Journal of Industrial Intelligence
https://www.acadlore.com/journals/JII

AG-CLRNet: A Real-Time Industrial Lane Perception Framework for Intelligent Driving Systems

Weiguo Ding[1], Jialin Ma[1*], Ashim Khadka[2]

[1] Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, 223003 Huaian, China
[2] Nepal College of Information Technology, Pokhara University, 44700 Lalitpur, Nepal

[*] Correspondence: Jialin Ma (majl@hyit.edu.cn)

Received: 04-30-2025 **Revised:** 06-17-2025 **Accepted:** 06-22-2025

Citation: W. G. Ding, J. L. Ma, and A. Khadka, "AG-CLRNet: A real-time industrial lane perception framework for intelligent driving systems," *J. Ind Intell.*, vol. 3, no. 2, pp. 125–136, 2025. https://doi.org/10.56578/jii030205.

Abstract: Reliable lane perception is a core enabling function in industrial intelligent driving systems, providing essential structural constraints for downstream tasks such as lane keeping assistance, trajectory planning, and vehicle control. In real-world deployments, lane detection remains challenging due to complex road geometries, illumination variations, occlusions, and the limited computational resources of on-board platforms. This study presents Attention-Guided Cross-Layer Refinement Network (AG-CLRNet), a real-time lane perception framework designed for industrial intelligent driving applications. Built upon an anchor-based detection paradigm, the framework integrates adaptive multi-scale contextual fusion, channel–spatial attention refinement, and long-range dependency modeling to improve feature discrimination and structural continuity while maintaining computational efficiency. The proposed design strengthens the representation of distant and slender lane markings, suppresses background interference caused by shadows and pavement textures, and enhances global geometric consistency in curved and fragmented scenarios. Extensive experiments conducted on the CULane benchmark demonstrate that AG-CLRNet achieves consistent improvements in precision, recall, and F1 score over representative state-of-the-art methods, while sustaining real-time inference performance suitable for practical deployment. Ablation studies further confirm the complementary contributions of the proposed modules to robustness and structural stability under challenging conditions. Overall, AG-CLRNet provides a practical and deployable lane perception solution for industrial intelligent driving systems, offering a balanced trade-off between accuracy, robustness, and real-time performance in complex road environments.

Keywords: Industrial intelligent driving; Lane perception; Real-time vision systems; Attention-based feature modeling; Edge deployment

1 Introduction

Lane detection is a fundamental component of environmental perception in intelligent transportation and autonomous driving systems, serving as a prerequisite for advanced driver-assistance functions such as lane keeping, adaptive cruise control, and forward collision warning, and thus playing a critical role in driving safety [1]. As vehicle automation continues to advance, lane detection systems are expected to achieve higher accuracy, robustness, and real-time performance. However, real-world road environments remain highly challenging. Illumination variations between day and night, adverse weather conditions (e.g., rain and snow), lane degradation, and frequent occlusions can severely deteriorate image features and reduce lane saliency. Moreover, complex urban road topologies, including sharp curves, ramps, and interchanges, impose stringent requirements on long-range contextual reasoning and spatial structure modeling [2].

Early lane detection approaches predominantly followed a semantic segmentation paradigm, extracting lane regions through pixel-wise classification. Representative methods include Spatial CNN (SCNN), EL-GAN, SAD, CurveLane-NAS, RESA, and LaneAF. Specifically, SCNN propagates spatial information along rows and columns to enhance lane continuity and alleviate fragmented segmentation outputs [3–6]. EL-GAN introduces adversarial learning to impose structural constraints on predictions, encouraging geometric consistency [4]. Structure-Aware Distillation (SAD) transfers structural knowledge from a teacher model to improve the performance of lightweight networks [5], while CurveLane-NAS leverages neural architecture search to automatically derive lane-sensitive

architectures with a favorable balance between accuracy and efficiency [6]. Recurrent Feature-Shift Aggregator (RESA) enhances global perception via recurrent feature shifting and aggregation, improving robustness to curved and fragmented lanes [7]. LaneAF formulates lane detection as an instance segmentation problem using vector-field regression to mitigate interference among overlapping lanes. Although segmentation-based methods provide fine-grained pixel-level representations, they are sensitive to occlusion and illumination variations and often rely on complex post-processing pipelines, which limits their robustness and real-time applicability.

To further improve inference efficiency, row-wise classification-based methods have been explored. Ultra Fast Lane Detection (UFLD) reformulates lane detection as a one-dimensional row-level classification problem, significantly reducing computational complexity and achieving inference speeds exceeding 300 FPS [8]. However, its strong reliance on regular lane patterns restricts generalization in scenarios involving sharp curves or complex road topology. CondLaneNet introduces conditional convolutions by dynamically generating instance-specific kernels for different lanes, enabling more adaptive feature extraction [9]. Building upon this framework, UFLDv2 incorporates adaptive feature fusion and multi-scale attention mechanisms to further improve robustness under curved and occluded conditions [10]. Despite their high efficiency, row-wise methods may still suffer from discontinuous predictions or local breaks when handling long-range and slender lane markings in complex scenes.

More recently, keypoint-based lane detection methods have attracted increasing attention. These approaches represent each lane as a set of discrete keypoints and perform end-to-end prediction via keypoint detection followed by association. PINet detects and clusters lane points to generate final lane instances [11], while FOLOLane leverages multi-level feature aggregation and keypoint-guided strategies to recover fine-grained lane structures, demonstrating strong robustness under occlusion and blur [12]. GANet further models lanes as graph structures and employs graph attention mechanisms to capture global dependencies among keypoints, thereby improving topological consistency [13]. Although keypoint-based methods offer improved geometric interpretability, the keypoint association and structure reconstruction processes introduce additional complexity, which often limits inference efficiency.

