ShapeCraft: LLM Agents for Structured, Textured and Interactive 3D Modeling

Shuyuan Zhang[1,*], Chenhan Jiang[2,*], Zuoou Li[1], Jiankang Deng[1]

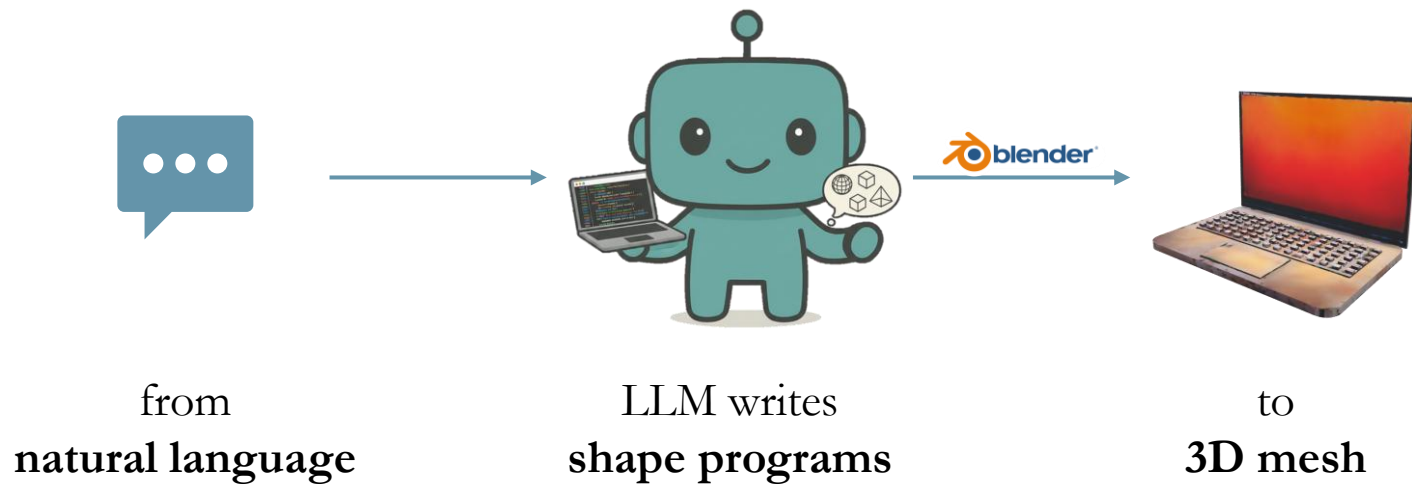[1]Imperial