ReVMap: Vectorized Global Mapping via Connectivity-Aware Local Map Fusion

Hongjae Shin* **Jungho Kim*** **Donghyuk Kwak** **Myeongjun Kim**
Seunghoon Yu **Jiyong Oh** **Jun Won Choi†**

Seoul National University

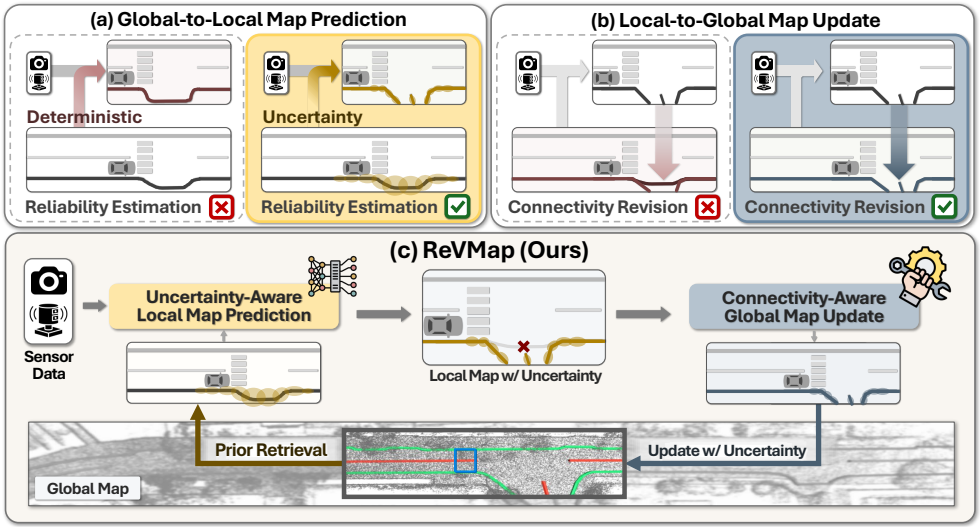


Figure 1: **Comparison of Map Update Paradigms and the ReVMap Framework.** (a) Reliability-agnostic prior reuse leads to over-trusting unreliable accumulated elements, whereas uncertainty-aware prediction modulates their influence. (b) Heuristic matching can lead to incorrect map connections, whereas connectivity-guided revision enables more robust integration of local predictions into the global map. (c) ReVMap iteratively constructs and refines the global map through uncertainty-aware local prediction and connectivity-guided global map revision.

Abstract

Large-scale annotation of vectorized high-definition (HD) maps is labor-intensive and costly, motivating local-to-global auto-labeling approaches that incrementally construct global maps from local vectorized predictions. However, existing approaches often overlook prediction uncertainty and rely on heuristic map integration, potentially propagating errors and introducing topological inconsistencies. To address these limitations, we propose ReVMap, a reliable local-to-global framework that iteratively refines local map predictions and global map construction through uncertainty-guided prior utilization and connectivity-guided global map revision. Specifically, ReVMap leverages point-wise localization uncertainty to incorporate global map priors into local map prediction, improving robustness against inaccurate prior information. It then predicts local-to-global connectivity and selectively updates the global map through uncertainty-aware fusion based on the reliability of local predictions. By iteratively coupling local prediction and

*Equal contribution.
†Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

global revision, ReVMap progressively improves both local map accuracy and global map consistency. Extensive experiments on the nuScenes and Argoverse 2 benchmarks demonstrate that ReVMap achieves state-of-the-art performance in both local map prediction and global map construction.

1 Introduction

HD maps are essential for safe autonomous driving, providing high-precision representations of lane boundaries, road edges, and road-element topology. Structured geometric and topological context from HD maps is widely used in downstream motion forecasting [6, 17] and trajectory planning [12–14]. A typical HD map construction pipeline consists of three stages: 1) global scene reconstruction from collected sensor data, 2) vectorized annotation of road elements, and 3) quality verification before deployment. Within this pipeline, the vectorized annotation stage is particularly labor-intensive, as accurate labels require precise map-element geometry and consistent topology across road structures. Recent HD map auto-labeling pipelines [3, 4, 33, 46] reduce manual labeling effort by producing vectorized HD map elements using learning-based map construction methods [7, 8, 18, 19, 26, 30], requiring human annotators only for minor corrections and quality verification.

For HD map auto-labeling in global coordinates, direct vector map generation over large areas is impractical due to high computational costs and the difficulty of optimizing long-range geometric and topological consistency. Previous works [3, 4, 31, 38, 39, 46] therefore adopt a local-to-global formulation, decomposing the task into local vector map generation and global map integration. This decomposition improves scalability by focusing each prediction on a local observation range, thereby reducing computational and memory demands while preserving fine-grained map geometry. During global map integration, newly predicted local elements either replace outdated global elements corresponding to the same map structures or are added as new elements in regions not yet covered by the global map.

Another line of research [31, 40, 44, 47] explores the use of global map priors to improve local map prediction. Unlike offboard auto-labeling, where the global map primarily serves as the final supervisory target, these methods leverage the accumulated global map as prior information to guide subsequent local map estimation. Such priors provide valuable geometric and topological information that may be difficult to infer from current sensor observations alone, particularly in scenarios involving limited perception range, occlusions, or ambiguous road structures.

However, these accumulated map priors may also contain errors from earlier local predictions. Existing methods [3, 31, 44, 47] that leverage these priors often treat accumulated map elements as deterministic without explicitly accounting for their reliability. Inaccurate global map elements may thus be over-trusted, and their errors may propagate into subsequent local predictions (see Figure 1(a)).

A related challenge arises when integrating newly predicted local elements into the accumulated global map. Existing approaches [3, 31] rely on heuristic rules, such as distance-based matching, to associate local predictions with existing global elements, which can lead to incorrect associations in the presence of prediction errors. They then update only the overlapping portions of matched global elements, leaving the remaining geometry unchanged. Consequently, incorrect associations may introduce topological inconsistencies, while partial updates may leave existing structural errors unresolved (see Figure 1(b)).

To address these challenges, we introduce ReVMap, a local-to-global framework for reliable vectorized global map construction, which can be used for HD map auto-label generation. As illustrated in Figure 1 (c), ReVMap iteratively couples uncertainty-guided prior utilization with connectivity-guided global map revision. The framework comprises two core components: Uncertainty-aware Local Map Generation (ULG) and Global Map Revision Network (GMR).

ULG leverages the accumulated global map as a reliability-aware prior to guide local map prediction. It also estimates point-wise localization uncertainty for newly predicted local map elements, which is propagated to the global map through GMR. This enables the accumulated global map to maintain uncertainty-aware representations for subsequent local predictions.

GMR updates the accumulated global map using local predictions from ULG. For local-to-global association, GMR predicts connectivity between current local instances and surrounding global segments outside the local prediction region, rather than relying on heuristic matching rules. The predicted connectivity guides map revision, while uncertainty-based filtering and fusion reduce the influence of unreliable local predictions on the updated geometry. The updated global map and its uncertainty are then reused by ULG in subsequent iterations, establishing a continuous feedback loop between local map prediction and global map refinement.

We evaluate ReVMap on large-scale HD map construction benchmarks, including nuScenes and Argoverse2. ReVMap achieves substantial performance gains over previous state-of-the-art methods on both benchmarks. On nuScenes, ReVMap establishes a new state of the art in both global map construction and local prediction, achieving 36.1% mGAP, 43.4% miECM, and 58.0% local mAP. On Argoverse2, ReVMap achieves 62.1% mGAP, 62.5% miECM, and 77.0% local mAP, establishing new state-of-the-art performance across both global map construction and local prediction.

- We propose ReVMap, a local-to-global mapping framework that constructs and progressively refines a global map through uncertainty-guided prior utilization and connectivity-guided map revision.
- We introduce ULG, which predicts vectorized local map elements and localization uncertainty to construct reliability-aware global map priors for subsequent local prediction.
- We present GMR that refines the global map through learned local-to-global connectivity estimation and uncertainty-aware map fusion.
- ReVMap achieves state-of-the-art performance in both local map prediction and global map construction on the nuScenes and Argoverse 2 benchmarks.

2 Related Work

Online Local Map Construction. Online local map construction [7, 8, 16, 18, 19, 21, 23, 26, 30, 32] predicts vectorized map elements around the ego vehicle from onboard sensor observations. Query-based approaches improve local map construction with structured map-element representations. MapTR [18] introduces permutation-equivalent point sets, and MapTRv2 [19] strengthens supervision and matching. Mask2Map [7] further incorporates BEV masks for instance-level semantics and point-level geometry. Since local prediction from current observations remains sensitive to limited perception range, occlusion, and temporal inconsistency, recent work incorporates accumulated map context into local prediction. NMP [40] utilizes global BEV features as a neural prior, HRMapNet [47] builds historical rasterized maps from previous predictions, and GlobalMapNet [31] constructs an online global vector map for subsequent local prediction.

