DUMAP: DIRICHLET EVIDENCE ALLOCATION UNCERTAINTY-GUIDED ONLINE VECTORIZED HD MAPPING UNDER ENVIRONMENTAL AND GEOGRAPHIC DOMAIN SHIFTS

Anonymous authors
Paper under double-blind review

ABSTRACT

Online vectorized HD mapping is important for autonomous driving, but existing methods suffer severe performance degradation under environmental domain shifts such as nighttime, rain, and low visibility. Existing methods mainly model uncertainty in geometric outputs, while the evidence-allocation process remains deterministic. Under distribution shifts, a map query can also be uncertain about which sampled BEV evidence should be used. To address this issue, we propose DUMap to model query-to-BEV evidence-allocation uncertainty before geometric prediction. Dirichlet Evidence Allocation (DEA) models the normalized evidence-allocation weights with a Dirichlet distribution. Its mean determines the relative evidence allocation, and its total concentration measures allocation reliability. Uncertainty-Guided Residual Fusion (UGRF) uses this reliability to combine local geometric predictions with global structural refinement. We also introduce Env-Split on nuScenes by excluding nighttime and rainy samples from training. Experiments show that DUMap achieves state-of-the-art (SOTA) performance under environmental OOD evaluation while maintaining strong results on region-based and city-based splits.

1 INTRODUCTION

Vectorized high-definition (HD) maps represent road geometry and semantics, including lane boundaries, drivable areas, and traffic infrastructure. They support vehicle localization (Ma et al., 2019), motion planning (Zeng et al., 2019), and decision making (Ma et al., 2024). Conventional approaches (Shan et al., 2020; Shan & Englot, 2018) rely on offline databases that require extensive surveying and are difficult to update. Online vectorized HD mapping (Li et al., 2022a) predicts map elements from onboard observations in real time and reduces reliance on prebuilt maps.

Early online mapping approaches (Li et al., 2022a; Zhang et al., 2024; Jia et al., 2024; Hao et al., 2024) predict pixel-level raster maps in bird's-eye-view (BEV) space. However, they require post-processing to cluster pixels and fit lane instances. Direct vectorization methods (Liao et al., 2023; 2025) use Transformer queries to predict sparse point sequences and structured map elements end-to-end. These query-based methods are efficient, but they remain vulnerable under complex environmental conditions. Recent work (Liu et al., 2024a; Shin et al., 2025; Monninger et al., 2026) improves robustness using semantic masks, temporal context, prediction uncertainty, or cross-view geometric consistency. However, generalization to unseen lighting and weather is still limited.

To further illustrate this problem, we examine the performance of classic HD map model MapTRv2 Liao et al. (2025) under different distribution shifts. Figure 1(a) shows that MapTRv2 (Liao et al., 2025) achieves 62.10 mAP on the Original split. Its mAP drops to 41.23 on Rain-only scenes, 23.1 on Night-only scenes, and 21.80 on the City-based split. Rain-only and Night-only represent environmental OOD shifts caused by weather and illumination. City-based evaluation represents a geographic domain shift between different urban layouts. Figure 1(b) also shows that MapTRv2 misses critical map elements in a rainy night scene, while DUMap produces more complete predictions.

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.

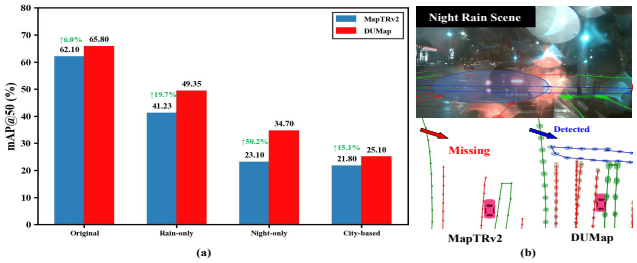

Figure 1: (a) Performance comparison between MapTRv2 and DUMap across driving scenarios. (b) In rainy night scenes, MapTRv2 shows limited generalization and misses critical elements.

Domain shifts make visual evidence unreliable. This leads to two problems. First, existing uncertainty-aware mapping methods mainly model uncertainty in predicted coordinates or map elements. They do not model uncertainty in the attention weights used to select sampled BEV features. Therefore, a map query cannot represent evidence-allocation uncertainty before geometric prediction. Second, existing evaluation protocols do not directly isolate environmental shifts caused by weather and illumination.

To address the modeling problem, we propose DUMap. Specifically, Dirichlet Evidence Allocation (DEA) models the normalized query-to-BEV allocation weights with a Dirichlet distribution because these weights lie on a probability simplex. The Dirichlet mean determines relative evidence allocation, and the total concentration measures allocation reliability. Uncertainty-Guided Residual Fusion (UGRF) uses this reliability to combine query-specific local predictions with global structural refinement.

To address the evaluation problem, we introduce Env-Split on nuScenes. It excludes nighttime and rainy scenes from training and uses them for environmental OOD evaluation. Experiments show that DUMap achieves SOTA performance on Env-Split and maintains strong results on city-based and region-based splits.

In summary, our main contributions are as follows:

- We model evidence-allocation uncertainty in query-based vectorized HD mapping. This allows each map query to estimate the reliability of its sampled BEV evidence before geometric prediction.
- We introduce DEA to model normalized evidence-allocation weights with a Dirichlet distribution. We also introduce UGRF to combine local prediction and global structural refinement using the resulting allocation reliability.
- We introduce Env-Split on nuScenes by excluding nighttime and rainy scenes from training for environmental OOD evaluation. DUMap achieves SOTA performance on Env-Split while maintaining strong results on city-based and region-based splits.

2 Related Work

Online mapping. Advances in 2D-to-BEV transformation technology enable perception systems to project multi-camera images onto a BEV representation (Lu et al., 2025). Many studies leverage this representation to formulate online map construction as a pixel-level semantic segmentation task (Zhang et al., 2024; Jia et al., 2024), generating grid-based outputs that encode road semantics. However, pixel-based representations lack geometric compactness and topological information. Therefore, recent works employ vector maps to provide more structured information for trajectory planning. VectorMapNet (Liu et al., 2023) introduces an end-to-end learnable framework that generates vector points sequentially through an autoregressive decoder. Although this enables direct prediction of structured map elements, it suffers from high inference latency and sensitivity to point ordering. MapTR (Liao et al., 2023) and MapTRv2 (Liao et al., 2025) address these issues with a query-based set prediction paradigm. InstaGraM (Shin et al., 2025) detects points and constructs connections between instances using an adjacency matrix. MGMap (Liu et al., 2024a) improves lane segmentation by using predicted semantic masks as attention priors, but still relies on fully observable

inputs. Performance on the original dataset splits does not directly establish geographic transfer: Lilja et al. (Lilja et al., 2024) show that geographic overlap can inflate online-mapping scores. Our MapTRv2 results in Figure 1(a) further show substantial degradation in rain and at night. Existing uncertainty-aware mapping mainly models uncertainty in coordinates or predictions (Gu et al., 2024). DUMap instead models uncertainty in the allocation of visual evidence.

Evidential learning uses Dirichlet distributions over class probabilities, whereas Bayesian attention places distributions over generic attention weights (Sensoy et al., 2018; Zhang et al., 2021; Cinquin et al., 2021). DUMap instead models deformable map-query allocation over sampled BEV evidence and uses query-level concentration to guide local–global geometric refinement for vectorized mapping under environmental and geographic shifts.

HD map datasets and distribution shifts. The nuScenes (Caesar et al., 2020) and Argoverse 2 (Wilson et al., 2021) datasets are standard benchmarks for online HD map construction. Their original splits suffer from geographic overlap with the training data (Lilja et al., 2024), which can cause models to overfit local map structures and inflate performance. Despite this, both datasets are widely used in online mapping. Methods evaluated on nuScenes include (Can et al., 2021; Dong et al., 2023; Hu et al., 2022; Jiang et al., 2023; Li et al., 2022b;a; Gu et al., 2024), while Argoverse 2 is used in (Liu et al., 2024a; Liao et al., 2023; 2025; Liu et al., 2025).