College London, [2]Hong Kong University of Science and Technology
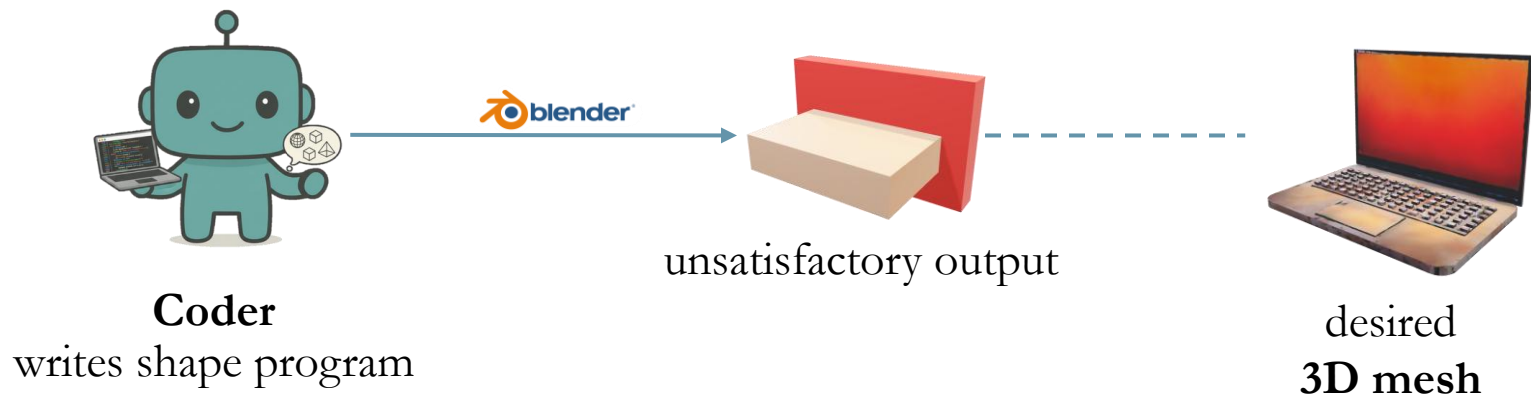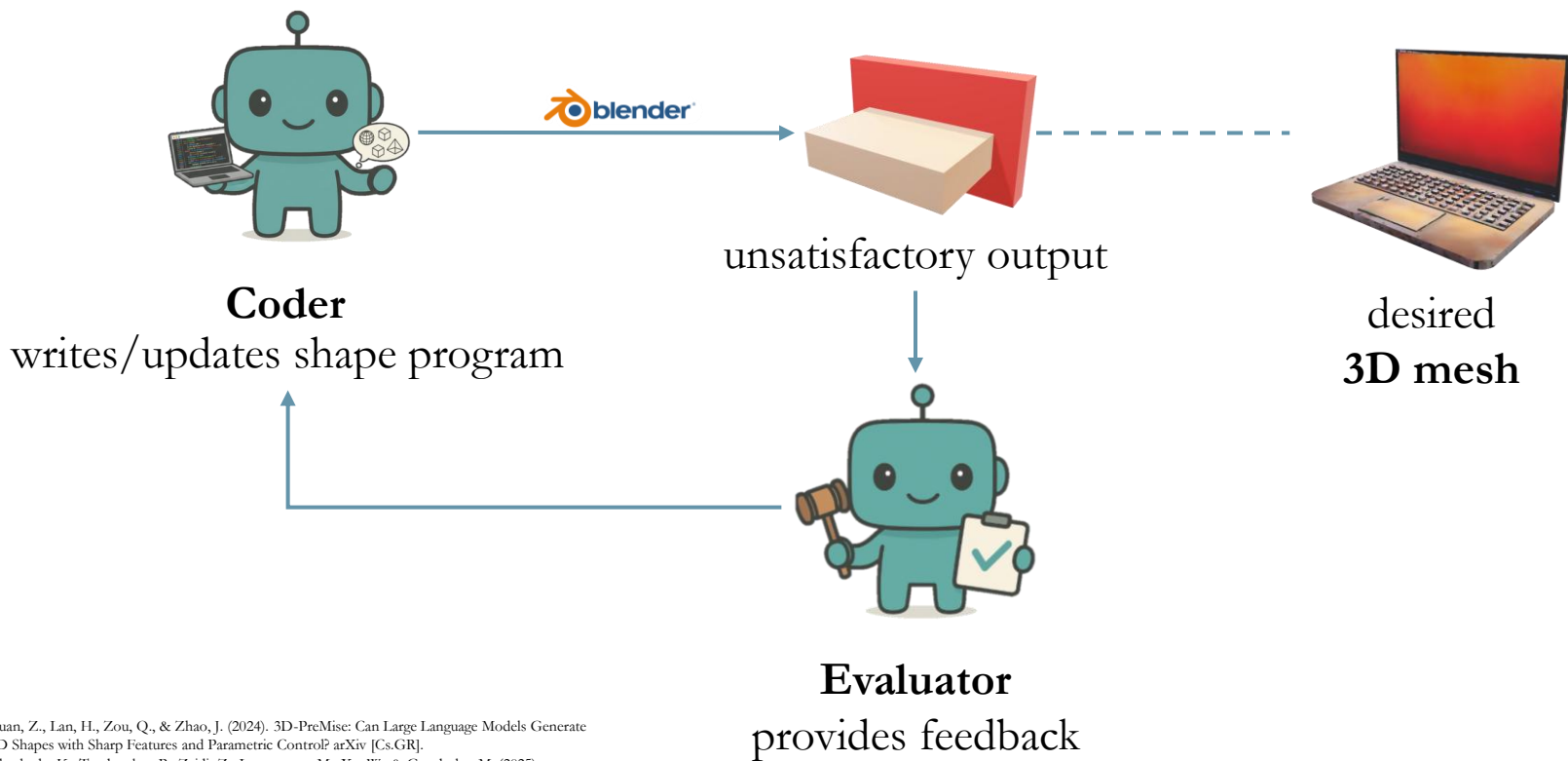*indicates equal contribution

