DAMap: Distance-aware MapNet for High Quality HD Map Construction

Supplementary Material

In the supplementary material, we first conduct more ablation studies. Then, we report the runtime analysis of our DAMap. Finally, we present the visual analysis of map element predictions on the nuScenes val set.

A. More Ablation Study

Hybrid Number	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
-	58.1	60.8	62.3	60.4
0	58.6	61.1	63.1	60.9
1	58.5	62.9	63.4	61.6
2	58.7	60.8	63.0	60.8
3	56.8	62.1	62.6	60.5

Table 1. Ablations on the hybrid number in Hybrid Loss Scheme.

Ablation on Hybrid Number in HLS. We analyze the effect of the hybrid number in HLS. The results are shown in Table 1. When using the Focal Loss [2] in the full decoder layer, the baseline performance is 60.4 mAP. When using our proposed DAFL in the full decoder layer, the performance of the model is 60.9 mAP. When Focal Loss is introduced as the hybrid loss in the first decoder layer, the model achieves 61.6 mAP. The result shows that using Focal Loss to obtain higher quality predictions than utilizing DAFL has better results. When we continue to introduce the Focal Loss into more decoder layers, the performance improvement of the model decreases. Thus, we adopt one decoder layer with Focal Loss in all our experiments.

λ	$AP_{ped.}$	$AP_{div.}$	$AP_{bou.}$	mAP
1.0	58.5	64.7	65.1	62.8
0.8	59.6	64.4	64.5	62.9
1.2	58.3	64.4	64.7	62.5

Table 2. Ablation on the hyper-parameter λ in the DAFL.

Ablation on Hyper-parameter λ. The λ is a hyper-parameter to adjust the distribution of loss to quality. To investigate the effect of the hyper-parameter λ in the DAFL, we conduct experiment of our method with different λ values. As shown in Table 2, setting different λ values has a small effect (0.3 margin) on the performance of the model. We set λ to 1.0 in all MapTR experiments.

Visualization of Attention Weights. To further understand that classification and localization tasks often have different feature preferences, we show the visualization of attention weights in Figure 1. It can be seen that the weight distribution of both tasks is different. The weight distribution for each task is different from the weight distribution when tasks are shared. These results show that our Task Modulated Deformable Attention can obtain different weights for different tasks.

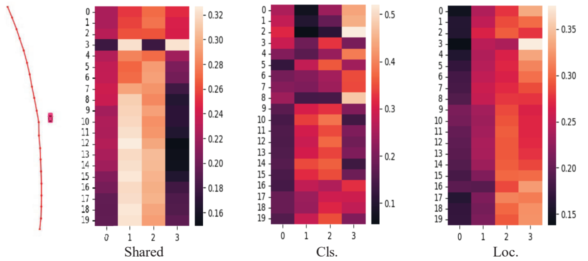

Figure 1. Example of learned attention weights for points in an instance when task-shared and task modulated.

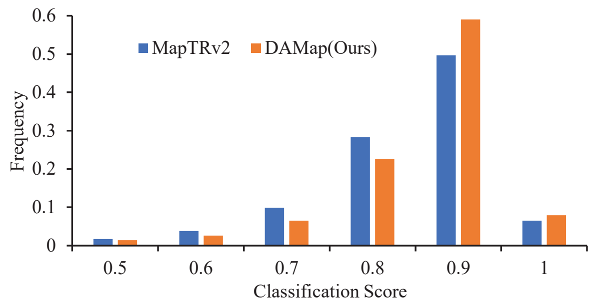

Figure 2. Frequency of the localization quality on the high classification score (0.9) with or without our method.

Quantitative Analysis of Localization Quality. To further illustrate that our method can produce high quality predictions, we count the localization quality of the predictions under the 0.9 classification score. As shown in Figure 2, our method improves the percentage of predictions with high localization quality under high classification scores. The results show that our method can encourage high classification samples to obtain more accurate localization. This phenomenon is further evidence that our method can produce high quality predictions.

Method	Para.	FPS	mAP
MapTRv2[1]	40M	14.1	68.7
MapTRv2†[1]	40M	12.6	68.3
MapTRv2+HLS(Ours)	40M	12.6	69.4
DAMap(Ours)	52M	12.1	70.4

Table 3. The analysis of Parameters and FPS of our DAMap. MapTRv2† is the inference speed in our environment.

