2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

DSSF-net: Dual-Task Segmentation and Self-supervised Fitting Network for End-to-End Lane Mark Detection

Wentao Du, Zhiyu Xiang, Yiman Chen, Shuya Chen

Abstract— **Lane mark detection is one of the key tasks for autonomous driving systems. Accurate detection of lane marks under complex urban environments remains a challenge. In this paper, an end-to-end lane mark detection network named DSSF-net, which is capable of directly outputting the accurate fitted lane curves, is proposed. First, a dual-task segmentation framework for jointing lane category prediction and spatial partition is presented. An IoU-based loss function is put forward to tackle the severely imbalanced category distribution problem. Then a fully self-supervised curve fitting network is proposed to directly output the parameters of lane line upon the probability map. To achieve better accuracy, the fitting network is trained with two sub-stages: coarse regression and confidence-based optimization. Finally the entire DSSF-net is implemented end-to-end. Comprehensive experiments conducted on challenging CULane dataset show that our model achieves 74.9% in F1-score and outperforms the state-of-the-art models.**

I. INTRODUCTION

Lane marks in urban roads or highways play an important role in regulating the traffic. Lane mark detection has been a hot research area for intelligent vehicles since the development of Assistant Driving Systems (ADS) in late 20th century. However, accurate and robust lane mark detection under complex urban roads and various illumination conditions still remains a challenge.

Conventional solutions attempt to grasp some invariant characteristics in different traffic scenes. For instance, white lane marks have the highest Y-channel value and yellow ones have the lowest Cb-channel value in YCbCr color space [1]. Second-derivative of Gaussian followed by thresholding operation is employed to extract edge positions with expected line width [2]. However, constrained on strong but practically unstable prior, such approaches barely work under complex conditions.

Data-driven algorithms have set off an upsurge thanks to the development of semiconductor industry and availability of massive data. Within computer vision domain, deep convolutional neural network provides a powerful tool which surpasses the performance of most algorithms delicately designed in conventional era. As a result, research on lane mark detection gradually resorts to neural networks for its better performance in various environments. Large-scale lane detection datasets like TuSimple [3], BDD100K [4] and CULane [5] have been set up for training and testing the CNN models. They annotate each lane line with a group of sampling points and connect them by smooth curves. Because of the slim and long shape, one challenge of the lane detection task lies in the small proportion of the lane mark pixels with respect to the entire image. It brings a serious category imbalance problem which constrains the performance of CNN based methods.

Generic objects with common shape have strong color similarity and local homogeneity so that post-processing algorithms like fully-connected CRFs [6] can remedy the defects on boundary regions. However, due to the slim shape and color inconsistency caused by shadow or severe occlusion, this strategy hardly takes effects in lane mark detection. Spatial CNN [5] imitates eyes scan process from different directions to simulate spatial continuity and retain integrity of lane mark. The spatial convolution in feature maps brings heavy computing burden to the network. StripNet [7] adopts an extra stage to regress the boundary after aligning RoI, which transforms strips into straight lines and effectively wipes off discontinuous segments to maintain spatial conformity. However, it heavily relies on the accurate localization of lane endpoints. Like SCNN or StripNet, most of current CNN based methods have in common that they only predict lane pixel candidates. The selection of the best candidates and fitting them into lane curves still rely on non-learning methods, which results in an inferior performance.

In this paper, an end-to-end lane mark detection network capable of directly outputting fitted parameters of the lane curves is proposed. The network is composed of two parts: lane mark segmentation and curve fitting. We decompose the first task into two sub-tasks, i.e., spatial partition for separating image into four subspaces and category prediction for recognizing lane mark area. By this way, each sub-task focuses on a single attribute so that specialized optimization techniques can be applied separately. Confronted with severe category imbalance problem, we propose an IoU-based loss function to raise the localization accuracy. In addition, we inherit the recurrent message passing scheme SCNN [5] and improve its computing efficiency by parallel transformation. For the second part, we design a fully self-supervised lane mark fitting network as a substitute for artificial algorithms. In training, we first roughly sample maximum response pixels from probability maps predicted by segmentation stage and distill the shape prior underneath the coarse sequences with neural networks. Aiming at establishing strong relationship between probability maps and sampling points, we provide a confidence-based method to precisely predict the

The work is supported by the National Natural Science Foundation of China-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709214) and the National Natural Science Foundation of China (NSFC) (61571390).

Wentao Du, Zhiyu Xiang, Yiman Chen and Shuya Chen are with College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310007, China, {duwentao, xiangzy, chenyiman, shuya_chen}@zju.edu.cn

parameters of lane mark without external supervision. With the help of this small network, the job of best candidates selection and curve fitting can be jointly finished. Finally, the entire lane mark detection network is implemented end-to-end and validated in popular urban road dataset CULane.