To reduce geographic leakage, Lilja et al. (Lilja et al., 2024) and StreamMapNet (Yuan et al., 2024) proposed geographically disjoint evaluation splits. These region-based and city-based splits provide a more realistic test of spatial generalization and geographic domain shift. However, they do not explicitly isolate environmental domain shifts. Many sequences captured at night, in rain, or on wet roads remain in the training set, allowing models to observe similar weather and illumination conditions before evaluation. This makes it difficult to measure whether a model generalizes to truly OOD environmental conditions such as low light or precipitation.

3 METHOD

3.1 PROBLEM FORMULATION

Online vectorized HD map generation takes multi-view images from calibrated cameras as input and predicts vectorized map elements in the ego-centric BEV space. Each map element, such as a lane divider, pedestrian crossing, or road boundary, is represented as an ordered 2D point sequence with a semantic category within a local BEV range of $[-15, 15]$ m and $[-30, 30]$ m along the X- and Y-axes.

We focus on environmental domain shifts and OOD conditions, where models are trained on favorable scenes but tested under unseen rain, nighttime, wet roads, or low-visibility conditions. These shifts change image appearance while largely preserving road topology, making robust structural reasoning under unreliable visual evidence the key challenge.

3.2 OVERVIEW

Figure 2 summarizes DUMap. A shared backbone extracts multi-view perspective-view (PV) features and projects them into BEV. In parallel, an auxiliary low-frequency (LF) guidance branch extracts coarse road-layout cues from the PV features and injects them into the BEV representation. The decoder then applies DEA to aggregate sampled BEV features and derive query-level Dirichlet attention uncertainty, after which UGRF combines local coordinate predictions with a global self-attention refinement to produce the vectorized map elements. BEV features are the primary representation for map decoding, but rain, nighttime, and low visibility can weaken coarse road structures during view transformation. The auxiliary LF branch retains complementary layout information in the PV space by adaptively combining the original feature with its smoothed counterpart.

Given a perspective-view feature map $\mathbf{F}$, we predict a structural confidence map $M = \sigma(\mathrm{Conv}(\mathbf{F}))$ and construct the low-frequency representation as

$$\mathbf{F}_{\mathrm{lf}} = M \odot \mathbf{F} + (1 - M) \odot \mathcal{G}(\mathbf{F}), \tag{1}$$

where $\mathcal{G}(\cdot)$ denotes Gaussian blur. The confidence map preserves reliable structural responses and replaces less reliable regions with their low-frequency content. A lightweight decoder predicts coarse

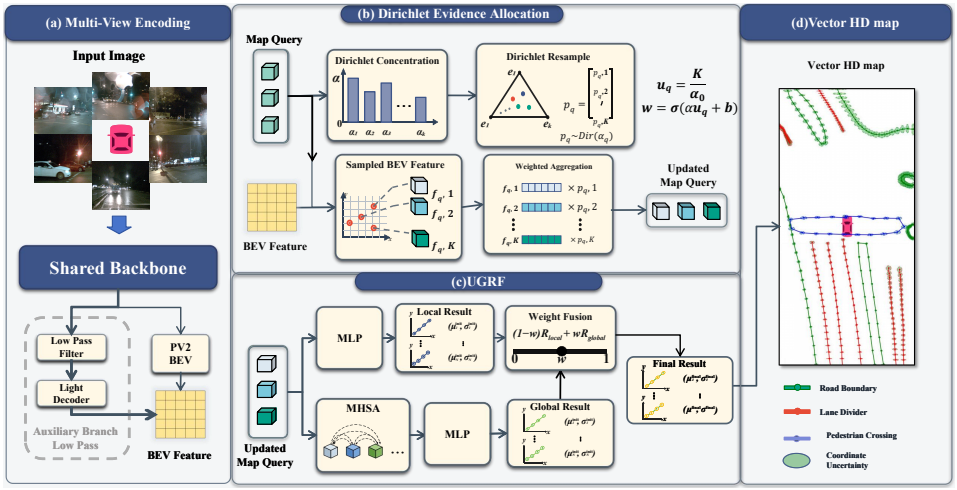

Figure 2: Overview of DUMap. (a) Multi-view images are encoded into BEV space, with a lightweight auxiliary low-pass branch. (b) Dirichlet Evidence Allocation (DEA) performs probabilistic evidence allocation over sampled BEV features and derives query-level Dirichlet attention uncertainty. (c) The Uncertainty-Guided Residual Fusion (UGRF) decoder uses this query-level signal to balance local predictions and global refinement. (d) The outputs form the final vectorized HD map.

PV map structures from $\mathbf{F}_{\text{lf}}$; these structures are projected into BEV and fused with the backbone features through projection-based fusion, providing layout cues for the uncertainty-aware decoder.

3.3 Dirichlet Evidence Allocation

Each map query aggregates K sampled BEV features with normalized non-negative weights on the $(K-1)$-dimensional probability simplex. Conventional deterministic attention (Xia et al., 2022) predicts a single weight vector, whereas DEA models the allocation distribution itself with a Dirichlet distribution. Dirichlet parameterization separates the expected evidence allocation from its concentration, allowing allocation reliability to be characterized in addition to the mean attention pattern.

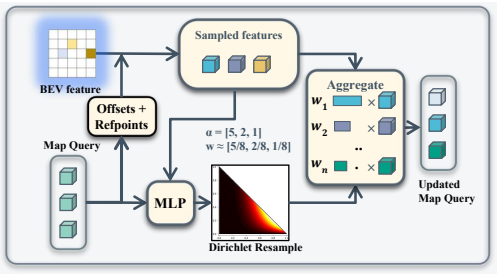

Figure 3: Illustration of DEA with $K = 3$ sampled BEV features. Given a map query, DEA predicts sampling offsets and Dirichlet concentration parameters $\boldsymbol{\alpha}$. Attention weights are sampled from $\mathrm{Dir}(\boldsymbol{\alpha})$ during training and used to aggregate sampled BEV features; the total concentration α_0 provides the query-level uncertainty $u_q = K/\alpha_0$.

For each map query q, DEA predicts K sampling locations $\mathbf{p}_k$ on the BEV feature map and non-negative concentration increments

$$e_k = \mathrm{softplus}(z_k), \qquad \alpha_k = e_k + 1. \tag{2}$$

This gives the concentration vector $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^\top$ with $\alpha_k \geq 1$. The relative magnitudes of α_k determine the expected allocation across the K sampled locations. Specifically, with $\alpha_0 = \sum_{k=1}^{K} \alpha_k$, the expected attention weight of the k-th location is $\mathbb{E}[w_k] = \alpha_k/\alpha_0$; hence, a larger relative α_k assigns more attention to its corresponding BEV support.

During training, the attention vector is sampled as

$$\mathbf{w} \sim \mathrm{Dir}(\boldsymbol{\alpha}), \tag{3}$$

where $\mathbf{w} = [w_1, \ldots, w_K]^\top$, $w_k \geq 0$, and $\sum_{k=1}^{K} w_k = 1$. We draw one differentiable reparameterized sample for each query. At inference, we use the Dirichlet mean

$$w_k = \frac{\alpha_k}{\alpha_0}, \tag{4}$$

to obtain deterministic attention weights.

The resulting weights aggregate the sampled BEV features as

$$\tilde{\mathbf{f}}_q = \sum_{k=1}^{K} w_k \, \mathbf{f}(\mathbf{p}_k), \tag{5}$$

where $\mathbf{f}(\mathbf{p}_k)$ is obtained by bilinear interpolation at the k-th sampling location. Each sampled feature is weighted by its Dirichlet attention and summed to update query q.

The total concentration $\alpha_0 = \sum_{k=1}^{K} \alpha_k$ characterizes the dispersion of the Dirichlet allocation. We define the query-level Dirichlet attention uncertainty as

$$u_q = \frac{K}{\alpha_0}. \tag{6}$$

Since $\alpha_k \geq 1$, we have $\alpha_0 \geq K$ and hence the global bound $0 < u_q \leq 1$. We use u_q as a normalized concentration-based uncertainty score because, for fixed K, it is strictly monotonic with the Dirichlet variance factor, $1/(\alpha_0 + 1) = u_q/(K + u_q)$. We next establish how the Dirichlet concentration governs the stability of BEV evidence aggregation.

