清華大學
Tsinghua University

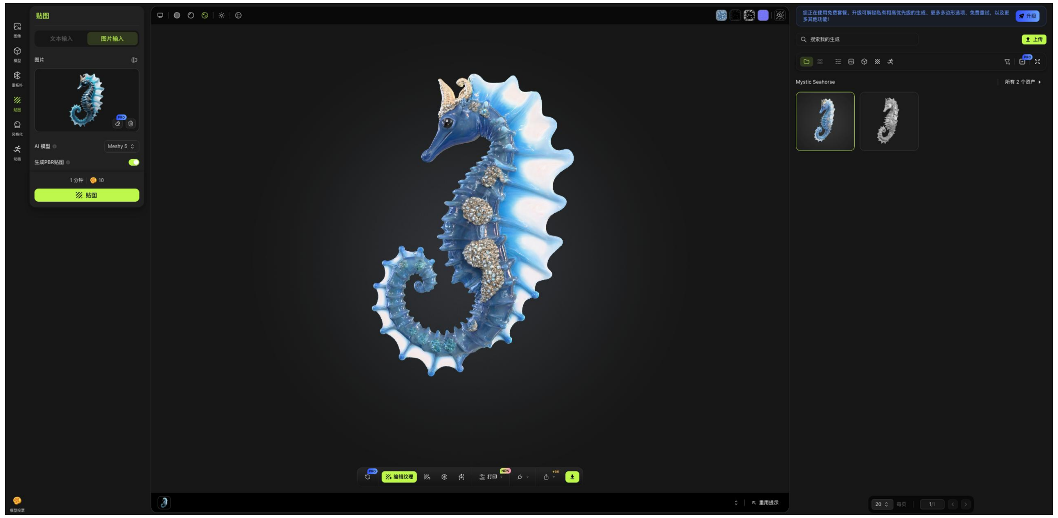
ICCV HONOLULU HAWAII
OCT 19-23, 2025

DeepMesh: Auto-Regressive Artist-mesh Creation with Reinforcement Learning

https://zhaorw02.github.io/DeepMesh/

Ruowen Zhao[*1,3] Junliang Ye[*1,3] Zhengyi Wang[*1,3]
Guangce Liu[3] Yiwen Chen[2] Yikai Wang[1] and Jun Zhu[†1,3]
[1]Tsinghua University [2]Nanyang Technological University [3]ShengShu
([*]Equal Contribution [†]Corresponding Author)

Motivation



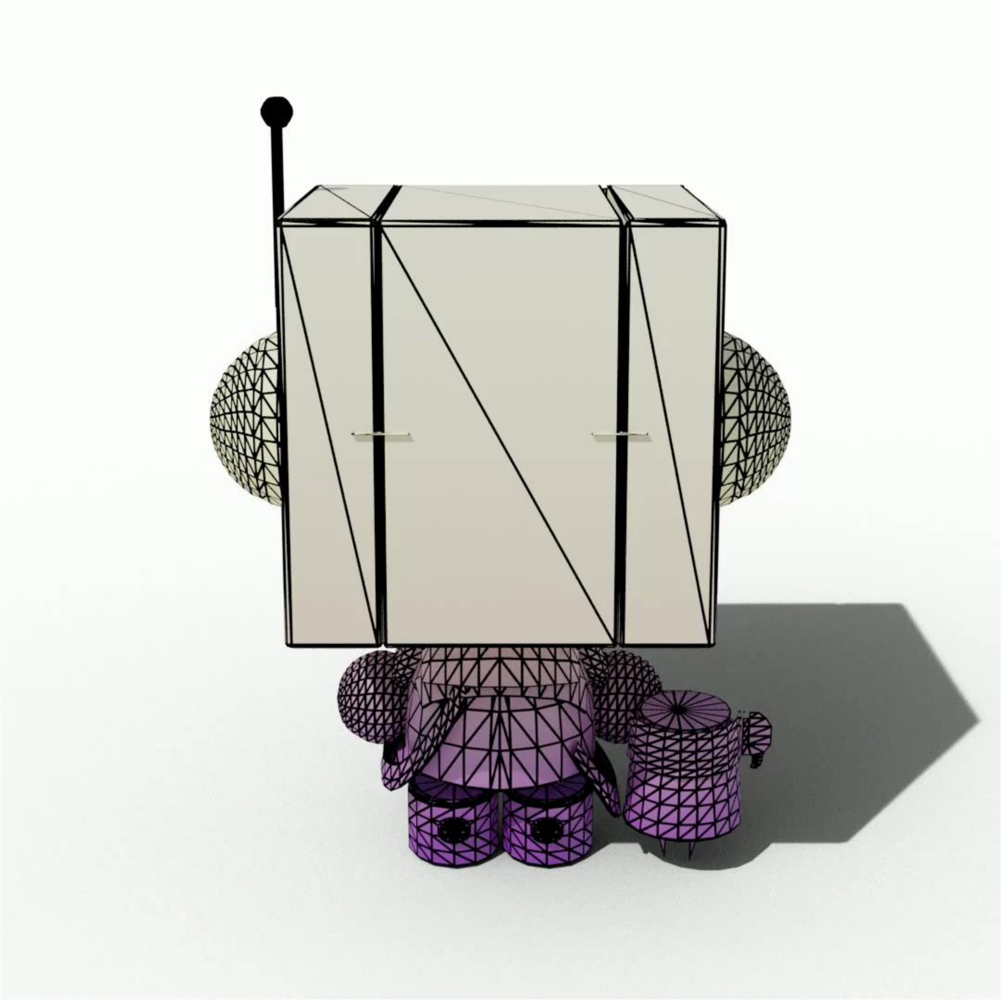
Meshy ai

Rodin

- Dense meshes with poor topology
- overly irregular structures
- Not suitable for downstream applications

