HSDF-Lane: Height-Aligned Signed Distance Field with Semantic Lane Prior for 3D Lane Detection

Jiyong Boo, Byeongin Joung, Hyemin Yang, and Kuk-Jin Yoon*

KAIST, Republic of Korea
{boojiyong, byeonginjoung, hyemin0806, kjyoon}@kaist.ac.kr

Abstract. Monocular 3D lane detection plays a critical role in autonomous driving, yet recovering reliable 3D geometry from a single image remains challenging due to inherent depth ambiguity. Prior methods project image features into Bird's-Eye-View (BEV) space under a flat-ground assumption, causing geometric distortion on real-world roads. Recent methods instead predict explicit height maps to capture non-planar surfaces, but still rely on sparse anchor-based regression and exploit the recovered geometry merely for spatial transformation rather than semantic understanding. To overcome these limitations, we propose **HSDF-Lane**, which implicitly models the road surface as a Height-aligned Signed Distance Field (HSDF) over a densely sampled 3D feature volume. Through differentiable rendering, the HSDF jointly produces an accurate height map and surface-aligned features. We further introduce Lane-aware Semantic Positional Encoding (LSPE), which injects a lane-existence prior derived from the surface-aligned features into the transformer queries, coupling geometric structure with semantic guidance. Extensive experiments on the OpenLane benchmark show that HSDF-Lane achieves state-of-the-art performance in both 3D lane detection and height map estimation. The code is available at `https://github.com/JiyongBoo/HSDF-Lane`.

Keywords: 3D Lane Detection · Height Estimation · Signed Distance Field

1 Introduction

3D lane detection is a key perception task for autonomous driving, as it provides essential information about road structure, vehicle localization, and trajectory planning. Monocular camera-based research is particularly attractive due to its cost-effective deployment and simple sensor configuration, while still providing rich appearance cues for identifying lane markings.

Recent advances in deep neural networks have significantly improved visual understanding from monocular images, enabling remarkable progress in 2D lane detection [5,15,16,18,22,28,30,39,42–44,59]. However, these approaches inherently lack the geometry information required for reliable 3D applications. Motivated by the need to recover 3D geometry from monocular images for downstream tasks, monocular 3D lane detection has been extensively studied in recent years. Early works primarily adopted Inverse Perspective Mapping (IPM) to transform image features into Bird's-Eye-View

* Corresponding author

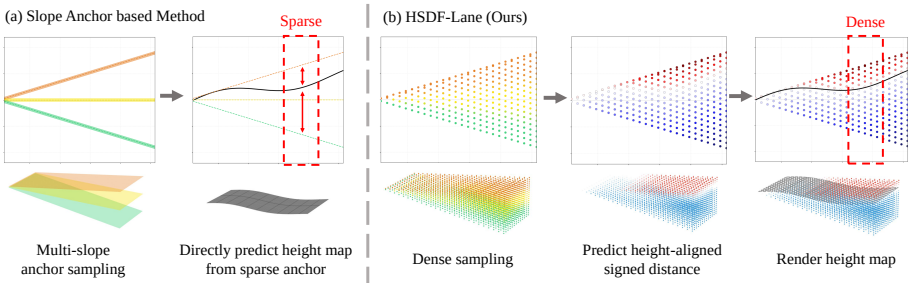

Fig. 1: Comparison between slope-anchor methods and HSDF-Lane. (a) Previous approaches [31, 32] rely on a small number of predefined slope anchors and directly regress the height map, which leads to sparse sampling and potential geometric misalignment. (b) HSDF-Lane densely samples the vertical space and predicts a height-aligned signed distance from each sample, enabling dense supervision and geometrically consistent rendering.

(BEV) representations [3, 6, 9, 17, 29, 34, 60]. However, BEV transformations relying on flat-ground assumptions often suffer from severe geometric misalignment on non-planar roads. To alleviate this limitation, subsequent methods introduced depth estimation branches or Lift-Splat-Shoot (LSS) [33] mechanisms to lift image features into 3D space [23, 26, 35]. Nevertheless, depth-based lifting remains unstable in far-range regions or rippled surfaces due to the scale ambiguity of monocular depth estimation and the amplification of projection errors during 3D reconstruction. Concurrently, lane query-based methods [24, 25, 36] directly predict 3D lanes from 2D features using 3D positional encoding (PE) to compensate for insufficient geometric cues.

Recent approaches, HeightLane [32] and SC-Lane [31], introduce a height map based architecture for monocular 3D lane detection. Unlike depth, which measures the line-of-sight distance from the camera, a height map represents the absolute vertical coordinate at each ground-plane location on the BEV grid. The height map-based modeling significantly alleviates geometric inconsistencies in BEV space. However, the framework still approximates complex road geometry using a sparse set of predefined planar anchors. Existing slope anchors rely on planar assumptions, which inherently introduce geometric distortion on complex roads. Furthermore, as distance increases, the spatial gaps between these discrete anchors widen significantly, making it inherently difficult to accurately regress the absolute road height, as shown in Fig. 1 (a).

To address the limitations discussed above, we propose HSDF-Lane, a novel 3D lane detection framework that combines dense spatial sampling with implicit surface modeling and effective utilization of spatially aligned features. Rather than relying on sparse slope anchors, HSDF-Lane densely samples the scene along the vertical axis to capture complex road surfaces. The resulting volume is modeled with a Height-aligned Signed Distance Field (HSDF), whose differentiable rendering naturally produces both an explicit height map and spatially consistent features, as illustrated in Fig. 1(b). To stabilize the HSDF representation and enforce physically consistent surfaces, a height-directional Eikonal loss is incorporated during optimization. Furthermore, to address the limited semantic exploitation in previous approaches, we introduce Lane-aware Se-

mantic Positional Encoding (LSPE). Specifically, a lane heatmap predicted from the rendered HSDF features is converted into a semantic embedding, which then augments the height-based positional encoding to provide semantic guidance for queries. By jointly leveraging dense surface representations and semantic priors, HSDF-Lane enables robust height estimation and accurate 3D lane detection.

In summary, the main contributions of this paper are as follows:

- We propose a dense 3D sampling framework with a Height-aligned Signed Distance Field (HSDF) representation and differentiable rendering pipeline, which replaces sparse anchor-based height regression to produce surface-aligned features and an accurate height map.
- We propose Lane-aware Semantic Positional Encoding (LSPE), which incorporates lane existence probabilities derived from rendered features as semantic priors for transformer queries.
- Extensive experiments on the OpenLane benchmark demonstrate that HSDF-Lane significantly improves both height map estimation and 3D lane detection accuracy, achieving state-of-the-art performance.

