ONCE-3DLanes: Building Monocular 3D Lane Detection: Supplementary material

Fan Yan[1*] Ming Nie[1*] Xinyue Cai[2] Jianhua Han[2] Hang Xu[2] Zhen Yang[2]
Chaoqiang Ye[2] Yanwei Fu[1] Michael Bi Mi[2] Li Zhang[1†]
[1]School of Data Science, Fudan University [2]Huawei Noah's Ark Lab
https://once-3dlanes.github.io

1. Dataset

Our dataset and code will be made publicly available at https://once-3dlanes.github.io.

2. Further analysis

We conduct ablation studies to show the rationality of our experiment settings including loss function, backbone network and regression method.

Loss function As shown in Table 1, for the spatial contextual branch, we study different loss functions. Results show the smooth L1 loss outperforms L1 and L2 loss functions at all metrics.

Backbone network We compare the SegFormer [4] with Unet [1] for the backbone network. Moreover, different attention mechanisms [2, 3] are added to Unet to help learn the global information of lane structures. Table 2 shows that model with SegFormer beat the variants of Unet by a clear margin.

Regression method We also conduct an ablation study to evaluate the way to predict the depth information in Table 3. Our method regress in a residual manner is referred as *relative* method, and the method directly regress the depth information without pre-defined shift and scale is called *absolute* method. Table 3 shows the *relative* outperforms *absolute* by a large margin.

3. More qualitative results

We present the qualitative results of *SALAD* lane prediction in Figure 1. 2D projections are shown in the left and 3D visualizations are presented in the right.

*The first two authors contributed equally to this work.

†Li Zhang (lizhangfd@fudan.edu.cn) is the corresponding author at School of Data Science and Shanghai Key Laboratory of Intelligent Information Processing, Fudan University

Loss function	F1(%)	Precision(%)	Recall(%)	CD error(m)
L1	63.47	75.08	54.97	0.101
L2	62.91	74.55	54.41	0.103
Smooth L1	**64.07**	**75.90**	**55.42**	**0.098**

Table 1. Ablation studies on loss functions in the spatial contextual branch.

Backbone	F1(%)	Precision(%)	Recall(%)	CD error(m)
Unet [1]	61.12	73.47	52.32	0.105
Unet+self-att. [2]	62.71	74.41	54.19	0.101
Unet+axial-att. [3]	63.15	74.81	54.64	0.101
SegFormer-B2 [4]	**64.07**	**75.90**	**55.42**	**0.098**

Table 2. Ablation studies on backbone networks.

Offset option	F1(%)	Precision(%)	Recall(%)	CD error(m)
absolute	62.37	74.07	53.86	0.104
relative	**64.07**	**75.90**	**55.42**	**0.098**

Table 3. Ablation studies on depth regression methods.

