

Dense Hybrid Proposal Modulation for Lane Detection

Yuejian Wu, Linqing Zhao, Jiwen Lu, *Senior Member, IEEE*, and Haibin Yan, *Member, IEEE*

Abstract—In this paper, we present a dense hybrid proposal modulation (DHPM) method for lane detection. Most existing methods perform sparse supervision on a subset of high-scoring proposals, while other proposals fail to obtain effective shape and location guidance, resulting in poor overall quality. To address this, we densely modulate all proposals to generate topologically and spatially high-quality lane predictions with discriminative representations. Specifically, we first ensure that lane proposals are physically meaningful by applying single-lane shape and location constraints. Benefitting from the proposed proposal-to-label matching algorithm, we assign each proposal a target ground truth lane to efficiently learn from spatial layout priors. To enhance the generalization and model the inter-proposal relations, we diversify the shape difference of proposals matching the same ground-truth lane. In addition to the shape and location constraints, we design a quality-aware classification loss to adaptively supervise each positive proposal so that the discriminative power can be further boosted. Our DHPM achieves very competitive performances on four popular benchmark datasets. Moreover, we consistently outperform the baseline model on most metrics without introducing new parameters and reducing inference speed. The codes of our method are available at <https://github.com/wuyuej/DHPM>.

Index Terms—Lane Detection, High-quality Proposal, Proposal Modulation, Hybrid Constraints

I. INTRODUCTION

Given a front-viewed image taken by a camera mounted on the vehicle, lane detection aims to distinguish and locate the lane markings in an image. As a traditional yet fast-growing computer vision task, lane detection is one of the most essential and safety-critical components in autonomous driving and advanced driver-assistance systems (ADAS). It is developed with the desired purpose of avoiding traffic accidents and improving traffic efficiency. Thus it has drawn the increasing attention of researchers from academia and industry. However, a variety of challenging issues that may interfere with the detection have not been fully addressed yet. For example, many complex-shaped lane lines are difficult to fit well with existing methods. Moreover, lane lines may be inapparent or even invisible due to illumination conditions or due to occlusion by nearby vehicles. There has been a number of recent attempts to overcome these challenges.

This work was supported in part by the National Natural Science Foundation of China under Grant 61976023, Grant U22B2050, and Grant 62125603. (Corresponding author: Haibin Yan).

Yuejian Wu and Haibin Yan are with the School of Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail: wuyuejian@bupt.edu.cn; eyanhaibin@bupt.edu.cn).

Linqing Zhao and Jiwen Lu are with the Department of Automation, Tsinghua University, Beijing 100084, China, and also with the Beijing National Research Center for Information Science and Technology (BNRist), Beijing 100084, China (e-mail: linqingzhao@tju.edu.cn; lujiwen@tsinghua.edu.cn).

To overcome these challenges, a variety of approaches have been proposed to detect lane lines with complex topologies, which can be roughly categorized into segmentation-based methods [1–4], anchor-based methods [2, 5–11], and curve-based methods [12–14]. Unlike general semantic segmentation methods, **segmentation-based** lane detectors are expected to be able to discriminate instance-level lanes. The disadvantage of these detectors is that they cannot be directly transferred to predict more lanes without modifying the model. In addition, they usually require heavy post-processing or post-clustering strategies [15]. To avoid trading performance for low latency, **anchor-based** methods [2, 5, 6] estimate the line shape by regressing the relative key coordinates. For example, some methods [2, 6] construct line-shape anchors with vertical lines, while recent works [5] build a one-stage detection approach by designing straight-line anchors with arbitrary direction. However, these predefined anchors cannot describe complex line shapes, which results in relatively inferior performance. To make good use of the shape priors, recent methods adopt row-wise anchors [7–11] and predict the line location for each row. Different from points regression, **curve-based** methods [12–14] formulate the lane curve with parameters and detect lanes by regressing these parameters. LSTR [13] and BézierLaneNet [14] predict the fitting parameters of each lane line, which can provide more proposals to improve the migration ability. Curve-based methods achieve high efficiency and superior performance benefitting from the end-to-end frameworks. However, curve-based methods are sensitive to the predicted parameters because the high-order coefficient may directly cause shape changes of lanes. As a brief summary, we attribute the unsatisfactory performance of existing methods to two fundamental but intractable challenges.

First, how to increase the number of TP (i.e., true positive) predictions? Many methods [1, 3, 7, 11, 16, 17] choose to predict a small number of proposals, which equals the maximum number of annotations of a single image. This indicates that the representations of these proposals must have enough generalization performance to adapt to various situations. However, there exist many complex-shaped lane lines in real scenes, even including some invisible lanes, which limit the overall detection performance. Therefore, we believe it is necessary to produce more proposals to deal with varied situations, which have been widely adopted by many object detection models [18, 19]. Nevertheless, existing methods [13, 14] only supervise a subset of proposals with high classification scores. In other words, these methods typically give no clear location or shape supervision to the low-scoring proposals, let alone the underlying relations between

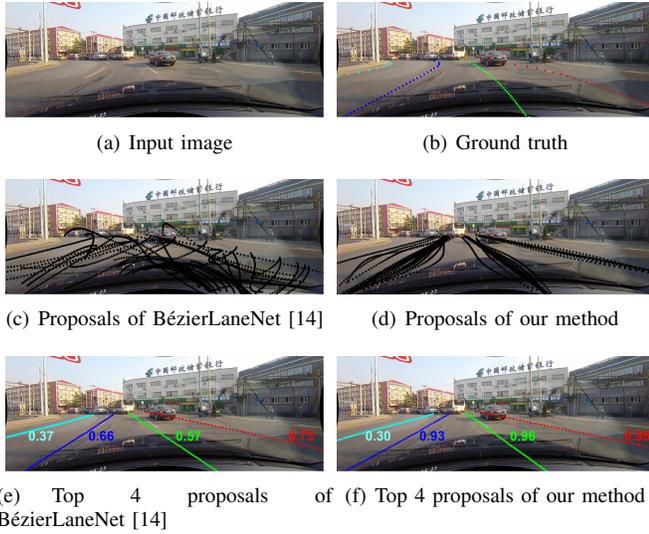


Fig. 1. The comparison between our DHPM and BézierLaneNet [14]. Compared with BézierLaneNet [14], our DHPM can produce topologically and spatially high-quality lane proposals. In addition, the confidence gap between high-quality and low-quality proposals is enlarged by our proposed quality-aware constraints, where a threshold can be easily found to distinguish between true and false predictions with less effort.

proposals. For example, we show the 50 proposals predicted by the top-performing BézierLaneNet [14] in Fig. 1 (c), where each proposal is represented as 100 key points. We observe that a vast majority of proposals are with unreasonable shapes or improper locations, which indicates that these proposals have a very low probability of being optimized during training, and these unreasonable proposals have no potentials to fit lane lines at the test phase. That means a small number of proposals will be supervised to fit all lanes, which is very challenging for the learned proposal representation. As a consequence, even the well-trained proposals cannot handle challenging samples and produce very confident lane proposals at the test phase, as shown in Fig. 1 (e). This observation motivates us to densely modulate all lane proposals w.r.t. ground truth labels since the increasing number of high-quality proposals are beneficial to raise the number of TP predictions.

Second, how to decrease the amount of FP (i.e., false positive) and FN (i.e., false negative) predictions? Most lane detection methods [1, 3, 7, 13, 14] restrain the proposal quality (shape and location accuracy) and instance-level discrimination individually. To be specific, they typically use a standard binary cross-entropy or multi-class cross-entropy loss as the discrimination constraint. However, these methods will inevitably produce some proposals with high scores but inaccurate shapes or locations, resulting in harmful effects on the discriminative power. To enhance the instance-level discrimination and decrease false classification, we propose to adaptively assign quality-aware classification labels to all proposals, which aims to give relatively lower classification labels to a low-quality proposal and vice versa.

To address these, we propose a dense hybrid proposal modulation method to generate topologically and locational high-quality lane proposals with discriminative representations. Specifically, our hybrid modulation can be divided into three aspects: 1) availability constraint, 2) diversity constraint, and

3) discrimination constraint. As mentioned above, only a tiny fraction of lane proposals are sufficiently supervised in shape and location, which leads to inaccurate predictions and poor generalization ability of most existing methods. Therefore, our prime objective is to make each proposal appear around the ground truth lanes with a reasonable shape. To avoid over-curved proposals in Fig. 1 (c), we define a simplified curvature metric to measure and control the degree of bending of curves. Then we force each proposal to approach its specific locating target assigned by our proposal-to-label matching algorithm. In addition to availability constraints, we encourage topological diversification of the proposals targeting the same ground-truth lanes, aiming to improve the fitting ability in the training phase and the generalization ability in the testing phase, respectively. To achieve this, we efficiently measure and regulate intra-cluster shape differences by extending our single-curve metric into a pairwise one (see Fig. 1 (d)). Last but not least, we propose to enhance the instance-level discrimination of proposals to reduce the number of false predictions. Instead of a standard cross-entropy loss, we build a quality-aware classification objective by explicitly taking shape and location accuracy into consideration. With this quality-aware design, our method can redistribute the probability distribution of existence classification, where a threshold can be easily found to classify all proposals into high- and low-quality groups. Fig. 1 (f) shows that our method tends to give relatively higher scores to good proposals while the score gap between good and bad proposals is enlarged. Since our method modulates the proposals with three constraints without modifying the network structure, it only increases the time of back-propagation, while the forward propagation remains unchanged. Our DHPM achieves very competitive performance without introducing new parameters and reducing inference speed on four widely-used datasets.

The main contributions can be summarized as follows:

- 1) Unlike existing lane detection methods, which either predict a few proposals or sparsely supervise a small fraction of proposals, we propose a dense hybrid proposal modulation method to improve the overall quality of proposals.
- 2) We first present the availability constraint to locate each proposal around the ground truth lanes with a reasonable shape. Then we enhance the inter-proposal topological diversification to exploit the underlying relations between proposals with our diversity constraint.
- 3) We propose a quality-aware discrimination constraint to build a bridge between the proposal quality (including topological and locational accuracy) and instance-level discrimination.
- 4) Experimental results on four popular datasets demonstrate that our DHPM achieves very competitive performance without introducing new parameters, and cross-dataset experiments show that DHPM significantly improves the generalization ability of proposals compared to the baseline.

