RIGS: Radar-Informed Gaussian Splatting for Uncertainty-Aware 3D Occupancy and Motion Prediction

Zeyu Han[1], Junzhe Wu[2], Fang Zhang[1](✉), Maani Ghaffari[2], and Jianqiang Wang[1](✉)

[1] Tsinghua University, Beijing, China
hanzy21@mails.tsinghua.edu.cn, {zhang_fang, wjqlws}@tsinghua.edu.cn
[2] University of Michigan, Ann Arbor, USA
{junzhewu, maani}@umich.edu

Abstract. 3D occupancy prediction provides dense semantic scene representations for autonomous driving, yet camera-only methods suffer from occlusion and adverse weather, while existing fusion approaches treat 4D radar as a generic point cloud and neglect its distinct physical measurements. We propose RIGS, a radar-informed Gaussian splatting framework that systematically exploits 4D radar across three stages of the pipeline. In scene modeling, we extract range and power from the 4D radar tensor for depth initialization and fuse them with images through image–radar–Gaussian tri-modal cross-attention. In occupancy refinement, we introduce evidential ellipsoidal BKI with anisotropic kernels and use RCS power as physical existence evidence for adaptive false-positive filtering, yielding uncertainty-aware predictions with Dirichlet epistemic and semantic uncertainty. In velocity estimation, we propose temporal displacement-guided Doppler de-aliasing to recover high-speed radial velocities and complement tangential components, supervised by a multi-level loss over sparse radar points, dense voxels, and connected components. Experiments on K-Radar show that RIGS achieves strong long-range occupancy among camera–radar methods with uncertainty-aware refinement and competitive velocity estimation, validated by systematic incremental ablations.

Keywords: 4D radar · Gaussian splatting · 3D occupancy prediction · Motion prediction

1 Introduction

Autonomous systems need to understand what space is occupied, what occupies it, and how it moves, motivating joint prediction of semantic 3D occupancy and a dense 3D motion field with uncertainty estimates [5, 22, 29]. This is critical in long-range, occluded, or adverse-weather scenarios, where camera-only models often miss distant structures or hallucinate occupancy [30, 37]. 4D radar

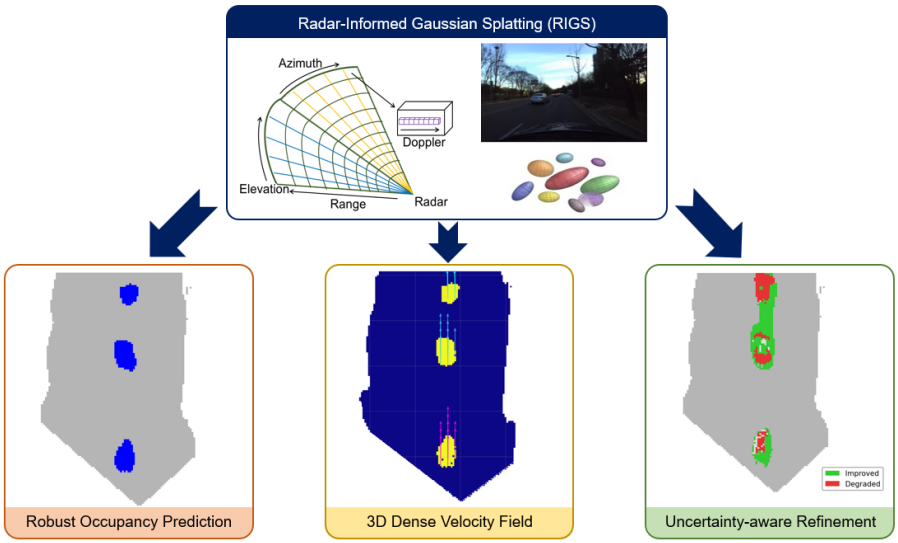

Fig. 1: Given a single camera image and a raw 4D radar tensor, Radar-Informed Gaussian Splatting (RIGS) constructs a radar-guided Gaussian scene representation to jointly predict semantic occupancy, dense 3D velocity, and voxel-wise uncertainty. The uncertainty signal further enables uncertainty-aware refinement, improving occupancy consistency and suppressing hallucinated predictions.

offers a complementary solution, providing weather-robust, direct measurements of range and motion [35, 48, 49]. Importantly, the raw 4D radar tensor preserves rich information across range, Doppler, and angular dimensions, which contain cues beyond point-based radar representations [9]. The key challenge is to exploit radar in a way that improves not only occupancy, but also uncertainty reasoning and motion learning, within a single coherent framework. Our goal is therefore to jointly predict semantic occupancy, dense 3D velocity, and voxel-wise uncertainty, with 4D radar contributing range, power, and Doppler cues at each stage.

Prior work typically addresses only part of this goal. Camera-only occupancy methods leverage strong semantics but degrade under weak geometric evidence [44]. Radar-only approaches are robust yet suffer from clutter and limited semantics [5], and point-based radar representations often discard informative structure [21, 32]. Existing camera–radar fusion pipelines commonly treat radar as auxiliary geometry or generic point clouds, while uncertainty is frequently reduced to heuristic confidence maps that do not distinguish lack of evidence from semantic conflict [18, 31]. Moreover, as the most distinctive cue for dynamics, radar Doppler is rarely integrated as a principled supervision signal for dense motion [25, 41].

We propose RIGS: Radar-Informed Gaussian Splatting for Uncertainty-Aware 3D Occupancy and Motion Prediction, which unifies occupancy, uncertainty, and

motion prediction through radar-informed 3D Gaussian scene anchors. As Fig. 1 shows, RIGS represents a scene as a set of learnable Gaussian ellipsoids that carry geometry and features, fuse camera appearance with radar measurement cues, and are differentiably splatted into a voxel grid to predict semantic occupancy and a dense 3D velocity field. Radar contributes consistently at three levels. (1) **Occupancy**: radar-informed depth priors and tri-modal fusion stabilize 3D anchoring and improve long-range geometry. (2) **Uncertainty**: we treat predicted Gaussians as spatial observations and perform evidential ellipsoid Bayesian kernel inference(BKI) at inference time to obtain a Dirichlet posterior, inject this prior into base predictions with confidence gating, and apply radar cross section(RCS) filtering to suppress implausible occupancy in low-evidence regions. (3) **Motion**: radar Doppler provides a physically grounded learning signal. We incorporate a robust de-aliasing strategy and regularization to learn coherent dense velocity fields aligned with the Gaussian scene structure.

Experiments on K-Radar demonstrate that RIGS improves occupancy prediction, produces meaningful uncertainty for refinement, and enables Doppler-supervised dense motion prediction; incremental ablations verify each component.

Our main contributions are: (1) A radar-informed Gaussian splatting framework for joint 3D occupancy and motion prediction. (2) An uncertainty-aware inference module via evidential ellipsoid BKI with interpretable uncertainty and refinement. (3) A Doppler-supervised motion learning strategy with robust de-aliasing. (4) K-Radar benchmark comparisons and systematic incremental ablations that jointly validate the full pipeline and isolate the contribution of each proposed component. Code will be released at `https://github.com/hanzy21/RIGS`.

2 Related Work

2.1 Semantic Occupancy Prediction

Semantic occupancy prediction has advanced along camera-only, radar-only, and multi-modal fusion approaches, with growing focus on scalability, long-range degradation, and adverse weather.