Vectorized Global Map Construction. Vectorized global map auto-annotation and large-scale map generation [3, 10, 31, 33, 38, 45, 49] aim to produce spatially consistent vector annotations over large regions, often by integrating local predictions into accumulated global maps. MapTracker [2] and IC-Mapper [49] propagate element identities through tracking-based association, but their single-sequence update logic lacks mechanisms for existing-map revision or multi-trip integration. GlobalMapNet [31] and VMA [3] instead accumulate local predictions through geometric association, where GlobalMapNet adopts Chamfer-distance-based one-to-one matching and VMA adds endpoint-distance constraints. These hand-crafted criteria limit flexible matching under branching or fragmented structures, and both methods update only overlapping portions of matched elements, leaving inconsistent non-overlapping geometry unchanged after accumulation. CAMA and CAMAv2 [4, 46] improve the consistency and accuracy of vectorized map auto-annotation from multi-camera data, but their VMA-style global update process inherits similar rule-based association and overlap-only update limitations.

Uncertainty Estimation in HD Map Construction. Semantic confidence quantifies the model's certainty in the predicted class of a map element but does not directly capture its localization accuracy. Explicit assessment of localization reliability is particularly important under occlusion and ambiguous road structures, motivating uncertainty estimation for map geometry. Prior work [11] leverages point-wise uncertainty to reduce the influence of unreliable map geometry on downstream trajectory prediction, and UIGenMap [22] introduces uncertainty-guided structure injection to improve the generalization of local map prediction. However, these approaches do not propagate uncertainty

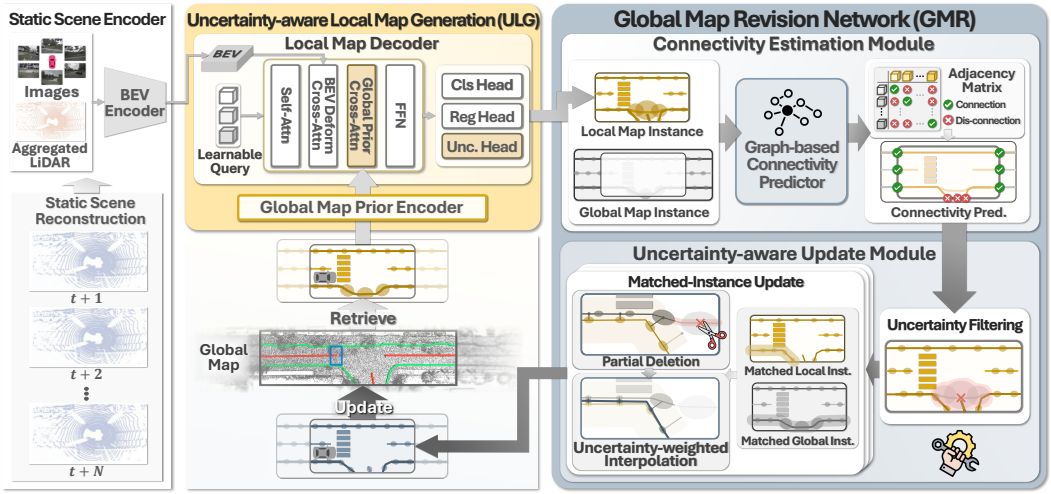


Figure 2: **Overall structure of ReVMap.** ReVMap consists of three main components: Static Scene Encoder, ULG, and GMR. The Static Scene Encoder generates a BEV feature map from multi-view images and, in the offboard setting, from a static 3D scene reconstructed from multi-timestep LiDAR sweeps. ULG retrieves global map priors and predicts local vectorized map instances with point-wise uncertainty. GMR then estimates local-to-global connectivity and updates the accumulated global map via uncertainty-aware filtering and fusion.

estimates into accumulated global maps. MV-Map [39] employs an uncertainty network to estimate the per-frame prediction reliability for map aggregation, while Kalman-filter-based HD map updating [24] incorporates uncertainty into the fusion of observations. However, these methods do not maintain point-wise localization uncertainty in an accumulated vectorized global map for reuse in subsequent local prediction.

3 Method

3.1 Overview

Each vectorized map instance consists of a semantic class and an ordered sequence of 2D BEV points, where the classes are lane dividers, pedestrian crossings, and road boundaries. Local instances use a fixed number of points, whereas accumulated global instances may contain a variable number of points after iterative merging and revision. As illustrated in Figure 2, ReVMap consists of three components, Static Scene Encoder, ULG, and GMR, and operates iteratively over local prediction regions. At each iteration, (1) the Static Scene Encoder encodes the reconstructed static scene and multi-camera images into BEV features. (2) ULG uses these features and the accumulated global map as a prior to predict local vectorized map instances with point-wise localization uncertainty. (3) GMR estimates local-to-global connectivity and revises the accumulated global map through uncertainty-aware updates. The updated global map and its uncertainty are reused as the prior for the next iteration.

3.2 Static Scene Encoder

Static Scene Reconstruction. A single LiDAR sweep provides sparse and partially occluded observations of the road layout, which is insufficient for reliable vectorized global map construction. We therefore reconstruct a dense static scene by aggregating multi-timestep LiDAR sweeps in a common global coordinate frame. For each sweep, points belonging to dynamic objects are removed using per-frame GT 3D boxes before aggregation, preventing movable objects from being fused into the static map representation. The filtered sweeps are merged into a scene-level static point cloud, which provides stable geometric evidence for BEV feature extraction and vectorized map prediction.

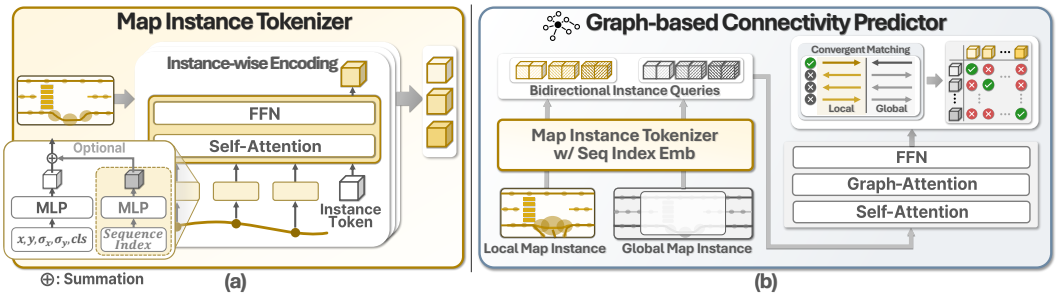


Figure 3: **Details of Map Instance Tokenizer and Graph-based Connectivity Predictor.** (a) The Map Instance Tokenizer aggregates per-point embeddings into a single instance token via self-attention. The sequence-index positional embedding is optional and applied only in the Graph-based Connectivity Predictor. (b) The Graph-based Connectivity Predictor first generates bidirectional instance queries from local and global instances using sequence-index positional embeddings, and then predicts their connectivity through self-attention, graph attention, and convergent matching.

BEV Feature Extractor. The BEV feature extractor produces a unified static-scene BEV representation from the reconstructed static point cloud and multi-camera images. We encode the static point cloud with a SECOND encoder [42] and transform image features into BEV space using LSS [29]. Since dynamic objects are removed only during LiDAR-based reconstruction, we exclude regions of dynamic objects during LSS pooling by identifying them using pretrained Mask2Former [5]. Following BEVFusion [25], the resulting camera BEV features are concatenated with the LiDAR BEV features and fused through a convolutional layer to obtain the final BEV feature map.

3.3 Uncertainty-aware Local Map Generation (ULG)

ULG consists of the Global Map Prior Encoder and Local Map Decoder. The Global Map Prior Encoder extracts reliability-aware prior embeddings from previously accumulated global map elements. The Local Map Decoder leverages global prior embeddings and the BEV feature map to predict local map instances with point-wise uncertainty.

Global Map Prior Encoder. The Global Map Prior Encoder tokenizes accumulated global map instances to provide global context to the Local Map Decoder. We apply the Map Instance Tokenizer illustrated in Figure 3 (a), which is inspired by the CLS-token-based aggregation mechanism [9]. Accumulated global instances are represented as vector geometries with varying numbers of points, since they are formed by merging multiple local predictions. Before tokenization, each global instance is cropped to the local prediction region and uniformly resampled into K points along its geometry, matching the fixed K-point format of local map instances. Each resampled point is first projected into a point embedding from its coordinate, class, and point-wise uncertainty using an MLP, without adding point-order positional embeddings. For each instance, a learnable instance token is concatenated to the point embeddings, and the combined tokens are processed by self-attention and FFN. The final instance token is used as the global prior embedding for the Local Map Decoder.

Local Map Decoder with Uncertainty Estimation. The Local Map Decoder predicts the class, point coordinates, and point-wise localization uncertainty for each local map instance utilizing global prior embeddings. Following MapTRv2 [19], learnable map queries are decoded through multiple layers. In each decoder layer, the map queries exchange information through self-attention and aggregate local BEV features via deformable attention. The updated queries then interact with the global prior embeddings through cross-attention within the same layer, incorporating accumulated map context. Since the global prior embeddings encode accumulated uncertainty with geometry and class, this cross-attention enables the decoder to selectively leverage reliable priors when refining local predictions.

