Stream Query Denoising for Vectorized HD-Map Construction Supplementary Material

Shuo Wang[1*†], Fan Jia[2*], Weixin Mao[3†], Yingfei Liu[2], Yucheng Zhao[2], Zehui Chen[1], Tiancai Wang[2], Chi Zhang[4], Xiangyu Zhang[2], and Feng Zhao[1✉]

[1] MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China
[2] MEGVII Technology [3] Waseda University [4] Mach Drive
shuowang2323@mail.ustc.edu.cn

Appendix

A More Detailed Implementation Settings

Suppose that a curve is composed of the point set $\{(p_1^x, p_1^y), (p_2^x, p_2^y), \cdots, (p_c^x, p_c^y)\}$, and n is the number of the point set. Then we define the approximate pseudo scale and pseudo center of the curve as:

$$L = \max(\{p_1^x, p_2^x, \cdots, p_c^x\}) - \min(\{p_1^x, p_2^x, \cdots, p_c^x\}), \quad (1)$$

$$W = \max(\{p_1^y, p_2^y, \cdots, p_c^y\}) - \min(\{p_1^y, p_2^y, \cdots, p_c^y\}), \quad (2)$$

$$C_x = \min(\{p_1^x, p_2^x, \cdots, p_c^x\}) + L/2, \quad (3)$$

$$C_y = \min(\{p_1^y, p_2^y, \cdots, p_c^y\}) + W/2. \quad (4)$$

For the original random noise scale $\{\Delta x, \Delta y, \Delta l, \Delta w\}$ in Dynamic Curve Noising, the identical shifts are subject to the constraint $|\Delta x| < \frac{\lambda_1 \cdot L}{2}$ and $|\Delta y| < \frac{\lambda_2 \cdot W}{2}$, where λ_1, λ_2 are the maximum scale of noise. Additionally, the scale shifts are within the constraint $|\Delta l| < \frac{\lambda_3 \cdot length}{2}$ and $|\Delta w| < \frac{\lambda_4 \cdot width}{2}$, respectively, where λ_3 and λ_4 denote the maximum scale of noise.

In our experiments, we set all the values of λ_1, λ_2, λ_3, λ_4 to 0.6. Like DN-DETR [2], we set the probability of the category noise perturbation to 0.5. Moreover, the number of denoising queries is set to 60.

B More Experiments On New Split

StreamMapNet [6] found an overlap of over 54% locations between the training and validation sets, including nuScenes [1] and Argoverse2 [5], and therefore proposed a new way to split the datasets. To verify the generalization of our method, we provide the results for the two datasets at the 30 m perceptual range under the new split.

† Work done during an internship at MEGVII Technology.
* Equal contribution. ✉ Corresponding author.

B.1 Performance on nuScenes.

We first compare the proposed SQD-MapNet with previous competitive vision-based counterparts on the nuScenes new validation set. As shown in Tab. 1, SQD-MapNet outperforms existing approaches on new split and achieves 35.9 mAP within only 24 epochs.

Method	AP_{ped}	AP_{div}	AP_{bound}	mAP
VectorMapNet [4]	15.8	17.0	21.2	18.0
MapTR [3]	6.4	20.7	35.5	20.9
StreamMapNet [6]	29.6	**30.1**	41.9	33.9
SQD-MapNet (Ours)	**33.7**	29.5	**44.5**	**35.9**

Table 1: Comparision with SOTAs on the new nuScenes [1] split at 30 m perception range. The results of VectorMapNet [4], MapTR [3] and StreamMapNet [6] are directly from [6].

B.2 Performance on Argoverse2.

As shown in Tab. 2, SQD-MapNet outperforms existing approaches by a significant margin. Specifically, SQD-MapNet achieves 61.2 mAP within only 30 epochs, surpassing the previous state-of-the-art method, StreamMapNet [6], by more than 3.0 mAP.

Method	AP_{ped}	AP_{div}	AP_{bound}	mAP
VectorMapNet [4]	35.6	34.9	37.8	63.1
MapTR [3]	48.1	50.4	55.0	51.1
StreamMapNet [6]	56.9	55.9	61.4	58.1
SQD-MapNet (Ours)	**61.2**	**58.1**	**64.3**	**61.2**

Table 2: Comparision with SOTAs on the new Argoverse2 [5] split at 30 m perception range. The results of VectorMapNet [4], MapTR [3] and StreamMapNet [6] are directly from [6].

References

1. Caesar, H., Bankiti, V., Lang, A.H., Vora, S., Liong, V.E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., Beijbom, O.: nuscenes: A multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11621–11631 (2020)

2. Li, F., Zhang, H., Liu, S., Guo, J., Ni, L.M., Zhang, L.: Dn-detr: Accelerate detr training by introducing query denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13619–13627 (2022)
3. Liao, B., Chen, S., Wang, X., Cheng, T., Zhang, Q., Liu, W., Huang, C.: Maptr: Structured modeling and learning for online vectorized hd map construction. arXiv preprint arXiv:2208.14437 (2022)
4. Liu, Y., Yuan, T., Wang, Y., Wang, Y., Zhao, H.: Vectormapnet: End-to-end vectorized hd map learning. In: International Conference on Machine Learning. pp. 22352–22369. PMLR (2023)
5. Wilson, B., Qi, W., Agarwal, T., Lambert, J., Singh, J., Khandelwal, S., Pan, B., Kumar, R., Hartnett, A., Pontes, J.K., et al.: Argoverse 2: Next generation datasets for self-driving perception and forecasting. arXiv preprint arXiv:2301.00493 (2023)
6. Yuan, T., Liu, Y., Wang, Y., Wang, Y., Zhao, H.: Streammapnet: Streaming mapping network for vectorized online hd map construction. arXiv preprint arXiv:2308.12570 (2023)