Camera-only Semantic Occupancy. While scalable, camera-only 3D lifting remains under-constrained in occluded or long-range regions [14, 34, 36]. Many works reduce dense voxel computational costs by factorizing or sparsifying 3D representations (TPVFormer [15], SparseOcc [42]) while transformer-based lifting further improves context aggregation (VoxFormer [23], OccFormer [51]). On the decoding side, compact representations such as COTR [27] and OctFormer [43] reduce redundancy while preserving semantics. RenderOcc [34] and SelfOcc [14] relax dense 3D supervision via rendering and self-supervised consistency, but these signals rely on cross-view or temporal coherence and degrade under occlusion or at distance. Adverse weather further weakens visual cues, motivating radar as an all-weather long-range cue [2, 40].

Radar Occupancy. Radar occupancy leverages measurement-domain cues, including range, Doppler spectra, and return strength. A common pipeline sparsifies radar into points for LiDAR-style learning [21, 28], but this collapse can drop weak spatio-spectral evidence and increase clutter sensitivity. RadarOcc [5] preserves raw 4D radar tensors to avoid point collapse, and 4D-ROLLS [25] explores weak supervision via LiDAR pseudo-label alignment. Even with raw tensors or weak supervision, radar-only occupancy remains semantically coarse and noisy, motivating fusion with complementary visual semantics.

Multi-modal Fusion Occupancy. LiDAR and camera fusion integrates appearance semantics with geometric constraints, with most designs differing mainly in fusion stage and bird's-eye view (BEV) transformation [24, 26, 33, 50]. OccFusion [29] and SurroundOcc [45] establish lift-and-fuse baselines with multi-sensor aggregation and dense refinement. Building on this line, SDGOcc [7] strengthens BEV formation via depth-semantic co-guidance and distillation, while Co-Occ [33] improves training signals with multi-view consistency regularization. REOcc [41] and MetaOcc [49] incorporate radar via radar-specific feature enrichment and attention, indicating radar is not a direct LiDAR replacement due to distinct measurement physics and sparsity. However, camera–radar fusion rarely exploits raw 4D radar tensors jointly with images [31], and direct adaptations of camera–LiDAR fusion architectures to camera–radar input often fail to outperform camera-only baselines; few works directly learn dense 3D velocity from Doppler, motivating our tensor-level radar processing with tri-modal Gaussian scene anchors and Doppler-supervised dense motion learning.

2.2 Gaussian Scene Representation

Gaussian scene representations (3DGS [19]) use explicit anisotropic Gaussian primitives as sparse continuous elements with efficient projection, enabling advances in novel view synthesis, SLAM, and scene reconstruction [10, 17, 47]. GS-SLAM [47] and SplaTAM [17] extend Gaussian primitives to online SLAM, supporting tracking and mapping via incremental optimization of a 3D Gaussian map with differentiable rendering. GaussianFormer [16] and GaussianFormer-2 [13] use Gaussians as perception tokens for semantic occupancy, with the latter improving token efficiency through probabilistic Gaussian superposition to better match scene sparsity. Following this idea, we enrich Gaussian primitives with radar cues so that each token jointly encodes geometry, semantics, motion, and reliability, enabling uncertainty-aware dynamic occupancy and dense 3D velocity inference.

2.3 Uncertainty-Aware Mapping

Uncertainty-aware semantic mapping studies how to represent semantic beliefs while propagating them coherently in space under sparse and noisy observations [11]. Evidential Deep Learning [39] models categorical predictions with Dirichlet evidence to separate low-evidence uncertainty from class ambiguity. In parallel, continuous mapping methods such as Hilbert Maps [38] and Bayesian

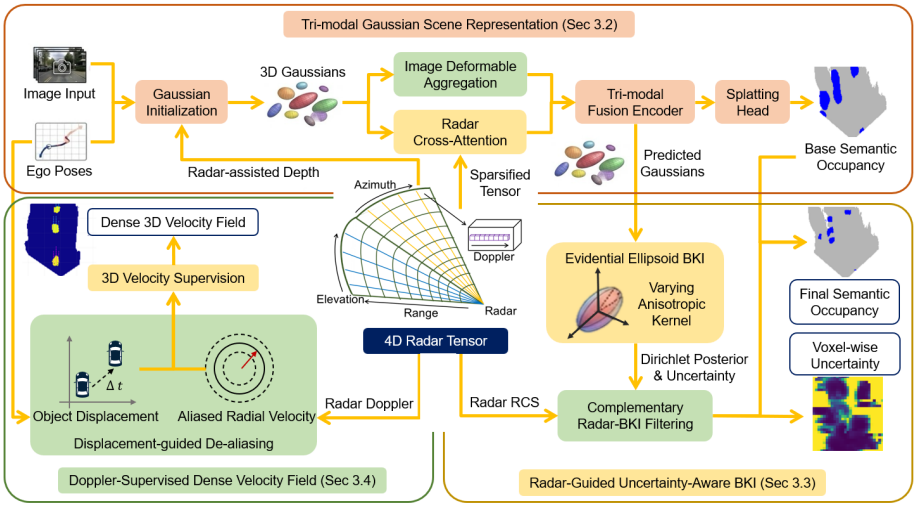

Fig. 2: Overview of our pipeline: tri-modal Gaussian scene representation, radar-guided uncertainty-aware BKI, and Doppler-supervised dense velocity field.

kernel inference [6, 8] avoid independent grids by using spatially correlated updates, extending from binary occupancy to multi-class semantics. ConvBKI [46] improves practicality by embedding kernel Bayesian updates into convolutional operators for efficient semantic mapping. E2-BKI [20] further bridges these lines by injecting evidential semantic uncertainty into ellipsoid-aligned kernel inference using anisotropic Gaussian primitives. In contrast, our refinement is additionally constrained by radar measurement-domain reliability cues, enabling robust inference-time updates for occupancy and motion prediction.

3 Method

3.1 Problem Setup and Overview

We jointly estimate (i) a semantic occupancy grid $\mathbf{O}_t$, (ii) a dense 3D velocity field $\mathbf{V}_t$, and (iii) voxel-wise uncertainty $\mathbf{U}_t$ from a front-view image, a raw 4D radar tensor, and short-term ego motion.

4D radar tensor. FMCW radar outputs a 4D radar tensor $\mathcal{R}_t$ indexed by range, azimuth, elevation, and Doppler; each cell stores reflected power (related to radar cross section). This representation preserves spatio-spectral structure that is lost when radar is collapsed into a generic point cloud. Our subsequent sparsification retains informative cells rather than discarding the tensor format.

Dirichlet uncertainty. We model semantic beliefs with a Dirichlet distribution $\mathrm{Dir}(\boldsymbol{\alpha})$, where total evidence $S = \sum_c \alpha^c$ reflects observation support and $\boldsymbol{\alpha}/S$ gives class probabilities. Low S indicates *epistemic uncertainty* from insufficient evidence, whereas a flat $\boldsymbol{\alpha}/S$ indicates *semantic ambiguity* from con-

flicting class support. Below, evidential BKI accumulates Gaussian observations into voxel-wise Dirichlet posteriors for interpretable refinement.

Given image $\mathbf{I}_t$, raw 4D FMCW radar tensor $\mathcal{R}_t$ (range–azimuth–elevation–Doppler), and ego poses $\mathbf{T}_{t-1:t} \in SE(3)$, we formalize the joint prediction as

$$(\mathbf{O}_t, \mathbf{V}_t, \mathbf{U}_t) = \mathcal{F}(\mathbf{I}_t, \mathcal{R}_t, \mathbf{T}_{t-1:t}), \tag{1}$$

where $\mathbf{O}_t \in \{0, 1, \ldots, C\}^{H \times W \times Z}$, $\mathbf{V}_t \in \mathbb{R}^{H \times W \times Z \times 3}$, and $\mathbf{U}_t$ encodes voxel-wise uncertainty derived from the Dirichlet posterior.

