

Article

Occlusion-Aware Lane Reconstruction and Lane–Vehicle Mask-Space Reasoning for Oversize-Vehicle Violation Detection

Jiajia Wu ^{1,2}, Yanyu Yan ³, Xiaojin Zhao ^{1,2}, Zhanlong Li ^{1,3}, Yufeng Li ³ and Gaoxu Deng ^{3,*} ¹ Shanxi Province Smart Transportation Laboratory Co., Ltd., Taiyuan 030032, China² Shanxi Province Smart Transportation Research Institute Co., Ltd., Taiyuan 030032, China³ School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China

* Correspondence: denggaoxu@tyust.edu.cn

Abstract

Background: Oversize transport vehicles pose substantial challenges to vision-based traffic monitoring because their large bodies can obscure lane markings and wheel–ground contact regions, while elevated oblique camera views introduce strong perspective distortion. To address these difficulties, this study presents an occlusion-aware lane reconstruction and lane–vehicle mask-space reasoning framework for oversize-vehicle violation detection. The lane-perception branch, termed CLRNet-ALD, extends CLRNet through adaptive cross-layer feature fusion, large-kernel attention, deformable convolution, and multi-loss supervision to improve the reconstruction of partially or severely occluded lane boundaries. In parallel, a YOLOv12-based instance-segmentation branch extracts vehicle contours and instance masks. The lane and vehicle masks are subsequently aligned in a common image coordinate system, and interpretable geometric rules based on row-matched boundary comparison, instance counting, and mask overlap are used to identify designated-lane deviation, simultaneous passage of multiple oversize vehicles, and lane-line crossing. On a 4000-image lane-focused dataset, CLRNet-ALD increased precision, recall, F1-score, and IoU from 86.3%, 83.9%, 85.1%, and 71.4% to 90.2%, 88.5%, 89.3%, and 77.8%, respectively. In a separate 600-image system-level evaluation, 157 of 167 confirmed violation cases were correctly recovered, corresponding to a positive-case recovery rate of 94.0%, while 141 of 143 confirmed multi-type vehicle cases were correctly identified (98.6%). These results demonstrate the practical feasibility of combining occlusion-aware lane reconstruction with interpretable lane–vehicle spatial reasoning. Because the lane-component and system-level experiments were conducted at different development stages, the reported system-level results are interpreted as validation of the complete reasoning pipeline rather than as a controlled measurement of the end-to-end improvement attributable specifically to CLRNet-ALD.

Keywords: oversize vehicle; traffic violation detection; lane detection; occlusion; instance segmentation; mask-space reasoning; intelligent transportation



Academic Editor: Ján Dižo

Received: 8 September 2026

Revised: 25 September 2026

Accepted: 2 October 2026

Published: 12 October 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Rapid infrastructure construction and the growth of large-equipment manufacturing have increased road transport demand for wind-turbine blades, transformers, heavy machinery, and other indivisible loads. Oversize vehicles are unusually long, wide, and high, so their lane position and passage mode are especially important on bridges, elevated roads, tunnels, and multi-lane facilities. Automated monitoring is attractive because fixed cameras can provide continuous evidence without requiring additional on-vehicle sensors [1].

The present framework is intended as an infrastructure-based roadside monitoring and violation-assessment system for fixed elevated cameras. It is not an onboard Lane Departure Avoidance (LDA) or Lane Keeping Assist (LKA) controller and, in its current form, does not directly transmit warning commands to the monitored vehicle or driver. References to onboard driver-assistance systems are included only as related visual-perception context.

Conventional violation-detection pipelines typically depend on visually complete vehicle boundaries, wheels, lane markings, or trajectories. In oversize-vehicle scenes, however, the vehicle body itself may cover a substantial portion of the lane boundary. From an elevated oblique viewpoint, the remaining lane segments may also be shortened, displaced, or distorted by perspective. These conditions make both perception and rule inference fragile: a missed lane segment can propagate directly into a false lane-position or lane-crossing decision.

Recent computer-vision studies have addressed traffic violation recognition with object detection, tracking, deep neural networks, and intelligent surveillance systems. In parallel, lane-detection research has advanced from spatial message passing and row-wise classification to cross-layer refinement and attention-guided approaches [2,3]. Yet the oversize-vehicle case requires both robust lane reconstruction and explicit spatial reasoning between a recovered lane and a large, partially occluding vehicle.

This study investigates the problem at two complementary experimental levels. First, CLRNNet is enhanced into CLRNNet-ALD to improve the reconstruction of long and interrupted lane markings under severe vehicle occlusion and perspective distortion [4–6]. Second, a complete lane–vehicle mask-space reasoning framework combines vehicle instance masks with reconstructed lane representations to recognize three practically relevant traffic violations [7–10]. The component-level experiment evaluates the effectiveness of the proposed lane-reconstruction enhancements, whereas the system-level experiment evaluates the operational feasibility of the complete perception-to-reasoning pipeline. Because these experiments were conducted at different stages of system development, their results are reported separately and are not interpreted as a controlled before-and-after comparison of the complete system.

A central methodological premise is that severe occlusion should be treated as a structured missing-information problem rather than as a conventional object-detection failure [11]. In the monitored elevated-road view, the vehicle body may remove direct evidence of the right lane boundary and wheel-ground contact points, but the remaining lane fragments, roadway perspective, vehicle contour, and opposite lane boundary still carry correlated geometric information. The proposed architecture, therefore, reconstructs the lane representation before applying an explicit spatial decision rule. This separation makes the inference chain interpretable: perception recovers the scene geometry, instance segmentation localizes the vehicle, and rule-based fusion converts their relative position into a violation label [12].

The resulting contribution is not merely the replacement of one lane detector with a stronger backbone. The combined framework establishes a task-specific coupling between an occlusion-aware lane-reconstruction model and a pixel-level violation-reasoning module. CLRNNet-ALD is designed to recover long interrupted lane structures under perspective distortion, while the Lane–Vehicle fusion stage consumes those reconstructed masks together with YOLOv12 vehicle masks. In this way, the upstream model is optimized for the exact type of structural evidence required by the downstream traffic-enforcement decision.

Accordingly, the main contributions of this study are summarized as follows:

1. An occlusion-oriented lane reconstruction approach is developed for fixed elevated-camera traffic scenes. Based on CLRNNet, adaptive cross-layer feature fusion, large-kernel attention, deformable convolution, and structural multi-loss supervision are

- incorporated to improve lane continuity and boundary localization under severe vehicle occlusion and perspective distortion.
2. A lane–vehicle mask-space representation is established for oversize-vehicle monitoring. Instead of relying on fully visible wheels or local lane fragments, reconstructed lane masks and vehicle instance masks are aligned in a common image coordinate system and used as explicit geometric evidence.
 3. Interpretable spatial rules are formulated for three practical violation scenarios. Designated-lane deviation is determined through row-matched boundary comparison, simultaneous oversize-vehicle passage through instance counting, and lane-line crossing through mask-overlap analysis.
 4. The proposed framework is evaluated at both component and system levels. A lane-focused experiment quantifies the effect of the CLNet-ALD enhancements, while a separate system-level experiment evaluates the complete perception, mask alignment, and rule-based reasoning pipeline. The two experimental levels are deliberately reported separately because they do not constitute a controlled end-to-end substitution experiment.

2. Related Work

2.1. Traffic Violation Detection

Vision-based traffic enforcement has progressed from handcrafted edge and Hough-transform pipelines toward deep detection and tracking [13]. Prior studies have used low-cost embedded vision for highway and toll-station rule monitoring, mobile AI-assisted surveillance, YOLO-based vehicle localization, DeepSORT-style tracking, and multi-task architectures for joint detection and behavior recognition [14]. These methods demonstrate the feasibility of automated enforcement but generally assume that decisive road and vehicle cues remain visible. Oversize vehicles violate that assumption because they can hide lane boundaries over long image regions. Recent YOLO-based methods have further improved vehicle perception under complex traffic conditions, while camera-based driver-assistance systems have combined lane detection, vehicle perception, and geometric risk assessment to support interpretable transportation decisions. These developments demonstrate the practical value of integrating road-structure and vehicle information rather than relying on vehicle appearance alone.

