SGC-Lane: Monocular 3D Lane Detection with Standard-Definition Map Guidance and Lane Completion

Fuqiang Jiang[1,2], Wei Li[1,2†], Tianyao Zhao[1,2], and Yu Hu[1,2]

[1] The Research Center for Intelligent Computing Systems, SKLP, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
{jiangfuqiang24s,liwei2019,huyu}@ict.ac.cn
zhaotianyao25@mails.ucas.ac.cn

Abstract. Monocular 3D lane detection offers a cost-effective solution for autonomous driving perception but is inherently challenged by the ill-posed problem of inferring 3D geometry from a 2D image. Standard-Definition (SD) maps offer a lightweight prior for road topology, promising enhanced accuracy for distant or occluded lanes. However, their inherent inaccuracies and representation as road-level centerlines, which misalign with the actual lanes to be perceived, pose significant challenges for direct integration. To tackle this, we propose SGC-Lane, a novel framework that advances monocular 3D lane detection through SD map guidance and lane completion. Specifically, we propose an FV-Guided Map Encoder that generates a lane probability map and structured map features, where the feature channels are aligned with potential lane instances. This probability map directs attention toward lane-specific regions in the front view (FV), refining image features and initializing lane queries by the Local Spatial Cross-Attention module for precise geometry decoding. In addition, to mitigate short-range truncation, we design a Step Consistency Completion (SCC) Head that is used to predict and cluster lane endpoints, effectively recovering missing segments to produce complete 3D lane representations. On the large-scale OpenLane dataset, our method achieves a 65.6% F1-score, showing improved performance against existing approaches with the assistance of only coarse road-level SD map priors. The improvements are especially evident in challenging conditions, such as roads with sharp curves and complex intersections. Code is released at `https://github.com/FuqingJIang/SGC-Lane`.

Keywords: 3D Lane Detection · Monocular Vision · Map Prior · Lane Completion · Autonomous Driving

1 Introduction

Accurate 3D lane detection is a fundamental task in the perception system for autonomous vehicles, providing essential references for planning and decision-

† Corresponding Author.

making. Among various sensing modalities, monocular vision offers a uniquely cost-effective solution. However, it inherently suffers from the ill-posed problem of inferring 3D geometry from a single 2D image. This challenge becomes particularly pronounced in complex scenarios such as heavy occlusion, steep road slopes, and sharp curves, where the visual evidence is often ambiguous or incomplete.

To mitigate this limitation, beyond-line-of-sight information from maps can serve as a powerful cue. Recent works in related tasks such as online vectorized map construction have demonstrated the benefits of leveraging High-Definition (HD) maps as a prior [11, 35, 36]. Standard-Definition (SD) maps offer a more practical alternative due to their wider coverage and lower cost. This has motivated methods like SMERF [18] and SEPT [22] to enhance Bird's-Eye-View (BEV) features with SD map priors via cross-attention mechanisms for lane-topology understanding. However, these methods primarily focus on detecting lane centerlines. Meanwhile, SD maps provide only road-level centerlines rather than the actual physical lane lines. Fundamentally, a road-level centerline represents the topology of a road segment, abstracting a group of adjacent lanes that share the same driving direction into a single geometric curve. This representation gap, along with the inaccuracies that can reach several meters in SD maps, creates a severe misalignment between the SD map prior and the actual lanes to be detected. Thus, effectively utilizing SD map as prior knowledge for accurate 3D lane detection is challenging.

Another challenge arises from prevalent use of sparse point sets to represent lanes [19,20,30]. In this paradigm, the original lane markings are often truncated at the endpoints during the generation of training annotations. This leads to an inability of models to predict the complete lane extent, resulting in fragmented detections, particularly at long range or under occlusion. Recently, EndPoint head [3] attempts to address this by predicting patching distance to endpoints. Nevertheless, this approach is susceptible to occasional inaccuracies in its patching distance predictions. Representing lanes as sparse-points typically enforces a fixed interval between consecutive points along the depth direction (Y-axis) as structural constraint. Leveraging this constraint could ensure better consistency in predicting the offsets from each preset lane point to endpoints, thereby establishing a more robust and geometrically-informed completion strategy.

To address the above challenges, we propose SGC-Lane, a novel framework that leverages SD map priors and an improved lane completion strategy. We design a Front-View Guided Map Encoder that dynamically predicts lane widths from front-view (FV) images to transform coarse road-level centerlines into lane-level boundary priors. The SD map is then encoded to produce a spatial lane probability map and lane-aware map features. This probability map guides the FV features to focus more on lane regions and, together with the map features, is utilized through a Local Spatial Cross-Attention (LSCA) module to initialize lane queries. For robust lane completion, we propose a Step Consistency Completion Head (SCC-Head) that exploits the structured constraint of fixed Y-interval sampling. It predicts consistent offsets from each lane point to the

endpoints, with a clustering mechanism aggregating all predictions to complete missing segments. Additionally, to handle the perspective projection effect in the front view, we employ non-uniform sampling along the lane-width direction in positional embedding. Our contributions are summarized as follows:

- We propose an SD map enhancement and fusion strategy. This includes an FV-Guided Map Encoder that predicts lane widths for each road-level centerline to expand them into multiple lane boundaries, thereby focusing FV features on these potential lane regions, alongside a Local Spatial Cross-Attention (LSCA) module that integrates the expanded lane-aware map features with image context to initialize lane queries.
- We introduce a Step Consistency Completion Head (SCC-Head) for lane completion. It leverages the intrinsic fixed Y-interval constraint of sparse-point representation to consistently predict endpoint offsets from all lane points, and subsequently employs a clustering mechanism for reliably completing truncated lanes.
- We present a complete framework, SGC-Lane, which integrates the above SD map guidance and lane completion. To support evaluation, we supplement the OpenLane [4] dataset with data from Waymo Open Dataset [28] and simulate real-world SD map characteristics. Extensive experiments on this large-scale dataset demonstrate that our method achieves F1-score of 65.6%, outperforming all compared methods in detection performance, with notable gains in challenging scenes like steep slopes and complex intersections.

