A. Video and Images

Please find the video attached with of our rendered depths and examples comparing our approach to the baselines. Fig. 6 and Fig. 7 illustrates the problem of street surface estimation and floating artifacts on nuScenes and Waymo, respectively.

B. Uncertainty Estimation

Please note that directly evaluating the uncertainty of the beliefs for each voxel is not possible, as no ground-truth is available.

To this end, we derive an uncertainty for the rendered depth instead, considering all the voxels through which a ray passes. Specifically, we use the BBA to estimate an upper and lower bound of the depth and take the difference as an uncertainty. Please recall that in section 4.1., we compared occupied and free belief to decide if a voxel i is occupied or not. This is equivalent to distributing the uncertainty $\mathrm{m_i}(\Omega)$ equally to the occupied and free hypothesis:

$$o_i = \begin{cases} 1 & \text{if } \mathrm{m_i}(\mathrm{o}) > \mathrm{m_i}(\mathrm{f}) \\ 0 & \text{else} \end{cases} \tag{15}$$

$$= \begin{cases} 1 & \text{if } \mathrm{m_i}(\mathrm{o}) + \frac{1}{2}\mathrm{m_i}(\Omega) > \mathrm{m_i}(\mathrm{f}) + \frac{1}{2}\mathrm{m_i}(\Omega) \\ 0 & \text{else} \end{cases} . \tag{16}$$

The depth d_j^{est} is then normally determined by finding the first occupied voxel o_i along the ray.

Minimal and Maximal Ray Length. Since the uncertainty mass $\mathrm{m_i}(\Omega)$ is compatible with all hypotheses, other assignments of $\mathrm{m_i}(\Omega)$ are also valid. Therefore, we estimate the lower and upper bound of the ray length by distributing the uncertainty mass to only one of both hypotheses. If we assign all uncertainty mass $\mathrm{m_i}(\Omega)$ to the occupied hypothesis $\mathrm{m_i}(\mathrm{o})$, we obtain more occupied voxels:

$$o_i^{\text{occ}} = \begin{cases} 1 & \text{if } \mathrm{m_i}(\mathrm{o}) + \mathrm{m_i}(\Omega) > \mathrm{m_i}(\mathrm{f}) \\ 0 & \text{else} \end{cases} . \tag{17}$$

This leads to the the minimal ray length $d_j^{\min}$. Contrary to this, assigning all uncertainty to the free hypothesis:

$$o_i^{\text{free}} = \begin{cases} 1 & \text{if } \mathrm{m_i}(\mathrm{o}) > \mathrm{m_i}(\mathrm{f}) + \mathrm{m_i}(\Omega) \\ 0 & \text{else} \end{cases} \tag{18}$$

will lead to less occupied voxels, and we obtain the maximum ray length $d_j^{\max}$. Overall, we obtain the lower bound of the ray length $d_j^{\min}$, the upper bound of the ray length $d_j^{\max}$, and the already computed estimated ray length d_j^{est}.

Rendered Depth Uncertainty. Given the bounds, we define the uncertainty per ray as maximal deviation from the estimation:

$$d_j^{\text{uncert}} = \max\left(|d_j^{\max} - d_j^{\text{est}}|, |d_j^{\text{est}} - d_j^{\min}|\right). \tag{19}$$

Note, that all ray lengths being equal corresponds to an uncertainty of zero. We compare the obtained uncertainty per ray d_j^{uncert} with the error to the LiDAR measurement

$$d_j^{\text{error}} = |d_j^{\text{est}} - d_j^{\text{lidar}}| . \tag{20}$$

Please find the results illustrated in Fig. 8.

C. Multi-Frame Temporal Aggregation

In Tab. 5 we show the impact of changing the number of frames for temporal aggregation. We observe that more frames for temporal aggregation increase the performance at the cost of computation time.

#frames	MAE ▲▼ in m	RMSE ▲▼ in m	$\delta < 1.25$ ▲▼ in %	$\delta < 1.25^2$ ▲▼ in %	$\delta < 1.25^3$ ▲▼ in %
5	0.99	3.09	90.0	94.8	97.2
10	0.94	2.99	91.8	95.6	97.4
50	0.92	2.96	92.9	96.2	97.7
100	0.92	2.95	93.0	96.3	97.7

Table 5. **Impact of Number of Frames Used for Temporal Aggregation.** We evaluate the generated GT for a varying number of frames used for temporal aggregation on the nuScenes mini dataset with a voxel size 0.2 m.

