4DSegStreamer: Streaming 4D Panoptic Segmentation via Dual Threads

Supplementary Material

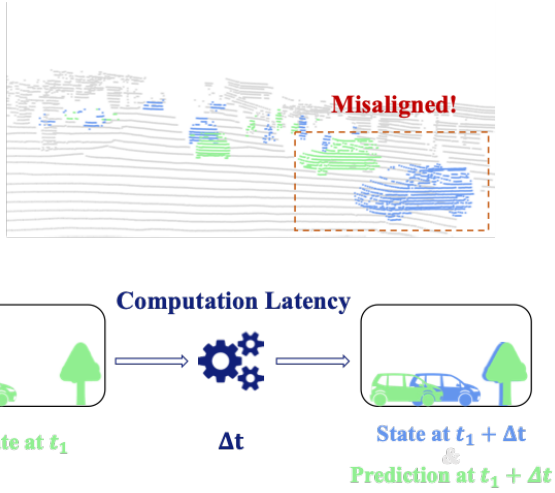

Figure S1. Streaming Perception Setting: Green points denote dynamic objects from the processed frame, whereas blue points represent the current frame at the time of prediction generated by the algorithm.

A. Streaming Perception Setting

Based on previous works [1–8] in streaming perception, our 4D streaming panoptic segmentation addresses a similar challenge by explicitly considering the impact of algorithmic processing latency on the final prediction and the scene at output time. As illustrated in Fig. S1, predictions from existing methods are misaligned with the actual scene due to this latency. This misalignment can lead to perception inaccuracies, posing potential risks when robotic systems operate in highly dynamic environments.

B. Forward Flow Iteration Proof

To find the flow between the current query point and history position in geometric memory, we use the forward flow iteration. The iteration converges if Eq. **??** holds, then the following equation holds

$$1 \geq L \geq \frac{|g(x_0 + \Delta x) - g(x_0 - \Delta x)|}{|(x_0 + \Delta x) - (x_0 - \Delta x)|} = \left| \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{(x_0 + \Delta x) - (x_0 - \Delta x)} \right| = |f'(x_0)|$$

For point x on a rigid object and the flow $f(x,t)$ representing velocity, the derivative $|f'(x)|$ can be expressed as:

$$|f'(x)| = \left| \frac{\partial f(x,t)}{\partial x} \right| = \left| \frac{\partial \left(v + \omega \times (x - x_c)\right)}{\partial x} \right| = \left| \frac{\partial(\omega \times x)}{\partial x} \right| = \left| [\omega]_\times \right| = |\omega|$$

where x_c is the rotation center of the rigid body, v is the translational velocity, ω is angular velocity, $[\omega]_\times$ is the cross-product matrix. The iteration converges when $|\omega| \leq 1$. In real-world scenarios, most rigid objects exhibit low angular velocity, allowing the iteration converges reliably.

While perfect convergence cannot be guaranteed in practice, our experiments show robust convergence in 97.4% of scenes in the SemanticKITTI dataset.

Table S1. Performance of different GPUs with different latency.

	4DSegStreamer(M4F)	PTv3	Mask4Former
sLSTQ$_{A40}$	0.681	0.526	0.501
sLSTQ$_{3090}$	0.688	0.536	0.504
sLSTQ$_{A100}$	0.702	0.561	0.538

C. Performance of different GPUs

Table S1 presents the performance of our method across different GPUs under streaming settings. Since the model's runtime speed and GPU processing capability significantly impact the metric performance, the choice of hardware plays a crucial role. Notably, the A40 and 3090 graphics cards exhibit comparable performance due to their similar computational efficiency. In contrast, the A100 demonstrates a substantial speed advantage over the 3090, leading to a 1.4% improvement in our model's performance on the A100.

References

[1] Jun-Yan He, Zhi-Qi Cheng, Chenyang Li, Wangmeng Xiang, Binghui Chen, Bin Luo, Yifeng Geng, and Xuansong Xie. Damo-streamnet: Optimizing streaming perception in autonomous driving. *arXiv preprint arXiv:2303.17144*, 2023. 1

[2] Yihui Huang and Ningjiang Chen. Mtd: Multi-timestep detector for delayed streaming perception. In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pages 337–349. Springer, 2023.

[3] Wonwoo Jo, Kyungshin Lee, Jaewon Baik, Sangsun Lee, Dongho Choi, and Hyunkyoo Park. Dade: delay-adaptive detector for streaming perception. *arXiv preprint arXiv:2212.11558*, 2022.

[4] Chenyang Li, Zhi-Qi Cheng, Jun-Yan He, Pengyu Li, Bin Luo, Hanyuan Chen, Yifeng Geng, Jin-Peng Lan, and Xuansong Xie. Longshortnet: Exploring temporal and semantic features fusion in streaming perception. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[5] Mengtian Li, Yu-Xiong Wang, and Deva Ramanan. Towards streaming perception. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 473–488. Springer, 2020.

[6] Xiaofeng Wang, Zheng Zhu, Yunpeng Zhang, Guan Huang, Yun Ye, Wenbo Xu, Ziwei Chen, and Xingang Wang. Are we ready for vision-centric driving streaming perception? the asap benchmark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9600–9610, 2023.

[7] Jinrong Yang, Songtao Liu, Zeming Li, Xiaoping Li, and Jian Sun. Real-time object detection for streaming perception. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5385–5395, 2022.

[8] Xiang Zhang, Yufei Cui, Chenchen Fu, Weiwei Wu, Zihao Wang, Yuyang Sun, and Xue Liu. Transtreaming: Adaptive delay-aware transformer for real-time streaming perception. *arXiv preprint arXiv:2409.06584*, 2024. 1