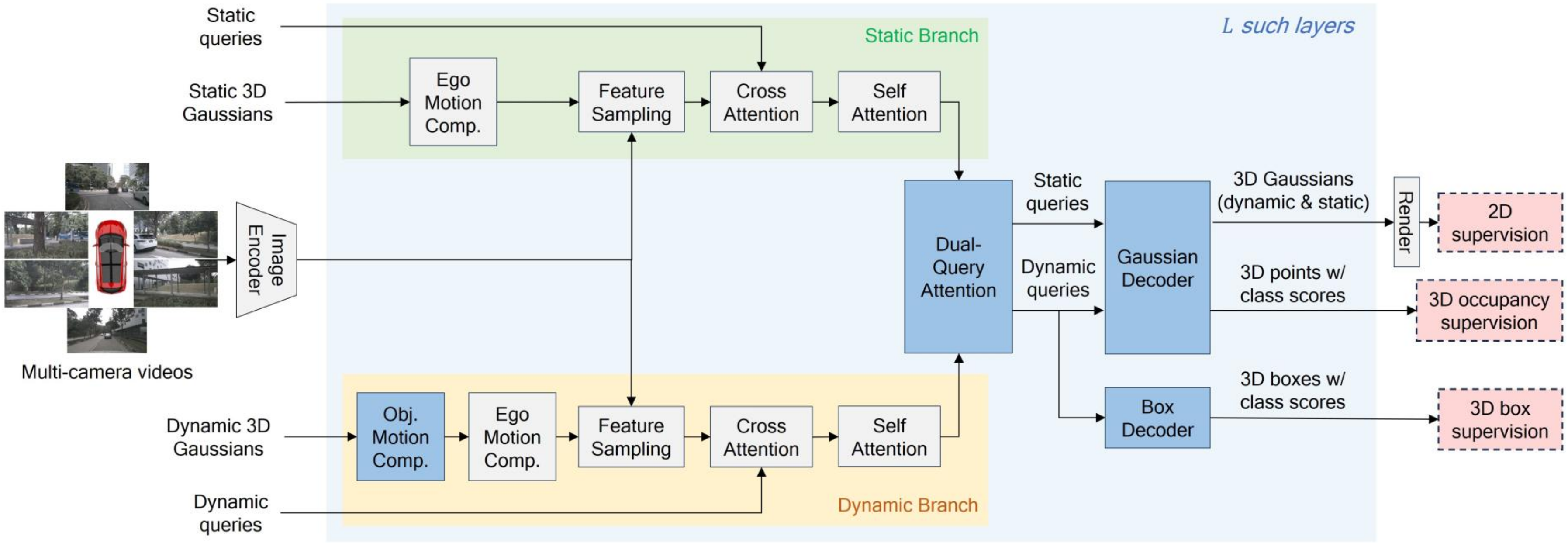
Dense grids



Point sets

Our Method: ODG

- Defines dual Gaussian queries to improve prediction of dynamic scenes
- Coarse-to-fine refinements of Gaussian properties
- Multi-stage rendering supervision leveraging 3D Gaussian splatting



*Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Evaluation

State-of-the-Art results on Occ3D-nuScenes and Occ3D-Waymo

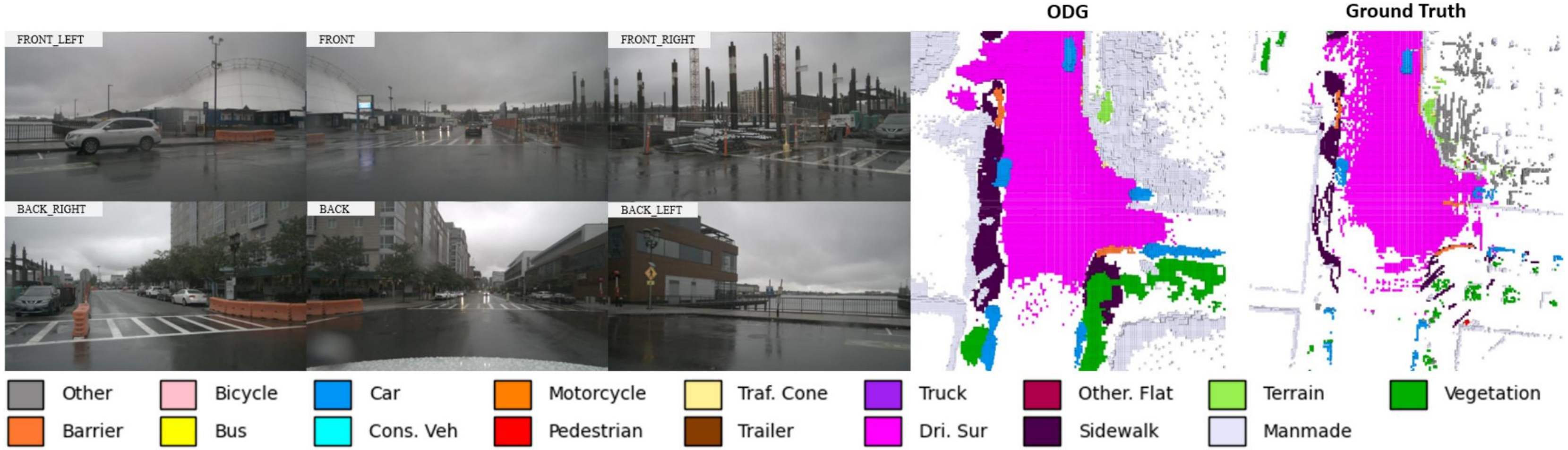
Method	mIoU	Others	Barrier	Bicycle	Bus	Car	Cons. Veh	Motorcycle	Pedestrian	Traffic Cone	Trailer	Truck	Dri. Sur	Other flat	Sidewalk	Terrain	Manmade	Vegetation	RayIoU	FPS
RenderOcc [41]	26.11	4.84	31.72	10.72	27.67	26.45	13.87	18.2	17.67	17.84	21.19	23.25	63.2	36.42	46.21	44.26	19.58	20.72	19.5	3.0
GaussRender [10]	30.38	8.87	40.98	23.25	43.76	46.37	19.49	25.2	**23.96**	19.08	25.56	33.65	58.37	33.28	36.41	33.21	22.76	22.19	37.5	-
GaussTR* [27]	12.27	-	6.5	8.54	21.77	24.27	6.26	15.48	7.94	1.86	6.1	17.16	36.98	-	17.21	7.16	21.18	9.99	-	-
SparseOcc (8f) [47]	30.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	34.0	17.3
SparseOcc (16f) [47]	30.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	35.1	12.5
OPUS-T (8f) [53]	33.2	10.72	39.82	21.27	39.76	45.25	23.41	21.80	17.81	19.26	27.48	33.20	71.61	37.12	45.13	43.59	33.80	33.18	38.4	**22.4**
OPUS-L (8f) [53]	36.2	11.95	43.45	25.51	40.95	47.24	23.86	25.89	21.26	**29.06**	30.13	35.28	73.13	**41.08**	**47.01**	45.66	37.40	**35.27**	41.2	7.2
GaussianFlowOcc* [4]	16.02	-	7.23	9.33	17.55	17.94	4.5	9.32	8.51	10.66	2.00	11.80	63.89	-	31.11	35.12	14.64	12.59	16.47	10.2
ODG-T (8f)	35.54	13.69	38.97	23.02	46.75	49.33	25.79	23.63	20.73	18.54	30.01	35.61	76.84	39.33	45.01	46.78	37.45	32.24	39.2	20.1
ODG-L (8f)	**38.18**	**14.11**	**46.62**	**27.09**	**48.77**	**52.09**	**26.79**	**28.05**	23.21	27.92	**30.86**	**38.17**	**77.13**	40.35	46.94	**47.37**	**40.01**	33.52	**42.3**	4.9

