Appendix for IFPNet: Integrated Feature Pyramid Network with Fusion Factor for Lane Detection

Zinan Lv[1] Dong Han[1] Wenzhe Wang[2,3] Cheng Chen[1]

[1]Harbin Engineering University [2]Zhejiang University [3]Westlake University

1. More Ablation Results

1.1. Ablation Study on IFP

1.1.1 Performance on other Structure

Compared to the traditional FPN [4], our IFP can balance the finer-scale features and global semantics for better prediction of lane lines, which can also be applied to other networks easily. To study the effectiveness and portability of our net, we add our IFP to other open-source networks, *e.g.* ConaLaneNet [5], CLRNet [11]. We conduct the experiments on the benchmark dataset CULane [6] and keep other training configurations remain unchanged with the same backbone ResNet18. As shown in Table 1, when IFP is added to ConaLaneNet and CLRNet, the total F1 score increases significantly, which demonstrates the effectiveness of the IFP module. In particular, the F1 scores of "Hlight", "Shadow", "Curve" and "Night" will show a great improvement, proving that our IFP can combine the finer-scale features and global semantics to detect lanes in a better manner. In conclusion, our IFP has two main advantages: (1) Capable of combining local and global information to balance the finer-scale features and global semantics. (2) A lightweight plug-and-play module with high portability.

1.1.2 Refining Shape

A key hyper-parameter in the IFP module is the refining shape when fusing the high- and low-scale information. In this subsection, we evaluate the performance of IFP with different refining shapes. Experiments on the benchmark dataset are conducted with other training configurations remaining the same. According to the results shown in Table 2, we set the refining shape to 20×50 in our manuscript.

1.1.3 Attention Method

In order to further capture the spatial dependencies of any two locations in the feature map and obtain long-range context dependency information, we employ a self-attentive mechanism [8] to further integrate the global contextual information. In this subsection, we study the influence of different self-attention mechanisms. As shown in Table 3, the non-local module performs the best on the benchmark dataset CULane and it is adopted in IFP. The performance of other attention modules is also evaluated in Table 3.

1.2. Ablation Study on RIoU

We propose RIoU to help the regression of lane lines. In order to further verify the effectiveness and portability of RIoU, we conduct experiments on the open-source networks LaneATT [7] and CLRNet [11]. Experiments are conducted on the benchmark dataset CULane [6] with other configurations remaining unchanged according to the original code provided by the authors. As shown in Table 4, RIoU loss improves the F1 score of our proposed IFPNet by 0.21%. When added to LaneATT and CLRNet, the performance of the networks is better than the two original open-source ones. In particular, the result of mF1 gets a great improvement of 0.85% on the LaneATT. All these results prove that it can better help the regression of lane lines. The effectiveness and portability of our proposed RIoU are thus evaluated.

2. Interesting Things about CULane

After careful calibration of the benchmark dataset CULane [6], we found some inappropriate aspects of it. The issues can be divided into two main aspects, which will be discussed in the following of this section.

2.1. Incongruous preceding and following images labeling

Images shown in Fig. 4 are obtained from the CULane [6] and the lines in the images are generated from the annotations of Ground Truth (GT) provided by the dataset. The background of each image is very similar and has certain lane characteristics, but the annotations of lane lines of each image are obviously different, which will cause more False Negatives (FNs) during metric evaluation.

Method	Neck	Total	Normal	Crowded	Hlight	Shadow	Noline	Arrow	Curve	Cross	Night
CondLaneNet	FPN	78.07	92.31	76.70	71.09	76.62	51.14	88.67	67.95	1137	72.82
	IFP	+0.29	**+0.43**	+0.11	+0.60	+1.92	**+0.90**	**+0.41**	+2.20	+166	+0.19
		(78.36)	(92.74)	(76.81)	(71.69)	(78.54)	(52.04)	(89.08)	(70.15)	(1303)	(73.01)
CLRNet	FPN	79.36	93.18	77.69	73.62	80.73	52.22	90.18	67.87	1101	74.93
	IFP	+0.30	+0.27	**+0.81**	+0.67	+1.05	+0.70	+0.12	+1.94	+162	+0.61
		(79.66)	(93.45)	**(78.50)**	(74.29)	(81.42)	(52.95)	(90.30)	(69.81)	(1263)	**(75.54)**
IFPNet	FPN	79.56	93.44	78.09	73.50	78.49	52.83	90.17	69.32	1033	74.82
	IFP	**+0.39**	+0.13	+0.04	**+2.25**	**+3.25**	+0.56	+0.33	**+2.37**	**-16**	**+0.72**
		(79.95)	**(93.57)**	(78.13)	**(75.78)**	**(81.74)**	**(53.39)**	**(90.50)**	**(71.69)**	**(1017)**	**(75.54)**

Table 1. Results of the ablation study on IFP. The configurations of ConaLaneNet [5] and CLRNet [11] are the same as the original code provided by the authors with the same backbone ResNet18 except for the neck of the net. We re-train the models on the same RTX3090 GPU.