To better exploit lane geometry, parameterized curve-based methods have been proposed. PolyLaneNet represents lanes as polynomial curves by regressing curve coefficients, reducing the reliance on complex post-processing [14]. Lane Shape Transformer (LSTR) introduces Transformer architectures to model long-range dependencies among lane points, improving prediction continuity and geometric consistency [15]. BezierLaneNet further adopts Bézier curve parameterization, fitting lanes using a small number of control points to enhance stability and interpretability. Despite their compact representations, parameterized methods may still face limitations under complex topologies such as lane splits and merges due to the fixed functional form of curve models.

In recent years, anchor-based (top-down) approaches have emerged as an effective direction for lane detection [16–19]. LineCNN introduces anchor lines to regress lane offsets in an end-to-end manner, though its performance can be sensitive to anchor design and hyperparameter settings [16]. Building upon this paradigm, LaneATT aggregates contextual features around anchors using attention mechanisms, improving modeling capability for curved lanes. SGNet dynamically adjusts prediction regions through spatial guidance to enhance geometric consistency [17], while SIIC-Net strengthens both global and local perception via cross-layer feature interaction and spatial correction modules [18]. Among these methods, CLRNet (Cross-Layer Refinement Network) achieves a favorable trade-off between detection accuracy and real-time performance by unifying global semantics and fine-grained spatial details through multi-level feature fusion [19]. As a result, anchor-based frameworks have become one of the mainstream solutions for lane detection.

Although CLRNet effectively alleviates the conflict between high-resolution details and high-level semantics through cross-layer refinement, it still encounters limitations in extreme scenarios. First, its feature fusion strategy primarily relies on direct aggregation and lacks adaptive multi-scale perception tailored to distant and slender lane markings, leading to weakened representations in low-texture regions. Second, conventional convolution operations exhibit limited capability in distinguishing subtle lane textures from visually similar patterns such as road cracks and shadows, which may result in false positives and boundary drift. Third, anchor-based regression focuses mainly on local features; consequently, the absence of explicit global geometric constraints can cause fragmented predictions under long-range curved or high-curvature scenarios.

To address these limitations, we propose AG-CLRNet, an enhanced lane detection framework built upon CLRNet. Specifically, an adaptive multi-scale contextual fusion module is introduced to capture distant and thin lane markings through scale-aware context modeling [20]. A dual channel–spatial attention module is incorporated to strengthen discriminative lane features while suppressing background interference caused by cracks and shadows [21]. In addition, a long-range dependency enhanced decoding head is designed to introduce stronger global geometric cues, thereby improving prediction continuity and reducing fragmentation in curved lane scenarios [22]. With these components, AG-CLRNet achieves improved global perception and structural consistency in challenging road environments, providing a more accurate and robust solution for lane detection.

From a system perspective, AG-CLRNet can be deployed as a front-end lane perception module in autonomous

driving or advanced driver-assistance systems (ADAS). The detected lane geometry and confidence information provide reliable structural constraints for downstream tasks such as lane keeping assistance, trajectory planning, and vehicle control. By maintaining a favorable balance between detection accuracy, robustness under complex conditions, and real-time efficiency, AG-CLRNet is well suited for practical integration into on-board perception pipelines and real-world intelligent transportation systems.

2 Proposed Method

Challenges in Lane Detection

The design of AG-CLRNet targets two fundamental challenges in lane detection. First, in complex road environments, lane markings are typically slender, continuous, and morphologically diverse. Under adverse conditions such as illumination variations, occlusions, lane wear, shadows, and severe weather (e.g., rain or snow), conventional convolutional neural networks (CNNs) struggle to capture long-range dependencies and global topological structures. As a result, lane predictions often suffer from fragmentation, drift, and false detections, leading to degraded geometric consistency. Second, many existing deep learning–based lane detection models pursue high accuracy at the cost of increased computational complexity and memory consumption, which hinders efficient deployment on vehicle-mounted embedded platforms and real-time autonomous driving systems.

Design Strategy of AG-CLRNet

To address these challenges, we propose AG-CLRNet, a lane detection framework derived from structural optimization and modular enhancement of the CLRNet architecture. While retaining CLRNet's efficient feature extraction and anchor-based detection paradigm, AG-CLRNet introduces three key improvements:

To explicitly address the key challenges in lane detection under complex driving scenarios, the proposed AG-CLRNet is designed following a problem-driven modular strategy. Specifically, three representative difficulties are considered: (1) distant and slender lane markings are easily weakened or lost during multi-scale feature fusion; (2) background clutter, shadows, and pavement textures introduce severe feature redundancy and false responses; and (3) local convolution-based decoding lacks sufficient global geometric awareness, leading to fragmented or discontinuous lane predictions. Correspondingly, three dedicated modules are introduced, namely AMCFM for enhancing multi-scale contextual perception of small and distant lanes, CSDA for suppressing background interference and emphasizing discriminative lane features, and LDED for strengthening long-range dependency modeling and global structural consistency.

(1) Adaptive Multi-scale Context Fusion Module (AMCFM): integrated into the feature extraction stage to aggregate multi-scale contextual information, thereby enhancing the perception of distant and slender lane markings as well as complex topological structures.

(2) Channel–Spatial Dual Attention (CSDA) Module: embedded in the intermediate feature enhancement stage to adaptively emphasize informative feature channels and spatial regions, suppressing background interference and reinforcing lane-related features.

