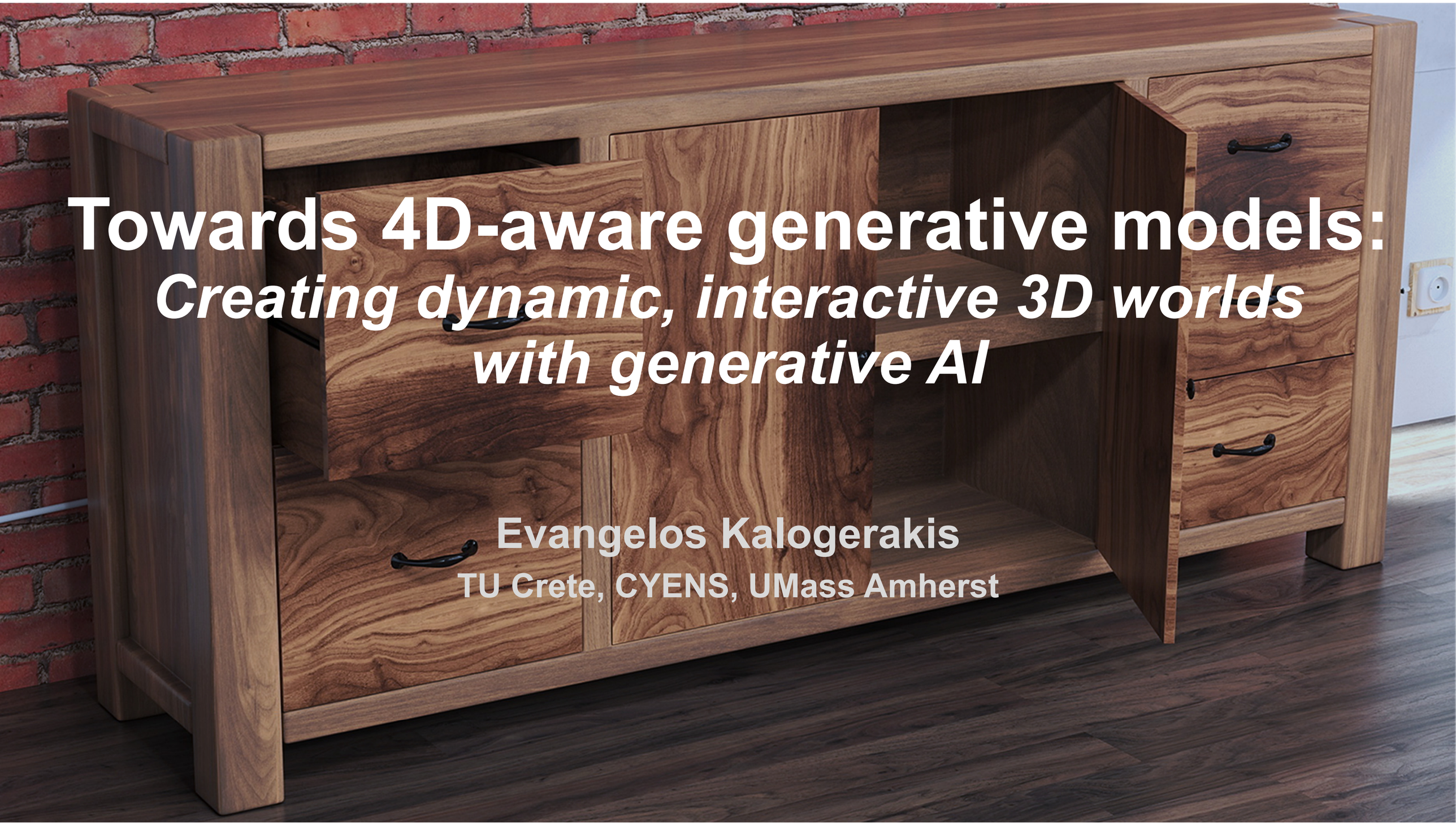
