Adaptive ROI Optimization Pyramid Network: Lane Detection for FSD under Data Uncertainty

Xu Cao, Weisheng Liu, and Zhijian Wang

Abstract—**To mitigate the challenges posed by data uncertainty in Full-Self Driving (FSD) systems. This paper proposes a novel feature extraction learning model called Adaptive Region of Interest Optimized Pyramid Network (ARO). Specifically, ARO introduces a novel cross-layer fusion attention mechanism that dynamically assigns weighted attention to feature maps across different levels, facilitating deep feature fusion and enabling the effective extraction and utilization of salient feature information. Furthermore, ARO incorporates a feature replication layer to duplicate and refine feature maps at multiple levels, thereby enhancing its capability to capture fine-grained details and generate richer feature representations. Additionally, a multi-path upsampling strategy preserves fine-grained features during upsampling. Extensive experimental evaluations conducted on the benchmark CuLane and Tusimple datasets demonstrate that ARO achieves an F1 score of 80.60% on the CuLane dataset, outperforming state-of-the-art methods, showcasing the effectiveness of the proposed approach in handling data uncertainty for robust lane detection in autonomous driving. Our code are available at https://github.com/caoxu0109/lane_detect_ldfr.**

Index Terms—**Lane detection, Multi-scale fusion network, Attention mechanisms, Culane, Tusimple.**

I. Introduction

THE rise of autonomous driving technology has made it a key focus in today's automotive industry and research. Lane detection, as a crucial component of environmental perception and interaction [1], [2], plays a key role in implementing Advanced Driver Assistance Systems (ADAS) [3], [4] and autonomous driving. It is designed to ensure vehicle stability within lanes and provide necessary road information for self-driving cars by identifying and analyzing lane markings on the road, helping drivers maintain proper driving.

Therefore, advancements in technologies capable of real-time and accurate lane detection [5]–[7] are crucial for enhancing both safety and passenger comfort. With the advent of Convolutional Neural Networks (CNNs), several lane detection methods [8]–[10] have achieved significant progress in real-time performance and accuracy. However, lane detection encounters numerous challenges in real-world road scenarios, with data uncertainty being a prominent issue. This uncertainty stems from several factors: Firstly, complex and variable road environments, including varying lighting conditions, weather patterns, and road surface materials, significantly impact image quality, leading to increased data noise and blurring. Secondly, the diversity of lane lines themselves, such as variations in color, shape, and width, as well as wear and occlusion, contribute to the difficulty of detection. Additionally, inherent sensor error and uncertainties in the data acquisition process are also significant factors. Traditional deterministic lane detection methods [11] often struggle to address these uncertainties effectively, resulting in erroneous detections or complete failures. Consequently, effectively mitigating data uncertainty caused by environmental interference is paramount to improving the robustness and reliability of lane detection algorithms.

For instance, CondLaneNet [12] introduced the Recurrent Instance Module (RIM) to capture more contextual information. Meanwhile, Resa [13] proposed an information dissemination mechanism. This mechanism aims to extract information from complex environments. Nevertheless, these methods do not fully explore the potential of integrating information from multi-scale features. Previous research indicates that features at different scales offer unique and valuable characteristics: deep features capture global structural information, while shallow features highlight finer details. Furthermore, the effectiveness of anchor frame design significantly influences target detection performance. A well-designed anchor frame can simplify the regression process and improve detection accuracy. However, existing anchor frame designs often lack adaptability to dynamic environments, limiting their performance in diverse scenarios. CLRnet [14] addresses this limitation by employing multi-level feature integration, initially determining the presence of lane lines at a higher level and subsequently refining their precise positions at a lower level. Moreover, the lane prior used for regressing lane line positions through point offset prediction has demonstrated advantages over other techniques. Despite its efficiency and accuracy, this method relies on only two feature layers, potentially overlooking valuable local information. This study aims to advance lane detection techniques by introducing a novel model named ARO. First, we propose an innovative cross-level fusion attention mechanism that enables the fusion of feature maps from different levels by assigning weighted attention. This mechanism facilitates more effective extraction and utilization of rich feature information. Second, to enhance the refinement process, we introduce a feature replication layer. This layer generates multi-hierarchical feature information by duplicating feature maps and processing them at distinct levels. This refinement improves the model's fine-grained representation of lane lines, leading to more accurate localization and identification.

Recognizing the common issue of information loss in traditional feature fusion [15]–[17], we propose a multipath

Manuscript received August 18, 2024; revised December 12, 2024.

X. Cao is a postgraduate student of the College of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: 222083500537@stu.ustl.edu.cn)

W. Liu is a professor of the College of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, Co 114051, China. (corresponding author to provide fax:0412-5929809; e-mail: succman@163.com)

Z. Wang is a postgraduate student of the College of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China. (e-mail: 222083500529@stu.ustl.edu.cn)

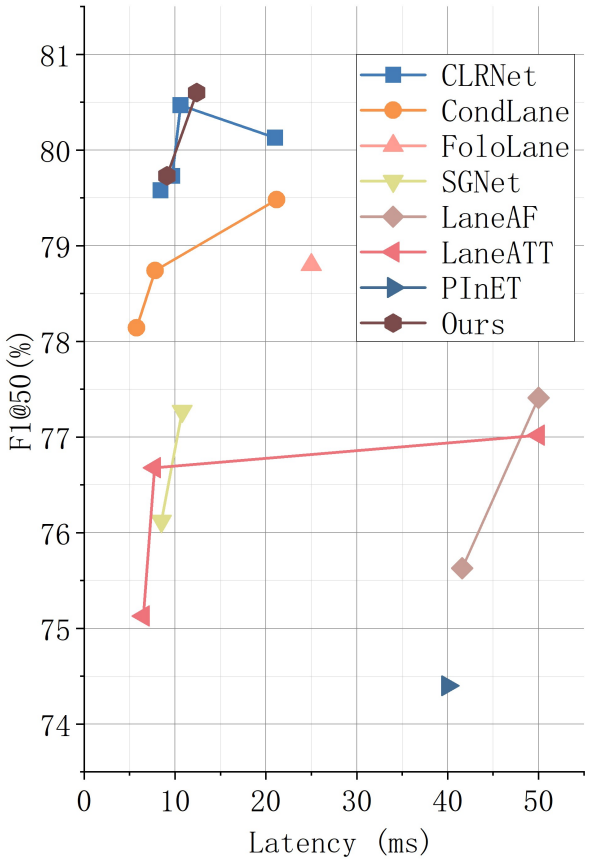

Fig. 1. Comparison Results of Mainstream Methods.

downsampling strategy. This strategy combines the parallel execution of various downsampling methods, capturing multi-scale representations of features more comprehensively and mitigating information loss. By integrating these techniques, along with the previously mentioned cross-level fusion attention mechanism, our model excels in enhancing the capture of context information and detail characteristics in complex scenes. In summary, the specific contributions of this paper can be outlined as follows:

1) The integration of the multi-scale attention fusion module within the Feature Pyramid Network (FPN) enhances the model's capacity to capture information across various scales.
2) We propose a multi-path up-sampling strategy designed to preserve more useful information during the up-sampling process, minimizing information loss.
3) The feature copy module and feature copy layer are introduced with creativity, wherein the original feature map is duplicated, providing a richer and multi-level feature representation for detection tasks.
4) The model underwent experimentation using the Culane dataset, wherein it demonstrated superior performance, surpassing the current benchmarks across various challenging categories.

