DISAMBIGUATING SEMANTIC RARITY FROM DISTRIBUTIONAL NOVELTY FOR OUT-OF-DISTRIBUTION 3D OCCUPANCY PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable out-of-distribution (OOD) 3D occupancy prediction requires distinguishing genuinely unfamiliar objects from rare but valid semantic patterns. However, long-tailed supervision and intra-class variation can leave rare known voxels weakly supported by the learned feature distribution, causing existing OOD detectors to conflate semantic rarity with distributional novelty and produce fragmented anomaly responses. To address this issue, we present UMPOcc, a plug-and-play framework for rare-vs-unknown disambiguation through complementary distributional and spatial reasoning. Its Uncertainty-Aware Multi-Prototype Occupancy Field (UMPOF) represents each semantic class with multiple modes characterized by empirical priors and feature dispersion, yielding more reliable source-domain support for rare known patterns. Peak-Preserving Spatial Calibration (PPSC) further exploits local occupancy support to reduce spatially inconsistent low-confidence responses while preserving high-confidence anomaly peaks. Extensive experiments across five benchmarks show strong occupancy and OOD performance, with consistent gains on tail classes and effective generalization to real-world scenes.

1 INTRODUCTION

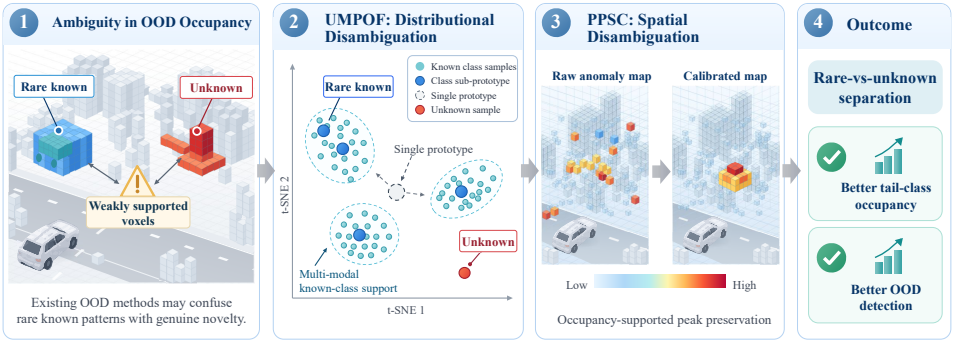

Figure 1: **Disambiguating semantic rarity from distributional novelty.** Long-tailed occupancy prediction can leave rare known voxels weakly supported, causing existing OOD methods to confuse them with genuinely unknown objects. UMPOF models multi-modal known-class support to improve distributional disambiguation, while PPSC calibrates fragmented anomaly responses through occupancy-supported peak preservation. Together, they improve tail-class occupancy prediction and OOD detection across different occupancy backbones.

Semantic occupancy prediction (Behley et al., 2019; Li et al., 2024b; Xue et al., 2026) provides a dense 3D description of the scene by assigning a semantic state to every voxel. Compared with object-centric detection (Li et al., 2025b), it exposes free space, object extent, and surrounding structure for planning in crowded urban environments. Camera-based occupancy models (Cao & de Charette, 2022; Huang et al., 2023a; Li et al., 2023b) have therefore become important for autonomous driving (Mei et al., 2024; Zuo et al., 2025; Huang et al., 2025). Their predictions, however,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

are usually restricted (Zhang et al., 2025b; Zhou et al., 2025) to classes observed during training (Yu et al., 2025; Zhou et al., 2026). An unfamiliar object can consequently receive a confident rare-class label despite geometrically plausible occupancy (Zhang et al., 2025a), causing an OOD object to be treated as a known obstacle with the wrong semantics (Nekrasov et al., 2025; Zhang et al., 2026b). This creates a central ambiguity: is a weakly supported voxel genuinely unknown, or merely a rare but valid known pattern?

This ambiguity is amplified by the long-tailed distribution of 3D occupancy labels (Lahoud et al., 2024; Chen et al., 2026). Dominant classes such as road and building provide abundant supervision, whereas small and distant categories contribute only a few occupied voxels (Zhang et al., 2025a; 2026b). Rare known features can therefore receive weak and poorly calibrated support despite being in-distribution (Guo et al., 2017; Heidrich et al., 2025; Kälble et al., 2025). Existing voxel-level OOD methods (Zhang et al., 2025a; 2026b) and uncertainty estimators (Heidrich et al., 2025; Beemelmanns et al., 2026; Chen & Ding, 2025) provide useful baselines, while classical softmax, energy, and feature-distance scores (Hendrycks & Gimpel, 2017; Liu et al., 2020; Lee et al., 2018) offer generic uncertainty signals. ProOOD (Zhang et al., 2026b) further combines class-wise prototypes with EchoOOD, showing that occupancy refinement and OOD scoring should be considered together. Yet the key question remains whether weak support reflects semantic rarity or genuine distributional novelty.

A major source of this confusion is that known-class support is itself multi-modal. A single prototype is a restrictive class representation (Kim et al., 2025; Li et al., 2025a; Zohrabi et al., 2025): visible surfaces, occluded interiors, and observations at different ranges can occupy distinct feature modes under the same semantic label (Li et al., 2025a). Collapsing them to one center can make legitimate rare features and nearby OOD features similarly unsupported (Zohrabi et al., 2025). The ambiguity is further compounded spatially, since voxel-wise anomaly maps are often noisy (Grcić et al., 2021; Heidrich et al., 2025). Uncertain logits and low-support voxels can create isolated peaks (Liu et al., 2020; Chen & Ding, 2025), whereas indiscriminate smoothing may erase the compact response of a small unknown object. Reliable rare-vs-unknown disambiguation therefore requires both faithful distributional support for known semantics and spatial evidence that separates coherent anomalies from isolated uncertainty.

Based on this view, we present UMPOcc, a lightweight modular framework that addresses rare-vs-unknown ambiguity through complementary distributional and spatial reasoning, as illustrated in Fig. 1. UMPOF replaces a single class center with multiple distributional modes and estimates source-domain support using their priors and feature dispersion. The resulting representation preserves intra-class variation, refines occupancy features, and yields a density-based uncertainty signal for tail and OOD voxels. PPSC complements this representation by calibrating voxel-wise anomaly responses with local score and occupancy support, attenuating spatially unsupported low-confidence fluctuations while preserving high-confidence local peaks. Both components attach to existing voxel features and logits without redesigning the occupancy backbone or changing its inference interface. Our contributions are summarized as follows:

- We introduce UMPOcc, a unified framework for rare-vs-unknown disambiguation in 3D occupancy prediction, jointly reasoning about distributional feature support and spatial occupancy evidence.
- UMPOF models multi-modal known-class support with mode priors and dispersion, while PPSC calibrates anomaly responses using occupancy-aware local support without erasing compact peaks.
- Extensive experiments show strong in-distribution and OOD performance, with consistent gains on tail classes and effective generalization to real-world scenes.

2 Related Work

3D Semantic Occupancy Prediction. 3D semantic occupancy prediction (Behley et al., 2019; Li et al., 2023a; 2024a;b; 2026) reconstructs a dense voxel volume from camera or LiDAR observations (Wang et al., 2025; Feng et al., 2026). Monocular methods such as MonoScene (Cao & de Charette, 2022) lift image evidence into a volumetric representation, while TPVFormer (Huang et al., 2023a), VoxFormer (Li et al., 2023b), HTCL(Li et al., 2024a) and SGN (Mei et al., 2024) im-

prove view transformation and sparse-to-dense reasoning. Recent methods (Leng et al., 2025; Chen et al., 2025; Zhang et al., 2026a) further explore contextual queries (Jiang et al., 2024a; Kim et al., 2025), Gaussian representations (Huang et al., 2024; Chambon et al., 2025; Lv et al., 2026), and efficient scene representations (Zhu et al., 2026; Gao et al., 2026). VoxDet (Li et al., 2025b) revisits occupancy prediction from the perspective of dense object detection and provides a strong complementary backbone. Our method is orthogonal to these architectural choices: it consumes the voxel features and logits of an existing occupancy model and does not require a new view transformer or voxel decoder.

OOD Detection for Dense 3D Perception. Classical OOD scores include maximum softmax probability (Hendrycks & Gimpel, 2017), energy (Liu et al., 2020), and distance to a feature distribution (Lee et al., 2018). Their direct use in dense prediction is challenging because a scene contains many correlated voxels, and a local anomaly may occupy only a small region (Grcić et al., 2021). OCCUQ (Heidrich et al., 2025), OccOoD (Zhang et al., 2025a), and the VAA benchmark suite (Zhang et al., 2025a) introduce uncertainty and anomaly evaluation for 3D occupancy. ProOOD (Zhang et al., 2026b) uses class prototypes and local/global semantic consistency to build EchoOOD. We retain the same benchmark and tolerance-aware evaluation, but model the class distribution with multiple modes and calibrate the inherited anomaly score using occupancy-aware local support.

Long-Tailed and Prototype-Based Representation. Long-tailed recognition methods improve rare-class representation through reweighting, resampling, or feature-level constraints (Lahoud et al., 2024). Prototype-based occupancy methods use class centers or prototype queries to provide semantic anchors (Kim et al., 2025; Zhang et al., 2026b). These approaches generally treat a class as a single center or use prototypes for recognition rather than uncertainty estimation (Li et al., 2025a; Zohrabi et al., 2025). UMPOF instead maintains a set of class-conditional modes with priors and dispersion estimates, allowing the same semantic class to occupy several regions in feature space. The resulting density provides an auxiliary voxel-level anomaly signal.

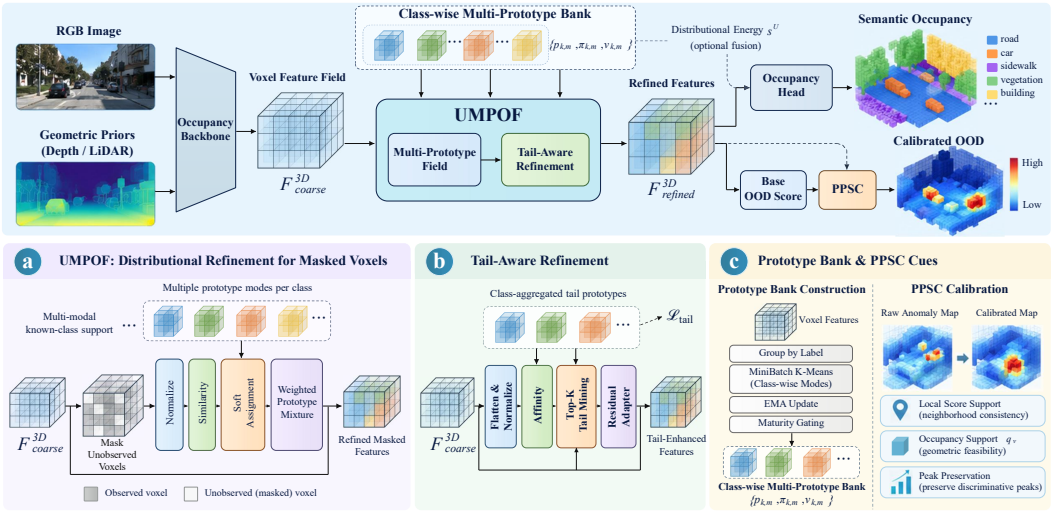

Figure 2: **Overview of the proposed UMPOcc framework.** Given RGB observations and geometric priors, the occupancy backbone first produces a coarse voxel feature field. UMPOF models each semantic class with a class-wise multi-prototype bank and refines voxel representations through multi-modal prototype aggregation and tail-aware enhancement, yielding refined features for semantic occupancy prediction. The same representation further provides distributional OOD evidence, while PPSC calibrates the inherited anomaly score using local score support, occupancy support, and peak preservation to obtain a spatially calibrated OOD map. The lower panels detail multi-prototype refinement, tail-aware refinement, prototype-bank construction (offline K-Means and online EMA updates), and PPSC calibration.