2 Related Work

2.1 3D Lane Detection

Motivated by the rapid progress in 2D lane detection [5, 7, 11, 13–16, 18, 22, 28, 30, 37, 39, 40, 42–44, 48, 50, 51, 58, 59, 62], recent works increasingly explore monocular 3D lane detection to meet the demands of autonomous driving.

Early approaches [6, 9, 17, 34] primarily utilized IPM to transform front-view (FV) features into a BEV representation. However, the planar-ground assumption and discretization inherent in IPM often introduce geometric bias and information loss. To mitigate such limitations, subsequent studies [3,46,60] incorporate learnable view transformations or feature alignment mechanisms. PersFormer [3] employs Deformable Cross-Attention (DCA) for spatial feature transformation, while BEV-LaneDet [46] introduces a virtual camera module to improve robustness under data augmentation. PVALane [60] proposes view-agnostic feature alignment to reduce the discrepancy between FV and BEV representations. Other studies [19, 35] adopt Lift-Splat-Shoot (LSS) [33] as an alternative strategy for constructing BEV representations. Recently, GLane3D [29] achieves strong performance among IPM-based methods by introducing non-uniform BEV sampling and a graph-based lane assembly scheme.

To overcome the geometric misalignment introduced by BEV transformations, anchor-based and query-based methods [1, 12, 24, 25] have been proposed to predict 3D lanes directly from FV images. Recent studies further improve direct 3D lane prediction through lane-aware attention and enhanced lane representations [2, 36]. In particular, SparseLaneSTP [36] demonstrates notable gains by leveraging spatio-temporal priors and a novel curve representation.

Recognizing the importance of geometric cues, several approaches explicitly incorporate depth estimation. SALAD [52] reconstructs 3D lanes by jointly predicting depth and lane segmentation in the image plane. More recently, Depth3DLane [26] distills

knowledge from foundation depth models [54], and DB3D-L [23] constructs BEV features with explicit depth supervision. Distinct from these depth-dependent approaches, HeightLane [32] introduces explicit height map estimation with multi-slope anchors and height-guided DCA for BEV transformation, and SC-Lane [31] further improves it with slope-aware adaptive features and temporal consistency.

2.2 Road Height Estimation

Road height estimation has been studied in several related contexts, including BEV-based perception, aerial elevation mapping, and ground height prediction. (i) *Object-centric height prediction*, exemplified by several existing approaches [49,53], which estimate height as an auxiliary cue for BEV-based perception and 3D object detection. Recent frameworks, including Height3D [57], HeightAware-BEV [61], and BEVSpread [47], further predict height distributions to facilitate view transformation and 2D-to-3D mapping. (ii) *Aerial imagery height estimation*, represented by HeightFormer [4], which focuses on large-scale elevation mapping and differs substantially from onboard driving scenarios. (iii) *Ground height estimation*, such as HeightMapNet [38], which leverages multi-camera inputs and map priors but relies on discretized formulations due to limited dense supervision.

HeightLane [32] and SC-Lane [31] extend road height estimation to monocular 3D lane detection by introducing dense BEV height maps for explicit road-surface modeling, while SC-Lane additionally establishes unified metrics for quantitative evaluation. However, both methods remain *anchor-based*, requiring one-shot regression of absolute height from sparse and slope-dependent projections, which becomes particularly challenging on distant or highly non-planar roads.

Meanwhile, implicit field representations [27] have shown strong capability for continuous 3D modeling [45,55] in driving scenes. SurroundSDF [21] adopts an SDF-based formulation with Eikonal-style regularization for camera-only surface perception, and InfiniDepth [56] models depth as a neural implicit field enabling continuous, arbitrary-resolution querying. Despite their effectiveness, these approaches primarily target general scene geometry, rather than explicitly optimizing dense road height maps on a BEV grid. Motivated by this gap, our work leverages an HSDF-based implicit formulation tailored to road height estimation for 3D lane detection.

3 Method

We define the vehicle coordinate system where the x-axis points to the forward (longitudinal) direction and the y-axis points to the left (lateral) direction in the BEV. The region of interest is quantized into a regular BEV grid $\mathcal{G}$ of size $H \times W$, where each grid cell (i, j) corresponds to a physical location (x_i, y_j). Given a single front-view image $\mathcal{I}$, our goal is to predict 3D lanes and a height map on the grid. Specifically, we estimate a dense height map $\mathcal{H} \in \mathbb{R}^{H \times W}$, where each value represents the vertical height z [32]. For 3D lane detection, we adopt the key-point based approach [46] to predict a lane segmentation map $\mathcal{S} \in \mathbb{R}^{H \times W \times 1}$, lateral offsets $\mathcal{O} \in \mathbb{R}^{H \times W \times 2}$, and embeddings $\mathcal{E} \in \mathbb{R}^{H \times W \times C_{\text{emb}}}$. These dense predictions are clustered to form lane instances, which are subsequently lifted into 3D space using the estimated road surfaces.

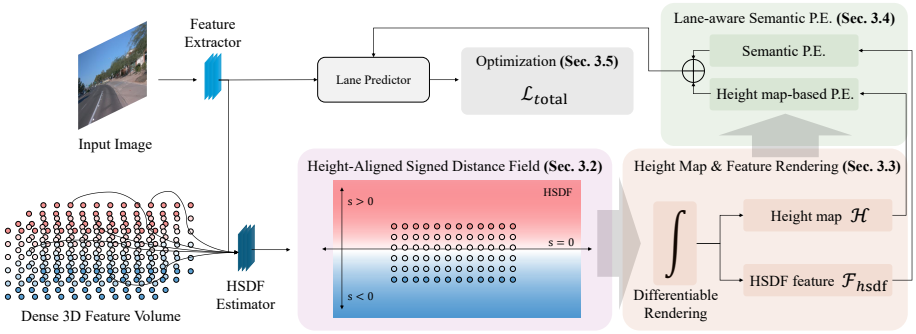

Fig. 2: Overview of HSDF-Lane. Given a monocular image, dense 3D features are constructed by vertically sampling the BEV grid. The HSDF estimator predicts a Height-aligned Signed Distance Field (HSDF) (Sec. 3.2), and following differentiable rendering (Sec. 3.3) produces a height map $\mathcal{H}$ and an HSDF-based feature $\mathcal{F}_{hsdf}$. The feature is further enhanced by Lane-aware Semantic Positional Encoding (LSPE) (Sec. 3.4), where a semantic positional encoding derived from a predicted HSDF feature $\mathcal{F}_{hsdf}$ is combined with height-based positional encoding to provide a semantic prior for queries. The refined queries are decoded by the lane predictor, and the entire framework is optimized end-to-end (Sec. 3.5).

