ISPRS Journal of Photogrammetry and Remote Sensing 212 (2024) 181–192

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Chat3D: Interactive understanding 3D scene-level point clouds by chatting with foundation model for urban ecological construction

Yiping Chen [a], Shuai Zhang [a,*], Ting Han [a], Yumeng Du [a], Wuming Zhang [a], Jonathan Li [b]

[a] School of Geospatial Engineering and Science, Sun Yat-sen University, ZhuHai 519082, GuangDong, China
[b] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Point cloud understanding
Large language model interact
Urban ecological construction
Prompt engineering
Thought chain

ABSTRACT

With the artificial intelligence technology development boom, large language models are demonstrating their potential in comprehension and creativity. Large language models such as GPT-4 and Gemini have been able to powerfully study for various professional-level exams. However, as a language model itself, its powerful comprehension can only be reflected in text sequences. Currently, although videos can be generated through the connection between 3D point clouds and large language models, there is currently no prompt project that directly interacts with one-dimensional through attribute calculation results. The point cloud data is also rich in information that can support various tasks of urban construction. For scene-level point cloud data, there has been a lot of research done on semantic segmentation, target detection, and other tasks. However, it is usually difficult to provide direct help to scene construction from the perception results. This paper presents a method for applying large language models to urban ecological construction by combining the results of 3D point cloud semantic segmentation. The objective is to integrate the prior knowledge and creative capabilities of Large Language Models (LLMs) within urban development with the outcomes derived from point cloud semantic segmentation results. This integration aims to establish an interactive point cloud intelligent analysis system, tailored for aiding decision-making processes in urban ecological civilization construction, thus presenting innovative perspectives for the advancement of smart city development.

1. Introduction

A new surge in the development of artificial intelligence models has erupted recently. Benefit to the popularization of the attention mechanism and the high performance of the transformer-based network architecture (Vaswani et al., 2017), large pre-trained models emerged and rapidly became popular. Because of the ability to learn general-purpose language representations unsupervised from large-scale data, pre-trained models are quite effective for many downstream tasks, especially NLP tasks. The successive emergence of large language models (Wei et al., 2022) such as GPT-3 (Brown et al., 2020), PaLM (Chowdhery et al., 2022), and Llama 2 (Touvron et al., 2023) which utilize their significant advantages in understanding the text, demonstrates the trend towards large pre-trained models. An enormous amount of research confirms that models mostly follow the scaling law (Kaplan et al., 2020) and that increasing the capacity of the model can significantly improve the modeling results. Meanwhile, the pre-trained model trained by a large amount of data has a strong generalization ability and approximate thinking logic ability. The pre-trained language model is able to give reasonable responses based on input data and textual prompts, combined with an analysis of the literature in the database.

LLMs have been used in a wide range of application situations motivated by their large knowledge base and powerful text comprehension capabilities. With reasonable prompts, LLMs can be motivated to make scientific responses and hypotheses that drive research and application efficiency. A lot of research on prompt engineering has also emerged (Liu et al., 2023b). However, the ability to understand high-dimensional image information or point cloud information is limited. Dealing with object-space relationships involved in high-dimensional information and recognizing texture patterns and target features requires more in-depth high-dimensional processing capabilities that are not available in language models. Inspired by natural language processing methods, transforming the spatial features of an image into textual signals for input to LLMs has become a common approach across modalities. Currently, such methods are used in medically related image processing tasks and better results are obtained.

Point clouds, as high-dimensional spatial data capable of representing depth information and three-dimensional vertical structures with high precision, are extensively utilized due to their distinctive features. Unlike remote sensing imagery, point cloud offers a three-dimensional

* Corresponding author.
E-mail address: zhangsh255@mail2.sysu.edu.cn (S. Zhang).

https://doi.org/10.1016/j.isprsjprs.2024.04.024
Received 21 February 2024; Received in revised form 8 April 2024; Accepted 22 April 2024
Available online 9 May 2024
0924-2716/© 2024 Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS).

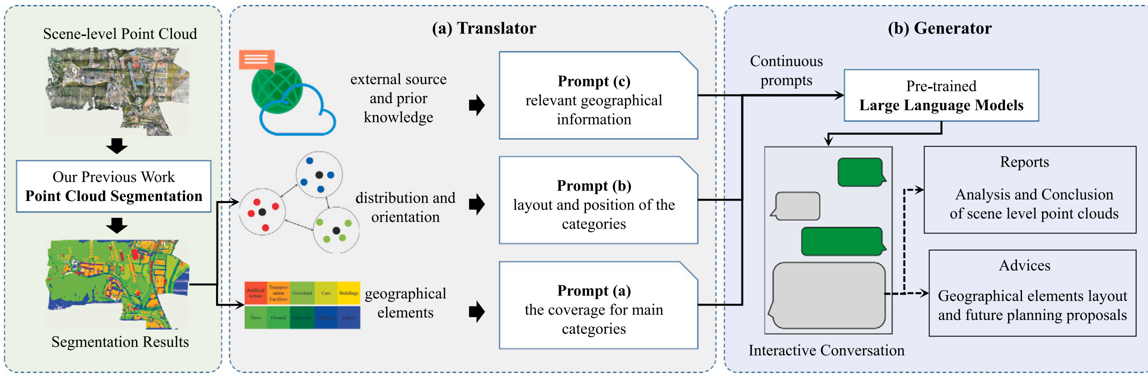

Fig. 1. An illustration of the Chat3D architecture. The architecture consists of two main modules: (a) Translator and (b) Generator, to provide a comprehensive analysis of the scene-level point cloud and to suggest layout optimizations. The Prompt (a), (b), and (c) denote the coverage for main categories from semantic segmentation, the layout and position of the categories from point clouds and text, and the relevant geographical information from external source and prior knowledge based on LLMs, respectively.

representation of the environment, enabling detailed spatial analysis and modeling for various applications such as urban planning and disaster management. Its high-dimensional nature and diverse acquisition methods make it valuable for capturing intricate terrain features and vertical structures accurately. Meanwhile, with the rise of the autonomous driving industry and the urgent need for smart city construction, point cloud sensing has become a research hotspot. Numerous techniques already exist for semantic segmentation of point clouds, including Qi et al. (2017a,b), Liu et al. (2019b), Zhao et al. (2021) and Zhang et al. (2023). These methods classify points in the scene into different categories combining local and global features for further research. However, the segmentation result of the network is only labeled data and cannot be directly used for subsequent research. Most of the city-building tasks, after obtaining the experimental results, also need to obtain environmentally relevant indexes based on a priori knowledge of the local geographic conditions and climate. The calculated indicators are then used to assess the probability of occurrence of ecological risks, which in turn generates a detailed report on the construction of the city. The urban construction report is able to assess the urban risk level, monitor the ecological index, and greatly assist in the construction of smart cities. However, urban construction reports require a large amount of information to be referenced, as well as a large amount of manual assistance to participate in order to ensure the accuracy of the report, and is highly influenced by artificiality. Relying solely on manual analysis would significantly diminish management efficiency. To the best of our knowledge, there are currently few large-scale 3D semantic scene understanding, ecological management and natural language models. Integrated paper publication.

Therefore, this paper proposes an interactive understanding scene-level point cloud by large language model for urban ecological construction method, called Chat3D. In this scheme, we first obtain urban scene-level point cloud semantic segmentation results. Further, we transform the 3D segmentation results into textual prompts for a pre-trained large language model as shown in Fig. 1. Simultaneously, we sequentially entered different levels of text prompts, which included information about the geographical elements of the city, the layout and position of category features in the city, etc. Then, LLMs subsequently condense the aforementioned prompts and conduct a comprehensive search of resources and prior knowledge, drawing from its native training database and an amalgamation of information procured through networking searches aligned with the prompt. This process entails the computation of environmental parameters and culminates in the generation of an eco-construction report tailored to the designated study area. The experiments demonstrate that large language models can accurately understand three-dimensional semantic segmentation results with effective textual prompts. It will accurately calculate the environmental parameters(environment index, EI) of the area based on the subsequent prompting engineering. Meanwhile, the generated report on urban ecological construction is very detailed, including the rationalization analysis of the construction of a variety of features and the corresponding adjustment methods with highly informative.