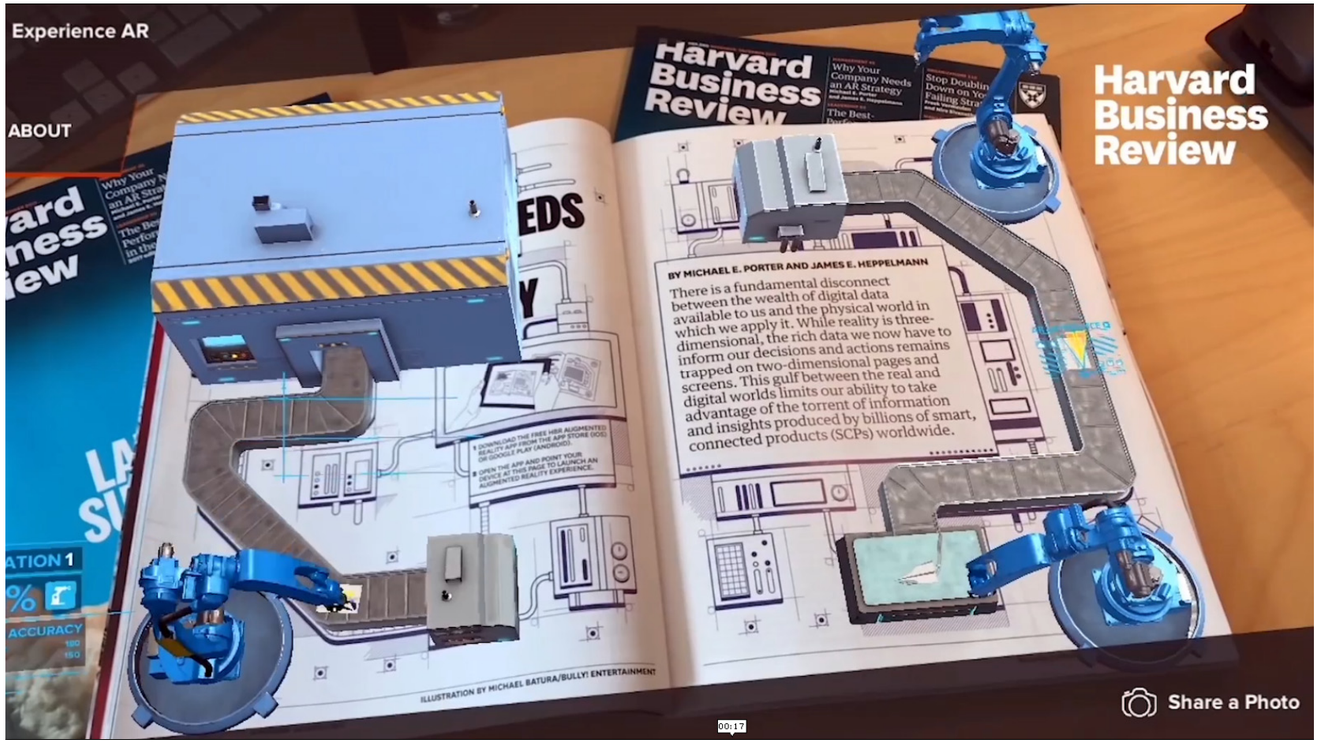