Theorem 1 (Dirichlet-Controlled Evidence Aggregation Stability) *For the current query q, let $\mathbf{w} \sim \mathrm{Dir}(\boldsymbol{\alpha})$ and define the expected allocation $\mu_k = \alpha_k/\alpha_0$, the expected aggregated feature*

$$\bar{\mathbf{f}}_q = \sum_{k=1}^{K} \mu_k \mathbf{f}(\mathbf{p}_k), \tag{7}$$

and the sampled-evidence disagreement

$$D_q = \sum_{k=1}^{K} \mu_k \left\| \mathbf{f}(\mathbf{p}_k) - \bar{\mathbf{f}}_q \right\|_2^2. \tag{8}$$

Then the allocation-induced aggregation instability satisfies

$$\mathcal{V}_q := \mathbb{E}\left[\left\| \tilde{\mathbf{f}}_q - \bar{\mathbf{f}}_q \right\|_2^2 \right] = \frac{D_q}{\alpha_0 + 1} = \frac{u_q}{K + u_q} D_q, \tag{9}$$

$$\frac{\partial \mathcal{V}_q}{\partial u_q} = \frac{K D_q}{(K + u_q)^2} \geq 0, \tag{10}$$

For fixed sampled BEV features and fixed expected allocation $\boldsymbol{\mu}$, varying the total concentration along the feasible ray $\boldsymbol{\alpha} = \alpha_0 \boldsymbol{\mu}$ leaves D_q unchanged. The monotonicity holds over $0 < u_q \leq K \min_k \mu_k$ and is strict whenever $D_q > 0$.

Under the theorem's fixed-evidence and fixed-expected-allocation condition, u_q monotonically controls only the concentration-induced component; absolute cross-query instability also depends on D_q. We therefore use u_q as a lightweight concentration-based reliability signal, not a complete cross-query estimator of $\mathcal{V}_q$. Appendix A provides the derivation and a downstream Lipschitz propagation bound.

3.4 Uncertainty-Guided Residual Fusion

Theorem 1 shows that, for fixed sampled evidence and expected allocation, concentration-induced aggregation instability is nondecreasing in u_q; accordingly, we design the UGRF gate to be nondecreasing in u_q, assigning more weight to global refinement as concentration uncertainty increases. UGRF implements the local and refinement paths in parallel. For each query, the two paths predict coordinate means $\boldsymbol{\mu}_{\mathrm{init}}, \boldsymbol{\mu}_{\mathrm{ref}} \in \mathbb{R}^{L \times 2}$ and unconstrained scale logits $\mathbf{s}_{\mathrm{init}}, \mathbf{s}_{\mathrm{ref}} \in \mathbb{R}^{L \times 2}$. The scale logits are mapped to positive Laplace scales by

$$\boldsymbol{\sigma}_a = \mathrm{softplus}(\mathbf{s}_a) + \epsilon, \qquad a \in \{\mathrm{init}, \mathrm{ref}\}, \tag{11}$$

where $\epsilon > 0$ is a small numerical constant. The refinement path applies multi-head self-attention across all map queries before its coordinate head.

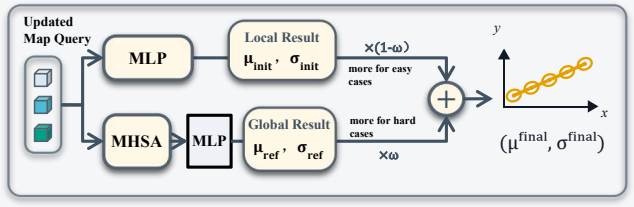

Figure 4: Structure of UGRF. The local path preserves query-specific geometric details, while the refinement path applies multi-head self-attention to capture global structural context. The query-level Dirichlet attention uncertainty from DEA is mapped through a learnable monotonic function to obtain the fusion weight ω_q, which adaptively balances the local and globally refined predictions.

The query-level Dirichlet attention uncertainty is mapped to a query-specific fusion weight through a learnable monotonic function:

$$\omega_q = \sigma(au_q + b), \tag{12}$$

where $a = \mathrm{softplus}(\tilde{a})$ is constrained to be non-negative and b is learnable. This learnable monotonic mapping adapts the scale and offset of the uncertainty signal for downstream fusion. Its derivative is

$$\frac{\partial \omega_q}{\partial u_q} = a\,\sigma(au_q + b)\,[1 - \sigma(au_q + b)] \geq 0. \tag{13}$$

This monotonic constraint implements the theorem-based gate direction: higher concentration uncertainty does not lead to a lower refinement weight.

The scalar ω_q is broadcast over the L points and two coordinate dimensions:

$$\boldsymbol{\mu}_{\text{final}} = (1 - \omega_q)\boldsymbol{\mu}_{\text{init}} + \omega_q \boldsymbol{\mu}_{\text{ref}}, \tag{14}$$

$$\boldsymbol{\sigma}_{\text{final}} = (1 - \omega_q)\boldsymbol{\sigma}_{\text{init}} + \omega_q \boldsymbol{\sigma}_{\text{ref}}. \tag{15}$$

3.5 Training Loss

We follow the bipartite matching and supervised training scheme of MapTR (Liao et al., 2023). Given N map queries and the ground-truth map elements, Hungarian matching establishes a one-to-one assignment using the predicted class scores and the final fused coordinates $\boldsymbol{\mu}_{\text{final}}$. The point-regression cost is the permutation-invariant Smooth L1 distance between predicted and ground-truth point sequences, considering both the original and reversed point orders; the Laplace coordinate scale is not included in the matching cost. Let $\mathcal{M}^*$ denote the resulting set of matched positive pairs (i, j), where prediction i is assigned to ground-truth element j, and let $N_{\text{pos}} = |\mathcal{M}^*|$.

After matching, the model is optimized with three terms: a classification loss, a coordinate regression loss, and a coordinate-uncertainty-aware regression loss:

$$\mathcal{L} = \lambda_{\text{nll}}\mathcal{L}_{\text{nll}} + \lambda_{\text{L1}}\mathcal{L}_{\text{L1}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}}. \tag{16}$$

Here, $\mathcal{L}_{\text{cls}}$ is the focal loss for map element classification, and $\mathcal{L}_{\text{L1}}$ is the Smooth L1 loss for point coordinate regression. To supervise the predicted coordinate uncertainty, we use a Laplace negative log-likelihood:

$$\mathcal{L}_{\text{nll}} = \frac{1}{N_{\text{pos}}} \sum_{(i,j)\in\mathcal{M}^*} \sum_{k=1}^{L} \sum_{d\in\{x,y\}} \left(\log(2\sigma_{\text{final},i,k,d}) + \frac{\left|\mu_{\text{final},i,k,d} - \hat{p}_{j,\pi^*_{ij}(k),d}\right|}{\sigma_{\text{final},i,k,d}} \right), \tag{17}$$

where $k \in \{1, \ldots, L\}$ indexes vector points, $d \in \{x, y\}$ indexes coordinate dimensions, and π^*_{ij} is the selected point ordering for pair (i, j). We average over matched map elements and sum over points and coordinate dimensions; $\sigma_{\text{final},i,k,d}$ is the positive Laplace scale of the corresponding coordinate. Across all experiments, we set $\lambda_{\text{nll}} = 0.05$, $\lambda_{\text{L1}} = 2.5$, and $\lambda_{\text{cls}} = 5.0$.

Table 1: Quantitative results on the nuScenes dataset under Env-Split, City-Based, and Region-Based evaluation protocols. All methods use ResNet-50 and are trained for 24 epochs.