Method	Backbone	Refining Shape	mF1	F1 score
IFPNet	ResNet18	10×25	55.36	79.61
IFPNet	ResNet18	20×50	55.43	79.95
IFPNet	ResNet18	40×100	55.04	79.78
IFPNet	ResNet101	10×25	55.85	80.12
IFPNet	ResNet101	20×50	**56.32**	80.33
IFPNet	ResNet101	40×100	55.93	**80.42**

Table 2. Results of the ablation study on refining shape. Experiments are conducted on the benchmark dataset CULane [6].

Method	Attention Method	mF1	F1 score
IFPNet	self-attention [8]	55.13	79.62
IFPNet	SENet [3]	55.21	79.56
IFPNet	DANet [2]	55.38	79.90
IFPNet	HANet [1]	54.95	79.46
IFPNet	CBAM [10]	**55.51**	79.83
IFPNet	non-local [9]	55.43	**79.95**

Table 3. Results of the ablation study on attention method. Experiments are conducted on CULane with the same backbone ResNet18.

Method	RIoU	mF1	F1 score
LaneATT		47.20	74.86
LaneATT	✓	+0.85(48.05)	+0.45(75.31)
CLRNet		54.95	79.36
CLRNet	✓	+0.72(55.67)	+0.30(79.66)
IFPNet		55.16	79.74
IFPNet	✓	+0.27(55.43)	+0.21(79.95)

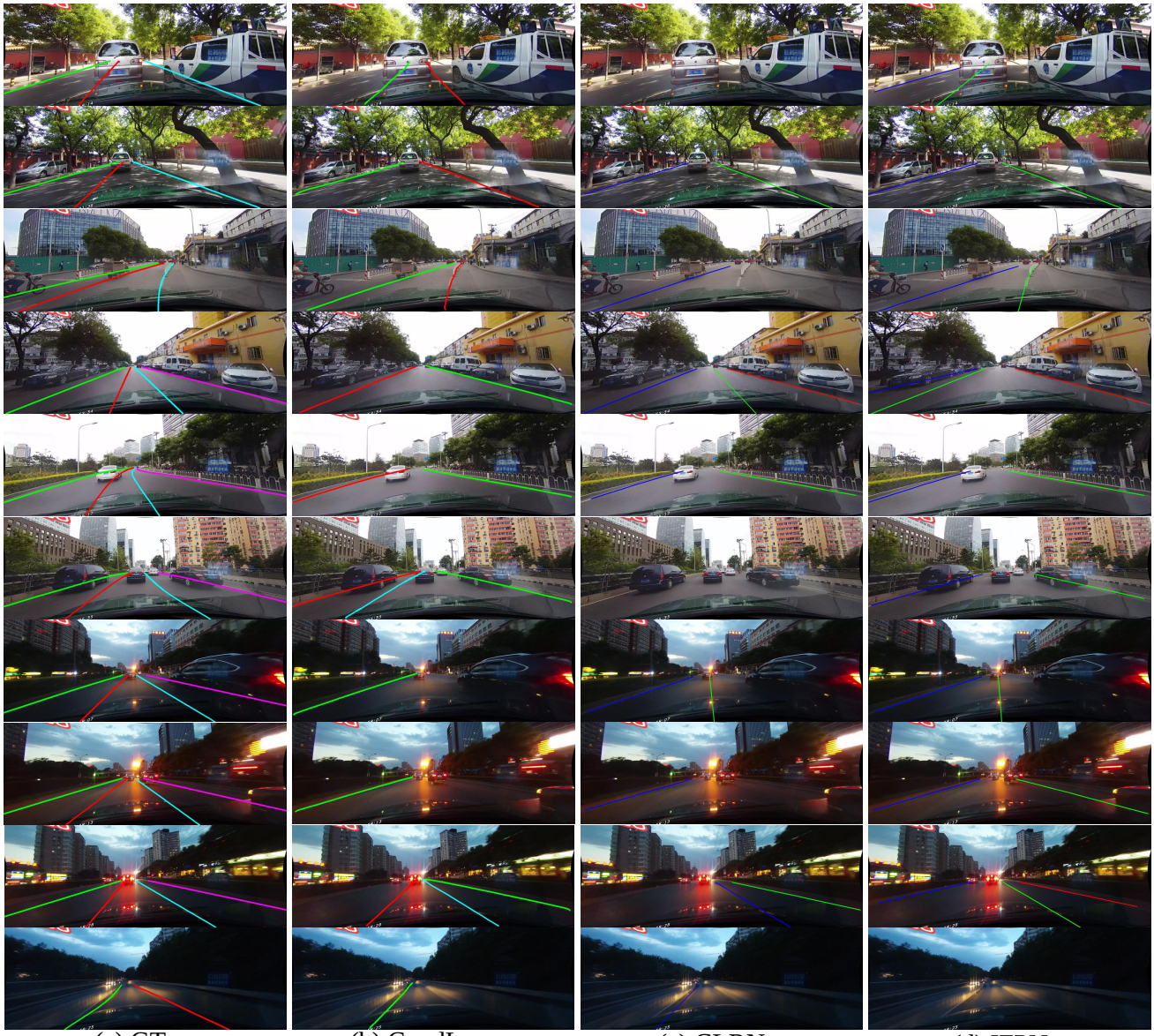
Table 4. Results of the ablation study on refine layer. Experiments are conducted on the benchmark dataset CULane [6] with the same backbone ResNet18.

