SFU

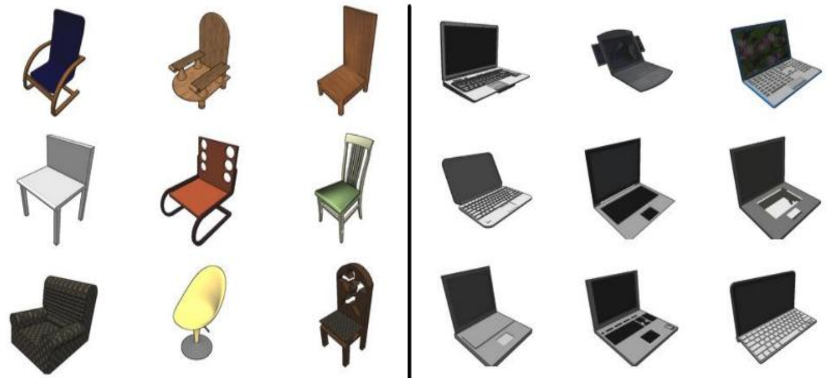
gruvi
{graphics} U {vision}

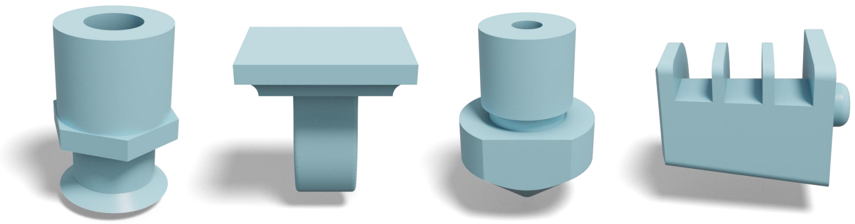
Learning Structured Representations of 3D CAD Models

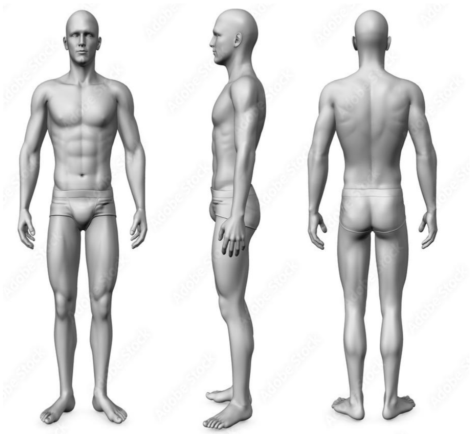
Fenggen Yu (余锋根)

Introduction

- What is Computer-Aided-Design (CAD) model?







3D human-made objects

3D organic models

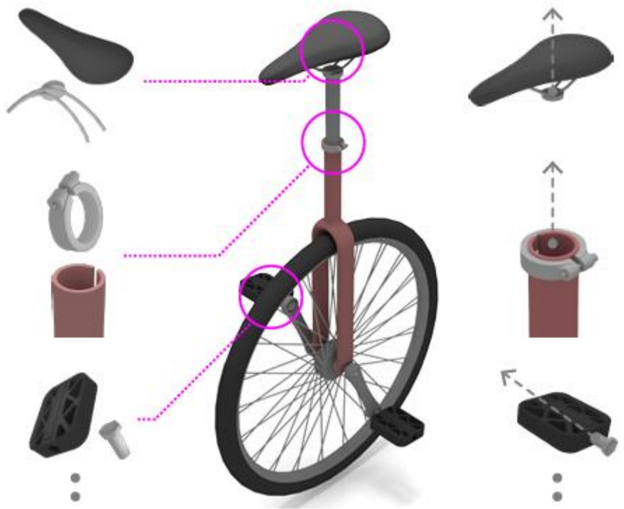
Complex structures, regular surfaces, and sharp edges Uniform structures with smooth surfaces and edges

Introduction

- Why do we need CAD models?

Video game, movie

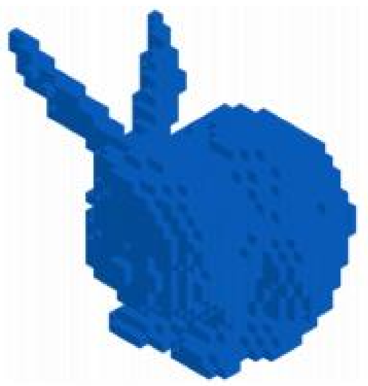
AR/VR, online e-commerce



Industrial design, manufacturing

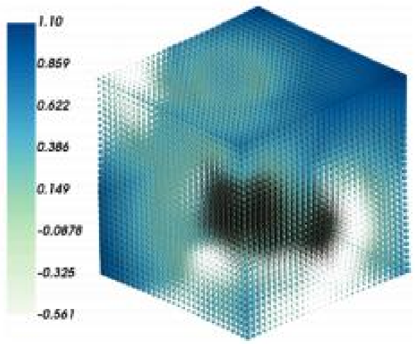
Introduction

- Unstructured 3D representations

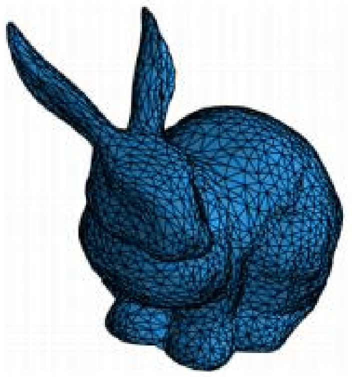


Voxel Grid

Point Cloud



Implicit Surface

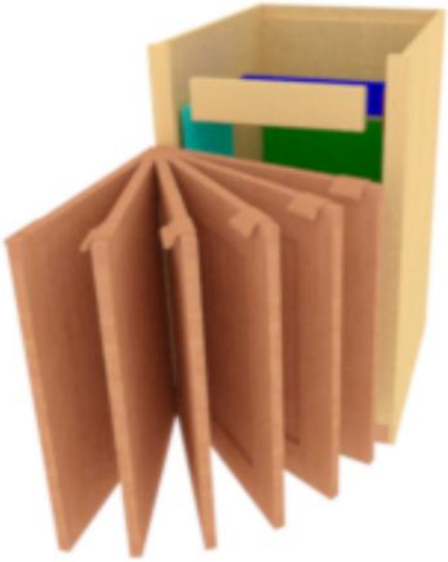


Triangle Mesh

Introduction

- Disadvantages of unstructured 3D representations

Imperfect shape surface
(Non-manifold, non-watertight)

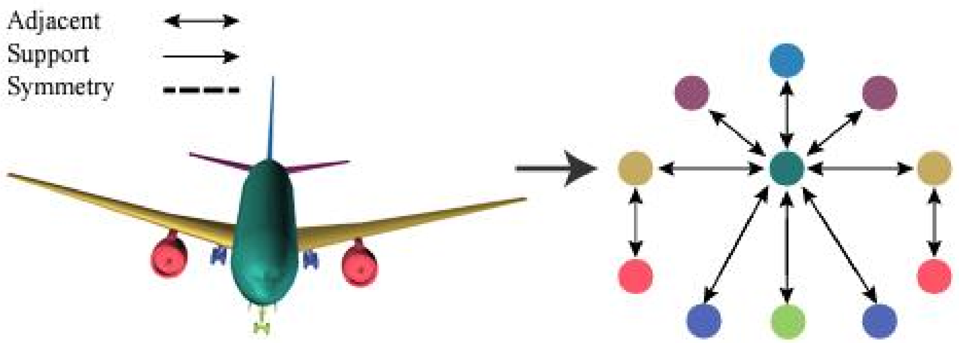


Not friendly for functionality and
semantic understanding

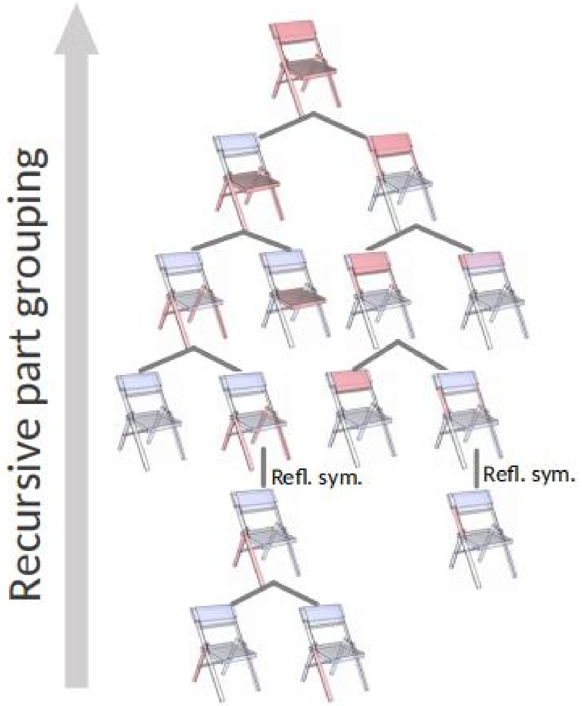
Not friendly for shape editing
(Additional segmentation needed)

Introduction

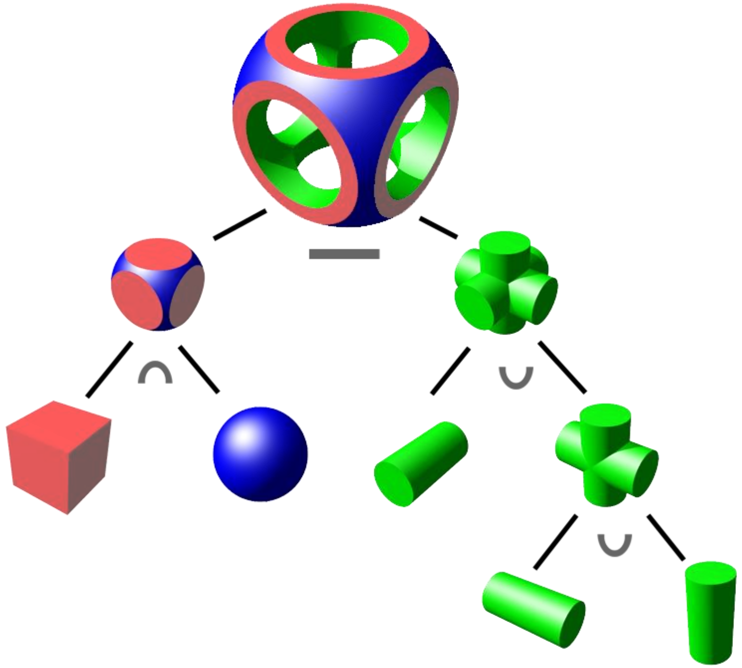
- Structured 3D representations: atomic elements and assembly patterns.



Graph [1]



Hierarchy-tree [2]

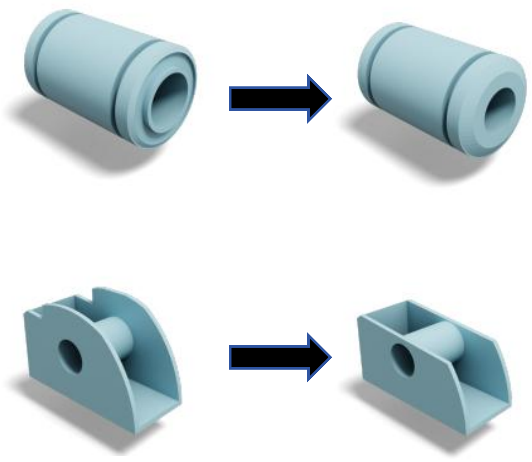


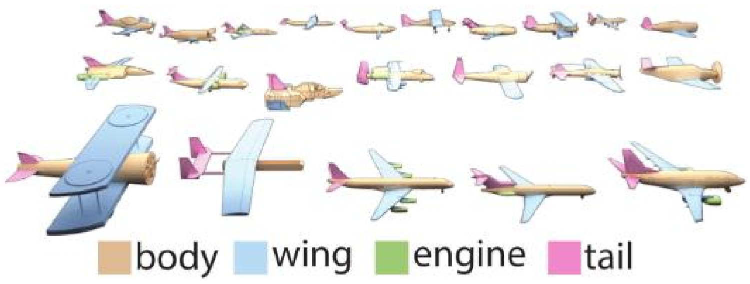
CSG-tree

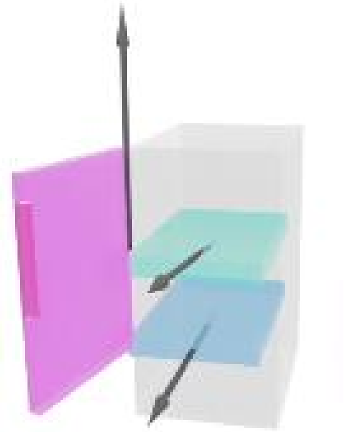
[1] SDM-NET: Deep Generative Network for Structured Deformable Mesh
[2] GRASS: Generative Recursive Autoencoders for Shape Structures

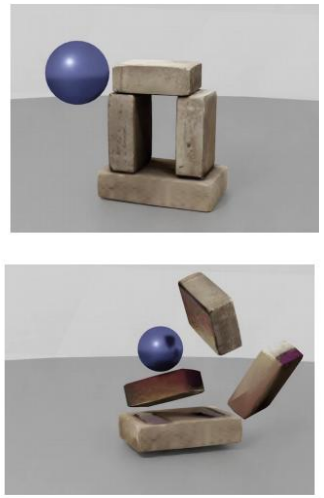
Introduction

- Advantages of structured 3D representations









Shape editing/manipulation

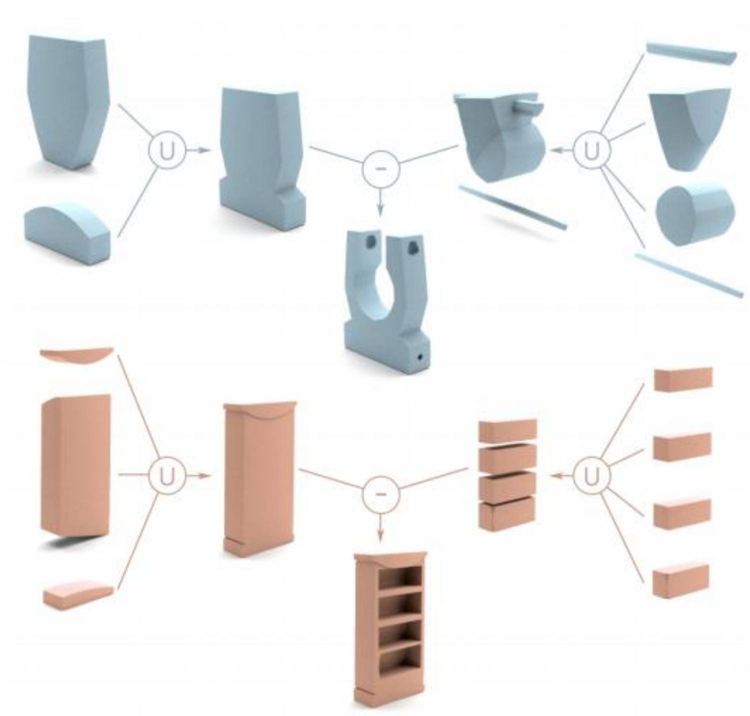
Semantic understanding

Part functionality annotation

Physics animation

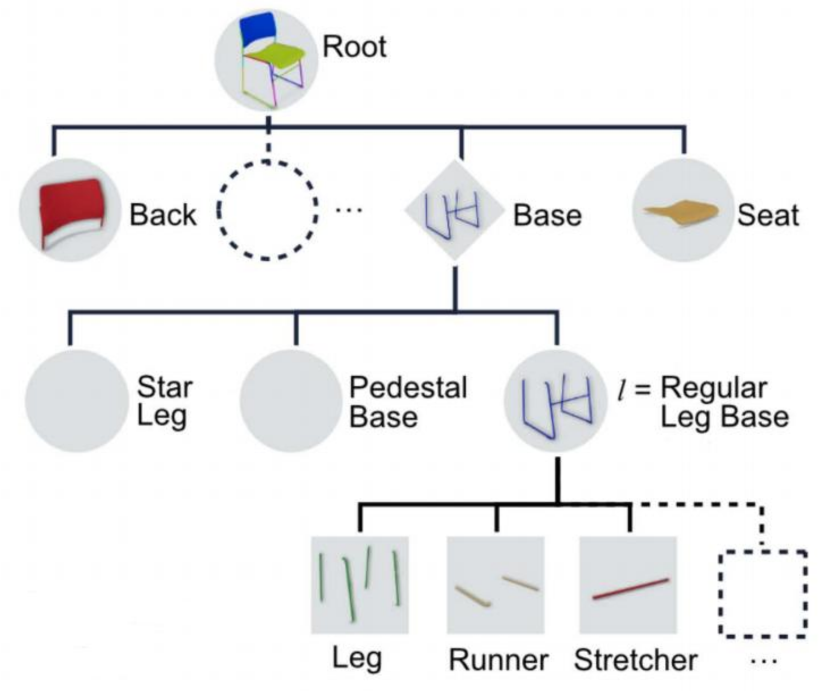
Introduction

- Talk topic: learning structured representations of 3D CAD models



Constructive Solid Geometry (CSG)

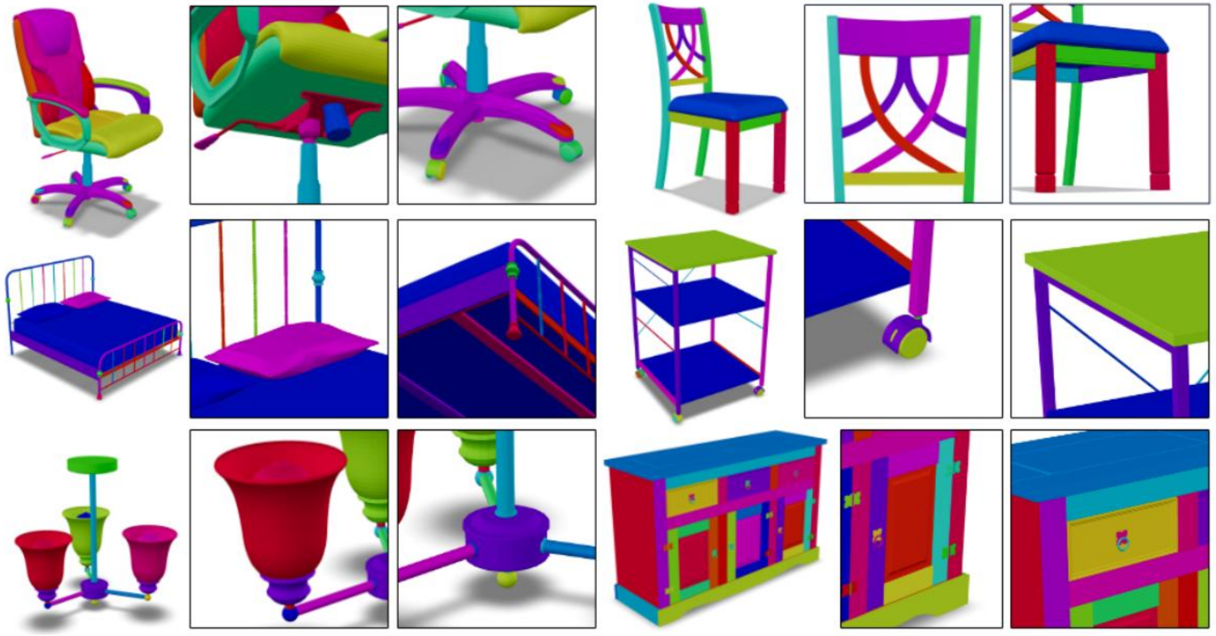
Part Set



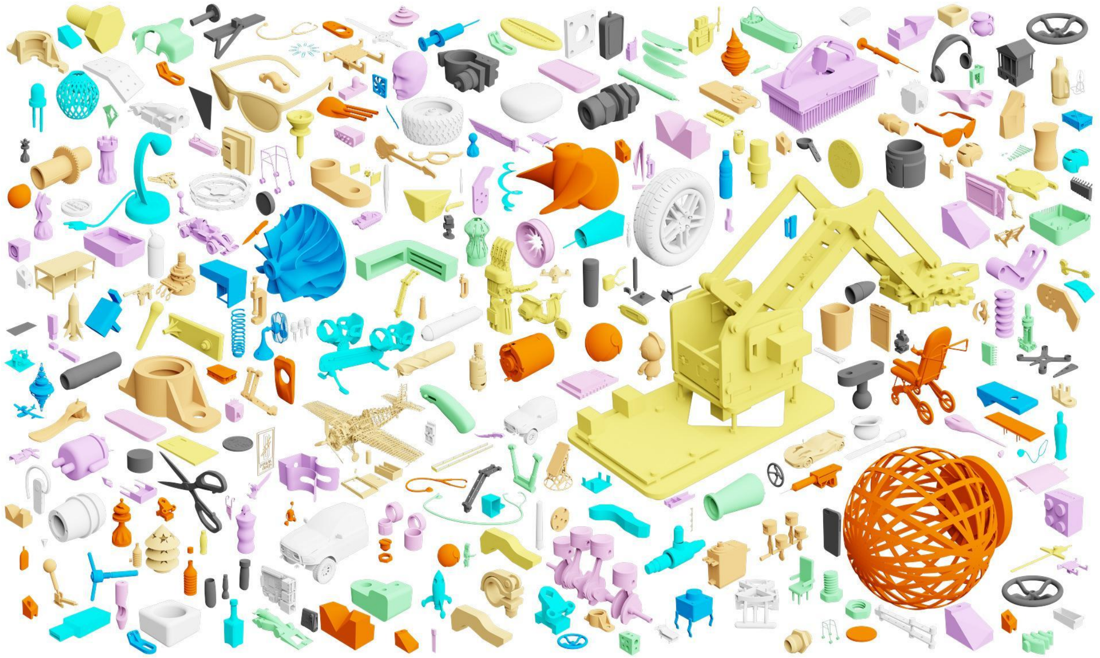
Hierarchical tree

Introduction

- Challenge: intricate 3D CAD model structure



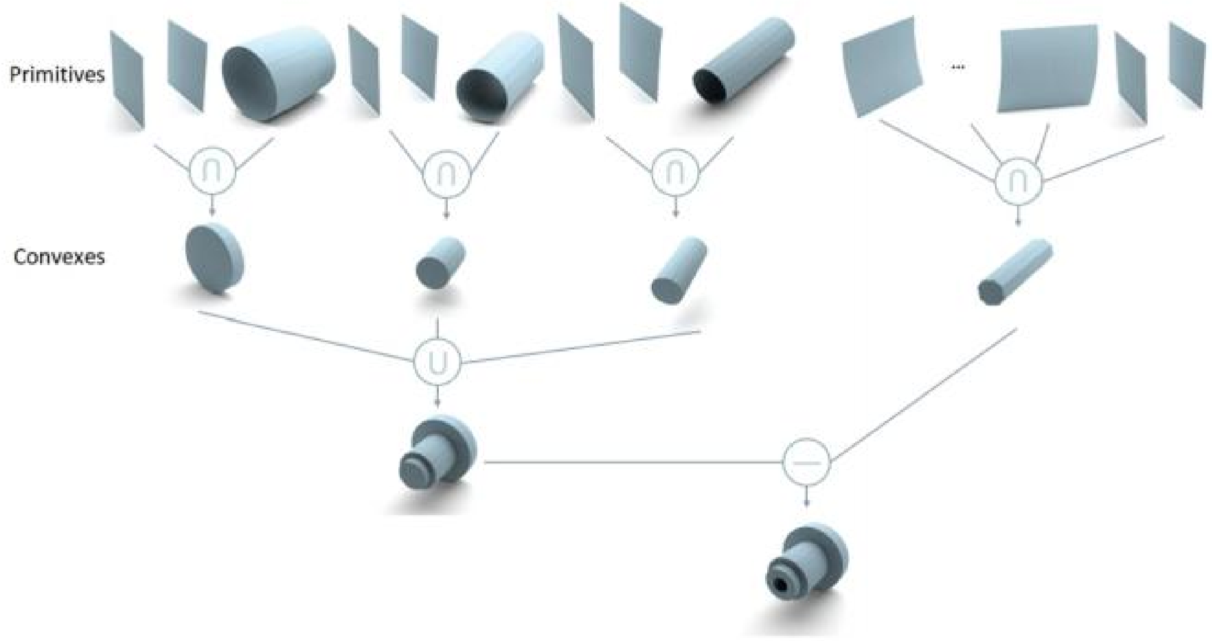
Small and different number of parts



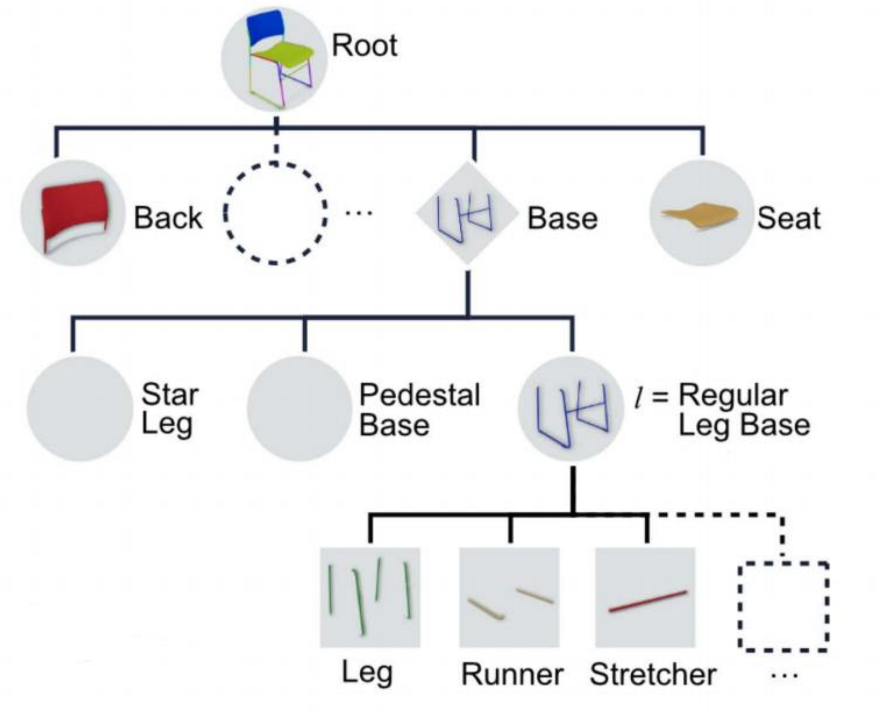
3D CAD components: complex and various topologies

Introduction

- Challenge: intricate 3D CAD model structure
- Our solution: hierarchical learning strategy to reduce learning complexity



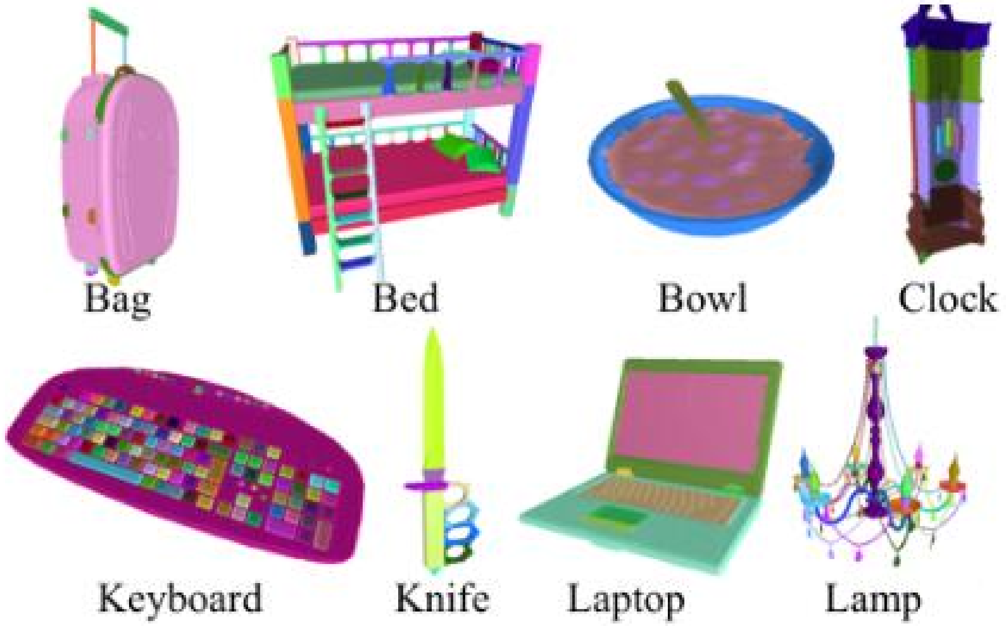
Hierarchical CSG tree structure



Hierarchical semantic tree structure

Introduction

- Challenge: limited training data in structured representations

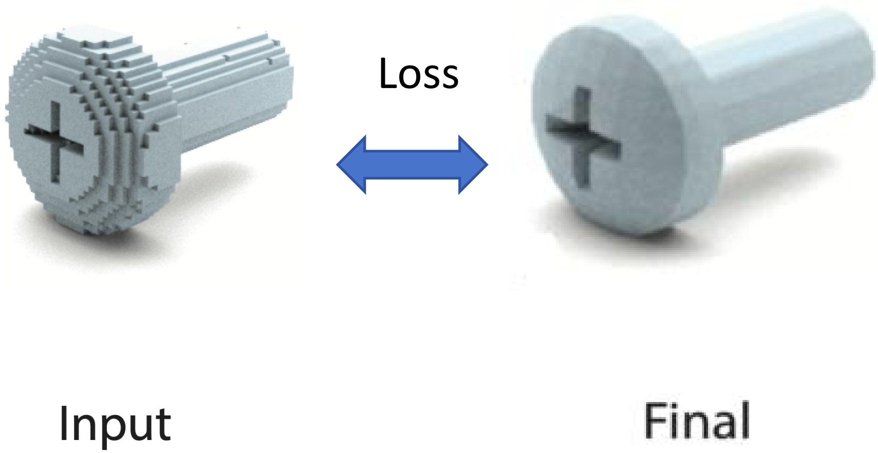


PartNet: less than 30K, most of categories have less than 1K shapes

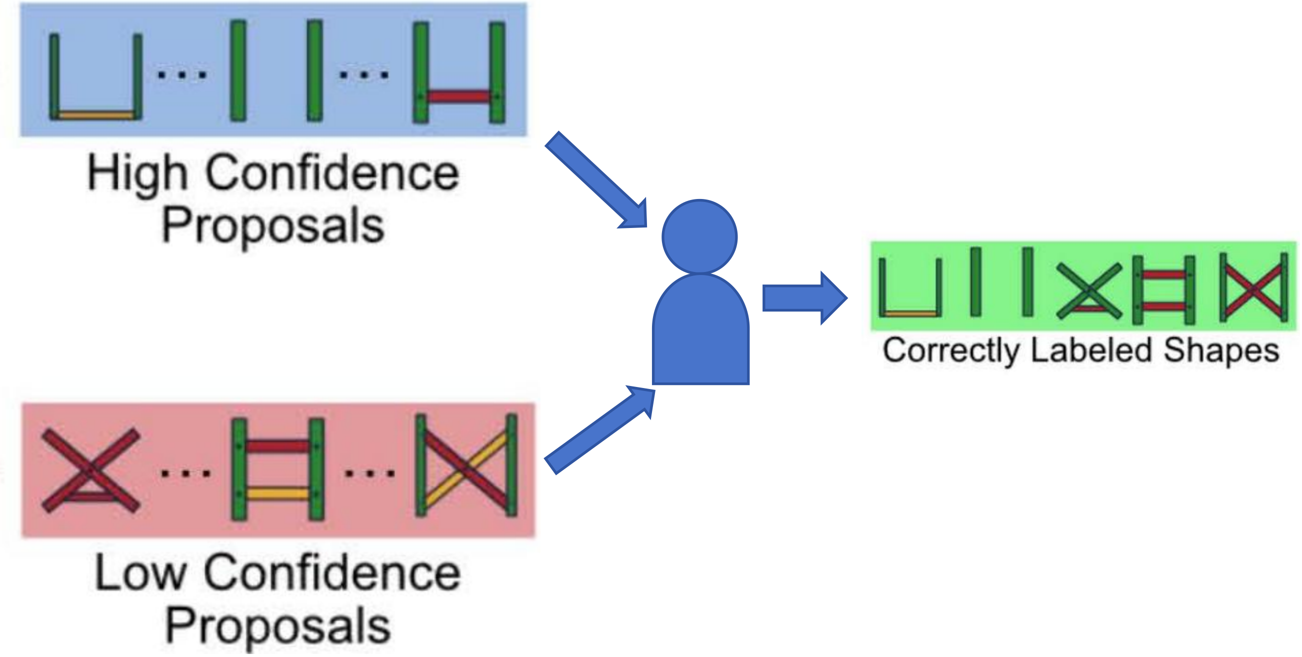
Objverse-XL: no category or part-level annotation

Introduction

- Challenge: limited training data in structured representations
- Our strategy: unsupervised learning and active learning



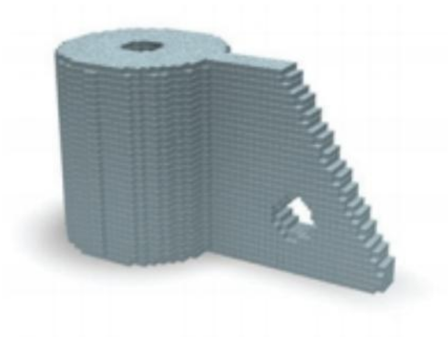
Unsupervised learning



Active learning

Introduction

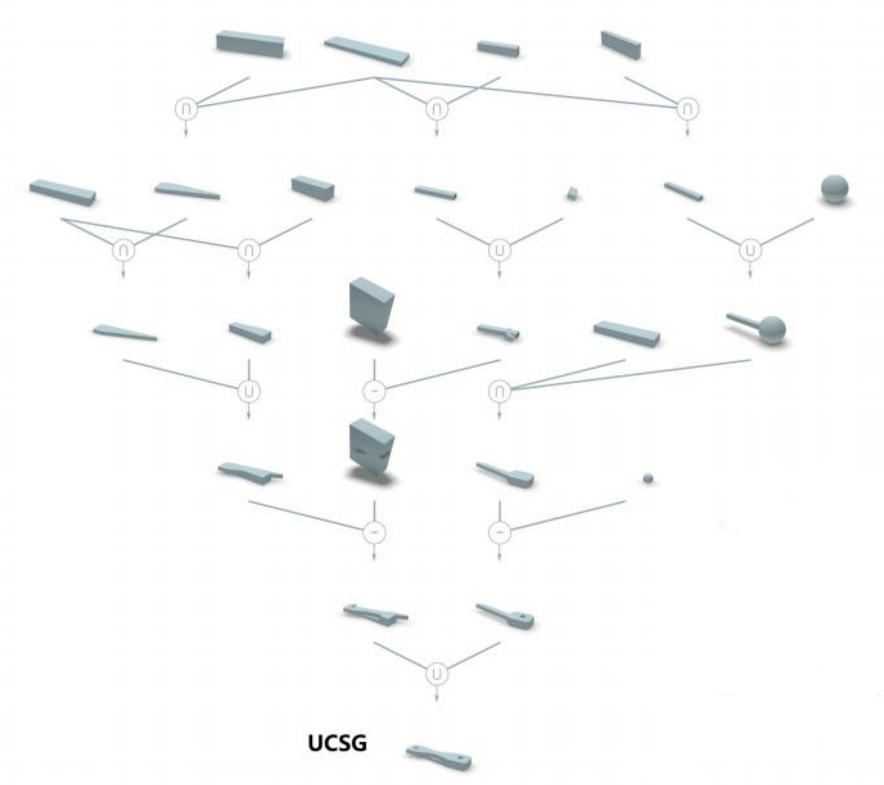
- Challenge: reconstructing CSG Representation by the neural network