Motivation

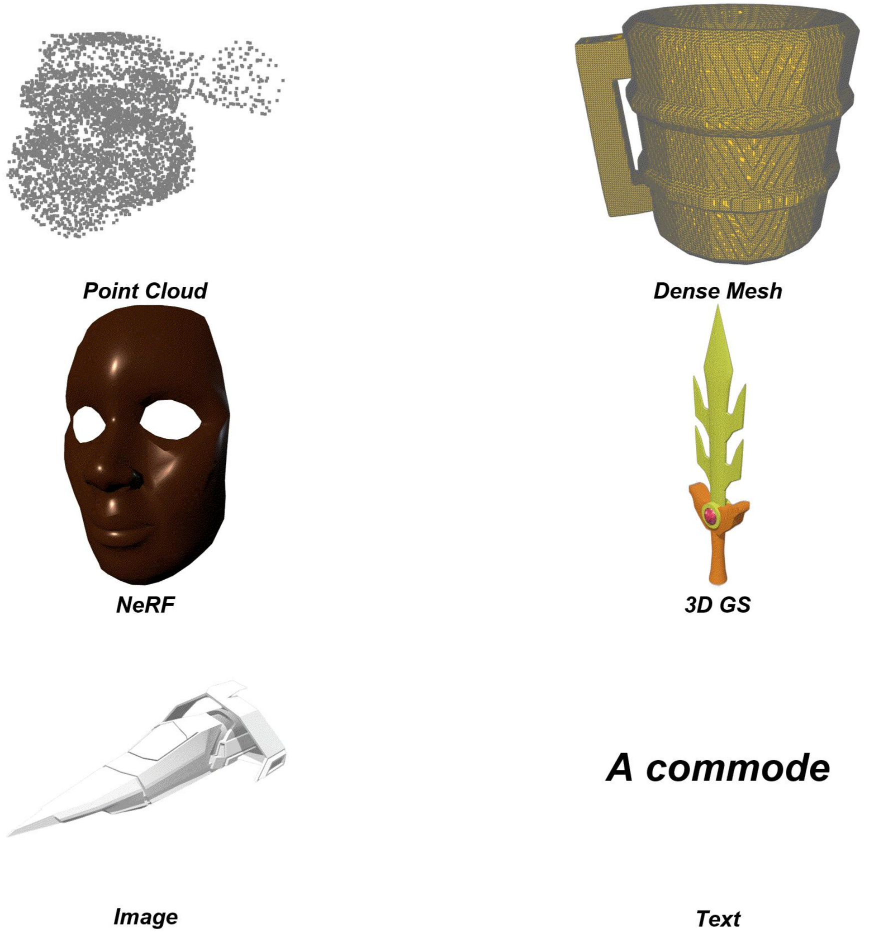
- Artist-created mesh has more optimal topology
- Suitable for downstream applications such as editing and rendering



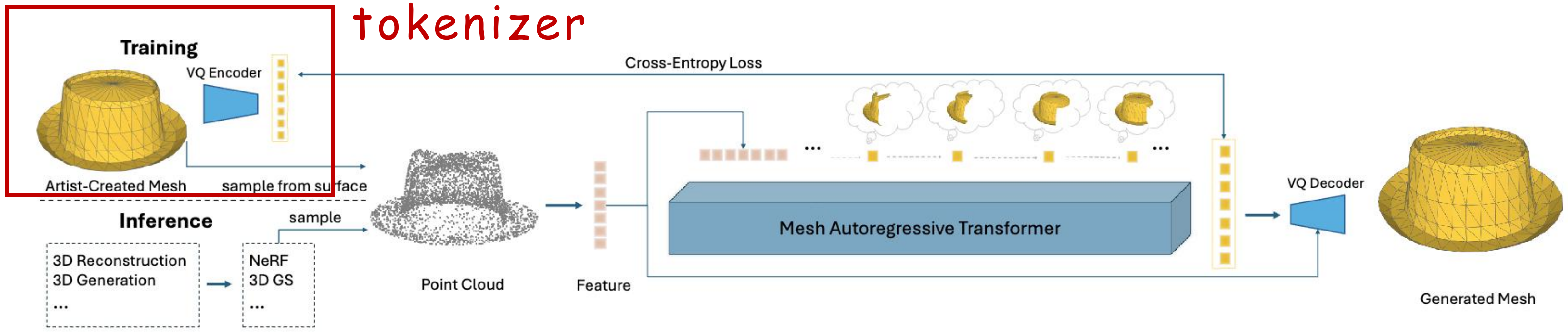
Previous method

- Existing methods learn to autoregressively predict mesh vertices and faces, preserving the structured and artistically optimized topology.

- Two stage pipeline: a mesh tokenizer and training an autoregressive transformer



MeshAnything



MeshAnything's pipeline

Challenges

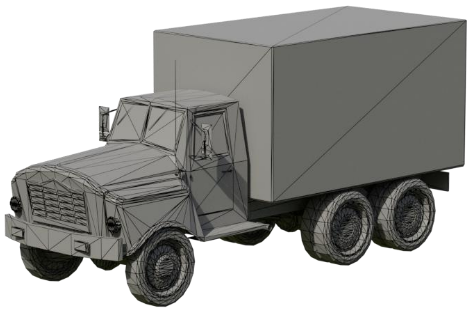
- Ineffcent tokenizer

More efficient tokenizer

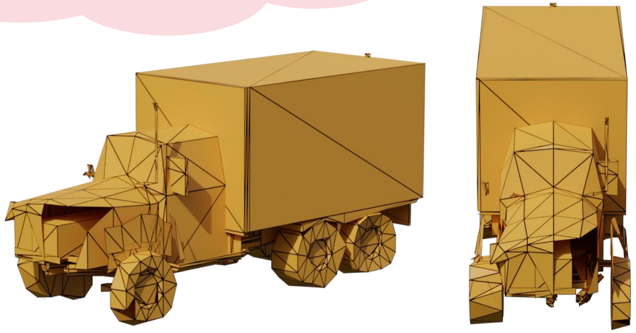
  - ➢ Native tokenization: 1 triangular face → (x,y,z) coordinates for 3 vertices → 9 tokens
  - ➢ AMT and EdgeRunner: compression rate of 46%, still long sequences
  - ➢ BPT: compression rate of 74% but 4w+ vocab size for 512 resolution

