IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

E2-BKI: Evidential Ellipsoidal Bayesian Kernel Inference for Uncertainty-aware Gaussian Semantic Mapping

Junyoung Kim[1], Minsik Jeon[2], Jihong Min[1], Kiho Kwak[1], Junwon Seo[2]

***Abstract*— Semantic mapping aims to construct a 3D semantic representation of the environment, providing essential knowledge for robots operating in complex outdoor settings. While Bayesian Kernel Inference (BKI) addresses discontinuities of map inference from sparse sensor data, existing semantic mapping methods suffer from various sources of uncertainties in challenging outdoor environments. To address these issues, we propose an uncertainty-aware semantic mapping framework that handles multiple sources of uncertainties, which significantly degrade mapping performance. Our method estimates uncertainties in semantic predictions using Evidential Deep Learning and incorporates them into BKI for robust semantic inference. It further aggregates noisy observations into coherent Gaussian representations to mitigate the impact of unreliable points, while employing geometry-aligned kernels that adapt to complex scene structures. These Gaussian primitives effectively fuse local geometric and semantic information, enabling robust, uncertainty-aware mapping in complex outdoor scenarios. Comprehensive evaluation across diverse off-road and urban outdoor environments demonstrates consistent improvements in mapping quality, uncertainty calibration, representational flexibility, and robustness, while maintaining real-time efficiency. Our project website:** https://e2-bki.github.io/

I. INTRODUCTION

Semantic mapping constructs a *what-is-where* map by estimating semantic labels at each 3D location, enabling robots to understand and operate in complex environments [1]–[7]. This process typically operates on sparse and noisy sensor data, along with semantic predictions from neural networks. As sparse observations often lead to discontinuous geometric reconstructions, *continuous mapping* approaches have been proposed to densify sparse regions. A representative approach is Bayesian Kernel Inference (BKI) [8], with its semantic extension, S-BKI [9], which leverages local information to infer the semantics of sparse regions. Through kernel-based spatial propagation, S-BKI addresses the discontinuity issues, enabling spatially consistent semantic mapping.

While S-BKI provides continuity in semantic map inference, it struggles in challenging outdoor conditions due to its inability to account for inherent uncertainties arising from various sources. *Semantic uncertainty* in neural network predictions can undermine mapping performance, as S-BKI treats all observations as equally reliable. *Spatial uncertainty* also arises from the use of static isotropic kernels, which apply uniform influence in all directions and can misalign with anisotropic scene structures such as roads or fences. Moreover, the point-wise processing in BKI fails to account for *observation uncertainty* as it processes points independently, making it vulnerable to noisy and sparse sensor measurements. Although recent works have separately addressed semantic [10], [11] or spatial uncertainty [12], [13], a unified approach has not been proposed.

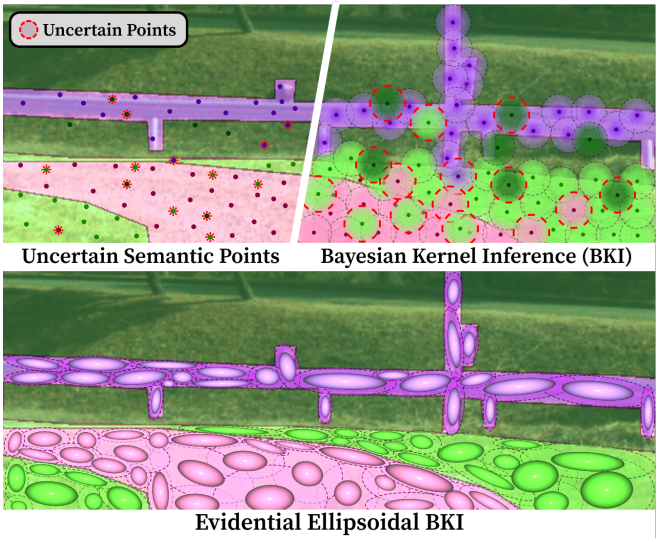

Fig. 1: **Evidential Ellipsoidal BKI (E2-BKI)** combines uncertainty-aware processing that prioritizes reliable observations with anisotropic kernels that align with local scene geometry, enabling the construction of accurate and reliable semantic maps from sparse, noisy, and uncertain semantic points.

To address these limitations, we extend the uncertainty-aware semantic mapping framework [10] by introducing *anisotropic* Gaussian primitives as a compact scene representation. Our key insight is that aggregating noisy observations into ellipsoidal primitives enables joint modeling of local geometry, semantics, and uncertainty, while improving robustness to noise. Building on this idea, we construct Gaussian primitives via uncertainty-aware spatial aggregation of neighboring observations, as illustrated in Fig. 1. These primitives undergo refinement through merging and pruning, and form the basis of our evidential ellipsoidal BKI formulation with anisotropic kernels that adapt to geometry and incorporate semantic uncertainty. Extensive experiments across both off-road and urban outdoor environments demonstrate superior performance, consistently outperforming existing methods in semantic accuracy and geometric completeness while maintaining real-time efficiency. In summary, our contributions include:

This work was supported by the Agency for Defense Development Grant funded by the Korean Government in 2025.

[1]Agency for Defense Development, Daejeon, Republic of Korea. {junyoung.kimv, happymin77, kkwak.add}@gmail.com

[2]Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA. {minsikj, junwonse}@andrew.cmu.edu

- An uncertainty-aware continuous 3D semantic mapping framework that extends BKI with anisotropic Gaussian primitives, jointly modeling local geometry, semantics, and uncertainty for robust semantic inference.
- Evidential Ellipsoidal BKI with geometry-aligned anisotropic kernels that conform to local scene structure and incorporate semantic uncertainty for adaptive fusion.
- Comprehensive experiments demonstrating improved accuracy, reliability, and robustness across diverse outdoor environments with real-time performance.

II. RELATED WORK

A. Continuous Semantic Mapping

Voxel representation has been widely adopted in semantic mapping for its simplicity and probabilistic compatibility [1], [3], [5]–[7], where each voxel independently estimates its occupancy and semantic distribution. However, this independence assumption often results in discontinuous maps under sparse sensor data, motivating continuous mapping approaches that incorporate spatial correlations between neighboring voxels. Semantic Bayesian Kernel Inference (S-BKI) [9] addresses this limitation via kernel-based probabilistic inference [14] to interpolate semantic information across space, leveraging neighboring observations to infer semantics at unobserved locations.