3.1 Overview

Fig. 2 illustrates the overall pipeline of HSDF-Lane. The input front-view image $\mathcal{I}$ is encoded by a backbone to obtain 2D features $\mathcal{F}_{FV}$. For each BEV cell in $\mathcal{G}$, we densely sample points along the vertical z-axis and project them onto the image plane using intrinsics $\mathbf{K}$ and extrinsics $\mathbf{T}$ to gather corresponding image features. Stacking the sampled features forms a dense 3D feature volume. An MLP then predicts the *Height-aligned Signed Distance Field (HSDF)*, which represents the vertical signed distance to the road surface. HSDF enables differentiable rendering of both the height map $\mathcal{H}$ and the HSDF-based feature map $\mathcal{F}_{hsdf}$. To incorporate semantic cues from $\mathcal{F}_{hsdf}$ into lane decoding, we introduce Lane-aware Semantic Positional Encoding (LSPE). A BEV lane heatmap $\mathcal{M} \in \mathbb{R}^{H\times W\times 1}$ is predicted from $\mathcal{F}_{hsdf}$ and used to augment BEV queries as an additional positional encoding for the transformer-based lane head. The enriched BEV queries are updated via deformable cross-attention using 3D reference points derived from $\mathcal{H}$. Finally, the updated queries are fed into the 3D lane head to produce the final 3D predictions, $(\mathcal{S}, \mathcal{O}, \mathcal{E})$.

3.2 Dense 3D Feature Sampling & HSDF Modeling

Unlike previous methods that rely on sparse slope anchors, we perform dense sampling along the vertical axis to capture detailed road geometry.

For each grid cell (x_i, y_j), the vertical search range $[z_{\min}, z_{\max}]$ is defined proportional to the longitudinal distance x_i:

$$z_{\min,\max} = x_i \cdot \tan(\pm\theta), \tag{1}$$

where θ denotes the predefined maximum slope angle. Within this vertical range, we sample N points along the z-axis. To achieve dense sampling across all distances by allocating fewer points at near regions and more points at far regions, we introduce an adaptive sampling strategy based on a target *Vertical Sampling Resolution* δ_z. The required number of samples N for a given cell (x_i, y_j) is determined as:

$$N(x_i, y_j) = \max\left(2,\ \left\lceil \frac{z_{\max} - z_{\min}}{\delta_z} \right\rceil + 1\right). \tag{2}$$

To facilitate the softmax operation during the subsequent differentiable rendering phase (detailed in Sec. 3.3), we strictly enforce a minimum of two samples ($N \geq 2$) in near-field regions. The adaptive sampling rule in Eq. 2 guarantees that the spacing between adjacent samples never exceeds δ_z. Consequently, we obtain a set of discrete 3D sampling points $\mathcal{P}_{sampled} = \{(x_i, y_j, z_k) \mid i \in [0, H), j \in [0, W), k \in [0, N)\}$, where the z-coordinates are uniformly partitioned:

$$z_k = z_{\min} + \frac{k}{N-1}(z_{\max} - z_{\min}), \quad k = 0, \ldots, N-1. \tag{3}$$

Next, we project the 3D sampling points $\mathcal{P}_{sampled}$ onto the image plane using the camera parameters to retrieve pixel-aligned features from the 2D backbone feature $\mathcal{F}_{FV}$. By stacking the retrieved features along the vertical z-axis up to N, we construct a dense 3D feature volume $\mathcal{V} \in \mathbb{R}^{H \times W \times N_{max} \times C}$. Invalid sample positions are explicitly handled via a padding mask during the rendering phase.

To implicitly model the road surface geometry within $\mathcal{V}$, we propose the *Height-aligned Signed Distance Field (HSDF)*. Unlike conventional SDF formulations that measure isotropic distances in 3D space, our HSDF is strictly *axis-aligned* to the vertical direction, making it directly interpretable for height estimation. We define the ground-truth HSDF value s at a point (x, y, z) as:

$$s(x, y, z) = z - \mathcal{H}_{GT}(x, y), \tag{4}$$

where $\mathcal{H}_{GT}(x, y)$ is the ground truth road height. A value of $s = 0$ indicates the exact road surface, while the sign represents the relative vertical position. Finally, we employ an MLP with a skip connection to predict the HSDF value $\hat{s}$ using the feature volume $\mathcal{V}$:

$$\hat{s}(x_i, y_j, z_k) = \text{MLP}\left(\text{cat}\left[\mathcal{V}(x_i, y_j, z_k),\ \gamma(x_i, y_j, z_k)\right]\right), \tag{5}$$

where $\gamma(\cdot)$ represents the positional encoding [27].

3.3 HSDF-based Differentiable Height Map & Feature Rendering

In our framework, the road surface is implicitly represented by the coordinates where the HSDF value is zero, $s(x, y, z) = 0$. To extract the explicit height map from the HSDF representation, we employ a differentiable soft argmin operation. For each grid cell (x_i, y_j), we compute probability weights w_k by applying a softmax function over the valid sampled indices k:

$$w_k(x_i, y_j) = \text{softmax}_k\left(-\frac{|\hat{s}(x_i, y_j, z_k)|}{\tau(x_i)}\right). \tag{6}$$

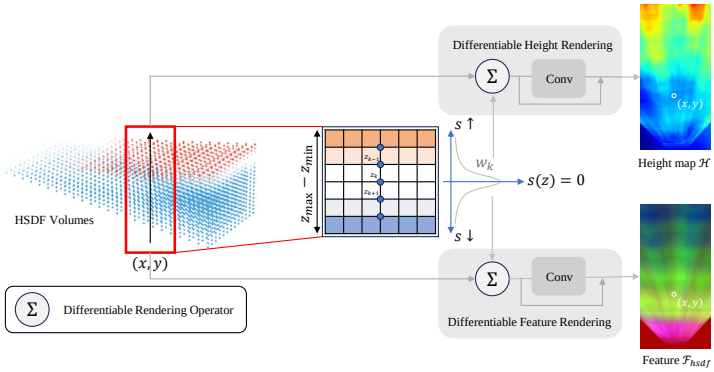

Fig. 3: HSDF-based differentiable rendering. For each BEV location (x, y), probability weights w_k are computed from the predicted HSDF values using a softmax operator. These weights are used to aggregate vertical samples along the ray, producing the rendered height map $\mathcal{H}$ and surface-aligned features $\mathcal{F}_{hsdf}$.

Invalid indices where the sample exceeds the upper boundary ($z_k > z_{\max}$) are explicitly masked out prior to the softmax operation. Here, the temperature $\tau(x_i)$ controls the sharpness of the weight distribution. To ensure consistent optimization behavior across varying vertical search ranges, we scale τ adaptively:

$$\tau(x_i) = \tau_0 \cdot \text{clip}\left(\frac{z_{\max} - z_{\min}}{\delta_z},\ 0.3,\ 3.0\right), \tag{7}$$

where τ_0 is a learnable base temperature, and δ_z is the fixed vertical sampling interval defined in Sec. 3.2. The clipping operation bounds the scaling factor to $[0.3, 3.0]$, preventing numerical instability from near-zero temperatures while avoiding excessive smoothing that would result from large temperatures.

