Scene Flow Prior Based Point Cloud Completion with Masked Transformer (Student Abstract)

Junzhe Ding[1], Yufei Que[2], Jin Zhang[3*], Cheng Wu[4*]

School of Rail Transportation, Soochow University, Suzhou, China
[1]jzding@stu.suda.edu.cn, [2]20225246039@stu.suda.edu.cn, [3]zhangjin1983@suda.edu.cn, [4]cwu@suda.edu.cn

Abstract

It is necessary to explore an effective point cloud completion mechanism that is of great significance for real-world tasks such as autonomous driving, robotics applications, and multi-target tracking. In this paper, we propose a point cloud completion method using a self-supervised transformer model based on the contextual constraints of scene flow. Our method uses the multi-frame point cloud context relationship as a guide to generate a series of token proposals, this priori condition ensures the stability of the point cloud completion. The experimental results show that the method proposed in this paper achieves high accuracy and good stability.

Introduction

As an important expression of 3D information, point cloud has many advantages; however, in the process of data acquisition, 3D sensors are susceptible to interference from multiple factors, which leads to sparsity and incompletion of point cloud. It greatly limits the performance of downstream tasks, such as detection, segmentation, and tracking. Therefore, completion and data enhancement of the point cloud are necessary. In recent years, research in the field of point cloud completion has gained significant momentum. Some early works (Han et al. 2017; Liu et al. 2019) complemented point clouds by voxelization and 3-dimensional convolution, and these methods faced the problem of high computational cost when dealing with large-scale point clouds. With the rise of PointNet (Qi et al. 2017a) and PointNet++ (Qi et al. 2017b), direct processing of point clouds has become the mainstream method in the field of 3D analysis. However, limited by the discontinuity of the dataset, the model can only learn a certain sample one by one, which is not fully compatible with practical application scenarios. Scene flow estimation (Zhai et al. 2021) is a typical research idea for the analysis of point cloud sequences, which can effectively address the discontinuity problem. In this study, inspired by the works of (Li, Kaesemodel Pontes, and Lucey 2021) and (Pang et al. 2022), we propose a point cloud completion method using a self-supervised transformer based on forward-reverse scene flow estimation. The overall architecture is shown in Figure 1.

*Corresponding authors: Cheng Wu and Jin Zhang.

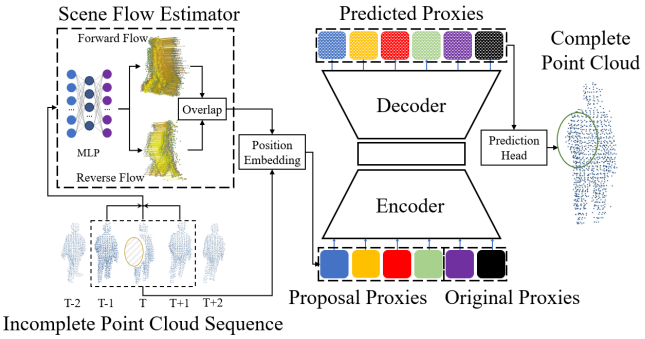

Figure 1: The Architecture of Proposed Method

Forward-reverse Scene Flow Estimation

We adopt NSFP (Li, Kaesemodel Pontes, and Lucey 2021) as the scene flow estimator. Given three point cloud frames:P_{t-1}, P_t and P_{t+1}, we estimate both the forward scene flow from P_{t-1} to P_t and the reverse scene flow from P_{t+1} to P_t. The overlapped forward and reverse scene flow estimation result is defined as P_{SF}, which is used as potential supplementary points for P_t frame. Figure 2 demonstrates the forward scene flow and the reverse scene flow.