2.2. Lane Detection Under Occlusion

Lane detection is a structured prediction problem. Spatial CNNs and recurrent feature-shift aggregation exploit context along the lane direction, while row-based methods reduce computation by reformulating detection as structured classification [15,16]. CLNet explicitly combines high-level semantics with low-level localization features and introduces line-level IoU optimization. For strong occlusion, however, fixed cross-layer weighting, limited receptive fields, and regular-grid convolution may still be insufficient to reconstruct a long missing lane segment [17,18]. The proposed CLNet-ALD, therefore, introduces adaptive fusion, large-receptive-field attention, deformable sampling, and additional structural losses.

2.3. Research Gap and Design Rationale

The design adopted here follows an information-complementarity principle. Long-range lane context is used to infer missing structure; adaptive fusion controls the contribution of shallow and deep features according to scene content; deformable sampling accommodates perspective-induced local shape variation and vehicle masks provide a complementary geometric representation of the traffic scene. The final decision does not rely

solely on latent visual classification but instead evaluates the spatial relationship between semantically meaningful lane and vehicle masks. This formulation is well aligned with the practical definitions of lane occupancy and lane-line crossing and provides an interpretable connection between visual perception and the final enforcement decision.

The technical contribution should, therefore, be understood at the level of task-specific integration rather than as the invention of each individual operator. Adaptive feature fusion, large-kernel attention, deformable convolution, and auxiliary structural losses are established techniques; here they are organized around a specific failure mode in fixed elevated-camera oversize-vehicle scenes: long lane segments may be hidden by the vehicle while the remaining evidence is strongly perspective-compressed. ACFM balances local boundary detail and global semantic context, LKA propagates long-range structural evidence, DCN adapts local sampling to perspective-induced deformation, and the multi-loss objective constrains continuity and edge placement. The reconstructed representation is then consumed directly by the downstream lane-vehicle mask-space reasoning stage. This coupling between occlusion-oriented reconstruction and explicit traffic-rule reasoning is the principal design contribution of CLRNNet-ALD within the proposed framework.

Existing violation-detection systems and lane-detection models generally optimize different objectives [19–21]. Traffic-violation systems emphasize object recognition, tracking, or event classification, whereas lane detectors focus on line localization under standard driving views [22–26]. Oversize-vehicle enforcement combines both difficulties: the lane that determines legality can be exactly the lane segment hidden by the vehicle, and the camera is often fixed at an elevated oblique viewpoint with strong perspective compression. Consequently, a detector with high average lane accuracy may still fail at the decisive occluded segment, while a strong vehicle detector alone cannot infer whether the vehicle crosses an invisible boundary.

3. Proposed Framework

3.1. Problem Formulation

Let an input frame be $I \in \mathbb{R}^{H \times W \times 3}$. The lane branch predicts a set of lane masks $M_L = \{M_L^1, M_L^2, \dots\}$, and the vehicle branch predicts instance masks $M_V = \{M_V^1, M_V^2, \dots\}$. For a monitored oversize vehicle v , the objective is to estimate a violation state $y = (y_d, y_m, y_c)$, where y_d denotes deviation from the designated lane, y_m denotes simultaneous passage of multiple oversize vehicles, and y_c denotes lane-line crossing. Instead of learning y directly from image appearance, the framework factorizes the mapping $I \rightarrow (M_L, M_V) \rightarrow y$. This factorization preserves interpretable geometry and allows the decision stage to use pixel-level evidence.

The factorized formulation is particularly useful under occlusion because F_L can reconstruct a hidden lane segment from context before G evaluates the vehicle-lane relation.

$$y = G(M_L, M_V), M_L = F_L(I, \theta_L), M_V = F_V(I, \theta_V) \quad (1)$$

Here F_L is CLRNNet-ALD, F_V is the YOLOv12 instance-segmentation branch, and G is the deterministic mask-fusion and violation-reasoning function. The framework processes each road image through two parallel perception branches. The first branch estimates left and right lane boundaries with CLRNNet-ALD and converts them to binary lane masks. The second branch segments oversize trucks and other vehicles with YOLOv12, producing instance-level vehicle masks. The masks are transformed to the same image coordinate system. A rule layer then evaluates relative coordinates, the number of oversize-vehicle instances, and lane-vehicle overlap to produce violation labels. This design deliberately separates learned visual perception from transparent geometric enforcement logic.

Figure 1 treats lane reconstruction, vehicle instance segmentation, and violation reasoning as one integrated model rather than three unrelated algorithms. The two learned branches share the same input coordinate system and produce complementary pixel-level masks. Their outputs are co-registered before coordinate comparison, instance counting, and mask-overlap rules generate the final violation label and retain visual evidence for verification.

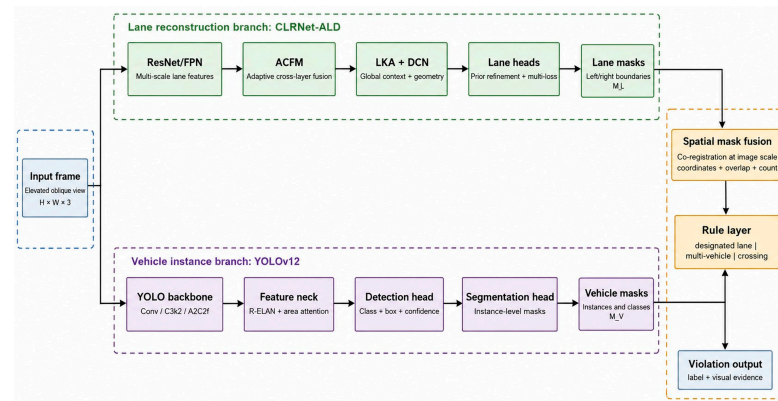


Figure 1. Integrated dual-branch network architecture.

3.2. CLRNet-ALD Lane Reconstruction

3.2.1. Adaptive Cross-Layer Feature Fusion

The adaptive cross-layer feature fusion module (ACFM) is designed to overcome the fixed feature utilization of the original CLRNet. Shallow features contain edges and local texture, whereas deep features provide semantic and global context. After 1×1 channel alignment, global response statistics and a lightweight attention mechanism estimate the relative importance of high- and low-level features for the current scene. The fused representation can be expressed as

$$F_{\text{fusion}} = \alpha F_{\text{high}} + \beta F_{\text{low}} \quad (2)$$

$$\alpha + \beta = 1 \quad (3)$$

The weights are normalized and learned adaptively. When a lane is locally occluded, the module can increase reliance on global semantic context while retaining available boundary details in visible regions.

Feature-response weighting can be interpreted as a reliability estimator for the two feature levels. Following the formulation, a response score $R(F)$ is obtained by averaging positive activations over the spatial domain. The normalized high-level weight is then computed relative to the sum of high- and low-level responses, and the low-level weight is its complement. This makes the fusion content-dependent rather than fixed across all images.

$$R(F) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \max(F_{i,j}, 0) \quad (4)$$

$$\alpha = \frac{R(F_{\text{high}})}{R(F_{\text{high}}) + R(F_{\text{low}})} \quad (5)$$

$$\beta = 1 - \alpha \quad (6)$$

Under heavy occlusion, high-level semantic context can dominate because it preserves the global lane hypothesis even when local pixels disappear. In lightly occluded

or well-illuminated regions, shallow features remain informative for boundary precision. This adaptive balance directly targets the two competing requirements of lane reconstruction: continuity across missing regions and accurate localization where visual evidence is available.