https://sanbingyouyong.github.io/shapecraft/

The Task

from
natural language

LLM writes
shape programs

to
3D mesh

The Approach

Coder
writes shape program

unsatisfactory output

desired
3D mesh

The Approach

unsatisfactory output

Coder
writes/updates shape program

desired
3D mesh

Evaluator
provides feedback

Yuan, Z., Lan, H., Zou, Q., & Zhao, J. (2024). 3D-PreMise: Can Large Language Models Generate
3D Shapes with Sharp Features and Parametric Control? arXiv [Cs.GR].
Alrashedy, K., Tambwekar, P., Zaidi, Z., Langwasser, M., Xu, W., & Gombolay, M. (2025).
Generating CAD Code with Vision-Language Models for 3D Designs. arXiv [Cs.LG].

The Approach: Scaling Up

final
3D mesh

The Approach: Scaling Out

keyboard modeling

touchpad modeling

...

final
3D mesh

The Approach: Scaling Out

combined
3D mesh

The Approach: Parsing

Parser

Graph-based **P**rocedural **S**hape

v_3
keyboard

v_1
screen

v_0
laptop

v_4
touchpad

v_2
top case

Geometric description n_i^g

Positional description n_i^p
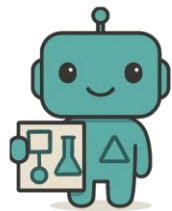
component
bounding volumes

The Approach: Parsing

Parser

Graph-based **P**rocedural **S**hape

v_3 keyboard			v_1 screen
	v_0 laptop		
v_4 touchpad			v_2 top case
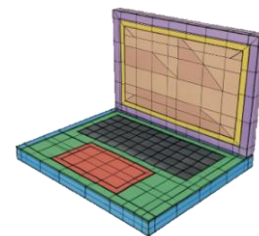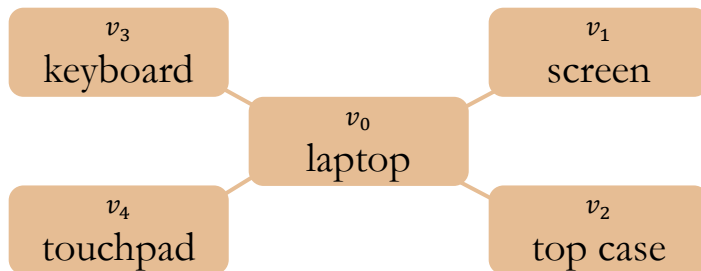
Geometric description n_i^g Code snippet p_i

Positional description n_i^p Bounding volume b_i
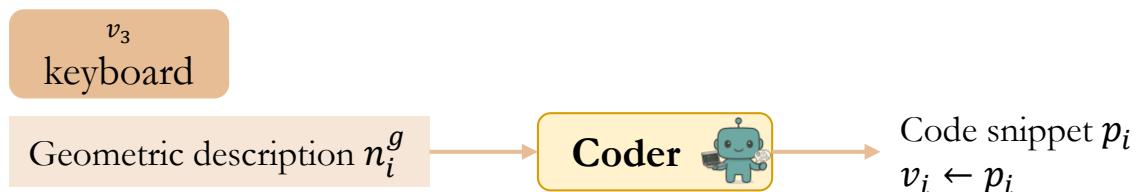
component **bounding volumes**

