# Persistent Autoregressive Mapping with Traffic Rules for Autonomous Driving

**Shiyi Liang**[1,2*†], **Xinyuan Chang**[3*], **Changjie Wu**[3*], **Huiyuan Yan**[1,2], **Yifan Bai**[5], **Xinran Liu**[3], **Hang Zhang**[3], **Yujian Yuan**[4], **Shuang Zeng**[1,2], **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
[5] DAMO Academy, Alibaba Group
sy_liang2023@stu.xjtu.edu.cn, {changxinyuan.cxy, wuchangjie.wcj}@alibaba-inc.com, weixing@mail.xjtu.edu.cn

## Abstract

Safe autonomous driving requires both accurate HD map construction and persistent awareness of traffic rules, even when their associated signs are no longer visible. However, existing methods either focus solely on geometric elements or treat rules as temporary classifications, failing to capture their persistent effectiveness across extended driving sequences. In this paper, we present **PAMR** (**P**ersistent **A**utoregressive **M**apping with Traffic **R**ules), a novel framework that performs autoregressive co-construction of lane vectors and traffic rules from visual observations. Our approach introduces two key mechanisms: **Map-Rule Co-Construction** for processing driving scenes in temporal segments, and **Map-Rule Cache** for maintaining rule consistency across these segments. To properly evaluate continuous and consistent map generation, we develop MapDRv2, featuring improved lane geometry annotations. Extensive experiments demonstrate that PAMR achieves superior performance in joint vector-rule mapping tasks, while maintaining persistent rule effectiveness throughout extended driving sequences.

**Code** — https://miv-xjtu.github.io/PAMR/

## 1 Introduction

Driving by the rules is fundamental to safe autonomous navigation. Inherently sequential in nature, driving requires continuous interpretation and application of traffic rules along the vehicle's trajectory. While existing High-Definition (HD) map construction (Liao et al. 2023a,b; Chen et al. 2024a; Ben Charrada et al. 2022) focuses on geometric elements like lane topology, it often overlooks traffic rules—semantic elements that govern driving behavior and persist beyond their signs' visibility. These rules and road geometry form an intrinsically interwoven, rule-governed space. Our core idea is that robust autonomous navigation demands **Persistent Driving by the Rules**, acknowledging these semantic guidelines' continuous influence across time and space, from initial observation through extended trajectories.

Current methods fail to capture this persistent nature. The key challenge lies in traffic rules' persistent effectiveness: a sign's influence extends well beyond its visible range (Fig 1). This creates a complex dependency between road geometry and rules that existing systems cannot model. Traditional vector-only methods (Liao et al. 2023a,b) generate geometric vectors but remain "semantically blind." While some approaches (Wang et al. 2023) incorporate rules, they reduce rule assignment to simple classification, failing to maintain consistency or interpret signs' lasting impact. Even recent advances like MapDR (Chang et al. 2025), despite addressing rule understanding, rely on pre-constructed vectorized maps (Fig 2 (a)), preventing end-to-end geometry-semantic co-construction. This fundamental inability to model persistent effectiveness fragments the driving environment, failing to create the coherent representation necessary for rule-compliant navigation.

In this paper, we propose **PAMR** (**P**ersistent **A**utoregressive **M**apping with Traffic **R**ules), a novel framework that performs autoregressive co-construction of lane vectors and traffic rules from visual observations. Our approach integrates geometric reasoning with traffic semantics, enabling the model to infer lane-level rules and maintain their validity over time. This design mirrors human drivers' ability to apply traffic rules beyond immediate visibility, bridging the gap between instantaneous perception and rule-aware decision-making in autonomous systems. Rather than detecting isolated elements, PAMR "narrates" the road scene by conditioning each map element on previously generated context. This sequential reasoning ensures consistency between lane structures and their governing rules, while enabling occluded lane inference through contextual understanding (Fig 2 (c)).

PAMR contains two key mechanisms: **Map-Rule Co-**

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

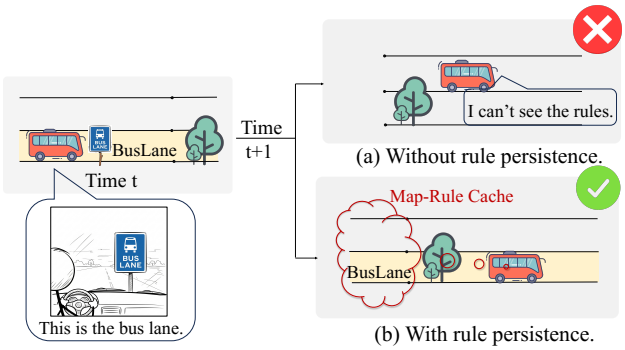


Figure 1: Schematic of Persistent Rule Effectiveness. (a) **Without rule persistence**: At time t, a bus lane sign is observed. At t+1, without remembering the rule, the vehicle attempts an illegal lane change. (b) **With rule persistence:** The system retains the bus lane rule, preventing the incorrect maneuver and ensuring continued adherence to traffic rules.