The remainder of this paper is organized as follows: Section II describes recent studies related to our work. Section III

introduces the crucial components of our DHPM. Extensive experimental results on various datasets are reported in Section IV, and the conclusion of this paper is in Section V.

II. RELATED WORK

In this section, we briefly review three related topics, including lane detection, high-quality proposals learning, and proposal relations learning.

A. Lane Detection Strategies

1) *Segmentation-based Methods*: Segmentation-based methods [1, 2, 10, 16, 17, 20–29] view lane detection as a per-pixel classification task. To handle the classification problem of lane line points, some prior arts [28] treat each lane line as a category for segmentation. In some methods, pixels are classified as either on lane or background to generate a binary segmentation mask. LaneNet [2] and LaneAF [24] predict the instance segmentation mask, and use post-clustering strategy [15] to determine the final lane instance. However, segmentation-based methods are limited by a predefined and fixed number of lanes, which is not robust to real driving scenarios where the number of lanes is unknown.

2) *Anchor-based Methods*: Anchor-based Methods [5–7, 11, 30–35] use the predefined anchor to help describe the lane line. Instead of predicting lane line points pixel by pixel, UFAST [7] manages to split the images into row-wise anchors, where the number of anchors is much smaller than the pixels, so it can achieve a good trade-off between performance and efficiency. To minimize the gap between anchor-based prediction and ground truth, some works [5, 6, 11, 30] also predict an offset map to refine the initial prediction. LaneATT [5] designs a line-shape anchor according to the slender and monotonous characteristics of the lane line, which is followed by many recent anchor-based methods [35]. CLRNet [32] leverages the high-level semantic features and low-level texture features to refine the proposal representations. However, the fixed anchor shape results in a low degree of freedom in describing the complex topology.

3) *Curve-based Methods*: Curve-based methods [12–14, 36, 37] directly output parametric lines expressed by curve equation. PolyLaneNet [12] fits all lane lines into polynomial functions, and directly predicts the fitting curve parameters through a deep network. LSTR [13] introduces the transformer with a similar structure to DETR [18] into the curve-based method, and redesigns the fitting curve to take the camera information into account. Different from segmentation-based methods and anchor-based methods, curve-based methods can avoid the separation between the predicted adjacent lane line points. Although the accuracy of curve-based methods still lags behind that of other methods, they have been proved [38] to be more robust due to the smoothness of the fitting curve.

B. High-Quality Proposals Learning

In the field of object detection, many methods have made progress in generating high-quality bounding box proposals.

Cai *et al.* [39] prove that high-quality proposals can bring high-quality detector. Some works improve the quality of proposal through ingeniously region proposal network design, for example, FPN [40] uses both high-resolution low-level features and high semantic information of high-level features to improve the quality of multi-scale proposals. Besides, LocNet [41] iteratively recognizes and locates the proposal boxes to generate higher-quality proposal boxes. Cascade RCNN [39] gradually improves the quality of proposal boxes through a designed cascade regressor, in which each regressor will optimize the output proposals according to the previous regressor. In weakly supervised object detection tasks, PG-PS [42] combines selective search and Grad-Cam [43] to perform coarse classification and fine classification in turn to obtain high-quality proposal boxes. Recently, many lane detection works [13, 14] predict proposals that are significantly larger than the number of lane lines. However, they only supervised a subset of high-scoring proposals, ignoring the overall quality. In this paper, we propose a dense hybrid modulation mechanism to improve the overall quality of proposals.

C. Proposal Relations Modeling

Modeling proposal relations has emerged in many fields, including 2D [44–46] and 3D object detection [47]. In the field of action localization, P-GCN [44] constructs a graph of proposals by establishing the edges and then applying GCNs to do message aggregation among proposals. For 3D object detection, Feng *et al.* [47] extract uniform appearance features for each 3D object proposal and construct a relation graph that exploits the 3D object-object relationships. In order to explore and utilize object-to-object relations, RDN [45] assembles and propagates object relations to enhance object features for video object detection. Despite the rapid development of modeling the proposal relations of 2D or 3D bounding boxes, the relations between lane proposals have not been well exploited yet. In our work, we aim to mine the location and score relations between proposals to generate diverse and discriminative proposals.

III. THE PROPOSED APPROACH

In this section, we first formulate the problem of lane detection and then describe our main idea. At last, we detail the overall objective function for lane detection.

A. Overview

Given an input image $I \in \mathbb{R}^{H \times W \times 3}$ with the corresponding ground truth $G = \{g_1, g_2, \dots, g_M\}$, our goal is to predict a collection of lane proposals $P = \{p_1, p_2, \dots, p_K\}$, and predict a confidence value for each lane lines $L = \{l_1, l_2, \dots, l_K\}$, where K is the total number of lane lines. Generally, a lane proposal p_k is represented by an ordered set of coordinates $p_k = \{(x_1^{(k)}, y_1^{(k)}), (x_2^{(k)}, y_2^{(k)}), \dots, (x_N^{(k)}, y_N^{(k)})\}$, where k is the index of proposal lane line and N is the max number of sampled points. Similarly, each ground truth lane line g_j is also represented as N points $g_j =$

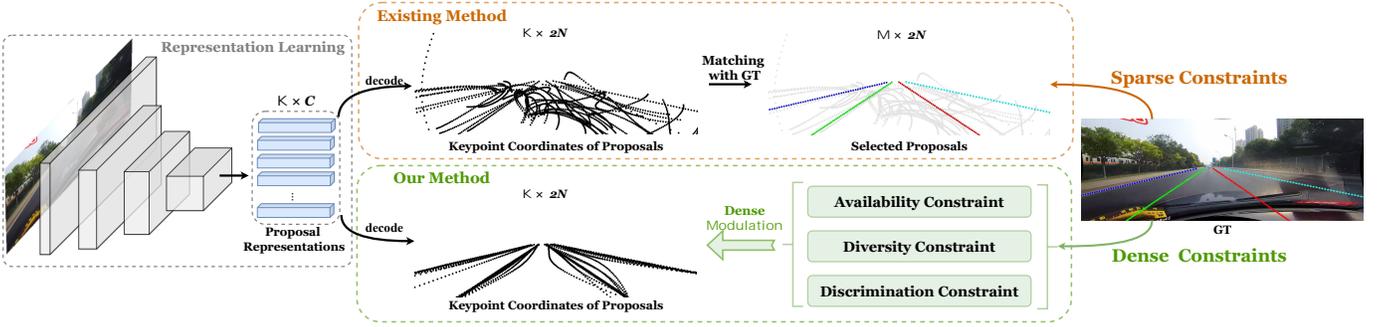


Fig. 2. Comparison of our overall framework with existing methods. Unlike existing methods that sparsely supervise a subset of proposals with high classification scores, we densely modulate all proposals to generate topologically and spatially high-quality lane proposals with three constraints, i.e., availability, diversity, and discrimination.

$\{(u_1^{(j)}, v_1^{(j)}), (u_2^{(j)}, v_2^{(j)}), \dots, (u_N^{(j)}, v_N^{(j)})\}$, where j is the index of ground truth lane line. The objective of a general curve-based method can be divided into three terms: 1) a coordinate regression term J_{reg} ; 2) an existence classification term J_{cls} ; and 3) an optional binary mask segmentation term J_{seg} . Therefore, the overall loss can be written as a weighted sum of all terms:

$$J_{sparse} = \lambda_1 J_{reg} + \lambda_2 J_{cls} + \lambda_3 J_{seg}, \quad (1)$$

where J_{reg} can be ℓ_1 , ℓ_2 , smooth ℓ_1 and other regression-based loss. Note that existing methods [13, 14] only apply J_{reg} to the positive proposals, which depend on the number of ground truth lanes M . To achieve this, they perform a one-to-one assignment between M labels and K predictions ($M < K$) using optimal bipartite matching [18]. Assuming that p_k is a positive proposal and its corresponding ground truth lane is g_j , we can represent the ℓ_1 -based regression loss J_{reg} as:

$$J_{reg} = \frac{1}{2N} \sum_{i=1}^N \|y_i^{(k)} - v_i^{(j)}\| + \|x_i^{(k)} - u_i^{(j)}\|. \quad (2)$$

Different from J_{reg} , J_{cls} and J_{seg} in a general curve-based method are calculated for each proposal, so we omit the proposal index for clarity. For example, J_{cls} can be written as follows:

$$J_{cls} = -l \log(p) - w(1-l) \log(1-p), \quad (3)$$

where l and p represent the existing label and predicted logits of a proposal, respectively, and w is a hyper-parameter for weighting negative samples. In addition, the binary segmentation loss J_{seg} shares the same format with J_{cls} .

While existing curve-based methods have made significant progress by extending the general model introduced above, their performance is suppressed by the following limitations:

- 1) **Overall poor rationality of lane proposals.** Optimal bipartite matching algorithms such as the Hungarian algorithm [48] only select a subset of the high-scoring lane proposals during training. Due to insufficient topological and locational supervision, most proposals become very curly and have a very chaotic layout.
- 2) **Lack of relations mining between proposals.** Existing methods independently supervise each matching proposal without considering the shape and location of other proposals. Therefore, there may exist some highly similar

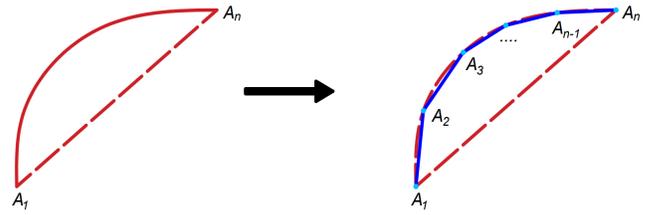


Fig. 3. Illustration of the shape loss S_k . The length of the red dotted line represents the straight-line distance between the starting point and the ending point, and the solid red line represents the curve. Referring to the calculus idea, we use the total length of the blue dashed line to approximate the length of the curve with dense sampling.

lane proposals, which have a detrimental effect on the generalization ability.

