

LLM-based Methodology for Expanding 3D Metadata in Point Cloud

Geon-Woo Kim*, Joo-Chang Kim**

* Department of AI Computer Science and Engineering, Kyonggi University, Suwon 16227, South Korea

** Contents Convergence Software Research Institute, Kyonggi University, Suwon 16227, South Korea

kgw20010916@kyonggi.ac.kr, kjc2232@naver.com

Abstract— This paper proposes a method for expanding the metadata of three-dimensional point cloud data using Large Language Models (LLMs). Currently, point cloud data plays a crucial role in various fields such as autonomous driving and medical image reconstruction, necessitating the expansion of metadata for efficient processing. Traditionally, metadata construction has relied on manual input, which is prone to errors. In this study, we propose a method that utilizes LLMs, particularly the Llama 3.1 model, to extract the center points of each class in the point cloud data and expand the metadata by adding these center points to the annotation files. By using center points, computational costs are reduced, and the performance of segmentation and detection models based on this data is improved.

Keywords— Point Cloud, Large Language Models (LLM), 3D Metadata Expansion, Attention Mechanism, Data Preprocessing

I. INTRODUCTION

Recently, research in the fields of artificial intelligence and natural language processing has been progressing rapidly. [1] In particular, Large Language Models (LLMs) in the field of natural language processing have demonstrated near-human-level comprehension across various domains. Consequently, the expansion of metadata using LLMs has become a significant research topic for AI researchers, as it enables a broader understanding of data. Among LLMs, models such as GPT and BERT are capable of learning patterns in textual data, allowing them to comprehend context and generate appropriate responses. However, these models are primarily focused on understanding two-dimensional text-based data or simple images. As a result, the application of LLMs to complex three-dimensional data, which is often encountered in real-world scenarios, remains in its early stages of development. [2]

In various modern industries, the importance of three-dimensional (3D) data is increasingly growing. 3D models, spatial information, and geographic data must be processed alongside vast amounts of metadata, necessitating effective methodologies for managing and utilizing this data. Moreover, the more metadata available to describe 3D data, the better the performance of deep learning models. Currently, the most common method for expanding metadata involves using textual data to describe the structure, attributes, features, and relationships of 3D objects, a process that requires human

input. However, this method of metadata expansion is prone to errors and missing values, as it relies on manual construction by humans.

Therefore, this paper proposes an objective and accurate method for expanding 3D metadata using LLMs capable of understanding data from various domains. The proposed methodology utilizes point cloud data as input to an LLM model and expands the metadata by adding the center point of the corresponding class as a new index in the annotation file. This approach enhances the performance of models that process point cloud data. The contributions of the proposed method are as follows:

- Proposing an approach for feature extraction from point cloud data using LLM models: By inputting point cloud data into an LLM, we propose a novel methodology that provides the model with the location of the class. This approach offers a new perspective on understanding point cloud data and opens up the possibility of exploring its application in various metadata expansion methods.
- Performance improvement of segmentation models based on LLM-driven point cloud expansion: The performance of the proposed methodology is objectively validated by comparing the performance of segmentation models on datasets with and without metadata expansion. This demonstrates the superiority of the proposed approach and highlights its contribution to improving segmentation model performance through metadata expansion.

The structure of this paper is as follows. In Section 2, we explain the existing methods for processing point cloud data and the data preprocessing techniques. Section 3 describes the method of applying point cloud datasets to LLMs to expand metadata and applying it to segmentation models. Finally, Section 4 concludes the paper.

II. RELATED WORK

A. Basic Methods for Processing Point Cloud Data

Point cloud data processing has become a significant research topic in fields where computers are tasked with understanding the real world, such as autonomous driving and medical image reconstruction, with scene graph segmentation being widely applied. This section provides an overview of

basic point cloud data processing methods and introduces widely adopted approaches and deep learning models in the field.

Point cloud data processing can be generally divided into projection, voxel, and point-based methods. The projection method involves projecting the point cloud onto a 2D image plane and processing it using 2D convolution operations. Similarly, the voxel method divides the 3D point cloud into uniform voxels and also utilizes 2D convolution operations. Both the projection and voxel methods have limitations as they project the continuous characteristics of point clouds into discrete spaces, leading to the loss of inherent features.

Currently, a point-based method is used, which inputs the point cloud directly into the model without projection or voxelization. Charles et al. [2] proposed PointNet, which processes raw point clouds and directly captures features by reflecting the non-canonical characteristics of point clouds. In a subsequent study, Charles et al. [3] added a hierarchical structure to the PointNet model, making the extraction of local features more efficient. However, both models have a limitation in that, while they perform scene segmentation without preprocessing the point cloud, they incur high computational costs.

B. Advanced Methods for Processing Point Cloud Data

The attention mechanism operates by calculating the weights of the whole in relation to a specific part. Hengshuang et al. [4] proposed Point Transformer, which applies the attention mechanism, unlike PointNet and PointNet++, for processing point clouds. To handle point cloud data, individual points are vectorized, and query vectors are calculated by designating points for comparison with key vectors. Finally, the key vector is used as the value vector to calculate similarity. However, a limitation of this approach is the high computational cost, as it calculates the similarity between all points and considers the positional coordinates of the points.

In this study, we propose a new approach based on existing point cloud data processing methods that utilizes LLMs to preprocess the data. This approach has the potential to offer faster processing speeds and better capture local features compared to traditional methods. Thus, this research is expected to provide a technological advancement and a new perspective on processing 3D data through LLMs.

