

Geometric Constrained Joint Lane Segmentation and Lane Boundary Detection

Jie Zhang, Yi Xu*, Bingbing Ni, and Zhenyu Duan

SJTU-UCLA Joint Center for Machine Perception and Inference, Shanghai Jiao Tong University, Shanghai 200240, China
{a135b920,xuyi,nibingbing,ol7650}@sjtu.edu.cn

Abstract. Lane detection is playing an indispensable role in advanced driver assistance systems. The existing approaches for lane detection can be categorized as lane area segmentation and lane boundary detection. Most of these methods abandon a great quantity of complementary information, such as geometric priors, when exploiting the lane area and the lane boundaries alternatively. In this paper, we establish a multiple-task learning framework to segment lane areas and detect lane boundaries simultaneously. The main contributions of the proposed framework are highlighted in two facets: (1) We put forward a multiple-task learning framework with mutually interlinked sub-structures between lane segmentation and lane boundary detection to improve overall performance. (2) A novel loss function is proposed with two geometric constraints considered, as assumed that the lane boundary is predicted as the outer contour of the lane area while the lane area is predicted as the area integration result within the lane boundary lines. With an end-to-end training process, these improvements extremely enhance the robustness and accuracy of our approach on several metrics. The proposed framework is evaluated on KITTI dataset, CULane dataset and RVD dataset. Compared with the state of the arts, our approach achieves the best performance on the metrics and a robust detection in varied traffic scenes.

Keywords: Lane Segmentation; Semantic Segmentation

1 Introduction

Trajectory planning [20] of autonomous driving is an challenging task in the field of computer vision. Lane area segmentation is a crucial issue in trajectory planning which classifies different lanes and generates definite driving areas.

Texture-based approaches are proposed in the early works. Texture features from different color spaces are aggregated to enhance the robustness of lane detection [3][22][24]. Generally, there are homogeneous regions in the lane area, so it is difficult to establish distinguishable feature descriptors for these regions.

Without adequate texture information to rely on, the supplement of lane boundaries is critical to precise detection of lane area. Traditional approaches

* Corresponding author: Yi Xu.

extract boundary information to tackle the problem of homogeneous regions in lane areas, where the high-pass filters are dominantly used [18][4][12][7]. With boundary information extracted, final lane areas are sketched by lane boundaries [9]. However, boundary information is frequently missing due to occlusions and dashing markers. Severely, surrounding shadows and vehicles often introduce irrelevant information that withers the detection performance. In recent years, fully convolutional neural networks (FCN) are put forward [23][6][10][25], where contextual features are self-learned with an encoder-decoder structure to enhance lane segmentation. FCN has achieved much better performance than traditional approaches [3][22][24]. Under ill-defined conditions, it is hard for FCN to effectively provide a significant, unique representation of lane areas.

In fact, there exists a geometric relationship between a lane area and its boundaries: lane areas always lie between lane boundaries, while a lane boundary consists of the outer contour of the lane area. To make use of this relationship, some prior models are advanced through a sequential processing strategy [2][30][35]. Some models extract lane boundaries to greatly reduce the search range for lane detection [8]. Given the lane boundaries, segmentation algorithms are applied to the bounded regions to refine lane labels. Reversely, some models segment lane areas first. Boundary information is then extracted by high-pass filters around segmented lane areas with a tolerance range [35][2]. However, these models treated lane area segmentation and lane boundary detection as two separate subprocesses, which share no information with each other, leading to a loss of geometrical dependency. Moreover, extremely poor performance could happen when the first sub-process is severely interfered by outliers.

To address abovementioned problems, we are motivated to provide a unified solution of lane area segmentation and lane boundary detection with a multi-task learning framework. Rather than simply fuse outputs of different tasks at final decision stage, we apply one shared encoder to the neural network for integrating complementary information of two tasks. Due to the lack of priors and loss of information, single-task approaches cannot achieve desired performance. Additionally, a novel structure called link encoder is appended, which can implicitly extract interrelationship information between lane area and its boundaries. Therefore, the flowed information between two tasks refines the performance for each other. At the classifier layer, the result is generated by the superposition of such a refinement over the original output. As shown in Figure 1, segmentation is interfered severely by outliers and it is unable to recover from error segmentation. In our approach, when lane area segmentation fails to segment hard examples, the other task of lane boundary detection with a good performance could provide valid features and recover the segmentation task from failure. It is the same vice versa. Furthermore, two geometric prior constraints are proposed in our model to regularize the problem of lane detection into well-posed formulation. Given lane boundary detection, we predict the lane area as the area integration result with the lane boundaries as the upper bound and the lower bound. Given extracted lane areas, lane boundaries are predicted as the outer contours. The differences between the prediction results and the ground truth are then formu-

lated as two loss terms to emphasize geometric priors during model training. The geometric constraints are differentiable due to pixel-wise convolution. The overall network is capable of joint-training as an end-to-end implementation. Experimental results on benchmarks demonstrate that our approach outperforms other state-of-art approaches under several metrics.