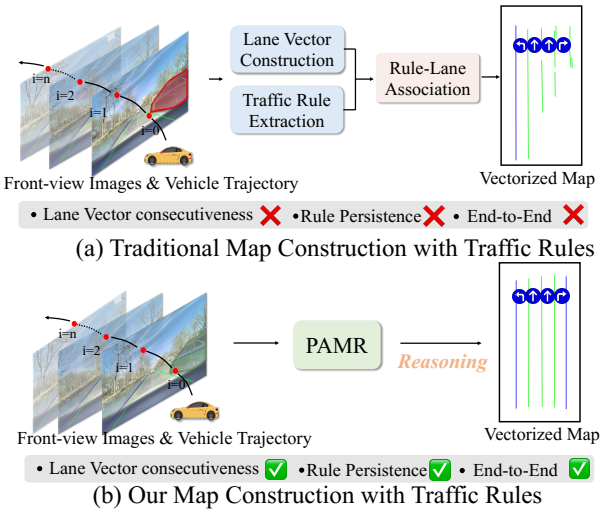


Figure 2: Comparison between traditional pipeline and PAMR. (a) Traditional methods adopt a multi-stage pipeline: constructing lane vectors, extracting and associating traffic rules. This separated approach fails to maintain vector consecutiveness, rule persistence, and end-to-end learning. (b) Our PAMR framework performs autoregressive co-construction of lanes and rules, enabling sequential reasoning. Through joint construction, PAMR achieves continuous vectors, persistent rule awareness, and end-to-end integration of geometric and semantic information.

**Construction** for processing driving scenes in temporal segments, and **Map-Rule Cache** for seamless propagation of rules and geometry between segments. However, evaluating continuous and consistent map generation requires an appropriate benchmark. The original MapDR (Chang et al. 2025) dataset, with its fragmented lane annotations, proves inadequate for this task. We therefore developed MapDRv2, featuring smooth and continuous lane geometries, to enable meaningful evaluation of our approach.

To sum up, our contributions are as follows:

- We develop MapDRv2, a re-annotated dataset featuring continuous lane geometries, providing a more suitable benchmark for evaluating models that focus on generating consistent and continuous HD maps with traffic rules.
- We propose PAMR, a novel framework that achieves persistent driving by the rules through autoregressive co-construction of lane vectors and traffic rules, addressing the fundamental challenge of maintaining rule effectiveness beyond immediate visibility.
- We introduce two key technical components: Map-Rule Co-Construction for processing driving scenes in temporal segments, and Map-Rule Cache for maintaining consistent rule propagation, enabling seamless integration of geometric and semantic information across extended driving sequences.

# 2 Related Work

## 2.1 HD Map Construction

Fueled by large-scale autonomous driving (Wilson et al. 2021; Caesar et al. 2020) and traffic sign datasets (Behrendt, Novak, and Botros 2017; Stallkamp et al. 2012; Yu et al. 2020; Fregin et al. 2018; Zhu et al. 2016), HD map construction is rapidly evolving towards end-to-end solutions (Li et al. 2024; Zhang et al. 2024, 2023; Yuan et al. 2025). Pioneering works like VectorMapNet (Liu et al. 2023) and MapTR (Liao et al. 2023a) established methods for sequential and permutation-equivalent vector prediction. However, despite architectural advances like PivotNet (Ding et al. 2023) for improving geometry, these methods remain "semantically blind" to traffic rules.

Recent attempts to incorporate semantics are limited. OpenLane-V2 (Wang et al. 2023) treats rule assignment as simple classification, while MapDR (Chang et al. 2025) relies on pre-constructed vectors, breaking the end-to-end paradigm. A critical challenge therefore remains: modeling the persistent effectiveness of traffic rules beyond their visible range while ensuring geometric-semantic consistency. This gap motivates our work towards unified geometric-semantic mapping.

## 2.2 MLLMs in Driving

Driving's sequential nature requires continuous interpretation of semantics, yet even advanced sequential models like VectorMapNet (Liu et al. 2023) and MapTR (Liao et al. 2023a) largely neglect the temporal persistence of traffic rules. This is where Multimodal Large Language Models (MLLMs) show great promise, known for their success in sequential reasoning and contextual awareness (Anthropic 2024; Bai et al. 2023; OpenAI 2024).

Their potential for understanding complex driving scenarios is being actively explored (Cui et al. 2024; Xu et al. 2024; Sima et al. 2024; Tian et al. 2024; Chen et al. 2024b; Ding et al. 2024; Choudhary et al. 2024; Cao, Wei, and Ma 2025; Zeng et al. 2025a, 2024; Xie et al. 2025a; Wei et al. 2024; Xie et al. 2025b; Zeng et al. 2024, 2025a,b). Particularly relevant is their ability to "narrate" sequential information while

maintaining long-term dependencies—a crucial capability for the persistent rule-aware mapping that current methods lack. This aligns with the trend towards unified architectures, offering a promising path to integrate deep geometric and semantic understanding in autonomous driving.

# 3 Method

## 3.1 MapDRv2

Persistent driving by the rules necessitates a benchmark that upholds the same principles of continuity and coherence. A model's ability to generate continuous outputs is fundamentally determined by the integrity of its training data. While the original MapDR dataset was pioneering in its integration of traffic rules, its geometric annotations suffer from significant fragmentation, frequently exhibiting discontinuities in scenarios involving occlusion or complex topologies (see Fig 3(a)). Consequently, this geometric fragmentation renders the dataset unsuitable for training or evaluating models whose primary objective is to generate a single, coherent representation of the road network.

To establish a robust foundation for this task, we performed a meticulous re-annotation of the lane vectors in the MapDR dataset. While preserving the original data scale and scenarios, we employed a human-in-the-loop methodology. This process involved projecting video frames into a unified BEV space, which enabled human annotators to leverage the full temporal context of each clip. By doing so, they could accurately extrapolate occluded lane segments based on contextual geometric cues and resolve topological ambiguities using logical priors. This procedure yielded a set of enhanced annotations characterized by smooth and continuous lane geometries that more faithfully represent real-world road structures, thereby providing a reliable ground truth for our restoration task (see Fig 3(b)).