- Incomplete geometry

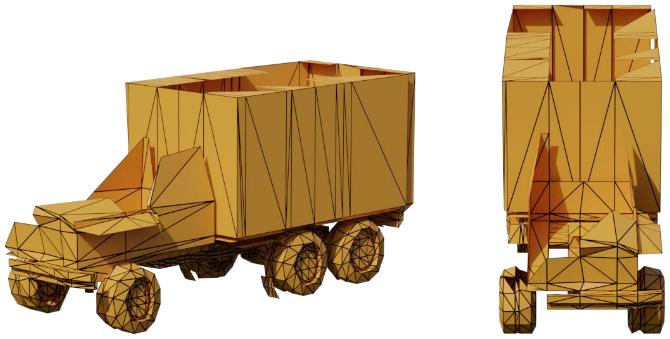
RL post-training



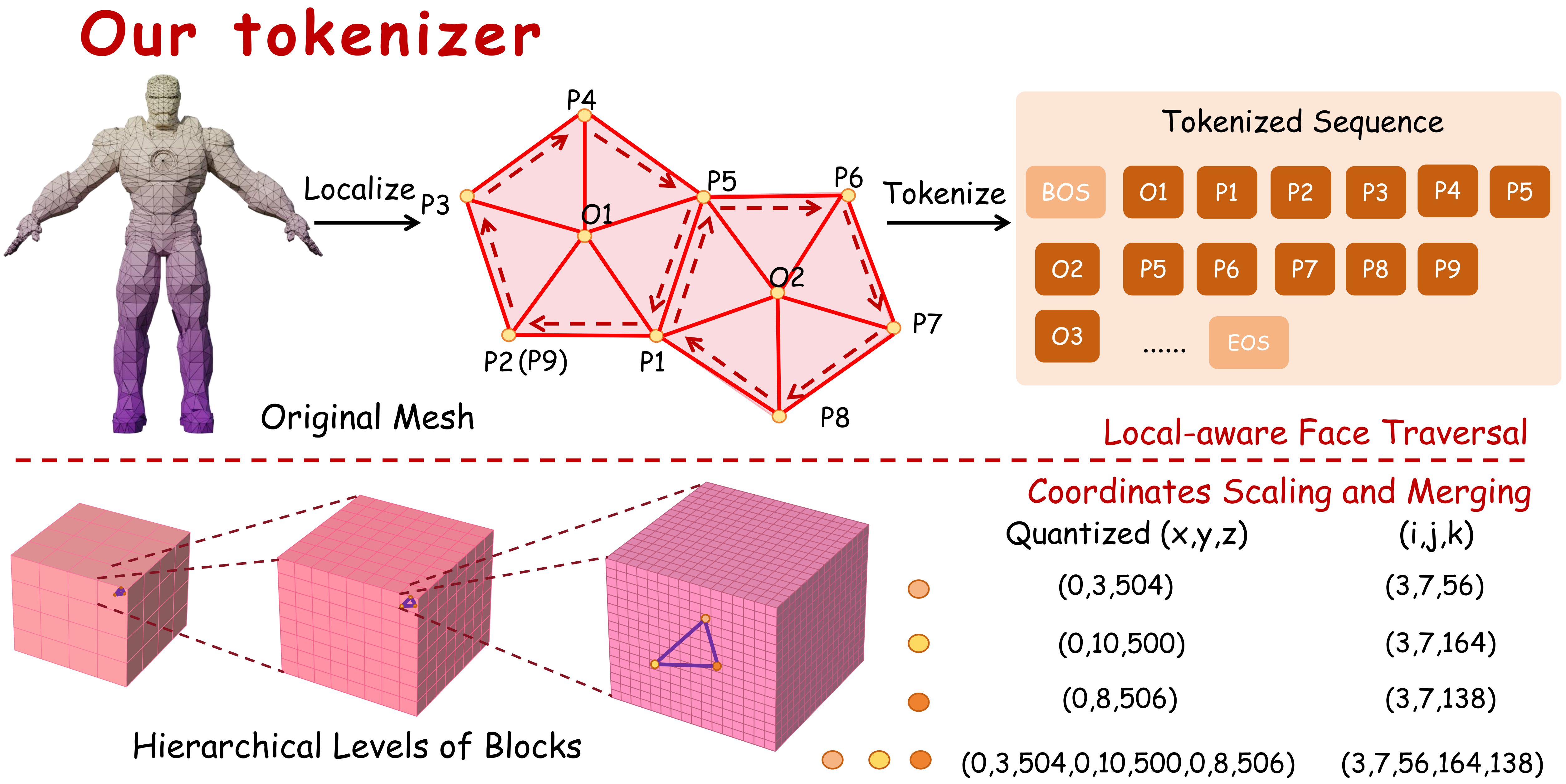
Original mesh



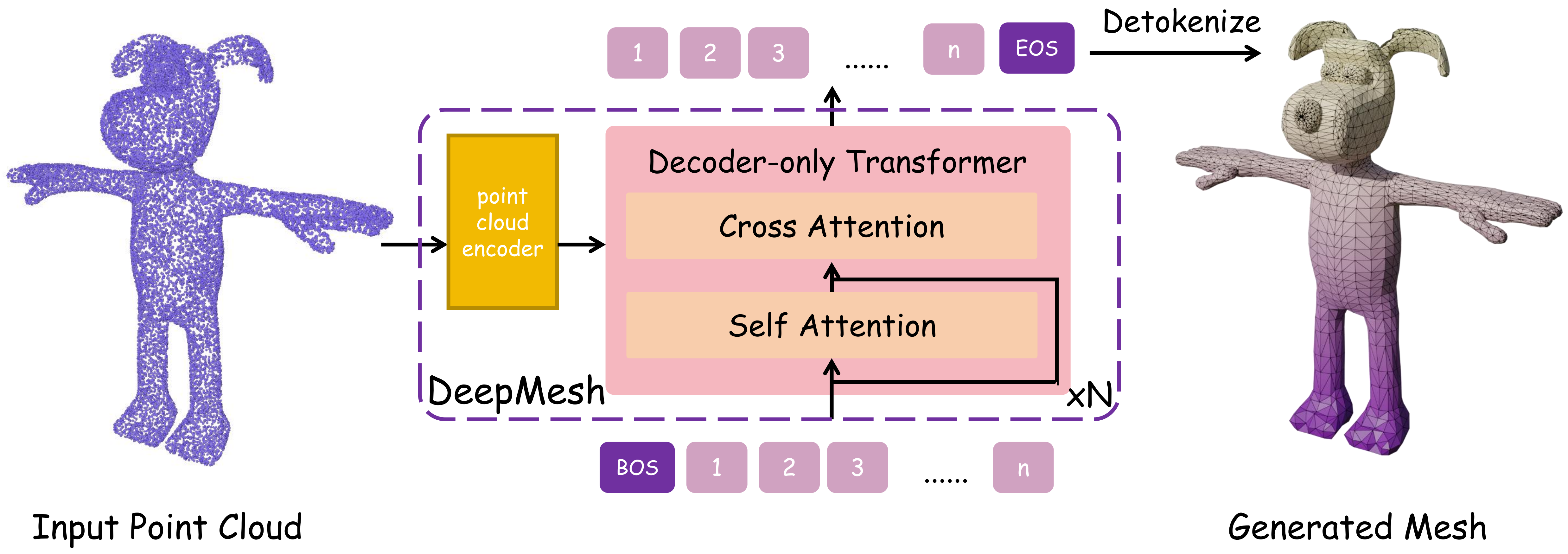
MeshAnythingv2

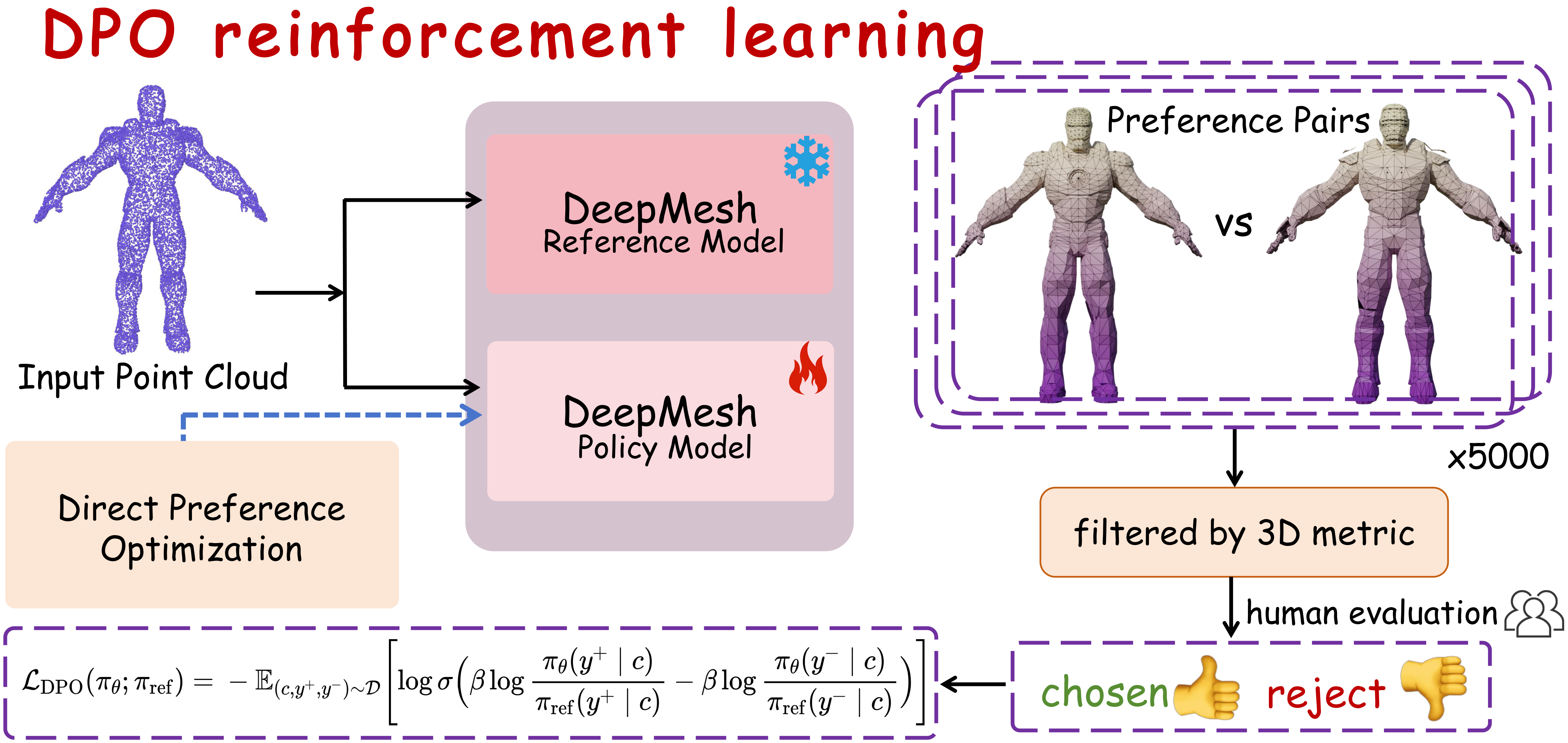