To validate our approach, confirmatory experiments were conducted on two standard datasets, Culane and Tusimple. As depicted in Figure 1, our approach surpasses many contemporary technologies in terms of both performance and efficiency.

II. Previous Work

This section primarily analyzes diverse methods of lane detection, categorizing them into three main groups: segmentation-based, anchor point-based, and parameter-based.

A. Anchor-Based methods

The lane detection methods based on anchor points [18], [19] can be categorized into point anchor-based and line anchor-based methods. The point anchor-based method utilizes predefined point anchors as references to precisely detect the position of lane lines through regression. This approach focuses on identifying discrete points in the image, allowing the network to predict lane line information at each point. GANet [18] adopts a global regression strategy, which proves to be more efficient than traditional local regression; however, the accuracy of global regression must also be upheld. CLRNet [14] seeks to optimize the training process by integrating high and low-level features, enhancing the learning of lane features through the ROIGather module, and employing LIoU losses. These improvements are designed to enhance the model's ability to understand the complex structure of lanes. While point anchor-based methods offer computational efficiency and scenario-specific lane detection, they still encounter difficulties while dealing with severe occlusions and challenging environmental conditions.

Line anchor-based methods focus on delineating the geometry of lane lines. For instance, CondLaneNet [12] intricately captures the geometric structure of continuous lane lines for precise regression, demonstrating a systematic approach to shape representation. While dynamic convolution and conditional lane detection frameworks contribute strategies for real-time performance, such approaches may encounter limitations in handling complex traffic scenarios. It highlights the need for more adaptive line anchor layouts and intelligent detection mechanisms to cope with dense traffic and road markings.

Anchor-based detection methods [20], [21] excel in reducing computational complexity compared to full-graph search or large-scale area detection techniques. They confine the search to predefined anchors that potentially contain the target. With well-designed anchors, this approach demonstrates adaptability in handling target scale variations and rotations. However, choosing an appropriate anchor configuration is crucial to ensure high detection accuracy. Part of this challenge involves determining the layout of these anchors. The anchor must mirror the typical features of the target to accurately capture it in various situations. This necessitates infusing expertise into the anchor design process, optimizing the size, shape, and number distribution of anchors through experimentation and practical experience. For instance, in lane detection applications, anchor points should cover lane lines of different widths and curvatures. Nonetheless, even with a properly designed anchor, the method still faces the challenge of information acquisition when the target is occluded. Effectively addressing the occlusion problem may require the integration of advanced occlusion recognition and processing techniques.

B. Segmentation-Based methods

Segmentation-based methods [22]–[24] frame lane segmentation as a pixel-by-pixel prediction task. For example, TwinLiteNet [25] employs a dual decoder structure, featuring two decoder blocks for driving area and lane splitting tasks separately. This enables independent optimization of segmentation performance for each task, enhancing overall segmentation accuracy. To fortify the model's capability to capture spatial and channel information, TwinLiteNet [25] introduces a dual attention mechanism, comprising a location attention module (PAM) and a Channel Attention Module (CAM). This inclusion aims to capture global dependencies in spatial and channel dimensions, thereby enhancing the model's perception of context information.

Moreover, YOLOPv2 [26] proposes a multi-task learning network to address panoramic car driving perception. This network simultaneously handles traffic object detection, drivable road area segmentation, and lane line detection. Researchers integrate data enhancement technologies such as Mosaic and Mixup, along with a hybrid loss function configuration combining Dice loss and Focal loss [27]. These strategies and mixed loss functions significantly improve the performance of all three tasks. ERFNet [28] introduces a novel lane detection method with a built-in Enhanced Spatial Attention module (ESA). The ESA module analyzes visual cues in the probability prediction graph to extract the accuracy confidence of the lane signal.

While these methods exhibit exceptional performance in handling sections with complex topological variability, such as twists and turns, intersections, or other intricate pavement structures, they often come with substantial computational load requirements. Due to their involvement in fine pixel-level classification of the entire image, these methods consume considerable computing resources and may introduce delays in real-time application scenarios. Additionally, they are sensitive to image noise and lighting changes, heightening the risk of false or missed detections. This sensitivity affects the reliability and accuracy of lane detection, especially under complex road conditions, such as intersecting, fractured, or low-visibility lane lines. Segmentation-based detection algorithms can suffer from a negative impact on accuracy, particularly when relying solely on GPU accelerators for a large number of calculations.

C. Parameter-Based methods

The lane detection method based on parameters [19], [29], [30] employs mathematical parameters to represent the shape of lane lines. This method typically characterizes lane lines in the image by fitting model parameters, such as those of a line or curve. BezierLaneNet [30] introduces a lane detection method based on parametric curves. Diverging from existing methods grounded in segmentation and point detection, this approach can inherently learn the overall lane representation without relying on heuristics to decode predictions or formulate numerous anchor points. To address optimization challenges inherent in existing polynomial curve methods, the authors propose the use of a parametric Bessel curve, chosen for its ease of calculation, stability, and high degree of freedom transformation. Additionally, the authors suggest a feature flipping fusion based on variable convolution to leverage lane symmetry in driving scenarios. This method has demonstrated state-of-the-art performance on the widely recognized LLAMAS [31] benchmark test.

Similarly, BSNet [32] presents a lane detection method based on B-spline curves. Initially, the B-spline curve serves as the representation of the lane line, with each point on the curve's position calculated through the weighted sum of control points. Subsequently, the residual network and pyramid feature network extract features from the lane line candidate region. Local features of the lane candidate region are then obtained using a self-attention mechanism and local feature pooling operation, and global features are fused. Finally, classification and regression modules predict the existence probability of lane lines and the coordinates of control points.

Parametric curve methods for detecting curved or continuously changing roads offer superior adaptability. They also have clear advantages in computational complexity compared to some methods based on whole image pixel segmentation. However, when confronted with complex traffic scenarios commonly encountered in practical applications, such as intersections or road conditions featuring multiple geometric structures, parameterization methods relying on simple geometric models may fall short in capturing and characterizing all the details of these scenarios. Additionally, lane features can be dynamic, with rapid lane changes or temporary road markings. A parameterized curve-based approach may require continuous adjustment and updating of model parameters to account for these changes. This poses challenges to the stability and accuracy of real-time performance.