D. Uncertainty Loss Weighting

To further analyze the influence of uncertainty weighting, we trained the models without it and report the results in Tab. 6. Uncertainty weighting is important for the final performance.

Loss	MAE ▲▼ in m	RMSE ▲▼ in m	$\delta < 1.25$ ▲▼ in %	$\delta < 1.25^2$ ▲▼ in %	$\delta < 1.25^3$ ▲▼ in %
BCE	1.44/1.70	3.60/4.04	88.0/82.7	94.0/91.1	96.7/94.6
CE	1.42/1.55	3.47/4.10	88.5/88.2	94.5/94.0	97.0/96.5

Table 6. **Uncertainty Loss Weighting.** Training results with/without uncertainty loss weighting.

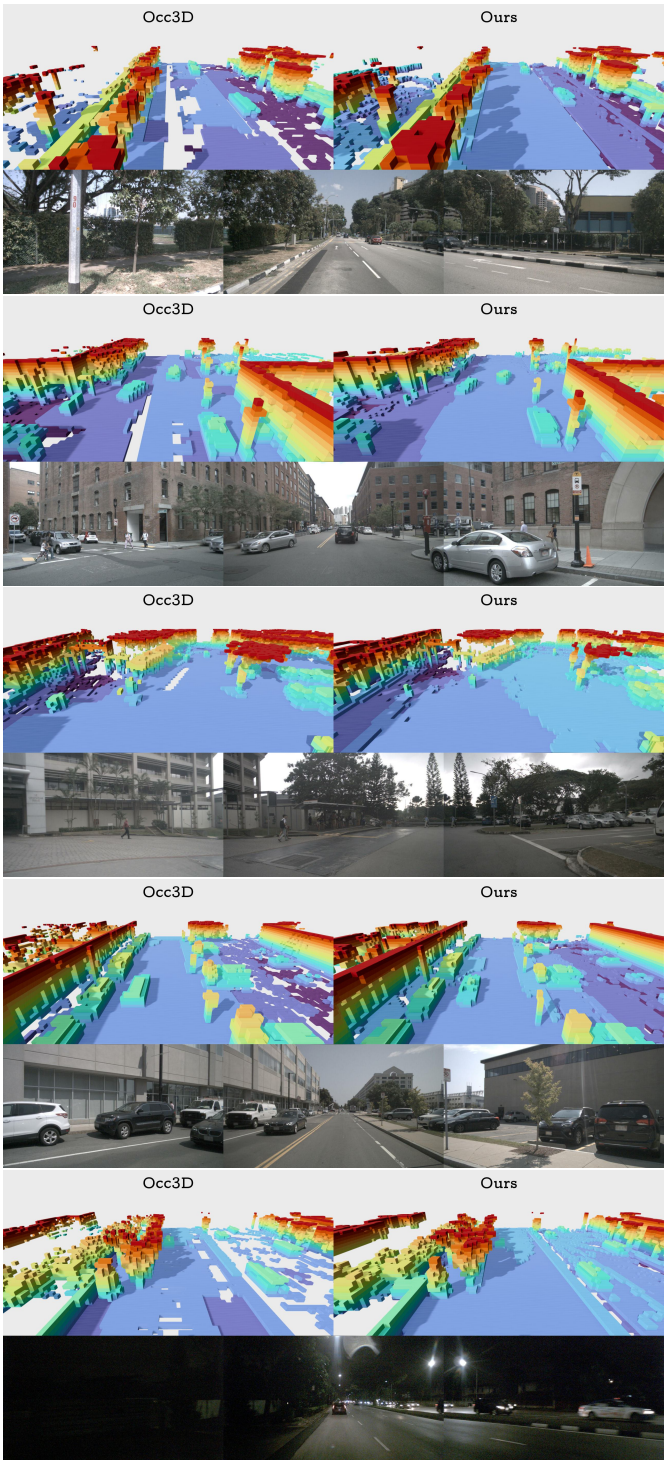

Figure 6. **Qualitative Comparison on the nuScenes Dataset.** We compare the occupancy maps generated by our method with the ones of Occ3D [22] on the nuScenes dataset [2]. Our method yields much better street surfaces compared to Occ3D in many scenarios. The voxel size of both methods is 0.4 m.

