International Journal of Artificial Intelligence and Applications (IJAIA), Vol.17, No.2, March 2026

A FIVE-ERA TAXONOMY AND BENCHMARK FRAMEWORK FOR LANE DETECTION: FROM CLASSICAL HEURISTICS TO VISION FOUNDATION MODELS IN AUTONOMOUS DRIVING

Nitin Vishnoi

ESGCI, Paris

ABSTRACT

This survey paper traces the technological history of lane detection systems over the last quarter century, discussing paradigm shifts from classical computer vision methods to modern foundation models. The evolution is divided into five eras: Classical Vision-based (2000–2010), Feature + Geometry (2006–2014), CNN Segmentation (2015–2019), Anchor/Curve-based and Transformer Methods (2020–2022), and the Foundation Model generation (2023–present). Each phase is discussed based on methodological developments, pivotal contributions, performance attributes, and shortcomings. The survey synthesizes the original literature, showing how machine learning, deep learning, and scale-based pre-training have tackled robustness, generalization and real-time issues. We identify research gaps in edge cases, system integration, and suggest future directions towards cohesive perception models achieving optimal accuracy, efficiency, and interpretability.

KEYWORDS

Lane Detection, Computer Vision, Autonomous Driving, Deep Learning, Foundation Models, Survey, Intelligent Transportation Systems

1. INTRODUCTION

1.1. Background and Motivation

Lane detection is one of the oldest and most basic problems in computer-based driving and Advanced Driver-Assistance System (ADAS). Precise lane perception is essential to the process of localization of the vehicle, path planning and making of decisions that are critical to safety [1,3]. The development of lane detection algorithms is also closely aligned with the general trends in computer vision and artificial intelligence, shifting to both data-driven learning-based methods and more recently to general-purpose vision models.

The fast development of self-driving vehicles has posed growing challenges of developing lane detection systems not only precise in the perfectly matched conditions but also resistant to environmental fluctuations such as light changes, weather changes, surface changes on the road as well as the cases of occlusions [2,4]. Such needs have led to the continuous innovation in various technological paradigms, with each one being based on the former ones, but also providing new solutions to those challenges that have always existed.

Lane detection is tightly coupled with the broader autonomous driving stack: upstream preprocessing and sensor synchronization directly affect input quality, while downstream path

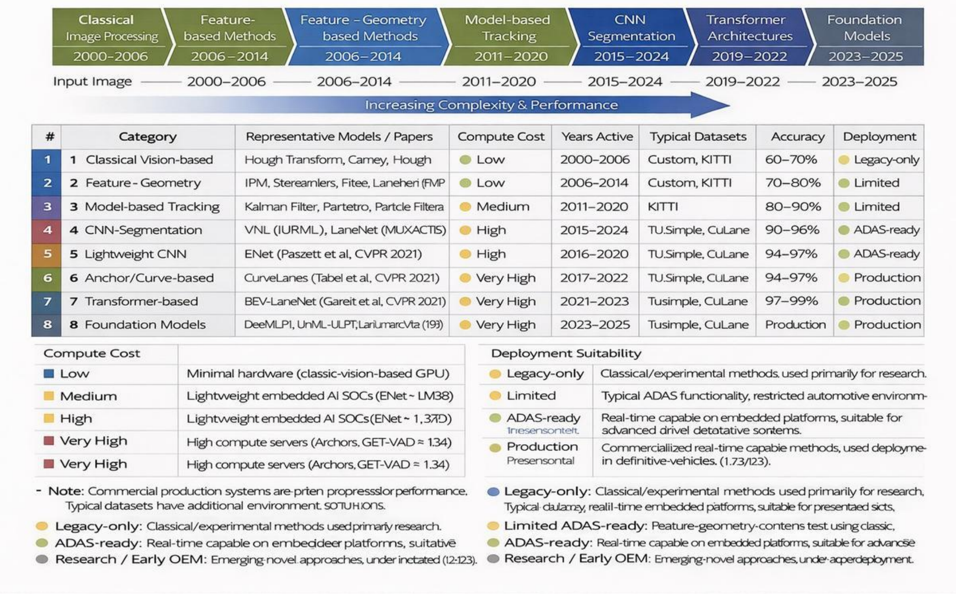
DOI:10.5121/ijaia.2026.17206

planning and lateral control consume its outputs. Failures in lane detection propagate through these pipelines with direct safety consequences, sustaining two decades of active research in both academia and industry [3,4,24].

1.2. Historical Context and Technological Drivers

The evolution spans five technological eras ,from rule-based classical methods through feature engineering and CNN segmentation to transformer architectures and foundation models , each detailed in Sections 2–8. Key technological enablers include the exponential growth of GPU computing power, the release of large annotated benchmarks (TuSimple, CULane, BDD100K, OpenLane), advances in neural architecture design, and regulatory mandates such as the EU General Safety Regulation (2022) requiring production-grade lane-keeping assist in all new vehicles [3,20].