III. Method

Data uncertainty is inevitable, as the large-scale datasets we currently apply are collected from the internet. On one hand, low-quality images are common in unconstrained large-scale application datasets. On the other hand, the image quality obtained by different acquisition methods is not the same, such as overexposure, underexposure, and blurriness. Furthermore, environmental occlusion, small targets, and low-quality acquisition exacerbate the problem of noisy data uncertainty. To address this issue, this chapter proposes an adaptive ROI optimization pyramid network. The network uses multi-scale fusion processing and feature replication techniques to extract key features, reducing the impact of data uncertainty on the detection results.

A. Adaptive ROI Optimization Pyramid Network

Our proposed lane detection model architecture is illustrated in Figure 2. The model is built on a ResNet backbone, integrating a Feature Pyramid Network (FPN) to process input images and create multi-level feature representations. The key innovation lies in the utilization of the FPN layer, embedding a multi-scale attention fusion structure to refine the feature maps. This ensures that feature maps at different levels prioritize attention to lane information. The notable advantage of this structure is its capacity to selectively extract features crucial for lane detection and enhance these features throughout the fusion process. To address the common issue of feature information loss in up-sampling, we introduce a

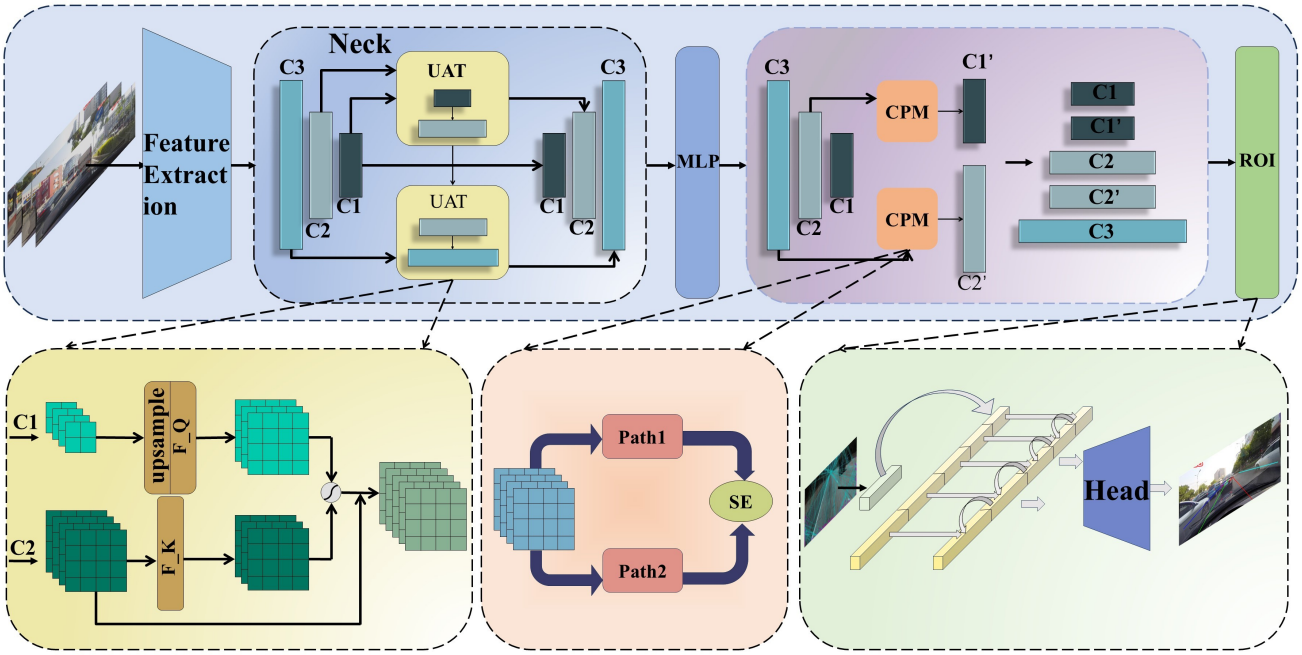

Fig. 2. ARO Model Overall Structure Diagram.

designed multi-path up-sampling architecture. This architecture not only effectively preserves rich information content but also maintains precise details of features, enhancing the model's performance. Additionally, a feature copy module is incorporated into the model, acting on the features of layers C1 and C2 and replicating them. This replication strategy contributes to generating more comprehensive and hierarchical feature maps, thereby boosting the overall expressive power of features. For the copied feature map, a method akin to Clrnet [14] is employed to determine the region of interest, with the detection head predicting the offset of the prior anchor point. This strategy enables high-precision localization of lane alignment. Ultimately, after the entire data flow processing, the model successfully outputs fine-grained lane detection results containing rich feature representations. By predicting the prior anchor points, our model efficiently performs the detection task.

The results are expressed as follows:

$$\mathrm{r}_k = \{cls_k, \mathrm{Sx}_k, Sy_k, \theta_k, Length_k, offset_k\} \quad (1)$$

$$offsets = \{l_k^1, l_k^2, l_k^3 ... l_k^{71}, l_k^{72}\} \quad (2)$$

In this study, k denotes the number of detected lane lines, and it is fixed at 192. In the context, cls signifies the classification score of the lane line, Sx and Sy denote the coordinates of the starting point of the lane line, θ represents the angle between the lane line and the horizontal plane, $Length$ represents the length of the lane line, $offset$ represents the offset of the predicted lane line point. A total of 72 points were employed to represent the anchor points of the lane lines.