2 Related Work

Traditional methods of lane segmentation dominantly utilize pixel-level and super-pixel level features. Among the pixel-level features [15][3][2][33], color features robust to shadow interference are extracted for lane segmentation using region growing [3]. Texture features from varied color space are described by histogram peaks and temporal filter responses, then lane areas are generated within flat regions [15]. Alon et al. compute dominant edges based on pixel-wise gradient map and form them as the lane boundaries [33]. With the guidance of these boundaries, a color-based region growing is followed to generate lane areas. Valente et al. [2] extract pixel-level color features and classify them into lane areas first. Then boundaries are introduced to constrain the refinement of lane areas. To handle outlier situations, super-pixel features are preferably used. Li et al. [22] train an AdaBoost classifier with super-pixel color features extracted by Orthogonal Matching Pursuit algorithm to enhance lane detection.

With the development of deep learning methods, semantic segmentation has got impressive results [23][5][27]. Several modified single-task networks are proposed, focusing on embedding extra knowledge into networks [6][10]. Gao et al. [10] advance contour priors and location priors to segment lane region elaborately. Due to the lack of priors and loss of information, multi-task approaches are proposed to tackle this problem by introducing more surrounding constraints. Oliveira et al. [25] train a joint classification, detection and semantic segmentation network with a shared encoder. With a joint-training manner, the final lane area is generated by better features containing more surrounding details. However, inherent connections are not mined between multiple outputs, making it difficult to explain the mechanism behind the network structure.

Plenty of works utilize high-pass filters for boundary detection [4][12]. Haloi et al. [12] combine responses from 2nd- and 4th-order filters to obtain lane boundary features with adaptive thresholding. Aly et al. filter inverse perspective mapping (IPM) image with 2D Gaussian kernels [4]. Kortli et al. [18] and Bergasa et al. [7] detect edges with a canny kernel and the Otsu method.

However, the precision of lane boundary detection suffers from illumination variations, noises and cluttered background. Some deep learning techniques are developed to improve the performance. Overfeat detector is proposed to integrate recognition, localization and detection using convolution [29]. Later, Huval et al. [14] modify Overfeat structure to handle lane boundary detection and vehicle detection at the same time. However, these methods are sensitive to surrounding objects. So, Li et al. [21] feed the features extracted by convolution layers into a recurrent neuron layer as a sequence, where the spatial continuity constraint is

used to regularize the result of lane detection. Kim et al. [16] propose a simpler but effective network structure. They finetune the network with a pretrained VGG network to generate detection results. These approaches require extra data for sufficient pretraining, and they are sensitive to cluttered background.



Fig. 1. Reciprocal constraints with geometric relation. **Left:** With an image input, traditional methods generate a binary segmentation mask for lane areas (green) or lane boundaries (red), which are severely affected by outlier situations. **Right:** Our approach introduces a geometric constraint into a multi-task network, which is capable to restore the missing lane area and lane boundaries (blue) mutually.

3 Methodology

3.1 Overview of Multi-task Framework

For human perception, lane area is always inseparable from the judgment of its boundaries. However, the existing lane detection methods dominantly rely on a single-task network to independently train lane segmentation and lane boundary detection, completely ignoring the inherent geometric constraints between two tasks. Simple multi-task networks like MultiNet [32] are developed to combine tasks together, such as classification, detection and segmentation, without investigating the inherent relationship between tasks. There are two major problems in these state-of-the-art methods: the loss of interrelationship between multiple training tasks and the lack of geometric priors for well-posed formulation. As a result, they are always stuck in detection failure on hard examples.

Inspired by this observation, we propose a multi-task learning framework to provide a unified solution of lane segmentation and lane boundary detection. The network architecture is illustrated in Figure 2, which consists of an encoder network and a decoder network, as a kind of fully convolutional network (FCN). Rather than conducting segmentation and detection with two separate networks, the proposed framework conducts two tasks with one shared encoder network and two separated decoders. To classify pixels into binary labels, each decoder is followed by a sigmoid classifier. Specifically, each decoder is connected to a link encoder to stream complementary information between two tasks and thus the features of two decoders could be reciprocally refined.

To achieve well-posed formulation of the multi-task learning framework, we propose a novel loss function by introducing the inherent geometric priors between tasks, as assumed that the lane boundary is predicted as the outer contour

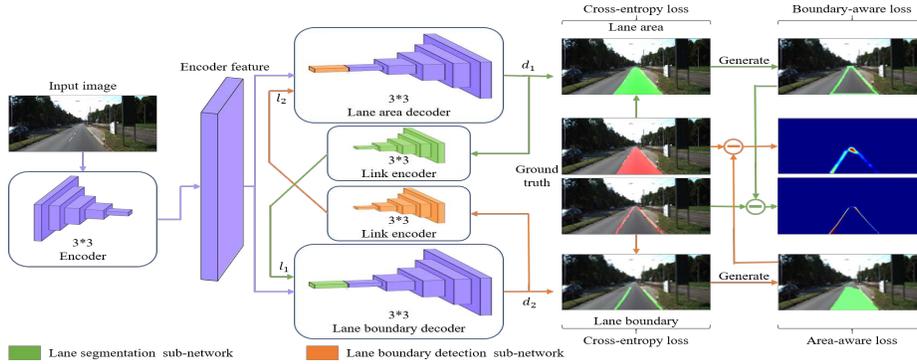


Fig. 2. The proposed multi-task framework. Input images are fed into a shared encoder, which extracts the critical features for lane segmentation and lane boundary detection. Two inter-link encoders connected to each decoder provide complementary information for tasks. The overall performance is enhanced by introducing a structure loss, assuming that the lane boundary is predicted as the outer contour of the lane area while the lane area is predicted as the area integration result within the lane boundaries.

of the lane area while the lane area is predicted as the area integration result within the lane boundaries. These geometric priors are critical to find a consistent solution of lane segmentation and lane boundary detection. With an end-to-end training process, these improvements extremely enhance the robustness and accuracy of our approach on several metrics

3.2 Critical Feature Extraction Using A Shared Encoder

To illustrate the activated regions by encoders of lane segmentation and lane boundary detection in a single task learning framework respectively, we visualize each activation map using a heat map. As shown in Figure 3, the lanes with the similar textures are all emphasized by lane segmentation [19], which incurs the ambiguity problem of lane detection. Also, some background regions are activated. In contrast, the edges in background incur a more severe outlier problem for lane boundary detection.

