PriorDrive: Enhancing Online HD Mapping with Unified Vector Priors

Shuang Zeng[1,3*†], **Xinyuan Chang**[3†], **Xinran Liu**[3],
Yujian Yuan[4], **Shiyi Liang**[1,2], **Zheng Pan**[3], **Mu Xu**[3], **Xing Wei**[1,2‡]
[1]State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Xi'an Jiaotong University
[2]School of Software Engineering, Xi'an Jiaotong University
[3]Amap, Alibaba Group
[4]The Hong Kong University of Science and Technology
{zengshuang, sy_liang2023}@stu.xjtu.edu.cn, weixing@mail.xjtu.edu.cn, yyuanbn@connect.ust.hk
{changxinyuan.cxy, tom.lxr, panzheng.pan, xumu.xm}@alibaba-inc.com

Abstract

High-Definition Maps (HD maps) are essential for the precise navigation and decision-making of autonomous vehicles, yet their creation and upkeep present significant cost and timeliness challenges. The online construction of HD maps using on-board sensors has emerged as a promising solution; however, these methods can be impeded by incomplete data due to occlusions and inclement weather, while their performance in distant regions remains unsatisfying. This paper proposes PriorDrive to address these limitations by directly harnessing the power of various vectorized prior maps, significantly enhancing the robustness and accuracy of online HD map construction. Our approach integrates a variety of prior maps uniformly, such as OpenStreetMap's Standard Definition Maps (SD maps), outdated HD maps from vendors, and locally constructed maps from historical vehicle data. To effectively integrate such prior information into online mapping models, we introduce a Hybrid Prior Representation (HPQuery) that standardizes the representation of diverse map elements. We further propose a Unified Vector Encoder (UVE), which employs fused prior embedding and a dual encoding mechanism to encode vector data. To improve the UVE's generalizability and performance, we propose a segment-level and point-level pre-training strategy that enables the UVE to learn the prior distribution of vector data. Through extensive testing on the nuScenes, Argoverse 2 and OpenLane-V2, we demonstrate that PriorDrive is highly compatible with various online mapping models and substantially improves map prediction capabilities. The integration of prior maps through PriorDrive offers a robust solution to the challenges of single-perception data, paving the way for more reliable autonomous driving.

Introduction

High-Definition Maps (HD maps) are essential for the accurate navigation and decision-making of autonomous vehicles, providing detailed vectorized representations of road elements (Elghazaly et al. 2023; Gao et al. 2020; Yao et al. 2024). Despite their significance, the traditional methods for creating and maintaining HD maps are often costly and labor-intensive. These methods can result in outdated maps that struggle to keep pace with the rapidly changing urban environments. In response, there is increasing interest in online HD map construction (Qiao et al. 2023), where maps are generated in real-time using on-board sensors. Although this approach reduces costs, it also introduces new challenges, particularly concerning incomplete and error-prone data caused by environmental occlusions (see red circle in Figure 1) or inclement weather conditions, coupled with unsatisfactory performance in distant regions.

Using prior maps to alleviate the problem of online mapping is a promising solution. Specifically, there are three common types of prior maps: 1) Standard Definition Maps (SD maps), 2) existing offline outdated HD maps, and 3) online historical prediction local maps, as illustrated in the upper left corner of Figure 1. SD maps provide crucial long-distance centerline skeletons, but with uncertain accuracy compared to HD maps. While existing HD maps (HD map-EX) offer high accuracy, their infrequent updates may lead to outdated information that fails to fully reflect current road conditions. Historical predicted local maps can offer insights by incorporating previous observations, but being a single mapping result, they may not guarantee the completeness and accuracy of the prior map.

Generally, there are two ways to encode vectors: 1) rasterized, 2) vectorized. Recent studies (Xiong et al. 2023; Jiang et al. 2024) have attempted to use rasterized or vectorized prior maps to alleviate the problem of online mapping. However, existing prior-based methods (Xiong et al. 2023; Jiang et al. 2024; Li et al. 2024) face several limitations. For instance, P-MapNet (Jiang et al. 2024), NMP (Xiong et al. 2023) and HRMapNet (Zhang et al. 2024b) leverage rasterized SD maps or historical prediction maps. The rasterized representation, limited by resolution, is lossy and redundant while lacking the detail necessary to capture vectorized instance-level information and express the type and direction of map elements effectively. Furthermore, P-MapNet and HRMapNet require complex post-processing to convert

*Work done during the internship at Amap, Alibaba Group.
†Equal contribution.
‡Corresponding author: Xing Wei.