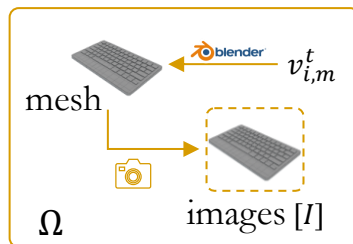# The Approach: Modeling

v_3
keyboard

Geometric description n_i^g → **Coder** → Code snippet p_i
$v_i \leftarrow p_i$

mesh → $v_{i,m}^t$

images $[I]$

Ω

$m \in \{1, M\}$ paths
$t \in \{0, T-1\}$ steps

The Approach: Modeling

$m \in \{1, M\}$ paths
$t \in \{0, T-1\}$ steps
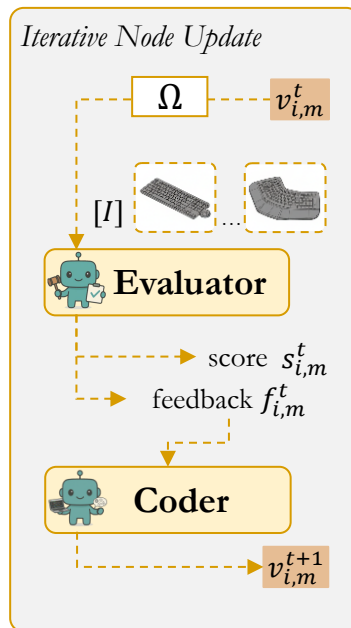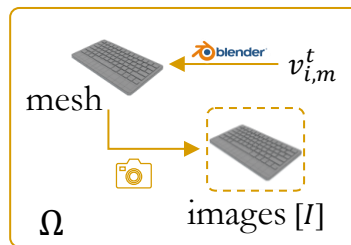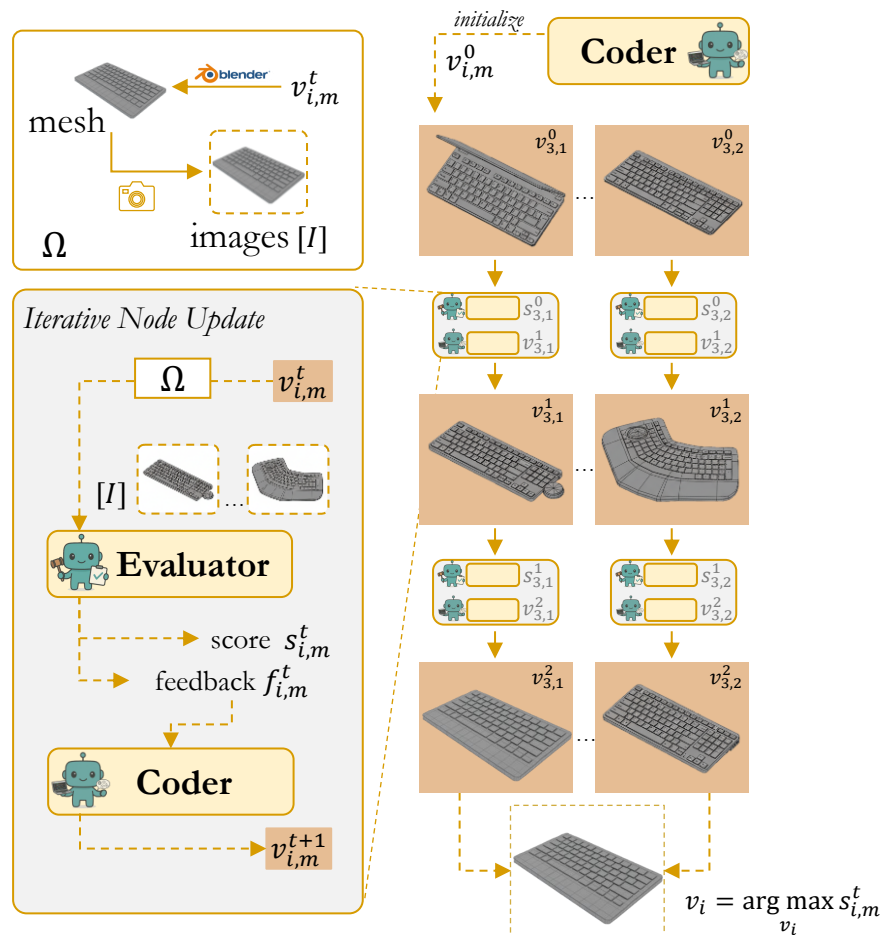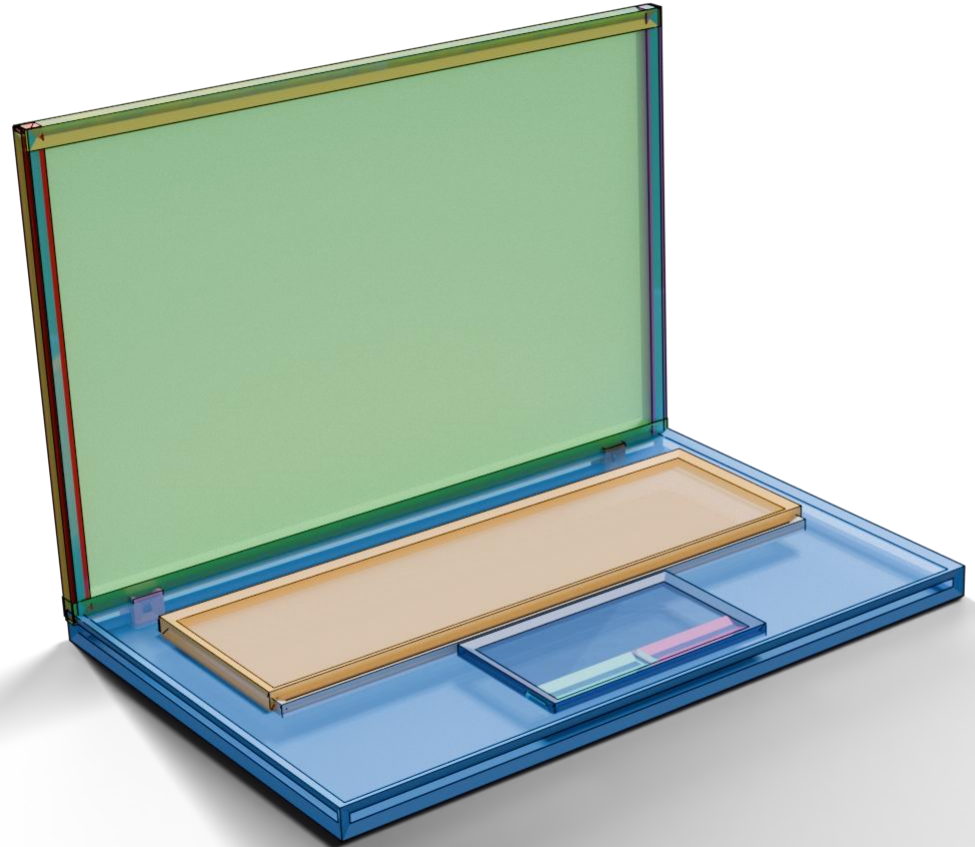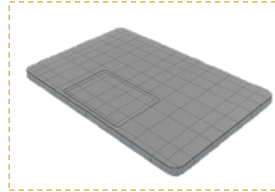
The Approach: Modeling