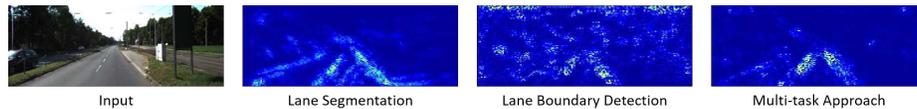


Fig. 3. Activation map of encoders. Activation maps are generated from final convolution layers of encoders. Pixel color indicates task-related saliency with respect to input images.

A shared encoder is proposed to greatly reduce the ambiguity problem and outliers, because the features critical to the performance improvement of both tasks have been emphasized during network training process. As compared in Figure 3, a clearer lane extraction is obtained while using the shared encoder.

3.3 Complementary feature extraction using an inter-link encoder



Fig. 4. Top: Original image and FP+FN area (red) generated by the initial network. **Bottom:** The image of absolute difference before and after the refinements by inter-link encoders. As noted from the heat maps, the initial false positive results are effectively suppressed and the originally missed lane pixels are well restored, as emphasized in warm color.

The shared encoder puts much attention to the features critical to the overall performance. However, some important features for one task might be suppressed once they are not so critical in the other task. For example, lane area segmentation puts much emphasis on the fine-grained texture features for accurate pixel-wise label, while lane boundary detection prefers edge-like features.

An inter-link encoder is put forward to stream complementary information between two tasks and thus the features of two decoders could be reciprocally refined. As shown in Figure 2, decoders initially receive the feature f and output preliminary results as the inputs of inter-link encoders. Then these decoders generate final results with the refined features, where features f are complemented with inter-link encoders outputs l_1 and l_2 using simple concatenation. Thus decoders actually do forward pass twice. These refined features enhance the representation of lanes. It is expected to improve the performance of lane segmentation and lane boundary detection in a unified way.

At the first row in Figure 4, segmentation results are generated without inter-link encoders, where the red regions indicate false positive and false negative results. The bottom row shows absolute difference image before and after refinements from inter-link encoders, where pixels are highlighted using heat map to indicate the absolute difference of segmentation confidence. As emphasized with warm color, with addition information provided by inter-link encoders, it is notable that the originally false positive results are then effectively suppressed, and the originally missed lane pixels are well restored.

3.4 Geometry Constrained Structure Loss

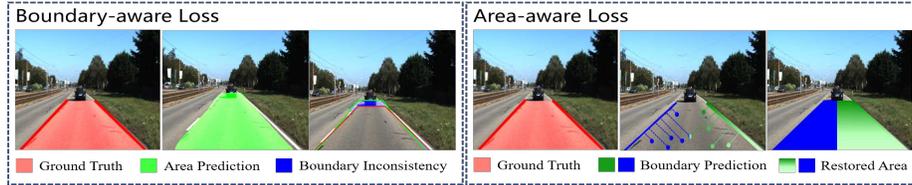


Fig. 5. Boundary-aware loss and area-aware loss. **Left:** An illustration of our boundary-aware loss. The blue area indicates boundary inconsistency. **Right:** An illustration of our area-aware loss. Different intensities in prediction areas indicate different prediction confidence. The difference between restored area and ground truth indicates the area aware loss.

Boundary Aware Loss for Lane Area Segmentation. With cross-entropy set up as a loss function for lane area segmentation, it results in groups of pixels with false labels due to high ambiguity. We introduce a boundary-aware loss for lane area segmentation, assuming that there exists a consistency between the boundaries of segmented lane areas with the ground truth of lane boundaries.

It is noted that a slight deviation of lane boundaries from ground truth could produce an extremely large loss with pixel-wise comparison, as illustrated in Figure 5. Therefore, we employ IoU loss [26] to measure boundary inconsistency. Accordingly a slight deviation would result in a small IoU loss, which ensures convergence. Let \mathcal{I} denote the set of pixels in the image. For every pixel p in the pixel set \mathcal{I} , y_p corresponds to its output probability. And $g = \{0, 1\}^{M \times N}$ is the ground truth for the set \mathcal{I} . Here, M and N are the height and width of the image. By masking lane segmentation results with the lane area bounded by the ground truth of lane boundaries, our boundary-aware loss l_{ba} can be defined as:

$$\text{IoU} = \frac{\sum_{p \in \mathcal{I}} (y_p \times g_p)}{\sum_{p \in \mathcal{I}} (y_p + g_p - y_p \times g_p)}, \quad (1)$$

$$l_{ba} = 1 - \text{IoU}, \quad (2)$$

where \times denotes a pixel-wise multiplication.