B. Multi Scale Attention Fusion

In the field of deep learning, feature fusion emerges as a pivotal technology enhancing the semantic comprehension capabilities of models. Profound deep features excel in capturing comprehensive semantic information, a critical aspect for comprehending the overall characteristics of an object. Nevertheless, the conventional feature pyramid network (FPN) [33] employs rudimentary upsampling and linear fusion, proving inadequate in leveraging the feature information of individual layers. To address this limitation, we propose an upsampled attention module (UAT) and devise an innovative feature fusion strategy. Subsequently, the fusion factor is derived by subjecting the similarity score to a nonlinear transformation through the activation function. This fusion factor allocates distinct weights to features at various spatial locations during the feature fusion process, thereby regulating the information flow. Ultimately, the upsampling attention module employs multiple paths to up-sample the C1 feature map, aligning its dimensions with C2. This not only enhances information flow, but also mitigates the information bottleneck associated with a single path. Through the weighted fusion of feature maps post-upsampling using fusion factors, a feature map with enhanced content and boundary information is obtained, optimizing lane detection performance in complex scenes. The UAT module not only effectively synthesizes the multi-layer feature information, but also delivers a more accurate feature representation for the lane detection task, thereby enhancing the overall performance of the model. The process can be formally described by the following formula:

$$C1' = \mathrm{Interpolation}(C1) \quad (3)$$

$$Query = fQ(C1') \quad (4)$$

$$Key = fK(C2) \quad (5)$$

$$SimMap = Sigmoid(Query \bullet Key) \quad (6)$$

$$out = (1 - SimMap) \bullet US(C1) + SimMap \bullet C2 \quad (7)$$

In this process, to determine the weights of the features intended for fusion, the computation is grounded in the attention of features of varying dimensions. Feature fusion follows Equation 1.7. By incorporating the upsampling attention module, the model integrates information from different

levels more effectively, focusing on boundary details and features of the region near the lane line. This significantly enhances the lane detection performance. As illustrated in

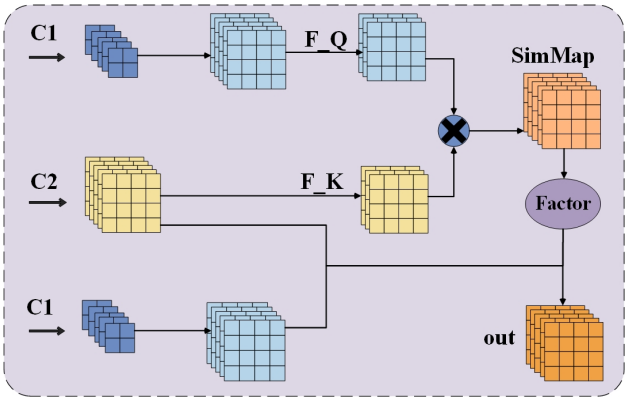

Fig. 3. Multi Scale Attention Feature Fusion Module.

Figure 3. The features of layers C1 and C2 are initially upsampled to the same dimension using linear interpolation. Subsequently, a multiscale attention mechanism is employed to compute the similarity score between these two layers of features and transform it into a fusion weight (scaling factor) through an activation function. In addition, the features of layer C1 undergo further upsampling via the multi-path upsampling module to enhance their combination with the features of layer C2. Finally, using the obtained fusion weights, the upsampled features of the C1 layer and the features of the C2 layer are linearly fused to perform feature integration.

C. Sampling on Multipath

Preserving Key information is crucial during feature fusion. Both interpolation algorithms and transposed convolution algorithms are frequently employed in upsampling processes; However, either approach has its own drawbacks. Interpolation algorithms enlarge the size by computing the value of a new pixel point, often resulting in information loss during the recovery of image details. While the transposed convolution reconstructs the features by training the upsampling parameters, it can potentially lead to artifacts and image blurring, which can affect the final fusion result.

To overcome these limitations, a two-path upsampling method, illustrated in Figure 4, is introduced. This approach seeks to incorporate the benefits of interpolation and transposed convolution while mitigating their respective drawbacks. Our module consists of two parallel paths: The first path employs a linear interpolation operation to upsample the input feature X, preserving the continuity and smoothness of the image while aiding the reconstruction of detailed information. The second path leverages transposed convolutions to learn upsampling efficiently, emphasizing the acquisition of feature transformations.

Following the combination of the two upsampled feature graphs, a convolution layer with a 1×1 convolution kernel is applied to diminish the dimension of the combined features. This step not only reduces the dimensionality of the features, but also combines the benefits of the two distinct upsampling strategies to optimize the feature expression and enhance the operational efficiency of the model. The upsampling formula can be expressed as:

$$upfea = ReLu(I(x) + CTBR(x)) \tag{8}$$

where $I(x)$ denote the upsampled feature map obtained by the interpolation algorithm and $CTBR(x)$ denotes the upsampled feature map obtained by the transposed convolution. The final upsampling result is obtained by summing them and subjecting them to a convolution operation.

The two-path sampling strategy thoroughly considers the benefits of the interpolation algorithm and the transposed convolution algorithm, while mitigating their drawbacks. By preserving feature continuity and smoothness while obtaining suitable upsampling parameters, we can optimize the retention of crucial information, thereby enhancing the quality and performance of feature fusion.

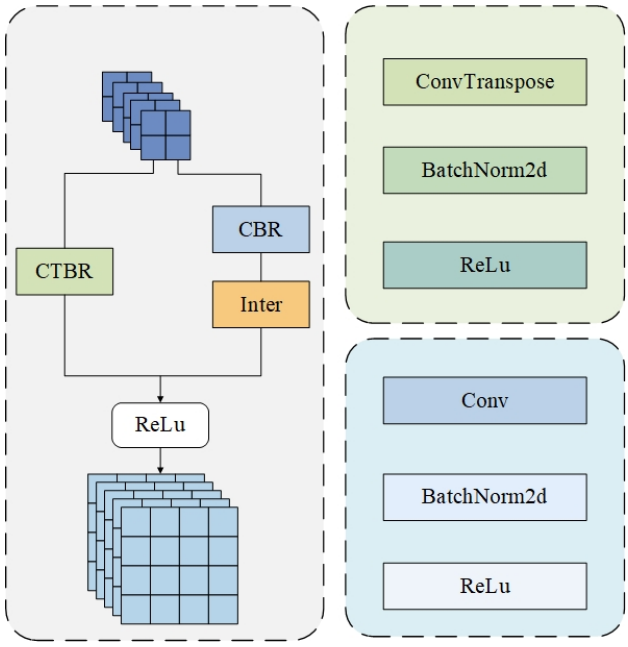

Fig. 4. Multi Path Upsampling Module.

D. Feature Copy Module

Drawing on the insights from Clrnet [14], to further enhance the accuracy of lane detection, this study introduces a method for intensifying the depth of feature refinement, proven through experiments to markedly enhance detection effectiveness.

We devise a new component known as the Feature Copy Module (FCM). This module downsamples the features of layer C3 to produce $C3^{'}$, and similarly processes the features of layer C2 to generate $C2^{'}$. In this way, features of three different scales and five levels can be obtained, which serves as a bridging mechanism in the elaboration of features across multiple levels.

The FCM module efficiently integrates high-level semantic features with low-level detailed features by transferring information and integrating features across different levels. This multi-scale and multi-level feature representation facilitates a more accurate capture of the detailed features and contours of lane lines, thus enhancing the accuracy and stability of lane detection.