(3) Long-range Dependency Enhanced Decoder (LDED): incorporated into the decoding stage to model global contextual dependencies, improving lane continuity and boundary smoothness while reducing fragmentation and misdetections.

The overall architecture of AG-CLRNet is illustrated in Figure 1.

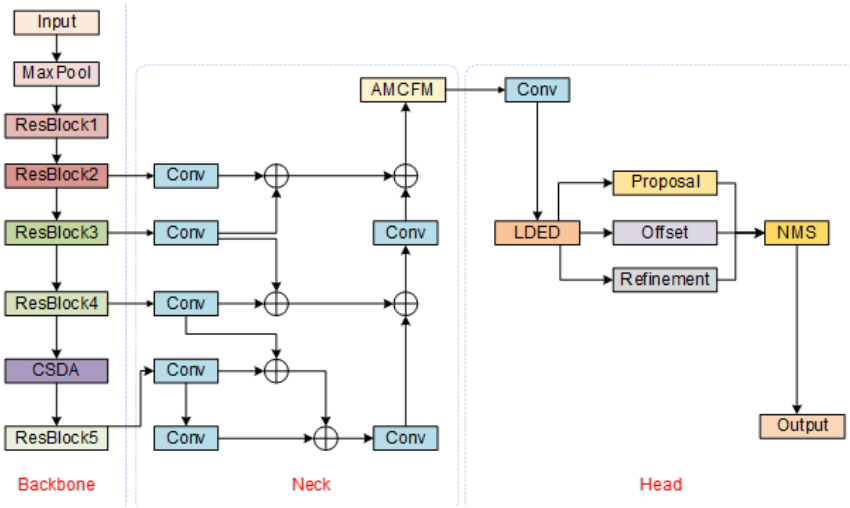

Figure 1. Network architecture of AG-CLRNet

2.1 The CLRNet Lane Detection Network

CLRNet is an end-to-end, anchor-based lane detection framework proposed by Tsinghua University in 2022, which achieves a favorable balance between detection accuracy and real-time performance, particularly in complex scenarios involving curved lanes, forks, occlusions, and illumination variations. Its architecture consists of three main components: a Backbone, a cross-layer feature fusion Neck, and a multi-level detection Head.

The Backbone, typically built upon the ResNet series, extracts multi-scale feature representations ranging from low-level textures and edges to high-level semantic information, providing a rich contextual foundation for lane detection. Unlike conventional backbones, CLRNet introduces cross-layer interactions among multi-level feature maps, enabling efficient fusion of shallow spatial details and deep semantic cues, which improves the perception of distant and faint lane markings.

The Neck adopts a Feature Pyramid Network (FPN) structure to fuse feature maps at different resolutions, thereby preserving both global structural information and local geometric characteristics. During this stage, CLRNet incorporates the Cross Layer Refinement Module (CLRM) to further enhance feature representations through context propagation and refinement, improving the continuity and robustness of lane predictions.

The Head employs an anchor-based detection strategy with multi-level decoder heads to predict lane parameters and corresponding confidence scores across different feature scales. Compared with anchor-free approaches, this strategy provides more stable localization and regression of complex lane topologies, such as splits, merges, and occlusions, resulting in improved detection precision and geometric consistency. Moreover, CLRNet jointly optimizes classification and localization in an end-to-end manner, maintaining high accuracy without sacrificing inference efficiency.

Overall, through cross-layer refinement and multi-scale feature fusion, CLRNet effectively enhances the spatial continuity and shape integrity of detected lanes. Its strong performance on benchmark datasets such as CULane and TuSimple establishes a solid baseline and provides a reliable foundation for further architectural improvements in lane detection.

2.2 Adaptive Multi-Scale Context Fusion Module (AMCFM)

Problem Statement and Motivation

Distant and slender lane markings usually occupy very limited pixel regions and are easily weakened or even lost during conventional multi-scale feature fusion, especially under complex illumination variations and occlusion conditions. Moreover, insufficient interaction between high-level semantic context and low-level fine-grained spatial details further degrades the representation of thin, curved, and low-contrast lane structures, leading to feature attenuation and prediction discontinuities in challenging scenarios.

Motivated by these limitations, we introduce an AMCFM into the Neck of the network to explicitly enhance scale-aware contextual perception and dynamic feature integration. By jointly leveraging multi-scale feature extraction, bidirectional feature propagation, dilated convolution-based receptive field expansion, and adaptive feature weighting, AMCFM enables effective fusion of local structural details and global contextual information, thereby strengthening the network's capability to accurately model distant, slender, and complex lane structures.

Technical Details

The overall architecture of the proposed AMCFM is illustrated in Figure 2.

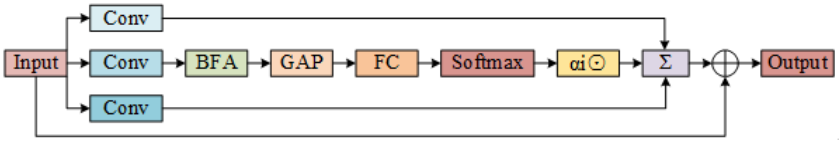

Figure 2. Structure of the AMCFM module

As illustrated in Figure 2, AMCFM operates through the following components:

Multi-scale Feature Capture

Feature maps from different backbone levels (ResNet34) are first processed by parallel convolutional branches with kernel sizes of 3×3, 5×5, and 7×7, allowing the network to capture contextual information under diverse receptive fields and improving sensitivity to both near and distant lane markings as well as curved boundaries.

Bidirectional Feature Aggregation

A bidirectional feature aggregation strategy is then employed, establishing both top-down and bottom-up information flows. This design allows high-level semantic cues to guide low-level feature refinement, while high-resolution spatial details from shallow layers enrich deeper semantic representations, facilitating effective interaction between spatial and semantic information.