Despite its effectiveness, S-BKI relies on a static isotropic kernel, which limits its ability to handle both spatial and semantic uncertainty. To overcome these limitations, several extensions have been proposed that modify the kernels. ConvBKI [13] and SEE-CSOM [12] both address spatial uncertainty caused by geometric misalignment. ConvBKI employs learnable class-wise kernels tailored to each class's geometry, and SEE-CSOM introduces label inconsistency measures to mitigate overinflation at semantic boundaries. Evidential approaches [10], [11] focus on semantic uncertainty in neural network predictions [15], [16], integrating uncertainty estimates into BKI through adaptive kernels. However, these methods primarily address a single type of uncertainty and process noisy observations independently, making them vulnerable to noise. These limitations motivate the design of intermediate representations that can capture local geometric structure, enable uncertainty-aware semantic fusion, and enhance robustness by aggregating local context.

B. Gaussian Scene Representation

Recent advances in Gaussian-based mapping provide a promising basis for such intermediate representations due to their capability for local context abstraction and geometric adaptation. Specifically, GMMap [17] and GIRA [18] demonstrate the effectiveness of Gaussians for efficient 3D scene reconstruction. However, these methods focus on geometric structure and do not incorporate semantics or uncertainty modeling. In parallel, recent efforts in semantic occupancy prediction have explored Gaussians [19] and evidential uncertainty modeling [20], but remain constrained by the lack of precise depth information, limiting their applicability to safety-critical scenarios that require accurate geometric reconstruction. Building on these insights, we propose to leverage anisotropic Gaussian primitives that jointly encode geometry, semantics, and uncertainty for robust continuous semantic mapping in complex environments.

III. SEMANTIC BAYESIAN KERNEL INFERENCE

We first revisit S-BKI [9], as a baseline for continuous semantic mapping, and highlight its key limitations that motivate our work. Building upon Evidential Semantic Mapping (EBS) [10] for semantic uncertainty modeling, we address these limitations through anisotropic Gaussian primitives that enable geometry-aligned inference. For more detailed BKI formulations, we refer readers to [8], [14].

A. Semantic Bayesian Kernel Inference

Let $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^{N}$ denote the set of input semantic points, where $\mathbf{x}_n \in \mathbb{R}^3$ is a 3D coordinate and $\mathbf{y}_n \in \{0,1\}^C$ is a one-hot semantic label over C categories, typically predicted by neural networks [10]–[13]. S-BKI leverages neighboring semantic points to estimate categorical distributions $\hat{\boldsymbol{\theta}}_m = [\hat{\theta}_m^1, ..., \hat{\theta}_m^C]$ at arbitrary query points $\hat{\mathbf{x}}_m$ through kernel inference. The posterior is modeled using the BKI framework [14], which extends the standard likelihood by incorporating spatial correlations through kernels. For online robotic applications with sequential semantic points, S-BKI recursively updates Dirichlet posterior parameters that accumulate evidence for each semantic class:

$$\begin{aligned} p(\hat{\boldsymbol{\theta}}_m|\hat{\mathbf{x}}_m, \mathcal{D}) &\propto p(\mathcal{D}|\hat{\boldsymbol{\theta}}_m, \hat{\mathbf{x}}_m) p(\hat{\boldsymbol{\theta}}_m|\hat{\mathbf{x}}_m) \\ &\propto \prod_{c=1}^{C} (\hat{\theta}_m^c)^{\alpha_0^c + \sum_{n=1}^{N} k(\hat{\mathbf{x}}_m, \mathbf{x}_n) y_n^c - 1}, \end{aligned} \tag{1}$$

where $\boldsymbol{\alpha}_0$ represents the Dirichlet prior, and k is the isotropic kernel [21] that prioritizes nearby observations, with spherical support of radius ℓ :

$$\begin{aligned} k(\hat{\mathbf{x}}_m, \mathbf{x}_n) &= k'(d, \ell) \\ &= \underset{d<\ell}{\mathbb{1}} \Big[\frac{2 + \cos(2\pi \frac{d}{\ell})}{3} (1 - \frac{d}{\ell}) + \frac{1}{2\pi} \sin(2\pi \frac{d}{\ell}) \Big], \end{aligned} \tag{2}$$

where $\mathbb{1}$ is the indicator function, $d = ||\hat{\mathbf{x}}_m - \mathbf{x}_n||$ is the distance. Collectively, the posterior parameters $\boldsymbol{\alpha_m}$ incrementally accumulate evidence at each time step t:

$$\alpha_{m,t}^c \leftarrow \alpha_{m,t-1}^c + \sum_{n=1}^{N} k(\hat{\mathbf{x}}_m, \mathbf{x}_n) \cdot y_n^c, \quad \alpha_{m,0}^c = \alpha_0^c, \tag{3}$$

where $\alpha_0^c \in \mathbb{R}^+$ is an initial hyperparameter for each class. The expectation and variance of $\hat{\boldsymbol{\theta}}_m$ given the Dirichlet parameters $\boldsymbol{\alpha}_m$ are calculated as:

$$S_m = \sum_{c=1}^{C} \alpha_m^c, \ \mathbb{E}[\hat{\theta}_m^c] = \frac{\alpha_m^c}{S_m}, \ \mathrm{Var}[\hat{\theta}_m^c] = \frac{\alpha_m^c (S_m - \alpha_m^c)}{S_m^2 (S_m + 1)}. \tag{4}$$

In this formulation, the semantic map estimate of a query is assigned as $\psi = \arg\max_c \mathbb{E}[\hat{\theta}_m^c]$, with its variance $\mathrm{Var}[\hat{\theta}_m^\psi]$ used as a proxy for the uncertainty of the estimate [9], [13].

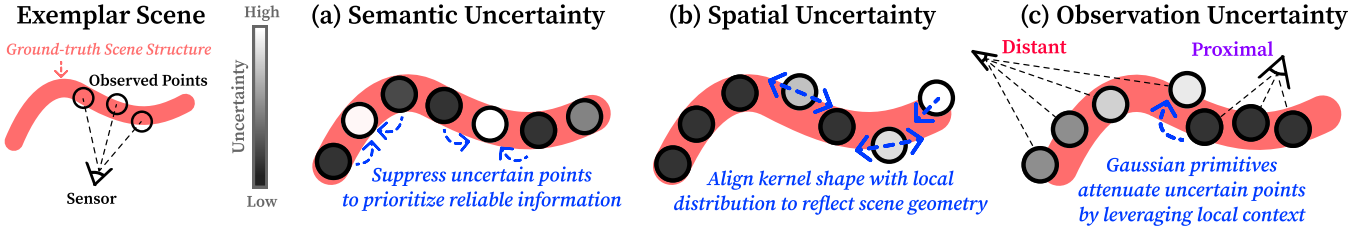

Fig. 2: **Key limitations of S-BKI and our solutions.** The static isotropic kernel in S-BKI does not account for (a) *semantic uncertainty* arising from neural network predictions, (b) *spatial uncertainty* caused by misalignment with local scene geometry, and (c) *observation uncertainty* from distant noisy measurements. Our method addresses these by (a) prioritizing reliable information via uncertainty estimates, (b) adapting kernels to local scene geometry, and (c) leveraging local context through Gaussian primitives.

