

# Evidential Semantic Mapping in Off-road Environments with Uncertainty-aware Bayesian Kernel Inference

Junyoung Kim<sup>\*1</sup>, Junwon Seo<sup>\*2</sup>, Jihong Min<sup>1</sup>

**Abstract**—Robotic mapping with Bayesian Kernel Inference (BKI) has shown promise in creating semantic maps by effectively leveraging local spatial information. However, existing semantic mapping methods face challenges in constructing reliable maps in unstructured outdoor scenarios due to unreliable semantic predictions. To address this issue, we propose an evidential semantic mapping, which can enhance reliability in perceptually challenging off-road environments. We integrate Evidential Deep Learning into the semantic segmentation network to obtain the uncertainty estimate of semantic prediction. Subsequently, this semantic uncertainty is incorporated into an uncertainty-aware BKI, tailored to prioritize more confident semantic predictions when accumulating semantic information. By adaptively handling semantic uncertainties, the proposed framework constructs robust representations of the surroundings even in previously unseen environments. Comprehensive experiments across various off-road datasets demonstrate that our framework enhances accuracy and robustness, consistently outperforming existing methods in scenes with high perceptual uncertainties.

## I. INTRODUCTION

Robotic mapping, which aims to construct an explicit environmental model using sensor measurements, is an essential component in various robotic systems. Conventional approaches, such as occupancy grid mapping [1], [2], construct 3D voxel maps by estimating the occupancy state of each voxel using sensor measurements intersecting with the voxel. However, due to their independent voxel assumptions, these approaches often produce discontinuous and sparse maps, especially when dealing with sparse sensor data like LiDAR [3]. Consequently, recent works have focused on enhancing the geometric completeness of occupancy estimation by leveraging local spatial information [3], [4]. Specifically, the adoption of Bayesian Kernel Inference (BKI) [5] has enabled the efficient utilization of neighboring measurements, resulting in more reliable continuous maps [4].

Semantic mapping has significantly improved the ability of autonomous systems to comprehend their surroundings across various robotic applications [6]–[12]. By estimating voxel-wise semantic states, it provides a richer understanding of the environments. Building upon the advancements in continuous mapping, S-BKI [13] extended BKI into

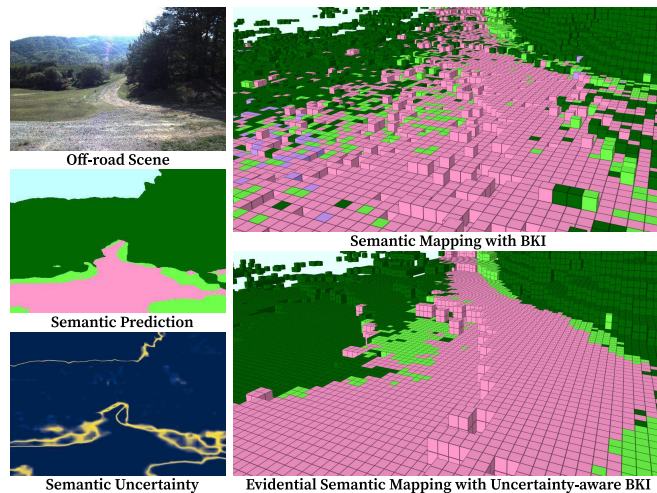


Fig. 1: In perceptually challenging unstructured outdoor environments, considering semantic uncertainty from the segmentation network is beneficial to produce more accurate semantic maps by prioritizing confident predictions. By incorporating the semantic uncertainty of the network, our uncertainty-aware semantic BKI mapping framework can produce reliable and accurate 3D semantic maps in off-road scenes.

3D continuous semantic mapping. Notably, recent semantic mapping approaches depend on the inference of Deep Neural Networks (DNNs) [14]–[18]. These networks derive semantic predictions from sensor measurements, which are then integrated into the mapping process. As a result, the performance of semantic mapping is highly dependent on the accuracy of semantic predictions generated by DNNs.

However, the predictions from DNNs often suffer from unreliability [19], [20], compromising the mapping performance, especially in off-road settings. The diverse appearances and unstructured nature of off-road environments, combined with illumination changes that increase intra-class variation, lead to unreliable operation of DNNs [21], [22]. Moreover, the common approach of compressing probabilistic predictions from DNNs into one-hot vectors [13], [23], [24] results in the loss of valuable information, which worsens the performance of mapping algorithms. Consequently, semantic mapping in perceptually challenging off-road environments frequently leads to unreliable outputs. To address this challenge, it is crucial to consider the uncertainty of semantic predictions from DNNs to construct semantic maps robustly, underscoring the necessity for advanced, uncertainty-aware mapping techniques, as illustrated in Fig. 1.

In this work, we propose an uncertainty-aware semantic BKI mapping framework for robust deployments in off-road environments with high perceptual uncertainties. To achieve this goal, we incorporate the uncertainty of semantic

This work was supported by the Agency for Defense Development Grant funded by the Korean Government in 2024.

<sup>1</sup>Agency for Defense Development, Daejeon, Republic of Korea. {junyoung.kimv, happymin77}@gmail.com

<sup>2</sup>Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. junwonse@andrew.cmu.edu

\*These authors contributed equally to this work.

Our project website can be found at <https://bit.ly/EvSemMap>

predictions into the mapping process, thereby improving the reliability and accuracy. To estimate the semantic uncertainty, we adopt Evidential Deep Learning [25], known for its reliable uncertainty estimates. Then, probabilistic semantic predictions are seamlessly integrated into our mapping process, enabling precise representations of uncertain environments. Furthermore, the mapping process is adapted to incorporate semantic uncertainty by not only filtering out uncertain predictions but also reinforcing the effect of confident predictions. Through validation with off-road datasets, including RELIS-3D and our own off-road datasets, we verify that our method enhances the accuracy and reliability of semantic mapping compared to existing methods.

## II. RELATED WORK

### A. 3D Semantic Mapping

Among various 3D geometric representations such as Gaussian [26], surfel [14], and mesh [7], voxel-based 3D model is commonly used in semantic mapping [6], [8], [10]–[12], [15], [16]. Since traditional voxel-based mapping techniques assume the statistical independence of each voxel, each sensor measurement is used to update only a restricted subset of voxels. As a result, sparse LiDAR measurements lead to the construction of discontinuous maps [3]. Therefore, methods have been proposed to relax the voxel independence assumption by incorporating local spatial information, demonstrating improved performance [3], [4].