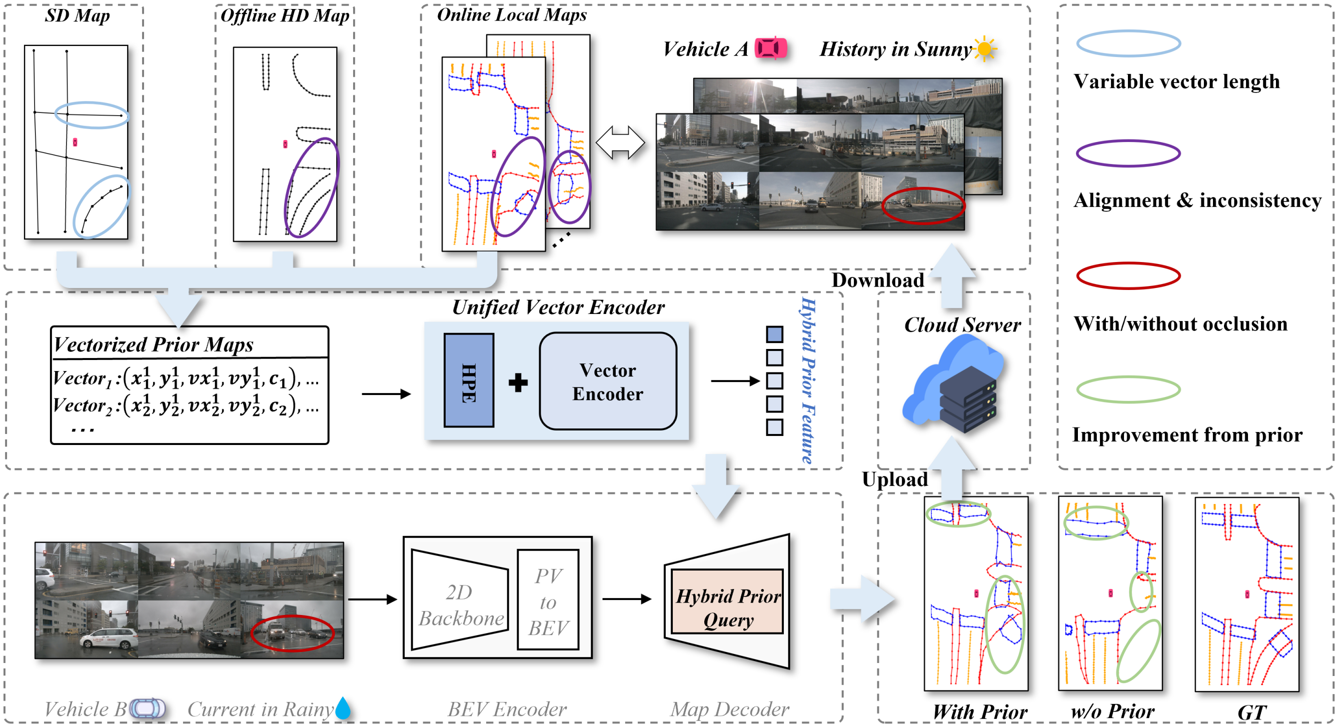

Figure 1: Overview of PriorDrive. PriorDrive seamlessly integrates diverse vectorized prior maps into existing online mapping frameworks, making final predictions more complete and accurate than those generated without priors. The optimized predictions can be uploaded to cloud servers for other vehicles to download and use as prior maps.

vectorized maps into rasterized maps, which also do not align with the native vector storage format of maps. On the other hand, MapEX (Sun et al. 2023a) exclusively relies on vectorized outdated HD maps, which struggle to reflect real-time changes in road structures, due to low update frequencies, making them unreliable for current navigation needs. Moreover, the encoding scheme of MapEX is overly simplistic, thereby failing to effectively capture instance-level information within vector representations.

It is worth mentioning that encoding vectors is not a straightforward task. There are several main challenges: 1) Different types of vector maps may contain varying vector types (e.g., points, lines, planes), 2) Vector lengths can be fixed or variable (see blue circle in Figure 1), and 3) Alignment and inconsistency issues exist between different vector maps (see purple circle in Figure 1). Previous methods were limited to efficiently encoding only a single type of prior map, but the complementary information contained across different prior map categories. For instance, NavMap (Schmidt et al. 2023) omitted road segments with variable lengths when integrating HD maps and SD maps, resulting in the loss of valuable information. Furthermore, previous methods generally encode only vector positions and categorical attributes, lacking the capability to capture finer-grained information such as direction, geometric shape, and topology.

In this paper, we overcome these challenges and introduce **PriorDrive**, a novel framework designed to directly and unifiedly integrate diverse vectorized prior maps into various mapping models, fully leveraging their respective strengths and complementary information. We propose a Unified Vector Encoder (UVE) that encodes various vector data and captures fine-grained vector information. It can extract fixed-length instance-level and point-level features from variable-length vectors. Additionally, we introduce for the first time a point-level and segment-level pre-training paradigm specifically tailored for vector data, aiming to enhance the model's understanding and representation of vector. Concurrently, we propose a hybrid prior representation (HPQuery) to represent all elements, facilitating the integration of unified vector features into mapping models. Experiments on nuScenes, Argoverse 2 and OpenLane-V2 demonstrate that PriorDrive is plug-and-play, seamlessly applicable to diverse online mapping models and vector tasks with significant performance improvements.

In summary, our contributions are as follows:

- We introduce a unified vector encoder that effectively captures diverse vector attributes and enables encoding of variable-length vectors into fixed-length features through fused prior embedding and dual encoding mechanism.
- We first proposed a point-level and segment-level pre-training paradigm for vector data, effectively capturing prior distribution and mitigating noise in historical maps.
- We propose a hybrid prior representation (HPQuery) to represent all elements and a PriorDrive framework with various vector prior maps to address the limitations of single-perception. Our comprehensive evaluation

demonstrates that PriorDrive is plug-and-play and significantly enhances the online mapping models.

Related Work

Online Vectorized HD Map Construction

Online vectorized HD map construction was initially treated as a segmentation task (Li et al. 2022; Zhang et al. 2024c; Sun et al. 2023b; Guo et al. 2025). To construct a vectorized HD map, HDMapNet (Li et al. 2021) requires complex post-processing of pixel-level rasterized maps. VectorMapNet (Liu et al. 2022b) introduced a coarse-to-fine, two-stage network leveraging keypoint representations. MapTR (Liao et al. 2023a) advanced this concept by modeling point sets in a permutation-equivalent manner using a DETR-like (Carion et al. 2020; Huang et al. 2022; Zeng et al. 2025b; Xiao et al. 2025) one-stage network. Subsequent studies (Zhou et al. 2024; Sun and Li 2025) have introduced various insightful approaches for further improvement. However, these methods rely solely on single-source perception data from vehicle-mounted sensors, limiting their effectiveness in challenging scenarios such as occlusion, inclement weather conditions or distant regions.

The release of OpenLane-V2 (Wang et al. 2023) dataset has sparked growing interest in topology reasoning within driving scenes. This task focuses on recognizing the topologies among lanes and lanes with traffic elements. TopoNet (Li et al. 2023) leverages GNN to update lane representation and lane topology. TopoMLP (Wu et al. 2024) utilizes PETR (Liu et al. 2022a) for centerline detection and employs simple MLPs to predict the relationships. In this paper, we also apply PriorDrive to topological inference tasks, demonstrating that PriorDrive is plug and play and widely applicable to other vector tasks.

Online Mapping Based on Prior Maps

Recent research has explored the integration of prior maps to enhance the performance of online mapping models. P-MapNet (Jiang et al. 2024) encodes Standard Definition Maps (SD map) as an additional conditional branch and employs a masked autoencoder to capture the prior distribution of HD maps. NMP (Xiong et al. 2023) introduces a global neural map prior that self-updates, thereby improving the performance of local map inference. MapEX (Sun et al. 2023a) utilizes existing HD maps and refines the query-based map estimation model's matching algorithm. Although these advancements have achieved considerable results, many studies have overlooked the potential of online historical prediction maps as a source of prior information. Moreover, previous methods (Zhang et al. 2024b; Wan et al. 2025; Yuan et al. 2025) typically could only utilize a single type of prior map. Our work can uniformly encode various types of vector prior maps to enhance current perception data, fully leveraging the complementary advantages among different prior maps.

