Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Reliable and Calibrated Semantic Occupancy Prediction by Hybrid Uncertainty Learning

Song Wang[1], **Zhongdao Wang**[2], **Jiawei Yu**[1], **Wentong Li**[3], **Bailan Feng**[2], **Junbo Chen**[4*] and **Jianke Zhu**[1*]
[1]Zhejiang University
[2]Huawei Noah's Ark Lab
[3]Nanjing University of Aeronautics and Astronautics
[4]Udeer.ai
{songw, jkzhu}@zju.edu.cn, junbo@udeer.ai

Abstract

Vision-centric semantic occupancy prediction plays a crucial role in autonomous driving, which requires accurate and reliable predictions from low-cost sensors. Although having notably narrowed the accuracy gap with LiDAR, there is still few research effort to explore the reliability and calibration in predicting semantic occupancy from camera. In this paper, we conduct a comprehensive evaluation of existing semantic occupancy prediction models from a reliability perspective for the first time. Despite the gradual alignment of camera-based models with LiDAR in terms of accuracy, a significant reliability gap still persists. To address this concern, we propose RELIOCC, a method designed to enhance the reliability of camera-based occupancy networks. RELIOCC provides a plug-and-play scheme for existing models, which integrates hybrid uncertainty from individual voxels with sampling-based noise and relative voxels through mix-up learning. Besides, an uncertainty-aware calibration strategy is devised to further improve model reliability in offline mode. Extensive experiments under various settings demonstrate that RELIOCC significantly enhances the reliability of learned model while maintaining the accuracy for both geometric and semantic predictions. Notably, our proposed approach exhibits robustness to sensor failures and out of domain noises during inference.

1 Introduction

The goal of semantic occupancy prediction is to obtain a comprehensive voxel-based representation of the 3D scene from either LiDAR point clouds [Roldao *et al.*, 2020; Xia *et al.*, 2023] or camera images [Cao and de Charette, 2022; Huang *et al.*, 2023; Li *et al.*, 2023], which is crucial for perception systems in autonomous driving and robotic platforms. Initially, LiDAR-based models [Roldao *et al.*, 2020; Cheng *et al.*, 2021; Yan *et al.*, 2021; Xia *et al.*, 2023] dominated the field due to their ability to provide accurate geometric cues. Researchers nowadays prefer to learn 3D occupancy information from images owing to the low cost and widespread availability of camera sensors. Recent progress [Cao and de Charette, 2022; Huang *et al.*, 2023; Li *et al.*, 2023; Yao *et al.*, 2023; Mei *et al.*, 2024] has significantly narrowed the gap in accuracy between camera and LiDAR-based approaches. However, their performance in terms of reliability remains under-explored, which becomes paramount in safety-critical scenarios.

Traditionally, the occupancy labeling relies on accumulated LiDAR scans and their corresponding point-wise semantic labels [Behley *et al.*, 2019; Tian *et al.*, 2024; Wang *et al.*, 2023; Wei *et al.*, 2023]. With the development of vision-centric approaches [Cao and de Charette, 2022; Huang *et al.*, 2023; Yao *et al.*, 2023; Li *et al.*, 2023; Mei *et al.*, 2024] using images, questions arise regarding the reliability of predictions solely derived from cameras without accurate depth information. Since the overall accuracy of occupancy networks is relatively low, exploring the reliability and uncertainty of their predictions can provide valuable reference information for downstream tasks in driving, such as decision-making and planning [Zheng *et al.*, 2024; Hu *et al.*, 2023; Albrecht *et al.*, 2021].

With the above considerations, we conduct a thorough evaluation of existing semantic occupancy prediction models based on a reliability standpoint. To achieve this, we introduce the misclassification detection and calibration metrics from both geometric and semantic dimensions for evaluating model that utilize camera or LiDAR data. Our findings reveal that camera-based models often lag behind their LiDAR-based counterparts in terms of reliability despite improvements in accuracy.

To mitigate this disparity, RELIOCC is proposed to improve the reliability in occupancy networks by a new hybrid uncertainty learning scheme. Our approach optimizes uncertainty by taking into consideration of perturbations in individual voxels (*absolute uncertainty*) and the relative relationships in mix-up voxels (*relative uncertainty*) during model training. By integrating multiple sources of information for uncertainty learning, our method enhances the reliability of

[*]Corresponding authors.

camera-based models without sacrificing inference speed or accuracy. Moreover, we provide an uncertainty-aware calibration strategy to utilize the learned uncertainty in offline mode, further enhancing model's reliability. Through extensive experiments across diverse configurations including online and offline modes, our method achieves competitive performance compared against the state-of-the-art models.

Our main contributions can be summarized as follows:

- A comprehensive evaluation is conducted on existing semantic occupancy prediction models from a reliability perspective, which provides a series of misclassification detection and calibration metrics across both geometric and semantic dimensions.
- We provide a systematic scheme for adapting the existing methods to achieve reliable and calibrated occupancy networks. Furthermore, RELIOCC is proposed to enhance the reliability of camera models. A novel hybrid uncertainty learning approach is presented to combine the variance from individual and mix-up voxels.
- Extensive experiments on online uncertainty learning and offline model calibration across diverse settings demonstrate the effectiveness of our approach in bridging the reliability gap between camera and LiDAR-based methods under general conditions, while showcasing robust performance in adverse scenarios such as sensor failures and noisy observations.

2 Related Work

Semantic Occupancy Prediction. Semantic occupancy prediction is also known as semantic scene completion and firstly explored in indoor scenes [2017; 2022]. In outdoor scenarios, SemanticKITTI [2019] stands as the first large-scale dataset, providing abundant data resources. Recently, several other datasets [2022; 2023; 2023; 2024] have been constructed to explore this task owing to its importance. LiDAR-based methods [2020; 2021; 2021; 2022; 2023] have dominated this field in accuracy. MonoScene [2022] is the first occupancy prediction method that utilizes single image as input. Subsequent studies [2023; 2023; 2023; 2024; 2024; 2024] have effectively improved the performance of camera-based models. However, there is a lack of research on reliability of occupancy predictions, posing potential risks to the safety of downstream tasks in driving [2021; 2023; 2024]. RELIOCC fills this gap by investigating the reliability of occupancy networks by uncertainty learning.

Uncertainty Learning and Model Calibration. The uncertainty in machine learning consists of aleatoric uncertainty from data noises and epistemic uncertainty from model parameters [2017; 2018]. Data uncertainty is widely explored in face field [2017; 2019; 2020; 2021]. Cai *et al.* [2023] propose a probabilistic embedding model to estimates the data uncertainty for point cloud. Model uncertainty is usually obtained from the statistics of multiple predictions through methods including model ensembling [2017], bootstrapping [2001], and bagging [1996]. Model calibration is another line to improve reliability in model prediction [2023], which provides a post-processing scheme applied to the non-probabilistic output from a trained model. Model calibration was initially studied in image classification [2017; 2019; 2021] and has since been widely applied to object detection [2020; 2022] and semantic segmentation [2021; 2025]. Our method adopts uncertainty as a learning objective and can support both online uncertainty estimation and offline model calibration simultaneously.

