PriorMaskMap: Robust Online Vectorized Map Construction with Biased Priors

Haoming Xu[1,2], Wei Li[1,3] *, and Yu Hu[1,3]

[1] The Research Center for Intelligent Computing Systems, SKLP, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2] Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China
[3] University of Chinese Academy of Sciences, Beijing 100049, China
xuhaoming24@mails.ucas.ac.cn
liwei2019@ict.ac.cn
huyu@ict.ac.cn

Abstract. Online vectorized map construction provides autonomous vehicles with essential, real-time semantic and geometric scene understanding. While leveraging prior maps presents an opportunity to improve accuracy, they are often biased due to outdated information or temporary road changes. Existing methods, which typically treat priors as fully reliable or only consider limited noise patterns, suffer severe performance degradation when these priors are imperfect. To address this, we propose PriorMaskMap, a robust framework for prior-guided vectorized map construction. Its first innovation is a Prior-Aware Confidence Estimator (PACE) that evaluates the input rasterized map against current visual observations. It assigns confidence scores to regions classified as reliable, erroneous, or blank, creating a foundational mask that guides two subsequent branches. In one branch, the Dual-Prior Query Generator (DPQG) selectively filters and embeds both rasterized and vectorized priors to initialize informative queries from trustworthy map structures. In the other, the Confidence-Guided Prior Fusion (CGPF) module spatially propagates and refines the mask, using the refined version to govern the fusion of BEV features with the prior map, thereby providing rich contextual representations for the map decoder. Evaluation on the nuScenes dataset shows that PriorMaskMap achieves superior mapping precision over existing approaches and exhibits marked robustness when provided with priors containing apparent errors. Our code is released at https://github.com/healenrens/PriorMaskMap.

Keywords: HD map construction · Prior map · Autonomous driving

1 Introduction

The online construction of vectorized high-definition (HD) maps is critical for autonomous driving, providing precise lane geometry and topological relationships for downstream tasks such as trajectory planning and decision-making.

* Corresponding author.

While current methods have made significant progress in building maps directly from multi-view images [5, 15, 16, 21, 38], their performance can degrade noticeably in challenging scenarios such as poor lighting, heavy occlusions, or at long distances. Fortunately, the increasing availability of offline maps allows them to provide beyond-line-of-sight information as a valuable prior, thereby enabling more accurate mapping in those situations where pure visual perception struggles.

Consequently, a series of prior-enhanced online mapping methods have been developed [11, 22, 27, 36, 41]. These approaches vary in terms of the prior map format and its fusion strategy. Among them, P-MapNet [11] designs a pipeline that integrates both standard-definition (SD) and HD map as priors. It fuses the SD prior features with sensor-derived features in the bird's-eye-view (BEV) space using an attention mechanism, while HD map priors are incorporated through a masked autoencoder for refining the output. Uni-PrevPredMap [22] employs HD maps as priors by storing them in a tile-indexed map processor. This design enables the model to selectively leverage the stored prior information, where retrieved map vectors are rasterized and fused with BEV features while simultaneously serving to initialize the decoder's queries.

However, prior maps can contain significant biases and errors due to outdated information, large-scale road renovations, or sensor misalignment. Such imperfections can severely degrade the performance of prior-enhanced methods that fully trust the map input [37]. Recent works like MapEX [27] and Uni-PrevPredMap [22] have recognized and attempted to mitigate this problem. MapEX [27] encodes the prior map into queries and interacts them with BEV features via cross-attention, serving as input to the map decoder. Uni-PrevPredMap [22] introduces a tri-mode operational optimization paradigm to reduce dependence on map fidelity assumptions. While both methods improve robustness to imperfect priors, explicit modeling and propagation of spatially varying prior reliability for confidence-guided feature fusion remain underexplored. In real-world driving scenarios, discrepancies between offline maps and actual observations are often more complex, involving structural deviations at the lane-level which may affect multiple adjacent lanes due to road reconstruction [12, 24]. In such cases, explicitly identifying reliable and erroneous regions in the prior map becomes even more valuable, but this critical capability has not been adequately explored in existing methods.

To address this limitation, we propose PriorMaskMap as Fig. 1, a robust framework for online vectorized map construction that selectively leverages trustworthy regions within prior maps. We first design a Prior-Aware Confidence Estimator (PACE), which classifies each grid of the rasterized prior map as reliable, erroneous, or blank based on current visual observations, and outputs a spatially-differentiated confidence map. This confidence map then acts as a soft mask in our Dual-Prior Query Generator (DPQG), selectively gating the influence of the prior during query initialization. The DPQG is designed with dedicated encoding branches for rasterized and vectorized prior maps, respectively. The resulting embeddings from both branches are merged and interact

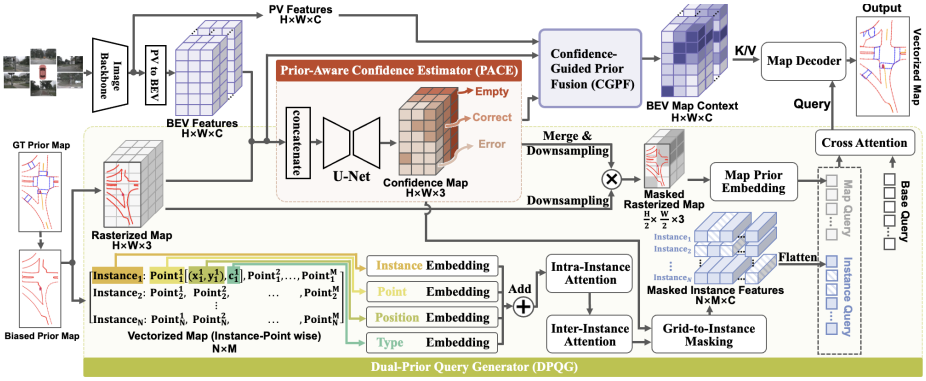


Fig. 1: Overall framework of PriorMaskMap. Given multi-view images and a biased prior map, PACE predicts grid-wise confidence over the rasterized prior. DPQG converts confident vector and raster structures into prior-aware queries, while CGPF diffuses and refines the confidence field to produce a BEV map context for the decoder.

through cross-attention with a set of learnable queries to decode desirable map elements. Furthermore, we introduce a Confidence-Guided Prior Fusion (CGPF) module, which refines the initial confidence through spatial propagation and uses it to adaptively fuse the prior map with BEV features. This process produces rich contextual representations that serves as Keys and Values for the transformer-based map decoder, effectively grounding the decoding process in both reliable prior knowledge and current sensor observations.