With this enhanced ground truth established, we developed a unified evaluation framework to holistically assess a model's performance. Our evaluation methodology is composed of two primary categories of metrics:

**Vectorized Lane Accuracy.** To assess predicted lane vectors $\hat{V}$ against ground-truth $V$, we implement topology-aware Intersection-over-Union (IoU (Zheng et al. 2020)) evaluation. For each lane polyline $\hat{l}_i \in \hat{V}$, we rasterize lanes into fixed-wide binary masks through polyline expansion. The IoU between each predicted mask $\hat{l}_i$ and ground-truth masks $\{l_i\}_{i=1}^{k}$ is computed, with matches established when IoU $>$ 0.5. We define the vector accuracy metric $\mathcal{F}_{\text{vec}}$ as Eq 1:

$$\mathcal{F}_{\text{vec}} = \frac{1}{|\mathcal{M}|} \sum (\hat{l}_i, l_j) \in \mathcal{M}_{\text{IoU}}(\hat{l}_i, l_j) \tag{1}$$

**Holistic Mapping Accuracy.** We propose a single, unified metric, HMA, that evaluates the joint correctness of lane vectors, traffic rules, and their association. A predicted pair $(\hat{r}_i, \hat{l}_j)$ is considered true positive (TP) iff: (1) The predicted lane$\hat{l}_i$ accurately matches a ground-truth lane $l_j$. (2) The extracted rule $\hat{r}_i$ correctly matches a ground-truth rule $r_i$. (3) The association between the matched pair $(r, l)$ is valid in the ground truth. The overall performance is then quantified by the holistic F1-score, $\mathcal{F}$, which harmonizes precision ($P$) and recall ($R$) calculated over these comprehensively validated true positives:

$$\mathcal{F} = \frac{2\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}} \tag{2}$$

## 3.2 PAMR

We propose **PAMR** (**P**ersistent **A**utoregressive **M**apping with Traffic **R**ules), an end-to-end framework for co-constructing lane vectors and traffic regulations. Our framework processes multimodal inputs and leverages MLLM (Anthropic 2024; OpenAI 2024; Team et al. 2024; Bai et al. 2023) for comprehensive HD map construction.

Specifically, our framework accepts a sequence of front-view images $I$ along with their corresponding vehicle trajectory $\mathcal{T}$ as input to generate a vectorized HD map, represented as a graph $G(V)$. Here $V = \{L, R\}$ consists of lane vectors $L = \{l_i\}_{i=1}^{k}$ and traffic rules $R = \{r_i\}_{i=1}^{m}$, This mapping process can be formally defined as Eq 3:

$$G(V) = f(I, \mathcal{T}, Prompt^*) \tag{3}$$

where $f$ denotes the MLLM, and $Prompt^*$ represents optional prompts that enable both continuous map-rule construction ($P_{con}$ in Sec 3.4) and interactive ($P_{rule}$ in Sec 3.5). These components will be thoroughly discussed in the following sections.

## 3.3 Map-Rule Co-Construction

We propose an integrated approach that jointly generates lanes and rules within a map-rule framework, capturing both spatial and temporal contexts along the driving trajectory. At each timestamp $t$, we establish a segment $G_t$ centered at the ego vehicle's current position $(x_i, y_i)$, which encompasses a fixed spatial region of size $W \times H$. As illustrated in Fig 4, each segment integrates the current frame with a sequence of $N$ historical frames to facilitate comprehensive vector and rule co-construction.

**Input Serialization.** Within each segment containing $N$ front-view images $\{I_t\}_{t=0}^{t=N}$, we first leverage a pre-trained visual backbone (e.g., Vision Transformer, ViT (Dosovitskiy et al. 2021)) to extract high-level feature maps. These features are subsequently flattened and projected into a sequence of visual tokens, denoted as $T_v = \{IMG^j\}$. Meanwhile, we sample the vehicle's trajectory to obtain a series of ego-poses, where each pose is characterized by its 2D coordinates and heading angle $(x_t, y_t, \theta_t)$ in a local coordinate frame. To align with MLLM's discrete token processing paradigm, we perform quantization on this continuous data. Specifically, the coordinate and angle values are first normalized and discretized according to the current segment's dimensions, then mapped to unique tokens within the model's vocabulary, yielding a sequence of pose tokens $T_{pose}$. Finally, we interleave the token sequences from all modalities into a unified input stream $T_{in} = [T_{v0}, T_{pose0}, ..., T_{vN}, T_{poseN}]$.

**Output Serialization.** Our framework implements a tightly coupled serialization scheme that integrates geometric information with semantic attributes at the generation level. Rather than producing geometries and rules separately, the