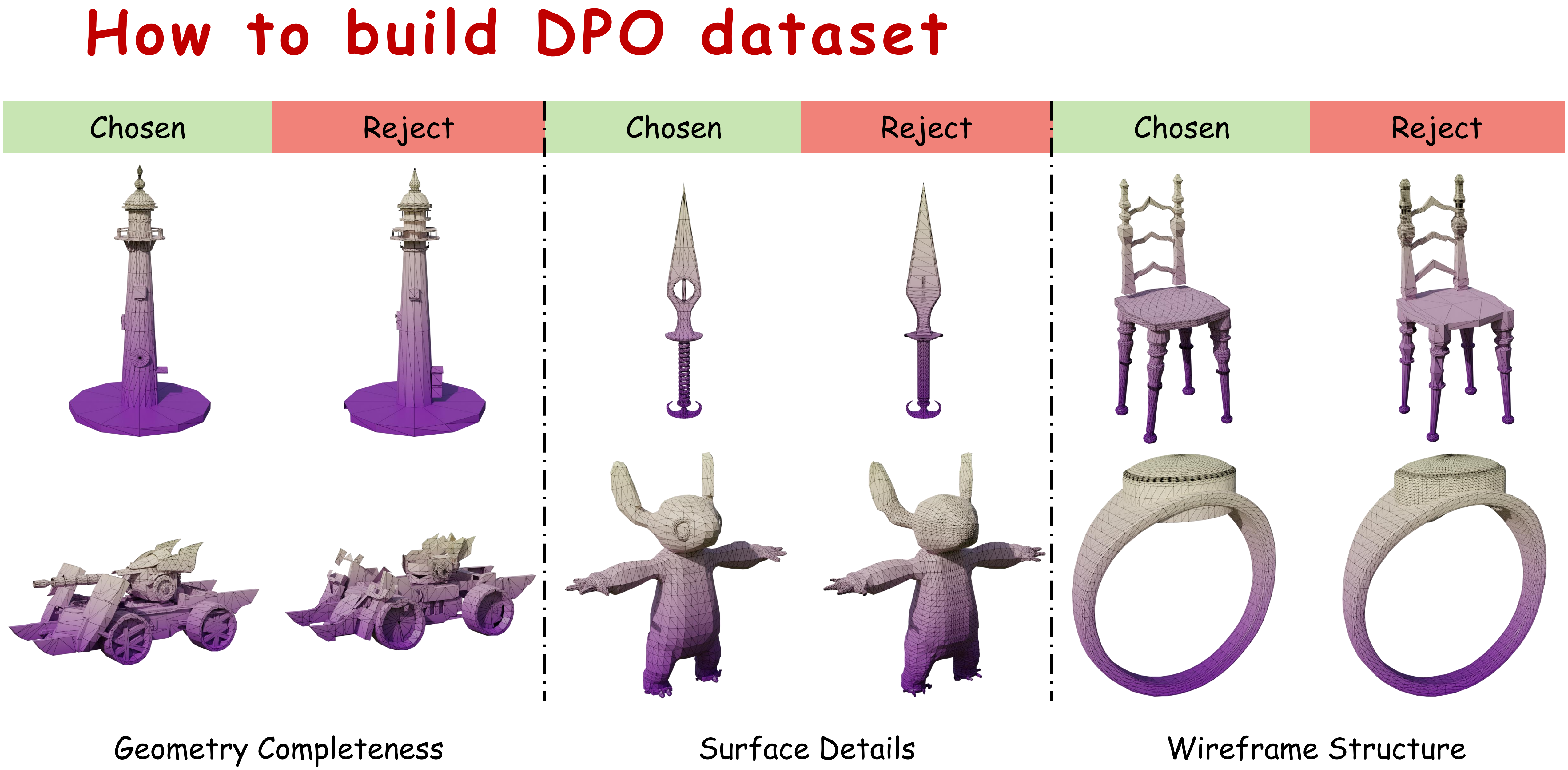
BPT


Our tokenizer
Localize
P4
P3
O1
P5
P6
O2
P7
P2 (P9)
P1
P8
Original Mesh
Tokenize
Tokenized Sequence
BOS
O1
P1
P2
P3
P4
P5
O2
P5
P6
P7
P8
P9
O3
......
EOS
Local-aware Face Traversal
Coordinates Scaling and Merging
Quantized (x,y,z)
(i,j,k)
(0,3,504)
(3,7,56)
(0,10,500)
(3,7,164)
(0,8,506)
(3,7,138)
(0,3,504,0,10,500,0,8,506)
(3,7,56,164,138)
Hierarchical Levels of Blocks