3 Preliminaries

3.1 Problem Formulation

Occupancy Prediction. Given inputs $\boldsymbol{x}$ from LiDAR or camera sensors, occupancy networks $V_\theta(\boldsymbol{x})$ generate dense features $\mathcal{V} \in \mathbb{R}^{d\times L\times W\times H}$ in a pre-defined volume, where L, W, and H represent the length, width, and height, respectively. d is the dimension of dense features. For any voxel $v_i \in \mathbb{R}^d$ within this volume, the prediction involves with two components. One is a binary indicator that specifies whether the voxel is occupied or not. The other is the semantic label of the voxel if the voxel is occupied. Generally, such process can be formulated by estimating the probability $p(y_i = y|v_i)$ for v_i, where $y \in \{0, 1, ..., S\}$. Here, 0 denotes that the voxel is unoccupied, and S is the total number of semantic classes.

Misclassification Detection. For a reliable classifier, we expect it to accurately reject those incorrect predictions with low-confidence. Therefore, misclassification detection is introduced to measure the gap between the actual trained model and the ideal one, which can be evaluated by *rejection curves* [Fumera and Roli, 2002; Hendrycks *et al.*, 2021]. To avoid the tendency of higher precision models, we adopt the same strategy as [Malinin *et al.*, 2019; de Jorge *et al.*, 2023] to normalize the area under the curve and deducts a baseline score.

Calibration. As a long-standing problem in machine learning, the goal of model calibration is to ensure that predicted confidence of a model aligns accurately with the actual likelihood of correctness [Niculescu-Mizil and Caruana, 2005; Guo *et al.*, 2017], thereby producing more reliable predictions. Within the framework of multi-class classification, a model is deemed perfectly calibrated if $p(y_i = y|c_i = c, v_i) = c$. Here, the model not only predicts a discrete label y but also generates a confidence score $c \in [0, 1]$. This score c should ideally reflect the true probability that the predictions are correct.

3.2 Evaluation Metrics

Canonical Metrics. Following common practice [Song *et al.*, 2017; Behley *et al.*, 2019], we employ the Intersection over Union (IoU) metric to assess the accuracy of geometric occupancy prediction that is typically treated as a binary classification task. Additionally, we utilize the mean Intersection over Union (mIoU) across multiple categories to evaluate the quality of semantic predictions. These two metrics are calculated using discrete predictions by applying `argmax` operation to the logits. Although IoU and mIoU effectively reflect the performance of model in accuracy aspect, they do not assess its reliability.

In this paper, we mainly evaluate the reliability of occupancy prediction on misclassification detection and calibration, which are assessed by following metrics.

Prediction Rejection Ratio (PRR). The Prediction Rejection Ratio (PRR) [Malinin *et al.*, 2019] is defined through *rejection curves* for misclassification detection. To construct a rejection curve, we initially sort predictions based on a specific criterion, such as predicted confidence or oracle confidence (where predictions are labeled 1 if correct and 0 otherwise). Subsequently, a threshold is set and predictions below this threshold are rejected, allowing us to calculate a rejection rate. As this threshold is incrementally adjusted, we obtain a rejection curve to illustrate how the classification error (depicted on the *y*-axis) decreases in tandem with the rejection rate (represented on the *x*-axis). The PRR metric is then quantitatively defined as follows

$$\texttt{PRR} = \frac{AUC_{\texttt{random}} - AUC_{\texttt{uncertainty}}}{AUC_{\texttt{random}} - AUC_{\texttt{oracle}}}, \quad (1)$$

where AUC represents the Area Under the Curve. Here, $AUC_{\texttt{random}} = 0.5$ corresponds to the AUC for randomly generated confidences. A perfectly reliable model would achieve $\texttt{PRR} = 1$. For occupancy networks, we report both $\texttt{PRR}_{\texttt{geo}}$ for geometric predictions and $\texttt{PRR}_{\texttt{sem}}$ for semantic predictions, respectively.

Expected Calibration Error (ECE). Expected Calibration Error (ECE) [Naeini *et al.*, 2015; Guo *et al.*, 2017] assesses the calibration of probabilistic predictions made by machine learning models. It measures the difference between predicted probabilities and observed frequencies across various confidence levels. Intuitively,

$$e_{\texttt{ECE}} = \mathbb{E}_{\hat{c}_i}[\,|\,p(\hat{y}_i = y_i \mid \hat{c}_i = c) - c\,|\,]\,. \quad (2)$$

A perfectly calibrated model yields $e_{\texttt{ECE}} = 0$. Eq. (2) is a continuous integration over $c \in [0, 1]$. Practically, we approximate this integration by discretizing c into M small bins. Denoting the set of samples falling into the m-th bin as B_m, the expectation can be calculated as

$$\texttt{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \,|\, \mathrm{acc}(B_m) - \mathrm{conf}(B_m)\,|\;, \quad (3)$$

where $\mathrm{acc}(\cdot)$ denotes the mean accuracy, and $\mathrm{conf}(\cdot)$ is mean confidence of B_m. N is the number of samples. We set the number of bins $M = 15$ by default. As with $\texttt{PRR}$, we report both $\texttt{ECE}_{\texttt{geo}}$ and $\texttt{ECE}_{\texttt{sem}}$ for geometric and semantic predictions, respectively.

4 Adaptation with Existing Methods for Occupancy Networks

Reliable predictions are paramount in occupancy networks, especially in critical applications such as autonomous driving and robotics where safety is a strict requirement. Despite their importance, methods for enhancing the reliability and calibration of occupancy networks are still under-explored in the existing literature. To address this gap, we begin by reviewing existing uncertainty learning and calibration methods, which are mostly developed for traditional tasks. Then, we adapt them for the recent occupancy prediction networks.

We categorize these methods into two paradigms. One is training uncertainty predictor $c_{\sigma|\phi}$ based on the dense features $\mathcal{V}$ concurrently with V_θ from scratch, which is termed *online uncertainty learning*. Another is training scaling factor $c_{f|\phi}$ on top of a fixed V_θ, which is termed *offline model calibration*. In the experimental section (see §6.2 and §6.3), we provide extensive evaluations on these methods to compare their effectiveness in boosting the reliability.

4.1 Online Uncertainty Learning

Uncertainty estimation is a long-standing problem in the context of Bayesian deep learning [Tishby and Solla, 1989; Denker and LeCun, 1990; Gal and Ghahramani, 2016]. Prior arts can be classified into ones concerning epistemic (model) uncertainty [Lakshminarayanan *et al.*, 2017; Jungo and Reyes, 2019] and ones concerning aleatoric (data) uncertainty [Kendall and Gal, 2017; Hüllermeier and Waegeman, 2021]. Although explicit uncertainty estimates are obtainable, we do not directly evaluate these estimates in online mode. Instead, since the uncertainty is learned concurrently with the model's predictions from scratch, we use them as a regularization term to help the model become more reliable.