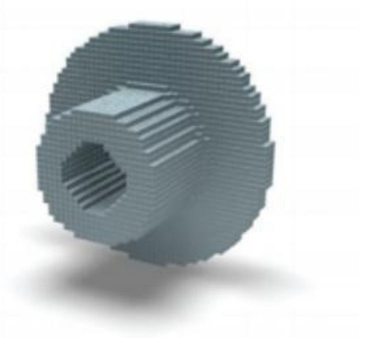
Input

BSP-Net

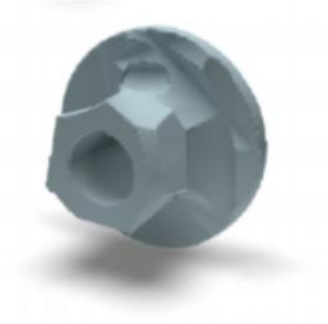
Generalization challenge



Learning compact and meaningful structure is challenging



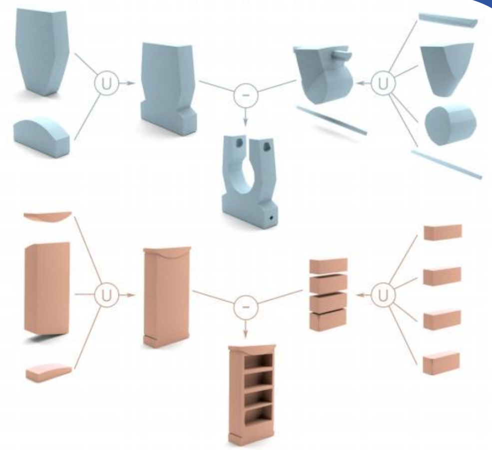
Input



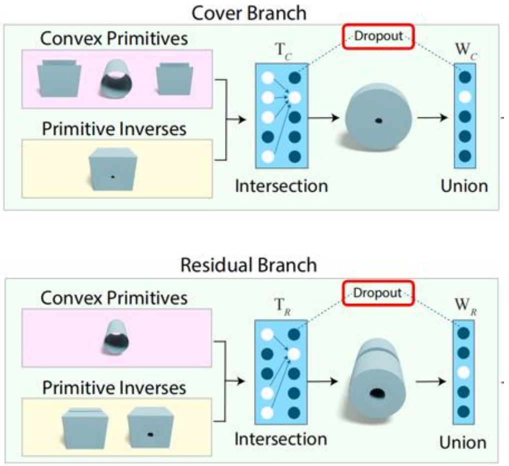
CSG-Stump

Various primitives and operations

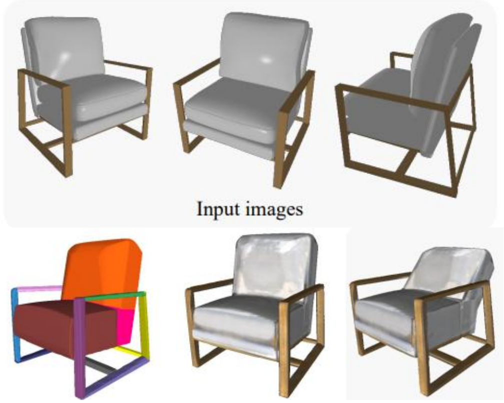
This Talk: Learning Structured 3D Representations



CAPRI-Net (CVPR 2022)

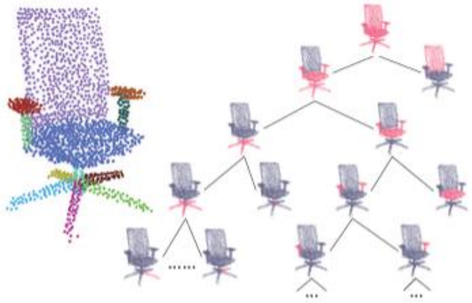


D^{2}CSG (NeurIPS 2023)

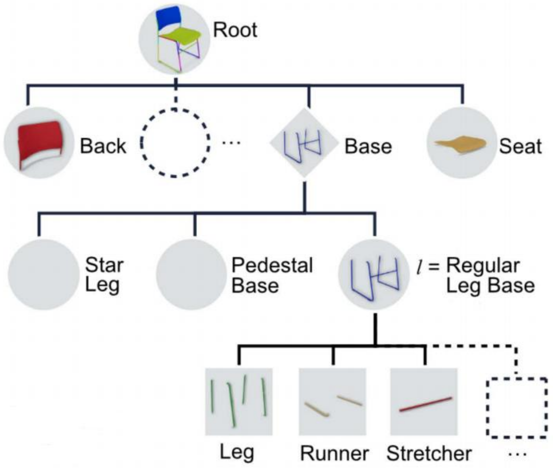


DPA-Net (ECCV 2024)

Sweep-Net (ECCV 2024)

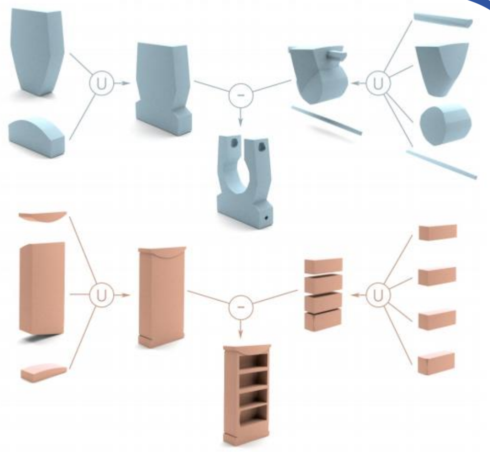


PartNet (CVPR 2019)

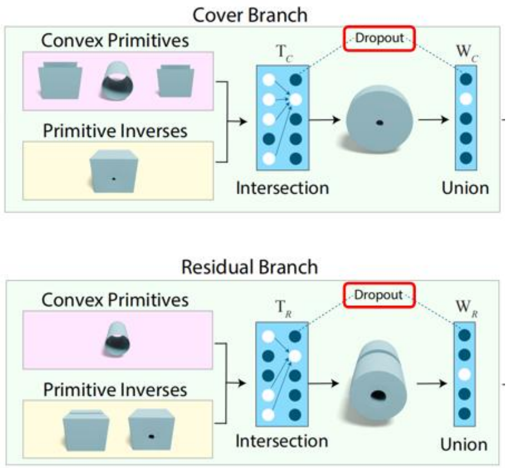


HAL3D (ICCV 2023)

This Talk: Learning Structured 3D Representations



CAPRI-Net (CVPR 2022)



D2CSG (NeurIPS 2023)

DPA-Net (ECCV 2024)

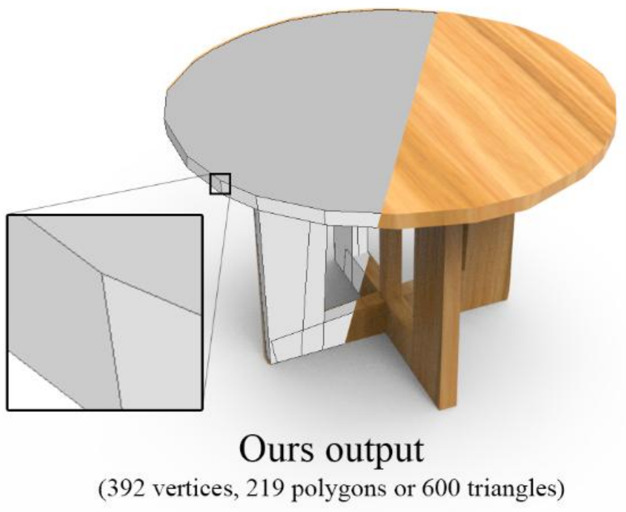
Sweep-Net (ECCV 2024)

PartNet (CVPR 2019)

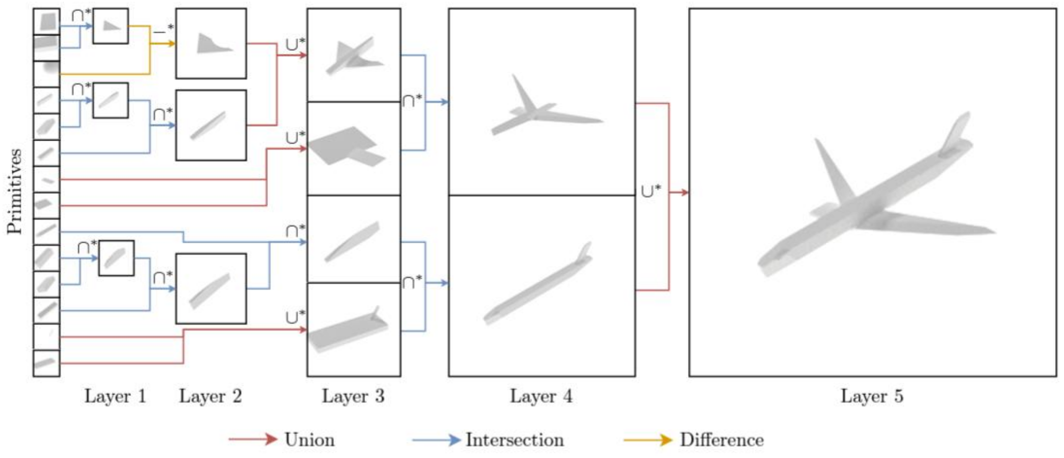
HAL3D (ICCV 2023)

CAPRI-Net: Learning Primitive Assembly for 3D CAD Models

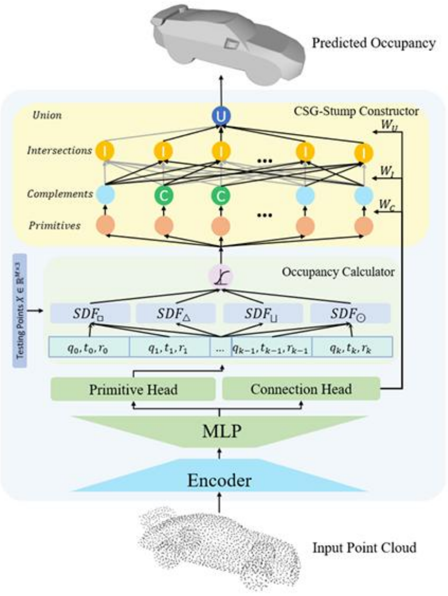
- Related works: unsupervised learning CSG representation



BSP-Net (CVPR 2020)

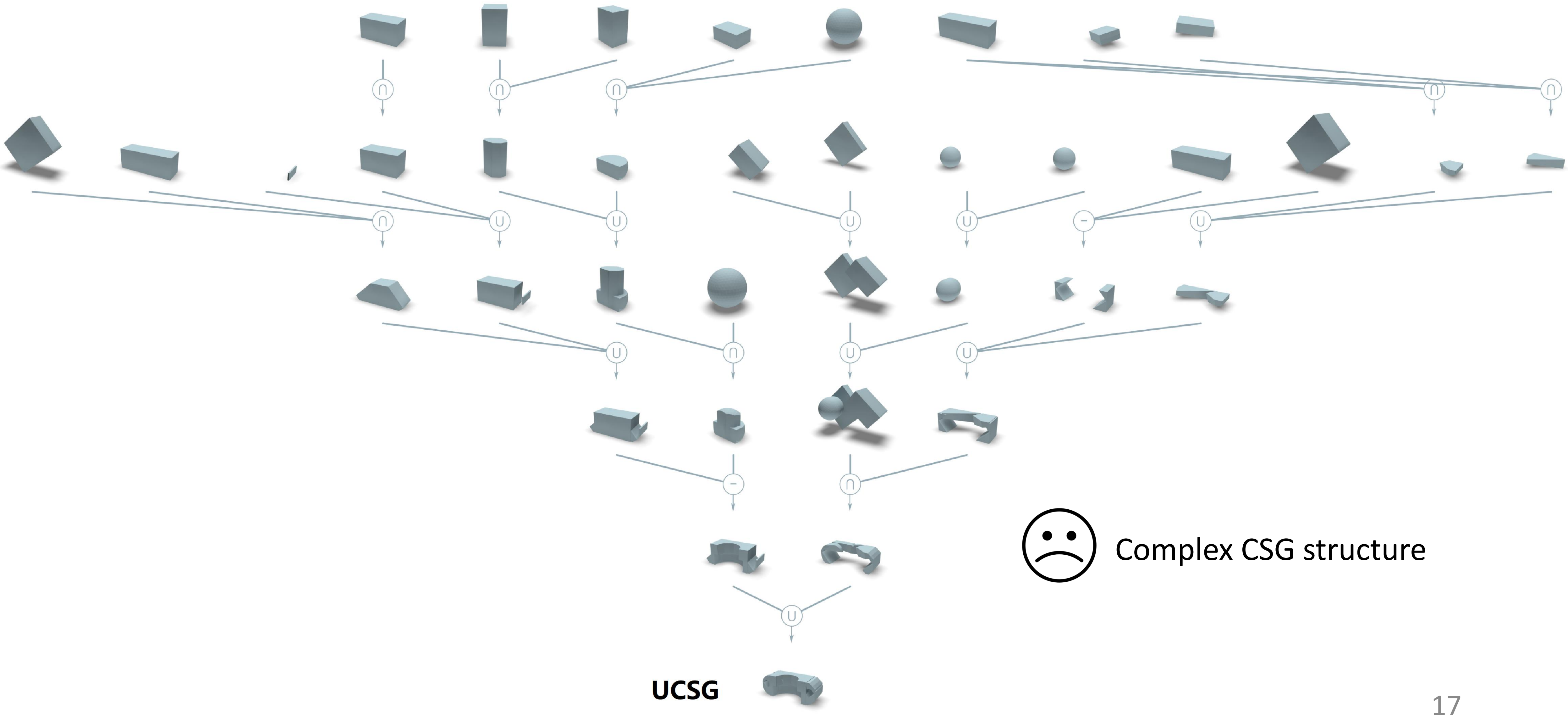


UCSG (NeurIPS 2020)

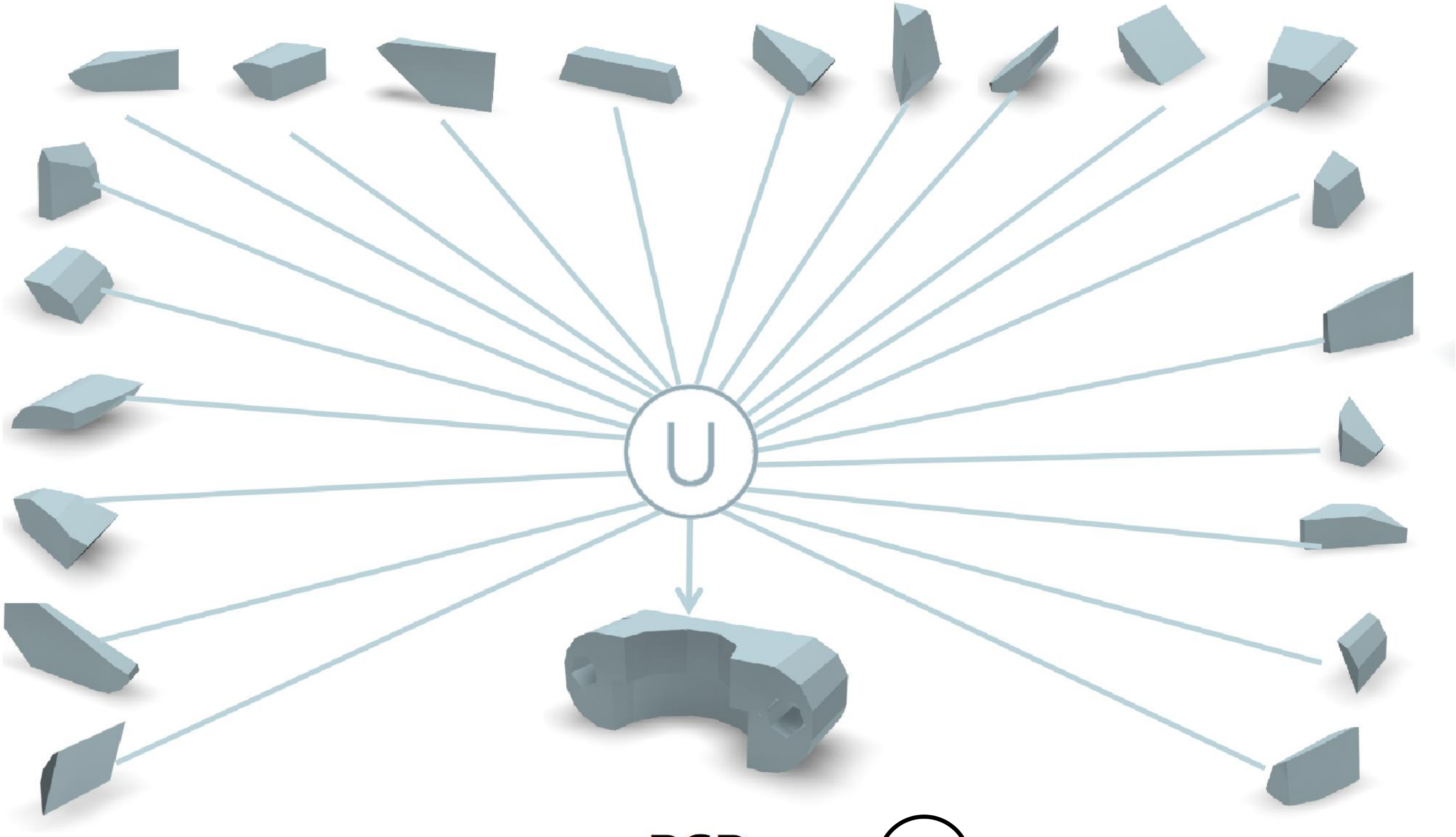


CSG-Stump (ICCV 2021)

CSG Tree Comparison

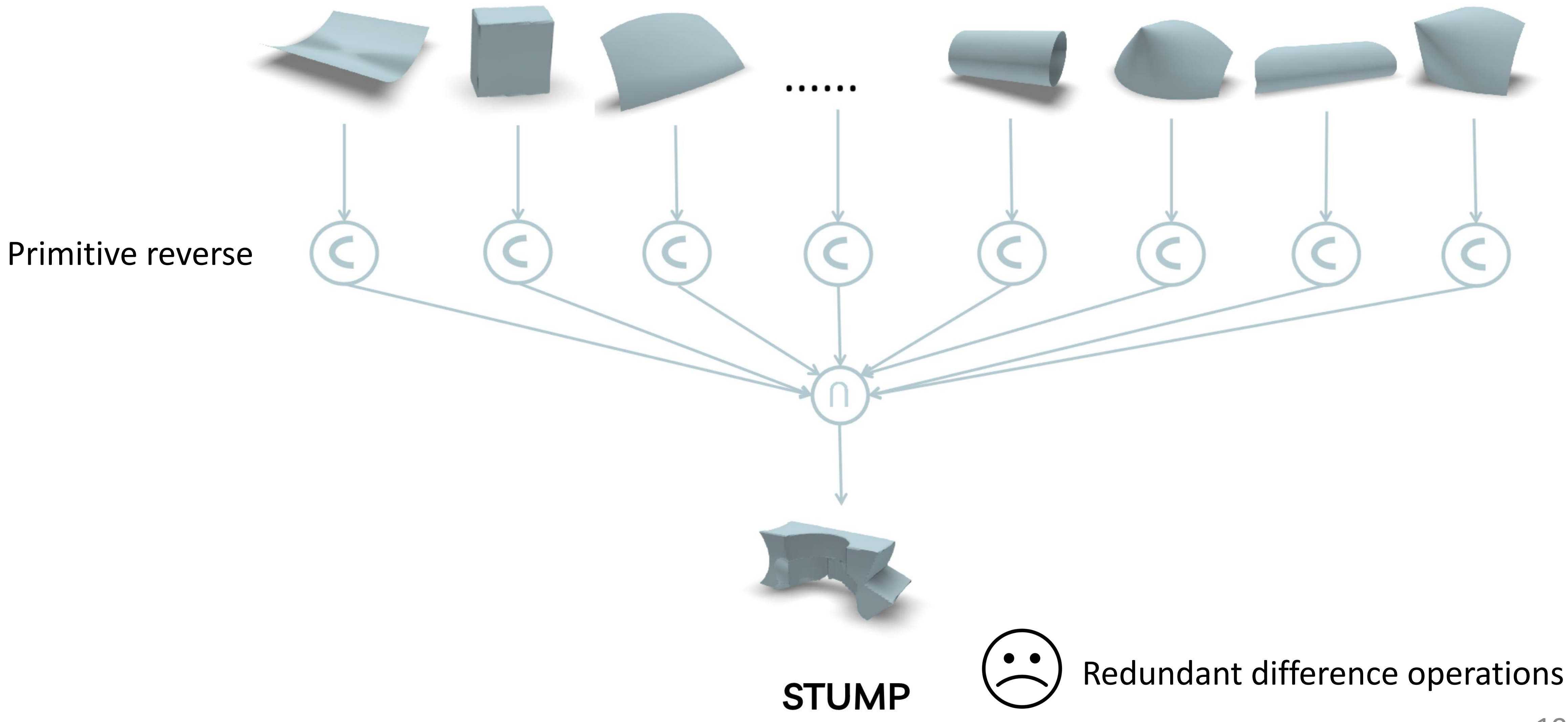


Complex CSG structure

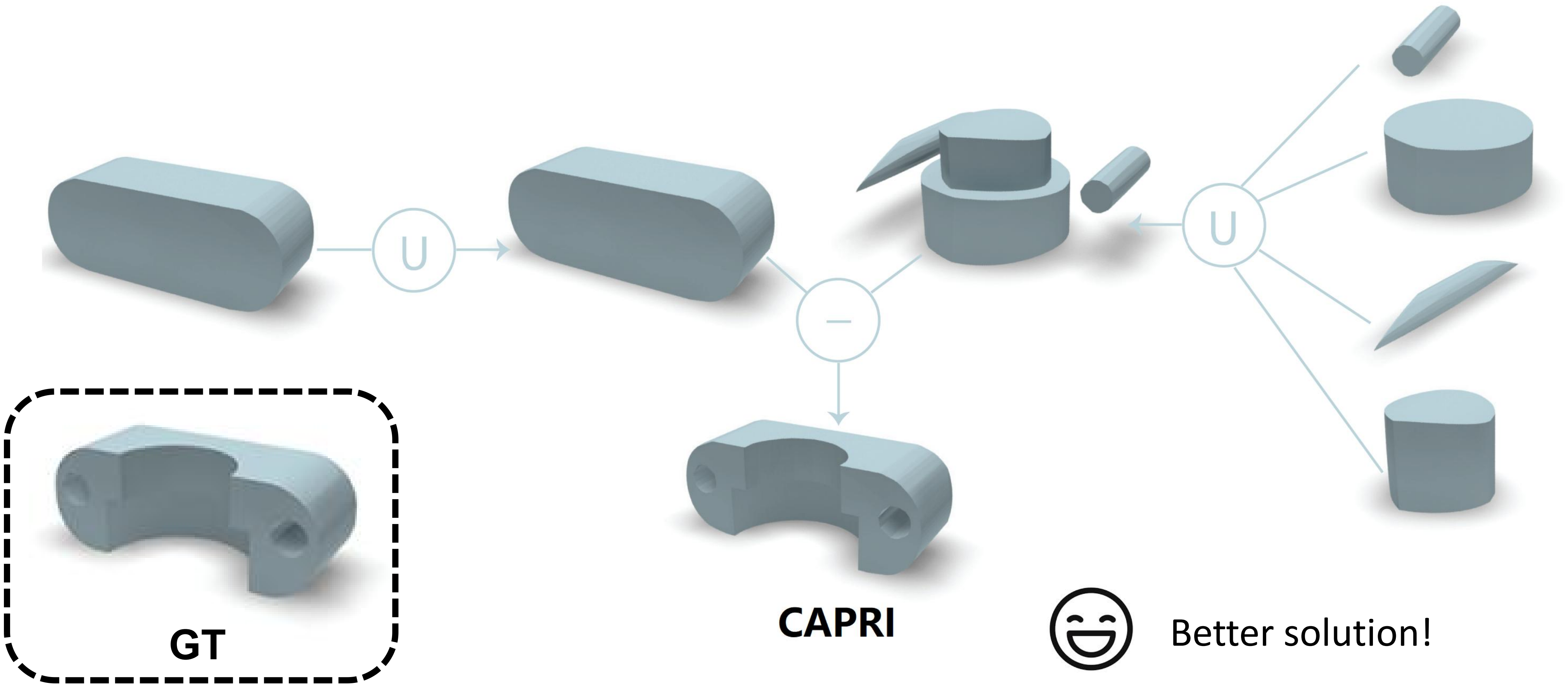


BSP

Incompact solution

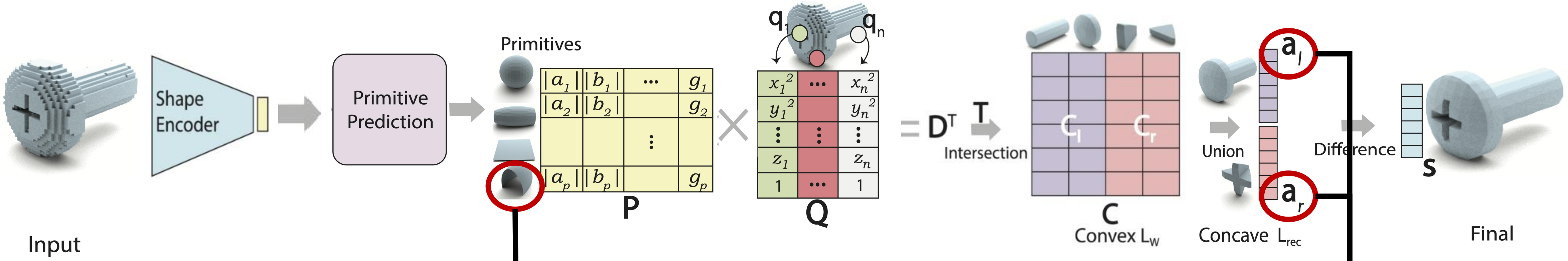


Primitive reverse
......
STUMP
Redundant difference operations



Better solution!

Method

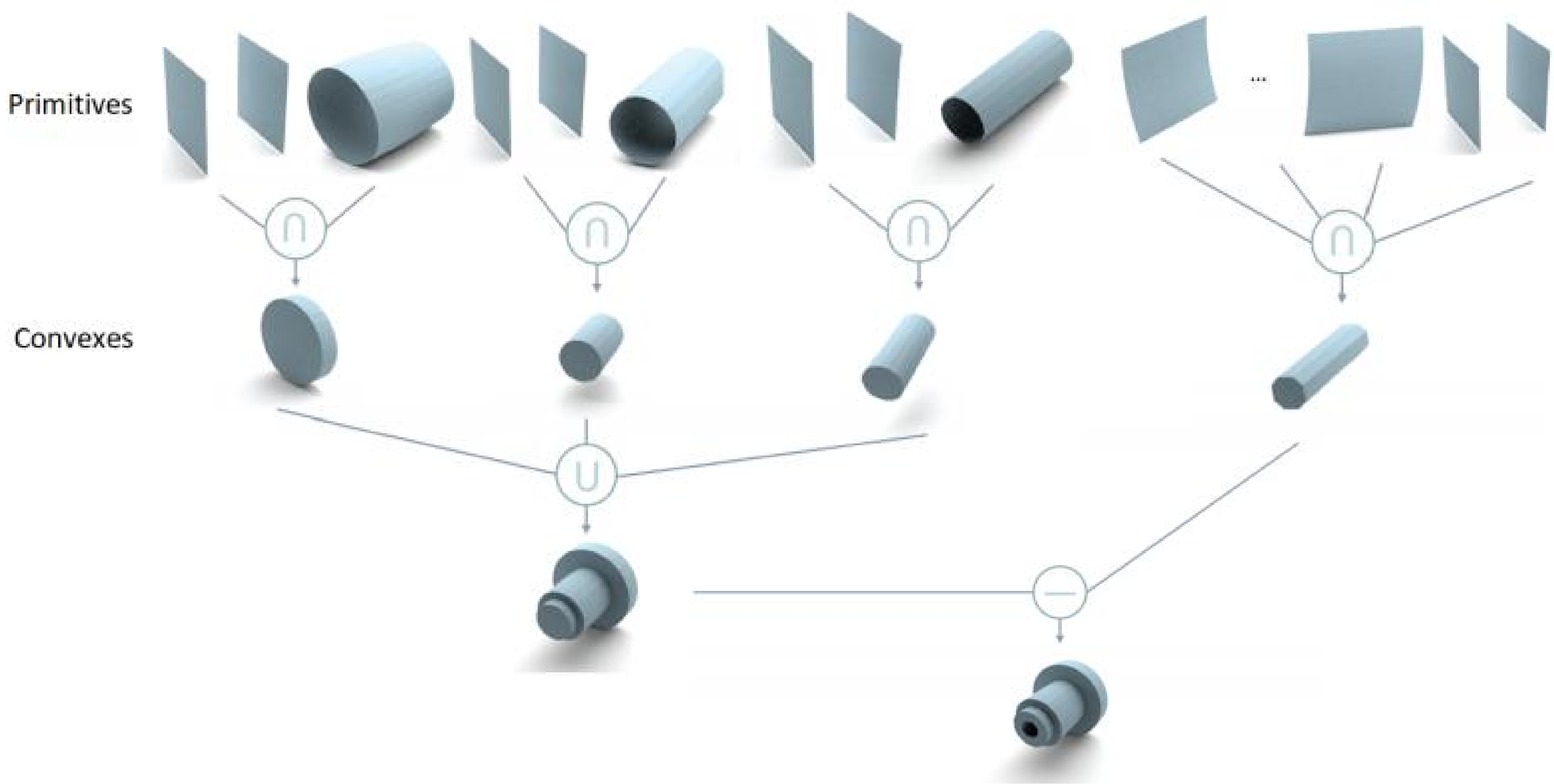


Quadric implicit representation

$|a_p|x^2+|b_p|y^2+|c_p|z^2+d_px+e_py+f_pz+g_p=0$

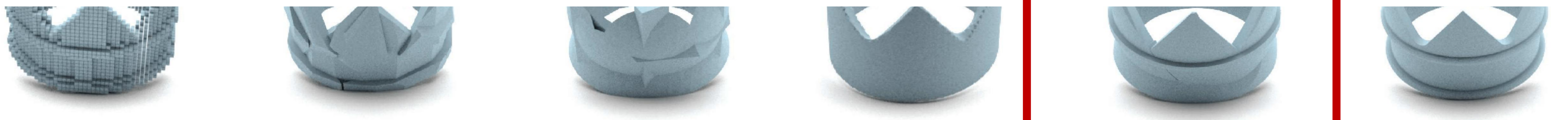
Weighted implicit reconstruction loss

CAD Model Meshing Process



ABC Model Reconstruction From Voxels

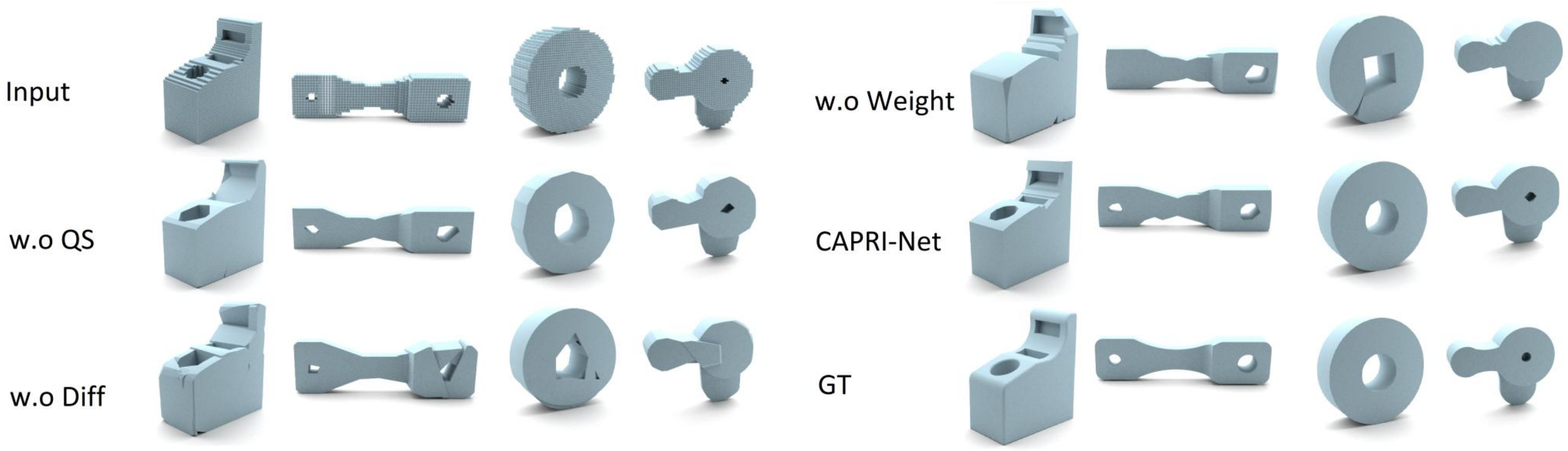
Methods	BSP-Net	UCSG	STUMP	Ours
CD ↓	0.491	0.300	1.180	**0.136**
NC ↑	0.868	0.877	0.829	**0.914**
ECD ↓	10.098	5.022	11.848	**2.208**
LFD ↓	1,342.7	1,494.8	2945.2	**800.2**
#Primitives (#P) ↓	114.44	-	-	**46.93**
#Convexes (#C) ↓	11.60	12.72	90.88	**6.03**



ShapeNet Model Reconstruction From Voxels

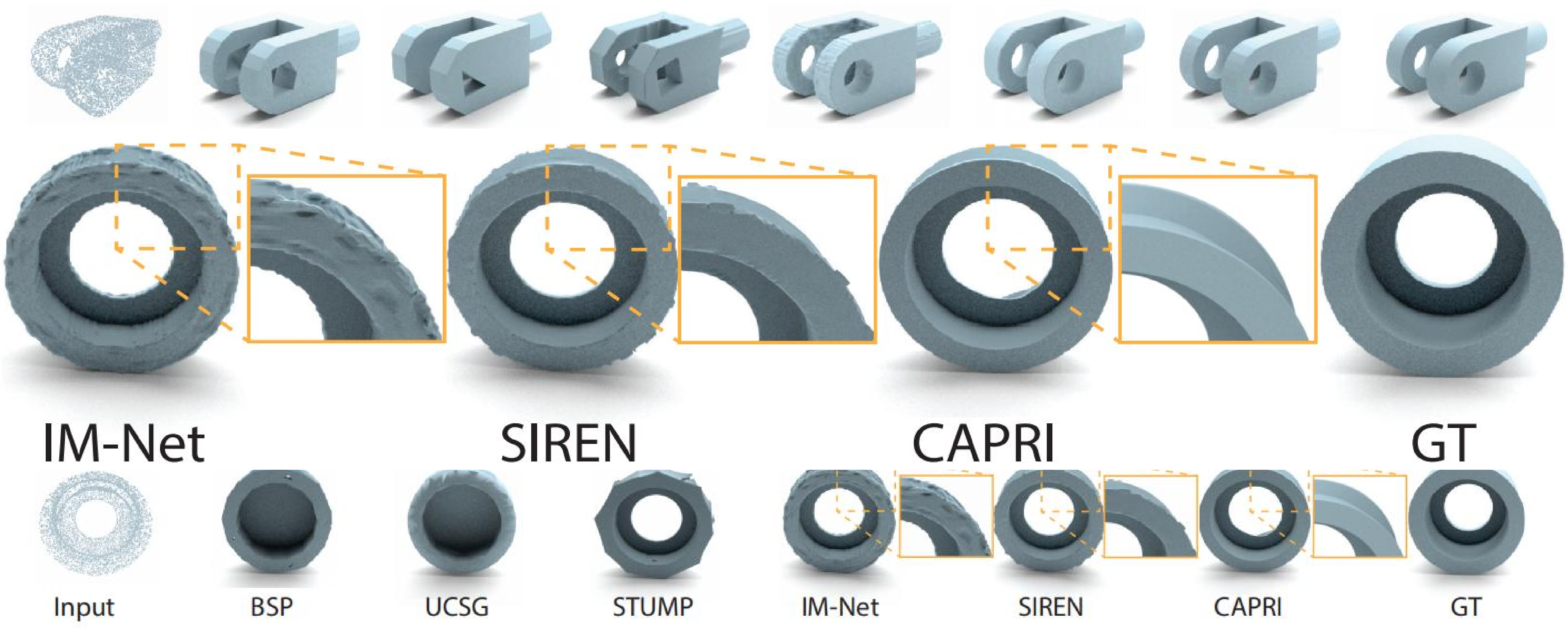
Methods	BSP-Net	UCSG	STUMP	Ours
CD ↓	0.220	1.317	2.288	**0.175**
NC ↑	0.869	0.815	0.792	**0.872**
ECD ↓	2.111	5.233	10.457	**2.101**
LFD ↓	2,254.4	3,582.5	5217.0	**1,824.1**
#Primitives (#P) ↓	214.70	-	-	**61.56**
#Convexes (#C) ↓	18.86	12.40	180.54	**8.71**