Our main contributions are summarized as follows:

- We propose a dual-task segmentation architecture for extracting lane mark area, which is able to better learn the continuity and slim shape prior of lane mark. An IoU-based loss is further proposed to tackle the category imbalance problem;
- We put forward a self-supervised lane fitting method that can perform best candidates selection and curve fitting jointly;
- Our entire DSSF-net is implemented end-to-end and tested with real data. It achieves state-of-the-art performance on the largest urban lane mark detection benchmark CULane with 74.9% F1-score.

II. Related Work

Conventional lane mark detection is composed of four basic procedures: image pre-processing, feature extraction, curve fitting and lane tracking [8] [9]. Among them, feature extraction is of great importance. Color-based and edge-based features are the most fundamental attributes in image processing which are also widely incorporated in lane mark detection solutions [1] [2]. However, they are troublesome when coping with changing illuminations caused by weather or unpredictable occlusions.

Similar to most other computer vision tasks, research on lane mark detection has been deeply influenced by the advancement of deep learning. Inspired by the flourishing development on object detection, Huval et al. [10] modify an early object detection model Overfeat to detect lane mark segments as groups of local bounding boxes and cluster with DBSCAN. VPGNet [11] prioritizes the prediction of vanishing point as a powerful structure indication, and detects both traffic lanes and road marks with a unified multi-branch network. Line-CNN [12] transfers the success of two-stage detection framework Faster R-CNN and puts forward an insightful line representation comprising discrete direction classification and horizontal coordinate offsets regression.

Lane mark detection by pixel-wise segmentation is another category of popular methods. Chen et al. [13] employ VGG and SegNet to obtain a dense prediction map and then group pixels according to distance-based connection cost. Neven et al. [14] propose an embedding branch and a segmentation branch for instance-level segmentation and design H-Net to adaptively predict the aerial view transformation matrix as a countermeasure against road plane changes. Hsu et al. [15] utilize the pairwise relationship between pixels to formulate a general learning objective for instance segmentation and apply it to lane mark detection. In order to model continuity of slim-shaped lines, Pan et al. [5] propose SCNN module to take the place of dense CRF and show the effectiveness in both lane mark detection and generic semantic segmentation.

Although neural networks have almost replaced hand-crafted features extraction in lane mark detection, fitting algorithms make little compelling progress during recent years. In the past, lane mark are often shaped with straight lines [1] [16] or higher-order polynomial curves [2] [13] while least mean square (LMS) [1] and random sample consensus (RANSAC) [17] [18] are universally adopted as fitting methods. After stepping into neural networks age, a typical method is to sample points by maximal probability principle and interpolate with cubic splines [5]. H-Net [14] intends to fit on dynamic bird-eye view for ease of processing. Inverse Perspective Mapping (IPM) is popular as a part of pre-processing for fitting both in conventional [2] [17] and neural networks [19] [20] schemes. Besides, some research makes efforts to end-to-end detection in different ways. Van Gansbeke et al. [21] introduce least-square as a differentiable matrix operation and directly supervise with ground truth parameters. Li et al. [12] describe a lane mark with a starting point, an orientation and a group of offsets which are predicted simultaneously by Line-CNN. PointLaneNet [22] predicts a lane line instance for each pixel in final feature map consisting of offsets, a starting point and corresponding confidence. Unlike the above methods, our fitting network is totally self-supervised and compatible with the current lane mark detection methods.

III. Proposed Method

The end-to-end framework of our lane mark detection scheme is displayed in Fig. 1. The input is first fed to a backbone network based on VGG-16 [23] with three top dilated convolution [24] layers and then enters into four directional message passing blocks [5] in parallel after reducing channels. The concatenated feature maps are separately sent to branches of category prediction and spatial partition, which jointly determine the lane mark segmentation results. Finally, a lane fitting network is deployed to predict curve parameters from combined dense score maps.

A. Dual-task Segmentation Architecture

Lane mark detection acts as a kind of scene perception to give potential guidance for driver assistance systems or autonomous driving. The current lane and two adjacent lanes are of the most importance in applications like lane keeping and lane changing. Based on these observation, we put emphasis on detecting four nearest neighboring lane lines which are the boundaries of ego lane and subsidiary lanes.