- 3) **Weak discrimination of classification scores.** During training, traditional detection methods [49–53] directly assign labels 1 to matching proposals and 0 to others. Recent lane detection methods continue to use such assignments, regardless of the quality (topological and locational accuracy) of matching proposals. This will inevitably generate some proposals with high confidence but inaccurate shapes or locations, increasing the difficulty of finding a suitable threshold.

To make lane proposals topologically and locational accurate and discriminative, we enforce three important criteria to modulate these proposal descriptors:

- 1) **The modulated lane proposals are topologically and spatially sound.** Realistic lane lines usually have strong prior knowledge, e.g., the ending points of different lane lines are usually close to each other, while the starting points are far apart. Therefore, forcing them to be reasonable can increase the number of true positive predictions.
- 2) **The modulated proposals are with good generalization ability.** There are a few lane lines with complex topology, including bumps and grooves, and some lane lines are even invisible. We need to ensure that there are sufficient and diverse proposals near ground-truth lanes as alternative predictions.
- 3) **The modulated proposal confidences are closely linked to the quality of proposals.** Since binary cross-entropy loss does not guarantee higher classification labels for

higher-quality proposals, it is important to build a bridge between the quality of proposals and the classification labels. By doing this, we can determine a threshold in the testing phase to distinguish between true and false predictions with less human effort.

To achieve these objectives, we present DHPM method to generate topologically and spatially high-quality lane proposals with discriminative representations, which is illustrated in Fig. 2. Without introducing new parameters, our DHPM achieves very competitive performance with three simple yet effective constraints. A light weight version of our method could achieve 200+ FPS with a higher performance over the baseline [14].

B. Availability Constraint

The goal of availability constraint is to make all proposal lane lines reasonable. Specifically, we separate the objective into two aspects. On the one hand, the topology of a lane proposal should not be too much complicated than a straight line, which is based on the structure priors in the driving scenarios. Therefore, we hope to find a way to measure and control the curvature of a lane curve efficiently. On the other hand, the lane proposals are expected to locate near the ground truth lanes evenly. At the same time, the endpoints and starting points should follow the layout priors in real scenarios. Thus, we propose to assign a label lane for each proposal lane to give a specific regression target. With such two constraints, we can make sure that the modulated proposals are topologically and spatially sound. Note that we densely apply these two constraints to each lane proposal, which is fundamentally different from existing sparse constraint methods. To sum up, we can formulate the full availability constraint as follows:

$$\begin{aligned} J_{ava} &= \lambda_{shape} J_{shape} + \lambda_{loc} J_{loc} \\ &= \lambda_{shape} \frac{1}{K} \sum_{k=1}^K S_k + \lambda_{loc} \frac{1}{K} \sum_{k=1}^K D_k, \end{aligned} \quad (4)$$

where λ_{shape} and λ_{loc} are weighting coefficients. S_k is the shape loss of k -th proposal, while D_k is the location loss.

In general, existing methods represent a lane proposal as an ordered set of coordinates. Therefore, the shape and location representations of a proposal are deeply coupled. We cannot directly apply our constraints on the coordinates because a coordinate perturbation will change the shape and location simultaneously. To disentangle shape attributes from coordinates, our first step is to densely sample n key points from a lane proposal, and we can re-write a lane representation as $\widehat{A_1 A_n} = \{A_1, A_2, \dots, A_{n-1}, A_n\}$, where we denote A_1 and A_n as the start and end points, respectively. Based on these dense points, we can approximate the curve length with the length of multiple line segments when $n \rightarrow \infty$. We present a straightforward and parallel-friendly way to efficiently measure the curvature of lane proposals. That is, the ratio of the length of the curve $|\widehat{A_1 A_n}|$ to the length of the straight line $|A_1 A_n|$, as shown in the Fig. 3. Mathematically,



(a) Proposals without J_{ava}



(b) Proposals with only J_{loc}



(c) Proposals with only J_{shape}



(d) Proposals with full J_{ava}

Fig. 4. Visual comparisons between different terms in the availability constraint. Compared to proposals without our J_{ava} (a), applying J_{loc} (b) or J_{shape} (c) alone can slightly improve the overall quality, still leaving some unreasonable proposals. Differently, the full availability constraint that includes both shape and location supervision significantly improves the rationality of proposals. The image is from the test split of CULane [1].

we can represent the simplified curvature term as:

$$\begin{aligned} S_k &= \frac{|\widehat{A_1 A_n}|}{|A_1 A_n|} \approx \frac{|A_1 A_2| + |A_2 A_3| + \dots + |A_{n-1} A_n|}{|A_1 A_n|} \\ &= \frac{\sum_{i=1}^{n-1} \sqrt{(x_{i+1}^{(k)} - x_i^{(k)})^2 + (y_{i+1}^{(k)} - y_i^{(k)})^2}}{\sqrt{(x_n^{(k)} - x_1^{(k)})^2 + (y_n^{(k)} - y_1^{(k)})^2}}, \end{aligned} \quad (5)$$

where the minimum value of S_k is 1, which is only satisfied when the lane proposal is a strictly straight line.

In addition to the shape properties, the spatial arrangement of the proposal should be explicitly guided by the layout prior. We propose to assign a target from ground truth lanes to each proposal so that the layout prior can be propagated to the lane proposals. To this end, we design a proposal-to-label matching algorithm to assign M ground truth lanes to K proposals. Instead of a slow clustering-based method, we only need to compute $M \times K$ ℓ_1 distances, from which each proposal will choose the closest ground truth lane as its target. Therefore, the ground truth lane that matches the k -th proposal can be

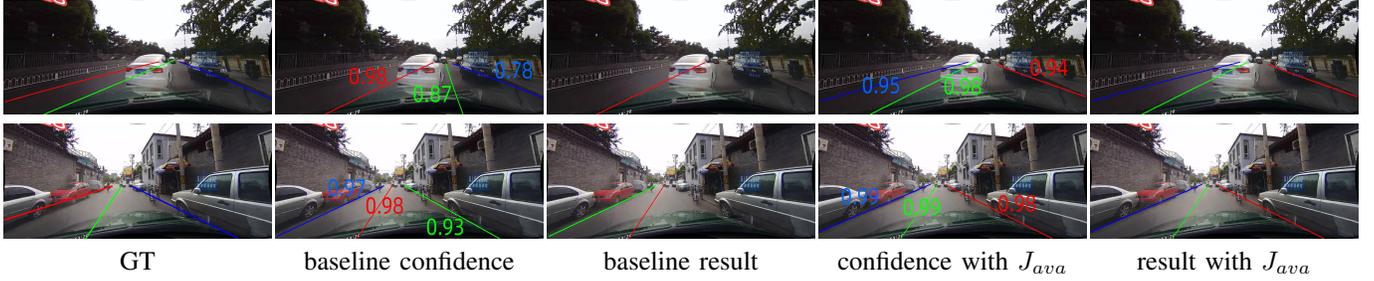


Fig. 5. Proposal comparisons between the baseline and the model with only J_{ava} . Our J_{ava} is used to increase the high-quality proposals (potential True Positive results), and reduce the bad proposals (useless predictions) at the same time.

computed as follows:

$$g^{(k)} = \arg \min_{g_i \in G} Q_{(i,k)}, \quad Q_{(i,k)} = 1 - L_1(g_i, p_k), \quad (6)$$

where $Q_{(i,k)}$ represents the distance between the i -th ground truth lane line with the k -th proposal. Note that all coordinates are normalized according to the image resolution. Having obtained the matched ground-truth lanes, we formulate the dense location constraints to all proposals as follows:

$$\begin{aligned} D_k &= D_s^{(k)} + D_e^{(k)} \\ &= (|x_1^{(k)} - u_1^{(g^{(k)})}| + |y_1^{(k)} - v_1^{(g^{(k)})}|) \\ &\quad + (|x_n^{(k)} - u_n^{(g^{(k)})}| + |y_n^{(k)} - v_n^{(g^{(k)})}|), \end{aligned} \quad (7)$$

where D_s and D_e represent the ℓ_1 distances of starting points and ending points, respectively. For some samples that have no ground truth lanes, we simply omit the location loss while the shape loss is reserved. It's worth noting that the proposed availability constraint is not the only constraint that supervises the location of proposals. The curve loss J_{reg} in basic constraints also provides sparse locational supervision to a subset of proposals. Differently, our location constraints supervise all proposals with clear locational targets.

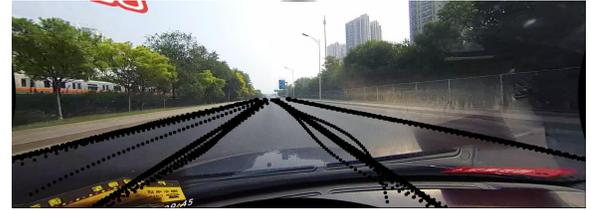
To illustrate the effect of each individual component of J_{ava} , we provide a detailed visual comparison in Fig. 4. J_{loc} or J_{shape} cannot solely improve the proposals since they are completely decoupled components. On the contrary, the full J_{ava} can effectively modulate most low-quality proposals to be topologically and spatially plausible. To sum up, our location constraint supervises the approximate location of the lane head and tail, while the shape constraint determines the lane neck. Therefore, it can be concluded that it is the combination of them that makes the lane have a reasonable location. Besides, Fig. 5 shows the confidence comparisons between the baseline and the model with only J_{ava} , which proves that J_{ava} successfully increases the high-quality proposals (potential True Positive results), and reduce the bad proposals (useless predictions) at the same time. Therefore, the proposals that exceed the threshold will have a greater probability of becoming a TP rather than an FP prediction, which proves that J_{ava} can directly improve the detection performance.

C. Diversity Constraint

Having obtained reasonable proposals by applying the availability constraint, we find that the proposals matching the same



(a) Great shape difference



(b) Small shape difference

Fig. 6. The comparison between great and small differences between proposals. Compared to (b), greater shape difference (a) can enhance the generalization ability of proposals by reducing the mutual information between proposals.

ground truth have a similar shape. The potential disadvantage can be summarized in the following two aspects. 1) The high similarity will reduce the generalization ability of the model since the unique information of each proposal is limited. 2) The discrimination of classification confidences will be suppressed because similar proposals may be given different binary labels.