Two consistency constraints are imposed to enhance the results of lane segmentation. The cross entropy loss term l_{lce} measures the consistency between the segmented area and its ground truth. Additionally, the loss term l_{ba} measures the consistency between the boundaries of the segmented area and the lane boundary ground truth. Correspondingly, the loss function l_{lt} to measure the total error of lane segmentation is updated as:

$$l_{lt} = l_{lce} + \lambda_1 \times l_{ba}, \quad (3)$$

where λ_1 is a constant for balancing two losses. Here we set λ_1 as 0.5. With only pixel-wise linear calculation involved, l_{ba} is fully differentiable.

Area-aware Loss For Lane Boundary. Compared with lane area segmentation, lane boundary detection suffers more from the higher missing rate due to the lower Signal Noise Ratio around boundaries. Motivated by the geometric prior that the lane area is the area integration result with lane boundaries as the upper and lower bounds, an area-aware loss is proposed to measure the difference between the lane area restored from detected lane boundary and lane area ground truth.

Our area-aware loss function is expressed as:

$$l_{aa} = \sum_{\mathcal{G}(p)=1} [1 - I_r(p)], \quad (4)$$

where \mathcal{G} is the pixel-wise label set of lane area ground truth, and $\mathcal{G}(p) = 1$ denotes that pixel p belongs to the lane area, and $I_r(p)$ is the calculated probability of pixel p belonging to the restored lane area. The loss function l_{mt} to measure the error of lane boundary detection is defined as

$$l_{mt} = l_{mce} + \lambda_2 \times l_{aa}, \quad (5)$$

where l_{mce} is the cross-entropy loss measuring the consistency between the detected lane boundary and its ground truth in a complementary way.

Pixels with strong spatial correlation always present similar intensity distribution, therefore we estimate pixel intensities in the restored lane area directly from the closest pixels on lane boundaries. Denote pixels of two boundary lane boundaries as pixel set \mathcal{B} . For pixel p between lane boundary ground truth, its probability belonging to lane area is equal to the probability of the closest pixel on lane boundaries, which is computed as:

$$I_r(p) = \frac{1}{n} \sum_{j=1}^n I_b(v_j), \quad (6)$$

$$v_j = \arg \min_{m_i} [d(p, m_i)] \quad m_i \in \mathcal{B}, \quad (7)$$

where $d(x, y)$ is the Euclidean distance between pixels x and y , $I_b(v)$ is the pixel probability in boundary detection map.

Computing the restored lane area from Equation 6 and 7, we reform the loss function l_{aa} as:

$$l_{aa} = \sum_{\mathcal{G}(p)=1} [1 - I_r(p)] = \sum_{\mathcal{G}(p)=1} \left\{ 1 - \frac{1}{n} \sum_{j=1}^n I_b \left\{ \arg \min_{m_i} [d(p, m_i)] \right\} \right\} \quad m_i \in \mathcal{B}, \quad (8)$$

Thus, the integrated loss function is finally formulated as below:

$$l = l_{lce} + \lambda_1 \times l_{ba} + l_{mce} + \lambda_2 \times l_{aa}. \quad (9)$$

3.5 Training Details

Our framework is designed to be fully-convolutional and differentiable, thus it could be trained in an end-to-end manner. In this section, we mainly focus on implementation details of training process.

The shared encoder network is initialized by ImageNet [28] with VGG structure [31]. First, we start with training single lane segmentation subnetwork. Secondly, we turn to both subnetworks of lane area segmentation and boundary detection, which is trained without inter-link encoder structure. Finally, our multi-task learning framework is overall retrained with inter-link encoders added.

We concatenate an all-zero tensor to the output of the shared encoder, so that the input feature dimension of decoders remains the same during the iterative training procedure. The overall framework utilizes a batch normalization with the batch size of 3. To avoid overfitting, a dropout layer [13] is adopted with a rate of 0.2. We use the Adam optimizer [17] and pretrain the lane segmentation and lane boundary detection subnetworks with a learning rate of 10^{-3} . For multi-task framework training process, learning rate is set as 10^{-4} until convergence.

4 Experiment

We evaluate our approach on two lane segmentation datasets: KITTI dataset [11], Road-Vehicle dataset (RVD) [8] and CULane dataset [34]. Approaches are coded and evaluated by Tensorflow [1]. Processing time is evaluated on GeForce GTX TITAN with $160 * 320$ input images.

4.1 Dataset and Evaluation

The KITTI dataset contains 289 training images and 290 testing images, including four subsets of road scenes: urban marked road (UM), urban multiple marked road (UMM), urban unmarked road (UU) and URBAN ROAD (the union of the former three). UM is defined as marked roads with two lanes, while UMM consists of the roads with multiple lanes. UU stands for roads without lane markings and contains one lane only.

The RVD dataset contains more than 10 hours of traffic scenarios with multiple sensors under different weather and road conditions, including highway scenes, night scenes and rain scenes. There are over 10,000 manually labeled images in this dataset, which are divided into different scenes with respect to surrounding conditions such as weather and illumination.

The CULane dataset contains 133,235 images extracted from 55 hours of traffic videos, which is divided into 88,880 images for training set, 9,675 for validation set and 34,680 for testing set. The test set is split into 8 subsets based on their scenes to demonstrate the robustness of different network structures. This newly released dataset contains lane boundary ground truth only, so we generate lane area ground truth according to bounded areas of lane boundaries.

To evaluate lane segmentation results, we follow the classical pixel-wise segmentation metrics with precision (P), recall (R), F1-measure and IoU score. The

metrics with the removal of foreshortening effects are not considered, because inverse perspective mapping incurs distortions in the ground truth.