nuScenes

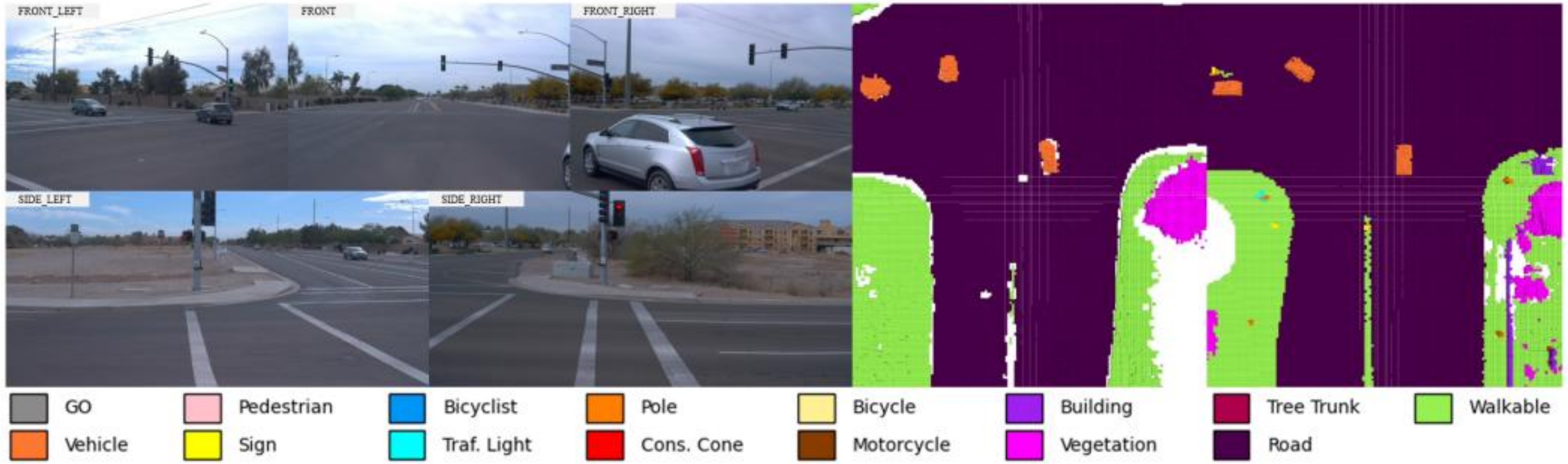
Method	mIoU	GO	Vehicle	Bicyclist	Pedestrian	Sign	Traf. Light	Pole	Cons. Cone	Bicycle	Motorcycle	Building	Vegetation	Tree Trunk	Road	Walkable	RayIoU	FPS
BEVDet [23]	9.88	0.13	13.06	2.17	10.15	7.80	5.85	4.62	0.94	1.49	0.00	7.27	10.06	2.35	48.15	34.12	-	-
BEVFormer [31]	16.76	3.48	17.18	13.87	5.9	13.84	2.7	9.82	12.2	13.99	0.00	13.38	11.66	6.73	74.97	51.61	-	-
TPVFormer [25]	16.76	3.89	17.86	12.03	5.67	13.64	8.49	8.90	9.95	14.79	0.32	13.82	11.44	5.8	73.3	51.49	-	4.6
CTF-Occ [50]	18.73	**6.26**	28.09	14.66	8.22	15.44	**10.53**	11.78	**13.62**	**16.45**	**0.65**	18.63	17.3	8.29	67.99	42.98	-	2.6
OPUS-L [53]	19.00	4.66	27.07	19.39	6.53	**18.66**	6.41	11.44	10.40	12.90	0.00	18.73	**18.11**	7.46	72.86	50.31	24.7	**8.5**
ODG-L	**21.35**	5.09	**31.34**	**22.4**	**19.06**	15.24	6.09	**12.51**	12.77	13.59	0.00	**21.49**	17.89	**8.37**	**78.19**	**56.28**	**25.9**	5.6

Waymo

Visualization



nuScenes



Waymo

Conclusion

- Expanding standard Gaussian queries to predict box attributes is an effective way to improve prediction of dynamic scenes.
- Coarse-to-fine prediction of Gaussian parameters stabilizes learning.
- Our method, ODG, achieves SotA results on challenging large-scale benchmarks.