Our pipeline has three stages (Fig. 2).

(1) **Radar–camera tri-modal Gaussian scene representations.** We represent the scene with learnable 3D Gaussians initialized from radar-assisted depth proposals, together with a temporal prior. A tri-modal fusion encoder injects image and radar into the Gaussians, which are then rendered into voxel logits for occupancy and velocity.

(2) **Uncertainty-aware inference via evidential ellipsoid BKI.** At inference time, we treat predicted Gaussians as observations and perform evidential ellipsoid BKI over voxel centers to obtain a Dirichlet posterior. We inject this semantic prior into base predictions with Dirichlet-confidence gating, and use RCS as a complementary physical cue for suppressing hallucinated predictions.

(3) **Doppler-supervised velocity learning.** We use radar Doppler for motion supervision and address Doppler aliasing and missing tangential components with displacement-guided de-aliasing, yielding robust supervision for learning a dense 3D velocity field.

3.2 Tri-modal Gaussian Scene Representation

Gaussian Parameterization. We represent the scene using a set of N Gaussian ellipsoids $\{G_i\}_{i=1}^{N}$. Each G_i is parameterized by its center $\boldsymbol{\mu}_i \in \mathbb{R}^3$, scale $\mathbf{s}_i \in \mathbb{R}^3_+$, rotation $\mathbf{q}_i$, opacity $\alpha_i \in [0, 1]$, latent feature $\mathbf{f}_i \in \mathbb{R}^D$, covariance $\boldsymbol{\Sigma}_i$, and prediction heads for semantics and motion.

Radar Tensor Sparsification. Following RadarOcc [5], we construct sparse radar observations from the raw 4D radar tensor with two refinements. The raw power tensor is stabilized using a clamped logarithmic transform, and Doppler responses are aggregated via mean–max fusion for spatial selection [4, 12]. For each range bin, the top-K_p azimuth–elevation cells are retained and projected into Cartesian coordinates, yielding sparse radar observations

$$\mathcal{P}_{\text{radar}} = \{(\mathbf{p}_j, \mathbf{a}_j)\}_{j=1}^{N_r},$$

where $\mathbf{p}_j \in \mathbb{R}^3$ denotes the 3D position, $\mathbf{a}_j$ is an 8-D Doppler descriptor composed of the top-K_d powers, their bin indices, and the mean and standard deviation of the Doppler spectrum.

Radar and Temporal Prior for Gaussian Initialization. To obtain geometrically stable Gaussian centers, we combine radar-derived depth cues with a temporal prior. The sparsified radar observations are projected to the image

plane to provide candidate depth measurements. For each pixel, nearby radar returns form a small set of depth hypotheses, which are back-projected into 3D to initialize Gaussian centers $\{\boldsymbol{\mu}_i^{\text{cur}}\}$.

For temporal consistency, we propagate previous-frame centers using ego-motion. Each current center is matched to its nearest propagated Gaussian and fused if the distance is below τ_{match}. The same correspondence updates the remaining Gaussian parameters.

Tri-modal Fusion Encoder. We extract multi-scale image features with a CNN and FPN backbone. Stacked fusion layers then refine Gaussian features by integrating image appearance and radar measurements, using Gaussians as explicit scene anchors. First, radar observations are encoded into motion-aware radar embeddings

$$\mathbf{V}_{\text{radar}} = \text{MLP}_{\text{radar}}\Big([\text{Norm}(\mathbf{P}_{\text{radar}}); \mathbf{A}_{\text{radar}}]\Big) \in \mathbb{R}^{B\times N_r\times D}, \tag{2}$$

where $\mathbf{P}_{\text{radar}}$ and $\mathbf{A}_{\text{radar}}$ stack $\{\mathbf{p}_j\}$ and $\{\mathbf{a}_j\}$, and $\text{Norm}(\cdot)$ normalizes coordinates within scene bounds. Let $\mathbf{F}_{\text{inst}} \in \mathbb{R}^{B\times N\times D}$ denote the current Gaussian features and $\mathbf{E}_{\text{anc}} \in \mathbb{R}^{B\times N\times D}$ an anchor embedding derived from Gaussian geometry. Multi-head cross-attention is applied with

$$\mathbf{Q} = \mathbf{F}_{\text{inst}} + \mathbf{E}_{\text{anc}}, \quad \mathbf{K} = \mathbf{V}_{\text{radar}}, \quad \mathbf{V} = \mathbf{V}_{\text{radar}}, \quad \mathbf{F}_{\text{radar}} = \text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}), \tag{3}$$

producing $\mathbf{F}_{\text{radar}} \in \mathbb{R}^{B\times N\times D}$. Each Gaussian attends only to its nearest K radar points, and computation is performed in chunks for efficiency.

Second, for each Gaussian anchor, 3D sampling points generated from fixed and learnable offsets are projected to the image plane to sample multi-scale FPN features. Geometry-aware weights predicted from Gaussian parameters aggregate features across sampling points and pyramid levels, yielding $\mathbf{F}_{\text{img}} \in \mathbb{R}^{B\times N\times D}$.

Third, the two modalities are fused via residual refinement:

$$\mathbf{F}_{\text{out}} = \mathbf{F}_{\text{in}} + \text{Proj}_{\text{img}}(\mathbf{F}_{\text{img}}) + \sigma(g)\,\text{Proj}_{\text{radar}}(\mathbf{F}_{\text{radar}}), \tag{4}$$

where g is a learnable scalar initialized such that $\sigma(g) = 0.5$. Each layer further refines Gaussian parameters using lightweight MLP heads and sparse 3D convolutions over aligned voxel features. Gaussians are then rendered onto a voxel grid by accumulating contributions at each voxel center $\hat{\mathbf{x}}_m$; a differentiable splatting head predicts voxel-wise semantic probabilities $p_m^{c,\text{base}}$ and velocity $\mathbf{v}_m^{\text{base}} \in \mathbb{R}^3$.

3.3 Radar-Guided Uncertainty-Aware BKI

Since the base voxel predictions from splatting cannot separate epistemic and semantic uncertainty sources, they are refined at inference by evidential ellipsoid BKI, which yields a Dirichlet posterior over voxel semantics. We adapt E2-BKI [20] to our radar–camera Gaussian representation with two extensions: (1) uncertainty-modulated anisotropic kernels for reliability-aware evidence propagation, and (2) RCS as a complementary cue to suppress hallucinated occupancy.

Gaussian Observations and Evidential BKI. We treat predicted Gaussians $G_j = (\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j, \mathbf{p}_j, u_j)$ as spatial observations and perform evidential kernel inference at voxel centers $\hat{\mathbf{x}}_m$. Specifically, we accumulate class evidence with a Dirichlet prior $\boldsymbol{\alpha}_0$:

$$\alpha_m^c = \alpha_0^c + \sum_j \tilde{k}(\hat{\mathbf{x}}_m, G_j)\, p_j^c, \qquad S_m = \sum_{c=1}^{C} \alpha_m^c, \qquad \mathbf{p}_m^{\text{bki}} = \boldsymbol{\alpha}_m / S_m. \tag{5}$$

Here S_m measures the amount of accumulated evidence, while $\mathbf{p}_m^{\text{bki}}$ is the posterior mean semantic distribution. The Dirichlet formulation exposes uncertainty through both the total evidence S_m and the class distribution $\boldsymbol{\alpha}_m$.