For lane boundary detection, we evaluate the performance with a pixel-wise metric. On KITTI dataset, when the distance of detected lane boundary and ground truth is smaller than a threshold (1.5% of image diagonal), the detected lane boundary is regarded as a true positive (TP). While on the CULane dataset, we follow its metric for fair comparison. When the IoU of detected lane boundary and ground truth is larger than 0.5 threshold, the detected boundary is regarded as a true positive (TP) [34]. The same for all the methods for comparison. The final results are evaluated with precision (P), recall (R) and F1-measure.

4.2 Results and Discussion

Our experiments are designed as two parts. First, we compare our lane area segmentation approach with state-of-the-art methods on KITTI, CULane and RVD dataset. Then, to demonstrate the effectiveness of our multi-task structure, lane boundary detection results are evaluated on KITTI and CULane dataset.

Lane Segmentation Results on KITTI. The proposed network is first compared with state-of-the-art approaches (including the SegNet [5], the U-Net [27] and the Up-Conv-Poly [25]) on KITTI dataset. Table 1 shows the overall results. Compared with the baseline approach [5], our methods are superior to it. Joint-training improves performance even without an inter-link encoder and structure loss functions. Benefited from the investigation of inherent inter-relationship between tasks, our multi-task framework obtains a better feature representation than a single-task network and boosts performance further.

Table 1. Lane segmentation results on URBAN_ROAD KITTI dataset. 'multi-task', 'loss', 'link' and 'link+loss' denote networks without losses or link structure, with losses only, with link structure only and with both losses and link structure.

Method	Runtime (ms)	F1	P	R	IoU
SegNet [5]	19.43	0.894	0.899	0.889	0.808
U-Net [27]	16.92	0.909	0.875	0.945	0.833
Up-Conv-Poly [25]	11.27	0.921	0.921	0.922	0.854
Ours (multi-task)	34.70	0.920	0.911	0.929	0.852
Ours (loss)	34.70	0.925	0.917	0.933	0.861
Ours (inter-link)	41.28	0.927	0.925	0.930	0.865
Ours (inter-link+loss)	41.28	0.933	0.936	0.930	0.874

Note that our approach also outperforms U-Net and Up-Conv-Poly with a gain of 4.1% and 2.0% on IoU score. Both approaches connect encoder layers with decoder layers, which make decoders receive the same scale information from encoders directly. Our multi-task network better captures the dependency of geometric structure of lanes and markers. We also evaluate approaches on several different traffic scenes in Table 2. Results show that our approach is robust to scenario changes.

Table 2. Lane segmentation results on KITTI subsets(UM/UMM/UU)

Method	UM Lane				UMM Lane				UU Lane			
	F1	P	R	IoU	F1	P	R	IoU	F1	P	R	IoU
SegNet [5]	0.926	0.920	0.933	0.863	0.927	0.905	0.950	0.864	0.859	0.905	0.817	0.752
U-Net [27]	0.924	0.876	0.977	0.859	0.929	0.877	0.988	0.867	0.904	0.907	0.902	0.826
Up-Conv-Poly [25]	0.936	0.929	0.944	0.881	0.953	0.943	0.963	0.910	0.918	0.936	0.901	0.849
Ours(Multi-task)	0.932	0.913	0.951	0.872	0.953	0.932	0.974	0.910	0.919	0.933	0.906	0.850
Ours(Loss)	0.950	0.940	0.961	0.906	0.952	0.932	0.972	0.908	0.922	0.927	0.917	0.855
Ours(Link)	0.948	0.942	0.954	0.901	0.958	0.947	0.970	0.926	0.921	0.932	0.911	0.855
Ours(Link+Loss)	0.954	0.957	0.951	0.912	0.962	0.957	0.967	0.927	0.931	0.947	0.916	0.871

We study the influence of inter-link encoders and structure loss functions in our model. Note that we achieve 86.1% IoU score on our structure-loss-only approach and 86.5% IoU score in our inter-link-only approach. With inter-link encoders and losses added, our final approach (link+loss) achieves 87.4% on IoU and 93.3% on F1-measure. The individually applied structure loss and inter-link encoders play a crucial role in promoting segmentation results. Figure 7 shows some lane area segmentation results obtained by our approach, Up-Conv-Net and U-Net approaches on the KITTI dataset. Our approach effectively handles hard cases such as vanishing boundaries on the first two columns of Figure 7.



Fig. 6. The IoU metric evaluated with a single image in KITTI dataset. The blue line is our approach with structure loss functions while the orange line is our approach without structure loss.

To demonstrate the efficiency of structure loss, our approaches with and without structure loss are evaluated by several examples of single images in Figure 6. We randomly pick 100 images in KITTI dataset and calculate the IoU score for both approaches. The evaluation results reveal that, the introduction of structure loss presents higher robustness to disturbance.

Lane Segmentation Results on CULane dataset. We also evaluate lane segmentation on a newly published CULane dataset. The test set is divided into 8 different scenes: Arrow, Crowded, Curve, Dazzle light, Night, No line, Normal and Shadow. The overall performance is also shown in the last column.