Pretrained Methods Based on Mask Modeling

In the fields of NLP and CV, masked pre-training has proven to be an effective strategy for self-supervised representation learning. In NLP, models like BERT (Devlin et al. 2019; Zeng et al. 2025a; Li et al. 2025) use masked language modeling to predict randomly masked tokens within a bidirectional text context. Similarly, methods such as MAE (He et al. 2021; Liu et al. 2023; Liang et al. 2025; Dai et al. 2025) mask random patches of input images and reconstruct them based on the remaining unmasked patches. Diverging from these approaches, which are predominantly tailored for text and image data, we introduce a pre-training paradigm specifically designed for vector data. Our method aims to learn the prior distribution of vector representations, offering a new avenue for pre-training in the context of vectorized data.

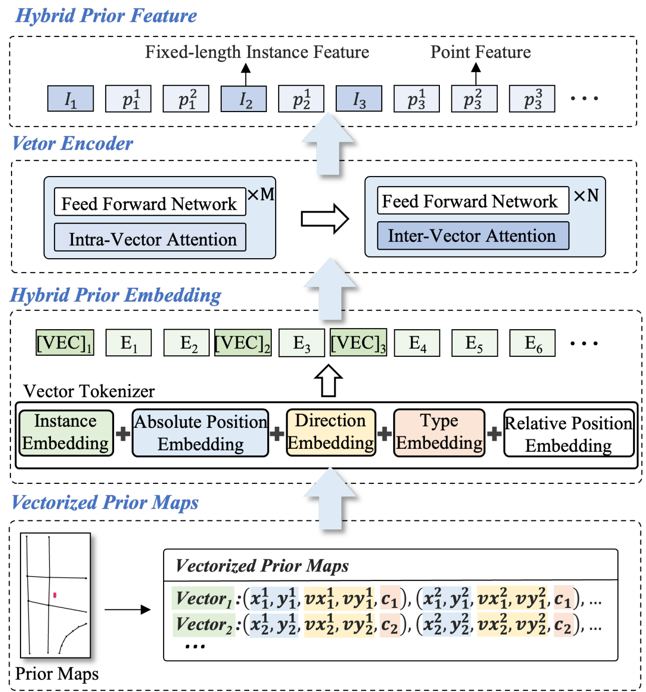

Figure 2: Structure of UVE. Fused prior embedding enhances vector representation. Intra and inter-instance attention refine local features and capture global context.

Proposed Method: PriorDrive

Problem Formulation

Formulation of Online Mapping. The goal of online mapping is to construct a local HD map using on-board sensor observations, such as the surrounding camera images. The online mapping process typically involves two main modules: the Bird's-Eye View (BEV) encoder and the map decoder. The BEV encoder generally includes a Perspective View (PV) feature extractor and a module for transforming PV features into BEV features. The map decoder usually comprises a Fully Convolutional Network (Shelhamer, Long, and Darrell 2015; Hu et al. 2024; Huang et al. 2025) or a query-based module similar to DETR (Carion et al. 2020; Xie et al. 2025; Tang et al. 2025), which outputs either rasterized perception results or vectorized map elements.

Let the set of images be denoted as $\{I_1, \dots, I_k\}$, and the

BEV encoder be represented as E_{bev}. The extraction of BEV features f_{bev} can then be expressed as:

$$f_{bev} = E_{bev}(\{I_1, \dots, I_K\}). \quad (1)$$

Let the map decoder be denoted as D_{map}. The final process of predicting map elements can be expressed as:

$$\mathcal{P} = D_{map}(f_{bev},\ opt(Q)), \quad (2)$$

where $\mathcal{P}$ represents the predicted map elements, Q denotes the learnable query, and *opt* indicates that this learnable query is optional, which is required in query-based models but not in other models.

Architecture of UVE

Inspired by BERT (Devlin et al. 2019; He and Hu 2025; Al Shafian, He, and Hu 2025) on text-related tasks, we propose unified vector encoder (UVE) by analogizing vector points to words and vector elements to sentences in Figure 2.

Extraction of Prior Map Features. Our proposed UVE serves as a unified vector encoder that can directly encode various vector data information, such as the position (x, y), direction (v_x, v_y), and type of points (c). Using $p = [x, y, v_x, v_y, c]$ to represent the point, a vector v typically consists of an ordered set of a variable number of points, $v = [p_1, p_2, \dots, p_n]$, where n varies. The map prior M consists of a set of vectors $M = \{v_1, v_2, \dots, v_m\}$, with m being variable. When employing different prior maps, all of prior maps were integrated into a cohesive whole, $M_{prior} = \{M_1, M_2, \dots, M_t\}$. Using E_{uve} to represent the UVE model, the extraction process of the features of the prior map, denoted as f_{prior}, can be represented as:

$$f_{prior} = E_{uve}(M_{prior}). \quad (3)$$

Fused Prior Embedding. As shown in Figure 2, for the given prior maps M_{prior}, we first construct the input for UVE, termed as Fused Prior Embedding (FPE). We use the method from (Tancik et al. 2020; Zeng et al. 2025c) for obtaining point position embedding and direction embedding for both (x, y) and (v_x, v_y). We concatenate these two embeddings as the point-level embedding. To obtain the fixed-length feature of the entire vector, we add a special *[VEC]* token at the beginning of each vector and use the embedding of this token as the instance-level embedding. Additionally, we introduce learnable instance embedding and type embedding to distinguish vector instances. To ensure the order of points, we also introduce 2D learnable position embedding. Finally, these embeddings are aggregated to form FPE.

Dual Encoding Mechanism. Due to the limited information content of vector points themselves, learning the concept of vector instances is very challenging, requiring more interactions among the points within the vectors to enhance each point's perceptual abilities towards others in the same vector. Therefore, UVE encoder utilizes M-layer intra-vector attention and N-layer inter-vector attention, using an attention masking mechanism to facilitate feature interactions within and between different vector instances. Each mask in the intra-vector encoder has a value of 1 at its own position and 0 elsewhere, while the inter-vector encoder employs a mask matrix filled entirely with 1s.

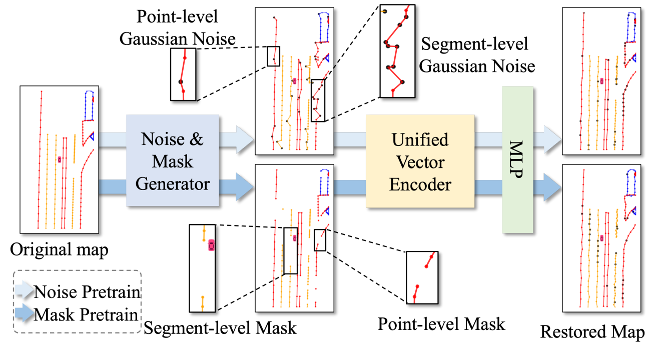

Figure 3: Pre-trained Pipeline. The noise & mask generator creates segment-level and point-level noise or mask, and then reconstructs the entire map via UVE and MLP.