Method	Env-Split				City-Based				Region-Based			
	AP_{ped}	AP_{div}	AP_{bou}	mAP	AP_{ped}	AP_{div}	AP_{bou}	mAP	AP_{ped}	AP_{div}	AP_{bou}	mAP
MapTR (Liao et al., 2023)	7.3	19.2	36.5	21.0	11.1	12.9	21.1	15.0	6.4	20.7	35.5	20.9
MGMap (Liu et al., 2024a)	11.2	25.4	37.8	24.8	10.9	17.8	24.7	17.8	9.4	27.0	38.1	24.8
MapQR (Liu et al., 2024b)	8.4	24.8	38.7	24.0	10.2	15.8	25.5	17.2	8.4	25.7	39.9	24.7
GeMap (Zhang et al., 2025)	11.3	24.9	42.9	26.4	12.9	16.3	26.7	18.6	11.8	26.6	43.6	27.3
MapTRv2 (Liao et al., 2025)	17.1	25.8	43.3	27.7	18.9	18.6	**27.9**	21.8	14.7	28.7	43.3	28.9
StreamMapNet (Yuan et al., 2024)	25.3	36.7	44.5	35.5	18.7	16.4	22.7	19.3	32.2	29.3	40.8	34.1
UIGenMap (Liu et al., 2025)	27.2	34.7	47.2	36.4	23.1	19.6	25.4	22.7	40.3	30.8	46.8	39.3
Ours	**32.6**	**37.4**	**47.7**	**39.2**	**25.4**	**22.2**	27.7	**25.1**	**41.7**	**33.9**	**48.1**	**41.2**

Table 2: Quantitative results on Argoverse 2 under Region-Based and City-Based evaluation protocols. All methods use ResNet-50 and are trained for 24 epochs; FPS is measured on an RTX A40.

Method	Region-Based				City-Based				FPS
	AP_{ped}	AP_{div}	AP_{bou}	mAP	AP_{ped}	AP_{div}	AP_{bou}	mAP	
MapTR (Liao et al., 2023)	48.1	50.4	55.0	51.1	34.7	41.7	34.7	37.9	**15.8**
GeMap (Zhang et al., 2025)	52.5	59.2	58.2	56.6	34.5	41.1	40.4	38.7	11.5
MapTRv2 (Liao et al., 2025)	53.3	58.7	58.8	56.9	37.4	42.2	41.9	40.5	12.9
MapQR (Liu et al., 2024b)	54.2	58.0	60.4	57.5	33.8	41.5	41.6	38.9	12.1
StreamMapNet (Yuan et al., 2024)	57.9	55.7	61.3	58.3	40.2	43.4	43.3	42.3	12.2
UIGenMap (Liu et al., 2025)	59.8	58.7	**62.7**	60.4	43.7	45.7	45.1	44.8	8.2
Ours	**61.2**	**62.4**	61.9	**61.8**	**45.1**	**47.8**	**48.3**	**47.1**	8.7

4 Experiments

4.1 Experimental Setup

We follow the evaluation protocol of StreamMapNet (Yuan et al., 2024). Performance is measured by mAP using bidirectional Chamfer distance over thresholds of 0.5 m, 1.0 m, and 1.5 m. All methods use a ResNet-50 backbone and are trained for 24 epochs under the same optimization setting.

4.2 Dataset

Datasets. We evaluate DUMap on two autonomous-driving benchmarks, nuScenes (Caesar et al., 2020) and Argoverse 2 (Wilson et al., 2021). nuScenes contains 1,000 driving scenes collected in Boston and Singapore with six calibrated surround-view cameras. Argoverse 2 contains 1,000 sequences from six cities. Following common protocols (Yuan et al., 2024), we downsample Argoverse 2 to 2 Hz for fair comparison.

Evaluation splits. We use three complementary evaluation protocols. The City-Based split evaluates transfer across held-out map locations in nuScenes and held-out cities in Argoverse 2 (Liu et al., 2025); the training and validation assignments are listed in Table 8. The Region-Based split separates geographically disjoint regions to evaluate spatial generalization (Lilja et al., 2024). Env-Split evaluates adverse weather and illumination conditions that do not occur in its training partition for this evaluation.

Env-Split. Env-Split is constructed from environmental attributes in the official nuScenes scene descriptions (Caesar et al., 2020). Scenes containing "rain", "night", "evening", or "wet" are assigned to validation and excluded from training. After deduplicating scene identifiers, the split contains 846 unique scenes: 598 favorable-condition training scenes and 248 adverse-condition validation scenes.

5 Quantitative Comparison

nuScenes. Table 1 reports 39.2, 25.1, and 41.2 mAP for DUMap on Env-Split, City-Based, and Region-Based, respectively. These scores exceed the strongest baseline in each setting by 2.8, 2.4, and 1.9 points. The improvement is larger on Env-Split and City-Based than on Region-Based, indicating that the advantage of DUMap becomes more evident under environmental and geographic distribution shifts. DUMap achieves the highest score on eight of nine class-specific metrics; City-Based boundary AP is 0.2 points below MapTRv2. DUMap runs at 8.7 FPS, maintaining practical inference efficiency while improving robustness.

Argoverse 2. On the Region-Based split in Table 2, DUMap achieves 61.8 mAP, outperforming UIGenMap by 1.4 points. It obtains the best pedestrian-crossing and divider AP of 61.2 and 62.4, respectively, while its boundary AP is 0.8 points lower than UIGenMap. On the more challenging City-Based split, DUMap reaches 47.1 mAP, improving over UIGenMap by 2.3 points. The gain increases from 1.4 points on Region-Based to 2.3 points on City-Based, suggesting that DUMap provides greater benefit when the geographic distribution shift becomes more pronounced. Moreover, DUMap achieves the best AP for all three map-element classes on City-Based, indicating that the improvement is consistent across different map structures rather than being concentrated in one category. DUMap also runs at 8.7 FPS, slightly faster than UIGenMap at 8.2 FPS. These results demonstrate consistent improvements across datasets and geographic evaluation protocols.

5.1 Ablation Studies

To evaluate the effect of each component, we conduct an ablation study in Table 3. We adopt the uncertainty-aware MapTRv2 variant of Gu et al. (Gu et al., 2024) as the baseline, including its Laplace coordinate head and NLL loss. Adding DEA increases mAP from 32.6 to 35.3. Adding LF and UGRF further increases mAP to 37.2 and 39.2, respectively. LF gives the largest improvement on road boundaries, while UGRF gives the largest improvement on dividers. Table 4 compares the adaptive gate with fixed fusion weights. The adaptive gate improves mAP from 38.1 to 39.2 over the best fixed-weight setting, mainly through gains on dividers and boundaries.

Table 3: Ablation study of key components in DUMap on the Env-Split validation set.

DEA	LF	UGRF	AP_{ped}	AP_{div}	AP_{bou}	mAP
			28.5	32.1	37.2	32.6
✓			31.0	33.3	41.7	35.3
✓	✓		31.9	33.7	45.9	37.2
✓		✓	32.1	36.9	44.8	37.9
✓	✓	✓	**32.6**	**37.4**	**47.7**	**39.2**

Table 4: Impact of fusion weight strategy in UGRF on nuScenes under Env-Split.

Fusion Strategy	AP_{ped}	AP_{div}	AP_{bou}	mAP
$\omega = 0.0$	31.9	33.7	45.9	37.2
$\omega = 0.3$	33.5	34.0	44.0	37.8
$\omega = 0.5$	**34.8**	34.2	45.2	38.1
$\omega = 0.7$	34.0	34.1	45.0	37.7
$\omega = 1.0$	29.8	31.5	40.0	35.1
Ours	32.6	**37.4**	**47.7**	**39.2**

5.2 Uncertainty Quantification

Table 5 reports correlations between uncertainty and geometric error. We evaluate point-level coordinate uncertainty and query-level DEA uncertainty with Pearson r and Spearman ρ. Coordinate uncertainty is the Laplace scale supervised by negative log-likelihood, and DEA uncertainty is the concentration-based score u_q for each map query's evidence allocation. For a controlled comparison, both the Dirichlet and Gaussian parameterizations use a fixed fusion weight of $\omega = 0.5$. Point-level coordinate uncertainty shows Pearson and Spearman correlations of 0.6355 and 0.7181, respectively. Query-level DEA uncertainty yields correlations of 0.3943 and 0.3828. In comparison, the Gaussian attention alternative obtains lower correlations of 0.2556 and 0.2433, together with a lower mAP of 37.2 versus 38.1 for the fixed-weight Dirichlet setting. The Dirichlet formulation therefore shows a stronger uncertainty–error correlation than the Gaussian alternative.