The resulting prior-enhanced queries are fed into classification and regression heads. For localization, the regression head parameterizes each point as a Gaussian distribution [11, 15, 22], where the mean $\mathbf{p} \in \mathbb{R}^2$ represents the point coordinate and the standard deviation $\boldsymbol{\sigma} \in \mathbb{R}^2$ represents its localization uncertainty. The predicted uncertainty is propagated to the global map and later used for uncertainty-aware integration and prior reuse.

3.4 Global Map Revision Network (GMR)

GMR updates the accumulated global map within the current local prediction region while preserving geometry outside it. It first estimates connectivity between current local instances and surrounding global segments outside this region, and then updates the global geometry using point-wise uncertainty.

Connectivity Estimation Module. To prevent potentially erroneous geometry within the local prediction region from affecting connectivity estimation, we extract candidate segments from accumulated global instances within a fixed-margin band immediately outside the region. The Connectivity Estimation Module then predicts connections between these global segments and the current local instances.

As illustrated in Figure 3 (b), the Graph-based Connectivity Predictor estimates connectivity based on instance orientation and distance. Since the same vector geometry can be traversed in two opposite directions, a single order-specific representation cannot capture all possible directed connections. We therefore encode each instance in both its original and reversed point orders using the Map Instance Tokenizer with sequence-index embeddings, producing two direction-aware instance tokens for each instance. Specifically, each point in a map instance is first projected from its coordinate, class, and point-wise uncertainty into a point embedding through an MLP. The sequence index is then embedded with another MLP and added to the point embeddings using the original and reversed point orders. The tokenizer outputs one directional instance token from each order, allowing the module to distinguish the two possible connection directions of each instance.

To explicitly model inter-instance distance, we update these instance tokens using self-attention and graph attention. Following [37, 48], we construct the graph with edges connecting nearby instance pairs, and define the edge feature as the minimum pairwise point distance between two instances. We predict a connectivity score for each directed local–global token pair and use these scores to determine local-to-global connections. For connectivity supervision, we adopt convergent matching, assigning a positive label only to the direction in which the two endpoints face each other. All unmatched pairs and non-converging directions are assigned negative labels.

Uncertainty-aware Map Update Module. To prevent noisy geometry from being accumulated into the global map, the Uncertainty-aware Map Update first filters local instances using point-wise uncertainty. A local instance is discarded when its mean point-wise uncertainty from ULG is above an uncertainty threshold. For each remaining local instance, we restrict the update of its matched global instance to the local prediction region. Within this region, we delete global points whose nearest distance to the local instance is larger than a distance threshold, eliminating accumulated geometry that is inconsistent with the current local geometry.

The remaining global geometry is then fused with the local instance through point-wise uncertainty-weighted interpolation. Since interpolation is performed point-wise, we first establish correspondences between local and global points. For each local point, we construct a normal line passing through the point and perpendicular to the local tangent direction, and take its intersection with the remaining global geometry as the corresponding global point. Since point-wise uncertainty is parameterized as the Gaussian standard deviation, interpolation weights are defined by inverse variance, $w = 1/(\sigma^2 + \epsilon)$, assigning larger weights to points with lower uncertainty. The corresponding local and global points are then interpolated with the normalized weights to update the global geometry.

3.5 Training Loss

The total loss $\mathcal{L}$ used to train ReVMap is given by

$$\mathcal{L} = \mathcal{L}_{\text{nll}} + \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{con}} + \mathcal{L}_{\text{aux}}, \tag{1}$$

where $\mathcal{L}_{\text{nll}}$ and $\mathcal{L}_{\text{cls}}$ supervise the local map predictions of ULG, $\mathcal{L}_{\text{con}}$ supervises the connectivity predictions of GMR, and $\mathcal{L}_{\text{aux}}$ provides auxiliary depth and segmentation supervision following [19]. The negative log-likelihood loss $\mathcal{L}_{\text{nll}}$ supervises the point-wise uncertainty by modeling each predicted point as a bivariate Gaussian with mean $\hat{\mathbf{p}}_i$ and standard deviation $\boldsymbol{\sigma}_i$:

$$\mathcal{L}_{\text{nll}} = \frac{1}{N_p} \sum_{i=1}^{N_p} \sum_{k \in \{x,y\}} \left(\log \sigma_i^k + \frac{(p_i^k - \hat{p}_i^k)^2}{2(\sigma_i^k)^2} \right), \tag{2}$$

Table 1: Comparison with SOTA methods on the nuScenes new split validation set. "C" and "C+L" denote camera-only onboard setting and multi-modal offboard setting, respectively. The "-" indicates that the method does not build a scene-level global map and is included for comparison with local prediction.

			GAP Metrics				iECM			
Method	Modality	mAP	*div*	*ped*	*bou*	*mean*	*div*	*ped*	*bou*	*mean*
Mask2Map[7]	C	29.4	-	-	-	-	-	-	-	-
HRMapNet[47]	C	37.1	-	-	-	-	-	-	-	-
UniPrevPredMap[28]	C	39.7	-	-	-	-	-	-	-	-
StreamMapNet[43]	C	32.9	13.4	15.5	12.8	13.9	32.8	16.3	40.8	30.0
GlobalMapNet[31]	C	34.8	16.3	18.5	18.0	17.6	36.0	18.8	42.4	32.4
MapTracker[2]	C	40.3	18.1	36.1	19.4	24.5	38.7	27.4	45.2	37.1
ReVMap (Ours)	C	**45.6**	18.1	**37.9**	**29.6**	**28.5**	**42.3**	**27.7**	**50.9**	**40.3**
ReVMap (Ours)	C + L	**58.0**	**18.2**	**46.8**	**43.2**	**36.1**	**46.1**	**28.6**	**55.6**	**43.4**

Table 2: Comparison with global map construction SOTA methods on Argoverse2 new split validation set. "C" and "C+L" denote camera-only onboard setting and multi-modal offboard setting, respectively.

			GAP Metrics				iECM			
Method	Modality	mAP	*div*	*ped*	*bou*	*mean*	*div*	*ped*	*bou*	*mean*
GlobalMapNet[31]	C	60.3	34.5	33.7	38.8	35.6	55.5	43.2	58.2	52.3
MapTracker[2]	C	71.3	29.3	52.8	45.0	42.4	46.9	50.7	63.1	53.6
ReVMap (Ours)	C + L	**77.0**	**55.3**	**59.0**	**71.9**	**62.1**	**56.3**	**55.3**	**75.9**	**62.5**

where $\mathbf{p}_i$ is the ground-truth coordinate of the i-th point and N_p is the number of points per instance. The classification loss $\mathcal{L}_{\text{cls}}$ and the connectivity loss $\mathcal{L}_{\text{con}}$ are defined as focal losses [20] for class score prediction in ULG and connectivity prediction in GMR, respectively.

4 Experiments

4.1 Experiment Setup

Datasets. We evaluate our method on nuScenes [1] and Argoverse2 [36], which contain 1,000 driving scenes each. nuScenes provides 20-second scenes at 2 Hz, whereas Argoverse2 provides approximately 15-second scenes at 10 Hz. To evaluate generalization to unseen areas, we follow the geographically non-overlapping new splits of StreamMapNet [43].

Evaluation Metrics. We evaluate local map prediction using *mean Average Precision* (mAP), following previous works [19, 43]. For global map construction, we adopt scene-level *mean Global Average Precision* (mGAP) [31]. We also report multi-trip mGAP, where overlapping predictions from different scenes are accumulated into an existing global map and evaluated as a unified map. We further evaluate whether predictions cover each GT element without fragmentation by adapting the Entropy-based Connectivity Metric (ECM) [3, 41] to the instance level. This penalizes fragmented predictions assigned to the same GT element, and the class-averaged score is reported as miECM. Details are provided in the Appendix.

Implementation Details. For fair comparison, static point cloud reconstruction and dynamic feature masking are used only in the offboard camera–LiDAR setting, not in the onboard camera-only setting. Dynamic-object point clouds are removed using ground-truth 3D boxes, which can be replaced by any 3D object detector. We use ResNet-50 as the image backbone. For the standard scene-level evaluation, we initialize each scene with a global map prior constructed from ReVMap's own local predictions across the scene. ReVMap is trained in three stages: local-map-only training, connectivity-module training with the local model frozen, and end-to-end training with global priors, for 24, 1, and 6 epochs, respectively. More implementation details are provided in the Appendix.

Table 3: Camera-only comparison on nuScenes with global map priors constructed using only past frames.

Method	mAP	mGAP	miECM
GlobalMapNet [31]	34.8	17.6	32.4
MapTracker [2]	40.3	24.5	37.1
ReVMap (Ours)	**43.2**	**26.2**	**38.7**

Table 4: Performance comparison of multi-trip global map construction on nuScenes. The "-" indicates that the method cannot merge maps across scenes. "$*$" denotes reproduced results.

Method	mAP	miECM	mGAP
GlobalMapNet* [31]	34.8	25.5	13.7
MapTracker [2]	40.3	-	-
ReVMap (Ours)	**45.6**	**34.1**	**21.1**

Table 5: Ablation study of main components of ReVMap.

