
COSE474-2023F: Final Project Proposal

3D Point Cloud Understanding using LLM

2020250028 김재현

1. Introduction

최근 방대한 이미지-텍스트 데이터에 학습된 Vision-Language foundation 모델을 통해 이미지를 이해하는 task에서 큰 발전이 이루어졌으며, 다양한 2D open-vocabulary task에 활용되고 있다. 그러나 2D 이미지 만큼의 대규모 3D 데이터-텍스트 pair는 부재함으로 인해, 이러한 2D domain에서의 pre-training 패러다임을 3D domain에서 직접적으로 적용하기는 어렵다. 하지만, 3D 데이터 또한 robot manipulation, virtual reality, human-machine interaction 등 현실에서 매우 다양한 곳에 활용되므로, 3D domain에서의 foundation 모델의 활용 역시 필수적이라 할 수 있다. 본 프로젝트에서는 Large Vision-Language model을 3D Point Cloud에 활용해 Understanding 중 Completion 및 Part Segmentation이라는 task를 수행함으로써, 3D domain에서의 LLM 활용과 동시에 좋은 성능을 달성하는 것을 목표로 한다.

2. Problem definition & challenges

Point Cloud Completion with LLM. Point cloud completion은 partial point cloud에 대한 생성 및 estimation 문제이다.(Fei et al., 2022) 현실에서 LiDAR 등의 센서를 통해 point cloud를 수집할 경우 sparse하게 수집되는 경우가 많아 알 수 없는 object에 대해서도 올바른 처리가 필요해 어려움이 있다. 이러한 경우 completion을 통해 대처해야 하므로, 3D 컴퓨터 비전의 적용 분야에서 중요한 역할을 한다. 이 point cloud completion에 LLM을 적용해보고자 한다.

Point Cloud Part Segmentation with LLM. Point Cloud segmentation은 point들을 다수의 homogeneous 영역으로 분류하는 문제이다. 높은 redundancy, 불균일한 샘플링 density, 명시적인 데이터 구조 결여 등으로 어려운 문제가 있다. navigation, 자율 mapping 등 다양한 분야에 활용된다. 이 문제를 LLM을 적용해 해결해보고자 한다.

3. Related Works

3D point cloud 데이터에 LLM을 활용하려는 연구들이 활발히 진행되고 있다. PLA(Ding et al., 2023)와 OpenScene(Peng et al., 2023)은 Language-Driven Open-Vocabulary를 통해 3D Scene Understanding을 수행한 연구

이며, Point-Bind(Guo et al., 2023)는 3D에서의 이해, 생성, instruction following을 위해 point cloud와 multi-modality를 연결한 연구이다.

4. Datasets

Dataset으로는 ShapeNet(Chang et al., 2015)을 사용한다. ShapeNet은 다양한 카테고리에 해당하는 물체에 대한 3D CAD model들의 dataset이다. 여러 주석을 포함하고 있으며, 3D 모델로부터 point cloud를 샘플링하여 변환할 수 있다. point를 샘플링하고, 이를 completion과 part segmentation task에 활용한다.

5. State-of-the-art methods and baselines

ShapeNet(Chang et al., 2015) 데이터셋에서 Point Cloud Completion task의 SOTA 모델은 AdaPointR(Yu et al., 2023)로, 6.53의 Chamfer Distance 성능을 보였다. ShapeNet(Chang et al., 2015) 데이터셋에 대해 3D Part Segmentation의 SOTA 모델은 GeomGCNN(Srivastava & Sharma, 2021)과 SPoTr(Park et al., 2023)로, 각각 instance average IoU 89.1과 97.2의 성능을 보였다. Point cloud 데이터에 대해 다양한 modality와 task를 적용한 Point-Bind Point-LLM(Guo et al., 2023)을 본 프로젝트의 baseline으로 삼는다. Baseline이 하나의 모델을 다양한 task에 적용한 방식에 기반해 completion과 part segmentation task로 확장해보고자 한다.

6. Schedule

- 현재 - 11월 중순 : Baseline 모델 이해, Point cloud completion 및 part segmentation에 LLM 적용
- 11월 중순 - 11월 말 혹은 12월 초 : 문제점 발생 시 분석 및 개선 방안 적용, 보고서 작성

References

- Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.
- Ding, R., Yang, J., Xue, C., Zhang, W., Bai, S., and Qi, X.

- Pla: Language-driven open-vocabulary 3d scene understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7010–7019, 2023.
- Fei, B., Yang, W., Chen, W.-M., Li, Z., Li, Y., Ma, T., Hu, X., and Ma, L. Comprehensive review of deep learning-based 3d point cloud completion processing and analysis. *IEEE Transactions on Intelligent Transportation Systems*, 2022.
- Guo, Z., Zhang, R., Zhu, X., Tang, Y., Ma, X., Han, J., Chen, K., Gao, P., Li, X., Li, H., et al. Point-bind & point-llm: Aligning point cloud with multi-modality for 3d understanding, generation, and instruction following. *arXiv preprint arXiv:2309.00615*, 2023.
- Park, J., Lee, S., Kim, S., Xiong, Y., and Kim, H. J. Self-positioning point-based transformer for point cloud understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21814–21823, 2023.
- Peng, S., Genova, K., Jiang, C., Tagliasacchi, A., Pollefeys, M., Funkhouser, T., et al. Openscene: 3d scene understanding with open vocabularies. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 815–824, 2023.
- Srivastava, S. and Sharma, G. Exploiting local geometry for feature and graph construction for better 3d point cloud processing with graph neural networks. In *2021 IEEE International conference on robotics and automation (ICRA)*, pp. 12903–12909. IEEE, 2021.
- Yu, X., Rao, Y., Wang, Z., Lu, J., and Zhou, J. Adapointr: Diverse point cloud completion with adaptive geometry-aware transformers. *arXiv preprint arXiv:2301.04545*, 2023.