The main contributions of our method are summarized as follows:

- We propose a comprehensive solution that utilizes semantic segmentation, distribution, geographic location, and other information derived from three-dimensional point cloud geographic features as prompts. Leveraging the extensive knowledge and automated search capabilities of large models, our approach enables accurate and rapid analysis of ecological environments, called Chat3D.
- During the interaction, we propose the utilization of various levels of prompts obtained from three-dimensional point cloud environment perception algorithms. Furthermore, the integration of the model's knowledge repository and online search functionality enhances the comprehensiveness and reliability of text generation.
- The experimental results show that Chat3D can accurately calculate the local eco-environmental index based on the input data combined with geographic prompts. Specifically, the environmental index of the study area calculated based on Gemini's Chat3D is 82.5, which represents an error of only 2.7 from the officially published result (EI = 85.8). The generated report on urban ecological construction can assess the probability of the occurrence of urban ecological risks, as well as evaluate the rationality of the city's functional structure and adjustment program. The generated reports are detailed, accurate, and adhere to sustainable development principles, thus significantly contributing to urban ecological services.

2. Related works

2.1. Large-language model

This study aims to design an intelligent ecological analysis system by integrating point clouds and Large-Language Models. At present, most combinations between point clouds and Large-Language Models require text as an intermediary. Therefore, the related work of this paper will be divided into four parts: Large-Language Model, Image-Text Model, Point-Text Model and Ecological Analysis Methodology.

Large-Language Models (LLMs) represented by transformer-based GPT series (Radford et al., 2018, 2019) and BERT (Devlin et al., 2018; Lan et al., 2019; Liu et al., 2019a) have dominated the field of natural language processing (NLP) in recent years. With supervised instruction tuning and reinforcement learning from human feedback, LLMs exhibit surprisingly effective zero- and few-shot generalization abilities to perform almost any NLP tasks (Su et al., 2023; Mishra et al., 2021; Ouyang et al., 2022; Victor et al., 2022). BERT model adopts bidirectional encoder architecture and masked language modeling techniques to process random text and reconstruct the original context, which is jointly conditioned on both the right and left context in all layers. On the contrary, the GPT series that utilizes the auto-regressive prediction approach is based on the decoder-only architecture. Based on the existing large corpus of texts, they can learn more language knowledge and produce more natural statements automatically.

As one of the most popular examples, ChatGPT (OpenAI, 2023a) has garnered extensive attention in several specific domains for its powerful ability to write, translate, analyze, and generate high-quality human-like text (Brown et al., 2020; Singhal et al., 2022). Many researchers in medicine (Kung et al., 2023; Sinha et al., 2023; Patel and Lam, 2023), environment (Biswas, 2023), engineering (Prieto et al., 2023), electronic (Tafferner et al., 2023), and other fields are conducting a variety of attempts to explore the potential of it across a diverse range of downstream tasks. For instance, ChatGPT has shown great intelligence and professionalism in various vocational examinations and can give specialized analysis for specific exam questions, which displays its latent ability to help users make reasonable and well-considered decisions. In the United States Medical Licensing Exam, ChatGPT even shows greater talent compared to PubMedGPT (Venigalla et al., 2022), a counterpart LLM with similar neural structure, but trained exclusively on biomedical domain literature (Kung et al., 2023). However, exam experiments also showed that ChatGPT is still a poor judge of its correctness. That being said, its confidence has little bearing on the correctness of its response (Jalil et al., 2023; Fijačko et al., 2023).

2.2. Image-text model

While ChatGPT can only accept text input, GPT-4 can adopt images as its inputs and generate captions, classifications, and analyses, which means it can realize images (OpenAI, 2023b). Some researchers also aim to combine vision alignment with LLMs for visually-grounded instruction following, such as LLaVa (Liu et al., 2023a) and Mini-GPT4 (Zhu et al., 2023). Although they possess numerous advanced vision-language capabilities, they still have limitations in recognizing detailed textual information from images and differentiating spatial localization. DetGPT (Pi et al., 2023) leverages state-of-the-art multi-modal models and open-vocabulary object detectors to perform reasoning within the context of the user's instructions and the visual scene. In addition to this, some domain-specific LLMs based on GPT models such as ChatCAD, which integrated LLMs into medical-image CAD networks, successfully merged the strengths of ChatGPT's medical domain knowledge and logical reasoning with the vision understanding capability (Wang et al., 2023). But its-generated reports are not very human-like in a certain way and more complex prompts needed to be tested based on this model.

What is more, the advances in LLMs also make great inspiration in the computer vision field. Motivated by the GPT series, VISORGPT applies the decoder-only architecture in modeling the visual probabilistic prior. By discretizing visual locations of objects, e.g., bounding boxes, human pose, and instance masks, into sequences, it can model visual prior through likelihood maximization (Xie et al., 2023). Image-GPT trains a sequence transformer to auto-regressively predict pixels without incorporating knowledge concerning the 2D input structure, exhibiting promising representation learning capabilities after pre-training (Chen et al., 2020). SAM takes advantage of the prompt idea in NLP tasks to quickly segment any object by providing a prompt to the image segmentation task, but its generality and breadth of use limits its accuracy when applied in a particular field (Kirillov et al., 2023).

2.3. Point-text model

At present, LLMs are not capable of understanding point cloud data directly. Some researchers aim to extend the concept of LLMs to point cloud processing. CLIP2 exploits naturally-existed correspondences in 2D and 3D scenarios and builds well-aligned and instance-based text-image-point proxies from those complex scenarios (Zeng et al., 2023). This method constructs intermediate 2D representations for the 3D data retaining its 3D geometry information, which can better adapt the success of 2D Vision-Language Models to the 3D space. mmPose-NLP, a novel NLP-inspired Sequence-to-Sequence (Seq2Seq) skeletal key-point estimator using millimeter-wave (mm-wave) radar data is also an example (Sengupta and Cao, 2022). It uses a similar procedure to NLP to process point cloud data. Moreover, PointGPT has improved classification accuracy on multiple datasets successfully by proposing a point cloud auto-regressive generation task to pre-train transformer models, which is inspired by the advancements of the GPT (Chen et al., 2023).

PointCLIP (Zhang et al., 2022b) is an early attempt to integrate multimodal models into the realm of point clouds. Like CLIP (Radford et al., 2021), it converts point cloud data into depth images and utilizes image-text pairs to predict image classifications, thereby enabling the mapping of textual content onto point cloud data. Cap3d (Luo et al., 2024) extends the methodology of PointCLIP by projecting the input 3D model into multiple views, creating multi-views. It then utilizes the BLIP+CLIP (Li et al., 2022) method to enhance the understanding of the relationship between images and text. Additionally, it combines the results from CLIP with manually crafted prompts and inputs them into the GPT language model to generate detailed point cloud-text paired data.

Similarly, UNiG3D (Sun et al., 2023) obtains multi-angle meshes and images from 3D models, in contrast to Cap3D. The aforementioned works focus on text understanding within individual 3D objects, whereas 3dLLM (Hong et al., 2024) distinguishes itself as the first multimodal model capable of comprehending entire three-dimensional scenes. It employs a multi-view approach to represent point clouds at the scene level, followed by the extraction of 2D features using an image encoder. Additionally, it generates three-dimensional reconstructions, SLAM (Simultaneous Localization and Mapping), Nerf (Neural Radiance Fields), and other three-dimensional features based on these 2D features. These three-dimensional features are subsequently fed into the LLM, which is prompted to analyze and understand the scene using a Q&A format, providing answers based on its comprehension.

PointLLM (Xu et al., 2023) is the pioneering model that enables LLMs to directly comprehend three-dimensional features. It processes point cloud data directly through a Point Encoder to obtain three-dimensional features, which are then inputted directly into the LLM for analysis. Moreover, the model is prompted to obtain textual analysis results through a Q&A format. However, its ability to understand three-dimensional semantic information is still limited to the relatively single model, which means it is not useful for large-scale complex scenes.