Towards 4D-aware generative models:
Creating dynamic, interactive 3D worlds with generative AI

Evangelos Kalogerakis
TU Crete, CYENS, UMass Amherst

Living in the 3D Digital Era!

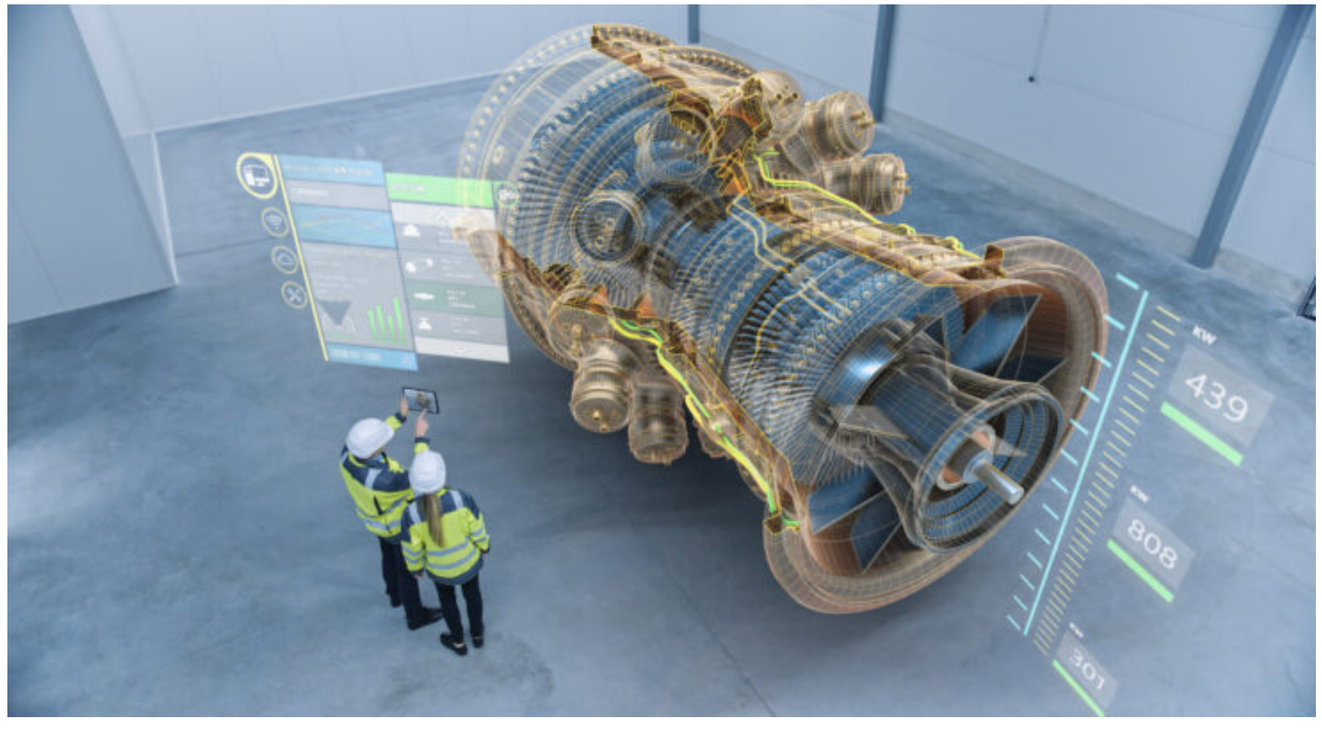
The Real Significance Of AR, VR And Mixed Reality [forbes.com]



Why Every Organization Needs an Augmented Reality Strategy [hbr.org]

3D **Interactive** Environments

A Simulated Part-based Interactive Environment [https://sapien.ucsd.edu/]



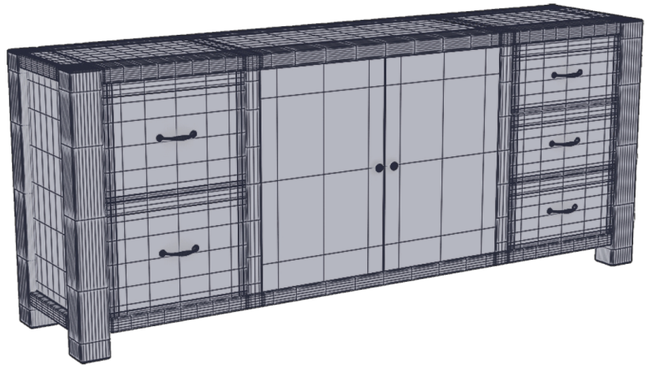
RoboDK - Guide to 3D Simulation Software [robodk.com]

