OmniPart++: Improving 3D Part-Level Reconstruction by Leveraging VLM-Enhanced Segmentation Masks

Final-terms Presentation

2025.12.01

Minseo Park, Jewoo Shin, Sangmin Lee

Team 2

Table of Contents

Table of Contents

Expectations

• Final-term project presentation

• Cover all the materials that you talked for your mid-term project

• Present your ideas that can address problems of those state-of-the-art techniques

• Give your qualitatively (or intuitive) reasons how your ideas address them

• Also, explain expected benefits and drawbacks of your approach

• (Optional) backup your claims with quantitative results collected by some implementations

• Explain roles of each members

—

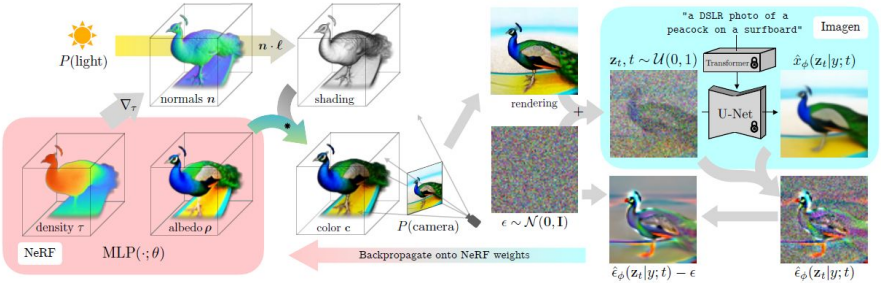
Team Roles

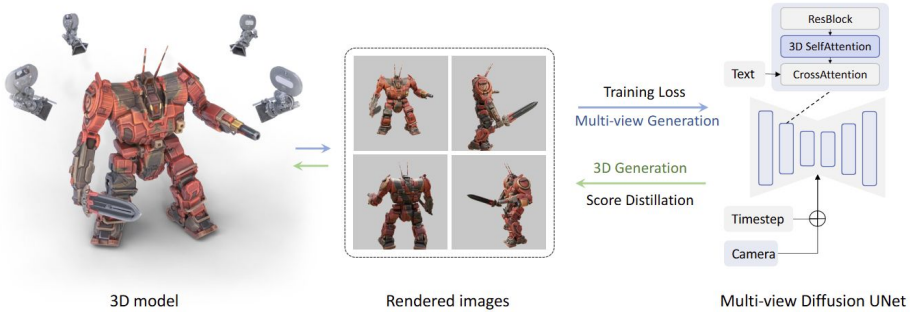
- **Minseo**
  - Test various segmentation
  - Preprocess pseudo segmentation mask to formatted segmentation mask
  - Apply segmentation to OmniPart
- **Sangmin**
  - Test various segmentation
  - Extract segmentation mask from VLM
  - Implement evaluation code
- **Jewoo**
  - Test various segmentation
  - Preprocess pseudo segmentation mask to formatted segmentation mask
  - Apply segmentation to OmniPart

Equal Contribution

Introduction

3D generative models



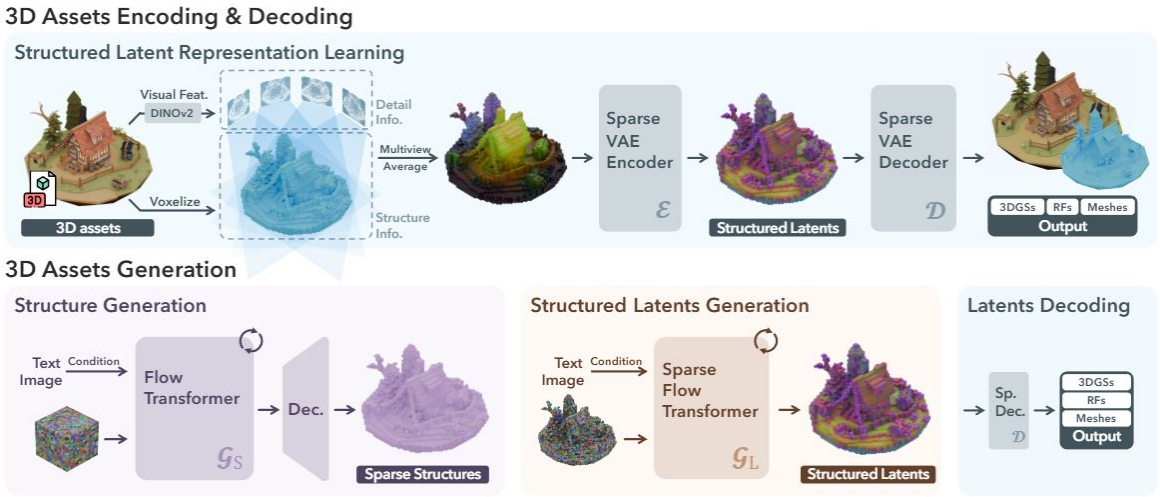


Single-view Image Generation Model based Distillation (DreamFusion)

Multi-view Image Generation Model based Distillation (MVDream)

Source: B. Poole et al., “DreamFusion: Text-to-3D using 2D Diffusion”, ICLR 2023.

