Chair of Computational Modeling and Simulation
TUM School of Engineering and Design
Technical University of Munich

Modeling:
Mathematics:
Programming:
Science:

Software Lab:

Using Multi-Modal Large Language Models for BIM/Bridges Point Cloud Data

Description:

This project aims to develop a point cloud-aware multi-modal language model (MLLM) by building a specialized point cloud encoder trained on BIM/bridge point cloud datasets.

MLLMs extend the capabilities of Large Language Models (LLMs) to process and understand non-textual modalities, including images, point clouds, and technical drawings. Recent works, such as CAD-MLLM [1] have successfully utilized modality-specific encoders to integrate diverse data types into a shared latent representation space, enabling advanced cross-modal reasoning and generation tasks.

In this project, the primary objective is to build a custom point cloud encoder tailored for BIM and bridge point clouds. The encoder will extract meaningful geometric and semantic features from the point cloud data, which will be integrated with a pre-trained language model backbone. The resulting pipeline will fine-tune the LLM to generate textual descriptions of point clouds, enabling tasks such as semantic classification, object detection, and segmentation. This approach builds upon the methodologies of works like PointLLM [2] but focuses specifically on bridge and BIM point clouds, addressing domain-specific challenges and constraints. The outputs of the model will be evaluated using standard benchmarks for classification and segmentation.

Tasks:

In this project you will have to perform the following tasks:

- Develop a pipeline for handling and pre-processing bridge/BIM specific point cloud data.
- Conduct an in-depth review of existing research on multi-modal language models and their application to point cloud data, identifying relevant architectures and methods.
- Experiment with state-of-the-art architectures tailored for bridge point cloud data. This includes designing, fine-tuning, and implementing a custom point cloud encoder. Writing efficient, scalable code for this task will be critical.
- Finally develop a pipeline to fine-tune the language model and evaluate the results based on different classification and segmentation benchmarks.

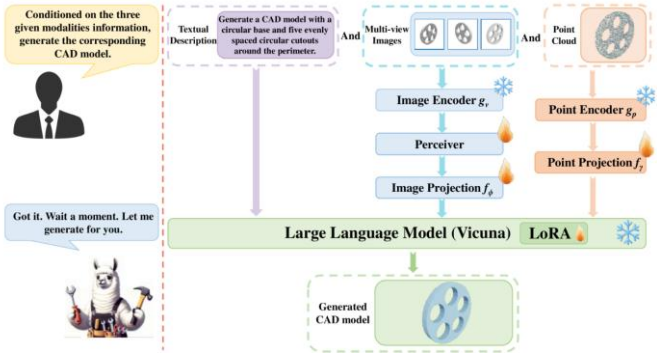

Figure 1: CAD-MLLM architecture [1]

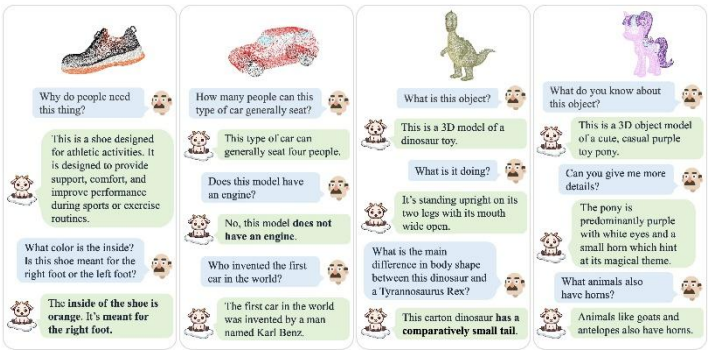

Figure 2: PointLLM [2]

Supervisors:

Syed Salman Ahmed (salman.ahmed@tum.de), Mansour Mehranfar (mansour.mehranfar@tum.de) | CMS | TUM

References:

[1] Xu, Jingwei, et al. "CAD-MLLM: Unifying Multimodality-Conditioned CAD Generation With MLLM." *arXiv preprint arXiv:2411.04954* (2024).

[2] Xu, Runsen, et al. "Pointllm: Empowering large language models to understand point clouds." *European Conference on Computer Vision*. Springer, Cham, 2025.