When using different prior maps and multiple online local maps, there are alignment problems between different prior maps and potential inconsistencies between online local maps generated across different trips or different timestamps in one trip. Our UVE mitigated this challenge by employing fused prior embedding and dual encoding mechanism to enhance the model's ability to learn instances, supporting adaptive focus on the superior map elements in different prior maps. The experiments presented in Table 4 substantiate this capability.

Pre-training UVE: Position Modeling

Due to the limited inference ability of online mapping models, there are errors in historical prediction maps. So we used position modeling to pre-train UVE to improve its encoding and noise reduction capabilities. Specifically, there are two methods: random noise and masking (see Figure 3).

Noise & Mask Generator. The noise is primarily categorized into segment-level and point-level. In point-level noise, random noise is added to 5% of the vector points across the entire map. For segment-level noise, random noise is added to all vector points within a sub-vector segment, after randomly selecting 10% of the map elements. Using M_{org} to represent the original map elements (with the same data structure as M), the process of adding Gaussian noise can be represented by the following formula:

$$M^*_{org} = M_{org}\textit{[random\_index]} + \epsilon, \quad (4)$$

where $\epsilon \sim \mathcal{N}(0, 1)$ is a random noise, which is only added to the horizontal and vertical coordinates of each point.

Similar to adding noise, the selection of points for adding the mask is also divided into point-level and segment-level. After selecting the points to add the mask, the coordinates of these points will be masked as -1.

$$M_{org}\textit{[random\_index]} = Mask. \quad (5)$$

Loss Function. We feed vector map to UVE encoding after passing through the noise mask generator, then decode

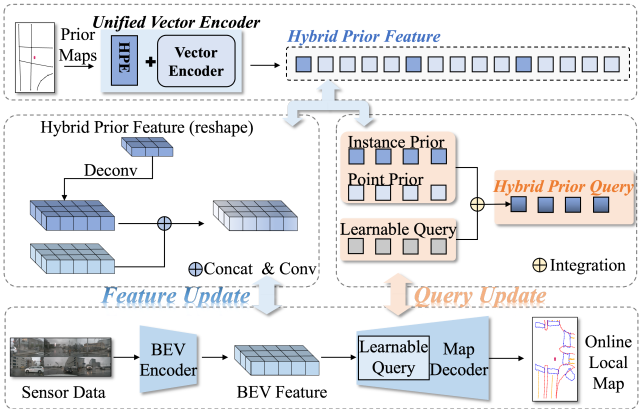

Figure 4: Our Integration Method. For query-based models, we enable interaction between the prior features and queries at both instance and point-level. For the other models, we integrate prior features directly with the BEV features.

the coordinates of all points using MLP, and calculate mean euclide distance from GT coordinates as supervision:

$$\mathcal{L} = RMSE(P,\ mlp(E_{uve}(M^*_{org}))), \tag{6}$$

where $P = \{(x_i, y_i)\}_{i=0}^{k}$ represent all the points in M_{org}, $RMSE$ on behalf of the root mean square error.

Integrated Approaches to BEV and HPQuery

We respectively proposed different integration methods for the query-based/non-query-based approaches to integrate the prior map features into the online map construction model. For methods without a learnable query Q, such as HDMapNet (Li et al. 2021), we directly reshape the prior feature f_{prior} and use $Deconv$ upsampling to match the shape of the BEV feature f_{bev}. Then connect it with the BEV feature f_{bev} and aligned by $Conv$, as shown in Figure 4 left.

$$f^*_{bev} = Conv(cat[f_{bev},\ Deconv(reshape(f_{prior}))]), \tag{7}$$

here, f^*_{bev} represents BEV feature after incorporating the prior map information, which can be input into the map decoder D_{map} to obtain the predicted results of map elements.

For methods based on learnable query Q, such as the vectorized model MapTR series (Liao et al. 2023b), we propose HPQuery of three operations: addition, replacement, and concatenation, as shown in Figure 4 right. Specifically, Q is typically composed of two parts, $Q = \{q_{ij}\}_{i=0,j=0}^{m,n} = \{q_i^{ins} + q_j^{pt}\}_{i=0,j=0}^{m,n}$, q_i^{ins} and q_i^{pt} representing the learnable queries at the instance-level and point-level respectively. The features f_{prior} of prior maps are also composed of instance-level features and point-level features, $f_{prior} = \{(f_i^{ins}, f_j^{pt})\}_{i=0,j=0}^{m',n'}$. So we interact prior features and queries at the instance-level and point-level, respectively. The HPQuery of addition, replacement, and concatenation operations can be expressed as:

$$q^*_{ij} = q_i^{ins}.add(f_i^{ins}) + q_j^{pt}.add(f_j^{pt}), \tag{8}$$

$$q^*_{ij} = q_i^{ins}.replace(f_i^{ins}) + q_j^{pt}.replace(f_j^{pt}), \tag{9}$$

$$q^*_{ij} = concat[q_i^{ins},\ f_i^{ins}] + concat[q_j^{pt},\ f_j^{pt}]. \tag{10}$$

We will detail the effectiveness of these feature fusion methods in supplementary material.

Experiments

Experimental Settings

Benchmark. We mainly validate our method on the popular nuScenes (Caesar et al. 2019) and Argoverse 2 (Wilson et al. 2021) following the previous methods (Zhang et al. 2024b; Zhou et al. 2025). NuScenes contains a total of 1,000 scenes, which was collected utilizing a 32-beam LiDAR operating and six cameras offering a 360-degree field of view. Argoverse 2 includes 1000 logs, which was captured from 7 cameras, along with a 3D vector map.

To demonstrate PriorDrive is plug-and-play and widely applicable to other vector tasks, we conducted topology inference on OpenLane-V2 (Wang et al. 2023) Subset A.

Evaluation Protocol. For vectorized results, we follow the standard metric used in previous works (Liao et al. 2023b; Li, Qiu, and Ke 2025; Li 2025). The perception range along the direction of vehicle travel is $[-15.0m, 15.0m]$ for X-axis and $[-30.0m, 30.0m]$ for Y-axis. We compute the average precision (AP) AP_τ under several Chamfer distance $D_{Chamfer}$ thresholds ($\tau \in T, T = \{0.5, 1.0, 1.5\}$), and then calculate mean across all thresholds as final metric:

$$\text{AP} = \frac{1}{|\text{T}|}\sum_{\tau \in \text{T}} \text{AP}_\tau. \tag{11}$$

For the rasterized results, we follow the previous works (Xiong et al. 2023; Ma et al. 2025) and employ IoU as the metric for segmentation results.