3D **Interactive** Social Media & Entertainment

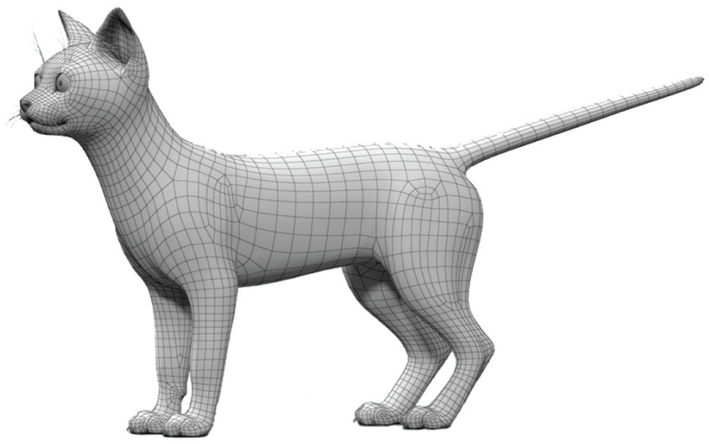
Entering The Metaverse [popularmechanics.com]

Animated Meerkat [Weta Digital / Epic Games]

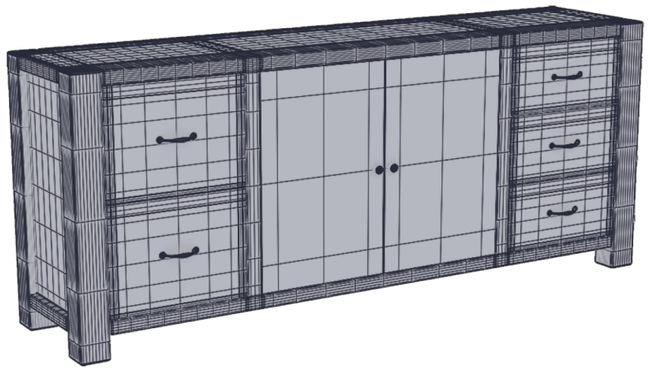
Crafting **interactive** 3D models



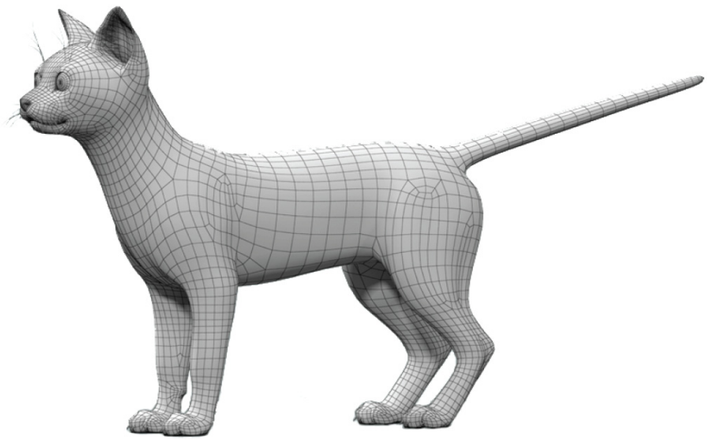
3D Geometry