Uncertainty-Aware Anisotropic Kernel. We use an anisotropic compact-support kernel over Gaussians, where the geometry-aware distance $d(\hat{\mathbf{x}}_m, G_j)$ is induced by $\boldsymbol{\Sigma}_j$. To prevent unreliable observations from propagating excessive evidence, we modulate the kernel radius using the predicted uncertainty u_j:

$$\tilde{k}(\hat{\mathbf{x}}_m, G_j) = k'\big(d(\hat{\mathbf{x}}_m, G_j),\ \ell \cdot \beta e^{1-u_j}\big), \tag{6}$$

where $k'(\cdot)$ denotes the compact-support kernel, ℓ is the base kernel radius, and β is a scaling factor. Reliable Gaussians therefore propagate evidence over larger spatial support, while uncertain ones contribute only locally. This adaptive mechanism improves robustness in sparse radar regions where observations are unevenly distributed.

BKI Semantic Prior Injection. For voxels with sufficient accumulated evidence ($S_m > S_{\text{thr}}$), we inject the BKI semantic prior into the base splatting prediction using Dirichlet confidence as a spatial attention weight. We define $c_m = \max_c \alpha_m^c / S_m$ and $w_m = w_{\text{bki}} \cdot c_m$, then blend

$$\hat{p}_m^c = (1 - w_m)\, p_m^{c,\text{base}} + w_m\, \alpha_m^c / S_m. \tag{7}$$

When neighboring Gaussians agree on a class, $c_m \to 1$ and the BKI prior dominates; in ambiguous regions, injection is attenuated.

Complementary Radar–BKI Filtering. After prior injection, neural networks may still produce high-confidence false positives, especially for distant vehicles or visually ambiguous structures. Radar provides a complementary physical existence cue through strong reflections. For each voxel center $\hat{\mathbf{x}}_m$, we search radar points within radius R:

$$\mathcal{N}_m = \{(\mathbf{p}_j, \mathbf{a}_j) \in \mathcal{P}_{\text{radar}} : \|\mathbf{p}_j - \hat{\mathbf{x}}_m\| < R\}. \tag{8}$$

We then compute the maximum reflection strength

$$P_{\text{max},m} = \max_{(\mathbf{p}_j, \mathbf{a}_j) \in \mathcal{N}_m} P_j. \tag{9}$$

Radar reflections and evidential BKI provide complementary cues with different reliability regimes: strong reflections give physical evidence of object existence, while evidential BKI enforces semantic consistency across neighboring observations, motivating a complementary filtering strategy.

If strong reflections are present ($P_{\max,m} > \tau$), we mark the voxel as *radar-supported* and apply a relaxed foreground threshold to $\hat{p}_m$ to preserve true objects. Otherwise, the voxel is *radar-unsupported*, where hallucinations are more likely, and we apply a stricter threshold to $\hat{p}_m$, jointly considering S_m and α_m^c/S_m to suppress unreliable predictions.

This mechanism leverages radar where available and evidential reasoning elsewhere, effectively reducing false positives without sacrificing recall.

3.4 Doppler–Supervised Dense Velocity Field

FMCW radar provides per-return Doppler as a precise but *radial-only* motion cue, yet its limited unambiguous range folds ego-compensated velocities and makes supervision unreliable for fast movers. We therefore build robust 3D pseudo-labels via displacement-guided Doppler de-aliasing to recover true radial motion and complete tangential components using temporal displacement, addressing single-view degeneracy.

Aliased Residual Radial Velocity. For each radar return with ray direction $\hat{\mathbf{u}}$, we denote the ego-compensated residual radial velocity measurement as $v_{\text{base}} \in [-v_{\max}, v_{\max})$, which is a wrapped version of the true radial velocity:

$$v_{\text{true}} = v_{\text{base}} + k \cdot V_{\text{unamb}}, \qquad k \in \mathbb{Z},\ V_{\text{unamb}} = 2v_{\max}. \tag{10}$$

The unknown folding number k varies across different returns and is the main obstacle for Doppler supervision.

Displacement–Guided De-aliasing and 3D Recovery. We estimate k using temporal displacement of dynamic objects extracted from occupancy labels. Let O_n^t denote a dynamic object at time t with centroid $\bar{\mathbf{p}}_n^t$, and $\bar{\mathbf{p}}_{m^*}^{t-1\to t}$ be its matched centroid from the previous frame after ego-motion warping. The displacement velocity is

$$\mathbf{v}_{\text{disp}} = \frac{\bar{\mathbf{p}}_n^t - \bar{\mathbf{p}}_{m^*}^{t-1\to t}}{\Delta t}. \tag{11}$$

For each radar return i inside O_n^t, we resolve the folding offset by comparing the radial projection of $\mathbf{v}_{\text{disp}}$ with $v_{\text{base},i}$:

$$k_i = \text{clamp}\left(\text{round}\left(\frac{\mathbf{v}_{\text{disp}}\cdot\hat{\mathbf{u}}_i - v_{\text{base},i}}{V_{\text{unamb}}}\right), -K, K\right), \quad v_i = v_{\text{base},i} + k_i V_{\text{unamb}}. \tag{12}$$

Given de-aliased radial targets $\{(\hat{\mathbf{u}}_i, v_i)\}$, we estimate the object 3D velocity by least squares:

$$\mathbf{v}_n^{\text{LS}} = \arg\min_{\mathbf{v}} \sum_{i\in O_n^t} \left(\mathbf{v}\cdot\hat{\mathbf{u}}_i - v_i\right)^2. \tag{13}$$

With a single radar, ray directions within an object can be nearly parallel, making tangential components ill-conditioned. We therefore keep the estimated radial component while injecting tangential motion from displacement:

$$\bar{\hat{\mathbf{u}}} = \text{mean}_{i\in O_n^t}\hat{\mathbf{u}}_i, \quad \mathbf{v}_\perp = \mathbf{v}_{\text{disp}} - (\mathbf{v}_{\text{disp}}\cdot\bar{\hat{\mathbf{u}}})\bar{\hat{\mathbf{u}}}, \quad \mathbf{v}_n^* = (\mathbf{v}_n^{\text{LS}}\cdot\bar{\hat{\mathbf{u}}})\bar{\hat{\mathbf{u}}} + \mathbf{v}_\perp. \tag{14}$$

This yields reliable pseudo-label supervision: Doppler provides high-resolution radial motion after de-aliasing, while displacement contributes coarse but unambiguous tangential cues.

Training Targets and Losses. We supervise the predicted voxel velocity field $\{\mathbf{v}_m\}$ using the object-level pseudo velocity $\mathbf{v}_n^*$. Specifically, we apply (i) a sparse point and voxel loss on voxels intersecting radar returns, (ii) a dense occupied-voxel loss within each dynamic object, and (iii) within-object consistency plus spatial smoothness regularization:

$$\mathcal{L}_{\text{vel}} = \lambda_p \sum_{i \in \mathcal{I}_p} \|\mathbf{v}_{m(i)} - \mathbf{v}^*_{n(i)}\|_1 + \lambda_v \sum_{m \in \mathcal{I}_{\text{occ}}} \|\mathbf{v}_m - \mathbf{v}^*_{n(m)}\|_1 \tag{15}$$

$$+ \lambda_c \sum_n \sum_{m \in O_n^t} \|\mathbf{v}_m - \bar{\mathbf{v}}_{O_n}\|_1 + \lambda_s \sum_m \|\nabla \mathbf{v}_m\|_1, \tag{16}$$

where $m(i)$ maps a radar return to its voxel, $n(i)$ (or $n(m)$) denotes the object assignment, and $\bar{\mathbf{v}}_{O_n}$ is the mean predicted velocity within object O_n^t. The final objective adds $\mathcal{L}_{\text{vel}}$ to the occupancy and semantic losses.