Table 5: Uncertainty–error correlation on the Env-Split validation set.

Model	Uncertainty signal	Pearson $r \uparrow$	Spearman $\rho \uparrow$	mAP ↑
DUMap ($\omega = 0.5$)	Coordinate uncertainty	**0.6355**	**0.7181**	**38.1**
	DEA uncertainty	**0.3943**	**0.3828**	
Gaussian alternative	Gaussian attention uncertainty	0.2556	0.2433	37.2

6 Qualitative Comparison

Robustness under ambiguous visual evidence.
In Figure 5(a), an intersection lies ahead and to the right under extremely low visibility. MapTRv2 and UiGenMap both miss a major turning branch and produce incomplete road structures. DUMap recovers the missing branch and assigns larger coordinate uncertainty to its ambiguous part. In Figure 5(b), a subtle road fork in the right-front view is nearly invisible because of poor lighting. Neither baseline predicts a coherent bifurcation, while DUMap approximates the split and places

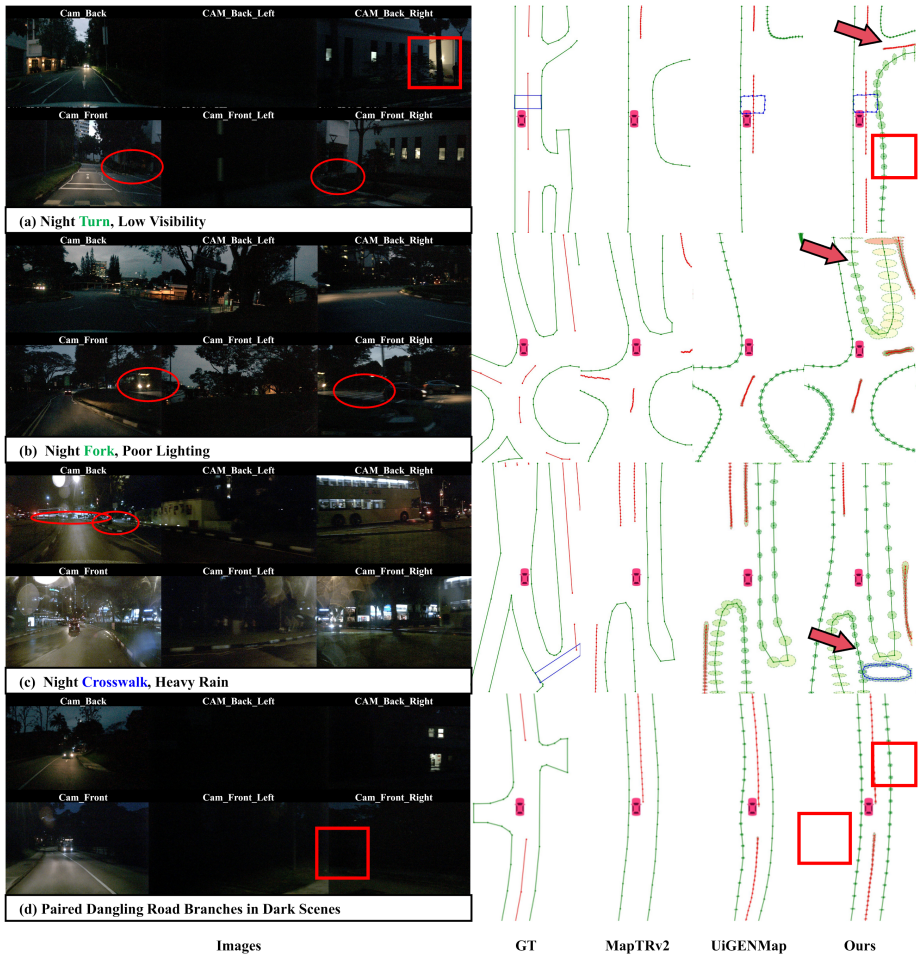

Figure 5: Qualitative comparison under nighttime conditions after Env-Split training. Columns show input images, ground truth, and predictions from MapTRv2 (Liao et al., 2025), UiGenMap (Liu et al., 2025), and DUMap. Red image-space ellipses mark poor visibility, green prediction-space ellipses show point-level Laplace coordinate uncertainty, and red boxes denote failure cases.

larger uncertainty along the uncertain branch. In Figure 5(c), heavy rain and limited sensor coverage obscure a rear-side pedestrian crossing. Both baselines omit this element, whereas DUMap approximates its position with high directional coordinate uncertainty. Across these cases, larger uncertainty ellipses mainly occur around visually weak or geometrically ambiguous predictions. This qualitative behavior is consistent with the uncertainty–error correlations reported in Table 5.

Failure cases. In Figure 5(a), DUMap recovers the rear-right lane boundary but still misses the fully occluded road branch. In Figure 5(d), near-black side views cause all methods to miss the paired dangling road branches. These cases show that DUMap remains limited when the input images provide insufficient evidence for the road geometry.

7 Conclusion

In this paper, we propose DUMap for online vectorized HD mapping under environmental and geographic domain shifts. DEA models query-to-BEV evidence allocation with a Dirichlet distribution and obtains query-level allocation reliability from its total concentration. UGRF uses this reliability to combine local prediction with global structural refinement. Experiments show that DUMap outperforms the evaluated baselines on Env-Split and achieves strong results on the region-based and city-based splits.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and code assistance. Specifically, generative AI was used to improve the clarity, grammar, and presentation of the manuscript, and to assist with code implementation and debugging. All AI-assisted text was carefully reviewed and revised by the authors to ensure technical accuracy and consistency with the intended claims. AI-assisted code was manually inspected, tested, and validated by the authors before being used in the experiments.

Generative AI was not used to generate experimental results, fabricate data, or make autonomous scientific conclusions. The authors take full responsibility for the final content of this work, including all text, claims, experimental results, and code produced with the aid of generative AI.

REFERENCES

Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2020. doi: 10.1109/cvpr42600.2020.01164. URL http://dx.doi.org/10.1109/cvpr42600.2020.01164.

Yigit Baran Can, Alexander Liniger, Danda Pani Paudel, and Luc Van Gool. Structured bird's-eye-view traffic scene understanding from onboard images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15661–15670, 2021.

Tristan Cinquin, Alexander Immer, Max Horn, and Vincent Fortuin. Pathologies in priors and inference for bayesian transformers. *arXiv preprint arXiv:2110.04020*, 2021.

Hao Dong, Xianjing Zhang, Jintao Xu, Rui Ai, Weihao Gu, Huimin Lu, Juho Kannala, and Xieyuanli Chen. Superfusion: Multilevel lidar-camera fusion for long-range hd map generation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2023.

Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, and Boris Ivanovic. Producing and leveraging online map uncertainty in trajectory prediction. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14521–14530, 2024. doi: 10.1109/CVPR52733.2024.01376.

Xiaoshuai Hao, Hui Zhang, Yifan Yang, Yi Zhou, Sangil Jung, Seung-In Park, and ByungIn Yoo. Mbfusion: A new multi-modal bev feature fusion method for hd map construction. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 15922–15928, 2024. doi: 10.1109/ICRA57147.2024.10609873.

Shengchao Hu, Li Chen, Penghao Wu, Hongyang Li, Junchi Yan, and Dacheng Tao. St-p3: End-to-end vision-based autonomous driving via spatial-temporal feature learning. In *European Conference on Computer Vision (ECCV)*, pp. 533–549. Springer, 2022.

Peijin Jia, Tuopu Wen, Ziang Luo, Mengmeng Yang, Kun Jiang, Ziyuan Liu, Xuewei Tang, Zhiquan Lei, Le Cui, Bo Zhang, Kehua Sheng, and Diange Yang. Diffmap: Enhancing map segmentation with map prior using diffusion model. *IEEE Robotics and Automation Letters*, 9(11):9836–9843, 2024. doi: 10.1109/LRA.2024.3455853.