In summary, our contributions are threefold:

- We propose PriorMaskMap, a novel framework for robust vectorized map construction, built around a core Prior-Aware Confidence Estimator (PACE) that adaptively identifies trustworthy regions in biased prior maps and generates a corresponding grid-wise confidence map.
- We introduce two key components guided by the confidence map to enhance both map element queries and feature fusion: a Dual-Prior Query Generator that produces gated query embeddings from both rasterized and vectorized priors, and a Confidence-Guided Prior Fusion module that spatially propagates and refines the initial confidence to guide the fusion of BEV features with the prior.
- We extend the nuScenes dataset with a supplementary set of biased vectorized maps that capture real-world discrepancies, specifically simulating structural deviations across multiple lanes. Experiments show that PriorMaskMap achieves higher accuracy and better robustness against inaccurate priors compared to existing methods.

2 Related Work

2.1 Online HD Map Construction

Early online HD mapping methods treated map construction as BEV segmentation followed by vectorization. HDMapNet [13] first performs BEV semantic segmentation and fits lane geometries, which is effective but computationally expensive. Later works move to query-based transformers that directly predict vectorized maps. VectorMapNet [20] uses a two-stage transformer that detects element boxes and then autoregressively decodes polylines. MapTR [15] adopts a DETR-style design to output each lane or boundary as an ordered point set, and MapTRv2 [16] enhances it with 3D geometric reasoning and multi-view cues. PivotNet [5] instead learns an ordered list of pivot points per lane to better capture complex shapes and topology.

Building on this line, recent models further boost accuracy and robustness [28]. MapNeXt [14] shows that scaling model capacity and training leads to strong gains on vectorized map detection. MGMap [19] learns BEV masks that highlight informative regions and uses them to guide query decoding, improving fine-grained lane geometry under sparse supervision. Mask2Map [4] combines segmentation and queries by predicting instance-level BEV masks and then generating vectors from them.

2.2 Context-Enhanced Online Map Construction

A complementary line of work exploits temporal context for stable online mapping. SQD-MapNet [30] reuses previous outputs via streaming query denoising, MapTracker [3] propagates latent queries across frames to fuse past BEV features and vectorized maps, ADMap [8] adopts a disturbance-rejection training paradigm to handle imperfect sensors and moving obstacles, and EAN-MapNet [35] designs efficient query layouts to accelerate inference while preserving detail.

Many methods further enhance mapping with priors. P-MapNet [11] injects fixed HD/SD map priors into the decoder to improve long-range structure, while PriorMapNet [29] and PriorDrive [41] encode vectorized map priors into the query space for lane retrieval and ordering. Temporal priors leverage the vehicle's own history: HRMapNet [42] accumulates past predictions into a lightweight raster to augment decoding and initialize queries, and PrevPredMap [23] and MemFusionMap [26] forward previous maps or features to handle occlusions and stabilize updates. External environment priors extend this idea further: PreSight [40] introduces city-scale neural radiance priors for geometric regularization, AerialHDMapper [34] uses aerial imagery for long-range and cross-view constraints, and ImagineMap [9] and PriorFusion [28] enhance structure from SD maps and fuse multiple prior sources.

However, real priors are often biased or outdated. Empirical studies [1] show that prior-informed models are sensitive to misalignment and map changes, motivating explicit change awareness and alignment. ExelMap [31] targets explainable element-level change detection, RTMap [7] couples online mapping

with recursive change detection for self-correction, and ArgoTweak [32] proposes structured prior-driven self-updating at the database level. These systems retain global and long-range benefits from priors while learning to align, down-weight, or override them, and our work follows this direction by focusing on biased prior maps and designing mechanisms that safely exploit their structure while actively correcting their errors.

3 Methodology

Given multi-view images and a biased prior map, our goal is to construct an online vectorized HD map that selectively uses the prior when reliable and otherwise relies on current observations. As shown in Fig. 1, PriorMaskMap consists of three main components. The Prior-Aware Confidence Estimator predicts a per-grid confidence map over the rasterized prior. The Dual-Prior Query Generator uses this confidence to select reliable vector and raster structures and converts them into map-aware queries for the decoder. The Confidence-Guided Prior Fusion diffuses and refines the confidence field on the BEV grid and fuses it with visual features to produce a prior-aware BEV context. We describe these components successively.

3.1 Prior-Aware Confidence Estimator (PACE)

Recent architectures often assume that HD or SD prior maps are correct [11,42], which can mislead downstream fusion under road reconstruction [37]. We relax this assumption with a Prior-Aware Confidence Estimator (PACE) that predicts a dense confidence map and measures spatial consistency between the prior and current observations. High-confidence regions of the prior are preserved, while low-confidence regions fall back to evidence from the current view.

To assess map reliability from actual observations we use BEV features as the decision anchor. The BEV feature map $F_{\mathrm{BEV}} \in \mathbb{R}^{H\times W\times C}$ provides a unified top-down representation, where H and W denote the spatial resolution and C is the channel dimension. We rasterize the prior map into $M_R \in \mathbb{R}^{H\times W\times 3}$ and concatenate F_{BEV} and M_R along the channel dimension. The concatenated tensor is fed into a U-Net head [25] that predicts reliability on the raster grid. For each grid u, the network outputs a confidence map $M_{\mathrm{conf}}(u) \in \mathbb{R}^3$ for the classes reliable, erroneous, and empty, forming a confidence map $M_{\mathrm{conf}} \in \mathbb{R}^{H\times W\times 3}$ that guides subsequent selection and fusion.