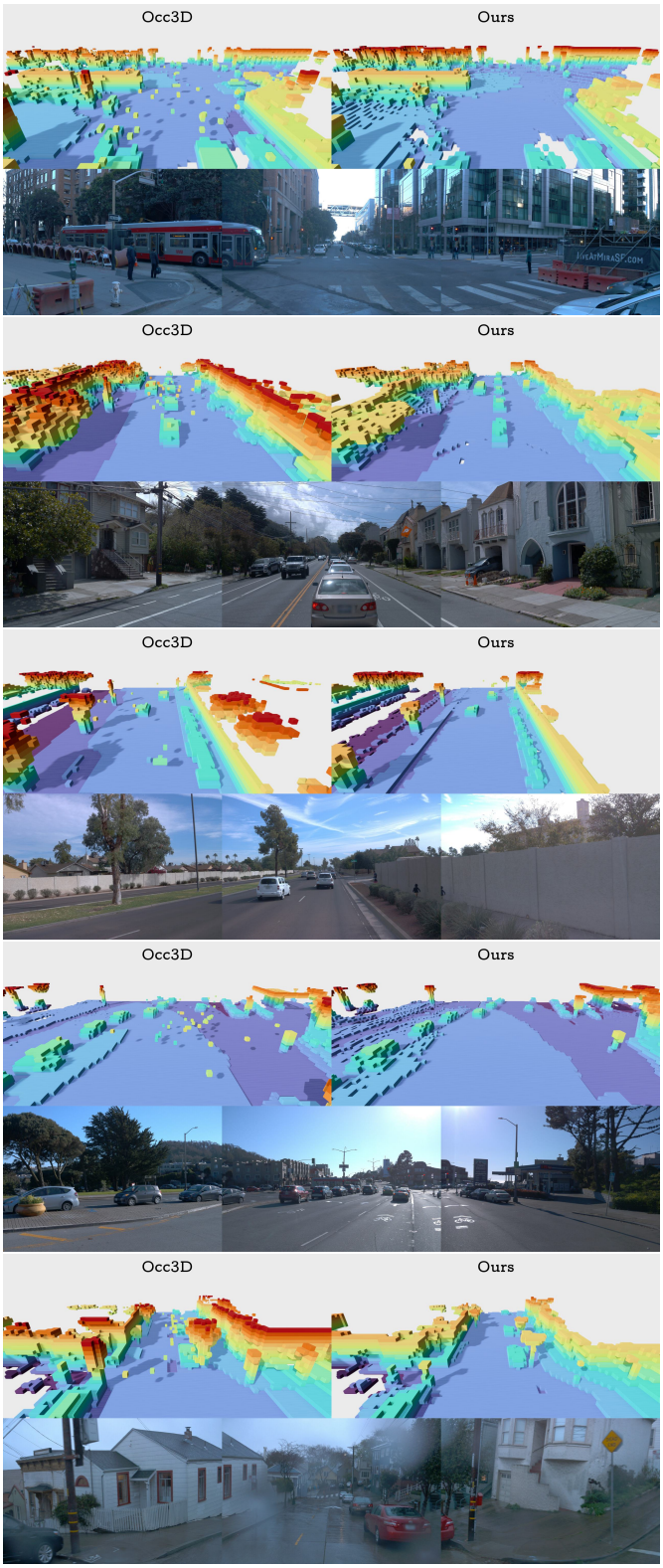

Figure 7. **Qualitative Comparison on the Waymo Dataset.** We compare the occupancy maps generated by our method with the ones of Occ3D [22] on the Waymo dataset [20]. Our method yields much less floating artifacts due to the explicit modeling of free space. The voxel size of both methods is 0.4 m.