Using the weights $w_k(x_i, y_j)$, we first compute a preliminary continuous height map $\tilde{\mathcal{H}}$. Instead of simply aggregating the discrete sample coordinates z_k, we treat the predicted HSDF value $\hat{s}$ as a local residual to refine the coarse sampling. Specifically, each sample z_k is shifted toward the surface using the predicted distance $\hat{s}$, and the refined coordinates are aggregated via a weighted sum:

$$\tilde{\mathcal{H}}(x_i, y_j) = \sum_{k=0}^{N-1} w_k(x_i, y_j) \cdot (z_k - \hat{s}(x_i, y_j, z_k)). \tag{8}$$

Because this ray-wise rendering operates independently for each grid cell, it may introduce local spatial discontinuities. To improve spatial smoothness while maintaining stable optimization in the early training stages, we apply a zero-initialized 3×3 residual convolution over $\tilde{\mathcal{H}}$ to obtain the final height map $\mathcal{H}$.

Concurrently, to extract semantic features corresponding to the road surface for lane understanding, we compute a preliminary feature map $\tilde{\mathcal{F}}_{hsdf}$ by aggregating the feature

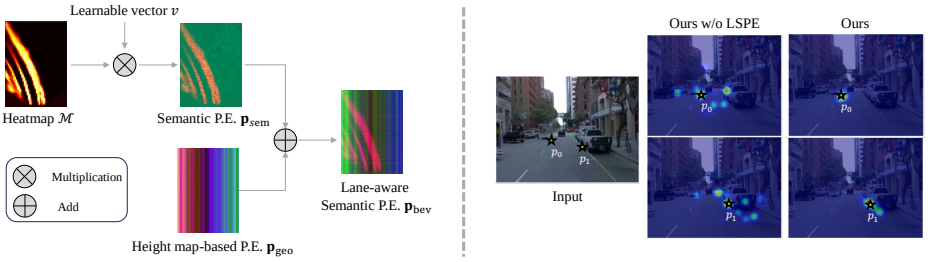

Fig. 4: (Left) Lane-aware Semantic Positional Encoding. A Lane heatmap $\mathcal{M}$ is multiplied by learnable vector $\mathbf{v}$ and builds a semantic embedding $\mathbf{p}_{sem}$ that is fused with height map–based positional encoding $\mathbf{p}_{geo}$ to produce the final Lane-aware Semantic positional embedding $\mathbf{p}_{bev}$. **(Right) Visualization of Effect of LSPE.** We sampled two reference points p_0 and p_1, and visualized weight of attention. With LSPE, our model shows robust sampling ability to align attention with the lane to enhance lane understanding.

volume $\mathcal{V}$ using the identical weights:

$$\tilde{\mathcal{F}}_{hsdf}(x_i, y_j) = \sum_{k=0}^{N-1} w_k(x_i, y_j) \cdot \mathcal{V}(x_i, y_j, z_k). \tag{9}$$

To further improve spatial coherence, the aggregated features $\tilde{\mathcal{F}}_{hsdf}$ are refined using a zero-initialized depthwise-pointwise residual block, producing the final HSDF-based feature map $\mathcal{F}_{hsdf}$. The overview of HSDF is illustrated in Fig. 3.

Furthermore, we introduce a *Height-directional Eikonal Regularization*. A property of an SDF is that the magnitude of its gradient is equal to 1, i.e., $\|\nabla f\| = 1$ [8, 45, 55]. While conventional approaches enforce the constraint on the full 3D gradient, our HSDF is designed specifically for height estimation along the vertical direction. Therefore, the regularization constrains only the partial derivative along the vertical axis, i.e., $\partial s/\partial z = 1$. Because the representation is defined on discrete samples, the derivative is approximated using finite differences:

$$\mathcal{L}_{\text{eik}} = \sum_{(x_i, y_j, z_k)} \left| \frac{\hat{s}(x_i, y_j, z_{k+1}) - \hat{s}(x_i, y_j, z_k)}{z_{k+1} - z_k} - 1 \right|. \tag{10}$$

This regularization encourages the network to learn a physically meaningful distance field rather than arbitrary scalar values, stabilizing zero-crossing localization and improving height rendering quality.

3.4 Lane-aware Semantic Positional Encoding (LSPE)

HSDF-based rendering produces a geometrically aligned feature map $\mathcal{F}_{hsdf}$, which provides a representation for extracting semantic priors. A lane existence heatmap $\mathcal{M} \in \mathbb{R}^{H \times W \times 1}$ is predicted directly from $\mathcal{F}_{hsdf}$ using a lightweight CNN. The auxiliary prediction task captures semantic lane information and encourages the features to concentrate on lane regions, thereby implicitly improving the spatial consistency of $\mathcal{F}_{hsdf}$.

As illustrated in the left panel of Fig. 4, the predicted heatmap $\mathcal{M}$ is converted into a positional encoding to incorporate the semantic prior into the transformer. Instead of adopting a full learnable embedding, which may introduce noise and interfere with the geometric positional embedding $\mathbf{p}_{\text{geo}}$ [32], a single learnable vector $\mathbf{v} \in \mathbb{R}^C$ is employed. The semantic encoding is obtained by modulating the vector $\mathbf{v}$ with the predicted heatmap probability $\mathcal{M}$. The resulting semantic embedding $\mathbf{p}_{\text{sem}}$ is combined with the geometric positional embedding $\mathbf{p}_{\text{geo}}$ derived from the heightmap $\mathcal{H}$ to construct the final Lane-aware Semantic Positional Encoding $\mathbf{p}_{\text{bev}}$:

$$\mathbf{p}_{\text{sem}}(x_i, y_j) = \mathcal{M}(x_i, y_j) \cdot \mathbf{v}, \tag{11}$$

$$\mathbf{p}_{\text{bev}}(x_i, y_j) = \mathbf{p}_{\text{geo}}(x_i, y_j) + \mathbf{p}_{\text{sem}}(x_i, y_j). \tag{12}$$

The composite embedding $\mathbf{p}_{\text{bev}}$ provides the transformer with a semantic prior indicating regions where lane existence is highly probable, while preserving the geometric structure encoded in $\mathbf{p}_{\text{geo}}$.

Finally, following the height-guided spatial transform framework [32], the rendered height map $\mathcal{H}$ and the composite positional embedding $\mathbf{p}_{\text{bev}}$ are used to generate refined BEV features. The refined features are subsequently processed by the lane detection head to estimate 3D lanes.