Some complex tasks are possible to be decomposed into small tasks and dealt with multi-task schemes. For example, object detection can be treated as bounding box regression and object classification respectively. Transferring the experience to this lane mark detection problem, we notice that two divisible attributes, i.e., spatial location and pixel category, describe the target from different perspectives and jointly determine each instance. Therefore, we propose a dual-task segmentation architecture for lane mark instance detection. One sub-task is to partition the lower part of image into four separate regions centered at each lane line, and the other is

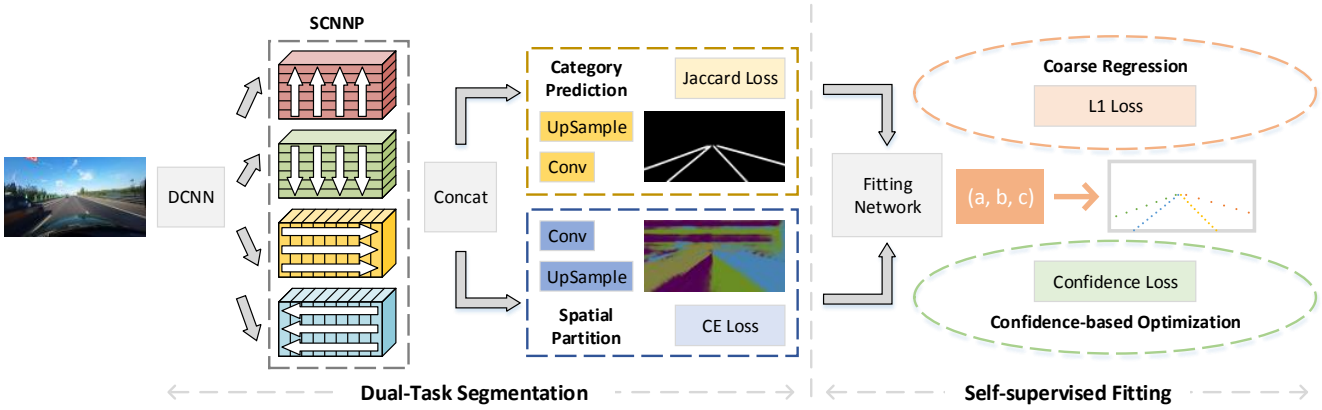

Fig. 1. An overview of our end-to-end DSSF-net. A modified VGG-16 module is used in DCNN as the skeleton for feature extraction. The features are then strengthened by a parallel message passing module SCNNP. Lane mark segmentation part has two branches, i.e., category prediction and spatial partition. The final self-supervised fitting network is responsible for outputting the fitted lane curves, which can be trained with two sub-stages, coarse regression and confidence-based optimization.

to recognize the entire lane mark area from the background. Combining the partition and classification results together, each lane line can be uniquely determined.

The benefits of this dual-task architecture are in several aspects. First, it decouples the supervisory label and forms a multi-task learning structure, which has been widely believed to boost performance of each task. Second, imbalanced category distribution is partially alleviated. In this problem, the proportion of lane mark pixels is small and the background usually takes up more than ninety-five percents of the entire frame, which is literally a disaster for machine learning methods. Although many researches have contributed to solving this problem by class re-sampling or cost-sensitive training, balanced dataset is still the preferred choice if it is possible. In our dual-task design, the spatial partition is relatively class-balanced and the imbalance problem is also largely relieved in category prediction task by combining all lane marks as a whole. Third, in category prediction branch we only need to deal with a simple two-class problem instead of previous five-class one. There are many good techniques to improve the performance of binary class problem. For example, since only two categories should be concerned, we can simply apply additional inter-class constraints or introduce hyper-parameters to raise performance without combinatorial explosion. Along this way, we deploy an IoU-based loss function to implicitly tackle the relationship between the foreground and background.

B. IoU-based Loss

Originated from information theory, cross entropy loss has almost been the standard loss function in supervised semantic segmentation tasks. Denote the probability of c-th class at i-th pixel as $p_{i,c}$ and ground truth of i-th pixel as g_i, the cross entropy loss can be expressed as

$$\mathrm{CE}(p,g) = \frac{1}{N}\sum_{i}^{N} -\log p_{i,g_i} \tag{1}$$

The average negative log-probability of corresponding class over all pixels is viewed as the overall loss, which aims to represent the expectation quality achieved by pixel-wise prediction. The drawback of the cross entropy loss mainly lies in the lack of supervision on pixel dependency or mutual correlations, which is the aspect that segmentation differs from simple combination of independent classifications.

