A Comprehensive Review of Techniques and Datasets for Lane Detection in Autonomous Systems

1.0 Introduction: The Evolving Landscape of Lane Detection

1.1 The Critical Role of Lane Detection

Lane detection is a foundational perception task for autonomous vehicles and advanced driver-assistance systems (ADAS).[1] The accurate localization and shape estimation of lane markings are crucial for a variety of downstream applications, including real-time trajectory planning, vehicle localization, and the construction of high-definition (HD) maps.[1] The task, however, is fraught with significant real-world challenges that extend far beyond simple object detection. The thin, long, and continuous nature of lane markings, combined with the dynamic complexities of the driving environment, requires a robust and highly specialized approach.

1.2 Core Challenges in Real-World Scenarios

The inherent difficulties of lane detection stem from both visual and geometric complexities. Visually, lane markings are often obscured by other traffic participants, such as vehicles and pedestrians, leading to severe occlusion.[1] Environmental factors present additional hurdles, including variations in lighting (e.g., night scenes, shadows, and the glare of dazzle light) and adverse weather conditions (e.g., rain and fog).[1] Compounding these issues is the geometric variability of lanes. Models must accurately perceive a wide range of lane shapes, from simple straight lines to complex curves, dense multi-lane configurations, and ambiguous topologies like merged or forked lanes.[1] These challenges are further exacerbated when lane markings are worn or completely absent, forcing a model to infer their position from contextual information.[1] The need to handle such scenarios has driven the field to develop increasingly sophisticated techniques.

1.3 From Heuristics to Deep Learning

Early lane detection methods relied on hand-crafted features and traditional computer vision algorithms, which were often brittle in complex, real-world conditions.[1] The advent of deep learning has revolutionized the field by enabling models to learn rich, high-level feature representations directly from data.[1] This transition has progressed through several distinct paradigms. Initially, methods treated the problem as a pixel-level segmentation task, attempting to classify each pixel as either a lane or a background element. This was followed by a shift towards more holistic, end-to-end approaches that directly regressed a parametric representation of the lane shape. The most recent innovations have incorporated transformer architectures, which excel at modeling long-range dependencies, and meta-learning frameworks that automate the design of optimal, task-specific networks. The following sections provide an in-depth analysis of these major paradigms, their key techniques, and the benchmark datasets that have shaped their development.

2.0 Foundational Methods: The Pixel-Level Paradigm

2.1 Spatial CNN (SCNN): A Pioneer in Structural Prior Modeling

The Spatial CNN (SCNN) introduced a foundational concept for lane detection by explicitly modeling the long, continuous shape of lane markings. Traditional convolutional neural networks (CNNs) process information hierarchically, building features layer-by-layer.[1] While effective for general object recognition, this approach struggles with objects that have a strong global structure but weak local appearance, such as occluded or faded lanes.[1] The SCNN addresses this by generalizing traditional layer-by-layer convolutions to a novel "slice-by-slice" convolution within a single feature map.[1]

This mechanism allows for a form of sequential message passing between pixels, both across rows and columns of a feature map.[1] An SCNN is composed of four directional convolution modules: downward, upward, rightward, and leftward. In this scheme, a slice of the feature map (e.g., a row) is updated only after receiving information from a previous slice, creating a cascade of information flow. This is a significant departure from standard CNNs, which lack this explicit spatial communication.[1]

The SCNN's approach offers several advantages over its contemporaries. It is substantially more computationally efficient than methods based on dense Conditional Random Fields (CRFs) or Markov Random Fields (MRFs), which require all-to-all message passing between pixels and are prohibitively expensive for real-time applications.[1] Furthermore, the SCNN

propagates messages as residuals, which makes the model more stable and easier to train compared to recurrent neural network (RNN)-based methods that can suffer from gradient issues over long sequences.[1] SCNN achieved state-of-the-art performance at the time on datasets like CULane and TuSimple, demonstrating the effectiveness of building architectural awareness of structural priors directly into the network.[1] The model's success established a new direction for research, moving beyond purely data-driven approaches to an approach that deliberately designs network architectures to understand the underlying geometry of the problem.

2.2 Insights and Nuances from the Pixel-Level Paradigm

The development of the SCNN fundamentally re-oriented the field's architectural philosophy. Prior to SCNN, many approaches relied on standard, off-the-shelf CNN backbones without special modifications for the lane detection task. The SCNN, conversely, introduced a bespoke architectural element—the slice-by-slice convolution—that was specifically engineered to capture the geometric properties of lanes.[1] This transition from a general-purpose, data-driven approach to one that is also guided by an architectural understanding of the problem's domain was a critical step. It demonstrated that superior performance on a specialized task could be achieved by embedding domain knowledge directly into the network's design.