3 METHOD

3.1 TASK AND FRAMEWORK OVERVIEW

We study voxel-level OOD detection together with semantic occupancy prediction (Zhang et al., 2025a; 2026b). Given an RGB observation and its geometric priors, an occupancy network produces a voxel feature field $\mathbf{F} \in \mathbb{R}^{C\times H\times W\times Z}$ and semantic logits $\mathbf{L} \in \mathbb{R}^{K\times H\times W\times Z}$, where class 0 denotes empty space. The semantic prediction is $\hat{y}_v = \arg\max_k L_{k,v}$. In addition to $\hat{\mathbf{y}}$, we seek an anomaly map $\mathbf{a} \in [0,1]^{H\times W\times Z}$ whose high values identify occupied voxels that are not supported by the source-domain semantic feature distribution.

Our framework, illustrated in Fig. 2, has two complementary components. **Uncertainty-Aware Multi-Prototype Occupancy Field (UMPOF)** replaces the single class center used by the original prototype branch (Zhang et al., 2026b) with a class-conditional set of modes. It provides distributional feature refinement during training and an auxiliary density signal that can be combined with the inherited anomaly score. **Peak-Preserving Spatial Calibration (PPSC)** uses the soft occupancy support of the predicted volume to regularize the anomaly map while retaining compact high-response regions. The occupancy decoder and its semantic logits remain unchanged; the proposed operations are attached to its voxel features and outputs.

3.2 PROTOTYPE-AWARE OCCUPANCY REPRESENTATION

The occupancy backbone first lifts image features into a coarse 3D field. A proposal head separates visible voxels from unobserved or masked voxels, and a lightweight prior mapping supplies features for the latter. Let $\mathbf{x}_v$ denote the resulting coarse feature at voxel v, and let $\mathcal{U}$ be the masked set. The original prototype branch uses one EMA prototype per non-empty class (Zhang et al., 2026b). We retain this interface so that the method can be applied to SGN (Mei et al., 2024) and VoxDet (Li et al., 2025b) without changing their view transformation or occupancy heads.

UMPOF changes the representation of each class from one point to a small mixture of modes. The resulting modes are used in two places: they provide semantic evidence for masked-feature refinement, and they provide an auxiliary source-domain density signal for OOD scoring. The tail-mining branch continues to select candidate voxels using class-level tail anchors, and injects a residual obtained from the selected tail anchors before the final occupancy decoder. Consequently, the proposed representation improves the features supplied to the semantic head while preserving the original binary occupancy and semantic prediction interfaces.

3.3 UNCERTAINTY-AWARE MULTI-PROTOTYPE OCCUPANCY FIELD

Multi-modal source representation. For every non-empty class k, UMPOF maintains M modes

$$\mathcal{P}_k = \{(\mathbf{p}_{k,m}, \pi_{k,m}, \nu_{k,m})\}_{m=1}^{M}, \tag{1}$$

where $\mathbf{p}_{k,m} \in \mathbb{R}^C$ is the mode center, $\pi_{k,m}$ is its empirical prior, and $\nu_{k,m}$ measures its intra-mode dispersion. For a voxel feature $\mathbf{f}$, we define the distance to mode m of class k as

$$d_{k,m}(\mathbf{f}) = 1 - \bar{\mathbf{f}}^{\top}\bar{\mathbf{p}}_{k,m}, \tag{2}$$

where bars denote ℓ_2 normalization. Each mode contributes the log-density evidence

$$\ell_{k,m}(\mathbf{f}) = \log(\pi_{k,m} + \epsilon) - \frac{d_{k,m}(\mathbf{f})}{\tau(\nu_{k,m} + \epsilon)}, \tag{3}$$

and the resulting distributional anomaly energy is

$$e_{\mathrm{U}}(\mathbf{f}) = -\,\mathrm{LSE}_{k,m}\left[\ell_{k,m}(\mathbf{f})\right], \tag{4}$$

where the log-sum-exp is taken over all class–mode pairs.

The bank is fitted independently for each class from source-domain training features using Mini-Batch K-Means (Sculley, 2010). The dispersion $\nu_{k,m}$ is estimated as the mean cosine distance of features assigned to the corresponding mode and is fixed after bank construction. Mode counts are converted into the empirical priors $\pi_{k,m}$. Empty and ignored voxels are excluded. Features supported by a reliable source mode receive low energy, whereas features outside all modes receive high energy. The energy is normalized per scene before score calibration; the exact normalization is specified in Appendix A.

Online prototype learning. During occupancy training, class-labeled voxel features are softly assigned to the modes of their ground-truth class. For a feature $\mathbf{f}_i$ with label k,

$$r_{i,m} = \frac{\exp\left(-d_{k,m}(\mathbf{f}_i)/\tau_a\right)}{\sum_{j=1}^{M} \exp\left(-d_{k,j}(\mathbf{f}_i)/\tau_a\right)}. \tag{5}$$

The mode center and prior are updated with stop-gradient EMA updates:

$$\widetilde{\mathbf{p}}_{k,m} = \frac{\sum_{i:y_i=k} r_{i,m}\mathbf{f}_i}{\sum_{i:y_i=k} r_{i,m} + \epsilon}, \qquad \mathbf{p}_{k,m} \leftarrow (1-\beta)\mathbf{p}_{k,m} + \beta\widetilde{\mathbf{p}}_{k,m}. \tag{6}$$

The mode prior is updated by the corresponding soft assignment mass:

$$\pi_{k,m} \leftarrow (1-\beta)\pi_{k,m} + \beta \frac{\sum_{i:y_i=k} r_{i,m}}{\sum_{j=1}^{M}\sum_{i:y_i=k} r_{i,j} + \epsilon}. \tag{7}$$

The same mode assignment is used in a distributional prototype contrastive objective. Let

$$g_k(\mathbf{f}) = \mathrm{LSE}_m\left(\frac{\bar{\mathbf{f}}^\top \bar{\mathbf{p}}_{k,m}}{\tau_p} + \log(\pi_{k,m} + \epsilon)\right). \tag{8}$$

For a non-empty voxel with label y, the distributional prototype loss is

$$\mathcal{L}_{\mathrm{dproto}} = -\log \frac{\exp(g_y(\mathbf{f}))}{\sum_k \exp(g_k(\mathbf{f}))}. \tag{9}$$

This objective encourages a feature to be close to at least one mode of its class without forcing all appearances of the class to collapse to a single center. Prototype updates start only after the warm-up period and are not differentiated through, which prevents early noisy features from destabilizing the occupancy backbone.

Distributional feature refinement and tail mining. For a masked voxel $v \in \mathcal{U}$, the refinement branch computes a soft mixture over all mature modes:

$$w_{v,k,m} = \mathrm{softmax}_{k,m}\left(\frac{\bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_{k,m}}{\tau_r}\right), \qquad \Delta\mathbf{x}_v = \sum_{k,m} w_{v,k,m}\mathbf{p}_{k,m}. \tag{10}$$

Only masked voxels predicted as non-empty by the auxiliary occupancy head are refined. The updated feature is

$$\mathbf{x}'_v = \mathbf{x}_v + \alpha_r \Delta\mathbf{x}_v. \tag{11}$$

This residual form preserves the visual/geometric evidence in $\mathbf{x}_v$ while supplying a class-consistent semantic direction.

Prototype refinement is activated only after a warm-up and maturity check. We call a mode mature when its class row has been initialized by a valid source-label update; the global gate additionally requires sufficient class coverage and prototype quality. Before the gate is active, the occupancy branch follows the original feature path. Once active, only initialized sub-modes are used. The inference bank is built offline from source training features and remains fixed during evaluation.

To protect rare categories, we also maintain a tail set $\mathcal{T}$. The class-level aggregate used by the tail branch is

$$\mathbf{p}_k = \sum_{m=1}^{M} \pi_{k,m}\mathbf{p}_{k,m}, \qquad \bar{\mathbf{p}}_k = \frac{\mathbf{p}_k}{\|\mathbf{p}_k\|_2 + \epsilon}. \tag{12}$$

For each candidate voxel, its tail affinity is

$$u_v = \max_{k\in\mathcal{T}} \bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_k. \tag{13}$$

We select the top K_t affinities and inject a tail-prototype residual through a lightweight adapter:

$$\mathbf{x}''_v = \mathbf{x}'_v + A\left(\sum_{k\in\mathcal{T}} \rho_{v,k}\mathbf{p}_k\right), \rho_{v,k} = \mathrm{softmax}_{k\in\mathcal{T}}\left(\eta\bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_k\right). \tag{14}$$

This aggregate is used only for tail voxel selection and refinement; the UMPOF density score still evaluates all modes separately. The adapter $A(\cdot)$ is a two-layer MLP consisting of $\mathrm{Linear}(C, C/4)$, ReLU, and $\mathrm{Linear}(C/4, C)$, while η is the tail-affinity temperature. The selected features receive an auxiliary classification loss $\mathcal{L}_{\mathrm{tail}}$. The tail branch is therefore supervised by source occupancy labels rather than OOD annotations, and its purpose is to reduce the tendency of rare known classes to be absorbed by dominant semantic classes (Lahoud et al., 2024).

3.4 PPSC: Peak-Preserving Spatial Calibration

The voxel-wise anomaly score is informative but spatially fragmented. A small unknown object may occupy only a few cells, so direct thresholding can yield isolated responses; indiscriminate smoothing, however, can erase the object. PPSC uses local occupancy support to correct this trade-off.

Let s_v be the normalized pre-calibration anomaly score at voxel v, $\mathcal{N}(v)$ a cubic neighborhood, and

$$q_v = 1 - \mathrm{softmax}(\mathbf{L}_v)_0 \tag{15}$$

the probability that the voxel is occupied. We first compute the local average, local maximum, and peakness:

$$\mu_v = \mathrm{Avg}_{u\in\mathcal{N}(v)}\, s_u, \qquad m_v = \max_{u\in\mathcal{N}(v)} s_u, \qquad r_v = \left(\frac{s_v}{m_v + \epsilon}\right)^{\gamma}. \tag{16}$$

The occupancy-weighted score and local occupancy support are

$$\mu_v^q = \frac{\mathrm{Avg}_{u\in\mathcal{N}(v)}(q_u s_u)}{\mathrm{Avg}_{u\in\mathcal{N}(v)}\, q_u + \epsilon}, \qquad h_v = \mathrm{Avg}_{u\in\mathcal{N}(v)}\, q_u. \tag{17}$$

The peak-preserving estimate is

$$s_v^{\mathrm{pk}} = r_v s_v + (1 - r_v)\left((1-\alpha)s_v + \alpha\mu_v\right). \tag{18}$$

The support branch is defined as

$$s_v^{\mathrm{sup}} = s_v + (1 - r_v)\left[b[\mu_v^q - s_v]_+ - c[s_v - \mu_v^q]_+\right]. \tag{19}$$

Finally, local support controls the interpolation:

$$g_v = \mathrm{clip}\left(\frac{h_v - q_{\min}}{\delta}, 0, 1\right)^{\rho}, s_v^{\mathrm{PPSC}} = (1 - g_v)s_v^{\mathrm{pk}} + g_v s_v^{\mathrm{sup}}. \tag{20}$$

When r_v is high, the original response is largely retained; when it is low, neighborhood evidence can attenuate or correct lower-confidence responses according to their local score and occupancy support. Local occupancy support determines how much of the correction is trusted. The result is clipped to $[0, 1]$, and voxels predicted as empty are assigned the minimum score in the scene. PPSC introduces no trainable parameters. Its hyperparameters are fixed before test-time evaluation and are not optimized per scene or using test labels. The concrete pre-calibration score and the score-composition variants used in the benchmark are specified in Appendix A.