To supervise PACE, we generate targets with a patch-level hybrid scheme based on rasterized maps. Let M_{gt} denote the ground-truth vectorized map and M_{err} the biased vectorized prior. We rasterize them with an operator $\mathcal{R}_r$ into three-channel rasters $M_{R\mathrm{gt}}, M_{R\mathrm{err}} \in \{0,1\}^{H\times W\times 3}$,

$$M_{R\mathrm{gt}} = \mathcal{R}_r(M_{\mathrm{gt}}), \tag{1a}$$

$$M_{R\mathrm{err}} = \mathcal{R}_r(M_{\mathrm{err}}). \tag{1b}$$

The subscript R indicates the rasterized domain. We partition the raster grid into non-overlapping patches h and assign each patch is composed of grids u. Grids label $L(u)$ is that summarizes spatial agreement between M_{Rgt} and M_{Rerr},

$$L(u) = \begin{cases} \texttt{empty}, & M_{R\text{gt}}(u) = 0 \ \wedge \ M_{R\text{err}}(u) = 0, \\ \texttt{erroneous}, & M_{R\text{gt}}(u) \neq M_{R\text{err}}(u), \\ \texttt{reliable}, & \text{otherwise.} \end{cases} \tag{2}$$

Empty grids contain no elements in either map, erroneous patches reflect disagreements, and all remaining occupied patches are treated as reliable.

During training the U-Net receives a single rasterized map as input for each patch. We sample a Bernoulli variable $B(h) \sim \text{Bernoulli}(\alpha)$ and construct a mixed raster M_R. It means that we mix in the patch level but the label is at the grid level.

$$M_R(h) = B(h)\, M_{R\text{err}}(h) + \big(1 - B(h)\big)\, M_{R\text{gt}}(h), \tag{3}$$

which randomly selects either the biased prior or the ground-truth prior in that patch. Labels are always computed from the pair $(M_{R\text{gt}}, M_{R\text{err}})$, so supervision is decoupled from input mixing and training remains stable. To address the class imbalance issue, we use focal loss and train with the main model. At inference time PACE only observes the available prior map, for example the biased prior in our experiments. We refer to M_{conf} as the confidence map and later derive raster masks and scalar confidence fields from it in DPQG and CGPF.

3.2 Dual-Prior Query Generator (DPQG)

The confidence estimator identifies unreliable regions in the prior map, but an imperfect prior still contains valuable topological structure, such as lane connectivity and divider layout. The Dual-Prior Query Generator (DPQG) extracts this structure and converts it into stable queries while avoiding regions that PACE marks as unreliable.

Vectorized map embedding. The biased vectorized map M_{err} contains N instances, each represented as a sequence of points. For the n-th point p_n, we denote its BEV coordinate by $\mathbf{g}_n = (x_n, y_n)$, its semantic type by c_n, the index of the instance it belongs to by instance i_n, and its order in that instance by s_n. We map $\mathbf{g}_n$ to the nearest raster grid and denote this grid location by u_n. Each point p_n is converted into a high-dimensional embedding

$$\mathbf{e}_n = E_{\text{inst}}(i_n) + E_{\text{point}}(s_n) + \text{PE}(u_n) + E_{\text{type}}(c_n), \tag{4}$$

where E_{inst} and E_{point} are learnable instance and order embeddings, PE is a 2D positional embedding shared with the raster branch as in HRMapNet, and E_{type} assigns a semantic code to each vector category.

Following PriorDrive [41], we apply hierarchical attention over these point embeddings. Intra-instance attention operates within each instance to capture spatial geometry, and inter-instance attention operates across instances to model

global topology. The output is a tensor $F_p^{\text{emb}} \in \mathbb{R}^{N\times M\times C}$, where M is the maximum number of points per instance (with padding when needed) and C is the feature dimension.

Confidence-guided masking. Standard encoders treat all points as equally reliable, which allows errors in the prior to propagate. DPQG instead uses the confidence map M_{conf} predicted by PACE. For each grid u, PACE predicts a confidence map $M_{\text{conf}} \in \mathbb{R}^{H\times W\times 3}$ with three channels for reliable, erroneous, and empty. After a channel-wise softmax, we denote these probabilities by $m_{\text{rel}}(u)$, $m_{\text{err}}(u)$, and $m_{\text{empty}}(u)$. We first compute an occupancy score

$$o(u) = 1 - m_{\text{empty}}(u), \qquad o \in \mathbb{R}^{H\times W},\ 0 \le o \le 1. \tag{5}$$

A grid is treated as a valid prior grid only when the occupancy is above a fixed threshold and the reliable probability exceeds the erroneous one, which yields a binary raster mask

$$C_{\text{mask}}^{R}(u) = \begin{cases} 1, & o(u) > \tau \text{ and } m_{\text{rel}}(u) > m_{\text{err}}(u), \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Grids that are empty or strongly biased are thus removed from the prior.

We map this raster mask to vector points through their grid locations. Each point coordinate $\mathbf{g}_n$ uses the mask value of its grid u_n, giving a grid-to-instance mask $C_{\text{mask}} \in \{0,1\}^{N\times M}$. Let i index instances, m index points within an instance, and c index channels. After the attention encoder we apply this mask to the point features by

$$\hat{F}_p^{\text{emb}}(i,m) = C_{\text{mask}}(i,m)\, F_p^{\text{emb}}(i,m). \tag{7}$$

So that points in unreliable or empty regions are zeroed out, while points in valid prior regions keep their features. Flattening along the instance and point dimensions yields a topology-aware vector feature matrix $F_{\text{vec}} \in \mathbb{R}^{N_{\text{vec}}\times C}$, where $N_{\text{vec}} = N \times M$, which serves as vectorized prior tokens.

Dual-prior fusion and queries. DPQG also uses a rasterized prior branch so that the decoder can access both vectorized and rasterized evidence. Following HRMapNet [42], we embed the rasterized prior into a feature map $F_{\text{Ras}} \in \mathbb{R}^{H\times W\times C}$ using map prior embedding. It has position and label embeddings at each vaild grid u, and $l(u)$ is label of grid u:

$$F_{\text{Ras}}(u) = \text{PE}(u) + \text{LE}(l(u)). \tag{8}$$

Here $\text{PE}(u)$ shares the positional function with the vector branch and keeps both in the same BEV frame, while $\text{LE}(u)$ is a per-class label embedding on raster grids, analogous to E_{type} for vector categories. The vaild grid is also obtained from C_{mask}^{R}. So only confident prior regions contribute to the raster feature map. We reshape F_{Ras} into a set of raster tokens named map query and concatenate

them with the vector tokens named instance query to form a fused prior feature $F_{\text{fuse}} \in \mathbb{R}^{N_{\text{fuse}} \times C}$, where N_{fuse} is the total number of tokens from both sources.