GPOctoMap [3] introduces Gaussian Process (GP) to account for local spatial correlations, relaxing the voxel independence assumption. However, the computational cost  $\mathcal{O}(n^3)$  with  $n$  being the number of sensor measurements makes it prohibitive for real-time applications. To address this gap, Bayesian Kernel Inference (BKI) [5] has been introduced as an efficient approximation to GP [4]. This significantly reduces the computational complexity to  $\mathcal{O}(\log n)$ , facilitating real-time updates and efficient integration of local spatial information into the mapping process. This efficient BKI was then adopted in S-BKI [13] for semantic mapping, extending the capability to include semantic information within the geometric model.

While traditional semantic mapping systems have utilized Conditional Random Field (CRF) [6]–[8], [27]–[29], subsequent improvements have increasingly favored Deep Neural Networks (DNNs) to incorporate semantic information into maps [14]–[18]. The semantic segmentation results obtained from DNNs at each time step are commonly utilized for the incremental update of the semantic map. After S-BKI [13] introduces the BKI-based continuous mapping into the semantic mapping field, various approaches are proposed to improve the accuracy of continuous semantic mapping. ConvBKI [24] proposes a trainable version of S-BKI using depth-wise convolution, encouraging the kernel to learn class-specific geometry. Meanwhile, SEE-CSOM [23] is proposed to address the challenges of overinflation and inefficiency by considering the label inconsistency of the voxels. However, the uncertainty of semantic predictions is yet to be introduced in continuous semantic mapping.

### B. Uncertainty Estimation of Deep Neural Networks

Uncertainty estimation [19], [20], considered an important aspect of deploying DNNs in safety-critical applications, is a process of estimating the uncertainty of the predictions from DNNs. The most widely used methods are sampling-based approaches [30], requiring multiple inferences for each input. For example, Monte Carlo dropout [31] performs inference multiple times with a stochastic dropout layer, and Deep Ensemble [32] uses an ensemble of networks to obtain multiple predictions to estimate uncertainty. Therefore, they are unsuitable for resource-constrained real-time applications [33]. Thus, methods that enable uncertainty estimation with a single forward pass, such as Evidential Deep Learning (EDL) [25], have been proposed.

EDL framework estimates uncertainty by modeling a second-order Dirichlet distribution based on the Dempster-Shafer Theory of Evidence [34]. The network is trained to compute parameters for the Dirichlet distribution from which the uncertainty of prediction is calculated. Since it can be easily applied to existing DNNs, it has been widely adopted in various fields for uncertainty estimation [35]–[38].

## III. METHODS

In this section, we present our uncertainty-aware semantic mapping framework. First, we provide an overview of the semantic BKI mapping process. Then, we incorporate the EDL into DNNs for semantic segmentation to estimate semantic uncertainty. Lastly, we introduce our uncertainty-aware BKI framework, which enhances the performance of semantic mapping by leveraging the predictive uncertainty of semantic predictions. Our entire framework is illustrated in Fig. 2.

### A. Semantic Bayesian Kernel Inference

In semantic BKI frameworks, the mapping objective is to estimate the probability distribution  $\theta_* = [\theta_*^1, \dots, \theta_*^C]$  across  $C$  semantic classes for arbitrary query point  $x_*$ , where  $\sum_{c=1}^C \theta_*^c = 1$  and  $\theta_*^c \in [0, 1]$ . The  $\theta_*$  is then used to predict semantic label  $\mathbf{y}_*$  based on the Categorical distribution  $p(\mathbf{y}_*|\theta_*)$ . Thus, the semantic points  $\mathcal{D} = \{x_i, \mathbf{y}_i\}_{i=1}^N$  is utilized to estimate the posterior distribution  $p(\theta_*|x_*, \mathcal{D})$ , where  $x_i$  is the coordinates of the  $i$ -th observation and  $\mathbf{y}_i$  is the corresponding semantic label. In practice, the posterior distribution is estimated only for the center of voxels during the mapping for efficiency, though queries on arbitrary locations can still be processed.

Initially, Bayes' rule is applied to derive the posterior in closed form equation:

$$\begin{aligned} p(\theta_*|x_*, \mathcal{D}) &\propto p(\mathcal{D}|\theta_*, x_*)p(\theta_*|x_*) \\ &\propto \left[ \prod_{i=1}^N \underbrace{p(\mathbf{y}_i|x_i, \theta_*, x_*)}_{\text{extended likelihood}} \right] \underbrace{p(\theta_*|x_*)}_{\text{prior}}. \end{aligned} \quad (1)$$

Then, BKI [5] is applied as it provides the relationship between the extended likelihood and the likelihood. In [5], the maximum entropy distribution  $g$  satisfying  $D_{\text{KL}}(g||f) \leq \rho(x_*, x)$  is proven to be  $g(y) \propto f(y)^{k(x_*, x)}$  with a kernel

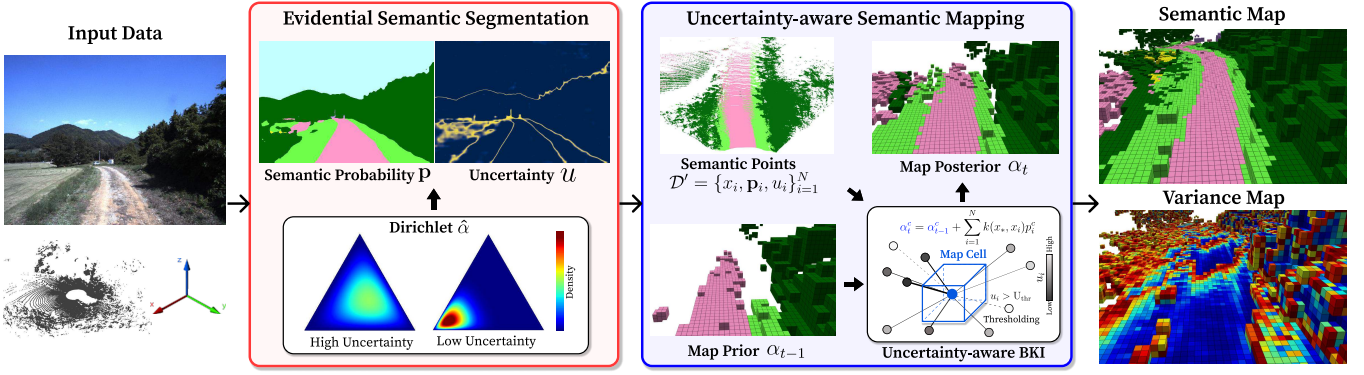


Fig. 2: Overview pipeline of our uncertainty-aware semantic BKI framework. With an evidential segmentation network trained by EDL, input data is processed to derive continuous semantic probability and uncertainty. These 3D semantic points are then integrated into the semantic map through Bayesian updates using the uncertainty-aware BKI, resulting in a dependable semantic map and variance map in uncertain off-road environments.