Long-range Context Modeling

To further enhance contextual perception, dilated convolutions are applied after multi-scale fusion to expand the receptive field and capture long-range dependencies, alleviating the limited response of standard convolutions to distant lane structures [20].

Adaptive Weight Fusion

A lightweight attention-based sub-network is introduced to dynamically learn fusion weights across different scales. This adaptive weighting mechanism emphasizes informative features while suppressing irrelevant responses, improving robustness under challenging conditions such as shadows and sharp curves.

Residual Learning

Finally, a residual connection is incorporated to stabilize feature distributions and mitigate gradient attenuation caused by deep feature stacking, thereby accelerating convergence and improving training stability [21].

Through this design, AMCFM significantly enhances contextual modeling and edge sensitivity, improving the robustness of CLRNet in detecting distant, intersecting, and blurred lane markings. The feature fusion process can be formulated as:

$$F_{out} = \sum_{i=1}^{n} \alpha_i \bullet DilatedConv(F_i) + F_{res} \tag{1}$$

where, F_i denotes the feature maps generated by convolutions at different scales, α_i represents the corresponding adaptive fusion weights, and F_{res} is the residual input feature.

Overall, AMCFM explicitly strengthens scale-aware contextual perception and dynamic feature integration, enabling more robust representation of distant, slender, and structurally complex lane markings.

2.3 Channel–Spatial Dual Attention Module (CSDA)

Problem Statement

In complex road environments, lane detection is frequently disturbed by background clutter, shadows, pavement wear, and strong illumination variations, which introduce substantial redundant responses in feature representations. Conventional convolutional operations treat all spatial locations and channels equally, lacking an explicit mechanism to selectively emphasize informative lane features while suppressing irrelevant background noise. To address this limitation, the CSDA module is introduced to adaptively reweight feature responses along both channel and spatial dimensions, thereby enhancing discriminative lane cues and improving robustness against background interference.

Dual Attention Mechanism

The overall structure of the proposed CSDA module is illustrated in Figure 3.

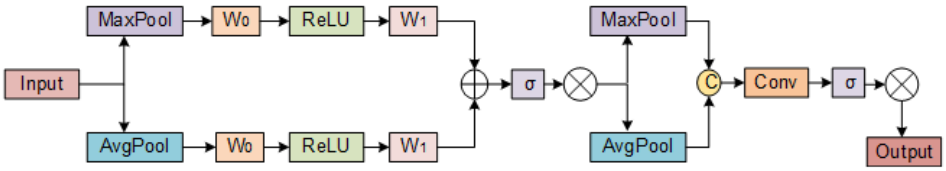

Figure 3. Structure of the CSDA module

To alleviate this issue, we integrate a CSDA module into the high-level feature extraction stage of the CLRNet backbone. CSDA sequentially models channel-wise and spatial-wise importance to enhance discriminative lane features while suppressing irrelevant background responses.

Given an input feature map $F \in R^{C\times H\times W}$, the channel attention branch first captures inter-channel dependencies. Specifically, global average pooling (GAP) and global max pooling (GMP) are applied to F to generate two channel descriptors:

$$F^c_{avg} = AvgPool(F) \tag{2}$$

$$F^c_{max} = MaxPool(F) \tag{3}$$

where, $F_{avg}, F_{max} \in R^{C\times 1\times 1}$.

These descriptors are fed into a shared multi-layer perceptron (MLP) consisting of two fully connected layers with a ReLU activation in between, following a standard attention modeling paradigm [22]. The outputs are summed and activated by a Sigmoid function to produce the channel attention map:

$$M_c = \sigma(W_1(\delta(W_0(F^c_{avg}))) + W_1(\delta(W_0(F^c_{max})))) \tag{4}$$

where, $\sigma(\cdot)$ and $\delta(\cdot)$ denote the Sigmoid and ReLU functions, respectively; $W_0 \in R^{\frac{C}{r}\times C}$ and $W_1 \in R^{C\times\frac{C}{r}}$ are learnable weight matrices, and r denotes the reduction ratio.

The channel-refined feature map is obtained via element-wise multiplication:

$$F' = M_c(F) \otimes F \tag{5}$$

where, $\otimes$ denotes element-wise multiplication.

Subsequently, the spatial attention branch focuses on informative spatial regions based on F'. Average pooling and max pooling are applied along the channel dimension to generate two 2D spatial maps, which are concatenated and processed by a convolution with a kernel size of 7×7. The spatial attention map is computed as:

$$M_s(F') = \sigma(f^{7\times7}([AvgPool_c(F'); MaxPool_c(F')])) \tag{6}$$

where, $[\cdot;\cdot]$ denotes channel-wise concatenation. Finally, the enhanced feature representation is obtained as:

$$F_{out} = M_s(F') \otimes F' \tag{7}$$

Through this sequential dual-attention mechanism, CSDA selectively strengthens salient channel-wise semantics and spatially informative regions while maintaining low computational overhead. This enables the network to more effectively extract lane edge and texture features under challenging conditions such as complex illumination, noisy backgrounds, and partial occlusions, providing more discriminative representations for subsequent multi-scale fusion and decoding stages. This dual-attention mechanism improves feature discriminability while maintaining low computational overhead, providing more reliable representations for subsequent multi-scale fusion and decoding stages.