4 Experiments

We evaluate RIGS on K-Radar [32] for the proposed camera–radar semantic occupancy and velocity prediction framework. We first describe the setup and protocol, then report quantitative and qualitative occupancy results, analyze uncertainty estimates derived from the Dirichlet posterior, and evaluate radar Doppler-supervised dense 3D velocity fields. Finally, we present incremental ablations to quantify the contribution of key components.

4.1 Experimental Setup

Dataset. We evaluate on the K-Radar dataset [32], which provides synchronized camera images and 4D FMCW radar tensors. Semantic occupancy ground truth is constructed from temporally aggregated LiDAR point clouds with motion compensation. Following prior work, we use a single front-view camera together with the 4D radar tensor as input. For controlled evaluation, we select daytime sequences under normal weather conditions for training, validation, and testing.

Implementation Details. Image features are extracted using a ResNet-101 backbone with a 4-level FPN. Our model maintains $N = 25600$ Gaussian primitives and predicts semantic occupancy on a $128 \times 128 \times 20$ voxel grid with voxel size 0.4 m. The Gaussian encoder has four fusion layers combining image aggregation, radar cross-attention, and sparse 3D convolution. The model is trained with AdamW for 20 epochs on a single NVIDIA A100 GPU.

Metrics. Occupancy performance is evaluated using mIoU, IoU_2, foreground IoU, and background IoU within three distance ranges (12.8 m, 25.6 m, and

Table 1: Occupancy and semantic segmentation results on K-Radar.

Method	Modality	IoU_2 [%]			mIoU [%]			BG IoU [%]			FG IoU [%]		
		12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m
SDGOcc	C+L	37.75	38.85	36.02	24.34	26.81	25.13	36.95	38.80	35.78	11.76	14.82	14.48
MonoScene	C	29.38	30.54	28.91	20.90	21.88	20.23	30.84	32.01	30.31	10.97	11.76	10.15
OccFormer	C	31.03	32.26	31.02	**21.50**	22.68	21.43	**31.47**	32.75	31.42	**11.53**	12.61	11.44
GaussianFormer-2	C	30.08	31.57	31.56	21.02	22.33	22.08	31.21	32.11	32.02	10.83	12.55	12.14
RadarOcc	R	**32.33**	32.29	32.11	20.34	21.88	21.24	30.19	32.06	31.88	10.48	11.71	10.60
SDGOcc	C+R	31.34	32.46	29.47	19.52	20.69	18.81	30.10	31.18	28.18	8.94	10.21	9.44
OccFusion	C+R	31.20	32.68	32.71	21.12	22.49	21.72	31.16	33.02	32.61	11.08	11.96	10.83
RIGS	C+R	30.61	**33.35**	**32.87**	21.21	**23.77**	**23.72**	31.25	**34.00**	**33.34**	11.17	**13.53**	**14.09**

Modality: L = LiDAR, C = camera, R = radar.

51.2 m). Velocity estimation is evaluated using the mean absolute error (MAE) and root mean squared error (RMSE).

Baselines. We compare against camera-only baselines (MonoScene [3], OccFormer [51], GaussianFormer-2 [13]), a radar-only baseline (RadarOcc [5]), and camera–radar fusion baselines obtained by adapting OccFusion [29] and SDGOcc [7] to radar input. We also include SDGOcc with 64-beam LiDAR input as a reference.

4.2 Occupancy Prediction and Semantic Segmentation

Quantitative Results Table 1 reports occupancy and semantic segmentation results on K-Radar. Among compared camera–radar methods, RIGS achieves the best mIoU and FG IoU at 25.6 m and 51.2 m. In particular, RIGS attains the highest mIoU, BG IoU, and FG IoU at 25.6 m and 51.2 m, demonstrating improved long-range semantic occupancy prediction.

At short range, camera-based methods such as OccFormer achieve strong semantic accuracy due to rich visual appearance cues, while RadarOcc obtains the highest IoU_2 because dense near-range radar returns provide reliable occupancy existence signals.

Existing camera–radar fusion methods such as SDGOcc and OccFusion do not consistently outperform camera-only models, indicating that directly adapting camera–LiDAR fusion architectures to 4D radar tensor inputs fails to properly account for radar sparsity and measurement uncertainty. In contrast, the proposed radar-informed representation enables radar observations to consistently benefit semantic occupancy prediction, particularly at longer distances.

Qualitative Evaluation under Adverse Weather. K-Radar includes heavy rain and snow sequences. Because adverse weather can also degrade LiDAR and reduce ground-truth reliability, we report qualitative comparisons for these challenging conditions only.

Fig. 3 shows that the camera-only OccFormer and the camera–radar fusion baseline OccFusion both struggle with distant occupancy under adverse weather, as image corruption and reduced visibility weaken visual cues, causing missed detections and blurred boundaries. Radar-only RadarOcc recovers part

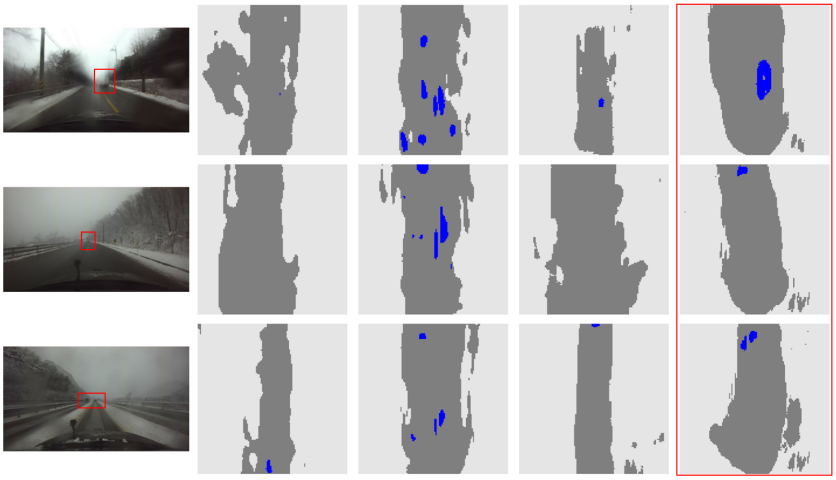

Fig. 3: Qualitative comparison under adverse weather. From left to right in each row: front-view image, OccFormer, RadarOcc, OccFusion, and our proposed method.

of the true foreground occupancy thanks to radarś robustness to weather, but often produces false positives due to sparse radar returns and limited semantic constraints. Our method maintains a better balance: radar provides geometric support when visual cues degrade, while BKI uncertainty modulation and radar RCS-based filtering help suppress spurious occupancy.

4.3 Uncertainty Evaluation

We evaluate whether the Dirichlet posterior produced by the proposed evidence ellipsoid-kernel BKI yields interpretable uncertainty signals that disentangle **epistemic** and **semantic** sources.

At each voxel center $\hat{\mathbf{x}}_m$, BKI outputs a Dirichlet posterior $\mathrm{Dir}(\boldsymbol{\alpha}_m)$ over C semantic classes ($C = 3$ in our experiment). We define the total evidence as $S_m = \sum_{c=1}^{C} \alpha_m^c$ and the predictive mean as $\mathbf{p}_m = \boldsymbol{\alpha}_m / S_m$. Epistemic uncertainty is measured by bounded vacuity $u_m = \frac{C}{C+S_m}$, while semantic uncertainty is measured by entropy $H(\mathbf{p}_m)$ [1]. All uncertainty statistics are computed over occupied voxels with $S_m > 0.01$ to exclude prior-dominated regions.