2 Related Work

2.1 Map Priors in Road Perception

Map priors have been increasingly adopted in 2D lane detection and road topology reasoning. PriorLane [26] employs an encoder-only transformer to integrate image features with grid map embeddings for improved 2D detection. SMERF [18] encodes SD map elements and fuses them with BEV features through cross-attention. Map priors have also gained attention in online vectorized mapping. P-MapNet [11] avoids strict alignment by fusing image features with SD priors via attention, further incorporating HD maps through an autoencoder for refinement. PriorDrive [36] unifies vectorized SD and HD maps as hybrid prior queries to enhance accuracy. HRMapNet [36] initializes queries with historical rasterized HD maps, while SQD-MapNet [31] and StreamMapNet [34] exploit temporal priors from previous frames. In practice, SD maps offer broader coverage and easier accessibility than HD maps. While well-suited for topology reasoning through centerline representations, SD maps lack direct correspondence with actual lanes. This misalignment, along with inherent inaccuracies, presents challenges for leveraging SD priors in accurate 3D lane detection.

2.2 Representations for Lane Detection

Existing lane representations can be broadly categorized into four types: mask-based [9, 25, 33], keypoint-based [12, 20, 27, 30], sparse-point [4, 6, 7, 10, 14, 19,

37], and curve-based [1, 5, 13, 16, 23, 29]. Mask-based approaches such as Once-3dlanes [33] require per-pixel depth prediction, which can yield inaccurate masks and thus hinder vectorized fitting. Keypoint-based methods like BEV-LanDet [30] first discretize the feature map into grids, each associated with a center, and then determine lane keypoints by predicting the offset and confidence of each center. Sparse-point methods exploit the elongated shape of lanes by pre-sampling along the Y-axis and regressing X-offsets and heights for those samples, with both BEV [4, 6, 7, 10, 14, 37] and non-BEV variants [19]. While these sparse-point representations are efficient, they struggle to fully capture lane geometry. Curve-based [1, 5, 13, 16, 23, 29] methods can describe the complete lane shape, but predicting an entire lane with only a few parameters often limits accuracy. The prevailing paradigm remains sparse-point representations. However, it does not fully leverage the inherent fixed Y-interval of sparse-point representation to regularize completion, which can lead to large endpoint prediction errors that are amplified over the long span of lanes along the Y-axis. Our improved completion strategy attempts to alleviate this concern.

2.3 Spatial Position in 3D Lane Detection

Reliable 3D positional cues are central to monocular 3D lane detection. One line projects FV features to fixed ground plane via IPM to form BEV representations, which regularize topology but are sensitive to extrinsics and road slope, suffering from long-range scale distortion [4, 6, 21, 24, 30]. Subsequent works improve these 3D positional cues through virtual image planes for robustness [30], deformable aggregation around projected points [4], and slope-adaptive height maps [21]. A second line avoids explicit BEV, injecting 3D awareness directly into FV features using learnable positional encodings. PETRv2 encodes latent 3D coordinates per pixel [15], while LATR samples learnable plane and uses its FV projections as positional cues [19]. Building on these insights, we adopt a perspective-aware non-uniform sampling strategy that matches FV projection density, offering stronger geometric cues without modifying the decoder.

3 Method

The overall architecture is shown in Fig. 1. Given a front-view image and prior SD map, the FV-Guided Map Encoder aligns the map priors centerlines with FV features and outputs a lane probability map and lane-aware map features. Using the lane probability map, our method performs gated weighting on the original FV features, generating attentive FV features. The query generator constructs per-point queries by combining instance queries with learnable point embeddings. Local Spatial Cross-Attention first obtains a ground position (x, y) for each query by predicting its x and indexing the corresponding y, then uses cross-attention to fuse nearby map prior with the query. A perspective-aware non-uniform sampling plane supplies positional embeddings, balancing high near-range density with extensive far-range coverage. The SCC-Head predicts point-wise completion

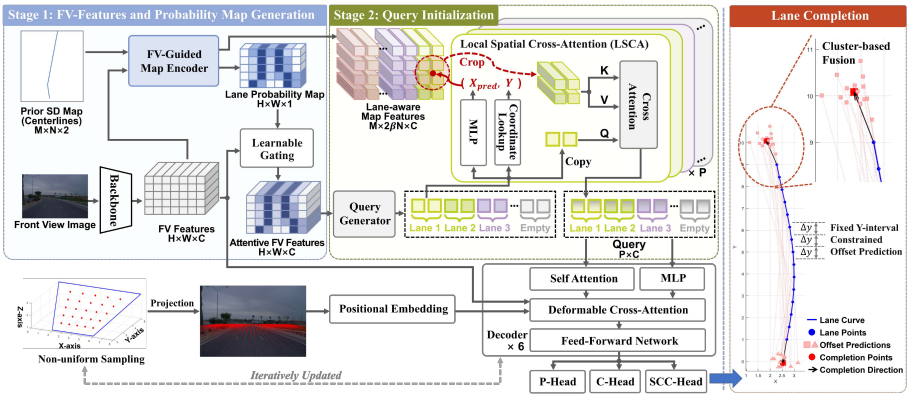

Fig. 1: The overall framework of SGC-Lane, comprising two stages to leverages SD maps. In stage 1, the FV-Guided Map Encoder predicts lane widths to convert SD prior centerlines, producing the lane probability map and map features. These outputs are fused with FV features via the LSCA module in stage 2, to generate refined queries for efficient lane search. Prediction heads output lane geometry, category, and endpoint-offset under a sequence-consistency loss. Lane completion is achieved by clustering the predicted endpoints.

offsets. These offsets are subsequently clustered in a separate decoding phase to derive the endpoints used for final 3D lane completion.