B. Limitations of S-BKI Framework

While S-BKI enables continuous semantic mapping, its failures in complex environments can be analyzed in terms of three types of uncertainty. We highlight these three aspects and propose corresponding solutions, as illustrated in Fig. 2.

1) Semantic Uncertainty: The update rule (3) assumes uniform reliability across all observations, neglecting semantic uncertainty from neural network predictions. In visually ambiguous, unfamiliar, or poorly illuminated scenes, the confidence of models varies substantially, and predictions can be inconsistent [15], [16]. Therefore, without uncertainty-aware processing, unreliable information is propagated throughout the map with the same weight as reliable predictions, compromising overall mapping quality. The problem intensifies in complex environments where prediction reliability varies widely. To address this limitation, we incorporate semantic uncertainty estimates that enable adaptive weighting of observations, prioritizing confident predictions while attenuating or filtering highly uncertain ones (Fig. 2a).

2) Spatial Uncertainty: The isotropic kernel in (2) applies uniform support across all directions, regardless of local scene geometry. We use the term *spatial uncertainty* to describe the resulting mismatch between this uniform support and the anisotropic nature of real-world structures. When processing elongated structures like roads or fences, spherical kernels propagate semantic information orthogonally to their structural directions, causing geometric misalignment and blurring semantic boundaries, as illustrated in Fig. 1. Moreover, the fixed radius ℓ cannot adapt to varying point densities across the scene, leading to excessive smoothing in dense regions and insufficient coverage in sparse areas. Our framework addresses these issues through geometry-aligned anisotropic kernels that adapt their spatial support to local scene structure, steering influence along principal geometric directions while providing adaptive coverage based on local point density (Fig. 2b).

3) Observation Uncertainty: We use the term *observation uncertainty* to describe the inherent unreliability arising from sensor noise, limited resolution, sparse point coverage, and calibration errors. Although the update rule (3) incorporates neighboring points for the posterior update, each point is processed independently without leveraging local context. This isolation makes the system susceptible to noise, as individual errors directly affect the semantic map without being corrected by nearby consistent observations. As illustrated in Fig. 3, this susceptibility is further exacerbated with increasing sensor distance, where points become sparser and less informative while calibration errors compound the unreliability of distant observations. These limitations motivate our design of spatially coherent primitives that capture local context, enabling noise reduction via local consistency while providing a more robust foundation for semantic inference (Fig. 2c).

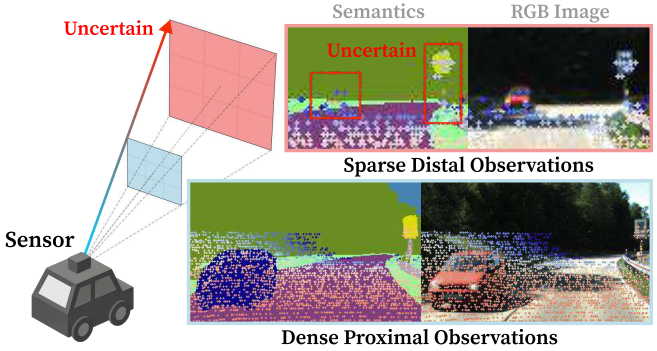

Fig. 3: Observation uncertainty varies with sensor distance. *Left:* sensor-agnostic schematic showing dense, high-resolution proximal observations and sparse, low-detail distal observations. *Right:* RGB image with semantic labels and LiDAR measurements overlaid; sparser returns at longer ranges lead to higher observation uncertainty.

IV. Evidential Ellipsoidal BKI

To overcome the limitations identified in Section III-B, we propose an uncertainty-aware mapping framework that simultaneously addresses semantic, spatial, and observation uncertainty through anisotropic Gaussian primitives. An overview of the full pipeline is illustrated in Fig. 4.

A. Evidential Semantic Segmentation

To quantify uncertainties of semantic predictions, we adopt Evidential Deep Learning (EDL) [22] to extend one-hot semantic predictions $\mathbf{y}_n$ into semantic probability $\mathbf{p}_n = [p_n^1, \ldots, p_n^C]$ with uncertainty estimates u_n, extending the input $\mathcal{D}$ to evidential points $\mathcal{D}' = \{(\mathbf{x}_n, \mathbf{p}_n, u_n)\}_{n=1}^N$. This formulation follows EBS [10], which presents EDL-based uncertainty estimates of semantic prediction and its integration into uncertainty-aware semantic mapping. Based on Dempster-Shafer Theory of evidence (DST) [23], EDL estimates per-class evidence in a single forward pass and converts it into class probability $\mathbf{p}_n$ and corresponding uncertainty u_n. These uncertainty estimates reflect the model's confidence in its predictions, which helps reduce the impact of uncertain predictions during fusion (see [10] for details).

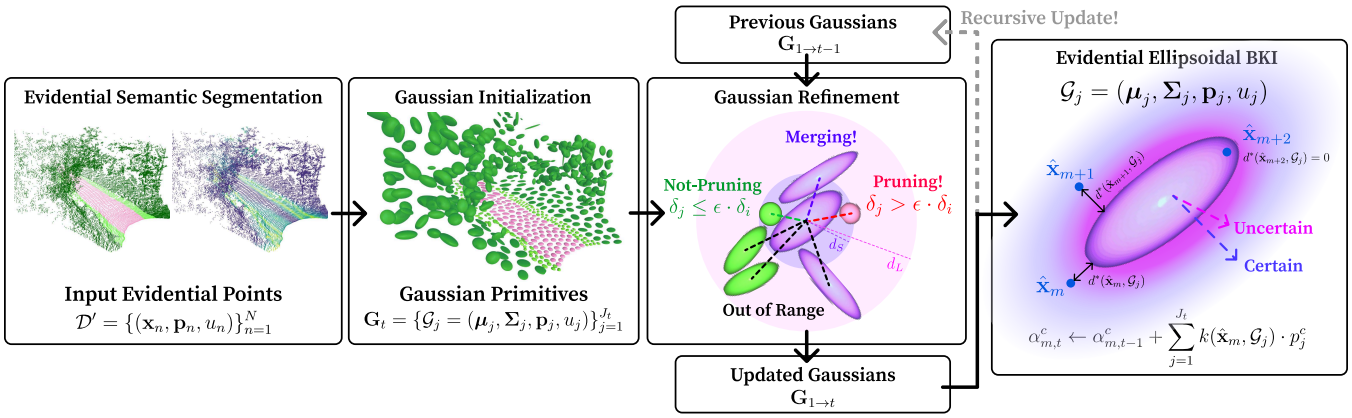

Fig. 4: **Overview of semantic mapping with Evidential Ellipsoidal BKI (E2-BKI).** Given evidential points with semantic probability $\mathbf{p}_n$ and uncertainty u_n (Section IV-A), our method operates through three key stages: (Section IV-B) Gaussian Initialization aggregates evidential points into anisotropic Gaussian primitives encoding local geometry and semantics; (Section IV-C) Gaussian Refinement merges spatially coherent primitives and prunes unreliable primitives; and (Sections IV-D and IV-E) Evidential Ellipsoidal BKI performs uncertainty-aware semantic mapping using Gaussian primitives.