B. Runtime Analysis

We analyze the Parameters and FPS of our DAMap, as can be seen from Table 3. MapTRv2† is the inference speed in our environment, slightly lower than the MapTRv2 paper due to hardware differences (such as CPU). All FPS are the average inference speed by repeating the test 3 times on the NVIDIA RTX 3090Ti GPU. Our approach adds only a small amount of inference speed. Furthermore, our method does not increase the inference speed when equipped only with our HLS.

C. Visualization of Map Element Predictions

As shown in Figure 3 and Figure 4, we provide comparisons of map element predictions with the score threshold of 0.4 between the baseline and our method on the nuScenes val set. It can be seen that our method can output more map elements overlooked by the baseline under the same score threshold. This result shows that our method can achieve higher classification and higher localization results, further validating our motivation.

References

[1] Bencheng Liao, Shaoyu Chen, Yunchi Zhang, Bo Jiang, Qian Zhang, Wenyu Liu, Chang Huang, and Xinggang Wang. Maptrv2: An end-to-end framework for online vectorized hd map construction. *IJCV*, pages 1–23, 2024. 1

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal Loss for Dense Object Detection . In *ICCV*, pages 2999–3007, 2017. 1

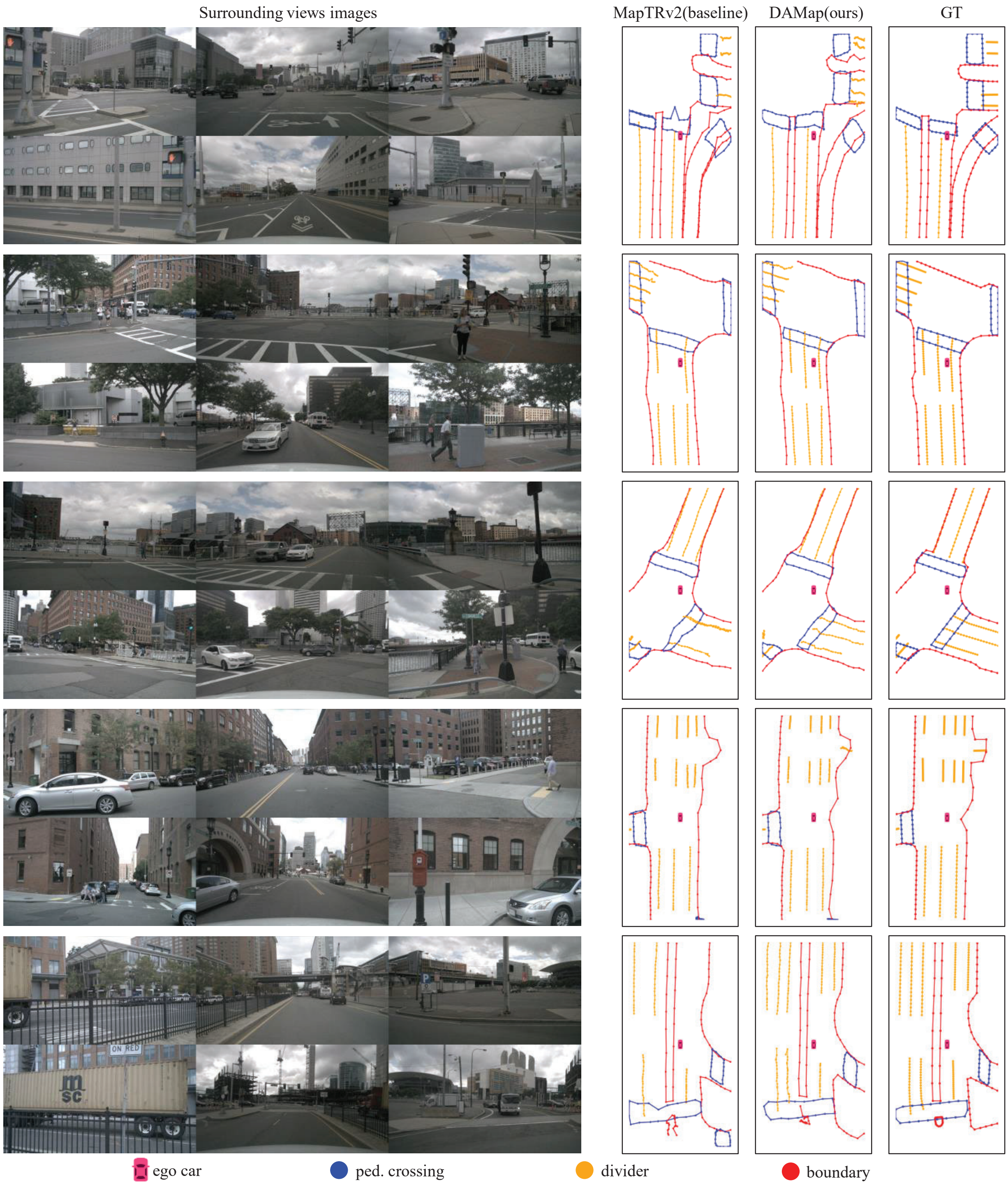

Figure 3. Comparison of map element predictions between our method and baseline with ResNet-50 and 24 epochs on the nuScenes val set. The score threshold is set to 0.4.