The fusion of 3D point clouds and LLM has attracted increasing attention. 3D captioning (Chen et al., 2021; Liu et al., 2024; Zhou et al., 2023) is employed to describe different objects within a 3D scene. Meanwhile, 3D LLM (Hong et al., 2024), Scene-LLM (Fu et al., 2024) and 3DMIT (Li et al., 2024) focus on the holistic understanding of the scene. However, these methods have not to explore applications in outdoor environments. Despite LiDAR-LLM (Yang et al., 2023) incorporating visual grounding to combine LLM and point clouds into autonomous driving tasks, it only considers object localization and overlooks the semantic understanding of point clouds.

2.4. Ecological analysis methodology

With the rapid development of remote sensing observation technologies, assessing ecological environmental quality using remote sensing data has become a crucial issue. A comprehensive understanding and scientific assessment of regional ecological environment can help us analyze landscapes and make plans for urban development (Popp et al., 2000; Wang and Zhong, 2017). Ecological Environment Index in Technical Criterion for Ecosystem Status Evaluation (HJ 192–2015) (TCESE 2015) has been proposed by the Ministry of Ecology and Environment of the People's Republic of China to build a reference standard for regional ecology assessment. Early studies (Sun et al., 2012; Larsson and Hanberger, 2016) often relied on statistical data or land use data. However, the objectivity of data and research periods are limited. Further, De Keersmaecker et al. (2014) and Li et al. (2017) referenced the land cover and normalized vegetation index to form ecosystem status evaluation. And Xu et al. (2013) and Hu and Xu (2018) used the Remote Sensing Ecological Index to evaluate the ecological quality. Zhang et al. (2022a) introduced more parameters such as normalized difference vegetation index, wetness, normalized differential build-up and bare soil index, and land surface temperature to construct a comprehensive index to evaluate the ecological quality. Shan et al. (2019) developed the remote sensing theoretical framework to provide a contrastive analysis between Remote Sensing Ecological Index and Ecological Index. According to aforementioned studies, Yu et al. (2022) proposed Ecological Livability Index, covering greenness, temperature, dryness, water-wetness, and atmospheric turbidity, to demonstrate that ecological environment index has an exemplary embodiment in urban ecological research. Das et al. (2023) based on these five ecological parameters developed a comprehensive urban ecological framework to achieve spatial landscape planning. Therefore, landscape ecological analysis provides a new impetus for urbanization research. Xu et al. (2021, 2022) considered the multi-source remote sensing data and the interactive coupling mechanism between urbanization and eco-environmental quality to provides a new perspective for the research on the urban sustainable development. However, existing studies based on multi-source remote sensing data overlook the crucial 3D point cloud, which can provide volumetric spatial observations and structural information. In addition, large language models (such as ChatGPT etc.) can both access networked ecological data to reduce measurement efficiency and provide insights into the connection between ecological analysis and urban planning when prompted by researchers. Therefore, integrating point cloud and large language models to conduct research on urbanization development using ecological parameters is valuable.

3. Chat3D

The overall architecture of Chat3D is illustrated in Fig. 1. The Chat3D is comprised of two components, the Translator and the Generator. The input to the pipeline is the segmented scene-level point clouds. The translator forms the textual understanding of semantic point clouds. Further, the generator concludes and analyzes the translation results using logical reasoning capabilities. The summary report will provide effective recommendations for optimizing the city's layout.

3.1. Point cloud to text translator

The critical aspect of the architecture design is how to utilize the powerful logical reasoning capabilities of the pre-trained LLMs to perform robust and intelligent analysis of the point cloud. We employ a translator to translate point cloud into text to connect the LLMs. The translator consists of two steps: (1) Following our previous work, the scene-level point cloud is semantically segmented into ten different classes;Note that the 10 categories are only properties of the study area and the number of categories is not a limiting condition for the algorithm. The algorithm can be applied to a variety of scenarios any category. In addition, the method is based on point cloud semantic segmentation results, and it can employ any kind of segmentation network to accomplish this step. (2) Translation of the segmented point cloud into textual information that can be recognized by LLMs. The strategy of the second step is described in more detail below.

Prompt (a): Scene-level point clouds are indicated by the classifier into ten different classes, i.e., artificial terrain, transportation facilities, buildings, cars, grassland, trees, ground, sidewalk, highways, and others. It is possible to calculate the percentage of points in the scene for each category based on the number of points in that category since the number of different classes of point clouds is stable before and after segmentation. This percentage is considered as the coverage of the category. The process is formulated as Eq. (1):

$$Coverage_i = \frac{N_i}{N_{total}} \tag{1}$$

where i is denoted as a category, N_i and N_{total} are indicated as the class i points and total points, respectively. The core of the second step is the translation of the point cloud segmentation results into the natural language to generate the prompt. The prompt (a) is designed to correlate with coverage. The coverage needs to be translated into a prompt for LLMs. The prompt (a) is the percentage of a particular category or all categories separately. And we do not need to normalize the form of the prompt, because even if it is a long paragraph, LLMs can extract critical information and translate it into an abbreviated tabular form.

Prompt (b): However, it is difficult for a single prompt to fully parse the point cloud and give a truly valuable report. Thus, prompt (b) is designed using geographical distribution, which assists in the analysis of the orientation of the categories. (Prompt(b) does not come from the semantic segmentation results, rather this is the user's a priori knowledge of the study area.) The distribution of natural geographic entities is computed and inferred from the class of points. For example, grasslands are concentrated in urban areas, while trees are located in suburban areas. In general, trees are surrounded by grasslands. Vegetation is more likely to occur in the nearby neighborhoods of rivers and lakes. Due to the uneven and scattered distribution of geographical entities in the different layouts, the orientation expresses the position of the subject of the category. You may be able to choose the center, south-east-north-west, south-east, north-east, etc., as orientation words. For example, the prompt could be written as "The buildings are mostly located in the campus center". Even the orientation relationships of different geographical entities can be employed as prompt, such as "The lake lies in the southeast of the woods, a long way from the buildings".

Prompt (c): Prompt (c) is networked data and prior knowledge obtained from an external source. The networked data and prior knowledge mainly include relevant geographical information such as local climatic conditions, weather conditions, and hydrological land conditions, and may also include statistics with time-series relationships. Prompt (c) benefits to generate better reports. Further, these data are used to assist the generator in making comprehensive analyses and predictions and giving reasonable recommendations.

3.2. Report and advises generator

The generator is employed to deduce on the basis of accurate translations via logical ability. The role of the generator is divided into three parts: (1) Firstly, the conversation about the understanding of scene descriptions is engaged in forming interactions; (2) Secondly, the generator utilizes LLMs to summarize the prompts and makes a conclusion; (3) Finally, based on the conversation, comprehensive reports and suggestions for layout optimization are deduced in the generator.

In order to generate more rational reports and valuable recommendations, we adopt a thought chain-like approach to prompt the LLM, rather than completing all prompts in a single step. Continuous prompts optimize the understanding ability of LLMs to adapt scene analysis. The process of question & answer structure in the form of conversation is seen as a domain-specific optimization process for LLMs. For example, we incorporate the point cloud into calculating the ecological environment index (EI), which is a composite of a series of indices that reflect the state of regional ecological quality. The vegetation coverage index (VI), river density index (RI), building coverage index (BCI), and land

Table 1
Evaluation indicators and calculation methods of ecological environment index.

Evaluation indicators	Definition	Calculation method
Biodiversity index	Evaluate the abundance of biological abundance in the region via difference of various ecological environments	$BI = 0.35 \times trees + 0.21 \times grasslands + 0.28 \times waters + 0.11 \times croplands + 0.04 \times buildings + 0.01 \times others$
Vegetation coverage index	Reflects the extent of vegetation coverage by calculating the proportion of forests, grasslands and croplands	$VI = 0.38 \times trees + 0.34 \times grasslands + 0.19 \times croplands + 0.07 \times buildings + 0.02 \times others$
River density index	Assessment of watershed areas and water percentage	$RI = \alpha \times length/area + \beta \times percentageoflakes + \gamma \times waterresources/area$
Land stress index	Reflect the extent of land degradation in the target area, including mild, moderate and severe	$LI = 0.05 \times mild + 0.25 \times moderate + 0.7 \times severe$
Pollution load index	Evaluation of environmental pollution pressures on the territory	–
Environmental restriction index	ERI is a constraining indicator that places limits on ecological status	–

stress index (LI) are calculated from the coverage during the conversation, where the land stress index is further transformed from building coverage index. The biodiversity index (BI) is acquired by calculating tress, grasslands, waters, croplands and buildings. Additional data such as pollution load index (PLI), and environmental restriction index (ERI) are acquired from networked data and prior knowledge. Each indicator is calculated as shown in Table 1.