3.1 Prior SD Map Format

To the best of our knowledge, existing public datasets cannot provide both 3D lane annotations and corresponding SD map priors. General autonomous driving datasets such as Argoverse 2 [32] and nuScenes [2] provide SD maps, they lack 3D lane annotations. Conversely, standard 3D lane benchmarks including OpenLane [4], Apollo Synthetic [7], and ONCE-3DLanes [33] do not include map priors. This data scarcity necessitates the synthesis of SD map priors. To bridge this gap, we supplemented the existing benchmark by constructing corresponding SD map. We specifically chose the OpenLane dataset because it is built upon the Waymo Open Dataset [28], which contains HD maps. By extracting and processing these surveyed maps, we successfully generated the SD map priors required for our task.

To construct realistic SD map priors from the HD maps, we employ a sequential degradation pipeline designed to simulate the quality and characteristics of real-world SD maps. Specifically, the pipeline consists of four key operations. **Mask Rendering:** we render the precise HD lane centerlines into regional road surface masks to simulate reduced spatial precision. **Skeleton Extraction:** we apply morphological skeletonization to these thickened masks to derive simplified, road-level centerlines. **Junction Separation:** we detect and disconnect junction points across the newly generated centerlines to handle complex road

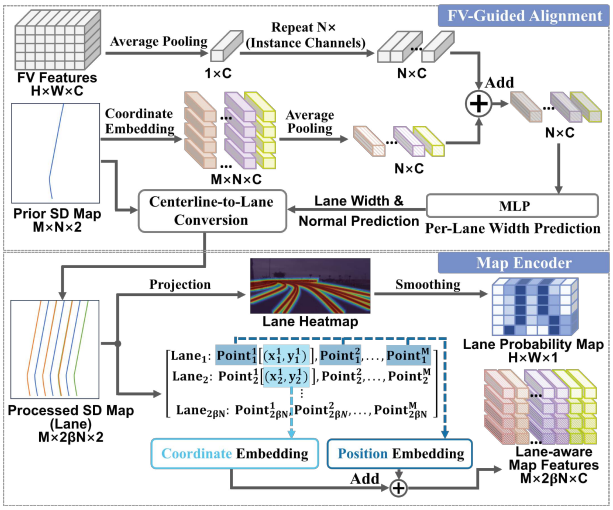

Fig. 2: FV-Guided Map Encoder: A cross-view alignment module that integrates prior SD map centerlines with front-view image features to generate a lane probability map and lane-aware representations.

intersections without topological artifacts. **Spline Resampling:** we resample the separated centerlines via spline interpolation to guarantee a smooth and uniform point distribution along the curves. This entire process naturally introduces geometric simplifications and positional errors at the lane level, effectively mirroring the typical uncertainties of real-world SD maps while preserving overall topological integrity. Detailed generation methodologies and the corresponding code are provided in the supplementary material.

To match the task setting, we retain only centerlines within the spatial window ($x \in [-15\,\text{m}, 15\,\text{m}]$, $y \in [3\,\text{m}, 103\,\text{m}]$), where x represents the lateral direction and y the longitudinal direction. Thus, in this paper, the prior SD map is defined as:

$$\mathcal{M} = \{C_i\}_{i=1}^{N}, \tag{1}$$

$$C_i = [\,p_{i,1}, \ldots, p_{i,M}\,], \tag{2}$$

$$p_{i,j} = (x_{i,j}, y_{i,j}) \in \mathbb{R}^2, \tag{3}$$

where N denotes the number of centerlines in the map, and M is the number of sampled points on each centerline.

3.2 FV-Guided Map Encoder

The prior SD map provides only road-level centerlines that capture the main road topology but contain no lane-level geometric details and can exhibit lateral

positional errors of several meters. To address these limitations and effectively leverage such coarse structural information for lane detection, we propose the FV-Guided Map Encoder. This module, as illustrated in Fig. 2, consists of two key components: FV-Guided Alignment and Map Encoder, which work to encode geometrically consistent map priors. The FV-Guided Alignment component translates road-level priors into lane-level structures. Specifically, it combines FV features with the coordinate embedding of the original centerline to predict a corresponding lane width. Using this predicted width, the component duplicates and offsets the original centerline along its normal direction, deriving left and right boundary for multiple parallel lanes within a unified coordinate space. The output of this alignment process is subsequently fed into the Map Encoder. This component produces two final outputs: a **lane probability map** to highlight regions likely to contain lanes, and **lane-aware map features** that provide essential semantic context.

FV-Guided Alignment. For the i-th centerline with sampled points $p_{i,j}$, we encode every sampled point with a coordinate encoder to obtain point-wise embeddings $f_{i,j}$. These embeddings capture local geometry along the curve and are aggregated by mean pooling described in Eq. (4):

$$\bar{F}_i = \mathrm{AvgPool}\big(\{f_{i,1}, \ldots, f_{i,M}\}\big). \tag{4}$$

Let $\mathbf{F}_{\mathrm{fv}}$ denote the FV features. We obtain a global context vector v_{ctx} by spatial average pooling and broadcast it to all centerlines. For the i-th centerline, We fuse them to obtain $h_i = \phi(\bar{F}_i, v_{\mathrm{ctx}})$ and predict the lane width with a lightweight head described in Eq. (5):

$$\hat{w}_i = \mathrm{MLP}(h_i). \tag{5}$$

During the centerline-to-lane conversion, we expand the single road-level centerline into a lane-level topology through a two-step geometric transformation. First, for each point $p_{i,j}$, we compute the unit tangent and its corresponding unit normal. By translating the original centerline along the normal direction, we generate β parallel lane-level centerlines. Second, we apply the same predicted width $\hat{w}_i$ to each of these newly generated lane-level centerlines, offsetting along their respective normals to derive the final left and right boundary points for all β candidate lanes. This progressive expansion process produces structured lane representations that maintain strict geometric consistency with both the road-level prior and the FV features.