3D generative models

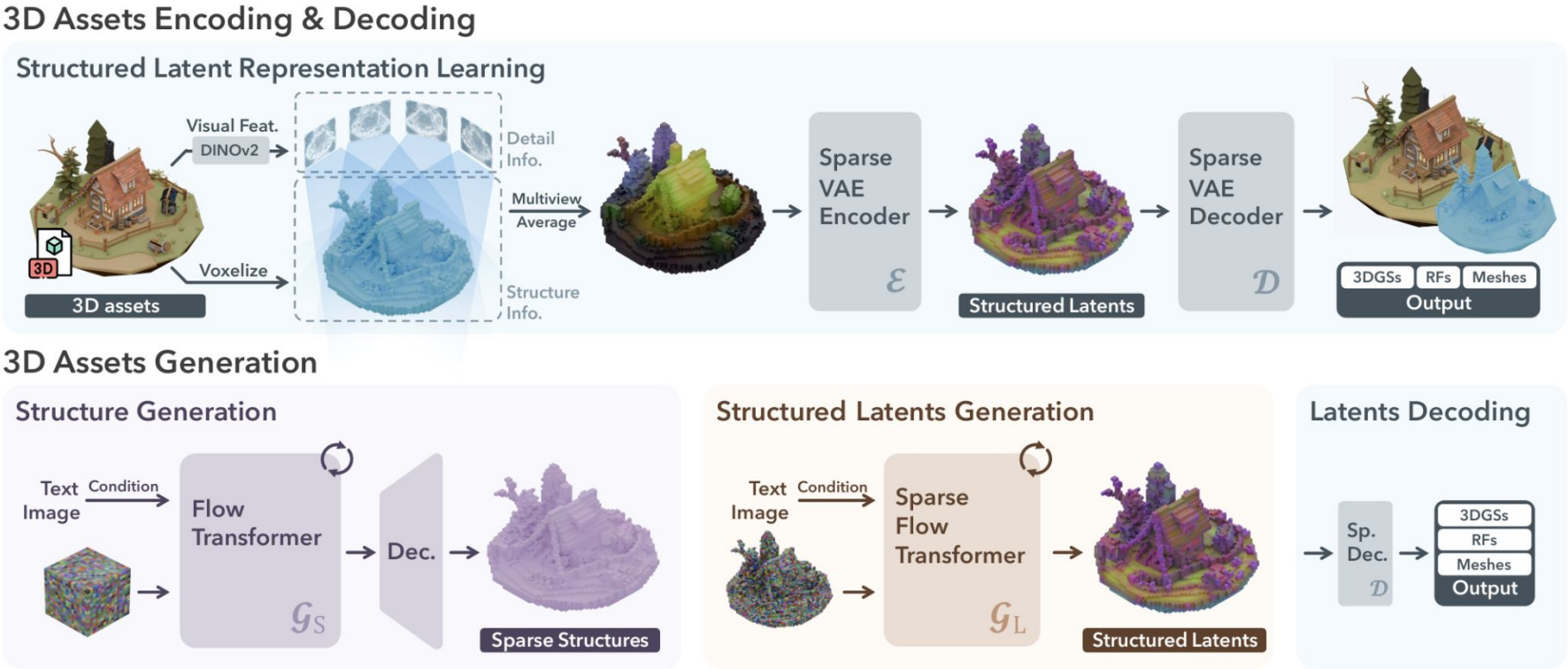


Feed-Forward Models (Trellis)

Source: Xiang et al., “Structured 3D Latents for Scalable and Versatile 3D Generation”, CVPR 2025.