function  $k : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ , where  $D_{\text{KL}}(\cdot || \cdot)$  is the Kullback-Leibler divergence and  $\rho(\cdot, \cdot)$  is a smoothness bound. Substituting  $g$  for the extended likelihood and  $f$  for the likelihood, we can obtain the following proportional relationship:

$$p(\mathbf{y}_i | x_i, \boldsymbol{\theta}_*, x_*) \propto p(\mathbf{y}_i | \boldsymbol{\theta}_*)^{k(x_*, x_i)}. \quad (2)$$

By combining (1) and (2), we have:

$$p(\boldsymbol{\theta}_* | x_*, \mathcal{D}) \propto \left[ \prod_{i=1}^N p(\mathbf{y}_i | \boldsymbol{\theta}_*)^{k(x_*, x_i)} \right] p(\boldsymbol{\theta}_* | x_*). \quad (3)$$

As  $\mathbf{y}_i$  represents the one-hot encoded semantic label and the Categorical likelihood  $p(\mathbf{y}_i | \boldsymbol{\theta}_*)$  is used in the frameworks, a Dirichlet prior is adopted to enable incremental Bayesian inference since it is the conjugate prior distribution. An initial Dirichlet prior  $Dir(C, \boldsymbol{\alpha}_0)$  over  $\boldsymbol{\theta}_*$  is parameterized by hyperparameters  $\boldsymbol{\alpha}_0 = [\alpha_0^1, \dots, \alpha_0^C]$ . Applying the definition of the prior and the likelihood, (3) becomes:

$$\begin{aligned} p(\boldsymbol{\theta}_* | x_*, \mathcal{D}) &\propto \left[ \prod_{i=1}^N \left( \prod_{c=1}^C (\theta_*^c)^{y_i^c} \right)^{k(x_*, x_i)} \right] \prod_{c=1}^C (\theta_*^c)^{\alpha_0^c - 1} \\ &\propto \prod_{c=1}^C (\theta_*^c)^{\alpha_0^c + \sum_{i=1}^N k(x_*, x_i) y_i^c - 1}. \end{aligned} \quad (4)$$

Consequently, a Dirichlet posterior  $Dir(C, \boldsymbol{\alpha}_t)$  at time  $t$  can be calculated based on the posterior at time  $t-1$ . This update process for the map cell can be recursively executed as:

$$\alpha_t^c = \alpha_{t-1}^c + \sum_{i=1}^N k(x_*, x_i) y_i^c, \quad (5)$$

and the expectation and variance of the posterior are calculated as:

$$S_t = \sum_{c=1}^C \alpha_t^c, \quad \mathbb{E}[\alpha_t^c] = \frac{\alpha_t^c}{S_t}, \quad \text{Var}[\alpha_t^c] = \frac{\alpha_t^c (S_t - \alpha_t^c)}{S_t^2 (S_t + 1)}. \quad (6)$$

It is worth noting that the variance of the predicted label  $\text{Var}[\alpha_t^\psi]$  is utilized as an uncertainty measure of the map cell in [13], [24], where  $\psi = \text{argmax}_c \mathbb{E}[\alpha_t^c]$ .

For the kernel function  $k(x_*, x_i)$  in (5), it is desired to prioritize information from nearby points and ignore

information from the points that are too distant. Among the functions with these characteristics, the sparse kernel [39] is commonly used to ensure smoothness and reduce computational complexity [4], [13], [23]:

$$\begin{aligned} k(x_*, x_i) &= k'(d, l, \sigma_0) \\ &= \mathbb{1}_{d < l} \sigma_0 \left[ \frac{2 + \cos(2\pi \frac{d}{l})}{3} \left(1 - \frac{d}{l}\right) + \frac{1}{2\pi} \sin(2\pi \frac{d}{l}) \right], \end{aligned} \quad (7)$$

where  $\mathbb{1}$  is the indicator function,  $d = \|x_* - x_i\|$  is the distance between  $x_*$  and  $x_i$ ,  $l > 0$  is the length-scale hyperparameter, and  $\sigma_0$  is a kernel scaling hyperparameter. As the value of the sparse kernel decreases smoothly as the distance increases up to  $l$ , the surrounding size of space considered by the BKI can be scaled by adjusting  $l$ .

## B. Evidential Semantic Segmentation

To utilize the predictive uncertainty of semantic predictions in the mapping process, we apply the EDL framework [25] into existing segmentation networks to estimate its uncertainty reliably. The output of the networks for  $i$ -th input is regarded as an evidence vector  $\mathbf{e}_i = [e_i^1, \dots, e_i^C]$  by replacing the last SoftMax( $\cdot$ ) of the network into  $\exp(\cdot)$  [40].

From the evidence vector  $\mathbf{e}_i$ , the uncertainty  $u_i$  is estimated as  $u_i = C / \hat{S}_i$ , where  $\hat{S}_i = \sum_{c=1}^C (e_i^c + 1)$ , enabling deterministic uncertainty estimation with a single-pass operation. Then, the semantic probability distribution  $\mathbf{p}_i = [p_i^1, \dots, p_i^C]$  can be modeled as  $p_i^c = \hat{\alpha}_i^c / \hat{S}_i$  by the Dirichlet distribution with concentration parameters of EDL  $\hat{\boldsymbol{\alpha}}_i = [\hat{\alpha}_i^1, \dots, \hat{\alpha}_i^C]$ , where  $\hat{\alpha}_i^c = e_i^c + 1$ .

We adopt the loss from [25], [37] to train the semantic segmentation network with EDL. The following evidential loss encourages collecting semantic evidence of its corresponding one-hot label  $\mathbf{y}_i$  from the training dataset  $\mathcal{D}_{\text{sem}} = \{x_i, \mathbf{y}_i\}_{i=1}^N$  for each pixel  $x_i$ :

$$\mathcal{L}_{\text{EDL}}(x_i, \mathbf{y}_i) = \sum_{c=1}^C y_i^c (\log(\hat{S}_i) - \log(\hat{\alpha}_i^c)). \quad (8)$$

While the EDL loss in (8) can theoretically encourage minimizing  $\hat{S}_i$  and maximizing the concentration parameter of ground-truth class, an additional regularization term is

commonly utilized to penalize incorrect evidence on non-ground-truth class. The most widely used KL regularization term [25] is adopted as follows:

$$\mathcal{L}_{\text{reg}}(x_i) = D_{\text{KL}}[\text{Dir}(\mathbf{p}_i|\tilde{\alpha}_i) \parallel \text{Dir}(\mathbf{p}_i|\langle 1, \dots, 1 \rangle)], \quad (9)$$

where  $\tilde{\alpha}_i = [\tilde{\alpha}_i^1, \dots, \tilde{\alpha}_i^C]$  is Dirichlet parameters after the removal of correct evidence from  $\hat{\alpha}_i$  to penalize the misleading evidence only:

$$\tilde{\alpha}_i^c = \begin{cases} 1 & \text{if } y_i^c = 1 \\ \hat{\alpha}_i^c & \text{if } y_i^c = 0 \end{cases}. \quad (10)$$

The overall loss for our EDL segmentation network is

$$\mathcal{L} = \mathcal{L}_{\text{EDL}} + \lambda_{\text{kl}}\mathcal{L}_{\text{reg}}, \quad (11)$$

where  $\lambda_{\text{kl}}$  is a hyperparameter.