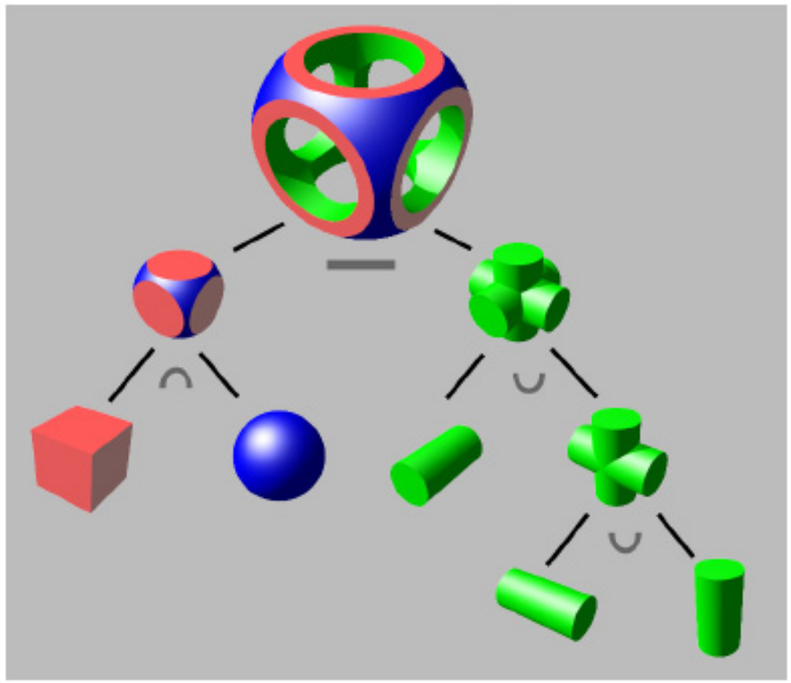


Crafting **interactive** 3D models

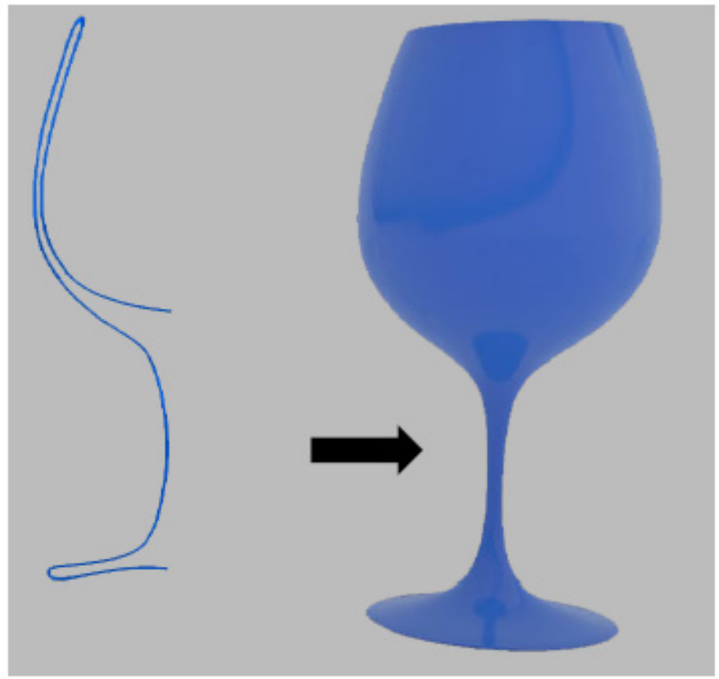


3D Geometry



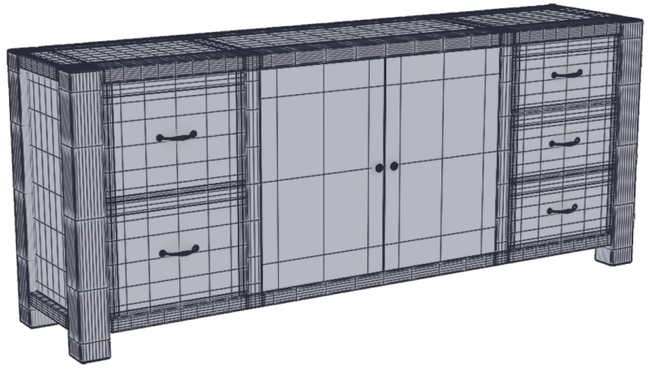


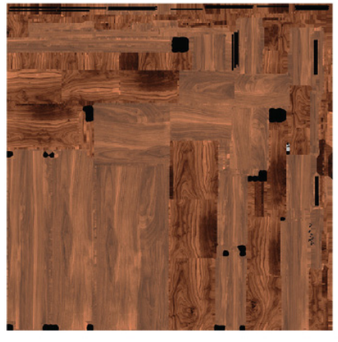
Constructive Solid Geometry



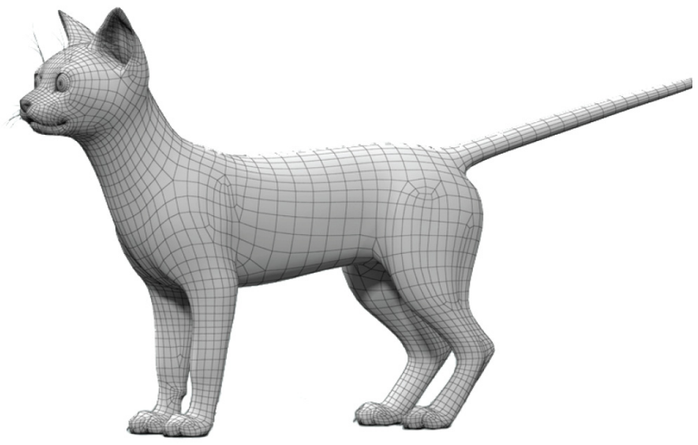
Parametric curves / surfaces