To enhance the generalization ability of proposals and further model the underlying relations between proposals, we propose to reduce the mutual information between proposal lane lines. The mutual information between p_i and p_j can be written as:

$$I(p_i, p_j) = I(p_i) - I(p_i|p_j), \quad (8)$$

where $I(p_i)$ means the information uncertainty of p_i , $I(p_i|p_j)$ means the uncertainty of p_i , given p_j has already appeared, and they are calculated as follows:

$$I(p_i) = \log \frac{1}{P(p_i)}, \quad I(p_i|p_j) = \log \frac{1}{P(p_i|p_j)}, \quad (9)$$

where we know the mutual information has a strong link with the probability between p_i and p_j .

As shown in Fig. 6 (b), availability constraints lead to the high similarity between proposals and large mutual information values I . There are some overlapping proposals, which means that their representations are very similar to each other.



(a) with the upper limit



(b) without the upper limit

Fig. 7. The comparison between the diversity constraint with and without an upper limit. The upper limit aims to make the distribution of lines more uniform, as shown in (a).

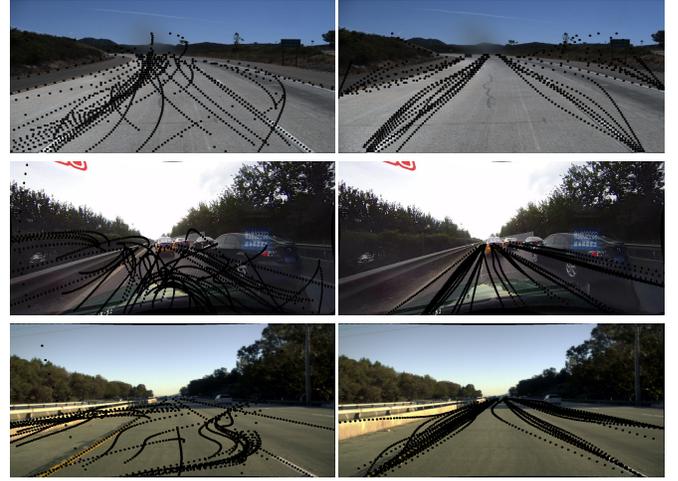
Therefore, it is important to make each proposal unique since it can help generate more effective proposals to adapt to different situations. Unique proposals are able to better fit lane lines in their unique scenarios, thus increasing TP and reducing FP. However, directly using the distance between every two lane representations as mutual information makes it difficult to control the shape of decoded proposals. Therefore, according to Equation (8) and (9), we hope to reduce the mutual information between proposals by 1) enlarging the shape differences of proposals; 2) redistributing the probabilities of proposals. Although both the availability constraint and the diversity constraint aim to modulate the shape and location of the lanes, there is an essential difference between the two. The availability constraint J_{ava} aims to supervise the location and shape attributes of a single lane proposal, while the diversity constraint J_{div} is proposed to restrain the relationship of multiple lane proposals. In this subsection, we will elaborate on how to enlarge the shape differences of proposals.

To enlarge the shape differences, we propose the diversity constraint to diversify the shapes of different proposal lane lines. Specifically, we first build a symmetric matrix with $K \times K$ elements, where each element represents a pairwise shape difference. A simple way to measure the shape difference is to calculate the curvature difference between different proposals. Therefore, we can raise the shape diversity by enlarging the mean curvature difference:

$$J_{div} = -\frac{2}{K(K-1)} \sum_{k=1}^K \sum_{i=k}^K |S_k - S_i|, \quad (10)$$

where we only count the elements in the upper triangular matrix, and the diagonal elements are all zero.

However, forcing all proposals to have different shapes may have strong side effects since proposals matching different ground-truth lanes should not be computed together. To solve this issue, we instead compute the intra-cluster shape differences, i.e., we only calculate the shape difference between



BézierLaneNet [14]

ours

Fig. 8. Proposals comparison between BézierLaneNet [14] and ours. From top to bottom, the images are from TuSimple [54], CULane [1], and LLA-MAS [55] dataset.

proposals that match the same ground truth according to Equation (6), and improve Equation (10) as follows:

$$J_{div} = -\frac{1}{K'} \sum_{g^{(k)}=g^{(i)}} |S_k - S_i|, \quad (11)$$

where $\text{card}(\{(k, i) | g^{(k)} = g^{(i)}\}) = K'$. To apply the intra-cluster diversity constraint for all groups, we generate a binary mask with $K \times K$ elements, of which only K' elements are 1 and 0 otherwise. To illustrate the effect of J_{div} , we show the visual comparison between great and small shape differences in Fig. 6.

While J_{div} achieves a promising overall arrangement of proposals, a few samples whose ground truth lanes are less than 3 will lead to unstable training. Specifically, few ground truth lanes will cause a large number of proposals to come together. In other words, one ground truth will match multiple proposal lane lines. Therefore, we limit the number of proposed lane lines that match a single ground truth. For example, when there are only two ground truth lane lines in Fig. 7 (b), all proposals will be forced to locate near these two lane lines by Equation (7). Setting a higher number will effectively solve this problem without degrading performance. We conduct J_{loc} and J_{div} only on the proposal lane lines in set $\{p_1, p_2, \dots, p_U | (p_1, p_2, \dots) = \text{sort}(Q_{(i,k)}), g^{(k)} = i\}$. The comparison between the proposals with and without an upper limit is shown in Figure 7. It can be concluded that when there is an upper limit, all proposed lane lines can still appear in reasonable locations. To avoid the generalization problem that may be caused by manually setting the upper limit, we set the upper limit value according to the number of proposals.

D. Discrimination Constraint

As mentioned above, existing methods separate confidence loss from proposal quality. Optimal binary matching algorithms such as the Hungarian algorithm [48] only match a relatively close proposal lane line for each ground truth in the

TABLE I
DETAILS OF DATASETS

| Dataset | #Total | Train | Validation | Test | Resolution | #Lines | #Scenarios | Environment |
|----------------|--------|--------|------------|-------|--------------------|--------|------------|-------------------|
| TuSimple [54] | 6408 | 3268 | 358 | 2782 | 1280×720 | ≤ 5 | 1 | highway |
| CULane [1] | 133235 | 88880 | 9675 | 34680 | 1640×590 | ≤ 4 | 9 | urban and highway |
| LLAMAS [55] | 100042 | 58269 | 20844 | 20929 | 1276×717 | ≤ 4 | 1 | highway |
| CurveLanes [6] | 150000 | 100000 | 20000 | 30000 | 2560×1440&1570×660 | ≤ 14 | 1 | urban and highway |

image, regardless of the distance between the matched proposal and the ground truth. Matching proposals will have the classification label set to 1, and the rest of the proposals will be assigned 0. This training strategy inevitably produces some high-confidence but low-quality proposals, which eventually become FP. During the testing phase, a proposal with a lower confidence value will be considered a negative prediction, even if it is close to the ground truth. In general, existing methods handle all positive proposals in a fixed manner, i.e., enforcing their confidence level to 1 while ignoring topological and spatial quality.

To address this, we reform the binary cross-entropy loss to boost the discriminative power. During training, different from existing methods that set each matched proposal confidence label to 1, we generate lane-aware ground truths as follows:

$$J_{dis} = -(l' \log(p) + w(1 - l') \log(1 - p)), \quad (12)$$

$$l' = \begin{cases} \gamma + (1 - \gamma) \times e^{-\beta J_{reg}}, & \text{if } l = 1, \\ 0, & \text{if } l = 0, \end{cases} \quad (13)$$

where l means the original existence label, and l' means the new confidence label. J_{reg} is the regression loss of selected proposal, which is set to be a constant variable. β and γ are used to balance J_{reg} , which are set to 10 and 0.5, respectively. When all proposals are insufficient to match the ground truth, this discriminative constraint will give lower confidence to the closest proposal. As shown in Fig. 1(e) and (f), due to the discrimination constraints during training, the three proposals on the right get higher confidence, while the left-most proposal gets lower confidence since their shape and location are not good enough. It proves that our discrimination constraint can effectively help screen out close but unqualified proposals, thereby effectively reducing FP. More importantly, we can determine a threshold in the testing phase to distinguish between true and false predictions with less human effort. Experimental results in Section IV show that J_{dis} successfully improved the performance from the perspective of the confidence distribution of proposals.

E. Overall Loss

Different from the general curve-based models [13, 14], we add the availability constraint, diversity constraint, and discrimination constraint together to modulate the quality of proposals. Our overall loss is a weighted sum of five losses:

$$J_{dense} = \lambda_1 J_{reg} + \lambda_2 J_{seg} + \lambda_3 J_{ava} + \lambda_4 J_{div} + \lambda_5 J_{dis}, \quad (14)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5$ are set to 1, 0.1, 0.0005, 0.0001, 0.75, respectively. These coefficients are determined by performing

5-fold cross-validation. Compared with the sparse constraints in existing methods, our dense hybrid modulation significantly improves lane proposals' topological and spatial qualities without introducing new parameters. We compare proposal quality with our baseline BézierLaneNet in Fig. 8, and the comparison shows that our proposals outperform the baselines on all datasets.

IV. EXPERIMENTS

In this section, we conduct various experiments to evaluate the proposed dense proposal modulation method. We perform an ablation study to show the effectiveness of each design through qualitative and quantitative analysis.

A. Experimental Setting

1) **Datasets:** To evaluate our methods, we conduct experiments on four commonly used benchmark datasets, including TuSimple [54], CULane [1], LLAMAS [55] and CurveLanes [6] datasets. TuSimple [54] is collected on the highway in the daytime with fair weather conditions. CULane [1] is collected by cameras installed on six different vehicles in Beijing, with 9 different scenarios such as night, crowded, and dazzle. LLAMAS [55] is a newly formed large dataset with over 100k annotated images, which is collected using Lidar maps. CurveLanes is a recently released dataset whose driving scenarios are full of complex topologies such as curve lines and dense lines. Detailed information about the datasets can be found in Table I.