According to the above indicators, the ecological environment index is mathematically expressed in Eq. (2):

$$EI = \alpha * BI + \beta * VI + \gamma * RI - \delta * LI + \epsilon * (100 - PLI) + ERI \tag{2}$$

where α, β, γ, δ, and ϵ are normalization parameters. The equation is referenced from Ecological Environment Index in Technical Criterion for Ecosystem Status Evaluation (HJ 192–2015) (TCESE 2015) proposed by the Ministry of Ecology and Environment of the People's Republic of China.

The obtained ecological environment index forms a conversation interaction with the generator. There are a total of five levels according to the specific values of the ecological environment index, including excellent, good, fair, poor, and very poor. The comparison of current data with historical data is categorized into four levels, which are none, slight, obvious, and significant, respectively. Then, the generator integrates all the prompts and combines them with inherent knowledge to conclude the analysis of the scene-level point cloud. An ecological analysis report is generated, as shown in Fig. 1(b). Moreover, we enrich the report by analyzing trends and ecological fluctuations in conjunction with time-series data. The ecological report is strong support for the planning and construction of the city. Finally, the generator will produce a proposal on how to adapt the layout and how to optimize the ecology under a sustainable development strategy.

4. Experience

We aim to use the powerful logic understanding capability of LLMs to analyze the semantic segmentation results of city-level 3D point clouds, and then get relevant suggestions for ecological or urban construction. The figures show an example where we use SYSU9 as the object of study. The dataset is obtained in the field at Sun Yat-sen University using drone scanning and contains nearly 200 million points, in total. A point cloud semantic segmentation network is used to segment the point cloud into a total of ten categories. Sun Yat-sen University covers an area of about 3.571 square kilometers. We acquired a total of 172,092,897 points and segmented the internal 171 819 904 points. Among them, there are 10,246,343 points for car-shaped roads, 5,072,780 points for sidewalks, 857,667 points for natural ground, 85,785,082 points for trees, 230,248,410 points for grass, 565,102 points for vehicles, 22,033,300 points for buildings. There are 40,843 points for traffic facilities (such as street lights, utility poles, traffic lights, etc.), 6,671,688 points for man-made terrain, and 9,833,289 points for other categories (including water). We hope to use LLMs' data processing power and logical ability to give opinions on campus ecology construction with the above information. "The higher the number of points, the greater the proportion of the feature" is the basic rule for this analysis. Then, the following template is used to convert the split result to language, "${category} point_number: ${point_number}" as the Prompt(a).

With the input of Prompt(a), LLMs can get the data specific to each category. The model can rush to death the proportion of each category. Figs. 2, 3 show the output report by entering Prompt(a). However, the specific number of points is not a common mode of analysis for ecological proposals, and a description of the location of the feature has also been added to obtain Prompt(b). The aim is to increase the model's spatial understanding of this data. Fig. 4 shows the output report by entering Prompt(b). Finally, we use language to suggest geographic information related to the model(Prompt(c)). Fig. 5 show the output report by entering Prompt(c).

To measure the LLMs' ability to understand the data, we set up the following three questions about the environment for testing. (1) Combining the local +biodiversity index, pollution load index, and environmental restriction index, please calculate the ecological environment index of this campus. And analyze whether the heat island effect will occur in the area. (2) A detailed analysis is given of the type and distribution of features on the campus and an assessment of the probability of risk from environmental problems. (3) Is the distribution of features on this campus reasonable, if so please analyze the rationality or not, how will it be adjusted? Please design a new and reasonable category distribution ratio with the functional positioning and structural characteristics of the campus, and give reasons.

4.1. Qualitative comparison of generated text

We evaluate the performance of our proposed method with two different LLMs, ChatGPT and New Bing(GPT-4 Online), which are easy-to-use and simple interactive language models. Of the three issues mentioned above, the quantitative evaluation of the issue of recommendation generation is more difficult to determine. Therefore, we have used question(1) for the quantitative evaluation of the method. The language model uses its powerful knowledge base and computing power to calculate an environmental index for the area, using the data and geographical prompts we input.

The foundational model, based on input point cloud segmentation results and prompts at different levels, combined with the backend database of the large model, can compute VI, RI, BCI, and LI. It is also

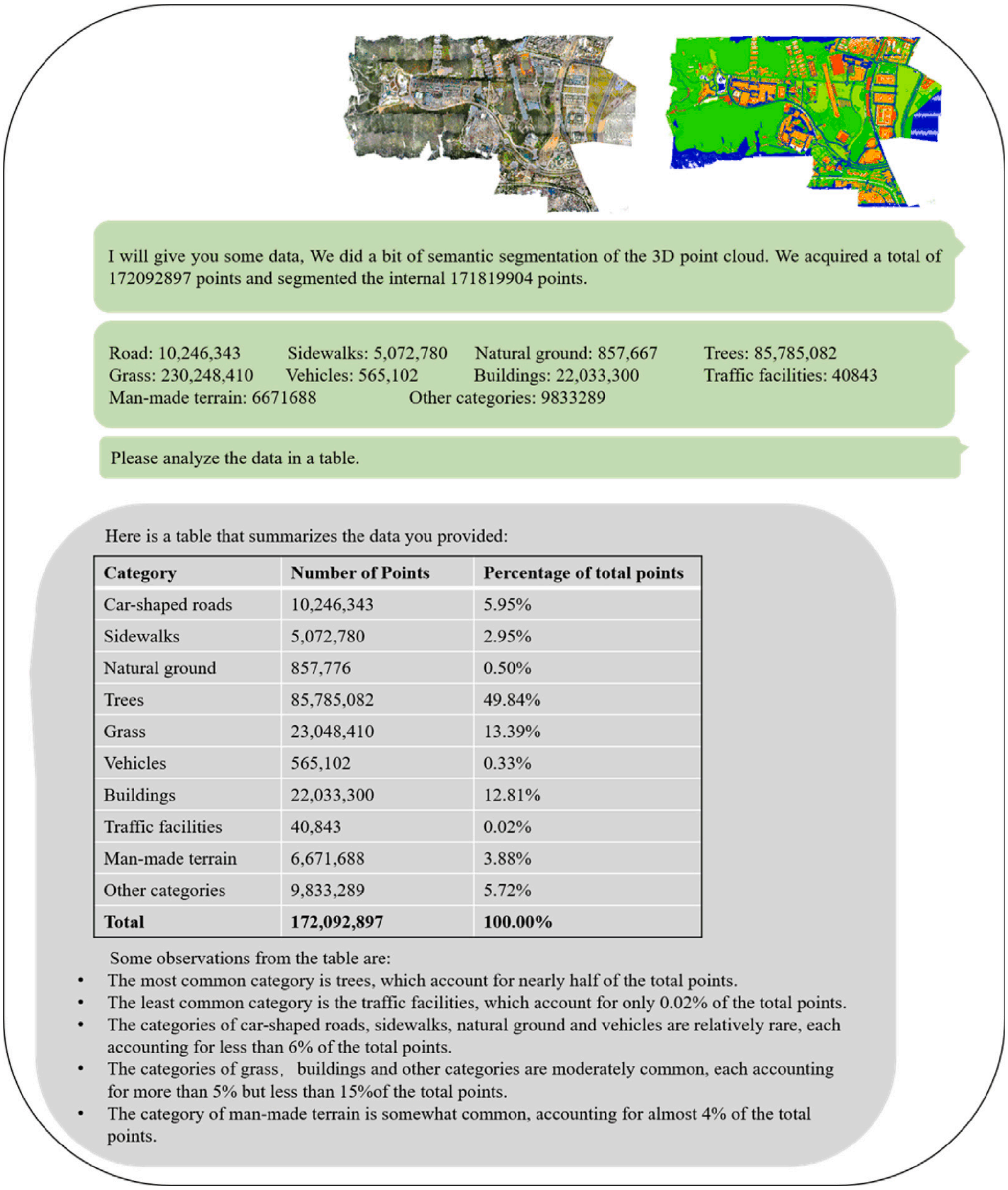

Fig. 2. Analysis of raw data by language models.

clear from Eq. (2) that the local BI, PLI, ERI, and regularization factor are required before the environmental index can be calculated. Various indices were obtained through field measurements in the different seasons in the region from the official environmental website. The generated results are compared as the true value with the model output. With the input data, the language model can accurately calculate the percentage of land features, such as the percentage of vegetation cover and the percentage of water system cover. The knowledge base and retrieval capabilities of the language model are also used to suggest the conditions for their selection in other parameters. At this point, the language model can calculate the corresponding environmental parameters.