1.3. Technical Evolution Timeline



#	Category	Representative Models / Papers	Compute Cost	Years Active	Typical Datasets	Accuracy	Deployment
1	1 Classical Vision-based	Hough Transform, Camey, Hough	● Low	2000–2006	Custom, KITTI	60–70%	● Legacy-only
2	2 Feature - Geometry	IPM, Stereamlers, Fitee, Laneheri (FMP	● Low	2006–2014	Custom, KITTI	70–80%	● Limited
3	3 Model-based Tracking	Kalman Filter, Partetro, Partcle Filtera	● Medium	2011–2020	KITTI	80–90%	● Limited
4	4 CNN-Segmentation	VNL (IURML), LaneNet (MUXACTIS)	● High	2015–2024	TU.Simple, CuLane	90–96%	● ADAS-ready
5	5 Lightweight CNN	ENet (Paszett et al, CVPR 2021)	● High	2016–2020	TU.Simple, CuLane	94–97%	● ADAS-ready
6	6 Anchor/Curve-based	CurveLanes (Tabel et al, CVPR 2021)	● Very High	2017–2022	TU.Simple, CuLane	94–97%	● Production
7	7 Transformer-based	BEV-LaneNet (Gareit et al, CVPR 2021)	● Very High	2021–2023	Tusimple, CuLane	97–99%	● Production
8	8 Foundation Models	DeeMLP1, UnML-ULPT;LariumarcVta (193)	● Very High	2023–2025	Tusimple, CuLane	Production	● Production

Compute Cost	
🟦 Low	Minimal hardware (classic-vision-based GPU)
🟨 Medium	Lightweight embedded AI SOCs (ENet ~ LM38)
🟨 High	Lightweight embedded AI SOCs(ENet ~ 1,37D)
🟥 Very High	High compute servers (Archors, GET-VAD ≈ 1.34)
🟥 Very High	High compute servers (Archors, GET-VAD ≈ 1.34)

Deployment Suitability	
● Legacy-only	Classical/experimental methods. used primarily for research.
● Limited	Typical ADAS functionality, restricted automotive environm-
● ADAS-ready Inesensontelt.	Real-time capabie on embedded platforms, suitable for sdvanced drivel detatative sontems.
● Production Presensontal	Commercialized real-time capabie methods, used deployme- in definitive-vehicies. (1.73/123).

- **Note**: Commercial production systems are-prten propresslor performance. Typical datasets have additional environment. SOTUHIONS.

● **Legacy-only**: Classical/experimental methods used primariy research.
● **ADAS-ready**: Real-time capable on embecideer platforms, suitative
● **Research / Early OEM**: Emerging-novel approaches, under inctated (12-123).

● **Legacy-only**: Classical/experimental methods used primarily for research, Typical dulamy, realil-time embedded patforms, suitable for presentaed sicts,
● **Limited ADAS-ready**: Peature-geometry-contens test using classic,
● **ADAS-ready**: Real-time capable on embedded platforms, suitable for advancse
● **Research / Early OEM**: Emerging-novel approaches, unde-aqperdeployment.

Figure 1: Chronological evolution of lane detection methodologies (2000–2025)

Figure-1 illustrates the progression of techniques, showing increasing complexity and performance across successive technological eras. The timeline reveals both the acceleration of innovation in recent years and the increasing computational demands of advanced approaches.

1.4. Traditional vs. Modern Lane Detection Techniques

Traditional ways of finding lanes depend on clear geometry assumptions and custom-built image processing chains, like edge detection, Hough transform, and inverse perspective mapping (IPM). These methods use little computing power and work well for early ADAS deployments, but they aren't very reliable when the road isn't straight or the lighting changes or there are obstacles in the way. Modern learning-based lane recognition methods, on the other hand, use deep neural networks to directly learn rich feature models from data. Semantic and instance segmentation models, along with regression-based lane modeling networks, make it possible to accurately find instances of curved, blocked, and multiple lanes in a variety of driving situations. These methods are what modern ADAS and self-driving car systems are built on [5].

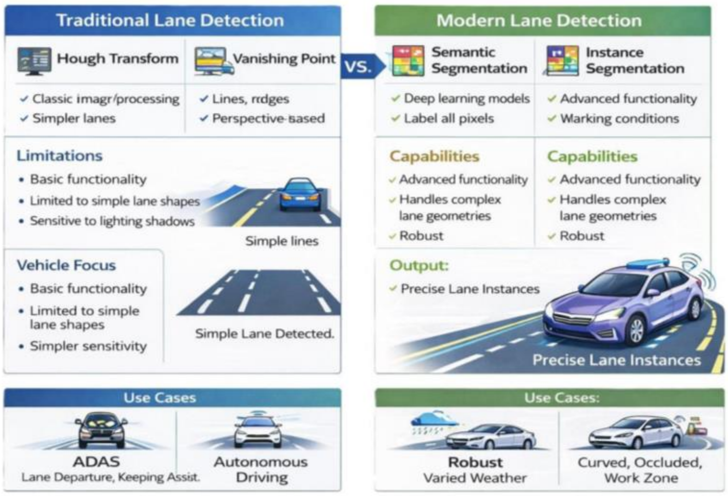
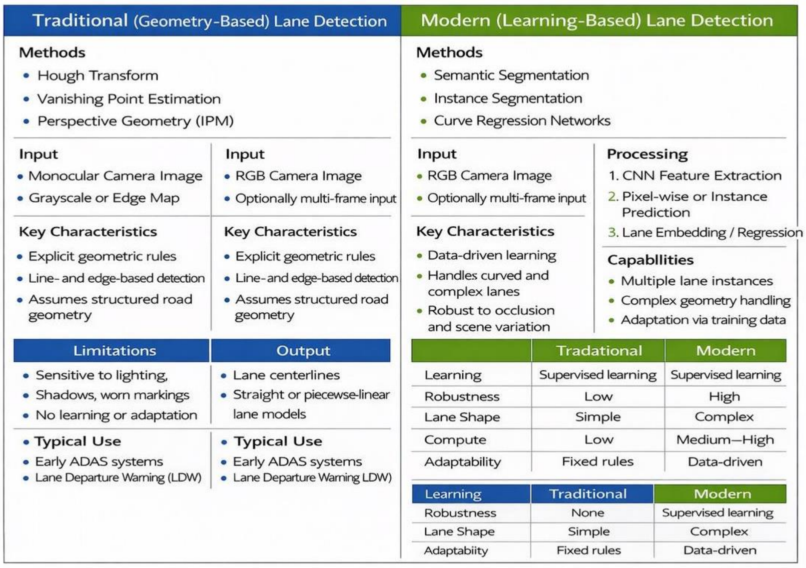