2) **Evaluation Metrics:** For the TuSimple [54] dataset, there are three official evaluation metrics: accuracy, false positive rate (FPR), and false negative rate (FNR). The accuracy is computed by:

$$accuracy = \frac{\sum_{clip} C_{clip}}{\sum_{clip} S_{clip}}, \quad (15)$$

where C_{clip} is the number of correctly predicted lane line points, and S_{clip} is the total number of ground truth points in each clip. Different from the TuSimple [54] dataset, the official evaluation metric for CULane [1], LLAMAS [55] and CurveLanes [6] datasets is F1 score, which is calculated as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (16)$$

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}. \quad (17)$$

The prediction is considered true positive (TP) when the intersection-over-union (IoU) between the prediction and ground truth exceeds 0.5.

TABLE II
QUANTITATIVE RESULTS ON THE CULANE [1] DATASET

| Method | Total | Normal | Crowded | Dazzle | Shadow | No line | Arrow | Curve | Crossroad ‡↓ | Night |
|---------------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Segmentation-based Method | | | | | | | | | | |
| SCNN (LargeFOV) [1] | 71.60 | 90.60 | 69.70 | 58.50 | 66.90 | 43.40 | 84.10 | 64.40 | 1990 | 66.10 |
| SAD (ENet) [16] | 70.80 | 90.10 | 68.80 | 60.20 | 65.90 | 41.60 | 84.00 | 65.70 | 1998 | 66.00 |
| SIM-CycleGAN (ERFNet) [56] | 73.90 | 91.80 | 71.80 | 66.40 | 76.20 | 46.10 | 87.80 | 67.10 | 2346 | 69.40 |
| RESA (ResNet-34) [3] | 74.50 | 91.90 | 72.40 | 66.50 | 72.00 | 46.30 | 88.10 | 68.60 | 1896 | 69.80 |
| Anchor-based Methods | | | | | | | | | | |
| UFLD (ResNet-18) [7] | 68.40 | 87.70 | 66.00 | 58.40 | 62.80 | 40.20 | 81.00 | 57.90 | 1743 | 62.10 |
| UFLD (ResNet-34) [7] | 72.30 | 90.70 | 70.20 | 59.50 | 69.30 | 44.40 | 85.70 | 69.50 | 2037 | 66.70 |
| CurveLanes-NAS(ResNet-18) [6] | 71.40 | 88.30 | 68.60 | 63.20 | 68.00 | 47.90 | 82.50 | 66.00 | 2817 | 66.20 |
| CurveLanes-NAS (ResNet-34) [6] | 73.50 | 90.20 | 70.50 | 65.90 | 69.30 | 48.80 | 85.70 | 67.50 | 2359 | 68.20 |
| UFLDv2 (ResNet-18) [8] | 74.70 | 91.70 | 73.00 | 64.60 | 74.70 | 47.20 | 87.60 | 68.70 | 1998 | 70.20 |
| UFLDv2 (ResNet-34) [8] | 75.90 | 92.50 | 74.90 | 65.70 | 75.30 | 49.00 | 88.50 | 70.20 | 1864 | 70.60 |
| CondLaneNet (ResNet-18) [11] | 78.14 | 92.87 | 75.79 | 70.72 | 80.01 | 52.39 | 89.37 | 72.40 | 1364 | 73.23 |
| CLRNet (ResNet-18) [32] | 79.58 | 93.30 | 78.33 | 73.71 | 79.66 | 53.14 | 90.25 | 71.56 | 1321 | 75.11 |
| LaneATT (ResNet-18) [5] | 75.13 | 91.17 | 72.71 | 65.82 | 68.03 | 49.13 | 87.82 | 63.75 | 1020 | 68.58 |
| ours-LaneATT (ResNet-18) | 75.42 | 91.12 | 72.88 | 66.49 | 70.61 | 48.74 | 86.84 | 65.29 | 979 | 69.73 |
| Curve-based Methods | | | | | | | | | | |
| LSTR(ResNet-18)* [13] | 68.72 | 86.78 | 67.34 | 56.63 | 59.82 | 40.10 | 78.66 | 56.64 | 1166 | 59.92 |
| BézierLaneNet (ResNet-18) [14] | 73.67 | 90.22 | 71.55 | 62.49 | 70.91 | 45.30 | 84.09 | 58.98 | 996 | 68.70 |
| BézierLaneNet (ResNet-34) [14] | 75.57 | 91.59 | 73.20 | 69.20 | 76.74 | 48.05 | 87.16 | 62.45 | 888 | 69.90 |
| ours-BézierLaneNet (ResNet-18) | 74.59 | 90.55 | 73.25 | 63.49 | 65.53 | 45.11 | 84.83 | 60.19 | 701 | 68.94 |
| ours-BézierLaneNet (ResNet-34) | 76.21 | 91.50 | 74.19 | 68.11 | 74.85 | 48.25 | 87.38 | 60.41 | 872 | 71.51 |

*Since the result of CULane [1] is not given in the LSTR [13] paper, we used the results in the paper of BézierLaneNet [14] here.

‡For the ‘‘Cross’’ category, only false positives are shown.

3) **Implementation Details:** We use the original architecture of BézierLaneNet [14] without changing any architecture, and we apply our DHPM with the three constraints mentioned in Section III during the training phase. The batch size is set to be 20 for all datasets, and we train for 400, 36, 20, and 36 epochs for TuSimple [54], CULane [1], LLAMAS [55] and CurveLanes [6], respectively. The input images are resized to 360×640 for TuSimple [54] and LLAMAS [55], 720×1280 for CurveLanes [6], while 288×800 for CULane [1]. Given the number of proposals K , the upper limits in our diversity constraint were set to $U = \lfloor \frac{4K}{25} \rfloor$. We use Adam optimizer and Cosine Annealing learning rate schedule [5] with an initial learning rate of 6×10^{-4} . All experiments were based on PyTorch. The data augmentation contains random affine transforms, the random horizontal flip of 50% probability, and color jitter, where random affine transforms include random rotation within 10 degrees, random sampling ratio from 0.8 to 1.2.

B. Main Results

In this subsection, we compare various state-of-the-art methods on four large-scale datasets. Following BézierLaneNet [14], we use ResNet-18 and ResNet-34 as our backbones.

1) **CULane [1]:** On the CULane dataset, we compare our DHPM with other 12 methods, including SCNN [1], SAD [16], UFLD [7], SIM-CycleGAN [56], CurveLanes [6], LaneATT [5], RESA [3], UFLDv2 [8], CondLaneNet [11], CLRNet [32], LSTR (ResNet-18)* [13], and

BézierLaneNet [14]. As shown in Table II, our DHPM achieves state-of-the-art performance among the curve-based methods. In addition, we consistently outperform the baseline models across different backbones. Specifically, our lightweight model achieves a 0.92 improvement on the F1 score with the same testing speed and without introducing new parameters to the model. In addition, we achieve the lowest FP on the ‘‘Crossroad’’ category across all kinds of methods, where 295 false positive predictions are reduced compared with the baseline. These results demonstrate the effectiveness of our dense hybrid modulation method. The visual comparisons between our DHPM and other methods are shown in Fig. 9. Besides, our method is not specific to curve-based methods. We also conduct experiments on LaneATT [5] as a new baseline. The experimental results show that our method also achieves a significant improvement over LaneATT, which demonstrates that our method has broad application prospects in the task of lane detection.

To prove the improvements in proposal quality, we compare the selected proposal statistics with the baseline on the test split of CULane. As shown in Fig. 10, our proposals participate more in the testing phase than the counterpart model. More proposals are improved to have reasonable locations and shapes, which is the basis for contributing their value.

2) **TuSimple [54]:** On the TuSimple dataset, we compare with current state-of-the-art methods and report the results in Table III. We achieve higher accuracy and lower FN metrics with comparable FP compared to BézierLaneNet [14]. Although the metrics of TuSimple are almost saturated, we

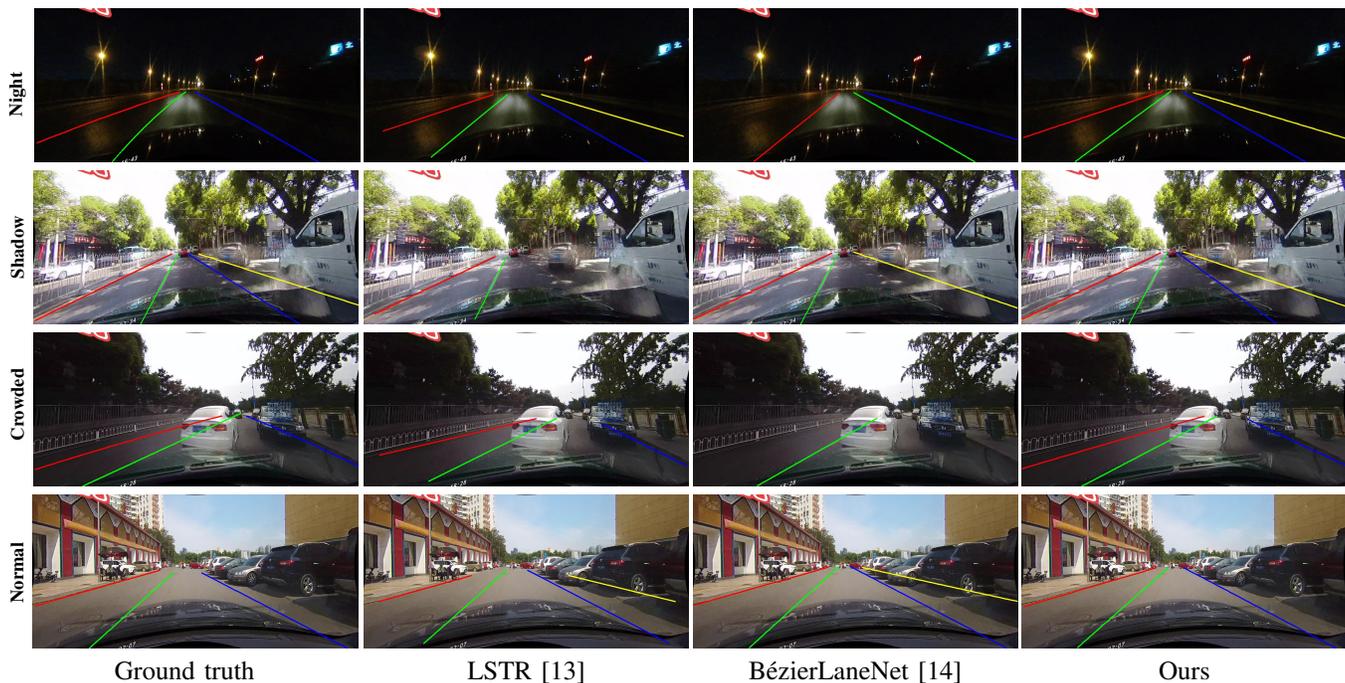
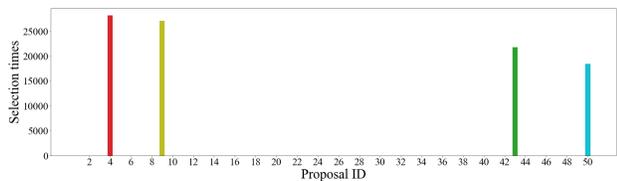
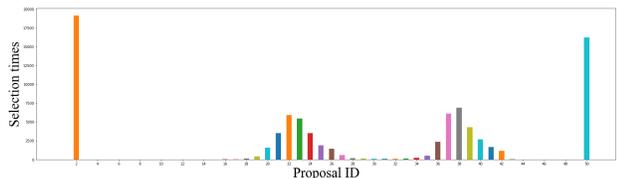


Fig. 9. Qualitative results on the CULane [1] dataset.