Map Encoder. We employ two lightweight pathways for SD map priors. A gating probability map highlights probable lane regions to mitigate occlusion in one pathway, whereas centerline geometry is encoded in the other to distinguish lanes and generate lane-aware features. This approach fully uses the map priors without, requiring only minimal computational resources.

Output 1: lane probability map. For the i-th road-level centerline, we expand it into β lane-level centerlines and denote the left and right boundaries of its k-th lane-level centerline as $L_{i,k}$ and $R_{i,k}$ respectively. By aggregating all boundaries across the N centerlines and their corresponding β lane-level

centerlines, we obtain the comprehensive sets $L = \bigcup_{i=1}^{N} \bigcup_{k=1}^{\beta} L_{i,k}$ and $R = \bigcup_{i=1}^{N} \bigcup_{k=1}^{\beta} R_{i,k}$. We then project these βN pairs of lane boundaries to the front-view image plane to obtain an FV heatmap representation:

$$\mathcal{B} = \Pi_{\text{gp}\to\text{fv}}\big(L \cup R\big), \tag{6}$$

here, $\Pi_{\text{gp}\to\text{fv}}$ denotes ground-to-image projection. Next, we rasterize the projected $\mathcal{B}$ with Gaussian smoothing to generate the lane probability map:

$$\mathcal{H} = \text{Rasterize}(\mathcal{B}) \in [0,1]^{H\times W\times 1}, \tag{7}$$

where, Rasterize$(\cdot)$ denotes the rasterization operation with Gaussian smoothing. The lane probability map $\mathcal{H}$ guides the generation of lane queries in subsequent processing stages.

Output 2: lane-aware map features. For each centerline, we convert it to left and right lanes through Centerline-to-Lane Conversion module and sample lane points $q_{i,j} = (x_{i,j}, y_{i,j})$. For each point, we construct feature by summing a coordinate embedding and a positional embedding as defined in Eq. (8):

$$\mathbf{M}_{i,j} = E_{\text{coord}}(x_{i,j}, y_{i,j}) + E_{\text{pos}}(j) \in \mathbb{R}^{C}. \tag{8}$$

We combine all embeddings to form the lane-aware map features $\mathbf{M}_{\text{map}}$, which serve as instance-conditioned geometric priors for subsequent processing.

Learnable Gating. We first obtain a channel-aligned prior $\mathcal{W} = \text{proj}(\mathcal{H}) \in \mathbb{R}^{H\times W\times C}$, where proj$(\cdot)$ denotes a learnable channel projection that maps the lane probability map to the image-feature dimension, and then apply a residual gate to $\mathbf{F}_{\text{fv}}$:

$$\tilde{\mathbf{F}}_{\text{fv}} \;=\; \mathbf{F}_{\text{fv}} \;+\; \alpha \cdot \mathcal{W}, \quad \alpha \in (0,1). \tag{9}$$

The attentive features $\tilde{\mathbf{F}}_{\text{fv}}$ are then consumed by the query generator module, providing coarse regional guidance with a learnable strength α.

3.3 Local Spatial Cross-Attention

Local Spatial Cross-Attention injects SD map priors into per-point queries at spatially matched locations. It takes as inputs per-point queries from the query generator and lane-aware map features $\mathbf{M}_{\text{map}} \in \mathbb{R}^{2\beta N\times M\times C}$ alongside their corresponding ground coordinates $\mathbf{C}_{\text{map}} \in \mathbb{R}^{2\beta N\times M\times 2}$ produced by the FV-Guided Map Encoder. Instead of attending to all map tokens globally, this module first predicts a query-specific ground location and then performs cross-attention exclusively over nearby map tokens. This localized mechanism attends precisely to relevant prior features, enhances the alignment between the expanded map priors and image evidence, and maintains minimal computational overhead.

The query generator module provides instance queries $\mathbf{Q}_{\text{lane}}$ and learnable point embeddings $\mathbf{Q}_{\text{point}}$. Per-point queries can be obtained using the following equation.

$$\mathbf{Q} \;=\; \mathbf{Q}_{\text{lane}} \;\oplus\; \mathbf{Q}_{\text{point}} \;\in\; \mathbb{R}^{P\times C}, \tag{10}$$

here, $\oplus$ denotes broadcasted summation, and q_p is the p-th per-point query in $\mathbf{Q}$. We first localize each per-point query by predicting its ground x and pairing it with the preset anchor-y at the query's point index:

$$\hat{x}_p = \mathrm{MLP}(q_p), \quad \hat{y}_p = y_{\mathrm{anchor}}[p]. \tag{11}$$

Given the target $(\hat{x}_p, \hat{y}_p)$, we retrieve locally relevant priors by selecting the K nearest map tokens according to their ground coordinates $\mathbf{C}_{\mathrm{map}}$:

$$\mathrm{idx}_p = \mathrm{TopK}_K\big(-\|\mathbf{C}_{\mathrm{map}} - (\hat{x}_p, \hat{y}_p)\|_2\big). \tag{12}$$

Finally, LSCA applies a cross-attention using the selected tokens as keys and values:

$$\tilde{q}_p = \mathrm{Attn}\big(Q{=}q_p,\ K{=}\mathbf{M}_{\mathrm{map,idx}_p},\ V{=}\mathbf{M}_{\mathrm{map,idx}_p}\big). \tag{13}$$

This way injects precise geometric cues while avoiding global mismatches.