Once the evidential model is trained to minimize the loss in (11), the probability  $\mathbf{p}_i$  and its uncertainty  $u_i$  is reliably estimated for each input  $x_i$ . By incorporating these estimates into the mapping process, we can improve the performance of semantic mapping even in perceptually challenging environments such as off-road settings.

### C. Uncertainty-aware Mapping Extension

Since the knowledge from the evidential semantic segmentation networks cannot be fully utilized through a squeezed one-hot prediction  $\mathbf{y}_i$ , we generalize the semantic points  $\mathcal{D} = \{x_i, \mathbf{y}_i\}_{i=1}^N$  to  $\mathcal{D}' = \{x_i, \mathbf{p}_i, u_i\}_{i=1}^N$ , where  $\mathbf{p}_i$  and  $u_i$  are provided from the evidential semantic segmentation network. Accordingly, the likelihood  $p(\mathbf{p}_i|\theta_*)$  is also generalized as the Continuous Categorical [41] by

$$p(\mathbf{p}_i|\theta_*) \propto \prod_{c=1}^C (\theta_*^c)^{p_i^c}. \quad (12)$$

As the Dirichlet distribution is also the conjugate prior of the Continuous Categorical, we can similarly derive the posterior distribution from (3):

$$\begin{aligned} p(\theta_*|x_*, \mathcal{D}') &\propto \left[ \prod_{i=1}^N \left( \prod_{c=1}^C (\theta_*^c)^{p_i^c} \right)^{k(x_*, x_i)} \right] \prod_{c=1}^C (\theta_*^c)^{\alpha_0^c - 1} \\ &\propto \prod_{c=1}^C (\theta_*^c)^{\alpha_0^c + \sum_{i=1}^N k(x_*, x_i) p_i^c - 1}. \end{aligned} \quad (13)$$

Hence, the formula for the recursive update of the Dirichlet posterior parameters in (5) becomes

$$\alpha_t^c = \alpha_{t-1}^c + \sum_{i=1}^N k(x_*, x_i) p_i^c. \quad (14)$$

Building upon the extension with the Continuous Categorical, semantic points can accurately represent the uncertain off-road environments compared to utilizing one-hot squeezed segmentation results, thereby enhancing semantic mapping performances.

Furthermore, we devise an uncertainty-aware adaptive kernel to prioritize reliable semantic predictions based on their semantic uncertainty. The kernel function in (7) is refined

to incorporate uncertainty estimates of semantic predictions into the mapping process as follows:

$$k(x_*, x_i) = \begin{cases} k'(d, l \cdot \beta e^{1-\gamma u_i}, \sigma_0) & \text{if } u_i \leq U_{\text{thr}} \\ 0 & \text{if } u_i > U_{\text{thr}}, \end{cases} \quad (15)$$

where  $\beta$  and  $\gamma$  are hyperparameters. This uncertainty-aware kernel can enhance semantic mapping accuracy and enable more reliable calculations of the variance of map cells through the following mechanisms.

Firstly, semantic predictions with excessive uncertainty are excluded based on the uncertainty threshold  $U_{\text{thr}}$ , which is dynamically calculated to filter out top- $\tilde{u}\%$  uncertain points. When the segmentation network lacks sufficient evidence for an input, the semantic prediction is disregarded in the mapping process. Consequently, this thresholding can prevent untrustworthy predictions from overriding accurate ones, improving the robustness against outliers in environments with perceptual uncertainties.

Then, the length of the sparse kernel is adjusted based on uncertainty to enhance the local influence of reliable predictions. The kernel length decreases exponentially as uncertainty increases, while the kernel length is increased for confident predictions. Increasing the kernel length for reliable predictions dynamically expands the surrounding space considered by the kernel, thereby reinforcing the local influence of reliable ones during BKI. This ensures that confident semantic predictions contribute more significantly during the recursive update of posterior parameters, enabling more reliable and consistent mapping in environments with perceptual uncertainties.

## IV. EXPERIMENTS

In this section, we validate that our uncertainty-aware semantic BKI mapping framework effectively enhances semantic mapping performance in off-road settings characterized by varied visual appearances and perceptual uncertainties. Using both publicly available off-road datasets and our own dataset, we verify that our framework can improve semantic mapping performance. Then, we show that our framework can operate reliably in unseen environments using a pre-trained semantic segmentation network, facilitating robust mapping across various off-road settings. Lastly, ablation studies are performed to assess the validity of each component of our framework on its overall performance.

### A. Datasets

We utilize two widely-used off-road datasets, *RELLIS-3D* [42] and *RUGD* [43]. *RELLIS-3D* contains RGB images and LiDAR point clouds with dense semantic annotations and precise robot poses, providing comprehensive information for semantic mapping. The dataset consists of 5 sequences, and we conduct a 5-fold evaluation where four sequences are designated for training the segmentation network and one for evaluation. *RUGD* dataset provides RGB images with pixel-wise semantic annotations. As it does not provide geometric information, it is only used for evidential segmentation network training.

To further validate our method across diverse off-road environments with varying visual appearances, we collect our dataset (*OffRoad*) using our platform equipped with an OS1-128 LiDAR and front-view RGB camera, accompanied by accurate robot poses recovered by SLAM [44]. This dataset is obtained through high-speed navigation and captured under varied conditions, including different seasons, lighting conditions, and terrain types, making reliable 3D mapping more challenging. For training and evaluation, RGB images are manually annotated by experts, and the evaluation dataset contains distinct trajectory sequences not present in the training dataset. The semantic classes of all datasets are remapped to seven categories, as shown in Table II.