Yanqin Jiang, Li Zhang, Zhenwei Miao, Xiatian Zhu, Jin Gao, Weiming Hu, and Yu-Gang Jiang. Polarformer: Multi-camera 3d object detection with polar transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1042–1050, 2023.

Qi Li, Yue Wang, Yilun Wang, and Hang Zhao. Hdmapnet: An online hd map construction and evaluation framework. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 4628–4634, 2022a. doi: 10.1109/ICRA46639.2022.9812383.

Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Yu Qiao, and Jifeng Dai. BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers. In *European Conference on Computer Vision (ECCV)*, pp. 1–18. Springer, 2022b.

Bencheng Liao, Shaoyu Chen, Xinggang Wang, Tianheng Cheng, Qian Zhang, Wenyu Liu, and Chang Huang. Maptr: Structured modeling and learning for online vectorized hd map construction. In *International Conference on Learning Representations*, 2023.

Bencheng Liao, Shaoyu Chen, Yunchi Zhang, Bo Jiang, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. MAPTRv2: An end-to-end framework for online vectorized hd map construction. *International Journal of Computer Vision (IJCV)*, 133:1352–1374, 2025. doi: 10.1007/s11263-024-02235-z.

Adam Lilja, Junsheng Fu, Erik Stenborg, and Lars Hammarstrand. Localization is all you evaluate: Data leakage in online mapping datasets and how to fix it. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22150–22159, 2024. doi: 10.1109/CVPR52733.2024.02091. URL https://openaccess.thecvf.com/content/CVPR2024/html/Lilja_Localization_Is_All_You_Evaluate_Data_Leakage_in_Online_Mapping_CVPR_2024_paper.html.

Xiaolu Liu, Song Wang, Wentong Li, Ruizi Yang, Junbo Chen, and Jianke Zhu. Mgmap: Mask-guided learning for online vectorized hd map construction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14812–14821, June 2024a.

Xiaolu Liu, Ruizi Yang, Song Wang, Wentong Li, Junbo Chen, and Jianke Zhu. Uncertainty-instructed structure injection for generalizable hd map construction. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22359–22368, 2025. doi: 10.1109/CVPR52734.2025.02083.

Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. VectorMapNet: End-to-end vectorized HD map learning. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22352–22369. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/liu23ax.html.

Zihao Liu, Xiaoyu Zhang, Guangwei Liu, Ji Zhao, and Ningyi Xu. Leveraging enhanced queries of point sets for vectorized map construction. In *European Conference on Computer Vision*, 2024b.

Shu-Wei Lu, Yi-Hsuan Tsai, and Yi-Ting Chen. Toward real-world bev perception: Depth uncertainty estimation via gaussian splatting. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17124–17133, 2025. doi: 10.1109/CVPR52734.2025.01596.

Wei-Chiu Ma, Ignacio Tartavull, Ioan Andrei Bârsan, Shenlong Wang, Min Bai, Gellért Máttyus, Namdar Homayounfar, Shrinidhi Kowshika Lakshmikanth, Andrei Pokrovsky, and Raquel Urtasun. Exploiting sparse semantic HD maps for self-driving vehicle localization. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5304–5311, 2019. doi: 10.1109/IROS40897.2019.8968122. URL https://arxiv.org/abs/1908.03274.

Yuexin Ma, Tai Wang, Xuyang Bai, Huitong Yang, Yuenan Hou, Yaming Wang, Yu Qiao, Ruigang Yang, and Xinge Zhu. Vision-centric bev perception: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10978–10997, 2024. doi: 10.1109/TPAMI.2024.3449912.

Thomas Monninger, Md Zafar Anwar, Stanislaw Antol, Steffen Staab, and Sihao Ding. Augmapnet: Improving spatial latent structure via bev grid augmentation for enhanced vectorized online hd map construction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 8541–8550, 2026.

Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Tixiao Shan and Brendan Englot. Lego-loam: Lightweight and ground-optimized lidar odometry and mapping on variable terrain. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4758–4765, 2018. doi: 10.1109/IROS.2018.8594299.

Tixiao Shan, Brendan Englot, Drew Meyers, Wei Wang, Carlo Ratti, and Daniela Rus. Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5135–5142, 2020. doi: 10.1109/IROS45743.2020.9341176.

Juyeb Shin, Hyeonjun Jeong, Francois Rameau, and Dongsuk Kum. Instagram: Instance-level graph modeling for vectorized hd map learning. *IEEE Transactions on Intelligent Transportation Systems*, 26(2):1889–1899, 2025. doi: 10.1109/TITS.2024.3518537.

Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, Deva Ramanan, Peter Carr, and James Hays. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021. URL https://www.argoverse.org/av2.html.

Zhuofan Xia, Xuran Pan, Shiji Song, Li Erran Li, and Gao Huang. Vision transformer with deformable attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4794–4803, June 2022.

Tianyuan Yuan, Yicheng Liu, Yue Wang, Yilun Wang, and Hang Zhao. Streammapnet: Streaming mapping network for vectorized online hd map construction. In *WACV*, pp. 7356–7365, 2024.

Wenyuan Zeng, Wenjie Luo, Simon Suo, Abbas Sadat, Bin Yang, Sergio Casas, and Raquel Urtasun. End-to-end interpretable neural motion planner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8660–8669, 2019. doi: 10.1109/CVPR.2019.00886. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Zeng_End-To-End_Interpretable_Neural_Motion_Planner_CVPR_2019_paper.html.

Shujian Zhang, Xinjie Fan, Bo Chen, and Mingyuan Zhou. Bayesian attention belief networks. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 12413–12426. PMLR, 2021.

Yiwei Zhang, Jin Gao, Fudong Ge, Guan Luo, Bing Li, Zhaoxiang Zhang, Haibin Ling, and Weiming Hu. Vq-map: bird's-eye-view map layout estimation in tokenized discrete space via vector quantization. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9798331314385.

Zhixin Zhang, Yiyuan Zhang, Xiaohan Ding, Fusheng Jin, and Xiangyu Yue. Online vectorized hd map construction using geometry. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 73–90, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72967-6.

A THEORETICAL ANALYSIS OF DIRICHLET EVIDENCE ALLOCATION

Proof. To characterize the instability introduced by stochastic Dirichlet evidence allocation, we first express the aggregation in a form that makes its dependence on the random allocation weights explicit. Collect the sampled BEV features into the feature matrix

$$\mathbf{F}_q = [\mathbf{f}(\mathbf{p}_1), \ldots, \mathbf{f}(\mathbf{p}_K)]. \tag{18}$$

Using the feature matrix in Eq. (18), the stochastic Dirichlet aggregation becomes

$$\tilde{\mathbf{f}}_q = \mathbf{F}_q \mathbf{w}. \tag{19}$$

Before measuring the variation of the aggregation in Eq. (19), we identify the center around which it fluctuates. For $\mathbf{w} \sim \mathrm{Dir}(\boldsymbol{\alpha})$, with $\alpha_0 = \sum_{k=1}^{K} \alpha_k$ and $\mu_k = \alpha_k/\alpha_0$, the Dirichlet mean is $\mathbb{E}[\mathbf{w}] = \boldsymbol{\mu}$. Combining this mean with Eq. (19) gives

$$\mathbb{E}[\tilde{\mathbf{f}}_q] = \mathbf{F}_q \boldsymbol{\mu} = \bar{\mathbf{f}}_q. \tag{20}$$

Thus the deterministic Dirichlet-mean aggregation used at inference corresponds exactly to the expected stochastic aggregation.