The evidential points are then aggregated into Gaussian primitives through uncertainty-aware fusion to construct robust semantic representations.

B. Gaussian Initialization

While evidential points contain valuable semantic information with uncertainty, individual semantic predictions are unstable and fail to capture spatial context. This limitation motivates the aggregation of evidential points into structured representations that capture local context and reduce sensitivity to noise. Specifically, we adopt 3D Gaussian primitives for their natural ability to capture spatial anisotropy in a compact and mathematically tractable form.

To this end, we begin by clustering the evidential points into J clusters $\{\mathcal{C}_j\}_{j=1}^{J}$. We use the K-Means++ algorithm [24] as a stable default choice, although other clustering strategies are also compatible with our framework. Each cluster $\mathcal{C}_j$ is then abstracted as a *Gaussian primitive* $\mathcal{G}_j = (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \mathbf{p}_j, u_j)$, where the local geometry is modeled by mean $\boldsymbol{\mu}_j$ and covariance $\boldsymbol{\Sigma}_j$. The semantic probability $\mathbf{p}_j$ and uncertainty u_j are derived by aggregating the semantic predictions $(\mathbf{p}_n, u_n)$ from all points in $\mathcal{C}_j$ through DST-based combination rules [25]. Specifically, we convert each probability $\mathbf{p}_n$ into belief mass $\mathbf{b}_n$ with $b^c_n = p^c_n - u_n/C$, and apply the combination rule for combining two belief masses:

$$b^c = \frac{1}{1-\eta}(b^c_1 b^c_2 + b^c_1 u_2 + b^c_2 u_1), \quad u = \frac{1}{1-\eta} u_1 u_2, \tag{5}$$

where C is the number of semantic classes, and $\eta = \sum_{x \neq y} b^x_1 b^y_2$ is a measure of the conflict between two belief masses. We iteratively apply this pairwise combination rule to fuse all $|\mathcal{C}_j|$ belief masses within the cluster, progressively accumulating evidence. Then, the merged belief mass $\mathbf{b}_j$ is converted back into the probability via $p^c_j = b^c_j + u_j/C$. This uncertainty-aware fusion effectively resolves semantic conflicts within each cluster and yields robust semantics with reliable uncertainty estimates (for more details, see [11]).

C. Gaussian Refinement

1) Merging: As BKI incrementally accumulates evidence over time (Section III-A), Gaussian primitives similarly accumulate through iterative initialization. This temporal accumulation can cause redundancy from overlapping regions, while some primitives may consist of insufficient points for stable geometric estimation. To address these issues, we merge nearby primitives with the same semantic label. Geometric components $(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ are merged by combining their statistical moments, while semantic components $(\mathbf{p}_j, u_j)$ are fused through DST-based combination rules as in (5). This merging stabilizes both geometric and semantic representations and reduces computational overhead by decreasing the total number of primitives.

2) Pruning: Additionally, accumulated primitives may exhibit semantic inconsistencies within local regions. These inconsistencies primarily arise from uncertain distant observations, which are both spatially sparse and semantically unreliable (Fig. 3). Inspired by human visual processing that maintains initial impressions until contradicted by more reliable evidence, we adopt a relative pruning strategy. We assess the reliability of each primitive $\mathcal{G}_j$ using its sensor distance δ_j, computed as the average distance from the sensor to the points aggregated in that primitive. Primitives are pruned only when a neighboring primitive $\mathcal{G}_i$ with significantly lower δ_i has conflicting semantics:

$$\operatorname{argmax} \mathbf{p}_i \neq \operatorname{argmax} \mathbf{p}_j \quad \text{and} \quad \delta_j > \epsilon \cdot \delta_i, \tag{6}$$

where ϵ controls the pruning sensitivity. This strategy selectively removes unreliable observations only when contradicted by more reliable ones, preserving uncertain yet potentially informative observations.

D. Evidential Ellipsoidal Kernel

To perform BKI with our Gaussian primitives, we design an evidential ellipsoidal kernel that exploits both the anisotropic geometry and semantic uncertainty of each primitive. The kernel adapts its spatial support to primitive geometry, while incorporating uncertainty estimates to filter unreliable primitives and prioritize reliable ones.

For geometry-aligned kernel support, we compute the distance $d^*(\hat{\mathbf{x}}_m, \mathcal{G}_j)$ by representing each primitive as an anisotropic ellipsoid and finding the closest point $\mathbf{v}^*$ on the

ellipsoid surface to query point $\hat{\mathbf{x}}_m$ through the following minimization over surface points $\mathbf{v}$:

$$d^* = \min_{\mathbf{v}\in\mathbb{R}^3} \|\hat{\mathbf{x}}_m - \mathbf{v}\|^2 \text{ s.t. } (\mathbf{v}-\boldsymbol{\mu}_j)^\top {\boldsymbol{\Sigma}_j}^{-1}(\mathbf{v}-\boldsymbol{\mu}_j) = \tau, \quad (7)$$

where τ controls the ellipsoid size and the resulting minimum $d^*(\hat{\mathbf{x}}_m, \mathcal{G}_j) := d^*$ used in our kernel. For queries inside the ellipsoid, we set $d^*(\hat{\mathbf{x}}_m, \mathcal{G}_j) := 0$. This formulation enables kernels to conform to elongated structures and propagate semantic information along principal geometric directions, addressing the spatial uncertainty of isotropic kernels.