Ablation Studies



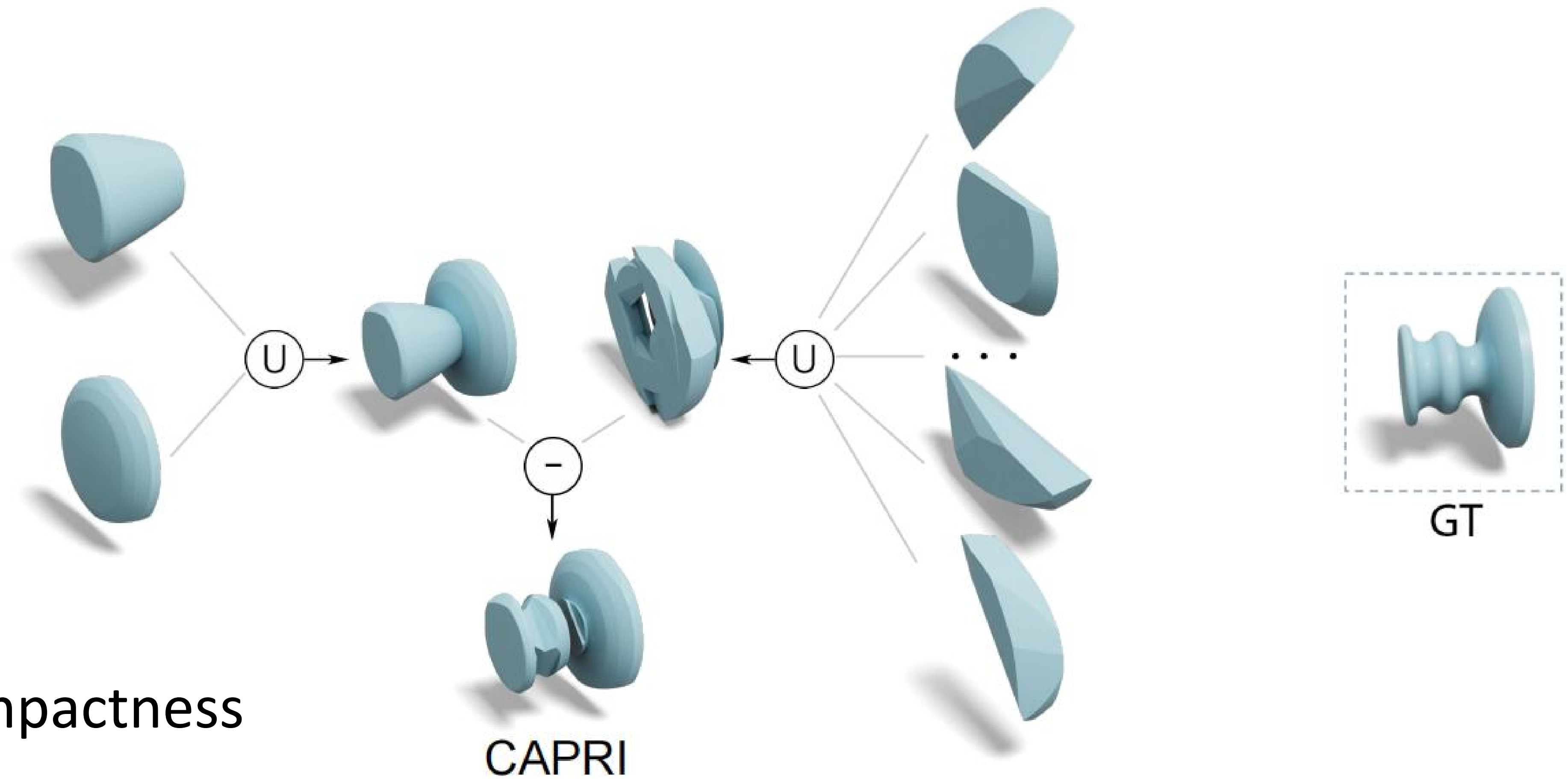
QS: quadric surface   Diff: difference operation   Weight: weighted implicit reconstruction loss

ABC Model Reconstruction From Point Clouds



Limitation of CAPRI-Net

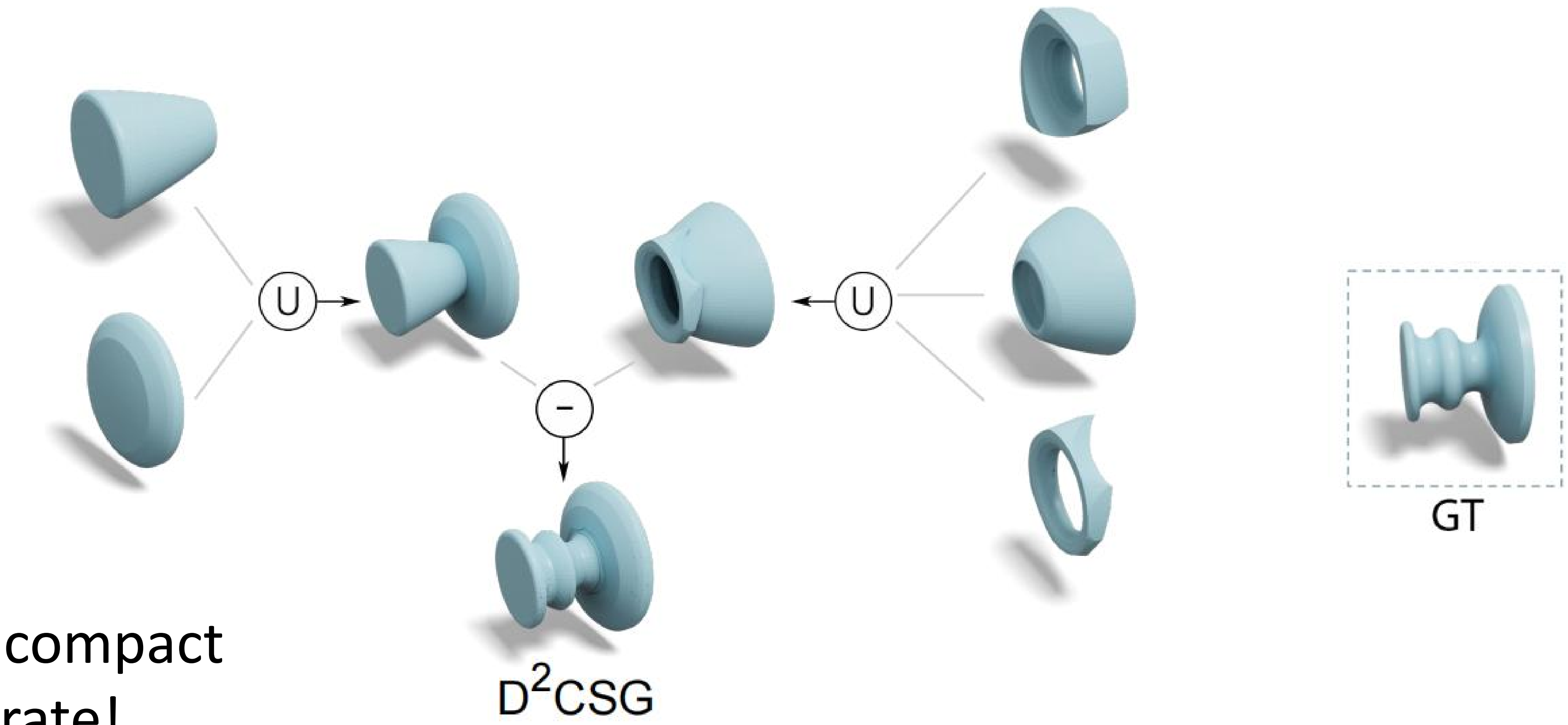
Convex shapes only



☹ Not general

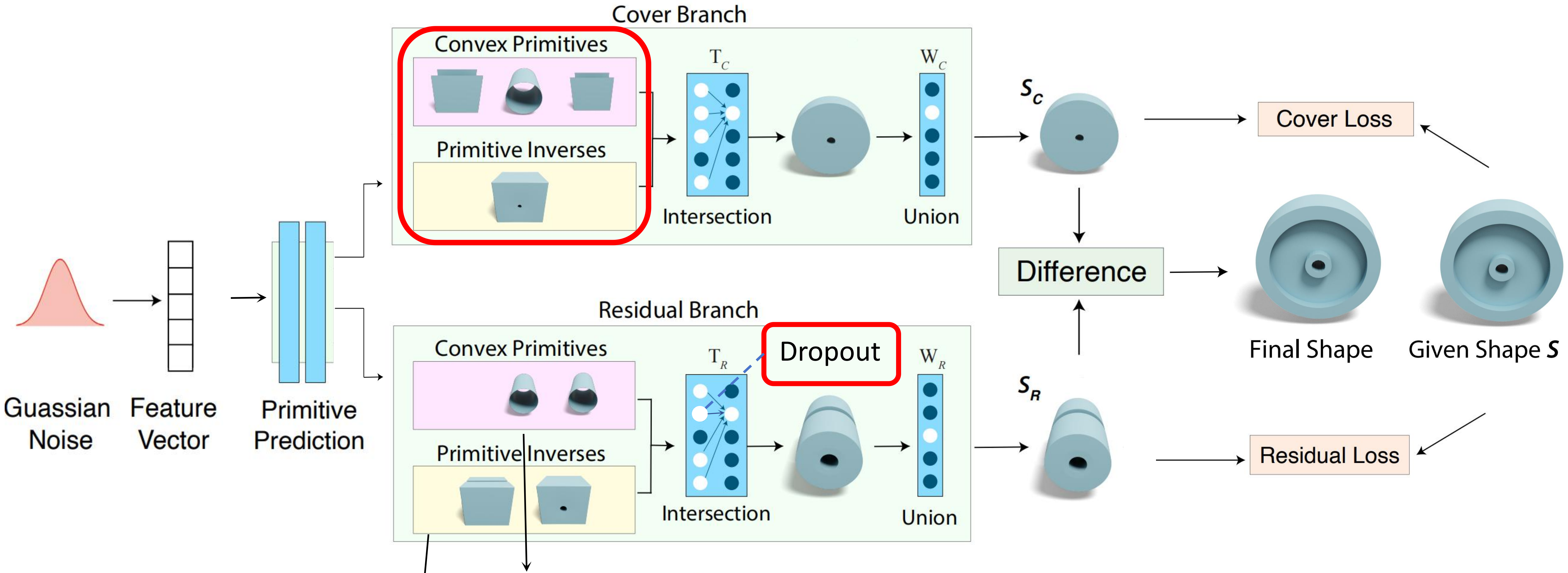
☹ No explicit compactness constraint

D²CSG: Unsupervised Learning of Compact CSG Trees with Dual Complements and Dropouts



☺ General, compact and accurate!

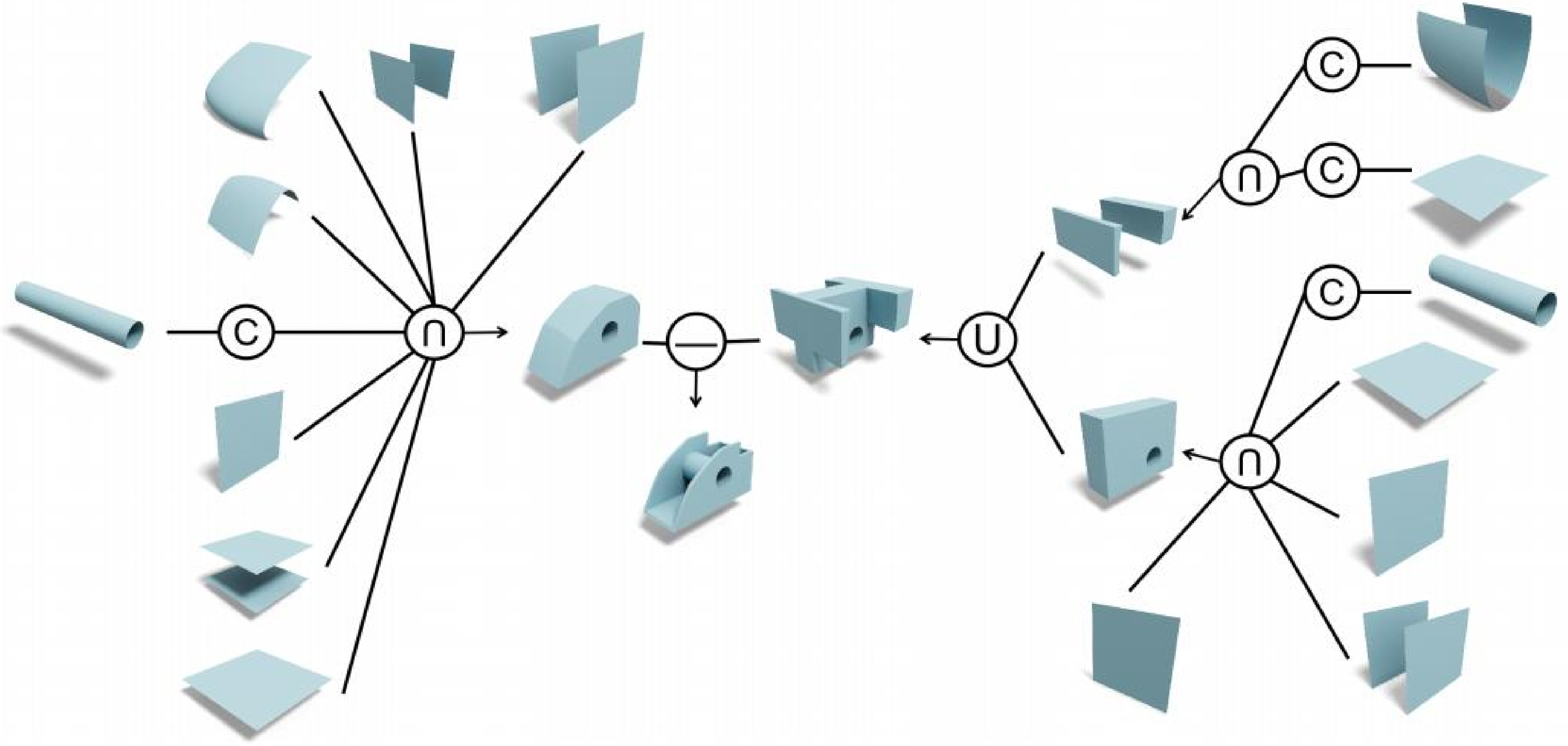
Network Overview



$$|a|x^2 + |b|y^2 + |c|z^2 + dx + ey + fz + g = 0$$

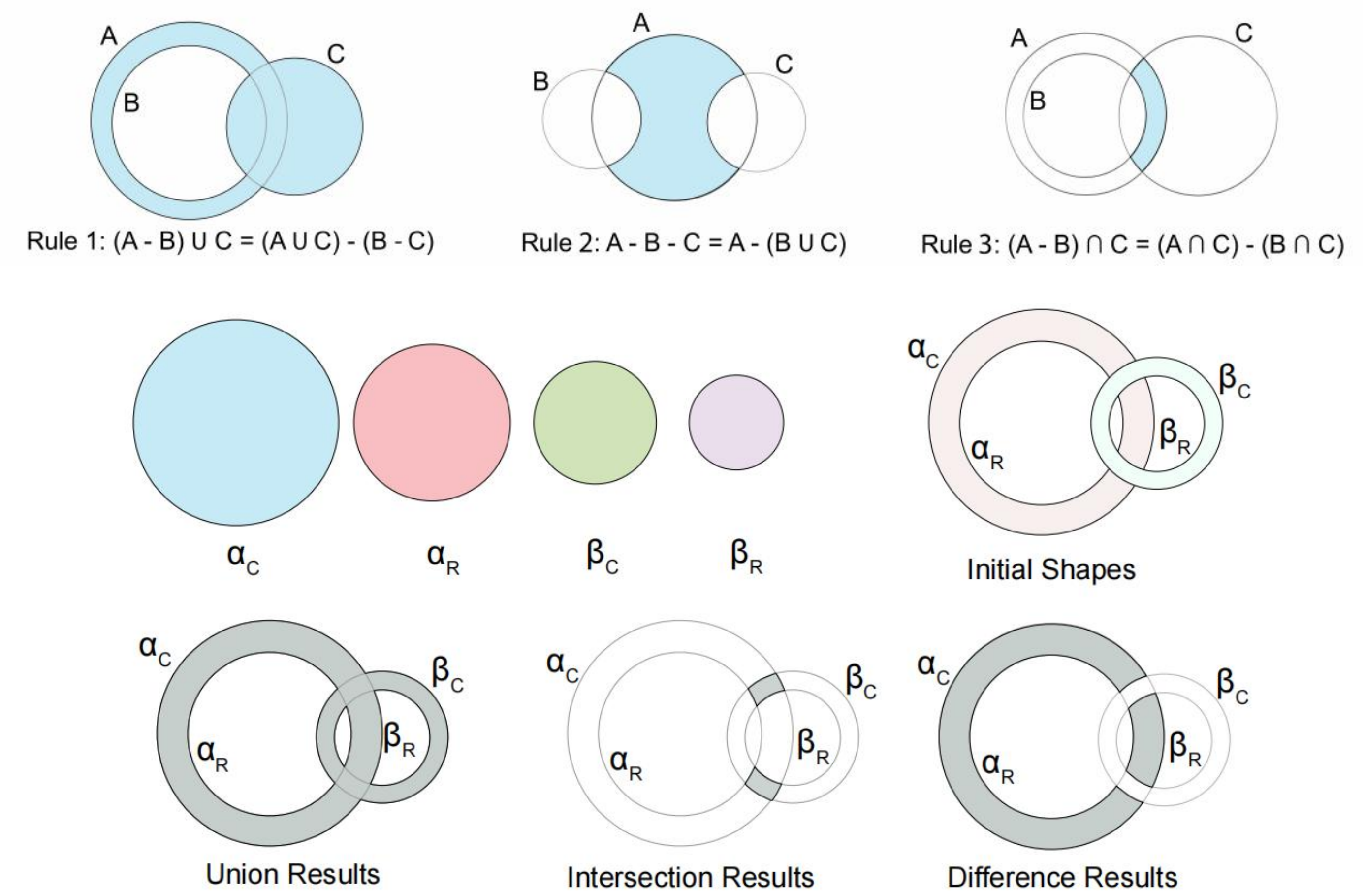
$$-|a|x^2 - |b|y^2 - |c|z^2 + dx + ey + fz + g = 0$$

Learned CSG Tree

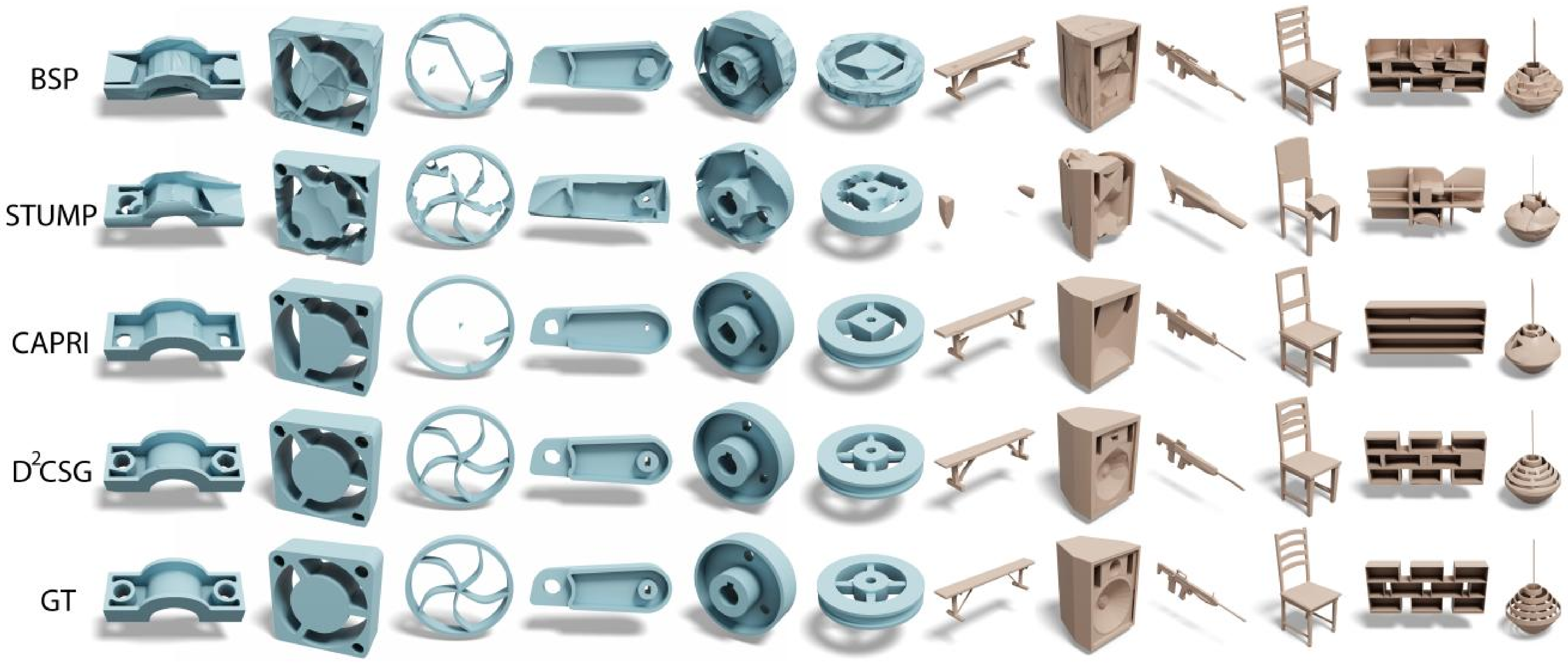


Generalization Proof

- The operation sequence in D²CSG is able to support any CSG sequence



Experiments: Mesh-to-CSG



ABC DataSet

ShapeNet

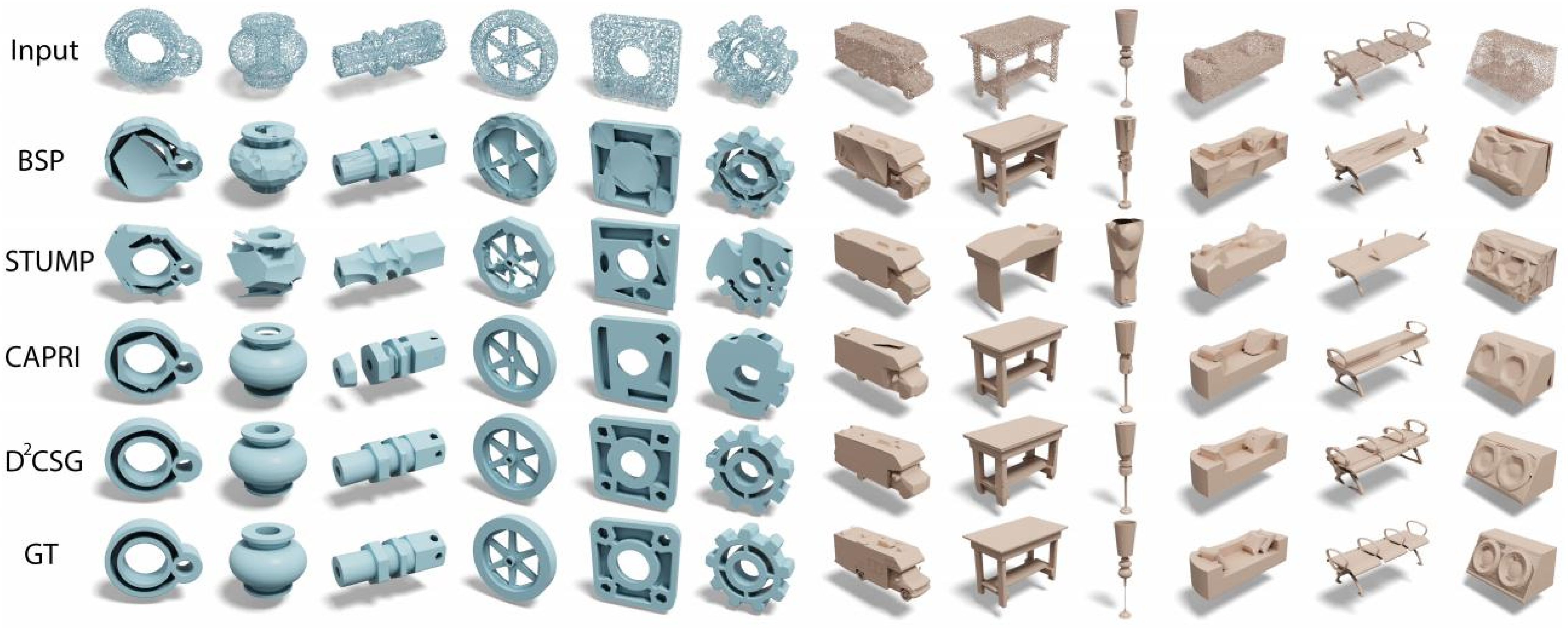
Experiments: Ablation Studies

Row ID	CP	DB	DO	CD ↓	NC ↑	ECD ↓	#P ↓	#IS ↓
1	-	-	-	0.183	0.907	3.92	77	9.2
2	-	✓	-	0.114	0.918	2.97	37	10.5
3	-	✓	✓	0.127	0.914	3.56	32	10.0
4	✓	-	-	0.073	0.935	3.12	38	5.8
5	✓	-	✓	0.088	0.926	3.48	**27**	**5.3**
6	✓	✓	-	**0.069**	**0.936**	**2.98**	53	6.8
7	✓	✓	✓	**0.069**	0.928	3.09	29	5.7

w.o CP w.o DB w.o DP Ours GT

CP: Complementary primitives DB: Dual branches DP: Dropout

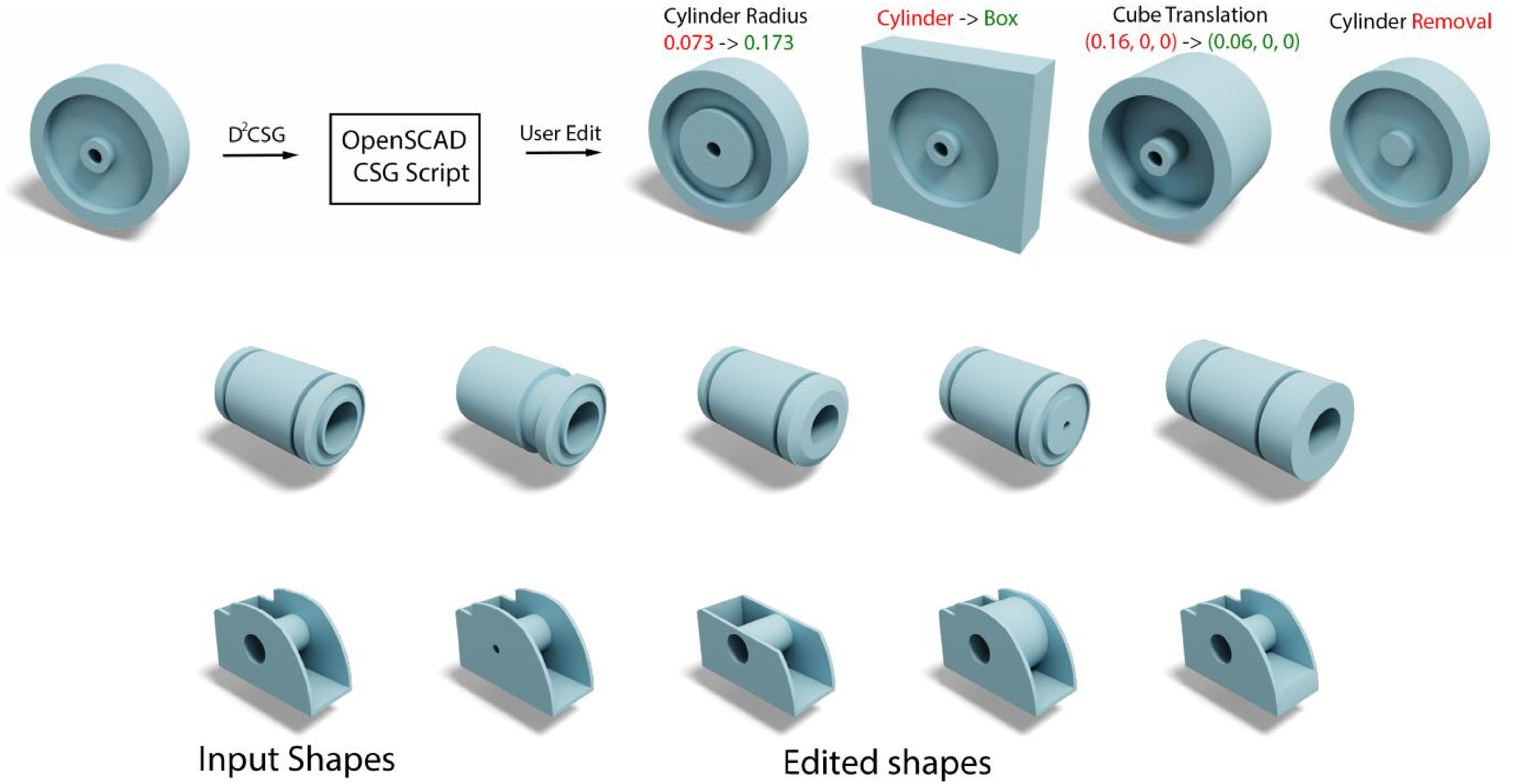
Application: PointCloud-to-CSG



ABC DataSet

ShapeNet

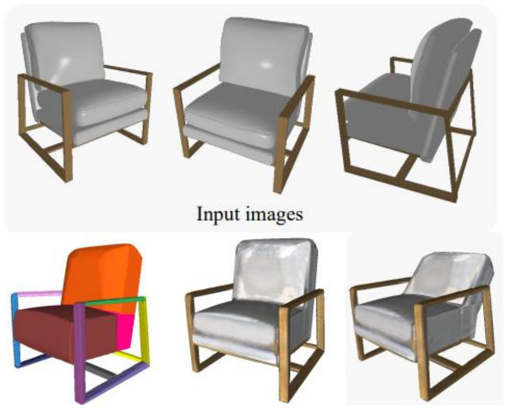
Application: Shape Editing



This Talk: Learning Structured 3D Representations

CAPRI-Net (CVPR 2022)

D²CSG (NeurIPS 2023)



DPA-Net (ECCV 2024)

Sweep-Net (ECCV 2024)

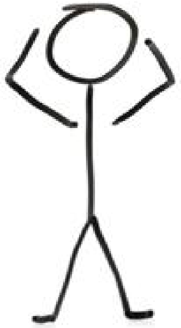
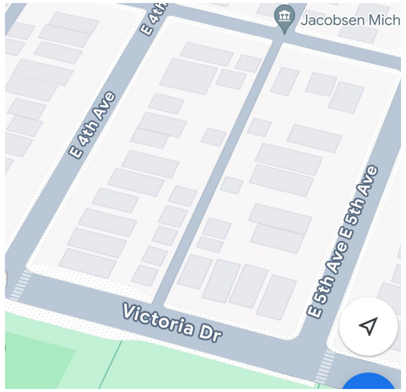
PartNet (CVPR 2019)

HAL3D (ICCV 2023)

3D Abstraction





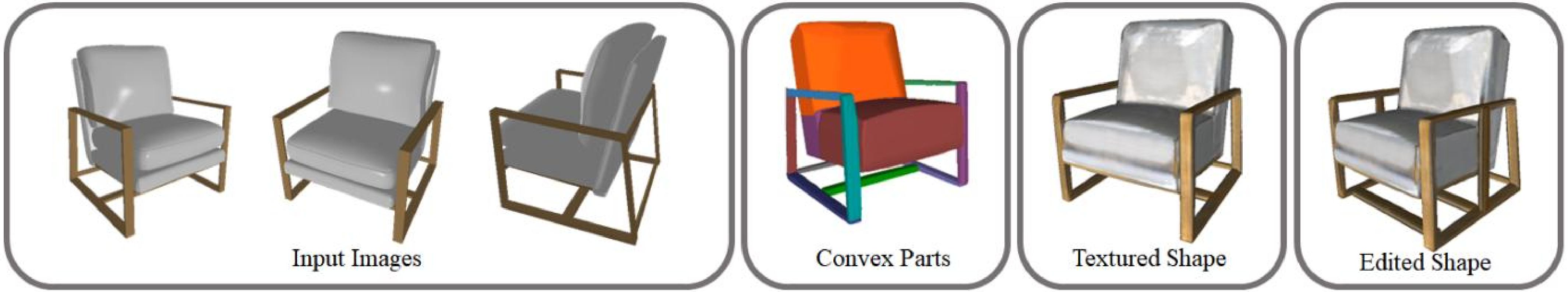
- Compress data
- Reduce computational cost

Facilitate high-level perception

On-target communication and visualization

Simplify complex shapes with **fundamental** and **manageable** primitives

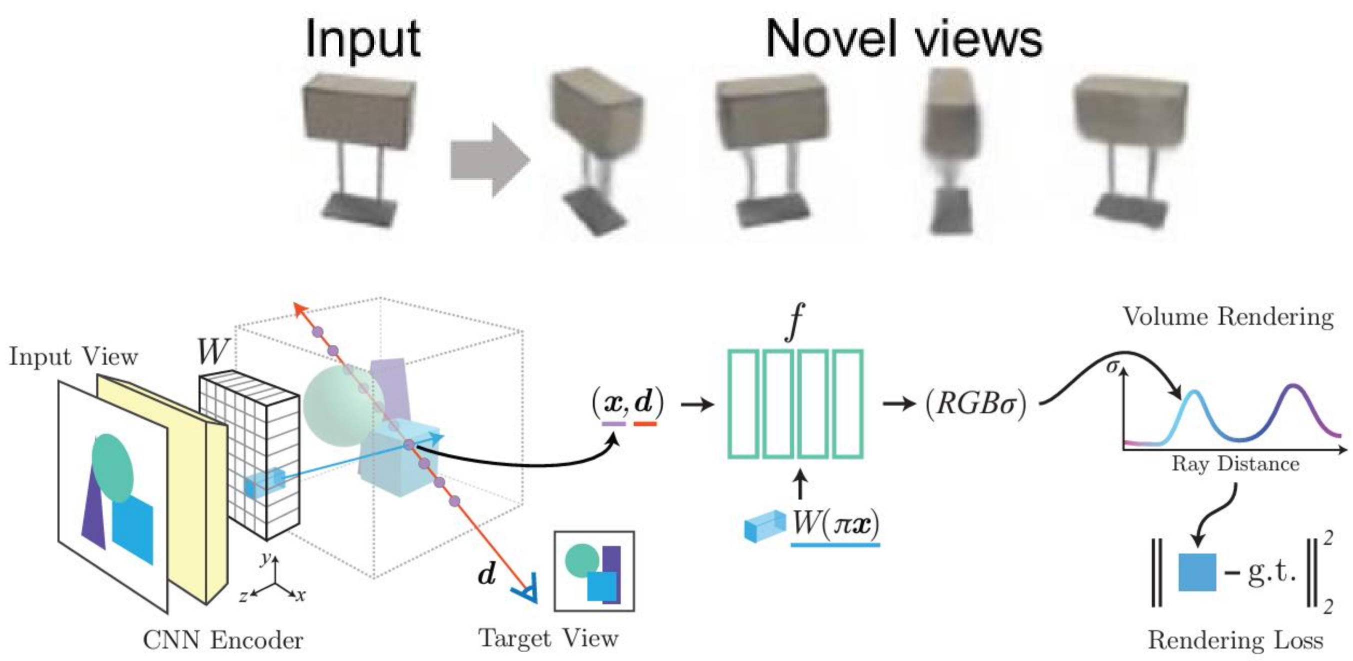
Structured 3D Abstraction from Sparse Views via Differentiable Primitive Assembly



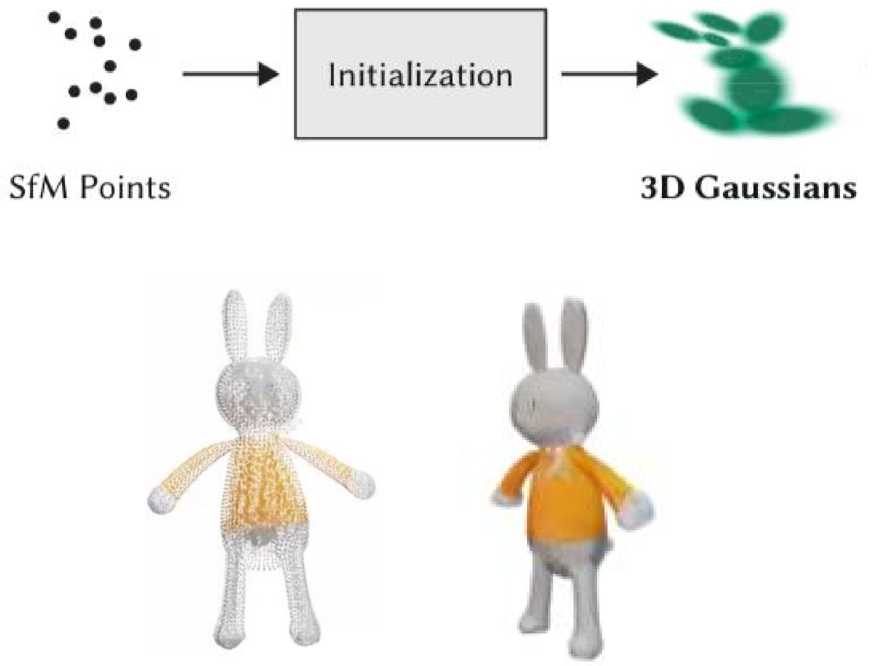
Structured 3D Abstraction from Sparse Views via Differentiable Primitive Assembly

- Related works

☹ Not friendly for manipulation and editing



pixelNerf (CVPR 2021)

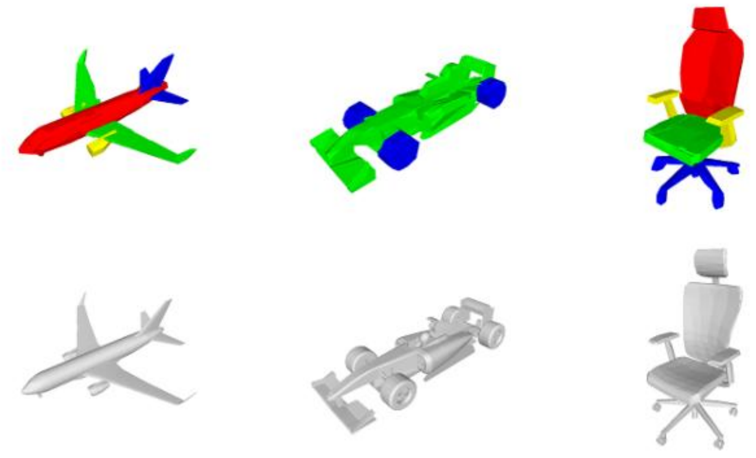


Gaussian splatting (SIGRRAPH 2023)