2.4 Long-distance Dependency Enhanced Decoding Module (LDED)

Lane lines typically exhibit strong geometric continuity and long spatial extents, which require effective modeling of global structural relationships during decoding. However, conventional decoding architectures mainly rely on local convolutional operations with limited receptive fields, making it difficult to capture long-range dependencies and global lane geometry, especially under high-curvature or long-distance scenarios. As a result, predictions may suffer from fragmentation, spatial offsets, or discontinuities. To overcome these limitations, the LDED module is introduced to jointly model global contextual relationships while preserving local geometric details, thereby improving structural consistency and prediction continuity.

As illustrated in Figure 4, the LDED module consists of three components: a Global Context Modeling Layer, a Relation-Enhanced Attention Layer, and a Multi-scale Fusion Prediction Layer. These components operate hierarchically to reinforce structural continuity and geometric consistency of lane predictions while progressively restoring spatial resolution during decoding.

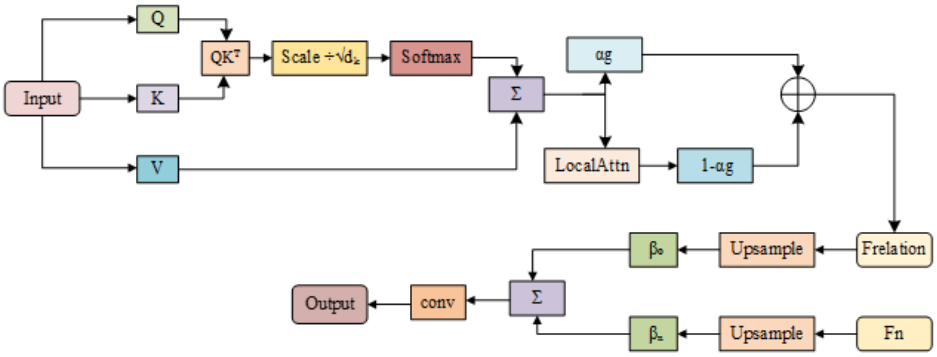

Figure 4. Structure of the LDED module

Global Context Modeling Layer

This layer is designed to alleviate prediction fragmentation by capturing global contextual relationships. Given an input feature map $F \in R^{C\times H\times W}$, linear projections are applied to obtain the **Query (Q)**, **Key (K)**, and **Value (V)** representations. Global context aggregation is then performed using a scaled dot-product attention mechanism:

$$Attention(Q, K, V) = Softmax(QK^T/\sqrt{d_k})V \tag{8}$$

where,d_k denotes the dimensionality of the key vectors and serves as a scaling factor to stabilize gradient propagation during training. This operation enables the decoder to incorporate global geometric cues into lane prediction.

Relation-Enhanced Attention Layer

To further improve localization accuracy while maintaining computational efficiency, a window-based attention mechanism is employed. Attention is computed within local feature windows to model fine-grained spatial relationships, thereby substantially reducing computational overhead. Meanwhile, a cross-window global guidance weight α_g is introduced to compensate for the limited receptive field of local attention:

$$F_r = \alpha_g F_g + (1 - \alpha_g) \cdot LocalAttn(F_g) \tag{9}$$

where, F_r denotes the refined output feature map, F_g represents the feature map from the preceding layer, $LocalAttn(\cdot)$ denotes the local window-based attention operation, and α_g is a learnable balancing coefficient. This formulation enables an effective trade-off between global structural awareness and local detail refinement.

Multi-Scale Fusion Prediction Layer

In the final stage, LDED integrates deep semantic consistency and shallow geometric details through cascaded upsampling and cross-layer feature fusion. Learnable weights β_i are used to adaptively regulate the contributions of multi-scale features:

$$F_{final} = \sum_{i=1}^{N} \beta i \cdot Upsample(F_i) \tag{10}$$

where, N denotes the number of decoding feature levels, F_i denotes feature maps from different decoding levels, and $Upsample(\cdot)$ represents a resolution alignment operation.

The fused representation F_{final} is subsequently processed by a 1×1 convolution to predict lane geometric parameters and confidence scores. Experimental results demonstrate that, by enforcing hierarchical constraints from global context to local details, the LDED module significantly improves prediction smoothness and continuity in scenarios involving high-curvature bends and fragmented lane markings. By jointly modeling global contextual dependencies and localized geometric refinement, LDED effectively mitigates fragmentation and enhances the structural continuity of lane predictions.

3 Experimental Results and Analysis

3.1 Dataset

Experiments are conducted on the publicly available CULane dataset [3], released by the Multimedia Laboratory of the Chinese University of Hong Kong and widely used as a benchmark for lane detection. The dataset covers diverse driving scenarios, including urban roads, highways, tunnels, curves, shadows, and traffic congestion.

To balance computational efficiency and training effectiveness, we randomly sampled approximately 13,000 representative images from the original CULane training set while preserving the overall data distribution. The test set strictly follows the official partitioning protocol. Specifically, 8,884 images are used for training, 968 for validation, and 3,468 for testing.

To improve generalization under complex conditions, the data were carefully screened and preprocessed. Abnormal frames were removed, and several data augmentation techniques—such as resizing, random cropping, brightness and contrast adjustment, and color jittering—were applied to simulate diverse real-world visual variations. Each image is annotated with a corresponding text file, where each line represents a lane instance as a sequence of (x, y) coordinates, effectively capturing lane geometry and spatial continuity.

3.2 Implementation Details and Hyperparameter Settings

The hardware and software configurations used for the experiments are summarized in Table 1

Table 1. Experimental environment configuration

Component	Specification
Operating system	Windows 11
Programming language	Python 3.9
Deep learning framework	Pytorch 2.2.2
CUDA	12.1
CPU	Intel i7-14700KF
GPU	RTX 3090 (24GB)
RAM	45GB

To ensure the fairness of the experimental comparisons, a consistent set of hyperparameters was applied across all training sessions. We employed the Stochastic Gradient Descent (SGD) optimizer coupled with a cosine annealing learning rate scheduler. The specific hyperparameters used during the training phase are detailed in Table 2.