			GAP Metrics			
Method	mAP	miECM	*div*	*ped*	*bou*	*mean*
Baseline	39.0	30.1	13.6	32.0	12.7	19.4
+ Connectivity Estimation	39.0	34.9	14.8	30.5	23.5	22.9
+ Global Map Prior	44.8	40.0	18.0	34.9	27.6	26.8
+ Uncertainty Estimation	**45.6**	40.2	17.9	36.2	28.1	27.4
+ Uncertainty Update	**45.6**	**40.3**	**18.1**	**37.9**	**29.6**	**28.5**

Table 6: Effect of replacing the GT boxes used for dynamic-object removal in the C+L setting on nuScenes.

Boxes	mAP↑	mGAP↑
GT boxes	**58.0**	**36.1**
Perturbed GT boxes	57.6	35.6
No removal	57.3	35.2

Table 7: Ablation study for evaluating the effect of Uncertainty Update. "Filter" denotes the Uncertainty Filtering and "Interp" refers to Uncertainty-weighted Interpolation.

			GAP Metrics			
Filter	Interp	miECM	*div*	*ped*	*bou*	*mean*
		40.2	17.9	36.2	28.1	27.4
✓		40.2	18.0	37.3	29.0	28.1
	✓	40.3	17.8	37.1	28.9	27.9
✓	✓	**40.3**	**18.1**	**37.9**	**29.6**	**28.5**

Table 8: Analysis of different local prediction and global construction method combinations.

Local Map	Global Map	mAP	miECM	mGAP
GlobalMapNet	GlobalMapNet	34.8	32.4	17.6
	Ours	34.8	**35.2** (+2.8)	**19.4** (+1.8)
MapTracker	MapTracker	40.3	37.1	24.5
	Ours	40.3	**37.8** (+0.7)	**26.5** (+2.0)
Ours	GlobalMapNet	45.6	38.2	23.5
	VMA	45.6	39.3	25.7
	Ours	45.6	**40.3** (+2.1)	**28.5** (+5.0)

4.2 Performance Comparison

Table 1 compares ReVMap with recent online mapping and global map construction methods [2, 7, 28, 31, 43, 47] on the nuScenes new split validation set, showing state-of-the-art performance in the onboard camera-only setting and substantial gains with offboard camera–LiDAR inputs, reaching 58.0 mAP, 36.1 mGAP, and 43.4 miECM. Table 2 shows that ReVMap also outperforms existing global map construction methods [2, 31] on Argoverse2 across all metrics.

Since the scene-level prior may contain elements observed only in later frames, Table 3 further compares camera-only methods with the same ResNet-50 backbone under a strictly causal protocol, where each scene starts from an empty map and the prior is built only from past frames. ReVMap still outperforms MapTracker [2] by 2.9% mAP, 1.7% mGAP, and 1.6% miECM, showing that its gain does not rely on future observations.

Table 4 evaluates multi-trip global map construction, a cross-scene setting that MapTracker [2] cannot handle due to its single-scene identity propagation scheme. ReVMap outperforms GlobalMapNet [31] in mAP, miECM, and mGAP, establishing state-of-the-art multi-trip global map construction.

4.3 Ablation Study

We conducted ablation studies to evaluate ReVMap's core components. Unless otherwise noted, these experiments used the onboard camera-only setting, with models trained on the nuScenes training new split and evaluated on the new split validation set.

Contributions of Main Components.

Table 5 ablates the main components of ReVMap. The baseline uses a hand-crafted endpoint-distance rule for local-to-global association instead of learned connectivity estimation. Connectivity Estimation improves mGAP by 3.5% and miECM by 4.8%, demonstrating more reliable association than rule-based matching. Adding the Global Map Prior further improves mAP, mGAP, and miECM by 5.8%, 3.9%, and 5.1%, showing the importance of accumulated map context. Uncertainty Estimation and

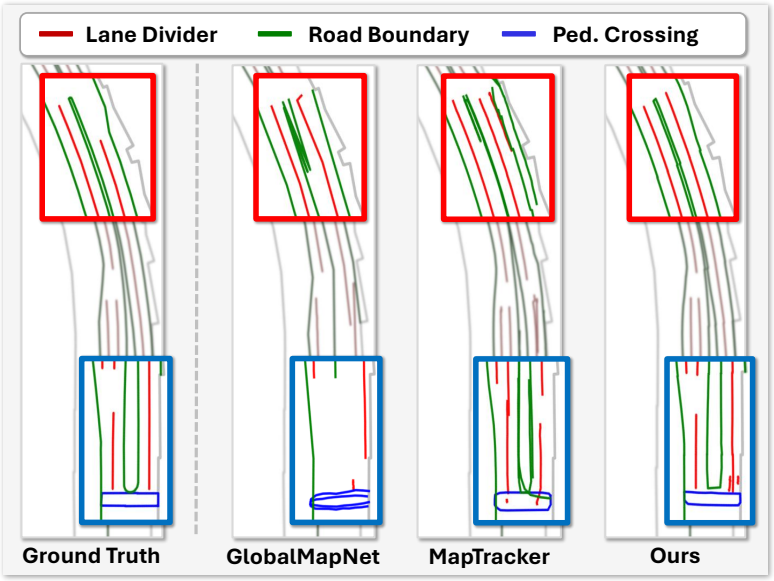


Figure 4: **Qualitative results on the nuScenes new split validation set.**

Update further improve mGAP by 0.6% and 1.1%, strengthening reliability-aware prediction and global map fusion.

Dependence on GT Boxes for Dynamic-object Removal. In the offboard camera–LiDAR setting, dynamic objects are removed from the aggregated LiDAR sweeps using GT 3D boxes. Since public nuScenes detectors are trained on scenes that overlap the new split validation set, Table 6 instead perturbs the GT boxes with missed detections and box noise so that their class-wise AP matches that of MambaFusion [34], and also evaluates a variant that removes no dynamic points. mAP and mGAP drop by only 0.4% and 0.5% with perturbed boxes and by 0.7% and 0.9% without removal, showing that ReVMap does not critically depend on GT boxes.

Effects of Uncertainty Update. Table 7 evaluates uncertainty-based Filtering and uncertainty-weighted Interpolation in the Uncertainty Update. Filtering alone improves mGAP by 0.7%, while Interpolation alone yields a 0.5% gain, and combining both operations raises mGAP to 28.5%, demonstrating their complementary roles in reliable global map update.

Plug-in Analysis of Global Construction. Table 8 analyzes our global map construction across different local predictors and existing global methods. GlobalMapNet and MapTracker use their own local predictors trained without an initial map, and the same ReVMap local predictions are processed by GlobalMapNet, VMA, and our method, so only the global construction method changes within each group. Combining our global construction with GlobalMapNet [31] and MapTracker [2] improves both mGAP and miECM, while replacing ours with GlobalMapNet or VMA [3] under the same local predictor degrades mGAP by 5.0% and 2.8%, respectively. This confirms that the gains generalize across local predictors and stem from the proposed connectivity estimation and uncertainty-aware update.

4.4 Qualitative Analysis

Qualitative Results Compared with SOTA Methods.

Figure 4 compares ReVMap with GlobalMapNet [31] and MapTracker [2]. While previous methods exhibit topological inconsistencies, ReVMap produces more coherent and structurally consistent maps. Additional qualitative results are provided in the Appendix.

Qualitative Analysis of the Update Process. Figure 5 illustrates the matching and update process of existing methods and the proposed method. VMA [3] discards valid local instances and retains erroneous global geometry because its rule-based matching rejects pairs that fail endpoint-distance criteria. GlobalMapNet [31] produces incorrect updates because it simply replaces only the overlap-

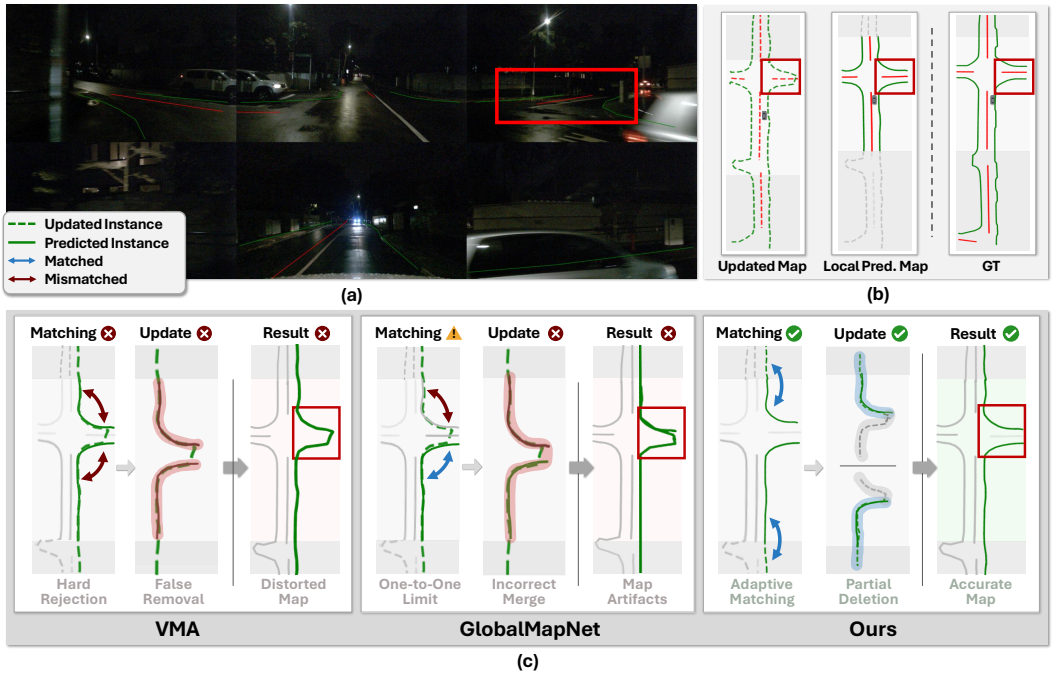


Figure 5: **Visualization of the Matching and Update Process.** (a) Multi-view camera inputs. (b) The accumulated global map, the current local prediction, and the ground-truth global map. (c) Comparison of update processes between existing methods and the proposed ReVMap.

ping regions between matched local and global instances. In contrast, ReVMap identifies revision targets through connectivity estimation and applies uncertainty-aware fusion, correcting erroneous segments without disrupting valid global structures.