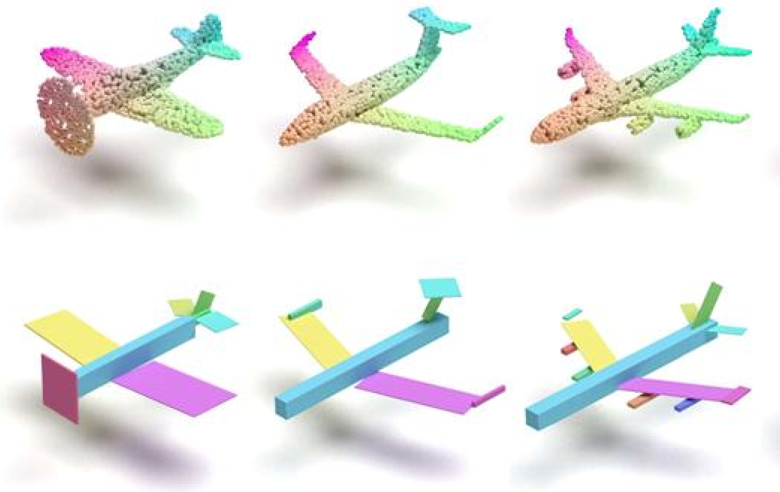
Structured 3D Abstraction from Sparse Views via Differentiable Primitive Assembly

- Related works

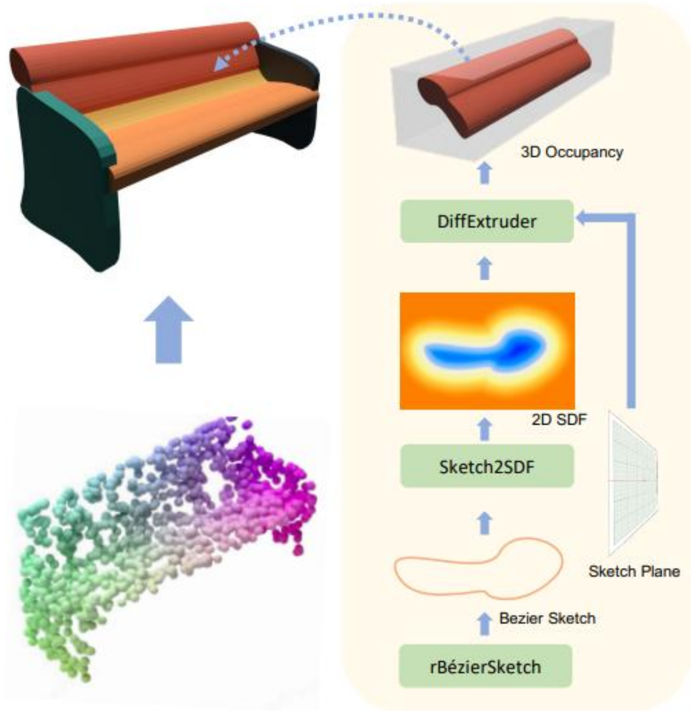
3D training data is needed



BSP-Net (CVPR 2020)



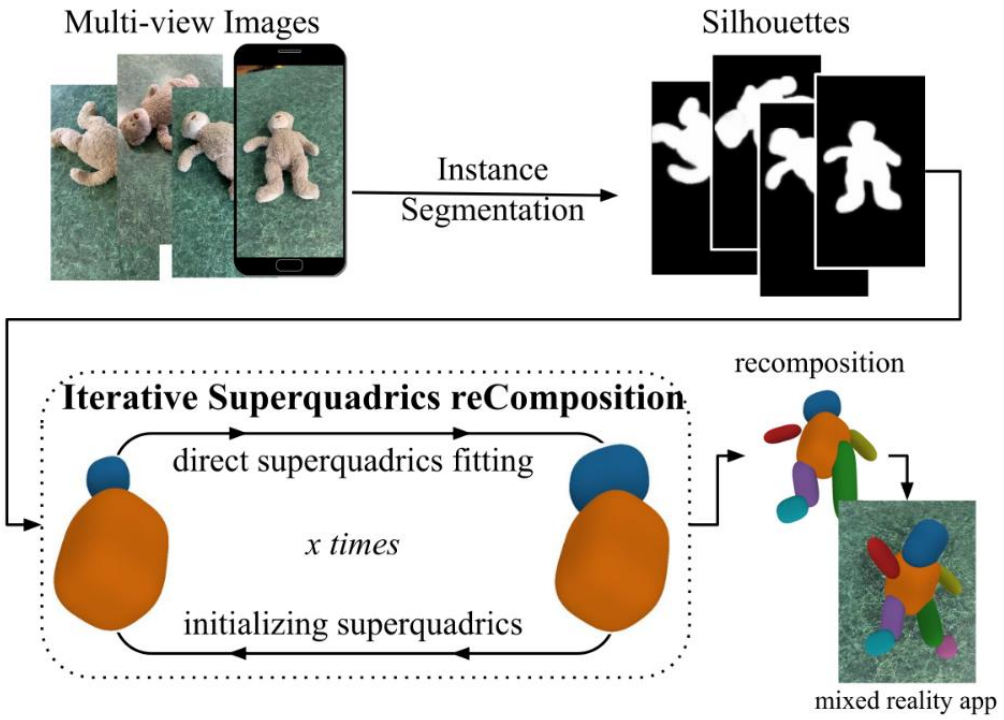
Cuboids Abstraction (SIGGRAPH 2021)



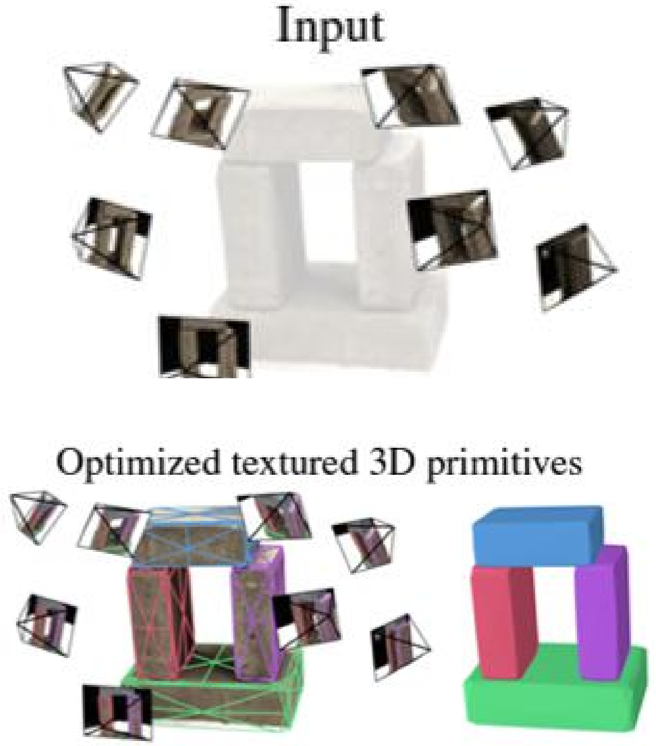
ExtrudeNet (ECCV 2022)

Structured 3D Abstraction from Sparse Views via Differentiable Primitive Assembly

- Related works ☹ Not general, require dense views

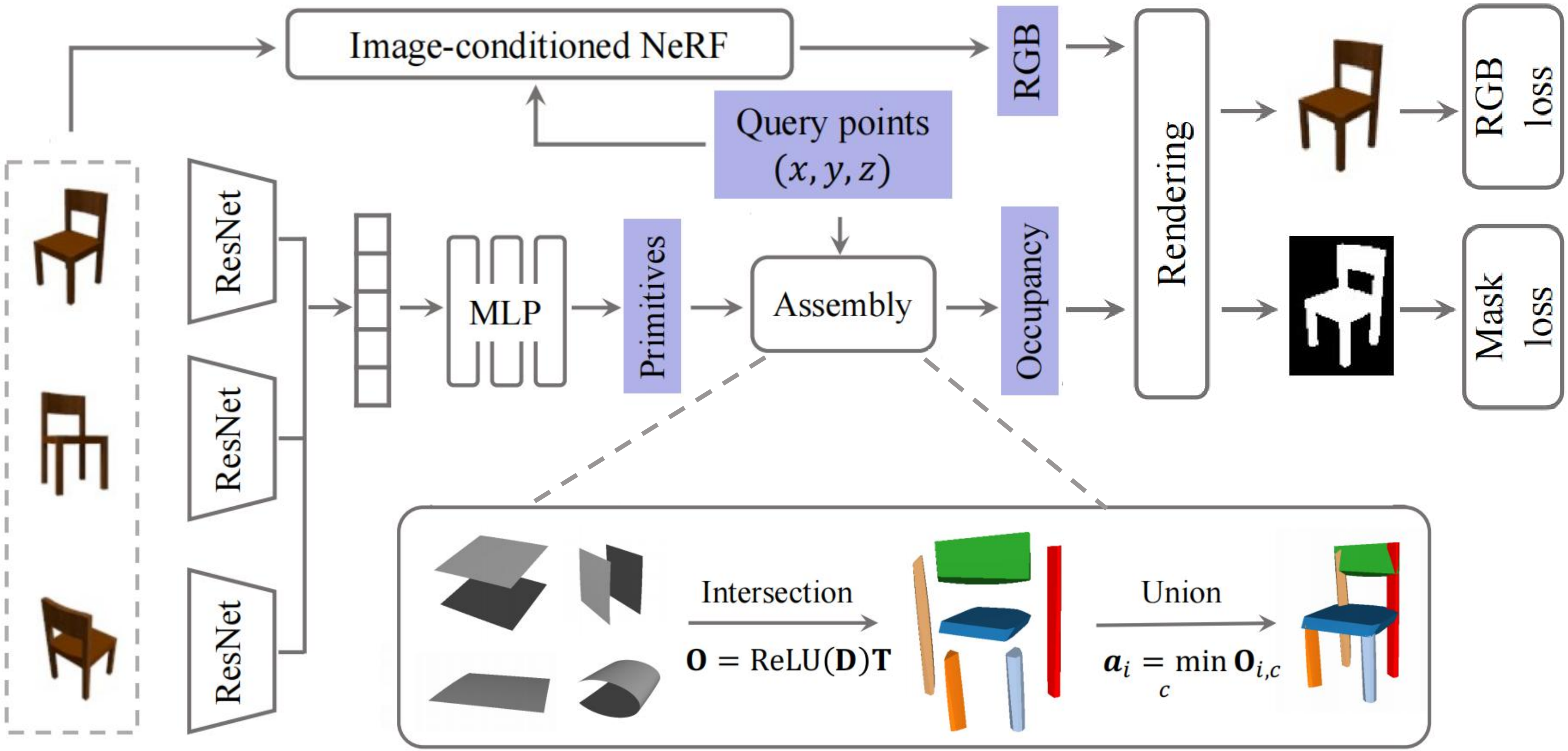


ISCO (ICCV 2023)



Differentiable block world (NeurIPS 2023)

Overview of DPA-Net



Improvements

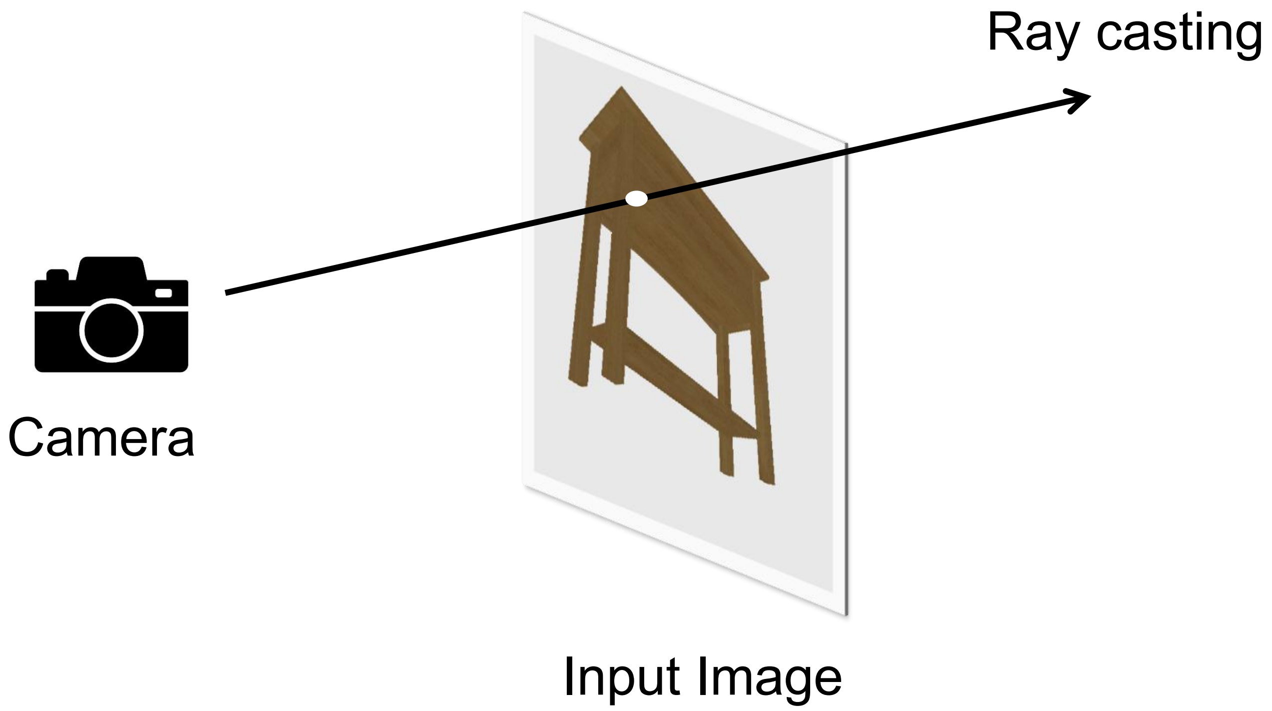
- Improve structure: overlapping loss and dropout strategy

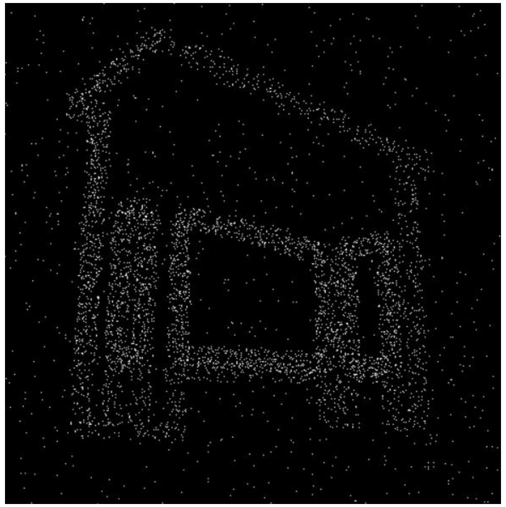
Phase	Type of $\mathbf{T}$	Type of $\mathbf{w}$	Occupancy	Opacity	Dropout	Loss	Model parameters
1	float	float	$\mathbf{a}^+$	$\mathbf{a}^+$	-	$\mathcal{L}_{ph} + \mathcal{L}_{\mathbf{T}} + \mathcal{L}_{\mathbf{w}}$	network, $\mathbf{T}$, $\mathbf{w}$
2	float	-	$\mathbf{a}^*$	$\exp(-10\mathbf{a}^*)$	-	$\mathcal{L}_{ph} + \mathcal{L}_{\mathbf{T}}$	network, $\mathbf{T}$
3	binary	-	$\mathbf{a}^*$	$\exp(-10\mathbf{a}^*)$	✓	$\mathcal{L}_{ph} + \boxed{\mathcal{L}_{over}}$	network

Details of the multi-stage fine-tuning

Improvements

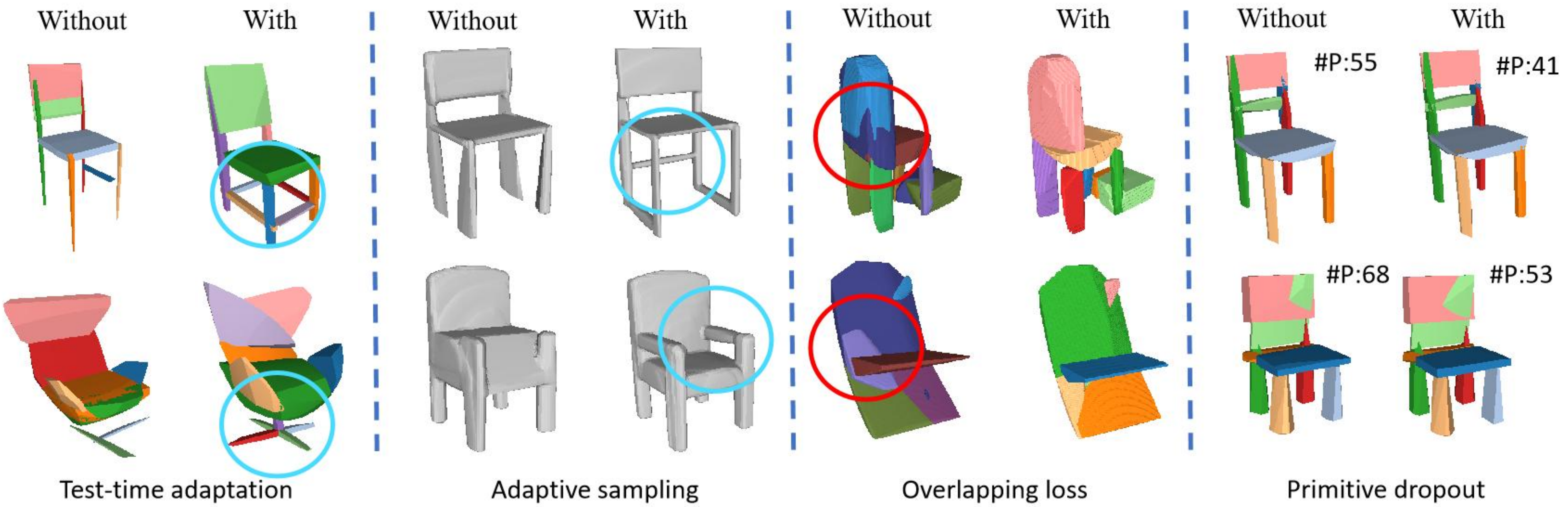
- Improve reconstruction accuracy: silhouette-aware adaptive sampling



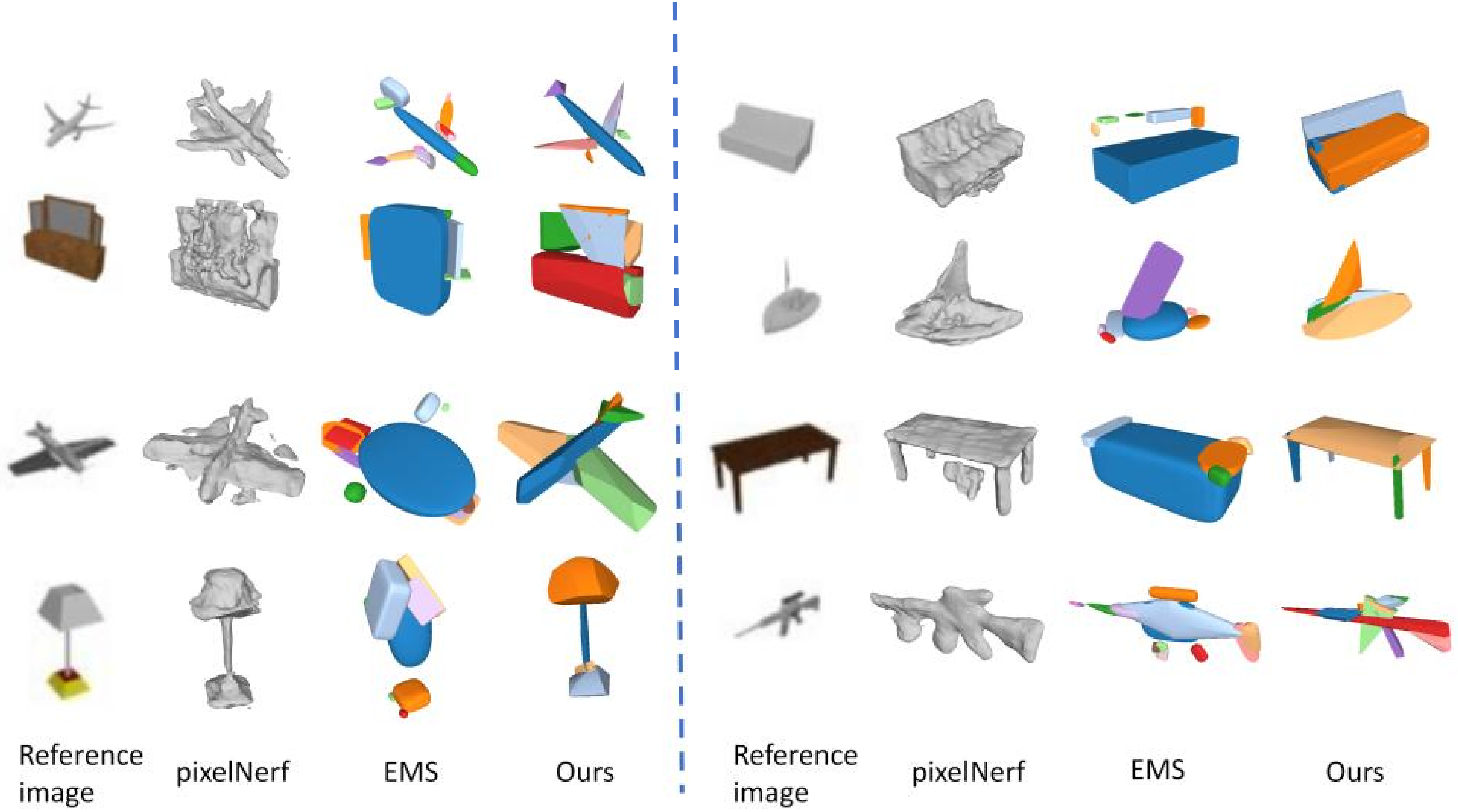


Adaptive camera ray direction sampling

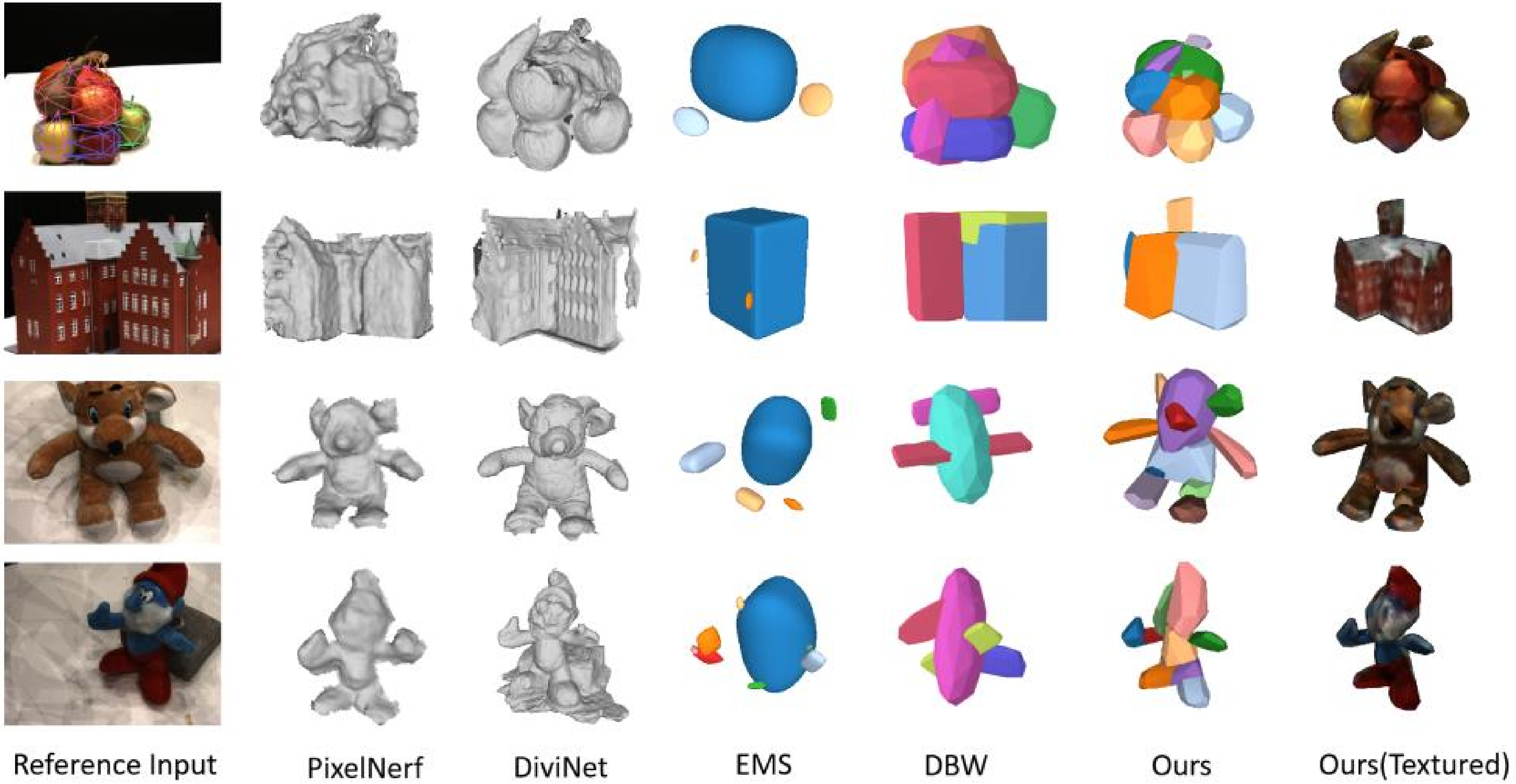
Ablation Studies



Results on ShapeNet Cross-categories

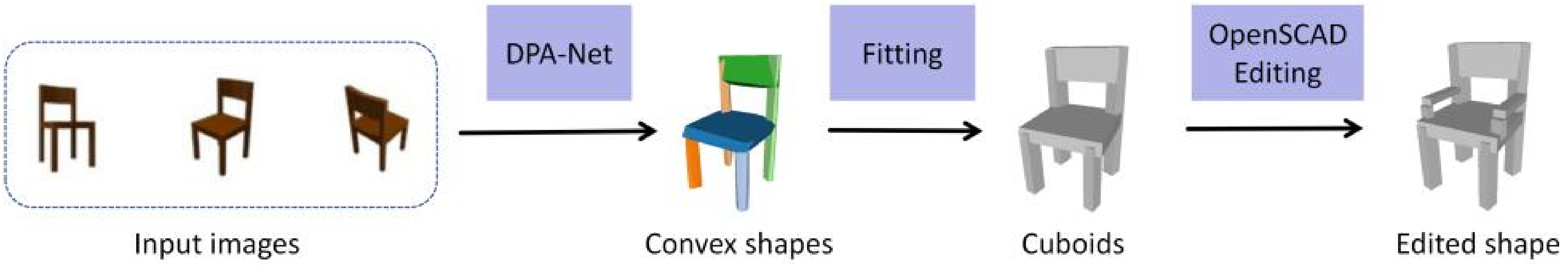


Results on Real Images



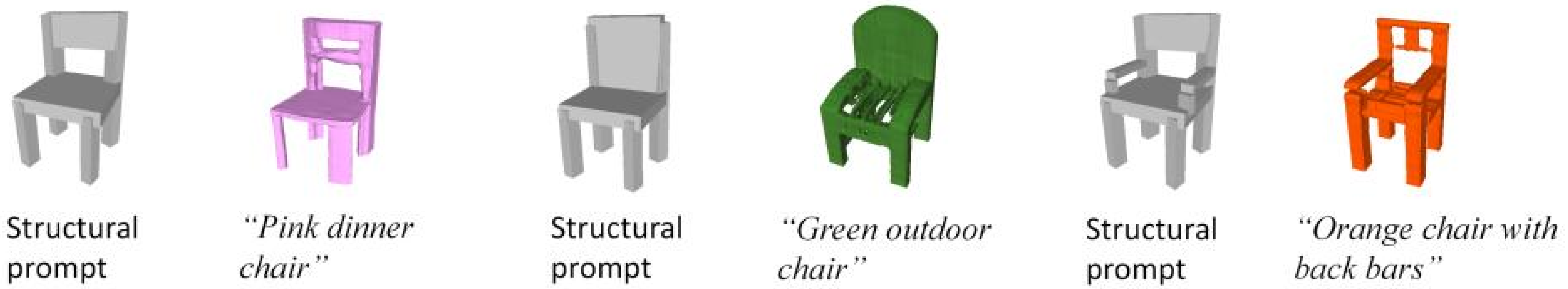
Application: Shape Editing

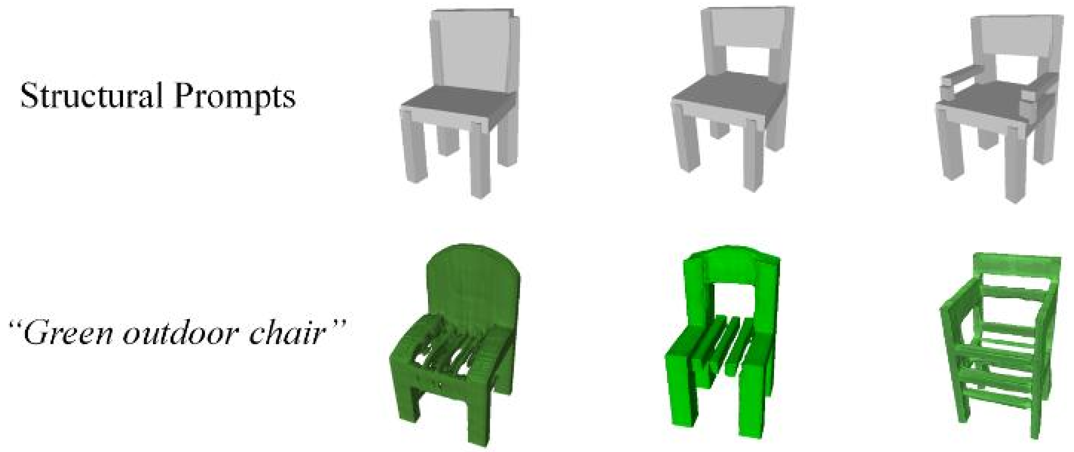
Shape editing in MeshLab



Shape editing in OpenSCAD

Application: Conditional Shape Generation

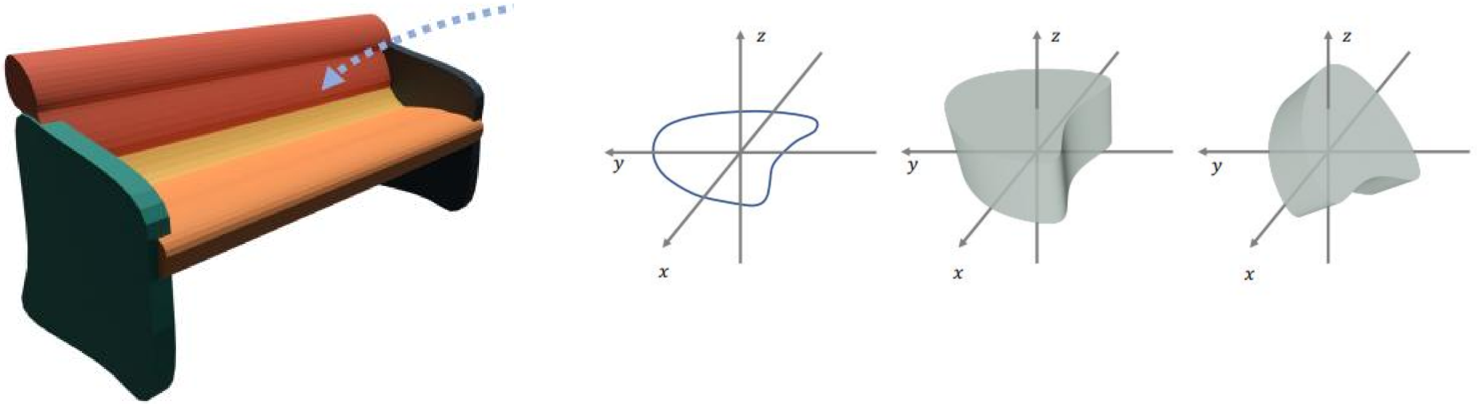




Limitation of Previous Representation

DPA-Net
(Convex shape)

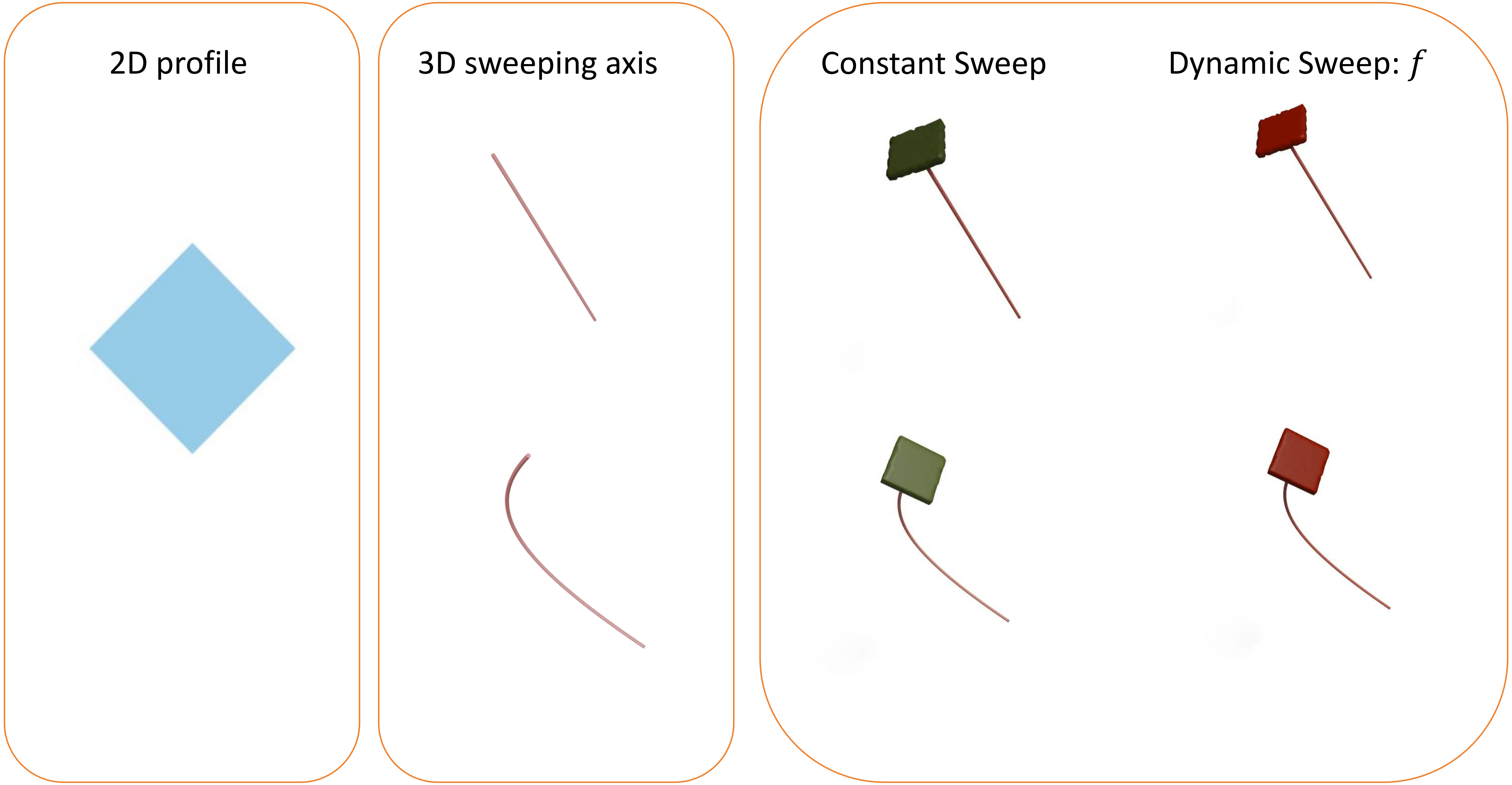
Extrude-Net
(Extrusion)



Not general for curvy objects

SweepNet: Unsupervised Learning of Shape Abstraction via Neural Sweepers

What is a sweep surface?

