3DLANEFORMER: RETHINKING LEARNING VIEWS FOR 3D LANE DETECTION SUPPLEMENTARY MATERIAL

Kun Dong *Jian Xue* *Xing Lan* *Ke Lu**

University of Chinese Academy of Sciences

A. Implementation Details

ResNet-18 is selected as the backbone of our proposed network. Input images are all resized to 360×480. We set $\lambda_{reg} = \lambda_{cls} = 1$, while α^h and γ are set to 0.5 and 2 respectively. Each ground-truth lane will be associated with 3 anchors. In addition, our network uses Adam optimizer, with a base learning rate of 2×10^{-4} and weight decay of 10^{-4}.

OpenLane. We sample 10 points for the OpenLane dataset [1] at y-coordinates of S_1 = {5, 10, 15, 20, 30, 40, 50, 60, 80, 100} to form the initial anchor features, which will be used to predict the coarse lane offsets at y-coordinates of S_2 = {5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100}. In the refinement stage, we will resample the FV and BEV point features in the FV and BEV feature maps, respectively. In particular, the resampling interval on the FV feature map is the same as S_1, while the sampling interval on the BEV feature map is consistent with S_2. Additionally, the training batch size is set to 64 and we train 3DLaneFormer on this dataset with four NVIDIA A100 GPUs. During training and inference, we predict the lanes of (-30m, 30m) in the x direction and (3m, 103m) in the y direction in the road ground coordinates.

ONCE-3DLanes. For the ONCE-3DLanes dataset [2], only lanes within 50 meters are required to be predicted. Therefore, we sample 10 points at y-coordinates of S = {2, 5, 8, 10, 15, 20, 25, 30, 40, 50}. The set of y-coordinates that need to be predicted for offset is the same as S. Other training settings are the same as those on the OpenLane dataset as mentioned above.

B. Qualitative Results

For visualization, we present the comparison between Anchor3DLane [3] and our 3DLaneFormer on OpenLane dataset. In addition, we show the qualitative results of our 3DLaneFormer on the ONCE-3DLanes benchmarks.

B.1 Visual Comparison on OpenLane

We compare our 3DLaneFormer with Anchor3DLane on the OpenLane dataset and the results are shown in Fig. 1. Our method can better fit lanes than Anchor3DLane, especially in the distant parts (row 1 and row 2). In addition, the number of the predicted lanes from our method is more consistent with the ground truth (row 3 and row 4). When the scene contains multiple lanes (row 5), 3DLaneFormer can also better capture the height changes than Anchor3DLane, which demonstrates the superiority of our proposed method.