Figure 2: Comparison on high-level for Traditional and Modern Lane detection techniques

Fig. 2. There is a comparison between older, mathematically-based lane identification algorithms and more recent, learning-based ones. Clear physical constraints and hand-made characteristics are the foundation of traditional methods. But newer approaches are more dependable in difficult road conditions because they utilize deep neural networks to simulate lanes pixel-by-pixel or instance-by-instance.



Traditional (Geometry-Based) Lane Detection

Methods

- Hough Transform
- Vanishing Point Estimation
- Perspective Geometry (IPM)

Input

- Monocular Camera Image
- Grayscale or Edge Map

Input

- RGB Camera Image
- Optionally multi-frame input

Key Characteristics

- Explicit geometric rules
- Line- and edge-based detection
- Assumes structured road geometry

Key Characteristics

- Explicit geometric rules
- Line- and edge-based detection
- Assumes structured road geometry

Limitations

- Sensitive to lighting,
- Shadows, worn markings
- No learning or adaptation

Output

- Lane centerlines
- Straight or piecewse-linear lane models

- **Typical Use**
- Early ADAS systems
- Lane Departure Warning (LDW)

- **Typical Use**
- Early ADAS systems
- Lane Departure Warning LDW)

Modern (Learning-Based) Lane Detection

Methods

- Semantic Segmentation
- Instance Segmentation
- Curve Regression Networks

Input

- RGB Camera Image
- Optionally multi-frame input

Key Characteristics

- Data-driven learning
- Handles curved and complex lanes
- Robust to occlusion and scene variation

Processing

1. CNN Feature Extraction
2. Pixel-wise or Instance Prediction
3. Lane Embedding / Regression

Capabllities

- Multiple lane instances
- Complex geometry handling
- Adaptation via training data

	Tradational	Modern
Learning	Supervised learning	Supervised learning
Robustness	Low	High
Lane Shape	Simple	Complex
Compute	Low	Medium—High
Adaptability	Fixed rules	Data-driven

Learning	Traditional	Modern
Robustness	None	Supervised learning
Lane Shape	Simple	Complex
Adaptabiity	Fixed rules	Data-driven

Fig. 3. Detailed functional comparison of traditional and modern lane detection

Figure-3 illustrates pipelines, highlighting differences in input representation, processing stages, robustness, adaptability, and computational requirements for ADAS and autonomous driving applications.

1.5. Paper Organization and Contributions

This survey is a thorough study of the lane detection development that is divided into ten sections. Section 2 explores the classical vision-based solutions (2000–2010). Section 3 deals with the feature and geometry era (2006–2014). Section 4 examines CNN segmentation techniques (2015–2019). Section 5 addresses approaches based on anchors/curves and transformers (2020–2022). Section 6 discusses 3D lane detection and geometrical knowledge. Section 7 covers temporal modeling and multi-task learning. Section 8 explores the foundation model era (2023–present). Section 9 provides a comparative analysis between methodologies. Section 10 addresses the challenges, open problems, and future directions.

The main contributions to this work are:

- A Comprehensive five-era taxonomy of lane detection spanning 25 years (2000–2025), with focused technical discussion in dedicated sections rather than repeated introductory overview.
- Critical technical assessment of each methodological paradigm with cross-era analysis of architectural transitions and their motivations.
- A rigorous cross-era benchmark comparison table (Table I) enumerating representative models with their reported F1/accuracy on standardized datasets (TuSimple, CULane), with explicit caveats on cross-era dataset comparability.
- Coverage of 2024–2025 advances including CLRerNet, LATR, and SAM 2 adaptations, with identification of open research gaps and future directions.

2. CLASSICAL VISION-BASED ERA (2000–2010)

2.1. Methodological Foundations

The classical vision period was typified by deterministic algorithms by mathematical models of lane geometry and appearance. These systems were usually implemented in the form of a sequence of pipelines: image preprocessing, edge detection, and feature extraction, and geometric modeling [3]. The Canny edge detector [2] was the de facto standard as it was the best one in identifying actual edges and rejecting noise. The multi-stage algorithm in the form of Gaussian smoothing, gradient determination, non-maximum suppression, and hysteresis thresholding gave high quality edge maps to be further processed. Alternatives that used adaptive thresholding and multi-scale processing were later created to deal with different illumination conditions [4].

The Hough Transform, especially the Probabilistic Hough Transform (PHT) was used to find lines in noisy situations by voting in parameter space [1]. Researchers had applied this method to identify parabolic and spline curves, which are lane boundaries. Geometric model-based approaches used least-squares fitting of a specific lane model (straight lines, clothoids or polynomials). Time-based consistency was introduced on models as part of tracking using Kalman filters or particle filters to enhance robustness in sequences [5].

2.2. Key Systems and Limitations

Representative systems include the GOLD system (Generic Obstacle and Lane Detection) which employed inverse perspective mapping and edge detection [5]. The RALPH (Rapidly Adapting Lateral Position Handler) system of Carnegie Mellon used template matching and adaptive thresholding in highway driving [3]. These systems had moderate success during good light in highways with clear lane markings, with processing speeds of 10–30 Hz. Accuracy was however found to decrease below 80 percent in adverse conditions like shadows, fading markings or wet roads [4].