Using this anisotropic distance, our kernel incorporates semantic uncertainty u_j to adaptively modulate kernel support and filter unreliable primitives. We extend the uncertainty-aware kernel function of the evidential semantic mapping framework (EBS [10]) with a geometry-aligned formulation:

$$\tilde{k}(\hat{\mathbf{x}}_m, \mathcal{G}_j) = \begin{cases} k'(d^*(\hat{\mathbf{x}}_m, \mathcal{G}_j), \ell \cdot \beta e^{1-u_j}) & \text{if } u_j \le \mathrm{U_{thr}} \\ 0 & \text{if } u_j > \mathrm{U_{thr}}, \end{cases} \quad (8)$$

where β controls uncertainty sensitivity and $\mathrm{U_{thr}}$ is dynamically set to exclude the most uncertain $\tilde{u}$ percentile of primitives. This formulation ensures that highly uncertain observations are filtered out while smoothly modulating spatial influence based on semantic uncertainty, enabling robust semantic inference (for more details, see [10]). When multiple primitives overlap, their kernel-weighted evidence is accumulated in $\boldsymbol{\alpha}_m$, so primitives with lower uncertainty naturally dominate the posterior semantics while uncertain primitives contribute only weak evidence.

E. Evidential Ellipsoidal BKI

We integrate the evidential ellipsoidal kernel into the BKI framework, enabling continuous semantic mapping with spatial and semantic adaptability. Building upon EBS [10], our method operates on Gaussian primitives with probabilistic semantics $\mathbf{p}_j$ rather than with discrete labels:

$$\alpha^c_{m,t} \leftarrow \alpha^c_{m,t-1} + \sum_{j=1}^{J} \tilde{k}(\hat{\mathbf{x}}_m, \mathcal{G}_j) \cdot p^c_j. \quad (9)$$

This formulation addresses all three limitations in Section III-B: *semantic uncertainty* is mitigated through evidential fusion, *spatial uncertainty* is resolved via geometry-aligned kernels, and *observation uncertainty* is reduced by leveraging local context through Gaussian primitives and uncertainty-aware pruning. These components collectively yield balanced semantic estimates with reduced noise sensitivity, enabling robust semantic inference in complex outdoor environments.

V. EXPERIMENTS

We validate our evidential ellipsoidal BKI framework through comprehensive experiments designed to answer four key research questions: (1) Does our unified uncertainty handling approach improve mapping quality across large-scale off-road and urban environments in terms of both semantic accuracy and uncertainty calibration? (2) How robust is our framework to challenging conditions such as sparse input data and unreliable semantic predictions? (3) Does our Gaussian primitive representation provide versatility across different map formats? (4) What is the individual contribution of each component, and does the framework maintain real-time efficiency?

TABLE I: Hyperparameters for mapping frameworks.

Symbol	α^c_0 (III-A)	ℓ (IV-D)	β (IV-D)	$\tilde{u}$ (IV-D)	ϵ (IV-C)	d_L (IV-C)	d_S (IV-C)
Value	0.001	0.2m	0.75	10%	2.5	5ℓ	ℓ

A. Datasets

1) Off-road Environments: Our evaluation includes two off-road datasets. The first, *RELLIS-3D* [26], provides RGB images and OS1-64 LiDAR scans with 2D semantic annotations and accurate robot poses. We conduct a five-fold evaluation by holding out each sequence once. The second is an extended version of *OffRoad* [10], offering broader spatial coverage and collected using a platform equipped with an OS1-128 LiDAR and RGB camera. It includes manually annotated RGB images and employs geographically disjoint train-test splits. Collected under diverse seasonal, lighting, and terrain conditions, *OffRoad* presents significant challenges for reliable semantic mapping.

2) Urban Environments: For comprehensive evaluation, we also include the *KITTI-360* [27] dataset, which spans 73.7 km of urban driving. We use front-view RGB images and HDL-64E LiDAR scans, along with 2D semantic annotations and GPS/IMU-based localization. We use sequences 00, 02-06 for training; 07, 09 for validation; and 10 for testing.

B. Experimental Setup

1) Implementation Details: We employ LRASPP [28] with EDL [22] for 2D semantic segmentation with semantic uncertainty estimation. The network is trained for 15 epochs on off-road datasets and 10 epochs on urban datasets. Other training details follow the configuration in EBS [10]. The 2D semantic predictions are projected onto 3D point clouds to construct the input $\mathcal{D}'$.

For 3D mapping, we perform a grid search to select the hyperparameters. The chosen values are listed in Table I, and we note that E2-BKI is not sensitive to these values. Gaussian initialization first partitions points by semantic class, then applies K-means++ clustering within each partition. We use 256 clusters for *RELLIS-3D* and 1024 for other datasets to account for variations in point density. During refinement, we first check for semantic consistency within d_L neighborhoods: if all neighboring primitives are semantically consistent, those within d_S are merged. When semantic conflicts exist within d_L and condition (6) is satisfied, the more distant primitive is pruned. The threshold τ is chosen to enclose 10% of the probability mass of each Gaussian.

2) Evaluation Metrics: For fair comparison, all methods are evaluated on a 0.2 m voxelized grid. We adopt metrics established in EBS [10], including per-class IoU and mIoU for semantic accuracy with extending the Accuracy to $\mathrm{Acc} = \sum_{c=1}^{C} \mathrm{TP}_c / \mathbf{Q}$ for joint geometric and semantic evaluation,

TABLE II: Quantitative results on off-road environments (*RELLIS-3D* and *OffRoad*). Semantic classes not present in the dataset are excluded from mIoU and BS and are indicated by a dash (-). Our method consistently outperforms or matches prior continuous semantic mapping approaches.

		Per-class IoU (%)									
Dataset	**Method**	puddle	object	paved	unpaved	dirt	grass	vegetation	mIoU [%]	Acc [%]	BS ↓ [%]
RELLIS 3D	*S-CSM [9]*	37.7	8.1	45.5	-	**19.5**	74.5	77.0	43.7	67.7	15.0
	S-BKI [9]	35.9	8.3	49.4	-	19.0	73.8	76.6	43.8	78.9	16.3
	SEE-CSOM [12]	36.5	10.0	51.5	-	18.6	74.0	77.2	44.6	80.0	14.1
	ConvBKI [13]	35.1	**11.4**	56.0	-	11.2	75.2	76.8	44.3	80.7	15.2
	EBS [10]	39.8	8.2	55.0	-	15.6	74.4	77.6	45.1	80.9	13.2
	Ours	**44.8**	10.7	**57.9**	-	16.1	**75.3**	**79.2**	**47.3**	**83.5**	**13.0**
OffRoad	*S-CSM [9]*	-	45.9	-	80.9	-	61.9	89.8	69.6	58.5	9.2
	S-BKI [9]	-	43.6	-	80.8	-	62.7	89.8	69.2	78.5	14.4
	SEE-CSOM [12]	-	42.9	-	80.2	-	62.2	89.9	68.8	80.3	8.7
	ConvBKI [13]	-	46.7	-	82.0	-	65.4	90.5	71.1	84.3	8.4
	EBS [10]	-	**48.1**	-	83.4	-	65.2	91.1	**71.9**	82.6	**7.0**
	Ours	-	43.9	-	**84.1**	-	**66.5**	**91.4**	71.5	**87.7**	7.4

where TP_c denotes true positives for class c and $\mathbf{Q}$ is the total number of queries. For uncertainty calibration, we employ Brier Score (BS ↓) with normalized $\mathrm{Var}[\hat{\theta}_m^{\psi}]$.