The SCNN's efficiency also highlights a perpetual trade-off in deep learning. While its sequential, directional message passing scheme is significantly more efficient than the dense, all-to-all communication of CRFs, it is also more constrained.[1] This constrained approach, while effective, paved the way for later transformer-based models that could model global context in a more flexible manner, albeit often with a higher computational cost. The SCNN's design thus exemplifies the continuous balancing act between computational efficiency and architectural generality in the pursuit of more robust and accurate solutions.

3.0 Modern Approaches: The Era of Transformers and Novel Representations

3.1 BézierFormer: A Unified End-to-End Architecture for 2D/3D Detection

BézierFormer represents a significant advancement by unifying the tasks of 2D and 3D lane detection within a single, end-to-end framework.[1] This is achieved through a core architectural principle: the representation of lane curves using Bézier curves. Unlike

pixel-based segmentation or rigid parametric models, Bézier curves can uniformly and efficiently represent both 2D and 3D curves using just a few control points, regardless of their orientation.[1]

The architecture is built upon a feature extractor and a Bézier curve decoder, inspired by detection transformers.[1] The decoder iteratively refines lane predictions, which are formulated as dynamic queries corresponding to Bézier control points. The central innovation is the **Bézier curve attention mechanism**, a multi-reference-points variant of deformable attention.[1] Instead of using a single reference point per object, this mechanism samples multiple sparse reference points along the curve based on the predicted control points.[1] This ensures that the network's receptive field comprehensively covers the entire slender lane, a crucial capability for accurately capturing features from a geometrically challenging object.[1] The extracted features are then fused and used to refine the curve embeddings and control points from the previous decoder layer.

For 3D lane detection, the model incorporates an optional perspective projection operation that uses camera parameters to project 3D reference points onto 2D image coordinates, allowing the same unified framework to be applied.[1] To optimize the model, a novel **Chamfer IoU-based loss** is proposed.[1] This loss function is more effective than simple point-based regression because it optimizes for the overall shape and location similarity between the predicted and ground truth curves, viewing each curve as a cohesive entity.[1] BézierFormer's state-of-the-art performance on datasets like CurveLanes and OpenLane validates the superiority of its architectural design and the use of Bézier curve representation.[1]

3.2 PriorLane: The Fusion of Vision Transformers and Prior Knowledge

PriorLane introduces a novel framework that enhances lane segmentation performance by fusing a vision transformer (MiT) with external, low-cost local prior knowledge.[1] This approach acknowledges that visual information alone may be insufficient for robust detection in challenging scenarios, such as when lanes are occluded or faded.[1]

The framework operates by representing prior knowledge, such as road networks or elevation data, as a grid map in a Bird's Eye View (BEV).[1] A key component is the **Knowledge Embedding Alignment (KEA)** module, which is designed to address the challenge of aligning this BEV data with the camera view, particularly when vehicle positions are only coarsely known.[1] The KEA module uses a Spatial Transformer Network (STN) to manipulate the prior knowledge embeddings and align them with the image features. This process is critical for ensuring the data from two different modalities is spatially consistent for effective fusion.[1] A **Fusion Transformer (FT)** then uses a self-attention mechanism to merge the aligned prior knowledge embeddings with the visual features extracted by the MiT backbone.[1] This creates a comprehensive feature representation that benefits from both image context and geometric

prior knowledge.

PriorLane's success, particularly on the Zjlab dataset, demonstrates that a significant performance gain can be achieved by moving beyond the paradigm of relying solely on visual input.[1] The framework suggests that future advancements in lane detection will likely involve multi-modal fusion, leveraging readily available data sources to improve model robustness.