We found out from the website where the official environmental index is published that the environmental index for the region is 85.8. As shown in Table 2, different base models provided estimates ranging from 62.7 to 93 based on their knowledge base and input data. The Gemini model achieved a computation result of 82.5 even without extensive prompts, demonstrating its capability for ecological assessment without significant guidance. This result is directly applicable to subsequent ecological evaluations. This result proves the effectiveness of our method.

Table 2
Comparison of environmental index by different foundation models.

Model	Environment index
ChatGPT	62.7
Newbing	93.0
Gemini	**82.5**
Ground truth	85.8

Three different language models are compared at the same time. The free version of ChatGPT has its knowledge base until April 2023. Therefore the calculated indices are relatively accurate before April 2023, and the data after April 2023 are all guesses and do not have reference values. Newbing provides an online GPT with the ability to search and update data, which allows it to access recent data and get more accurate results. The Gemini model is trained on a multi-modal and

Please give a detailed report analysis of the above data.

The data shows the results of a semantic segmentation of a 3D point cloud, which is a collection of points that represent the shape and appearance of a scene or an object. Semantic segmentation is a process of assigning a label or a category to each point in the point cloud, such as car-shaped road, sidewalk, tree, etc. Semantic segmentation can be useful for applications such as autonomous driving, urban planning, environmental monitoring, etc.

The data consists of 172,092,897 points, which is a large amount of data that covers a wide range of categories. The data was obtained from an unknown source and location, so it is not clear what kind of scene or object the point cloud represents. However, based on the categories and their proportions, it is possible to make some assumptions and interpretations.

Based on these assumptions, some interpretations can be made about the data and its implications. For example:

- The data shows that trees are the most dominant category in the point cloud, accounting for nearly half of the total points. This could imply that the scene or object is heavily influenced by nature and has a high degree of greenery and biodiversity. This could also imply that the scene or object is environmentally friendly and has a positive impact on the air quality and climate.
- The data shows that traffic facilities are the least dominant category in the point cloud, accounting for only 0.02% of the total points. This could imply that the scene or object is not very congested or crowded by traffic and has a low degree of noise and pollution. This could also imply that the scene or object is safe and accessible for pedestrians and cyclists.
- The data shows that car-shaped roads, sidewalks, natural ground and vehicles are relatively rare categories in the point cloud, each accounting for less than 6% of the total points. This could imply that the scene or object is not very spacious or expansive and has a limited amount of surface area and mobility. This could also imply that the scene or object is not very diverse or dynamic and has a low degree of variation and activity.
- The data shows that grass, buildings and other categories (including water) are moderately common categories in the point cloud, each accounting for more than 5% but less than 15%of the total points. This could imply that the scene or object is moderately complex or interesting and has a balanced amount of contrast and harmony. This could also imply that the scene or object is moderately attractive or appealing and has a moderate degree of aesthetics and functionality.
- The data shows that man-made terrain is somewhat common category in the point cloud, accounting for nearly 4% of the total points. This could imply that the scene or object is somewhat artificial or modified and has some degree of human intervention and influence. This could also imply that the scene or object is somewhat innovative or creative and has some degree of novelty and originality.

These are some possible ways to analyze the data in more detail. However, these are only based on assumptions and interpretations that may not be accurate or valid. Therefore, it is important to verify the data with other sources and methods before drawing any conclusions or making any decisions based on it.

Fig. 3. Initial report generated by the language model (Prompt(a)).

multi-lingual dataset. Leveraging larger datasets and increased computational resources has endowed it with greater potential in multi-modal understanding.

4.2. Quantitative comparison of generated text

In this section, we quantitatively compare the performance of different language models for report generation and recommendation generation. Three different prompts are entered in the order. A detailed report was requested to be produced, addressing the distribution of features in the study area. The quality of reports generated is compared with different prompts and the quality of reports generated by different models with the same prompts. Because the evaluation criteria for report generation are relatively subjective, we used report length and coverage for comparison instead.

The length of the reports generated at different prompts is compared first. Without any restrictions, other language prompts will generate reports of different lengths. In this section, the output results of

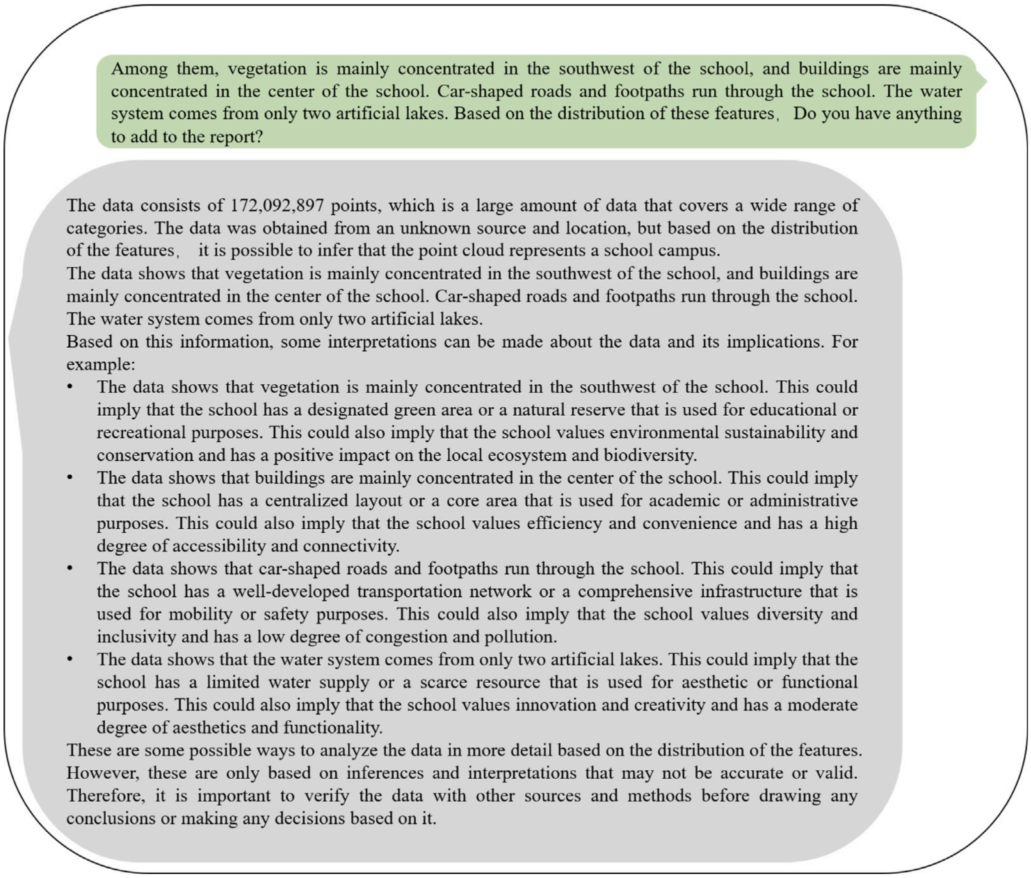

Fig. 4. Additions to the report after adding spatial distribution (Prompt(b)).

the Newbing model are presented. The report generated based on Prompt(a) analyses each analogy, comparing the proportion of them and the role that each category of feature plays in the environment. As shown in Fig. 3, the same question is repeated several times and the language model gave the same analysis results. However, the report lengths generated by different models vary slightly, with ChatGPT approximately 550 words, Newbing around 750 words, and the Gemini model providing only 370 words (see Fig. 6).