For OpenLane-V2's lane centerline perception task, DET_l averages the Frechet distance to quantify similarity. DET_t uses IoU and averages it across various traffic categories. TOP_{ll} and TOP_{lt} compute the topology matrix similarity between lanes and between lanes and traffic elements, respectively. The overall metric is called OLS:

$$\text{OLS} = \frac{1}{4}[\text{DET}_l + \text{DET}_t + f(\text{TOP}_{ll}) + f(\text{TOP}_{lt})], \tag{12}$$

where f is the square root function.

Implementation Details. For fair comparisons, experiments were conducted on all baseline models using their original hyperparameters. All experiments were trained with 8 NVIDIA RTX A6000. For pre-training UVE, we trained the 24-epoch on nuScenes vector data. This one-time offline pre-training (12h) yields significant improvements, and its cost is amortized through plug-and-play reuse across various models. The sources of the three types of prior maps are as follows: The SD map is from OpenStreetMap (Haklay and Weber 2008); as for the existing HD prior maps, we follow MapEX (Sun et al. 2023a). For each map element localization, we add noise from a Gaussian distribution with a standard deviation of 1 meter. This has the effect of applying a uniform translation to each map element; as for the predicted online local maps, in actual scenarios, the online local maps are generated by vehicles passing through the same area, and

Method	Prior Map	Backbone	Epoch	Metric	Ped.	Div.	Bou.	Mean
HDMapNet [ICRA21] (Li et al. 2021)	×	Effi-B0	30	IoU	18.7	40.6	39.5	32.9
P-MapNet [RAL24] (Jiang et al. 2024)	✓	Effi-B0	30	IoU	22.6	44.1	43.8	36.8
NMP [CVPR23] (Xiong et al. 2023)	✓	Effi-B0	30	IoU	21.0	44.2	46.1	37.1
PriorDrive (Ours)	✓	Effi-B0	30	IoU	**25.1**	**46.7**	**48.4**	**40.1**
MapTR [ICLR23] (Liao et al. 2023a)	×	R50	24	AP	41.2	49.5	51.1	47.3
P-MapNet [RAL24] (Jiang et al. 2024)	✓	R50	24	AP	43.7	50.9	53.5	49.4
StreamMapNet [WACV24] (Yuan et al. 2024)	Temporal	R50	24	AP	60.4	61.9	58.9	60.4
PriorDrive (Ours)	✓	R50	24	AP	**60.9**	**65.3**	**63.8**	**63.3**
MapTRv2 [IJCV24] (Liao et al. 2023b)	×	R50	24	AP	59.8	62.4	62.4	61.5
PrevPredMap [WACV25] (Peng et al. 2025)	✓	R50	24	AP	64.5	66.9	67.6	66.3
HRMapNet [ECCV24] (Zhang et al. 2024b)	✓	R50	24	AP	65.8	67.4	68.5	67.2
HRMapNet [ECCV24] (Zhang et al. 2024b)	✓	R50	110	AP	72.0	72.9	75.8	73.6
InteractionMap [CVPR25] (Wu, Yang, and Li 2025)	Temporal	R50	24	AP	69.7	72.7	73.0	71.8
InteractionMap [CVPR25] (Wu, Yang, and Li 2025)	Temporal	R50	110	AP	75.3	75.6	77.0	76.0
PriorDrive (Ours)	✓	R50	24	AP	**71.8**	**81.4**	**74.1**	**75.8**
PriorDrive (Ours)	✓	R50	110	AP	**78.0**	**82.0**	**81.6**	**80.5**

Table 1: Performance of various online mapping models on nuScenes. All methods use camera input. In each block, the first row represents the baseline, and other methods are modifications upon the baseline. "Temporal" means using temporal information. The results of StreamMapNet and HRMapNet are from HRMapNet, while the other results are from the respective papers.

Method	Epoch	AP_{ped}	AP_{div}	AP_{bou}	mAP
MapTRv2 [IJCV24]	6	60.7	68.9	64.5	64.7
HRMapNet [ECCV24]	30	65.1	71.4	68.6	68.3
PrevPredMap [WACV25]	6	64.1	71.4	67.4	67.6
InteractionMap [CVPR25]	6	66.6	75.6	72.7	71.6
PriorDrive (Ours)	6	**68.3**	**75.8**	**74.2**	**72.8**

Table 2: Results of 3D map elements on Argoverse 2.

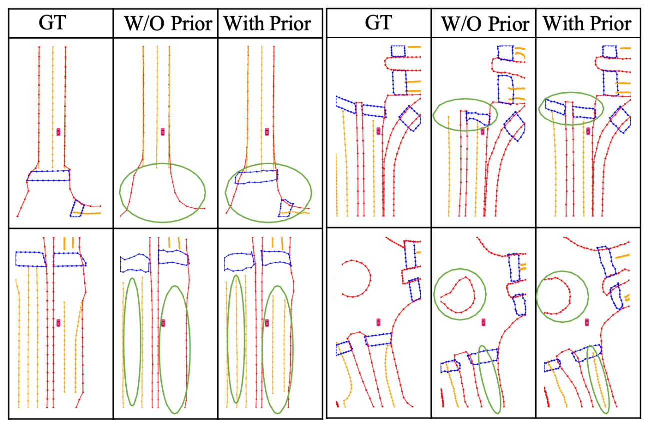

Figure 5: Qualitative results with unified prior maps.

their data is stored on cloud servers for other vehicles to retrieve in the future. For datasets collected by a single vehicle such as nuScenes, certain road sections are captured through multiple drives. Therefore, we use each baseline model to infer the entire dataset, thereby generating online local prior maps at different trips or different times in the same trip, simulating this actual process. Due to the different occlusion conditions of different trips, the online local prior map contains more abundant supplementary information for the current driving scene. So we first search the historical prediction of different trips, and then search the past time in the same trip as the local priors.

Main Results

Results on nuScenes. We evaluate our method across different model, evaluation metrics in Table 1. Our PriorDrive achieved a 7.2 mIoU improvement over baseline HDMapNet, which demonstrates that our vector prior remains effective even for rasterized models and metrics. In single-variable comparisons with prior/temporal-based methods employing MapTR as baseline, PriorDrive has demonstrated the most substantial improvement of 16 mAP. We further benchmarked our approach against the mainstream MapTRv2 baseline, which demonstrated an average performance gain greater than those achieved on HDMapNet, demonstrating that our HPQuery achieves more comprehensive and fine-grained exploitation of prior information. Furthermore, PriorDrive outperforms existing prior/temporal-based SOTA methods that adopt the MapTRv2 baseline, demonstrating the effectiveness of PriorDrive in fully utilizing complementary information from different prior maps. Figure 5 shows that using the prior maps can effectively recover the occluded elements, as well as more accurate mapping results. Overall, our method demonstrated consistent performance improvements across all baselines, highlighting its generalizability and potential applicability to other frameworks.