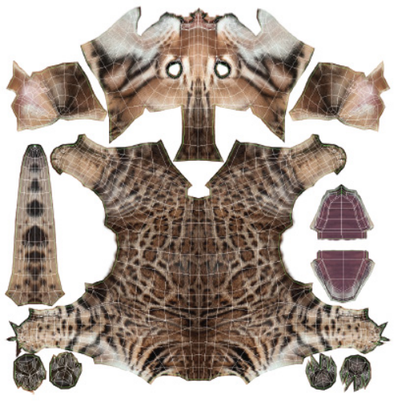
Crafting **interactive** 3D models



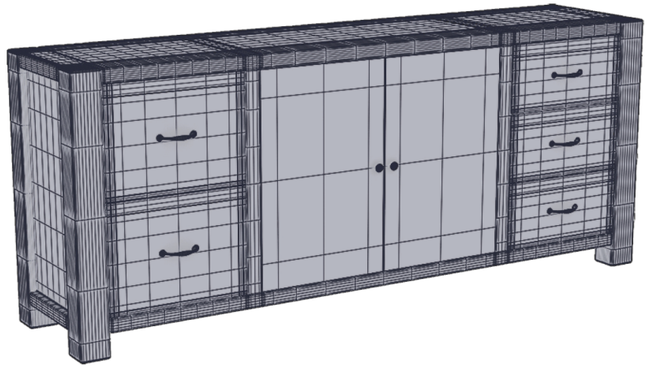


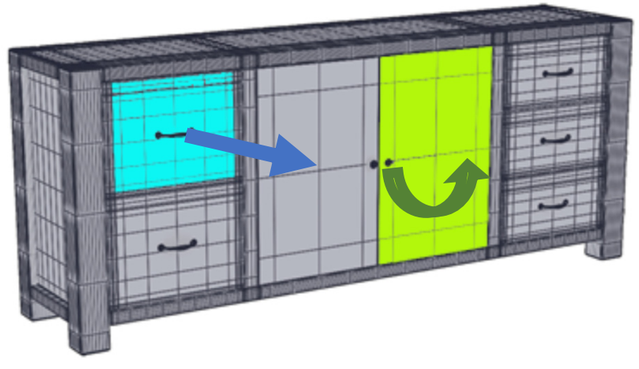
3D Geometry → **Textures**



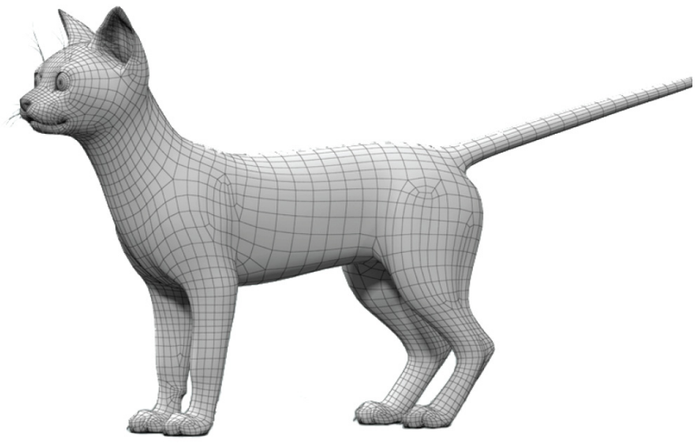


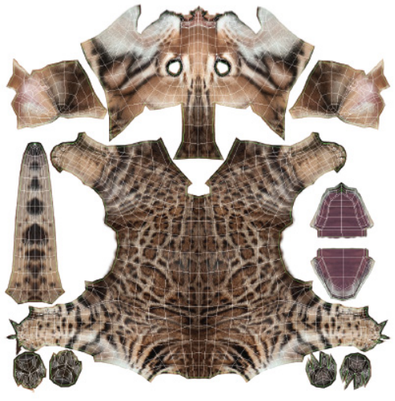
Crafting **interactive** 3D models



