Generalizing Visual Geometry Priors to Sparse Gaussian Occupancy Prediction

Supplementary Material

6. More experimental results

6.1. Ablation studies

Joint effect of τ and K. We evaluate different opacity thresholds (0.01–0.10) under varying K values $1, 2, 4, 8, 16, 32$ (Figure 6). The results confirm a consistent trend: lower thresholds and larger K improve accuracy, but with diminishing returns. We adopt $K = 16, \tau = 0.01$ as default.

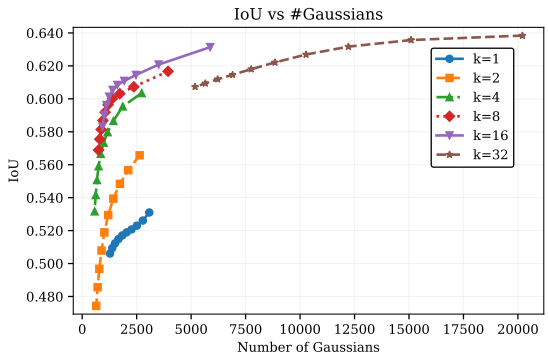

Figure 6. **IoU vs number of Gaussians.** Each curve corresponds to a fixed number of sampled points K. Moving from left to right along a curve reflects decreasing τ from 1.0 to 0.01.

Hyper-parameters of Incremental Update Strategy By default, we use $\epsilon = 0.04$ (half of the voxel size) to aggregate Gaussian primitives and $\gamma = 0.1$ to weight Gaussian attributes. To gain deeper insight, we study the influence of these hyper-parameters in the proposed incremental update strategy, evaluating them on the EmbodiedOcc-ScanNet benchmark. The results are reported in Tables 6 to 8. Using the top-1 prediction probability during fusion yields slightly better results than without as shown in Table 6 (mIoU improves by 0.21), indicating that confidence-aware weighting stabilizes updates. For the aggregation radius ϵ, a smaller value (0.02) achieves the highest mIoU and IoU, whereas a larger value (0.06) degrades performance Table 7 due to oversmoothing. Regarding the temporal weight γ, the performance remains relatively stable, likely because the dataset contains only static scenes. This observation raises an open question of how to effectively handle dynamic objects, which we leave for future research. Overall, these ablations provide a more comprehensive understanding of the proposed strategy.

Table 6. **Effect of using top-1 probability for weighting.**

	mIoU	IoU
w/	55.39	61.41
w/o	55.18	61.13

Table 7. **Effect of the aggregation radius ϵ.**

ϵ	mIoU	IoU
> 0.02	55.65	61.82
> 0.04	55.39	61.41
> 0.06	54.54	60.31

Table 8. **Effect of the temporal weight γ.**

γ	mIoU	IoU
0.1	55.39	61.41
0.3	55.36	61.36
0.5	55.37	61.38

6.2. More Quality Results

We present more visualization results in Figure 7 and Figure 8.

7. Limitation

While our method achieves strong performance, we note two limitations. First, the model is less accurate on large flat objects such as floors. Second, in the incremental update process, the number of Gaussian primitives keeps increasing, potentially causing inefficiency over long sequences. Both issues present promising directions for further improvement.