Results on Argoverse 2. In Table 2, we present the results of 3 dim maps on Argoverse 2. PriorDrive improves performance by 8.1 mAP and outperforms the previous prior/temporal-based SOTA methods, demonstrating excellent generalization performance.

Method	Prior Map	DET$_l$	DET$_t$	v1.0.0 TOP$_{ll}$	v1.0.0 TOP$_{lt}$	v1.0.0 OLS	v2.1.0 TOP$_{ll}$	v2.1.0 TOP$_{lt}$	v2.1.0 OLS
TopoNet + OSMR [IROS24] (Zhang et al. 2024a)	✓	30.6	44.6	7.7	22.9	37.7	-	-	-
RoadPainter [ECCV24] (Ma et al. 2024)	✓	**36.9**	47.1	12.7	25.8	42.6	-	-	-
SMERF [ICRA24] (Luo et al. 2024)	✓	33.4	**48.6**	7.5	23.4	39.4	15.4	25.4	42.9
TopoLogic [NeurIPS24] (Fu et al. 2024)	×	29.9	47.2	18.6	21.5	41.6	23.9	25.4	44.1
PriorDrive (Ours)	✓	31.5	48.4	**24.6**	**30.7**	**46.2**	**31.4**	**33.7**	**48.5**

Table 3: Performance comparison with SOTA non-prior/prior-based methods on OpenLane-V2 subset_A set.

Method	AP$_{ped}$	AP$_{div}$	AP$_{bou}$	mAP
PriorDrive	**71.8**	**81.4**	**74.1**	**75.8**
w/o UVE	67.3	74.0	67.8	69.7
w/o Pretrain	69.4	76.5	69.6	71.5
w/o SD Prior	69.5	77.1	72.2	72.9
w/o Online Local	68.4	73.9	69.5	70.6
w/o HD Prior	67.3	72.0	68.6	69.3

Table 4: The ablation of each component of PriorDrive.

Method	AP$_{ped}$	AP$_{div}$	AP$_{bou}$	mAP
VectorMapNet	15.8	17.0	21.2	18.0
MapTR	6.4	20.7	35.5	20.9
StreamMapNet	29.6	30.1	41.9	33.9
HRMapNet	36.9	30.3	44.0	37.1
MapTracker	**45.9**	30.0	45.1	40.3
PriorDrive (Ours)	44.7	**35.5**	**47.3**	**42.5**

Table 5: Comparisons on the new nuScenes split datasets.

Results on OpenLane-V2. We compared PriorDrive on OpenLane-V2 when using ResNet-50 for training 24 epochs in Table 3. Compared to TopoLogic baseline, our method achieves better topological reasoning accuracy. And our OLS reached 46.2 and 48.5, outperforming previous prior-based approaches. Overall, PriorDrive is a plug-and-play solution that can be widely applied to other vector tasks.

Ablation Study

we conducted ablation experiments on nuScenes based on MapTRv2 unless otherwise specified.

Ablation Study of PriorDrive. The ablation study in Table 4 investigates the contribution of each component. Specifically, replacing the UVE with separate MLPs for each prior degrades performance by 6.1 mAP, demonstrating UVE's effectiveness in mitigating the inconsistency and weak alignment among different priors. indicating it can also alleviate these issues and enhance generalization. The performance degradation observed when ablating any single prior highlights that they contain complementary information, which PriorDrive is the first to leverage simultaneously.

Comparison on New Split. The original split of nuScenes exhibits geographical overlap between training and validation sets. Therefore, we adopt new split proposed by StreamMapNet in Table 5, which training and validation data are separated in locations. PriorDrive achieves 42.5 mAP, which outperforms the previous SOTA methods.

Method	Range	AP$_{ped}$	AP$_{div}$	AP$_{bou}$	mAP
P-MapNet	120×60m	22.0	27.2	19.5	22.9
StreamMapNet	120×60m	37.2	42.3	30.2	36.6
HRMapNet	120×60m	43.1	47.7	34.9	41.9
PriorDrive (Ours)	120×60m	**45.0**	**49.5**	**37.3**	**43.9**
P-MapNet	240×120m	16.3	22.7	10.5	16.5
StreamMapNet	240×120m	22.4	31.3	16.0	23.3
HRMapNet	240×120m	**29.4**	34.2	19.7	27.8
PriorDrive (Ours)	240×120m	28.8	**36.3**	**21.6**	**28.9**

Table 6: Results of large perception ranges on nuScenes.

Comparison of Large Perception Ranges. Existing online mapping methods are constrained in prediction range due to the low resolution of distant areas in the captured images which leads to a significant decline in performance. However, PriorDrive leverages various prior maps to enhance accuracy in long-range predictions in Table 6.

Conclusion

In this paper, we introduced PriorDrive, a novel framework that effectively and uniformly leverages various types of prior maps to enhance the accuracy and robustness of online HD map construction. UVE, a new paradigm, unifies diverse priors via dual encoding mechanism and fused embeddings. We pioneered a pre-training paradigm for vector. PriorDrive is plug-and-play and scalable. Through comprehensive experiments, we demonstrated that PriorDrive significantly improves performance of SOTA models in various vector tasks. PriorDrive not only addresses the challenges associated with online HD map construction in complex environments but also offers a scalable solution that continually improves accuracy over time. The iterative use of historical maps as priors leads to progressively refined map outputs, providing latest information for autonomous driving.

Acknowledgments

This work was support by the National Natural Science Foundation of China No. 62572385, the Fundamental Research Funds for the Central Universities No. xxj032023020, and CAAI-CANN Open Fund, developed on OpenI Community.

References

Al Shafian, S.; He, C.; and Hu, D. 2025. DamageScope: An Integrated Pipeline for Building Damage Segmentation, Geospatial Mapping, and Interactive Web-Based Visualization. *Remote Sensing*.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2019. nuScenes: A Multimodal Dataset for Autonomous Driving. In *CVPR*.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *ECCV*.

Dai, R.; Li, C.; Yan, Y.; Mo, L.; Qin, K.; and He, T. 2025. Unbiased Missing-modality Multimodal Learning. In *ICCV*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL*.