3.4 Step Consistency Completion Head

Building on the EP-Head [3] addressing information loss in sparse-point representations, we propose the Step Consistency Completion Head to enhance sequence coherence. While existing refinements process anchor points independently, our method incorporates a lightweight sequence consistency prior to preserve structural information and stabilize predictions as illustrated in Fig. 3. We employ clustering-based completion to fuse candidates and recover truncated segments.

Offset Predictions. We sample M preset ground-plane anchors $\mathcal{P} = \{p_k\}_{k=1}^{M}$ with monotonically increasing y. Each anchor p_k is parameterized by its position (x_k, y_k, z_k) and an associated feature v_k, the head then regresses two endpoint displacement offsets s_k and e_k relative to p_k. Given point-aware features $P \in \mathbb{R}^{M \times C}$, SCC-Head outputs per-anchor endpoint distances:

$$\hat{s}_k,\ \hat{e}_k\ =\ \mathrm{MLP}(\mathbf{p}_k)\,. \tag{14}$$

We propose a consistency loss that utilizes the inherent fixed Y-interval constraint in sparse point representation. Instead of specifying a step size, we encourage the increments of successive endpoint Y-offsets to be constant along the sequence by minimizing their second-order differences:

$$\mathcal{L}_{\text{sy-cons}} = \frac{1}{M-2} \sum_{k=2}^{M-1} \left| \Delta\hat{s}_k^y - \Delta\hat{s}_{k-1}^y \right|, \tag{15}$$

$$\mathcal{L}_{\text{ey-cons}} = \frac{1}{M-2} \sum_{k=2}^{M-1} \left| \Delta\hat{e}_k^y - \Delta\hat{e}_{k-1}^y \right|, \tag{16}$$

where we define the increments inline as $\Delta\hat{s}_k^y = \hat{s}_{k+1}^y - \hat{s}_k^y$ and $\Delta\hat{e}_k^y = \hat{e}_{k+1}^y - \hat{e}_k^y$. The base objective $\mathcal{L}_{\mathrm{base}}$ is the average L1 distance between predicted and

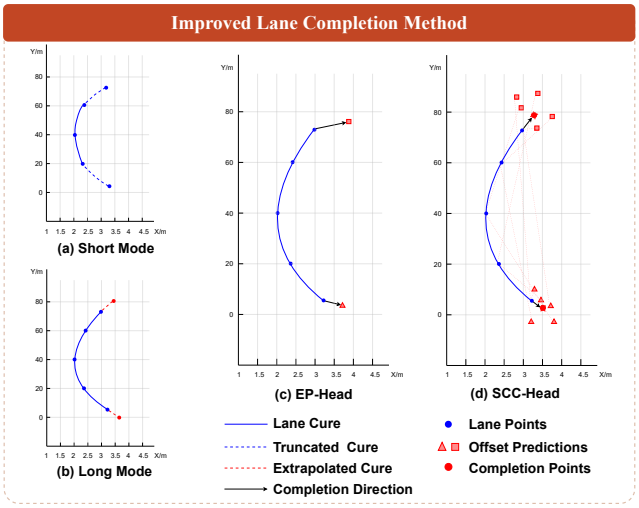

Fig. 3: Improved Lane Completion Method. (a) Short mode: Uses sparse ground truth with truncated ends. (b) Long mode: Extends ends through extrapolation but introduces redundancy. (c) EP-Head: Performs endpoint refinement. (d) SCC-Head (our method): Achieves lane completion through sequence consistency with clustering-based refinement.

ground-truth offsets. With the weighting coefficients λ_{sy} and λ_{ey} as hyperparameters, the total objective is defines as:

$$\mathcal{L}_{\text{SCC}} = \mathcal{L}_{\text{base}} + \lambda_{\text{sy}}\mathcal{L}_{\text{sy-cons}} + \lambda_{\text{ey}}\mathcal{L}_{\text{ey-cons}}. \tag{17}$$

Lane completion. From the predicted per-anchor offsets, we form endpoint candidates by adding them to their anchors, as $p_k + \hat{s}_k$ and $p_k + \hat{e}_k$. We then estimate robust endpoints by clustering and taking the cluster centers:

$$\hat{\mathbf{s}} = \text{Centr}\big(\mathcal{K}(\{\mathbf{c}_k^s\})\big), \quad \hat{\mathbf{e}} = \text{Centr}\big(\mathcal{K}(\{\mathbf{c}_k^e\})\big), \tag{18}$$

where $\mathcal{K}$ is a lightweight clustering. The fused endpoints are combined with fixed Y-interval samples to complete truncated segments.

3.5 Decoder

We adopt the standard LATR [19] decoder, which comprises stacked layers with self-attention, deformable cross-attention to FV features, and feed-forward networks. The per-point queries $\mathbf{Q} \in \mathbb{R}^{P\times C}$ are refined layer-by-layer, with iterative reference-point updates and shared prediction heads at each stage.

Non-uniform sampling plane. Our improvement here is a non-uniform, perspective-aware ground sampling plane. Instead of a uniform grid, we allocate denser Y-samples at near ranges and sparser ones further away, while widening

the horizontal span with distance. This naturally projects more valid ground samples into the FV, enriching the 3D spatial cues for deformable cross-attention without altering the lightweight decoder architecture. Exact construction details are provided in the supplementary material.

4 Experiment

Datasets. We conduct our primary evaluation on the large-scale OpenLane [4] dataset for two key reasons. First, our method is predicated on the availability of map priors. OpenLane is uniquely suitable for our study because it is built upon the Waymo Open Dataset [28], which provides the necessary underlying survey-grade maps. Second, its status as a large-scale, real-world benchmark containing 200K images and numerous challenging scenes provides a rigorous and practical testbed for evaluating perception models. While the Apollo Synthetic [7] dataset relies on synthetically generated environments and the ONCE-3DLanes [33] dataset presents a smaller-scale real-world challenge without corresponding map priors, we still demonstrate the generalizability of our approach on these benchmarks. Extensive evaluation results are provided in the supplementary material. Throughout our main experiments, we follow the standard OpenLane [4] dataset splits and evaluation metrics, and the SD map prior for each frame is constructed as detailed in Sec. 3.1.