Table 2. Training hyperparameters

Parameter	Value
Learning rate	0.01
Image size	640×640
Optimizer	SGD
Batch size	32
Epochs	300
Weight decay	0.0005

3.3 Evaluation Metrics

In this study, the following evaluation metrics are employed:

Precision (P): Represents the proportion of true positive samples among those predicted as positive by the model. The formula is shown in Eq. (11):

$$P = \frac{TP}{TP + FP} \times 100\% \tag{11}$$

Recall (R): Represents the proportion of ground-truth positive samples that are successfully detected. The formula is shown in Eq. (12):

$$R = \frac{TP}{TP + FN} \times 100\% \tag{12}$$

Mean Average Precision (mAP): The Average Precision (AP) for each category is calculated as in Eq. (13), and the mAP is obtained by averaging these AP values, as shown in Eq. (14):

$$AP = \int_0^1 P(R)dR \tag{13}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{14}$$

where, TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively; n is the number of target categories, and AP_i is the AP of the i-th category.

Frames Per Second (FPS): A key metric to measure the real-time processing capability of the model.

Parameters (Param): Used to measure the size and complexity of the model.

Table 3. Ablation study results on the CULane dataset.

Model	AMCFM	CSDA	LDED	mAP50 (%)	P (%)	R (%)	F_1 (%)	Param (M)
CLRNet				77.4	79.2	75.1	77.1	25.6
Model A	√			78.2	79.8	76.0	77.9	26.1
Model B		√		78.5	80.9	75.5	78.1	26.5
Model C			√	78.3	80.1	76.2	78.1	26.3
Model D	√	√		79.3	81.3	76.8	79.0	27.0
Model E	√		√	79.0	80.6	77.1	78.8	27.2
Model F		√	√	79.1	81.1	76.9	79.0	27.4
Ours	√	√	√	79.7	81.7	77.4	79.5	27.7

As shown in Table 3, introducing AMCFM alone improves Recall by 0.9%, indicating enhanced capability in detecting distant and slender lane markings through multi-scale contextual aggregation.When CSDA is applied independently, Precision increases by 1.7%, demonstrating the effectiveness of the attention mechanism in suppressing background noise and reducing false positives. Incorporating LDED leads to an F_1 score of 78.1%, validating the role of long-range dependency modeling in improving lane continuity. When all three modules are integrated (Ours), the proposed model achieves an F_1 score of 79.5%, representing a 2.4% improvement over the baseline CLRNet. Although the inference speed decreases from 124 FPS to 112 FPS, the model still satisfies real-time requirements ($> 30FPS$). These results demonstrate that AG-CLRNet effectively enhances detection performance through coordinated feature reinforcement and contextual modeling while maintaining a manageable computational cost.

Table 4 compares the detection performance of AG-CLRNet with several state-of-the-art (SOTA) lane detection methods on the CULane dataset.Compared with UFLD, which emphasizes extreme inference speed, AG-CLRNet

achieves a substantial 7.2% improvement in F_1 score, while maintaining real-time performance with only a moderate reduction in FPS. When compared with strong baseline methods using the same ResNet34 backbone, including LaneATT and CondLaneNet, AG-CLRNet outperforms them by 3.9% and 1.7% in F_1 score, respectively.Notably, relative to the original CLRNet, AG-CLRNet achieves balanced improvements in both Precision and Recall, with gains of 2.5% and 2.3%, respectively. These results demonstrate the effectiveness of the proposed modules in handling complex driving scenarios and confirm that AG-CLRNet achieves a favorable trade-off between detection accuracy and inference efficiency.

Table 4. Comparison of detection performance among different models.

Model	Backbone	Precision (%)	Recall (%)	F_1 (%)	FPS
SCNN	VGG16	64.2	81.4	71.8	7.5
RESA	ResNet34	68.1	76.8	72.2	23
UFLD	ResNet34	76.0	68.4	72.3	175
LaneATT	ResNet34	75.1	76.1	75.6	140
CondLaneNet	ResNet34	80.1	75.6	77.8	102
CLRNet	ResNet34	79.2	75.1	77.1	124
AG-CLRNet	ResNet34	81.7	77.4	79.5	112

In addition to accuracy improvement, practical deployment requires a careful balance between inference speed, model complexity, and hardware constraints. As reported in Table 4, AG-CLRNet achieves 112 FPS on an RTX 3090 GPU with a parameter size of approximately 27.7M, indicating a moderate computational footprint. Although embedded automotive platforms usually provide lower peak computing capability than desktop GPUs, this level of efficiency still provides sufficient margin for real-time perception when deployed on modern edge accelerators through optimized inference engines and hardware acceleration.

Compared with extremely lightweight methods such as UFLD, which prioritize inference speed at the expense of detection accuracy, AG-CLRNet maintains significantly higher precision and robustness while preserving real-time performance. Compared with heavier anchor-based methods, the proposed model avoids excessive parameter growth and achieves a more favorable balance between accuracy, robustness, and computational efficiency, making it more suitable for practical on-board deployment scenarios.

Although the experimental evaluation in this work is mainly conducted on the CULane dataset, this benchmark covers a wide range of real-world driving scenarios, including urban roads, highways, tunnels, curves, shadows, and traffic congestion, which provides a reasonable approximation of diverse road conditions. Moreover, the proposed AG-CLRNet emphasizes multi-scale contextual modeling, attention-guided feature refinement, and long-range dependency reasoning, which are generally beneficial for improving robustness under variations in illumination, weather conditions, road topology, and partial occlusions.