Interpretability Analysis. We validate interpretability by checking whether the two signals respond to region splits that isolate observation coverage (epistemic) and class ambiguity (semantic).

(1) **Coverage Split (Epistemic).** We partition voxels into *radar-supported* and *radar-unsupported* regions according to nearby radar returns. As shown in Table 2, radar-supported voxels accumulate more evidence ($S = 0.846$ vs. 0.693) and exhibit lower vacuity (0.828 vs. 0.842), indicating reduced epistemic uncertainty when radar observations are available.

(2) **Ambiguity Split (Semantic).** We further separate voxels into a 1–2 voxel *boundary band* around ground-truth class transitions and *interior* regions.

Table 2: Interpretability analysis of Dirichlet uncertainty.

Region split	Voxels	Strength $S \uparrow$	Vacuity $u \downarrow$	Entropy $H(\mathbf{p}) \downarrow$
Radar-supported	1,560,803	**0.846**	**0.828**	–
Radar-unsupported	5,502,980	0.693	0.842	–
Boundary band	4,013,014	–	–	0.0925
Interior	3,050,769	–	–	**0.0900**

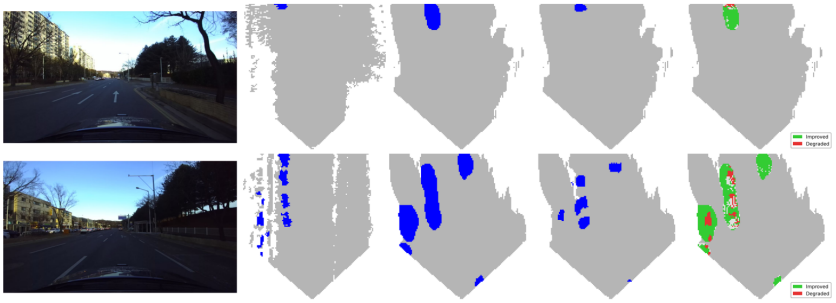

Fig. 4: Qualitative results of uncertainty-aware refinement in BEV. From left to right in each row: front-view image, ground truth, base prediction, refined prediction, and prediction difference (green: improved, red: degraded).

Entropy is higher near boundaries (0.0925 vs. 0.0900), consistent with increased semantic ambiguity at class interfaces.

Overall, the observed trends match the intended decomposition: vacuity primarily reflects evidence coverage, while entropy concentrates around semantic transitions.

Qualitative Effect of Uncertainty–Aware Refinement. Fig. 4 shows qualitative BEV results of our uncertainty-aware refinement. It improves long-range localization and removes hallucinated occupancy from the base model, with most new errors confined to thin boundary bands, indicating a favorable precision–recall trade-off. These results suggest the uncertainty signals effectively correct unreliable regions while preserving confident predictions.

4.4 3D Velocity Field Estimation

In this section, we evaluate dense 3D velocity estimation enabled by radar Doppler supervision.

Quantitative Results. To analyze different motion regimes, we report results on two voxel subsets defined by ground-truth semantic labels: (i) static voxels ($\mathcal{V}_{sta}$) and (ii) dynamic voxels ($\mathcal{V}_{dyn}$). We compare our full model with two simplified baselines that isolate individual motion cues. **DispCentroid** estimates velocity from temporal centroid displacement, using only temporal cues without Doppler supervision. **DopplerSup** directly supervises velocity from

Table 3: 3D velocity prediction errors under voxel-level evaluation (m/s). Higher RMSE is caused by a small portion of challenging dynamic cases.

Method	$\mathcal{V}_{sta}$		$\mathcal{V}_{dyn}$	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
DispCentroid	0.25	1.47	2.62	6.79
DopplerSup	0.43	1.86	3.48	8.34
Ours	**0.06**	**1.21**	**2.37**	**5.95**

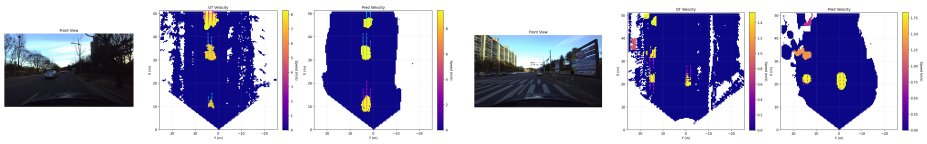

Fig. 5: Qualitative velocity field visualization. (a)-(b) show representative scenes. Left to right: front-view image, ground-truth BEV velocity field, and our prediction.

radar Doppler but does not resolve Doppler aliasing. Our full model combines both cues with the proposed displacement-guided de-aliasing.

Table 3 reports voxel-level velocity errors. While static voxels ($\mathcal{V}_{\text{sta}}$) dominate the evaluation set and therefore yield very small errors, dynamic voxels ($\mathcal{V}_{\text{dyn}}$) provide a more informative evaluation of motion prediction. Our method achieves the lowest error across both subsets, showing that combining Doppler supervision with displacement-guided de-aliasing improves motion estimation.

Qualitative Results. Fig. 5 visualizes predicted velocity fields in BEV with corresponding front-view camera and ground-truth velocity fields. Our method produces coherent motion directions and magnitudes within each object while maintaining near-zero motion in static background regions.

4.5 Ablation Studies

We adopt an incremental ablation protocol under identical training and evaluation settings, where each setting extends the previous one by a single major component, mirroring the three-stage pipeline in Fig. 2 from camera-only perception to the full RIGS model:

A1 Camera only.
A2 A1 + 4D radar tensor (radar-assisted init + tri-modal fusion).
A3 A2 + velocity learning (Doppler sup. + de-alias + multi-level losses).
A4 A3 + evidential BKI refinement (ellipsoid kernel + RCS filtering; full).
A5 A4 but radar points (replacing the sparsified 4D radar tensor).

Ablation results are shown in Table 4. Compared with the camera-only baseline (A1), introducing the 4D radar tensor (A2) yields larger gains at longer ranges, most noticeably on FG IoU at 51.2 m, indicating that radar provides

Table 4: Ablation results of the proposed components. A1: camera only; A2: A1 + 4D radar tensor; A3: A2 + velocity learning; A4: A3 + evidential BKI refinement (RIGS); A5: A4 with radar points instead of the sparsified 4D radar tensor.

Setting	IoU_2 [%]			mIoU [%]			BG IoU [%]			FG IoU [%]		
	12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m	12.8 m	25.6 m	51.2 m
A1	29.56	31.68	30.13	20.40	21.32	21.08	31.27	32.42	31.97	9.53	10.22	10.19
A2	29.69	32.25	31.39	20.47	21.74	21.76	31.19	32.78	32.32	9.76	10.69	11.20
A3	30.15	32.50	31.57	20.44	22.38	22.40	30.63	33.36	32.44	10.26	11.41	12.36
A4 (RIGS)	**30.61**	**33.35**	**32.87**	**21.21**	**23.77**	**23.72**	**31.25**	**34.00**	**33.34**	**11.17**	**13.53**	**14.09**
A5	30.08	31.30	31.21	20.52	22.79	22.56	30.47	33.07	32.52	10.57	12.51	12.61

complementary evidence for far-range foreground occupancy. Adding Doppler-supervised velocity learning (A3) further improves foreground prediction, increasing FG IoU at 51.2 m from 11.20% to 12.36%. Incorporating evidential BKI refinement with ellipsoid kernels and radar RCS filtering (A4, full model) consistently boosts all metrics, with a pronounced improvement on FG IoU at 25.6 m (+2.12% over A3), suggesting reduced hallucination and sharper boundaries. Finally, replacing the sparsified 4D radar tensor with point-based cues (A5) degrades long-range performance (FG IoU drops 1.48% at 51.2 m), confirming the benefit of the proposed tensor representation beyond point measurements.