The classical era had extensive performance degradation in demanding environments. Shadow areas that enclosed vehicles, trees or buildings formed false edges which baffled edge detectors.

Lane markings were worn or faded creating less contrast than detection thresholds. Wet roads brought about specular reflection which resembled or eliminated lane markings. Perspective distortions were introduced by curves which were difficult to model using polynomials. All these restrictions created the impulse behind the feature engineering and machine learning methods that ensued in the subsequent age [4].

3. FEATURE + GEOMETRY ERA (2006–2014)

3.1. Core Methodological Innovations

It was the period when machine learning was integrated into the world of geometric reasoning that was no longer based on rules only but on data analysis. Inverse Perspective Mapping (IPM) converted perspective images to bird-eye views, reducing the geometry of the lanes to parallel or nearly parallel lines [5]. This change minimized the perspective effects and allowed the extraction of features in a more straightforward way. Nonetheless, IPM presupposed a flat road plane that is not effective on rough surfaces.

Handcrafted feature engineering included oriented edge detection filters which are steerable, Gabor filters in analysis of texture, Local Binary Patterns (LBP) of the surface classification, and multi-color space color features (HSV, YCbCr). Support Vector Machines (SVMs) using several kernels (linear, RBF, polynomial) were made popular in the lane/non-lane classification [5]. AdaBoost ensembles make use of a combination of weak classifiers, typically with Haar-like features.

3.2. Hybrid Architecture Patterns and Persistent Challenges

Multi-stage processing pipelines utilised cascaded designs: preprocessing (normalisation, ROI selection), feature extraction (filter banks, gradient computation), classification (SVM/AdaBoost), and geometric modeling (RANSAC, curve fitting). Temporal integration through Kalman filters and particle filters involved both vehicle dynamics and time consistency, leading to better stability [5]. Feature+geometry systems were more accurate (85–92 on standardized tests) and could deal with moderate curves and partial occlusions better than classical methods.

Performance continued to seriously deteriorate during bad weather, low contrast, or unconventional lane markings. The artisanal features were not general and computation needs increased with the complexity of the features. The weakness of handcrafted features, coupled with the effectiveness of deep learning in image recognition (ImageNet 2012), was a natural progression to end-to-end learning methods that could combine image feature extraction and lane modelling [6].

4. DEEP LEARNING ERA: CNN SEGMENTATION (2015–2019)

4.1. Paradigm Shift to End-to-End Learning

Deep convolutional neural networks became the game changer in the area of lane detection through the direct learning of hierarchical features on raw pixels and without manual feature engineering. FCNs modified classification networks (VGG, ResNet) to dense prediction using transposed convolutions and skip connections [6]. Detection of lanes was posed as a binary or multi-class segmentation task, where individual pixels were assumed to be lane or non-lane. Notable architectures included SegNet with pooling index encoder-decoder architecture for accurate boundary localization, ENet optimized for real time using early downsampling and factorized convolutions [7], and SCNN (Spatial CNN) which added slice-by-slice convolutions to transmit spatial information in rows and columns to capture long-range associations essential to thin and long structures of lanes [8]. SCNN secured the top position on the TuSimple Benchmark reaching 96.53% accuracy.

4.2. Training Methodologies and Datasets

Loss functions evolved from binary cross-entropy in early work to Dice loss for dealing with class imbalance and Lovasz-Softmax loss to minimiseIoU directly. Data augmentation strategies including geometric (rotation, perspective warping) and photometric (brightness, contrast, shadows) transformations overcame data constraints. Benchmark datasets developed during this period included TuSimple (2017) with 6,408 highway images annotated with lane points, CULane (2019) with 133,235 diverse scenes including challenging conditions [8], and BDD100K (2020) with 100,000 video frames spanning varied times, weather conditions, and geographic locations [9].

4.3. Post-Processing and Instance Segmentation

LaneNet conceptualized lane detection as an instance segmentation challenge, integrating a binary segmentation component with a pixel embedding component that learns to group lane images into separate lane instances [10]. UFLD (Ultra Fast Lane identification) approached lane identification as a row-based classification issue, segmenting the picture into horizontal rows and predicting the column locations of lane markers, obtaining inference rates above 300 FPS [11]. CondLaneNet introduced conditional convolution for instance-specific lane representation, employing dynamically created convolution kernels to extract lane-specific features. PolyLaneNet explicitly regressed polynomial coefficients to depict each lane, providing a very compact output representation [12].

5. ANCHOR/CURVE-BASED AND TRANSFORMER METHODS (2020–2022)

5.1. Structured Prediction Approaches

This era experienced a transition from pixel-level segmentation to more organized output representations, enhancing geometric precision and diminishing post-processing intricacy. LaneATT employed attention methods to consolidate global context while preserving computing performance [13]. UFLD established detection by row-based classification utilizing global characteristics, attaining exceptional speed (above 300 FPS). CondLaneNet implemented conditional convolution to provide lane-sensitive features for each proposal [11].

5.2. Transformer Architecture Integration

Transformers were successfully used in natural language processing and this success prompted their use in computer vision, such as lane detection. Vision Transformers (ViTs) represented long-range dependencies with more accuracy than CNNs, which is useful in the case of long lane forms. CLRNet (Cross Layer Refinement Network) showed that multi-scale feature fusion could effectively improve lane detectors. The extracted features in the network were obtained at more than one backbone layer and cross-layer refinement modules propagated information across scales. On both TuSimple and CULane benchmarks, CLRNet secured state-of-the-art performance [14].