3.2.2. Large-Kernel Attention and Deformable Convolution

Lane markings are elongated structures and may remain invisible across a substantial image region. A large-kernel attention (LKA) block combines depthwise convolution, dilated depthwise convolution, and 1×1 channel mixing to enlarge the effective receptive field without a prohibitive increase in parameters. This enables the network to infer lane direction from distant visible context and suppress irrelevant background responses. Deformable convolution (DCN) complements LKA by learning sampling offsets around the regular convolution grid, allowing local filters to align with lane curvature and perspective-induced deformation. LKA, therefore, strengthens global structural reasoning, while DCN improves local geometric adaptation.

The LKA block increases the effective receptive field without requiring a prohibitively large dense convolution. A depthwise convolution extracts local spatial detail, a dilated depthwise convolution expands the contextual field, and a 1×1 convolution performs channel interaction. The resulting attention map reweights features along structures that are consistent with the global lane trajectory.

$$A(x) = \text{Conv}_{1 \times 1}(\text{DilatedDWConv}(\text{DWConv}(x))) \quad (7)$$

DCN complements this global mechanism at the local level. For an output position p , standard convolution samples a fixed grid p_k , whereas deformable convolution adds a learned offset Δp_k . Sampling locations, therefore, move with the local lane geometry, which is useful when oblique cameras compress lane spacing and cause apparent curvature or nonuniform line width.

LKA and DCN thus address different error sources: LKA strengthens long-range structural continuity, while DCN reduces local misalignment caused by perspective distortion and irregular visible fragments [27–30]. Their combination is expected to increase IoU more strongly than a purely semantic enhancement because the downstream fusion module depends on precise lane boundaries, not only lane existence.

$$y(p) = \sum_k w_k \cdot x(p + p_k + \Delta p_k) \quad (8)$$

3.2.3. Multi-Loss Optimization

To balance pointwise localization and line-level continuity, the lane module is trained using a weighted combination of classification loss, regression loss, Line IoU loss, geometric-consistency loss, and edge loss. The total objective is

$$\mathcal{L}_{\text{total}} = \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}} + \lambda_{\text{IoU}} \mathcal{L}_{\text{LineIoU}} + \lambda_{\text{geo}} \mathcal{L}_{\text{geo}} + \lambda_{\text{edge}} \mathcal{L}_{\text{edge}} \quad (9)$$

In Equation (9), $\mathcal{L}_{\text{cls}}$ denotes the classification loss for lane existence, $\mathcal{L}_{\text{reg}}$ the regression loss for lane/key-point localization, $\mathcal{L}_{\text{LineIoU}}$ the line-level IoU loss, $\mathcal{L}_{\text{geo}}$ the geometric-consistency loss, and $\mathcal{L}_{\text{edge}}$ the edge-alignment loss. The corresponding λ coefficients are weighting factors controlling the relative contribution of these terms. Classification and regression supervise lane existence and key-point location. Line IoU optimizes the lane as a complete structure. The geometric term encourages smoothness and perspective consistency, while the edge term penalizes discrepancies between predicted and annotated

boundaries. Together, these terms reduce discontinuity and positional drift in regions where direct visual evidence is weak.

Each loss term constrains a different failure mode. Classification loss suppresses false lane hypotheses; regression loss refines anchor or key-point coordinates; Line IoU loss couples predictions along the full lane extent; geometric consistency penalizes implausible discontinuities or abrupt curvature changes and edge loss emphasizes alignment between predicted and annotated lane boundaries. Because an occluded lane contains few directly observable pixels, relying on only pointwise regression would underconstrain the hidden segment. The combined objective introduces both local evidence and structural priors into optimization.

The loss terms are combined through weighted summation to balance local localization accuracy and global structural continuity. In the present evaluation, the multi-loss strategy is assessed as a complete optimization component. Because separate coefficient-level ablation records are not available, the manuscript does not attempt to attribute the observed improvement to individual loss weights. The reported results, therefore, support the practical utility of the combined loss formulation rather than the independent contribution of each coefficient.

3.3. YOLOv12 Vehicle Instance Segmentation

The vehicle branch adopts a YOLOv12-based instance-segmentation model to obtain category-specific vehicle masks. In this application, the principal output required by the downstream reasoning stage is the instance mask rather than the bounding box, because violation decisions depend on vehicle contours, relative boundary positions, and mask overlap. Separate instance masks are retained for oversize vehicles and other vehicle categories, enabling both contour extraction and vehicle-instance counting.

3.4. Lane–Vehicle Mask Fusion and Violation Logic

Let M_L and M_R denote the left- and right-lane masks and M_V the visible vehicle mask. The masks are aligned pixel-wise in the original image coordinate system. The system evaluates three rules in parallel.

Designated-lane rule: the lower-left vehicle boundary point $A(x_1, y_1)$ is compared with the rightmost lane-mask coordinate $C(x_3, y_1)$ at the same row, while the lower-right vehicle boundary point $B(x_2, y_2)$ is compared with the corresponding right-lane coordinate $D(x_4, y_2)$. A vehicle is considered inside the designated corridor when $x_1 > x_3$, $x_2 < x_4$. If the right vehicle boundary exceeds the right-lane boundary under the specified configuration, the system flags non-designated-lane travel.

Multiple-vehicle rule: because YOLOv12 generates an instance mask for each detected vehicle, the presence of two or more oversize-vehicle instances in the monitored bridge-passage region triggers a multi-vehicle passage violation, consistent with the single-vehicle passage requirement used in the application scenario.

Lane-crossing rule: after isolating the relevant lane boundary, the method computes the binary intersection $M_{\text{cross}} = M_V \cap M_L$. If the intersection contains no foreground pixels, the vehicle contour does not cross the monitored lane region. Otherwise, lane-line crossing or leftward encroachment is flagged. The number or area of overlapping pixels can additionally quantify crossing severity, separating marginal contact from extensive overlap.

For a binary vehicle mask M_V and the relevant lane mask M_L , define the overlap area $S_{\cap} = |M_V \cap M_L|$ and vehicle area $S_V = |M_V|$. A normalized crossing score can be written as $\rho = \frac{S_{\cap}}{S_V}$. The implemented rule uses the zero/nonzero intersection condition for violation detection, while ρ provides a natural extension for severity grading because it is invariant

to vehicle-mask area. Small isolated overlap can correspond to line touching, whereas a larger ratio indicates more substantial lane occupation.

The reported implementation uses the binary zero/nonzero intersection rule without an additional minimum-overlap threshold. We recognize that this choice can be sensitive to segmentation noise, lane-mask thickness, or boundary-touching cases. The normalized overlap ratio ρ is, therefore, presented as a natural extension for tolerance or severity grading rather than as an additional experimentally validated threshold in the present study. Controlled sensitivity analysis of overlap thresholds, coordinate tolerance, mask thickness, and optional morphological filtering is left for future work.

For designated-lane compliance, let $A(x_1, y_1)$ and $B(x_2, y_2)$ be the lower-left and lower-right points of the vehicle contour. Let $C(x_3, y_1)$ and $D(x_4, y_2)$ be the corresponding lane-boundary coordinates on the same image rows. The vehicle is geometrically contained when its lateral boundary remains between the two lane limits. The implementation expresses this through coordinate inequalities such as $x_1 > x_3$, $x_2 < x_4$. A violation is triggered when the vehicle boundary passes beyond the corresponding lane limit.