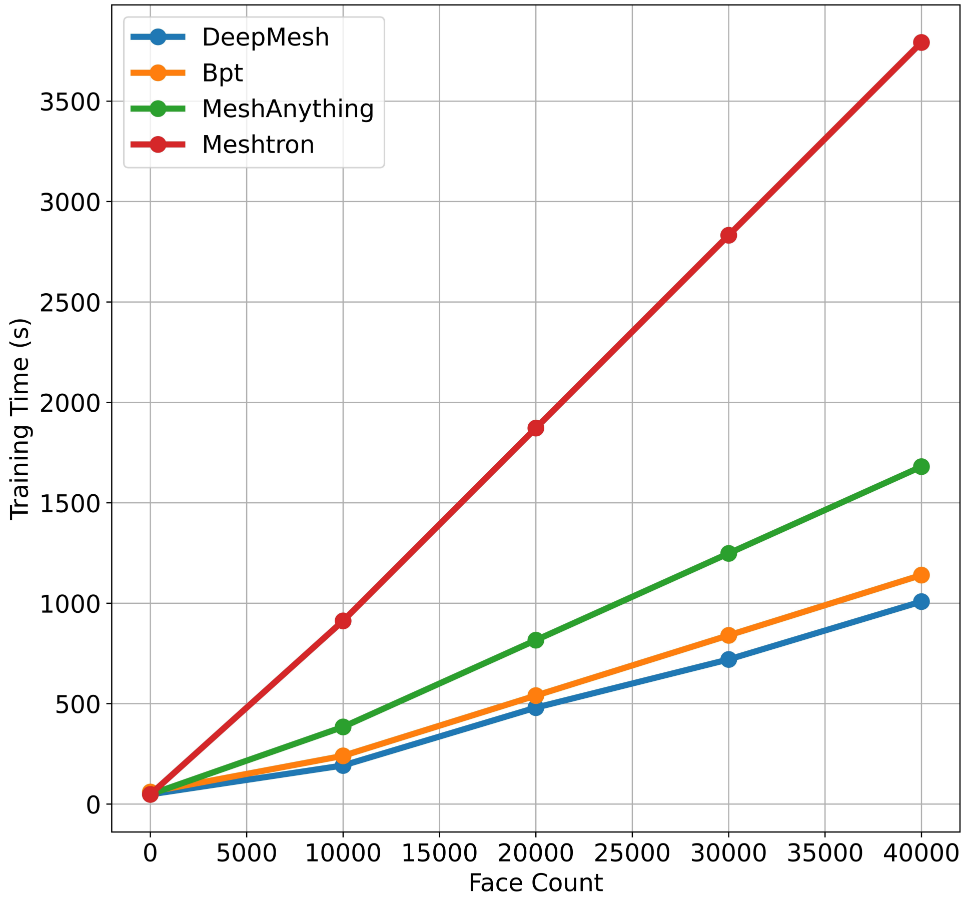
Pre-training architecture




DPO reinforcement learning
Input Point Cloud
DeepMesh
Reference Model
DeepMesh
Policy Model
Direct Preference Optimization
Preference Pairs
vs
x5000
filtered by 3D metric
human evaluation
chosen
reject
$\mathcal{L}_{\mathrm{DPO}}(\pi_\theta;\pi_{\mathrm{ref}}) = -\mathbb{E}_{(c,y^+,y^-)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y^+\mid c)}{\pi_{\mathrm{ref}}(y^+\mid c)} - \beta\log\frac{\pi_\theta(y^-\mid c)}{\pi_{\mathrm{ref}}(y^-\mid c)}\right)\right]$


How to build DPO dataset
Chosen
Reject
Chosen
Reject
Chosen
Reject
Geometry Completeness
Surface Details
Wireframe Structure

Expeiments



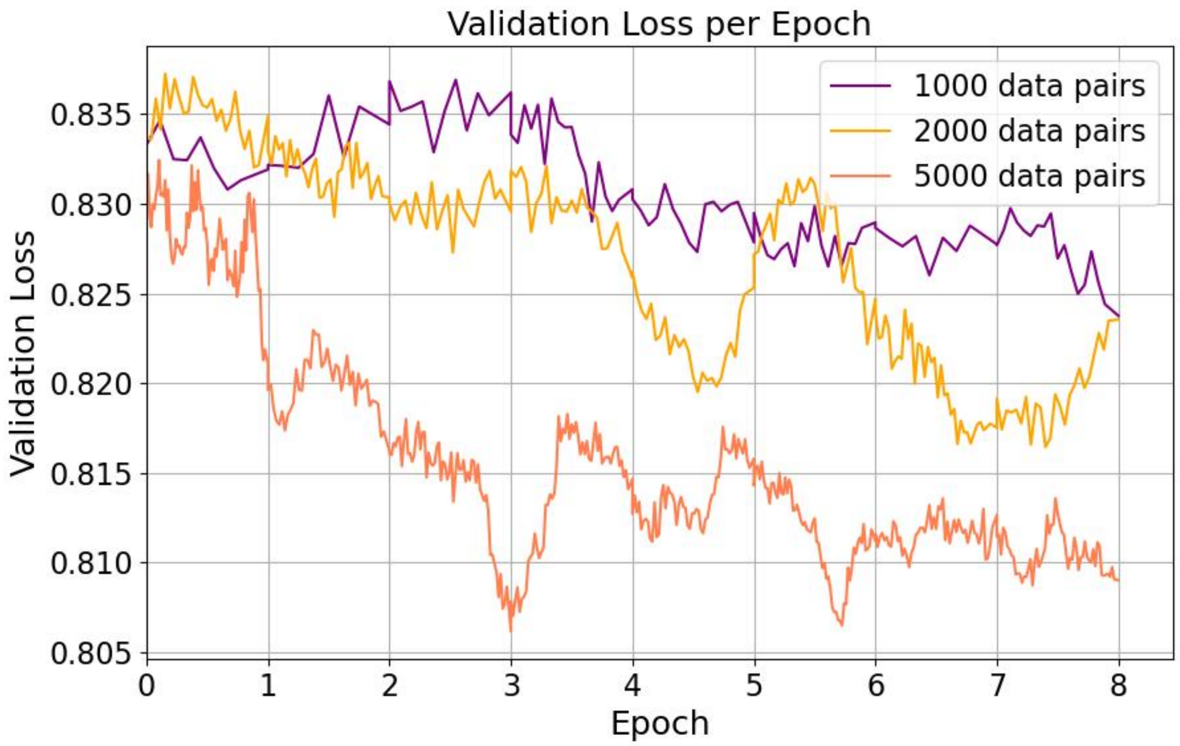
Metrics	C.Dist. ↓	H.Dist. ↓	User Study ↑
MeshAnythingv2 [6]	0.1249	0.2991	10%
BPT [67]	0.1425	0.2796	19%
Ours w/o DPO	0.1001	0.1861	34%
Ours w DPO	**0.0884**	**0.1708**	**37%**