Nevertheless, it should be acknowledged that domain gaps may still exist when deploying the model in different countries, camera configurations, or unseen environmental conditions. Differences in lane marking standards, camera mounting height, field of view, and sensor calibration may affect the data distribution and consequently influence detection performance. Future work will focus on cross-dataset evaluation and domain adaptation to further validate and enhance the generalization capability of the proposed method in broader real-world deployment scenarios.

3.4 Visualization Analysis

To further evaluate the detection performance and feature representation capabilities of the proposed AG-CLRNet, we select typical road scenarios from the CULane dataset for a qualitative analysis of detection results and intermediate feature responses. As illustrated in Figure 5, comparing AG-CLRNet with the baseline CLRNet, it is evident that our improved model exhibits superior lane localization precision and spatial continuity.

From the qualitative results, CLRNet still suffers from certain missed and false detections in complex road environments. On one hand, in segments featuring distant slender lanes, low-light conditions, or severe shadow occlusions, CLRNet is prone to lane fragmentation, discontinuous responses, or even complete missed detections. On the other hand, in cluttered regions such as curbs, road cracks, and surface markings (e.g., crosswalks and arrow signs), the baseline model occasionally generates false lane detections. Overall, the visualization results of AG-CLRNet in complex road environments are markedly superior to those of CLRNet, demonstrating higher reliability, particularly in detecting distant slender lanes, handling heavy occlusions, and restoring lane continuity.

As illustrated in Figure 6, we compare the mAP50 and Loss curves of CondLaneNet, LaneATT, RESA, UFLD, CLRNet, and the proposed AG-CLRNet during the training process. It can be observed that all models gradually stabilize after approximately 50 epochs, indicating a well-converged training process without significant gradient explosion or severe oscillations.

(a)

(b)

Figure 5. Visualization of detection results: (a) CLRNet; (b) AG-CLRNet

Regarding the mAP50 curves, AG-CLRNet exhibits a steeper upward trend during the early stages of training compared to the baseline CLRNet and other methods. This suggests that the proposed multi-scale feature fusion mechanism can capture key lane features more rapidly, thereby accelerating the initial learning process. As the number of iterations increases, the mAP50 curve of AG-CLRNet consistently remains above those of the competing models, eventually achieving the highest detection accuracy at the convergence stage. This significantly outperforms mainstream algorithms such as CondLaneNet and LaneATT, validating the effectiveness of the AMCFM and CSDA modules in enhancing feature representation.

In terms of the Loss curves, AG-CLRNet shows a slightly faster decline than the original CLRNet, reaching a lower final convergence value with fewer fluctuations in the later stages. This demonstrates that the LDED provides stronger fitting capability and optimization stability when regressing lane geometric parameters, enabling a more accurate approximation of the ground-truth labels. In summary, while maintaining training stability, AG-CLRNet not only significantly improves the final detection accuracy but also exhibits superior convergence efficiency, further confirming the rationality and feasibility of our proposed strategies for complex lane detection tasks.

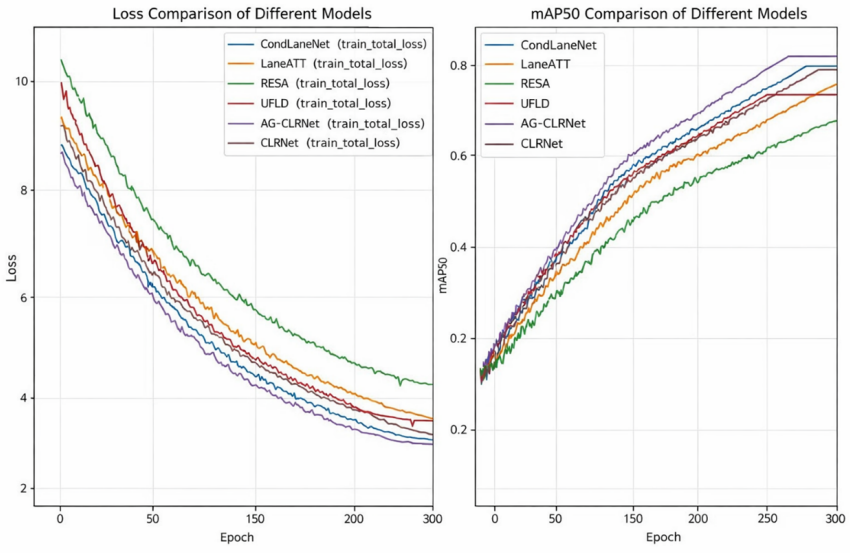

Figure 6. Detection curves

4 Conclusion

In this paper, addressing the challenges in lane detection under complex road scenarios—such as inconspicuous distant slender lanes, strong background interference, and insufficient modeling of lane topological structures—we propose an improved lane detection network, AG-CLRNet, based on the classical CLRNet framework. This network optimizes the original feature encoding and decoding processes from three levels: multi-scale context modeling, salient region enhancement, and long-range structural modeling.

Specifically, we first introduce the AMCFM, which effectively improves the response intensity of distant and slender lanes and enhances lane-background discrimination through cross-scale receptive field aggregation and semantic-guided channel redistribution. Secondly, the CSDA module is designed to reinforce the saliency of lane candidate regions while suppressing interference from curbs and road markings, thereby improving the discriminativeness and robustness of feature representations. Finally, the LDED is constructed to explicitly capture long-range dependencies and structural constraints along the main direction of the lanes. This optimizes the overall connectivity and geometric consistency of the detected lanes, mitigating issues such as fragmentation, jitter, and unstable topological identification.