For each voxel feature $\mathbf{v_i}$, we compute a logit vector $\mathbf{z_i} \in \mathbb{R}^{S+1}$ using a linear layer, where S represents the number of semantic classes.

Heteroscedastic Aleatoric Uncertainty (HAU) [Kendall and Gal, 2017] is a data-dependent uncertainty learning method. We employ the classification form, which modifies upon a deterministic model by placing a Gaussian over the logit: $\hat{\mathbf{z}}_i|\phi \sim \mathcal{N}(\mathbf{z}_i, (\sigma_i^\phi)^2)$[1]. The sampled logit vector $\hat{\mathbf{z}}_i$ is then passed through a *softmax* operator and cross entropy loss is computed. Here, σ_i^ϕ is the predicted uncertainty parameterized by ϕ. Optimization of ϕ can be done with back-propagation using the re-parameterization trick [Kingma and Welling, 2013]: $\hat{\mathbf{z}}_i = \mathbf{z}_i + \sigma_i^\phi \epsilon, \epsilon \in \mathcal{N}(\mathbf{0}, \mathbf{I})$. Note that the uncertainty predictions vary for different voxel i.

Data Uncertainty Learning (DUL) [Chang *et al.*, 2020] shares a similar spirit with HAU with two distinctions. Instead of using the logit $\hat{\mathbf{z}}_i$, DUL models the feature $\hat{\mathbf{v}}_i$ as a Gaussian distribution by $\hat{\mathbf{v}}_i = \mathbf{v}_i + \boldsymbol{\sigma}_i^{\mathbf{W}} \epsilon$. Moreover, DUL introduces a regularization term in the loss function that minimizes the Kullback–Leibler (KL) divergence between the predicted Gaussian and a standard Gaussian.

MC Dropout (MCD) [Jungo and Reyes, 2019] is proposed to explore the epistemic (model) uncertainty. Differently from the above methods, MCD does not require additional parameters to learn uncertainty. Instead, it incorporates multiple dropout layers into the original network during training. For inference, the occupancy prediction of each voxel is obtained by $\hat{\mathbf{z}}_i = \frac{1}{K}\sum_{k=1}^{K} \mathbf{z}_{k,i}$, where $\mathbf{z}_{k,i}$ is the model output at the k-th test. The normalized entropy of K predictions is adopted as the model uncertainty. To fully explore the uncertainty within the model, we set $K = 40$ in our experiments.

For above online uncertainty learning methods, the calibrated confidence is set as the *softmax* output of the sampled

[1] We omit predicting the mean $\mu^\phi(\mathbf{z}_i)$ and use $\mathbf{z}_i$ for simplicity. Empirical results are similar.

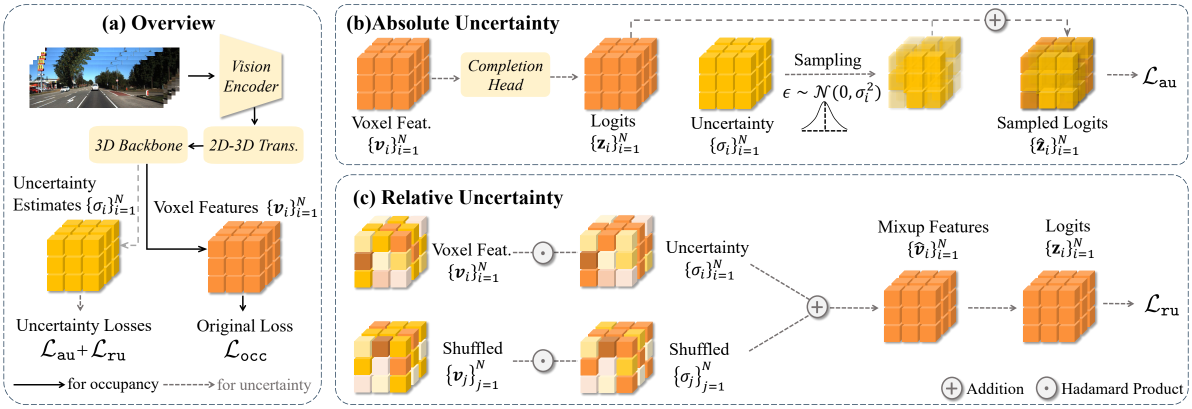

Figure 1: (a) *Overview of proposed* RELIOCC. Besides the original objective of an occupancy network, we introduce an uncertainty estimation branch and supervise it with absolute and relative uncertainty learning losses. (b) *Absolute uncertainty learning*. Deterministic logits are replaced with ones sampled from predicted distributions. (c) *Relative uncertainty learning*. We leverage the relative relationships between uncertainty pairs to further enhance uncertainty learning.

logit and then determined by taking the maximum probability across all classes: $c_i = \max_s \mathcal{S}(\hat{\mathbf{z}}_i)^{(s)}$, where $\mathcal{S}$ denotes the *softmax* function.

4.2 Offline Model Calibration

Offline calibration methods build on pre-trained V_θ and need to learn a scaling function, which typically employ the following formulation

$$c_i \equiv c_{f|\phi}(\mathbf{z}_i) = \max_s f_\phi(\mathbf{z}_i)^{(s)} , \tag{4}$$

where f_ϕ is the scaling function applied to $\mathbf{z}_i$ parameterized by the learnable parameters ϕ. In the absence of explicit uncertainty estimation, uncertainty σ_i is set to $1 - c_i$.

Temperature Scaling (TempS) [Guo *et al.*, 2017] employs a scalar parameter T, termed as temperature, to scale the logits $\mathbf{z_i}$, by $f_\phi(\mathbf{z_i}) = \mathcal{S}(\frac{\mathbf{z_i}}{T})$. T is data-independent, which is shared across all classes.

Dirichlet Scaling (DiriS) [Kull *et al.*, 2019] assumes that the model's output follows a Dirichlet distribution. Based on this assumption, they propose the Dirichlet scaling, $f_\phi(\mathbf{z_i}) = \mathcal{S}(\mathbf{W} \cdot \log(\mathcal{S}(\mathbf{z_i})) + \mathbf{b})$. Here, learnable parameters ϕ include weight $\mathbf{W} \in \mathbb{R}^{(S+1)\times(S+1)}$ and bias $\mathbf{b} \in \mathbb{R}^{S+1}$.

Meta-Calibration (MetaC) [Ma and Blaschko, 2021] proposes to employ the entropy of model prediction $-c_i \log(c_i)$ to select different calibrators. Specifically, an identical $f_\phi(\mathbf{z_i}) = \mathcal{S}(\frac{\mathbf{z_i}}{T})$ as in TempS is used when $-c_i \log(c_i)$ is smaller than the predefined threshold η. Otherwise, the calibration function $f_\phi(\mathbf{z_i})$ is set to the constant value $\frac{1}{S+1}$. MetaC introduces new randomness into predictions, which lead to variations in accuracy, making it less practical for safety-critical tasks such as occupancy prediction.