### B. Experimental Setup

**Implementation Details** In all experiments, semantic segmentation is conducted on RGB images to leverage their richer semantic information, and the 2D semantic predictions are projected onto paired 3D point clouds. We adopt DeepLabV3 [45] with the EDL framework for semantic segmentation, and the network is trained with the loss in (11) for 100 epochs. For each dataset, the network is separately trained with the training subset. During the training, the hyperparameter  $\lambda_{kl} = 0.5 \times (\tau/100)$  is progressively increased, where  $\tau$  is the current training epoch.

TABLE I: Hyperparameters for the semantic mapping experiments.

Dataset	Resolution	$\alpha_0^c$	$l$	$\beta$	$\gamma$	$\tilde{u}$
<i>RELLIS-3D</i>	0.2m	1.0	0.2	0.5	10.0	10%
<i>OffRoad</i>	0.3m	1.0	0.3	1.0	3.0	30%

The mapping hyperparameters are summarized in Table I. In both datasets, the kernel scaling parameter  $\sigma_0$  is set to 1.0, to ensure the validity of the BKI framework [5]. To evaluate the performance of semantic mapping techniques, we first generate 3D ground truths by projecting 2D semantic labels onto a 3D point cloud. Evaluations are then conducted by querying the semantic map for each point location of the 3D ground truth and assessing the accuracy of the obtained semantic labels.

**Evaluation Metrics** Intersection over Union (IoU) metric is employed to evaluate the semantic mapping performance. The IoU for the  $c$ -th class is computed as:

$$\text{IoU}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c}, \quad (16)$$

where  $\text{TP}_c$ ,  $\text{FP}_c$ , and  $\text{FN}_c$ , represent the number of true positive, false positive, and false negative queries from the semantic map, respectively. The calculation is restricted only to queries on occupied cells, excluding any queries where the semantic map identifies them as unoccupied. Then, the mean IoU (mIoU) is calculated by averaging the IoU scores for all classes to evaluate the overall performance. Moreover, we also report voxel semantic accuracy to measure the collective accuracy of queries, which is defined as:

$$\text{Acc} = \frac{\sum_{c=1}^C \text{TP}_c}{\sum_{c=1}^C (\text{TP}_c + \text{FP}_c)}. \quad (17)$$

TABLE II: Quantitative results on *RELLIS-3D* and our *OffRoad* dataset. Semantic classes not present in the evaluation dataset are excluded when calculating mIoU and denoted with a dash (-). Our method shows superior performance compared to other 3D continuous semantic mapping methods.

Dataset	Method	puddle	object	paved	unpaved	dirt	grass	vegetation	mIoU [%]	Acc [%]
		■	■	■	■	■	■	■		
<i>RELLIS-3D</i>	<i>S-CSM</i> [13]	29.6	<b>34.9</b>	52.9	-	<b>18.4</b>	<b>77.2</b>	<u>73.9</u>	<b>47.8</b>	<b>84.3</b>
	<i>S-BKI</i> [13]	27.0	28.5	<u>53.8</u>	-	<u>13.4</u>	76.4	73.8	45.5	83.7
	<i>ConvBKI</i> [24]	<b>30.6</b>	22.4	45.5	-	6.3	69.6	68.4	40.5	79.7
	<i>SEE-CSOM</i> [23]	27.7	33.4	53.4	-	6.7	76.4	<b>74.0</b>	45.3	83.6
	<i>Ours</i>	<u>29.7</u>	<u>34.6</u>	<b>55.8</b>	-	12.1	<u>76.7</u>	73.8	<u>47.1</u>	<u>84.0</u>
<i>OffRoad</i>	<i>S-CSM</i> [13]	-	<u>43.2</u>	-	72.3	-	52.3	91.5	<u>64.8</u>	89.7
	<i>S-BKI</i> [13]	-	38.0	-	68.6	-	49.9	90.9	61.9	89.0
	<i>ConvBKI</i> [24]	-	34.7	-	61.2	-	36.8	91.3	56.0	88.2
	<i>SEE-CSOM</i> [23]	-	41.4	-	<u>72.3</u>	-	<u>52.3</u>	<u>91.8</u>	64.4	<u>90.1</u>
	<i>Ours</i>	-	<b>44.2</b>	-	<b>76.4</b>	-	<b>57.9</b>	<b>92.5</b>	<b>67.8</b>	<b>91.3</b>

**Comparison Methods** We perform comparative analyses with existing 3D semantic mapping methods. *S-CSM* and *S-BKI* [13] are adopted as baselines for discrete and continuous semantic mapping, respectively. We also include results from *ConvBKI* [24], which utilizes a learnable class-wise kernel. We use the pre-trained kernel derived from *RELLIS-3D* across all experiments. Lastly, we compare the results of *SEE-CSOM* [23], which incorporates a simple class entropy during the mapping process.

### C. Quantitative and Qualitative Results

The quantitative results with *RELLIS-3D* and *OffRoad* are presented in Table II. Our method shows superior performance among continuous mapping methods in both datasets. While the discrete semantic mapping approach, *S-CSM*, exhibits better performance in *RELLIS-3D*, this is primarily because the evaluation is conducted only on occupied voxels from discontinuous maps. In our dataset, characterized by a higher level of sparsity and perceptual uncertainty, our method surpasses *S-CSM* while also constructing a continuous map. Methods adopting label inconsistency during the mapping (*SEE-CSOM*) lead to enhanced performance in off-road datasets since otherwise inconsistent predictions can be negatively propagated during BKI. Our method shows better results than *SEE-CSOM*, implying the relevance of EDL for quantifying semantic uncertainty and effectively utilizing this uncertainty through the adaptive kernel. Moreover, *ConvBKI* leads to even poorer performance, suggesting that pre-trained kernels for specific semantic classes would result in inaccuracies in uncertain unstructured environments.

Fig. 3 provides a qualitative comparison of 3D semantic maps generated by different methods. Semantic mapping methods that do not incorporate perceptual uncertainty tend to produce noisy semantic maps. Our method achieves more accurate semantic map generation by conducting uncertainty quantification and integrating this uncertainty into the mapping process.