Further, the approximate orientation of the above feature category in the scene is provided for the language model, Prompt(b). As shown in Fig. 4, the analysis of not only the percentage of features and the function of the different features in the generated report thus far. A discussion of the reasons for the distribution has also been added. For example, the language model speculates that the distribution of buildings is due to the layout of a core area used for academic or administrative purposes and that the distribution of vegetation and water systems implies the school's focus on environmental sustainability, etc. At this juncture, the report generated by ChatGPT comprises approximately 700 words, Newbing's report spans nearly 1000 words, while the Gemini model yields a report of around 570 words (see Fig. 6).

Finally, other relevant information, combined as Prompt(c) as shown in Fig. 5, was incorporated, including the areas and means of data acquisition inputted into the language model. This was done to incorporate local characteristics and generate detailed reports. At this point the report will be analyzed in more detail, combining the model's knowledge base with the ecology of the campus, the school's motto, and its mission. The result is a complete and comprehensive report. With this additional information, the current report length for ChatGPT is approximately 1050 words, while Newbing's report extends to 1330 words, and the Gemini model's output comprises around 860 words (see Fig. 6). The experiment indicated that after different prompts, the language model produced different reports. And the more detailed the prompt, the more comprehensive the report generated and the more words. At the same time, our method can summarize the patterns in the data well and can use the knowledge base of the language model, and the corresponding prompting works, to get better reports for decision-making.

The more accessible language models ChatGPT, NewBing (GPT-4 Online), and Gemini are used for comparison to explore the data comprehension capabilities of the language models. The same questions are asked for the different language models while entering the same prompt data and text. Prompt(abc) is inputted and generates a report using the language model. The experimental results show that the data can be collated and analyzed in detail in two different large languages. The difference is that the reports generated by ChatGPT are prompted texts for very detailed answers, with extensive analysis for each type of feature. However, ChatGPT uses its own very small knowledge base and has no way of extending it in response to the information given. Conversely, NewBing has similarly generalized and analyzed the input data and used its strong knowledge base to extend the input prompts considerably, yielding reports with greater coverage. In addition to accurately understanding the input data and textual prompts, the Gemini model can remarkably infer the ecological environment index of the region with high precision. Furthermore, it provides more specific and rational ecological improvement measures.

5. Conclusion

In this work, we have designed a complete, city-level 3D point cloud semantic segmentation result using a large language model to understand the results and assist in urban construction and ecological

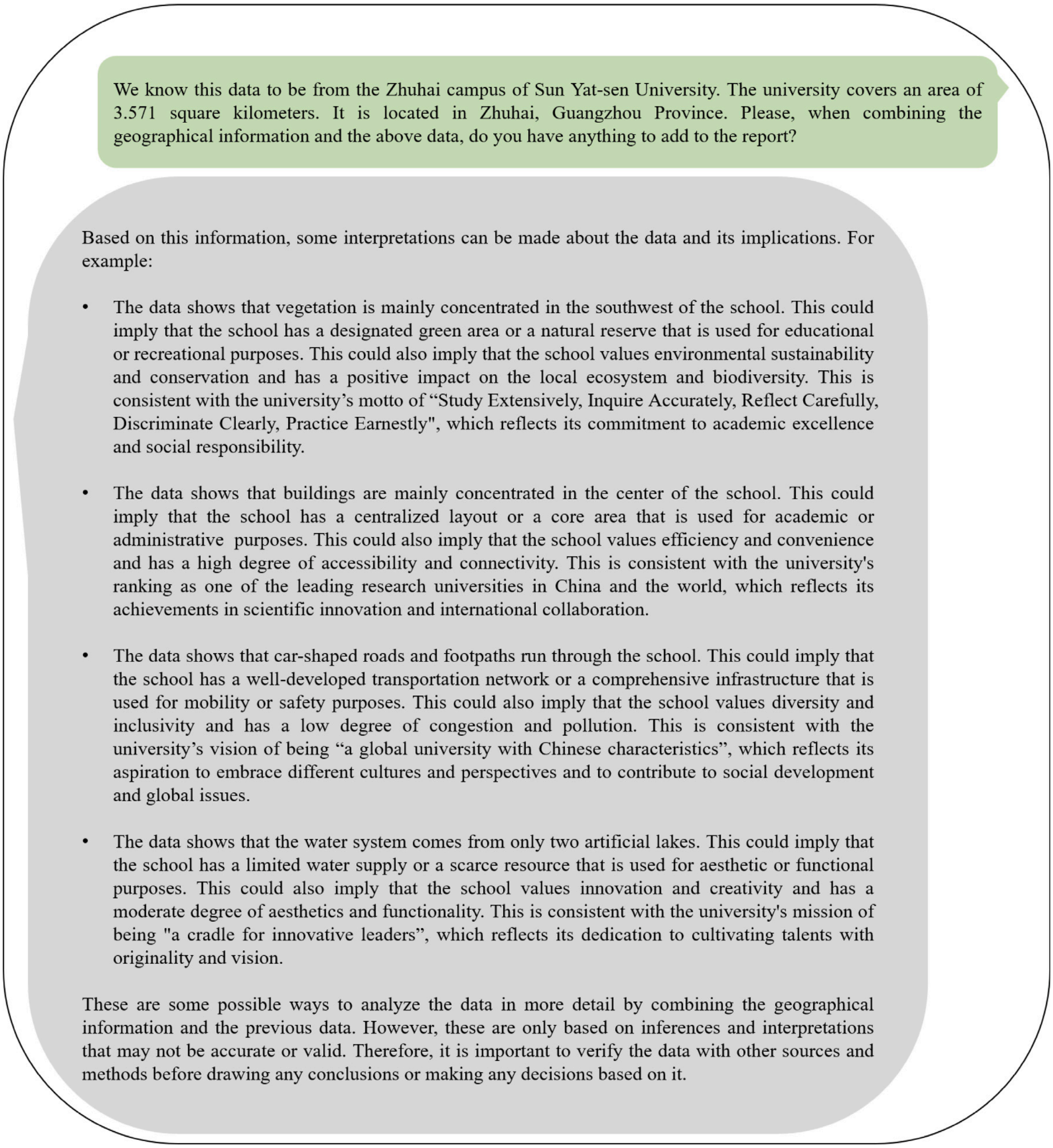

Fig. 5. Additional report after adding geographic information (Prompt(c)).

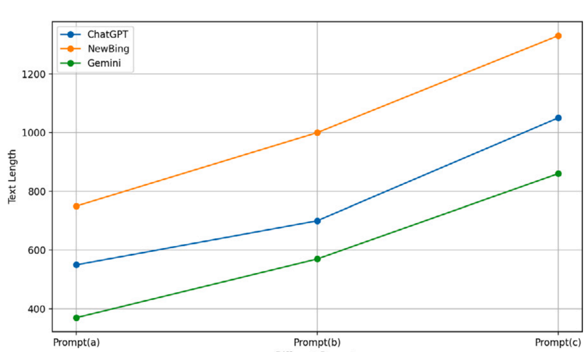

Fig. 6. Comparison of text length by different model.

assessment programs. The language model is used firstly to understand the initial results of semantic segmentation of 3D point clouds and analyze the data. The relevant feature distribution information and geographic location information are then processed into textual prompting signals, which are input to the large language model as prompting works. At this time, based on the data information and spatial geographic information, combined with its own knowledge reserves, the large language model is able to accurately calculate the environmental index of the region, and assess the potential ecological risk of the region, giving a detailed report on urban ecological construction. The scheme breaks the limitation of the application of language models on high-dimensional information and better utilizes the intelligence and chain-of-thought features of language models to obtain quasi-acquisitive computations and detailed reports by prompting engineering. The proposed method constructs an interactive point cloud intelligent analysis system for urban ecological civilization construction decision-making, which provides new ideas for smart city development. In future work, we will continue to explore more effective ways to integrate point clouds and LLMs to achieve applications in a wider range of tasks.

CRediT authorship contribution statement

Yiping Chen: Funding acquisition, Supervision. **Shuai Zhang:** Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ting Han:** Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Yumeng Du:** Validation, Writing – original draft, Writing – review & editing. **Wuming Zhang:** Funding acquisition, Resources, Supervision. **Jonathan Li:** Conceptualization, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (No. 42371343).