Depth-Aware Scaling (DeptS) [Ma and Blaschko, 2021] is an improved variant upon MetaC, which is specially designed for LiDAR segmentation. Depth d_i of each point or voxel is encoded into the calibration function f_ϕ by a linear mapping $\alpha_i = k_1 \cdot d_i + k_2$, where k_1 and k_2 are learnable parameters. When prediction entropy $-c_i \log(c_i)$ is greater than the threshold η, $f_\phi(\mathbf{z_i}) = \mathcal{S}(\frac{\mathbf{z_i}}{\alpha \cdot T_1})$. Otherwise, $f_\phi(\mathbf{z_i}) = \mathcal{S}(\frac{\mathbf{z_i}}{\alpha \cdot T_2})$ Both T_1 and T_2 are temperature parameters, where T_1 is initially set higher than T_2.

5 RELIOCC

Method Overview. Inspired by investigation on prior arts, we propose RELIOCC, a plug-and-play method tailored for the 3D semantic occupancy prediction task. RELIOCC has two main improvements over existing methods. Firstly, beyond traditional uncertainty learning, we propose to utilize the relative relationships between voxel pairs of uncertainty estimates in order to further refine the uncertainty estimation process. Secondly, we present a unified framework that integrates the methodologies of uncertainty learning with scaling-based calibration, which demonstrates that their synergy offers substantial benefits. As shown in Figure 1(a), we predict a scalar uncertainty σ_i from a voxel feature $\boldsymbol{v_i}$ using an MLP, which receives supervision from both the individual and relative voxels.

Absolute Uncertainty. By *absolute* here we mean the uncertainty σ_i is only determined by the individual $\boldsymbol{v}_i$ itself while ignoring relative relations between a pair of $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. We adopt a similar formulation as in HAU [Kendall and Gal, 2017]. Specifically, We randomly sample the logit $\hat{\mathbf{z}}_i$ based on the predicted uncertainty σ_i by $\hat{\mathbf{z}}_i = \mathbf{z}_i + \sigma_i \epsilon, \epsilon \in \mathcal{N}(\mathbf{0}, \mathbf{I})$ as illustrated in Figure 1(b). We denote this absolute uncertainty loss as $\mathcal{L}_{\text{au}}$, which is computed with the re-sampled logit $\hat{\mathbf{z}}_i$ and its corresponding ground truth.

Relative Uncertainty. A potential drawback of absolute uncertainty is that the optimization of σ_i tends to plateau once it reaches a small scale. To address this issue, we introduce the concept of relative uncertainty learning for occupancy prediction. The fundamental principle of relative uncertainty learning involves enforcing comparisons between uncertainty pairs $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. This approach ensures that optimization does not plateau, even when σ_i and σ_j are small.

Concretely, we shuffle voxel features in $\mathcal{V}$, paired the shuffled features with the original ones and obtain random pairs $(\boldsymbol{v}_i, \boldsymbol{v}_j)$ at each iteration. Inspired by the mix-up [Zhang *et al.*, 2018] learning principle, we blend the paired voxel fea-

Method	Modality	Semantics			Geometry		
		mIoU (%)↑	$\mathtt{PRR}_{\mathtt{sem}}$(%)↑	$\mathtt{ECE}_{\mathtt{sem}}$(%)↓	IoU(%)↑	$\mathtt{PRR}_{\mathtt{geo}}$(%)↑	$\mathtt{ECE}_{\mathtt{geo}}$(%)↓
SSCNet [CVPR17] [Song *et al.*, 2017]	LiDAR	16.41	46.77	1.61	50.75	42.92	0.97
LMSCNet [3DV20] [2020]	LiDAR	17.27	**48.89**	**0.79**	**54.91**	**48.01**	**0.67**
JS3C-Net [AAAI21] [Song *et al.*, 2017]	LiDAR	22.77	41.09	2.94	53.08	37.04	1.64
SSC-RS [IROS23] [Mei *et al.*, 2023]	LiDAR	24.75	45.04	0.87	58.62	44.29	0.72
SCPNet* [CVPR23] [Xia *et al.*, 2023]	LiDAR	**35.06**	38.35	2.52	49.06	-	-
MonoScene [CVPR22] [2022]	Camera	11.30	41.95	6.65	36.79	38.39	5.95
TPVFormer [CVPR23] [Huang *et al.*, 2023]	Camera	11.30	38.83	7.10	35.62	32.10	6.32
NDCScene [ICCV23] [Yao *et al.*, 2023]	Camera	12.70	43.29	7.24	37.24	**40.17**	6.45
VoxFormer [CVPR23] [Li *et al.*, 2023]	Camera	13.17	42.97	5.90	43.96	36.56	5.02
SGN [TIP24] [Mei *et al.*, 2024]	Camera	**15.52**	**44.72**	**5.69**	**45.45**	39.78	**4.85**

Table 1: The accuracy and reliability evaluation of state-of-the-art semantic occupancy prediction models on the `validation set` of SemanticKITTI. $*$ indicates that the output of SCPNet is a sparse representation and does not contain confidence score for empty voxels, making it infeasible to evaluate the corresponding geometric metrics in reliability.

tures with $\hat{\boldsymbol{v}} = \lambda \boldsymbol{v}_i + (1-\lambda)\boldsymbol{v}_j$. Correspondingly $\hat{\boldsymbol{v}}$ is trained with a blend of the label pairs using cross-entropy loss. The blended label $y = y_i + y_j$, where y_i, y_j are one-hot label encodings[2]. We employ the predicted uncertainty for the weighting: $\lambda = \frac{\sigma_i}{\sigma_i+\sigma_j} \in [0,1]$. We denote the loss computed with blended label and mixed output from the shared completion head as the relative uncertainty loss $\mathcal{L}_{\mathtt{ru}}$, as shown in Figure 1(c).

An intuitive understanding emerges when considering that relative uncertainty adaptively modulates the learning dynamics between the feature pair $(\boldsymbol{v}_i, \boldsymbol{v}_j)$. Specifically, if the model exhibits greater confidence in the prediction associated with $\boldsymbol{v}_i$, it suffices for the mixup feature to incorporate a smaller portion of $\boldsymbol{v}_i$ while still achieving a reduced loss $\mathcal{L}_{\mathtt{ru}}$. In contrast, a lower confidence in $\boldsymbol{v}_j$ necessitates a greater inclusion of $\boldsymbol{v}_j$ within the mixup to diminish the loss. Consequently, this process enables the model to effectively differentiate between the uncertainties σ_i and σ_j, typically resulting in a smaller σ_i and a larger σ_j. Importantly, this differentiation does not hinge on the absolute magnitudes of $\sigma_{i,j}$. Rather, it is the relative relationship between them that is central to the learning process. In driving scenarios, there are rich relative relationships between voxels, including *distance*, *occupancy*, *surface* and *interior* properties. This focus on relative differences ensures that the model's adjustments are robust to the absolute scales of the uncertainties.