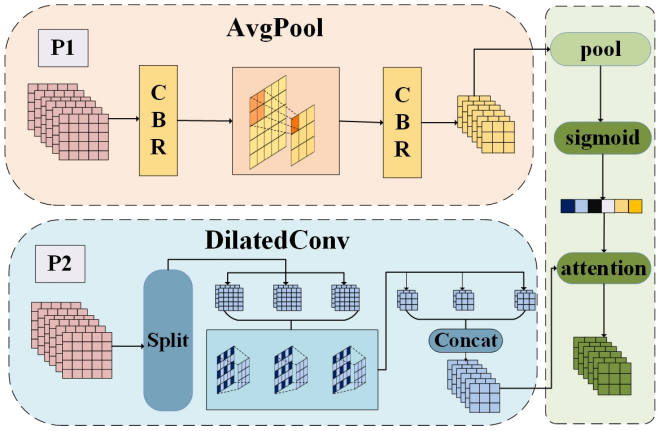

Fig. 5. Feature Replication Enhancement Module.

This article presents a valuable exploration of lane detection techniques and demonstrates that combining high-level semantics with low-level details, accompanied by the implementation of FCM modules, is able to effectively enhance detection performance. The design of the FCM is illustrated in Figure 5 and is performed through the following steps: Path1: The input X undergoes a convolutional layer to yield Conv(X), followed by undersampling through the pooling layer to obtain Path1(X). Path2: The input X is partitioned into channel groups, with each group undergoing dilated convolution, denoted as DConv(X), and the resulting outputs are concatenated to form Path2(X).

Finally, the output of Path 1 and Path 2 are fused and dimensionality reduced through the SE (Squeeze-and-Excitation) module. The SE module enhances feature recognition by autonomously learning channel weights, optimizing the fusion of outputs from both paths, and simultaneously reducing the number of channels. This design strategy combines convolution, pooling, and grouped- dilated convolutions to achieve subsampling that preserves more comprehensive feature information, enables extraction of features of varying sizes, and enhances the model's understanding and processing power of the data.

In the process depicted in Figure 5, the input feature map undergoes downsampling through two distinct paths. Path1 uses a pooling layer for downsampling, whereas Path2 uses dilated convolution to achieve the same. While both methods aim to reduce the spatial dimension of feature maps, they employ different techniques to preserve crucial image information and features. Finally, the Channel Attention Module is applied to weight the feature maps from Path1 and Path2, enabling the model to prioritize critical information and disregard less relevant details for the final task. Following this procedure, we obtain downsampling results that combine the benefits of pooling layer and dilated convolution, emphasizing the most crucial feature information.

IV. Experiments

A. Dataset

Our model underwent evaluation on two extensively employed datasets: the CULane [31] dataset and the Tusimple dataset. The CULane [31] dataset serves as a comprehensive dataset for Lane Detection, initially introduced in the paper titled "Scale-Aware CNN for Real-Time Traffic Lane Detection". The dataset consists of more than 55 hours of video data captured by onboard cameras on various vehicles driven by different individuals in Beijing, yielding 133, 235 image frames. The dataset was partitioned into a training set (88,880 frames), a validation set (9,675 frames), and a test set (34,680 frames). Due to its substantial data volume, the dataset enhances the robustness of our model. In addition to the regular scenarios, the test set incorporates eight challenging scenarios including Crowded, Dazzle, Light, Shadow, NoLine, Arrow, Curve, Cross, and Night, which are designed to assess the model's capacity to generalize in complex environments. The Tusimple dataset, which is more focused on highway scenes, is widely used in the lane detection field. The dataset encompasses 3,268 images for training, 358 for verification, and 2,782 for testing, all maintaining a consistent resolution of 1280×720 pixels. This dataset facilitates the evaluation of model performance in a highway environment.

B. Experimental Design

For the experiments, ResNet [34] and DLA were selected as the fundamental backbone networks for image feature extraction. These networks were selected based on their proficiency in learning feature representations at various levels, enabling efficient capture of complex features in images. To strike a balance between model performance and computational efficiency, the input image resolution is consistently set to 320×800. This ensures the model's ability to maintain high performance in image processing while judiciously allocating computational resources. The model was constructed and trained using the PyTorch deep learning framework, with computation accelerated by a single NVIDIA GeForce RTX 3090 graphics card. Depending on the dataset, the training cycles vary: the CULane [31] dataset underwent 11 training cycles, whereas the Tusimple dataset underwent up to 70 training cycles, illustrating the differences in complexity and sample size between the two. AdamW optimizer is employed to optimize the model. Derived from the classical Adam algorithm, it introduces a weight decay mechanism to enhance model generalization. The initial learning rate is set to 0.001 and the cosine decay learning rate strategy is implemented with a power of 0.9. This ensures that the learning rate gradually decreases during training, promoting more stable convergence. To optimize the model, the classification loss, xyz loss, and regression loss formulas are defined as follows:

$$L_{total} = w_1 \bullet L_{cls} + w_2 \bullet L_{xyz} + w_3 \bullet L_{Liou} \quad (9)$$

where L_{cls} represents the focal loss between the prediction and label [27], L_{xyz} is the smooth L1 loss of the starting point coordinates, θ represents the angle, and lane length regression, while L_{Liou} is the Line IoU loss between the predicted lane and ground truths

C. Evaluation Indicators

The evaluation metrics for the model on the CULane [31] dataset consist of the F1 score and mean F1 score (mF1), providing a comprehensive assessment of the model's performance across eight challenging categories. For individual F1 score calculations, an intersection ratio (IoU) threshold of

0.5 is employed to determine the proximity of the predicted lane line to the actual lane line. Notably, the F1 score is computed as the harmonic average of prediction accuracy and recall.

Precision is defined as the ratio of correctly detected lane-line points to predicted lane-line points, while recall represents the ratio of correctly detected lane-line points to actual lane-line points. In this context, the Intersection over Union (IoU) threshold is utilized to ascertain the correctness of lane detection. The mF1 scores are derived by computing the F1 scores across various IoU thresholds and subsequently averaging them. This encompasses F1 scores at different thresholds ranging from 0.5 to 0.95, separated by 0.05, denoted as F1@50, F1@55, ..., F1@95. This approach ensures a comprehensive and rigorous evaluation of the model performance beyond the limitations of a specific IoU threshold. The specific calculation formula is shown as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{10}$$

$$mF1 = (F1@50 + F1@55 + ... + F1@95)/10 \tag{11}$$

By employing the $mF1$ score as the final evaluation metric, we can precisely gauge the proficiency of the model in lane-line detection, ensuring its continued high accuracy and recall across diverse and complex scenarios.

