Semantic Query for 3D Fine-grained Object Segmentation

Codey Sun, Yin-Li Liu
Electrical Engineering, Stanford University

Motivation

- This project aims to improve open-vocabulary 3D segmentation on the part level, allowing users to perform precise fine-grain segmentations from language prompts.
- The inverse problem we are trying to solve is:

image $y = Ax$ semantic of scene
???

- Current method: model A with per-image **CLIP**
- Our method: model A with a segmenter + **CoT VLM**

Related Work

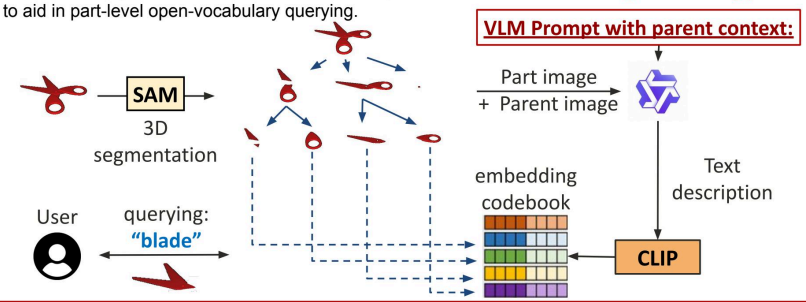
- LERF [1] embeds semantic CLIP features into a radiance field, but struggles with fine-level segmentation and querying.
- Ultrametric feature fields [2] provides hierarchical 3D segmentations, but does not support open-vocabulary querying.

References

[1] Kerr, Justin, et al. "Lerf: Language embedded radiance fields." ICCV 2023.
[2] He, Haodi, et al. "View-Consistent Hierarchical 3D Segmentation Using Ultrametric Feature Fields." ECCV 2024.

New Technique

Our method exploits the hierarchical structure of parts to embed **chain-of-thought reasoning priors** to aid in part-level open-vocabulary querying.



Experimental Results