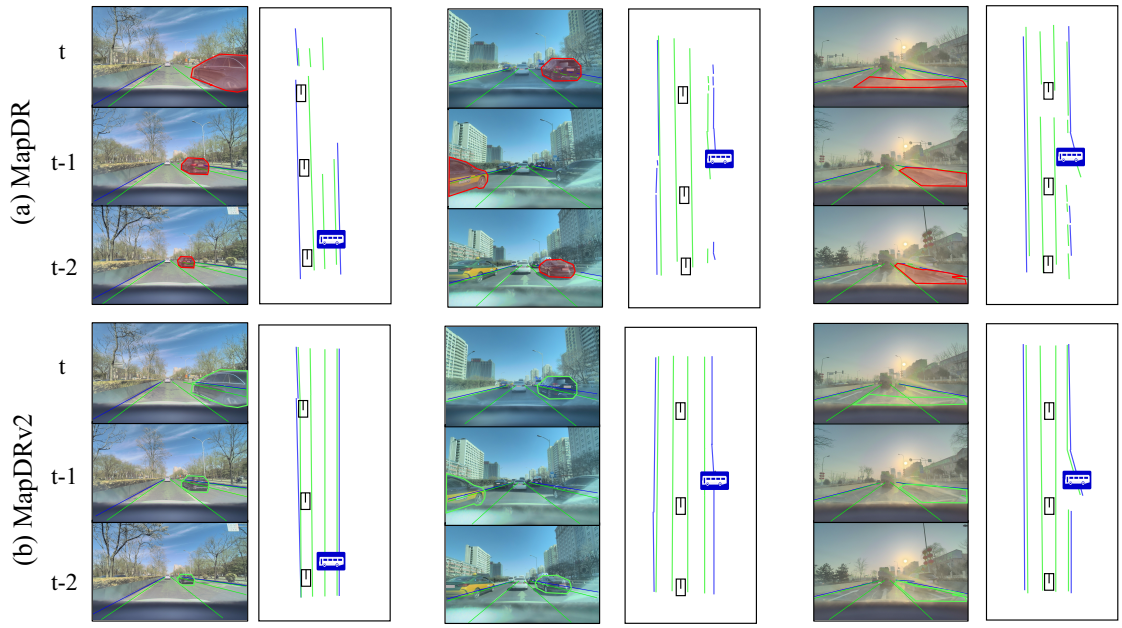


Figure 3: Comparison with MapDR in occluded scenarios. For each example, the left three images show the input PV views, while the right images present the corresponding ground-truth HD maps. Green lines and blue lines represent the dividers and borderlines, respectively. In the case of occlusion and glare, MapDR generates fragmented vector representations. In contrast, MapDRv2 provides more complete results.

MLLM is trained to generate a single, coherent sequence where each lane vector is immediately followed by its associated traffic rule.

The generation of a single lane vector $l_i$ follows a structured format: it begins with a start token `[lane]`, continues with a sequence of quantized 2D coordinate tokens representing its polyline, and terminates with an end token `[\lane]`.

Traffic regulations are encoded as key-value pairs linked to their corresponding geometric elements. Following a similar structure, each rule begins with `[rule]`, contains predefined rule template tokens, and ends with `[\rule]`. For lanes without associated rules, a special token `[None]` is used. The final outputs are organized as $output$ = `[lane]`...`[\lane][rule]`...`[\rule]`....

A deterministic parser processes this output string through a systematic procedure: first segmenting the sequence based on predefined delimiters, then de-quantizing the coordinate tokens into metric polylines, and finally associating each polyline with its corresponding rule attributes. This parsing process directly constructs the map graph $G(V)$.

### 3.4 Map-Rule Cache

As discussed in the Introduction, traffic rules exhibit persistent effects that extend well beyond their initial points of observation. Local segment consistency alone proves insufficient, as critical context may be lost when frames containing traffic signs move out of view. To address this limitation, we propose a map-rule cache that enables effective information propagation across consecutive segments.

As illustrated in Fig 4, our caching mechanism operates as follows: after processing each segment, we extract vector points near its boundary and project them into new map dimensions as cache data. These caches then serve as initialization states for subsequent segments, denoted as $p_{con}$. To ensure smooth transitions, we design caches with an overlapping region of length $\delta$. This process can be formalized:

$$G_t(V) = f(X_t, T_t, P_{t-1}^{con}) \tag{4}$$

Where $G_t$ represents the vectorized HD map within the current segment, $X_t$ denotes the PV images, and $\mathcal{T}_t$ indicates the trajectory. The term $\mathcal{P}_{t-1}^{con}$ represents the cache inherited from the previous segment. During concatenation, we resolve overlapping regions by retaining the prediction from the succeeding segment while discarding the corresponding region from the preceding one to ensure seamless integration.

This iterative process of propagation and concatenation enables the generation of arbitrarily long, continuous maps. Crucially, it ensures persistent rule awareness by maintaining historical context, even when the corresponding visual cues have moved beyond the current observation window.

### 3.5 Interactive Prompt

Although driving scenarios typically involve multiple co-located traffic signs, and our PAMR model is fully capable of processing them simultaneously, we align our evaluation framework with the established MapDR protocol by focusing on individual signs. This selective evaluation is enabled by PAMR's interactive feature, which allows users or evaluation