This row-matched comparison reduces sensitivity to perspective scaling because vehicle and lane coordinates are evaluated at approximately the same image height. It also avoids estimating real-world lane width from camera calibration, which simplifies deployment when the monitoring camera is fixed but intrinsic/extrinsic calibration is unavailable.

$$\rho = \frac{S_{\cap}}{S_V} = \frac{|M_V \cap M_L|}{|M_V|} \quad (10)$$

3.5. End-to-End Inference Procedure

The complete inference procedure integrates learned perception with deterministic spatial reasoning, as summarized in Figure 2. A monitoring frame is processed by the CLRNNet-ALD lane branch and the YOLOv12 vehicle-instance branch in a shared image coordinate system [31,32]. Their mask outputs are spatially aligned, geometric primitives are extracted, and the designated-lane, multiple-vehicle, and lane-crossing rules are evaluated in parallel. The rule outputs are then aggregated while retaining the corresponding image and mask evidence for verification.

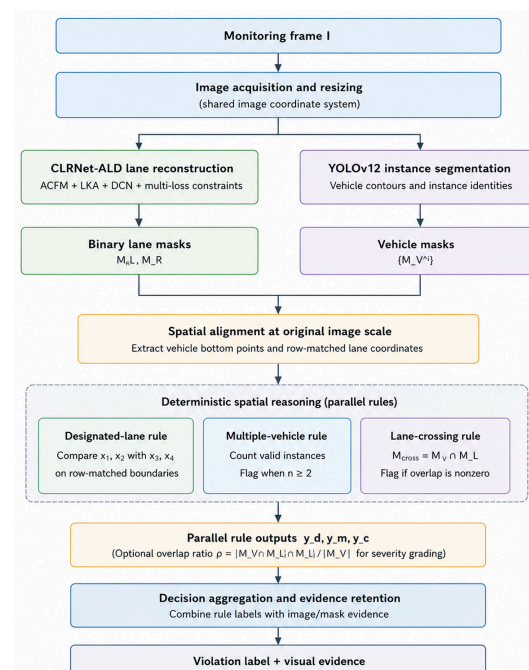


Figure 2. End-to-end inference procedure of the proposed Lane–Vehicle violation detection framework.

4. Experimental Setup

4.1. Datasets and Annotation

The first dataset is a lane-focused dataset containing 4000 static road images collected from elevated-camera oversize-vehicle monitoring scenes. The images cover different illumination and weather conditions, including sunny, cloudy, and dusk scenes, as well as different degrees of vehicle-induced lane occlusion, ranging from partial occlusion to large-area and near-complete occlusion. Lane regions were manually annotated using LabelMe. Visible lane regions were labeled according to directly observable image evidence, whereas lane segments hidden by large vehicles were completed according to the continuity of the surrounding lane geometry and road context. All annotations were manually reviewed before model training and evaluation. The dataset was divided into training, validation, and test subsets at a ratio of 8:1:1, corresponding to 3200, 400, and 400 images, respectively.

The second dataset is a system-level dataset containing 6000 road images at a resolution of 4096×2160 pixels. The images were acquired across different recording periods, weather conditions, and real driving scenarios in elevated-road oversize-vehicle monitoring environments. This dataset was used for training and evaluating the vehicle-instance perception and lane-vehicle reasoning components of the complete system. Lane boundaries and vehicles were manually annotated using LabelMe. The left and right lane boundaries were assigned separate identifiers, while oversize vehicles and other vehicle categories were assigned separate class labels and independent instance identifiers to support downstream contour extraction, instance counting, and spatial reasoning.

The 6000-image system-level dataset was divided into training, validation, and test subsets using an 8:1:1 ratio, corresponding to 4800 training images, 600 validation images, and 600 test images. The 600-image test subset constitutes the primary system-level evaluation set used in Section 5 and Table 1. This test subset was excluded from model training and was used to evaluate the complete perception-to-reasoning pipeline, including vehicle instance perception, lane and vehicle mask alignment, and rule-based violation detection.

Table 1. Datasets used in the component-level and system-level evaluations.

Dataset	Primary Task	Images	Split/Evaluation	Key Annotation
Lane-focused set	Occluded lane reconstruction	4000	8:1:1 train/validation/test	Pixel-level lane labels with contextual completion of occluded segments
System set	Vehicle masks and violation reasoning	6000	8:1:1 train/validation/test; 600-image system test	Left/right lane masks and vehicle instances
Closed-loop verification set	Supplementary system verification	200	Previously unseen images	Violation-related lane and vehicle evidence
On-site field set	Supplementary field validation	600	Additional field images	Violation-related visual evidence

In addition to the above datasets, 200 previously unseen images were used for supplementary closed-loop system verification, and a further 600 on-site images were used for supplementary field validation. These additional image sets were not included in the primary quantitative results reported in Table 4 and are treated separately from the 600-image system test subset.

The two main datasets, therefore, serve distinct but complementary purposes. The 4000-image lane-focused dataset is used to evaluate whether CLRNNet-ALD improves lane

reconstruction under vehicle occlusion and perspective distortion, whereas the 6000-image system-level dataset is used to develop and evaluate the complete vehicle-perception and lane–vehicle mask-space reasoning framework. Accordingly, the component-level and system-level results are reported separately and are not interpreted as a single jointly controlled benchmark.

For lane annotation, directly visible pixels were labeled from observable image evidence, while fully or partially hidden segments were completed from surrounding lane continuity and road context and were manually reviewed. The historical annotation record does not contain a formal inter-annotator agreement study or a visibility-stratified annotation-confidence score. Consequently, no retrospective agreement statistic is reported, and context-completed labels in fully occluded regions are interpreted as structural reference annotations rather than directly observed ground truth.

The retained dataset records confirm the 8:1:1 image-level partitions reported above, but they do not preserve sufficiently detailed passage-level or video-sequence grouping information to retrospectively verify that temporally adjacent frames from the same vehicle passage were always confined to a single subset, shown as Table 1. This is, therefore, treated as a limitation of the present benchmark. Future dataset construction will use event-, sequence-, or recording-period-level partitioning to prevent potential temporal leakage.

4.2. Training Configuration

All experiments were conducted using an Intel Core i7-14700 CPU, an NVIDIA GeForce RTX 4060 Ti GPU with 16 GB memory, and 32 GB system memory under Ubuntu 22.04 LTS. The lane-component experiment and the system-level experiment were conducted at different stages of system development and, therefore, used closely related but not completely identical software environments and optimization settings. To preserve the validity of the reported results, each experimental result is associated with the configuration under which it was originally obtained. Comparisons within each controlled experiment use consistent hardware, data partitions, and parameter settings, while results across the two experimental stages are treated as complementary rather than directly interchangeable, shown as Table 2.

Table 2. Training configurations used in the component-level and system-level experiments.

Module	Optimizer	Initial Learning Rate	Epochs	Schedule	Weight Decay
CLRNet-ALD	AdamW	0.001	100	Decay at epochs 30 and 60; factor 0.1	0.0001
YOLOv12 system branch	SGD	0.02	80	Decay at epochs 50 and 70; factor 0.1	0.0005
CLRNet system baseline	AdamW	1×10^{-4}	100	Decay at 40, 70, and 90; factor 0.5	0.01

4.3. Evaluation Metrics

Lane reconstruction is evaluated using pixel-level precision, recall, F1-score, and intersection over union (IoU). Precision measures the proportion of predicted lane pixels that are correct, recall measures the proportion of annotated lane pixels recovered by the model, F1-score balances precision and recall, and IoU measures the spatial agreement between the predicted and annotated lane regions. IoU is particularly relevant to the present application because downstream lane–vehicle reasoning depends on boundary placement and mask overlap rather than only on the presence of a lane.

For the system-level evaluation, the available experimental record provides the numbers of confirmed positive cases and correctly recovered positive cases. Therefore, the directly computable system-level quantity is reported as the positive-case recovery rate.

$$R_p = \frac{N_{\text{correct_positive}}}{N_{\text{confirmed_positive}}}. \quad (11)$$

In the 600-image evaluation, 157 of 167 confirmed violation cases were correctly recovered, yielding $R_p = 94.0\%$. Similarly, 141 of 143 confirmed multi-type vehicle cases were correctly identified, yielding $R_p = 98.6\%$.