(a) Statistics of times each proposal is selected in BézierLaneNet [14]



(b) Statistics of times each proposal is selected in our method



(c) Selected proposal statistics of BézierLaneNet [14] (d) Selected proposal statistics of our methods

Fig. 10. The comparison of selected proposal statistics between BézierLaneNet [14] and our method. Statistics are tested using the ResNet-18 backbone on the test set of CULane [1]. The comparison demonstrates that our proposals participate more in the testing phase than its counterpart model.

still achieve very competitive performances among all kinds of methods. For both backbones, we consistently improve the performance over the baseline counterparts, which proves the effective design of our method. The visual comparisons between our DHPM and other methods are shown in Fig. 11.

TABLE III
QUANTITATIVE RESULTS ON THE TUSIMPLE [54] DATASET

| Method | Accuracy | FP | FN |
|---------------------------------------|----------|-------|------|
| Segmentation-based Method | | | |
| SCNN (LargeFOV) [1] | 96.53 | 6.17 | 1.80 |
| EL-GAN [57] | 96.39 | 4.12 | 3.36 |
| SAD (ENet) [16] | 96.64 | 6.02 | 2.05 |
| Anchor-based Method | | | |
| PointLaneNet (MobileNet-v2) [30] | 96.34 | 4.67 | 5.18 |
| LaneATT (ResNet-18) [5] | 95.57 | 3.56 | 3.01 |
| UFLD (ResNet-18) [7] | 95.82 | 19.05 | 3.92 |
| UFLD (ResNet-34) [7] | 95.86 | 18.91 | 3.75 |
| UFLDv2 (ResNet-18) [8] | 95.50 | 3.06 | 4.82 |
| UFLDv2 (ResNet-34) [8] | 95.56 | 3.18 | 4.37 |
| CondLaneNet (ResNet-18) [11] | 95.48 | 2.18 | 3.80 |
| CLRNet (ResNet-18) [32] | 96.84 | 2.28 | 1.92 |
| Curve-based Method | | | |
| PolyLaneNet (EfficientNet-b0) [12] | 93.36 | 9.42 | 9.33 |
| LSTR (ResNet18) [13] | 96.18 | 2.91 | 3.38 |
| BézierLaneNet (ResNet-18) [14] | 95.41 | 5.30 | 4.60 |
| BézierLaneNet (ResNet-34) [14] | 95.65 | 5.10 | 3.90 |
| ours-BézierLaneNet (ResNet-18) | 95.61 | 5.30 | 3.50 |
| ours-BézierLaneNet (ResNet-34) | 95.87 | 5.00 | 3.40 |

3) *LLAMAS* [55]: We also compare DHPM with recent top-performing approaches on the online benchmark of LLAMAS. As shown in Table IV, we achieve very competitive results on F1 and Recall metrics. Compared to the baseline model, we consistently improve all metrics by a relatively clear advantage. Our ResNet-18 model is comparable to BézierLaneNet’s ResNet-34 model, indicating that our proposal has better overall quality than BézierLaneNet. The visual comparisons between DHPM and others are shown in Fig. 12.

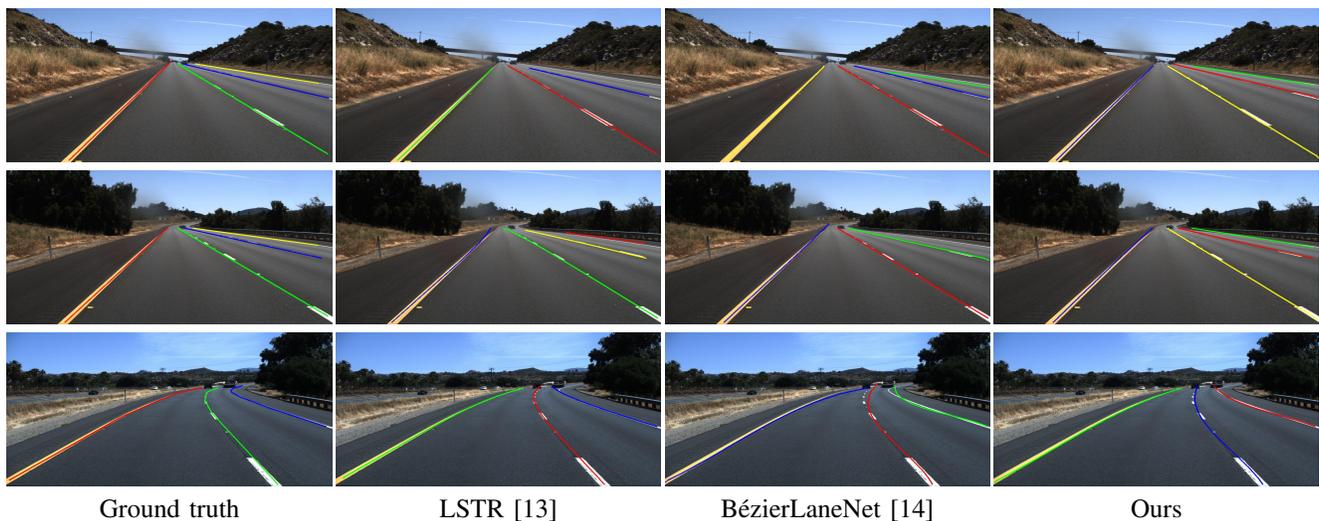


Fig. 11. Qualitative results on the TuSimple [54] test set.

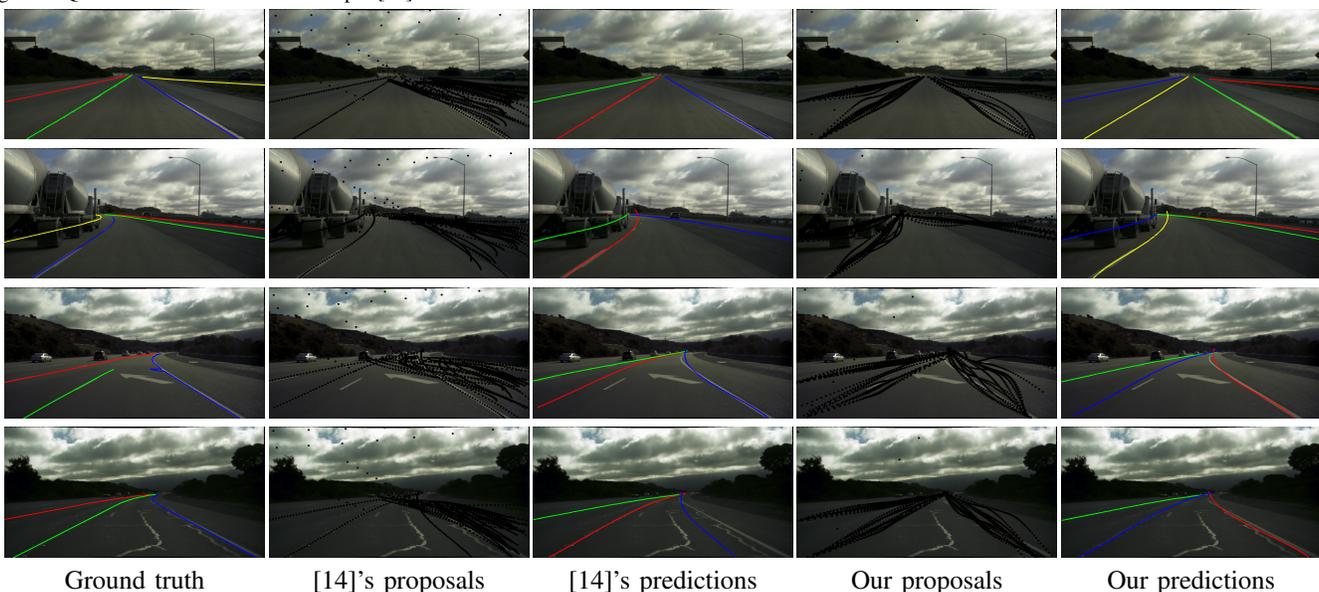


Fig. 12. Qualitative comparison on the LLAMAS [55] validation set.

4) *CurveLanes* [6]: We conduct experiments on the CurveLanes [6] dataset to better prove the effectiveness of our method on curves. The results are shown in Table V. Since the CurveLanes dataset [6] does not release the labels of the test set, we only report the evaluation results on the validation set. The CurveLanes [6] dataset has more than 90% curve samples, so the improvement of our method on it can well prove that our method has a positive effect on curve modeling. The visual comparisons between DHPM and others are shown in Fig. 13.

C. Cross-Dataset Generalization Results

To verify the generalization of our method, we conduct cross-dataset experiments between every two datasets. We use the model trained on one dataset as a starting point to train and test on another dataset. When re-training on the target datasets, we use the same loss function as BézierLaneNet [14], which is illustrated in Equation (1). As comparative experiments, we also conduct cross-dataset re-training experiments on the baseline model, utilizing the same training settings. Consistent

improvements in Table VI show that the starting checkpoint provided by our method is generally better, which demonstrates that our densely modulated proposals have stronger generalization ability and can be used as a good feature initialization.