Having established the mean aggregation in Eq. (20), we next characterize how much the stochastic aggregation varies around it. Since Eq. (19) expresses $\tilde{\mathbf{f}}_q$ as a linear transformation of $\mathbf{w}$, its covariance is

$$\mathrm{Cov}(\tilde{\mathbf{f}}_q) = \mathbf{F}_q \mathrm{Cov}(\mathbf{w}) \mathbf{F}_q^\top. \tag{21}$$

At this point, the required Dirichlet covariance identity is

$$\mathrm{Cov}(\mathbf{w}) = \frac{1}{\alpha_0 + 1} \left[\mathrm{Diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}^\top \right]. \tag{22}$$

Substituting Eq. (22) into Eq. (21) gives

$$\mathrm{Cov}(\tilde{\mathbf{f}}_q) = \frac{1}{\alpha_0 + 1} \mathbf{F}_q \left[\mathrm{Diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}^\top \right] \mathbf{F}_q^\top. \tag{23}$$

We now convert the feature covariance in Eq. (23) into the scalar aggregation instability in the theorem. For a random vector $\mathbf{x}$ with finite second moment, $\mathbb{E}[\|\mathbf{x} - \mathbb{E}[\mathbf{x}]\|_2^2] = \mathrm{tr}(\mathrm{Cov}(\mathbf{x}))$. Applying this trace identity to $\tilde{\mathbf{f}}_q$ and using Eq. (23) yields

$$\mathcal{V}_q := \mathbb{E}\Big[\|\tilde{\mathbf{f}}_q - \bar{\mathbf{f}}_q\|_2^2 \Big] = \mathrm{tr}\left(\mathrm{Cov}(\tilde{\mathbf{f}}_q) \right) \tag{24}$$

$$= \frac{1}{\alpha_0 + 1} \left[\sum_{k=1}^{K} \mu_k \|\mathbf{f}(\mathbf{p}_k)\|_2^2 - \|\bar{\mathbf{f}}_q\|_2^2 \right]. \tag{25}$$

The remaining bracket is precisely the weighted disagreement among the sampled BEV features. Indeed, because $\sum_{k=1}^{K} \mu_k = 1$ and $\bar{\mathbf{f}}_q = \sum_{k=1}^{K} \mu_k \mathbf{f}(\mathbf{p}_k)$,

$$\sum_{k=1}^{K} \mu_k \|\mathbf{f}(\mathbf{p}_k) - \bar{\mathbf{f}}_q\|_2^2 = \sum_{k=1}^{K} \mu_k \|\mathbf{f}(\mathbf{p}_k)\|_2^2 - \|\bar{\mathbf{f}}_q\|_2^2. \tag{26}$$

With

$$D_q = \sum_{k=1}^{K} \mu_k \|\mathbf{f}(\mathbf{p}_k) - \bar{\mathbf{f}}_q\|_2^2, \tag{27}$$

Eq. (26) shows that the bracket in Eq. (25) equals the disagreement D_q in Eq. (27). The instability therefore reduces to

$$\boxed{\mathcal{V}_q = \frac{D_q}{\alpha_0 + 1}} \tag{28}$$

To express the concentration-controlled instability in terms of the uncertainty quantity used by DEA, we use

$$u_q = \frac{K}{\alpha_0}, \qquad \alpha_0 = \frac{K}{u_q}, \tag{29}$$

Substituting Eq. (29) into Eq. (28) gives

$$\mathcal{V}_q = \frac{D_q}{K/u_q + 1} = \frac{u_q}{K + u_q} D_q. \tag{30}$$

For fixed sampled BEV features and fixed expected allocation $\boldsymbol{\mu}$, varying the total concentration along $\boldsymbol{\alpha} = \alpha_0 \boldsymbol{\mu}$ leaves D_q unchanged. The constraint $\alpha_k = \mu_k \alpha_0 \geq 1$ restricts this ray to $\alpha_0 \geq 1/\min_k \mu_k$, or equivalently $0 < u_q \leq K \min_k \mu_k$. Differentiating Eq. (30) over this feasible range gives

$$\frac{\partial \mathcal{V}_q}{\partial u_q} = \frac{K D_q}{(K + u_q)^2} \geq 0. \tag{31}$$

The derivative in Eq. (31) is non-negative for $D_q \geq 0$ and strictly positive for $D_q > 0$. Thus, u_q monotonically characterizes the concentration-induced component under fixed evidence and allocation; across queries, absolute aggregation instability also depends on D_q.

Finally, with the feature-level relation in Eq. (28) established, we examine how this instability propagates to the downstream vector-map prediction. Because $\mathbf{w}$ and $\boldsymbol{\mu}$ lie on the probability simplex, both aggregations belong to the convex hull

$$\tilde{\mathbf{f}}_q, \bar{\mathbf{f}}_q \in \mathcal{H}_q := \operatorname{conv}\{\mathbf{f}(\mathbf{p}_1), \ldots, \mathbf{f}(\mathbf{p}_K)\}. \tag{32}$$

Let $\mathcal{C}_q$ denote all other fixed decoder context, and assume the downstream prediction mapping is L_g-Lipschitz on $\mathcal{H}_q$ conditional on $\mathcal{C}_q$:

$$\|g(\mathbf{a}; \mathcal{C}_q) - g(\mathbf{b}; \mathcal{C}_q)\|_2 \leq L_g \|\mathbf{a} - \mathbf{b}\|_2, \qquad \forall\, \mathbf{a}, \mathbf{b} \in \mathcal{H}_q. \tag{33}$$

Because the two aggregations lie in $\mathcal{H}_q$ by Eq. (32), applying Eq. (33) to this pair gives

$$\|g(\tilde{\mathbf{f}}_q; \mathcal{C}_q) - g(\bar{\mathbf{f}}_q; \mathcal{C}_q)\|_2^2 \leq L_g^2 \|\tilde{\mathbf{f}}_q - \bar{\mathbf{f}}_q\|_2^2. \tag{34}$$

Taking expectations in Eq. (34) and substituting the feature-level result in Eq. (28) yields

$$\boxed{\mathbb{E}\left[\|g(\tilde{\mathbf{f}}_q; \mathcal{C}_q) - g(\bar{\mathbf{f}}_q; \mathcal{C}_q)\|_2^2\right] \leq L_g^2 \mathcal{V}_q = L_g^2 \frac{D_q}{\alpha_0 + 1}.} \tag{35}$$

This gives a downstream bound on allocation-induced prediction instability around the deterministic mean-attention prediction. This completes the proof.

B Sensitivity Analysis of Gaussian Kernel Size

The LF branch uses a 7×7 Gaussian kernel by default. Table 6 compares kernel sizes from 3×3 to 9×9 on Env-Split with all other settings fixed.

The 5×5, 7×7, and 9×9 settings differ by at most 0.8 mAP; the 7×7 kernel obtains the highest score of 39.2.

Table 6: Ablation study on the Gaussian kernel size in the LF branch on the Env-Split validation set.

Kernel Size	AP_{ped}	AP_{div}	AP_{bou}	mAP
3×3	30.1	35.8	45.2	37.0
5×5	31.8	36.5	46.8	38.4
7×7	**32.6**	**37.4**	**47.7**	**39.2**
9×9	32.0	37.0	47.1	38.7

C Dataset Partitions

Table 7 lists the training and validation sample counts for the original, region-based, and environment-based partitions. Table 8 gives the map-location and city assignments used by the City-Based protocol Liu et al. (2025). Argoverse 2 is sampled at 2 Hz in all experiments.

After scene-identifier deduplication, Env-Split contains 598 unique training scenes and 248 unique validation scenes, for 846 scenes in total. These scenes provide 25,800 favorable-condition training samples and 8,360 adverse-condition validation samples (Table 9). The partition is based on rain, wet-road, night, and evening metadata.

Table 7: Training and validation sample counts for the dataset partitions.

Dataset	Split	Train. #	Val. #
nuScenes	Ori.	27068	6019
	Region	27846	5981
	Env	**25800**	**8360**
Argoverse2	Ori.	21794	4704
	Region	23434	4676

Table 8: Training and validation assignments for the City-Based partitions. nuScenes entries use the official map identifiers.

Dataset	Split	Map locations / cities	Sample #
nuScenes	Train	boston-seaport; singapore-onenorth	25,926
	Val.	singapore-queenstown; singapore-hollandvillage	8,056
Argoverse 2	Train	Miami; Pittsburgh	21,975
	Val.	Austin; Detroit; Palo Alto; Washington	9,232

Table 9: Data collections on environment-based partition (Env-Split).