3) Comparison Methods: We include representative baselines where S-CSM [9] serves as the discrete mapping baseline, and S-BKI [9], ConvBKI [13], SEE-CSOM [12], and EBS [10] as continuous mapping approaches. SEE-CSOM [12] addresses spatial uncertainty by incorporating label inconsistency measures to mitigate overinflation at the semantic boundary. ConvBKI [13] learns class-specific kernels, necessitating dataset-specific training. Consequently, we train separate models for each dataset.

C. Evaluations on Semantic Mapping Performance

1) Semantic Accuracy: Our evidential ellipsoidal BKI framework consistently outperforms existing methods across diverse environments (Table II, Table III). Across all datasets, our approach achieves the highest Acc, indicating superior performance in both geometric completeness and semantic accuracy. These improvements stem from geometry-aligned ellipsoidal kernels and primitive-based processing that mitigate spatial uncertainty, while comprehensive uncertainty handling ensures robust semantic representations by prioritizing reliable predictions. Our method also achieves the best mIoU on *RELLIS-3D* and *KITTI-360*, with competitive results on *OffRoad*. The slight mIoU reduction on *OffRoad* relative to EBS reflects an inherent trade-off: improving geometric completeness increases the number of difficult queries in sparse regions, making semantic classification more challenging. Nevertheless, the overall Acc gains outweigh this mIoU difference in practical robotic applications where both geometry and semantics matter.

These quantitative improvements translate into clear visual advantages across all environments, as shown in Fig. 5. On *RELLIS-3D*, we preserve puddle boundaries and produce clean traversable paths by suppressing uncertain predictions, while baselines blur edges and yield incorrect labels. In *OffRoad*, the unpaved road remains clearly separated from grass with noise effectively suppressed, whereas baselines exhibit small isolated errors near boundaries. For urban *KITTI-360*, road geometry remains continuous and free from noisy artifacts, owing to our comprehensive handling of local geometric structures and uncertainties.

TABLE III: Quantitative results on *KITTI-360*. Our method shows superior performance over prior continuous semantic mapping approaches.

	Per-class IoU (%)													
Method	road	sidewalk	building	fence	pole	sign/light	vegetation	terrain	person	car	truck	mIoU [%]	Acc [%]	BS ↓ [%]
S-CSM [9]	88.8	61.5	68.3	47.1	**15.8**	11.4	68.9	59.0	8.0	**67.1**	23.9	47.3	59.5	14.7
S-BKI [9]	88.3	61.3	66.7	47.7	15.0	10.8	68.6	60.4	8.2	65.5	25.9	47.1	74.9	17.4
SEE-CSOM [12]	88.5	61.8	66.8	47.8	14.2	9.0	68.5	60.2	8.0	65.6	25.0	46.9	76.1	13.4
ConvBKI [13]	87.6	64.1	**69.4**	48.6	13.2	13.3	69.9	63.6	8.7	66.9	26.6	48.3	68.0	13.5
EBS [10]	**90.5**	**67.1**	69.3	50.2	14.0	11.9	**70.4**	63.4	9.2	65.6	27.2	49.0	77.8	12.5
Ours	90.1	67.0	68.7	**50.8**	14.3	**14.2**	69.9	**63.7**	**9.7**	64.9	**30.2**	**49.4**	**80.0**	**12.1**

2) Uncertainty Calibration: E2-BKI achieves well-calibrated uncertainty estimates across all datasets, achieving the best BS on *RELLIS-3D* and *KITTI-360* and competitive performance on *OffRoad*. Fig. 7 illustrates how our uncertainty-aware processing addresses unreliable semantic predictions. In the boxed region, S-BKI produces incorrect semantics with weakly informative uncertainty, whereas our framework generates more reliable semantic assignments and preserves higher uncertainty in areas where the evidence is insufficient. Consequently, the predicted uncertainty better reflects the difficulty of each prediction and clearly identifies high-risk regions for downstream modules, enabling accurate semantic maps together with well-calibrated and trustworthy uncertainty information across diverse environments.

D. Evaluations on Robustness and Versatility

1) Robustness under Input Sparsity: To evaluate robustness under sparse sensor data, we reduce the mapping input by using only a subset of available frames while maintaining the same evaluation protocol. Fig. 6a illustrates the robustness of our method to sparse input on *RELLIS-3D*, consistently outperforming baselines as input availability decreases to 4%. This resilience stems from Gaussian primitives that aggregate sparse observations into coherent structures, capturing local continuity despite missing observations. In contrast, point-based methods do not leverage local context and degrade significantly under sparse conditions.

2) Robustness to Unreliable Semantic Predictions: To evaluate the robustness of mapping performance under degraded semantic predictions, we train semantic segmentation networks on progressively smaller training datasets. Fig. 6b presents results on *OffRoad* with training data reduced to 2%. While all methods experience performance degradation, ours consistently achieves the highest Acc across all settings. This robustness stems from our uncertainty-aware processing, which mitigates the impact of unreliable predictions.

3) Scene Representation Versatility: Our Gaussian primitives offer greater flexibility and versatility compared to traditional voxel-based representations. For instance, they can be directly projected onto 2D planes to form BEV semantic maps. Table IV demonstrates that our E2-BKI framework effectively transfers to this different map format,