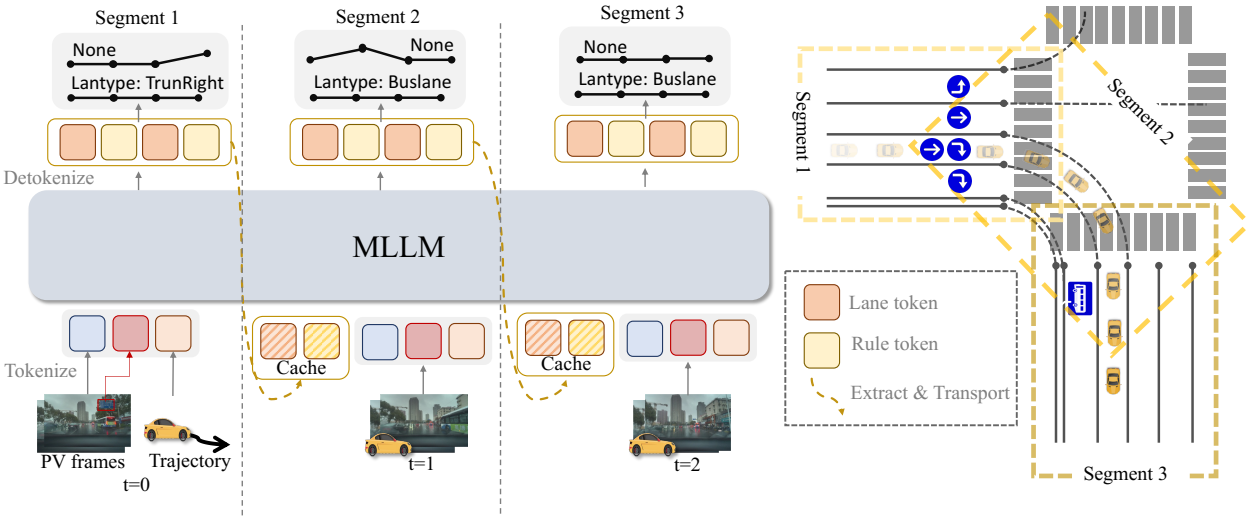


Figure 4: Overview of the PAMR framework. **Left**: The sequential processing of map-rules, where each segment takes PV frames and trajectory as input, tokenizes them along with cache from previous segment (if any), and feeds them into MLLM. The MLLM outputs are then detokenized into lane vectors with associated rules. **Right**: Bird's-eye view visualization of the map-rule construction process, showing how information is propagated across consecutive segments through the caching mechanism. The cache ensures continuous rule awareness even as the vehicle moves forward, enabling consistent map-rule generation across the entire trajectory.

scripts to specify target traffic signs for map construction. This approach facilitates systematic, fine-grained assessment of the model's performance on each individual sign. Importantly, this single-sign focus during evaluation reflects a methodological choice rather than a model constraint - PAMR maintains the capability to process any combination of visible traffic signs concurrently.

To enable this interactivity, we introduce three distinct prompting strategies, formulated in Eq 5:

$$\mathcal{P}_{\text{rule}} \in \left\{ \begin{array}{ll} \texttt{[COORD]} \oplus (u_k^{(i)}, v_k^{(i)}) & \text{(Coordinate)} \\ \texttt{[BBOX]} \oplus b_k & \text{(Visual Highlight)} \\ \texttt{[ROI]} \oplus X_k^{\text{ROI}} & \text{(Visual Rule)} \end{array} \right\}, \tag{5}$$

where `[COORD]` leverages discretized 3D coordinates of traffic signs projected onto the PV as $(u_k^{(i)}, v_k^{(i)})$, `[BBOX]` highlights the target traffic sign by drawing a bounding box $b_k$on the input image; and `[ROI]` extracts and encodes visual features from the traffic sign's region of interest.

# 4 Experiments

## 4.1 Implementation Details

**Dataset and Metric.** We evaluate PAMR on MapDRv2, with metrics defined in Sec 3.1. The dataset encompasses diverse scenarios, weather conditions, and traffic situations, comprising over 10k traffic scene segments, 18k driving rules, and 400k images with $1960 \times 1240$ resolution. Each segment is set to $224 \times 224$ ($W \times H$, corresponding to a $22.4m \times 22.4m$ real-world area), with an overlap ratio $\delta$ of 10%. The dataset is split into *train* and *test* sets with a $9:1$ ratio for evaluation.

**Training Strategy** We use Qwen2-VL-2B (Bai et al. 2023) as our MLLMs model. For training, we set the batch size a 256, and the models are optimized using AdamW (Loshchilov and Hutter 2017) with a weight decay of 0.1. The learning rate is set to $2 \times 10^{-5}$ and a linear warm-up of 100 steps. The training process comprises 20 epochs. 32 NVIDIA H20 are used in total. In order to reduce the occupation of GPU memory, we uniformly sample each group of input PV images and retain less than 10 images and each image is resized to $644 \times 364$. Please refer to the supplementary materials for more training details.

## 4.2 Main Result

Table 1 presents a comprehensive evaluation of our proposed PAMR. While PAMR shows lower performance in rule extraction compared to RuleVLM (Chang et al. 2025), this is primarily because RuleVLM processes PV images containing **only** target rules, whereas PAMR needs to identify and evaluate specific signs from multiple traffic rules. Although this selective evaluation requirement affects PAMR's performance in single-rule metrics compared to MapDR, our method significantly outperforms both RuleVLM and VLE-MEE in joint rule extraction and association tasks. This superior performance in unified tasks demonstrates PAMR's effectiveness in joint vector-rule mapping, enabled by our co-construction approach that enables deep semantic understanding of rules rather than simple matching-based association.

## 4.3 Ablation Study

We conduct ablation studies to evaluate our key design choices. All experiments use identical settings except for

| Methods | Vec | Rule Extract | | | HMA | | |
|---|---|---|---|---|---|---|---|
| | $\mathcal{F}_{vec}$ | $P_{\mathcal{H}}(\%)$ | $R_{\mathcal{H}}(\%)$ | $\mathcal{F}_{\mathcal{H}}$ | $P_{\mathcal{H}}(\%)$ | $R_{\mathcal{H}}(\%)$ | $\mathcal{F}_{\mathcal{H}}$ |
| *single task* | | | | | | | |
| MapDR(VLE-MEE) | / | 76.67 | 74.54 | 75.58 | 63.35 | 67.37 | 65.29 |
| MapDR(RuleVLM) | / | 89.28 | 89.44 | 89.3 | 64.16 | 64.25 | 64.20 |
| *comprehensive task* | | | | | | | |
| PAMR | 0.46 | 84.52 | 82.22 | 83.39 | 71.32 | 69.33 | **70.37** |