5 Discussion and Conclusion

Limitations and Future Work. Despite reducing manual effort in map auto-labeling, ReVMap remains supervised and relies on ground-truth vectorized maps to train its local map prediction and connectivity estimation modules. In addition, the map update process relies on empirically selected thresholds for uncertainty filtering and geometric revision, which may require tuning for different datasets. Future work could reduce reliance on ground-truth maps by generating pseudo-labels from unlabeled sensor data, potentially leveraging neural reconstruction methods such as Gaussian splatting to provide geometric cues.

Conclusion. We presented ReVMap, a local-to-global framework for reliable vectorized global map construction. By combining uncertainty-guided prior reuse with connectivity-guided global map revision, ReVMap enables the correction of accumulated map errors while limiting their propagation into subsequent local predictions. Experiments on nuScenes and Argoverse2 demonstrate state-of-the-art performance in both local map prediction and global map construction, highlighting the effectiveness of reliability-aware map evolution.

Acknowledgments and Disclosure of Funding

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [No. RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)], the Technology Innovation Program (RS-2025-25448249, E2E Autonomous Driving Reference Data Construction and Core Technology Development) funded by the Ministry of Trade, Industry & Resources (MOTIR, Korea), and Hyundai Motor Company. The authors declare no competing interests.

References

[1] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[2] Jiacheng Chen, Yuefan Wu, Jiaqi Tan, Hang Ma, and Yasutaka Furukawa. Maptracker: Tracking with strided memory fusion for consistent vector hd mapping. In *European Conference on Computer Vision*, pages 90–107. Springer, 2024.

[3] Shaoyu Chen, Yunchi Zhang, Bencheng Liao, Jiafeng Xie, Tianheng Cheng, Wei Sui, Qian Zhang, Chang Huang, Wenyu Liu, and Xinggang Wang. Vma: Divide-and-conquer vectorized map annotation system for large-scale driving scene. *arXiv preprint arXiv:2304.09807*, 2023.

[4] Shiyuan Chen, Jiaxin Zhang, Ruohong Mei, Yingfeng Cai, Haoran Yin, Wei Sui, and Cong Yang. Camav2: A vision-centric approach for static map element annotation. *IEEE Transactions on Intelligent Transportation Systems*, 2026.

[5] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022.

[6] Sehwan Choi, Jungho Kim, Junyong Yun, and Jun Won Choi. R-pred: Two-stage motion prediction via tube-query attention-based trajectory refinement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8525–8535, 2023.

[7] Sehwan Choi, Jungho Kim, Hongjae Shin, and Jun Won Choi. Mask2map: Vectorized hd map construction using bird's eye view segmentation masks. In *European Conference on Computer Vision*, pages 19–36. Springer, 2024.

[8] Wenjie Ding, Limeng Qiao, Xi Qiu, and Chi Zhang. Pivotnet: Vectorized pivot learning for end-to-end hd map construction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3672–3682, 2023.

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[10] Zebang Feng, Miao Fan, Bao Liu, Shengtong Xu, and Haoyi Xiong. End-to-end generation of city-scale vectorized maps by crowdsourced vehicles. *arXiv preprint arXiv:2507.08901*, 2025.

[11] Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, and Boris Ivanovic. Producing and leveraging online map uncertainty in trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14521–14530, 2024.

[12] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17853–17862, 2023.

[13] Bo Jiang, Shaoyu Chen, Qing Xu, Bencheng Liao, Jiajie Chen, Helong Zhou, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Vad: Vectorized scene representation for efficient autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8350, 2023.

[14] Jungho Kim, Jiyong Oh, Seunghoon Yu, Hongjae Shin, Donghyuk Kwak, and Jun Won Choi. Safedrive: Fine-grained safety reasoning for end-to-end driving in a sparse world. *arXiv preprint arXiv:2602.18887*, 2026.

[15] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

[16] 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)*, pages 4628–4634. IEEE, 2022.

[17] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *European Conference on Computer Vision*, pages 541–556. Springer, 2020.

[18] 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. *arXiv preprint arXiv:2208.14437*, 2022.

[19] 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*, 133(3):1352–1374, 2025.

[20] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[21] 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*, pages 14812–14821, 2024.

[22] Xiaolu Liu, Ruizi Yang, Song Wang, Wentong Li, Junbo Chen, and Jianke Zhu. Uncertainty-instructed structure injection for generalizable hd map construction. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 22359–22368, 2025.

[23] Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. Vectormapnet: End-to-end vectorized hd map learning. In *International conference on machine learning*, pages 22352–22369. PMLR, 2023.

[24] Yinqi Liu, Mengxuan Song, and Yafeng Guo. An incremental fusing method for high-definition map updating. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 4251–4256. IEEE, 2019.

[25] Zhijian Liu, Haotian Tang, Alexander Amini, Xinyu Yang, Huizi Mao, Daniela L Rus, and Song Han. Bevfusion: Multi-task multi-sensor fusion with unified bird's-eye view representation. In *2023 IEEE international conference on robotics and automation (ICRA)*, pages 2774–2781. IEEE, 2023.

[26] 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*, pages 461–477. Springer, 2024.

[27] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[28] Nan Peng, Xun Zhou, Mingming Wang, Guisong Chen, and Wenqi Xu. Uni-prevpredmap: Extending prevpredmap to a unified framework of prior-informed modeling for online vectorized hd map construction. *arXiv preprint arXiv:2504.06647*, 2025.

[29] Jonah Philion and Sanja Fidler. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In *European conference on computer vision*, pages 194–210. Springer, 2020.

[30] Limeng Qiao, Wenjie Ding, Xi Qiu, and Chi Zhang. End-to-end vectorized hd-map construction with piecewise bezier curve. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13218–13228, 2023.

[31] Anqi Shi, Yuze Cai, Xiangyu Chen, Jian Pu, Zeyu Fu, and Hong Lu. Globalmapnet: An online framework for vectorized global hd map construction. *arXiv preprint arXiv:2409.10063*, 2024.

[32] 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.

[33] Kun Tang, Xu Cao, Zhipeng Cao, Tong Zhou, Erlong Li, Ao Liu, Shengtao Zou, Chang Liu, Shuqi Mei, Elena Sizikova, et al. Thma: Tencent hd map ai system for creating hd map annotations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15585–15593, 2023.

[34] Hanshi Wang, Jin Gao, Weiming Hu, and Zhipeng Zhang. MambaFusion: Height-fidelity dense global fusion for multi-modal 3D object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 26664–26674, 2025.

[35] Huijie Wang, Tianyu Li, Yang Li, Li Chen, Chonghao Sima, Zhenbo Liu, Bangjun Wang, Peijin Jia, Yuting Wang, Shengyin Jiang, et al. Openlane-v2: A topology reasoning benchmark for unified 3d hd mapping. *Advances in Neural Information Processing Systems*, 36:18873–18884, 2023.

[36] Benjamin Wilson, William Qi, Tanmay Agarwal, John Lambert, Jagjeet Singh, Siddhesh Khandelwal, Bowen Pan, Ratnesh Kumar, Andrew Hartnett, Jhony Kaesemodel Pontes, et al. Argoverse 2: Next generation datasets for self-driving perception and forecasting. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

[37] Wei Wu, Xiaoxin Feng, Ziyan Gao, and Yuheng Kan. Smart: Scalable multi-agent real-time motion generation via next-token prediction. *Advances in Neural Information Processing Systems*, 37:114048–114071, 2024.

[38] Deguo Xia, Weiming Zhang, Xiyan Liu, Wei Zhang, Chenting Gong, Jizhou Huang, Mengmeng Yang, and Diange Yang. Dumapnet: An end-to-end vectorization system for city-scale lane-level map generation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6015–6024, 2024.

[39] Ziyang Xie, Ziqi Pang, and Yu-Xiong Wang. Mv-map: Offboard hd-map generation with multi-view consistency. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8658–8668, 2023.

[40] Xuan Xiong, Yicheng Liu, Tianyuan Yuan, Yue Wang, Yilun Wang, and Hang Zhao. Neural map prior for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17535–17544, 2023.

