Supplemental Materials: Point Primitive Transformer for Long-Term 4D Point Cloud Video Understanding

Hao Wen[1]*, Yunze Liu[1]*, Jingwei Huang[2], Bo Duan[2], and Li Yi[1,3]

[1] Tsinghua University
wenh19@mails.tsinghua.edu.cn, liuyzchina@gmail.com
[2] Huawei Technologies {huangjingwei6,duanbo5}@huawei.com
[3] Shanghai Qi Zhi Institute ericyi0124@gmail.com

This document provides a list of supplemental materials to support the main paper.

- **Does primitive play the role of label smoothing?** - We conduct experiments to explore if primitive play the role of label smoothing.
- **Curves of Training process and Occupied memory.** - We provide training and memory-clip curves to verify the efficiency of our more optimization friendly PPTr.
- **Visualization** - We provide some visualization results of 4D semantic segmentation and primitive fitting.

A Does primitive play the role of label smoothing?

We conducted experiments on Sythia4D to determine if the primitive plays the role of label smoothing. We directly vote the label with the most as the primitive label, then calculate mIoU. However, we can only achieve 44.59 mIoU in this way, which shows that simple label smoothing is not sufficient to achieve better performance. This also proves that primitive does not play the role of label smoothing, but acts as an intermediate representation to capture long-term spatio-temporal information more effectively.

B Curves of Training process and Occupied memory.

To verify the model efficiency and whether our PPTr reduces the optimization difficulty of the transformer, we provide the network training curve along with a memory growth curve as the clip length increases in figure 1. Our model can quickly reach accuracy above 35, while P4Transformer can only reach about 5 when getting only one epoch trained. The results also show that our model reduces the memory requirements for training significantly when clip length is relatively long.

*Equal contribution.

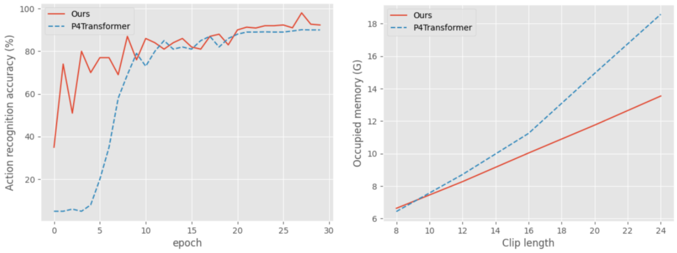

Fig. 1. Curve of Training process and Occupied memory.

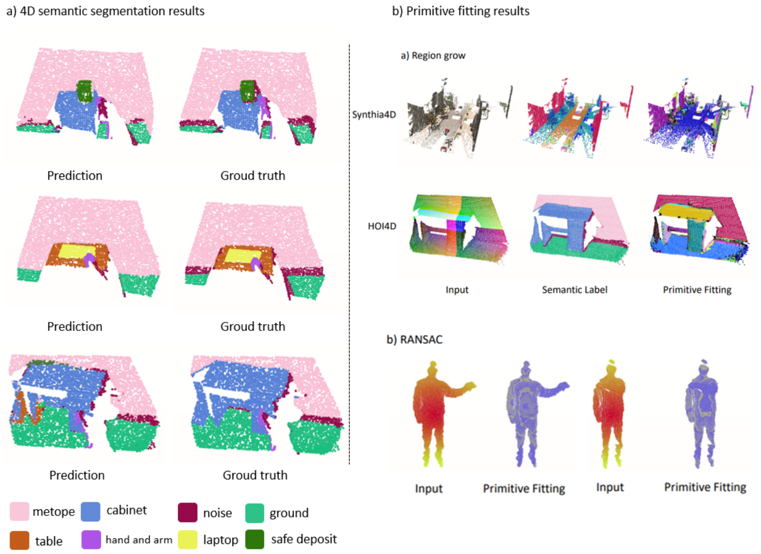

Fig. 2. Visualizations of HOI4D 4D semantic segmentation and some primitive fitting results

C Visualization

We provide visualizations of HOI4D 4D semantic segmentation and some primitive fitting results in figure 2.