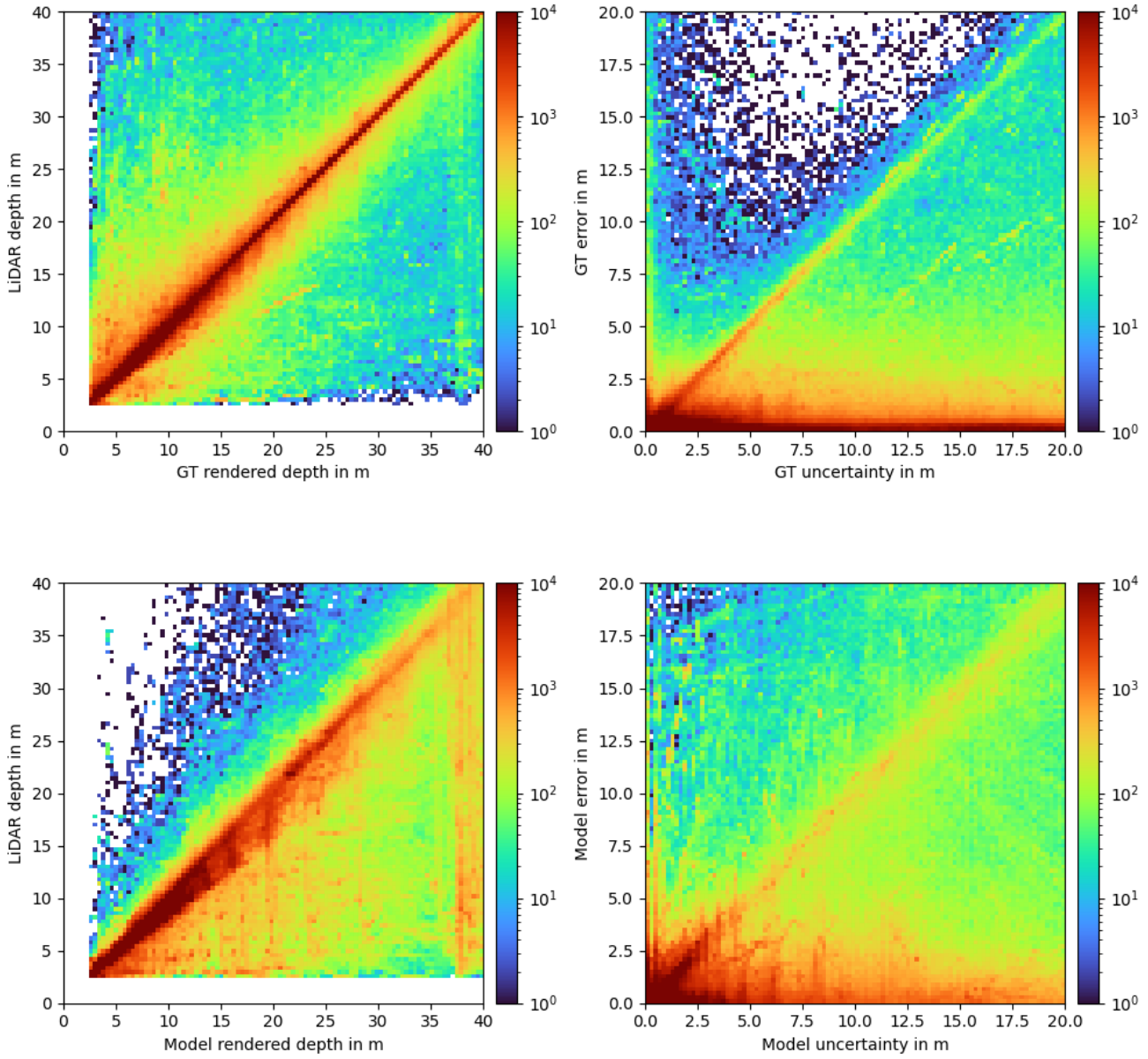

Figure 8. **Uncertainty Estimates.** We compare rendered depths and corresponding uncertainties of our occupancy ground-truth data with our model predictions. The top row contains data from our ground-truth occupancy, while the bottom row contains data from model predictions. We plot rendered depths d_j^{est} against the LiDAR measurements on the left side. Therefore, we create a heatmap containing the absolute frequency of $(d_j^{\text{est}}, d_j^{\text{lidar}})$ pairs. Perfect predictions lie on the diagonal such that $d_j^{\text{est}} = d_j^{\text{lidar}}$. The right column shows the estimated uncertainty d_j^{uncert} on the x-axis and the estimation error $d_j^{\text{error}} = |d_j^{\text{est}} - d_j^{\text{lidar}}|$ on the y-axis. We follow the same procedure to create the heatmap of $(d_j^{\text{uncert}}, d_j^{\text{error}})$ pairs. As visible by the strong diagonals, our method provides meaningful uncertainty estimates with a slight tendency to overestimate them.