3.5 Training Objective and Inference

The occupancy backbone is optimized with the original semantic, geometric, and binary occupancy objectives. Denoting their sum by $\mathcal{L}_{\mathrm{occ}}$, the additional representation terms give

$$\mathcal{L} = \mathcal{L}_{\mathrm{occ}} + \lambda_d \mathcal{L}_{\mathrm{dproto}} + \lambda_t \mathcal{L}_{\mathrm{tail}}. \tag{21}$$

The auxiliary occupancy head and the tail head use the same source labels and are removed from the reported inference interface. To make the score path explicit, let the inherited pre-calibration score be

$$s^{\mathrm{base}} = \mathcal{S}_{\mathrm{base}}(\mathbf{F}, \mathbf{L}). \tag{22}$$

We obtain the PPSC output and normalized UMPOF energy as

$$\widetilde{s}_v = \mathrm{PPSC}\left(s^{\mathrm{base}}, \mathbf{L}\right)_v, \qquad s_v^{\mathrm{U}} = \mathrm{Norm}_{\mathrm{scene}}\left(e_{\mathrm{U}}(\mathbf{f}_v)\right). \tag{23}$$

The final anomaly score is

$$a_v = \begin{cases} \widetilde{s}_v, & \text{PPSC-only path,} \\ \mathrm{clip}\left(\widetilde{s}_v + \lambda_g [s_v^{\mathrm{U}} - \widetilde{s}_v]_+, 0, 1\right), & \text{UMPOF–PPSC fusion path.} \end{cases} \tag{24}$$

The occupancy logits are never altered by PPSC. The main text uses the first path when discussing the PPSC component; the fusion rule and all score-mode assignments are detailed in Appendix A.

At inference, a bank with K source classes and M modes requires $\mathcal{O}(HWZKM)$ similarity operations. We evaluate the density in chunks to bound memory, and PPSC requires only fixed-kernel 3D pooling. No target-domain test labels or per-scene optimization are used at inference. UMPOcc can be attached to a pretrained occupancy model while preserving its semantic predictions, and its extra computation is confined to a lightweight feature-density and spatial-calibration path.

Table 1: Quantitative results on the SemanticKITTI test set (Behley et al., 2019) and VAA-KITTI (Zhang et al., 2025a) (Single-Frame Evaluation for Fair Comparison).

Method	IoU	road (15.30%)	sidewalk (11.13%)	parking (1.12%)	other-grnd. (0.56%)	building (14.1%)	car (3.92%)	truck (0.16%)	bicycle (0.03%)	motorcycle (0.03%)	other-veh. (0.20%)	vegetation (39.3%)	trunk (0.51%)	terrain (9.17%)	person (0.07%)	bicyclist (0.07%)	motorcyclist (0.05%)	fence (3.90%)	pole (0.29%)	traf.-sign (0.08%)	mIoU	*Tail* mIoU
SemanticKITTI																						
TPVFormer (Huang et al., 2023b)	34.25	55.10	27.20	27.40	6.50	14.80	19.20	3.70	1.00	0.50	2.30	13.90	2.60	20.40	1.10	2.40	0.30	11.00	2.90	1.50	11.26	2.25
OccFormer (Zhang et al., 2023)	34.53	55.90	30.30	31.50	6.50	15.70	21.60	1.20	1.50	1.70	3.20	16.80	3.90	21.30	2.20	1.10	0.20	11.90	3.80	3.70	12.32	2.64
VoxFormer (Li et al., 2023b)	42.95	53.90	25.30	21.10	5.60	19.80	20.80	3.50	1.00	0.70	3.70	22.40	7.50	21.30	1.40	2.60	0.20	11.10	5.10	4.90	12.20	3.29
HASSC (Wang et al., 2024)	43.40	54.60	27.70	23.80	6.20	21.10	22.80	4.70	1.60	1.00	3.90	23.80	8.50	23.30	1.60	**4.00**	0.30	13.10	5.80	5.50	13.34	3.92
Symphonies (Jiang et al., 2024b)	42.19	58.40	29.30	26.90	11.70	24.70	23.60	3.20	3.60	2.60	5.60	24.20	10.00	23.10	3.20	1.90	**2.00**	16.10	7.70	8.00	15.04	5.41
SHTOcc (Yu et al., 2026)	-	57.80	28.90	26.30	12.00	23.40	24.60	3.60	3.50	1.80	4.60	24.90	11.00	22.50	**3.70**	3.80	0.10	17.10	8.70	9.40	15.07	5.65
CGFormer (Yu et al., 2024)	44.41	64.30	34.20	34.10	**12.10**	25.80	26.10	4.30	3.70	1.30	2.70	24.50	11.20	29.30	1.70	3.60	0.40	18.70	8.70	9.30	16.63	5.36
SGN (Mei et al., 2024)	41.88	**57.80**	**29.20**	27.70	5.20	**23.90**	24.90	2.70	0.40	0.30	4.00	24.20	10.00	25.80	1.10	**2.50**	0.30	14.20	7.40	4.40	14.01	3.48
ProOOD-S (Zhang et al., 2026b)	43.14	57.10	28.00	**28.00**	**7.10**	23.80	24.70	4.20	**1.30**	**0.90**	2.40	25.20	**11.40**	24.80	2.10	2.00	**0.30**	**16.20**	8.10	**8.00**	14.51	4.35
UMPOcc (+SGN)	**44.09**	59.78	28.22	20.71	0.08	21.60	**31.55**	**14.63**	0.82	**0.90**	**5.57**	**26.18**	8.49	**34.27**	**2.89**	2.40	0.10	9.80	**10.57**	6.90	**15.02**	**4.85**
VoxDet (Li et al., 2025b)	46.69	65.80	35.50	33.90	11.40	27.10	27.50	5.20	**5.10**	**5.60**	4.50	27.70	11.90	31.20	2.30	2.70	**0.50**	21.10	9.50	9.20	17.77	6.17
ProOOD-V (Zhang et al., 2026b)	46.75	**66.20**	**36.40**	34.90	**11.40**	**27.70**	27.60	4.30	4.10	5.00	7.20	28.40	**12.90**	32.20	2.50	2.80	0.40	**21.30**	9.60	**9.50**	18.12	6.34
UMPOcc (+VoxDet)	**47.09**	65.36	34.41	**36.08**	0.30	25.56	**34.39**	**15.07**	4.19	4.00	**14.44**	**29.67**	10.18	**42.03**	**5.08**	**5.49**	0.10	12.09	**12.63**	8.60	**18.93**	**7.28**
VAA-KITTI																						
SGN (Mei et al., 2024)	39.07	46.62	26.54	15.88	0.12	18.34	**20.51**	0.62	0.44	0.35	1.23	24.76	6.40	21.78	0.48	1.29	0.00	12.25	8.70	4.57	11.10	2.20
ProOOD-S (Zhang et al., 2026b)	40.26	49.34	27.14	16.47	0.15	20.07	19.94	3.32	0.43	**0.94**	**1.31**	26.08	6.50	23.10	**1.57**	1.18	0.00	12.89	9.06	5.87	11.86	2.76
UMPOcc (+SGN)	**40.32**	**49.44**	**27.25**	**16.92**	**0.22**	**20.31**	20.09	**3.58**	**0.46**	0.84	1.26	**26.12**	**6.72**	**23.16**	1.54	**1.30**	0.00	**12.91**	**9.34**	**5.97**	**11.97**	**2.84**
VoxDet (Li et al., 2025b)	**41.60**	55.93	27.12	21.33	0.02	**22.81**	20.38	5.77	**3.27**	1.08	4.14	**27.36**	6.03	23.13	1.09	1.70	0.00	15.73	8.83	4.68	13.18	3.33
ProOOD-V (Zhang et al., 2026b)	40.88	56.41	**28.20**	23.77	**0.19**	21.73	**21.32**	**7.00**	2.89	**3.78**	3.57	27.07	6.41	21.21	**2.62**	**1.98**	0.00	15.96	9.06	6.05	13.63	3.96
UMPOcc (+VoxDet)	41.47	**57.17**	28.03	**27.68**	0.10	21.70	18.43	6.37	0.64	2.59	**9.22**	25.83	**6.77**	**23.45**	1.77	1.76	0.00	**15.97**	**9.61**	**6.82**	**13.89**	**4.15**

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets and metrics. We evaluate UMPOcc on five benchmarks following the ProOOD protocol (Zhang et al., 2026b). Semantic occupancy is evaluated on SemanticKITTI (Behley et al., 2019), SSCBench-KITTI-360 (Li et al., 2024b), VAA-KITTI, and VAA-KITTI-360; OOD detection is evaluated on VAA-KITTI, VAA-KITTI-360, and the real-world VAA-STU benchmark. We report IoU, mIoU, and tail mIoU for occupancy prediction. Following ProOOD, tail classes contain less than 0.6% and 0.3% of voxels in SemanticKITTI and SSCBench-KITTI-360, respectively. For OOD detection, we report tolerance-aware AuPRC_r at 0.8, 1.0, and 1.2 m together with AuROC.

Implementation details. We evaluate UMPOcc with SGN (Mei et al., 2024) and VoxDet (Li et al., 2025b) under single-frame settings. For SGN, we initialize from the corresponding ProOOD checkpoints and adapt UMPOF for eight epochs using AdamW with learning rate 2×10^{-5}, cosine annealing, 200 warm-up iterations, and $\lambda_d = 0.1$. We use four modes for SemanticKITTI and SSCBench-KITTI-360, and 2/4/2 modes for VAA-KITTI/VAA-KITTI-360/VAA-STU. PPSC uses a 7^3 neighborhood with $\alpha = 0.75$ and $\gamma = 2$. Remaining hyperparameters and implementation details are provided in the Appendix.

Table 2: **Quantitative OOD detection results** on VAA-KITTI, VAA-KITTI-360, and VAA-STU. We report tolerance-aware AuPRC_r at 0.8 m, 1.0 m, and 1.2 m, together with AuROC. Best results are shown in **bold**.