These values should not be interpreted as conventional all-sample classification accuracy because the available aggregate records do not contain the complete false-positive and true-negative counts required to reconstruct a full confusion matrix. Consequently, system-level precision, specificity, conventional accuracy, and per-class F1-score are not claimed in the present manuscript. A complete confusion-matrix-based evaluation is identified as an important extension of the current study.

5. Results

The experiments are interpreted against three predictions derived from Section 3. Prediction 1 is that ACFM should improve recovery of missing lane evidence, primarily reflected in recall. Prediction 2 is that LKA, DCN, and structural supervision should improve continuity and boundary placement, primarily reflected in IoU and qualitative reconstruction under occlusion. Prediction 3 is that distinct vehicle instances and co-registered lane/vehicle masks should support traceable rule decisions, reflected in multi-type detection, end-to-end violation recovery, and visualized intermediate evidence. The following subsections test these predictions in sequence; component-level and system-level results are not merged into an unmeasured causal claim.

5.1. Lane-Line Detection

CLRNet-ALD consistently outperforms the original CLRNet on the lane-focused test set. Precision increases from 86.3% to 90.2%, recall from 83.9% to 88.5%, F1-score from 85.1% to 89.3%, and IoU from 71.4% to 77.8%. The simultaneous improvement in recall and precision indicates that the model recovers more annotated lane pixels without merely expanding the predicted lane region. The 6.4-percentage-point increase in IoU further indicates improved spatial agreement between the reconstructed and annotated lane regions, which is particularly relevant to the downstream mask-overlap and boundary-comparison rules.

The gain is balanced rather than concentrated in one metric: precision rises by 3.9 percentage points, recall by 4.6 points, and F1 by 4.2 points. Higher recall is consistent with reconstruction of lane pixels hidden by the vehicle, while the simultaneous precision increase shows that this recovery is not achieved simply by expanding the predicted lane region. The larger IoU improvement indicates better continuity and lateral placement, reducing the risk that a visually plausible but shifted boundary changes a designated-lane or crossing decision.

The metric pattern is consistent with the complementary roles proposed in Section 3. ACFM selects between local detail and global semantics according to scene reliability; LKA propagates long-range lane context across missing regions; DCN adapts sampling to perspective-induced deformation and the structural losses constrain continuity and edge placement. Figure 3 confirms stable optimization under the scheduled learning-rate decay, while Table 3 confirms that the converged model improves both detection completeness and geometric agreement. Training convergence alone does not prove the mechanism, but

together with the controlled comparison and ablation evidence in Section 5.6, it supports the theoretical account.

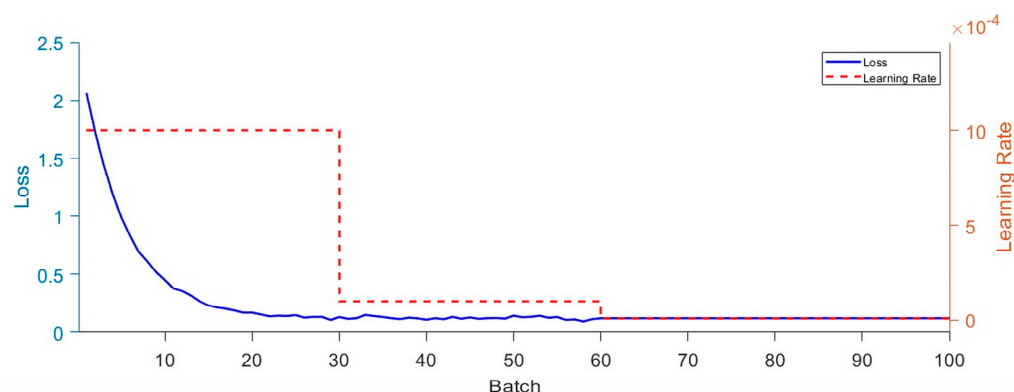


Figure 3. Training convergence of CLRNNet-ALD: loss evolution and scheduled learning-rate decay over 100 epochs.

Table 3. Lane-detection performance on the 4000-image lane-focused dataset.

Method	Precision (%)	Recall (%)	F1 (%)	IoU (%)
CLRNNet	86.3	83.9	85.1	71.4
CLRNNet-ALD	90.2	88.5	89.3	77.8

In Figure 4, the large vehicle removes a substantial portion of the visible right-lane evidence. The baseline result contains disconnected lane fragments and noticeable boundary drift, whereas CLRNNet-ALD produces a smoother continuation through the occluded region. This qualitative observation is consistent with the intended roles of adaptive feature fusion, long-range contextual modeling, and deformable geometric adaptation. Together with the quantitative improvement in recall and IoU, the visualization provides supporting evidence that the proposed modifications improve lane continuity under occlusion.

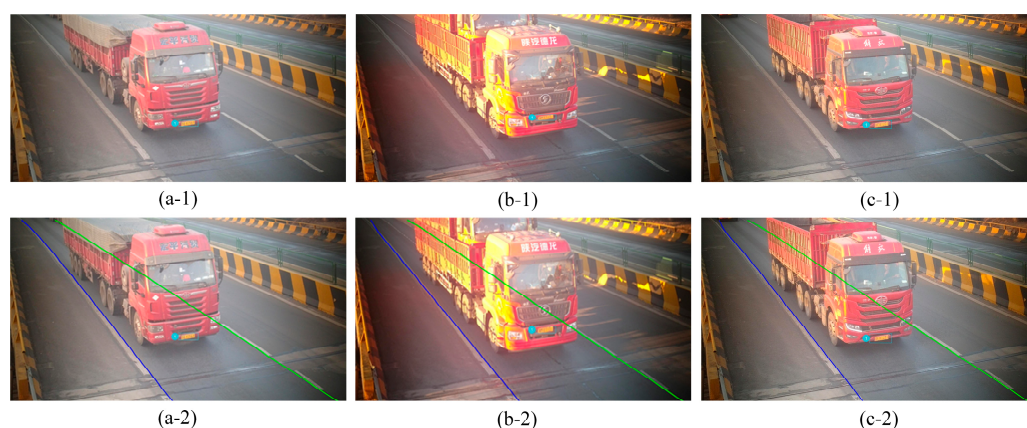


Figure 4. Qualitative lane-detection results. (a-1–c-1), different scenarios to be detected; (a-2–c-2), lane line detection result.

5.2. Vehicle Instance Detection

The second stage tests the vehicle-side premise of the Lane–Vehicle theory: violation reasoning requires separate, category-specific vehicle instances rather than a single merged foreground region. YOLOv12 preserves distinct masks for oversize vehicles, heavy trucks, and passenger cars in both single- and multi-vehicle scenes, enabling reliable instance counting and contour extraction.

Figure 5 illustrates that adjacent vehicles can be retained as separate instances rather than merged into a single foreground region, which is necessary for the multiple-vehicle rule. In the available system-level record, 141 of 143 confirmed multi-type vehicle cases were correctly identified, corresponding to a positive-case recovery rate of 98.6%. This result suggests that vehicle-instance separation is effective in the evaluated scenarios. However, because a complete instance-segmentation confusion matrix and class-wise mask AP statistics are not available in the current record, the result should not be interpreted as a comprehensive evaluation of vehicle segmentation quality.



Figure 5. Vehicle instance-detection results in multi-vehicle scenes. (a-1,b-1), different vehicles to be detected; (a-2,b-2), detected result.

5.3. Violation Detection

The third stage tests the central mask-space reasoning premise. Vehicle masks and reconstructed lane regions are aligned at the original image scale; designated-lane, multiple-vehicle, and lane-crossing rules then operate on row-matched coordinates, instance counts, and binary overlap. If the theory is correct, the intermediate masks should preserve the evidence used by the final decision rather than behaving as unrelated visual outputs.