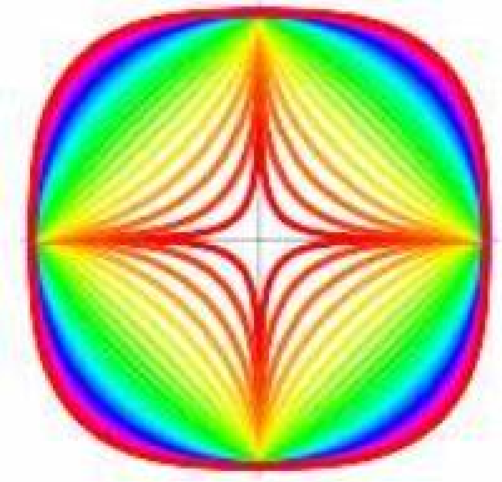


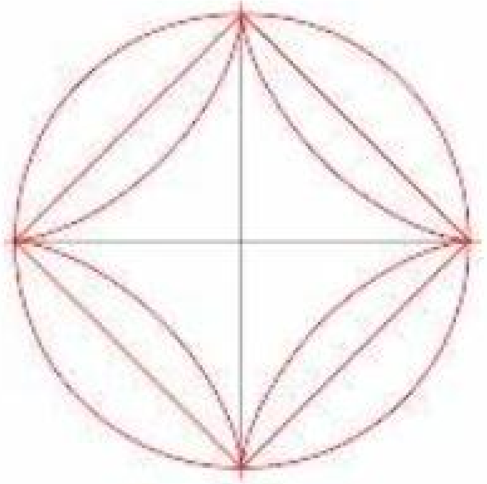
Sweep Surface Parametrization – Profile

- Efficient parametrization
- Wide shape vocabulary
- Guaranteed close-loop without self-intersection

Superellipse

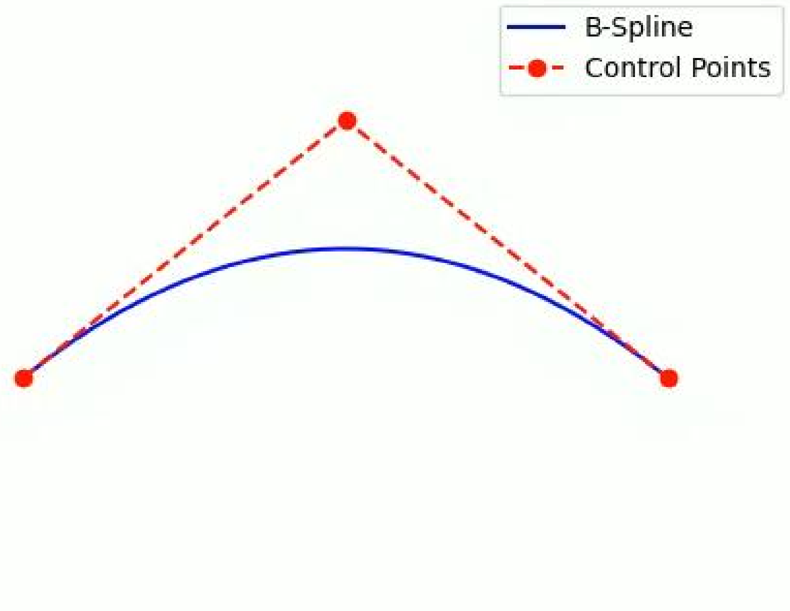
$$\begin{cases} x(\theta) = a \cdot |\cos(\theta)|^{\frac{2}{d}} \cdot \operatorname{sgn}(\cos(\theta)), \\ y(\theta) = b \cdot |\sin(\theta)|^{\frac{2}{d}} \cdot \operatorname{sgn}(\sin(\theta)), \end{cases}$$



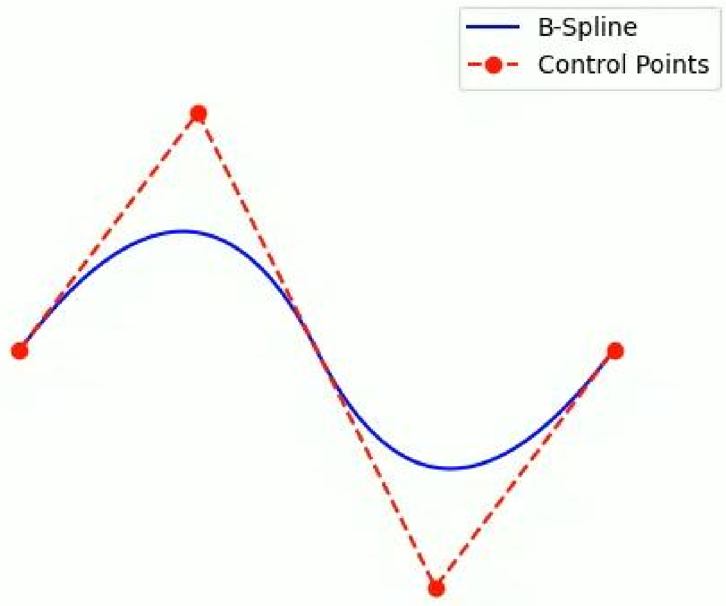


Sweep Surface Parametrization – Sweeping Axis

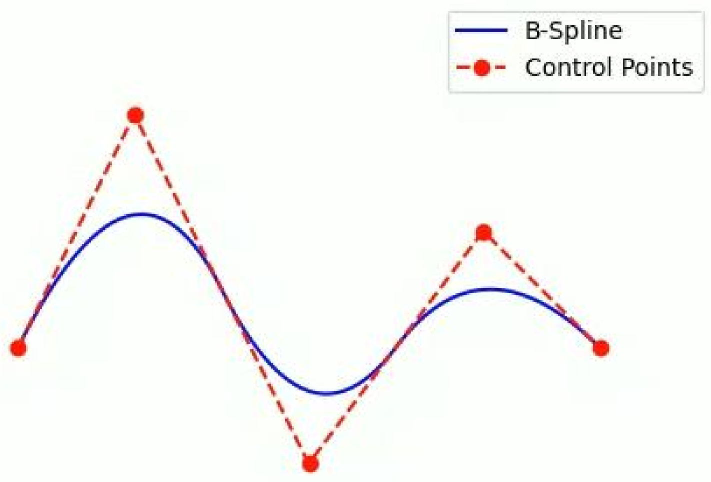
B-spline curves



3 control points



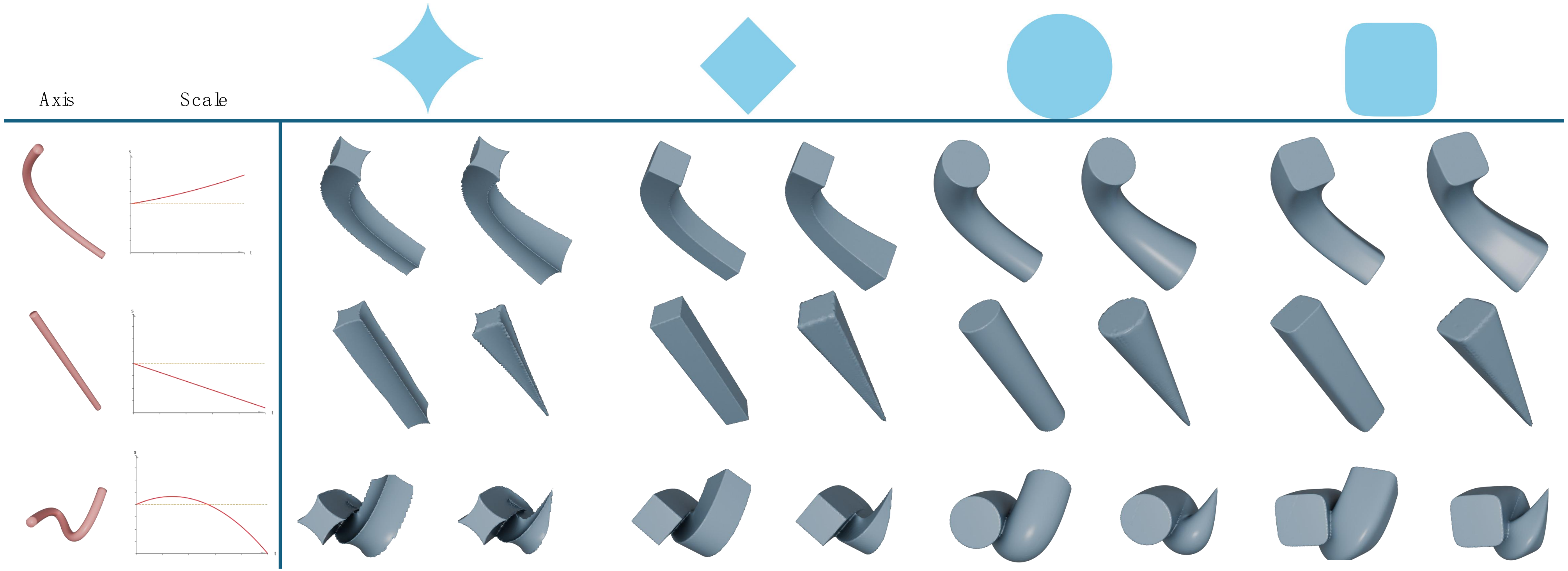
4 control points



5 control points

Sweep Surface Parametrization: Scaling function

Quadratic scaling function $s(t) = at^2 + bt + 1$



Sweep Surface Parametrization

$$S = [\underbrace{c_1, \cdots, c_n}_{\text{Sweeping axis}}, \underbrace{a, b, d}_{\text{Profile}}, \underbrace{f_0, \cdots, f_k}_{\text{Scaling function}}] \in \mathbb{R}^{3n+k+3}$$

With 3 control points B-spline and fixed-constant quadratic scaling function
A sweep surface only need **14** float numbers to represent

Sweep Surface Construction

Parameter space

$$S = [c_1, \cdots, c_n, a, b, d, f_0, \cdots, f_k] \in \mathbb{R}^{3n+k+3}$$

A 3D representation

3D realm

Occupancy field

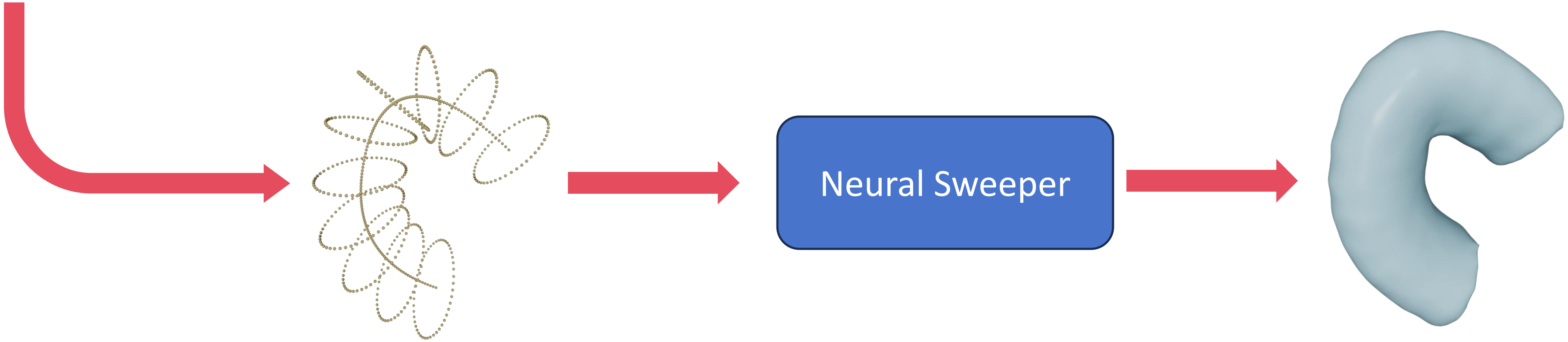
- Resolution-invariant – allow dense sampling
- Robust against union operations, suitable for primitive assembly
- Gradient friendly – easy to train

Neural Sweeper

$$[\underbrace{c_1, \cdots, c_n}_{\text{Sweeping axis}}, \underbrace{a, b, d}_{\text{Profile}}, \underbrace{f_0, \cdots, f_k}_{\text{Scaling function}}]$$

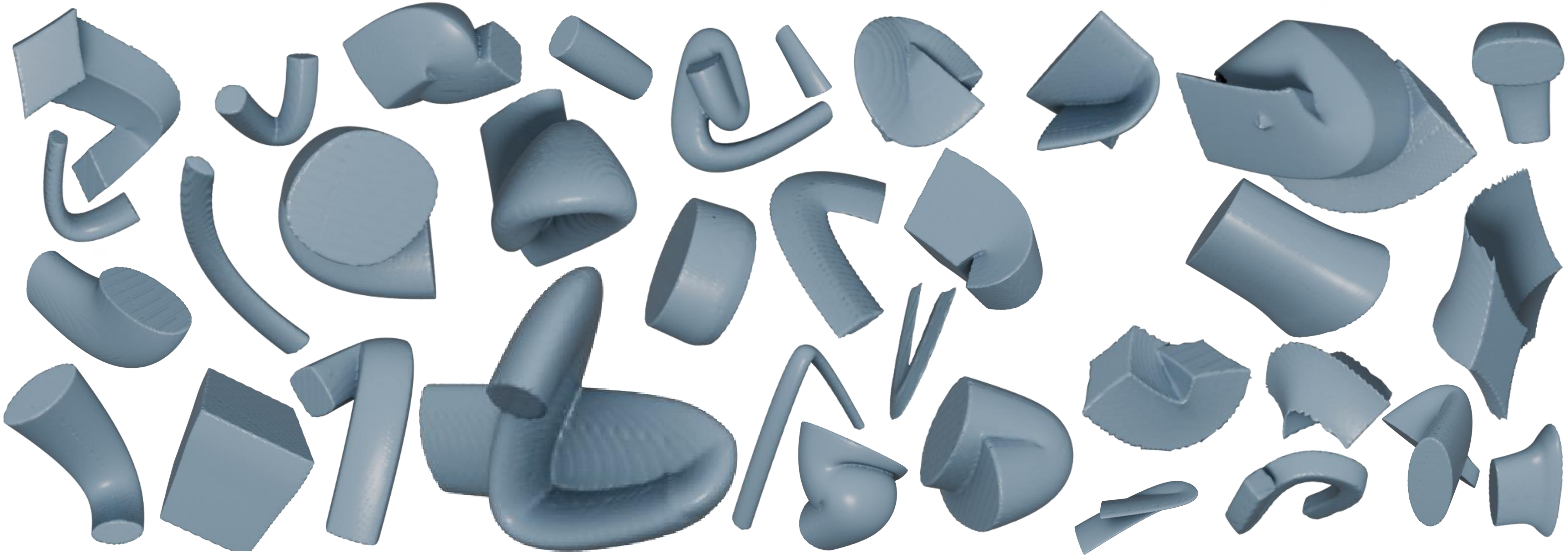
Complex implicit fomula

Non-differentiable
Significant computation cost



A differentiable surrogate!

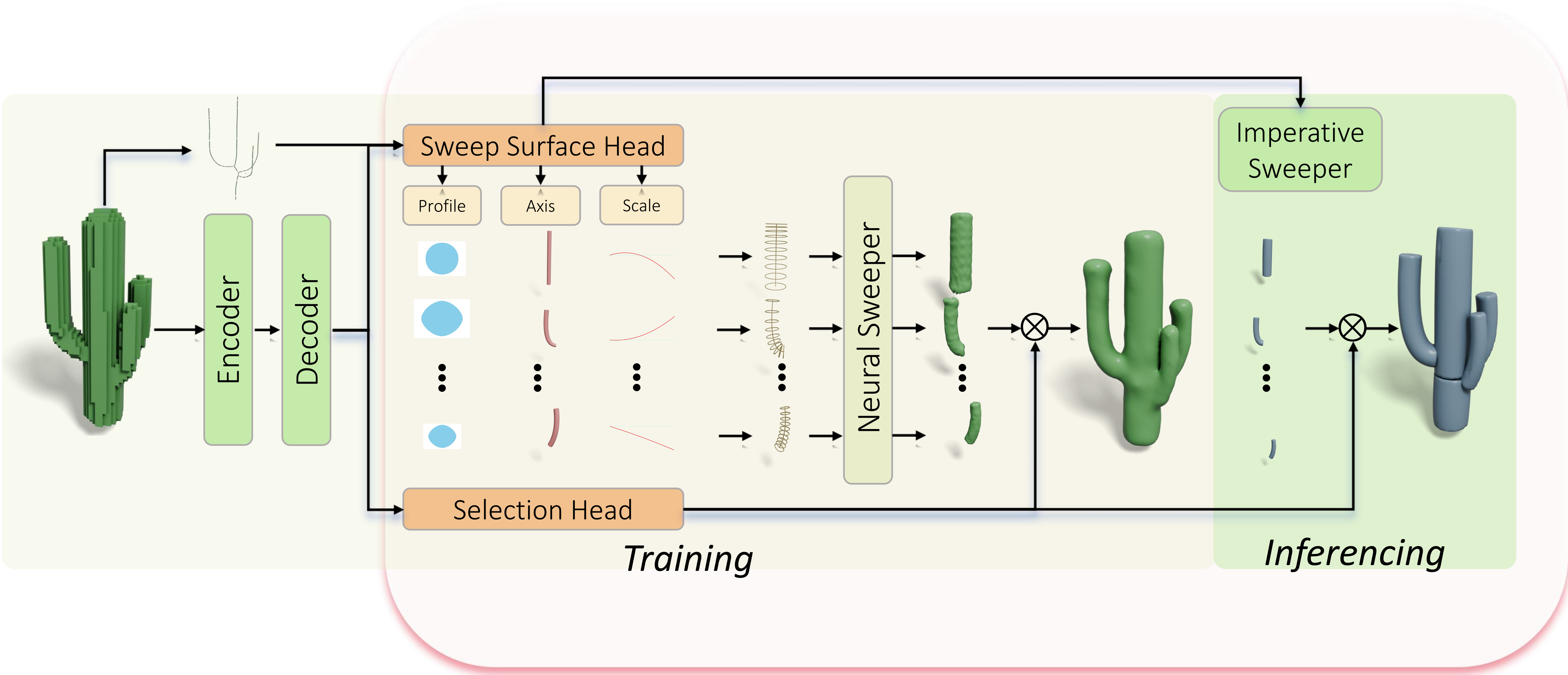
Neural Sweeper



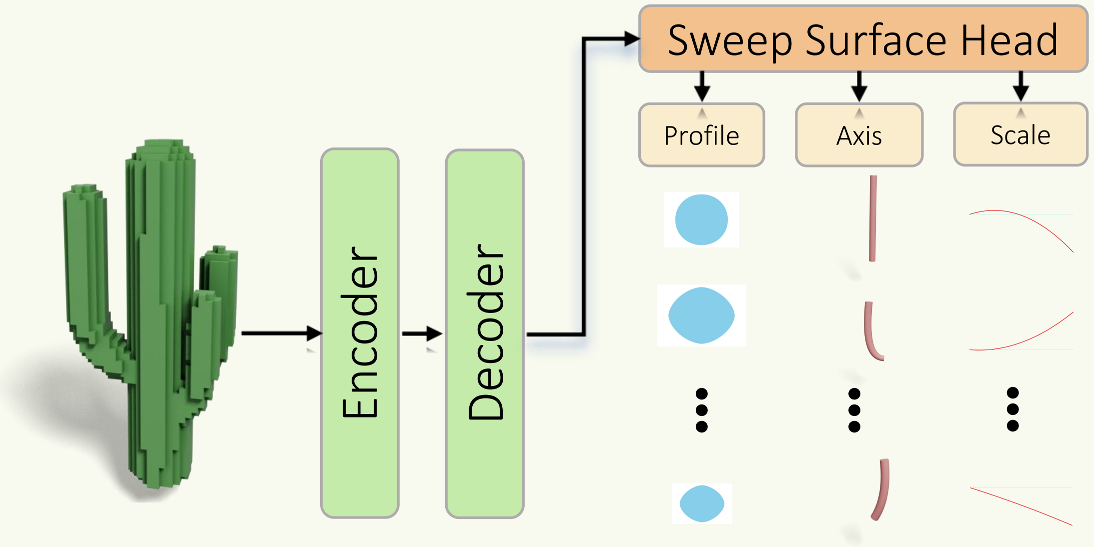
Train on 20,000 sweep surfaces


HELLO THIS IS

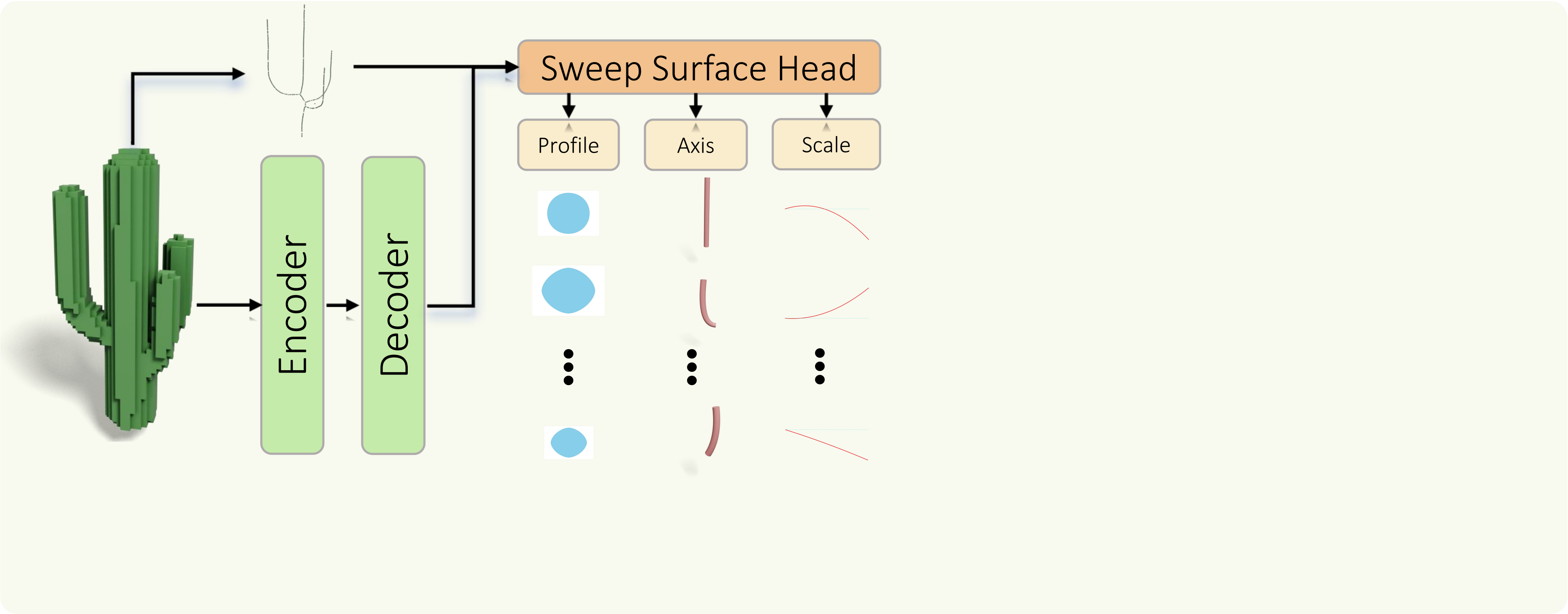
SweepNet


Encoder
Decoder
Sweep Surface Head
Profile
Axis
Scale
Neural Sweeper
Imperative Sweeper
Selection Head
Training
Inferencing

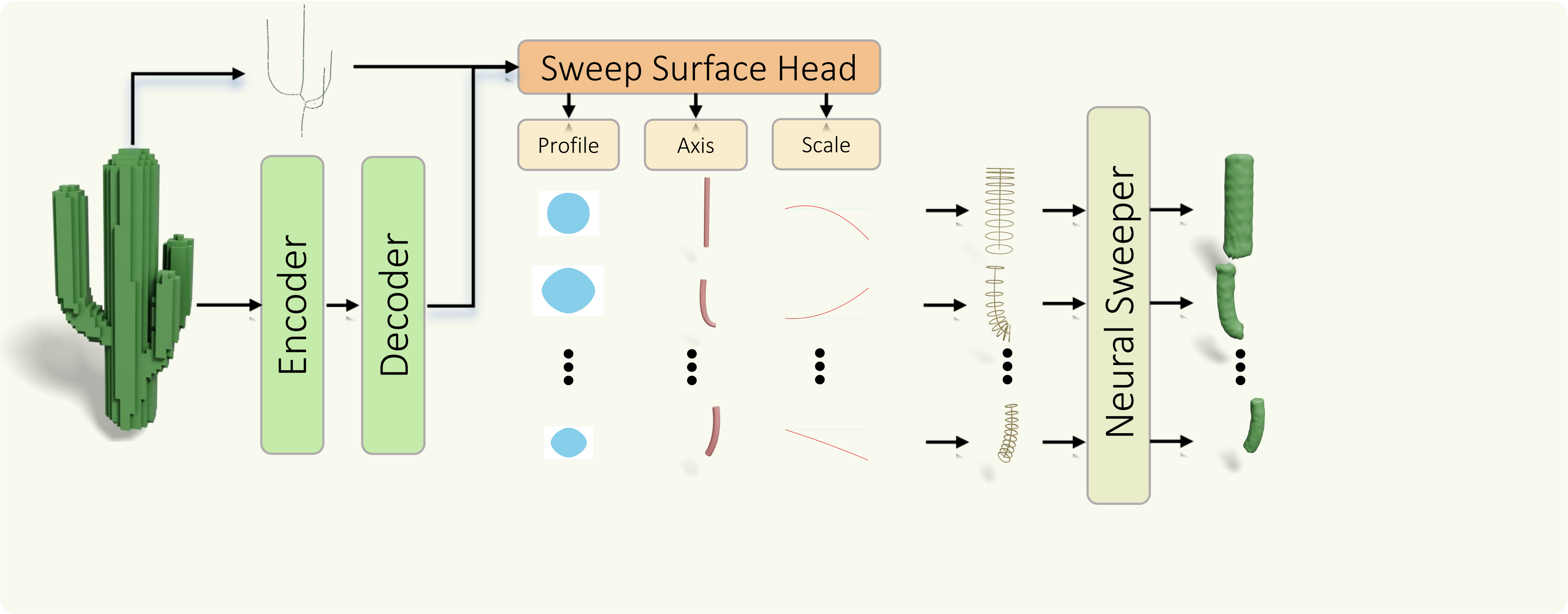
SweepNet



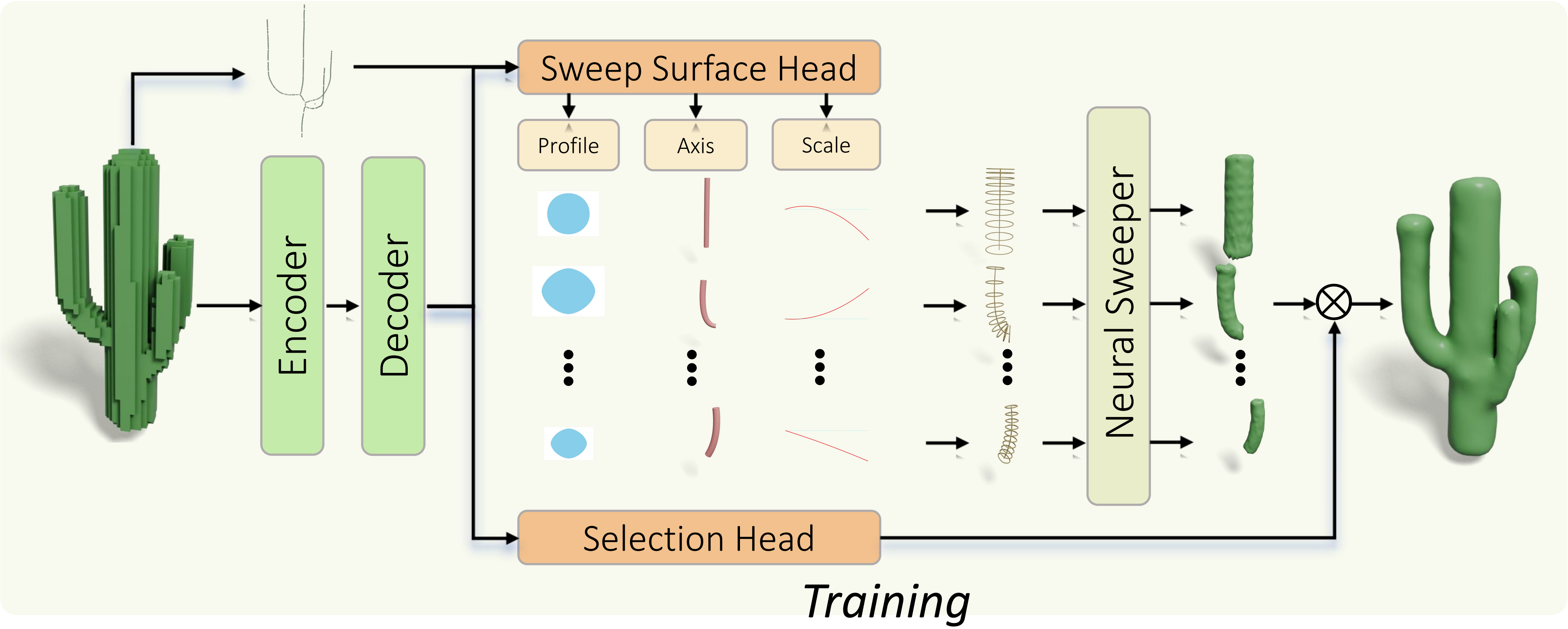
SweepNet



SweepNet

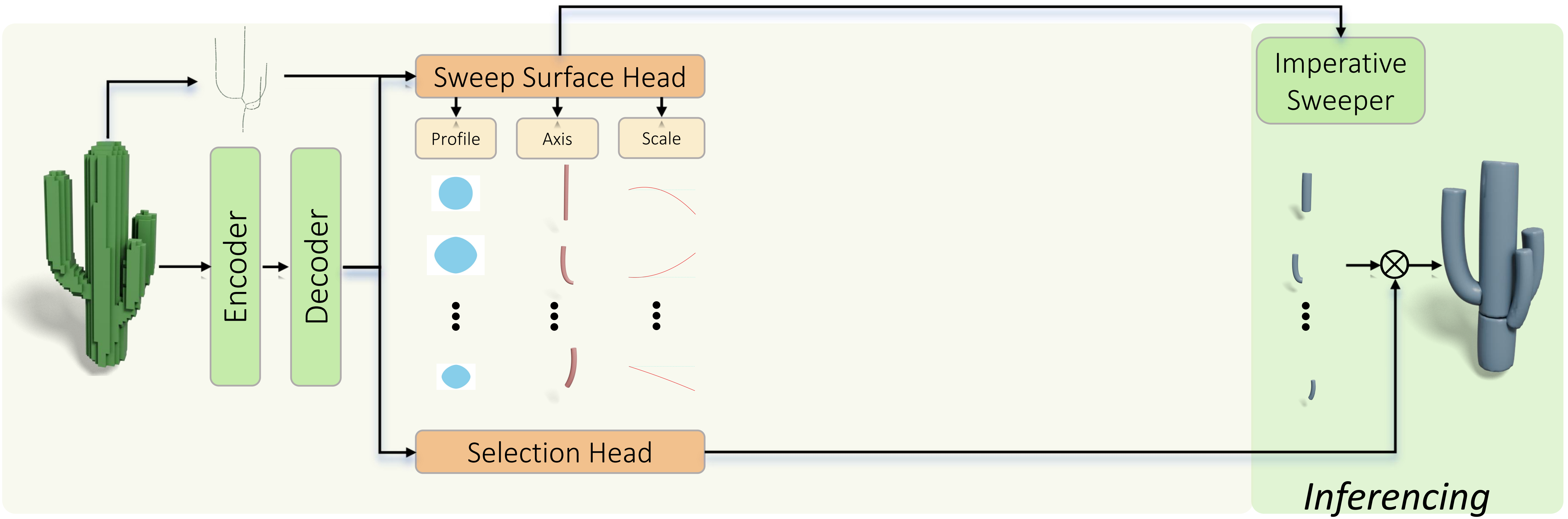

Sweep Surface Head
Profile
Axis
Scale
Encoder
Decoder
Neural Sweeper

SweepNet



Sweep Surface Head
Profile
Axis
Scale
Encoder
Decoder
Neural Sweeper
Selection Head
Training

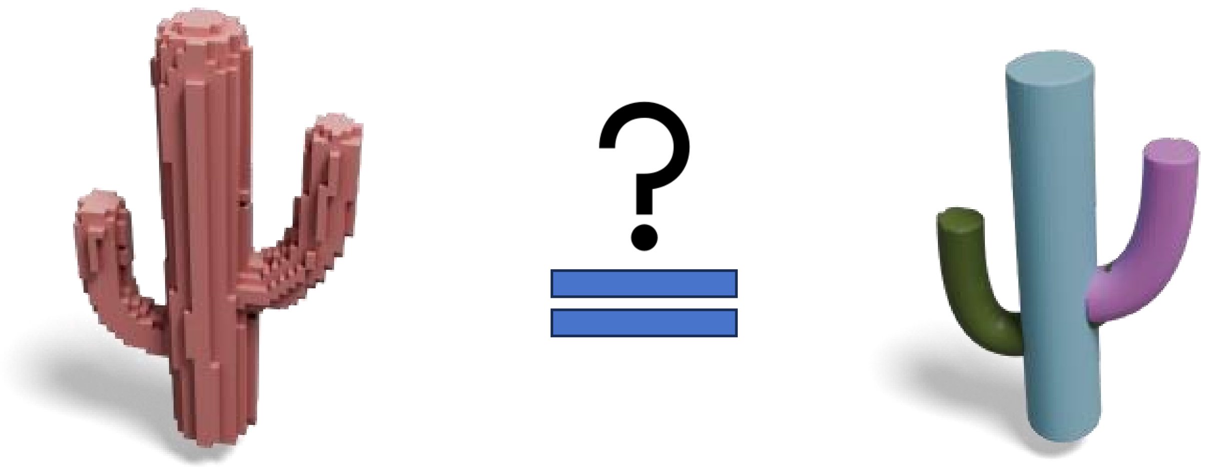
SweepNet


Sweep Surface Head
Profile
Axis
Scale
Encoder
Decoder
Selection Head
Imperative Sweeper
Inferencing

Loss function

Reconstruction Loss

$$\mathcal{L}_{recon} = \mathbb{E}_{t\sim T}\left[\left\|O_{GT}(t) - \frac{\sum_{i=1}^{q} O_i(t)e^{\alpha O_i(t)}}{\sum_{i=1}^{q} e^{\alpha O_i(t)}}\right\|_2^2\right]$$



Loss function

Overlap Loss

$$\mathcal{L}_{ol} = \mathbb{E}_{t\sim T}\left[\min(\sum_{i=1}^{q} O_i(t) - \beta, 0)\right]$$