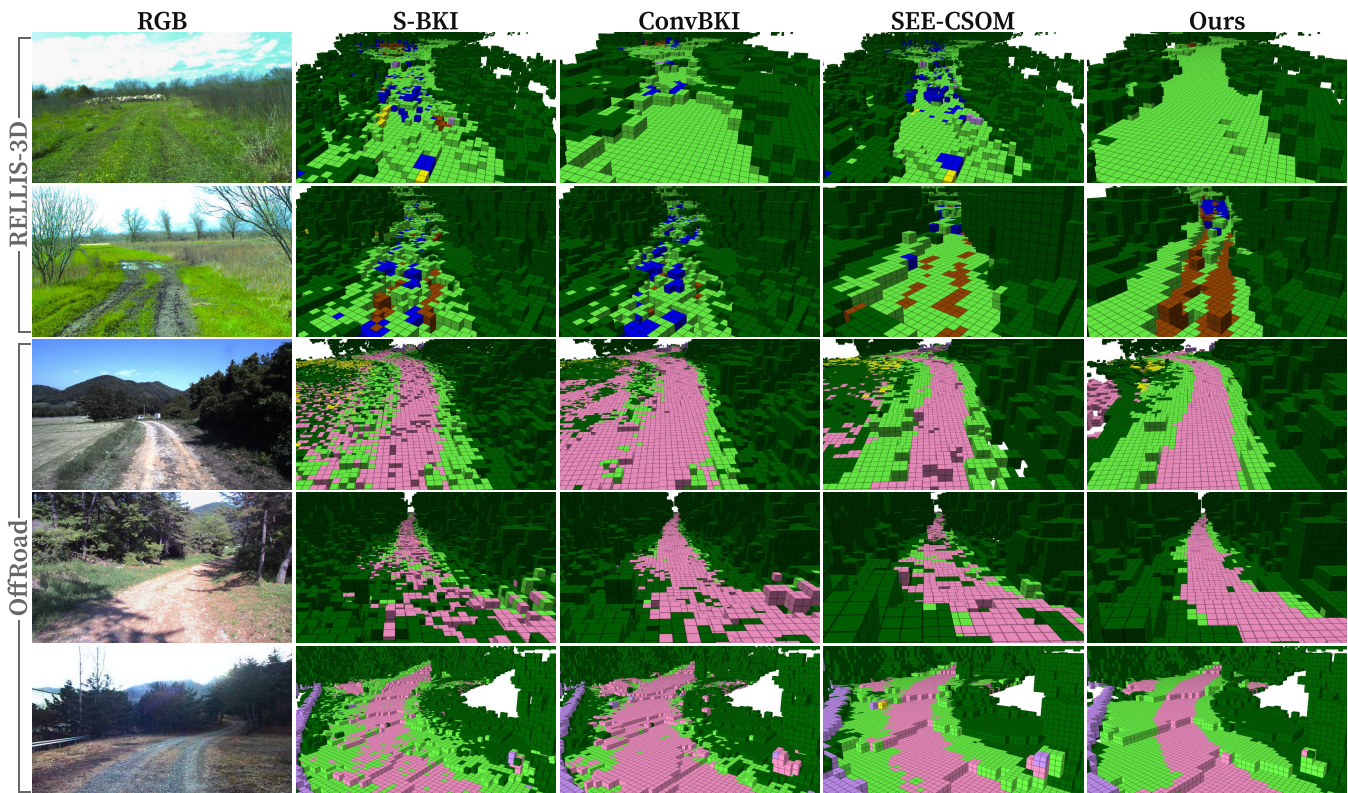


Fig. 3: Qualitative results of 3D semantic mapping methods. Compared to others, our method generates reliable and accurate maps that preserve semantic details while excluding noisy predictions. In *RELLIS-3D*, only our method reconstructs **grass** consistently (First row), and **dirt roads** and **puddles** in detail (Second row). In our *OffRoad* dataset, our method accurately reconstructs the boundaries of **unpaved roads**, **grass**, and **vegetation**, compared to others.

TABLE III: Quantitative results on our *OffRoad* dataset using semantic segmentation networks pre-trained on different datasets.

<i>Training Data</i>	<i>OffRoad</i>		<i>OffRoad-P</i>		<i>RUGD</i>		<i>RELLIS-3D</i>	
<b>Method</b>	mIoU	Acc	mIoU	Acc	mIoU	Acc	mIoU	Acc
<i>S-CSM</i> [13]	64.8	89.7	60.4	87.0	41.2	71.1	21.9	65.9
<i>S-BKI</i> [13]	61.9	89.0	57.7	86.1	39.4	71.0	<b>22.0</b>	65.7
<i>ConvBKI</i> [24]	56.0	88.2	52.4	85.1	39.3	71.2	20.8	64.2
<i>SEE-CSOM</i> [23]	64.4	<u>90.1</u>	59.2	86.9	<b>42.1</b>	<u>71.4</u>	21.6	<b>66.4</b>
<i>Ours</i>	<b>67.8</b>	<b>91.3</b>	<b>61.0</b>	<b>87.7</b>	<u>42.0</u>	<b>73.1</b>	21.8	<b>66.4</b>

#### D. Robust Semantic Mapping in Off-road

We show that our uncertainty-aware 3D semantic mapping method can be robustly deployed in various off-road settings. Given that robots operating in off-road scenes frequently encounter unseen surroundings, we simulate deployments in unseen environments. To assess the capability of our method to perform robustly in a zero-shot manner, we utilize a pre-trained semantic segmentation network trained on distinct datasets. Specifically, the semantic segmentation network is trained on four different datasets: our off-road dataset (*OffRoad*), a subset of our off-road dataset (*OffRoad-P*), *RUGD*, and *RELLIS-3D*. Then, the performance of semantic mapping is evaluated without adaptation on our off-road dataset.

The quantitative results are summarized in Table III. Our method outperforms other methods, even when utilizing segmentation models trained with different training datasets. This highlights the robustness of our approach, which main-

tains stable performance under variations in the training data distribution. Despite using a subset of the training data (*OffRoad-P*), which introduces higher uncertainty in semantic prediction, our method remains superior to others. Additionally, methods integrating label inconsistency into the mapping process (*SEE-CSOM*) exhibit robust performance. Nevertheless, our method performs better than *SEE-CSOM*, suggesting the capability to more effectively mitigate untrustworthy predictions and handle predictive uncertainties by incorporating EDL and uncertainty-aware BKI. Due to the substantial domain gap with the evaluation dataset, the performance gap observed in experiments with *RELLIS-3D* trained models is less evident than others.

Fig. 4 shows the semantic mapping results using the segmentation network trained on *RUGD*. While other methods struggle to reconstruct accurate semantic maps or produce discontinuous maps, our method successfully creates a clear, continuous map with unambiguous boundaries. *SEE-CSOM* fails to generate fully continuous semantic maps, indicating inefficacy in reliably addressing temporal inconsistency and uncertainties in semantic predictions. Additionally, *ConvBKI*, which requires dense 3D ground truths for learning the class-wise adaptive kernel, shows limited adaptability to unseen environments due to potential overfitting. In contrast, our method consistently produces continuous maps with clear boundaries without requiring additional 3D ground truth.