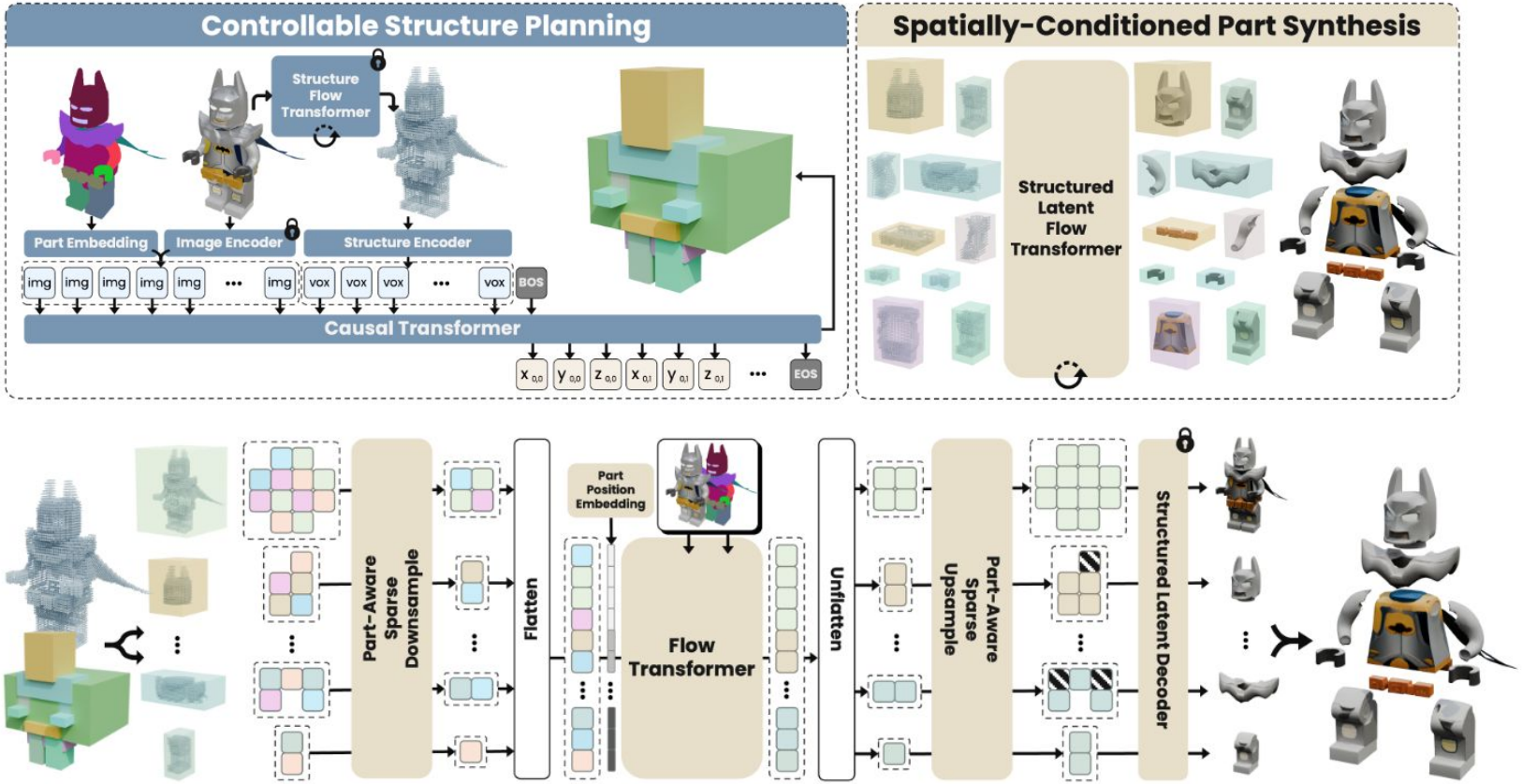
Related Works

Trellis: Structured 3D Latents for Scalable and Versatile 3D Generation



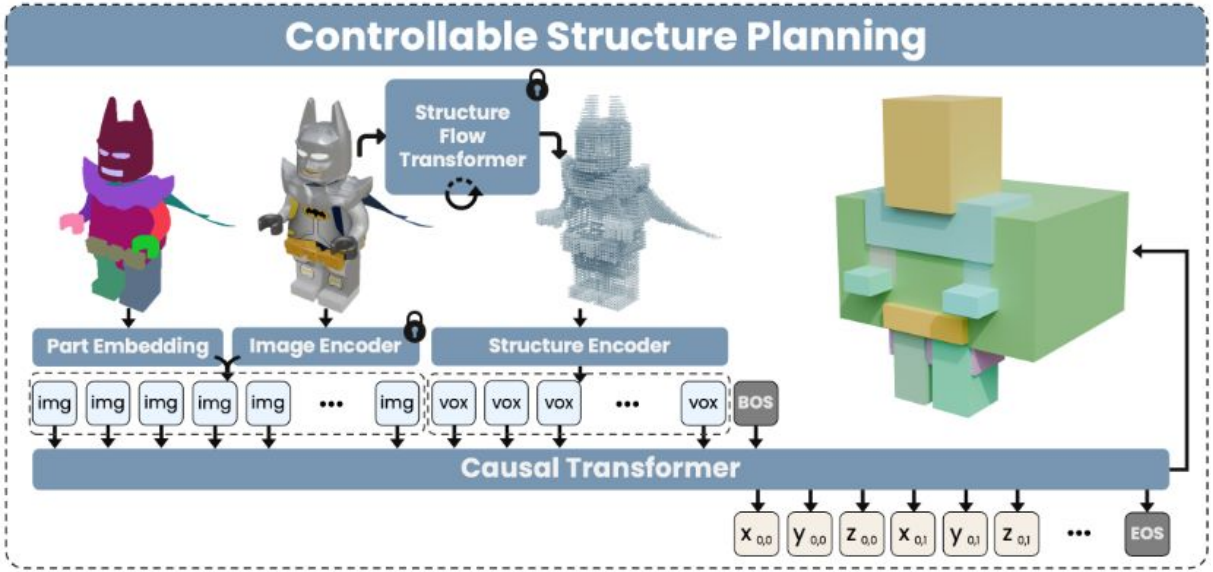
Source: Xiang et al., "Structured 3D Latents for Scalable and Versatile 3D Generation", CVPR 2025.

OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion



Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

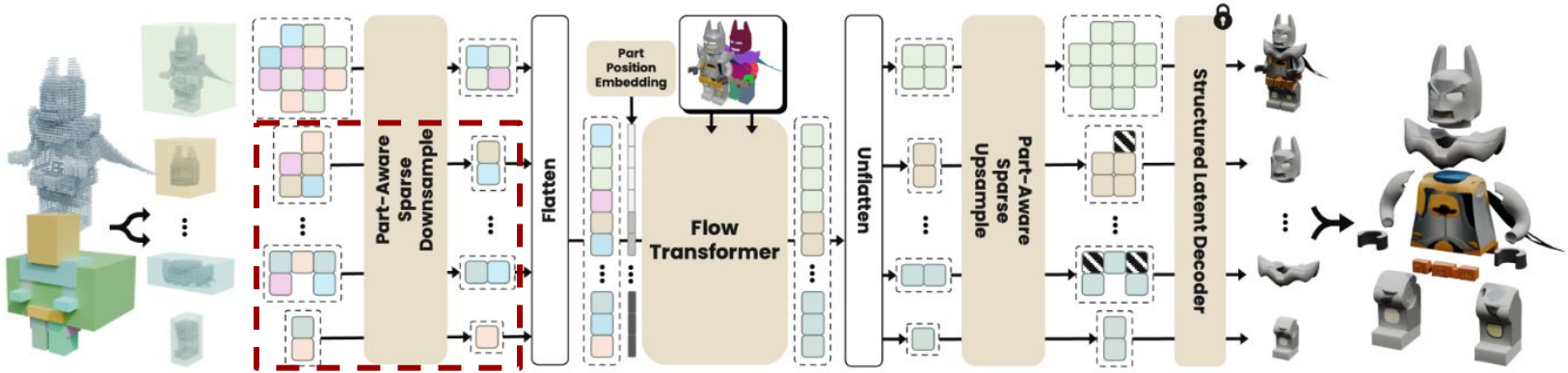
OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion



Part-wise embeddings from 2D segmentation, image embeddings, and structural information used for generating bounding boxes.

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

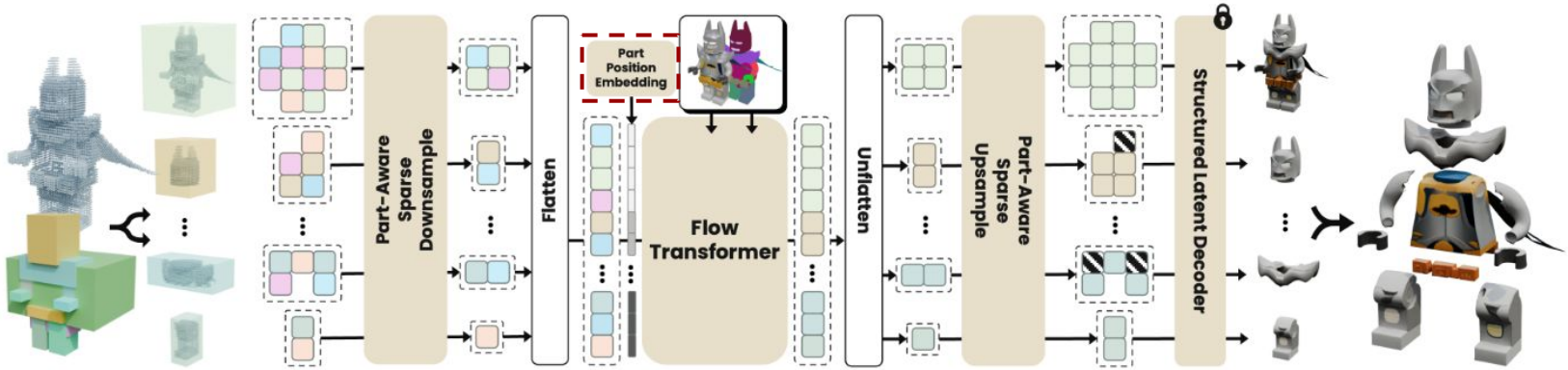
OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion



Decomposing the shape into part-wise sparse latents through a part-aware downsampling process

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

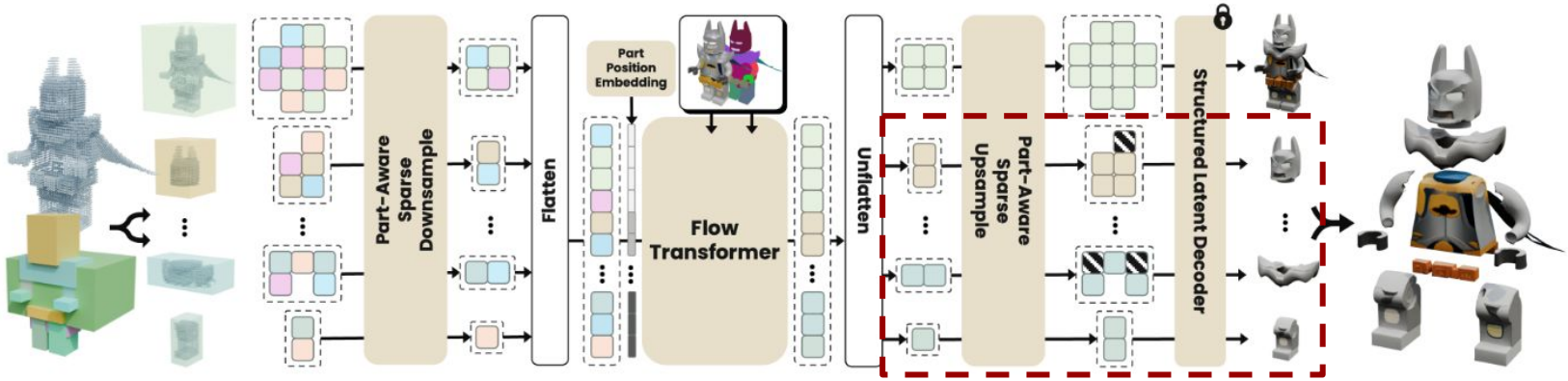
OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion



Part-level positional features (location, size, and ordering) are injected into the latent sequence to inform the Flow Transformer of each part's spatial context

Source: Yang et al., "OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion", SIGGRAPH 2025.

OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion

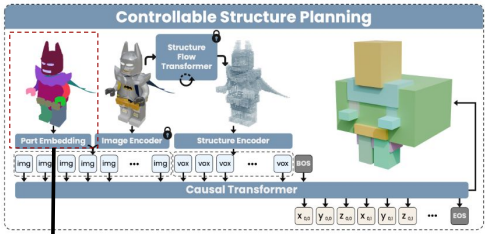


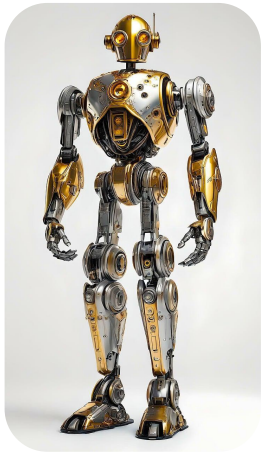
Each part latent is independently upsampled and decoded into its own 3D geometry, enabling part-wise generation rather than a single output (**Trellis**)

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Problem

Problem: Failure of Existing SAM Segmentation for Part Generation

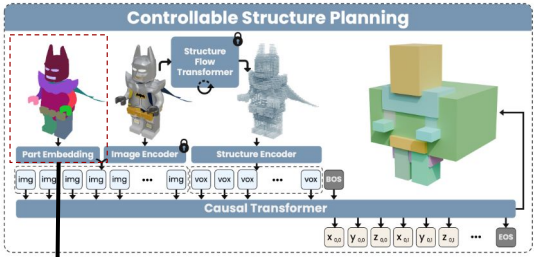


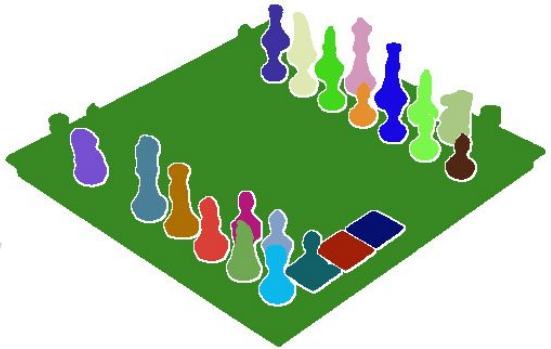


Decent at Single-Object Segmentation

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Problem: Failure of Existing SAM Segmentation for Part Generation