B.2 Qualitative Results on ONCE-3DLanes

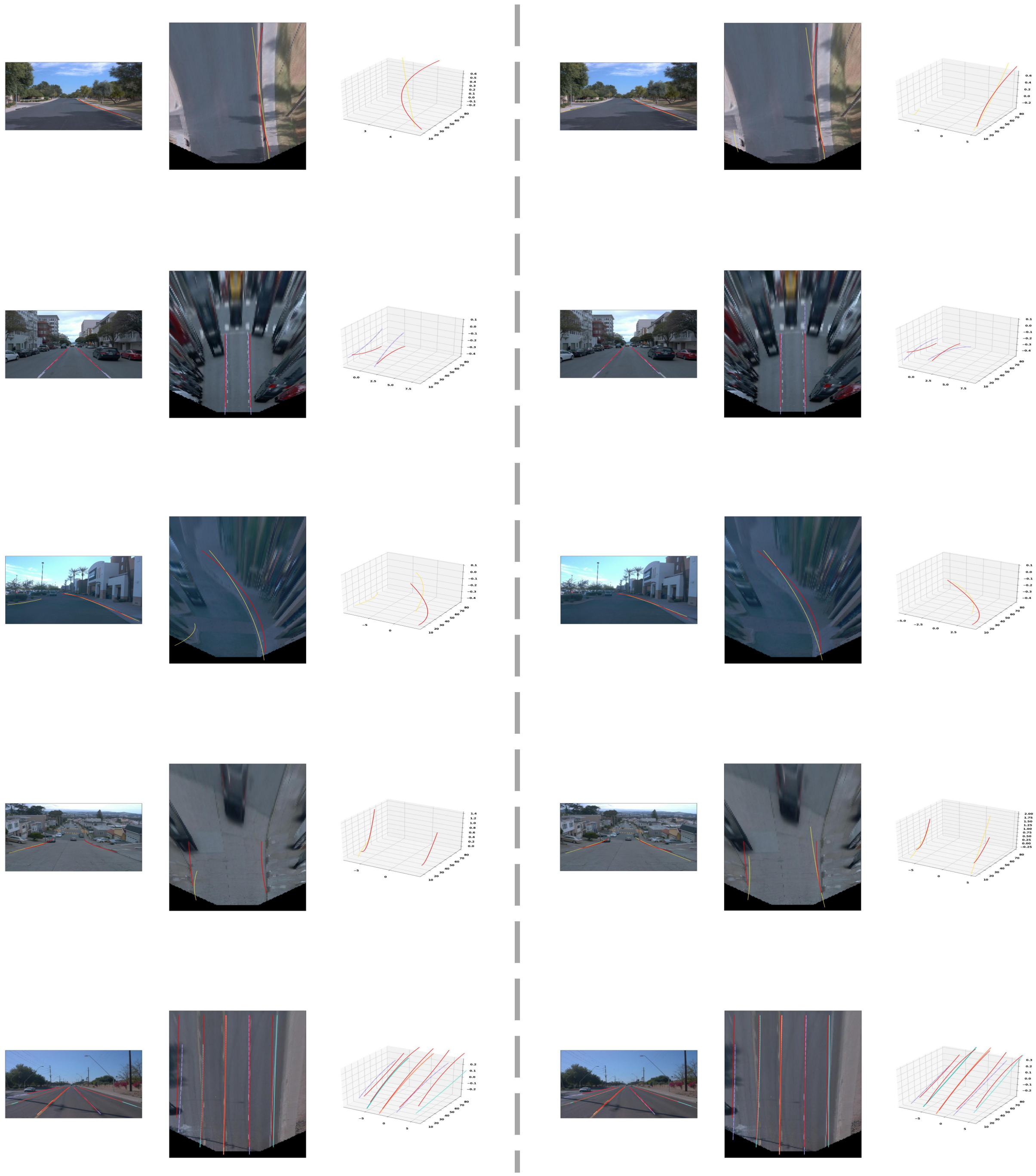
We further present the qualitative results of our 3DLaneFormer on the ONCE-3DLanes benchmark. As illustrated in Fig. 2, our method performs well in various scenes, such as bad weather like rainy days (row 1). Note that ONCE-3DLanes utilizes a lane detection model to generate 2D annotations, which are inaccurate or incomplete in some cases. However, our 3DLaneFormer can still produce excellent predictions.

1. REFERENCES

[1] Li Chen, Chonghao Sima, Yang Li, Zehan Zheng, Jiajie Xu, Xiangwei Geng, Hongyang Li, Conghui He, Jianping Shi, Yu Qiao, et al., "Persformer: 3d lane detection via perspective transformer and the openlane benchmark," in *European Conference on Computer Vision*, 2022, pp. 550–567.

[2] Fan Yan, Ming Nie, Xinyue Cai, Jianhua Han, Hang Xu, Zhen Yang, Chaoqiang Ye, Yanwei Fu, Michael Bi Mi, and Li Zhang, "Once-3dlanes: Building monocular 3d lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17143–17152.

[3] Shaofei Huang, Zhenwei Shen, Zehao Huang, Zi-han Ding, Jiao Dai, Jizhong Han, Naiyan Wang, and Si Liu, "Anchor3dlane: Learning to regress 3d anchors for monocular 3d lane detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17451–17460.



(a) Anchor3DLane (b) 3DLaneFormer

Fig. 1: Comparison between Anchor3DLane and our 3DLaneFormer on the OpenLane dataset. (a): Qualitative results of Anchor3DLane. (b): Qualitative results of our 3DLaneFormer. Red: Ground-truth. Other color: Prediction.