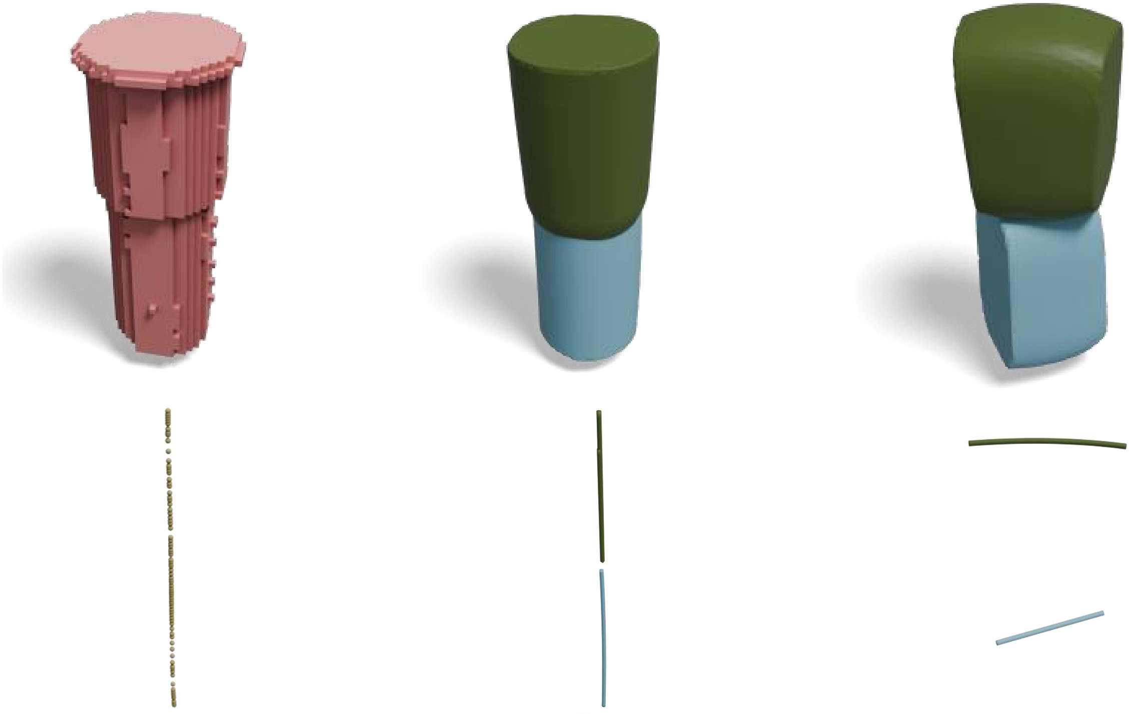
Parsimony Loss

$$\mathcal{L}_{pars} = \sqrt{q}.$$

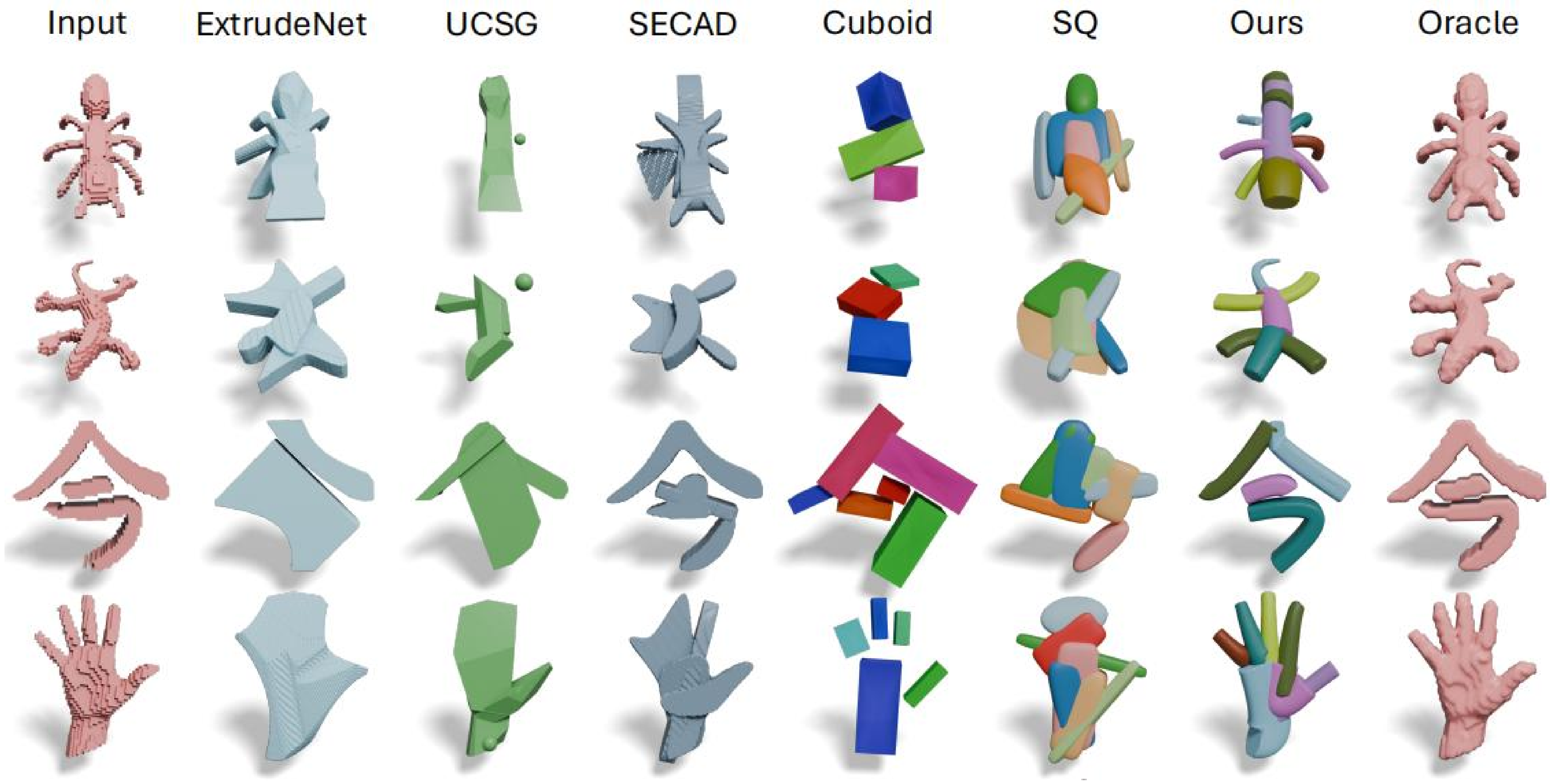
Loss function

Axis Loss

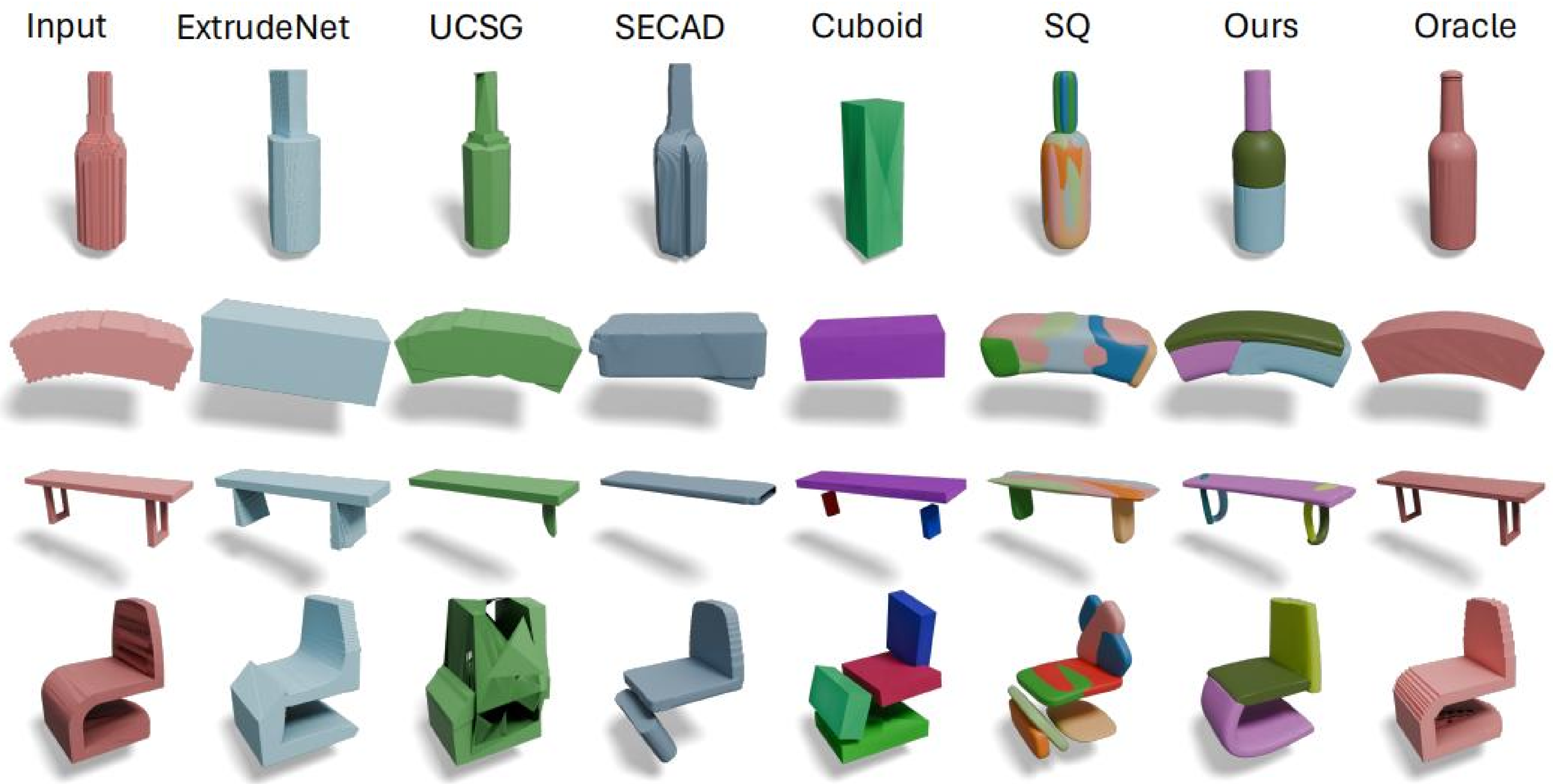
$$\mathcal{L}_{axis} = \mathbb{E}_{m \sim M}\left[\min_{s \in S} dist(m, s)\right]$$



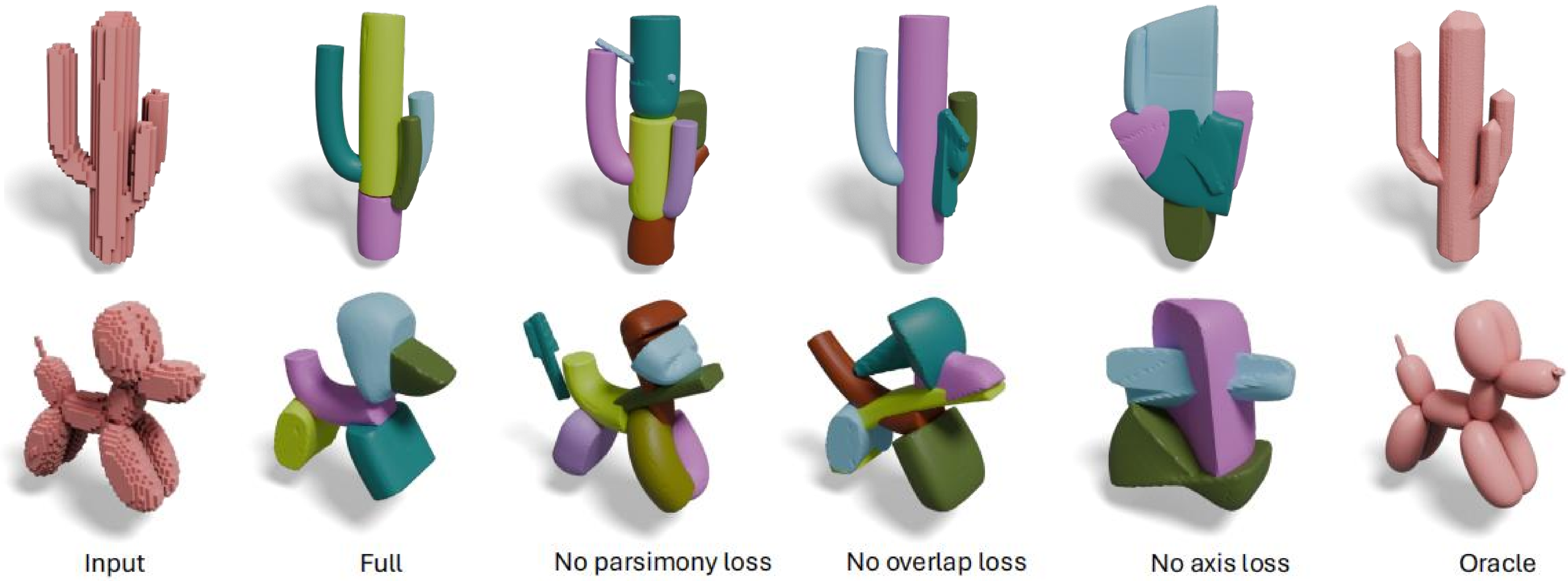
Results



Results



Ablation Study



Editability

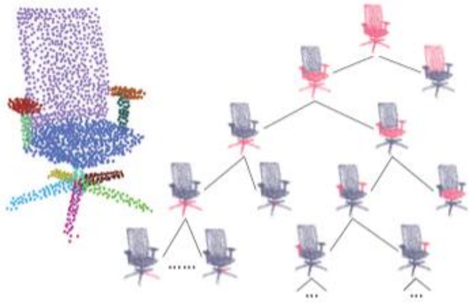
This Talk: Learning Structured 3D Representations

CAPRI-Net (CVPR 2022)

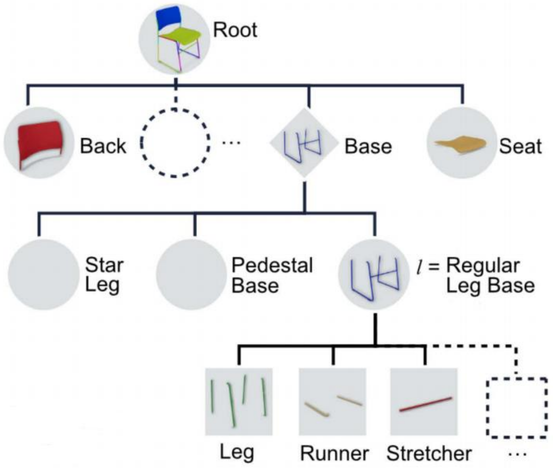
D²CSG (NeurIPS 2023)

DPA-Net (ECCV 2024)

Sweep-Net (ECCV 2024)



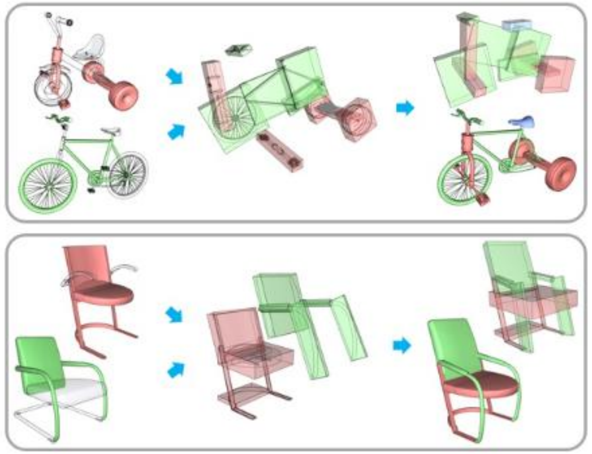
PartNet (CVPR 2019)

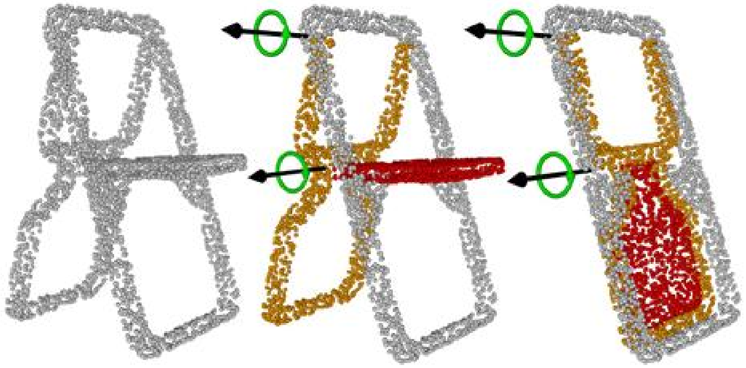


HAL3D (ICCV 2023)

Understanding Parts

- Advantages of part segmentation





Shape editing

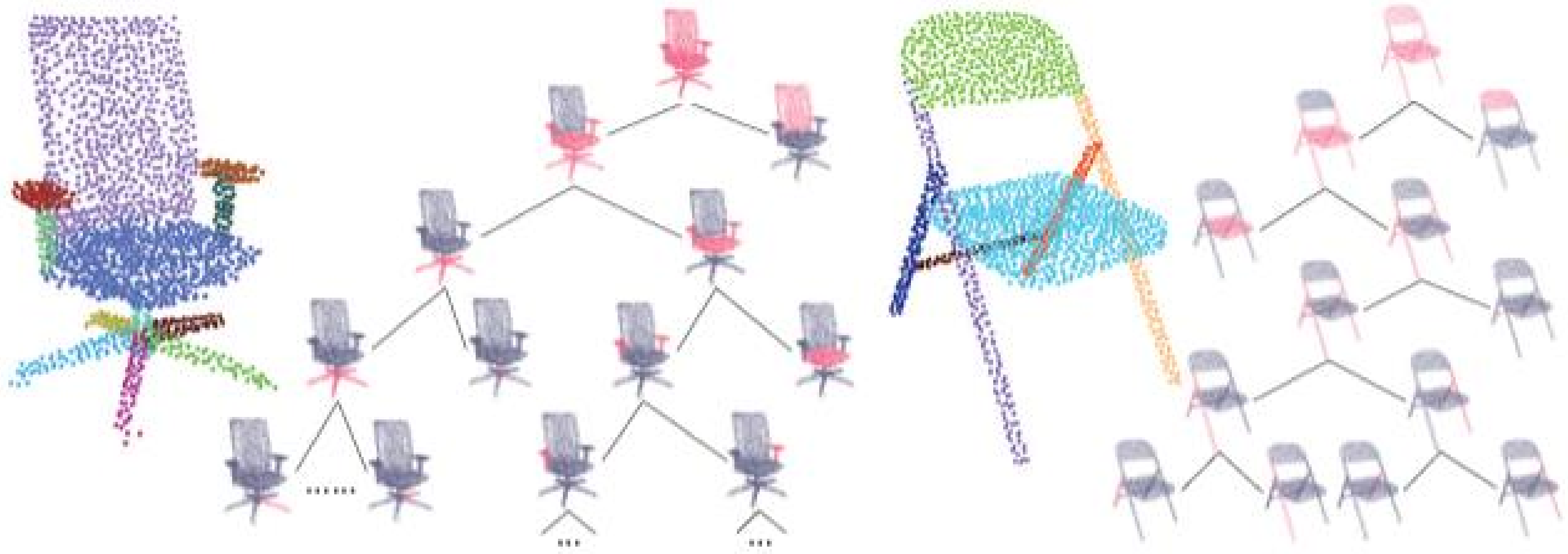
Part motion

Part texture editing

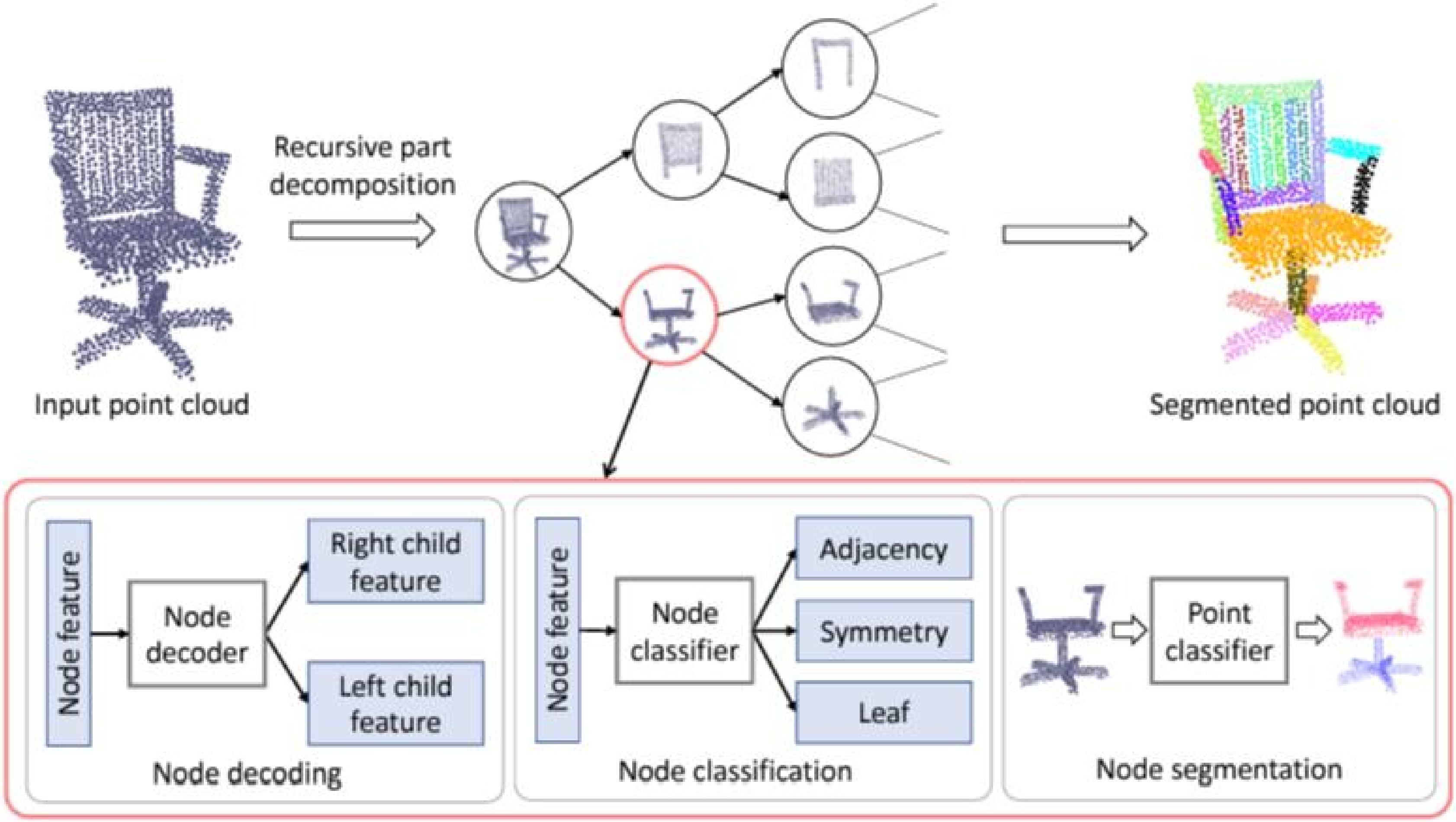
[1]Zhu et al 2018, SCORES: Shape Composition with Recursive Substructure Priors
[2]Wang et al 2019, Shape2Motion:Joint Analysis of Motion Part sand Attributes from 3D Shapes
[3]Yawar Siddiqui et al 2022, Texturify: Generating Textures on 3D Shape Surfaces

PartNet: A Recursive Part Decomposition Network for Fine-grained and Hierarchical Shape Segmentation

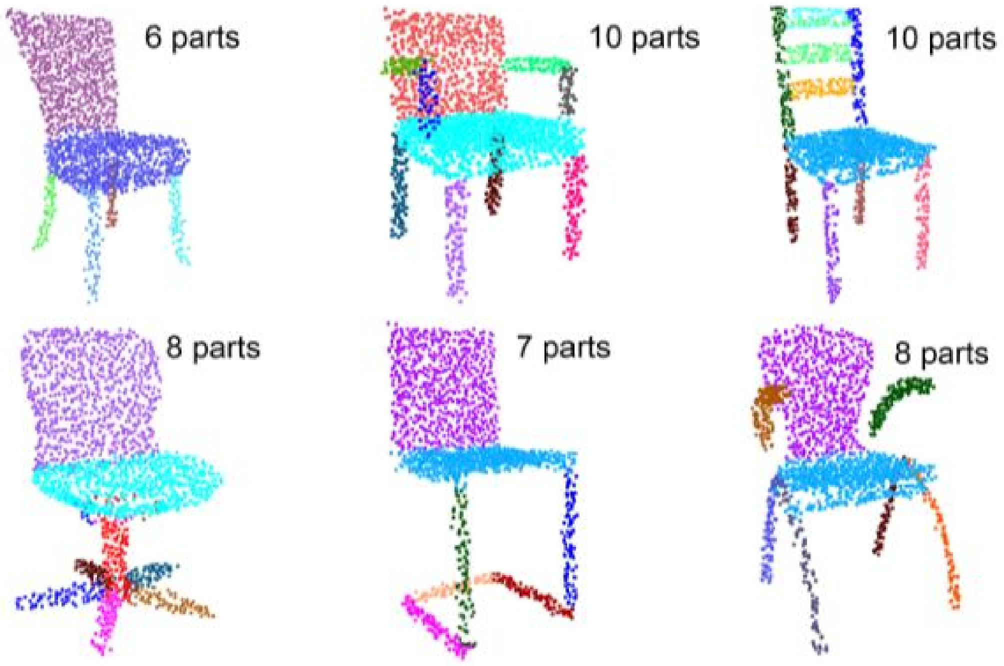
- Input: 3d point cloud
- Output: fine-grained part instance segmentation and part relations

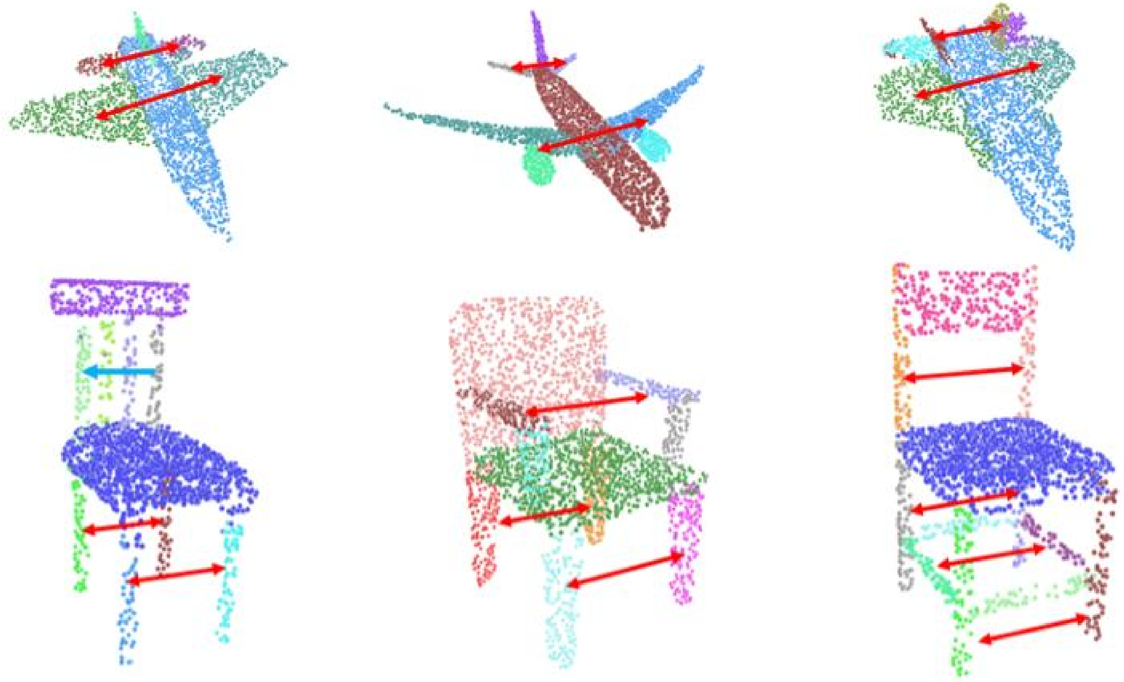


PartNet: Method



Part Segmentation Results

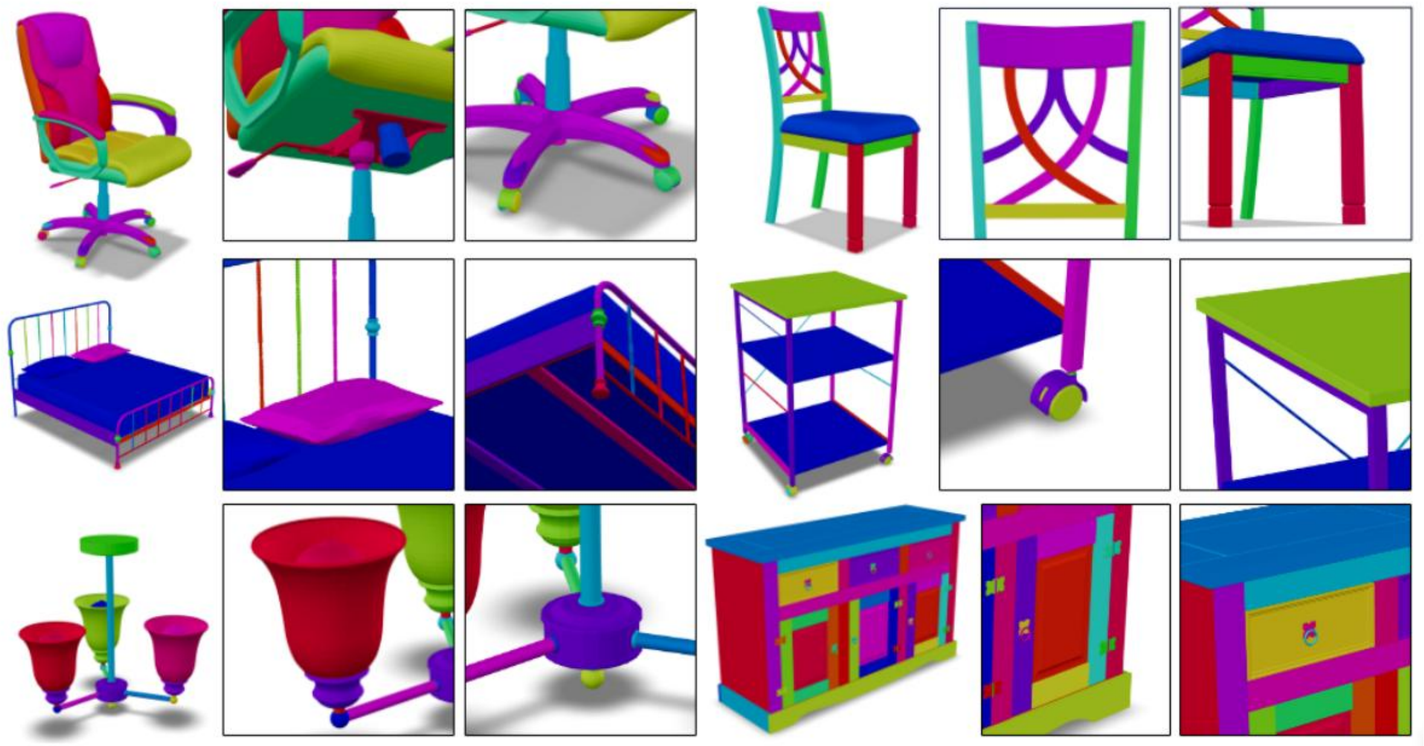




HAL3D: Hierarchical Active Learning for Fine-Grained 3D Part Labeling

- Online 3D assets created by human artists usually are made by connected components

Connected components in the ABO dataset

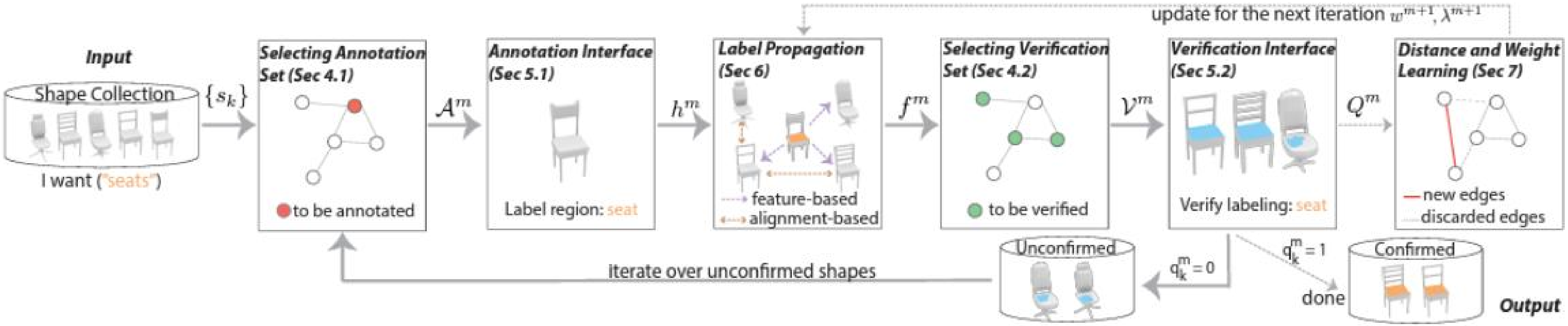


Unlabeled abstraction from DPA-Net

[1] ABO: Dataset and Benchmarks for Real-World 3D Object Understanding, CVPR 2022

HAL3D

- Related works

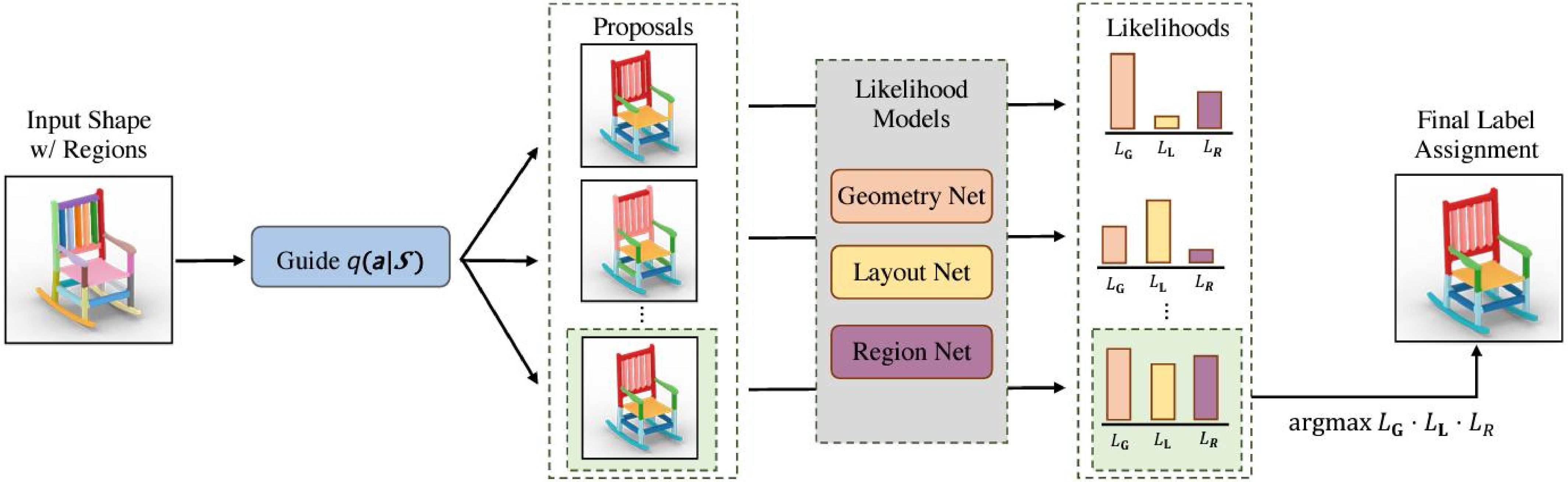


The active learning framework for high-level semantic segmentation [1]

[1] A Scalable Active Framework for Region Annotation in 3D Shape Collections, SIGGRAPH Asia 2016