Finally, we update the learnable base queries Q_{init} by cross attention over this fused prior feature,

$$Q_{\text{map}} = \text{CrossAttention}\big(Q_{\text{init}},\, F_{\text{fuse}},\, F_{\text{fuse}}\big), \tag{9}$$

obtaining prior-aware queries Q_{map} that encode both vector topology and raster layout while being restricted to confident prior regions. These queries are then fed to the map decoder, which can attend to informative prior structures while staying robust to biased parts of the prior map.

3.3 Confidence-Guided Prior Fusion (CGPF)

While DPQG focuses on vector queries that capture topology, nearby grids on the BEV grid also provide strong cues. Errors in the prior often appear as small gaps, offsets, or short spatial shifts. The Confidence-Guided Prior Fusion (CGPF) module uses the confidence map to enlarge reliable regions in a controlled way and to decide how prior maps and visual features contribute to the final BEV context as show in Fig. 2a.

We operate on three BEV aligned inputs. The BEV feature map from the image backbone is $F_{\text{BEV}} \in \mathbb{R}^{H \times W \times C}$. Multi-view perspective features are processed and yields another BEV aligned feature map $F_{\text{PV}} \in \mathbb{R}^{H \times W \times C}$. The biased rasterized prior is $M_{\text{Rerr}} \in \mathbb{R}^{H \times W \times 3}$.

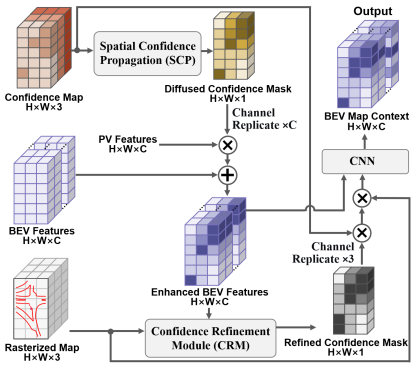
Spatial Confidence Propagation (SCP). SCP builds a diffused confidence mask through a dilation-like operation on the BEV grid. We first ignore grids that are likely to be empty and retain only those where m_{empty} is below a fixed threshold. On these grids, we measure the fraction of reliable mass among all non-empty states,

$$r(u) = \begin{cases} \dfrac{m_{\text{rel}}(u)}{m_{\text{rel}}(u) + m_{\text{err}}(u) + \varepsilon}, & m_{\text{empty}}(u) < \tau_{\text{empty}}, \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

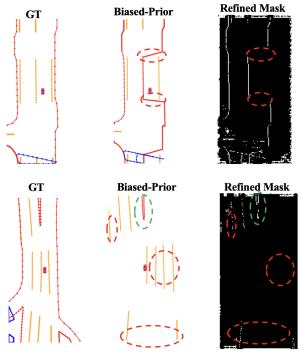
where τ_{empty} is a fixed empty threshold and ε is a small constant for numerical stability. The scalar field $r \in \mathbb{R}^{H \times W}$, with $0 \le r \le 1$, summarizes the degree of correctness conditioned on the prior being non-empty.

To capture different reliability levels, we split r into T confidence bands. Each band is described by a selector $S_\ell(r) \in \mathbb{R}^{H \times W}$, with $0 \le S_\ell(r) \le 1$, which highlights grids with similar reliability for $\ell = 1, \ldots, T$. For each band, we apply a diffusion operator D_{κ_ℓ} with kernel parameter κ_ℓ and assign a non-negative weight w_ℓ. The diffused confidence mask is

$$D_m(u) = \sum_{\ell=1}^{T} w_\ell \left[D_{\kappa_\ell}(S_\ell(r)) \right](u). \tag{11}$$



(a) The architecture of CGPF



(b) Illustration of the binarized refined confidence mask.

Fig. 2: (a) Architecture of the Confidence-Guided Prior Fusion (CGPF) module. It contains a Spatial Confidence Propagation (SCP) block that produces a diffused confidence mask and a Confidence Refinement Module (CRM) that refines prior weights and outputs the BEV map context. (b) Illustration of the binarized refined confidence mask. For clarity, it is presented as a binary map where white indicates reliable regions of the prior and black indicates erroneous ones. Compared to the Ground Truth (GT), erroneous regions in the Biased Prior (red dashed circles) are excluded as black in the map, while reliable regions (green dashed circles) are preserved as white.

Here, D_{κ_ℓ} is implemented as convolution with a compact averaging kernel. Higher reliability bands use tighter kernels and larger weights, whereas lower reliability bands use wider kernels to enlarge the spatial search range and smaller weights for conservative propagation. Thus, the lower magnitude of D_m in unreliable regions is a consequence of conservative weighting rather than the purpose of using wider kernels. This soft morphological diffusion enlarges reliable support around the prior and bridges small gaps while retaining a conservative shape.

The diffused mask D_m then controls how multi-view evidence is injected into the BEV representation. We broadcast D_m along the channel dimension and compute

$$F_{\text{enh}} = F_{\text{BEV}} + D_m \odot F_{\text{PV}}, \qquad F_{\text{enh}} \in \mathbb{R}^{H\times W\times C}, \tag{12}$$

where $\odot$ denotes element-wise multiplication. Importantly, D_m modulates only the additional F_{PV} branch, while F_{BEV} remains unscaled. A small $D_m(u)$ reduces the additional support from $F_{\text{PV}}(u)$, whereas a large $D_m(u)$ injects more support; in both cases, the online BEV evidence is preserved.

Confidence Refinement Module (CRM). SCP propagates spatial reliability, but it does not determine whether the prior content agrees with the visual evidence. The Confidence Refinement Module (CRM) refines the prior weights by evaluating this agreement.