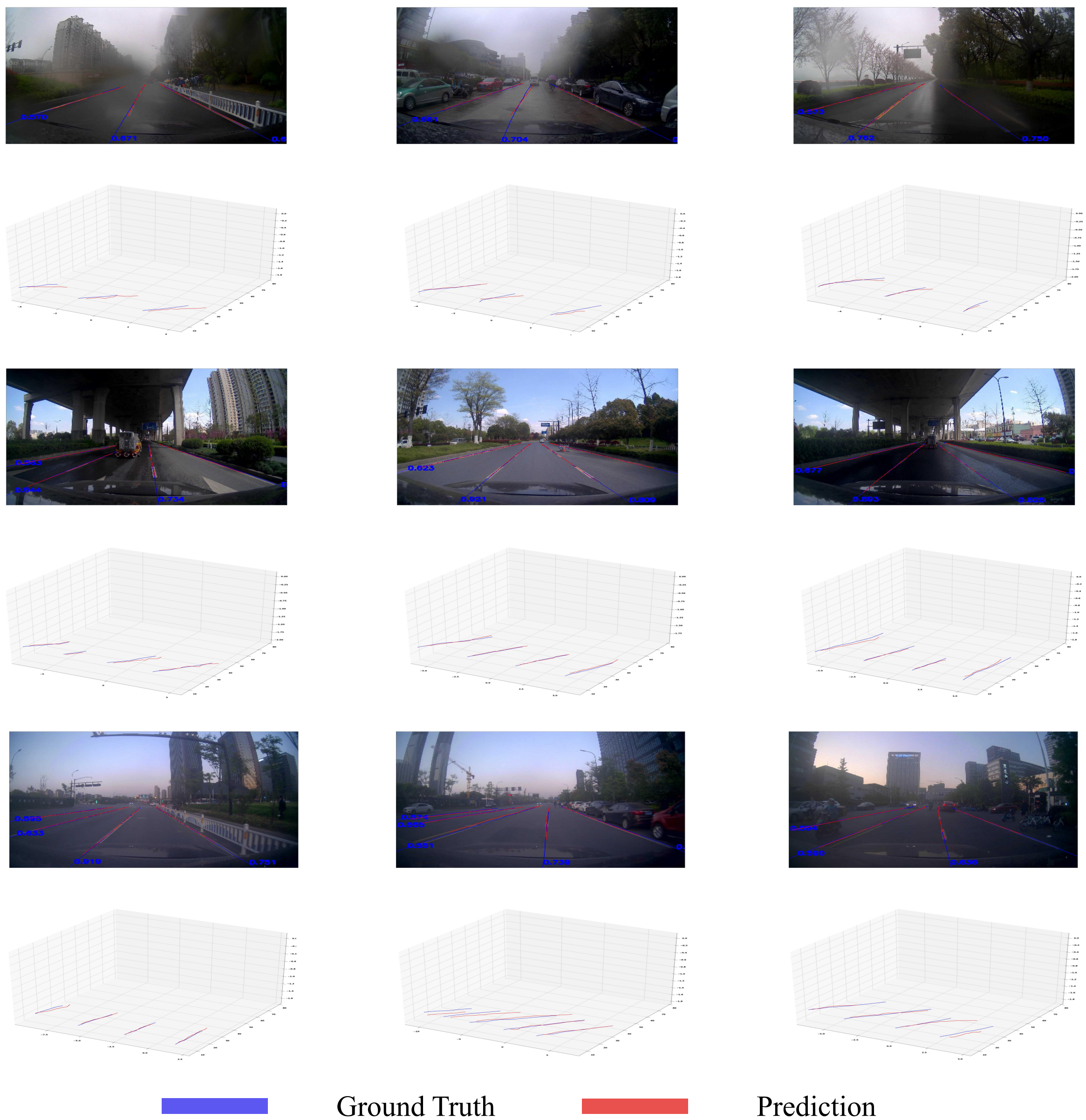


Fig. 2: Qualitative results on the ONCE-3DLanes dataset.