On the other hand, intersection over union (IoU) is an acknowledged metric in segmentation tasks. For each specific class, it counts the number of pixels that both labeling and prediction are positive as intersection and either positive as union. Then IoU is defined as the ratio between intersection and union, which reveals the segmentation quality in a global scale and the numeric value coincides well with human visual intuitions. Dice loss [25] is put forward in the task of medical image segmentation based on a similar concept Dice similarity coefficient (DSC). With the same set of annotations mentioned before, dice loss of c-th class can be written as

$$\mathrm{D}(p,g) = 1 - \frac{2\sum_i^N p_{i,c} * \mathcal{I}(g_i,c)}{\sum_i^N \left[p_{i,c}^2 + \mathcal{I}(g_i,c)^2\right]} \tag{2}$$

where the value of $\mathcal{I}(i,j)$ is 1 if i equals to j and otherwise 0. Numerical multiplication and addition replace set intersection and union respectively while the result still lies in $[0,1]$ on the basis of arithmetic and geometric means (AM–GM) inequality. Although not being a strict IoU metric, dice loss does well in keeping a balance between precision and recall in medical image processing. It is actually equivalent to F1 score in statistical analysis.

We believe that keeping full consistence between loss functions and assessment criteria will benefit the neural network in supervised learning. Since IoU is an ideal evaluation metric in the task of lane mark detection, we propose a strict IoU-based objective function termed Jaccard loss derived from Jaccard similarity coefficient (JSC). It is defined as

$$\mathrm{J}(p,g) = 1 - \frac{\sum_i^N p_{i,c} * \mathcal{I}(g_i,c)}{\sum_i^N \left[p_{i,c}^2 + \mathcal{I}(g_i,c)^2 - p_{i,c} * \mathcal{I}(g_i,c)\right]} \tag{3}$$

Jaccard loss inherits the numerical multiplication and addition expression and reproduces the form expressed in IoU. We observe IoU-based loss function also has a drawback that categories with dominant pixel quantity can be less numerical sensitive than those covering limited area. This is due to the magnitude difference in denominators and disturbs especially in multi-class cases. However, it won't be a problem when we only focus on the small lane mark area in our binary category prediction sub-task. In this circumstances, IoU-based loss succeeds in implicitly handling the imbalanced distribution problem. It does not require any additional hyper-parameters which is usually necessary for weighted cross entropy loss, and outperforms the simple aggregation of pixel-wise losses.

C. Self-supervised Lane Mark Fitting

After obtaining the probability map (probmap) predicted by the segmentation network, what should be done next is to distill precise coordinates of lane mark from probmap. Although deep neural networks have dominated image feature extraction in recent years, tasks equipped with developed mathematical models like line fitting have little improvement. Most of the CNN based lane detection methods still rely on artificially designed post-processing algorithm to get the fitted lane curves. However, picking out the best lane candidate pixels in the probmap is not a simple problem. Lots of outliers still exist in the probmap. Therefore, considering only the pixels with highest probability in probmap is not enough for accurate curve fitting. The neighboring pixels with lower probability should all be considered together. Furthermore, the candidate selecting and curving fitting should be solved jointly to achieve better performance. As a result, we propose a four-layer fully-connected lane fitting network to solve this problem. Despite of the simple structure, the training of this network is nontrivial. We develop a self-supervised training process for it, which can be divided into two sub-stages, i.e., coarse regression and confidence-based optimization.

1) Coarse Regression: Coarse regression is a necessary preparation for fine-grained optimization. It teaches the network what is the common shape or locations of lane curve and how probmap interacts with spatial coordinates. It attempts to make predictions closely related with the pattern of probmap and exploit the lane position information that probmap depicts. To reach a coarse initial state, the fitting network is supervised with low precision coordinates which can be collected by max-response principle or any other strategies directly from the predicted probmap.

Considering the shape constraint, it's not reliable to explicitly predict all sampling points with a regression network. Since lane lines within the field of view can be simply modeled by parabolic curves $y = ax^2 + bx + c$, we take low-resolution probmap as input and then predict three coefficients of each lane line, which determine the unique instance in image space. We utilize L1 loss of coordinates prediction instead of those parabolic coefficients as supervision signal, and all coordinates are normalized within $[-1, 1]$ so as to eliminate the scale effects. Through this regression process, our fitting network is able to remove some numeric deviation and distill valid information from inaccurate statistics.

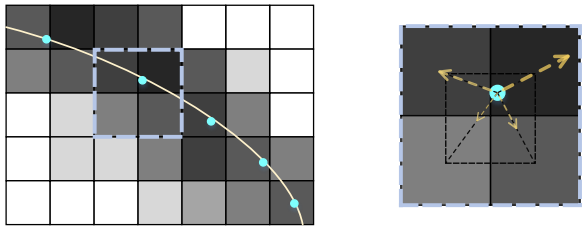

Fig. 2. Confidence-based Optimization. *Left:* The darker grids mean pixels with higher probability and the cyan points are sampled from the predicted lane mark curve. *Right:* Each sampling point is pulled by four nearest neighbor pixels. The longer arrowed lines represent stronger pull effects depending on position offsets and likelihoods in probmap.