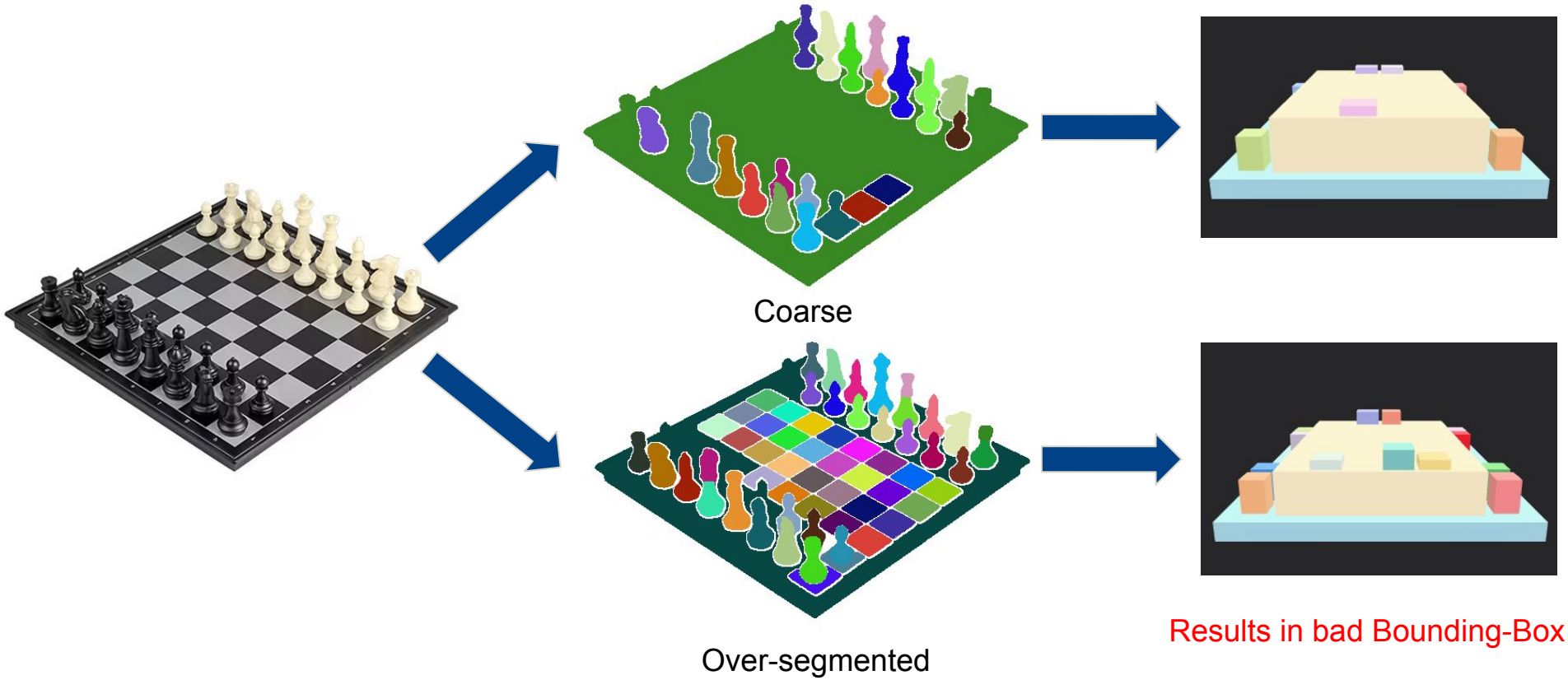
Coarse

Poor at Multi-Object(Part-level) Segmentation

Over-segmented

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Problem: Failure of Existing SAM Segmentation for Part Generation



Results in bad Bounding-Box

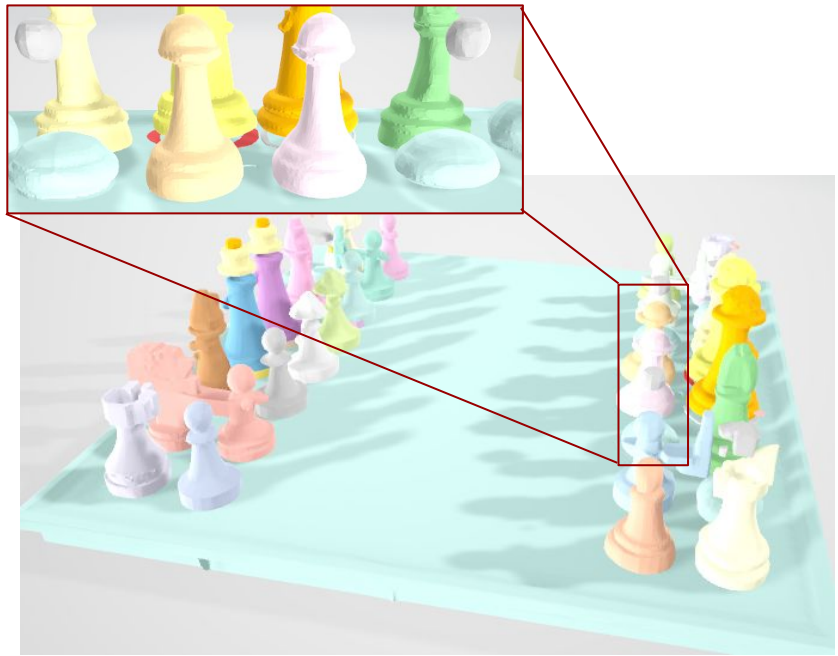
Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Problem: Failure of Existing SAM Segmentation for Part Generation

- SAM roughly separates board vs. pieces
  - Adjacent pieces are fused into large blobs.
  - Portions of the board and pieces are incorrectly grouped.
- With finer granularity:
  - Shadows, board texture, and tiny color shifts produce numerous meaningless masks.
- Segmentation fails to capture semantic instance boundaries.
- SAM is strong at segmenting single objects,
  - but weak at segmenting many repeated objects in cluttered scenes.
- SAM cannot recover individual chess pieces cleanly.
  - preventing correct instance-level, part-aware structure generation.

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Problem: Failure of Existing SAM Segmentation for Part Generation



Geometry reconstruction fails due to incorrect masks.

Source: Yang et al., “OmniPart: Part-Aware 3D Generation with Semantic Decoupling and Structural Cohesion”, SIGGRAPH 2025.

Methodology

Idea

- Human knows, if the image of chessboard is given, everyone could distinguish the image into board and pieces. But **self-supervised segmentation models** and **segmentation models without text grounding** like DINO, SAM doesn't consider these common knowledge.

- However, Visual language models like Nano Banana would consider desired behavior from contextual information and able to manipulate given image with the knowledge.