Appendix

A.1. Detail report by ChatGPT

Ecological Environment Report

1. Overview:

The point cloud data, divided into nine categories, offers insights into the campus's ecological landscape. These categories and their respective point counts are as follows:

- Roads: 10,246,343 points
- Sidewalks: 5,072,780 points
- Natural land: 857,667 points
- Trees: 85,785,082 points
- Grasslands: 230,248,410 points
- Vehicles: 565,102 points
- Buildings: 22,033,300 points
- Traffic facilities: 40,843 points
- Artificial facilities: 6,671,688 points
- Other categories (including water system): 9,833,289 points

2. Environmental Features:

The campus spans an area of 3.571 square kilometers and is characterized by diverse topography. Roads crisscross the campus, facilitating transportation within and around the university premises. Buildings are predominantly clustered in the central region, housing academic and administrative facilities. The campus's water system comprises two artificial lakes, contributing to its aesthetic appeal and ecological balance.

3. Ecological Significance:

The abundance of trees and grasslands enhances the campus's biodiversity and ecological resilience. The presence of water bodies, including artificial lakes, adds to the campus's environmental diversity and provides habitats for various aquatic species. Green spaces contribute to air quality improvement, carbon sequestration, and the overall well-being of the campus community.

4. Environmental Sustainability:

Efforts to preserve and enhance the campus's ecological integrity are essential for ensuring a sustainable and conducive learning environment. Initiatives such as tree planting, green infrastructure development, and sustainable transportation practices contribute to the campus's environmental sustainability goals. Collaboration with local authorities and community engagement initiatives can further strengthen the university's commitment to environmental stewardship.

5. Conclusion: The ecological environment report underscores the importance of maintaining and enhancing the ecological balance of Sun Yat-sen University's campus. Through sustainable practices and community involvement, the university can continue to serve as a model for environmental stewardship in higher education.

This comprehensive assessment provides valuable insights into the ecological landscape of Sun Yat-sen University, guiding future conservation and sustainability efforts on campus.

A.2. Detail report by Gemini

Ecological Report: Sun Yat-sen University

1. Introduction:

The Zhuhai Campus of Sun Yat-sen University, situated in Guangdong Province, covers an approximate area of 3.571 square kilometers. This report aims to assess the ecological condition of the campus-based on provided data regarding land use and semantic segmentation of the region.

2. Ecological Assessment:

- Vegetation Cover: The campus exhibits rich vegetation cover, with high densities of trees and grasslands. Tree density measures approximately 240,133.98 points per square kilometer, while grass density stands at around 64,498,024.89 points per square kilometer. Such dense vegetation contributes positively to air quality improvement, soil protection, and habitat enhancement.
- Water Bodies: Two artificial lakes are present within the campus, contributing to a water body density of approximately 2,754,361.92 points per square kilometer. Regular monitoring of water quality and pollution control measures are essential for maintaining the health of aquatic ecosystems.
- Building Density: The density of buildings is relatively high, with approximately 6,165,294.34 points per square kilometer. Increased building density may lead to alterations in land cover, ecological fragmentation, and habitat destruction, necessitating careful urban planning and management.
- Other Artificial Facilities: The density of other artificial facilities, including roads and infrastructural elements, is comparatively lower, at around 1,868,997.18 points per square kilometer. While the presence of such facilities may have some impact on the ecological environment, their density remains relatively low.

3. Recommendations:

To safeguard and enhance the ecological quality of the Zhuhai Campus, the following measures are recommended:

Strengthen vegetation protection and greenery initiatives, including the expansion of tree and grassland coverage. Regular monitoring of water bodies and effective management strategies for maintaining water cleanliness. Rational urban planning to minimize the ecological impact of building development. Promote environmental awareness and practices such as energy conservation, emissions reduction, and green transportation methods to mitigate human-induced ecological disturbances.

4. Conclusion:

In conclusion, the Zhuhai Campus of Sun Yat-sen University exhibits a generally favorable ecological condition, characterized by rich vegetation cover and water resources. However, measures should be taken to address concerns regarding building density and ensure sustainable environmental management for a healthier and more balanced campus ecosystem.

A.3. Environment index calculation by Gemini

Summary of the Calculation Process
1. Indicator Selection

- Green coverage rate
- Per capita green area
- Building density
- Road density
- Air quality
- Water quality

2. Weight Determination
Assign weights to each indicator based on its impact on environmental quality.
3. Data Acquisition

- Green coverage rate, per capita green area, building density, and road density data come from point cloud data analysis.
- Air quality data comes from Zhuhai Environmental Monitoring Station.
- Water quality data comes from Zhuhai Water Affairs Bureau.

4. Calculation Formula

$$EI = \frac{P_1 * I_1 + P_2 * I_2 + \cdots + P_n * I_n}{P_1 + P_2 + \cdots + P_n} \tag{3}$$

5. Calculation Result

$$EI = 82.5 \tag{4}$$

6. Conclusion

$$EI \geq 80 \tag{5}$$

Excellent level. The ecological environment quality of the campus is good.
Data Sources

- Point cloud data analysis
- Zhuhai Environmental Monitoring Station
- Zhuhai Water Affairs Bureau

References

Biswas, S.S., 2023. Potential use of chat gpt in global warming. Ann. Biomed. Eng. 51 (6), 1126–1127.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.
Chen, Z., Gholami, A., Nießner, M., Chang, A.X., 2021. Scan2cap: Context-aware dense captioning in rgb-d scans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3193–3203.
Chen, M., Radford, A., Child, R., Wu, J., Jun, H., Luan, D., Sutskever, I., 2020. Generative pretraining from pixels. In: International Conference on Machine Learning. PMLR, pp. 1691–1703.
Chen, G., Wang, M., Yang, Y., Yu, K., Yuan, L., Yue, Y., 2023. PointGPT: Auto-regressively generative pre-training from point clouds. arXiv preprint arXiv:2305.11487.
Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S., et al., 2022. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311.
Das, M., Das, A., Pereira, P., 2023. Developing an integrated urban ecological efficiency framework for spatial ecological planning: A case on a tropical mega metropolitan area of the global south. Geosci. Front. 14 (1), 101489.
De Keersmaecker, W., Lhermitte, S., Honnay, O., Farifteh, J., Somers, B., Coppin, P., 2014. How to measure ecosystem stability? An evaluation of the reliability of stability metrics based on remote sensing time series across the major global ecosystems. Glob. Change Biol. 20 (7), 2149–2161.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Fijačko, N., Gosak, L., Štiglic, G., Picard, C.T., Douma, M.J., 2023. Can ChatGPT pass the life support exams without entering the American heart association course? Resuscitation 185.
Fu, R., Liu, J., Chen, X., Nie, Y., Xiong, W., 2024. Scene-LLM: Extending language model for 3D visual understanding and reasoning. arXiv preprint arXiv:2403.11401.
Hong, Y., Zhen, H., Chen, P., Zheng, S., Du, Y., Chen, Z., Gan, C., 2024. 3D-llm: Injecting the 3d world into large language models. Adv. Neural Inf. Process. Syst. 36.
Hu, X., Xu, H., 2018. A new remote sensing index for assessing the spatial heterogeneity in urban ecological quality: A case from Fuzhou City, China. Ecol. Indic. 89, 11–21.
Jalil, S., Rafi, S., LaToza, T.D., Moran, K., Lam, W., 2023. Chatgpt and software testing education: Promises & perils. In: 2023 IEEE International Conference on Software Testing, Verification and Validation Workshops. ICSTW, IEEE, pp. 4130–4137.
Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D., 2020. Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.
Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., et al., 2023. Segment anything. arXiv preprint arXiv:2304.02643.
Kung, T.H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., et al., 2023. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLoS Digit. Health 2 (2), e0000198.
Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R., 2019. Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
Larsson, M., Hanberger, A., 2016. Evaluation in management by objectives: A critical analysis of Sweden's national environmental quality objectives system. Evaluation 22 (2), 190–208.
Li, Y., Cao, Z., Long, H., Liu, Y., Li, W., 2017. Dynamic analysis of ecological environment combined with land cover and NDVI changes and implications for sustainable urban–rural development: The case of Mu Us Sandy Land, China. J. Clean. Prod. 142, 697–715.
Li, J., Li, D., Xiong, C., Hoi, S., 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In: International Conference on Machine Learning. PMLR, pp. 12888–12900.
Li, Z., Zhang, C., Wang, X., Ren, R., Xu, Y., Ma, R., Liu, X., 2024. 3DMIT: 3D multi-modal instruction tuning for scene understanding. arXiv preprint arXiv:2401.03201.
Liu, D., Huang, X., Hou, Y., Wang, Z., Yin, Z., Gong, Y., Gao, P., Ouyang, W., 2024. Uni3D-LLM: Unifying point cloud perception, generation and editing with large language models. arXiv preprint arXiv:2402.03327.
Liu, H., Li, C., Wu, Q., Lee, Y.J., 2023a. Visual instruction tuning. arXiv preprint arXiv:2304.08485.
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019a. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
Liu, Z., Tang, H., Lin, Y., Han, S., 2019b. Point-voxel cnn for efficient 3d deep learning. Adv. Neural Inf. Process. Syst. 32.
Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G., 2023b. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Comput. Surv. 55 (9), 1–35.
Luo, T., Rockwell, C., Lee, H., Johnson, J., 2024. Scalable 3d captioning with pretrained models. Adv. Neural Inf. Process. Syst. 36.
Mishra, S., Khashabi, D., Baral, C., Hajishirzi, H., 2021. Natural instructions: Benchmarking generalization to new tasks from natural language instructions. pp. 839–849, arXiv preprint arXiv:2104.08773.
OpenAI, 2023a. Chatgpt: Optimizing language models for dialogue.
OpenAI, 2023b. Gpt-4 technical report.
Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al., 2022. Training language models to follow instructions with human feedback. Adv. Neural Inf. Process. Syst. 35, 27730–27744.
Patel, S.B., Lam, K., 2023. ChatGPT: the future of discharge summaries? Lancet Digit. Health 5 (3), e107–e108.
Pi, R., Gao, J., Diao, S., Pan, R., Dong, H., Zhang, J., Yao, L., Han, J., Xu, H., Zhang, L.K.T., 2023. DetGPT: Detect what you need via reasoning. arXiv preprint arXiv:2305.14167.
Popp, J.H., Hyatt, D.E., Hoag, D., 2000. Modeling environmental condition with indices: a case study of sustainability and soil resources. Ecol. Model. 130 (1–3), 131–143.
Prieto, S.A., Mengiste, E.T., García de Soto, B., 2023. Investigating the use of ChatGPT for the scheduling of construction projects. Buildings 13 (4), 857.
Qi, C.R., Su, H., Mo, K., 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. pp. 652–660.
Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Adv. Neural Inf. Process. Syst. 30.
Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning. PMLR, pp. 8748–8763.
Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al., 2018. Improving language understanding by generative pre-training.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al., 2019. Language models are unsupervised multitask learners. OpenAI Blog 1 (8), 9.