$m \in \{1, M\}$ paths
$t \in \{0, T-1\}$ steps

The Approach: Modeling
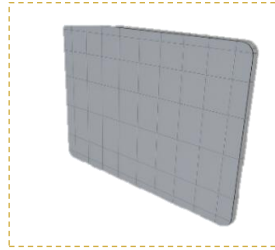
The Approach: Modeling
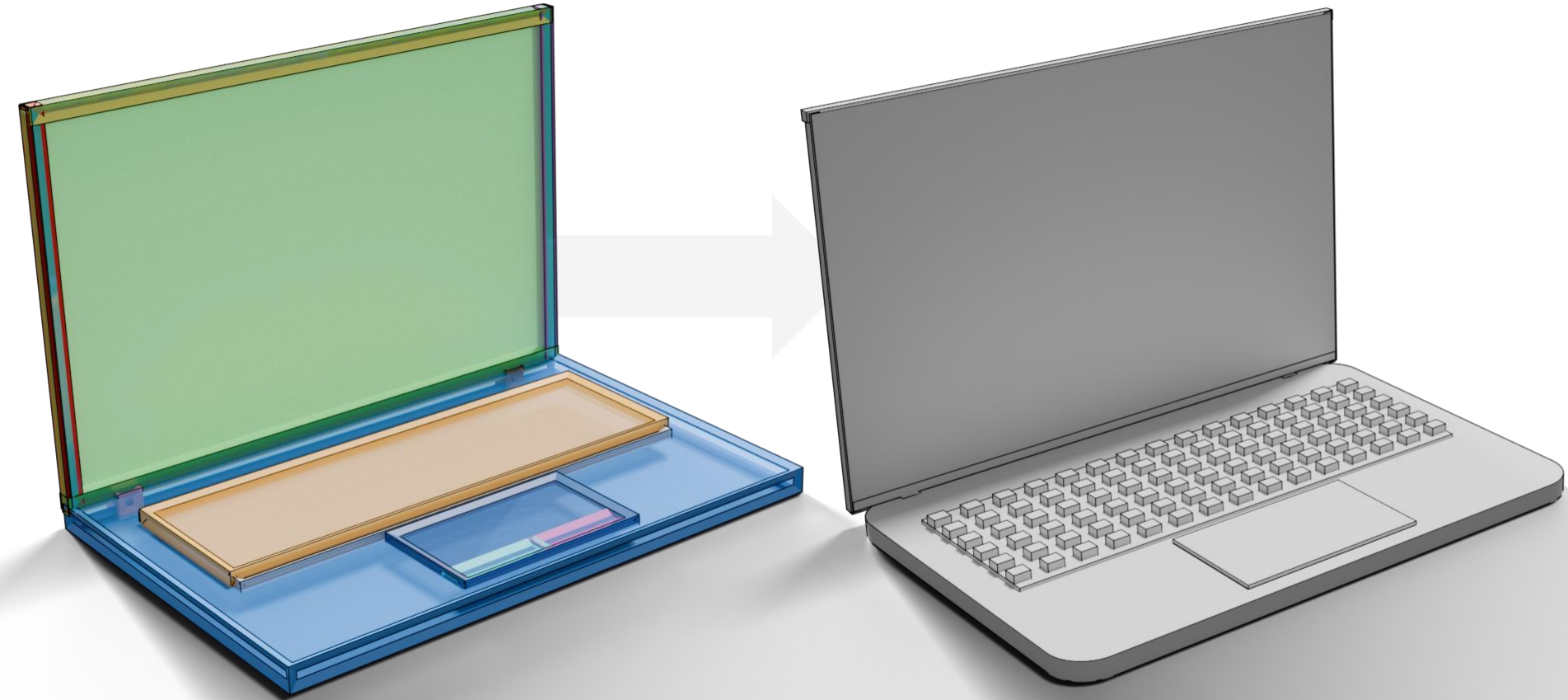
The Approach: Painting

Painting: *Texture Generation through Compositional Score Distillation*

Texturer Field Training

$\mathcal{G}^*$

v_i

Rendered Image

Global Description x

Forward / Backward — Texturer Field (Shared) — Stable Diffusion

Part Description n_1^q

Forward / Backward — Texturer Field (Shared) — Stable Diffusion

Part Description n_2^q

Forward / Backward — Texturer Field (Shared) — Stable Diffusion

Part Description n_i^q

Forward / Backward — Texturer Field (Shared) — Stable Diffusion

Texture Upsampling

Textured Shape Generation

Textured Shape
(Albedo, Metallic, Roughness)

Texturer Field
(Shared)

UV Coordinates
$\mathbf{p}$

The Approach: Painting

| BlenderLLM | 3D-PreMise | CADCodeVerify | L3GO | LLaMA-Mesh | **ShapeCraft** |

The *Stichter Banjo* is a traditional musical instrument with a **round body** and a **long neck**. The body has a **metallic ring** around it, holding a **drum-like** surface. The neck extends to a **headstock** with **tuning pegs** ... The **fretboard** has **evenly spaced frets**.