The lane and vehicle masks are aligned at the original image scale before the three rule-based decisions are evaluated. The intermediate visualizations in Figure 6 show the vehicle instance masks, lane–vehicle mask alignment, and the resulting violation decision. These intermediate representations provide an interpretable evidence chain from perception to decision and allow the final output to be inspected in terms of explicit geometric relations rather than only a latent classification score.

5.4. Quantitative Results

The integrated procedure was evaluated on 600 road images. Among 167 confirmed violation cases, 157 were correctly recovered and 10 were missed or incorrectly handled, corresponding to a positive-case recovery rate of 94.0%. In the multi-type vehicle subset, 141 of 143 confirmed positive cases were correctly identified, corresponding to 98.6%. Because the available aggregate record does not include the complete false-positive and true-negative counts, these percentages are reported as positive-case recovery rates rather than conventional classification accuracy.

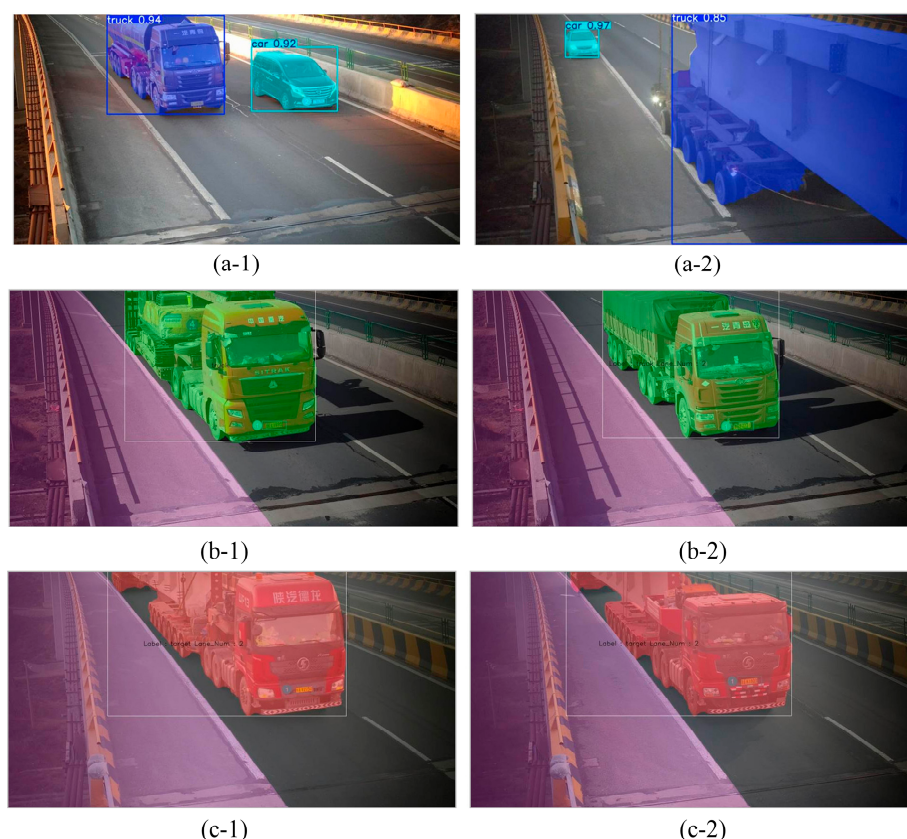


Figure 6. End-to-end violation-detection results: (a-1,a-2),vehicle detection, (b-1,b-2) lane/vehicle mask fusion, (c-1,c-2) violation visualization.

The difference between 98.6% multi-type detection and 94.0% violation recovery localizes the more difficult part of the integrated system to lane reconstruction and spatial relationship reasoning rather than vehicle-instance separation alone. The value of 6.0% represents the proportion of confirmed positive violation cases that were not correctly recovered (10/167), rather than the error rate over all 600 images. This is diagnostic evidence, not a formal causal decomposition, because per-module error traces are unavailable. Moreover, because false-positive and true-negative counts for all 600 inputs are not reported, both percentages are interpreted as positive-case recovery measures under the definitions rather than independently reconstructed conventional all-image accuracy.

Tables 4 and 5 report the same 1281-s elapsed time, so it is treated as one shared batch-level procedure time and is not added across tasks. The derived average is 2.135 s per input (0.468 image/s), which characterizes the documented offline workflow, including unseparated processing overhead. It should not be equated with the separately reported 45-FPS model-inference figure. A deployment study should instrument lane inference, vehicle inference, spatial fusion, rule evaluation, file I/O, and evidence generation independently.

Table 4. Experimental results.

Task	Images	Confirmed Target Cases	Correct	Incorrect	Elapsed Time (s)	Positive-Case Recovery Rate
Violation detection	600	167	157	10	1281	94.0%
Multi-type vehicle detection	600	143	141	2	1281	98.6%

Table 5. Derived operational measures from the experiments.

Task	Correct/Target	Positive-Case Miss Rate	Time Per Input (s)	Throughput (Image/s)
Violation detection	157/167	6.0%	2.135	0.468
Multi-type vehicle detection	141/143	1.4%	2.135	0.468

5.5. Reference Comparison with Alternative Approaches

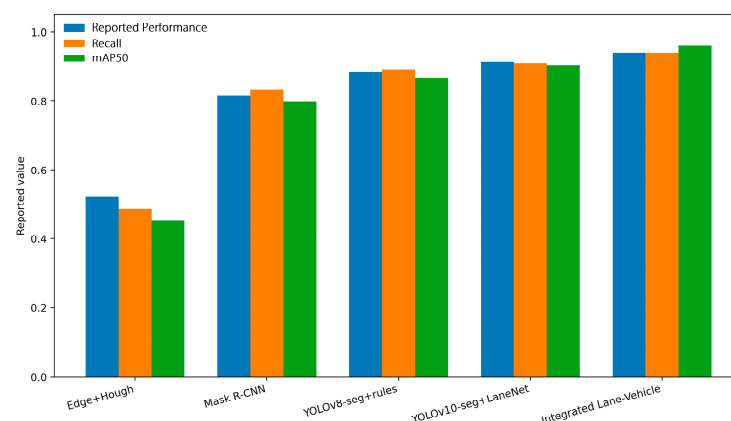
To provide contextual reference for the proposed framework, Table 6 summarizes the performance of several representative alternative approaches, including Edge detection + Hough transform, Mask R-CNN, YOLOv8-seg with traditional rules, and YOLOv10-seg + LaneNet. These approaches represent progressively more advanced levels of traffic-scene perception, ranging from handcrafted geometric processing to learned instance segmentation and combined lane–vehicle perception.

Table 6. Reference comparison with representative alternative approaches.

Method	Reported Performance	Recall	mAP50
Edge detection + Hough transform	0.523 (Accuracy)	0.487	0.452
Mask R-CNN	0.815 (Accuracy)	0.832	0.798
YOLOv8-seg + traditional rules	0.884 (Accuracy)	0.891	0.865
YOLOv10-seg + LaneNet	0.912 (Accuracy)	0.908	0.902
Integrated Lane–Vehicle network	0.940 (Positive-case recovery rate)	0.940	0.960

The performance measure in the second column follows the metric available for each method. For the proposed Integrated Lane–Vehicle network, 0.940 denotes the positive-case recovery rate (157 correctly recovered cases among 167 confirmed positive violation cases), rather than conventional all-sample classification accuracy. Because complete false-positive and true-negative counts were not available, conventional accuracy could not be reconstructed for the proposed system. Therefore, the values in this column are provided for contextual reference and should not be interpreted as directly equivalent metrics under identical evaluation conditions.