Methodology

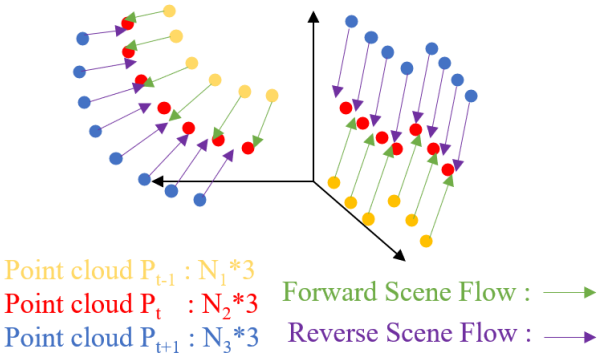

Figure 2: Forward-reverse Scene Flow

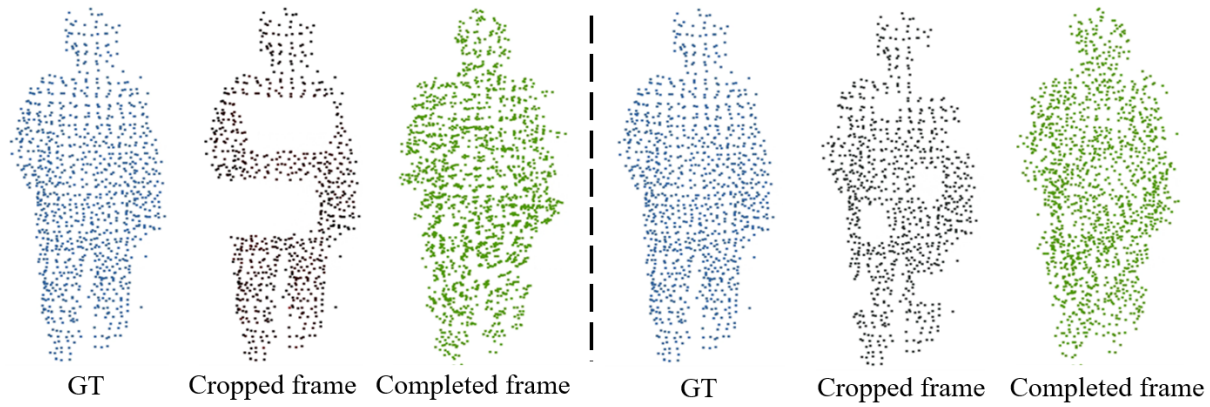

Figure 3: Results of Point Cloud Completion

Point Cloud Embedding

The P_{SF} and the P_t have a large number of overlapping parts, we need to process these parts first. The average point cloud distance D of the P_t is used as the threshold for deleting points. For each point in the P_t, we delete the points of P_{SF} within the distance D and keep the remaining points.

To generate the point cloud proxies, we are consistent with the method in (Pang et al. 2022) to get embeddings and tokens. The visible tokens are derived from P_t and the mask tokens are from P_{SF} and so as the positional embeddings.

Transformer-based Point Cloud Completion

We use the standard transformer model to complete the point cloud. The model only inputs the visible tokens and their positional embedding at the Encoder, and only at the Decoder are visible tokens and Mask tokens with all the positional embeddings fed into the network for reconstruction.

The final layer of the model uses a fully connected layer as the prediction head to map the output of the decoder to the same dimension as the input point cloud proxy, and the reconstructed point cloud is obtained by the reshape operation and used to calculate the reconstruction loss.

Experiments

Experimental setup The hyperparameters of the scene flow model and transformer model are the same as in their original code. We set the learning rate of NSFP to 8e-4 and the number of epochs to 1000 to avoid overfitting problems and the transformer model is pretrained on our pedestrian point cloud dataset collected in the laboratory with 500 rounds.

Results As shown in Figure 3, our method can effectively complete the missing part of the point cloud while keeping the overall shape. To evaluate the effectiveness of the method proposed in this paper, Fidelity Distance(Chamfer Distance between corresponding input and output) and Consistency(Chamfer Distance between different outputs) are used as evaluation metrics for the effect of point cloud completion and it scores 0.01765 and 0.01097 respectively.

Conclusions

In this paper, we innovatively propose a point cloud completion method with a self-supervised transformer model based on forward-reverse scene flow constraints. This method relies on a self-supervised model, which requires no manual labeling and can be flexibly migrated between different scenes. In addition, this method uses forward-reverse scene flow as a guide to generate a series of token proposals, which ensures the stability of point cloud completion. The experimental results show that the proposed method can achieve high accuracy and good stability on our dataset.

References

Han, X.; Li, Z.; Huang, H.; Kalogerakis, E.; and Yu, Y. 2017. High-resolution shape completion using deep neural networks for global structure and local geometry inference. In *Proceedings of the IEEE international conference on computer vision*, 85–93.

Li, X.; Kaesemodel Pontes, J.; and Lucey, S. 2021. Neural scene flow prior. *Advances in Neural Information Processing Systems*, 34: 7838–7851.

Liu, Z.; Tang, H.; Lin, Y.; and Han, S. 2019. Point-voxel cnn for efficient 3d deep learning. *Advances in Neural Information Processing Systems*, 32.

Pang, Y.; Wang, W.; Tay, F. E.; Liu, W.; Tian, Y.; and Yuan, L. 2022. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, 604–621. Springer.

Qi, C. R.; Su, H.; Mo, K.; and Guibas, L. J. 2017a. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 652–660.

Qi, C. R.; Yi, L.; Su, H.; and Guibas, L. J. 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Zhai, M.; Xiang, X.; Lv, N.; and Kong, X. 2021. Optical flow and scene flow estimation: A survey. *Pattern Recognition*, 114: 107861.