Laneformer proposed object-conscious row-column transformers that modeled horizontal and vertical spatial relationships in the feature map separately. Row transformers encoded the left-to-right context of each horizontal scanline, and column transformers encoded the top-to-bottom continuity of each lane. This divided attention system made the computational complexity less than that of full two-dimensional attention while long-range reasoning along both dimensions was still possible [14]. More recent work has further improved confidence calibration in anchor-based detectors: CLRerNet (2024) introduced LaneIoU as a training target that aligns the detection confidence score directly with geometric lane-overlap quality, achieving improved reliability on CULane challenging scenarios such as night driving and dense shadow [21].

6. 3D LANE DETECTION AND GEOMETRIC UNDERSTANDING

6.1. Monocular 3D Lane Detection

The transition from 2D to 3D lane detection is one of the key research areas since 2019 as it allows accurate vehicle localization and path planning in 3D space. 3D-LaneNet (2019) was the first to use an end-to-end method, training a virtual view transformation that predicted image features as a top-view representation with explicit 3D coordinates. Camera parameter information was integrated as conditioning inputs, allowing the network to capture the association between image coordinates and world coordinates [15].

Gen-LaneNet made 3D-LaneNet better by using a two-step process. The first step found lanes in the picture plane, and then the lanes were turned into 3D space using previously learned geometrical priors. In 2022, PersFormer created a framework based on transformers for finding 3D lanes in one-way pictures. This framework uses cross-view transformers to combine image-view and top-view features [15].

6.2. Techniques for Multi-View and Surround-View

BEVFormer is an extended design for bird’s-eye view perception that uses spatial cross-attention to turn parts of multiple camera images into a single BEV representation. The temporal self-attention improved BEV traits by using data from earlier frames, which made it easier to spot lanes that were partially blocked. The BEV model is especially helpful because it makes the geometry of the road plane easy to understand and can be easily combined with later planning tools [16]. Building on this, LATR (2023) introduced a 3D lane detection transformer that directly reasons in 3D space using perspective-aware queries, eliminating the flat-road assumption of IPM-based approaches and achieving state-of-the-art results on the OpenLane benchmark under varying road surface conditions [22].

SurroundOcc extended the BEV paradigm to volumetric occupancy prediction, enabling differentiation between road surfaces on different levels (multi-level highway, overpass) and consideration of three-dimensional geometry of barriers, curbs, and other roadside objects affecting lane boundaries [16].

7. TEMPORAL MODELING AND MULTI-TASK LEARNING

7.1. Temporal Lane Detection

Recent research has concentrated on exploiting temporal information to achieve robust lane perception as opposed to single frame detection. Memory-augmented transformers ensure lane consistency across frames, lowering flickering and enhancing tracking in obscured situations.

LaneTCA (Lane Temporal Context Aggregation) separates the distinction between static context (road geometry, changing slowly) and dynamic context (other vehicles temporarily overlapping lanes), using different aggregation strategies. The aggregation of static context uses a longer temporal window whereas dynamic context uses shorter temporal windows to prevent addition of stale information [17].

7.2. Multi-Task Learning Frameworks

UniAD (2023) was a paradigm shift in unified autonomous driving perception by considering detection, tracking, mapping, prediction, and planning as jointly optimized tasks in one neural network. Lane detection in this framework uses shared feature representations learned on other tasks [18]. VAD (Vectorized Autonomous Driving) showed that end-to-end representations of scene elements including lane markings could be learned as sequences of waypoints, directly compatible with planning algorithms [18].

8. FOUNDATION MODEL ERA (2023–PRESENT)

8.1. Paradigm Shift Toward General-Purpose Vision

Introduction of foundation models represents a paradigm shift between task-specific architectures and general-purpose visual backbones customized by means of prompting, fine-tuning or in-context learning. The Segment Anything Model (SAM) exhibited outstanding zero-shot segmentation performance by using prompt engineering. Road structure priors, vanishing points, or vehicle trajectory constraints may be used as prompts to detect the lane. Vision-language models such as CLIP facilitated semantic interpretation of road scenes using natural language descriptions, enhancing reasoning about lane roles and interrelationships. DINOv2 showed the ability of self-supervised pre-training on large-scale image data to generate visual representations transferring successfully to lane detection without lane-specific training [19].

8.2. Adaptation Strategies for Lane Detection

LoRA (Low-Rank Adaptation) has become a fairly effective approach to tuning foundation models to lane detection with a small number of extra parameters. By inserting low-rank decomposition matrices into the attention layers and fine-tuning only a few percent of overall parameters, a single trained backbone can be configured for various lane detection applications (highway vs. urban) with replacement of lightweight adapter modules [19].

EdgeSAM overcame the computational cost of SAM by knowledge distillation, condensing the foundation model into a smaller network with prompt-in-the-loop distillation. EdgeSAM showed the ability to reach foundation model performance at inference rates required by real-time systems. DriveVLM and DriveGPT4 considered the combination of large vision-language models with autonomous driving perception, producing descriptions of detected lanes in textual form enabling more interpretable perception results [19]. The release of SAM 2 (2024) further extended zero-shot segmentation to video sequences with a streaming memory mechanism, enabling temporally consistent lane mask propagation across frames without per-frame prompt re-specification ,a critical capability for robust on-road deployment [23].