3.3 CurveLane-NAS: Meta-Learning for Optimal Architecture

The CurveLane-NAS framework employs a meta-learning approach, Neural Architecture Search (NAS), to automatically design a lane-sensitive network.[1] This approach moves beyond hand-crafted model design by using a search algorithm to find an optimal architecture, particularly for difficult curve lane detection.[1] The framework's core principle is a holistic optimization that considers the entire pipeline, from feature extraction to post-processing. The framework consists of three integrated search modules:

1. **Elastic Backbone Search:** This module explores different configurations of network width, depth, and down-sampling locations to find a backbone that balances strong semantic feature extraction with low computational latency.[1]
2. **Feature Fusion Search:** This module automatically discovers the best way to combine multi-level features from the backbone. This is crucial for capturing both long-range coherent information (from high-level features) and accurate short-range details (from low-level features).[1]
3. **Adaptive Point Blending:** This is a novel and critical innovation. Instead of relying on a fixed, hand-designed post-processing step, this module searches for an optimal strategy to refine predictions from different feature levels. It uses adaptive score masking and a point blending technique to combine accurate local points with a consistent, long-range prediction.[1] This co-design of the model and its post-processing is unique and highly effective, as it avoids the pitfall of a great model being undermined by a suboptimal final step.[1]

CurveLane-NAS achieves state-of-the-art performance on the CULane and CurveLanes datasets, confirming that an automated, holistic design process can surpass architectures designed by human experts.[1]

3.4 Insights and Nuances from the Modern Paradigm

The modern era of lane detection models reflects several fundamental shifts in design philosophy. One is the move from a dense, pixel-level output to a sparse, geometric representation. SCNN's output is a segmentation mask that requires a separate, often heuristic, post-processing step to convert it into a usable geometric line. In contrast, BézierFormer and CurveLane-NAS directly regress a sparse set of parameters (e.g., Bézier control points or line proposals) that define a geometric curve.[1] This fundamental re-framing

of the problem space provides an output that is immediately consumable by a vehicle's planning and control systems, representing a more end-to-end approach.

Another key development is the understanding that a robust perception system may require more than just visual data. PriorLane's success with its Knowledge Embedding Alignment (KEA) module demonstrates the power of multi-modal fusion.[1] The model's recognition that coarse external data requires a specialized alignment mechanism to be useful is a crucial technical point, as it shows that modality fusion is not a trivial task. This suggests that the future of autonomous perception systems lies in integrating various data sources, such as visual, Lidar, and map data.

Finally, CurveLane-NAS introduced the concept of co-designing the network architecture and the post-processing pipeline.[1] By viewing the entire system as a single, optimizable entity, it moves beyond the traditional approach where a model's output is treated as a final product to be refined by a separate, often non-differentiable, algorithm. This paradigm highlights that a model's true effectiveness is not just in its predictions but in its ability to produce a final output that is both accurate and robust.

4.0 A Comparative Analysis of Lane Detection Techniques

4.1 Comparative Table

The evolution of lane detection techniques can be understood by comparing their core architectural paradigms, lane representations, and key innovations. The table below provides a concise overview of the models discussed in this report, illustrating their respective strengths and the design trade-offs they represent.

Table 1: Key Lane Detection Techniques

Technique	Architectural Paradigm	Lane Representation	Key Innovation	Strengths	Limitations
SCNN	CNN-based	Pixel-level Segmentation	Slice-by-slice convolution for spatial message passing	Efficient; effective at capturing structural priors for long objects	Requires post-processing to get geometric lines; limited to 2D output
BézierFormer	Transformer-b	Bézier Curve	Bézier curve	Unified 2D/3D	Requires

	ased	(Parametric)	attention for comprehensive feature extraction; Chamfer IoU-based loss for shape-fitting	framework; end-to-end regression of geometric shapes	specialized attention mechanism and loss function
PriorLane	Hybrid (CNN + Transformer)	Pixel-level Segmentation	Knowledge Embedding Alignment (KEA) module for fusing BEV prior knowledge	Enhanced robustness to occlusion and difficult scenes through data fusion	Requires external prior knowledge; still pixel-based, necessitating post-processing
CurveLane-NAS	Meta-Learning (NAS)	Point-based (Parametric)	Automated search for optimal backbone, feature fusion, and post-processing (Adaptive Point Blending)	Automatically finds an efficient, task-specific architecture; co-designs network and post-processing	High computational cost for the search phase

4.2 Paradigms: From CNN to Transformers

The architectural shift from CNNs to transformers reflects a change in how a model perceives the scene. CNNs, as exemplified by the SCNN, are inherently good at capturing local, hierarchical features. While SCNN's slice-by-slice convolution extends this to better handle long structures, its message passing is still constrained compared to a transformer's global attention mechanism.[1] Transformers, as used in BézierFormer, possess an inherent ability to model global context and long-range dependencies, making them particularly well-suited for understanding the relationship between distant parts of a lane or inferring the location of an occluded segment from a visible one.[1] This global perspective often yields superior performance in visually challenging, cluttered environments.