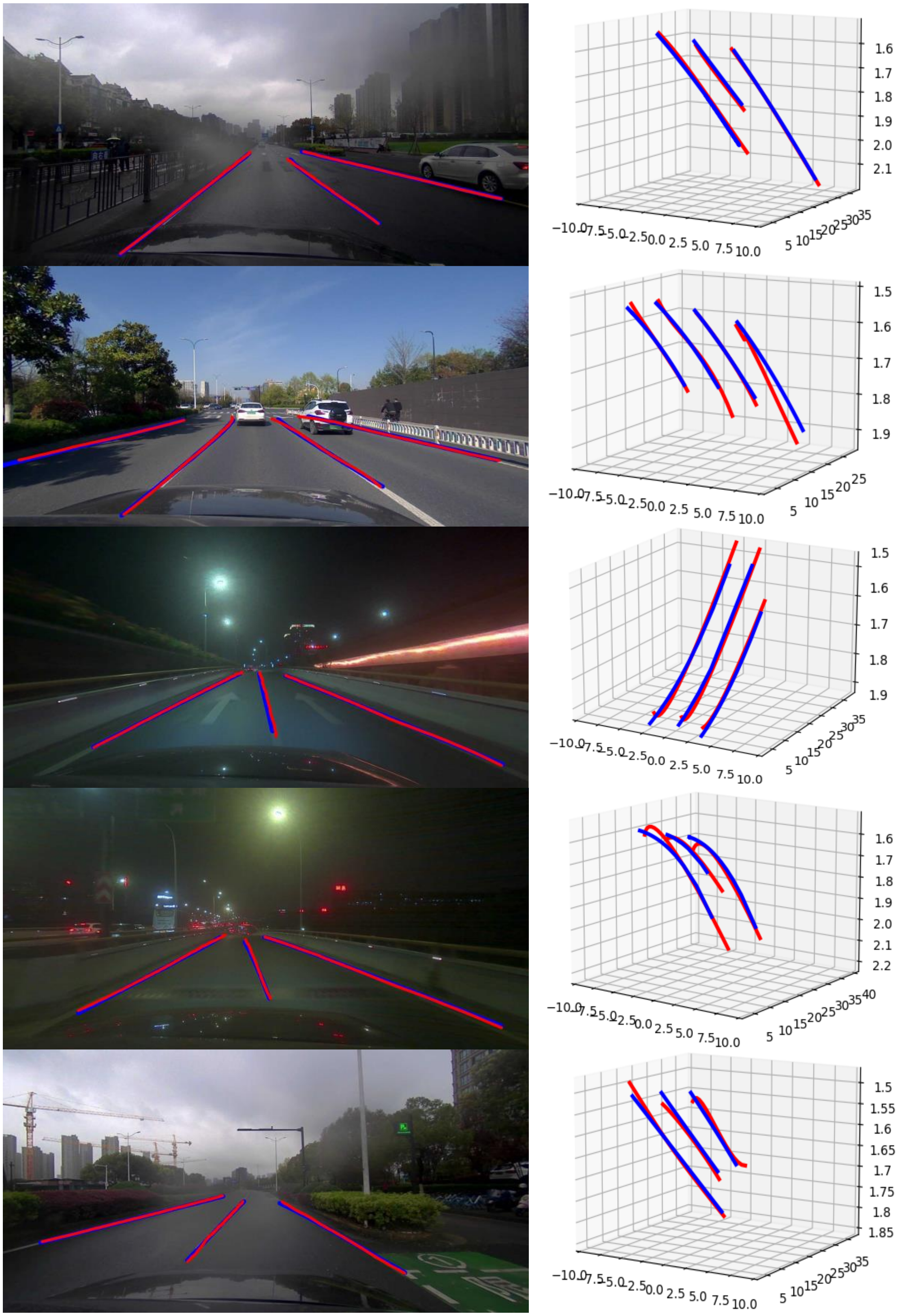
References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 1

[2] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017. 1

[3] Huiyu Wang, Yukun Zhu, Bradley Green, Hartwig Adam, Alan Yuille, and Liang-Chieh Chen. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation. In *ECCV*, 2020. 1

[4] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *arXiv preprint arXiv:2105.15203*, 2021. 1


1.6
1.7
1.8
1.9
2.0
2.1
−10.0−7.5−5.0−2.5 0.0 2.5 5.0 7.5 10.0
5 10 15 20 25 30 35
1.5
1.6
1.7
1.8
1.9
−10.0−7.5−5.0−2.5 0.0 2.5 5.0 7.5 10.0
5 10 15 20 25
1.5
1.6
1.7
1.8
1.9
−10.0−7.5−5.0−2.5 0.0 2.5 5.0 7.5 10.0
5 10 15 20 25 30 35
1.6
1.7
1.8
1.9
2.0
2.1
2.2
−10.0−7.5−5.0−2.5 0.0 2.5 5.0 7.5 10.0
5 10 15 20 25 30 35 40
1.5
1.55
1.6
1.65
1.7
1.75
1.8
1.85
−10.0−7.5−5.0−2.5 0.0 2.5 5.0 7.5 10.0
5 10 15 20 25 30 35