Figure 7 visualizes the reported values in Table 6 only as a contextual reference. The first series is labeled as the reported primary metric value because its mathematical definition is not uniform across methods; specifically, the proposed framework uses positive-case recovery rate rather than conventional accuracy. Recall and mAP50 are shown as reported in the available comparison record. Accordingly, the figure is not interpreted as a controlled superiority test.

**Figure 7.** Reference-level comparison of reported performance among representative approaches.

5.6. Mechanism Verification, Robustness, and Reproducibility

The evidence across Sections 5.1–5.5 makes the proposed dependency chain explicit. Lane reconstruction determines the permitted corridor; vehicle segmentation supplies contours and instance counts; co-registration enables row-matched comparison and overlap and the rule layer produces an inspectable decision. A lane error can shift the corridor, a fragmented vehicle mask can create a false overlap, and an instance-count error can activate or suppress the multiple-vehicle rule. Conversely, the jointly observed lane gains, separated vehicle instances, and traceable fusion outputs are mutually consistent with the mechanism proposed in Section 3.

The available ablation is cumulative and reflects the historical development sequence of the model. Relative to CLNet, adding ACFM increases recall; subsequently adding LKA+DCN further improves IoU and activating the combined multi-loss objective yields the final reported metrics, shown as Figure 8. These trends are consistent with the intended functions of the components, but the current design does not isolate the independent effects of LKA versus DCN, individual loss terms, or loss coefficients. Separate repeated runs, parameter/FLOP profiling, and coefficient-level ablations were not retained and are, therefore, not retrospectively inferred.

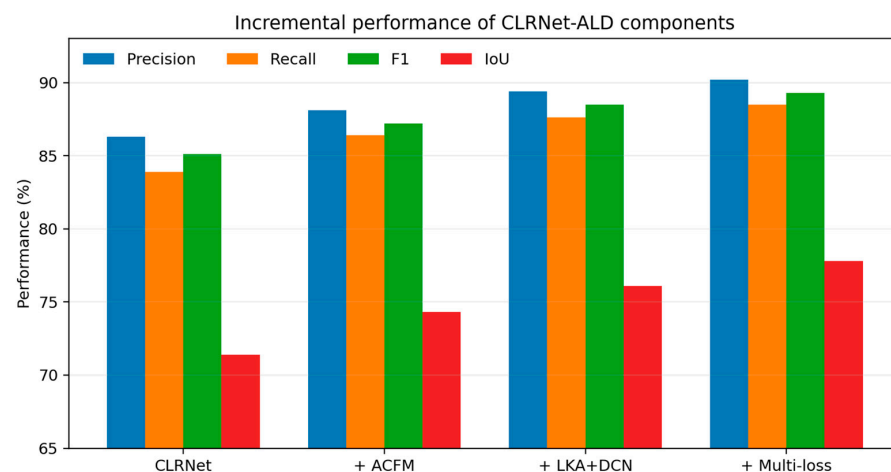


Figure 8. Incremental ablation of the CLNet-ALD components.

The component-level and system-level experiments support two distinct conclusions. The lane-focused experiment demonstrates that CLNet-ALD improves lane reconstruction relative to the original CLNet under the evaluated conditions, while the system-level experiment demonstrates that the complete lane–vehicle mask-space reasoning pipeline can produce interpretable violation decisions on real road images. Because the system-level experiment was not conducted as a controlled substitution of CLNet by CLNet-ALD under otherwise identical conditions, the 94.0% positive-case recovery rate is not attributed specifically to the CLNet-ALD replacement.

6. Discussion

The principal advantage of the proposed framework is the explicit separation between scene reconstruction and traffic-rule reasoning. Severe vehicle occlusion first affects the availability of lane evidence; therefore, the perception stage attempts to reconstruct a continuous lane representation from visible context. The reconstructed lane mask is then compared with vehicle instance masks in a common image coordinate system. This formulation avoids relying exclusively on wheel–ground contact regions, which are frequently hidden by oversize vehicles.

A second advantage is interpretability. The system can retain the reconstructed lane mask, vehicle instance mask, mask-overlap region, and boundary coordinates used by the rule layer. These intermediate representations provide inspectable evidence for the final decision and facilitate manual verification and error analysis.

The results nevertheless have several important limitations. First, the lane-component experiment and the complete system experiment were conducted at different stages of development and do not constitute a controlled before-and-after substitution of the complete system; therefore, the system-level positive-case recovery rate cannot be attributed specifically to replacing CLRNNet with CLRNNet-ALD. Second, the current rule layer assumes a fixed elevated-camera configuration and application-specific lane-passage policies, so substantial changes in camera pose, road geometry, lane configuration, or infrastructure-induced occlusion may require adaptation. Third, complete FP and TN counts are unavailable for each violation category, preventing reconstruction of conventional accuracy, specificity, precision, false-positive rate, and per-class F1-score. Fourth, context-completed annotations of fully occluded lane regions contain unavoidable uncertainty, and formal inter-annotator agreement records are unavailable. Fifth, the retained split records do not permit retrospective verification of passage-level or sequence-level separation. Sixth, controlled sensitivity studies for camera/image resolution, heavy rain, snow, fog, overlap thresholds, mask thickness, vehicle speed, or onboard LDA/LKA status were not part of the present evaluation. The system should, therefore, be interpreted as an infrastructure-based roadside monitoring and violation-assessment framework rather than an onboard active-safety controller or a vehicle-dynamics stability estimator.

Future work should, therefore, focus on a unified end-to-end benchmark in which the original CLRNNet and CLRNNet-ALD are substituted within the same complete reasoning pipeline under identical training and test conditions. Priority items include event- or sequence-level dataset partitioning, complete class-wise confusion matrices, visibility-stratified and adverse-weather evaluation, controlled image-resolution and camera-viewpoint studies, threshold/morphology sensitivity analysis, independent and repeated ablation runs with parameter/FLOP/memory profiling, and module-wise latency measurement. Temporal tracking could further support vehicle-speed and directional-stability analysis, while additional views from escort-vehicle cameras or unmanned aerial vehicles could complement fixed roadside monitoring in severely occluded scenes.

7. Conclusions

This study presents an occlusion-aware lane reconstruction and lane–vehicle mask-space reasoning framework for oversize-vehicle violation detection under elevated-camera traffic monitoring conditions. The framework combines CLRNNet-ALD for lane reconstruction, YOLOv12-based vehicle instance segmentation, and explicit geometric reasoning based on boundary comparison, vehicle-instance counting, and mask overlap.

On the 4000-image lane-focused dataset, CLRNNet-ALD improved precision, recall, F1-score, and IoU from 86.3%, 83.9%, 85.1%, and 71.4% to 90.2%, 88.5%, 89.3%, and 77.8%, respectively. In the separate 600-image system-level evaluation, 157 of 167 confirmed violation cases were correctly recovered, corresponding to a positive-case recovery rate of 94.0%, while 141 of 143 confirmed multi-type vehicle cases were correctly identified, corresponding to 98.6%.

The results indicate that occlusion-aware lane reconstruction can improve lane-structure recovery and that aligned lane and vehicle masks can support interpretable violation reasoning in the evaluated real-road monitoring scenarios. However, the component-level and system-level experiments were conducted at different development stages and should, therefore, be interpreted as complementary rather than as a controlled end-to-end

comparison. Complete system-level confusion statistics, uniformly retrained baselines, sequence-level split verification, and independent submodule ablations are not available in the current record. Future work will, therefore, emphasize controlled end-to-end benchmarking, complete class-wise performance analysis, robustness under different image resolutions, adverse weather, occlusion and camera conditions, temporal vehicle-motion analysis and deployment-oriented latency profiling.