Better generation results

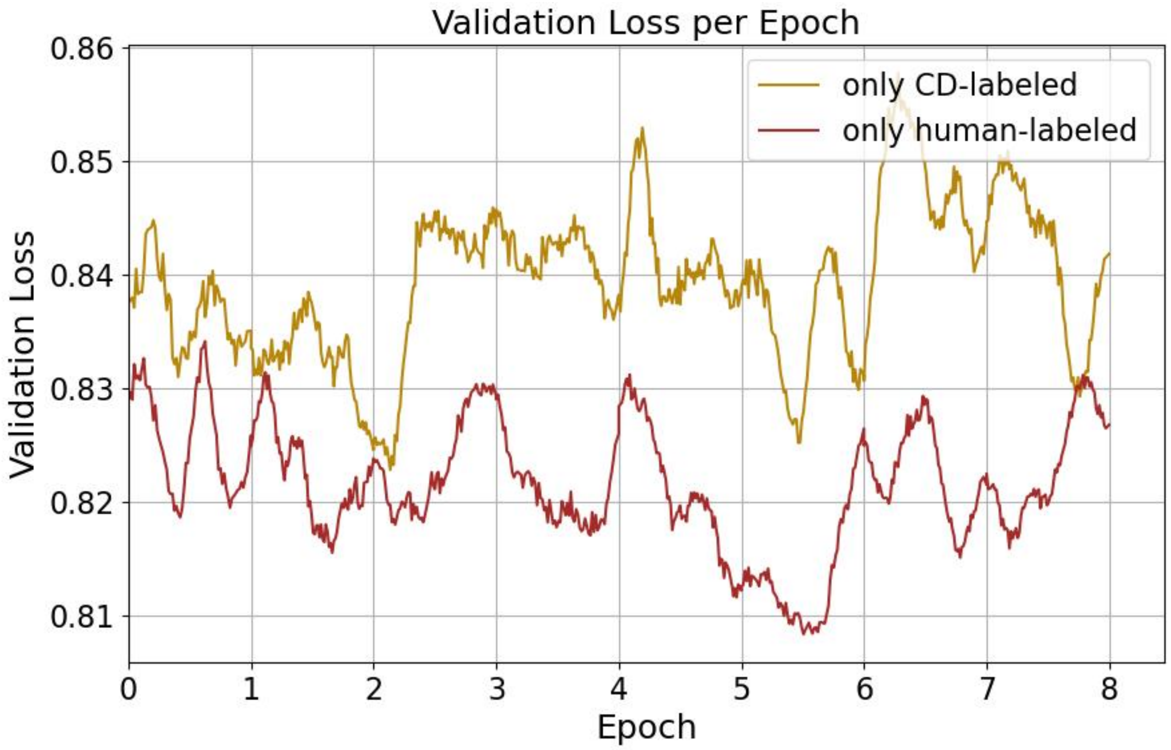
Metrics	AMT	EdgeRunner	BPT	Ours
Comp Ratio ↓	0.46	0.47	**0.26**	0.28
Vocal Size ↓	**512**	**512**	40960	4736
Time (s) ↓	816	-	540	**480**

More efficient tokenizer

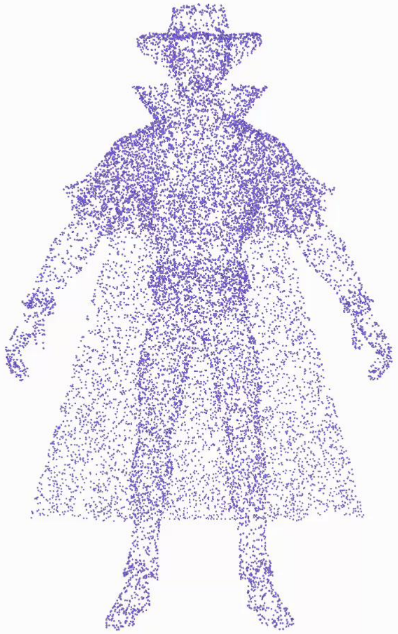
Expeiments



Scaling up data pairs leads to greater reductions in validation loss, indicating better DPO generalization.



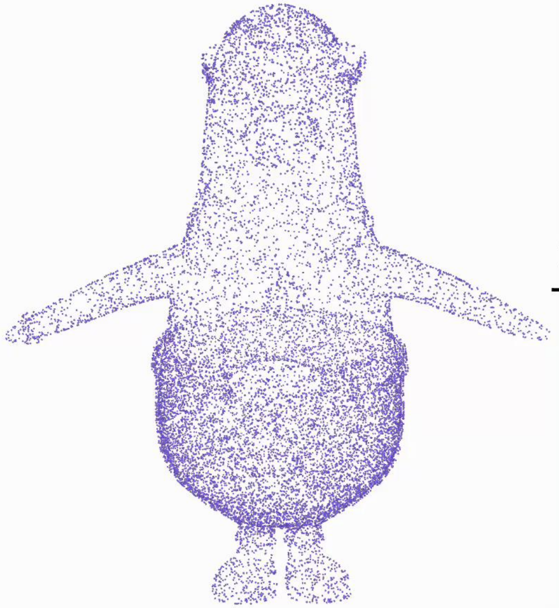
Post-training on human-labeled data yields lower validation loss, highlighting the necessity of human-annotation beyond geometric metrics.



