Camera-Only 3D Panoptic Scene Completion for Autonomous Driving through Differentiable Object Shapes

Supplementary Material

A. Detailed results and additional qualitative examples

This section provides more detailed results on the Occupancy task of Occ3D-nuScenes. Tables 5 and 6 provide IoU numbers for each individual *things* and *stuff* class, respectively. We also provide additional qualitative results in Figure 5.

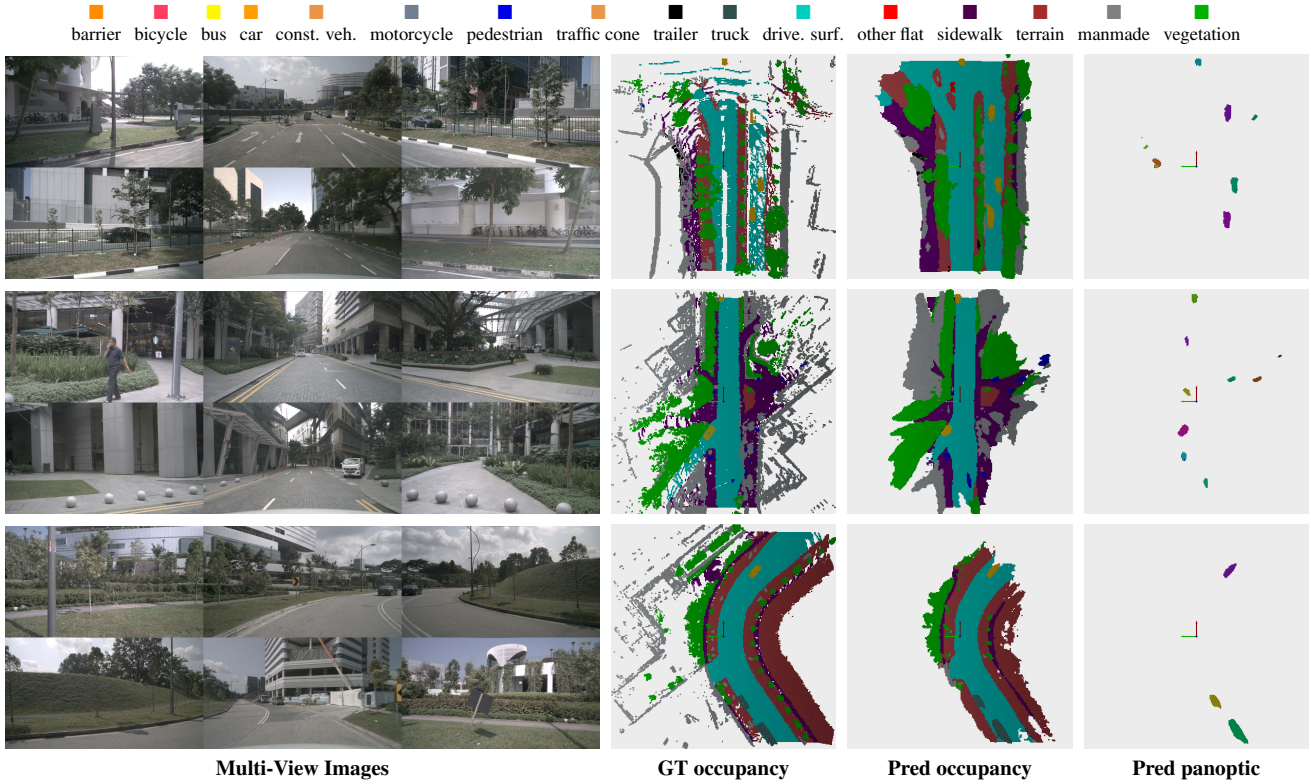

Figure 5. Additional qualitative results of our model OffsetOcc.