A pair of **eyeglasses** with a **thin, rectangular frame**. The **lenses** are slightly larger than the frame, and the **temples** are **slim and long rectangles**, currently both opened at **90 degrees vertically** to the frame.
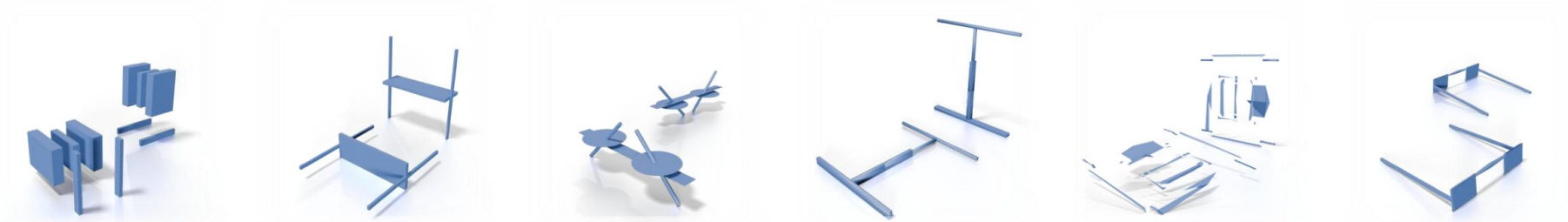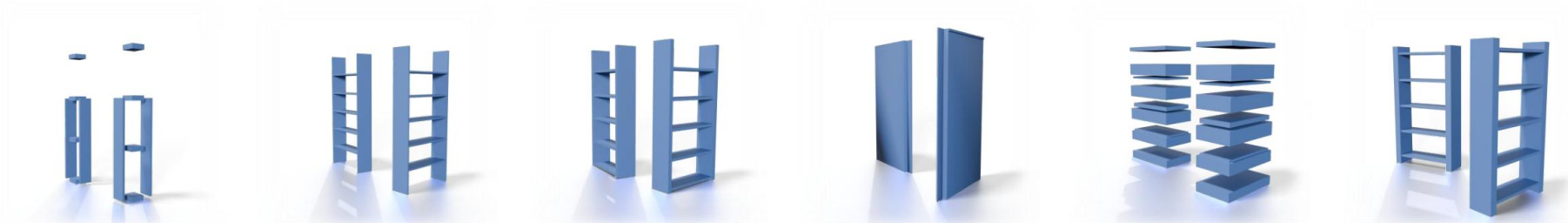
Bookshelf: A tall, **vertical frame** composed of **two thin, rectangular side panels**. **Horizontal shelves** — simple flat boxes — span between them at regular intervals. The overall structure is a **narrow, side-open box** with clean, right-angled lines.

Du, Y., Chen, S., Zan, W., Li, P., Wang, M., Song, D., ... Wang, B. (2024). BlenderLLM: Training Large Language Models for Computer-Aided Design with Self-improvement. arXiv [Cs.HC].
Yamada, Y., Chandu, K., Lin, Y., Hessel, J., Yildirim, I., & Choi, Y. (2024). L3GO: Language Agents with Chain-of-3D-Thoughts for Generating Unconventional Objects. arXiv [Cs.AI].
Wang, Z., Lorraine, J., Wang, Y., Su, H., Zhu, J., Fidler, S., & Zeng, X. (2024). LLaMA-Mesh: Unifying 3D Mesh Generation with Language Models. arXiv [Cs.LG].

The **air conditioner** is a **rectangular** model with ... **large circular fan grille** on ... **rust and dirt spots** ... color is **off-white** ... material has a **rough**, matte finish.
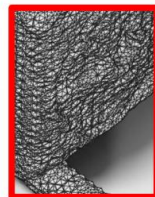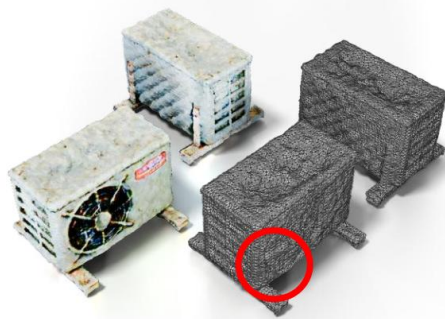
A **rectangular sofa** ... a **wide, flat box**. A **thinner box** ... **two narrow, vertical boxes** as armrests ... **four short cylindrical legs, slightly inset** ... floating appearance.
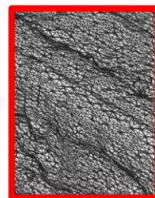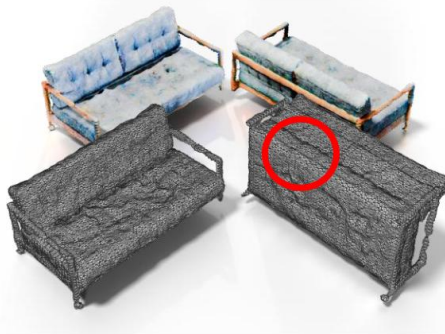
A **tall, upright rectangular fridge** ... **Two thinner rectangles** form the doors ... A **subtle horizontal groove** separates them and serves as a handle.
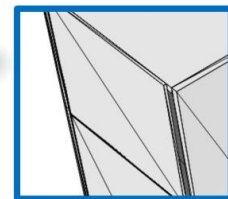
MVDream   **ShapeCraft**