-> We leverage VLM to enhance segmentation map and finally enhance the 3D part-level generation

Pipeline

- **Nano banana pro pseudo segmentation -> Remove watermark -> Kmeans/AgglomerativeClustering -> Inference mask for OmniPart**

Source: Xiang et al., “Structured 3D Latents for Scalable and Versatile 3D Generation”, CVPR 2025.

Model Pipeline

생각하는 과정 표시 (Nano Banana Pro)

Pseudo segmentation Mask

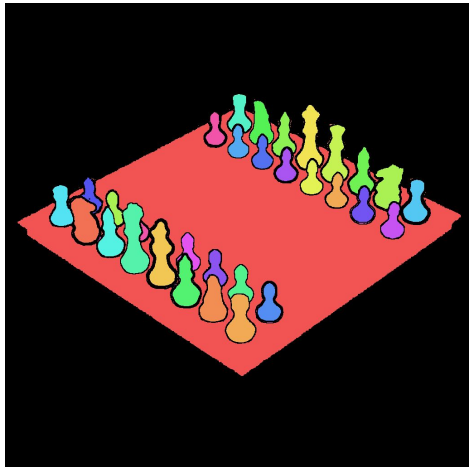
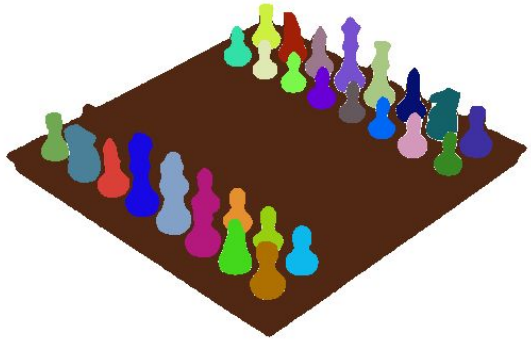


Image Processing



Segmentation mask

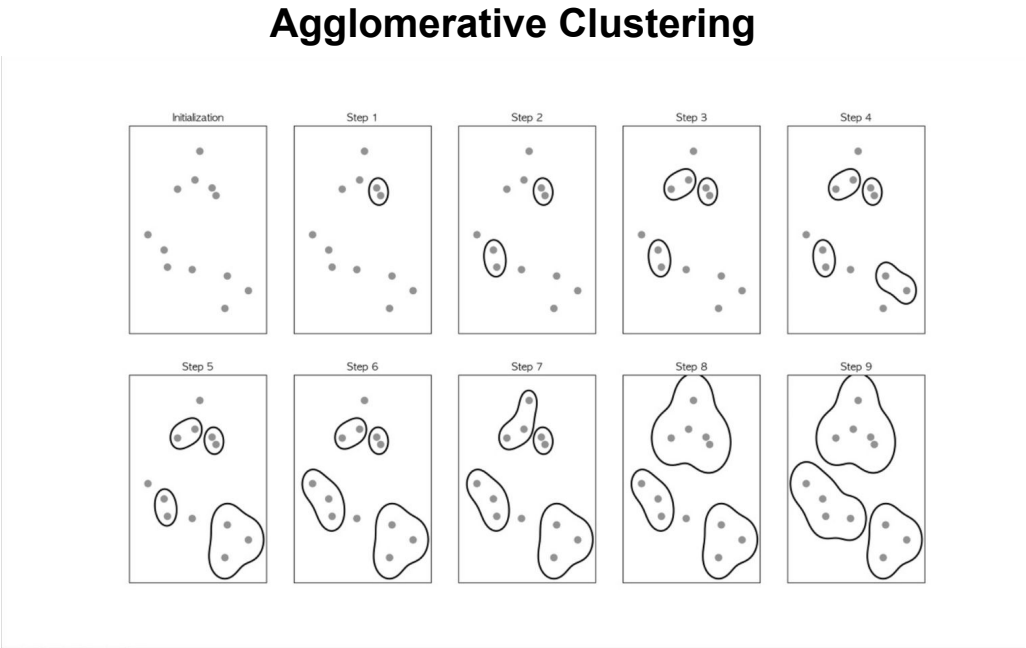
Pseudo Segmentation Mask Details

- prompt: “**You are a human level or even better segmentation model. Give us a segmentation mask based on general knowledge(ex. book is composed of cover and papers, then even though image doesn't show the full book and show only little bit of papers you should still predict unsure parts of the book image is paper and give us the correct mask). You should give part level segmentation mask, not only the whole object but with the part-level decomposition knowledge. Give different color mask for every different part objects. But give the same black segmentation for the background.**”

You are a human level or even better segmentation model. Give us a segmentation mask based on general knowledge(ex. book is composed of cover and papers, then even though image doesn't show the full book and show only little bit of papers you should still predict...

생각하는 과정 표시 (Nano Banana Pro)

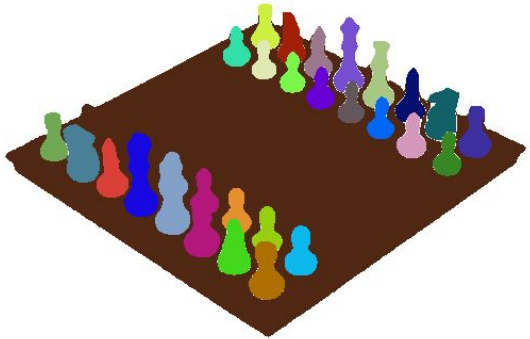
Image Processing

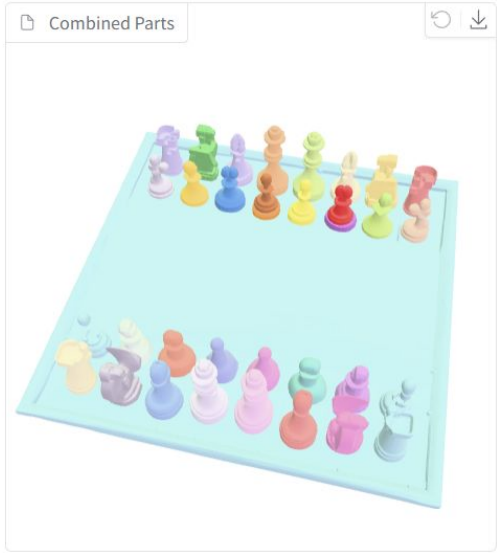


Extracting segmentation masks from VLM-generated maps

Image Source:
https://creatorjo.tistory.com/entry/%ED%8C%8C%EC%9D%B4%EC%8D%AC%EC%9C%BC%EB%A1%9C-%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-%EB%B0%B0%EC%9A%B0%EA%B8%B0-%EB%B9%84%EC%A7%80%EB%8F%84%ED%95%99%EC%8A%B5AgglomerativeClustering