3.5 Training Loss

The training objective of HSDF-Lane is formulated as a weighted sum of all losses:

$$\begin{aligned}\mathcal{L}_{total} = \lambda_{\text{seg}}\mathcal{L}_{\text{seg}} + \lambda_{\text{off}}\mathcal{L}_{\text{off}} + \lambda_{\text{emb}}\mathcal{L}_{\text{emb}} \\ + \lambda_{\text{2d}}\mathcal{L}_{\text{2d}} + \lambda_{\text{geo}}(\mathcal{L}_{\mathcal{H}} + \lambda_{\text{eik}}\mathcal{L}_{\text{eik}}) + \lambda_{\text{hm}}\mathcal{L}_{\text{hm}}.\end{aligned} \tag{13}$$

Following [32, 46], the segmentation loss $\mathcal{L}_{\text{seg}}$ is implemented as a combination of Binary Cross Entropy (BCE) and IoU losses. The offset loss $\mathcal{L}_{\text{off}}$ is optimized using a BCE loss, while the embedding loss $\mathcal{L}_{\text{emb}}$ employs the discriminative push–pull loss. The auxiliary 2D supervision loss is denoted as $\mathcal{L}_{\text{2d}}$. Geometry-related losses $\mathcal{L}_{\mathcal{H}}$, $\mathcal{L}_{\text{eik}}$, and the lane heatmap loss $\mathcal{L}_{\text{hm}}$ are employed for geometry learning. Detailed formulations of the losses and the weighting coefficients λ are provided in the supplementary material.

4 Experiments

4.1 Experimental Setup

Datasets. To validate our framework, we conducted experiments on the extensive OpenLane benchmark [3], which is derived from the Waymo Open Dataset [41]. The OpenLane benchmark encompasses 1,000 driving scenes, totaling 200,000 frames split into 150,000 for training and 50,000 for validation, alongside 880,000 spatial lane annotations in 3D coordinates. The evaluation set is categorized into six distinct scenarios: Up & Down, Curve, Extreme Weather, Night, Intersection, and Merge & Split. Furthermore, ablation experiments are carried out on OpenLane-300, a representative subset

Table 1: Quantitative comparison on OpenLane Benchmark [3]. The best results are highlighted in **bold**, while the second-best results are underlined. (†): HSDF-Lane equipped with FPN.

Method	Backbone	F1-Score(%)↑	X-error (m) ↓		Z-error (m) ↓	
			near	far	near	far
PersFormer [3]	EfficientNet-B7	50.5	0.485	0.553	0.364	0.431
BEV-LaneDet [46]	ResNet-34	58.4	0.309	0.659	0.244	0.631
LATR [25]	ResNet-50	61.9	0.219	0.259	0.075	0.104
PVALane [60]	Swin-B	63.4	0.226	0.257	0.093	0.119
GroupLane [19]	ConvNext-Base	64.1	0.320	0.441	0.233	0.402
HeightLane [32]	ResNet-50	62.7	0.240	0.266	0.116	0.165
SC-Lane [31]	ResNet-50	64.3	0.227	0.251	0.088	0.128
Rethinking [2]	ResNet-50	64.7	0.205	0.255	0.074	0.105
GLane3D-B [29]	ResNet-50	63.9	0.193	0.234	**0.065**	**0.090**
SparseLaneSTP [36]	ResNet-50	66.1	0.203	0.240	0.066	0.092
HSDF-Lane (Ours)	ResNet-50	66.3	0.201	**0.223**	0.088	0.114
HSDF-Lane† (Ours)	ResNet-50	**66.9**	**0.186**	0.226	0.084	0.114

Table 2: Quantitative results comparison by scenario on the OpenLane [3] validation set using F1-score (%). The best results for each scenario are highlighted in **bold** and second-best results are underlined. (†): HSDF-Lane equipped with FPN.

Method	**All**	**Up & Down**	**Curve**	**Extreme Weather**	**Night**	**Intersection**	**Merge & Split**
BEVLaneDet [46]	58.4	48.7	63.1	53.4	53.4	50.3	53.7
LATR [25]	61.9	55.2	68.2	57.1	55.4	52.3	61.5
HeightLane [32]	62.7	53.6	69.3	55.4	54.6	54.1	61.1
PVALane [60]	62.7	54.1	67.3	**62.0**	57.2	53.4	60.0
SC-Lane [31]	64.3	54.6	70.7	58.1	56.5	56.1	63.0
GLane3D-B [29]	63.9	58.2	71.1	60.1	**60.2**	55.0	64.8
SparseLaneSTP [36]	66.1	57.3	73.0	60.1	58.3	58.2	66.5
HSDF-Lane (Ours)	66.3	**58.9**	73.7	60.7	57.4	58.3	65.7
HSDF-Lane† (Ours)	**66.9**	58.2	**74.2**	60.5	59.6	**59.2**	**66.8**

consisting of 300 scenes. For height map training, we directly utilize the ground-truth dense height maps provided by HeightLane [32], which were generated by accumulating Waymo LiDAR point clouds. These height maps are structured in a 200×48 BEV grid format, covering a physical range of $x \in [3\text{m}, 103\text{m}]$ and $y \in [-12\text{m}, 12\text{m}]$ with a spatial resolution of 0.5m per cell.

Baselines. We compare our method with several strong monocular 3D lane detection baselines, including PersFormer [3], BEV-LaneDet [46], LATR [25], PVALane [60], GroupLane [19], and HeightLane [32], as well as recent state-of-the-art methods such as SC-Lane [31], Rethinking [2], GLane3D-B [29], and SparseLaneSTP [36]. For height estimation, we follow the evaluation protocol of SC-Lane and report comparisons with HeightLane and SC-Lane. For qualitative comparisons, we visualize results from HSDF-Lane, HeightLane, and LATR. Other recent methods, such as GLane3D-B and Sparse-

LaneSTP, are not included in the qualitative comparison due to the lack of publicly available code.

Evaluation Metrics. We utilized the established metrics from Gen-LaneNet [9] to assess 3D lane detection performance, reporting the F1-score alongside positional errors along the X and Z axes. Here, the X- and Z-errors denote lateral and vertical deviations, differing from the vehicle coordinate system in Sec. 3. Specifically, a predicted lane is treated as a true positive only when a minimum of 75% of its constituent points fall within a 1.5-meter distance threshold from the corresponding ground truth. Spatial deviations (X and Z errors) are computed over two distinct longitudinal intervals: the near range (0 to 40 meters) and the far range (40 to 100 meters). For height map evaluation, we follow the metrics proposed in SC-Lane [31]. Specifically, we report Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and threshold-based accuracies at 0.05m, 0.1m, and 0.2m.

4.2 Implementation Details

Network Architecture. We use ResNet-50 [10] as the backbone network to extract image features $\mathcal{F}_{FV}$ with a stride of 16 and a channel size of 1024. The input images are resized to 600×800. We adopt the transformer decoder and 3D lane detection head from HeightLane [32]. To ensure computational efficiency, the transformer decoder is configured with only two layers. We employ four attention heads, with four sampling points per head for the Deformable Cross-Attention (DCA).