2) Confidence-based Optimization: Our ultimate goal is to bridge the gap between probmap and coefficients of fitted curves. But until now, we have merely trained the fitting network with coarse coordinates sampled roughly from probmap. So we intend to further optimize in this stage.

Confidence-based optimization inherits the network from coarse regression and borrows its model weights as the initial state. The main motivation of this stage is to maximize the aggregated likelihood of pixels in the probmap where curve regression indicates. One noticeable problem is that indexing operation itself is not differentiable with respect to the index, that is to say, the forward function defined by taking the value $T_{<i,j>}$ from a tensor T at i-th row and j-th column cannot be propagated backward to i and j. Inspired by [26], we introduce the bilinear interpolation as an auxiliary tool, which takes integer part as the index and fractional part as an offset. The diagram in Fig. 2 illustrates the core mechanism of how to lead the network to a desired state, where the feedback based on gradient descent comes from the local difference of four nearest pixels in probmap.

With regard to the loss function, we ignore those noise-dominated predictions with confidence lower than a threshold τ, which we set 0.4 in experiments. Apart from that, two other factors are taken into account to mitigate disturbance from numerically unstable fluctuation. First, the optimal lane curve should be largely dependent on candidate pixels with high probability. Second, the probability of the pixels in the fitted curve cannot exceed the highest probability at the same row. The gap between them should also be taken as a component in loss function. Finally, the confidence loss function can be formed as

$$\text{ConfLoss} = \frac{1}{N}\sum_{i}^{N} p_i * (\max(P_i) - p_i) * \mathcal{I}(p_i > \tau) \quad (4)$$

where p_i and P_i are the probability at i-th fitted point and the set of all values at the same row in probmap, respectively. $\mathcal{I}(x)$ equals to 1 if statement x is true and otherwise 0.

After analyzing the pattern of the predicted probability, we find that there is usually a wide and flat value peak with two sharp edges for the pixels in the same row, as

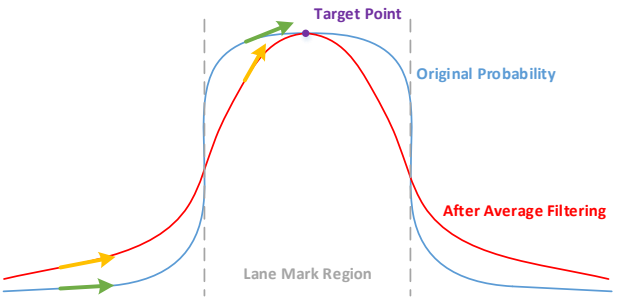

Fig. 3. Illustration of applying average filtering to the predicted probability map. Arrowed lines represent probability slopes at given positions, which shows the filtered curve is beneficial for converging from distant positions and approaching the target point more rapidly and accurately.

shown in Fig. 3. In this circumstance, the gradient is small at most places except for the two narrow transition zone, which will lead the learning process prone to be confined at flat zones and cause a problem when initial value is not good. Furthermore, the ideal target point should be at the center of lane mark belt instead of an uncertain position within the belt area. Therefore it is necessary to make the top region more distinguishable at a local scale. In other words, we need the grey value of the probability band to be more Gaussian-like. As Fig. 3 depicts, we apply average filtering to the probmap before calculating the confidence-based loss so that the flat top is shrunk and the transition area gets expanded. So for those initial positions outside of the lane mark region, they are more inclined to break through the lower barrier and approach the target point. When the position is inside of the lane mark region, the narrower peak of the grey distribution is also helpful to guide the prediction to reach the center of lane mark region.

IV. Experiments

A. Implementation

Dataset: CULane is the largest challenging dataset on lane mark detection in urban environments. It consists of 133,235 monocular images in total where 34680 frames are divided into 9 categories serving as test set, including crowded urban scenes, dazzling or dark conditions, shadow covered and even no lane mark roads. Considering the variety of challenging traffic scenarios this dataset covers, we think it is ideal for evaluation of our method.

Training Settings: In training process, we set batch size to 8 and use SGD with base learning rate 0.01, momentum 0.9, weight decay 0.0001. Following [27][28], we also employ a poly learning rate policy with power 0.9 and maximal iteration 100K. All images are resized to 288×800 with bicubic interpolation in pre-processing stage. For the fitting network, batch size is adjusted to 16 and the base learning rates are 0.01 and 0.001 for two sub-stages respectively. The probmap is interpolated to 72×200 in self-supervised fitting as a trade-off between complexity and precision.