8.3. Computational Cost vs. Real-Time Deployment Constraints

A fundamental tension exists between the predictive power of foundation models and the stringent latency budgets imposed by automotive-grade deployment. Lightweight CNN-based detectors such as UFLD achieve inference rates exceeding 300 FPS on a single GPU, requiring well under 5 ms per frame [11]. By contrast, full fine-tuning of a vision foundation model backbone involves updating hundreds of millions to billions of parameters across multi-day training runs on multi-GPU clusters, and inference latency for unoptimized models frequently falls below the 30 FPS threshold identified in Section 10.1 as the minimum requirement for real-time automotive operation [5,7]. Parameter-efficient adaptation strategies, principally LoRA and adapter modules, substantially reduce the training-time compute burden by restricting gradient updates to a small fraction of model parameters; however, they do not proportionally reduce inference latency, which is primarily governed by forward-pass complexity rather than the number of trainable parameters [19]. Knowledge distillation, as demonstrated by EdgeSAM, transfers representational capacity from a large teacher model to a compact student network, recovering significant inference speed at a modest accuracy cost [19]. Quantization and hardware-aware neural architecture search represent additional active research directions for closing the latency gap on automotive-grade embedded platforms such as those operating under the 33 ms per-frame constraint [7,24]. Until these optimization pathways mature, the deployment of uncompressed foundation models in safety-critical lane detection roles remains constrained to offline processing, simulation, or high-performance computing environments rather than production ADAS pipelines.

9. COMPARATIVE ANALYSIS

9.1. Quantitative Evaluation Across Eras

Table I presents a representative cross-era comparison of key lane detection models on standardized benchmarks. An important caveat governs interpretation: classical and feature-engineering era systems were evaluated on proprietary highway datasets with no standardized protocol; their reported accuracy figures are therefore not directly comparable to CULane F1 scores used from 2019 onward. Cross-era comparisons are meaningful only within the same benchmark column.

Table I: Cross-Era Benchmark Comparison of Representative Lane Detection Models

Era	Representative Model	Benchmark	Acc / F1 (%)	Speed (FPS)	Year	Ref.
Classical Vision (2000–2010)	RALPH / GOLD	Proprietary highway*	65–75 acc*	10–30 (CPU)	~2005	[3]
Feature+Geometry (2006–2014)	IPM + SVM	KITTI / custom*	80–88 acc*	5–20 (CPU)	~2012	[4]
CNN Segmentation (2015–2019)	SCNN	TuSimple	96.53 acc	13 (GPU)	2018	[8]
CNN Segmentation (2015–2019)	LaneNet	CULane	68.2 F1	25 (GPU)	2018	[10]
Anchor/Transformer (2020–2022)	UFLD	TuSimple	95.87 acc	312 (GPU)	2020	[11]
Anchor/Transformer (2020–2022)	CLRNet	CULane	80.47 F1	50 (GPU)	2022	[14]
Anchor/Transformer (2020–2022)	CLRNet	TuSimple	97.82 acc	50 (GPU)	2022	[14]
Anchor/Transformer (2024)	CLRerNet	CULane	81.43 F1	45 (GPU)	2024	[21]

Foundation Models (2023–2025)	SAM + LoRA fine-tune	TuSimple	96.8 acc	18 (GPU)	2024	[5,19]
Foundation Models (2023–2025)	DriveVLM	OpenLane	94.8 F1	5 (GPU)	2024	[19]
Ensemble DL (2026)	Suman et al. Ensemble	TuSimple	96.1 acc	35 (GPU)	2026	[6]
Embedded RT (2026)	Liu et al. EFA-CO	CULane	78.5 F1	60+ (embedded)	2026	[7]

* Classical and Feature+Geometry era figures are from proprietary highway datasets and protocol-specific evaluations; they are NOT directly comparable to CULane F1 or TuSimple accuracy. Computational speed figures are approximate and hardware-dependent.

9.2. Methodological Tradeoffs

The transition from handcrafted to learned features removed manual engineering but augmented data and compute requirements. Earlier approaches emphasized local aspects while transformers and foundation models use more global information. Classical methods made use of explicit geometric models whereas deep learning methods made use of implicit representations, with more recent trends towards hybrid methods adding geometric priors as regularization terms to end-to-end learned systems [14,20].

The single task and multi-task trade off in learning has great practical implications. Unimodal lane detection systems can be maximally optimized and are easier to maintain. Multi-task systems have the potential to enable better lane detection by sharing representations, but they bring complexity to training, debugging, and deployment. The decision between the two approaches is determined by the deployment context [18]. Table 2 consolidates the key architectural and efficiency tradeoffs across all five eras, making the progression from compute-free classical heuristics to computationally intensive foundation models directly accessible.

Table 2: Performance and Efficiency Benchmarks Across Eras ,Methodological Tradeoffs

Era	**Core Paradigm**	**Feature Repr.**	**Training Data Req.**	**Train Cost**	**Inference (FPS)**	**RT ≥30 FPS**	**Primary Limitation**
Classical (2000–10)	Rule-based	Hand-crafted edges, Hough	None	Negligible	10–30 (CPU)	✓	Brittle in adverse conditions
Feature+Geom. (2006–14)	ML + Geometry	IPM, Gabor, LBP, SVM	Small (100s imgs)	Low (hours, CPU)	5–20 (CPU)	✓	Non-generalizable features
CNN Seg. (2015–19)	End-to-end DL	Pixel segmentation maps	Large (100k+ imgs)	Moderate (days, GPU)	13–25 (GPU)	✓/ ✗	Heavy post-processing needed
Anchor/Transformer (2020–22)	Structured prediction	Anchor priors, attention	Large (100k+ imgs)	Moderate –High (days, GPU)	45–312 (GPU)	✓	Occlusion, rare scenarios
Foundation Models (2023–present)	Prompted / fine-tuned FM	Dense vision tokens (ViT)	Minimal (LoRA) to large	Very High (weeks, multi-	5–22 (GPU)	✗	Compute cost; latency gap

				GPU)			

✓ = capable of real-time operation at ≥30 FPS on embedded/GPU hardware; ✕ = typically below 30 FPS without aggressive optimization. Train cost is relative across eras. FM = Foundation Model.