Split	Unique scenes	Samples
Train	598	25,800
Val	248	8,360
Total	846	34,160

C.1 Validity Checks for Env-Split

Table 10 compares traffic density and map statistics between the 598 unique training scenes and 248 unique validation scenes. The average pedestrian count changes from 7.68 to 3.73 per sample, while vehicle counts and total map elements differ by 0.18 and 0.08 per sample, respectively.

Table 10: Validity checks for Env-Split. Differences in proportions are reported in percentage points (pp).

(a) Scene-level covariates

Statistic	Train	Adverse val.	Difference
Vehicles per sample	18.62	18.79	+0.18
Pedestrians per sample	7.68	3.73	-3.95
Map elements per sample	9.58	9.50	-0.08

(b) Map-element class composition

Map-element class	Train (%)	Adverse val. (%)	Difference
Pedestrian crossing	14.62	14.51	-0.11 pp
Divider	49.19	51.57	+2.38 pp
Boundary	36.19	33.92	-2.27 pp

(c) Composition of the adverse validation set

Adverse condition	Share (%)
Night only	33.40
Rain/wet only	60.19
Night and rain/wet	6.41

The largest class-composition difference is 2.38 percentage points. The adverse validation scenes comprise 33.40% night only, 60.19% rain or wet road only, and 6.41% combined night and rain or wet road.

D Uncertainty Diagnostics and Reliability Analysis

We analyze two distinct uncertainty signals on 89,023 matched predictions from Env-Split. Prediction error is the point-wise Chamfer distance to the matched ground truth. Query-level Dirichlet attention uncertainty u_q is derived from the total concentration of DEA, while point-level coordinate uncertainty is the Laplace scale supervised by the negative log-likelihood and averaged over vector points and coordinate dimensions.

D.1 Quantitative Uncertainty Diagnostics

Table 11 groups matched predictions by geometric error. Mean coordinate uncertainty increases monotonically from 0.1161 in the lowest-error interval to 0.5875 in the highest-error interval. At the individual prediction level, it has Pearson correlation $r = 0.6355$ and Spearman rank correlation $\rho = 0.7181$ with Chamfer error (Table 12). Query-level attention-uncertainty results for the Gaussian and Dirichlet parameterizations appear in Table 5.

Table 11: Coordinate uncertainty across Chamfer-error intervals on Env-Split.

Error interval (m)	Matched predictions	Mean error (m)	Mean coordinate uncertainty
$[0.0, 0.1)$	27,738	0.0633	0.1161
$[0.1, 0.3)$	38,655	0.1792	0.2067
$[0.3, 0.5)$	12,804	0.3849	0.3570
$[0.5, 0.7)$	5,449	0.5880	0.4836
$[0.7, 1.0)$	4,377	0.8365	0.5875

Table 12: Sample-level association between point-level coordinate uncertainty and Chamfer error on Env-Split.

Uncertainty signal	Pearson $r \uparrow$	Spearman $\rho \uparrow$	Matched predictions
Point-level coordinate uncertainty	0.6355	0.7181	89,023

We define failures as the top 20% of matched predictions by Chamfer error, with a threshold of 0.35 m. Coordinate uncertainty obtains 0.8797 AUROC and 0.6673 AUPR for failure detection (Table 13). Removing predictions from high to low coordinate uncertainty gives 0.1269 AURC and 0.0295 AUSE.

Table 13: Failure-warning and selective-risk results for point-level coordinate uncertainty on 89,023 matched predictions.

AUROC ↑	AUPR ↑	AURC ↓	AUSE ↓	Positive / negative
0.8797	0.6673	0.1269	0.0295	17,805 / 71,218

D.2 Application-Relevant Omission Analysis

We also report class-wise recall after one-to-one Chamfer matching. Unlike AP, this measure counts whether each ground-truth element is recovered without confidence ranking. The baseline is the uncertainty-aware MapTRv2 variant without DEA, LF, or UGRF from Table 3.

DUMap improves recall for every class and condition in Table 14. The largest change is for nighttime pedestrian crossings: recall increases by 8.08 percentage points and false negatives fall from 695 to

582. Nighttime divider and boundary recall increase by 1.37 and 2.15 points; under combined night and rain, the gains are 3.25 and 2.54 points.

Table 14: Chamfer-matched omission analysis under adverse conditions. Recall and its change are reported in percentage points (pp).

Condition	Class	Baseline recall	DUMap recall	Change	False negatives
Night	Pedestrian crossing	50.32	58.40	+8.08	695 → 582
Night	Divider	76.15	77.52	+1.37	2,117 → 1,995
Night	Boundary	74.27	76.42	+2.15	2,440 → 2,236
Night + rain	Pedestrian crossing	50.98	52.21	+1.23	200 → 195
Night + rain	Divider	72.94	76.19	+3.25	816 → 718
Night + rain	Boundary	78.02	80.56	+2.54	476 → 421

Directional coordinate uncertainty. Figure 6 shows camera-view projections and the corresponding BEV prediction for a low-light scene. Green curves are predicted map elements; ellipses are the coordinate scales predicted for their points.

For each predicted map point, DUMap predicts not only its location $p_k = (x_k, y_k)$, but also its directional coordinate uncertainty $\sigma_k = (\sigma_{k,x}, \sigma_{k,y})$. We visualize this uncertainty as an ellipse centered at the predicted point:

$$\mathcal{E}_k = \left\{ z \in \mathbb{R}^2 \;\middle|\; \frac{(z_x - x_k)^2}{(r\sigma_{k,x})^2} + \frac{(z_y - y_k)^2}{(r\sigma_{k,y})^2} \le 1 \right\},$$

where r is shared across all points; the ellipse semi-axes are $r\sigma_{k,x}$ and $r\sigma_{k,y}$.

The blue arrow marks a right-side boundary that curves into a dark, partially occluded area beside the building. Its predicted scales are approximately $\sigma_x = 0.14$–0.19 and $\sigma_y = 0.17$–0.18, compared with 0.07–0.10 in clearer regions.

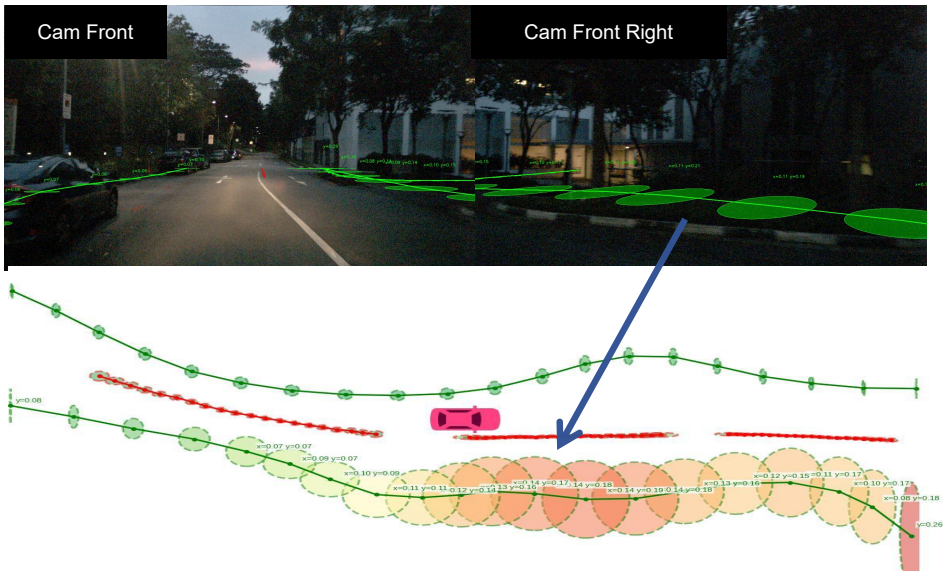

Figure 6: Directional coordinate uncertainty in a low-light scene. The upper panels project predictions into two camera views; the lower panel shows the BEV map. Ellipse axes are proportional to the predicted x- and y-coordinate scales. The blue arrow marks a dark, partially occluded boundary segment.