III. LLM-BASED METHODOLOGY FOR EXPANDING 3D METADATA IN POINT CLOUD

Point cloud data is collected through LiDAR sensors and contains points that represent objects in indoor or outdoor environments. Point clouds consist of files containing points and annotation files that describe the position and features of those points. To expand the information within 3D metadata, this study proceeds in three stages. In the first stage, the data used for the experiment is input into the Llama 3.1 model. At this stage, the Llama 3.1 model identifies the coordinates of the center point for each class point set in the point cloud. In the second stage, the center point coordinates are stored in the

feature index of the annotation. Finally, in the third stage, the structure is modified so that the point cloud processing model can input the data expanded by Llama 3.1. Figure 1 illustrates the overall process of the methodology proposed in this paper.

A. Center Point Extraction

The data used in this study consists of the bin files of point cloud data and their corresponding annotations. The bin files are composed of a collection of light reflections captured by LiDAR sensors and include multiple classes, such as walls and curtains. By inputting the data into the Llama 3.1 model, the coordinates of the center points for the classes are extracted. Instead of inputting all points into the model and calculating the similarity between them, assigning weights to specific center points reduces computational costs. Therefore, the data is preprocessed by following these steps.

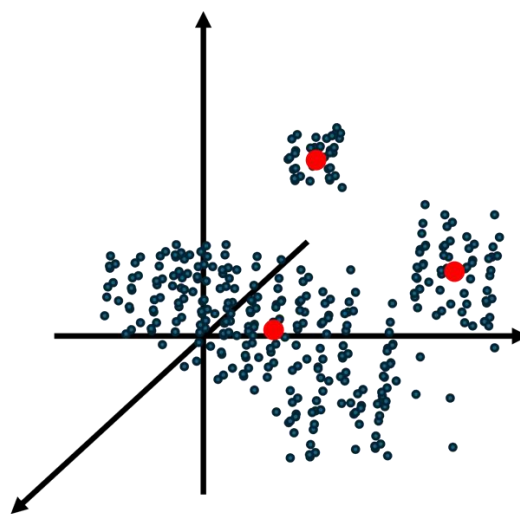


Figure 1. Center Point Analysis Results

Figure 1 shows the results of the Llama 3.1 model analyzing the bin file of the point cloud. The best way to approximate the locations of the classes is by finding their center points. Once the center points are identified, using algorithms like KNN to perform tasks such as segmentation or detection from these center points can improve model performance while reducing computation time. The Llama 3.1 model receives the input command “analyze the bin file and extract the coordinates of the central point in areas where points are densely clustered,” and from this, it extracts the center points.

B. Concat Center Point

This stage involves linking the extracted center points to a suitable format for application to the point cloud processing model. Through this step, metadata expansion is achieved, amplifying the informational content of the original data.

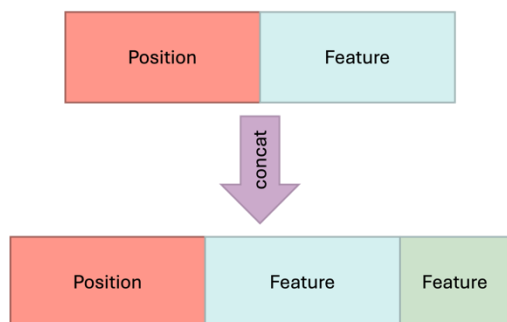


Figure 2. Example of an unacceptable low-resolution image



Figure 2 illustrates the concatenation of center points with the existing annotation file. In the case of non-center points, their positions may overlap with those of other points, potentially causing confusion for the model. Therefore, all center points are added at the end of the feature index.

C. Modifying the Model Structure

In this section, we explore the methodology for applying the concatenated data from Section 3.2. The traditional Point Transformer uses the value vector as the key vector. To calculate similarity, the key vector being compared is simply multiplied twice. While this method has the advantage of calculating more accurate similarities, it also has the limitation of computing potentially meaningless vectors. To address this, our study sets the center point as the value vector in the Point Transformer, while using the points to be compared as the query and the surrounding vectors as the key. This approach aims to better segment local features while reducing computational costs.

IV. CONCLUSIONS

This study explored an approach where point cloud data is preprocessed using an LLM model to provide the model with prior information before performing tasks. The proposed method is expected to outperform existing point cloud processing techniques. We introduced a novel preprocessing approach for three-dimensional data, which best represents real-world scenarios. Future experiments will validate the

generalization performance of this method and explore new possibilities to enhance its versatility and applicability.

ACKNOWLEDGMENT

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No: RS-2023-00248899).

REFERENCES

- [1] K. Chowdhary and K. R. Chowdhary, "Natural language processing," in *Fundamentals of Artificial Intelligence*, 2020, pp. 603-649.
- [2] M. Callieri, P. Cignoni, F. Ganovelli, G. Impoco, C. Montani, P. Pingi, and R. Scopigno, "Visualization and 3D data processing in the David restoration," *IEEE Comput. Graph. Appl.*, vol. 24, no. 2, pp. 16-21, 2004.
- [3] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 652-660.
- [4] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [5] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point Transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 16259-16268.

Geon-Woo Kim has bachelor's degree Division of Computer Science and Engineering. Kyonggi University, South Korea, since 2020. He has worked as a researcher at the Data Mining Lab., Kyonggi University. His research interests include Data Mining, Big Data, Anomaly Detection, Deep Learning.

Joo-Chang Kim has received B.S. and M.S. degrees from the School of Computer Information Engineering, Sangji University, South Korea in 2014 and 2016, respectively. He has received Ph.D. from Department of Computer Science, Kyonggi University, South Korea in 2021. He is currently a research professor in Contents Convergence Software Research Institute, Kyonggi University. Since 2021, he is currently a lecturer in the Department of Software Convergence Engineering, Inha University, South Korea. His research interests include data mining, data management, knowledge systems, machine learning, deep learning, big data, healthcare, and recommendation systems.