Training details. We use ResNet-50 [8] as our backbone and resize inputs to 720×960. Optimization uses AdamW [17] with a weight decay of 0.01. The learning rate is initialized to 2×10^{-4} and follows a cosine annealing schedule. We train for 24 epochs on OpenLane [4] dataset with a global batch size of 32 distributed across $8 \times$ NVIDIA RTX4090 GPUs. We set $\lambda_{\text{sy}} = 1.0$ and $\lambda_{\text{ey}} = 1.0$. All other optimization and hyperparameters follow the baseline LATR [19].

4.1 Quantitative Results

Results on OpenLane. As summarized in Tab. 1, SGC-Lane achieves state-of-the-art detection performance on the OpenLane validation set by attaining an F1-score of **65.6%** with a standard ResNet-50 backbone. This represents a 0.9% F1 improvement over the recent EP-Head [3] at 64.7%, while also surpassing the F1-scores of other leading methods including GLane3D [20] at 63.9% and LATR [19] at 61.9%. Furthermore, with a stronger Swin-B backbone, SGC-Lane achieve a higher F1-score of 68.0%. In terms of localization precision, SGC-Lane maintains competitive geometric accuracy, achieving a near-range error of 0.204m and a far-range error of 0.254m along the X-axis. Notably, when compared directly to the baseline LATR [19], SGC-Lane not only delivers a substantial 3.7% improvement in F1-score but also reduces the X error at both distance ranges. Ultimately, the significant gain in detection capability alongside robust geometrical precision confirms the effectiveness and advancement of our proposed approach.

Table 1: Quantitative results on OpenLane [4] dataset. **Best performance** and second best are highlighted.

Methods	Backbone	F1 ↑	X error (m) ↓		Z error (m) ↓	
			near	*far*	*near*	*far*
PersFormer [4] ECCV 2022	EfficientNet	50.5	0.485	0.553	0.364	0.413
Anchor3DLane [10] CVPR 2023	ResNet-18	53.1	0.300	0.311	0.103	0.139
BEV-LaneDet [30] CVPR 2023	ResNet-34	58.4	0.309	0.659	0.244	0.631
LATR [19] ICCV 2023	ResNet-50	61.9	0.219	0.259	0.075	0.104
LaneCPP [24] CVPR 2024	EfficientNet	60.3	0.264	0.310	0.077	0.177
GLane3D [20] CVPR 2025	ResNet-50	63.9	**0.193**	**0.234**	**0.065**	**0.090**
SC-Lane [21] ICCV 2025	ResNet-50	64.3	0.227	0.251	0.088	0.128
Chang *et al.* [3] CVPR 2025	ResNet-50	64.7	0.205	0.255	0.074	0.105
SGC-Lane(Ours)	ResNet-50	**65.6**	0.204	0.254	0.076	0.107
GLane3D [20] CVPR 2025	Swin-B	66.0	0.170	0.203	0.063	0.087
SGC-Lane(Ours)	Swin-B	68.0	0.197	0.237	0.072	0.101

Table 2: Quantitative results per scenario on the OpenLane [4] dataset. **Best performance** and second best are highlighted. Performance gains are highlighted with blue arrows.

Methods	All	Up& Down	Curve	Extreme Weather	Night	Intersection	Merge& Split
PersFormer [4]	50.5	42.4	55.6	48.6	46.6	40.0	50.7
Anchor3DLane [10]	53.1	45.5	56.2	51.9	47.2	44.2	50.5
BEV-LaneDet [30]	58.4	48.7	63.1	53.4	53.4	50.3	53.7
LATR [19]	61.9	55.2	68.2	57.1	55.4	52.3	61.5
LaneCPP [24]	60.3	53.6	64.4	56.7	54.9	52.0	58.7
GLane3D [20]	63.9	54.1	67.3	**62.0**	57.2	53.4	60.0
SC-Lane [21]	64.3	54.6	70.7	58.1	56.5	56.1	63.0
Chang *et al.* [3]	64.7	55.3	71.5	58.5	58.5	55.5	**65.5**
SGC-Lane(Ours)	**65.6**	**56.9**	**72.4**	58.7	**58.6**	**57.1**	**65.5**
Improvement	↑0.9	↑1.6	↑0.9	-	↑0.1	↑1.0	↑2.5

Scenario-wise Analysis. Beyond overall performance, we further evaluate SGC-Lane across challenging scenarios in the OpenLane [4] validation set. As shown in Tab. 2, our method delivers superior F1-scores against existing approaches across multiple complex driving environments. It achieves peak F1-scores in geometrically challenging scenes including Curve at 72.4% and Merge& Split at 65.5%, while maintaining strong detection in topologically complex situations like Intersection at 57.1% and Up&Down at 56.9%. Additionally, it secures the leading position under Night conditions at 58.6%. In extreme weather, our method records a highly competitive second-place F1-score of 58.7%, slightly trailing GLane3D [20] due to architectural differences. Under such conditions with fragmented lane segments, GLane3D's bottom-up methodology—first detecting dense keypoints then grouping them—proves more effective at capturing these elements. Overall, these findings demonstrate our method's robust capability in handling diverse driving situations.