2.1.1 Chosen On Fusion Proportion Between the Multi Layers

For the reason that the performance of FPN is affected by the fusion proportion α, α denotes the classification-based fusion factor (CFF). CFF is a plug-and-play module

Figure 1. Examples of inappropriate lane markings. The images are generated from annotations of Ground Truth (GT).

added to the FPN for the calculation of the ratio of high- and low-scale information in the fusion. When set to 1, it degenerates to the traditional FPN. In the ablation studies, IFP refers to the feature fusion structure, which is a plug-and-play integration module added after the traditional FPN multilayer output for further hierarchical information fusion. If removed, the unfused multi-scale information is directly fed to the head of the IFPNet. In order to choose the reasonable fusion proportion, we apply an unsupervised classification structure to calculate the reasonable value of the proportion. In this section, we study the statistics of prediction for α, and the tendency of the data distribution can be seen in Fig. 3. The distribution of α is concentrated in three regions, from left to right, representing "cross" (no lane lines in the picture), the case where more high-scale information is needed, and the case where more low-scale information is needed. The data distribution is reasonable,



(a) GT (b) CondLane (c) CLRNet (d) IFPNet

Figure 2. Failure cases of CondLaneNet [5] (CondLane in the figure), CLRNet [11], and our proposed method IFPNet on the benchmark dataset CULane [6]. The results are generated with ResNet18 with the same structure.

reflecting the role of the fusion factor in balancing hierarchical information.

2.2. Inappropriate lane markings

As shown in Fig. 1, Some of the annotations in the images may be inappropriate. In Fig. 1 (a), the features of the distal curve lane line are not obvious and do not conform to the features of the bottom-up lane line. In Fig. 1 (b), the markings in the image are dense and not fully located in the lane lines. In Fig. 1 (c), the rightmost lane line should be the long white line on the ground. In Fig. 1 (d), there is a fork in the road ahead, and it is hard to tell which road the car is

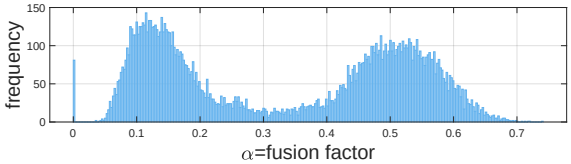


Figure 3. The statistics of prediction for α, which shows the tendency of the data distribution.

Figure 4. Examples of incongruous preceding and following images labeling. The images are generated from annotations of Ground Truth (GT).

about to take just by the information in the current image.

3. Possible Limitations of IFPNet and Future work

Though our method has achieved state-of-art results on benchmark datasets and performs well in some challenging and complex road conditions, we still find some failure cases, which are shown in Fig. 2. Even if better than the former studies in such cases, our method still misses some lines especially when there is no white line on the ground, leading to large recognition errors. In such cases, it is hard for lane detection models to decide whether there is a line on the empty road and how to accurately regress the line. As shown in Table 1, lane detection models commonly perform the worst on the type "noline" among all the categories.

Without thinking about the annotation issues mentioned in the former section, considering the similarity between images in similar scenes, the number of lane lines in these images should be similar. Therefore, we have tried to add a classification module in the head of the IFP, so as to train the model to identify the number of lane lines in the images. But the result does not yet meet our satisfaction. In the future study, we will continue to improve the classification module to alleviate the issue of missing lines.

References

[1] Sungha Choi, Joanne T Kim, and Jaegul Choo. Cars can't fly up in the sky: Improving urban-scene segmentation via height-driven attention networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9373–9383, 2020. 2

[2] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, et al. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3146–3154, 2019. 2

[3] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018. 2

[4] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, et al. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125, 2017. 1

[5] Lizhe Liu, Xiaohao Chen, Siyu Zhu, and Ping Tan. CondLaneNet: a top-to-down lane detection framework based on conditional convolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3773–3782, 2021. 1, 2, 3

[6] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018. 1, 2, 3

[7] Lucas Tabelini, Rodrigo Berriel, Thiago M Paixao, Claudine Badue, Alberto F De Souza, et al. Keep your eyes on the lane: Real-time attention-guided lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 294–302, 2021. 1

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, et al. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017. 1, 2

[9] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7794–7803, 2018. 2

[10] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018. 2

[11] Tu Zheng, Yifei Huang, Yang Liu, Wenjian Tang, Zheng Yang, et al. CLRNet: Cross layer refinement network for lane detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 898–907, 2022. 1, 2, 3