[41] Zhenhua Xu, Yuxiang Sun, and Ming Liu. Topo-boundary: A benchmark dataset on topological road-boundary detection using aerial images for autonomous driving. *IEEE Robotics and Automation Letters*, 6(4):7248–7255, 2021.

[42] Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018.

[43] Tianyuan Yuan, Yicheng Liu, Yue Wang, Yilun Wang, and Hang Zhao. Streammapnet: Streaming mapping network for vectorized online hd map construction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7356–7365, 2024.

[44] Tianyuan Yuan, Yucheng Mao, Jiawei Yang, Yicheng Liu, Yue Wang, and Hang Zhao. Presight: Enhancing autonomous vehicle perception with city-scale nerf priors. In *European Conference on Computer Vision*, pages 323–339. Springer, 2024.

[45] Yujian Yuan, Changjie Wu, Xinyuan Chang, Sijin Wang, Hang Zhang, Shiyi Liang, Shuang Zeng, and Mu Xu. Unimapgen: A generative framework for large-scale map construction from multi-modal data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 12277–12285, 2026.

[46] Jiaxin Zhang, Shiyuan Chen, Haoran Yin, Ruohong Mei, Xuan Liu, Cong Yang, Qian Zhang, and Wei Sui. A vision-centric approach for static map element annotation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15861–15867. IEEE, 2024.

[47] Xiaoyu Zhang, Guangwei Liu, Zihao Liu, Ningyi Xu, Yunhui Liu, and Ji Zhao. Enhancing vectorized map perception with historical rasterized maps. In *European Conference on Computer Vision*, pages 422–439. Springer, 2024.

[48] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu. Hivt: Hierarchical vector transformer for multi-agent motion prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8823–8833, 2022.

[49] Jiangtong Zhu, Zhao Yang, Yinan Shi, Jianwu Fang, and Jianru Xue. Ic-mapper: Instance-centric spatio-temporal modeling for online vectorized map construction. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 9961–9969, 2024.

Supplementary Materials for ReVMap

A Additional Details for ReVMap

A.1 Evaluation Metrics

Chamfer-based AP Metrics. We evaluate mAP, mGAP, and multi-trip mGAP using the same Chamfer-distance-based AP computation. AP is computed at distance thresholds of 0.5, 1.0, and 1.5 meters for lane divider, pedestrian crossing, and road boundary classes. The final score is obtained by averaging AP over thresholds and classes.

Instance-level ECM. ECM [41] is designed to penalize cases where a single GT element is covered by multiple fragmented predictions. To compute this, predicted map elements are rasterized, and each prediction mask is assigned to the GT mask closest to most of its pixels. It then scores each GT mask using the coverage and entropy of its assigned prediction masks. However, overlapping fragmented predictions are merged into a single component after rasterization, causing ECM to treat multiple fragments as a single prediction and thereby under-penalize fragmented predictions. We therefore rasterize each predicted polyline separately, keeping fragments as individual prediction masks, and compute ECM in the same manner. For each GT mask M_j, C_j measures the entropy of the normalized sizes of its assigned prediction masks, while α_j measures their coverage of M_j. The per-class instance-level ECM is

$$\mathrm{iECM}_c = \sum_j \frac{|M_j|}{\sum_{j'} |M_{j'}|} \alpha_j e^{-C_j}.$$

We report miECM by averaging over classes, for both scene-level and multi-trip global maps.

TOP Metric. To evaluate the local-to-global connectivity prediction performance in the Connectivity Estimation Module, we adapt the TOP metric [35], proposed for lane topology reasoning. The original TOP metric uses Fréchet-distance-based matching to align predicted instances with ground-truth instances before computing per-instance mean average precision. We instead evaluate the binary connectivity matrix between local predictions and global instances using Chamfer-distance-based matching with a distance threshold of 1.5 m. The connectivity ground truth is defined based on matched ground-truth instances in each region, assigning connectivity when a local instance and a global instance are matched to the same ground-truth instance. It is defined over all four directional combinations between the original and reverse directions of local and global instances, with only the convergent matching pair labeled as positive.

A.2 Implementation Details

For nuScenes, input images are resized by 0.5. For Argoverse2, seven camera images are padded to 2048×2048 and resized by 0.3. For image-to-BEV transformation, we use an LSS-based image-to-BEV module with a BEV grid resolution of 0.3 m. The local perception range is set to $[-30, 30, -15, 15]$m, and global instances for connectivity estimation are retrieved within $[-45, 45, -30, 30]$m, following the longitudinal and lateral order. Mask2Former [5] is pretrained on Mapillary Vistas V2 [27] for dynamic-region masking. ULG uses 100 instance queries, each with 20 point queries, so each local map instance contains 20 points. Global instances in the Global Prior Encoder and Connectivity Module are also resampled to 20 points. The Global Prior Encoder and Local Map Decoder consist of 2 and 6 layers, respectively. The Connectivity Module consists of an encoder with a single layer and a decoder with 6 layers. For uncertainty filtering, the uncertainty thresholds for lane divider, pedestrian crossing, and road boundary are empirically set to 1.6, 1.4, and 2.8, respectively. Before fusing a matched local-global instance pair, we trim the matched global instance by discarding segments farther than 1.5 m from the local instance. We use AdamW with an initial learning rate of 6×10^{-4} and weight decay of 0.01, cosine learning-rate decay with 500 warm-up iterations, and four A6000 GPUs with a total batch size of 16.

B Supplementary Experimental Results

Table 9: Impact of ground-truth preprocessing on local map construction. All models are evaluated under the camera-only onboard setting.

Method	*div*	*ped*	*bou*	mAP
MapTRv2 (merged GT)	26.5	37.2	46.1	36.6
MapTRv2 (split GT)	29.3	42.0	45.7	39.0
ReVMap (Ours)	33.7	49.6	53.4	45.6

Table 10: Ablation study for evaluating the effect of the Connectivity Estimation submodules. "Encoder" denotes the bidirectional Map Instance Tokenizer and "Decoder" refers to the Graph-based Connectivity Predictor.

		TOP Metrics			
Encoder	Decoder	*div*	*ped*	*bou*	*mean*
		58.2	58.3	56.9	57.8
✓		61.0	62.5	57.1	60.2
	✓	61.8	63.4	57.0	60.7
✓	✓	**62.7**	**65.8**	**57.4**	**61.9**

B.1 Baseline Analysis with Consistent Pedestrian Crossing Ground Truth

Recent local map construction methods use different ground-truth definitions for pedestrian crossings. Several methods [7, 18, 19, 26] merge overlapping pedestrian crossing polygon fragments into a single polygon and resample it into a fixed number of points. In contrast, StreamMapNet [43] and MapTracker [2] use a different preprocessing strategy that avoids indiscriminate merging of pedestrian crossing fragments and produces more consistent pedestrian crossing instances.

Following this StreamMapNet-style preprocessing, we conduct an ablation study to clarify whether the performance gain comes from ground-truth preprocessing or from the proposed architecture. As shown in Table 9, changing the MapTRv2 baseline from merged GT to split GT notably improves pedestrian crossing AP from 37.2% to 42.0%, showing that the pedestrian crossing annotation definition has a significant impact on this category. Under the same split-GT setting, our model further improves the strengthened baseline and achieves the best performance across all categories, increasing mAP from 39.0% to 45.6%.

B.2 Effect of the Connectivity Estimation submodules.

Table 10 evaluates the two submodules of the Connectivity Estimation. In the baseline configuration, both submodules are replaced with simple MLPs. Adding either submodule alone improves the TOP metric by 2.4% and 2.9%, respectively, and applying both together yields a further gain to 61.9%, showing that the two submodules are complementary. The bidirectional encoder captures the point ordering direction at each end of a polyline, while the graph-based decoder models geometric proximity between candidate pairs.

C Extensive Qualitative Results

Figure 6 visualizes the predicted local HD map together with its point-wise uncertainty on the nuScenes new split validation set. The uncertainty estimates concentrate on visually ambiguous and occluded regions, showing that ULG captures localization reliability. Figures 7 and 8 present additional qualitative results on the nuScenes and Argoverse2 new split validation sets, respectively, showing the consistency of the proposed method across scene layouts. Figures 9 and 10 provide comparisons between the reconstructed multi-trip global maps and the corresponding ground-truth maps, demonstrating that the proposed method produces globally coherent map structures over large-scale driving areas.



