CLRerNet: Improving Confidence of Lane Detection with LaneIoU
— Supplementary Material —

Hiroto Honda Yusuke Uchida
GO Inc.
{hiroto.honda, yusuke.uchida}@goinc.jp

1. Removing redundant frames

As is described in the Subsection 4.1 of the main paper, the CULane dataset [1] includes a non-negligible amount of redundant frames where the ego-vehicle is stationary and the lane annotations do not change. We remove the frames whose average pixel value difference from the previous frame is below a threshold T_{diff}. Fig. 1 shows the F1$_{50}$ *val* score and the number of train data with different T_{diff}. There is a trade-off between the data cleanness and data size, thus a peak of F1$_{50}$ can be seen at $T_{diff} \approxeq 15$. When the redundant data is removed, the spikes in the θ distribution of ground-truth lanes (Fig. 2) disappear, which results in avoiding overfitting to the spikes. From the experiment above, the optimal threshold (=15) is chosen and the remaining 55,698 (62.7%) frames are utilized for training. The F1 *test* score of CLRNet-DLA34 is improved from 80.30±0.05 to 80.86±0.06 ($N = 5$ each) with the same 15-epoch training.

2. Improved LaneATT with LaneIoU

LaneATT [2] assigns the anchors according to the distance metric D. The distance is calculated as the average of x-coordinate distance at the common horizontal lines between the anchors X_a and ground-truths (GTs) X_{GT}. Unlike CLRNet [3], assignment between anchors and GTs does not change throughout the training as the anchors are not learnable. The positive and negative anchors are determined as:

$$\text{positive: } D(X_a, X_{GT}) < T_{dist}^{pos}, \tag{1}$$

$$\text{negative: } D(X_a, X_{GT}) > T_{dist}^{neg}, \tag{2}$$

where T_{dist}^{pos} and T_{dist}^{neg} are the threshold parameters and set to 15 and 20 respectively in the (640, 360) resolution. For the positive and negative anchors, the learning target of the classification logit is set to positive and negative, otherwise ignored. The regression loss is imposed only on the positive anchors. We keep the above assignment algorithm only for the regression target, and separately assign the positive and negative anchors for the classification target leveraging LaneIoU. We use *predictions* X_p after refined by the regression-head output for IoU calculation:

$$\text{positive: } LaneIoU(X_p, X_{GT}) > T_{iou}^{pos}, \tag{3}$$

$$\text{negative: } LaneIoU(X_p, X_{GT}) < T_{iou}^{neg}, \tag{4}$$

where the thresholds T_{iou}^{pos} and T_{iou}^{neg} are set to 0.7 and 0.5 respectively. With the LaneIoU-based assignment for classification, the predictions that have high IoU with the corresponding GTs are prioritized and the confidence scores get close to the metric IoU with GTs. The LaneIoU-based assignment improves the F1$_{50}$ score from 77.51±0.10 to 78.19±0.06 as is shown in Table 2 of the main paper.

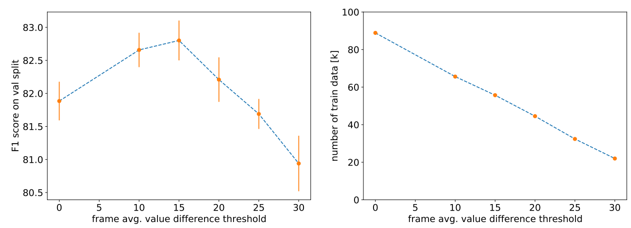

Figure 1. F1$_{50}$ *val* score (left) and number of train data with different thresholds to filter redundant *train* data.

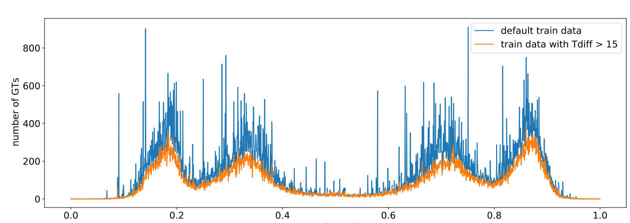

Figure 2. θ distribution of *train* data before and after redundant data removal.

References

[1] Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In *AAAI*, February 2018. 1, 2, 3

[2] Lucas Tabelini, Rodrigo Berriel, Thiago M. Paix ao, Claudine Badue, Alberto Ferreira De Souza, and Thiago Oliveira-Santos. Keep your Eyes on the Lane: Real-time Attention-guided Lane Detection. In *CVPR*, 2021. 1

[3] Tu Zheng, Yifei Huang, Yang Liu, Wenjian Tang, Zheng Yang, Deng Cai, and Xiaofei He. Clrnet: Cross layer refinement network for lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 898–907, June 2022. 1