We additionally report HSDF-Lane†, a variant that adopts a Feature Pyramid Network (FPN) [20] following LATR [25]. Instead of relying on a single-scale feature, the FPN aggregates multi-scale features from the backbone at strides of 8, 16, and 32, each with a channel size of 256. The resulting multi-scale features are jointly fed into the transformer decoder, which is configured with four layers for this variant. All other settings remain identical to the base model.

Spatial and SDF Settings. The BEV space is quantized into a 200×48 grid, representing a physical range of $x \in [3\text{m}, 103\text{m}]$ in the longitudinal direction and $y \in [-12\text{m}, 12\text{m}]$ in the lateral direction, which yields a grid resolution of 0.5m per cell. For the dense 3D sampling, the predefined maximum slope angle used to define the vertical search range $[z_{\min}, z_{\max}]$ is set to $\theta = 5^{\circ}$. We set the target vertical sampling resolution δ_z to 1.5m. Accordingly, the maximum number of sample points along the z-axis is bounded by $N_{\max} = 13$. The learnable temperature for the soft argmin rendering operation is initialized to $\tau_0 = 0.3$.

4.3 Quantitative Results

Table 1 presents the quantitative comparison with existing monocular 3D lane detection methods on the OpenLane [3] validation set. HSDF-Lane achieves the best F1-score among existing methods, and its FPN-equipped variant HSDF-Lane† further improves it, yielding the top two results and outperforming both anchor-based approaches such as HeightLane [32] and SC-Lane [31] and query-based methods including Sparse-LaneSTP [36]. Both models also achieve competitive geometric accuracy, with HSDF-Lane obtaining the best far-range X-error and HSDF-Lane† the best near-range X-

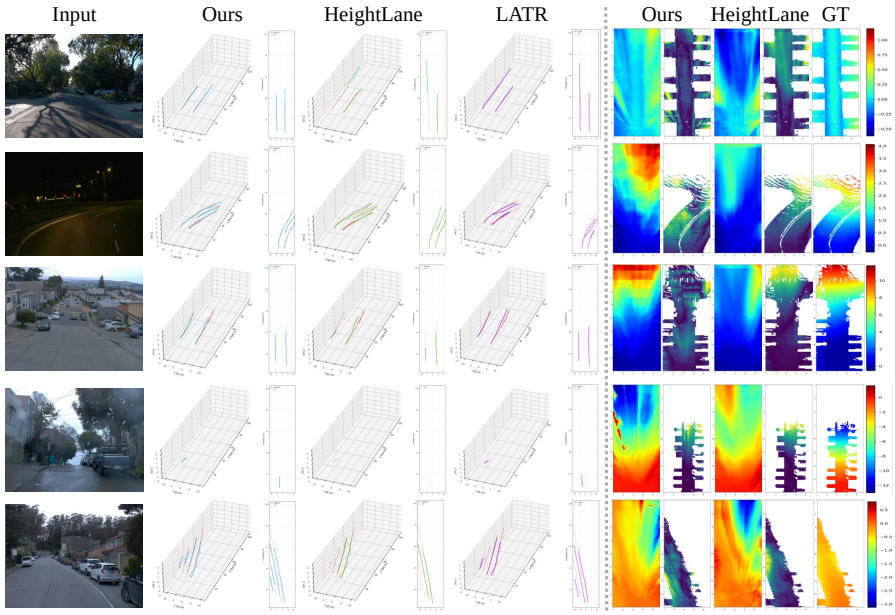

Fig. 5: Qualitative comparison of 3D lane detection and road height estimation. From left to right: input image $\mathcal{I}$, and 3D lane predictions of HSDF-Lane (ours), HeightLane [32], and LATR [25]. The ground truth (GT) is indicated by red dashed lines. The right panel compares the estimated height maps $\mathcal{H}$ and their corresponding error maps (masked by valid GT regions) for our method and HeightLane, alongside the GT height map $\mathcal{H}_{\text{GT}}$. Other recent methods [29, 36] lack public code and are thus excluded. More qualitative results are in the supplementary material.

error. We note that Z-error is influenced by the representation used for lane estimation. Height-map-based methods [31, 32], including our HSDF-Lane, predict and supervise the height over the entire BEV grid. In contrast, other recent methods such as LATR [25], SparseLaneSTP [36], and GLane3D [29] directly regress the height at each lane point and apply the Z-loss only at those points. Because the supervision is concentrated on lane locations, such methods tend to report lower Z-error, so the values are not always directly comparable across methods. Nevertheless, HSDF-Lane remains competitive through the proposed HSDF formulation.

To analyze performance under different environmental conditions, Table 2 reports F1-score across the scenario-based subsets of OpenLane. HSDF-Lane performs best in the *Up & Down* scenario, while HSDF-Lane† attains the best results in *Curve*, *Intersection*, and *Merge & Split*, as well as the best overall F1-score. These results demonstrate that the proposed HSDF-based representation and LSPE effectively improve lane detection under complex road geometries and diverse environments.

Table 3 presents the quantitative comparison of road height estimation. Both variants outperform HeightLane and SC-Lane across most metrics, with HSDF-Lane† obtaining the lowest MAE and the highest threshold-based accuracies. The only exception is RMSE, where HSDF-Lane is slightly higher than SC-Lane. The anchor-based re-

Table 3: Comparison of height estimation performance. The best results are highlighted in **bold** and second-best results are underlined. (†): HSDF-Lane equipped with FPN.

Method	Error Metrics (↓)		Accuracy Metrics (↑)		
	MAE	RMSE	@0.05	@0.1	@0.2
HeightLane [32]	0.235	0.343	0.253	0.442	0.673
SC-Lane [31]	0.176	**0.259**	0.293	0.507	0.756
HSDF-Lane	0.158	0.313	0.294	0.512	0.763
HSDF-Lane†	**0.149**	0.272	**0.307**	**0.527**	**0.776**

gression and temporal consistency loss adopted in SC-Lane suppress outliers and yield lower RMSE, but constrain the representable height range. In contrast, the HSDF-based formulation captures a substantially wider height range, resulting in a marginally higher RMSE. Overall, these results confirm that the proposed HSDF-based representation yields more accurate height estimation than previous anchor-based methods.