Method	Image Backbone	Temporal	Train w/ mask	Evaluate w/ mask	mIoU	IoU	bicycle	bus	car	const. veh.	motorcycle	pedestrian	traffic cone	trailer	truck
MonoScene [2]	R101-DCN	✗	✗	✓	6.1	-	4.3	4.9	9.4	5.7	4.0	3.0	5.9	4.4	7.2
BEVDet [9]	R101-DCN	✗	✗	✓	19.4	-	0.2	32.3	34.5	13.0	10.3	10.4	6.3	8.9	23.6
OccFormer [46]	R101	✗	✗	✓	21.9	-	12.3	34.4	39.2	14.4	16.4	17.2	9.3	13.9	26.4
BEVFormer [20]	R101-DCN	✓	✗	✓	26.9	-	17.9	40.4	42.4	7.4	23.9	21.8	21.0	22.4	30.7
TPVFormer [10]	R101-DCN	✓	✗	✓	27.8	-	13.7	40.8	45.9	17.2	20.0	18.8	14.3	26.7	34.2
CTF-Occ [34]	R101-DCN	✗	✗	✓	28.5	-	20.6	38.3	42.2	16.9	24.5	22.7	21.0	23.0	31.1
SparseOcc [24]	R50	✓	✗	✓	30.9	-	-	-	-	-	-	-	-	-	-
PanoOcc [40]	R101-DCN	✓	✗	✓	32.5	-	27.2	43.5	48.7	23.0	31.2	27.6	28.6	26.6	38.3
TPVFormer‡ [10]	R50	✓	✓	✓	34.2	66.8	17.7	40.9	47.0	15.1	20.5	24.7	24.7	24.3	29.3
OccFormer‡ [46]	R50	✗	✓	✓	37.4	70.1	18.2	42.8	50.3	24.0	20.8	22.9	21.0	31.9	38.1
BEVFormer [20]	R101-DCN	✓	✓	✓	39.2	-	24.9	47.6	54.5	20.2	28.8	28.0	25.7	33.0	38.6
PanoOcc [40]	R101-DCN	✓	✓	✓	**44.5**	**75.0**	**29.6**	**49.4**	**55.5**	**23.3**	**33.3**	**30.6**	**31.0**	**34.4**	**42.6**
OffsetOcc (Ours)	R101	✗	✗	✓	28.0	43.9	21.6	39.0	43.3	18.3	21.8	20.2	14.2	19.9	30.3
OffsetOcc (Ours)	R101	✗	✗	✗	17.2	24.9	2.3	27.9	30.6	11.2	12.2	13.3	6.2	10.6	21.3

Table 5. **3D Occupancy prediction performance on the Occ3D-nuScenes dataset *things* classes.** "Temporal" indicates that the model uses past frames when generating predictions. "Train w/ mask" and "Evaluate w/ mask" indicate whether the model has been trained using the camera mask and whether the performance has been measured using the camera mask, respectively. ‡ indicates performance measured by [26]. Best performance is **bolded** and second best is underlined.

Method	Image Backbone	Temporal	Train w/ mask	Evaluate w/ mask	mIoU	IoU	others	barrier	drive. suf.	other flat	sidewalk	terrain	manmade	vegetation
MonoScene [2]	R101-DCN	✗	✗	✓	6.1	-	1.8	7.2	14.9	6.3	7.9	7.4	1.0	7.6
BEVDet [9]	R101-DCN	✗	✗	✓	19.4	-	4.4	30.3	52.3	24.6	26.1	22.3	15.0	15.1
OccFormer [46]	R101	✗	✗	✓	21.9	-	5.9	30.3	51.0	31.0	34.7	22.7	6.8	7.0
BEVFormer [20]	R101-DCN	✓	✗	✓	26.9	-	5.8	37.8	55.4	28.4	36.0	28.1	20.0	17.7
TPVFormer [10]	R101-DCN	✓	✗	✓	27.8	-	7.2	38.9	55.6	35.5	37.6	30.7	19.4	16.8
CTF-Occ [34]	R101-DCN	✗	✗	✓	28.5	-	8.1	39.3	53.3	33.8	38.0	33.2	20.8	18.0
SparseOcc [24]	R50	✓	✗	✓	30.9	-	-	-	-	-	-	-	-	-
PanoOcc [40]	R101-DCN	✓	✗	✓	32.5	-	10.8	46.9	58.0	38.9	38.2	32.3	15.6	16.4
TPVFormer‡ [10]	R50	✓	✓	✓	34.2	66.8	7.7	44.0	79.3	40.6	48.5	49.4	32.6	29.8
OccFormer‡ [46]	R50	✗	✓	✓	37.4	70.1	9.2	45.8	80.1	38.2	50.8	54.3	**46.4**	40.2
BEVFormer [20]	R101-DCN	✓	✓	✓	39.2	-	10.1	47.9	82.0	40.6	50.9	53.0	43.9	37.2
PanoOcc [40]	R101-DCN	✓	✓	✓	**44.5**	**75.0**	**11.7**	**50.5**	**83.3**	**44.2**	**54.4**	**56.0**	45.9	**40.4**
OffsetOcc (Ours)	R101	✗	✗	✓	28.0	43.9	3.7	35.7	61.2	30.5	38.1	36.4	19.3	22.6
OffsetOcc (Ours)	R101	✗	✗	✗	17.2	24.9	2.3	19.5	36.9	20.1	22.9	17.8	11.3	16.8

Table 6. **3D Occupancy prediction performance on the Occ3D-nuScenes dataset *stuff* classes.** "Temporal" indicates that the model uses past frames when generating predictions. "Train w/ mask" and "Evaluate w/ mask" indicate whether the model has been trained using the camera mask and whether the performance has been measured using the camera mask, respectively. ‡ indicates performance measured by [26]. Best performance is **bolded** and second best is underlined.