On the Tusimple dataset, we employed accuracy (ACC) and F1 scores as performance indicators. These metrics provide insights into the overall effectiveness of the model in the lane detection task and its accuracy in recognizing intricate details. Further details are provided below:

$$acc = \frac{\sum Cclip}{\sum Sclip} \tag{12}$$

$Cclip$ represents the number of correctly predicted points in the image, and $Sclip$ represents the number of points in the ground truth. In F1 scores, a correct prediction is defined when more than 85% of the predicted lane points fall within 20 pixels of the actual lane point. This criterion enables a precise evaluation of the performance of the model on the lane detection task. Employing these evaluation criteria facilitates a comprehensive comparison of model performance on diverse lane detection tasks across different datasets. This ensures that the model not only exhibits overall proficiency, but also maintains enhanced accuracy in identifying individual components.

D. Compare with the most Advanced models

1) Performance on CULane Dataset: To assess the effectiveness of our approach, we conduct a comparative analysis with 12 widely used models and show that our model outperforms them [35]–[38], achieving superior results in terms of performance. Specifically, when ResNet is employed as the backbone network, as shown in Table I our model achieved an F1 metric of 80.6% with an IOU threshold set to 0.5. Moreover, across eight challenging scenarios, our approach outperforms current mainstream methods, including the top 12 models. For example, in scenarios such as Normal, Crowded, Dazzle and Noline, our model demonstrated superior performance compared to other methods [24], [39]–[41].

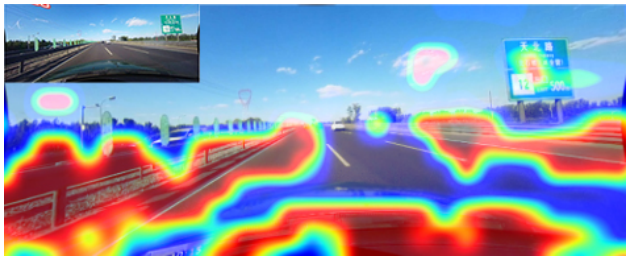

Fig. 6. Visual heat map.

These findings indicate that our model excels in handling object detection tasks within complex scenes. Considering the feature extraction and fusion capabilities of the multi-scale attention fusion module, our model can achieve more accurate localization and recognition of target objects.

2) Performance on Tusimple Dataset: Table II shows the method's performance on the Tusimple dataset. Although the performance metrics of this dataset are close to saturation, our approach exhibits substantial improvements. Remarkably, the proposed method achieves state-of-the-art results in the F1 metric, demonstrating an excellent balance between accuracy and recall.

Tusimple, being a widely recognized benchmark dataset, poses specific challenges in lane detection. Despite extensive research efforts on this dataset, technological advances have led to incremental improvements, thus limiting further enhancements. Nevertheless, the introduction of innovative structures and modules, such as the Feature Replication Module (CPM) and the Unified Attention Fusion Module (UAT), has allowed us to achieve a certain level of performance improvement on this challenging dataset. This outcome not only validates the effectiveness of our approach, but also demonstrates that introducing novel methods and techniques can yield substantial improvements even on well-established datasets. There is still potential to push the boundaries of performance. In the process of verifying the practicality of our approach, it establishes a new research direction and objective for lane detection. Specific improvements in F1 scores and other related metrics are shown in Table III.

E. Ablation Experiment

1) The Whole Ablation Experiment: To examine the contributions of the multi-scale attention fusion module and the feature replication layer in our model, we performed ablation experiments on the CULane [31] dataset using a scaled-down version of the model. The results for each scenario are presented in the following table. The first row shows the results of the baseline model. Subsequently, the second line introduces the incorporation of the multi-scale attention fusion module, accompanied by the respective outcomes. Finally, the third row incorporates the feature copy layer and depicts the corresponding results. By comparing the results, the following conclusions can be drawn: Integrating the multi-scale attention fusion module improves the F1 value across IoU thresholds. Adding the feature replication layer further enhances the F1 value, particularly at the 0.5 threshold. The Feature Replication Layer enhances the capacity of

TABLE I
RESULTS ON CULANE DATASET.

Method	Backbone	F1@50	F1@75	Normal	Crowded	Dazzle	Shadow	Noline	Arrow	Curve	Cross	Night
SCNN	VGG16	71.60	39.84	90.60	69.70	58.50	66.90	43.40	84.10	64.40	1990	66.10
RESA	ResNet34	74.50	-	91.90	72.40	66.50	72.00	46.30	88.10	68.60	1896	69.80
RESA	ResNet50	75.30	53.39	92.10	73.10	69.20	72.80	47.70	88.30	70.30	1503	69.90
FastDraw	ResNet50	-	-	85.90	63.60	57.00	69.90	40.60	79.40	65.20	7013	57.80
E2E	ERFNet	74.00	-	91.00	73.10	64.50	74.10	46.60	85.80	71.90	2022	67.90
UFLD	ResNet18	68.40	40.01	87.70	66.00	58.40	62.80	40.20	81.00	57.90	1743	62.10
UFLD	ResNet34	72.30	-	90.70	70.20	59.50	69.30	44.40	85.70	69.50	2037	66.70
PINet	Hourglass	74.40	51.33	90.30	72.30	66.30	68.40	49.80	83.70	65.20	1427	67.70
LaneATT	ResNet18	75.13	51.29	91.17	72.71	65.82	68.03	49.13	87.82	63.75	1020	68.58
LaneATT	ResNet34	76.68	54.34	92.14	75.03	66.47	78.15	49.39	88.38	67.72	1330	70.72
LaneATT	ResNet122	77.02	57.50	91.74	76.16	69.47	76.31	50.46	86.29	64.05	1264	70.81
LaneAF	ERFNet	75.63	54.53	91.10	73.32	69.71	75.81	50.62	86.86	65.02	1844	70.90
LaneAF	DLA34	77.41	56.79	91.80	75.61	71.78	79.12	51.38	86.88	72.70	1360	73.03
SGNet	ResNet18	76.12	-	91.42	74.05	66.89	72.17	50.16	87.13	67.02	1164	70.67
SGNet	ResNet34	77.27	-	92.07	75.41	67.75	74.31	50.90	87.97	69.65	1373	72.69
FOLOLane	ERFNet	78.80	-	92.70	77.80	75.20	79.30	52.10	89.00	69.40	1569	74.50
CondLane	ResNet18	78.14	57.42	92.87	75.79	70.72	80.01	52.39	89.37	72.40	1364	73.23
CondLane	ResNet34	78.74	59.39	93.38	77.14	71.17	79.93	51.85	89.89	73.88	1387	73.92
CondLane	ResNet101	79.48	61.23	93.47	77.44	70.93	80.91	54.13	90.16	75.21	1201	74.80
CLRNet	ResNet18	79.58	62.21	93.30	78.33	73.71	79.66	53.14	90.25	71.56	1321	75.11
CLRNet	ResNet34	79.73	62.11	93.49	78.06	74.57	79.92	54.01	90.59	72.77	1216	75.02
CLRNet	ResNet101	80.13	62.96	93.85	78.78	72.49	82.33	54.50	89.79	75.57	1262	75.51
CLRNet	DLA34	80.47	62.78	93.73	79.59	75.30	82.51	54.58	90.62	74.13	1155	75.37
Ours	ResNet18	**79.73**	62.02	**93.34**	**78.38**	**76.00**	79.21	**53.17**	**90.44**	68.55	1034	74.96
Ours	DLA34	**80.60**	62.70	**93.82**	**79.68**	**76.81**	82.00	**54.60**	89.75	71.26	1027	75.33

TABLE II
RESULTS ON TUSIMPLE DATASET.