Method	Score	VAA-KITTI				VAA-KITTI-360				VAA-STU			
		$\mathrm{AuPRC}_r \uparrow$			AuROC↑	$\mathrm{AuPRC}_r \uparrow$			AuROC↑	$\mathrm{AuPRC}_r \uparrow$			AuROC↑
		0.8m	1.0m	1.2m		0.8m	1.0m	1.2m		0.8m	1.0m	1.2m	
SGN (Mei et al., 2024)	ASS	8.22	14.45	22.93	60.51	13.27	28.57	52.99	47.27	2.91	5.74	11.32	64.51
OccOoD (Zhang et al., 2025a)	ASS	10.80	16.05	23.42	61.96	14.26	29.72	54.35	47.84	3.05	6.15	11.39	65.37
ProOOD (+SGN) (Zhang et al., 2026b)	EchoOOD	27.86	43.55	62.65	64.31	**14.56**	**30.83**	**57.50**	46.91	**7.98**	**14.54**	24.21	62.11
UMPOcc (+SGN)	EchoOOD+PPSC	**31.59**	**50.44**	**74.44**	**89.63**	13.39	27.71	51.09	**52.40**	4.96	11.72	**24.53**	**77.15**

4.2 Main Results

Semantic occupancy prediction. Tables 1 and 3 compare UMPOcc with existing occupancy and OOD-aware methods across SemanticKITTI, SSCBench-KITTI-360, and their shifted VAA counterparts. UMPOcc consistently improves the corresponding SGN and VoxDet baselines, with consistent gains on tail categories. On SemanticKITTI, UMPOcc reaches 15.02/4.85 and 18.93/7.28 mIoU/tail mIoU with SGN and VoxDet, respectively. The improvement transfers to SSCBench-KITTI-360, where UMPOcc (+SGN) obtains 18.08/8.44 and the VoxDet-based variant reaches 22.03/13.53. Under distribution shift, the same trend remains on VAA-KITTI and VAA-KITTI-360, where UMPOcc improves both overall and tail-class accuracy over the corresponding OOD-aware baselines. These results demonstrate that modeling multi-modal source support improves rare-class recognition and generalizes across datasets and occupancy backbones.

Out-of-distribution detection. Table 2 reports OOD detection results across three benchmarks. On VAA-KITTI, UMPOcc consistently surpasses ProOOD at all spatial tolerances, improving AuPRC_r from 27.86/43.55/62.65 to 31.59/50.44/74.44 and AuROC from 64.31 to 89.63. The large AuROC gain indicates substantially stronger separation between known and novel voxels. On VAA-KITTI-360, UMPOcc further improves AuROC from 46.91 to 52.40, although ProOOD remains stronger under tolerance-aware AuPRC_r. A similar trend appears on the real-world VAA-STU benchmark: UMPOcc raises AuROC from 62.11 to 77.15 and achieves the best AuPRC_r at 1.2 m. Overall, UMPOcc provides markedly stronger global OOD discrimination, while its localization gains are most pronounced on VAA-KITTI.

Table 3: Quantitative results on the SSCBench-KITTI-360 test set (Li et al., 2024b) and VAA-KITTI-360 (Zhang et al., 2025a) (Single-Frame Evaluation for Fair Comparison). † denotes our local reproduction using the released VoxDet checkpoint.

Method	IoU	car (2.85%)	bicycle (0.01%)	motorcycle (0.01%)	truck (0.16%)	other-veh. (5.75%)	person (0.02%)	road (14.98%)	parking (2.31%)	sidewalk (6.43%)	other-grnd. (2.05%)	building (15.67%)	fence (0.96%)	vegetation (41.99%)	terrain (7.10%)	pole (0.22%)	traf.-sign (0.06%)	other-struct. (4.33%)	other-obj. (0.28%)	mIoU	*Tail* mIoU
SSCBench-KITTI-360																					
MonoScene (Cao & de Charette, 2022)	37.87	19.34	0.43	0.58	8.02	2.03	0.86	48.35	11.38	28.13	3.32	32.89	3.53	26.15	16.75	6.92	5.67	4.20	3.09	12.31	3.61
TPVFormer (Huang et al., 2023b)	40.22	21.56	1.09	1.37	8.06	2.57	2.38	52.99	11.99	31.07	3.78	34.83	4.80	30.08	17.52	7.46	5.86	5.48	2.70	13.64	4.13
VoxFormer (Li et al., 2023b)	38.76	17.84	1.16	0.89	4.56	2.06	1.63	47.01	9.67	27.21	2.89	31.18	4.97	28.99	14.69	6.51	6.92	3.79	2.43	11.91	3.44
OccFormer (Zhang et al., 2023)	40.27	22.58	0.66	0.26	9.89	3.82	2.77	54.30	13.44	31.53	3.55	36.42	4.80	31.00	19.51	7.77	8.51	6.95	4.60	13.81	4.92
Symphonies (Jiang et al., 2024b)	44.12	30.02	1.85	5.90	25.07	12.06	8.20	54.94	13.83	32.76	6.93	35.11	8.58	38.33	11.52	14.01	9.57	14.44	11.28	18.58	10.84
SGN (Mei et al., 2024)	46.22	**28.20**	2.09	3.02	11.95	3.68	4.20	59.49	**14.50**	36.53	4.24	39.79	7.14	36.61	**23.10**	**14.86**	**16.14**	8.24	4.95	17.71	8.17
ProOOD-S (Zhang et al., 2026b)	**46.62**	27.23	**2.75**	4.23	12.06	5.76	4.83	60.07	14.02	36.57	5.03	**40.92**	7.13	36.96	22.84	14.15	15.47	8.82	4.93	17.99	8.35
UMPOcc (+SGN)	**46.62**	27.46	2.72	**4.34**	**12.16**	**5.81**	**4.94**	**60.15**	14.28	**36.71**	**5.08**	**40.92**	**7.22**	**37.00**	22.89	14.24	15.62	**8.88**	**5.03**	**18.08**	**8.44**
VoxDet (Li et al., 2025b)	48.22	28.94	**3.88**	5.23	16.20	6.58	6.15	62.02	16.27	39.69	5.34	43.36	9.10	38.73	**22.79**	17.22	**19.74**	10.27	7.99	19.97	10.92
ProOOD-V (Zhang et al., 2026b)	48.23	28.89	3.77	7.23	18.83	6.49	5.83	62.31	16.23	**40.08**	5.36	**43.75**	9.29	38.66	22.51	17.28	19.42	10.67	8.25	20.27	11.52
UMPOcc (+VoxDet)†	**48.91**	**33.76**	1.01	**9.25**	**24.28**	**15.00**	**12.52**	**63.19**	**17.91**	37.57	**9.32**	39.13	**11.44**	**41.78**	14.82	**18.04**	14.60	**17.93**	**15.03**	**22.03**	**13.53**
VAA-KITTI-360																					
SGN (Mei et al., 2024)	**39.68**	20.68	**0.05**	**0.00**	**0.00**	**2.92**	0.00	47.59	14.66	24.71	2.56	32.83	**3.52**	**28.68**	11.63	7.44	8.67	4.29	**5.52**	11.99	3.10
ProOOD-S (Zhang et al., 2026b)	39.38	22.17	**0.05**	**0.00**	**0.00**	0.05	2.10	48.19	16.50	**25.48**	**3.62**	33.75	2.89	28.08	12.86	**8.78**	**10.29**	5.57	4.71	12.52	3.71
UMPOcc (+SGN)	39.42	**22.24**	**0.05**	**0.00**	**0.00**	0.06	**2.53**	**48.21**	**16.66**	**25.48**	3.60	**33.81**	2.64	28.38	**12.90**	8.74	9.89	**5.62**	4.88	**12.54**	**3.73**
VoxDet (Li et al., 2025b)	42.22	21.38	**0.99**	0.00	0.00	**0.37**	2.61	51.41	12.19	**28.84**	2.68	36.19	3.50	**29.86**	12.54	9.30	15.83	5.46	8.35	13.42	5.29
ProOOD-V (Zhang et al., 2026b)	42.54	23.11	0.17	0.00	0.00	0.07	3.60	51.49	14.49	28.51	**3.60**	37.31	4.01	29.79	12.47	**10.64**	**19.01**	7.17	**9.83**	14.18	6.18
UMPOcc (+VoxDet)†	**42.79**	**23.37**	0.41	**0.69**	0.00	0.16	**4.19**	**55.51**	**17.36**	27.17	3.58	**37.33**	**6.81**	29.52	**12.79**	10.47	18.32	**7.43**	9.67	**14.71**	**6.25**

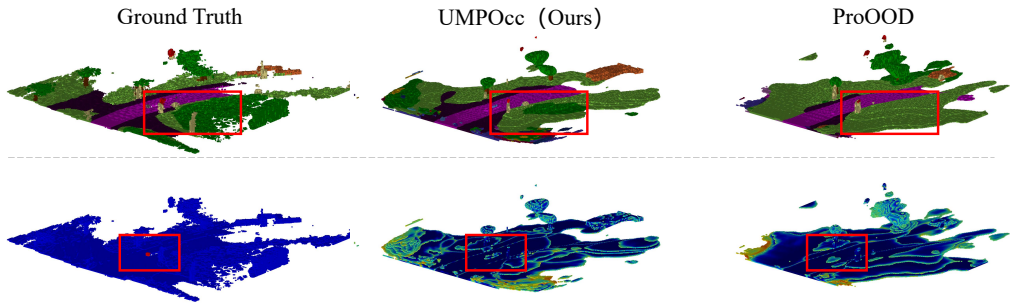

Figure 3: **Qualitative comparison.** The top row compares semantic occupancy predictions, while the bottom row visualizes OOD responses. Red boxes highlight challenging regions. Compared with ProOOD, UMPOcc better preserves the semantic structure of rare known regions and produces a more localized anomaly response around the unknown object.

Table 4: Component ablation of UMPOcc on VAA-KITTI under the matched local setting.

UMPOF	PPSC	mIoU↑	*Tail* mIoU↑	AuPRC_r ↑		AuROC↑
				1.0m	1.2m	
–	–	11.86	2.76	37.80	55.52	64.50
✓	–	**11.97**	**2.84**	37.82	55.69	64.59
–	✓	11.86	2.76	**50.47**	74.32	89.24
✓	✓	**11.97**	**2.84**	50.44	**74.44**	**89.63**

Table 5: Effect of PPSC with the adapted model fixed across OOD benchmarks.

Dataset	Score	AuPRC_r@1.0m↑	AuPRC_r@1.2m↑	AuROC↑
VAA-KITTI	EchoOOD	37.82	55.69	64.59
	+ PPSC	**50.44**	**74.44**	**89.63**
VAA-KITTI-360	EchoOOD	26.57	49.86	45.35
	+ PPSC	**27.71**	**51.09**	**52.40**
VAA-STU	EchoOOD	9.60	20.62	67.38
	+ PPSC	**11.72**	**24.53**	**77.15**

Qualitative results. Figure 3 qualitatively compares UMPOcc with ProOOD. For semantic occupancy prediction (top), UMPOcc better preserves the structure and semantic extent of the highlighted region, producing a prediction closer to the ground truth, whereas ProOOD shows incomplete or less accurate reconstruction. For OOD detection (bottom), UMPOcc yields a more concentrated response around the annotated anomaly, while reducing diffuse activations in the surrounding region. These observations are consistent with the quantitative improvements in tail-class recognition and OOD discrimination.

4.3 Analysis and Ablation

Component ablation. Table 4 disentangles the roles of UMPOF and PPSC. UMPOF mainly improves semantic prediction, increasing mIoU/tail mIoU from 11.86/2.76 to 11.97/2.84, while only marginally affecting the uncalibrated EchoOOD score. In contrast, PPSC accounts for most of the OOD gain, raising AuPRC_r@1.2m from 55.52 to 74.32 and AuROC from 64.50 to 89.24. Combining both retains the occupancy gains and yields the best 1.2m AuPRC_r and AuROC.