Figure 6: **Qualitative visualization of uncertainty-aware local vectorized maps on the nuScenes new split validation set.**

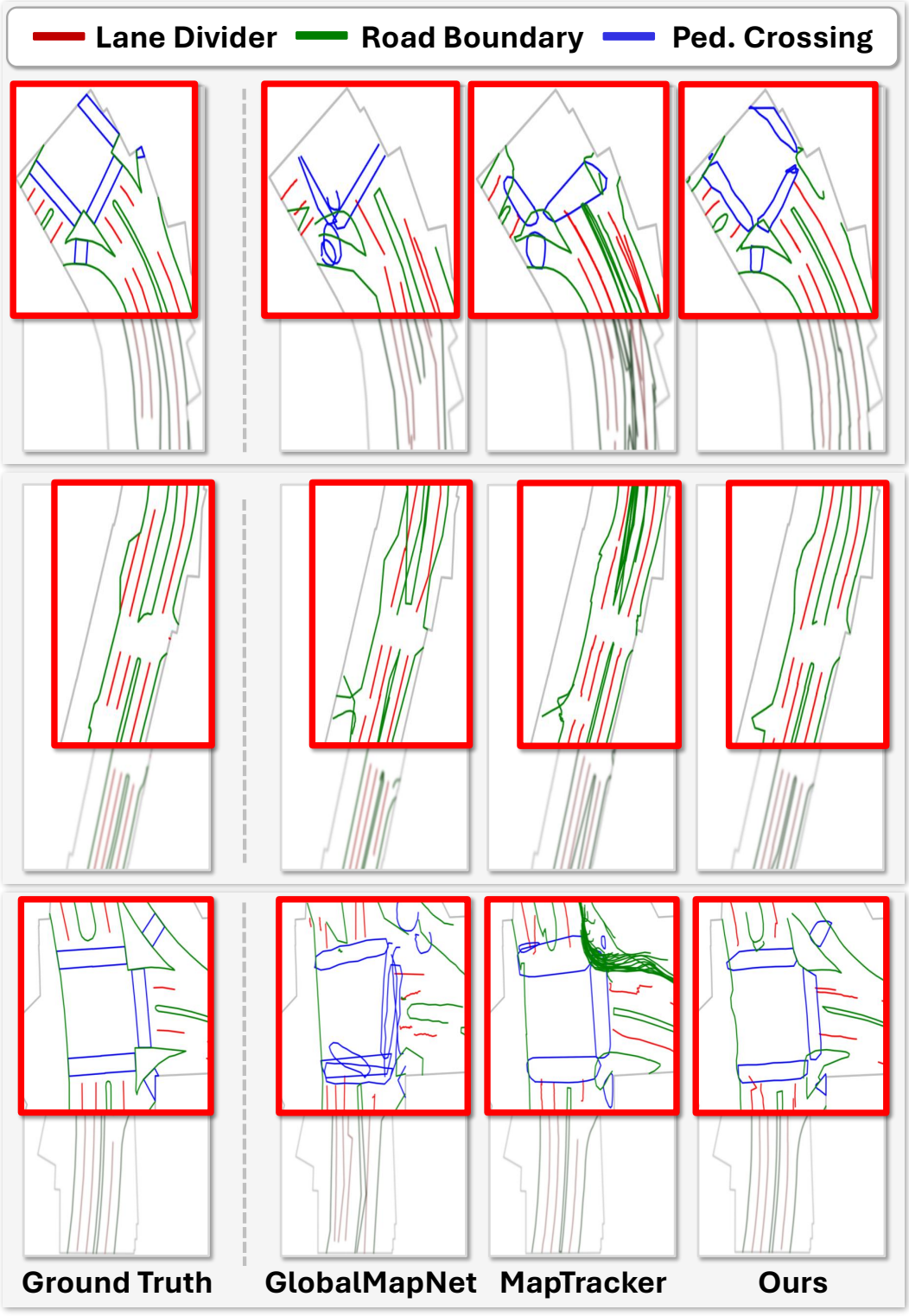


Figure 7: **Additional qualitative results on the nuScenes new split validation set.**