Table 1: Evaluation on MapDRv2 test set. Comparison of different methods across three metrics: **HMA** (holistic quality of rule extraction and contextual association), **Rule Extraction** (extracting key-value pairs from traffic sign rules), **Vec** (lane vector reconstruction accuracy). **Notably, MapDR directly provides target traffic signs, while PAMR must first identify the relevant signs from multiple candidates before performing rule extraction.**

| Segment Size | Vec | HMA | | |
|---|---|---|---|---|
| | $\mathcal{F}_{vec}$ | $P_{\mathcal{H}}(\%)$ | $R_{\mathcal{H}}(\%)$ | $\mathcal{F}_{\mathcal{H}}$ |
| $122 \times 224$ | **0.46** | 63.12 | 46.58 | 53.60 |
| $224 \times 224$ | **0.46** | **71.32** | **69.33** | **70.37** |
| $448 \times 224$ | 0.41 | 64.34 | 55.59 | 59.64 |

Table 2: Segment Size. Performance comparison across different Segment widths with a fixed height. All configurations are designed to ensure complete coverage of local map segments.

| Map-Rule Cache | HMA | | |
|---|---|---|---|
| | $P_{R.E.}(\%)$ | $R_{R.E.}(\%)$ | $\mathcal{F}_{rule}$ |
| w/o cache | 41.24 | 38.73 | 39.93 |
| w cache | **71.32** | **69.33** | **70.37** |

Table 3: Effectiveness Evaluation of Map-Rule Cache.

| Strategy | HMA | | |
|---|---|---|---|
| | $P_{R.E.}(\%)$ | $R_{R.E.}(\%)$ | $\mathcal{F}_{rule}$ |
| w/o prompt | 44.52 | 39.30 | 41.74 |
| `[COORD]` | 58.52 | 53.99 | 56.16 |
| `[BBOX]` | 63.78 | 55.10 | 59.12 |
| `[ROI]` | **71.32** | **69.33** | **70.37** |

Table 4: Interactive Prompt.

| Number | Vec | HMA | | |
|---|---|---|---|---|
| | $\mathcal{F}_{vec}$ | $P_{\mathcal{H}}(\%)$ | $R_{\mathcal{H}}(\%)$ | $\mathcal{F}_{\mathcal{H}}$ |
| 5 | 0.42 | 67.15 | 62.35 | 64.66 |
| 10 | **0.46** | **71.32** | **69.33** | **70.37** |
| 15 | 0.43 | 68.43 | 61.33 | 64.68 |

Table 5: Numbers of PV Images. Each experiment uniformly samples frames from the complete image sequence.

the components under comparison, ensuring controlled evaluation. Configurations used in our final model are highlighted in gray.

**Map-Rule Cache.** We validate the importance of Map-Rule Cache through ablation studies, as shown in Table 3. Without the cache mechanism, all metrics show significant degradation. This significant degradation occurs because, in the absence of caching, rule association is limited to segments containing traffic signs, and rule propagation between segments is disrupted. Consequently, vector measurements become discontinuous and inconsistent across segments.

**Segment Size.** Our method employs local map-rule co-construction, where the segment dimensions critically influence the information density of local reconstruction. As shown in Table 2, our size variation experiments reveal a clear trade-off: smaller segments excel in lane vector construction due to their focus on local details but compromise rule understanding due to limited context. Conversely, larger segments struggle with vector generation due to detail loss while introducing excessive noise that impairs rule interpretation. Medium-sized segments achieve optimal performance, maintaining sufficient detail for vector construction while capturing adequate context for rule understanding.

**Interactive Prompt.** In Eq. 5, we propose three distinct prompting strategies $\mathcal{P}_{rule}$ to direct the model's attention to specific traffic signs. Table 4 compares their effectiveness. The results demonstrate that directly incorporating visual ROI features achieves optimal performance. Alternative approaches, providing explicit coordinates or highlighting signs with bounding boxes, show reduced effectiveness. Without any prompting guidance, the model defaults to predicting all visible traffic signs, resulting in redundant outputs.

**Numbers of PV Images.** We investigate the impact of input sequence length by varying the number of PV images. As shown in Table 5, model performance improves with increasing input frames, suggesting that richer visual information enhances output quality. However, performance slightly degrades when exceeding 10 frames, indicating that excessive inputs may introduce redundant information that adversely affects model performance.

**Overlapping Region.** Table 6 presents ablation studies on the overlap region size, with map-rule dimensions held constant across experiments. Here, $\delta\%$ represents the ratio of