CRM takes the enhanced BEV feature F_{enh} and the rasterized prior M_{Rerr} as input. We concatenate them along the channel dimension and apply a shallow convolutional network ϕ, followed by a sigmoid function, to produce a

single-channel weight:

$$G(u) = \sigma\big(\phi([\,F_{\text{enh}},\, M_{\text{Rerr}}\,])(u)\big), \qquad G \in \mathbb{R}^{H\times W}. \tag{13}$$

Unlike the early, grid-wise reliability estimate m_{rel} used by SCP, G is a late-stage fusion weight that incorporates local spatial context through ϕ. It is high where the prior geometry is consistent with the enhanced BEV feature and low where they disagree. We combine it with $m_{\text{rel}}(u)$ to form the refined confidence mask

$$\beta(u) = m_{\text{rel}}(u)\, G(u), \qquad \beta \in \mathbb{R}^{H\times W}. \tag{14}$$

Grids that are reliable but inconsistent with the images are downweighted, whereas grids that are both reliable and consistent retain large weights.

Finally, the refined mask shown in Fig. 2b reweights only the rasterized prior branch. We broadcast β along the channel dimension, concatenate the weighted prior $\beta \odot M_{\text{Rerr}}$ with the unscaled F_{enh}, and fuse them using a shallow CNN to obtain the BEV map context $F_{\text{ctx}} \in \mathbb{R}^{H\times W\times C}$. This map context serves as the key and value for the map decoder. Consequently, low-confidence or conflicting regions reduce only the contribution of the prior map, while the online visual evidence in F_{enh} remains preserved. In this way, CGPF provides a confidence-aware BEV representation that can be readily integrated into standard decoders.

4 Experiments

4.1 Datasets

Base datasets. We evaluate on the nuScenes [2] benchmarks for online vectorized map construction. nuScenes contains 1 000 urban driving scenes (about 20 s each) recorded at 2 Hz with six surrounding RGB cameras that cover a full 360° field of view. we use the official split of scenes 700/150/150 for training, validation, and testing.

Biased maps. To train models that are robust to biased priors, we consider both real and synthetic biased maps. Although Argoverse2 [33] includes a small set of biased maps, their scale is insufficient for our study, so we mainly rely on synthetic priors. Among synthetic options, MapEX [27] and Syn-D-Maps [37] target nuScenes. MapEX defines several error categories (e.g., missing content, global translation, local warping) but does not model explicit topological changes and is not publicly available. Syn-D-Maps, in contrast, is an open-source pipeline that supports topology-altering operations such as lane widening/narrowing and lane addition/reduction. Following its procedure on nuScenes, we generate four topology-altering biased types (lane widening, lane narrowing, lane addition, lane reduction) along with rotation biased type, yielding one biased prior per types for every frame.

4.2 Implementation Details

Metric. For the nuScenes dataset, we use the official split of scenes. Following previous work, we evaluate on validation set and report performance on three categories. We report class-wise AP and mAP following the official nuScenes protocol, which matches polylines using Chamfer-distance thresholds of $\{0.5, 1.0, 1.5\}$ m. The prediction map range is $[-15, 15]$ m along the X-axis and $[-30, 30]$ m along the Y-axis.

Generate Biased Maps. We adopt Syn-D-Maps [37] as our corruption framework to introduce five types of map bias. These consist of four topology-altering biases (lane addition, lane reduction, lane widening, and lane narrowing) and a rotation bias. When they modify maps, all related instances (e.g. road boundaries) are modified accordingly to preserve topological consistency. In accordance with the Syn-D-Maps setup, we applied a corruption ratio of 0.6 for four topology-altering biased types, defined as the fraction of total lanes subject to modification. For rotation biased type, it controls bias by varying the rotation angles. We also follow the setup from Syn-D-Maps. During training, each sample is paired with one biased types drawn uniformly from four topology-altering biased types. The rotation biased type is used in ablation study.

Model. We adopt a ResNet-50 backbone with a total batch size of 32. The base learning rate is set to 6×10^{-4} and we use a learning rate of 1.2×10^{-3} for the U-Net branch. We follow the loss function from MapTRv2 [16], which combines classification, point-to-point, and edge direction losses, along with auxiliary one-to-many loss [10], dense prediction, and depth losses. We add Focal Loss [18] to train the U-Net.

In training, we set $\alpha = 0.5$, $N = 35$ and $M = 20$. All retrained models (ours and † baselines) are trained for 24 epochs on eight NVIDIA A100 GPUs and the remaining training settings follow HRMapNet. For the three baseline methods that rely on priors or temporal history, we keep their original training protocols and only replace the original priors or historical maps with the biased priors. All models share the same data mix strategy described in the Methods section, and during evaluation they access only the biased maps.

4.3 Main Results

In Tab. 1, we summarizes the comparison on nuScenes. The first block contains methods that do not use priors or temporal history and rely only on current sensor observations. These models already achieve strong performance with mAP up to 72.6. In contrast, the last block evaluates models that receive only biased priors at test time under our mixed bias setting. PriorMaskMap reaches 81.4 mAP in this regime and still outperforms all no-prior baselines, which shows that the network can extract useful structural cues even when the prior is systematically corrupted.

Table 1: Comparison with state-of-the-art methods on the nuScenes dataset. The first block lists methods that do not rely on priors or history for map construction, while the second block lists methods that leverage various priors. The third block lists models retrained and evaluated only with biased priors under our mixed-bias setting. * Err. denotes that, during evaluation, only biased prior maps are used as inputs. † The results are trained with our mixed data from scratch by replacing history or prior maps.