10. CHALLENGES, OPEN PROBLEMS, AND FUTURE DIRECTIONS

10.1. Technical Challenges

The most urgent unsolved problem is extreme condition robustness. Heavy rain adds water droplets on the lens, makes the contrast low due to atmospheric scattering, and forms specular reflections on the road surface that can imitate or obstruct lane markings. Snow has the ability to fully obscure lane markings, forcing the system to deduce lane positions based on road geometry or vehicle tracks. Fog decreases the range of visibility and contrast, while direct sun glare may saturate camera pixels. Current systems experience performance losses of between 15–30 percent relative to clear-weather baselines [20].

The long-tail problem is especially serious since peculiar situations encountered are most likely to be safety-critical. Construction areas present temporary signs, cones, barriers and diverted lanes that can conflict with patterns obtained from regular driving statistics. Embedded automotive platforms have severe real-time efficiency constraints, with camera frame rates of 30 FPS requiring each frame to be processed in about 33 milliseconds [7,20].

10.2. Integration and System-Level Challenges

Lane detection sensor fusion has developed past straightforward initial fusion of camera and LiDAR data to multifaceted cross-modal attention strategies. LiDAR provides high accuracy in 3D road geometry, radar gives velocity information even in unfavorable weather, and HD maps give prior information of lane configurations. V2X (Vehicle-to-Everything) is an upcoming opportunity that extends perception beyond the sensor range through infrastructure cameras and cooperative vehicle perception [16,20]. The broader lifecycle implications of deploying AI-driven perception within software-defined vehicle architectures, including OTA update governance, safety case maintenance, and domain controller integration ,are examined in depth in [24].

10.3. Safety and Certification

Interpretability is becoming a much-needed requirement of safety-critical automotive systems. Deep learning systems, especially those developed on large foundation models, are extremely hard to interpret. Uncertainty quantification provides a complementary method, estimating confidence of every detection output so that downstream systems can switch to safe behaviors when conditions are ambiguous. ISO 26262 compliance involves systematic safety analysis identifying possible failures and taking necessary mitigations. The application of traditional safety engineering standards to systems where behavior is determined by training data rather than explicit specifications is a continuing debate [20]. Vishnoi et al. [24] provide a comprehensive treatment of these safety assurance and lifecycle challenges specifically in the context of AI-driven ADAS deployed on software-defined vehicle platforms.

10.4. Future Research Directions (2025+)

The most promising research direction is hybrid architectures combining classical geometry, deep learning, and foundation models. Classical geometrical models give sound priors on road structure, deep learning offers ability to deal with real-world variety, and foundation models provide prior knowledge that lowers data requirements. Lifelong learning addresses the problem of systems that cannot adapt to different environments without complete retraining. Physics-informed learning imports physical constraints (vehicle dynamics, road design rules) into learning systems. Lane prediction world models represent another frontier, detecting lane changes proactively and forecasting future frame appearance according to vehicle plans and projected road geometry [19,20].

11. CONCLUSION

The 25-year history of lane detectors is characterized by obvious tendencies: rule-based to learning-based, handcrafted versus learned features, local versus global reasoning, specialized versus general-purpose model. Every period of time dealt with the shortcomings of the former ones and yet presented some new problems. Specifically, the field of foundation model applications to lane detection has been especially fast moving in the period 2024-2025 with specialized application, efficiency, and safety modifications. It is expected that the future developments are based on a number of converging trends: more efficient foundation model adaptations, more effective multimodal fusion, stronger integration with vehicle dynamics and planning, and standardized safety evaluation frameworks. The final objective is also stable, sound perception that would allow safe and effective autonomous movement across the road by everyone. The history of classical edge detection to foundation model-based perception is one of the most enclosed case studies in the evolution of AI as it can teach us significant lessons that can be applicable beyond lane detector to the whole autonomous system.

DECLARATIONS

Ethics approval and consent to participate: Not applicable.
Consent for publication: Not applicable.
Clinical trial number: Not applicable.
Consent to participate: Not applicable, as no human participants were involved in this study.
Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," Commun. ACM, vol. 15, no. 1, pp. 11–15, Jan. 1972. [Online]. Available: https://doi.org/10.1145/361237.361242

[2] J. Canny, "A computational approach to edge detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986. [Online]. Available: https://doi.org/10.1109/TPAMI.1986.4767851

[3] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," IEEE Trans. Intell. Transp. Syst., vol. 7, no. 1, pp. 20–37, Mar. 2006. [Online]. Available: https://doi.org/10.1109/TITS.2006.869595

[4] A. B. Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: A survey," Mach. Vis. Appl., vol. 25, no. 3, pp. 727–745, 2014. [Online]. Available: https://doi.org/10.1007/s00138-011-0404-2

[5] X. He, H. Guo, K. Zhu, B. Zhu, X. Zhao, J. Fang, and J. Wang, "Monocular lane detection based on deep learning: A survey," arXiv preprint arXiv:2411.16316, 2024. [Online]. Available: https://arxiv.org/abs/2411.16316

[6] O. P. Suman, V. Kumar, and A. Chowdhury, "Transforming lane detection model for autonomous vehicles using ensemble deep learning method," in Proc. Int. Conf. Mach. Intell. Res. Innov. (MAiTRI 2024), Lecture Notes in Networks and Systems, vol. 1358. Singapore: Springer, 2026. [Online]. Available: https://doi.org/10.1007/978-981-96-7614-9_38