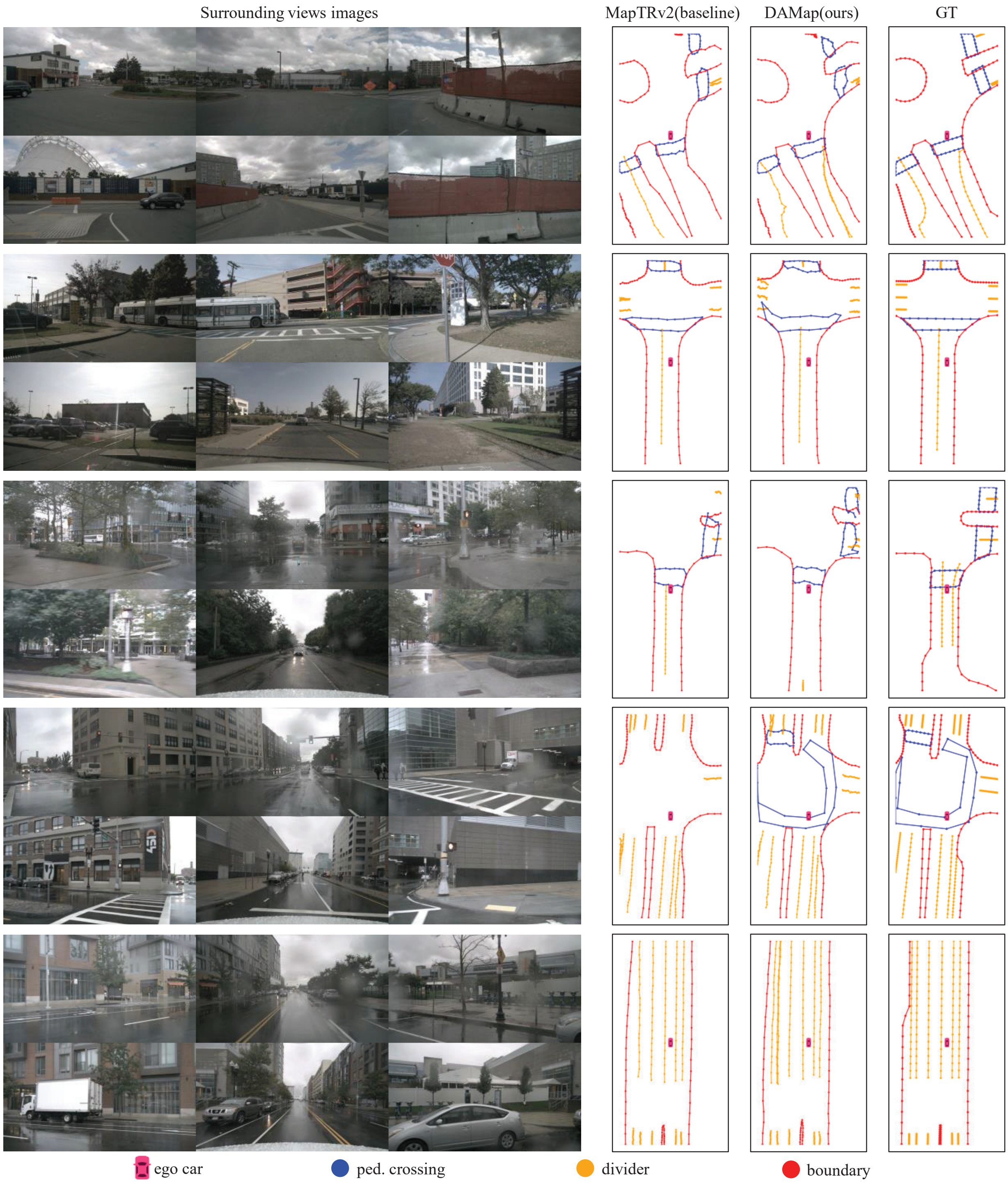

Figure 4. Comparison of map element predictions between our method and baseline with ResNet-50 and 24 epochs on the nuScenes val set. The score threshold is set to 0.4.