Experiment results are shown in Table 3. It is significant that our method outperforms state-of-the-art methods on all 8 subsets and achieves 90.2% F1-measure and 82.4% IoU on the overall dataset, demonstrating that our method is more robust to handle various traffic scenes than state-of-the-art methods. Also, our method achieves a remarkable improvement on 4 subsets (Arrow, Crowd,

Shadow and Normal). This is because our method could capture lane boundary structure from cluttered backgrounds. The well-extracted boundary features provide complementary information to effectively suppress error segmentation.

Table 3. Lane segmentation results on CULane dataset

Method	Arrow		Crowd		Curve		Dazzle light		Night	
	F1	IoU								
U-Net [27]	0.909	0.832	0.906	0.828	0.865	0.763	0.873	0.775	0.875	0.779
SegNet [5]	0.906	0.829	0.899	0.818	0.846	0.733	0.867	0.765	0.847	0.735
Up-Conv-Poly [25]	0.899	0.816	0.903	0.823	0.864	0.761	0.886	0.796	0.871	0.772
Ours	0.922	0.856	0.913	0.841	0.868	0.765	0.891	0.803	0.876	0.781

Method	No line		Normal		Shadow		Total	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU
U-Net [27]	0.802	0.671	0.934	0.887	0.825	0.703	0.892	0.810
SegNet [5]	0.800	0.668	0.932	0.873	0.835	0.711	0.883	0.794
Up-Conv-Poly [25]	0.813	0.685	0.935	0.878	0.832	0.711	0.892	0.807
Ours	0.814	0.686	0.949	0.903	0.872	0.774	0.902	0.824

Lane Segmentation Results on RVD dataset. Furthermore, we evaluate lane segmentation on the RVD dataset. As mentioned in 4.1, this dataset contains three different scenes: Highway, Night and Rainy & Snowy Day. Besides SegNet, U-Net and Up-Conv-Poly, we evaluate the performance of CMA method [8].

Table 4. Lane segmentation results on RVD dataset

Method	Highway				Night				Rainy & Snowy Day			
	F1	P	R	IoU	F1	P	R	IoU	F1	P	R	IoU
U-Net [27]	0.972	0.972	0.973	0.947	0.946	0.915	0.980	0.898	0.971	0.963	0.979	0.944
SegNet [5]	0.978	0.969	0.987	0.957	0.968	0.962	0.975	0.940	0.971	0.965	0.978	0.944
Up-Conv-Poly [25]	0.979	0.977	0.982	0.960	0.974	0.972	0.977	0.950	0.971	0.979	0.963	0.944
CMA [8]	0.989	0.989	0.989	/	0.976	0.976	0.977	/	0.974	0.968	0.980	/
Ours	0.990	0.987	0.992	0.979	0.988	0.989	0.987	0.981	0.977	0.973	0.981	0.956

Overall results are presented in Table 4. Note that CMA only extracts end-points of two lane boundaries to segment lane area. It enforces a rigid geometric assumption, and thus fails to segment curve lanes. In contrast, the geometric priors introduced in our network are more applicable in various scenes, achieving a significant improvement over all the metrics. Although the performance on Highway is similar due to clear background, we dramatically improve the performance in other scenes, especially in night scenarios. With a better representation of boundary information and geometry constraints, illumination variation and image degradation are well-handled by our approach.

Lane boundary Detection Results. In addition to lane segmentation, we also evaluate the effectiveness of lane boundary detection on KITTI and CULane dataset. We test several approaches with manually labeled lane boundary ground truth, and present the performance of SegNet [5], SegNet-Ego-Lane [16] and SCNN [34]. We also report our approach only with the cross-entropy loss to emphasize the effectiveness of structure loss function on KITTI dataset. Some lane boundary detection results are also shown in Figure 7.

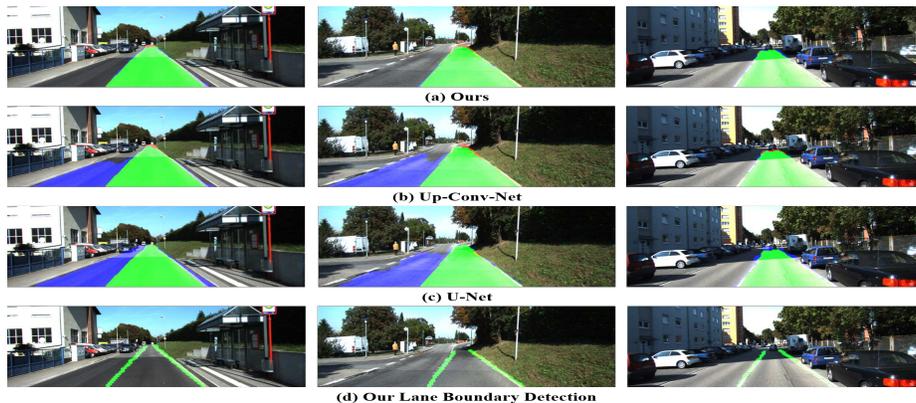


Fig. 7. Lane area segmentation and lane boundary detection results on KITTI dataset. Green corresponds to true positives, blue to false positives and red to false negatives.

Table 5. Lane boundary results on KITTI dataset

Method	F1	P	R
SegNet [5]	0.691	0.865	0.575
SegNet-Ego-Lane [16]	0.522	0.411	0.714
Ours (inter-link)	0.788	0.806	0.771
Ours (inter-link+loss)	0.831	0.838	0.824

The results on KITTI dataset are provided in Table 5. As for precision, SegNet yields slightly better than us. However, SegNet has a extremely low recall rate, which indicates that SegNet misses plenty of true positives. Our approach achieves the highest recall rate than other approaches, as well as the F1 measure. The ablation analysis of our approaches indicates that the lane area-aware loss function dramatically improves the performance of lane boundary detection. We gain 3.2% on precision, 5.3% on recall and 4.3% on F1 measure.

