

EvOcc: Accurate Semantic Occupancy for Automated Driving Using Evidence Theory

Supplementary Material

| Parameters | Value |
|------------------------|---|
| Spherical grid extent: | $\rho = 2.5 \text{ m to } 60 \text{ m}$ $\theta = 75^\circ \text{ to } 125^\circ$ $\phi = -180^\circ \text{ to } 180^\circ$ |
| Spherical cell size: | $\Delta\rho = 0.1 \text{ m}, \Delta\theta = \Delta\phi = 0.5^\circ$ |
| Cartesian grid extent: | $x = -40 \text{ m to } 40 \text{ m}$ $y = -40 \text{ m to } 40 \text{ m}$ $z = -1 \text{ m to } 5.4 \text{ m}$ |
| Cartesian voxel size: | $\Delta x = \Delta y = \Delta z = 40 \text{ cm or } 20 \text{ cm}$ |
| Reflection scale: | $\alpha_r^{40 \text{ cm}} = 10^{-1}$ $\alpha_r^{20 \text{ cm}} = 5 \times 10^{-2}$ |
| Transmission scale: | $\alpha_t^{40 \text{ cm}} = 10^{-2}$ $\alpha_t^{20 \text{ cm}} = 10^{-2}$ |

Table I. **Evidential mapping parameter overview.** We report the parameters used for the evidential grid mapping.

A. Spherical Grid Mapping Details

We normalize the reflections and transmissions by the voxel volume ratio $s(\rho, \phi, \theta)$ between the Cartesian voxel volume $V_{\text{cartesian}}$ and the spherical voxel volume $V_{\text{spherical}}(\rho, \phi, \theta)$ [15]:

$$\hat{r}_{c_i}^{\text{sph}}(\rho, \phi, \theta) = s(\rho, \phi, \theta) \cdot r_{c_i}^{\text{sph}}(\rho, \phi, \theta), \quad (16)$$

$$s(\rho, \phi, \theta) = \frac{V_{\text{cartesian}}}{V_{\text{spherical}}(\rho, \phi, \theta)}, \quad (17)$$

$$V_{\text{spherical}}(\rho, \phi, \theta) = \int_{\rho - \frac{\Delta\rho}{2}}^{\rho + \frac{\Delta\rho}{2}} \int_{\theta - \frac{\Delta\theta}{2}}^{\theta + \frac{\Delta\theta}{2}} \int_{\phi - \frac{\Delta\phi}{2}}^{\phi + \frac{\Delta\phi}{2}} \tilde{\rho}^2 \sin(\tilde{\theta}) d\tilde{\rho} d\tilde{\theta} d\tilde{\phi}, \quad (18)$$

$$V_{\text{cartesian}} = \Delta x \Delta y \Delta z. \quad (19)$$

This normalization is necessary since the volume of the spherical grid cells $V_{\text{spherical}}(\rho, \phi, \theta)$ is not constant and depends on the spherical coordinates (ρ, ϕ, θ) . The integral for the spherical volume $V_{\text{spherical}}(\rho, \phi, \theta)$ corresponds to the volume of a spherical voxel located around (ρ, ϕ, θ) with a voxel size of $(\Delta\rho, \Delta\phi, \Delta\theta)$. The volume of the Cartesian grid cells $V_{\text{cartesian}}$ is constant and can be calculated by multiplying the grid cell dimensions $(\Delta x, \Delta y, \Delta z)$.

We report the used parameters for the grid mapping in Tab. I. They are selected to match the LiDAR's field of view on the nuScenes[2] dataset.

B. Subset Matrix

We explicitly calculate the matrix \mathbf{S} from Sec. 4.1 given the hypotheses defined in Tab. 1. Each row and column of the matrix is associated with the hypothesis written next to it. A matrix element \mathbf{S}_{ij} is one if the hypothesis of the column j is a subset of the hypothesis of row i and zero otherwise:

$$\mathbf{S} \in \mathbb{R}^{n \times n} : \mathbf{S}_{ij} = 1\{\omega_j \subseteq \omega_i\}, \quad (20)$$

$$\mathbf{S} = \begin{matrix} & \mathcal{F} & c_1 & \dots & c_C & \mathcal{O} & \Omega \end{matrix} \begin{matrix} \mathcal{F} \\ c_1 \\ \vdots \\ c_C \\ \mathcal{O} \\ \Omega \end{matrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (21)$$

C. Inverse of Subset Matrix

In the following we prove that \mathbf{S} is invertible for any choice of hypotheses. Without loss of generality we sort the hypothesis ω by cardinality $|\omega|$, such that

$$|\omega_i| \leq |\omega_j| \quad \forall i < j, \quad (22)$$

which puts smaller subsets first, followed by larger ones towards the end. Next, we evaluate Eq. (20) for the diagonal and the upper triangular part of matrix \mathbf{S} . By construction of \mathbf{S} , the diagonal elements are always equal to one $\mathbf{S}_{ii} = 1$, since a hypothesis is always equal to or is a subset of itself $\omega_i \subseteq \omega_i$. Furthermore, there are the following cases to consider for the elements \mathbf{S}_{ij} with $i < j$ in the upper right triangular part based on the cardinalities of the hypotheses:

1. $|\omega_i| < |\omega_j|$: ω_j has a greater cardinality than ω_i and can't be equal to or subset of ω_i . Evaluating Eq. (20) leads to $\mathbf{S}_{ij} = 0$.
2. $|\omega_i| = |\omega_j|$: Since ω_j and ω_i have the same cardinality, ω_j can't a true subset of ω_i . Furthermore, ω_i and ω_j can't be equal since all hypotheses differ in at least one element. This also leads to $\mathbf{S}_{ij} = 0$.
3. $|\omega_i| > |\omega_j|$: This case is impossible due to the sorting of the hypotheses described in Eq. (22).