Method	Backbone	F1	Acc	FP	FN
SCNN	VGG16	95.97	96.53	6.17	1.80
RESA	ResNet34	96.93	96.82	3.63	2.48
PolyLaneNet	EfficientNetB0	90.62	93.36	9.42	9.33
E2E	ERFNet	96.25	96.02	3.21	4.28
UFLD	ResNet18	87.87	95.82	19.05	3.92
UFLD	ResNet34	88.02	95.86	18.91	3.75
LaneATT	ResNet18	96.71	95.57	3.56	3.01
LaneATT	ResNet34	96.77	95.63	3.53	2.92
LaneATT	ResNet122	96.06	96.10	5.64	2.17
FOLOLane	ERFNet	96.59	96.92	4.47	2.28
CondLaneNet	ResNet18	97.01	95.48	2.18	3.80
CondLaneNet	ResNet34	96.98	95.37	2.20	3.82
CondLaneNet	ResNet101	97.24	96.54	2.01	3.50
CLRNet	ResNet18	97.89	96.84	2.28	1.92
CLRNet	ResNet34	97.82	96.87	2.27	2.08
CLRNet	ResNet101	97.62	96.83	2.37	2.38
Outs	ResNet18	97.90	96.86	2.24	1.95

TABLE III
RESULT OF ABLATION EXPERIMENT.

Method	UAT	CP	MF1	F1@50
ARO			55.23	79.58
ARO	✓		55.21	79.60
ARO	✓	✓	55.29	79.60

the model to provide richer feature representations, further improving the overall performance of lane detection.

Ablation experiment results demonstrate that both the Multi-scale Attention Fusion Module and the Feature Replication Layer contribute positively to enhancing the performance of our model on the CULane [31] dataset. These components augment the perceptual power of features at various scales, yielding a more comprehensive feature representation and thus enhancing the accuracy and robustness of lane detection.

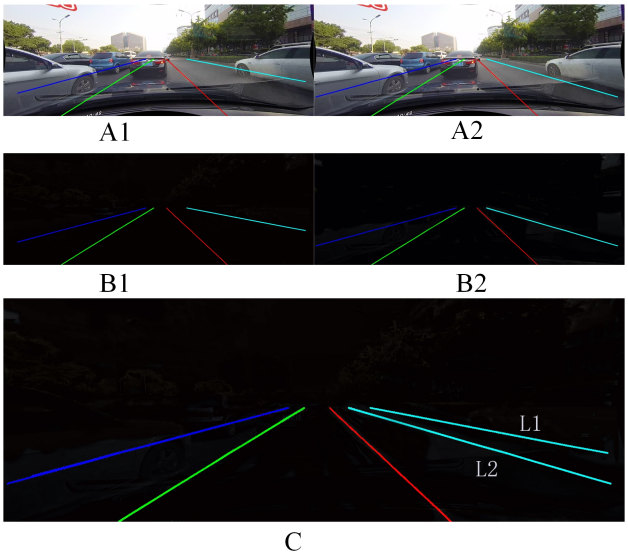

Fig. 7. Visual detection results.

2) An Experiment on Multi-scale Attention Fusion Module: To substantiate the practical impact of the multi-scale attention fusion module, we employ a visual approach to depict the attention scores, providing a visual representation of the regions that receive heightened attention during feature fusion. In Figure 6, brighter colors correspond to elevated attention scores, indicating heightened significance of the respective regions in the feature fusion process. Consequently, the distribution of attention in distinct regions can be visually observed, elucidating the role played by different regions in the feature fusion process. Drawing insights from the

experimental results, we can infer that the Unified Attention Fusion Module (UAT) structure allocates greater attention to the lower part of the image during the feature fusion stage, with a particular emphasis on the local lane line and the nearby vehicle. Such a fusion approach is imperative to capture the key regions of interest and thereby substantially enhance the accuracy of lane detection.

3) Experiments for Feature Replication: As shown in Figure 7. In the visual analysis of the impact of our Feature Copy Module, it is observed that the detection results of the L1 lane lines displayed in the left image significantly deviate from the real scene in the absence of the feature copy layer. In turn, upon incorporating the Feature Replication Layer into the model, the right image shows a significantly improved detection of the L1 lane lines, which significantly enhances the agreement with the actual lane line positions. Through a comparative analysis of the results, it can be affirmed that the Feature Replication Layer significantly enhances the detection accuracy in lane detection task.

V. Conclusion

This study tackles the inherent data uncertainty encountered in lane detection by proposing an adaptive pyramid network. Recognizing that the data acquisition and utilization pipeline is susceptible to unavoidable errors and uncertainties, our network employs adaptively learned feature fusion weights to enhance the model's feature extraction capabilities. Central to our approach are two novel modules: the Feature Copy Module (FCM), designed to enrich feature representation, and the Unified Attention Fusion Module (UAT), which facilitates effective multi-scale feature integration. The FCM aims to improve lane detection performance by facilitating information transfer and integration between different feature layers, combining high-level semantic features with detailed features. Simultaneously, we designed the Unified Attention Fusion Module (UAT) to perform feature fusion by calculating adaptive weights. To maximize the retention of information at different scales during feature fusion, we designed a multi-path upsampling method that reduces feature loss rate by incorporating transposed convolutions. Finally, we validated the superiority of our model on the CuLane and Tusimple datasets. On the CuLane dataset, the model achieved an F1 score of 80.60%, while on the Tusimple dataset, it achieved an F1 score of 97.90%, demonstrating high accuracy and robustness in various challenging scenarios. Overall, the proposed lane detection model in this study effectively improves the accuracy and robustness of lane detection by combining high-level semantic information with low-level detailed features, and through innovative feature fusion and replication strategies. These achievements not only provide new solutions for lane detection problems in autonomous driving and driver assistance systems but also offer new directions and goals for future research in this field.

References

[1] D. Strawser and B. Williams, "Motion planning under uncertainty with complex agents and environments via hybrid search," *Journal of Artificial Intelligence Research*, vol. 75, pp. 1–81, 2022.