Table 6. Lane boundary results on CULane dataset (F1-measure)

Method	Normal	Crowded	Night	No line	Shadow	Arrow	Dazzle light	Curve	Total
SegNet [5]	0.792	0.617	0.496	0.145	0.294	0.710	0.389	0.456	0.572
SegNet-Ego-Lane [16]	0.754	0.620	0.578	0.177	0.310	0.714	0.476	0.430	0.584
SCNN [34]	0.883	0.753	0.686	0.365	0.593	0.821	0.531	0.594	0.720
Ours	0.897	0.765	0.687	0.351	0.655	0.822	0.674	0.632	0.731

The results on CULane dataset are shown in Table 6. Note that our method outperforms state-of-the-art methods on 7 subsets. State-of-the-art methods have worse performance mainly due to the image degradation and the missing of lane boundaries. As for the image degradation problem, our subnetwork is able to extract better boundary features with our area-aware loss. So we dramatically improve performance on Night, Dazzle Light and Shadow subsets, where image quality is affected severely by illumination conditions. For unseen

lane boundary problem, it is extremely difficult to extract enough boundary features for boundary detection. Although SCNN introduces context information for boundary detection, various scenes contain extremely different context information, resulting in inaccuracy of lane detection results. Meanwhile, our inter-link structure utilizes more robust geometric relationship between lane areas and boundaries, which constrains each other for better performance.

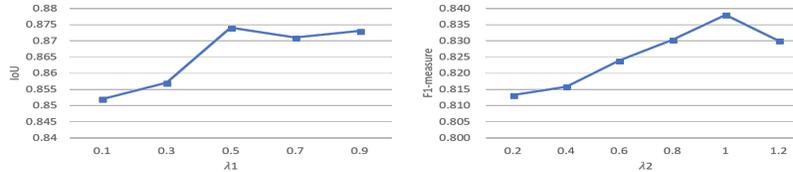


Fig. 8. Parameters λ_1 and λ_2 validation experiments on KITTI

4.3 Parameter Study

To choose optimal parameters λ_1 and λ_2 , parameter study is performed on 10-fold cross validation set. The performance of different λ_1 values is compared by IoU score of lane area segmentation while λ_2 performance are evaluated by F1-measure of lane boundary detection. The final results are shown in Figure 8, where both parameters are chosen at regular intervals. Although λ_1 larger than 0.5 achieves similar IoU score, the experiment shows large λ_1 is sensitive to hyper-parameters. So λ_1 is set to 0.5. And λ_2 is set to 1.0 for the best performance.

5 Conclusion

We propose a multi-task learning framework to jointly address the problems of lane segmentation and lane boundary detection. In this framework, a shared encoder and an inter-link encoder structure are proposed, whose benefits for the boost of detection precision have been proved by experiments. In addition, we come up with two novel loss functions which are established to be applicable to a more general traffic scene. The proposed method is compared with state-of-the-art ones on KITTI and RVD dataset, and shows a leading performance.

6 Acknowledgement

This work was supported by National Science Foundation of China 61671298 and STCSM 17511105400, 18DZ2270700. This work was supported by SJTU-UCLA Joint Center for Machine Perception and Inference. The work was also partially supported by NSFC (U1611461, 61502301, 61521062), Chinas Thousand Youth Talents Plan, the 111 project B07022 and MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University, China.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), <http://tensorflow.org/>, software available from tensorflow.org
2. Alon, Y., Ferencz, A., Shashua, A.: Off-road path following using region classification and geometric projection constraints. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA. pp. 689–696 (2006), <https://doi.org/10.1109/CVPR.2006.213>
3. Álvarez, J.M., López, A.M., Baldrich, R.: Shadow resistant road segmentation from a mobile monocular system. In: Pattern Recognition and Image Analysis, Third Iberian Conference, IbPRIA 2007, Girona, Spain, June 6-8, 2007, Proceedings, Part II. pp. 9–16 (2007)
4. Aly, M.: Real time detection of lane markers in urban streets. CoRR **abs/1411.7113** (2014), <http://arxiv.org/abs/1411.7113>
5. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2017)
6. Barnes, D., Maddern, W., Posner, I.: Find your own way: Weakly-supervised segmentation of path proposals for urban autonomy. In: 2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017. pp. 203–210 (2017), <https://doi.org/10.1109/ICRA.2017.7989025>
7. Bergasa, L.M., Almeria, D., Almazán, J., Torres, J.J.Y., Arroyo, R.: Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, June 8-11, 2014. pp. 240–245 (2014), <https://doi.org/10.1109/IVS.2014.6856461>
8. Chen, S., Zhang, S., Shang, J., Chen, B., Zheng, N.: Brain inspired cognitive model with attention for self-driving cars. CoRR **abs/1702.05596** (2017), <http://arxiv.org/abs/1702.05596>
9. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981), <http://doi.acm.org/10.1145/358669.358692>
10. Gao, J., Wang, Q., Yuan, Y.: Embedding structured contour and location prior in siamesed fully convolutional networks for road detection. In: 2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017. pp. 219–224 (2017), <https://doi.org/10.1109/ICRA.2017.7989027>
11. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The KITTI dataset. *I. J. Robotics Res.* **32**(11), 1231–1237 (2013), <https://doi.org/10.1177/0278364913491297>
12. Haloi, M., Jayagopi, D.B.: Vehicle local position estimation system. CoRR **abs/1503.06648** (2015), <http://arxiv.org/abs/1503.06648>
13. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. CoRR **abs/1207.0580** (2012), <http://arxiv.org/abs/1207.0580>