Uncertainty-Aware Calibration. Using the above uncertainty estimation objectives, RELIOCC is capable of learning uncertainty with existing occupancy models in an online setting. We introduce a scaling-based calibration objective to make it also compatible with the offline one. A variant form of TempS [Guo *et al.*, 2017] is adopted, and the uncertainty-aware temperature T_σ is a linear transform of σ_i:

$$T_\sigma = k_1 \cdot \sigma_i + k_2, \quad f_\phi(\mathbf{z_i}) = \mathcal{S}\left(\frac{\mathbf{W}\cdot\mathbf{z_i}+\mathbf{b}}{T_\sigma}\right), \qquad (5)$$

where k_1 and k_2 are learnable parameters, and $\mathbf{b}$ is the bias. $\mathbf{W}$ is initialized as the identity matrix and only the elements on the diagonal are optimized. The calibration loss is denoted as $\mathcal{L}_{\mathtt{calib}}$.

Training and Inference. RELIOCC supports both online uncertainty learning and offline model calibration settings. In the online setting, the uncertainty predictor is trained concurrently with the occupancy network from scratch. The total loss function comprises $\mathcal{L}_{\mathtt{occ}}$, $\mathcal{L}_{\mathtt{au}}$, and $\mathcal{L}_{\mathtt{ru}}$. Here, $\mathcal{L}_{\mathtt{occ}}$ represents the primary loss for the occupancy network. During inference, the model operates consistently with the original network design. In the offline setting, the occupancy network is well trained and frozen, eliminating the need for $\mathcal{L}_{\mathtt{occ}}$ and introducing the calibration loss $\mathcal{L}_{\mathtt{calib}}$ instead. The inference process incurs a minimal increase in computational overhead due to the addition of the calibrator.

[2]Different from the original mix-up [Zhang *et al.*, 2018] paper, we omit weighting the labels with λ for stable training.

6 Experiments

6.1 Benchmark Results

Datasets and Evaluation. SemanticKITTI [Behley *et al.*, 2019] is the first large-scale outdoor dataset for semantic occupancy prediction containing 64-beam LiDAR scans and front camera images as inputs [Geiger *et al.*, 2012]. The dataset comprises 22 sequences, where 00-10 (excluding 08) are used as the training set, 08 is the validation set, and 11-21 are the test set. Since the ground truth for the test set is not publicly available, we cannot measure our newly introduced metrics on it. Therefore, we primarily evaluate the existing methods on the validation set (`val.`). As described in §3.2, mIoU and IoU are used to measure the model's accuracy in semantic and geometric completion, respectively. For misclassification detection and calibration metrics including `PRR` and `ECE`, we also report the corresponding results from both geometric and semantic perspectives.

Re-evaluated Methods. We reproduce and evaluate existing publicly available methods on the SemanticKITTI benchmark, including five LiDAR-based models [Song *et al.*, 2017; Roldao *et al.*, 2020; Yan *et al.*, 2021; Mei *et al.*, 2023; Xia *et al.*, 2023] and five camera-based models [Cao and de Charette, 2022; Huang *et al.*, 2023; Li *et al.*, 2023; Yao *et al.*, 2023; Mei *et al.*, 2024]. All results are obtained using the official implementation and the configurations are kept consistent for inference. As shown in Tab. 1, we find that although the accuracy of camera-based methods has been continuously improved and gradually approaches the baseline accuracy of LiDAR methods, their reliability metrics, particularly the `ECE`, have not shown corresponding improvements. In cases of lower accuracy compared to LiDAR, the camera-based models' reliability is also quite poor, which undoubtedly poses significant safety risks for autonomous driving.

Method	Semantics			Geometry		
	mIoU (%)↑	PRR$_{sem}$(%)↑	ECE$_{sem}$(%)↓	IoU(%)↑	PRR$_{geo}$(%)↑	ECE$_{geo}$(%)↓
VoxFormer Framework						
VoxFormer [CVPR23] [Li *et al.*, 2023]	13.17	42.97	5.90	43.96	36.56	5.02
VoxFormer+HAU [NIPS17] [Kendall and Gal, 2017]	**13.43**	45.38	5.26	43.57	40.72	4.47
VoxFormer+DUL [CVPR20] [Chang *et al.*, 2020]	13.29	43.57	6.09	**44.10**	38.66	5.17
VoxFormer+MCD [MICCAI19] [Jungo and Reyes, 2019]	13.28	42.21	5.83	43.90	37.43	4.99
VoxFormer+RELIOCC (Ours)	**13.43**	**47.75**	**2.84**	43.28	**44.58**	**2.57**
SGN Framework						
SGN [TIP24] [Mei *et al.*, 2024]	15.52	44.72	5.69	45.45	39.78	4.85
SGN+HAU [NIPS17] [Kendall and Gal, 2017]	15.50	46.51	5.08	45.07	44.24	4.34
SGN+DUL [CVPR20] [Chang *et al.*, 2020]	**15.81**	44.00	5.78	45.75	39.56	4.95
SGN+MCD [MICCAI19] [Jungo and Reyes, 2019]	15.62	44.70	6.02	45.50	40.34	5.11
SGN+RELIOCC (Ours)	15.65	**50.72**	**3.75**	**45.78**	**49.61**	**3.07**

Table 2: Quantitative results of online uncertainty learning (§6.2) on SemanticKITTI (`validation set`).

Method	Semantics			Geometry		
	mIoU(%)↑	PRR$_{sem}$(%)↑	ECE$_{sem}$(%)↓	IoU(%)↑	PRR$_{geo}$(%)↑	ECE$_{geo}$(%)↓
VoxFormer Framework						
VoxFormer [CVPR23] [Li *et al.*, 2023]	13.17	42.97	5.90	43.96	36.56	5.02
VoxFormer+TempS [ICML17] [Guo *et al.*, 2017]	13.17	43.63	2.61	43.96	33.59	2.28
VoxFormer+DiriS [NeurIPS19] [Kull *et al.*, 2019]	13.17	48.12	2.38	43.96	42.78	2.42
VoxFormer+MetaC [ICML21] [Ma and Blaschko, 2021]	11.86	43.06	4.11	34.73	34.80	3.67
VoxFormer+DeptS [arXiv24] [Kong *et al.*, 2025]	13.17	41.29	2.27	43.96	30.31	**1.63**
VoxFormer+RELIOCC (Ours)	13.17	**48.17**	**2.05**	43.96	**44.34**	2.57
SGN Framework						
SGN [TIP24] [Mei *et al.*, 2024]	15.52	44.72	5.69	45.45	39.78	4.85
SGN+TempS [ICML17] [Guo *et al.*, 2017]	15.52	46.90	2.68	45.45	37.25	2.35
SGN+DiriS [NeurIPS19] [Kull *et al.*, 2019]	15.52	**48.20**	2.61	45.45	43.04	2.51
SGN+MetaC [ICML21] [Ma and Blaschko, 2021]	14.71	46.38	4.06	40.37	37.97	3.65
SGN+DeptS [WACV25] [Kong *et al.*, 2025]	15.52	45.45	2.14	45.45	34.99	**1.42**
SGN+RELIOCC (Ours)	15.52	47.40	**2.09**	45.45	**43.80**	2.43

Table 3: Quantitative results of offline model calibration (§6.3) on SemanticKITTI (`validation set`).