Author Contributions: Conceptualization, X.Z.; Methodology, Y.Y.; Software, J.W.; Validation, Y.L.; Writing—original draft, G.D.; Writing—review & editing, G.D.; Supervision, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Joint Funding Project of the Basic Research Program of Shanxi Province (Transportation Control), Grant No. 202303011222008, and the Innovation and Development Program of Shanxi Transportation Research Institute Group Co., Ltd., Grant No. 23-JKCF-02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in this study were collected from real-road oversize-vehicle monitoring scenarios and contain application-specific roadside traffic imagery. The datasets are not currently deposited in a public repository. Additional information regarding the data and annotation protocol may be made available by the corresponding author upon reasonable request, subject to applicable project and data-management restrictions.

Conflicts of Interest: Authors Jiajia Wu, Xiaojin Zhao and Zhanlong Li were employed by the company Shanxi Province Smart Transportation Laboratory Co., Ltd. Authors Jiajia Wu and Xiaojin Zhao were employed by the company Shanxi Province Smart Transportation Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li, X.; Li, J.; Hu, X.; Yang, J. Line-CNN: End-to-End Traffic Line Detection with Line Proposal Unit. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 248–258. [CrossRef]
2. Pan, X.; Shi, J.; Luo, P.; Wang, X.; Tang, X. Spatial as Deep: Spatial CNN for Traffic Scene Understanding. In *Thirty-Second AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2018; Volume 32.
3. Zheng, T.; Fang, H.; Zhang, Y.; Tang, W.; Yang, Z.; Liu, H.; Cai, D. RESA: Recurrent Feature-Shift Aggregator for Lane Detection. In *AAAI-21 Technical Tracks 4*; AAAI Press: Palo Alto, CA, USA, 2021; Volume 35, pp. 3547–3554.
4. Zheng, T.; Huang, Y.; Liu, Y.; Tang, W.; Yang, Z.; Cai, D.; He, X. CLNet: Cross Layer Refinement Network for Lane Detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 898–907.
5. Qin, Z.; Wang, H.; Li, X. Ultra Fast Structure-Aware Deep Lane Detection. *arXiv* **2020**, arXiv:2004.11757.
6. 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 *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2021; pp. 294–302.
7. Liu, L.; Chen, X.; Zhu, S.; Tan, P. CondLaneNet: A Top-to-Down Lane Detection Framework Based on Conditional Convolution. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021; pp. 3773–3782.
8. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2017; pp. 764–773.
9. Ding, X.; Zhang, X.; Han, J.; Ding, G. Scaling Up Your Kernels to 31×31 : Revisiting Large Kernel Design in CNNs. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 11963–11975.
10. Seif, G.; Androutsos, D. Edge-Based Loss Function for Single Image Super-Resolution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: New York, NY, USA, 2018; pp. 1468–1472.
11. Tian, Y.; Ye, Q.; Doermann, D. YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv* **2025**, arXiv:2502.12524.
12. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE: New York, NY, USA, 2017; pp. 2961–2969.
13. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. In *Advances in Neural Information Processing Systems 37*; Curran Associates, Inc.: Red Hook, NY, USA, 2024.

14. Nagrani, A.; Yang, S.; Arnab, A.; Jansen, A.; Schmid, C.; Sun, C. Attention Bottlenecks for Multimodal Fusion. In *Advances in Neural Information Processing Systems 34*; Curran Associates, Inc.: Red Hook, NY, USA, 2021; pp. 14200–14213.
15. Purohit, M.R.; Yadav, A.R.; Kumar, R.; Kumar, G.; Dhariwal, S.; Kumar, J. Next-Gen Traffic Rule Violation Detection Using Optimum Feature Extraction Techniques on Highway and Toll Tax Using Raspberry-Pi Hardware. In *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*; IEEE: New York, NY, USA, 2022; pp. 1–5.
16. Han, J.; Deng, X.; Cai, X.; Yang, Z.; Xu, H.; Xu, C.; Liang, X. Laneformer: Object-Aware Row-Column Transformers for Lane Detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 799–807. [[CrossRef](#)]
17. Dede, D.; Sarsıl, M.A.; Shaker, A.; Altıntaş, O.; Ergen, O. Next-Gen Traffic Surveillance: AI-Assisted Mobile Traffic Violation Detection System. *arXiv* **2023**, arXiv:2311.16179.
18. Feng, Z.; Guo, S.; Tan, X.; Xu, K.; Wang, M.; Ma, L. Rethinking Efficient Lane Detection via Curve Modeling. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 17062–17070.
19. M, R.D.S.; Manasa, A.; Deepti, K.; Kumar, Y.R. A Real-Time Multiple Traffic Violation Detection System Using DeepSORT. In *2024 4th International Conference on Artificial Intelligence and Signal Processing (AISP)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
20. Sun, X.; Zhang, Y.; Huang, X.; Wang, F.; Mo, Z. Vehicle Violation Detection System Based on Improved YOLOv5 Algorithm. In *2022 18th International Conference on Computational Intelligence and Security (CIS)*; IEEE: New York, NY, USA, 2022; pp. 148–152.
21. Zhu, R.; Hu, S.; Helal, S.; Song, J.; Wang, J.; Chen, Y. BTDS: Blockchain-Enabled Trusted Vehicle Violation Detection by Self-Supervision. *IEEE Internet Things J.* **2025**, *12*, 20156–20173. [[CrossRef](#)]
22. Badue, C.; Guidolini, R.; Carneiro, R.V.; Azevedo, P.; Cardoso, V.B.; Forechi, A.; Jesus, L.; Berriel, R.; Paixão, T.M.; Mutz, F.; et al. Self-Driving Cars: A Survey. *Expert Syst. With Appl.* **2021**, *165*, 113816. [[CrossRef](#)]
23. Chen, Z.; Liu, Y.; Gong, M.; Du, B.; Qian, G.; Smith-Miles, K. Generating Dynamic Kernels via Transformers for Lane Detection. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 6835–6844.
24. Zhou, K. Lane2Seq: Towards Unified Lane Detection via Sequence Generation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 16944–16953.
25. Honda, H.; Uchida, Y. CLRerNet: Improving Confidence of Lane Detection With LaneIoU. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New York, NY, USA, 2024; pp. 1176–1185.
26. Yang, Z.; Shen, C.; Shao, W.; Xing, T.; Hu, R.; Xu, P.; Chai, H.; Xue, R. LDTR: Transformer-Based Lane Detection with Anchor-Chain Representation. *Comput. Vis. Media* **2024**, *10*, 753–769. [[CrossRef](#)]
27. Xiao, L.; Li, X.; Yang, S.; Yang, W. ADNet: Lane Shape Prediction via Anchor Decomposition. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 6404–6413.
28. Yang, Z.; Yang, X.; Wu, L.; Hu, J.; Zou, B.; Zhang, Y.; Zhang, J. Pre-Inpainting Convolutional Skip Triple Attention Segmentation Network for AGV Lane Detection in Overexposure Environment. *Appl. Sci.* **2022**, *12*, 10675. [[CrossRef](#)]
29. Abualsaud, H.; Liu, S.; Lu, D.; Situ, K.; Rangesh, A.; Trivedi, M.M. LaneAF: Robust Multi-Lane Detection With Affinity Fields. *IEEE Robot. Autom. Lett.* **2021**, *6*, 7477–7484. [[CrossRef](#)]
30. 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 *Computer Vision—ECCV 2022*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 550–567.
31. Yang, Y.; Peng, H.; Li, C.; Zhang, W.; Yang, K. LaneFormer: Real-Time Lane Exaction and Detection via Transformer. *Appl. Sci.* **2022**, *12*, 9722. [[CrossRef](#)]
32. Na, H.; Kim, D.G.; Kang, J.M.; Lee, C. The Development of a Lane Identification and Assessment Framework for Maintenance Using AI Technology. *Appl. Sci.* **2025**, *15*, 7410. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.