Evaluation Metrics: We adopt two metrics to evaluate the robustness and accuracy of our model, namely area-metric and distance-metric. Area-metric is provided by CULane dataset, which converts sampling points to smooth curves by cubic spline interpolation and defines the region within 30 pixels as lane mark area. The instance will be counted as true positive (TP) if IoU > 0.5 holds. Distance-metric is based on the distance between sampling points on the prediction and ground truth. The distance threshold is 10 pixels in our experiment. The lane lines with more than half of all points correctly predicted are judged as TP. Finally, we calculate F1 score for both metrics, which is defined as the harmonic mean of precision and recall. These two metrics evaluate lane mark prediction from different perspectives. The former pays more attention to overall robustness while the latter focuses on local accuracy.

B. Ablation Study

In this part, we begin with the plain baseline and evaluate our proposed elements step by step. Due to the limited space, we choose two representative results for both area-metric (A) and distance-metric (D), the most common type of road conditions in CULane dataset (Normal) and the overall performance under all test samples (Total).

(1) *Dual-task Segmentation Architecture.* Here we intend to compare dual-task segmentation architecture (T2) with the original single-task scheme (T1), which treats four lane and background as five different categories. The plain base model (Base) only uses the DCNN in Fig. 1 as feature extraction module while SCNNP adds the parallel SCNN into it. In order to make a fair comparison, cross-entropy loss is applied to all segmentation branches. Single-task architecture utilizes weighted cross entropy with background weight 0.4 and category prediction branch is weighted to 3 in dual-task architecture. Experimental results are depicted in TABLE I. It is obvious that dual-task segmentation outperforms the single-task method with distinguishable improvement under the same loss function type, which validates the effectiveness of task decomposition. With the help of spatial convolution, the model with SCNNP performs further better.

TABLE I
EVALUATION ON DUAL-TASK SEGMENTATION ARCHITECTURE

Model	Normal(D)	Total(D)	Normal(A)	Total(A)
Base(T1)	61.89	43.73	80.49	59.73
Base(T2)	**63.98**	**45.91**	**82.55**	**61.82**
SCNNP(T1)	65.93	48.40	90.01	70.90
SCNNP(T2)	**67.80**	**50.15**	**90.56**	**71.47**

(2) *IoU-based Loss.* To verify the effectiveness of IoU-based loss function, we use the dual-task network with SCNNP and compare the networks trained by three different loss functions respectively, i.e., weighted cross entropy loss, dice loss and jaccard loss we proposed. As TABLE II reveals, both dice and jaccard loss strikingly boost performance compared with the cross-entropy loss, and jaccard loss slightly outperforms dice loss. In fact, dice and jaccard loss are substantially homogeneous but have diverse tolerance to TP

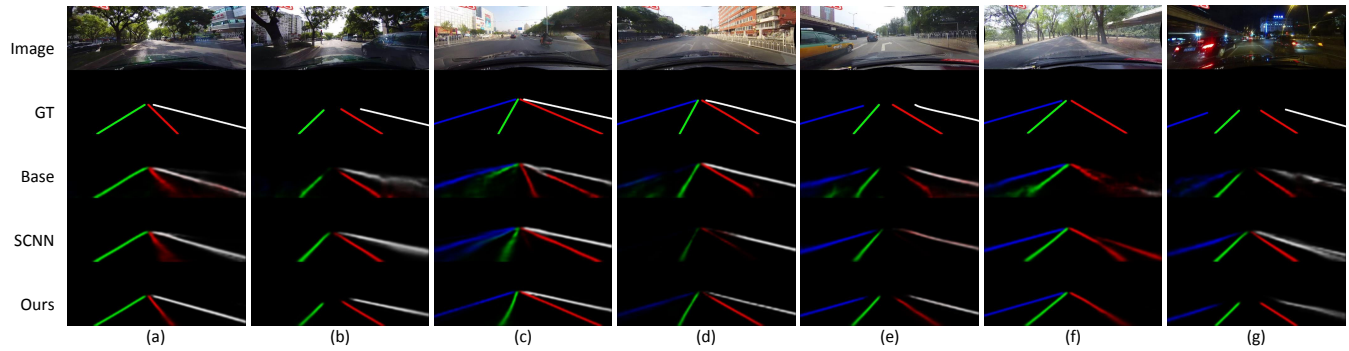

Fig. 4. Visualization results of lane mark segmentation on CULane challenging samples. From left to right, four lane marks are colored with blue, green, red and white respectively while the saturation represents predicted probability.

and false prediction, which contributes to the difference in lane mark segmentation. Compared with the weighted cross entropy, IoU-based loss deals with class imbalance implicitly without any hyper-parameter and efficiently improves lane mark segmentation quality.