PPSC generalization. Table 5 evaluates PPSC on the same adapted model across three OOD benchmarks. For each benchmark, the UMPOF-adapted occupancy model is kept fixed and EchoOOD is used as the inherited base scorer; only PPSC calibration is toggled. PPSC consistently improves both tolerance-aware localization and global discrimination. In particular, AUROC increases from 64.59 to 89.63 on VAA-KITTI, from 45.35 to 52.40 on VAA-KITTI-360, and from 67.38 to 77.15 on VAA-STU. These consistent gains indicate that occupancy-supported peak calibration generalizes across different anomaly distributions.

5 Conclusion

We presented UMPOcc for disambiguating semantic rarity from distributional novelty in OOD 3D occupancy prediction. UMPOF models multi-modal known-class support to better preserve rare semantic patterns, while PPSC uses occupancy-aware peak preservation to calibrate fragmented anomaly responses. Extensive experiments across datasets and backbones demonstrate improved tail-class occupancy prediction and OOD discrimination, showing the benefit of complementary distributional and spatial reasoning.

AI USE STATEMENT

In this work, we used generative AI tools to assist with research planning and experimental design, implementation and debugging, translation, and manuscript drafting and editing. We did not use generative AI to generate synthetic datasets, construct theoretical models or mathematical proofs, formulate the core scientific claims, or perform qualitative or thematic data analysis. All AI-assisted code and experimental outputs were manually reviewed and tested by the authors, and any identified errors were corrected before use. All experimental results, numerical claims, and conclusions were verified against the corresponding implementations and evaluation outputs. The authors take full responsibility for the final content of this work, including text, claims, code, and other artifacts produced with the assistance of generative AI.

ETHICS STATEMENT

This work does not involve human-subject experiments or the collection of new personally identifiable information. All experiments are conducted on existing public research benchmarks and follow their established evaluation protocols. The proposed method is developed for research on 3D scene perception and OOD detection. We are not aware of additional ethical concerns specific to the method beyond those generally associated with perception systems for autonomous driving.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide an anonymous code repository containing the implementation, configuration files, and evaluation scripts used in this work. The Appendix documents the training and inference environments, dataset protocols, optimization settings, prototype-bank construction, PPSC configuration, and additional implementation details. We also report the main hyperparameters and evaluation settings required to reproduce the quantitative results in the paper. The anonymous repository is available at `https://anonymous.4open.science/r/UMPOcc-ICLR/`.

REFERENCES

Till Beemelmanns, Alexey Nekrasov, Stefan Vilceanu, Jonas Steinhaus, Timo Woopen, Bastian Leibe, and Lutz Eckstein. Query2uncertainty: Robust uncertainty quantification and calibration for 3d object detection under distribution shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4677–4686, 2026.

Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9297–9307, 2019.

Anh-Quan Cao and Raoul de Charette. Monoscene: Monocular 3d semantic scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3991–4001, 2022.

Loick Chambon, Eloi Zablocki, Alexandre Boulch, Mickael Chen, and Matthieu Cord. Gaussrender: Learning 3d occupancy with gaussian rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27010–27020, 2025.

Dubing Chen, Huan Zheng, Yucheng Zhou, Xianfei Li, Wenlong Liao, Tao He, Pai Peng, and Jianbing Shen. Semantic causality-aware vision-based 3d occupancy prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 24878–24888, 2025.

Qi Chen and Hu Ding. Dual energy-based model with open-world uncertainty estimation for out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25728–25737, 2025.

Shiyuan Chen, Wei Sui, Bohao Zhang, Zeyd Boukhers, John See, and Cong Yang. Unleashing semantic and geometric priors for 3d scene completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 3020–3028, 2026.

Yi Feng, Junwu E, Zizhan Guo, Yu Ma, Hanli Wang, and Rui Fan. An instance-centric panoptic occupancy prediction benchmark for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14219–14228, 2026.

Yang Gao, Wuyang Li, Po-Chien Luan, and Alexandre Alahi. Deformable gaussian occupancy: Decoupling rigid and nonrigid motion with factorized distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28588–28598, 2026.

Matej Grcić, Petra Bevandić, Zoran Kalafatić, and Siniša Šegvić. Dense anomaly detection by robust learning on synthetic negative data. *arXiv preprint arXiv:2112.12833*, 2021.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1321–1330. PMLR, 2017.

Severin Heidrich, Till Beemelmanns, Alexey Nekrasov, Bastian Leibe, and Lutz Eckstein. Occuq: Exploring efficient uncertainty quantification for 3d occupancy prediction. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4884–4891, 2025.

Dan Hendrycks and Kevin Gimpel. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu. Tri-perspective view for vision-based 3d semantic occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9223–9232, 2023a.

Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu. 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. IEEE, 2023b.

Yuanhui Huang, Wenzhao Zheng, Yunpeng Zhang, Jie Zhou, and Jiwen Lu. Gaussianformer: Scene as gaussians for vision-based 3d semantic occupancy prediction. In *Computer Vision – ECCV 2024*, pp. 376–393. Springer, 2024.

Yuanhui Huang, Amonnut Thammatadatrakoon, Wenzhao Zheng, Yunpeng Zhang, Dalong Du, and Jiwen Lu. Gaussianformer-2: Probabilistic gaussian superposition for efficient 3d occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27477–27486, 2025.

Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, and Xinggang Wang. Symphonize 3d semantic scene completion with contextual instance queries. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20258–20267, 2024a.

Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, and Xinggang Wang. Symphonize 3d semantic scene completion with contextual instance queries. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20258–20267. IEEE, 2024b.

Jonas Kälble, Sascha Wirges, Maxim Tatarchenko, and Eddy Ilg. Evocc: Accurate semantic occupancy for automated driving using evidence theory. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27467–27476, 2025.

Jungho Kim, Changwon Kang, Dongyoung Lee, Sehwan Choi, and Jun Won Choi. Protoocc: Accurate, efficient 3d occupancy prediction using dual branch encoder-prototype query decoder. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 4284–4292, 2025.

Jean Lahoud, Fahad Shahbaz Khan, Hisham Cholakkal, Rao Muhammad Anwer, and Salman Khan. Long-tailed 3d semantic segmentation with adaptive weight constraint and sampling. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5037–5044, 2024.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In *Advances in Neural Information Processing Systems*, volume 31, pp. 7167–7177, 2018.

Ziyang Leng, Jiawei Yang, Wenlong Yi, and Bolei Zhou. Occupancy learning with spatiotemporal memory. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 26569–26578, 2025.

Bohan Li, Yasheng Sun, Zhujin Liang, Dalong Du, Zhuanghui Zhang, Xiaofeng Wang, Yunnan Wang, Xin Jin, and Wenjun Zeng. Bridging stereo geometry and bev representation with reliable mutual interaction for semantic scene completion. *arXiv preprint arXiv:2303.13959*, 2023a.

Bohan Li, Jiajun Deng, Wenyao Zhang, Zhujin Liang, Dalong Du, Xin Jin, and Wenjun Zeng. Hierarchical temporal context learning for camera-based semantic scene completion. In *European Conference on Computer Vision*, pp. 131–148. Springer, 2024a.

Bohan Li, Jiajun Deng, Yasheng Sun, Xiaofeng Wang, Xin Jin, and Wenjun Zeng. Hierarchical context alignment with disentangled geometric and temporal modeling for semantic occupancy prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Shawn Li, Huixian Gong, Hao Dong, Tiankai Yang, Zhengzhong Tu, and Yue Zhao. Dpu: Dynamic prototype updating for multimodal out-of-distribution detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10193–10202, 2025a.

Wuyang Li, Zhu Yu, and Alexandre Alahi. Voxdet: Rethinking 3d semantic scene completion as dense object detection. In *Advances in Neural Information Processing Systems*, volume 38, 2025b.

Yiming Li, Zhiding Yu, Christopher Choy, Chaowei Xiao, Jose M. Alvarez, Sanja Fidler, Chen Feng, and Anima Anandkumar. Voxformer: Sparse voxel transformer for camera-based 3d semantic scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9087–9098, 2023b.

Yiming Li, Sihang Li, Xinhao Liu, Moonjun Gong, Kenan Li, Nuo Chen, Zijun Wang, Zhiheng Li, Tao Jiang, Fisher Yu, Yue Wang, Hang Zhao, Zhiding Yu, and Chen Feng. Sscbench: A large-scale 3d semantic scene completion benchmark for autonomous driving. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13333–13340, 2024b.

Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. Energy-based out-of-distribution detection. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

Chengxin Lv, Yihui Li, Hongyu Yang, and YunHong Wang. Gau-occ: Geometry-completed gaussians for multi-modal 3d occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14198–14207, 2026.

Jianbiao Mei, Yu Yang, Mengmeng Wang, Junyu Zhu, Jongwon Ra, Yukai Ma, Laijian Li, and Yong Liu. Camera-based 3d semantic scene completion with sparse guidance network. *IEEE Transactions on Image Processing*, 33:5468–5481, 2024.

Alexey Nekrasov, Malcolm Burdorf, Stewart Worrall, Bastian Leibe, and Julie Stephany Berrio Perez. Spotting the unexpected (stu): A 3d lidar dataset for anomaly segmentation in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11875–11885, 2025.

David Sculley. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, pp. 1177–1178, 2010.

Song Wang, Jiawei Yu, Wentong Li, Wenyu Liu, Xiaolu Liu, Junbo Chen, and Jianke Zhu. Not all voxels are equal: Hardness-aware semantic scene completion with self-distillation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14792–14801. IEEE, 2024.

Yuping Wang, Xiangyu Huang, Xiaokang Sun, Mingxuan Yan, Shuo Xing, Zhengzhong Tu, and Jiachen Li. Uniocc: A unified benchmark for occupancy forecasting and prediction in autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 25560–25570, 2025.

Yu Xue, Longjun Gao, Yuanqi Su, HaoAng Lu, and Xiaoning Zhang. Sparsity-aware voxel attention and foreground modulation for 3d semantic scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5751–5761, June 2026.

Qiucheng Yu, Yuan Xie, and Xin Tan. Shtocc: Effective 3d occupancy prediction with sparse head and tail voxels. *Pattern Recognition*, pp. 114340, 2026.

Zhu Yu, Runmin Zhang, Jiacheng Ying, Junchen Yu, Xiaohai Hu, Lun Luo, Si-Yuan Cao, and Hui-Liang Shen. Context and geometry aware voxel transformer for semantic scene completion. *Advances in Neural Information Processing Systems*, 37:1531–1555, 2024.

Zhu Yu, Bowen Pang, Lizhe Liu, Runmin Zhang, Qiang Li, Si-Yuan Cao, Maochun Luo, Mingxia Chen, Sheng Yang, and Hui-Liang Shen. Language driven occupancy prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7548–7558, 2025.

Fengyi Zhang, Xiangyu Sun, Huitong Yang, Zheng Zhang, Zi Huang, and Yadan Luo. Test-time 3d occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35691–35701, 2026a.

Yuheng Zhang, Mengfei Duan, Kunyu Peng, Yuhang Wang, Ruiping Liu, Fei Teng, Kai Luo, Zhiyong Li, and Kailun Yang. Out-of-distribution semantic occupancy prediction. *arXiv preprint arXiv:2506.21185*, 2025a.

Yuheng Zhang, Mengfei Duan, Kunyu Peng, Yuhang Wang, Di Wen, Danda Pani Paudel, Luc Van Gool, and Kailun Yang. Proood: Prototype-guided out-of-distribution 3d occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14241–14252, 2026b.

Yunpeng Zhang, Zheng Zhu, and Dalong Du. Occformer: Dual-path transformer for vision-based 3d semantic occupancy prediction. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9399–9409. IEEE, 2023.