4.3 Lane Representations: A Critical Design Choice

The choice of lane representation is a defining characteristic of a model. The segmentation-based approach, used by SCNN and PriorLane, provides a dense, pixel-by-pixel map of the lanes.[1] While intuitive, this method is computationally intensive and requires an additional post-processing step to convert the segmentation mask into a usable geometric shape, such as a spline or a set of points. The parametric approach, adopted by BézierFormer and CurveLane-NAS, directly regresses a sparse set of parameters that define the lane geometry.[1] This is a more efficient and direct approach, as the output is immediately a geometric primitive that can be fed into downstream systems for planning and control. The success of these parametric models suggests that this is a more effective and pragmatic solution for the lane detection task.

4.4 External Data and Modality Fusion

The effectiveness of PriorLane's KEA module highlights a crucial point: that the most robust solutions will likely be multi-modal.[1] A model relying solely on camera imagery is at a disadvantage when dealing with severe occlusion or low visibility. By fusing visual data with a low-cost, non-visual data source like a BEV map, PriorLane demonstrates that a system can become more resilient to these issues.[1] The existence and complexity of the KEA module also illustrate that fusing data from different modalities is a non-trivial challenge that requires a dedicated architectural component to ensure proper spatial and semantic alignment.[1]

5.0 Benchmark Datasets: A Landscape Analysis

5.1 Comparative Table

The evolution of lane detection research has been shaped by the release of increasingly challenging and diverse benchmark datasets. Each new dataset has addressed the limitations of its predecessors, pushing the field to develop more robust and generalized models.

Table 2: Major Lane Detection Datasets

Dataset	Year	Images/Clips	Resolution	Scenes	Key Challenge
TuSimple	2017	6,408 images	1280×720	Highway (US)	Simple, daytime, clear conditions
CULane	2018	133,235 images	1640×590	Urban,	Scale, diverse

				Highway (Beijing)	conditions, occlusion, and difficult scenarios
LLAMAS	2019	100k images	1276×717	Highway (Germany)	Automated Lidar-based annotations, high-quality, real-world data
CurveLanes	2021	150k images	1640×590	Challenging curved roads	High proportion of curve lanes (>90%), complex topologies, S-curves
VIL-100	2022	12k video clips	1920×1080	Urban (China)	Temporal consistency, video-based detection
OpenLane	N/A	200k images	Varies	Diverse	3D annotations, scale

5.2 Analysis of Dataset Evolution and Its Impact

The history of lane detection datasets demonstrates a clear progression from simplified environments to complex, real-world scenarios. The **TuSimple** dataset served as a crucial starting point, providing a benchmark with clear, straightforward highway driving conditions.[1] However, its limited scope quickly made it insufficient for developing truly robust models. The subsequent release of the
CULane dataset was a major leap forward, offering a significantly larger scale and a much broader range of challenging scenes, including crowded roads, night driving, and varied lighting and weather conditions.[1] This dataset was instrumental in proving the effectiveness of models like the SCNN, which were designed to handle these complexities.
The **LLAMAS** dataset introduced the concept of automated, high-quality annotation using Lidar maps, providing a valuable resource for evaluating models with a different form of ground truth.[1] Recognizing that curve lanes are a particularly difficult and crucial problem for autonomous navigation, the

CurveLanes dataset was specifically created to address the lack of these challenging topologies in existing benchmarks. With over 90% of its images containing at least one curve, it became a powerful tool for evaluating models like CurveLane-NAS that were designed to handle such complexities.[1]

The latest generation of datasets, **VIL-100** and **OpenLane**, push the boundaries further by adding new dimensions to the problem. VIL-100, a video-based dataset, shifts the focus from frame-by-frame detection to temporal consistency across a sequence of frames, reflecting the reality of a moving vehicle's camera.[1] Meanwhile, OpenLane introduces 3D annotations, acknowledging that for true autonomous navigation, a 2D projection is insufficient; understanding the lane geometry in 3D world coordinates is essential.[1]

This evolution of datasets is not a passive process of collection but an active force that shapes the direction of research. Each new benchmark identifies a specific, unsolved problem and forces the community to develop new architectures and techniques to address it. This feedback loop, where new datasets necessitate new models and those models in turn reveal new problems to be addressed by even more challenging datasets, is a defining characteristic of progress in this field.