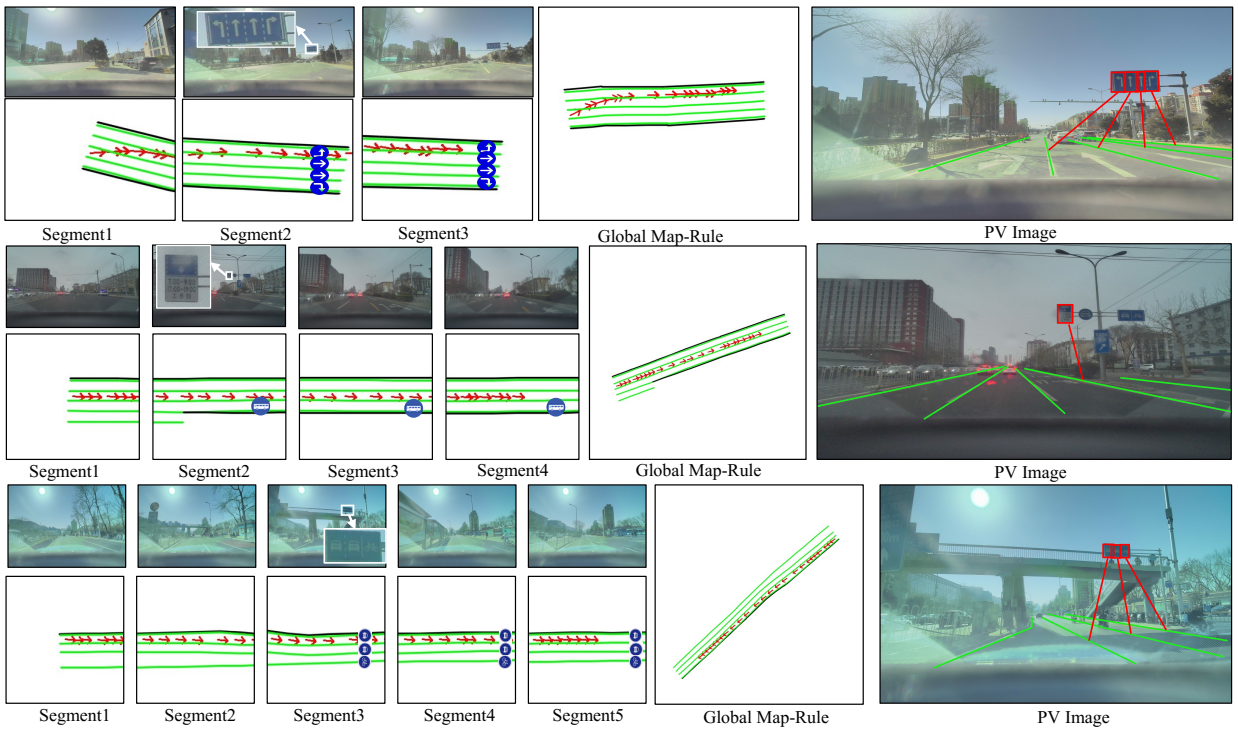


Figure 5: Visualization of map-rule construction. Segment 1-5 demonstrate the sequential processing results within individual segments, while the HD map shows the final integrated output after segments are concatenation. The green lines represent the constructed lane vectors while the black lines indicate border lines, and the red arrow shows the vehicle trajectory within each segment.

| $\delta\%$ | Vec | HMA | | |
|---|---|---|---|---|
| | $\mathcal{F}_{vec}$ | $P_{\mathcal{H}}(\%)$ | $R_{\mathcal{H}}(\%)$ | $\mathcal{F}_{\mathcal{H}}$ |
| 5 | 0.37 | 59.50 | 56.96 | 58.20 |
| 10 | **0.46** | **71.32** | **69.33** | **70.37** |
| 15 | 0.39 | 59.96 | 45.63 | 51.82 |

Table 6: Overlapping Region. $\delta\%$ represents the proportion of segments that overlap.

overlap width to map-rule width. The results show that both insufficient and excessive overlap adversely affect vector construction and rule understanding performance.

### 4.4 Visualization

Our framework demonstrates robust performance across various challenging scenarios, as illustrated in Fig 5. The map-rule mechanism effectively addresses several key challenges in HD map construction. First, it maintains consistency during lane-changing scenarios and generates smooth, accurate curved lane vectors, demonstrating precise vector modeling capabilities. Second, our cache mechanism successfully preserves rule awareness even when traffic signs move beyond the current map-rule. Additionally, the model exhibits strong semantic understanding by accurately associating multiple rules from individual traffic signs with corresponding lanes.

## 5 Conclusion

We propose PAMR, a novel framework for persistent rule-aware HD map construction. By integrating map-rule co-construction with a cache mechanism, PAMR successfully maintains rule awareness across extended driving sequences, addressing a critical challenge in autonomous navigation. Additionally, we introduce MapDRv2, featuring re-annotated continuous vector labels, to enable comprehensive evaluation of rule-aware mapping systems. Extensive experiments on MapDRv2 demonstrate PAMR's effectiveness in joint vector-rule mapping and consistent rule propagation.

## 6 Limitations

Our current evaluation is limited to the MapDRv2 benchmark. Although we envision PAMR as a universal plug-and-play solution for rule-aware HD mapping, its broader applicability requires validation. Future research will focus on extending PAMR to diverse mapping scenarios and benchmarks to verify its effectiveness as a general-purpose solution.

## 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


Anthropic. 2024. Claude-3. https://www.anthropic.com/news/claude-3-family.

Bai, J.; Bai, S.; Yang, S.; Wang, S.; Tan, S.; Wang, P.; Lin, J.; Zhou, C.; and Zhou, J. 2023. Qwen-VL: A Frontier Large Vision-Language Model with Versatile Abilities. *arXiv preprint arXiv:2308.12966*.

Behrendt, K.; Novak, L.; and Botros, R. 2017. A deep learning approach to traffic lights: Detection, tracking, and classification. In *ICRA*.

Ben Charrada, T.; Tabia, H.; Chetouani, A.; and Laga, H. 2022. Toponet: Topology learning for 3d reconstruction of objects of arbitrary genus. In *Computer Graphics Forum*.

Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuScenes: A Multimodal Dataset for Autonomous Driving. In *CVPR*.

Cao, A.; Wei, X.; and Ma, Z. 2025. FLAME: Frozen Large Language Models Enable Data-Efficient Language-Image Pre-training. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 4080–4090.