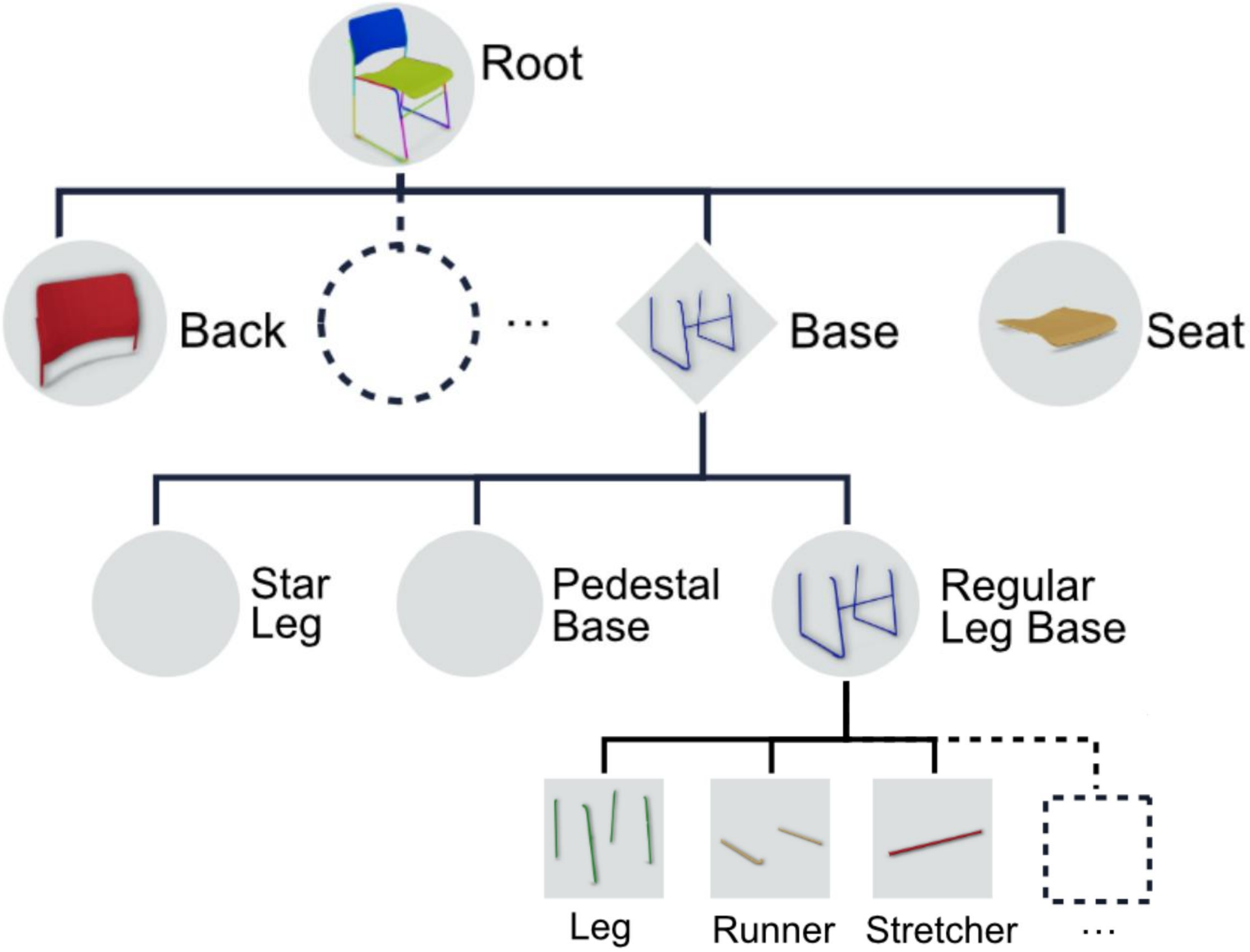
HAL3D

- Related works

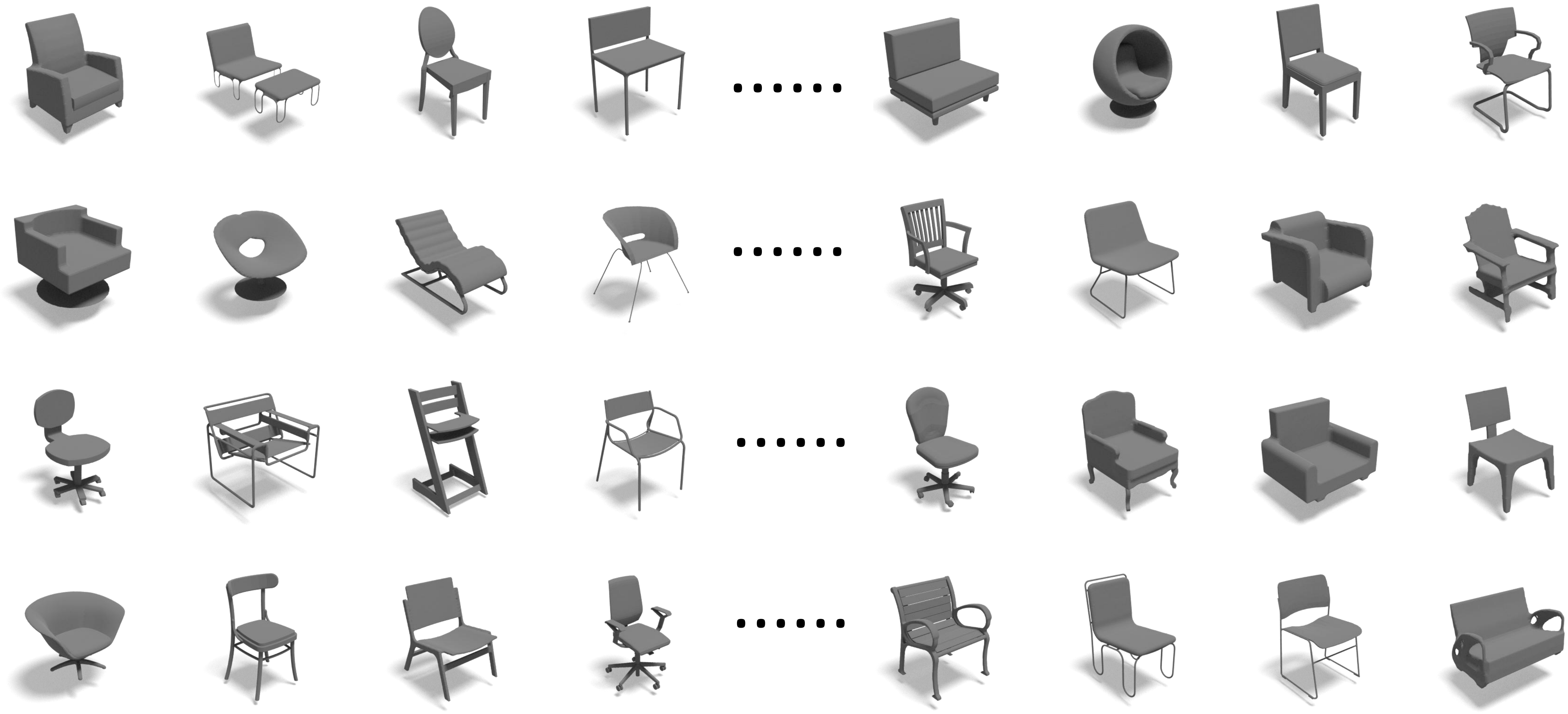


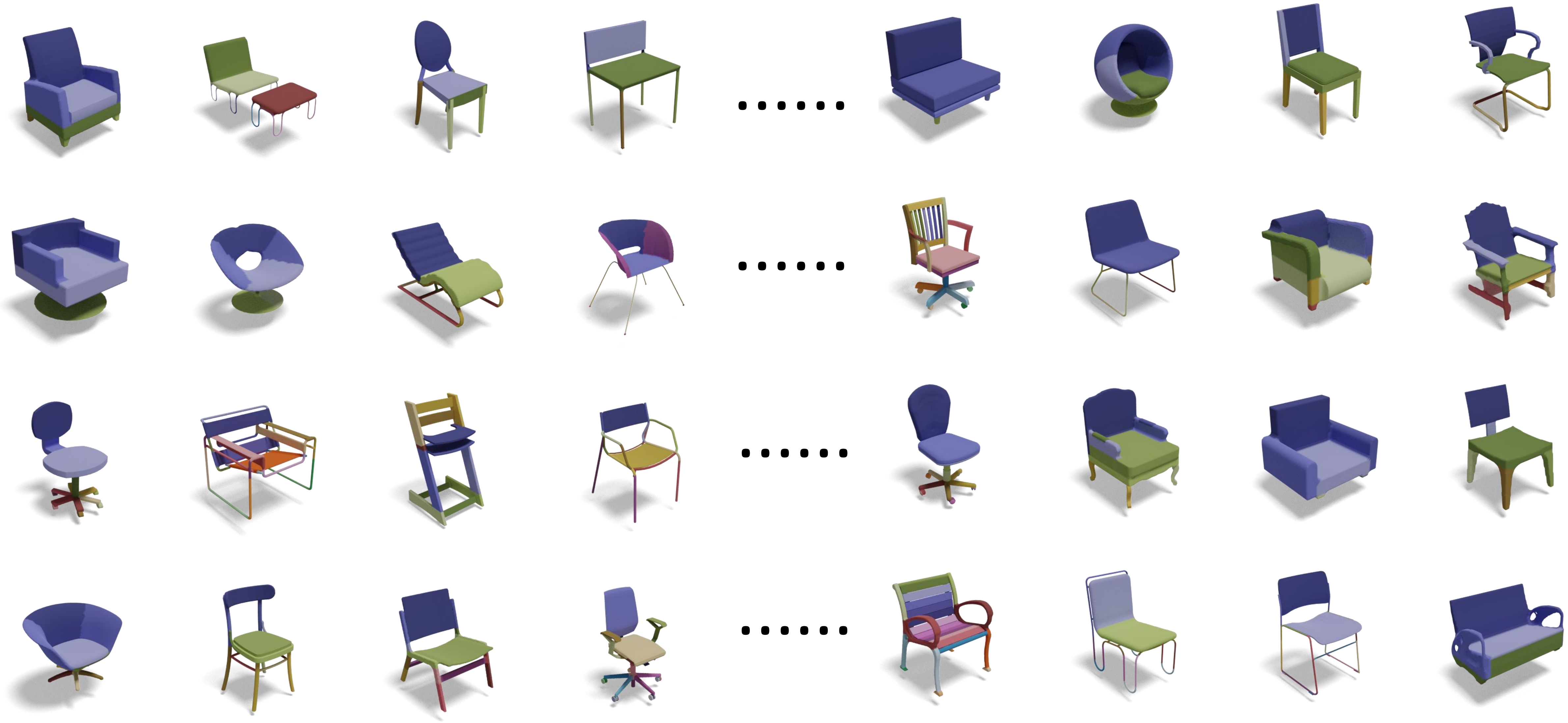
The fine-grained 3D part labeling challenges even the most advanced deep learning (DL) methods

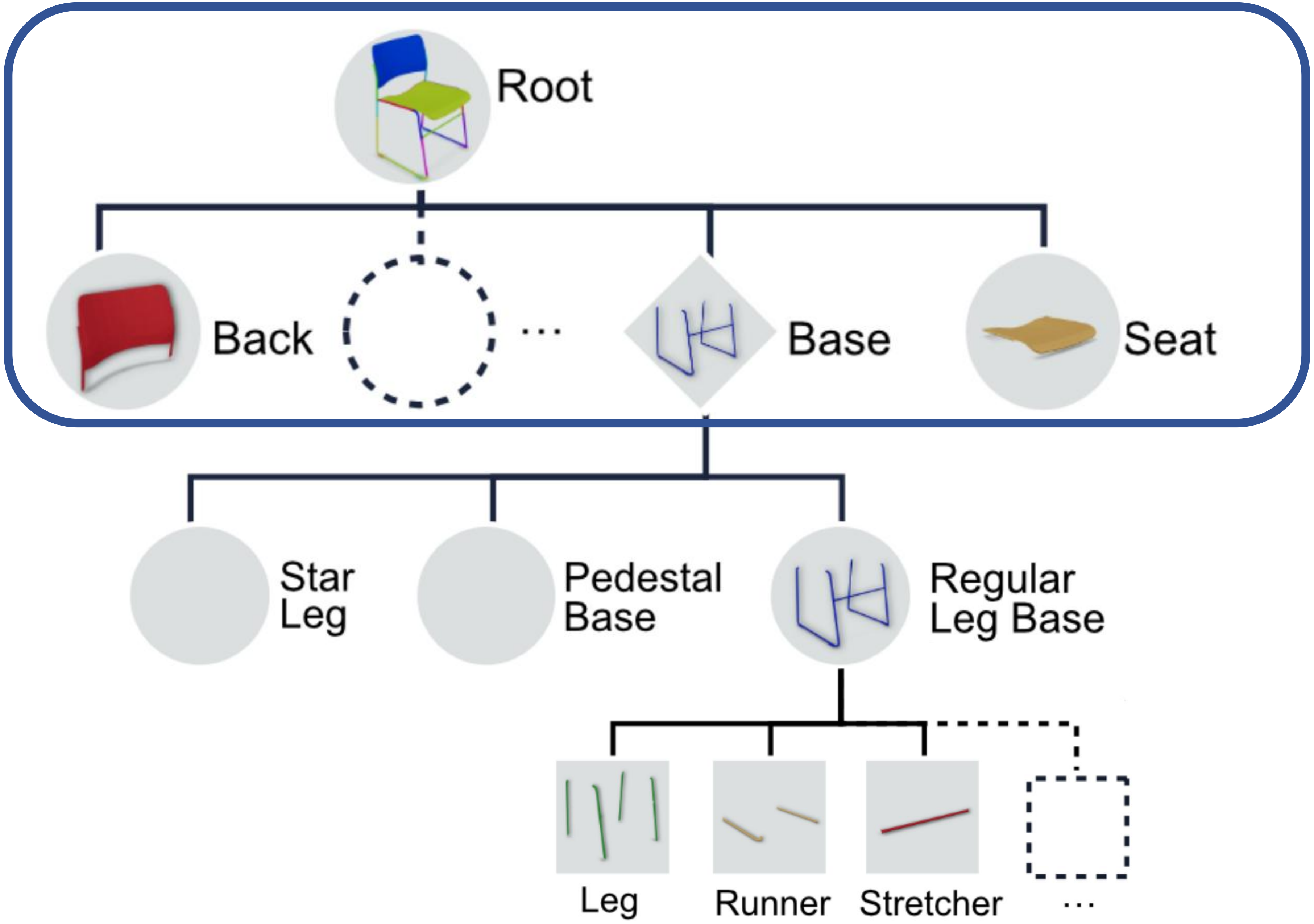
[1] The Neurally-Guided Shape Parser: Grammar-Based Labeling of 3D Shape Regions With Approximate Inference,CVPR 2022



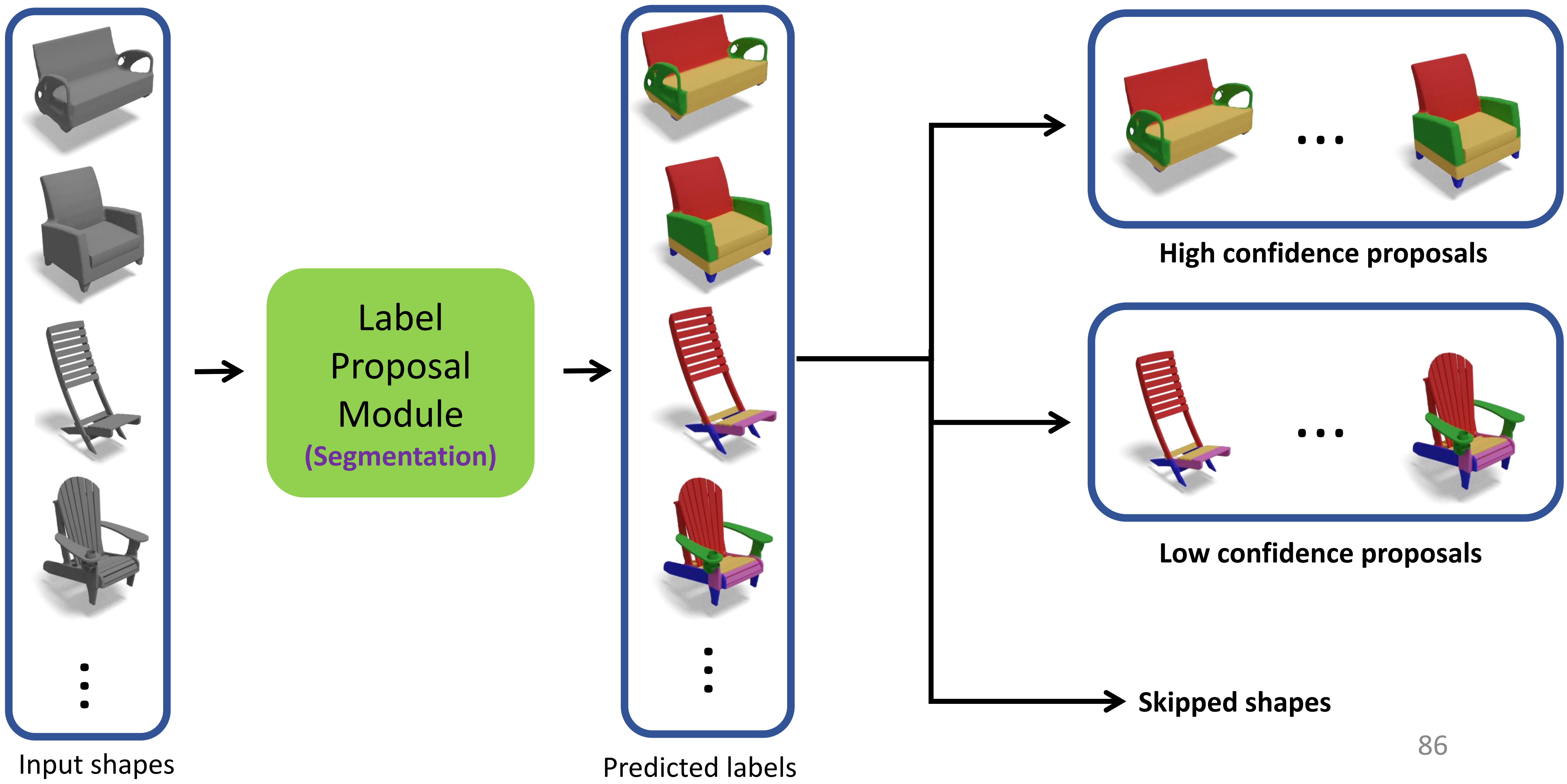
The first active learning framework for fine-grained 3D part labeling

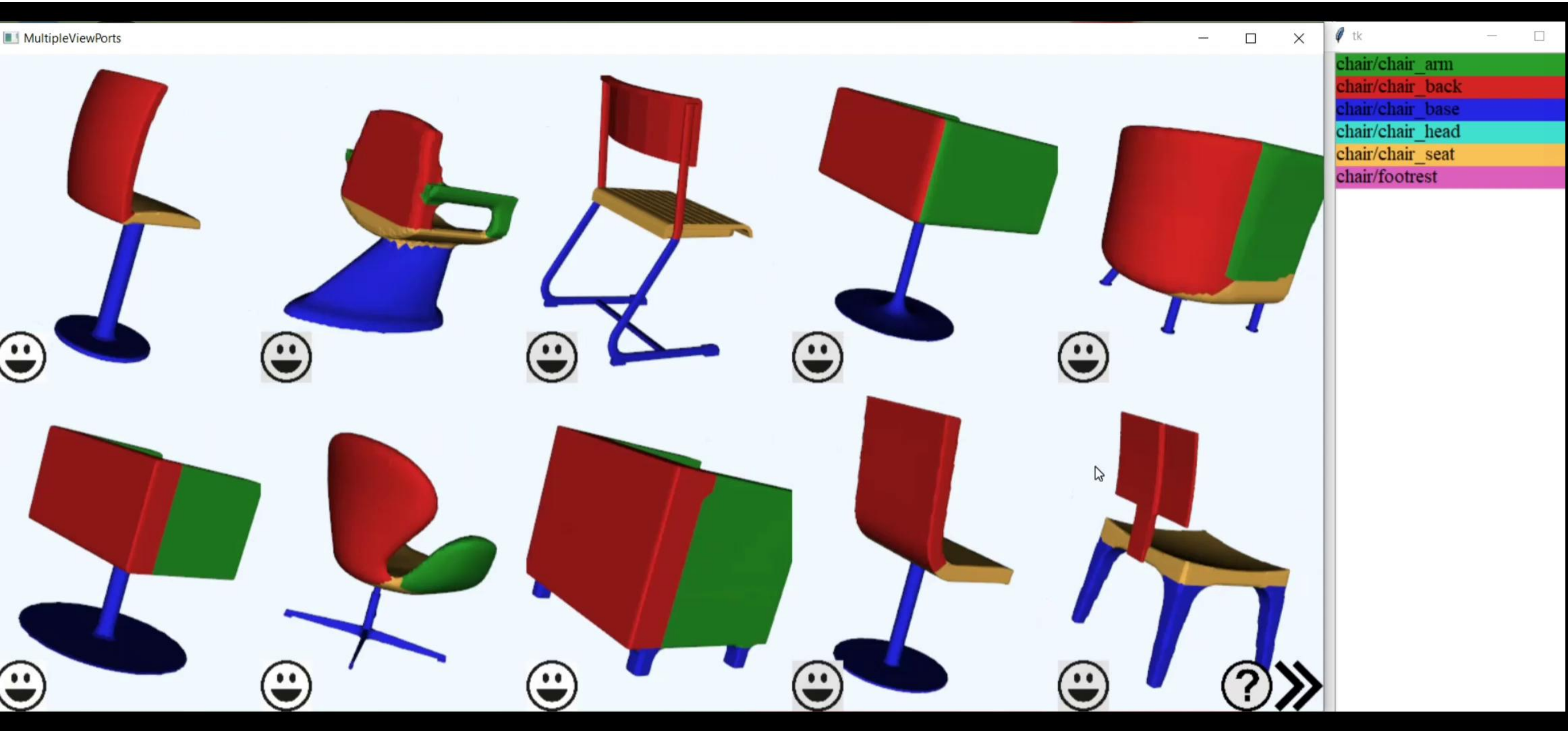


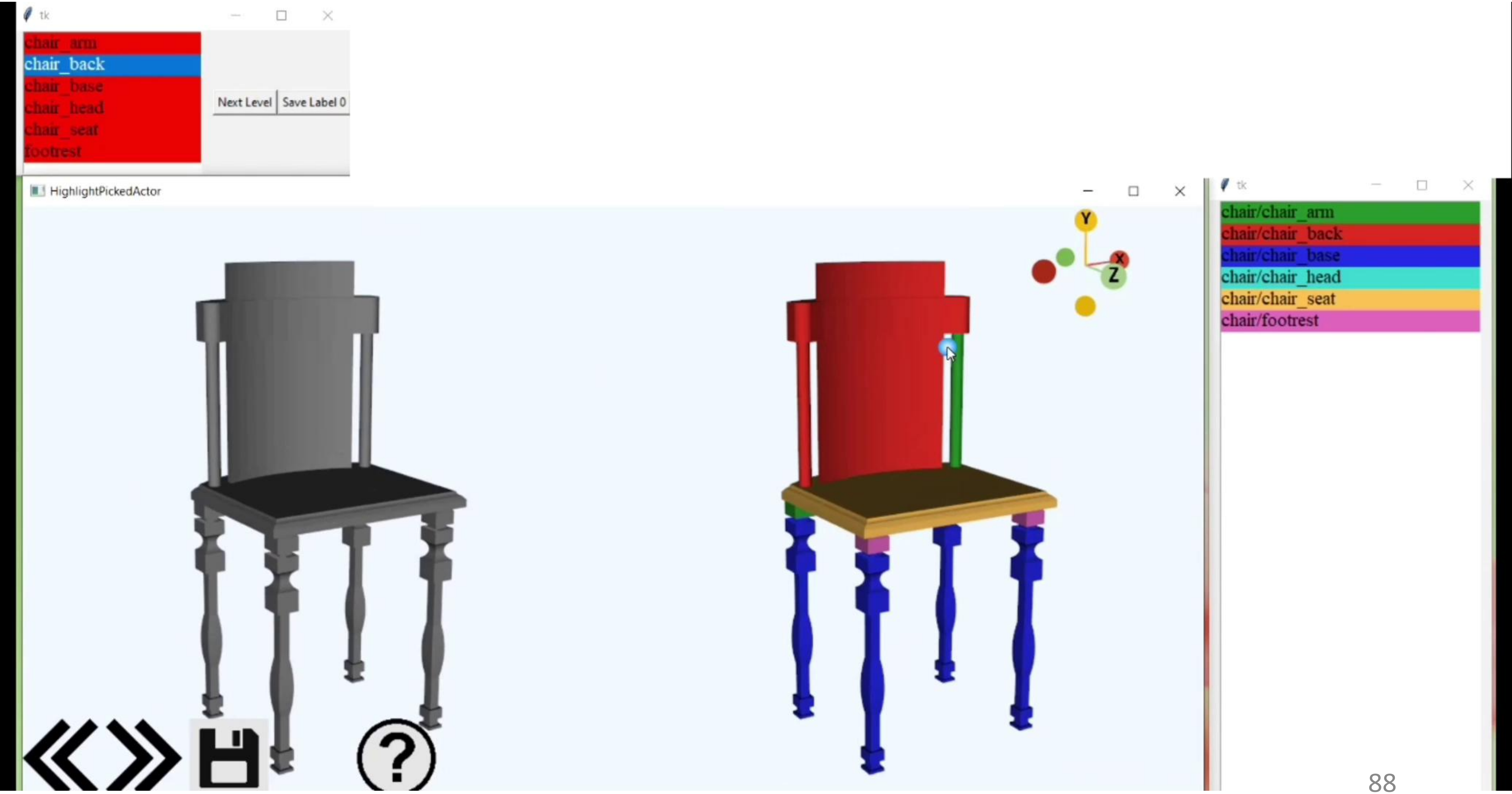



Root
Back
...
Base
Seat
Star Leg
Pedestal Base
Regular Leg Base
Leg
Runner
Stretcher
...

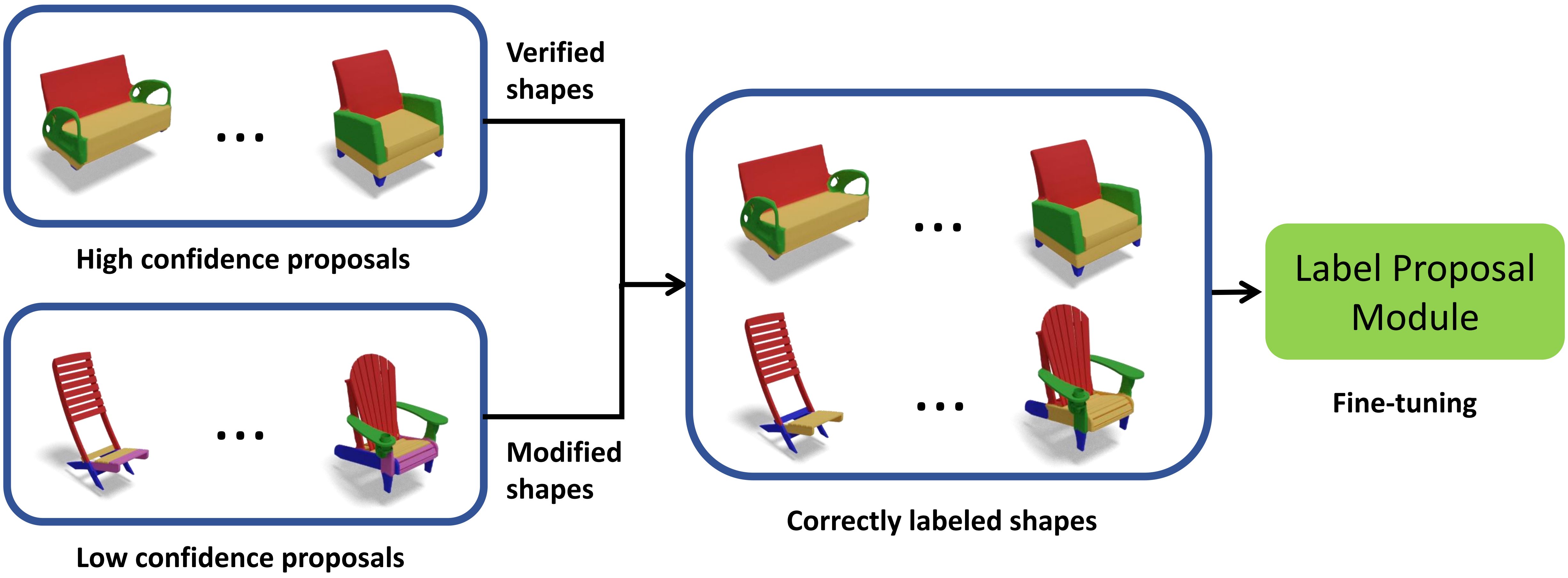
The first iteration at root node of chair category



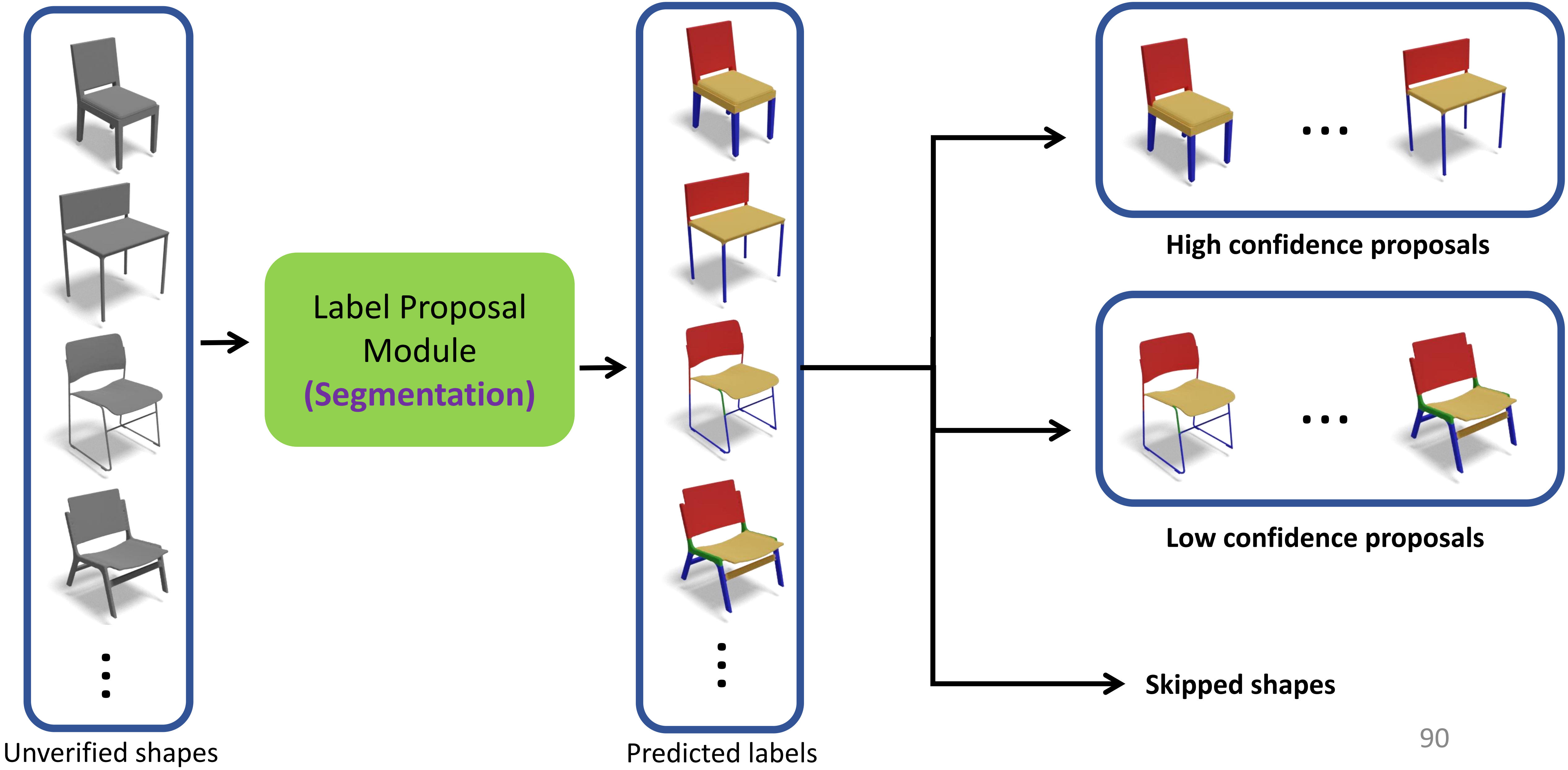

MultipleViewPorts
tk
chair/chair_arm
chair/chair_back
chair/chair_base
chair/chair_head
chair/chair_seat
chair/footrest


tk
chair_arm
chair_back
chair_base
chair_head
chair_seat
footrest
Next Level
Save Label 0
HighlightPickedActor
Y
X
Z
tk
chair/chair_arm
chair/chair_back
chair/chair_base
chair/chair_head
chair/chair_seat
chair/footrest

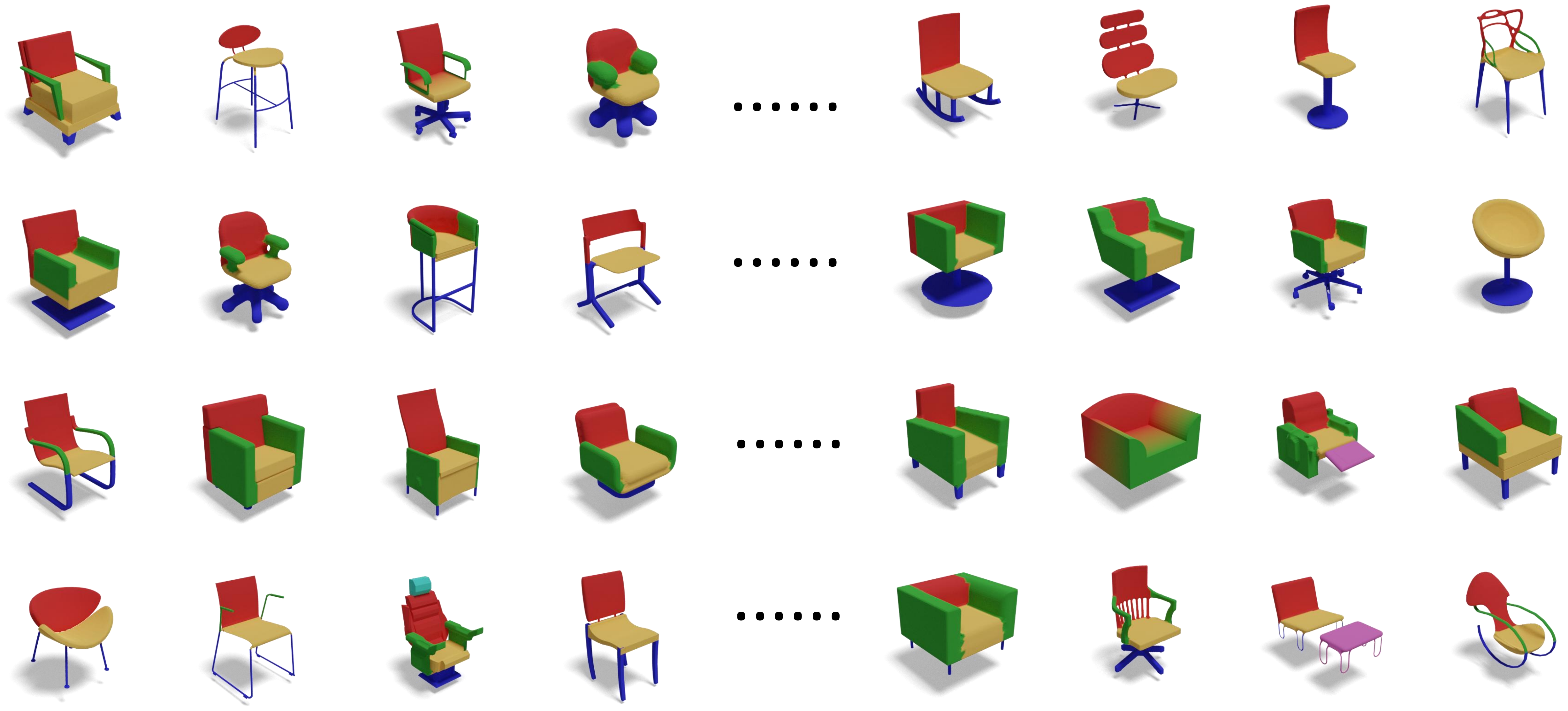
The first iteration is completed after fine-tuning