D. Speed and Parameters

Inference speed and model size are important for autonomous driving algorithms. We report our inference speed and parameters in Table VII, where all FPS results are tested with 590×1640 random inputs on the same device, i.e., a single NVIDIA GTX 3090 GPU. Since our method only re-design the training method and does not modify the architecture, the testing speed and parameters are the same as BézierLaneNet. To sum up, our approach improves model performance and generalization without increasing new parameters and reducing inference speed.



Fig. 13. Qualitative results on the CurveLanes [6] dataset.

TABLE IV
QUANTITATIVE RESULTS ON THE LLAMAS [55] DATASET

| Method | F1 | Precision | Recall |
|---------------------------------------|-------|-----------|--------|
| Anchor-based Method | | | |
| PolyLaneNet [12] | 88.40 | 88.87 | 87.93 |
| CLRNet (ResNet-18) [32] | 96.00 | - | - |
| LaneATT(ResNet-18) [5] | 93.46 | 96.92 | 90.24 |
| UFLDv2 (ResNet-18) [8] | 94.58 | 95.29 | 93.88 |
| UFLDv2 (ResNet-34) [8] | 94.95 | 95.75 | 94.17 |
| Curve-based Method | | | |
| PolyLaneNet [12] | 88.40 | 88.87 | 87.93 |
| BézierLaneNet (ResNet-18) [14] | 94.91 | 95.71 | 94.13 |
| BézierLaneNet (ResNet-34) [14] | 95.17 | 95.89 | 94.46 |
| ours-BézierLaneNet (ResNet-18) | 95.15 | 96.05 | 94.26 |
| ours-BézierLaneNet (ResNet-34) | 95.30 | 96.16 | 94.46 |

TABLE V
QUANTITATIVE RESULTS ON THE VALIDATION SET OF CURVELANES [6]

| Method | F1 | Precision | Recall |
|----------------------------------|-------|-----------|--------|
| Anchor-based Method | | | |
| SCNN (LargeFOV) [1] | 65.02 | 76.13 | 56.74 |
| SAD (ENet) [16] | 50.31 | 63.60 | 41.60 |
| PointLaneNet (MobileNet-v2) [30] | 78.47 | 86.33 | 72.91 |
| CurveLanes-NAS (ResNet-18) [6] | 81.12 | 93.58 | 71.59 |
| CondLaneNet (ResNet-18) [11] | 85.09 | 87.75 | 82.58 |
| Curve-based Method | | | |
| BézierLaneNet (ResNet-18) [14] | 74.56 | 83.27 | 67.50 |
| ours-BézierLaneNet (ResNet-18) | 75.03 | 82.75 | 68.62 |

E. Ablation Study

We conduct ablation experiments on the CULane [1] dataset to verify the effectiveness of each component. All the ablation experiments are performed based on the ResNet-18 version of

TABLE VI
QUANTITATIVE COMPARISON OF CROSS-DATASET GENERALIZATION

| Target \ Source | TuSimple [54] | | CULane [1] | | LLAMAS [55] | |
|-----------------|---------------|--------------|------------|--------------|-------------|--------------|
| | baseline | ours | baseline | ours | baseline | ours |
| TuSimple [54] | 95.01 | 95.44 | 94.98 | 95.13 | 95.51 | 95.70 |
| CULane [1] | 73.48 | 73.91 | 73.36 | 74.33 | 72.53 | 73.56 |
| LLAMAS [55] | 95.36 | 95.78 | 95.52 | 95.66 | 95.42 | 95.75 |

TABLE VII
INFERENCE SPEED AND MODEL SIZE. THE INFERENCE SPEED WAS TESTED ON AN NVIDIA GTX 3090 GPU.

| Method | FPS (image/s) | Params (M) |
|---------------------------------------|---------------|------------|
| LaneATT (ResNet-18) [5] | 205 | 12.02 |
| LaneATT (ResNet-34) [5] | 183 | 22.13 |
| CLRNet (ResNet-18) [32] | 130 | 11.77 |
| CLRNet (ResNet-34) [32] | 106 | 21.88 |
| CondLaneNet (ResNet-18) [11] | 201 | 11.93 |
| CondLaneNet (ResNet-34) [11] | 149 | 22.04 |
| UFLDv2 (ResNet-18) [8] | 320 | 206.30 |
| UFLDv2 (ResNet-34) [8] | 162 | 216.41 |
| BézierLaneNet (ResNet-18) [14] | 212 | 4.10 |
| BézierLaneNet (ResNet-34) [14] | 179 | 9.48 |
| ours-BézierLaneNet (ResNet-18) | 206 | 4.10 |
| ours-BézierLaneNet (ResNet-34) | 183 | 9.48 |

our method. The results are shown in Table VIII, which are the average values of four independent experiments with the same training settings. The visualization of ablation experiments is shown in Fig. 14. We emphasize that the availability constraint J_{ava} aims to supervise the location and shape attributes of a **single** lane proposal, while the diversity constraint J_{div} is proposed to restrain the relationship of **multiple** lane proposals. In other words, the availability constraint is a unitary constraint about S_k , while the diversity constraint is a dualistic constraint. They make their own contributions and assist each other in

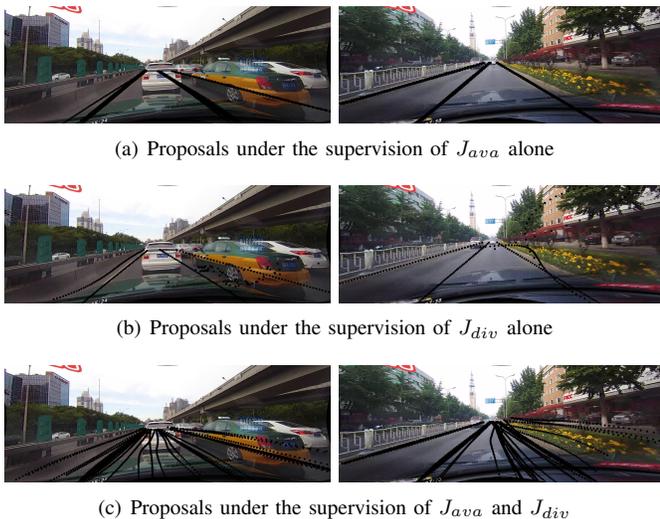


Fig. 14. The visualization of ablation experiments. The proposals sparsely supervised by the availability constraint (a) would cause the overlapping problem. Only applying the diversity constraint (b) would result in low-quality predictions, where most proposals are no longer located in the image. When we add the diversity constraint based on the availability constraint (c), proposals are more dispersed in the image space and of higher diversity.

TABLE VIII
ABLATION STUDY OF THE IMPROVEMENT STRATEGIES ON CULANE

| Model | Basic constraints | J_{ava} | J_{div} | J_{dis} | F1 Score |
|-------|-------------------|-----------|-----------|-----------|----------|
| (a) | ✓ | | | | 73.36 |
| (b) | ✓ | ✓ | | | 73.14 |
| (c) | ✓ | | ✓ | | 73.28 |
| (d) | ✓ | | | ✓ | 73.92 |
| (e) | ✓ | ✓ | ✓ | | 73.72 |
| (f) | ✓ | ✓ | | ✓ | 73.66 |
| (g) | ✓ | | ✓ | ✓ | 73.85 |
| (h) | ✓ | ✓ | ✓ | ✓ | 74.33 |

jointly optimizing the model.

From Table VIII (a-c), we observe that availability constraint or diversity constraint do not positively impact training and even lead to performance degradation. From models (a-d), we observe that the models trained with only J_{ava} or J_{div} fail to bring a performance boost while J_{dis} successfully improved the performance from the perspective of the confidence distribution of proposals. The reason for this phenomenon can be attributed to the fact that J_{ava} and J_{div} are designed to regulate the spatial distributions of proposals from different orientations. On the one hand, J_{ava} is proposed to locate each proposal around the ground truth lanes with a reasonable shape. On the other hand, J_{div} enhances the inter-proposal topological diversification to exploit the underlying relations between proposals. Therefore, the combination of two losses (i.e., model (e)) significantly improves the detection results of our method compared to the baseline. After adding the discrimination constraint, the average F1 score of our method can reach 74.33, which proves the feasibility of improving the quality of all proposals.

The goal of availability constraint is to improve the quality of all proposals. To achieve that, we need to constrain their

TABLE IX
ABLATION STUDY OF AVAILABILITY CONSTRAINT

| Baseline | Location constraint | Shape constraint | F1 Score |
|----------|---------------------|------------------|----------|
| ✓ | | | 73.85 |
| ✓ | ✓ | | 73.80 |
| ✓ | | ✓ | 74.06 |
| ✓ | ✓ | ✓ | 74.33 |

TABLE X
ABLATION STUDY OF DIVERSITY CONSTRAINT

| Baseline | Difference Constraint | Upper Limit | F1 Score |
|----------|-----------------------|-------------|----------|
| ✓ | | | 73.66 |
| ✓ | ✓ | | 74.01 |
| ✓ | ✓ | ✓ | 74.33 |

shapes and locations to a reasonable range. Table IX presents the ablation experiment to evaluate each component (i.e., shape and location) in the availability constraint. We can observe that none of them can improve performance alone, because the attributes they constrain are orthogonal, so constraining the shape and location of the proposal at the same time will bring obvious performance improvement.

As described above, the diversity constraint is divided into two parts: the constraint to increase shape differences and the upper limit of proposal-to-label matching. We conduct three groups of experiments to verify their effectiveness, and results are shown in Table X. By comparing the first two rows in Table X, we can see that the proposed difference constraint has successfully improved the performance. Comparing the results of the last two rows, we can see the positive effect of an upper limit. Comparing these three groups of experiments proves that both parts of the diversity constraint are essential.

Compared with the first two constraints, the discriminative constraints are relatively independent. The conclusion can be drawn by comparing (a-d) and (e-h) of Table VIII. We find that experiments with discriminative constraints have higher F1 scores than experiments without discriminative constraints, leading to the conclusion that discriminative constraints alone can have a positive effect.