Sengupta, A., Cao, S., 2022. Mmpose-nlp: A natural language processing approach to precise skeletal pose estimation using mmwave radars. IEEE Trans. Neural Netw. Learn. Syst..

Shan, W., Jin, X., Ren, J., Wang, Y., Xu, Z., Fan, Y., Gu, Z., Hong, C., Lin, J., Zhou, Y., 2019. Ecological environment quality assessment based on remote sensing data for land consolidation. J. Clean. Prod. 239, 118126.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., et al., 2022. Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138.

Sinha, R.K., Roy, A.D., Kumar, N., Mondal, H., Sinha, R., 2023. Applicability of ChatGPT in assisting to solve higher order problems in pathology. Cureus 15 (2).

Su, Y., Lan, T., Li, H., Xu, J., Wang, Y., Cai, D., 2023. Pandagpt: One model to instruction-follow them all. arXiv preprint arXiv:2305.16355.

Sun, Q., Li, Y., Liu, Z., Huang, X., Liu, F., Liu, X., Ouyang, W., Shao, J., 2023. UniG3D: A unified 3D object generation dataset. arXiv preprint arXiv:2306.10730.

Sun, D., Zhang, J.-x., Zhu, C., Hu, Y., Zhou, L., 2012. An assessment of China's ecological environment quality change and its spatial variation. Acta Geogr. Sin. 67 (12), 1599–1610.

Tafferner, Z., Illés, B., Krammer, O., Géczy, A., 2023. Can ChatGPT help in electronics research and development? A case study with applied sensors. Sensors 23 (10), 4879.

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al., 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Venigalla, A., Frankle, J., Carbin, M., 2022. PubMed GPT: A domain-specific large language model for biomedical text.

Victor, S., Albert, W., Colin, R., Stephen, B., Lintang, S., Zaid, A., Antoine, C., Arnaud, S., Arun, R., Manan, D., et al., 2022. Multitask prompted training enables zero-shot task generalization. In: International Conference on Learning Representations.

Wang, S., Zhao, Z., Ouyang, X., Wang, Q., Shen, D., 2023. Chatcad: Interactive computer-aided diagnosis on medical image using large language models. arXiv preprint arXiv:2302.07257.

Wang, J., Zhong, L., 2017. Problems and suggestion for developing ecological construction in land management work. Trans. Chin. Soc. Agric. Eng. 33 (5), 308–314.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., et al., 2022. Emergent abilities of large language models. arXiv preprint arXiv:2206.07682.

Xie, J., Ye, K., Li, Y., Li, Y., Lin, K.Q., Zheng, Y., Shen, L., Shou, M.Z., 2023. VisorGPT: Learning visual prior via generative pre-training. arXiv preprint arXiv:2305.13777.

Xu, D., Cheng, J., Xu, S., Geng, J., Yang, F., Fang, H., Xu, J., Wang, S., Wang, Y., Huang, J., et al., 2022. Understanding the relationship between China's eco-environmental quality and urbanization using multisource remote sensing data. Remote Sens. 14 (1), 198.

Xu, R., Wang, X., Wang, T., Chen, Y., Pang, J., Lin, D., 2023. Pointllm: Empowering large language models to understand point clouds. arXiv preprint arXiv:2308.16911.

Xu, D., Yang, F., Yu, L., Zhou, Y., Li, H., Ma, J., Huang, J., Wei, J., Xu, Y., Zhang, C., et al., 2021. Quantization of the coupling mechanism between eco-environmental quality and urbanization from multisource remote sensing data. J. Clean. Prod. 321, 128948.

Xu, H., et al., 2013. A remote sensing index for assessment of regional ecological changes. China Environ. Sci. 33 (5), 889–897.

Yang, S., Liu, J., Zhang, R., Pan, M., Guo, Z., Li, X., Chen, Z., Gao, P., Guo, Y., Zhang, S., 2023. Lidar-llm: Exploring the potential of large language models for 3d lidar understanding. arXiv preprint arXiv:2312.14074.

Yu, J., Li, X., Guan, X., Shen, H., 2022. A remote sensing assessment index for urban ecological livability and its application. Geo-Spat. Inf. Sci. 1–22.

Zeng, Y., Jiang, C., Mao, J., Han, J., Ye, C., Huang, Q., Yeung, D.-Y., Yang, Z., Liang, X., Xu, H., 2023. CLIP2: Contrastive language-image-point pretraining from real-world point cloud data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15244–15253.

Zhang, R., Guo, Z., Zhang, W., Li, K., Miao, X., Cui, B., Qiao, Y., Gao, P., Li, H., 2022b. Pointclip: Point cloud understanding by clip. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8552–8562.

Zhang, S., Huang, W., Chen, Y., Zhang, W., Li, J., 2023. Spdc: a super-point and point combining based dual-scale contrastive learning network for point cloud semantic segmentation. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 48, 571–578.

Zhang, M., Kafy, A.-A., Ren, B., Zhang, Y., Tan, S., Li, J., 2022a. Application of the optimal parameter geographic detector model in the identification of influencing factors of ecological quality in guangzhou, China. Land 11 (8), 1303.

Zhao, H., Jiang, L., Jia, J., Torr, P.H., Koltun, V., 2021. Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16259–16268.

Zhou, J., Wang, J., Ma, B., Liu, Y.-S., Huang, T., Wang, X., 2023. Uni3d: Exploring unified 3d representation at scale. arXiv preprint arXiv:2310.06773.

Zhu, D., Chen, J., Shen, X., Li, X., Elhoseiny, M., 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models. arXiv preprint arXiv:2304.10592.