This means, that each element in the upper right triangle of \mathbf{S} is equal to zero: $\mathbf{S}_{ij} = 0 \quad \forall i < j$. Therefore, \mathbf{S} is a square lower triangular matrix with unit diagonal, which is always invertible.

| Occupancy GT | Loss | MAE in m \uparrow | RMSE in m \uparrow | $\delta < 1.25$ in % \uparrow | $\delta < 1.25^2$ in % \uparrow | $\delta < 1.25^3$ in % \uparrow | $\delta_{\text{cls}} < 1.25$ in % \uparrow | $\delta_{\text{cls}} < 1.25^2$ in % \uparrow | $\delta_{\text{cls}} < 1.25^3$ in % \uparrow | mIoU in % \uparrow |
|--------------|-------------------|------------------------|-------------------------|------------------------------------|--------------------------------------|--------------------------------------|---|---|---|-------------------------|
| Occ3D | CE | 2.33 | 4.31 | 58.8 | 89.6 | 94.3 | 50.4 | 74.3 | 76.4 | 43.1 |
| Ours | CE | 2.30 | 5.92 | 86.0 | 91.9 | 94.7 | 75.4 | 78.0 | 78.7 | 46.7 |
| Ours | CE _{BBA} | 1.66 | 4.18 | 88.6 | 94.3 | 96.7 | 76.6 | 79.5 | 80.4 | 48.6 |

Table II. **Ablation of loss function and occupancy GT.** The model in the first row is trained with classical cross-entropy loss with the occupancy data from Occ3D [31]. The second row shows the same model trained with the same loss function but with occupancy data from our method. The last row shows the model trained with our proposed loss function (see Sec. 4.2) and occupancy data. Both the occupancy data and the loss function have a significant impact on the model’s performance.

| | MAE in m \uparrow | RMSE in m \uparrow | $\delta < 1.25$ in % \uparrow | $\delta < 1.25^2$ in % \uparrow | $\delta < 1.25^3$ in % \uparrow | $\delta_{\text{cls}} < 1.25$ in % \uparrow | $\delta_{\text{cls}} < 1.25^2$ in % \uparrow | $\delta_{\text{cls}} < 1.25^3$ in % \uparrow | mIoU in % \uparrow |
|-------|------------------------|-------------------------|------------------------------------|--------------------------------------|--------------------------------------|---|---|---|-------------------------|
| Occ3D | 3.23 | 6.52 | 75.6 | 83.1 | 87.6 | 68.0 | 71.3 | 72.5 | 32.9 |
| Ours | 1.97 | 5.07 | 88.0 | 91.9 | 94.0 | 81.4 | 83.0 | 83.5 | 49.4 |

Table III. **Evaluation on the Waymo dataset.** We compare our semantic occupancy mapping method with Occ3D [31] on the Waymo dataset with a voxel size of 40 cm. Our method yields superior occupancy grid maps and outperforms Occ3D in all metrics.

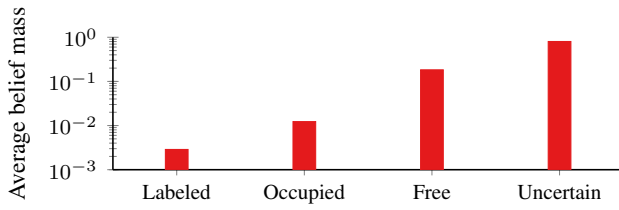


Figure I. **Belief mass distribution.** The bar plot shows the distribution of the belief masses $m_{\text{mean}}^{\text{gt}}(\omega)$ in log-scale for the different hypotheses averaged over the training set of nuScenes [2]. The belief mass of all labeled classes is summed up since individual classes are even less frequent. The belief mass for the uncertain hypothesis Ω is significantly higher than for all other hypotheses. This is due to the fact that the most voxels have never been measured and are therefore assigned to the uncertain hypothesis.

D. Distribution of Belief Masses

We visualize the belief mass distribution in Fig. 1 and observe that it is very unbalanced. Therefore, we introduced the per voxel weighting described in Eq. (13).

E. Impact of Direct BBA Supervision

To evaluate the impact of the direct BBA supervision (see Sec. 4.2), we compare the performance of the model trained with and without the direct BBA supervision. We follow the same training procedure as described in Sec. 5.2 and evaluate the model on the nuScenes dataset [2]. We report results in Tab. II. Both the occupancy data and the loss function have a significant impact on the model’s performance.

F. Evaluation on Waymo

We evaluate our method on the Waymo dataset [30] to demonstrate the generalization capabilities of our semantic occupancy mapping approach. We compare the semantic occupancy maps generated by our method with those of Occ3D [31] for a voxel size of 40 cm. Tab. III shows that our method produces superior occupancy grid maps compared to Occ3D.

G. Compute Resources

The spherical grid mapping and multi-frame aggregation processing takes around a day to process all training and validation scenes (850 total scenes) of the nuScenes dataset [2] on eight Nvidia V100. The training of a single model (48 epochs) roughly takes two days on two Nvidia A100 80GB. The computational overhead of predicting basic belief masses instead of class probabilities is negligible compared to the overall computational cost of the model, since only the number of outputs in the last layer change slightly.

H. Limitations

Our method relies on the poses of the ego vehicle and given 3D bounding boxes for motion compensation and multi-frame aggregation. Therefore, the quality of the generated semantic occupancy maps is limited by the quality of the ego poses and the 3D bounding boxes. Misaligned poses or inaccurate bounding boxes may lead to artifacts in the generated semantic occupancy maps. Furthermore, 3D bounding boxes may not represent the true shape of an object. Additionally, the voxel size is a limiting factor since the memory requirements increase cubically with decreasing voxel sizes.