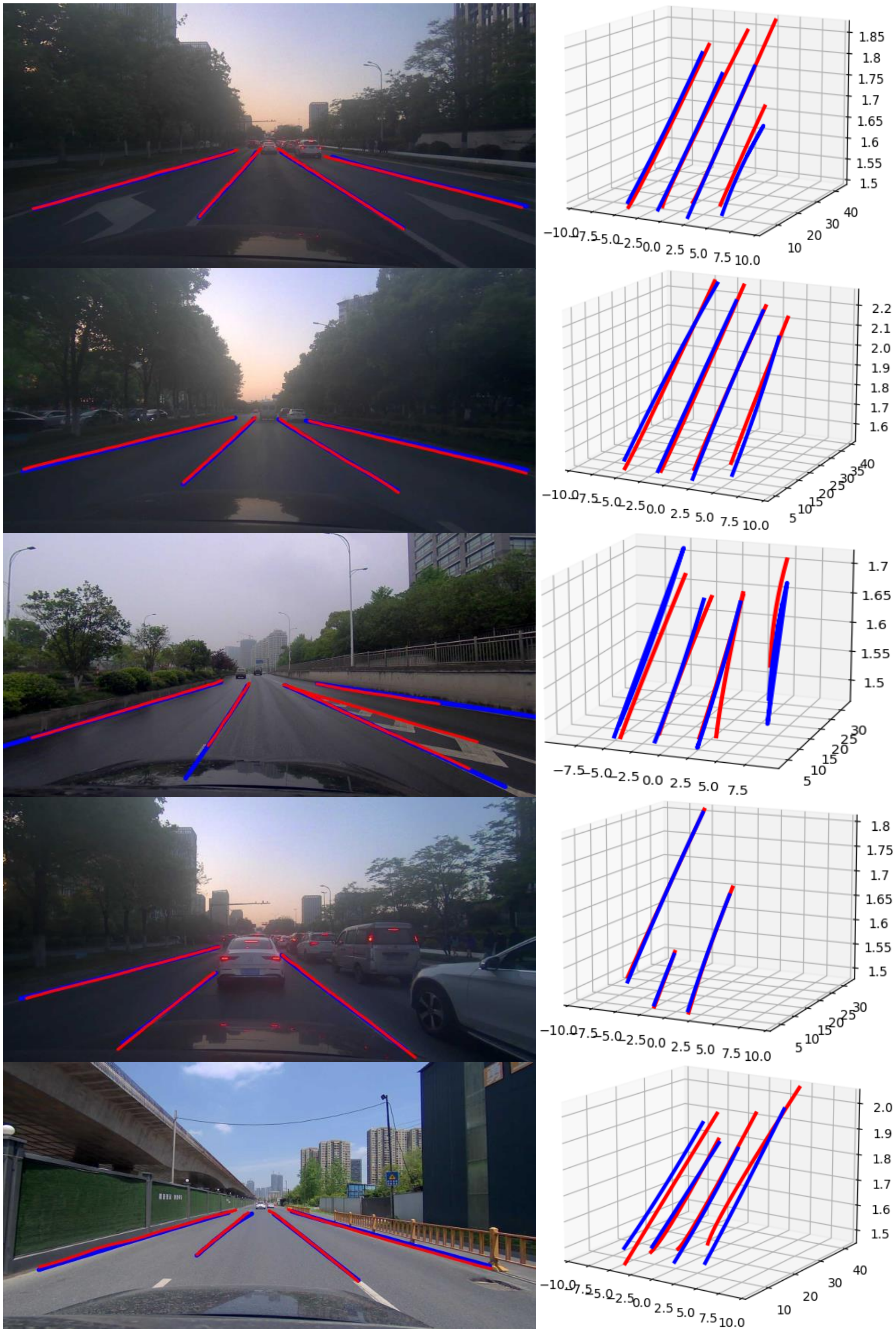


Figure 1. Visualization of *SALAD* on ONCE-3DLanes test set. The ground-truth lanes are colored in red while the predicted lanes are colored in blue. 2D projections are shown in the left and 3D visualizations in the right.