6.0 Findings, Tips, and Future Directions

6.1 Performance Findings on Key Benchmarks

Empirical data from the research demonstrates the clear performance gains of modern, transformer-based approaches over earlier CNN-based methods. The following table consolidates key performance metrics from the provided research, highlighting the state-of-the-art results for various models on popular benchmarks.

Table 3: Performance Findings on Key Benchmarks

Model	Backbone	Dataset	F1 Score (%)	FPS	Note
SCNN	VGG16	CULane	71.6	8	SOTA at the time; based on official code
CLRNet	ResNet101	CULane	86.48	74	A modern, high-performance method
BézierFormer	Swin-Tiny	CurveLanes	91.31	84	SOTA for 2D curve detection

BézierFormer	Swin-Tiny	OpenLane	56.43	90	SOTA for 3D lane detection
PriorLane	MiT-B5	Zjlab	73.78 (mIoU)	N/A	SOTA on the Zjlab dataset
CurveLane-L	Searched	CULane	74.8	86.5	New SOTA at the time; a result of NAS

6.2 Key Takeaways and Practical Tips

The analysis of modern lane detection techniques and datasets provides several critical takeaways for achieving superior performance and developing next-generation models.
- **Embrace Parametric Representations:** The success of BézierFormer and CurveLane-NAS on difficult datasets provides a strong argument that directly regressing geometric primitives is more effective than the two-step process of pixel segmentation followed by post-processing.[1] This paradigm shift simplifies the problem and produces a more useful output for autonomous systems.
- **Design for the Problem:** The SCNN and BézierFormer demonstrate the value of a domain-specific architectural design. By engineering a network to inherently understand the geometry of a lane, these models achieve superior results over generic, off-the-shelf architectures.[1]
- **The Loss Function is a Critical Component:** The Chamfer IoU-based loss function is a key element of BézierFormer's success.[1] It teaches a valuable lesson that optimizing for a holistic metric, such as overall shape similarity, is more effective for complex curves than relying on simple point-based distance metrics.
- **Fuse External Data for Robustness:** PriorLane's results highlight the potential of multi-modal fusion.[1] A model can become significantly more robust to visual occlusions and noise by intelligently fusing visual data with external, non-visual information, such as BEV maps. The challenge, however, lies in properly aligning the data from different modalities.
- **Optimize the Entire Pipeline:** The CurveLane-NAS framework's most profound contribution is its recognition that the post-processing step should not be a separate, heuristic phase.[1] By including the adaptive point blending module in its search space, the framework demonstrates that a holistic approach to design can yield significant performance gains by optimizing the entire system from input to final output.

6.3 Future Directions

Based on the current state of the art, several promising avenues for future research are apparent. The temporal and 3D challenges posed by VIL-100 and OpenLane are often addressed separately. A unified architecture that can handle both temporal consistency and 3D geometric reasoning simultaneously would be a major breakthrough. Additionally, while significant progress has been made in handling challenging visual conditions, performance in edge cases like extreme weather and low-visibility scenes remains a frontier. Finally, the ultimate goal is not just to detect lanes but to predict their future trajectory. This will require models that can understand and anticipate the temporal dynamics of the road, moving from a static perception task to a dynamic prediction task.

7.0 Conclusion

The journey of lane detection, as evidenced by the progression from simple CNNs to sophisticated transformer-based frameworks and meta-learning, reflects a field that is maturing rapidly. What began as a pixel-level segmentation problem has evolved into a quest for robust, end-to-end solutions that can directly produce geometric primitives usable by autonomous systems. The evolution of benchmark datasets has been a primary catalyst for this progress, with each new dataset raising the bar and forcing new architectural innovations. The most promising trends involve designing models that are inherently aware of the problem's structure, fusing information from multiple data modalities, and embracing automated design processes that optimize the entire system pipeline. The future of lane detection will likely be defined by models that can unify temporal and 3D perception, demonstrating true autonomy and resilience in even the most complex, real-world conditions.

Works cited

1. curvelane.pdf
2. Unsupervised LLAMAS - FMI Data Index, accessed September 14, 2025, https://fmi-data-index.github.io/llamas.html
3. Unsupervised Llamas Lane Marker Dataset, accessed September 14, 2025, https://unsupervised-llamas.com/llamas/
4. VIL-100: A New Dataset and a Baseline Model for Video Instance Lane Detection, accessed September 14, 2025, https://openaccess.thecvf.com/content/ICCV2021/papers/Zhang_VIL-100_A_New_Dataset_and_a_Baseline_Model_for_Video_ICCV_2021_paper.pdf