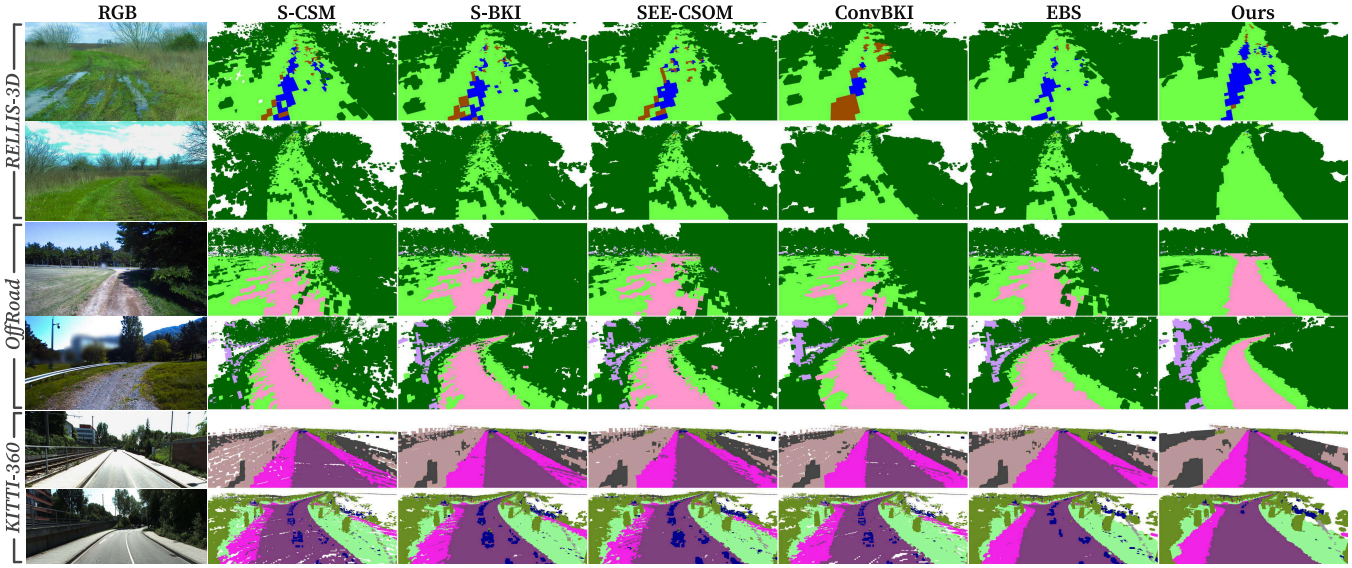

Fig. 5: Qualitative comparison of semantic mapping results on *RELLIS-3D*, *OffRoad*, and *KITTI-360*. Compared to baselines, our method produces more accurate and visually consistent semantic reconstructions across diverse scenes. The color scheme for semantic classes follows Table II and Table III.

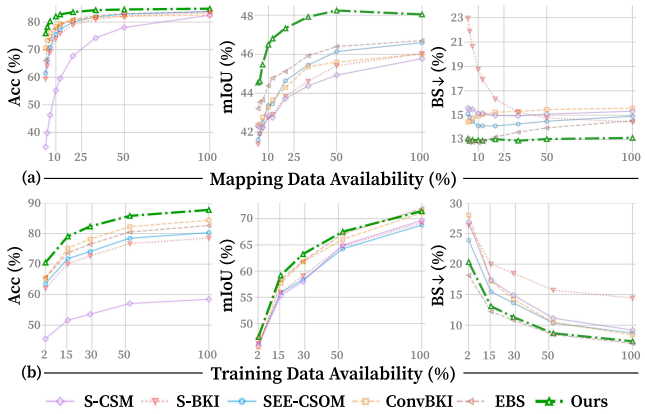

Fig. 6: Robustness evaluation under data limitations. (a) Performance under varying input sparsity on *RELLIS-3D*. (b) Performance with semantic networks trained on limited data of *OffRoad*. E2-BKI maintains the most robust performance under uncertainties arising from data limitations.

achieving consistent improvements across all datasets in BEV projection mode. Beyond BEV mapping, our compact Gaussian representation enables continuous evaluation at arbitrary query points through direct semantic inference. This is achieved by applying our evidential ellipsoidal BKI directly to nearby primitives at any 3D location, eliminating pre-defined voxel grids. Since this continuous mode avoids discretization artifacts by directly querying primitives, further performance improvements are observed across all metrics in Table IV. This flexibility supports diverse mapping applications, including multi-resolution structures, adaptive sampling strategies, and custom query patterns, highlighting the practical advantages of our primitive-based framework.

E. Ablation Studies

We conduct ablation studies on *RELLIS-3D* to isolate the contribution of each component in our framework, as summarized in Table V. Starting from the S-BKI baseline, EBS addresses semantic uncertainty and demonstrates substantial improvements. The introduction of isotropic Gaussian primitives transitions the framework to primitive-based operations, allowing local observations to be aggregated into compact, semantically coherent units. Extending primitives to anisotropic Gaussians enables geometry-aligned kernels, mitigating spatial uncertainty. Local merging improves mapping stability and computational efficiency, while adaptive pruning significantly enhances semantic performance.

TABLE IV: Results on BEV semantic mapping. Our method consistently improves performance in both BEV projection and continuous evaluation modes, with the latter achieving the best performance via direct inference at ground-truth query points without discretization.

	RELLIS-3D			*OffRoad*			*KITTI-360*		
Method	mIoU	Acc	BS ↓	mIoU	Acc	BS ↓	mIoU	Acc	BS ↓
S-CSM [9]	42.5	73.5	15.2	65.1	80.6	11.3	41.6	63.5	14.3
S-BKI [9]	41.9	78.5	18.1	65.6	85.8	18.0	40.2	75.3	20.3
SEE-CSOM [12]	41.3	79.0	14.5	64.8	85.8	10.8	39.4	76.5	14.3
ConvBKI [13]	42.3	80.2	14.2	68.4	86.7	9.5	41.9	71.1	13.2
EBS [10]	44.0	80.4	12.9	66.9	86.4	9.3	42.0	78.1	14.7
Ours (BEV)	46.5	82.9	12.7	70.2	89.5	8.0	43.5	82.5	10.5
Ours (Cont.)	**46.7**	**83.1**	**12.6**	**70.7**	**89.6**	**7.9**	**43.6**	**82.6**	**10.4**

TABLE V: Ablation study results. Each row indicates the mapping performance after cumulative removal of components.

Component(s)	**Section**	mIoU	Acc	BS ↓
Ours	–	**47.3**	**83.5**	**13.0**
- *Gaussian Refinement (Pruning)*	IV-C	45.9	83.2	13.9
- *Gaussian Refinement (Merging)*	IV-C	45.8	83.2	14.3
- *Anisotropic Gaussian Primitive*	IV-B	45.5	83.1	14.4
- *Isotropic Gaussian Primitive* (EBS [10])	IV-B	45.1	80.9	13.2
- *Evidential BKI* (S-BKI [9])	IV-A	43.8	78.9	16.3

F. Runtime Analysis

Finally, we evaluate the computational efficiency of our framework, averaging over five runs on a laptop with an Intel i7-12700H CPU. Gaussian construction achieves 9.73 ± 0.21 Hz, and evidential ellipsoidal BKI operates at 12.10 ± 0.23 Hz. The full framework runs at 5.39 ± 0.11 Hz, faster than