Model Pipeline

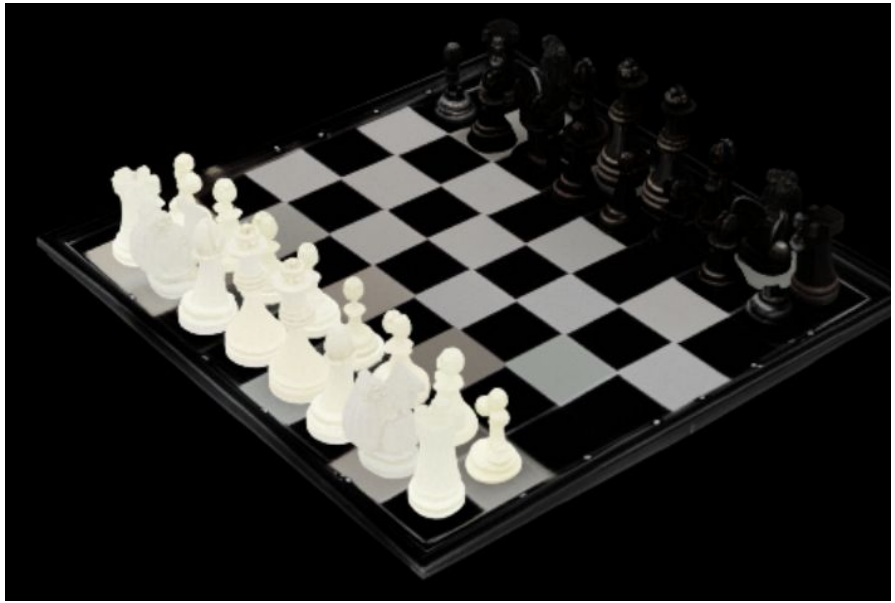




Segmentation mask

Bounding Box generation / Part-level geometry reconstruction

Model Pipeline



Decoded Structured Latents(3D Gaussians)

Experiment Results

3D Reconstruction Results



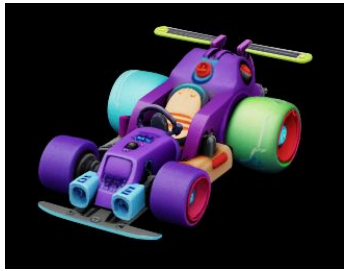
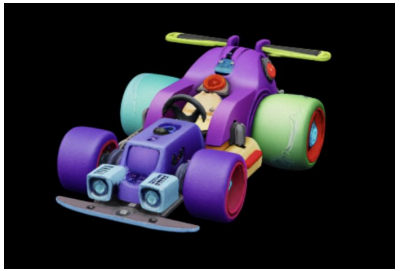
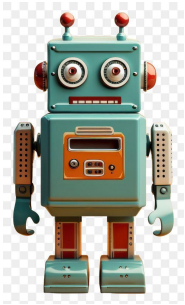





Image Prompts | Trellis | OmniPart | OmniPart++

3D Reconstruction Results



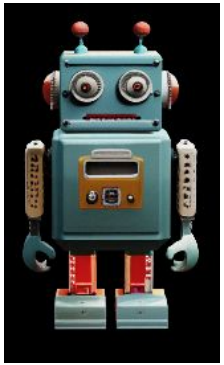
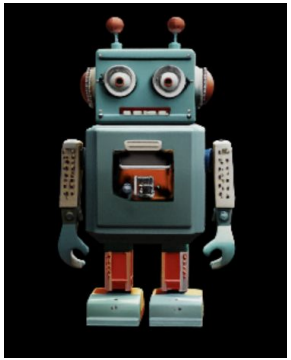
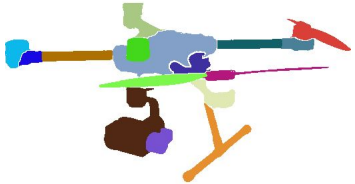
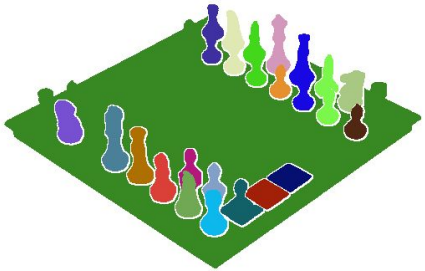