[2] M. Prédhumeau, L. Mancheva, J. Dugdale, and A. Spalanzani, "Agent-based modeling for predicting pedestrian trajectories around an autonomous vehicle," *JOURNAL of Artificial Intelligence Research*, vol. 73, pp. 1385–1433, 2022.

[3] S. Kannan, A. Thangavelu, and R. Kalivaradhan, "An intelligent driver assistance system (i-das) for vehicle safety modelling using ontology approach," *International JOURNAL of UbiComp*, vol. 1, no. 3, pp. 15–29, 2010.

[4] S. Velarde-Gomez and E. Giraldo, "Robust state space embedded control of a 3d printed permanent magnet synchronous motor." *IAENG International Journal of Applied Mathematics*, vol. 54, no. 2, pp. 255–261, 2024.

[5] H.-B. Chen, L. Pan, L.-J. Lin, and L.-Q. Huang, "Non-local spatial information based lane detection," *2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, pp. 1–3, 2020.

[6] Y. Chen, Z. Xiang, and W. Du, "Improving lane detection with adaptive homography prediction," *The Visual Computer*, vol. 39, no. 2, pp. 581–595, 2023.

[7] Y. Liu, P. Xu, L. Zhu, M. Yan, and L. Xue, "Reinforced attention method for real-time traffic line detection," *JOURNAL of Real-Time Image Processing*, vol. 19, no. 5, pp. 957–968, 2022.

[8] G. Ras, N. Xie, M. Van Gerven, and D. Doran, "Explainable deep learning: A field guide for the uninitiated," *JOURNAL of Artificial Intelligence Research*, vol. 73, pp. 329–396, 2022.

[9] T. Tornede, A. Tornede, J. Hanselle, F. Mohr, M. Wever, and E. Hüllermeier, "Towards green automated machine learning: Status quo and future directions," *JOURNAL of Artificial Intelligence Research*, vol. 77, pp. 427–457, 2023.

[10] C. Jiqing, W. Depeng, L. Teng, L. Tian, and W. Huabin, "All-weather road drivable area segmentation method based on cyclegan," *The Visual Computer*, vol. 39, no. 10, pp. 5135–5151, 2023.

[11] M. Jafarlou and H. Naderi, "Improving fuzzy-logic based map-matching method with trajectory stay-point detection," *The International MultiConference of Engineers and Computer Scientists 2023*, pp. 48–57, 2023.

[12] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3773–3782, 2021.

[13] T. "Zheng, H. Fang, Y. Zhang, W. Tang, Z. Yang, H. Liu, and D. Cai, "Resa: Recurrent feature-shift aggregator for lane detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 3547–3554, 2021.

[14] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "Clrnet: Cross layer refinement network for lane detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 898–907, 2022.

[15] H. Xu, L. Yao, W. Zhang, X. Liang, and Z. Li, "Auto-fpn: Automatic network architecture adaptation for object detection beyond classification," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6649–6658, 2019.

[16] Z. Lv, D. Han, W. Wang, and C. Chen, "Ifpnet: Integrated feature pyramid network with fusion factor for lane detection," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1888–1897, 2023.

[17] J. Chen, H. Mai, L. Luo, X. Chen, and K. Wu, "Effective feature fusion network in bifpn for small object detection," *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 699–703, 2021.

[18] J. Wang, Y. Ma, S. Huang, T. Hui, F. Wang, C. Qian, and T. Zhang, "A keypoint-based global association network for lane detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1392–1401, 2022.

[19] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Polylanenet: Lane estimation via deep polynomial regression," *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 6150–6156, 2021.

[20] H. Xu, S. Wang, X. Cai, W. Zhang, X. Liang, and Z. Li, "Curvelanenas: Unifying lane-sensitive architecture search and adaptive point blending," *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XV 16*, pp. 689–704, 2020.

[21] Y.-B. Liu, M. Zeng, and Q.-H. Meng, "Heatmap-based vanishing point boosts lane detection," *arXiv preprint arXiv:2007.15602*, vol. 8, 2020.

[22] Z. Yang, C. Shen, W. Shao, T. Xing, R. Hu, P. Xu, H. Chai, and R. Xue, "Canet: Curved guide line network with adaptive decoder for lane detection," *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023.

[23] H. Fang, J. Zhu, and Y. Fang, "Continuitylearner: Geometric continuity feature learning for lane segmentation," *arXiv preprint arXiv:2108.03507*, vol. 2, 2021.

[24] H. Abualsaud, S. Liu, D. B. Lu, K. Situ, A. Rangesh, and M. M. Trivedi, "Laneaf: Robust multi-lane detection with affinity fields," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7477–7484, 2021.

[25] Q.-H. Che, D.-P. Nguyen, M.-Q. Pham, and D.-K. Lam, "Twinlitenet: An efficient and lightweight model for driveable area and lane segmentation in self-driving cars," *2023 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, pp. 1–6, 2023.

[26] C. Han, Q. Zhao, S. Zhang, Y. Chen, Z. Zhang, and J. Yuan, "Yolopv2: Better, faster, stronger for panoptic driving perception," *arXiv preprint arXiv:2208.11434*, vol. 5, 2022.

[27] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.

[28] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Efficient convnet for real-time semantic segmentation," *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1789–1794, 2017.

[29] S. Mane and R. K. Lodhi, "Cubic b-spline technique for numerical solution of singularly perturbed convection-diffusion equations with discontinuous source term," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp. 402–407, 2023.

[30] Z. Feng, S. Guo, X. Tan, K. Xu, M. Wang, and L. Ma, "Rethinking efficient lane detection via curve modeling," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17 062–17 070, 2022.

[31] K. Behrendt and R. Soussan, "Unsupervised labeled lane markers using maps," *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 832–839, 2019.

[32] H. Chen, M. Wang, and Y. Liu, "Bsnet: Lane detection via draw b-spline curves nearby," *arXiv preprint arXiv:2301.06910*, vol. 4, 2023.

[33] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2403–2412, 2018.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[35] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

[36] S. Yoo, H. S. Lee, H. Myeong, S. Yun, H. Park, and J. Cho, "End-to-end lane marker detection via row-wise classification," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1006–1007, 2020.

[37] J. Philion, "Fastdraw: Addressing the long tail of lane detection by adapting a sequential prediction network," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11 582–11 591, 2019.

[38] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," *Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXIV 16*, pp. 276–291, 2020.

[39] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Keep your eyes on the lane: Real-time attention-guided lane detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 294–302, 2021.

[40] J. Su, C. Chen, K. Zhang, J. Luo, X. Wei, and X. Wei, "Structure guided lane detection," *arXiv preprint arXiv:2105.05403*, vol. 13, 2021.

[41] Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz, "Key points estimation and point instance segmentation approach for lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8949–8958, 2021.