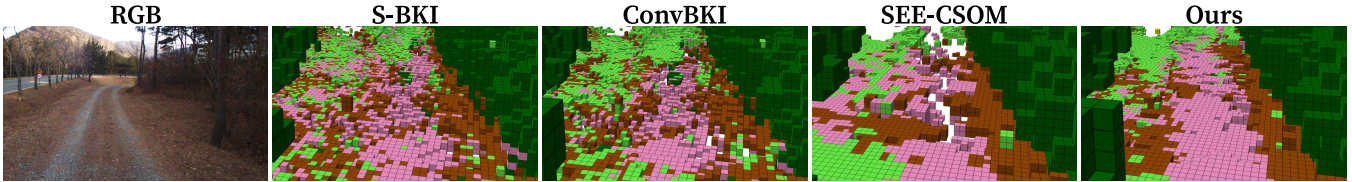


Fig. 4: Zero-shot semantic mapping results on our *OffRoad* dataset using a semantic segmentation network pre-trained on *RUGD*. Our method robustly constructs semantic maps despite prediction uncertainties in unseen environments, whereas other methods struggle to produce clear maps.

### E. Ablation Studies

We conduct ablation studies on our uncertainty-aware BKI framework to evaluate the impact of its key components described in Section III-C. Specifically, we examine the contribution of the Continuous Categorical extension (**Prob.**) (14), the adaptive kernel length (**Adap.**) (15), and the uncertainty thresholding (**Thr.**) (15) to the overall performance.

To comprehensively evaluate the impact of each component, we further investigate Voxel Occupancy Accuracy (O.Acc), defined as the proportion of queries with occupied voxels, measuring the geometric completeness of semantic maps. Moreover, the Brier Score (BS) is calculated to evaluate the reliability of the variance map, which can be interpreted as representing the uncertainty of the map cells. The confidence of a map cell’s semantics is defined as  $\text{Conf}_i = 1 - \text{Var}[\alpha_t^\psi]$  by normalizing the variance in (6) to  $\hat{\text{Var}}[\alpha_t^\psi]$ . BS is then defined as follows:

$$\text{BS} = \frac{1}{M} \sum_{i=1}^M (\mathbb{1}_{\text{correct}} - \text{Conf}_i)^2, \quad (18)$$

where  $M$  is the number of occupied map cells, and the indicator function  $\mathbb{1}_{\text{correct}}$  is assigned 1 if the semantic label of a map cell  $i$  is correctly estimated as the ground truth label of the cell and 0 otherwise. This metric becomes lower if the confidence of correctly estimated cells is high and the confidence of incorrectly estimated map cells is low.

The results are summarized in Table IV. In both datasets, the Continuous Categorical extension and the uncertainty threshold improve mIoU, whereas the adaptive kernel length enhances the reliability of the uncertainty measure with low BS. Although the uncertainty threshold significantly improves the mapping performances, it exacerbates geometric completeness with decreased O.Acc. Combining the adaptive kernel length with the uncertainty threshold enhances both the semantic mapping performance and the uncertainty’s reliability while maintaining geometric completeness. The

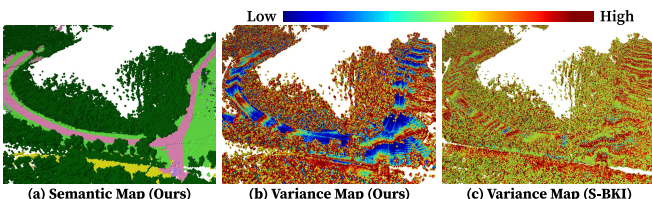


Fig. 5: Visualization of variance maps. Through uncertainty-aware BKI, our method generates more informative variance maps capable of reliably quantifying semantic map uncertainty.

TABLE IV: Results of the ablation studies.

Prob.	Adap.	Thr.	<i>RELLIS-3D</i>				<i>OffRoad</i>			
			mIoU	Acc	O.Acc	BS(↓)	mIoU	Acc	O.Acc	BS(↓)
-	-	-	45.5	83.7	96.7	35.7	61.9	89.0	99.7	39.7
✓	-	-	46.4	83.6	96.7	40.1	63.4	89.5	<b>99.9</b>	51.8
✓	✓	-	46.3	83.8	<b>96.9</b>	<u>27.4</u>	64.9	90.2	<b>99.9</b>	<b>12.5</b>
✓	-	✓	<b>47.8</b>	<b>84.4</b>	95.5	43.1	<b>68.9</b>	<b>91.9</b>	95.4	51.8
✓	✓	✓	<u>47.1</u>	<u>84.0</u>	<b>96.9</b>	<b>24.1</b>	<u>67.8</u>	<u>91.3</u>	98.2	<b>12.5</b>

variance map illustrated in Fig. 5 shows that the variance of our map holds more meaningful uncertainty information, as quantitatively described with improved BS.

## V. CONCLUSIONS

This paper presents an uncertainty-aware Bayesian Kernel Inference framework for reliable semantic mapping in uncertain environments. To address the challenges of high perceptual uncertainties, our framework incorporates uncertainty of semantic predictions from DNNs into the mapping process. Specifically, we leverage Evidential Deep Learning for semantic segmentation and introduce the adaptive kernel and uncertainty threshold mechanism to address the semantic uncertainty. Through extensive experiments, we demonstrate the effectiveness of our framework in handling semantic uncertainty, resulting in accurate and reliable semantic maps in perceptually challenging unstructured outdoor environments. Our method is robust in unseen environments with higher semantic uncertainty, constructing continuous semantic maps with precise semantic boundaries. We believe our framework has the potential to improve the accuracy and reliability of semantic mapping in a variety of contexts with perceptual uncertainties beyond off-road environments.

Future works include enhancing the real-time capability of our framework with more efficient geometric models. Additionally, we intend to extend the framework through uncertainty-aware multi-modal fusion, enabling effective integration of information. Furthermore, we are interested in advanced techniques for uncertainty calibration to ensure more reliable mapping. Exploring the potential of our semantic map and its uncertainty for active exploration also presents an interesting research avenue.

## REFERENCES

- [1] A. Elfes, “Using occupancy grids for mobile robot perception and navigation,” *Computer*, vol. 22, no. 6, pp. 46–57, 1989.
- [2] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, “Octomap: An efficient probabilistic 3d mapping framework based on octrees,” *Autonomous Robots*, vol. 34, pp. 189–206, 2013.