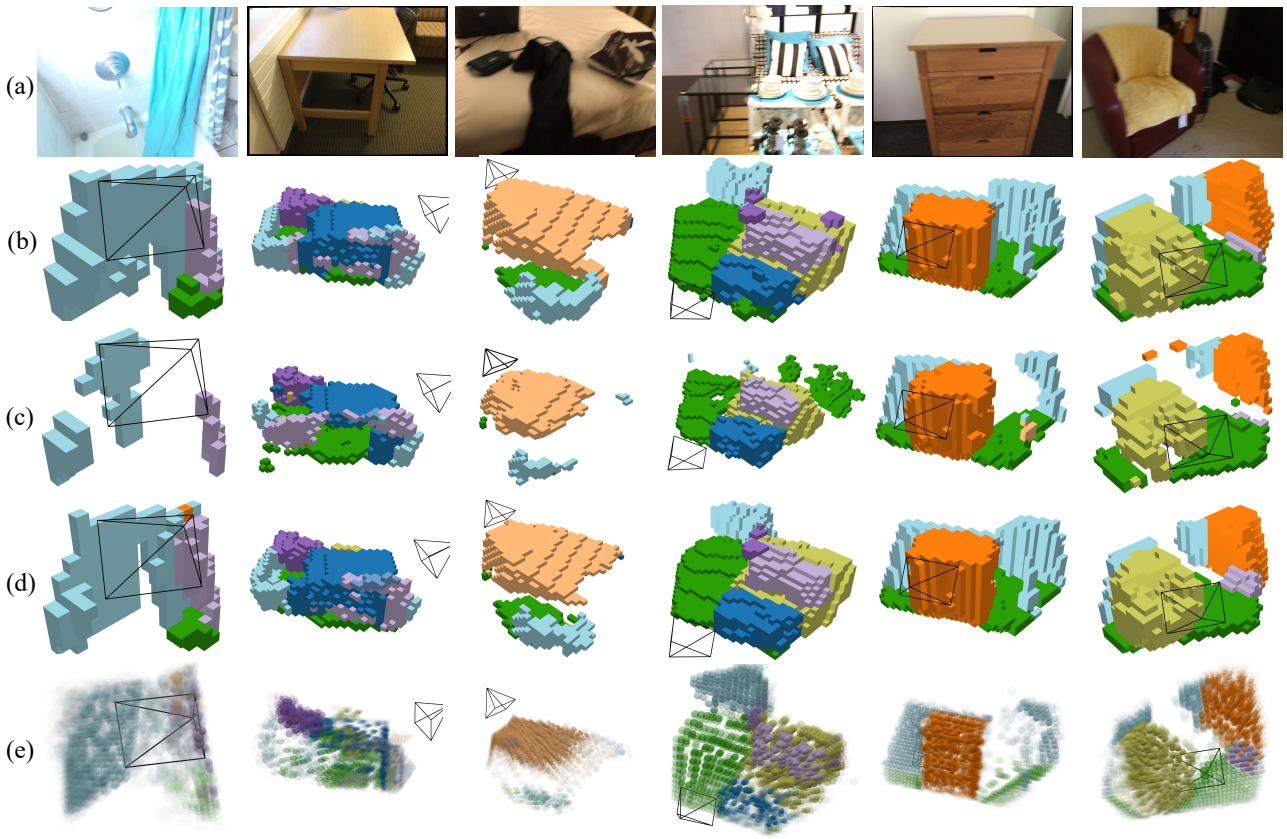

Figure 7. **Qualitative comparison on monocular occupancy prediction.** (a) shows the input RGB images, (b) the ground-truth occupancy, (c) the predictions of EmbodiedOcc [33], (d) the predictions of our method, and (e) the visualization of the Gaussian primitives predicted by our method.

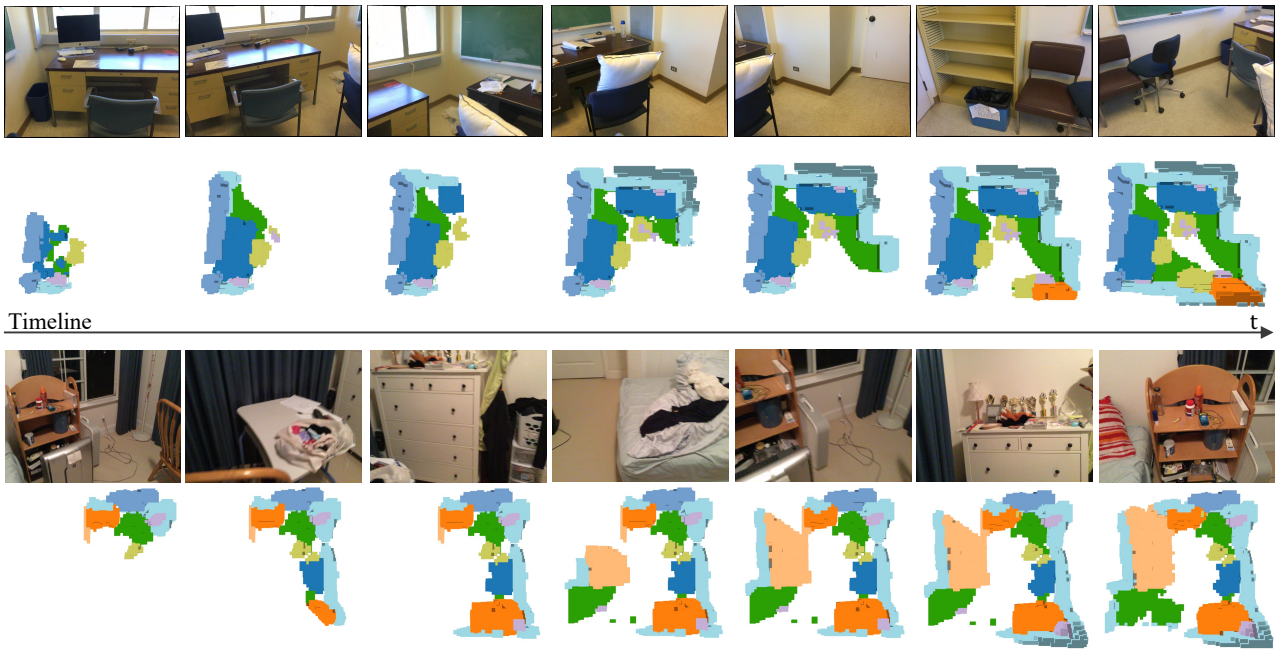

Figure 8. **Qualitative results on streaming inputs.**