6.2 Online Uncertainty Learning

Base Architectures and Competing Methods. Considering the potential applications of camera-based methods and their current limitations, we adopt the state-of-the-art vision-based methods including VoxFormer [Li *et al.*, 2023] and SGN [Mei *et al.*, 2024] as our base architectures to conduct relevant experiments. We report the results of some existing methods on the two baseline frameworks for comparison, including HAU, DUL, and MCD (see §4.1).

Implementation Details. We follow the original training setting and add additional uncertainty learning parameters without altering the network structure. The inputs consist of the current image from the left camera and previous 4 frames. The image size is copped into 1220×370. VoxFormer [Li *et al.*, 2023] and SGN [Mei *et al.*, 2024] with online uncertainty estimation are trained for 20 epochs and 40 epochs, respectively. The loss coefficients (α, β) for $\mathcal{L}_{\texttt{au}}$ and $\mathcal{L}_{\texttt{ru}}$ are set to (4.0, 6.0) for both frameworks.

Evaluation Results. In Tab. 2, we provide the comparison among the methods with the same framework for fairness. Compared to data uncertainty-based HAU and DUL, as well as model uncertainty-based MCD, our method shows significant improvements in the new evaluation metrics for reliability. The calibration errors (`ECE`) in both semantic and geometric aspects are significantly reduced compared to the existing uncertainty estimation methods. The improvement in `PRR` also indicates a notable enhancement in model reliability. Our method maintains stability in terms of the original accuracy (mIoU and IoU) across the two different frameworks and even surpasses LiDAR-based methods in `PRR`. Furthermore, we visualize the error map with corresponding uncertainty map of SGN and our approach. The uncertainty for vanilla SGN is obtained by subtracting the confidence from 1. As shown in Figure 2, when the network's predictions exhibit large areas of error, SGN's uncertainty map still shows low uncertainty, indicating high confidence in prediction. In contrast, our proposed RELIOCC displays high uncertainty in most of the error regions, providing more reliable information for downstream tasks.

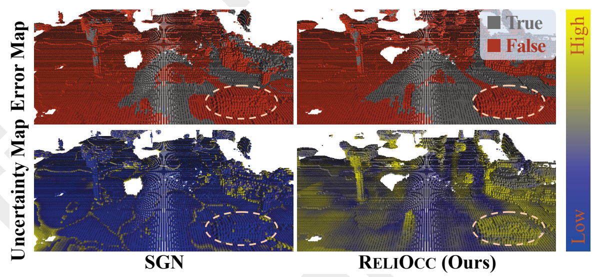

Figure 2: Visual results of the error map and uncertainty map from the prediction by SGN and RELIOCC. In the uncertainty map, a closer proximity to *yellow* indicates a higher level of uncertainty.

6.3 Offline Model Calibration

In this section, VoxFormer and SGN are also adopted as baseline frameworks. We primarily compare our method with scaling-based model calibration approaches including TempS, DiriS, MetaC, and DeptS (see §4.2).

Implementation Details. We select the best-performing

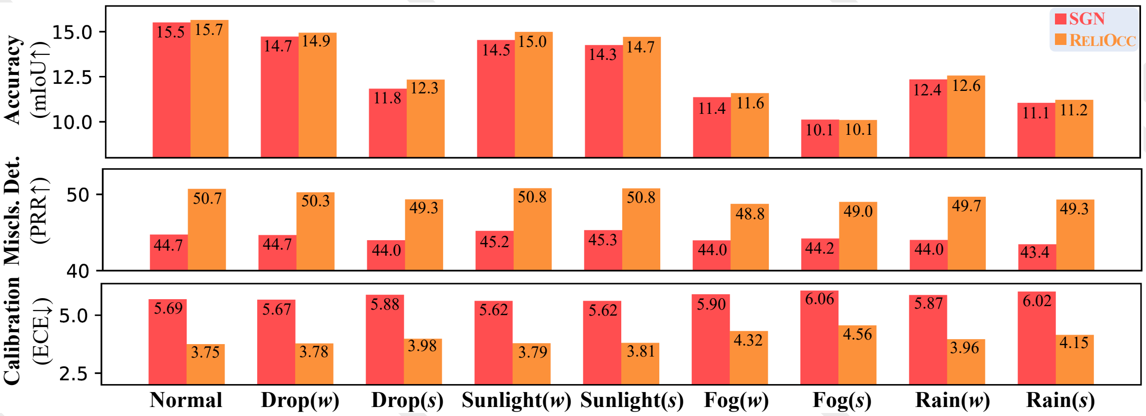

Figure 3: The comparison of accuracy and reliability performance between SGN and RELIOCC under four out-of-domain conditions.

Absolute Unc.	Relative Unc.	$\text{PRR}_{\text{sem}}\uparrow$	$\text{ECE}_{\text{sem}}\downarrow$	$\text{PRR}_{\text{geo}}\uparrow$	$\text{ECE}_{\text{geo}}\downarrow$
		42.97	5.90	36.56	5.02
✓		45.47	5.41	41.54	4.62
	✓	46.58	4.02	42.65	4.31
✓	✓	**47.75**	**2.84**	**44.58**	**2.57**

Table 4: Ablation of our online uncertainty learning.

Scaling Calib.	Relative Unc.	Absolute Unc.	$\text{PRR}_{\text{sem}}\uparrow$	$\text{ECE}_{\text{sem}}\downarrow$	$\text{PRR}_{\text{geo}}\uparrow$	$\text{ECE}_{\text{geo}}\downarrow$
			42.97	5.90	36.56	5.02
✓			43.63	2.61	33.59	2.28
✓	✓		45.13	**1.75**	39.66	**2.19**
✓	✓	✓	**48.17**	2.05	**44.34**	2.57

Table 5: Ablation of our offline model calibration.

model checkpoints on the validation set from the pre-trained VoxFormer and SGN as the targets for calibration. During the calibration process, the parameters of the original network are frozen, and only the parameters ϕ in the calibration function f_ϕ and uncertainty learning layers are trainable. For both frameworks, these methods are trained on 8 GPUs for 20 epochs with a learning rate as 0.001 and AdamW optimizer [Zhao *et al.*, 2022]. The batch size is set to 1 per GPU. For our method, the loss weights (α, β, γ) for uncertainty learning $(\mathcal{L}_{\texttt{au}}, \mathcal{L}_{\texttt{ru}})$ and model calibration $(\mathcal{L}_{\texttt{calib}})$ are set to 1.5, 1.0, and 4.0, respectively.