4.4 Qualitative Results

Fig. 5 presents qualitative comparisons with representative monocular 3D lane detection methods on the OpenLane dataset [3]. Compared with HeightLane [32] and LATR [25], HSDF-Lane produces more geometrically consistent lane predictions, particularly in challenging scenarios such as curved roads, nighttime conditions, and complex urban environments. In addition to improved lane detection, the proposed method also provides accurate road height estimation. As shown in the right panel of Fig. 5, the predicted height maps of HSDF-Lane exhibit smoother and more consistent surface structures, while the corresponding error maps indicate smaller deviations from the ground truth compared with HeightLane. These results demonstrate that modeling the road surface using the proposed HSDF representation enables more reliable geometric understanding, which further benefits downstream 3D lane detection.

4.5 Ablation Studies

To evaluate the contributions of the proposed components of HSDF-Lane, we conduct ablation studies on the OpenLane-300 subset [3]. The quantitative results are reported in Table 4.

Effect of Explicit Height Map Supervision. Using only the Eikonal regularization (Exp. A) fails to capture accurate road elevations, resulting in poor lane detection performance. Introducing explicit height map supervision $\mathcal{L}_{\mathcal{H}}$ (Exp. B) substantially reduces MAE and RMSE while significantly improving threshold-based accuracies. This improved geometric alignment leads to a higher F1-score, confirming that direct supervision on the rendered height map is essential for learning the correct road topology.

Effect of Height-directional Eikonal Regularization. Comparing Exp. B and Exp. C shows that adding the height-directional Eikonal regularization $\mathcal{L}_{\text{eik}}$ further improves the stability of HSDF-based height estimation. In particular, $\mathcal{L}_{\text{eik}}$ substantially reduces

RMSE and slightly lowers MAE, indicating fewer extreme height deviations and a more physically consistent vertical distance field. While the threshold-based accuracies change only marginally, the improved global consistency leads to a gain in F1-score.

Effect of LSPE. Comparing Exp. C and Exp. D demonstrates the benefit of incorporating LSPE, which introduces the heatmap supervision $\mathcal{L}_{\text{hm}}$. LSPE enhances the F1-score and achieves the best threshold-based accuracies, suggesting that the lane heatmap prior helps the model focus on lane-relevant regions and improve height estimation in a way that is more aligned with lane perception. Although RMSE increases, the overall gains in detection accuracy and threshold-based height metrics indicate that LSPE provides complementary semantic guidance beyond purely geometric supervision. This observation highlights that semantic guidance reinforces geometric supervision and plays a key role in improving 3D lane detection performance.

Table 4: Ablation study on the key components of HSDF-Lane evaluated on the OpenLane-300 subset [3]. **Bold** and underlined values indicate the best and second-best performances.

Exp.	Preserve (✓)			F1-score (↑)	Error (↓)		Accuracy (↑)		
	$\mathcal{L}_{\text{eik}}$	$\mathcal{L}_{\mathcal{H}}$	LSPE		MAE	RMSE	Acc@0.05	Acc@0.1	Acc@0.2
A	✓			71.8	0.512	0.648	0.052	0.105	0.197
B		✓		74.2	0.161	0.333	<u>0.283</u>	<u>0.498</u>	<u>0.755</u>
C	✓	✓		<u>74.7</u>	**0.157**	**0.286**	0.275	0.487	0.750
D	✓	✓	✓	**75.6**	<u>0.158</u>	<u>0.326</u>	**0.288**	**0.503**	**0.767**

5 Conclusion

In this paper, we presented HSDF-Lane, a novel monocular 3D lane detection framework that addresses the limitations of sparse anchor-based height regression and insufficient semantic utilization. To capture complex road geometry, we replaced predefined slope anchors with a dense feature volume modeled by a Height-aligned Signed Distance Field (HSDF). Stabilized by a Height-directional Eikonal regularization, this implicit formulation enables accurate differentiable rendering of a dense height map and surface-aligned features. We further proposed Lane-aware Semantic Positional Encoding (LSPE), which predicts a lane existence heatmap from the surface-aligned features and embeds it into the transformer queries to integrate semantic priors with the rendered geometry. Extensive experiments on the OpenLane benchmark demonstrate that HSDF-Lane achieves state-of-the-art performance in monocular 3D lane detection while improving road height estimation accuracy.

Limitations. Despite these results, several limitations remain. The height map is predicted over a fixed BEV grid without identifying the drivable road surface, training requires dense LiDAR-derived height-map supervision, and experiments are limited to a monocular setting without multi-view, temporal, or multi-modal extensions. Overcoming these limitations remains future work.

Acknowledgements

This research was supported by the Challengeable Future Defense Technology Research and Development Program through the Agency For Defense Development (ADD) funded by the Defense Acquisition Program Administration (DAPA) in 2026 (No.915102201).

References

1. Bai, Y., Chen, Z., Fu, Z., Peng, L., Liang, P., Cheng, E.: Curveformer: 3d lane detection by curve propagation with curve queries and attention. arXiv preprint arXiv:2209.07989 (2022)
2. 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 Computer Vision and Pattern Recognition Conference. pp. 6802–6811 (2025)
3. 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)
4. Chen, Z., Zhang, Y., Qi, X., Mao, Y., Zhou, X., Wang, L., Ge, Y.: Heightformer: A multi-level interaction and image-adaptive classification–regression network for monocular height estimation with aerial images. Remote Sensing **16**(2), 295 (2024)
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. Ghafoorian, M., Nugteren, C., Baka, N., Booij, O., Hofmann, M.: El-gan: Embedding loss driven generative adversarial networks for lane detection. In: proceedings of the european conference on computer vision (ECCV) Workshops. pp. 0–0 (2018)
8. Gropp, A., Yariv, L., Haim, N., Atzmon, M., Lipman, Y.: Implicit geometric regularization for learning shapes. arXiv preprint arXiv:2002.10099 (2020)
9. 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)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. 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)
12. 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)
13. Huval, B., Wang, T., Tandon, S., Kiske, J., Song, W., Pazhayampallil, J., Andriluka, M., Rajpurkar, P., Migimatsu, T., Cheng-Yue, R., et al.: An empirical evaluation of deep learning on highway driving. arXiv preprint arXiv:1504.01716 (2015)
14. 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)