Zongkai Zhang, Bin Gao, Jingrui Ye, Huan Jin, Lihui Jiang, and Wenming Yang. Clip prior-guided 3d open-vocabulary occupancy prediction. *Pattern Recognition*, 162:111347, 2025b.

Changqing Zhou, Yueru Luo, Han Zhang, Zeyu Jiang, and Changhao Chen. Monocular open vocabulary occupancy prediction for indoor scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21627–21637, 2026.

Xiaoyu Zhou, Jingqi Wang, Yongtao Wang, Yufei Wei, Nan Dong, and Ming-Hsuan Yang. Autoocc: Automatic open-ended semantic occupancy annotation via vision-language guided gaussian splatting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3367–3377, 2025.

Xubo Zhu, Haoyang Zhang, Fei He, Rui Wu, Yanhu Shan, Wen Yang, and Huai Yu. Dr.occ: Depth- and region-guided 3d occupancy from surround-view cameras for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28610–28619, 2026.

Reihaneh Zohrabi, Hosein Hasani, Mahdieh Soleymani Baghshah, Anna Rohrbach, Marcus Rohrbach, and Mohammad Hossein Rohban. Spurious-aware prototype refinement for reliable out-of-distribution detection. In *Advances in Neural Information Processing Systems*, volume 38, pp. 49813–49857, 2025.

Sicheng Zuo, Wenzhao Zheng, Yuanhui Huang, Jie Zhou, and Jiwen Lu. Gaussianworld: Gaussian world model for streaming 3d occupancy prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6772–6781, 2025.

APPENDIX CONTENTS

A OOD Score Construction . . . 15
A.1 Inherited Base Anomaly Score . . . 15
A.2 Scene-Level Energy Normalization . . . 15
A.3 Score Composition and Evaluation Convention . . . 16

B Additional Method Details and Analytical Properties . . . 17
B.1 Offline Multi-Prototype Bank Construction . . . 17
B.2 Online Prototype Learning and EMA Updates . . . 17
B.3 Prototype Initialization and Maturity Gating . . . 17
B.4 Distributional Feature Refinement . . . 18
B.5 Tail-Aware Refinement . . . 18
B.6 Analytical Properties of UMPOF and PPSC . . . 18

C Implementation and Reproducibility Details . . . 19
C.1 Experimental Protocols . . . 19
C.2 Key Implementation Settings . . . 19
C.3 Code and Reproducibility . . . 20

D Additional Experimental Analysis . . . 20
D.1 Comparison of Anomaly Scores . . . 20
D.2 UMPOF–PPSC Score-Composition Analysis . . . 20
D.3 Offline Prototype-Bank Capacity . . . 21
D.4 Calibration Diagnostic . . . 21
D.5 Computational Efficiency . . . 21

E Additional Qualitative Results . . . 21
F Limitations and Future Work . . . 21
F Limitations . . . 21
F Future Work . . . 22

A OOD SCORE CONSTRUCTION

This section specifies the anomaly-score construction used throughout our OOD experiments. We separate score construction from the main method description because the base anomaly scorer is inherited from the prototype-based OOD framework (Zhang et al., 2026b), whereas UMPOF and PPSC are the proposed representation and spatial-calibration components. This separation also makes explicit which score is used for the main results and which variants are used only for diagnostic analysis.

A.1 INHERITED BASE ANOMALY SCORE

Let $\mathbf{F}$ denote the refined voxel feature field produced by the occupancy network after UMPOF-based representation refinement, and let $\mathbf{L}$ denote the corresponding semantic logits. Following ProOOD (Zhang et al., 2026b), we use EchoOOD as the inherited voxel-level anomaly scorer. EchoOOD combines local logit coherence with local and global prototype consistency and produces a normalized pre-calibration anomaly map, which we denote by

$$\mathbf{s}^{\text{base}} = \mathcal{S}_{\text{base}}(\mathbf{F}, \mathbf{L}). \tag{25}$$

Here, a larger value of s_v^{base} indicates that voxel v is less supported by the source-domain semantic representation.

Importantly, EchoOOD is treated as the *base score generator* rather than as part of PPSC. PPSC operates on the resulting anomaly map and does not modify the semantic logits or the occupancy prediction. For the main OOD experiments, the calibrated score is

$$\mathbf{a}^{\text{PPSC}} = \text{PPSC}\left(\mathbf{s}^{\text{base}}, \mathbf{L}\right). \tag{26}$$

Thus, the primary evaluation path is

$$(\mathbf{F}, \mathbf{L}) \xrightarrow{\mathcal{S}_{\text{base}}} \mathbf{s}^{\text{base}} \xrightarrow{\text{PPSC}} \mathbf{a}^{\text{PPSC}}. \tag{27}$$

This convention avoids conflating improvements due to representation learning, spatial calibration, and score-level fusion.

A.2 SCENE-LEVEL UMPOF ENERGY NORMALIZATION

In addition to the inherited anomaly score, UMPOF provides an independent distributional uncertainty signal. Recall from Sec. 3.3 that the multi-prototype energy for voxel feature $\mathbf{f}_v$ is

$$e_U(\mathbf{f}_v) = -\,\text{LSE}_{k,m}\left[\ell_{k,m}(\mathbf{f}_v)\right], \tag{28}$$

where $\ell_{k,m}$ incorporates the empirical prior and dispersion of mode (k, m). Features well supported by at least one source-domain mode therefore receive lower energy, whereas features lying away from all known modes receive higher energy.

Because the absolute scale of e_U can vary across scenes, we normalize the energy independently for each scene before using it as an anomaly signal:

$$s_v^U = \text{Norm}_{\text{scene}}\left(e_U(\mathbf{f}_v)\right), \tag{29}$$

where

$$\text{Norm}_{\text{scene}}(z_v) = \begin{cases} \dfrac{z_v - \min_u z_u}{\max_u z_u - \min_u z_u}, & \max_u z_u > \min_u z_u, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

The resulting score satisfies $s_v^U \in [0, 1]$, with larger values indicating weaker support under the source-domain multi-prototype field.

We emphasize that $\mathbf{s}^U$ is an *auxiliary distributional signal*. It is not required by the main PPSC-only evaluation path in Eq. 27. This distinction is useful because it allows us to separately analyze whether gains arise from the learned multi-modal representation or from explicit score-level fusion.

Table 6: OOD score constructions used in our experiments. The PPSC-only configuration is used for the main reported OOD results; the remaining variants are used only when explicitly stated.

Score mode	Definition	Usage
Base	$\mathbf{s}^{\text{base}}$	Diagnostic
UMPOF energy	$\mathbf{s}^{U}$	Diagnostic
Max fusion	$\max(\mathbf{s}^{\text{base}}, \mathbf{s}^{U})$	Diagnostic
PPSC	$\text{PPSC}(\mathbf{s}^{\text{base}}, \mathbf{L})$	**Main**
Residual fusion	Eq. 34	Diagnostic

A.3 Score Composition and Evaluation Convention

Main PPSC-only configuration. Unless otherwise stated, all main OOD results use the PPSC-calibrated base score in Eq. 26. In particular, the results reported as "PPSC" in the main OOD comparison do not implicitly combine multiple anomaly maps. The final pre-masking anomaly score is therefore

$$\widetilde{s}_v = \text{PPSC}\left(\mathbf{s}^{\text{base}}, \mathbf{L}\right)_v . \tag{31}$$

Direct density fusion. For diagnostic experiments, we additionally consider a simple maximum rule that combines the inherited anomaly score with UMPOF density evidence:

$$a_v^{\max} = \max\left(s_v^{\text{base}}, s_v^{U}\right). \tag{32}$$

This variant tests whether adding distributional energy directly to the inherited score provides complementary OOD evidence.

PPSC-first residual fusion. We also evaluate a conservative residual fusion strategy. We first calibrate the inherited score:

$$\widetilde{s}_v^{\text{base}} = \text{PPSC}\left(\mathbf{s}^{\text{base}}, \mathbf{L}\right)_v , \tag{33}$$

and then allow UMPOF energy to increase the calibrated response only when it provides stronger anomaly evidence:

$$a_v^{\text{fuse}} = \text{clip}\left(\widetilde{s}_v^{\text{base}} + \lambda_g \left[s_v^{U} - \widetilde{s}_v^{\text{base}}\right]_+, 0, 1\right), \tag{34}$$

where $[x]_+ = \max(x, 0)$ and λ_g controls the contribution of the distributional residual. This one-sided construction prevents the UMPOF density signal from suppressing an anomaly already assigned a high calibrated score by PPSC.

Empty-voxel handling. PPSC is designed to identify anomalous *occupied* voxels. Following the main implementation, the calibrated response is first clipped to $[0, 1]$. Voxels predicted as empty by the semantic occupancy head are then assigned the minimum anomaly score of the scene. Denoting $\hat{y}_v = \arg\max_k L_{k,v}$ and class 0 as empty, the score used for evaluation can be written as

$$a_v = \begin{cases} \text{clip}(\widetilde{s}_v, 0, 1), & \hat{y}_v \neq 0, \\ \min\limits_u \text{clip}(\widetilde{s}_u, 0, 1), & \hat{y}_v = 0, \end{cases} \tag{35}$$

where $\widetilde{s}$ denotes the score produced by the selected scoring mode. This prevents predicted free space from receiving a high OOD response solely because of uncertain semantic evidence.

Evaluation modes. For clarity, Table 6 summarizes the score constructions used in the paper. The PPSC-only path is the default configuration for the main OOD results, while the remaining variants are reserved for supplementary diagnostic analyses.

The supplementary experiments in Appendix D compare these alternatives explicitly. Keeping the main evaluation convention fixed is important: improvements attributed to PPSC are measured from the calibrated inherited anomaly map, rather than from an unspecified combination of EchoOOD and UMPOF density.

B Additional Method Details and Analytical Properties

This section complements the main method description with implementation-level definitions of the multi-prototype representation and several simple analytical properties of UMPOF and PPSC. We focus only on details that are useful for understanding or reproducing the method; numerical hyperparameters are deferred to Appendix C.

B.1 Offline Multi-Prototype Bank Construction

For each non-empty semantic class k, we collect source-domain training features with ground-truth label k and partition them into M modes using class-wise MiniBatch K-Means (Sculley, 2010). Let $\mathcal{I}_{k,m}$ denote the set of features assigned to mode m of class k. The empirical mode prior is

$$\pi_{k,m} = \frac{|\mathcal{I}_{k,m}|}{\sum_{j=1}^{M} |\mathcal{I}_{k,j}|}, \qquad \sum_{m=1}^{M} \pi_{k,m} = 1. \tag{36}$$

Given the corresponding mode center $\mathbf{p}_{k,m}$, we estimate its dispersion using the mean cosine distance

$$\nu_{k,m} = \frac{1}{|\mathcal{I}_{k,m}|} \sum_{i \in \mathcal{I}_{k,m}} \left(1 - \bar{\mathbf{f}}_i^{\top} \bar{\mathbf{p}}_{k,m}\right), \tag{37}$$

where bars denote ℓ_2 normalization. Empty and ignored voxels are excluded from these statistics.

Together, $(\mathbf{p}_{k,m}, \pi_{k,m}, \nu_{k,m})$ specify the mode evidence in Eq. (3) of the main paper. The prior reflects how frequently the mode occurs, while $\nu_{k,m}$ adapts its effective support to the observed intra-mode variation. The inference density bank is constructed only from source-domain training features and remains frozen during evaluation.