TABLE II
EXPERIMENTAL RESULTS OF DIFFERENT LOSS FUNCTIONS

Loss	Normal(D)	Total(D)	Normal(A)	Total(A)
WCE	67.80	50.15	90.56	71.47
Dice	**70.78**	53.36	91.65	73.77
Jaccard	70.63	**53.54**	**91.75**	**74.03**

(3) *Lane Mark Fitting.* We compare our self-supervised lane mark fitting method with the simple max-sampling strategy on the basis of dual-task segmentation network. The baseline method directly samples the pixels with the highest probabilities at each row then applies fitting while self-supervised lane fitting is able to directly output the fitting results. We compare our network trained by only first stage (train_1) or full stages (train_12) with the baseline. Area-metric results of these methods are recorded in TABLE III. Our fitting method raises the overall performance from 74.3% to 74.7% by only first stage training and further to 74.9% after the two-stage full training. It verifies the effectiveness of our simple fitting network and novel self-supervised training strategy, including coarse regression and confidence-based optimization.

TABLE III
AREA-METRIC RESULTS OF LANE LINE FITTING METHODS

Method	Normal	Crowd	Hlight	Shadow	Noline	Arrow	Curve	Night	Total
baseline	91.8	72.4	65.2	**73.2**	46.7	**87.2**	**68.2**	69.7	74.3
train_1	**92.4**	72.1	**66.3**	72.6	49.3	87.2	63.3	70.8	74.7
train_12	91.5	**72.4**	65.6	73.2	**49.4**	87.1	63.1	**71.0**	**74.9**
samples(%)	27.8	23.4	1.4	2.6	11.7	2.6	1.2	20.3	-

C. Comparison to State of the Art

(1) *Lane Mark Segmentation.* The comparing results of IoU metric with the state-of-the-art method SCNN are shown in TABLE IV. It can be observed that in all scenarios our dual-task segmentation network surpasses SCNN by a considerable gap. Some visualization results of the predicted probability maps under challenging circumstances are illustrated in in Fig. 4. Our result has higher localization accuracy with comparatively thinner contours and more centralized region prediction, as shown in the red area in (a) and the white in (b). Besides, our results demonstrate great robustness in confusing scenarios like newly-repaired road patch in (c), deteriorated lane mark in (d), (e), and puzzling curb in (f). Furthermore, we surprisingly notice that our segmentation network excels in adaptively adjusting prediction shape by noticing stop line in the image. For example, in the case of (b), (e) and (g), the annotation clearly confines the extension of lane line according to stop lines ahead and our method is able to grasp this implicit rule with high precision.

TABLE IV
IOU METRIC RESULTS OF LANE MARK SEGMENTATION

Model	Normal	Crowd	Hlight	Shadow	Noline	Arrow	Curve	Night	Total
SCNN [5]	62.2	43.8	41.1	46.1	31.4	56.7	44.4	42.0	47.1
Ours	67.4	47.9	44.9	48.3	32.4	61.7	47.1	44.3	50.9
Δ	+5.2	+4.0	+3.7	+2.1	+1.0	+4.9	+2.7	+2.3	+3.8

(2) *Lane Mark Detection with Fitting.* At the end, we further compare our entire DSSF-net model with several existing methods evaluated on CULane dataset. The F1-score with area-metric on each test subset is shown in TABLE V. With the help of dual-task segmentation and self-supervised fitting network, our DSSF-net outperforms its counterparts in almost all of the scenes. Especially in some difficult scenes like highlight, shadow or night, we achieve more than 4 to 5 percentages of gains. The overall F1 score is 74.9%, which is much higher than the state-of-the-art models. Besides the excellent performance produced, another superiority of our model lies in the end-to-end manner of directly outputting the fitted detection results. The entire processing algorithm can run in a single GTX 1080Ti GPU with the real time performance at about 20Hz. To the best of our knowledge, DSSF-net has achieved the top performance in the large and challenging CULane dataset.

TABLE V

COMPARISONS OF OUR PROPOSED DSSF-NET WITH OTHER EXISTING LANE MARK DETECTION APPROACHES

Method	Normal	Crowd	Hlight	Shadow	Noline	Arrow	Curve	Cross(FP)	Night	Total
FastDraw [29]	85.9	63.6	57.0	59.9	40.6	79.4	65.2	7013	57.8	-
SCNN [5]	90.6	69.7	58.5	66.9	43.4	84.1	64.4	**1990**	66.1	71.6
StripNet [7]	90.8	69.9	60.0	69.7	44.5	85.3	**66.1**	2020	66.9	72.2
R-101-SAD [30]	90.7	70.0	59.9	67.0	43.5	84.4	65.7	2052	66.3	71.8
DSSF-net	**91.5**	**72.4**	**65.6**	**73.2**	**49.4**	**87.1**	63.1	2056	**71.0**	**74.9**

V. CONCLUSION

We present a robust end-to-end lane mark detection network with the ability of outputting the fitted lane lines. We decouple the lane segmentation procedure into two related tasks and apply the IoU-based loss function to tackle the severe category imbalance problem. A simple fitting network equipped with a novel self-supervised training strategy is proposed to replace the traditional hand-crafted fitting algorithms. These designs make the entire inference end-to-end and enable the model with better generalization performance to various road scenes. Comprehensive experimental results in the challenging CULane dataset show that our DSSF-net outperforms the state-of-the-art methods, which verified the effectiveness of our method.