5 Conclusion

We presented RIGS, a radar-informed Gaussian splatting framework for joint 3D semantic occupancy, uncertainty, and dense motion prediction. Integrating camera and 4D radar tensor within a tri-modal Gaussian representation improves long-range occupancy prediction while enabling interpretable uncertainty via evidential BKI. Furthermore, Doppler-supervised learning with displacement-guided de-aliasing yields coherent dense velocity fields. Experiments on K-Radar show that exploiting radar range, power, and Doppler improves occupancy, uncertainty-aware refinement, and motion prediction, but our evaluation is confined to this benchmark with a single front-view camera. Extending to additional datasets and more efficient uncertainty refinement remains future work.

References

1. Bao, W., Yu, Q., Kong, Y.: Evidential deep learning for open set action recognition. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13349–13358 (2021)
2. Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., Heide, F.: Seeing Through Fog Without Seeing Fog: Deep Multimodal Sensor Fusion in Unseen Adverse Weather. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11679–11689 (Jun 2020). https://doi.org/10.1109/CVPR42600.2020.01170

3. Cao, A.Q., de Charette, R.: MonoScene: Monocular 3D Semantic Scene Completion. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3981–3991 (Jun 2022). https://doi.org/10.1109/CVPR52688.2022.00396
4. Cong, X., Han, Y., Sheng, W., Guo, S., Sun, H.: Range-Doppler domain spatial alignment for networked radars. EURASIP Journal on Advances in Signal Processing **2022**(1), 12 (Feb 2022). https://doi.org/10.1186/s13634-022-00849-4
5. Ding, F., Wen, X., Zhu, Y., Li, Y., Lu, C.X.: RadarOcc: Robust 3D occupancy prediction with 4D imaging radar. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. NIPS '24, vol. 37, pp. 101589–101617. Curran Associates Inc., Red Hook, NY, USA (Dec 2024)
6. Doherty, K., Shan, T., Wang, J., Englot, B.: Learning-Aided 3-D Occupancy Mapping With Bayesian Generalized Kernel Inference. IEEE Transactions on Robotics **35**(4), 953–966 (Aug 2019). https://doi.org/10.1109/TRO.2019.2912487
7. Duan, Z., Dang, C., Hu, X., An, P., Ding, J., Zhan, J., Xu, Y., Ma, J.: SDGOCC: Semantic and Depth-Guided Bird's-Eye View Transformation for 3D Multimodal Occupancy Prediction. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6751–6760 (Jun 2025). https://doi.org/10.1109/CVPR52734.2025.00633
8. Gan, L., Zhang, R., Grizzle, J.W., Eustice, R.M., Ghaffari, M.: Bayesian Spatial Kernel Smoothing for Scalable Dense Semantic Mapping. IEEE Robotics and Automation Letters **5**(2), 790–797 (Apr 2020). https://doi.org/10.1109/LRA.2020.2965390
9. Han, Z., Wang, J., Xu, Z., Yang, S., He, L., Xu, S., Wang, J., Li, K.: 4d millimeter-wave radar in autonomous driving: A survey. arXiv preprint arXiv:2306.04242 (2023)
10. Hess, G., Lindström, C., Fatemi, M., Petersson, C., Svensson, L.: SplatAD: Real-Time Lidar and Camera Rendering with 3D Gaussian Splatting for Autonomous Driving. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11982–11992 (Jun 2025). https://doi.org/10.1109/CVPR52734.2025.01119
11. Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., Burgard, W.: OctoMap: An efficient probabilistic 3D mapping framework based on octrees. Auton. Robots **34**(3), 189–206 (Apr 2013). https://doi.org/10.1007/s10514-012-9321-0
12. Huang, T., Prabhakara, A., Chen, C., Karhade, J., Ramanan, D., O'toole, M., Rowe, A.: Towards foundational models for single-chip radar. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 24655–24665 (2025)
13. Huang, Y., Thammatadatrakoon, A., Zheng, W., Zhang, Y., Du, D., Lu, J.: GaussianFormer-2: Probabilistic Gaussian Superposition for Efficient 3D Occupancy Prediction. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 27477–27486 (Jun 2025). https://doi.org/10.1109/CVPR52734.2025.02559
14. Huang, Y., Zheng, W., Zhang, B., Zhou, J., Lu, J.: SelfOcc: Self-Supervised Vision-Based 3D Occupancy Prediction. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19946–19956 (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.01885
15. Huang, Y., Zheng, W., Zhang, Y., Zhou, J., Lu, J.: Tri-Perspective View for Vision-Based 3D Semantic Occupancy Prediction. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9223–9232 (Jun 2023). https://doi.org/10.1109/CVPR52729.2023.00890

16. Huang, Y., Zheng, W., Zhang, Y., Zhou, J., Lu, J.: GaussianFormer: Scene as Gaussians for Vision-Based 3D Semantic Occupancy Prediction. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) Computer Vision – ECCV 2024. pp. 376–393. Springer Nature Switzerland, Cham (2025). https://doi.org/10.1007/978-3-031-73383-3_22
17. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21357–21366 (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.02018
18. Kendall, A., Gal, Y.: What uncertainties do we need in Bayesian deep learning for computer vision? In: Proceedings of the 31st International Conference on Neural Information Processing Systems. pp. 5580–5590. NIPS'17, Curran Associates Inc., Red Hook, NY, USA (Dec 2017)
19. Kerbl, B., Kopanas, G., Leimkuehler, T., Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Trans. Graph. **42**(4), 139:1–139:14 (Jul 2023). https://doi.org/10.1145/3592433
20. Kim, J., Jeon, M., Min, J., Kwak, K., Seo, J.: E2-BKI: Evidential Ellipsoidal Bayesian Kernel Inference for Uncertainty-Aware Gaussian Semantic Mapping. IEEE Robotics and Automation Letters **11**(3), 2378–2385 (Mar 2026). https://doi.org/10.1109/LRA.2026.3653367
21. Kong, S.H., Paek, D.H., Cho, S.: RTNH+: Enhanced 4D Radar Object Detection Network using Combined CFAR-based Two-level Preprocessing and Vertical Encoding (Oct 2023). https://doi.org/10.48550/arXiv.2310.17659
22. Li, X., Li, P., Zheng, Y., Sun, W., Wang, Y., chen, y.: Semi-supervised vision-centric 3D occupancy world model for autonomous driving. In: Yue, Y., Garg, A., Peng, N., Sha, F., Yu, R. (eds.) International Conference on Learning Representations. vol. 2025, pp. 62563–62580 (2025)
23. Li, Y., Yu, Z., Choy, C., Xiao, C., Alvarez, J.M., Fidler, S., Feng, C., Anandkumar, A.: VoxFormer: Sparse Voxel Transformer for Camera-Based 3D Semantic Scene Completion. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9087–9098 (Jun 2023). https://doi.org/10.1109/CVPR52729.2023.00877
24. Li, Y., Ge, Z., Yu, G., Yang, J., Wang, Z., Shi, Y., Sun, J., Li, Z.: BEVDepth: Acquisition of Reliable Depth for Multi-View 3D Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence **37**(2), 1477–1485 (Jun 2023). https://doi.org/10.1609/aaai.v37i2.25233
25. Liu, R., Wu, X., Chen, X., Hu, L., Lou, Y.: 4D-ROLLS: 4D Radar Occupancy Learning via LiDAR Supervision. In: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 13602–13609 (Oct 2025). https://doi.org/10.1109/IROS60139.2025.11246471
26. Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D.L., Han, S.: BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 2774–2781 (May 2023). https://doi.org/10.1109/ICRA48891.2023.10160968
27. Ma, Q., Tan, X., Qu, Y., Ma, L., Zhang, Z., Xie, Y.: COTR: Compact Occupancy TRansformer for Vision-Based 3D Occupancy Prediction. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19936–19945 (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.01884