B.2 Online Prototype Learning and EMA Updates

During occupancy adaptation, UMPOF maintains an online multi-mode representation for feature refinement. For a source feature $\mathbf{f}_i$ of class k, its soft responsibility for mode m is

$$r_{i,m} = \frac{\exp(-d_{k,m}(\mathbf{f}_i)/\tau_a)}{\sum_{j=1}^{M} \exp(-d_{k,j}(\mathbf{f}_i)/\tau_a)}. \tag{38}$$

The assignment-weighted batch estimate is

$$\widetilde{\mathbf{p}}_{k,m} = \frac{\sum_{i:y_i=k} r_{i,m} \mathbf{f}_i}{\sum_{i:y_i=k} r_{i,m} + \epsilon}, \tag{39}$$

and the center is updated by

$$\mathbf{p}_{k,m} \leftarrow (1-\beta)\mathbf{p}_{k,m} + \beta \widetilde{\mathbf{p}}_{k,m}. \tag{40}$$

The corresponding prior is updated from the normalized soft-assignment mass, as described in Eq. (7) of the main paper. Prototype statistics are updated with stop-gradient and only after the warm-up stage, preventing unstable early features from directly perturbing the occupancy backbone.

We distinguish this training-time update from the offline inference-bank construction above. Online modes support representation learning and feature refinement during adaptation, whereas the density bank used for evaluation is estimated from source training features and frozen at test time.

B.3 Prototype Initialization and Maturity Gating

Prototype refinement is enabled only after sufficient source-domain prototype statistics have been accumulated. A mode is regarded as initialized after its class row receives a valid source-label update. A global maturity gate further checks class coverage and prototype quality before enabling multi-prototype feature refinement.

Before this gate becomes active, the original occupancy feature path is used without prototype residuals. After activation, only initialized modes participate in the mixture. Consequently, an unavailable or unreliable mode cannot introduce an arbitrary residual into the occupancy representation. The concrete warm-up setting used in our implementation is reported in Appendix C.2.

B.4 Distributional Feature Refinement

For a masked voxel $v \in \mathcal{U}$, UMPOF aggregates semantic support from the mature modes according to

$$w_{v,k,m} = \mathrm{softmax}_{k,m}\left(\frac{\bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_{k,m}}{\tau_r}\right), \qquad \Delta\mathbf{x}_v = \sum_{k,m} w_{v,k,m}\mathbf{p}_{k,m}. \tag{41}$$

Only masked voxels predicted as non-empty by the auxiliary occupancy head are refined:

$$\mathbf{x}'_v = \mathbf{x}_v + \alpha_r \Delta\mathbf{x}_v. \tag{42}$$

The residual form is deliberate: the original visual/geometric feature remains the dominant representation, while the prototype mixture provides a class-consistent semantic correction. Unmasked voxels and masked voxels not selected by the occupancy gate retain their original feature path.

B.5 Tail-Aware Refinement

The tail branch operates on class-level aggregates rather than individual sub-modes. For a tail class $k \in \mathcal{T}$, its aggregate prototype is

$$\mathbf{p}_k = \sum_{m=1}^{M} \pi_{k,m}\mathbf{p}_{k,m}, \qquad \bar{\mathbf{p}}_k = \frac{\mathbf{p}_k}{\|\mathbf{p}_k\|_2 + \epsilon}. \tag{43}$$

Candidate voxels are ranked by their maximum affinity to the tail anchors,

$$u_v = \max_{k\in\mathcal{T}} \bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_k, \tag{44}$$

and the top-K_t candidates receive a residual constructed from tail prototypes:

$$\mathbf{x}''_v = \mathbf{x}'_v + \mathcal{A}\left(\sum_{k\in\mathcal{T}} \rho_{v,k}\mathbf{p}_k\right), \qquad \rho_{v,k} = \mathrm{softmax}_{k\in\mathcal{T}}\left(\eta\, \bar{\mathbf{x}}_v^\top \bar{\mathbf{p}}_k\right). \tag{45}$$

Here, $\mathcal{A}(\cdot)$ is the two-layer residual adapter described in the main paper. The tail branch is supervised using source occupancy labels through $\mathcal{L}_{\mathrm{tail}}$; it does not require OOD annotations. Importantly, the class aggregation in Eq. 43 is used only for tail-aware selection and refinement. The UMPOF density in Eq. (4) still evaluates every class–mode pair individually.

B.6 Analytical Properties of UMPOF and PPSC

We finally state several simple properties that clarify the behavior of the two proposed components. These results are not additional assumptions of the method; they follow directly from the definitions in the main paper.

Property 1: UMPOF energy is a smooth best-mode score. Let

$$\ell_{\max}(\mathbf{f}) = \max_{k,m} \ell_{k,m}(\mathbf{f}), \tag{46}$$

and let N be the number of valid class–mode pairs. Then

$$-\ell_{\max}(\mathbf{f}) - \log N \le e_U(\mathbf{f}) \le -\ell_{\max}(\mathbf{f}). \tag{47}$$

Proof. The standard log-sum-exp inequality gives

$$\ell_{\max} \le \mathrm{LSE}_{k,m}(\ell_{k,m}) \le \ell_{\max} + \log N.$$

Multiplying by -1 yields Eq. 47. □

Hence, UMPOF energy behaves as a smooth approximation to the negative evidence of the best supporting known mode, while still aggregating evidence from all available modes.

Property 2: Limiting behavior of the occupancy-support gate. Recall

$$g_v = \mathrm{clip}\left(\frac{h_v - q_{\min}}{\delta}, 0, 1\right)^{\rho}. \tag{48}$$

For $\delta > 0$ and $\rho > 0$,

$$h_v \le q_{\min} \quad \Longrightarrow \quad g_v = 0, \qquad s_v^{\mathrm{PPSC}} = s_v^{\mathrm{pk}}, \tag{49}$$

whereas

$$h_v \ge q_{\min} + \delta \quad \Longrightarrow \quad g_v = 1, \qquad s_v^{\mathrm{PPSC}} = s_v^{\mathrm{sup}}. \tag{50}$$

Proof. Both statements follow directly from clipping the normalized support to $[0, 1]$ before applying the positive exponent ρ. □

Thus, PPSC falls back to its peak-preserving branch when local occupancy evidence is weak and fully trusts occupancy-supported correction only when the local support is sufficiently strong.

Property 3: Local anomaly peaks are stable under PPSC. Assume the normalized anomaly scores lie in $[0, 1]$, $b, c, \alpha \ge 0$, and voxel v is a positive local maximum, i.e., $s_v = m_v > 0$. Its peakness is

$$r_v = \left(\frac{s_v}{s_v + \epsilon}\right)^{\gamma}. \tag{51}$$

Then the perturbation introduced by PPSC satisfies

$$\left|s_v^{\mathrm{PPSC}} - s_v\right| \le (1 - r_v)\max\{\alpha, b, c\}. \tag{52}$$

Proof. Because s_v is a local maximum, $|s_v^{\mathrm{pk}} - s_v| \le (1 - r_v)\alpha$. The occupancy-weighted local score also lies in $[0, 1]$, giving $|s_v^{\mathrm{sup}} - s_v| \le (1 - r_v)\max\{b, c\}$. Since $s_v^{\mathrm{PPSC}} = (1 - g_v)s_v^{\mathrm{pk}} + g_v s_v^{\mathrm{sup}}$ with $g_v \in [0, 1]$, the convex combination satisfies Eq. 52. □

For a confident peak with $s_v \gg \epsilon$, r_v approaches one and the bound approaches zero. This formalizes the intended peak-preserving behavior: strong compact anomaly responses are minimally altered, while lower-confidence responses remain available for neighborhood-based correction.

C Implementation and Reproducibility Details

This section summarizes the experimental protocol and the key implementation settings required to interpret our results. Full configuration files, training scripts, and evaluation code are provided in the anonymous repository, and we therefore avoid duplicating implementation details that are directly available from the released code.

C.1 Experimental Protocols

We follow the single-frame evaluation protocol of ProOOD (Zhang et al., 2026b). Semantic occupancy prediction is evaluated on SemanticKITTI (Behley et al., 2019), SSCBench-KITTI-360 (Li et al., 2024b), VAA-KITTI, and VAA-KITTI-360. OOD detection is evaluated on VAA-KITTI, VAA-KITTI-360, and the real-world VAA-STU benchmark.

For occupancy prediction, we report IoU, mIoU, and tail mIoU. Following ProOOD, classes occupying less than 0.6% of source-domain voxels in SemanticKITTI and less than 0.3% in SSCBench-KITTI-360 are treated as tail classes. For OOD evaluation, we report tolerance-aware AuPRC_r at 0.8, 1.0, and 1.2 m together with AuROC. All reported OOD evaluations use the score convention defined in Appendix A.

C.2 Key Implementation Settings

We evaluate UMPOcc with both SGN (Mei et al., 2024) and VoxDet (Li et al., 2025b). For the SGN branch, training is initialized from the corresponding ProOOD checkpoint and UMPOF is adapted for eight epochs. The principal settings used in our experiments are summarized in Table 7.

Table 7: Key training and inference settings used in our experiments. Dataset-specific and secondary hyperparameters are provided in the anonymous configuration files.

Setting	Configuration
Optimizer	AdamW
Training epochs	8
Initial learning rate	2×10^{-5}
LR schedule	Cosine annealing
Prototype warm-up	200 iterations
λ_d	0.1
M (SemanticKITTI)	4
M (SSCBench-KITTI-360)	4
M (VAA-KITTI)	2
M (VAA-KITTI-360)	4
M (VAA-STU)	2
PPSC neighborhood	$7 \times 7 \times 7$
PPSC α	0.75
PPSC γ	2

The online prototype statistics are updated only after the warm-up stage, whereas the inference density bank is constructed from source-domain training features and kept fixed during evaluation. PPSC has no trainable parameters and is applied after the inherited anomaly score is produced. No target-domain test labels or per-scene optimization are used at inference.

C.3 Code and Reproducibility

Our anonymous repository contains the UMPOF and PPSC implementations, experiment configuration files, training scripts, and evaluation code used to produce the reported results: `https://anonymous.4open.science/r/UMPOcc-ICLR/`.

D Additional Experimental Analysis

This section provides supplementary analyses of anomaly-score construction, prototype-bank capacity, calibration, and computational efficiency. Evaluation protocols that differ from the main tolerance-aware OOD benchmark are explicitly identified below.

D.1 Comparison of Anomaly Scores

Table 8: Comparison of OOD scoring pipelines on VAA-KITTI. We report tolerance-aware AuPRC$_r$ and AuROC.

Score	AuPRC$_r$@0.8m↑	AuPRC$_r$@1.0m↑	AuPRC$_r$@1.2m↑	AuROC↑
Post-processing	5.81	10.02	16.22	63.85
Energy	6.80	12.11	20.08	66.69
Entropy	7.38	12.91	21.05	66.24
ASS	14.47	20.96	29.65	64.20
EchoOOD	27.86	43.55	62.65	64.31
EchoOOD + PPSC (ours)	**31.59**	**50.44**	**74.44**	**89.63**

Table 8 compares representative OOD scoring pipelines on VAA-KITTI. Controlled fixed-model comparisons are provided in Tables 5 and 9.

D.2 UMPOF–PPSC Score-Composition Analysis

Table 9 studies whether the auxiliary UMPOF density signal provides additional benefits when combined with EchoOOD or EchoOOD+PPSC. For each dataset, the underlying model is kept fixed and only the score composition is changed. Explicit density fusion provides no consistent benefit after PPSC, supporting our choice to use UMPOF primarily for representation refinement and PPSC for score calibration.