14. Huval, B., Wang, T., Tandon, S., Kiske, J., Song, W., Pazhayampallil, J., Andriluka, M., Rajpurkar, P., Migimatsu, T., Cheng-Yue, R., Mujica, F., Coates, A., Ng, A.Y.: An empirical evaluation of deep learning on highway driving. CoRR **abs/1504.01716** (2015), <http://arxiv.org/abs/1504.01716>
15. Katramados, I., Crumpler, S., Breckon, T.P.: Real-time traversable surface detection by colour space fusion and temporal analysis. In: Computer Vision Systems, 7th International Conference on Computer Vision Systems, ICVS 2009, Liège, Belgium, October 13-15, 2009, Proceedings. pp. 265–274 (2009)
16. Kim, J., Park, C.: End-to-end ego lane estimation based on sequential transfer learning for self-driving cars. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, Honolulu, HI, USA, July 21-26, 2017. pp. 1194–1202 (2017), <https://doi.org/10.1109/CVPRW.2017.158>
17. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. CoRR **abs/1412.6980** (2014), <http://arxiv.org/abs/1412.6980>
18. Kortli, Y., Marzougui, M., Bouallegue, B., Bose, J.S.C., Rodrigues, P., Atri, M.: A novel illumination-invariant lane detection system. In: 2017 2nd International Conference on Anti-Cyber Crimes (ICACC). pp. 166–171 (March 2017)
19. Kotikalapudi, R., contributors: keras-vis. <https://github.com/raghakot/keras-vis> (2017)
20. Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J.Z., Langer, D., Pink, O., Pratt, V.R., Sokolsky, M., Stanek, G., Stavens, D.M., Teichman, A., Werling, M., Thrun, S.: Towards fully autonomous driving: Systems and algorithms. In: IEEE Intelligent Vehicles Symposium (IV), 2011, Baden-Baden, Germany, June 5-9, 2011. pp. 163–168 (2011)
21. Li, J., Mei, X., Prokhorov, D.V., Tao, D.: Deep neural network for structural prediction and lane detection in traffic scene. IEEE Trans. Neural Netw. Learning Syst. **28**(3), 690–703 (2017), <https://doi.org/10.1109/TNNLS.2016.2522428>
22. Li, J., Jin, L., Fei, S., Ma, J.: Robust urban road image segmentation. In: Proceeding of the 11th World Congress on Intelligent Control and Automation. pp. 2923–2928 (June 2014)
23. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. CoRR **abs/1411.4038** (2014), <http://arxiv.org/abs/1411.4038>
24. Lu, K., Li, J., An, X., He, H.: A hierarchical approach for road detection. In: 2014 IEEE International Conference on Robotics and Automation, ICRA 2014, Hong Kong, China, May 31 - June 7, 2014. pp. 517–522 (2014), <https://doi.org/10.1109/ICRA.2014.6906904>
25. Oliveira, G.L., Burgard, W., Brox, T.: Efficient deep models for monocular road segmentation. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2016, Daejeon, South Korea, October 9-14, 2016. pp. 4885–4891 (2016), <https://doi.org/10.1109/IROS.2016.7759717>
26. Rahman, M.A., Wang, Y.: Optimizing intersection-over-union in deep neural networks for image segmentation. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., Isenberg, T. (eds.) Advances in Visual Computing. pp. 234–244. Springer International Publishing, Cham (2016)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). LNCS, vol. 9351, pp. 234–241. Springer (2015), <http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a>, (available on arXiv:1505.04597 [cs.CV])

28. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* **115**(3), 211–252 (2015)
29. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. *CoRR* **abs/1312.6229** (2013), <http://arxiv.org/abs/1312.6229>
30. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. *CoRR* **abs/1312.6034** (2013), <http://arxiv.org/abs/1312.6034>
31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *CoRR* **abs/1409.1556** (2014), <http://arxiv.org/abs/1409.1556>
32. Teichmann, M., Weber, M., Zöllner, J.M., Cipolla, R., Urtasun, R.: Multinet: Real-time joint semantic reasoning for autonomous driving. *CoRR* **abs/1612.07695** (2016), <http://arxiv.org/abs/1612.07695>
33. Valente, M., Stanciulescu, B.: Real-time method for general road segmentation. In: *IEEE Intelligent Vehicles Symposium, IV 2017*, Los Angeles, CA, USA, June 11–14, 2017. pp. 443–447 (2017), <https://doi.org/10.1109/IVS.2017.7995758>
34. Xingang Pan, Jianping Shi, P.L.X.W., Tang, X.: Spatial as deep: Spatial cnn for traffic scene understanding. In: *AAAI Conference on Artificial Intelligence (AAAI)* (February 2018)
35. Zhang, G., Zheng, N., Cui, C., Yang, G.: An efficient road detection method in noisy urban environment. In: *Intelligent Vehicles Symposium*. pp. 556–561 (2009)