15. Ko, Y., Lee, Y., Azam, S., Munir, F., Jeon, M., Pedrycz, W.: Key points estimation and point instance segmentation approach for lane detection. IEEE Transactions on Intelligent Transportation Systems **23**(7), 8949–8958 (2021)
16. Lee, S., Kim, J., Shin Yoon, J., Shin, S., Bailo, O., Kim, N., Lee, T.H., Seok Hong, H., Han, S.H., So Kweon, I.: Vpgnet: Vanishing point guided network for lane and road marking detection and recognition. In: Proceedings of the IEEE international conference on computer vision. pp. 1947–1955 (2017)
17. Li, C., Shi, J., Wang, Y., Cheng, G.: Reconstruct from top view: A 3d lane detection approach based on geometry structure prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4370–4379 (2022)
18. Li, X., Li, J., Hu, X., Yang, J.: Line-cnn: End-to-end traffic line detection with line proposal unit. IEEE Transactions on Intelligent Transportation Systems **21**(1), 248–258 (2019)
19. Li, Z., Han, C., Ge, Z., Yang, J., Yu, E., Wang, H., Zhang, X., Zhao, H.: Grouplane: End-to-end 3d lane detection with channel-wise grouping. IEEE Robotics and Automation Letters **9**(11), 10487–10494 (2024)
20. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
21. Liu, L., Wang, B., Xie, H., Liu, D., Liu, L., Tian, Z., Yang, K., Wang, B.: Surroundsdf: Implicit 3d scene understanding based on signed distance field. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21614–21623 (2024)
22. Liu, R., Yuan, Z., Liu, T., Xiong, Z.: End-to-end lane shape prediction with transformers. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3694–3702 (2021)
23. Liu, Y., Xu, X., Wang, Z., Yao, Y.: Db3d-l: Depth-aware bev feature transformation for accurate 3d lane detection. arXiv preprint arXiv:2505.13266 (2025)
24. 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)
25. 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)
26. Lyu, D., Huang, H., Tan, C., Li, Z.: Depth3dlane: Monocular 3d lane detection via depth prior distillation. arXiv preprint arXiv:2504.18325 (2025)
27. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
28. Neven, D., De Brabandere, B., Georgoulis, S., Proesmans, M., Van Gool, L.: Towards end-to-end lane detection: an instance segmentation approach. In: 2018 IEEE intelligent vehicles symposium (IV). pp. 286–291. IEEE (2018)
29. Öztürk, H.İ., Kalfaoğlu, M.E., Kilinc, O.: Glane3d: Detecting lanes with graph of 3d keypoints. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 27508–27518 (2025)
30. Pan, X., Shi, J., Luo, P., Wang, X., Tang, X.: Spatial as deep: Spatial cnn for traffic scene understanding. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
31. 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)

32. Park, C., Seo, E., Lim, J.: Heightlane: Bev heightmap guided 3d lane detection. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1692–1701. IEEE (2025)
33. Philion, J., Fidler, S.: Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In: European conference on computer vision. pp. 194–210. Springer (2020)
34. 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)
35. 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)
36. Pittner, M., Janai, J., Faigle, M., Condurache, A.P.: Sparselanestp: Leveraging spatio-temporal priors with sparse transformers for 3d lane detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 29099–29109 (2025)
37. Pizzati, F., Allodi, M., Barrera, A., García, F.: Lane detection and classification using cascaded cnns. In: International conference on computer aided systems theory. pp. 95–103. Springer (2019)
38. Qiu, W., Pang, S., Zhang, H., Fang, J., Xue, J.: Heightmapnet: Explicit height modeling for end-to-end hd map learning. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 6022–6031. IEEE (2025)
39. 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)
40. Su, J., Chen, C., Zhang, K., Luo, J., Wei, X., Wei, X.: Structure guided lane detection. arXiv preprint arXiv:2105.05403 (2021)
41. 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)
42. Tabelini, L., Berriel, R., Paixao, T.M., Badue, C., De Souza, A.F., Oliveira-Santos, T.: Keep your eyes on the lane: Real-time attention-guided lane detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 294–302 (2021)
43. Tabelini, L., Berriel, R., Paixao, T.M., Badue, C., De Souza, A.F., Oliveira-Santos, T.: Polylanenet: Lane estimation via deep polynomial regression. In: 2020 25th international conference on pattern recognition (ICPR). pp. 6150–6156. IEEE (2021)
44. Wang, J., Ma, Y., Huang, S., Hui, T., Wang, F., Qian, C., Zhang, T.: A keypoint-based global association network for lane detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1392–1401 (2022)
45. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. arXiv preprint arXiv:2106.10689 (2021)
46. 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)
47. Wang, W., Lu, Y., Zheng, G., Zhan, S., Ye, X., Tan, Z., Wang, J., Wang, G., Li, X.: Bevspread: Spread voxel pooling for bird's-eye-view representation in vision-based roadside 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14718–14727 (2024)
48. Wang, Z., Ren, W., Qiu, Q.: Lanenet: Real-time lane detection networks for autonomous driving. arXiv preprint arXiv:1807.01726 (2018)

49. Wu, Y., Li, R., Qin, Z., Zhao, X., Li, X.: Heightformer: Explicit height modeling without extra data for camera-only 3d object detection in bird's eye view. IEEE Transactions on Image Processing **34**, 689–700 (2024)
50. Xu, H., Wang, S., Cai, X., Zhang, W., Liang, X., Li, Z.: Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In: European Conference on Computer Vision. pp. 689–704. Springer (2020)
51. Xu, S., Cai, X., Zhao, B., Zhang, L., Xu, H., Fu, Y., Xue, X.: Rclane: Relay chain prediction for lane detection. In: European conference on computer vision. pp. 461–477. Springer (2022)
52. 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)
53. Yang, L., Yu, K., Tang, T., Li, J., Yuan, K., Wang, L., Zhang, X., Chen, P.: Bevheight: A robust framework for vision-based roadside 3d object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21611–21620 (2023)
54. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. Advances in Neural Information Processing Systems **37**, 21875–21911 (2024)
55. Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. Advances in neural information processing systems **34**, 4805–4815 (2021)
56. Yu, H., Lin, H., Wang, J., Li, J., Wang, Y., Zhang, X., Wang, Y., Zhou, X., Hu, R., Peng, S.: Infinidepth: Arbitrary-resolution and fine-grained depth estimation with neural implicit fields. arXiv preprint arXiv:2601.03252 (2026)
57. Zhang, Z., Sun, C., Wang, B., Guo, B., Wen, D., Zhu, T., Ning, Q.: Height3d: A roadside visual framework based on height prediction in real 3-d space. IEEE Transactions on Intelligent Transportation Systems (2025)
58. Zheng, T., Fang, H., Zhang, Y., Tang, W., Yang, Z., Liu, H., Cai, D.: Resa: Recurrent feature-shift aggregator for lane detection. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3547–3554 (2021)
59. 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)
60. Zheng, Z., Zhang, X., Mou, Y., Gao, X., Li, C., Huang, G., Pun, C.M., Yuan, X.: Pvalane: Prior-guided 3d lane detection with view-agnostic feature alignment. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 7597–7604 (2024)
61. Zhou, R., Li, J., Su, Z., Lu, C., Wang, Z.: Heightaware-bev: Height-aware feature mapping for efficient bird's-eye-view perception. In: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 6172–6178. IEEE (2025)
62. Zou, Q., Jiang, H., Dai, Q., Yue, Y., Chen, L., Wang, Q.: Robust lane detection from continuous driving scenes using deep neural networks. IEEE transactions on vehicular technology **69**(1), 41–54 (2019)