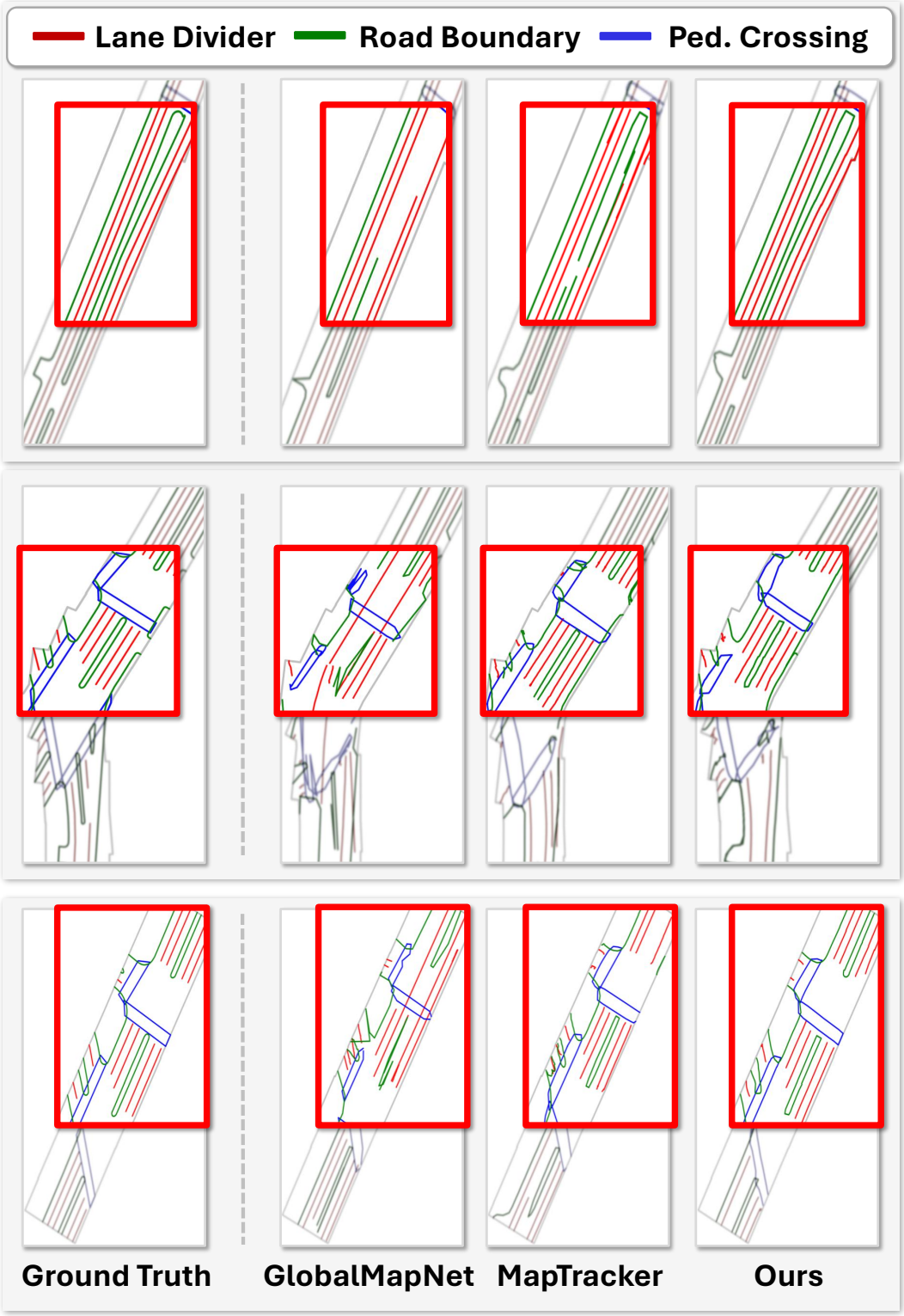


Figure 8: **Qualitative results on the Argoverse2 new split validation set.**

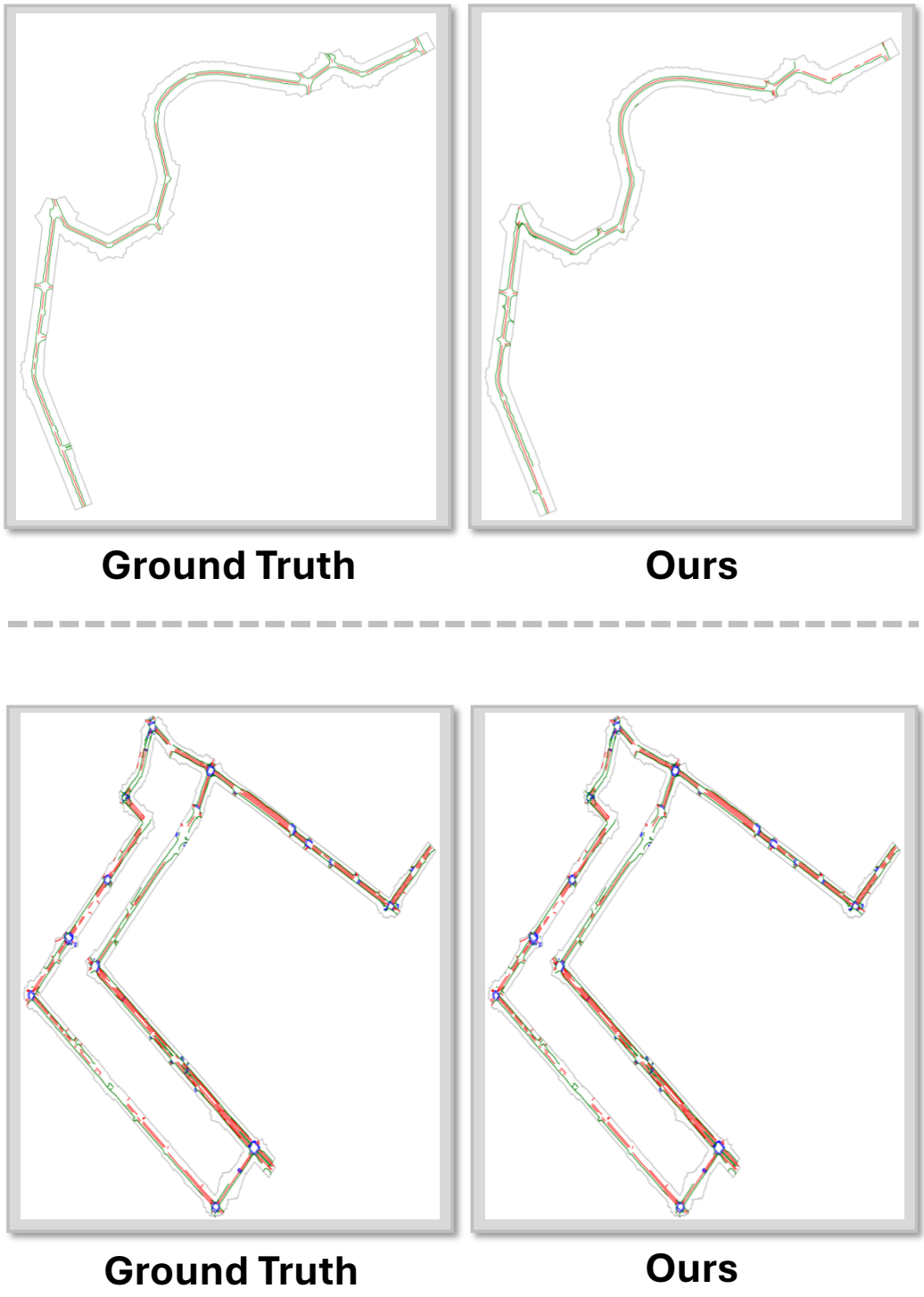


Figure 9: **Qualitative results of multi-trip vectorized global map construction on the nuScenes new split validation set.**

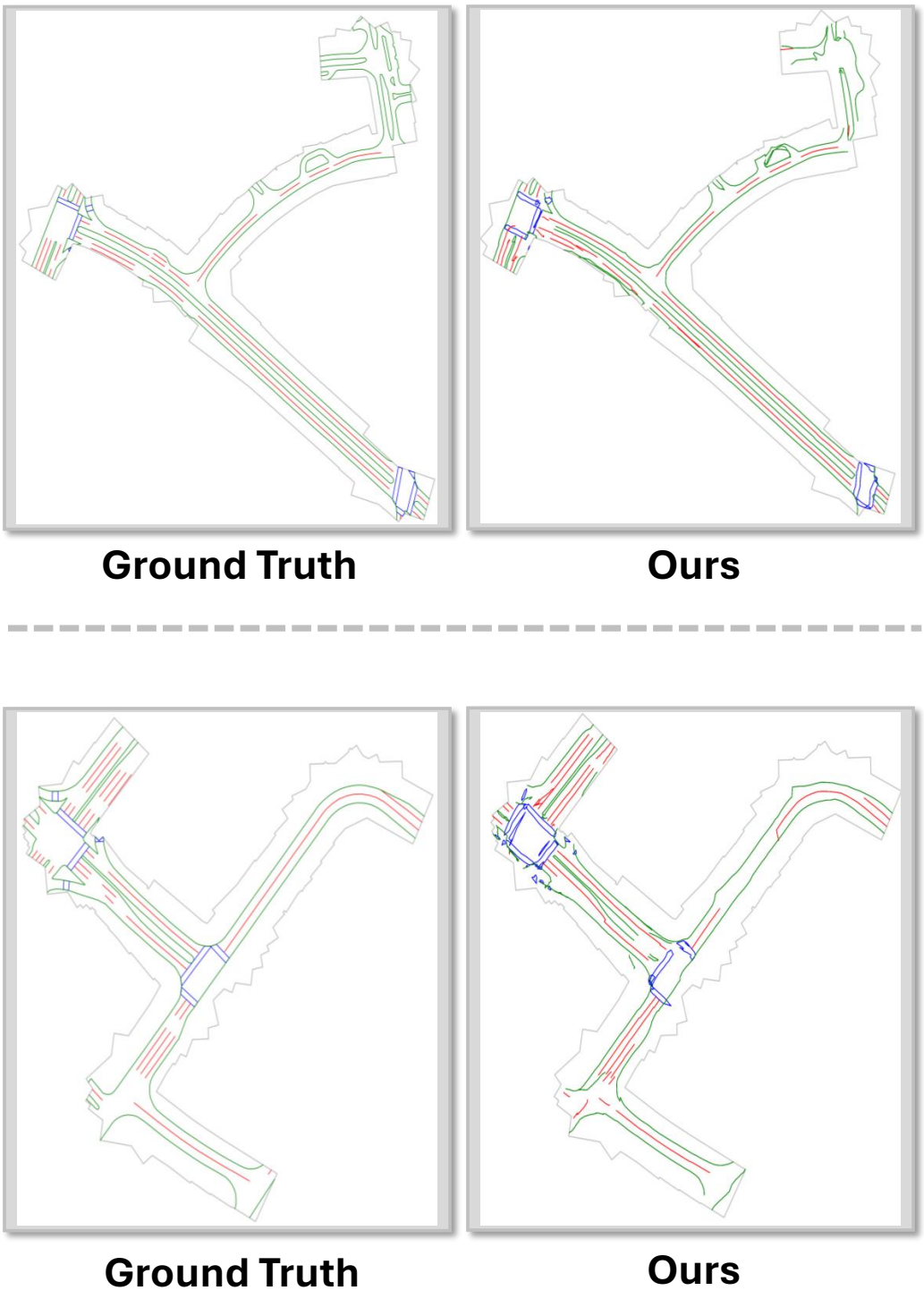


Figure 10: **Qualitative results of multi-trip vectorized global map construction on the nuScenes new split validation set.**

D Licenses

We summarize the URLs, citations, and licenses of the external assets used in this work in Table 11.

Table 11: **URLs, citations, and licenses of external assets used in this work.**

URL	Citation	License
https://www.nuscenes.org/terms-of-use	[1]	CC BY-NC-SA 4.0 license
https://www.argoverse.org/about.html	[36]	CC BY-NC-SA 4.0 license
https://github.com/nutonomy/nuscenes-devkit	[1]	Apache-2.0 license
https://github.com/argoverse/av2-api	[36]	MIT license
https://www.mapillary.com/dataset/vistas	[27]	CC BY-NC-SA 4.0 license
https://github.com/hustvl/MapTR	[18, 19]	MIT license
https://github.com/mit-han-lab/bevfusion	[25]	Apache-2.0 license
https://github.com/facebookresearch/Mask2Former	[5]	MIT license[a]
https://github.com/yuantianyuan01/StreamMapNet	[43]	GPL-3.0 license
https://github.com/Honminden/GlobalMapNet	[31]	GPL-3.0 license
https://github.com/woodfrog/maptracker	[2]	GPL-3.0 license[b]
https://github.com/hustvl/VMA	[3]	MIT license
https://github.com/SehwanChoi0307/Mask2Map	[7]	Apache-2.0 license
https://github.com/HXMap/HRMapNet	[47]	MIT license
https://github.com/pnnnnnnn/Uni-PrevPredMap	[28]	MIT license

[a]Mask2Former is predominantly MIT-licensed; specified components are licensed under Apache-2.0.
[b]MapTracker applies GPLv3 terms to research use and prohibits commercial use of its code, data, and model weights.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims in the abstract and introduction are supported by the proposed method description in Section 3 and the experimental results in Section 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations of our method in Section 5.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

Justification: The paper does not include theoretical results, theorems, or formal proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide the model architecture, dataset details, evaluation metrics, and implementation details necessary to reproduce the main experimental results in Section 4.1 and the supplementary material A.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: The code is not publicly available because the research was conducted under an industry collaboration agreement that restricts code release.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: We specify the dataset splits, training hyperparameters, optimizer, learning rate schedule, and evaluation settings in Section 4.1 and the supplementary material A.2.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: We do not report error bars, confidence intervals, or statistical significance tests due to the computational cost of repeated training runs. We provide comparisons under consistent experimental settings and ablation studies.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We report the compute resources used for our experiments. The main experiments were conducted using 4 NVIDIA RTX A6000 GPUs with 48GB memory each; training took approximately 24 hours, and evaluation took approximately 30 minutes.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and believe that our research conforms to it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

Justification: ReVMap can reduce human annotation effort for large-scale HD map construction and improve the reliability of auto-labeled maps. However, inaccurate map predictions could negatively affect downstream autonomous driving systems if used without careful validation.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release high-risk generative models, scraped datasets, or models with direct misuse potential requiring special safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The URLs, citations and licenses of external assets used in this work are provided in Table 11 of supplementary material.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We do not release new assets. Code release is restricted by the industry collaboration agreement described in our response to Checklist 5.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The paper does not involve human subjects research and therefore does not require IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as an important, original, or non-standard component.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.