- [3] J. Wang and B. Englot, "Fast, accurate gaussian process occupancy maps via test-data octrees and nested bayesian fusion," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1003–1010.
- [4] K. Doherty, J. Wang, and B. Englot, "Bayesian generalized kernel inference for occupancy map prediction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 3118–3124.
- [5] W. R. Vega-Brown, M. Doniec, and N. G. Roy, "Nonparametric bayesian inference on multivariate exponential families," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 27, 2014.
- [6] B.-s. Kim, P. Kohli, and S. Savarese, "3d scene understanding by voxel-crf," in *IEEE/CVF International Conference on Computer Vision (CVPR)*, 2013, pp. 1425–1432.
- [7] J. P. Valentin, S. Sengupta, J. Warrell, A. Shahrokni, and P. H. Torr, "Mesh based semantic modelling for indoor and outdoor scenes," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 2067–2074.
- [8] S. Sengupta and P. Sturgess, "Semantic octree: Unifying recognition, reconstruction and representation via an octree constrained higher order mrf," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 1874–1879.
- [9] D. Paz, H. Zhang, Q. Li, H. Xiang, and H. I. Christensen, "Probabilistic semantic mapping for urban autonomous driving applications," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 2059–2064.
- [10] A. Asgharivaskasi and N. Atanasov, "Semantic octree mapping and shannon mutual information computation for robot exploration," *IEEE Transactions on Robotics*, 2023.
- [11] D. Morilla-Cabello, L. Mur-Labadia, R. Martinez-Cantin, and E. Montijano, "Robust fusion for bayesian semantic mapping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.
- [12] J. M. C. Marques, A. J. Zhai, S. Wang, and K. Hauser, "On the overconfidence problem in semantic 3d mapping," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 11 095–11 102.
- [13] L. Gan, R. Zhang, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "Bayesian spatial kernel smoothing for scalable dense semantic mapping," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 790–797, 2020.
- [14] J. McCormac, A. Handa, A. Davison, and S. Leutenegger, "Semantifusion: Dense 3d semantic mapping with convolutional neural networks," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 4628–4635.
- [15] S. Yang, Y. Huang, and S. Scherer, "Semantic 3d occupancy mapping through efficient high order crfs," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 590–597.
- [16] Y. Xiang and D. Fox, "Da-rnn: Semantic mapping with data associated recurrent neural networks," *Robotics: Science and Systems (RSS)*, 2017.
- [17] L. Sun, Z. Yan, A. Zaganidis, C. Zhao, and T. Duckett, "Recurrent-octomap: Learning state-based map refinement for long-term semantic mapping with 3-d-lidar data," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3749–3756, 2018.
- [18] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Field and Service Robotics: Results of the 11th International Conference*, 2018, pp. 335–350.
- [19] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *International Conference on Machine Learning (ICML)*, 2017, pp. 1321–1330.
- [20] P. de Jorge, R. Volpi, P. H. Torr, and G. Rogez, "Reliability in semantic segmentation: Are we on the right track?" in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 7173–7182.
- [21] Y. Jin, D. Han, and H. Ko, "Memory-based semantic segmentation for off-road unstructured natural environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 24–31.
- [22] J. Seo, S. Sim, and I. Shim, "Learning off-road terrain traversability with self-supervisions only," *IEEE Robotics and Automation Letters*, 2023.
- [23] Y. Deng, M. Wang, Y. Yang, D. Wang, and Y. Yue, "See-csom: Sharp-edged and efficient continuous semantic occupancy mapping for mobile robots," *IEEE Transactions on Industrial Electronics*, 2023.
- [24] J. Wilson, Y. Fu, J. Friesen, P. Ewen, A. Capodiecici, P. Jayakumar, K. Barton, and M. Ghaffari, "Conybki: Real-time probabilistic semantic mapping network with quantifiable uncertainty," *arXiv preprint arXiv:2310.16020*, 2023.
- [25] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.
- [26] P. Z. X. Li, S. Karaman, and V. Sze, "Gmmmap: Memory-efficient continuous occupancy map using gaussian mixture model," *IEEE Transactions on Robotics*, 2024.
- [27] S. Sengupta, E. Greveson, A. Shahrokni, and P. H. Torr, "Urban 3d semantic modelling using stereo vision," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 580–585.
- [28] A. Kundu, Y. Li, F. Dellaert, F. Li, and J. M. Rehg, "Joint semantic segmentation and 3d reconstruction from monocular video," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 703–718.
- [29] V. Vineet, O. Miksik, M. Lidegaard, M. Nießner, S. Golodetz, V. A. Prisacariu, O. Köhler, D. W. Murray, S. Izadi, P. Pérez *et al.*, "Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 75–82.
- [30] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, "Hands-on bayesian neural networks—a tutorial for deep learning users," *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 29–48, 2022.
- [31] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning (ICML)*, 2016, pp. 1050–1059.
- [32] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.
- [33] T. Kim, J. Mun, J. Seo, B. Kim, and S. Hong, "Bridging active exploration and uncertainty-aware deployment using probabilistic ensemble neural network dynamics," *Robotics: Science and Systems (RSS)*, 2023.
- [34] R. R. Yager and L. Liu, *Classic works of the Dempster-Shafer theory of belief functions*. Springer, 2008, vol. 219.
- [35] Y. Chang, F. Xue, F. Sheng, W. Liang, and A. Ming, "Fast road segmentation via uncertainty-aware symmetric network," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022, pp. 11 124–11 130.
- [36] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2551–2566, 2022.
- [37] K. Sirohi, S. Marvi, D. Büscher, and W. Burgard, "Uncertainty-aware panoptic segmentation," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2629–2636, 2023.
- [38] X. Cai, S. Ancha, L. Sharma, P. R. Osteen, B. Bucher, S. Phillips, J. Wang, M. Everett, N. Roy, and J. P. How, "EVORA: Deep Evidential Traversability Learning for Risk-Aware Off-Road Autonomy," *IEEE Transactions on Robotics*, vol. 40, pp. 3756–3777, 2024.
- [39] A. Melkumyan and F. T. Ramos, "A sparse covariance function for exact gaussian process inference in large datasets," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2009.
- [40] D. S. Pandey and Q. Yu, "Learn to accumulate evidence from all training samples: theory and practice," in *International Conference on Machine Learning (ICML)*, 2023, pp. 26 963–26 989.
- [41] E. Gordon-Rodriguez, G. Loaiza-Ganem, and J. Cunningham, "The continuous categorical: a novel simplex-valued exponential family," in *International Conference on Machine Learning (ICML)*, 2020, pp. 3637–3647.
- [42] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 1110–1116.
- [43] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5000–5007.
- [44] J. Seo, T. Kim, S. Ahn, and K. Kwak, "Metaverse: Meta-learning traversability cost map for off-road navigation," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.
- [45] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.