28. Meyer, M., Kuschk, G.: Deep Learning Based 3D Object Detection for Automotive Radar and Camera. In: 2019 16th European Radar Conference (EuRAD). pp. 133–136 (Oct 2019)
29. Ming, Z., Stephany Berrio, J., Shan, M., Worrall, S.: OccFusion: Multi-Sensor Fusion Framework for 3D Semantic Occupancy Prediction. IEEE Transactions on Intelligent Vehicles **10**(5), 3421–3433 (May 2025). https://doi.org/10.1109/TIV.2024.3453293
30. Mohan, R., Hurtado, J.V., Mohan, R., Valada, A.: ForecastOcc: Vision-based Semantic Occupancy Forecasting (Feb 2026). https://doi.org/10.48550/arXiv.2602.08006
31. Nabati, R., Qi, H.: CenterFusion: Center-based Radar and Camera Fusion for 3D Object Detection. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1526–1535. IEEE, Waikoloa, HI, USA (Jan 2021). https://doi.org/10.1109/WACV48630.2021.00157
32. Paek, D.H., Kong, S.H., Wijaya, K.T.: K-Radar: 4D radar object detection for autonomous driving in various weather conditions. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. pp. 3819–3829. NIPS '22, Curran Associates Inc., Red Hook, NY, USA (Nov 2022)
33. Pan, J., Wang, Z., Wang, L.: Co-Occ: Coupling Explicit Feature Fusion With Volume Rendering Regularization for Multi-Modal 3D Semantic Occupancy Prediction. IEEE Robotics and Automation Letters **9**(6), 5687–5694 (Jun 2024). https://doi.org/10.1109/LRA.2024.3396092
34. Pan, M., Liu, J., Zhang, R., Huang, P., Li, X., Xie, H., Wang, B., Liu, L., Zhang, S.: RenderOcc: Vision-Centric 3D Occupancy Prediction with 2D Rendering Supervision. In: 2024 IEEE International Conference on Robotics and Automation (ICRA). pp. 12404–12411 (May 2024). https://doi.org/10.1109/ICRA57147.2024.10611537
35. Peng, X., Tang, M., Sun, H., Bierzynski, K., Servadei, L., Wille, R.: 4D mmWave radar for sensing enhancement in adverse environments: Advances and challenges. In: 2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC). pp. 10:50–11:10. IEEE (Nov 2025)
36. Philion, J., Fidler, S.: Lift, Splat, Shoot: Encoding Images from Arbitrary Camera Rigs by Implicitly Unprojecting to 3D. In: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV. pp. 194–210. Springer-Verlag, Berlin, Heidelberg (Aug 2020). https://doi.org/10.1007/978-3-030-58568-6_12
37. Pitropov, M., Garcia, D.E., Rebello, J., Smart, M., Wang, C., Czarnecki, K., Waslander, S.: Canadian Adverse Driving Conditions dataset. Int. J. Rob. Res. **40**(4-5), 681–690 (Apr 2021). https://doi.org/10.1177/0278364920979368
38. Ramos, F., Ott, L.: Hilbert maps: Scalable continuous occupancy mapping with stochastic gradient descent. In: Robotics: Science and Systems XI. vol. 11 (Jul 2015)
39. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. pp. 3183–3193. NIPS'18, Curran Associates Inc., Red Hook, NY, USA (Dec 2018)
40. Sheeny, M., De Pellegrin, E., Mukherjee, S., Ahrabian, A., Wang, S., Wallace, A.: RADIATE: A Radar Dataset for Automotive Perception in Bad Weather. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 1–7 (May 2021). https://doi.org/10.1109/ICRA48506.2021.9562089

41. Song, C., Kim, S., Jeong, H., Shin, J., Lim, J., Kum, D.: REOcc: Camera-Radar Fusion with Radar Feature Enrichment for 3D Occupancy Prediction. In: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 15367–15374 (Oct 2025). https://doi.org/10.1109/IROS60139.2025.11246246
42. Tang, P., Wang, Z., Wang, G., Zheng, J., Ren, X., Feng, B., Ma, C.: SparseOcc: Rethinking Sparse Latent Representation for Vision-Based Semantic Occupancy Prediction. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15035–15044 (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.01424
43. Wang, P.S.: OctFormer: Octree-based Transformers for 3D Point Clouds. ACM Trans. Graph. **42**(4), 155:1–155:11 (Jul 2023). https://doi.org/10.1145/3592131
44. Wang, T., Zhu, X., Pang, J., Lin, D.: FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). pp. 913–922. IEEE, Montreal, BC, Canada (Oct 2021). https://doi.org/10.1109/ICCVW54120.2021.00107
45. Wei, Y., Zhao, L., Zheng, W., Zhu, Z., Zhou, J., Lu, J.: SurroundOcc: Multi-Camera 3D Occupancy Prediction for Autonomous Driving. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21672–21683 (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.01986
46. Wilson, J., Fu, Y., Friesen, J., Ewen, P., Capodieci, A., Jayakumar, P., Barton, K., Ghaffari, M.: ConvBKI: Real-Time Probabilistic Semantic Mapping Network With Quantifiable Uncertainty. IEEE Transactions on Robotics **40**, 4648–4667 (2024). https://doi.org/10.1109/TRO.2024.3453771
47. Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., Li, X.: GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting. In: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19595–19604 (Jun 2024). https://doi.org/10.1109/CVPR52733.2024.01853
48. Yang, B., Guo, R., Liang, M., Casas, S., Urtasun, R.: RadarNet: Exploiting Radar for Robust Perception of Dynamic Objects. In: Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII. pp. 496–512. Springer-Verlag, Berlin, Heidelberg (Aug 2020). https://doi.org/10.1007/978-3-030-58523-5_29
49. Yang, L., Zheng, L., Ai, W., Liu, M., Li, S., Lin, Q., Yan, S., Bai, J., Ma, Z., Huang, T., Zhu, X.: MetaOcc: Spatio-Temporal Fusion of Surround-View 4D Radar and Camera for 3D Occupancy Prediction with Dual Training Strategies (Aug 2025). https://doi.org/10.48550/arXiv.2501.15384
50. Zhang, Y., Tu, C., Gao, K., Wang, L.: Multisensor information fusion: Future of environmental perception in intelligent vehicles. Journal of Intelligent and Connected Vehicles **7**(3), 163–176 (2024). https://doi.org/10.26599/JICV.2023.9210049
51. Zhang, Y., Zhu, Z., Du, D.: OccFormer: Dual-path Transformer for Vision-based 3D Semantic Occupancy Prediction. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 9399–9409 (Oct 2023). https://doi.org/10.1109/ICCV51070.2023.00865