Methods	Backbone	Epoch	Prior	Err.*	AP_{ped}	AP_{div}	AP_{boun}	mAP
StreamMapNet [39] WACV24	R50	30	–	–	56.9	55.9	61.4	58.1
MGMap [19] CVPR24	R50	24	–	–	61.8	65.0	67.5	64.8
HIMap [43] CVPR24	R50	30	–	–	62.6	68.4	69.1	66.7
MapQR [21] ECCV24	R50	24	–	–	63.4	68.0	67.7	66.4
Mask2Map [4] ECCV24	R50	24	–	–	70.6	71.3	72.9	71.6
MapTRv2 [16] IJCV25	R50	24	–	–	59.8	62.4	62.4	61.5
DAMap [6] ICCV25	R50	24	–	–	70.7	72.3	74.9	72.6
SQD-MapNet [30] ECCV24	R50	24	✓	–	63.0	62.5	63.3	63.9
P-MapNet [11] RAL24	EN-B0	10	✓	–	41.3	54.2	63.7	53.1
MapTracker [3] ECCV24	R50	72	✓	–	80.0	74.1	74.1	76.1
PriorMapNet [29] arXiv24	R50	24	✓	–	64.0	69.0	68.2	67.1
PriorDrive [41] arXiv24	R50	24	✓	–	63.6	67.3	66.2	65.7
HRMapNet [42] ECCV24	R50	24	✓	–	65.8	67.4	68.5	67.2
PrevPredMap [23] WACV25	R50	24	✓	–	68.7	66.0	68.3	67.6
UniPrevPred [22] arXiv25	R50	24	✓	–	76.2	72.3	73.6	74.0
PriorFusion [28] arXiv25	R50	110	✓	–	68.3	72.0	70.9	70.4
SQD-MapNet†	R50	24	–	✓	55.4	53.3	57.9	55.5
PrevPredMap†	R50	24	–	✓	63.5	62.9	66.4	64.3
HRMapNet†	R50	24	–	✓	**78.5**	82.0	78.5	79.7
PriorMaskMap (ours)	R50	24	–	✓	78.1	**85.1**	**80.9**	**81.4**
HRMapNet†	R50	110	–	✓	**81.8**	85.0	83.3	83.4
PriorMaskMap (ours)	R50	110	–	✓	81.7	**88.9**	**84.5**	**85.0**

The second block reports approaches that are designed for clean priors or propagated maps and are evaluated in their original configurations with mAP around 70 to 76. To place prior-based models in the same biased regime we retrain SQD-MapNet†, PrevPredMap† and HRMapNet† with the mixed data which gives the results in the last block. Among these baselines HRMapNet† is the strongest with 79.7 mAP while PriorMaskMap further improves mAP to 81.4 without access to any clean prior or temporal history. Additionally, we compare our model against HRMapNet† over a longer training duration. Even at epoch 110, our approach continues to sustain an advantage of 1.6 mAP. This suggests that explicitly modeling prior reliability recovers most of the benefit of a prior and even surpasses prior-based baselines once the maps are biased, some examples are shown in Fig. 3.

Impact of topology-altering biased types. We further analyze the impact of each topology-altering bias in Tab. 2. Compared with HRMapNet†, PriorMaskMap consistently achieves higher mAP across all four biased types, with modest gains for lane addition and reduction and noticeably larger improvements for lane

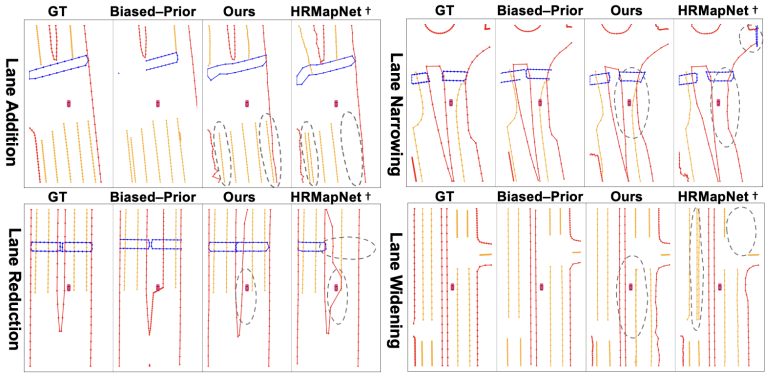


Fig. 3: Visualization comparing our proposed model against the HRMapNet†.

widening and narrowing. Most of the benefit comes from higher AP on lane dividers and road boundaries, while pedestrian-crossing AP stays within about ±1 point relative to HRMapNet†. Meanwhile, SQD-MapNet† and PrevPredMap† show almost unchanged performance across different biased types, which we attribute to their conservative use of prior information: when the bias becomes large, the models tend to largely ignore the prior. Overall, these results indicate that the proposed confidence-guided diffusion and topology-aware aggregation can effectively recover missing or distorted topology under diverse structure-changing corruptions instead of overfitting to a specific biased pattern.

Result of spatially non-overlapping splits Recent work suggests that the official nuScenes split is suboptimal for the online vectorized map construction task, as overlapping geographical regions inherently lead to model overfitting. Therefore, to more comprehensively evaluate our model's performance in real-world scenarios, we conduct experiments on two widely adopted alternative splits proposed by StreamMapNet [39] and the Localization Is All You Evaluate [17] (Geo-split) in Tab. 3. While a performance decline is observed across the board, our model continues to outperform HRMapNet†. Specifically, we achieve a 3.4 mAP advantage under the Geo-split, which further expands to a 4.9 mAP margin on the StreamMapNet split. Overall, these results demonstrate our model's robust generalization capabilities and consistent superiority in unseen environments.

4.4 Ablation study

We analyzes the contribution of each component in Tab. 4. Compared to topology-altering biased types, the rotation biased type shifts all map elements from their original positions, thereby better highlighting the performance gaps between different component. The result shows that removing vectorized prior in DPQG while keeping SCP and CRM leads to the largest degradation, reducing mAP from 76.7 to 74.8, which shows that topology-aware prior queries are

Table 2: Performance comparison on the nuScenes dataset with mixed bias prior. † as in Tab. 1 and the first two results are not typos but follow from their respective processing mechanisms. Bold numbers indicate the best mAP for each biased type.