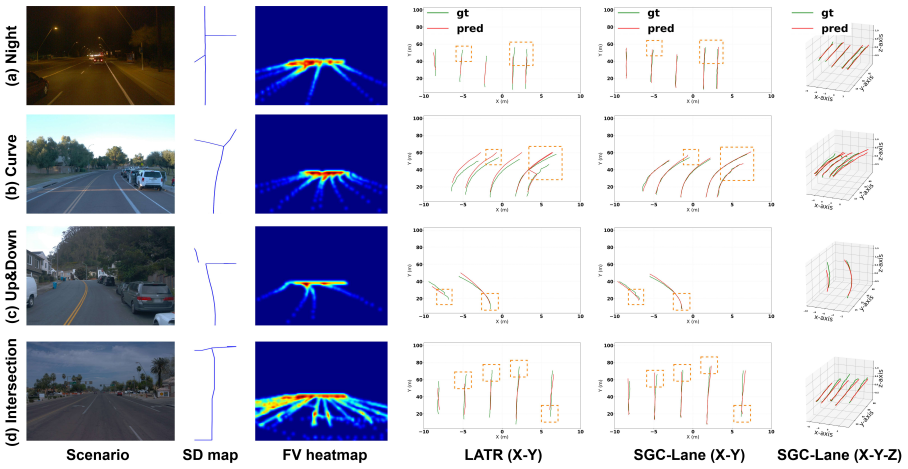

Fig. 4: Qualitative results on the OpenLane [4] validation set. The first column shows four typical driving scenes: (a) Night, (b) Curve, (c) Up&Down, and (d) Intersection. The second and third columns present the input SD map and the FV heatmap. The fourth and fifth columns compare the X-Y plane predictions of LATR [19] and our SGC-Lane. The last column visualizes the 3D estimation results of SGC-Lane. Across all visualizations, ground truth is shown in green and predictions in red.

4.2 Qualitative Comparison

Fig. 4 illustrates typical cases across Night, Curve, Up&Down, and Intersection scenes. The FV heatmap shows that, despite biased SD maps, FV-Guided alignment concentrates attention on plausible lane regions. From the 2D projections, introducing the SD map prior achieves more accurate positions than LATR [19]. Regarding completeness, our regressed lanes are distinctly more continuous, evidencing the SCC-Head's effectiveness for endpoint completion. Finally, the 3D view confirms our predictions remain highly accurate in the 3D domain.

4.3 Ablation Studies

We conduct comprehensive designed ablation studies on OpenLane [4] dataset to validate the effectiveness of our proposed components under the same experimental settings as our main training procedure. The detailed analysis of model complexity and additional experiments are provided in the supplementary. Tab. 3 presents the contributions of individual components to the overall performance. Through systematic ablation studies, we have obtained key findings.

Evaluation of Individual Module Contributions. Introducing the SD map prior significantly enhances detection performance, increasing an F1-score to 63.4% (1.5% ↑), while enhancing localization accuracy, with the near-range X error reduced to 0.204m (0.015 ↓). Although SD maps can be biased or loosely aligned, our design, comprising FV-Guided alignment, learnable gating, and local

Table 3: Ablation of components with error metrics. **Best performance** and second best are highlighted.

Exp	SD-Map	SCC-Head	NUS-Plane	F1 ↑	Gain ↑	X error (m) ↓		Z error (m) ↓	
						near	*far*	*near*	*far*
1				61.9	-	0.219	0.259	0.075	0.104
2	✓			63.4	1.5	0.208	**0.241**	**0.072**	**0.103**
3		✓		64.4	2.5	0.231	0.284	0.079	0.111
4	✓	✓		65.3	3.4	0.217	0.256	0.078	0.108
5		✓	✓	64.7	2.8	0.233	0.275	0.078	0.109
6	✓	✓	✓	**65.6**	**3.7**	**0.204**	0.254	0.076	0.107

Table 4: Results comparison between EP-Head [3] and SCC-Head on the OpenLane [4] validation set, with LATR [19] as the baseline.

Methods	F1 ↑
LATR [19]	61.9
LATR+EP-Head [3]	63.1
LATR+SCC-Head (Ours)	64.4

injection via LSCA, uses the prior as soft guidance rather than hard labels, amplifying reliable geometric cues and suppressing misaligned ones. As a result, the model attains both higher recall and lower positional error.

Used SCC-Head alone, it improves F1-score to 64.4% (2.5% ↑) but increases localization errors (X: 0.231/0.284m, Z: 0.079/0.111m) due to completion mechanism, highlighting a trade-off between completeness and precision. Under a controlled swap on LATR [19] backbone, as shown in Tab. 4, SCC-Head attains higher detection performance than EP-Head [3](64.4% vs 63.1%) while preserving the same training recipe, indicating stronger completion effectiveness.

Complementary effects between modules. By integrating complementary SD map prior with SCC-Head, we achieve F1-score of 65.3% while maintaining baseline-level localization accuracy. The prior narrows the endpoint search and makes completion more reliable, so F1-score gains of SCC-Head turn into improvements without hurting geometric accuracy. Adding perspective-aware non-uniform sampling plane results in the full model with F1-score of 65.6%, X error 0.204/0.254m, and Z error 0.076/0.107m, further reducing lateral error.

5 Conclusion

In this paper, we present SGC-Lane, a novel and effective framework that incorporates SD map guidance and lane completion for accurate monocular 3D lane detection. We design an FV-Guided Map Encoder that predicts lane widths from image features to expand road-level map centerlines into lane boundaries. These expanded geometric structures are then fused with the image evidence through the Local Spatial Cross-Attention module to generate prior-enhanced

lane queries. Furthermore, the proposed Step Consistency Completion Head effectively leverages the fixed Y-interval constraint of sparse-point representation to reliably complete truncated lanes by predicting offsets to the lane endpoints and aggregating predictions via clustering. On the supplemented OpenLane dataset, our method achieves a state-of-the-art 65.6% F1-score, demonstrating improved performance especially in challenging scenarios despite relying on only coarse road-level priors from SD maps. For the future work, we plan to explore temporal fusion of sequential predictions and incorporate more robust depth estimation techniques to further enhance detection accuracy and robustness.