Experimental results on typical lane detection datasets, such as CULane, demonstrate that AG-CLRNet maintains high inference efficiency while achieving various degrees of improvement over the baseline CLRNet in key metrics including F_1, Precision, and Recall. Furthermore, it exhibits superior detection stability and structural restoration capabilities in challenging scenarios like backlighting, shadow occlusion, and dense traffic. Ablation studies further verify the synergistic gains of the AMCFM, CSDA, and LDED modules in multi-scale context enhancement, salient region modeling, and long-range structural constraints. Overall, AG-CLRNet provides an improved solution for lane detection in complex road environments that balances accuracy and speed with strong generalization capability, offering significant engineering value for enhancing the reliability of lane perception in autonomous driving systems.

Data Availability

The data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] S. D. Pendleton, H. Andersen, X. Du, X. Shen, M. Meghjani, Y. H. Eng, D. Rus, and M. H. Ang, "Perception, planning, control, and coordination for autonomous vehicles," *Machines*, vol. 5, no. 1, p. 6, 2017. https://www.mdpi.com/2075-1702/5/1/6

[2] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Found. Trends Comput. Graph. Vis.*, vol. 12, no. 1–3, pp. 1–308, 2020. https://www.nowpublishers.com/article/Details/CGV-079

[3] D. Neven, B. De Brabandere, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: An instance segmentation approach," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, 2018, pp. 286–291. https://ieeexplore.ieee.org/document/8500547

[4] M. Ghafoorian, C. Nugteren, N. Baka, O. Booij, and M. Hofmann, "El-gan: Embedding loss driven generative adversarial networks for lane detection," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, Munich, Germany, 2018. https://doi.org/10.48550/arXiv.1806.05525

[5] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, 2018. https://ieeexplore.ieee.org/document/8063438

[6] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, and R. Benenson, "The Cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 3213–3223. https://ieeexplore.ieee.org/document/7780719

[7] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic lane markings in urban streets," in *Proceedings of the IEEE International Intelligent Transportation Systems Conference (ITSC)*, The Hague, Netherlands, 2013, pp. 576–581. https://ieeexplore.ieee.org/document/6728299

[8] N. Garnett, D. Cohen, R. Pe'er, and D. Lahav, "Real-time lane detection using deep neural networks," in *Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV)*, Paris, France, 2019, pp. 1474–1479. https://ieeexplore.ieee.org/document/8814091

[9] A. Gurghian, T. Koduri, S. Bailur, K. Carey, and V. Murali, "Deeplanes: End-to-end lane position estimation using deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern*

Recognition Workshops (CVPRW), Las Vegas, NV, USA, 2016, pp. 38–45. https://ieeexplore.ieee.org/document/7789502

[10] Z. Qin, P. Zhang, and X. Li, "Ultra fast deep lane detection with hybrid anchor driven ordinal classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 1133–1145, 2022. https://ieeexplore.ieee.org/document/9811225

[11] G. Sivakumar, E. Almehdawe, and G. Kabir, "Development of a collaborative decision-making framework to improve the patients' service quality in the intensive care unit," in *Proceedings of the 2020 International Conference on Decision Aid Sciences and Application (DASA)*, Sakhir, Bahrain, 2020, pp. 597–600. https://ieeexplore.ieee.org/document/9317286

[12] Z. Qu, H. Jin, Y. Zhou, Z. Yang, and W. Zhang, "Focus on local: Detecting lane marker from bottom up," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 2021, pp. 14 122–14 130. https://ieeexplore.ieee.org/document/9578529

[13] J. Philion, A. Karpathy, and S. Fidler, "Learning to evaluate perception models using planner-centric metrics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 12 214–12 223. https://ieeexplore.ieee.org/document/9157054

[14] N. Dahiya, Y. Fan, S. Bignardi, R. Sandhu, and A. Yezzi, "Polylanenet: Lane estimation via deep polynomial regression," in *Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021, pp. 6150–6156. https://ieeexplore.ieee.org/document/9413305

[15] B. Parsa and A. G. Banerjee, "A multi-task learning approach for human activity segmentation and ergonomics risk assessment," in *Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2021, pp. 2351–2361. https://ieeexplore.ieee.org/document/9423367

[16] V. Kalogeiton, P. Weinzaepfel, V. Ferrari, and C. Schmid, "Joint learning of object and action detectors," in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2001–2010. https://ieeexplore.ieee.org/document/8237481

[17] J. Kim and J. Lee, "Robust lane detection and tracking using deep neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 748–761, 2018. https://ieeexplore.ieee.org/document/7967948

[18] L. Tabelini, R. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "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 (CVPR)*, Nashville, TN, USA, 2021, pp. 294–302. https://ieeexplore.ieee.org/document/9577584

[19] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "Clrnet: Cross layer refinement network for lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 898–907. https://doi.org/10.48550/arXiv.2203.10350

[20] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proceedings of the International Conference on Learning Representations (ICLR)*, San Juan, Puerto Rico, 2016. https://arxiv.org/abs/1511.07122

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. https://ieeexplore.ieee.org/document/7780459

[22] L. Baraldi, M. Douze, R. Cucchiara, and H. Jegou, "Lamv: Learning to align and match videos with kernelized temporal layers," in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 7804–7813. https://ieeexplore.ieee.org/document/8578912

[23] K. Zhou, "Lane2seq: Towards unified lane detection via sequence generation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023, pp. 17 635–17 644. https://doi.org/10.48550/arXiv.2402.17172