V. CONCLUSION

In this paper, we have proposed dense hybrid proposal modulation to generate high-quality proposals for lane detection. The availability constraint makes all proposal lane lines have suitable locations and shapes while the diversity constraint avoids all the proposals under the supervision of availability constraint overlapping together and improves the diversity of proposals. The discriminative constraints improve the discriminative ability of detectors and filter out proposals that are not good enough. Benefitting from these constraints, we achieve very competitive results on four popular datasets.

REFERENCES

- [1] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," in *AAAI*, 2018, pp. 7276–7283.

- [2] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans, and L. V. Gool, "Towards end-to-end lane detection: an instance segmentation approach," in *IV*, 2018, pp. 286–291.
- [3] J. Zhang, T. Deng, F. Yan, and W. Liu, "Resa: Recurrent feature-shift aggregator for lane detection," in *AAAI*, 2021, pp. 3547–3554.
- [4] Y. Ko, Y. Lee, S. Azam, F. Munir, M. Jeon, and W. Pedrycz, "Key points estimation and point instance segmentation approach for lane detection," *TITS*, pp. 8949–8958, 2022.
- [5] 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," in *CVPR*, 2021, pp. 294–302.
- [6] H. Xu, S. Wang, X. Cai, W. Zhang, X. Liang, and Z. Li, "Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending," in *ECCV*, 2020, pp. 689–704.
- [7] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in *ECCV*, 2020, pp. 276–291.
- [8] Z. Qin, P. Zhang, and X. Li, "Ultra fast deep lane detection with hybrid anchor driven ordinal classification," *TPAMI*, pp. 1–14, 2022.
- [9] J. Phillion, "Fastdraw: Addressing the long tail of lane detection by adapting a sequential prediction network," in *CVPR*, 2019, pp. 11 582–11 591.
- [10] S. Yoo, H. Seok Lee, H. Myeong, S. Yun, H. Park, J. Cho, and D. Hoon Kim, "End-to-end lane marker detection via row-wise classification," in *CVPRW*, 2020, pp. 4335–4343.
- [11] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," in *ICCV*, 2021, pp. 3753–3762.
- [12] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. D. Souza, and T. Oliveira-Santos, "Polylanenet: Lane estimation via deep polynomial regression," in *ICPR*, 2021, pp. 6150–6156.
- [13] R. Liu, Z. Yuan, T. Liu, and Z. Xiong, "End-to-end lane shape prediction with transformers," in *WACV*, 2021, pp. 3693–3701.
- [14] Z. Feng, S. Guo, X. Tan, K. Xu, M. Wang, and L. Ma, "Rethinking efficient lane detection via curve modeling," in *CVPR*, 2022, pp. 17 062–17 070.
- [15] B. D. Brabandere, D. Neven, and L. V. Gool, "Semantic instance segmentation with a discriminative loss function," *arXiv preprint arXiv:1708.02551*, 2017.
- [16] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," in *ICCV*, 2019, pp. 1013–1021.
- [17] J. Yang, L. Zhang, and H. Lu, "Lane detection with versatile atrousformer and local semantic guidance," *arXiv preprint arXiv:2203.04067*, 2022.
- [18] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *ECCV*, 2020, pp. 213–229.
- [19] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for object detection," in *CVPR*, 2018, pp. 3588–3597.
- [20] F. Pizzati, M. Allodi, A. Barrera, and F. García, "Lane detection and classification using cascaded cnns," in *EUROCAST*, 2019, pp. 95–103.
- [21] Z. M. Chng, J. Lew, and J. A. Lee, "Ronelvd2: A faster, improved lane tracking method," *arXiv preprint arXiv:2202.13137*, 2022.
- [22] D. Wu, M. Liao, W. Zhang, and X. Wang, "Yolop: You only look once for panoptic driving perception," *arXiv preprint arXiv:2108.11250*, 2021.
- [23] H. Fang, J. Zhu, and Y. Fang, "Continuitylearner: Geometric continuity feature learning for lane segmentation," *arXiv preprint arXiv:2108.03507*, 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," *RA-L*, pp. 7477–7484, 2021.
- [25] S. Jung, S. Choi, M. A. Khan, and J. Choo, "Towards lightweight lane detection by optimizing spatial embedding," *arXiv preprint arXiv:2008.08311*, 2020.
- [26] W. Cheng, H. Luo, W. Yang, L. Yu, and W. Li, "Structure-aware network for lane marker extraction with dynamic vision sensor," *arXiv preprint arXiv:2008.06204*, 2020.
- [27] P. Lu, C. Cui, S. Xu, H. Peng, and F. Wang, "Super: A novel lane detection system," *TIV*, pp. 583–593, 2021.
- [28] Y. Hou, Z. Ma, C. Liu, T.-W. Hui, and C. C. Loy, "Inter-region affinity distillation for road marking segmentation," in *CVPR*, 2020, pp. 12 483–12 492.
- [29] F. Yan, M. Nie, X. Cai, J. Han, H. Xu, Z. Yang, C. Ye, Y. Fu, M. B. Mi, and L. Zhang, "Once-3dlanes: Building monocular 3d lane detection," in *CVPR*, 2022, pp. 17 122–17 131.
- [30] Z. Chen, Q. Liu, and C. Lian, "Pointlanenet: Efficient end-to-end cnns for accurate real-time lane detection," in *IV*, 2019, pp. 2563–2568.
- [31] D. Jin, W.-K. Park, S.-G. Jeong, H. Kwon, and C.-S. Kim, "Eigenlanes: Data-driven lane descriptors for structurally diverse lanes," in *CVPR*, 2022, pp. 17 163–17 171.
- [32] T. Zheng, Y. Huang, Y. Liu, W. Tang, Z. Yang, D. Cai, and X. He, "Clrnet: Cross layer refinement network for lane detection," in *CVPR*, 2022, pp. 898–907.
- [33] J. Su, C. Chen, K. Zhang, J. Luo, and X. Wei, "Structure guided lane detection," in *IJCAI*, 2021, pp. 997–1003.
- [34] M. Lee, J. Lee, D. Lee, W. Kim, S. Hwang, and S. Lee, "Robust lane detection via expanded self attention," in *WACV*, 2022, pp. 1949–1958.
- [35] 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 *ECCV*, 2022, pp. 550–567.
- [36] W. Van Gansbeke, B. De Brabandere, D. Neven, M. Proesmans, and L. Van Gool, "End-to-end lane detection through differentiable least-squares fitting," in *ICCVW*, 2019, pp. 905–913.
- [37] C.-B. Wu, L.-H. Wang, and K.-C. Wang, "Ultra-low complexity block-based lane detection and departure warning system," *TCSVT*, pp. 582–593, 2019.
- [38] T. Sato and Q. A. Chen, "On robustness of lane detection models to physical-world adversarial attacks in autonomous driving," *arXiv preprint arXiv:2107.02488*, 2021.
- [39] Z. Cai and N. Vasconcelos, "Cascade r-cnn: High quality object detection and instance segmentation," *TPAMI*, pp. 1483–1498, 2021.
- [40] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 936–944.
- [41] G. Spyros and K. Nikos, "Locnet: Improving localization accuracy for object detection," in *CVPR*, 2016, pp. 789–798.
- [42] G. Cheng, J. Yang, D. Gao, L. Guo, and J. Han, "High-quality proposals for weakly supervised object detection," *TIP*, pp. 5794–5804, 2020.
- [43] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *ICCV*, 2017, pp. 618–626.
- [44] R. Zeng, W. Huang, C. Gan, M. Tan, Y. Rong, P. Zhao, and J. Huang, "Graph convolutional networks for temporal action localization," in *ICCV*, 2019, pp. 7093–7102.
- [45] J. Deng, Y. Pan, T. Yao, W. Zhou, H. Li, and T. Mei, "Relation distillation networks for video object detection," in *ICCV*, 2019, pp. 7022–7031.
- [46] X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian, "Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis," *TCSVT*, pp. 1758–1770, 2020.
- [47] M. Feng, S. Z. Gilani, Y. Wang, L. Zhang, and A. Mian, "Relation graph network for 3d object detection in point clouds," *TIP*, pp. 92–107, 2021.
- [48] J. Wang, L. Song, Z. Li, H. Sun, J. Sun, and Z. Nanning, "End-to-end object detection with fully convolutional network," in *CVPR*, 2021, pp. 15 849–15 858.

- [49] X. Lu, C. Ma, B. Ni, and X. Yang, “Adaptive region proposal with channel regularization for robust object tracking,” *TCSVT*, pp. 1268–1282, 2021.
- [50] J. U. Kim, J. Kwon, H. G. Kim, and Y. M. Ro, “Bbc net: Bounding-box critic network for occlusion-robust object detection,” *TCSVT*, pp. 1037–1050, 2020.
- [51] X. Chen, H. Li, Q. Wu, K. N. Ngan, and L. Xu, “High-quality r-cnn object detection using multi-path detection calibration network,” *TCSVT*, pp. 715–727, 2021.
- [52] X. Zhang, C. Zhu, S. Wang, Y. Liu, and M. Ye, “A bayesian approach to camouflaged moving object detection,” *TCSVT*, pp. 2001–2013, 2017.
- [53] Y. Lin, Y. Tong, Y. Cao, Y. Zhou, and S. Wang, “Visual-attention-based background modeling for detecting infrequently moving objects,” *TCSVT*, no. 6, pp. 1208–1221, 2017.
- [54] TuSimple, “Tusimple benchmark.” <https://github.com/TuSimple/tusimple-benchmark/>, Accessed September, 2020.
- [55] B. Karsten and S. Ryan, “Unsupervised labeled lane markers using maps,” in *ICCVW*, 2019, pp. 832–839.
- [56] T. Liu, Z. Chen, Y. Yang, Z. Wu, and H. Li, “Lane detection in low-light conditions using an efficient data enhancement: Light conditions style transfer,” in *IV*, 2020, pp. 1394–1399.
- [57] M. Ghafoorian, C. Nugteren, N. Baka, O. Booiij, and M. Hofmann, “El-gan: Embedding loss driven generative adversarial networks for lane detection,” in *ECCVW*, 2018, pp. 256–272.