Evaluation Results. As illustrated in Tab. 3, all calibration methods demonstrate improvements compared to the baselines, particularly in calibration error (ECE). MetaC [Ma and Blaschko, 2021] loses the characteristic of maintaining accuracy in calibration due to the introduction of new random classifications. Our approach with uncertainty-aware design achieves competitive performance on both ECE and PRR metrics without depth information even compared with the state-of-the-art DeptS [Kong *et al.*, 2025].

6.4 Diagnostic Experiments

Ablation of Online Uncertainty Learning. We provide ablation experiments on the effect of absolute uncertainty and relative uncertainty during the whole model training. The experiments are conducted with VoxFormer [Li *et al.*, 2023] on the validation set of SemanticKITTI. As shown in Tab. 4, the first row presents the baseline results. The inclusion of individual absolute uncertainty and relative one both contribute to the improvement of the model's reliability, albeit with a modest enhancement. When our proposed hybrid uncertainty learning module is incorporated, the PRR and ECE metrics of model's prediction achieve the best results.

Ablation of Offline Model Calibration. Further ablations are also conducted in offline mode. With the pre-trained VoxFormer, we found that employing standard scaling strategies such as TempS can achieve good calibration results as illustrated in second row of Tab. 5. However, it impacts the improvement of misclassification detection metrics (PRR) and even leads to a decline in geometry. Our introduced relative uncertainty learning can further improve calibration performance and enhance misclassification detection. Moreover, the combination of absolute and relative uncertainties achieves the best performance in misclassification detection, although it is marginally less effective in calibration due to the distinct focus of the PRR and ECE metrics.

Robustness Analysis. Figure 3 presents the robustness analysis results of RELIOCC compared to the baseline model SGN [Mei *et al.*, 2024]. We simulate four potential out-of-domain scenarios during the inference, including sensor failures (frames drop), strong sunlight, foggy and rainy conditions, to evaluate the model's robustness [Dong *et al.*, 2023]. Each adverse scenario provides *weak*(w) and *strong*(s) modes of perturbation. As the noise increases in various conditions, our method not only maintains stability in reliability metrics but also demonstrates more improvement in accuracy compared to the baseline.

7 Conclusion

In this paper, we address the issue of assessing reliability in semantic occupancy prediction for the first time. The reliability is evaluated from two aspects including misclassification detection and calibration. Extensive evaluation on existing LiDAR and camera-based methods is provided. Besides, we propose a new scheme RELIOCC that integrates hybrid uncertainty from the individual and relative voxels into existing occupancy networks without affecting accuracy or inference speed. Both online and offline modes are designed to illustrate the generalization capability of our learned uncertainty. Extensive experiments are conducted under various settings, demonstrating RELIOCC is effective in improving the reliability and robustness of semantic occupancy models.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No. 62376244) and Program of China Scholarship Council (Grant No. 202406320418). It is also supported by Information Technology Center and State Key Lab of CAD&CG, Zhejiang University.

References

[Albrecht *et al.*, 2021] Stefano V Albrecht, Cillian Brewitt, John Wilhelm, Balint Gyevnar, Francisco Eiras, Mihai Dobre, and Subramanian Ramamoorthy. Interpretable goal-based prediction and planning for autonomous driving. In *ICRA*, 2021.

[Behley *et al.*, 2019] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. In *ICCV*, 2019.

[Breiman, 1996] Leo Breiman. Bagging predictors. *Machine learning*, 1996.

[Cai *et al.*, 2023] Kaiwen Cai, Chris Xiaoxuan Lu, and Xiaowei Huang. Uncertainty estimation for 3d dense prediction via cross-point embeddings. *IEEE RA-L*, 2023.

[Cao and de Charette, 2022] Anh-Quan Cao and Raoul de Charette. MonoScene: Monocular 3D semantic scene completion. In *CVPR*, 2022.

[Chang *et al.*, 2020] Jie Chang, Zhonghao Lan, Changmao Cheng, and Yichen Wei. Data uncertainty learning in face recognition. In *CVPR*, 2020.

[Cheng *et al.*, 2021] Ran Cheng, Christopher Agia, Yuan Ren, Xinhai Li, and Liu Bingbing. S3CNet: A sparse semantic scene completion network for LiDAR point cloud. In *CoRL*, 2021.

[de Jorge *et al.*, 2023] Pau de Jorge, Riccardo Volpi, Philip HS Torr, and Grégory Rogez. Reliability in semantic segmentation: Are we on the right track? In *CVPR*, 2023.

[Denker and LeCun, 1990] John Denker and Yann LeCun. Transforming neural-net output levels to probability distributions. In *NIPS*, 1990.

[Dong *et al.*, 2023] Yinpeng Dong, Caixin Kang, Jinlai Zhang, Zijian Zhu, Yikai Wang, Xiao Yang, Hang Su, Xingxing Wei, and Jun Zhu. Benchmarking robustness of 3d object detection to common corruptions. In *CVPR*, 2023.

[Fumera and Roli, 2002] Giorgio Fumera and Fabio Roli. Support vector machines with embedded reject option. In *Pattern Recognition with Support Vector Machines: First International Workshop*, 2002.

[Gal and Ghahramani, 2016] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *ICML*, 2016.

[Geiger *et al.*, 2012] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *CVPR*, 2012.

[Gong *et al.*, 2017] Sixue Gong, Vishnu Naresh Boddeti, and Anil K Jain. On the capacity of face representation. *arXiv preprint arXiv:1709.10433*, 2017.

[Guo *et al.*, 2017] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *ICML*, 2017.

[Hendrycks *et al.*, 2021] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *CVPR*, 2021.

[Hu *et al.*, 2023] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *CVPR*, 2023.

[Huang *et al.*, 2023] Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu. Tri-perspective view for vision-based 3D semantic occupancy prediction. In *CVPR*, 2023.

[Hüllermeier and Waegeman, 2021] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 2021.

[Johnson, 2001] Roger W Johnson. An introduction to the bootstrap. *Teaching statistics*, 2001.

[Jungo and Reyes, 2019] Alain Jungo and Mauricio Reyes. Assessing reliability and challenges of uncertainty estimations for medical image segmentation. In *MICCAI*, 2019.

[Kassapis *et al.*, 2021] Elias Kassapis, Georgi Dikov, Deepak K. Gupta, and Cedric Nugteren. Calibrated adversarial refinement for stochastic semantic segmentation. In *ICCV*, 2021.

[Kendall and Gal, 2017] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NIPS*, 2017.

[Kendall *et al.*, 2018] Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *CVPR*, 2018.

[Khan *et al.*, 2019] Salman Khan, Munawar Hayat, Syed Waqas Zamir, Jianbing Shen, and Ling Shao. Striking the right balance with uncertainty. In *CVPR*, 2019.

[Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[Kong *et al.*, 2025] Lingdong Kong, Xiang Xu, Jun Cen, Wenwei Zhang, Liang Pan, Kai Chen, and Ziwei Liu. Calib3d: Calibrating model preferences for reliable 3d scene understanding. In *WACV*, 2025.

[Kull *et al.*, 2019] Meelis Kull, Miquel Perello Nieto, Markus Kängsepp, Telmo Silva Filho, Hao Song, and Peter Flach. Beyond temperature scaling: Obtaining well-calibrated multi-class probabilities with dirichlet calibration. In *NeurIPS*, 2019.

[Kuppers *et al.*, 2020] Fabian Kuppers, Jan Kronenberger, Amirhossein Shantia, and Anselm Haselhoff. Multivariate confidence calibration for object detection. In *CVPRW*, 2020.

[Lakshminarayanan *et al.*, 2017] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NIPS*, 2017.

[Li *et al.*, 2023] Yiming Li, Zhiding Yu, Christopher Choy, Chaowei Xiao, Jose M Alvarez, Sanja Fidler, Chen Feng, and Anima Anandkumar. VoxFormer: Sparse voxel transformer for camera-based 3D semantic scene completion. In *CVPR*, 2023.

[Li *et al.*, 2024] Bohan Li, Yasheng Sun, Zhujin Liang, Dalong Du, Zhuanghui Zhang, Xiaofeng Wang, Yunnan Wang, Xin Jin, and Wenjun Zeng. Bridging stereo geometry and bev representation with reliable mutual interaction for semantic scene completion. In *IJCAI*, 2024.

[Ma and Blaschko, 2021] Xingchen Ma and Matthew B. Blaschko. Meta-cal: Well-controlled post-hoc calibration by ranking. In *ICML*, 2021.

[Malinin *et al.*, 2019] Andrey Malinin, Bruno Mlodozeniec, and Mark Gales. Ensemble distribution distillation. In *ICLR*, 2019.

[Mei *et al.*, 2023] Jianbiao Mei, Yu Yang, Mengmeng Wang, Tianxin Huang, Xuemeng Yang, and Yong Liu. SSC-RS: Elevate LiDAR semantic scene completion with representation separation and BEV fusion. In *IROS*, 2023.

[Mei *et al.*, 2024] Jianbiao Mei, Yu Yang, Mengmeng Wang, Junyu Zhu, Jongwon Ra, Yukai Ma, Laijian Li, and Yong Liu. Camera-based 3d semantic scene completion with sparse guidance network. *IEEE TIP*, 2024.

[Munir *et al.*, 2022] Muhammad Akhtar Munir, Muhammad Haris Khan, M. Sarfraz, and Mohsen Ali. Towards improving calibration in object detection under domain shift. In *NeurIPS*, 2022.

[Naeini *et al.*, 2015] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. In *AAAI*, 2015.

[Niculescu-Mizil and Caruana, 2005] Alexandru Niculescu-Mizil and Rich Caruana. Predicting good probabilities with supervised learning. In *ICML*, 2005.

[Roldao *et al.*, 2020] Luis Roldao, Raoul de Charette, and Anne Verroust-Blondet. LMSCNet: Lightweight multiscale 3D semantic completion. In *3DV*, 2020.

[Song *et al.*, 2017] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. Semantic scene completion from a single depth image. In *CVPR*, 2017.

[Tang *et al.*, 2022] Jiaxiang Tang, Xiaokang Chen, Jingbo Wang, and Gang Zeng. Not all voxels are equal: Semantic scene completion from the point-voxel perspective. In *AAAI*, 2022.

[Tian *et al.*, 2024] Xiaoyu Tian, Tao Jiang, Longfei Yun, Yucheng Mao, Huitong Yang, Yue Wang, Yilun Wang, and Hang Zhao. Occ3D: A large-scale 3D occupancy prediction benchmark for autonomous driving. In *NeurIPS*, 2024.

[Tishby and Solla, 1989] Levin Tishby and Solla. Consistent inference of probabilities in layered networks: predictions and generalizations. In *IJCNN*, 1989.

[Wang *et al.*, 2023] Xiaofeng Wang, Zheng Zhu, Wenbo Xu, Yunpeng Zhang, Yi Wei, Xu Chi, Yun Ye, Dalong Du, Jiwen Lu, and Xingang Wang. OpenOccupancy: A large scale benchmark for surrounding semantic occupancy perception. In *ICCV*, 2023.

[Wang *et al.*, 2024] Song Wang, Jiawei Yu, Wentong Li, Wenyu Liu, Xiaolu Liu, Junbo Chen, and Jianke Zhu. Not all voxels are equal: Hardness-aware semantic scene completion with self-distillation. In *CVPR*, 2024.

[Wei *et al.*, 2023] Yi Wei, Linqing Zhao, Wenzhao Zheng, Zheng Zhu, Jie Zhou, and Jiwen Lu. SurroundOcc: Multi-camera 3D occupancy prediction for autonomous driving. In *ICCV*, 2023.

[Wilson *et al.*, 2022] Joey Wilson, Jingyu Song, Yuewei Fu, Arthur Zhang, Andrew Capodieci, Paramsothy Jayakumar, Kira Barton, and Maani Ghaffari. Motionsc: Data set and network for real-time semantic mapping in dynamic environments. *IEEE RA-L*, 2022.

[Xia *et al.*, 2023] Zhaoyang Xia, Youquan Liu, Xin Li, Xinge Zhu, Yuexin Ma, Yikang Li, Yuenan Hou, and Yu Qiao. SCPNet: Semantic scene completion on point cloud. In *CVPR*, 2023.

[Yan *et al.*, 2021] Xu Yan, Jiantao Gao, Jie Li, Ruimao Zhang, Zhen Li, Rui Huang, and Shuguang Cui. Sparse single sweep LiDAR point cloud segmentation via learning contextual shape priors from scene completion. In *AAAI*, 2021.

[Yao *et al.*, 2023] Jiawei Yao, Chuming Li, Keqiang Sun, Yingjie Cai, Hao Li, Wanli Ouyang, and Hongsheng Li. Ndc-scene: Boost monocular 3d semantic scene completion in normalized device coordinates space. In *ICCV*, 2023.

[Zhang *et al.*, 2018] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018.

[Zhang *et al.*, 2021] Yuhang Zhang, Chengrui Wang, and Weihong Deng. Relative uncertainty learning for facial expression recognition. In *NeurIPS*, 2021.

[Zhao *et al.*, 2022] Borui Zhao, Quan Cui, Renjie Song, Yiyu Qiu, and Jiajun Liang. Decoupled knowledge distillation. In *CVPR*, 2022.

[Zheng *et al.*, 2024] Wenzhao Zheng, Weiliang Chen, Yuanhui Huang, Borui Zhang, Yueqi Duan, and Jiwen Lu. OccWorld: Learning a 3D occupancy world model for autonomous driving. In *ECCV*, 2024.