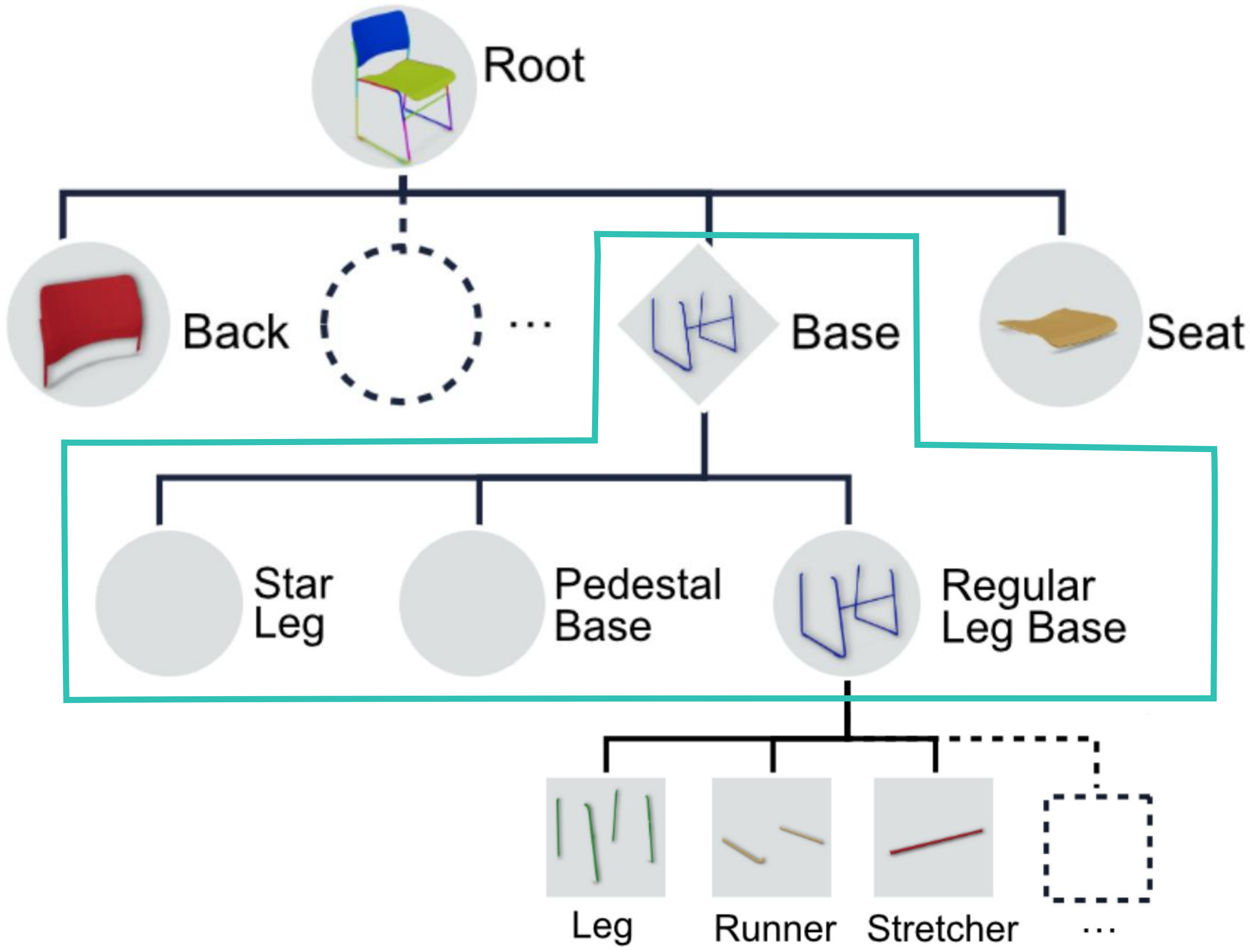


The second iteration at root node of chair category

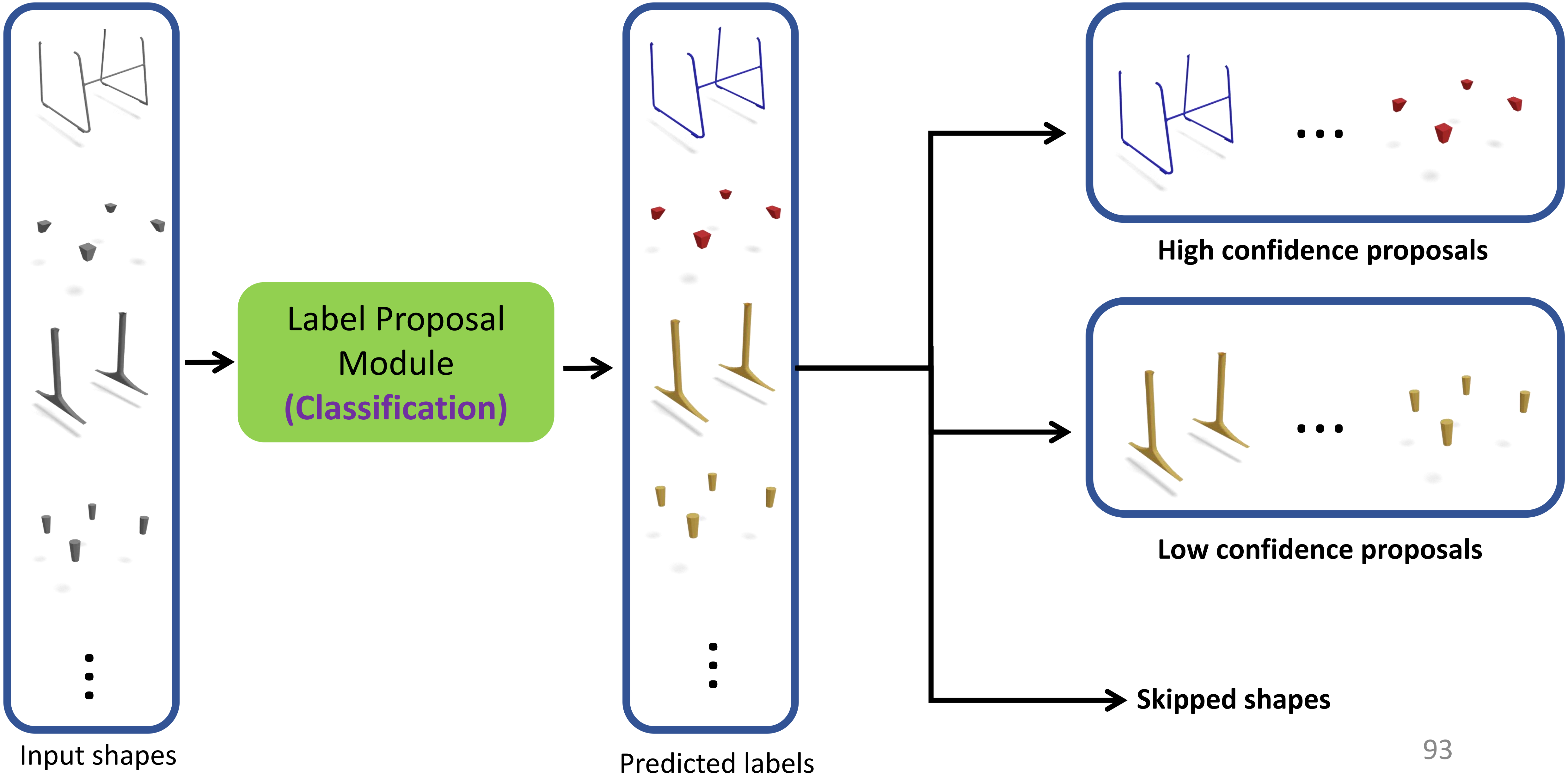


Final labeled results at root node

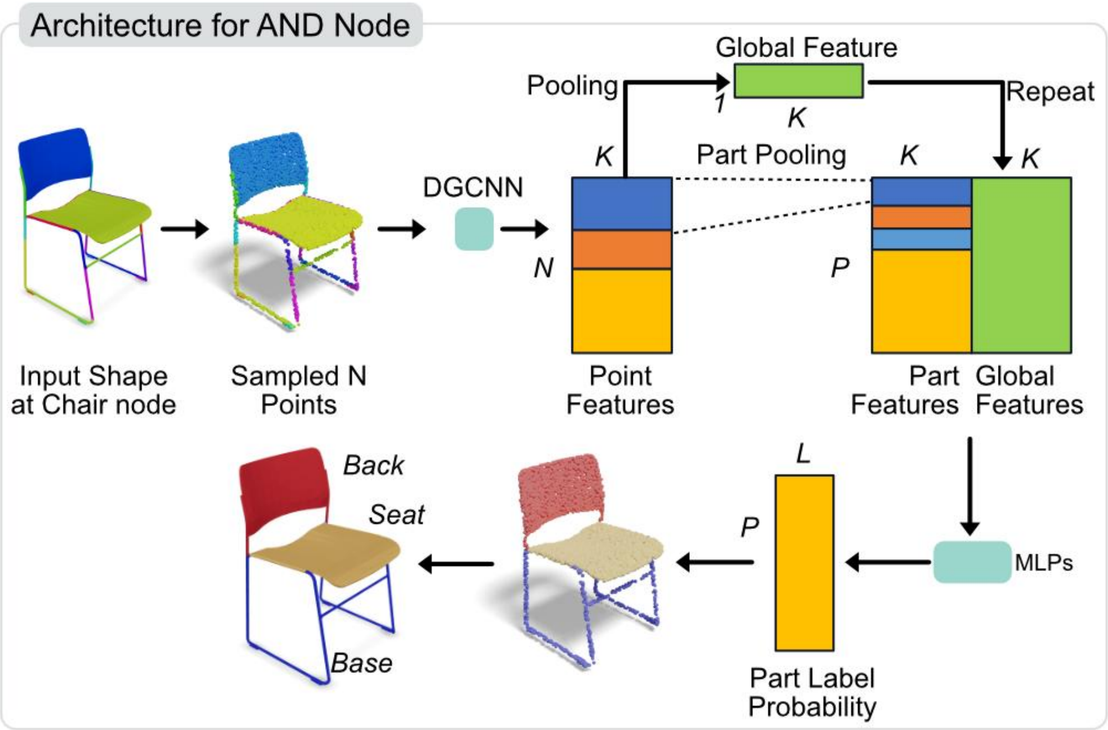



Root
Back
···
Base
Seat
Star Leg
Pedestal Base
Regular Leg Base
Leg
Runner
Stretcher
···

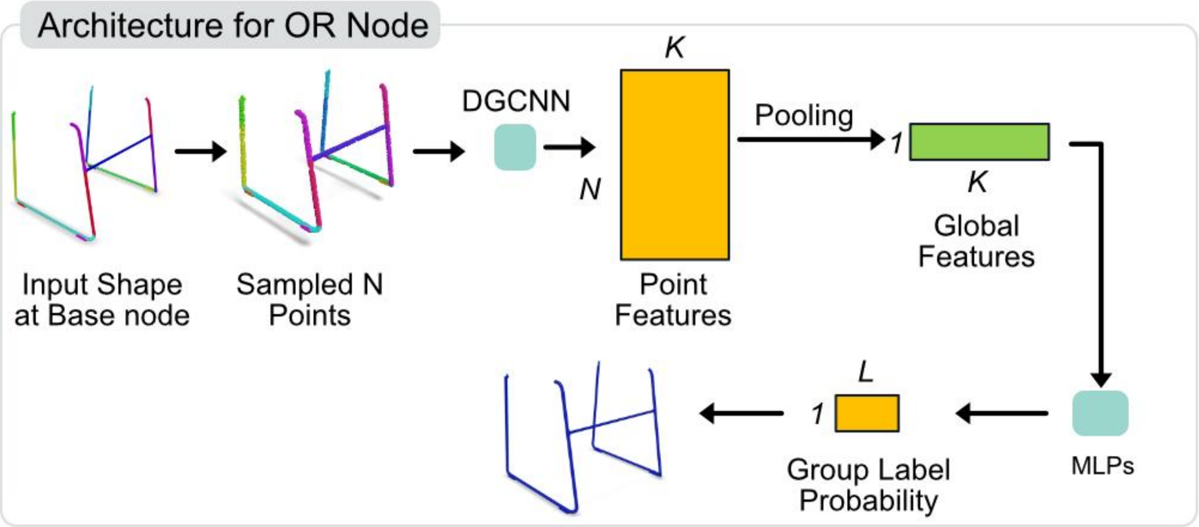
The first iteration at chair base node



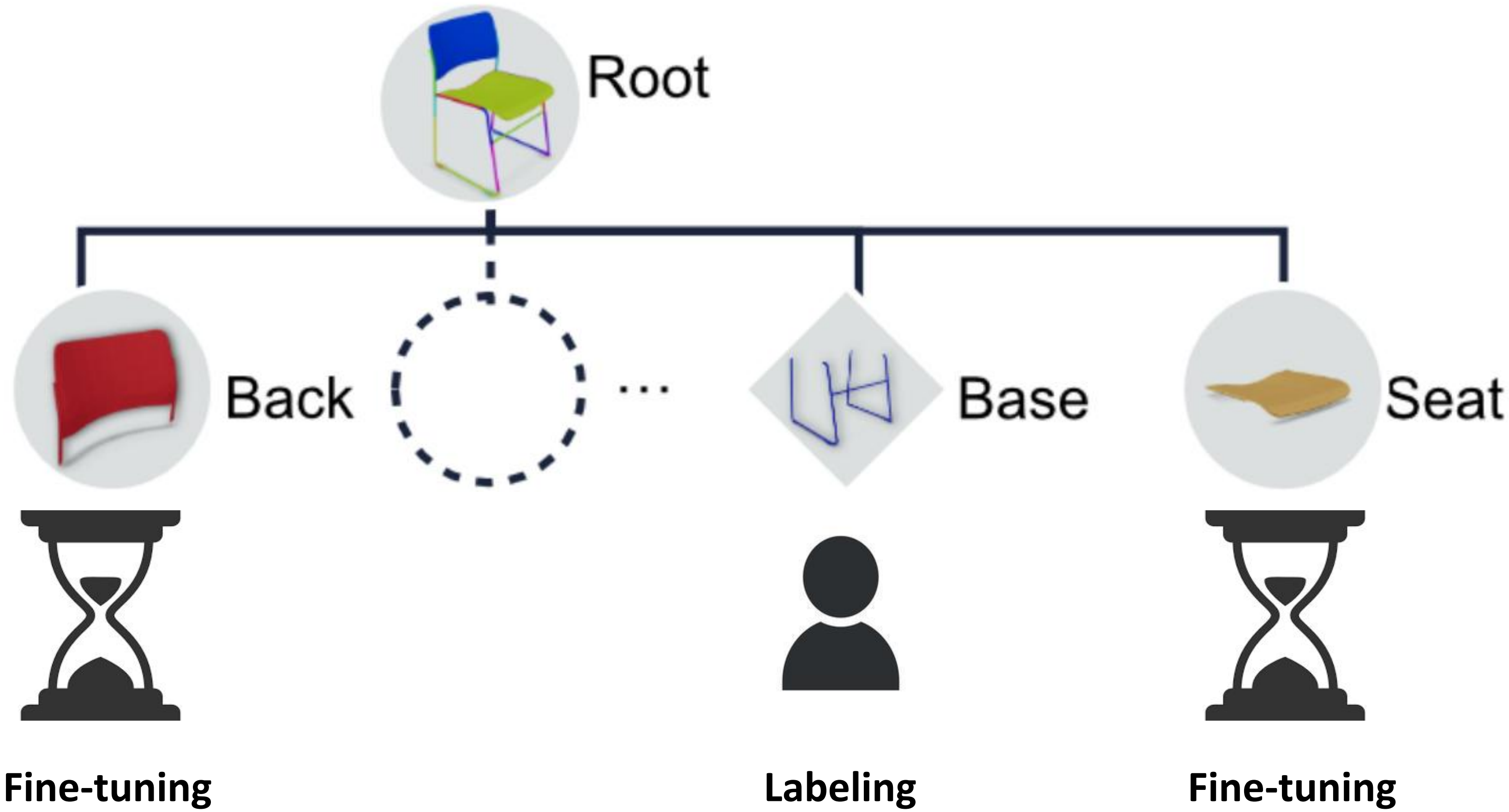
Proposal Modules



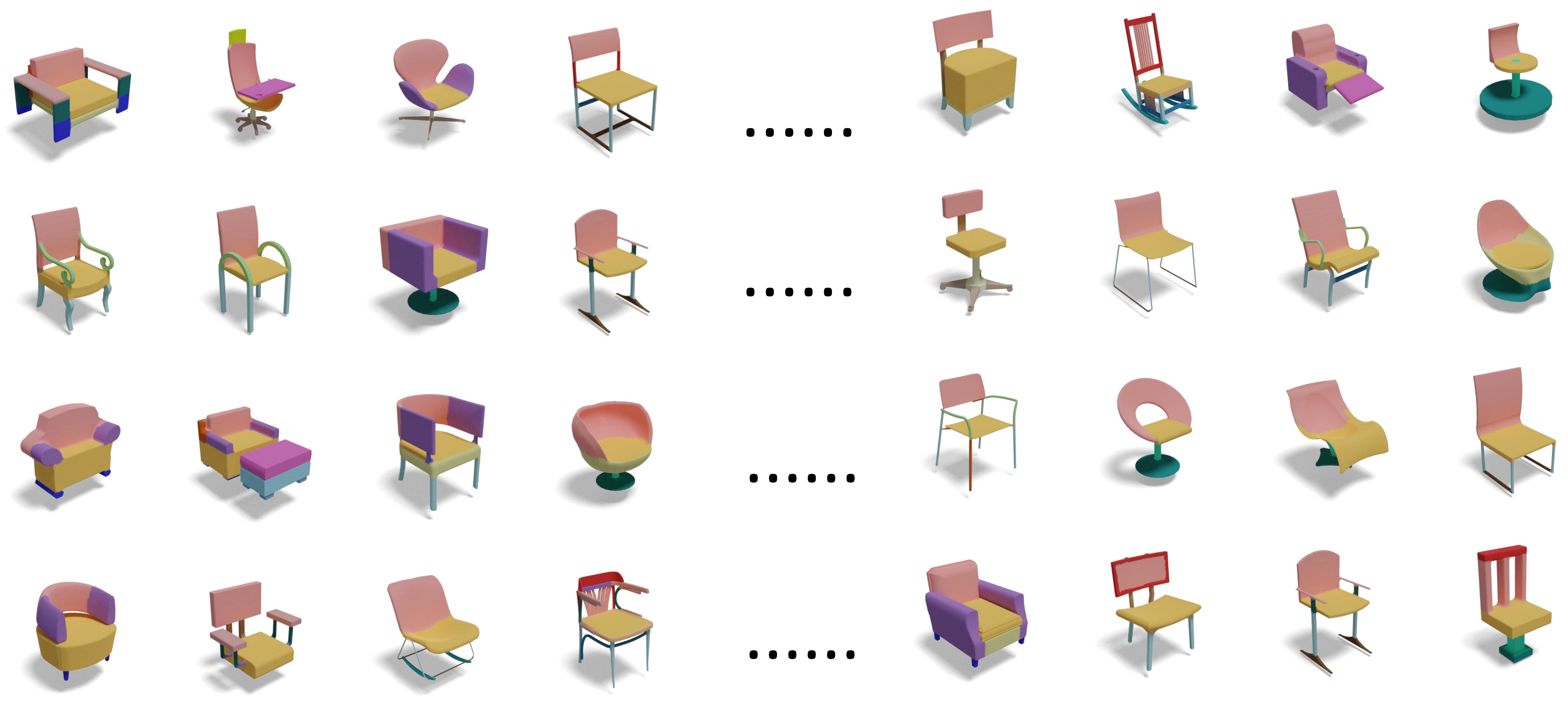
Segmentation



Classification


Root
Back
...
Base
Seat
Fine-tuning
Labeling
Fine-tuning

Final labeled results after finishing labeling at all internal nodes



Results

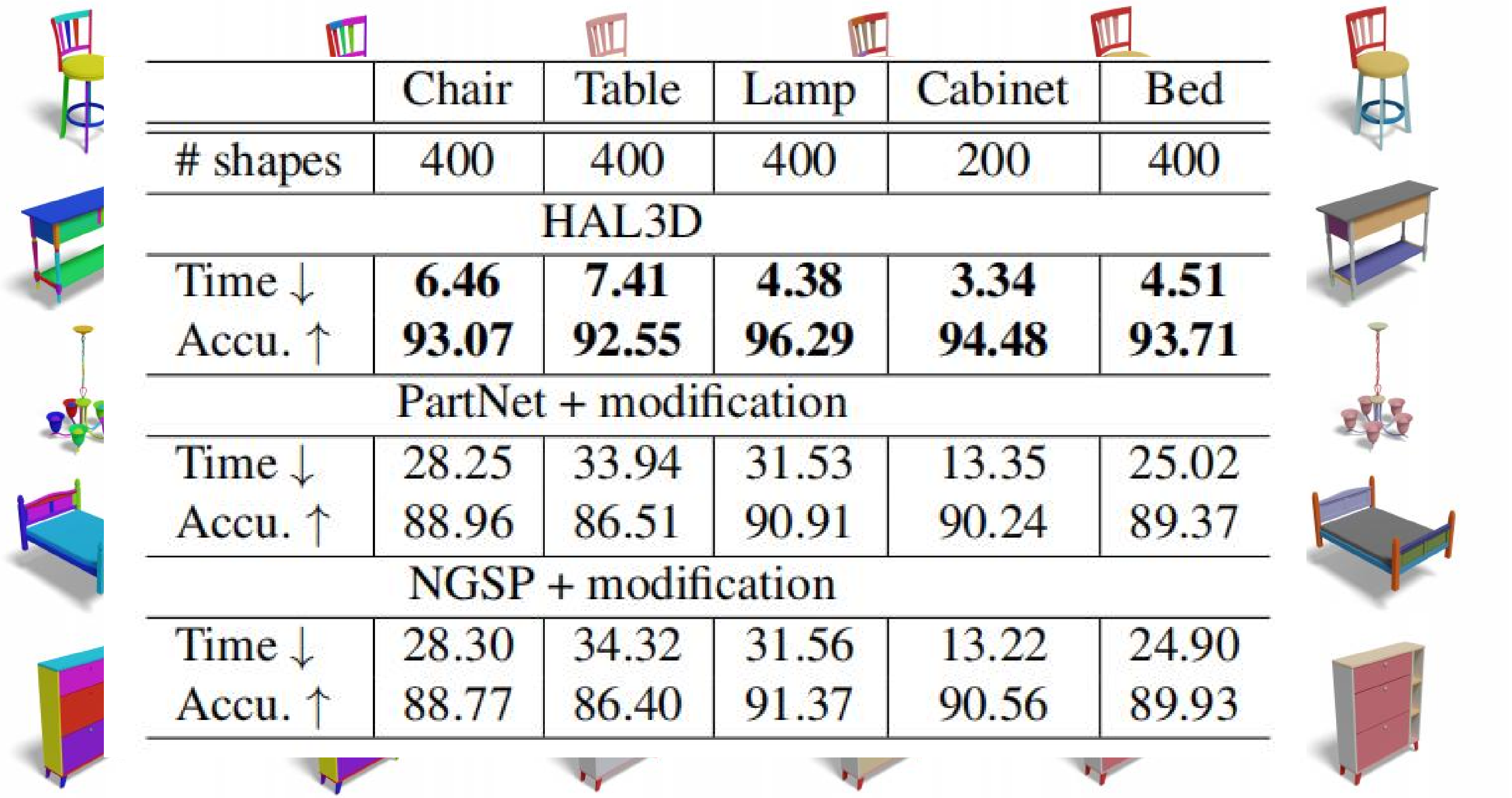
- Ablation study on Stanford PartNet chair dataset

Row ID	Prop.	Hier.	Sym.	AL	Lab-T↓	Accu↑
2nd	-	-	-	-	22.05	89.16
3rd	✓	-	-	-	8.65	88.53
4th	✓	✓	✓	-	6.37	93.87
5th	✓	-	✓	✓	5.99	89.84
6th	✓	✓	-	✓	5.21	93.45
7th	✓	✓	✓	✓	**4.34**	**94.13**

Prop: proposal module
Hier: hierarchical labeling
Sym: symmetry constraint
AL: active learning
Lab-T: human labeling time
Accu: labeling accuracy

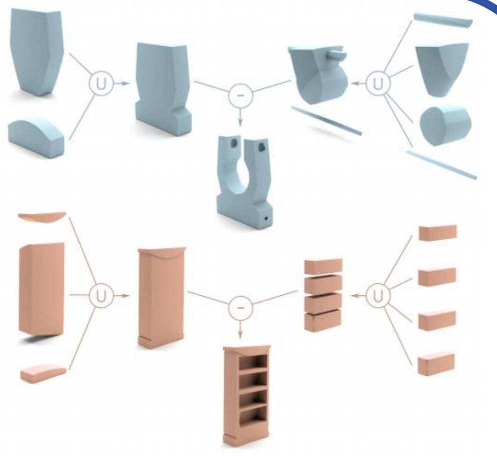
Results

- Results on the ABO dataset

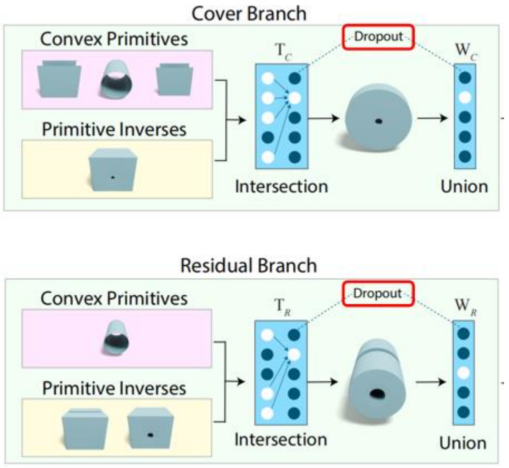


	Chair	Table	Lamp	Cabinet	Bed
# shapes	400	400	400	200	400
HAL3D					
Time ↓	**6.46**	**7.41**	**4.38**	**3.34**	**4.51**
Accu. ↑	**93.07**	**92.55**	**96.29**	**94.48**	**93.71**
PartNet + modification					
Time ↓	28.25	33.94	31.53	13.35	25.02
Accu. ↑	88.96	86.51	90.91	90.24	89.37
NGSP + modification					
Time ↓	28.30	34.32	31.56	13.22	24.90
Accu. ↑	88.77	86.40	91.37	90.56	89.93

Limitations



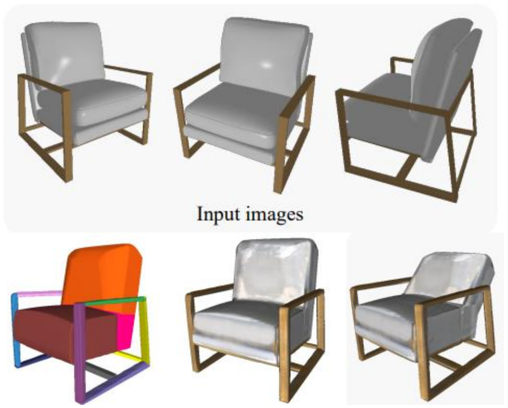
CAPRI-Net (CVPR 2022)



D²CSG (NeurIPS 2023)

Limited primitives

Learning surface prior

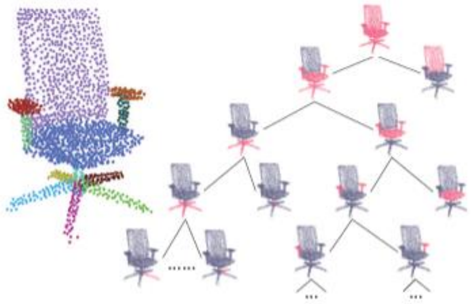


DPA-Net (ECCV 2024)

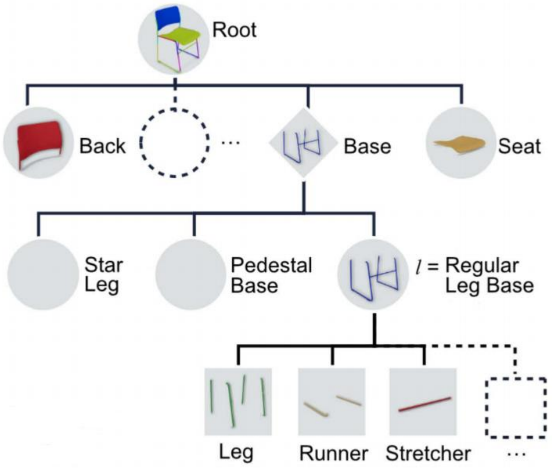
Sweep-Net (ECCV 2024)

Camera pose required

Test-time fine-tuning



PartNet (CVPR 2019)



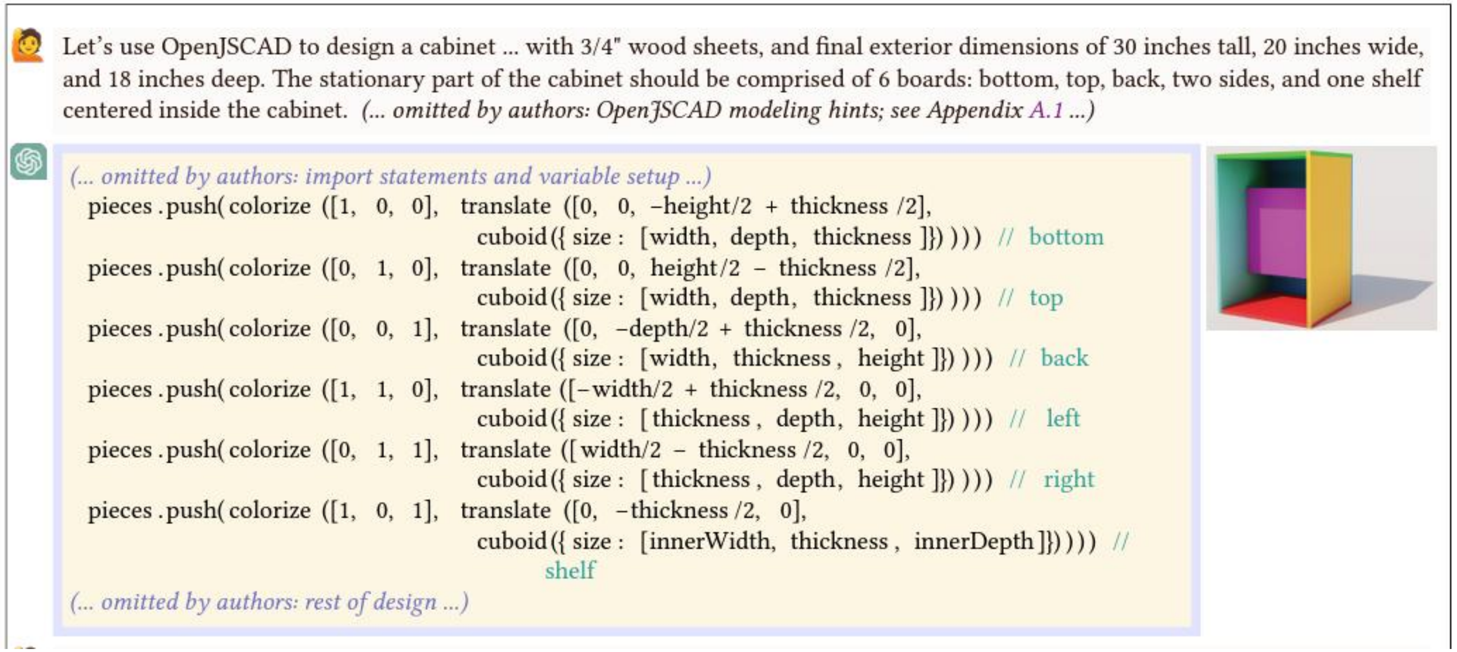
HAL3D (ICCV 2023)

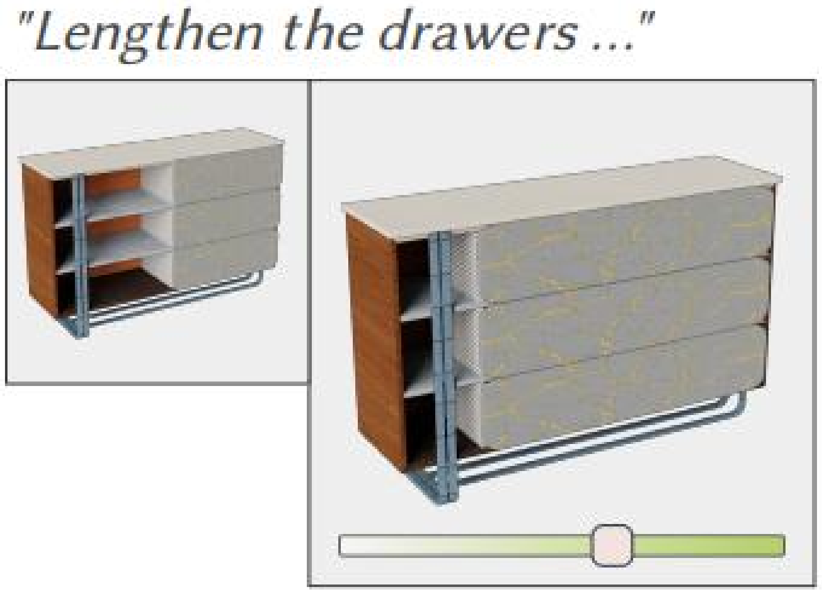
No split operation

Inter-shape correspondence

Future Direction

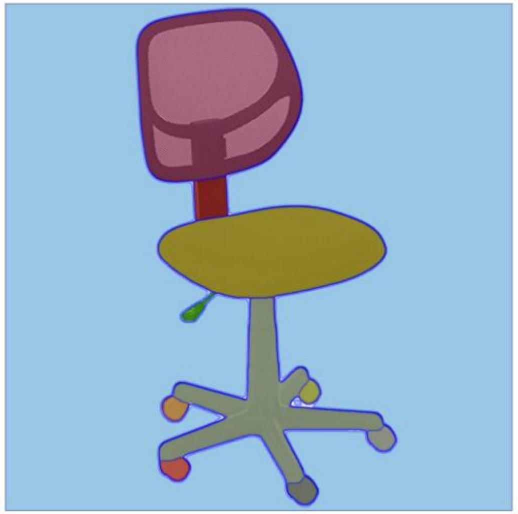
- Large language model + CAD [1]



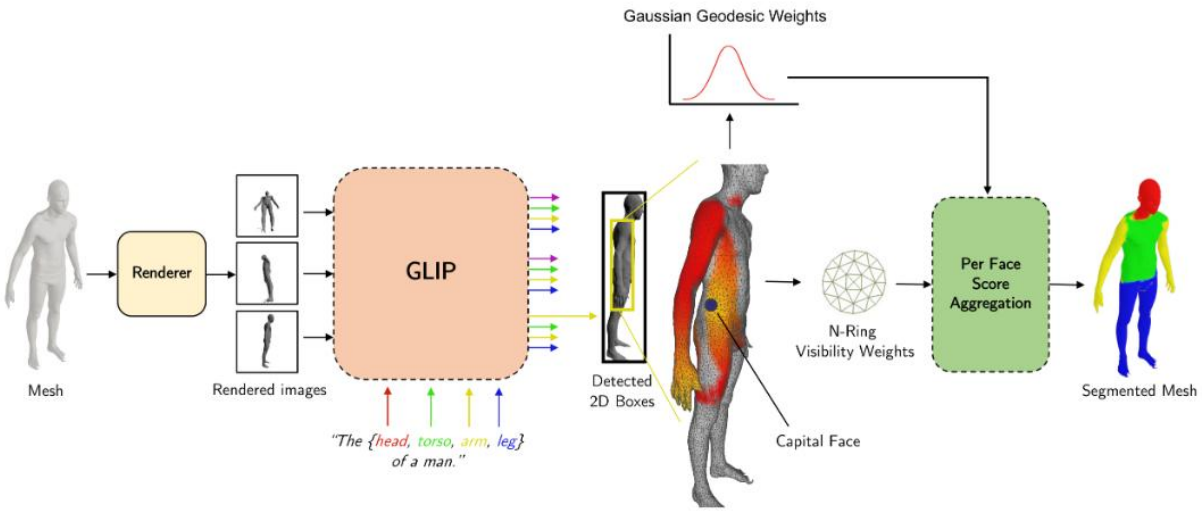


Future Direction

- Large image foundation model + 3D segmentation

SAM [1]

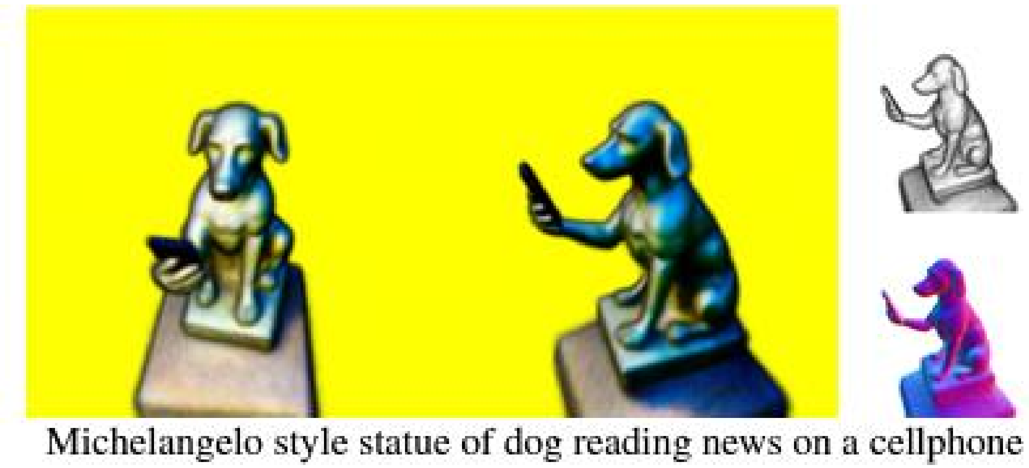


SATR [2]

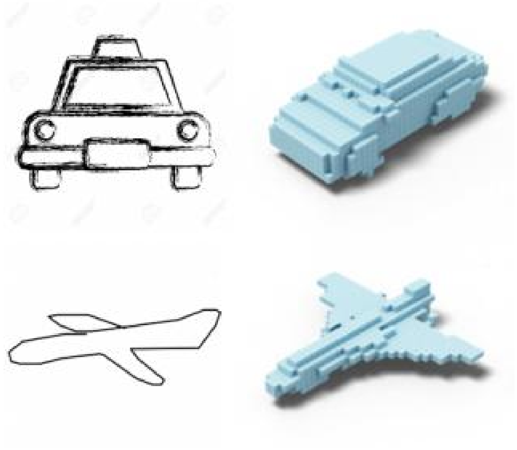
[1] Segment anything
[2] SATR: Zero-Shot Semantic Segmentation of 3D Shapes

Future Direction

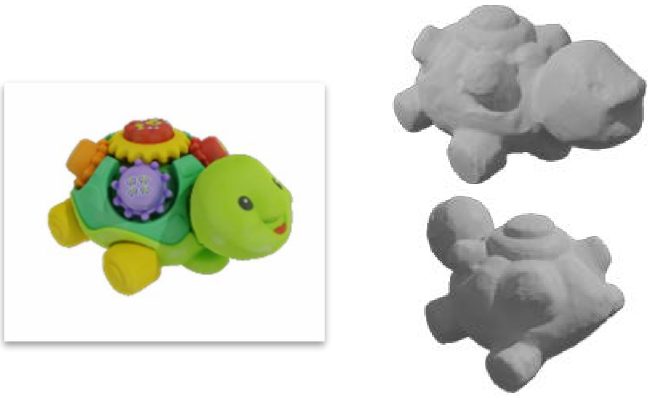
- 3D CAD model generation from multi-modality data



Text-to-3D [1]



Sketch-to-3D [2]



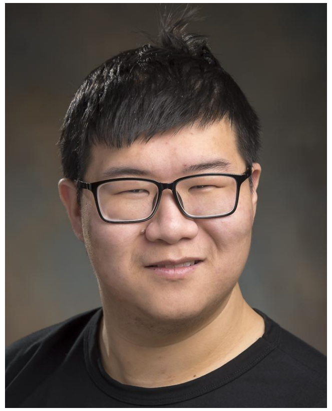
Single image-to-3D [3]

[1] Dreamfusion: Text-to-3d using 2d diffusion
[2] Sketch-A-Shape: Zero-Shot Sketch-to-3D Shape Generation
[3] MVDiffusion++: A Dense High-resolution Multi-view Diffusion Model for Single or Sparse-view 3D Object Reconstruction

Acknowledgement

Hao (Richard) Zhang

Ali Mahdavi Amiri



Zhiqin Chen

Mingrui Zhao

Manyi Li

Qimin Chen

Maham Tanveer

Yizhi Wang

Gruvi Lab, Simon Fraser University

Acknowledgement


Brian Jackson

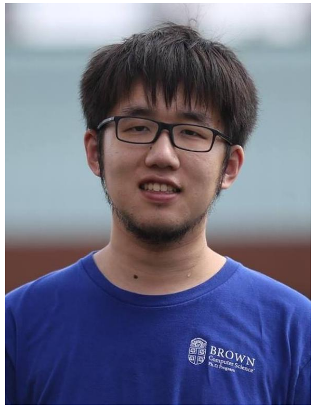

Eric Bennett


Hooman Shayani


Yiming Qian


Francisca Gil-Ureta


Xu Zhang



Kai Wang


Aditya Sanghi

Amazon

Autodesk AI Lab

Thank you!