Elghazaly, G.; Frank, R.; Harvey, S.; and Safko, S. 2023. High-Definition Maps: Comprehensive Survey, Challenges, and Future Perspectives. *IEEE OJ-ITS*.

Fu, Y.; Liao, W.; Liu, X.; xu, H.; Ma, Y.; Dai, F.; and Zhang, Y. 2024. TopoLogic: An Interpretable Pipeline for Lane Topology Reasoning on Driving Scenes. *NeurIPS*.

Gao, J.; Sun, C.; Zhao, H.; Shen, Y.; Anguelov, D.; Li, C.; and Schmid, C. 2020. VectorNet: Encoding HD Maps and Agent Dynamics From Vectorized Representation. In *CVPR*.

Guo, T.; Lu, B.; Wang, F.; and Lu, Z. 2025. Depth-Aware Super-Resolution via Distance-Adaptive Variational Formulation. *arXiv preprint arXiv:2509.05746*.

Haklay, M. M.; and Weber, P. 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*.

He, C.; and Hu, D. 2025. Social Media Analytics for Disaster Response: Classification and Geospatial Visualization Framework. *Applied Sciences*.

He, K.; Chen, X.; Xie, S.; Li, Y.; Doll'ar, P.; and Girshick, R. B. 2021. Masked Autoencoders Are Scalable Vision Learners. In *CVPR*.

Hu, C.; Zhao, C.; Shao, H.; Deng, J.; and Wang, Y. 2024. TMFF: Trustworthy Multi-Focus Fusion Framework for Multi-Label Sewer Defect Classification in Sewer Inspection Videos. *IEEE Transactions on Circuits and Systems for Video Technology*.

Huang, S.; Shen, G.; Kang, Y.; and Song, Y. 2025. Immersive Augmented Reality Music Interaction through Spatial Scene Understanding and Hand Gesture Recognition. *Preprints*.

Huang, Z.; Yang, S.; Zhou, M.; Li, Z.; Gong, Z.; and Chen, Y. 2022. Feature Map Distillation of Thin Nets for Low-Resolution Object Recognition. *IEEE Transactions on Image Processing*.

Jiang, Z.; Zhu, Z.; Li, P.; Gao, H.; Yuan, T.; Shi, Y.; Zhao, H.; and Zhao, H. 2024. P-MapNet: Far-seeing Map Generator Enhanced by both SDMap and HDMap Priors. *IEEE Robotics and Automation Letters*.

Li, Q.; Wang, Y.; Wang, Y.; and Zhao, H. 2021. HDMapNet: An Online HD Map Construction and Evaluation Framework. In *ICRA*.

Li, S.; Li, B.; Sun, B.; and Weng, Y. 2024. Towards Visual-Prompt Temporal Answer Grounding in Instructional Video. *IEEE transactions on pattern analysis and machine intelligence*, 46(12): 8836–8853.

Li, T.; Chen, L.; Wang, H.; Li, Y.; Yang, J.; Geng, X.; Jiang, S.; Wang, Y.; Xu, H.; Xu, C.; Yan, J.; Luo, P.; and Li, H. 2023. Graph-based Topology Reasoning for Driving Scenes. *arXiv preprint arXiv:2304.05277*.

Li, W.; Zhang, R.; Shao, R.; He, J.; and Nie, L. 2025. Cogvla: Cognition-aligned vision-language-action model via instruction-driven routing & sparsification. *arXiv preprint arXiv:2508.21046*.

Li, Z. 2025. Knowledge-Grounded Detection of Cryptocurrency Scams with Retrieval-Augmented LMs. In *Knowledgeable Foundation Models at ACL 2025*.

Li, Z.; Qiu, S.; and Ke, Z. 2025. Revolutionizing Drug Discovery: Integrating Spatial Transcriptomics with Advanced Computer Vision Techniques. In *CVPR Workshop*.

Li, Z.; Wang, W.; Li, H.; Xie, E.; Sima, C.; Lu, T.; Qiao, Y.; and Dai, J. 2022. BEVFormer: Learning Bird's-Eye-View Representation from Multi-camera Images via Spatiotemporal Transformers. In *ECCV*.

Liang, S.; Chang, X.; Wu, C.; Yan, H.; Bai, Y.; Liu, X.; Zhang, H.; Yuan, Y.; Zeng, S.; Xu, M.; et al. 2025. Persistent Autoregressive Mapping with Traffic Rules for Autonomous Driving. *arXiv preprint arXiv:2509.22756*.

Liao, B.; Chen, S.; Wang, X.; Cheng, T.; Zhang, Q.; Liu, W.; and Huang, C. 2023a. MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction. In *ICLR*.

Liao, B.; Chen, S.; Zhang, Y.; Jiang, B.; Zhang, Q.; Liu, W.; Huang, C.; and Wang, X. 2023b. MapTRv2: An End-to-End Framework for Online Vectorized HD Map Construction. *IJCV*.

Liu, J.; Huang, X.; Zheng, J.; Liu, Y.; and Li, H. 2023. MixMAE: Mixed and Masked Autoencoder for Efficient Pretraining of Hierarchical Vision Transformers. In *CVPR*.

Liu, Y.; Wang, T.; Zhang, X.; and Sun, J. 2022a. PETR: Position Embedding Transformation for Multi-View 3D Object Detection. *ECCV*.

Liu, Y.; Wang, Y.; Wang, Y.; and Zhao, H. 2022b. VectorMapNet: End-to-end Vectorized HD Map Learning. In *ICML*.

Luo, K. Z.; Weng, X.; Wang, Y.; Wu, S.; Li, J.; Weinberger, K. Q.; Wang, Y.; and Pavone, M. 2024. Augmenting Lane Perception and Topology Understanding with Standard Definition Navigation Maps. *ICRA*.

Ma, Z.; Liang, S.; Wen, Y.; Lu, W.; and Wan, G. 2024. RoadPainter: Points Are Ideal Navigators for Topology transformER. *ECCV*.

Ma, Z.; Luo, Y.; Zhang, Z.; Sun, A.; Yang, Y.; and Liu, H. 2025. Reinforcement Learning Approach for Highway Lane-Changing: PPO-Based Strategy Design. *Preprints*.

Peng, N.; Zhou, X.; Wang, M.; Yang, X.; Chen, S.; and Chen, G. 2025. PrevPredMap: Exploring Temporal Modeling with Previous Predictions for Online Vectorized HD Map Construction. *WACV*.

Qiao, L.; Ding, W.; Qiu, X.; and Zhang, C. 2023. End-to-End Vectorized HD-map Construction with Piecewise Bézier Curve. In *CVPR*.