Shi, Y., Wang, P., Ye, J., Long, M., Li, K., & Yang, X. (2024). MVDream: Multi-view Diffusion for 3D Generation. arXiv [Cs.CV].
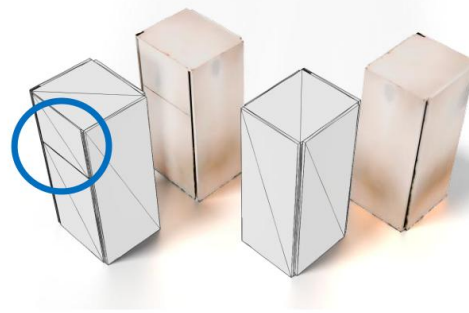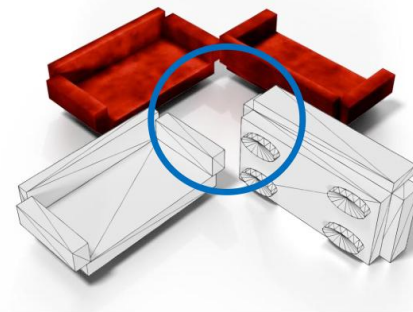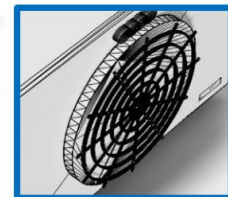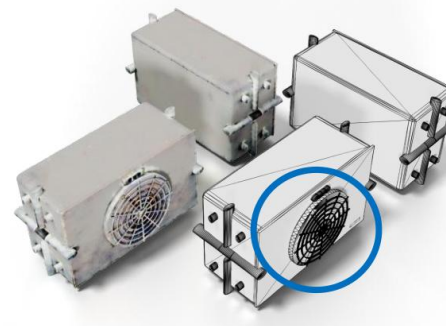
Table 1: Quantitative comparison of geometry quality and text-3D consistency on MARVEL subset.

Methods	IoGT↑	Hausdorff dist.↓	CLIP Score↑	VQA Pass Rate↑	Run Time↓	API Calls↓
3D-PREMISE [65]	0.385	0.527	26.76	0.33	2.81 min.	6
CADCodeVertify [1]	0.334	0.511	25.94	0.34	3.06 min.	9
BlenderLLM [14]	0.455	0.511	26.99	0.43	5.11 min.	N.A
LlaMA-Mesh [57]	0.346	0.464	25.72	0.28	15.64 min.	N.A
MVDream [51]	0.427	**0.411**	26.84	0.42	32.10 min.	N.A
ShapeCraft	**0.471**	0.415	**27.27**	**0.44**	11.68 min.	21

Table 2: **Ablation studies on sampled paths M and iterative updates T in shape modeling.** Lower Hausdorff and runtime are better, and higher IoGT and CLIP Score are better. ShapeCraft demonstrates a strong balance between exploration and efficiency.

Metric	M=1, T=1	M=3, T=1	M=1, T=3	**ShapeCraft** (M=3, T=3)	M=3, T=5
Hausdorff ↓	0.485	0.444	0.494	<u>0.415</u>	**0.360**
IoGT ↑	0.436	**0.535**	<u>0.492</u>	0.471	0.431
CLIP Score ↑	25.75	25.90	26.20	**27.27**	<u>26.39</u>
Run Time (min) ↓	**1.62**	<u>3.71</u>	3.90	11.68	18.04

Table 3: **Ablation study on hierarchical shape parsing in GPS representation.** We compare with advanced LLMs operating with thinking mode. The results show our GPS representation constrains the reasoning space of LLMs, leading to more reliable and interpretable.

Metrics	ChatGPT-o3	ChatGPT-o4-mini-high	Deepseek-R1-0528	Gemini-2.5-Pro	**ShapeCraft**
IoGT ↑	0.177	0.244	0.326	0.102	**0.471**
Hausdorff ↓	0.708	0.493	0.489	0.586	**0.415**
CLIP ↑	25.48	26.30	**29.01**	27.31	27.27
Compile Rate ↑	60%	80%	80%	60%	**100%**

Zhou, T., Mei, S., Li, X., Liu, Z., Xiong, C., Liu, Z., … Yu, G. (2024). MARVEL: Unlocking the Multi-Modal Capability of Dense Retrieval via Visual Module Plugin. arXiv [Cs.IR].