Method	Biased Type	Epoch	AP_{ped}	AP_{div}	AP_{boun}	mAP
SQD-MapNet†	Addition	24	55.4	53.3	57.9	55.5
	Reduction		55.4	53.3	57.9	55.5
	Widening		55.4	53.4	57.9	55.6
	Narrowing		55.4	53.3	57.9	55.5
PrevPredMap†	Addition	24	63.1	60.6	65.9	63.2
	Reduction		63.1	60.6	65.9	63.2
	Widening		63.1	60.6	65.9	63.2
	Narrowing		63.1	60.6	65.9	63.2
HRMapNet†	Addition	24	68.2	77.6	68.9	71.6
	Reduction		85.4	93.8	87.6	88.9
	Widening		76.7	75.1	74.2	75.3
	Narrowing		77.7	75.6	75.2	76.2
Ours	Addition	24	67.3	79.9	69.5	**72.2**
	Reduction		86.3	94.6	87.3	**89.4**
	Widening		76.3	81.3	79.5	**79.0**
	Narrowing		76.7	80.9	78.1	**78.9**

Table 3: Quantitative comparison with the baseline on StreamMapNet and Geo-split datasets. † as in Tab. 1. Best results are highlighted in **bold**. Both models are trained for 24 epochs.

Method	StreamMapNet				Geo-split			
	AP_{ped}	AP_{div}	AP_{boun}	mAP	AP_{ped}	AP_{div}	AP_{boun}	mAP
HRMapNet†	38.0	70.0	71.3	59.7	51.9	65.7	64.7	60.8
Ours	40.2	78.7	74.8	**64.6**(+4.9)	49.9	75.5	67.3	**64.2**(+3.4)

crucial for exploiting biased maps. However, retaining only SCP and CRM still yielded a 3.4 mAP improvement. This clearly demonstrates the effectiveness of both modules.

Starting from the DPQG-only configuration, enabling either CRM or SCP alone brings consistent improvements of around 0.6-0.7 mAP and slightly boosts both divider and boundary AP. Combining all three modules yields the best performance with 76.7 mAP, indicating that SCP and CRM provide complementary spatial diffusion and local refinement on top of DPQG. Overall, DPQG supplies informative prior tokens, while SCP and CRM further regulate where and how these priors influence the BEV representation.

5 Conclusion

This work presents PriorMaskMap, a robust framework for online vectorized map construction that enhances mapping precision by selectively leveraging biased prior maps. Our approach is built upon a core Prior-Aware Confidence

Table 4: Ablation study of DPQG, SCP, and CRM.

Vectorized Prior	SCP	CRM	AP_{ped}	AP_{div}	AP_{boun}	mAP
✗	✗	✗	74.5	69.3	70.3	71.4
✓	✗	✗	79.1	74.4	73.9	75.8 (+4.4)
✓	✗	✓	80.2	73.5	75.9	76.5
✓	✓	✗	79.7	74.9	74.5	76.4
✗	✓	✓	80.5	72.6	71.6	74.8 (+3.4)
✓	✓	✓	79.6	75.3	75.1	**76.7** (+5.3)

Estimator (PACE) that identifies trustworthy regions in prior maps and produces a grid-wise confidence map. This output directs the subsequent generation of map element queries and feature fusion. Specifically, the Dual-Prior Query Generator (DPQG) creates query embeddings through separate rasterized and vectorized prior branches, while the Confidence-Guided Prior Fusion (CGPF) spatially propagates confidence scores to regulate BEV-prior fusion.

To support evaluation, we extend the nuScenes dataset with synthesized biased prior maps that simulate diverse and realistic structural deviations. The experiments demonstrate that PriorMaskMap achieves superior accuracy while maintaining remarkable robustness under significantly inaccurate priors. Our future work will explore temporal fusion of sequential predictions and updating offline priors with online mapping results to further improve performance.

Acknowledgement.

This work was supported by Beijing Natural Science Foundation (L243008).

References

1. Bateman, S.M., Xu, N., Zhao, H.C., Ben Shalom, Y., Gong, V., Long, G., Maddern, W.: Exploring real world map change generalization of prior-informed hd map prediction models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4568–4578 (2024)
2. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11621–11631 (2020)
3. Chen, J., Wu, Y., Tan, J., Ma, H., Furukawa, Y.: Maptracker: Tracking with strided memory fusion for consistent vector hd mapping. In: European Conference on Computer Vision. pp. 90–107. Springer (2024)
4. Choi, S., Kim, J., Shin, H., Choi, J.W.: Mask2map: Vectorized hd map construction using bird's eye view segmentation masks. In: European Conference on Computer Vision. pp. 19–36. Springer (2024)
5. Ding, W., Qiao, L., Qiu, X., Zhang, C.: Pivotnet: Vectorized pivot learning for end-to-end hd map construction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3672–3682 (2023)

6. Dong, J., Li, C., Lin, Y., Fu, J., Zhou, S., Zheng, N.: Damap: Distance-aware mapnet for high quality hd map construction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5285–5294 (2025)
7. Du, Y., Yang, S., Wang, L., Hou, Z., Cai, C., Tan, Z., Chen, M., Huang, S.S., Li, Q.: Rtmap: Real-time recursive mapping with change detection and localization. arXiv preprint arXiv:2507.00980 (2025)
8. Hu, H., Wang, F., Wang, Y., Hu, L., Xu, J., Zhang, Z.: Admap: Anti-disturbance framework for reconstructing online vectorized hd map. arXiv preprint arXiv:2401.13172 (2024)
9. Ji, Y., Li, Z., Lu, T.: Imaginemap: Enhanced hd map construction with sd maps. arXiv preprint arXiv:2412.16938 (2024)
10. Jia, D., Yuan, Y., He, H., Wu, X., Yu, H., Lin, W., Sun, L., Zhang, C., Hu, H.: Detrs with hybrid matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 19702–19712 (2023)
11. Jiang, Z., Zhu, Z., Li, P., Gao, H.a., Yuan, T., Shi, Y., Zhao, H., Zhao, H.: P-mapnet: Far-seeing map generator enhanced by both sdmap and hdmap priors. IEEE Robotics and Automation Letters (2024)
12. Lambert, J., Hays, J.: Trust, but verify: Cross-modality fusion for HD map change detection. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2) (2021)
13. Li, Q., Wang, Y., Wang, Y., Zhao, H.: Hdmapnet: An online hd map construction and evaluation framework. In: 2022 International Conference on Robotics and Automation (ICRA). pp. 4628–4634. IEEE (2022)
14. Li, T.: Mapnext: Revisiting training and scaling practices for online vectorized hd map construction. arXiv preprint arXiv:2401.07323 (2024)
15. Liao, B., Chen, S., Wang, X., Cheng, T., Zhang, Q., Liu, W., Huang, C.: Maptr: Structured modeling and learning for online vectorized hd map construction. In: International Conference on Learning Representations (2023)
16. Liao, B., Chen, S., Zhang, Y., Jiang, B., Zhang, Q., Liu, W., Huang, C., Wang, X.: Maptrv2: An end-to-end framework for online vectorized hd map construction. International Journal of Computer Vision **133**(3), 1352–1374 (2025)
17. Lilja, A., Fu, J., Stenborg, E., Hammarstrand, L.: 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. pp. 22150–22159 (2024)
18. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)
19. Liu, X., Wang, S., Li, W., Yang, R., Chen, J., Zhu, J.: Mgmap: Mask-guided learning for online vectorized hd map construction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14812–14821 (2024)
20. Liu, Y., Yuan, T., Wang, Y., Wang, Y., Zhao, H.: Vectormapnet: End-to-end vectorized hd map learning. In: International Conference on Machine Learning. pp. 22352–22369. PMLR (2023)
21. Liu, Z., Zhang, X., Liu, G., Zhao, J., Xu, N.: Leveraging enhanced queries of point sets for vectorized map construction. In: European Conference on Computer Vision. pp. 461–477. Springer (2024)
22. Peng, N., Zhou, X., Wang, M., Chen, G., Xu, W.: Uni-prevpredmap: Extending prevpredmap to a unified framework of prior-informed modeling for online vectorized hd map construction. arXiv preprint arXiv:2504.06647 (2025)