Acknowledgements

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

References

1. Bai, Y., Chen, Z., Liang, P., Song, B., Cheng, E.: Curveformer++: 3d lane detection by curve propagation with temporal curve queries and attention. IEEE Transactions on Intelligent Transportation Systems (2025)
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. Chang, Y., Huang, J., Wang, X., Ye, Y., Liang, Z., Shan, Y., Du, D., Wang, X.: Rethinking lanes and points in complex scenarios for monocular 3d lane detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6802–6811 (2025)
4. Chen, L., Sima, C., Li, Y., Zheng, Z., Xu, J., Geng, X., Li, H., He, C., Shi, J., Qiao, Y., et al.: Persformer: 3d lane detection via perspective transformer and the openlane benchmark. In: European Conference on Computer Vision. pp. 550–567. Springer (2022)
5. Feng, Z., Guo, S., Tan, X., Xu, K., Wang, M., Ma, L.: Rethinking efficient lane detection via curve modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17062–17070 (2022)
6. Garnett, N., Cohen, R., Pe'er, T., Lahav, R., Levi, D.: 3d-lanenet: end-to-end 3d multiple lane detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2921–2930 (2019)
7. Guo, Y., Chen, G., Zhao, P., Zhang, W., Miao, J., Wang, J., Choe, T.E.: Gen-lanenet: A generalized and scalable approach for 3d lane detection. In: European Conference on Computer Vision. pp. 666–681. Springer (2020)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
9. Hou, Y., Ma, Z., Liu, C., Loy, C.C.: Learning lightweight lane detection cnns by self attention distillation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1013–1021 (2019)

10. Huang, S., Shen, Z., Huang, Z., Ding, Z.h., Dai, J., Han, J., Wang, N., Liu, S.: Anchor3dlane: Learning to regress 3d anchors for monocular 3d lane detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17451–17460 (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. Jin, D., Park, W., Jeong, S.G., Kwon, H., Kim, C.S.: Eigenlanes: Data-driven lane descriptors for structurally diverse lanes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17163–17171 (2022)
13. Kim, Z.: Robust lane detection and tracking in challenging scenarios. IEEE Transactions on Intelligent Transportation Systems **9**(1), 16–26 (2008)
14. Liu, L., Chen, X., Zhu, S., Tan, P.: Condlanenet: a top-to-down lane detection framework based on conditional convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3773–3782 (2021)
15. Liu, Y., Yan, J., Jia, F., Li, S., Gao, A., Wang, T., Zhang, X.: Petrv2: A unified framework for 3d perception from multi-camera images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3262–3272 (2023)
16. Liu, Y., Chen, H., Shen, C., He, T., Jin, L., Wang, L.: Abcnet: Real-time scene text spotting with adaptive bezier-curve network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9809–9818 (2020)
17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
18. Luo, K.Z., Weng, X., Wang, Y., Wu, S., Li, J., Weinberger, K.Q., Wang, Y., Pavone, M.: Augmenting lane perception and topology understanding with standard definition navigation maps. In: IEEE International Conference on Robotics and Automation. pp. 4029–4035. IEEE (2024)
19. Luo, Y., Zheng, C., Yan, X., Kun, T., Zheng, C., Cui, S., Li, Z.: Latr: 3d lane detection from monocular images with transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7941–7952 (2023)
20. Öztürk, H.İ., Kalfaoğlu, M.E., Kilinc, O.: Glane3d: Detecting lanes with graph of 3d keypoints. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27508–27518 (2025)
21. Park, C., Seo, E., Hwang, J., Lim, J.: Sc-lane: Slope-aware and consistent road height estimation framework for 3d lane detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 28407–28416 (2025)
22. Pei, M., Shan, J., Li, P., Shi, J., Huo, J., Gao, Y., Shen, S.: Sept: Standard-definition map enhanced scene perception and topology reasoning for autonomous driving. IEEE Robotics and Automation Letters **10**(7), 7126–7133 (2025). https://doi.org/10.1109/LRA.2025.3574978
23. Pittner, M., Condurache, A., Janai, J.: 3d-splinenet: 3d traffic line detection using parametric spline representations. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 602–611 (2023)
24. Pittner, M., Janai, J., Condurache, A.P.: Lanecpp: Continuous 3d lane detection using physical priors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10639–10648 (2024)
25. Qin, Z., Wang, H., Li, X.: Ultra fast structure-aware deep lane detection. In: European Conference on Computer Vision. pp. 276–291. Springer (2020)
26. Qiu, Q., Gao, H., Hua, W., Huang, G., He, X.: Priorlane: A prior knowledge enhanced lane detection approach based on transformer. In: IEEE International Conference on Robotics and Automation. pp. 5618–5624. IEEE (2023)

27. Qu, Z., Jin, H., Zhou, Y., Yang, Z., Zhang, W.: Focus on local: Detecting lane marker from bottom up via key point. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14122–14130 (2021)
28. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2446–2454 (2020)
29. Van Gansbeke, W., De Brabandere, B., Neven, D., Proesmans, M., Van Gool, L.: End-to-end lane detection through differentiable least-squares fitting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 0–0 (2019)
30. Wang, R., Qin, J., Li, K., Li, Y., Cao, D., Xu, J.: Bev-lanedet: An efficient 3d lane detection based on virtual camera via key-points. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1002–1011 (2023)
31. 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)
32. 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)
33. Yan, F., Nie, M., Cai, X., Han, J., Xu, H., Yang, Z., Ye, C., Fu, Y., Mi, M.B., Zhang, L.: Once-3dlanes: Building monocular 3d lane detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17143–17152 (2022)
34. 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)
35. 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)
36. 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)
37. Zheng, T., Huang, Y., Liu, Y., Tang, W., Yang, Z., Cai, D., He, X.: Clrnet: Cross layer refinement network for lane detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 898–907 (2022)