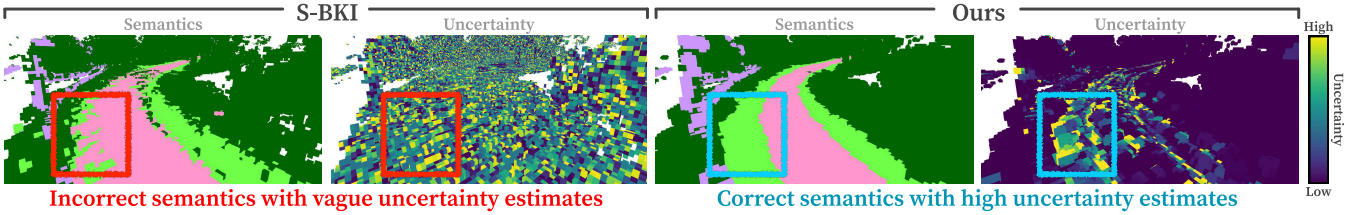

Fig. 7: Semantic and uncertainty map (the third row of Fig. 5). While S-BKI produces incorrect semantics with uncalibrated uncertainty estimates, E2-BKI recovers the correct semantic labels and explicitly flags these challenging regions with higher uncertainty.

the S-BKI baseline (3.15 ± 0.04 Hz). This advantage arises from (1) the abstraction of individual points into Gaussians, which reduces the number of BKI computations, and (2) local merging, which further reduces redundant computation, as demonstrated by the throughput drop to 2.70 ± 0.08 Hz without refinement steps.

VI. CONCLUSIONS

We propose an uncertainty-aware semantic mapping framework that addresses the limitations of S-BKI through anisotropic Gaussian primitives. Our approach tackles semantic uncertainty through evidential deep learning, spatial uncertainty through ellipsoidal kernels, and observation uncertainty through local context aggregation and pruning. By representing local geometry and semantics through Gaussian primitives, our method enables uncertainty-aware semantic propagation of sparse and noisy observations. Extensive experiments demonstrate superior performance in mapping quality and uncertainty calibration. The framework's robustness and versatility across different map representations highlight its practicality for autonomous systems. Future work includes improving uncertainty estimation to obtain more calibrated predictive uncertainties, refining the formal definitions of the uncertainties, and extending the uncertainty-aware mapping framework to open-set settings.

REFERENCES

[1] B.-s. Kim, P. Kohli, and S. Savarese, "3d scene understanding by voxel-crf," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1425–1432.

[2] J. P. Valentin, S. Sengupta, J. Warrell, A. Shahrokni, and P. H. Torr, "Mesh based semantic modelling for indoor and outdoor scenes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2067–2074.

[3] S. Sengupta and P. Sturgess, "Semantic octree: Unifying recognition, reconstruction and representation via an octree constrained higher order mrf," in *Proc. Int. Conf. Robot. Automat.*, 2015, pp. 1874–1879.

[4] D. Paz, H. Zhang, Q. Li, H. Xiang, and H. I. Christensen, "Probabilistic semantic mapping for urban autonomous driving applications," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 2059–2064.

[5] A. Asgharivaskasi and N. Atanasov, "Semantic octree mapping and shannon mutual information computation for robot exploration," *IEEE Trans. Robot.*, 2023.

[6] D. Morilla-Cabello, L. Mur-Labadia, R. Martinez-Cantin, and E. Montijano, "Robust fusion for bayesian semantic mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2023.

[7] J. M. C. Marques, A. J. Zhai, S. Wang, and K. Hauser, "On the overconfidence problem in semantic 3d mapping," in *Proc. Int. Conf. Robot. Automat.*, 2024, pp. 11 095–11 102.

[8] K. Doherty, J. Wang, and B. Englot, "Bayesian generalized kernel inference for occupancy map prediction," in *Proc. Int. Conf. Robot. Automat.*, 2017, pp. 3118–3124.

[9] L. Gan, R. Zhang, J. W. Grizzle, R. M. Eustice, and M. Ghaffari, "Bayesian spatial kernel smoothing for scalable dense semantic mapping," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 790–797, 2020.

[10] J. Kim, J. Seo, and J. Min, "Evidential semantic mapping in off-road environments with uncertainty-aware bayesian kernel inference," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2024, pp. 1420–1427.

[11] J. Kim and J. Seo, "Uncertainty-aware semantic mapping in off-road environments with dempster-shafer theory of evidence," in *Proc. ICRA Workshop on Resilient Off-road Autonomy*, 2024.

[12] 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 Trans. Ind. Electron.*, 2023.

[13] J. Wilson, Y. Fu, J. Friesen, P. Ewen, A. Capodieci, P. Jayakumar, K. Barton, and M. Ghaffari, "Convbki: Real-time probabilistic semantic mapping network with quantifiable uncertainty," *IEEE Trans. Robot.*, 2024.

[14] W. R. Vega-Brown, M. Doniec, and N. G. Roy, "Nonparametric bayesian inference on multivariate exponential families," *Proc. Conf. Neural Inf. Process. Syst.*, vol. 27, 2014.

[15] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1321–1330.

[16] P. de Jorge, R. Volpi, P. H. Torr, and G. Rogez, "Reliability in semantic segmentation: Are we on the right track?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7173–7182.

[17] P. Z. X. Li, S. Karaman, and V. Sze, "Gmmap: Memory-efficient continuous occupancy map using gaussian mixture model," *IEEE Trans. Robot.*, 2024.

[18] K. Goel and W. Tabib, "Gira: Gaussian mixture models for inference and robot autonomy," in *Proc. Int. Conf. Robot. Automat.*, 2024, pp. 6212–6218.

[19] Y. Huang, W. Zheng, Y. Zhang, J. Zhou, and J. Lu, "Gaussianformer: Scene as gaussians for vision-based 3d semantic occupancy prediction," in *Proc. Eur. Conf. Comput. Vis.*, 2024, pp. 376–393.

[20] J. Kälble, S. Wirges, M. Tatarchenko, and E. Ilg, "Evocc: Accurate semantic occupancy for automated driving using evidence theory," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2025, pp. 27 467–27 476.

[21] A. Melkumyan and F. T. Ramos, in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2009.

[22] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," *Proc. Conf. Neural Inf. Process. Syst.*, vol. 31, 2018.

[23] G. Shafer, "Dempster-shafer theory," *Encyclopedia of artificial intelligence*, vol. 1, pp. 330–331, 1992.

[24] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[25] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2551–2566, 2022.

[26] P. Jiang, P. Osteen, M. Wigness, and S. Saripalli, "Rellis-3d dataset: Data, benchmarks and analysis," in *Proc. Int. Conf. Robot. Automat.*, 2021, pp. 1110–1116.

[27] Y. Liao, J. Xie, and A. Geiger, "Kitti-360: A novel dataset and benchmarks for urban scene understanding in 2d and 3d," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3292–3310, 2022.

[28] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for mobilenetv3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.