Chang, X.; Xue, M.; Liu, X.; Pan, Z.; and Wei, X. 2025. Driving by the Rules: A Benchmark for Integrating Traffic Sign Regulations into Vectorized HD Map. In *CVPR*.

Chen, J.; Wu, Y.; Tan, J.; Ma, H.; and Furukawa, Y. 2024a. Maptracker: Tracking with strided memory fusion for consistent vector hd mapping. In *European Conference on Computer Vision*, 90–107. Springer.

Chen, L.; Sinavski, O.; Hünermann, J.; Karnsund, A.; Willmott, A. J.; Birch, D.; Maund, D.; and Shotton, J. 2024b. Driving with LLMs: Fusing Object-Level Vector Modality for Explainable Autonomous Driving. In *ICRA*.

Choudhary, T.; Dewangan, V.; Chandhok, S.; Priyadarshan, S.; Jain, A.; Singh, A. K.; Srivastava, S.; Jatavallabhula, K. M.; and Krishna, K. M. 2024. Talk2bev: Language-enhanced bird's-eye view maps for autonomous driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 16345–16352. IEEE.

Cui, Y.; Huang, S.; Zhong, J.; Liu, Z.; Wang, Y.; Sun, C.; Li, B.; Wang, X.; and Khajepour, A. 2024. DriveLLM: Charting the Path Toward Full Autonomous Driving With Large Language Models. *IEEE TIV*.

Ding, W.; Qiao, L.; Qiu, X.; and Zhang, C. 2023. Pivotnet: Vectorized pivot learning for end-to-end hd map construction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3672–3682.

Ding, X.; Han, J.; Xu, H.; Liang, X.; Zhang, W.; and Li, X. 2024. Holistic autonomous driving understanding by bird's-eye-view injected multi-modal large models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13668–13677.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *ICLR*.

Fregin, A.; Müller, J.; Krebel, U.; and Dietmayer, K. 2018. The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets. In *ICRA*.

Li, Z.; Wang, W.; Li, H.; Xie, E.; Sima, C.; Lu, T.; Yu, Q.; and Dai, J. 2024. Bevformer: learning bird's-eye-view representation from lidar-camera via spatiotemporal transformers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

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. *arXiv preprint arXiv:2308.05736*.

Liu, Y.; Yuan, T.; Wang, Y.; Wang, Y.; and Zhao, H. 2023. VectorMapNet: End-to-end Vectorized HD Map Learning. In *ICML*.

Loshchilov, I.; and Hutter, F. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

OpenAI. 2024. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.

Sima, C.; Renz, K.; Chitta, K.; Chen, L.; Zhang, H.; Xie, C.; Beißwenger, J.; Luo, P.; Geiger, A.; and Li, H. 2024. DriveLM: Driving with Graph Visual Question Answering. In *ECCV*.

Stallkamp, J.; Schlipsing, M.; Salmen, J.; and Igel, C. 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks*.

Team, G.; Anil, R.; Borgeaud, S.; Wu, Y.; Alayrac, J.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A. M.; Hauth, A.; and et al. 2024. Gemini: A Family of Highly Capable Multimodal Models. *arXiv preprint arXiv:2312.11805*.

Tian, X.; Gu, J.; Li, B.; Liu, Y.; Hu, C.; Wang, Y.; Zhan, K.; Jia, P.; Lang, X.; and Zhao, H. 2024. DriveVLM: The Convergence of Autonomous Driving and Large Vision-Language Models. *arXiv preprint arXiv:2402.12289*.

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*.

Wei, Y.; Wang, Z.; Lu, Y.; Xu, C.; Liu, C.; Zhao, H.; Chen, S.; and Wang, Y. 2024. Editable scene simulation for autonomous driving via collaborative llm-agents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15077–15087.

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*.

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

Xie, M.; Zeng, S.; Chang, X.; Liu, X.; Pan, Z.; Xu, M.; and Wei, X. 2025b. SeqGrowGraph: Learning Lane Topology as a Chain of Graph Expansions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 27166–27175.

Xu, Z.; Zhang, Y.; Xie, E.; Zhao, Z.; Guo, Y.; Wong, K. K.; Li, Z.; and Zhao, H. 2024. DriveGPT4: Interpretable End-to-End Autonomous Driving Via Large Language Model. *IEEE Robotics Autom. Lett.*

Yu, F.; Chen, H.; Wang, X.; Xian, W.; Chen, Y.; Liu, F.; Madhavan, V.; and Darrell, T. 2020. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *CVPR*.

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.; Liu, X.; Pan, Z.; and Wei, X. 2024. Driving with Prior Maps: Unified Vector Prior Encoding for Autonomous Vehicle Mapping. *arXiv preprint arXiv:2409.05352*.

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. *arXiv preprint arXiv:2505.17685*.

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*.

Zhang, G.; Lin, J.; Wu, S.; Luo, Z.; Xue, Y.; Lu, S.; Wang, Z.; et al. 2023. Online map vectorization for autonomous driving: A rasterization perspective. *Advances in Neural Information Processing Systems*, 36: 31865–31877.

Zhang, Z.; Zhang, Y.; Ding, X.; Jin, F.; and Yue, X. 2024. Online vectorized hd map construction using geometry. In *European Conference on Computer Vision*, 73–90. Springer.

Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; and Ren, D. 2020. Distance-IoU loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 12993–13000.

Zhu, Z.; Liang, D.; Zhang, S.; Huang, X.; Li, B.; and Hu, S. 2016. Traffic-Sign Detection and Classification in the Wild. In *CVPR*.