Table 9: Diagnostic comparison of UMPOF density and PPSC score composition. All rows within each dataset use the same underlying model.

Dataset	Score	AuPRC_r@1.2m↑	AuROC↑
VAA-KITTI	EchoOOD	55.69	64.59
	EchoOOD + UMPOF density	57.00	64.12
	EchoOOD + PPSC	**74.44**	**89.63**
	EchoOOD + PPSC + density residual	74.01	88.96
VAA-KITTI-360	EchoOOD	49.86	45.35
	EchoOOD + UMPOF density	49.93	45.37
	EchoOOD + PPSC	51.09	**52.40**
	EchoOOD + PPSC + density residual	**51.15**	52.39

Table 10: Sensitivity to the number of modes in the offline prototype bank. This diagnostic evaluates standalone UMPOF density using bounded standard AUPR rather than tolerance-aware AuPRC_r.

Dataset	M	Standard AUPR↑	AuROC↑
VAA-KITTI	1	2.25	55.15
	2	2.54	55.24
	4	**3.32**	55.24
	8	2.74	**55.28**
VAA-KITTI-360	1	1.24	46.82
	2	1.26	46.87
	4	1.26	46.89
	8	**1.27**	**47.08**

D.3 Offline Prototype-Bank Capacity

Table 10 evaluates the sensitivity of the offline UMPOF density bank to the number of modes per semantic class. This diagnostic uses a frozen model and reports bounded standard AUPR, rather than the tolerance-aware AuPRC_r used in the main OOD experiments.

D.4 Calibration Diagnostic

Table 11: Calibration diagnostic on VAA-KITTI under the matched local SGN setting. Lower ECE is better.

Method	$\mathrm{ECE}_{sem}\downarrow$	$\mathrm{ECE}_{tail}\downarrow$	$\mathrm{ECE}_{geo}\downarrow$	mIoU↑	Tail↑	AuROC↑
Matched ProOOD	51.71	52.95	**4.64**	11.86	2.76	64.61
UMPOcc	**51.53**	**52.90**	4.66	**11.97**	**2.84**	**89.63**

Table 11 examines the calibration of semantic occupancy predictions under the matched local SGN setting on VAA-KITTI. We report semantic, tail-class, and geometric expected calibration error (ECE) together with occupancy and OOD performance. Here, calibration refers to prediction ECE and is complementary to PPSC's spatial score calibration.

D.5 Computational Efficiency

Table 12 reports the computational overhead of UMPOF and PPSC under matched local settings. FPS and peak inference memory are measured in the same environment for each baseline–UMPOcc pair, while the reported network FLOPs exclude score-level post-processing.

E Additional Qualitative Results

Figure 4 and Figure 5 provide additional qualitative comparisons of semantic occupancy prediction and OOD detection, further illustrating the behavior of UMPOcc across different scenes and anomaly configurations.

F Limitations and Future Work

Limitations. UMPOcc still depends on the quality of the underlying occupancy representation and inherited anomaly score. The current prototype-bank capacity and PPSC neighborhood are fixed

Table 12: Computational overhead of UMPOF/PPSC. FPS and peak inference memory are measured locally on the same shared RTX 4090 environment for each matched pair.

Method	Params (M)	FLOPs (G)	FPS↑	Memory (GiB)↓
Local ProOOD + EchoOOD	28.40	580.66	2.09	4.04
UMPOcc (+SGN)	28.40	580.66	2.08	4.05
VoxDet checkpoint	55.34	1311.10	5.99	1.55
VoxDet + UMPOF/PPSC	55.34	1311.10	5.96	1.79

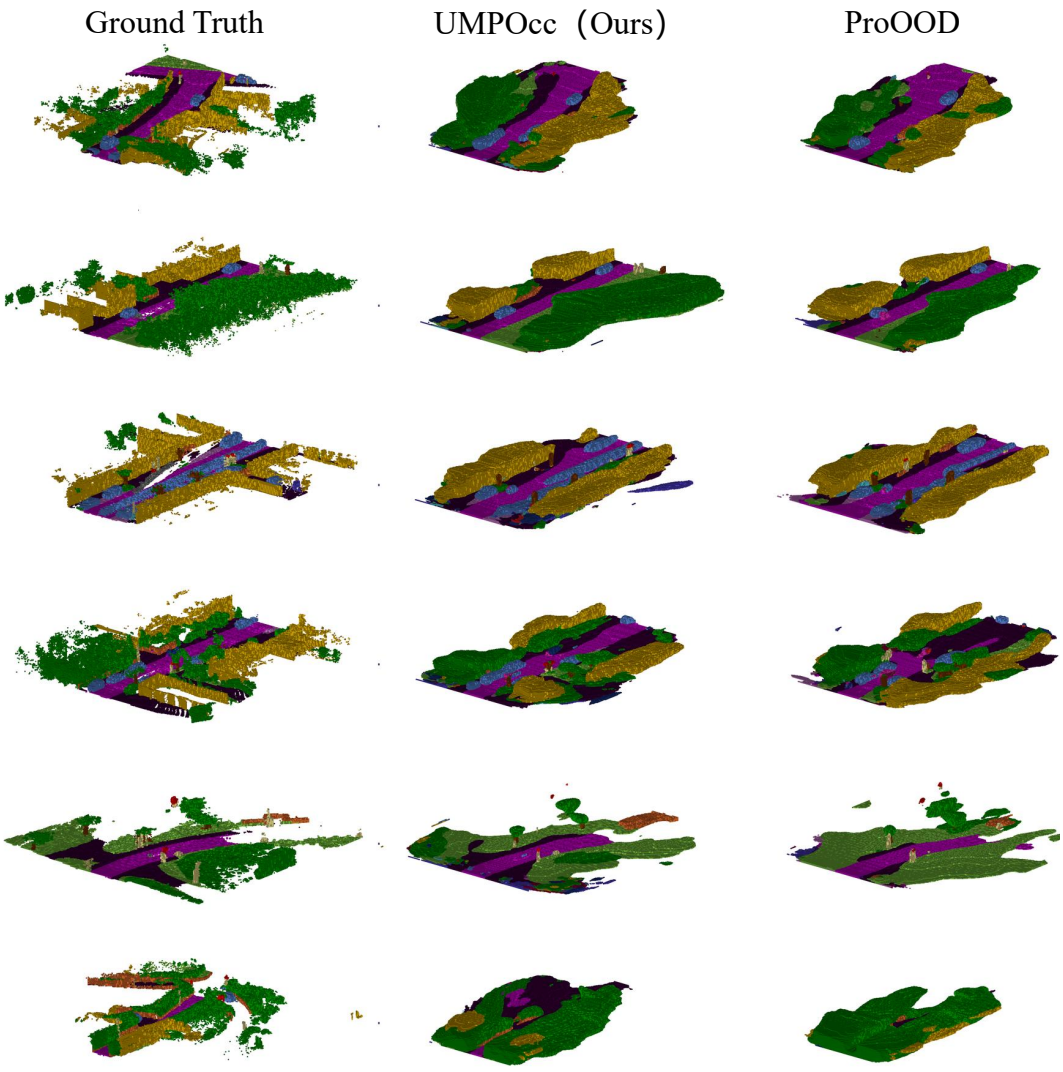

Figure 4: Additional qualitative results of 3D Occupancy.

for each experimental setting, and extremely sparse, small, or spatially ambiguous anomalies may therefore remain challenging. In addition, the current evaluation focuses on single-frame occupancy prediction and a limited set of driving-domain benchmarks.

Future work. Future work can investigate adaptive prototype allocation and scale-aware spatial calibration, allowing the representation and calibration range to adjust automatically to scene complexity. Extending UMPOcc to temporal or multi-frame occupancy prediction, stronger occupancy backbones, and broader open-world settings would also provide a more comprehensive evaluation of its rare-vs-unknown disambiguation capability.

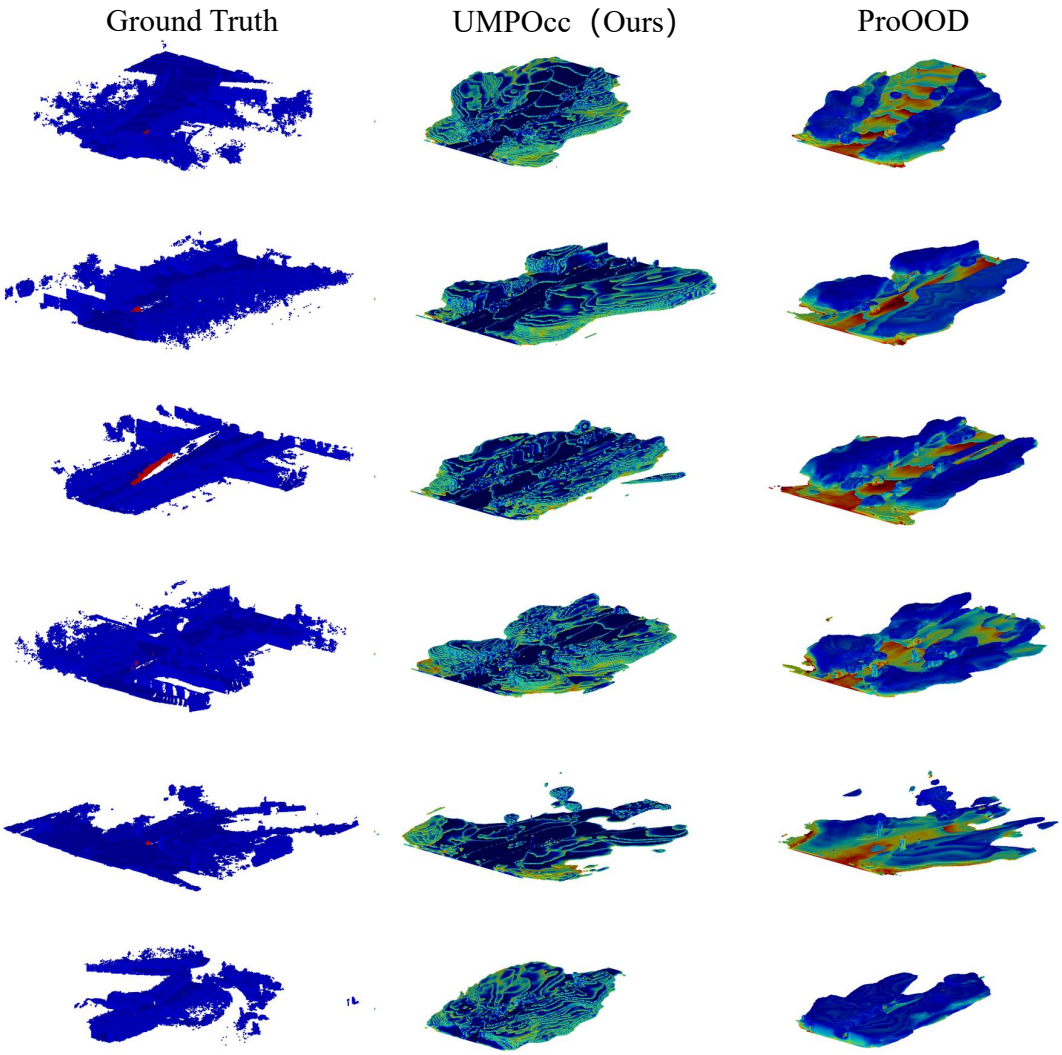

Figure 5: Additional qualitative results of OOD detection.