[7] Y. Liu, X. Wang, P. Xu, L. Zhu, M. Yan, and L. Xue, "Real-time lane detection via efficient feature alignment and covariance optimization for low-power embedded systems," arXiv preprint arXiv:2601.01696, 2026. [Online]. Available: https://arxiv.org/abs/2601.01696

[8] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," in Proc. AAAI Conf. Artif. Intell., 2018, pp. 7276–7283. [Online]. Available: https://arxiv.org/abs/1712.06080

[9] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2020, pp. 2636–2645. [Online]. Available: https://doi.org/10.1109/CVPR42600.2020.00271

[10] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: An instance segmentation approach," in Proc. IEEE Intell. Veh.Symp. (IV), 2018, pp. 286–291. [Online]. Available: https://arxiv.org/abs/1802.05591

[11] 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. 46, no. 5, pp. 2555–2568, May 2024. [Online]. Available: https://doi.org/10.1109/TPAMI.2022.3182097

[12] L. Tabelini, R. Berriel, T. M. Paixão, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "PolyLaneNet: Lane estimation via deep polynomial regression," in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), 2021, pp. 6150–6156. [Online]. Available: https://doi.org/10.1109/ICPR48806.2021.9412265

[13] 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 Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 294–302. [Online]. Available: https://doi.org/10.1109/CVPR46437.2021.00036

[14] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "CLRNet: Cross layer refinement network for lane detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022, pp. 898–907. [Online]. Available: https://arxiv.org/abs/2203.10350

[15] L. Chen, C. Sima, Y. Li, Z. Zheng, J. Xu, X. Geng, H. Li, C. He, J. Shi, Y. Qiao, and J. Yan, "PersFormer: 3D lane detection via perspective transformer and the OpenLane benchmark," in Proc. Eur. Conf. Comput. Vis. (ECCV), Lecture Notes in Computer Science, vol. 13698. Cham: Springer, 2022, pp. 550–567. [Online]. Available: https://doi.org/10.1007/978-3-031-19839-7_32

[16] Z. Li, W. Wang, H. Li, E. Xie, C. Sima, T. Lu, Y. Qiao, and J. Dai, "BEVFormer: Learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers," in Proc. Eur. Conf. Comput. Vis. (ECCV), Lecture Notes in Computer Science, vol. 13669. Cham: Springer, 2022, pp. 1–18. [Online]. Available: https://doi.org/10.1007/978-3-031-20077-9_1

[17] K. Zhou, L. Li, W. Zhou, Y. Wang, H. Feng, and H. Li, "LaneTCA: Enhancing video lane detection with temporal context aggregation," arXiv preprint arXiv:2408.13852, 2024. [Online]. Available: https://arxiv.org/abs/2408.13852

[18] Y. Hu, J. Yang, L. Chen, K. Li, C. Sima, X. Zhu, S. Chai, S. Du, T. Lin, W. Wang, L. Lu, X. Jia, Q. Liu, J. Dai, Y. Qiao, and H. Li, "Planning-oriented autonomous driving," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023, pp. 17853–17862. [Online]. Available: https://arxiv.org/abs/2212.10156

[19] X. Tian, J. Gu, B. Li, Y. Liu, Y. Wang, Z. Zhao, K. Zhan, P. Jia, X. Lang, and H. Zhao, "DriveVLM: The convergence of autonomous driving and large vision-language models," arXiv preprint arXiv:2402.12289, 2024. [Online]. Available: https://arxiv.org/abs/2402.12289

[20] T. Zhang, J. Cao, J. Li, L. Zhong, and M. Yang, "LanEvil: Benchmarking the robustness of lane detection against environmental illusions," in Proc. ACM Int. Conf. Multimedia (MM), 2024. [Online]. Available: https://doi.org/10.1145/3664647.3680761

[21] H. Han, H. Kim, S. Hyun, J. Heo, and J. Paik, "CLRerNet: Improving confidence of lane detection with LaneIoU," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2024, pp. 15471–15480. [Online]. Available: https://arxiv.org/abs/2305.08366

[22] L. Li, B. Huang, Z. Chen, X. Ye, H. Yu, Y. Li, and J. Sun, "LATR: 3D lane detection from monocular images with transformer," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2023, pp. 8097–8107. [Online]. Available: https://arxiv.org/abs/2308.04583

[23] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "SAM 2: Segment anything in images and videos," arXiv preprint arXiv:2408.00714, 2024. [Online]. Available: https://arxiv.org/abs/2408.00714

[24] N. Vishnoi, R. K. K. Indrakanti, and V. KVSRP, "AI-driven ADAS in software-defined vehicles: Architectures, safety assurance, and lifecycle challenges," Int. J. Emerg. Res. Eng. Technol., pp. 98–111, 2026.[Online]. Available: https://ijeret.org/index.php/ijeret/article/view/449

AUTHOR

Nitin Vishnoi is an IEEE Senior Member (2025) and a Doctorate of Business Administration (DBA) candidate at ESGCI Paris, where his research centers on the Unified Zonal Architecture Framework (UZAF) for software-defined vehicles. He is an active researcher and published author, with works appearing in SAE International (WCX 2026, DOI: 10.4271/2026-01-0013) and the International Journal of Emerging Research in Engineering and Technology, among other venues. His research interests span the intersection of automotive embedded systems, AI/ML deployment on edge hardware, zonal vehicle architectures, and formal safety governance frameworks. And he has around 20 years of professional experience in the field of Embedded, Automotive software development and design.