23. Peng, N., Zhou, X., Wang, M., Yang, X., Chen, S., Chen, G.: Prevpredmap: Exploring temporal modeling with previous predictions for online vectorized hd map construction. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 8134–8143. IEEE (2025)
24. Plachetka, C., Maier, N., Fricke, J., Termöhlen, J.A., Fingscheidt, T.: Terminology and analysis of map deviations in urban domains: Towards dependability for hd maps in automated vehicles. In: 2020 IEEE Intelligent Vehicles Symposium (IV). pp. 63–70. IEEE (2020)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
26. Song, J., Chen, X., Lu, L., Li, J., Skinner, K.A.: Memfusionmap: Working memory fusion for online vectorized hd map construction. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 9248–9257. IEEE (2025)
27. Sun, R., Yang, L., Lingrand, D., Precioso, F.: Mind the map! accounting for existing maps when estimating online hdmaps from sensors. In: Proceedings of the Winter Conference on Applications of Computer Vision (WACV). pp. 1671–1681 (2025)
28. Tang, X., Yang, M., Wen, T., Jia, P., Cui, L., Luo, M., Sheng, K., Zhang, B., Jiang, K., Yang, D.: Priorfusion: Unified integration of priors for robust road perception in autonomous driving. Communications in Transportation Research **5**, 100229 (2025)
29. Wang, R., Lu, X., Liu, X., Zou, X., Cao, T., Li, Y.: Priormapnet: Enhancing online vectorized hd map construction with priors. arXiv preprint arXiv:2408.08802 (2024)
30. Wang, S., Jia, F., Mao, W., Liu, Y., Zhao, Y., Chen, Z., Wang, T., Zhang, C., Zhang, X., Zhao, F.: Stream query denoising for vectorized hd-map construction. In: European Conference on Computer Vision. pp. 203–220. Springer (2024)
31. Wild, L., Ericson, L., Valencia, R., Jensfelt, P.: Exelmap: Explainable element-based hd-map change detection and update. arXiv preprint arXiv:2409.10178 (2024)
32. Wild, L., Valencia, R., Jensfelt, P.: Argotweak: Towards self-updating hd maps through structured priors. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6091–6100 (2025)
33. Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., Hartnett, A., Pontes, J.K., et al.: Argoverse 2: Next generation datasets for self-driving perception and forecasting. arXiv preprint arXiv:2301.00493 (2023)
34. Xie, H., Hu, X., Jiang, H., Dai, H., Dong, P., Wang, P.: Aerialhdmappper: High-definition map construction from aerial imagery. IEEE Transactions on Geoscience and Remote Sensing (2025)
35. Xiong, H., Shen, J., Zhu, T., Pan, Y.: Ean-mapnet: Efficient vectorized hd map construction with anchor neighborhoods. arXiv preprint arXiv:2402.18278 (2024)
36. Xiong, X., Liu, Y., Yuan, T., Wang, Y., Wang, Y., Zhao, H.: Neural map prior for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17535–17544 (2023)
37. Xu, H., Xiao, Y., Li, W., Hu, Y.: Generating synthetic deviation maps for prior-enhanced vectorized hd map construction. In: 2025 IEEE Intelligent Vehicles Symposium (IV). pp. 419–426 (2025)
38. Yang, J., Jiang, M., Yang, S., Tan, X., Li, Y., Ding, E., Wang, H., Wang, J.: Mgmapnet: Multi-granularity representation learning for end-to-end vectorized hd map construction. aarXiv preprint arXiv:2410.07733 (2024)

39. Yuan, T., Liu, Y., Wang, Y., Wang, Y., Zhao, H.: Streammapnet: Streaming mapping network for vectorized online hd map construction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7356–7365 (2024)
40. Yuan, T., Mao, Y., Yang, J., Liu, Y., Wang, Y., Zhao, H.: Presight: Enhancing autonomous vehicle perception with city-scale nerf priors. In: European Conference on Computer Vision. pp. 323–339. Springer (2024)
41. Zeng, S., Chang, X., Liu, X., Pan, Z., Wei, X.: Driving with prior maps: Unified vector prior encoding for autonomous vehicle mapping. arXiv preprint arXiv:2409.05352 (2024)
42. Zhang, X., Liu, G., Liu, Z., Xu, N., Liu, Y., Zhao, J.: Enhancing vectorized map perception with historical rasterized maps. In: European Conference on Computer Vision. pp. 422–439. Springer (2024)
43. Zhou, Y., Zhang, H., Yu, J., Yang, Y., Jung, S., Park, S.I., Yoo, B.: Himap: Hybrid representation learning for end-to-end vectorized hd map construction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15396–15406 (2024)