REFERENCES

[1] J. Son, H. Yoo, S. Kim, and K. Sohn, "Real-time illumination invariant lane detection for lane departure warning system," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1816–1824, 2015.

[2] M. Aly, "Real time detection of lane markers in urban streets," in *Intelligent Vehicles Symposium, 2008 IEEE*. IEEE, 2008, pp. 7–12.

[3] Tusimple, "Tusimple Benchmark," 2017. [Online]. Available: http://benchmark.tusimple.ai/#/

[4] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving video database with scalable annotation tooling," *arXiv preprint arXiv:1805.04687*, 2018.

[5] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[6] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.

[7] G. Qu, W. Zhang, Z. Wang, X. Dai, J. Shi, J. He, F. Li, X. Zhang, and Y. Qiao, "StripNet: Towards Topology Consistent Strip Structure Segmentation," in *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, 2018, pp. 283–291.

[8] A. B. Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine vision and applications*, vol. 25, no. 3, pp. 727–745, 2014.

[9] S. Yenikaya, G. Yenikaya, and E. Düven, "Keeping the vehicle on the road: A survey on on-road lane detection systems," *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, p. 2, 2013.

[10] B. Huval, T. Wang, S. Tandon, J. Kiske, W. Song, J. Pazhayampallil, M. Andriluka, P. Rajpurkar, T. Migimatsu, and R. Cheng-Yue, "An empirical evaluation of deep learning on highway driving," *arXiv preprint arXiv:1504.01716*, 2015.

[11] S. Lee, J. Kim, J. S. Yoon, S. Shin, O. Bailo, N. Kim, T.-H. Lee, H. S. Hong, S.-H. Han, and I. S. Kweon, "Vpgnet: Vanishing point guided network for lane and road marking detection and recognition," in *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017, pp. 1965–1973.

[12] X. Li, J. Li, X. Hu, and J. Yang, "Line-CNN: End-to-End Traffic Line Detection With Line Proposal Unit," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2019.

[13] P.-R. Chen, S.-Y. Lo, H.-M. Hang, S.-W. Chan, and J.-J. Lin, "Efficient Road Lane Marking Detection with Deep Learning," *arXiv preprint arXiv:1809.03994*, 2018.

[14] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: an instance segmentation approach," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 286–291.

[15] Y.-C. Hsu, Z. Xu, Z. Kira, and J. Huang, "Learning to cluster for proposal-free instance segmentation," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–8.

[16] H. Jung, J. Min, and J. Kim, "Efficient Lane Detection Algorithm For Lane Departure Detection," in *2013 IEEE Intelligent Vehicles Symposium*. IEEE, 2013, pp. 976–981.

[17] A. Borkar, M. Hayes, and M. T. Smith, "A novel lane detection system with efficient ground truth generation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 365–374, 2012.

[18] H. Tan, Y. Zhou, Y. Zhu, D. Yao, and K. Li, "A novel curve lane detection based on Improved River Flow and RANSA," in *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 2014, pp. 133–138.

[19] B. He, R. Ai, Y. Yan, and X. Lang, "Accurate and robust lane detection based on dual-view convolutional neutral network," in *Intelligent Vehicles Symposium (IV), 2016 IEEE*, vol. 2016-Augus, no. Iv. IEEE, 2016, pp. 1041–1046.

[20] W. Song, Y. Yang, M. Fu, Y. Li, and M. Wang, "Lane Detection and Classification for Forward Collision Warning System Based on Stereo Vision," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5151–5163, 2018.

[21] 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 *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[22] Z. Chen and C. Lian, "PointLaneNet : Efficient end-to-end CNNs for Accurate Real-Time Lane Detection," *Intelligent Vehicles Symposium*, no. Iv, pp. 0–5, 2019.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[24] L.-C. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[25] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

[26] M. Jaderberg, K. Simonyan, A. Zisserman, and Others, "Spatial transformer networks," in *Advances in neural information processing systems*, 2015, pp. 2017–2025.

[27] W. Liu, A. Rabinovich, and A. C. Berg, "Parsenet: Looking wider to see better," *arXiv preprint arXiv:1506.04579*, 2015.

[28] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

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

[30] Y. Hou, Z. Ma, C. Liu, and C. C. Loy, "Learning lightweight lane detection cnns by self attention distillation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 1013–1021.