Image Prompts | Trellis | OmniPart | OmniPart++

Segmentation Results





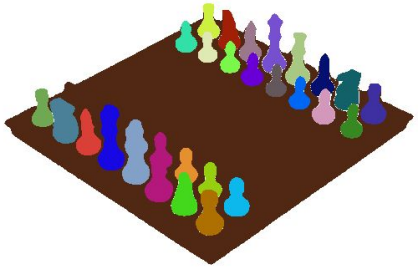


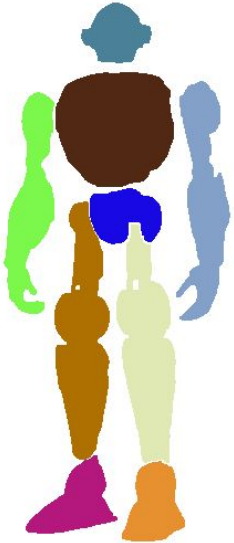
Image Prompts

OmniPart

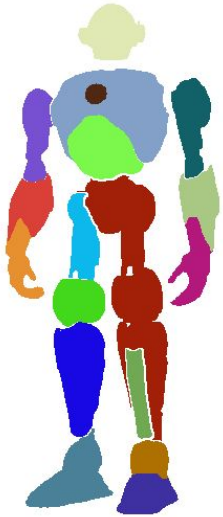
OmniPart++

Segmentation Results

Image Prompts



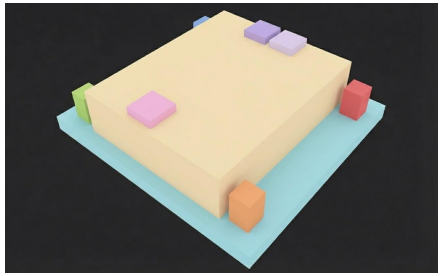
OmniPart



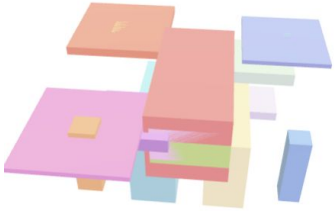
OmniPart++

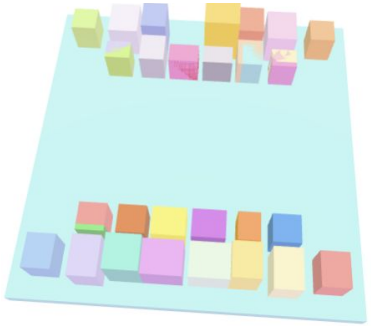
Bounding-Box Results

Image Prompts



OmniPart





OmniPart++

Bounding-Box Results

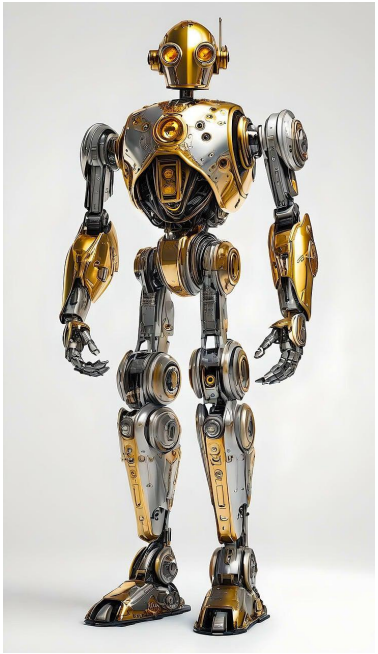
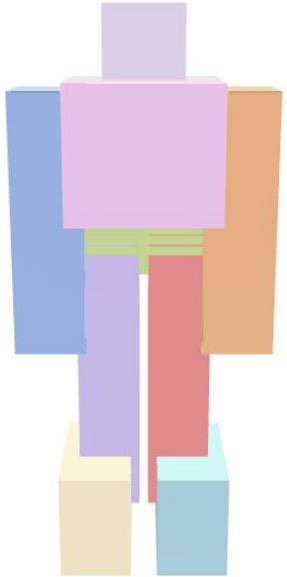


Image Prompts



OmniPart

OmniPart++

3D reconstruction Results

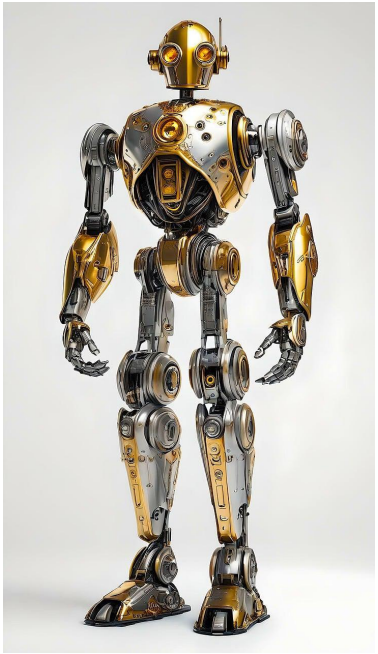


Image Prompts

OmniPart

OmniPart++

Quatitative Results

VisionReward:

- 설명

- **결과**

Quatitative Results

ImageReward:

- ImageReward is the general-purpose text-to-image **reward model** trained on expert human feedback to effectively evaluate and optimize generative models for better **alignment with human preferences**.
- Objects: batman, car, chessboard, chessboard2, digger, drone, hammer, robot, robot2, robotdog

```
Trellis Average Score: 0.5703
OmniPart Average Score: 0.4985
OmniPartPlusPlus Average Score: 0.6518
```

```
Difference (OmniPart++ - Trellis): 0.0815
Difference (OmniPart++ - OmniPart): 0.1533
```

```
[Trellis Results]
----------------------------------------
       batman: 1.6381
          car: 0.1735
   chessboard: 0.2891
  chessboard2: 0.3267
       digger: 0.9739
        drone: -0.7923
       hammer: 0.0916
        robot: 0.3074
       robot2: 1.8449
    robotdog: 0.8500
```

```
[OmniPart Results]
----------------------------------------
       batman: 1.6448
          car: 0.1991
   chessboard: 0.3338
  chessboard2: 0.2292
       digger: 0.6191
        drone: -0.4710
       hammer: 0.1137
        robot: -0.1262
       robot2: 1.0643
    robotdog: 1.3779
```

```
[OmniPartPlusPlus Results]
----------------------------------------
       batman: 1.7455
          car: 0.1146
   chessboard: 0.0865
  chessboard2: 0.1787
       digger: 0.5746
        drone: -0.4710
       hammer: 1.2690
        robot: 0.0607
       robot2: 1.6211
    robotdog: 1.3382
```

Conclusion

Limitations:

- Could not perfectly segment/reconstruct unseen area.
- Even with better segmentation mask, bounding box didn’t improve. Bounding box prediction model was bottleneck for the poor quality.
- Lack of generalizability.

Further Study:

- Improve structure of bounding box generation model to strictly follow segmentation mask & More tight bounding theme(box, sdf, pointcloud etc)
    - We have tried to train the bounding model of our own but didn’t have enough result until the final presentation

- Text-guided embeddings for part-level sampler, expand to text-to-3D model.

Thank you