Post-modeling interactions

Global shape program

Given the following shape program that generates "A vertically positioned notebook with …", please modify it based on user request: **Reduce the number of metal rings to 3.**

To reduce the number of metal binding rings from 5 to 3, only the `binding_rings` function needs to be updated. Specifically, **change the count parameter** in the Modifiers.array call **from 5 to 3.** Here's the modified version of that function:
```python
…
```

```python
# Use array modifier to create 5 rings with proper vertical spacing
ring = Modifiers.array(
    ring,
    count=3,   # Reduced from 5 to 3
    relative_offset=(0, 0, 1.2)  # Spacing along Z-axis (vertical)
)
```

Updated shape program
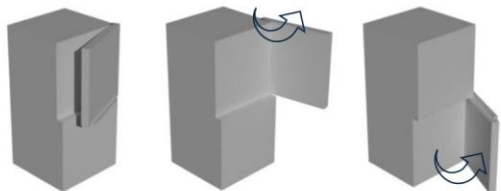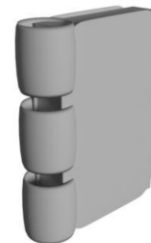
"Given the shape program, please write a Python method to add the following animation and export as video:
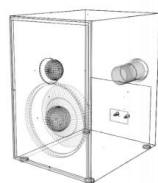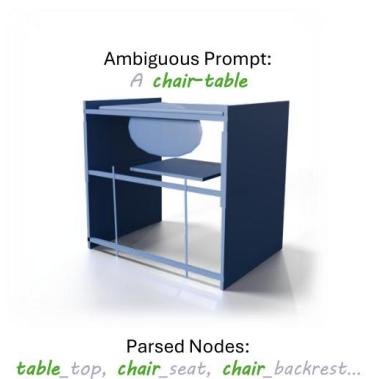- Fridge door opens and closes
- Freezer door opens and closes

"Given the shape program, please write a Python method to add the following animation and export as video:
- Laptop screen is turned to 135 degrees
- Laptop screen is closed

Limitations

Ambiguous Prompt:
A chair-table

Parsed Nodes:
table_top, chair_seat, chair_backrest...

Actual
Geometry

Brief Prompt:
A simple speaker

Visual
Exterior

Interior
Parts

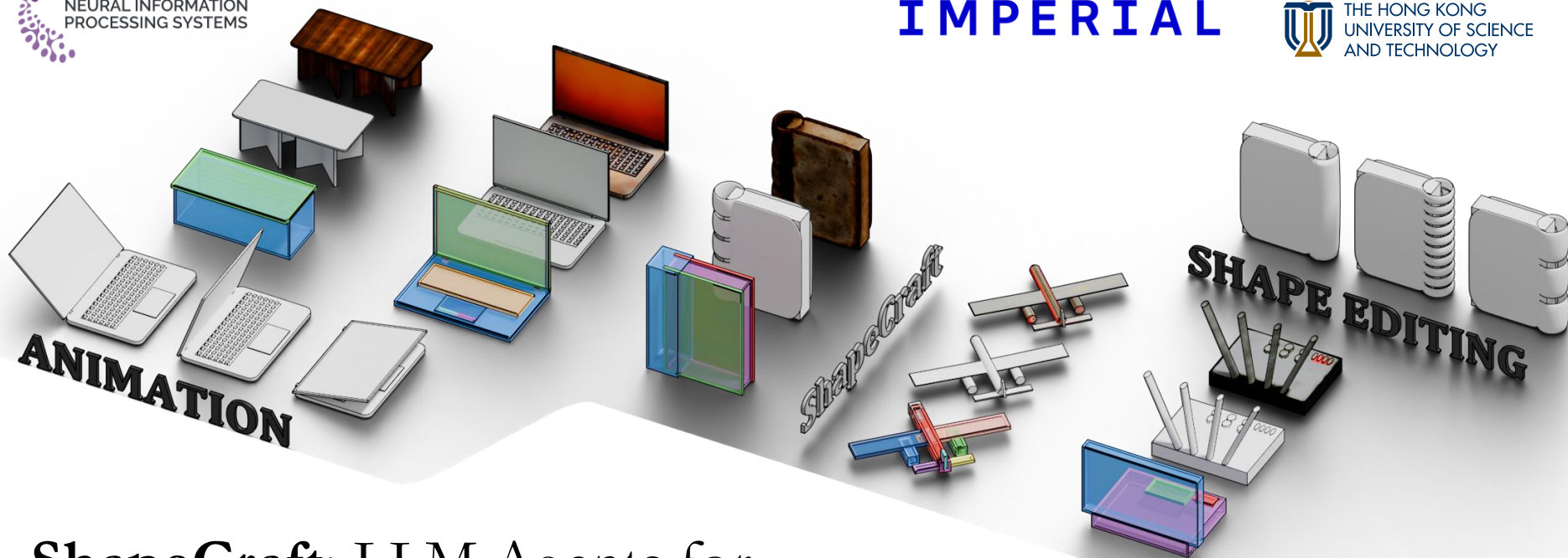Creative Prompt:
A sofa that looks like an apple

Floating
Stem

Interior
Cushions

ShapeCraft: LLM Agents for Structured, Textured and Interactive 3D Modeling

Shuyuan Zhang[1,*], Chenhan Jiang[2,*], Zuoou Li[1], Jiankang Deng[1]

[1]Imperial College London, [2]Hong Kong University of Science and Technology
*indicates equal contribution

https://sanbingyouyong.github.io/shapecraft/