DeepMesh

Point Cloud

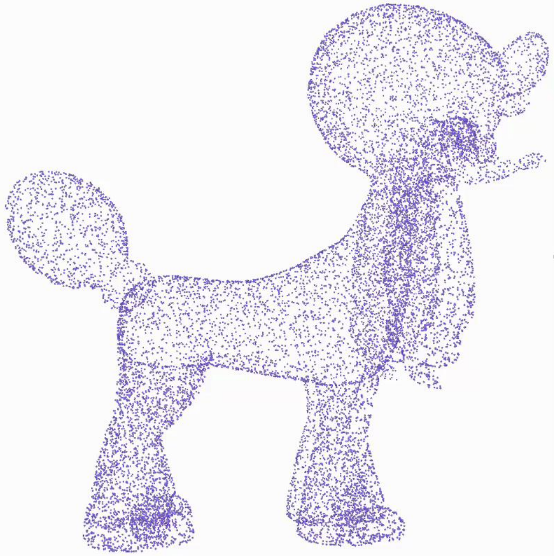
Generated Mesh and Its Generation Process



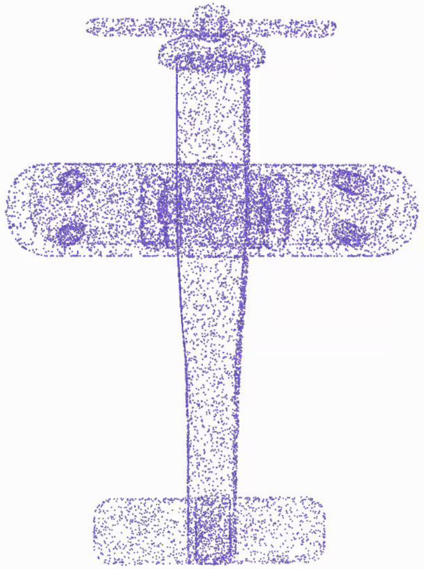
DeepMesh

Point Cloud

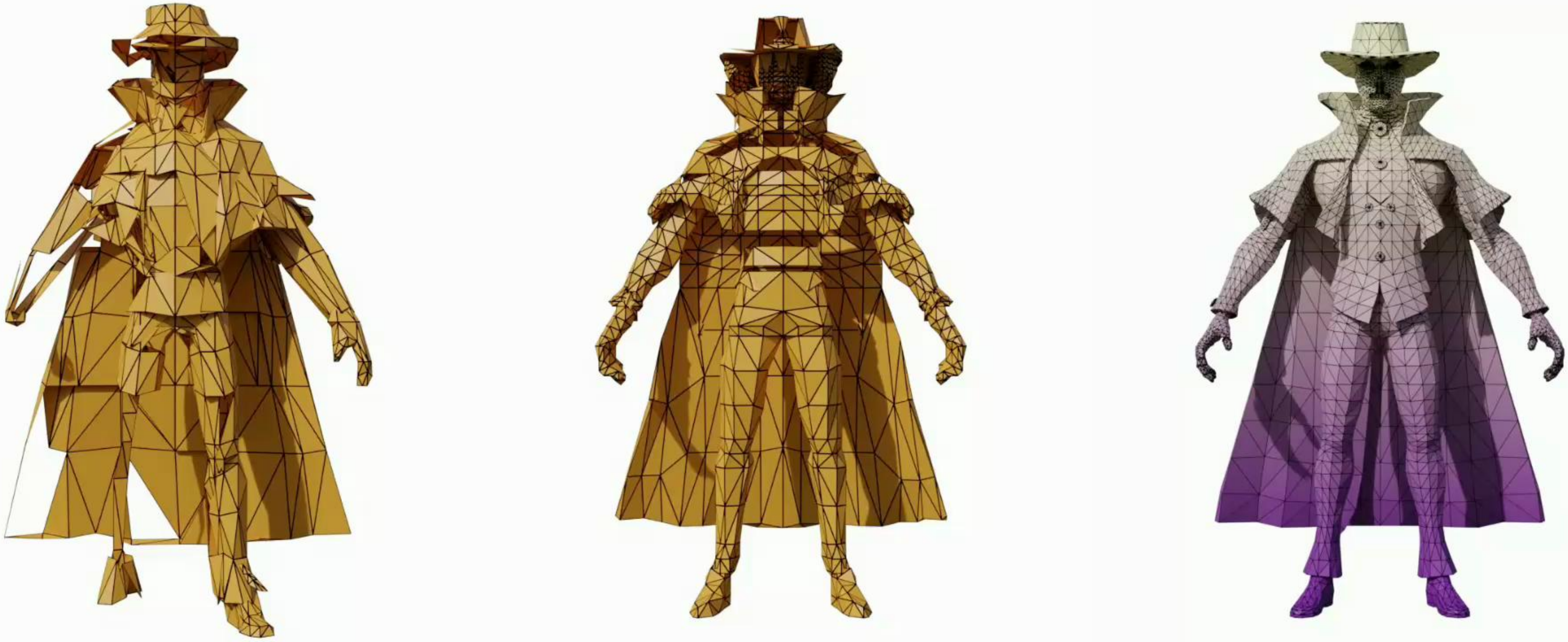
Generated Mesh and Its Generation Proce



DeepMesh



DeepMesh



MeshAnythingv2 BPT DeepMesh (ours)

MeshAnythingv2

BPT

DeepMesh (ours)


Point Cloud
Our Diverse Results

Thank you

Project page: https://zhaorw02.github.io/DeepMesh/