Schmidt, J.; Jordan, J.; Gritschneder, F.; Monninger, T.; and Dietmayer, K. 2023. Exploring Navigation Maps for Learning-Based Motion Prediction. In *ICRA*.

Shelhamer, E.; Long, J.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*.

Sun, Q.; and Li, J. 2025. A Lightweight YOLOv4-SVM Model for Automated Waste Monitoring in Smart Cities. *TechRxiv preprint*.

Sun, R.; Yang, L.; Lingrand, D.; and Precioso, F. 2023a. Mind the map! Accounting for existing map information when estimating online HDMaps from sensor data. *arXiv preprint arXiv:2311.10517*.

Sun, S.; Ren, Y.; Ma, C.; and Zhang, X. 2023b. Large language models as topological structure enhancers for text-attributed graphs. *arXiv preprint arXiv:2311.14324*.

Tancik, M.; Srinivasan, P. P.; Mildenhall, B.; Fridovich-Keil, S.; Raghavan, N.; Singhal, U.; Ramamoorthi, R.; Barron, J. T.; and Ng, R. 2020. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. In *NeurIPS*.

Tang, L.; Huang, K.; Chen, C.; Yuan, Y.; Li, C.; Tu, X.; Ding, X.; and Huang, Y. 2025. Dissecting Generalized Category Discovery: Multiplex Consensus under Self-Deconstruction. *arXiv preprint arXiv:2508.10731*.

Wan, J.; Wang, X.; Xie, M.; Chang, X.; Liu, X.; Pan, Z.; Xu, M.; and Yuan, D. 2025. Driving by Hybrid Navigation: An Online HD-SD Map Association Framework and Benchmark for Autonomous Vehicles. *arXiv preprint arXiv:2507.07487*.

Wang, H.; Li, T.; Li, Y.; Chen, L.; Sima, C.; Liu, Z.; Wang, B.; Jia, P.; Wang, Y.; Jiang, S.; Wen, F.; Xu, H.; Luo, P.; Yan, J.; Zhang, W.; and Li, H. 2023. OpenLane-V2: A Topology Reasoning Benchmark for Unified 3D HD Mapping. In *NeurIPS*.

Wilson, B.; Qi, W.; Agarwal, T.; Lambert, J.; Singh, J.; Khandelwal, S.; Pan, B.; Kumar, R.; Hartnett, A.; Pontes, J. K.; Ramanan, D.; Carr, P.; and Hays, J. 2021. Argoverse 2: Next Generation Datasets for Self-Driving Perception and Forecasting. In *NeurIPS*.

Wu, D.; Chang, J.; Jia, F.; Liu, Y.; Wang, T.; and Shen, J. 2024. TopoMLP: An Simple yet Strong Pipeline for Driving Topology Reasoning. *ICLR*.

Wu, K.; Yang, C.; and Li, Z. 2025. InteractionMap: Improving Online Vectorized HDMap Construction with Interaction. *CVPR*.

Xiao, J.; Yang, Y.; Chang, X.; Chen, R.; Xiong, F.; Xu, M.; Zheng, W.-S.; and Zhang, Q. 2025. World-Env: Leveraging World Model as a Virtual Environment for VLA Post-Training. *arXiv preprint arXiv:2509.24948*.

Xie, M.; Zeng, S.; Chang, X.; Liu, X.; Pan, Z.; Xu, M.; and Wei, X. 2025. SeqGrowGraph: Learning Lane Topology as a Chain of Graph Expansions. *ICCV*.

Xiong, X.; Liu, Y.; Yuan, T.; Wang, Y.; Wang, Y.; and Zhao, H. 2023. Neural Map Prior for Autonomous Driving. In *CVPR*.

Yao, S.; Guan, R.; Wu, Z.; Ni, Y.; Huang, Z.; Liu, R. W.; Yue, Y.; Ding, W.; Lim, E. G.; Seo, H.; Man, K. L.; Ma, J.; Zhu, X.; and Yue, Y. 2024. WaterScenes: A Multi-Task 4D Radar-Camera Fusion Dataset and Benchmarks for Autonomous Driving on Water Surfaces. *IEEE T-ITS*.

Yuan, T.; Liu, Y.; Wang, Y.; Wang, Y.; and Zhao, H. 2024. StreamMapNet: Streaming Mapping Network for Vectorized Online HD Map Construction. In *WACV*.

Yuan, Y.; Wu, C.; Chang, X.; Wang, S.; Zhang, H.; Liang, S.; Zeng, S.; and Xu, M. 2025. UniMapGen: A Generative Framework for Large-Scale Map Construction from Multi-modal Data. *arXiv preprint arXiv:2509.22262*.

Zeng, S.; Chang, X.; Xie, M.; Liu, X.; Bai, Y.; Pan, Z.; Xu, M.; and Wei, X. 2025a. FutureSightDrive: Thinking Visually with Spatio-Temporal CoT for Autonomous Driving. *NeurIPS*.

Zeng, S.; Qi, D.; Chang, X.; Xiong, F.; Xie, S.; Wu, X.; Liang, S.; Xu, M.; and Wei, X. 2025b. JanusVLN: Decoupling Semantics and Spatiality with Dual Implicit Memory for Vision-Language Navigation. *arXiv preprint arXiv:2509.22548*.

Zeng, X.; Wang, H.; Lin, J.; Wu, J.; Cody, T.; and Zhou, D. 2025c. LENSLLM: Unveiling Fine-Tuning Dynamics for LLM Selection.

Zhang, H.; Paz, D.; Guo, Y.; Das, A.; Huang, X.; Haug, K.; Christensen, H. I.; and Ren, L. 2024a. Enhancing Online Road Network Perception and Reasoning with Standard Definition Maps. *IROS*.

Zhang, X.; Liu, G.; Liu, Z.; Xu, N.; Liu, Y.; and Zhao, J. 2024b. Enhancing Vectorized Map Perception with Historical Rasterized Maps. In *ECCV*.

Zhang, Y.; Mo, K.; Shen, F.; Xu, X.; Zhang, X.; Yu, J.; and Yu, C. 2024c. Self-Adaptive Robust Motion Planning for High DoF Robot Manipulator using Deep MPC. *arXiv preprint arXiv:2407.12887*.

Zhou, X.; Zhang, M.; Lee, Z.; Ye, W.; and Zhang, S. 2025. Hademif: Hallucination detection and mitigation in large language models. In *ICLR*.

Zhou, Y.; Zhang, H.; Yu, J.; Yang, Y.; Jung, S.; Park, S.-I.; and Yoo, B. 2024. HIMap: HybrId Representation Learning for End-to-end Vectorized HD Map Construction. In *CVPR*.