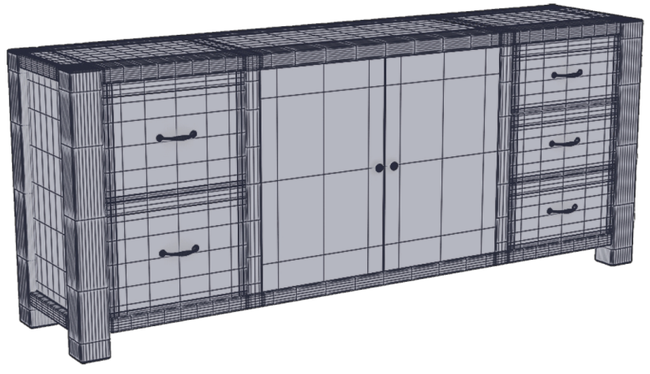


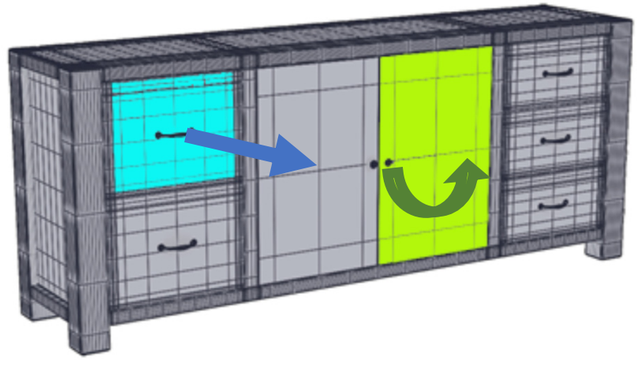
3D Geometry → **Textures** → **Motion controller & attributes**



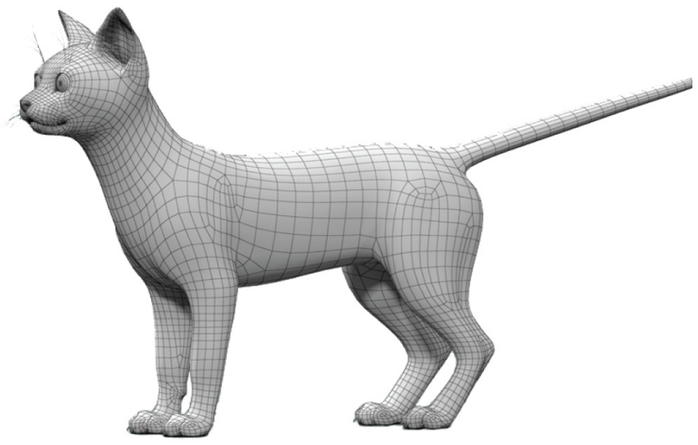


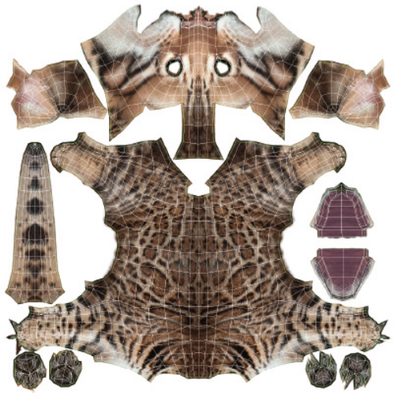
Crafting **interactive** 3D models

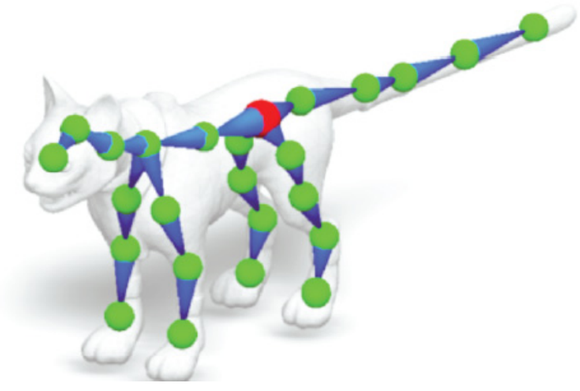




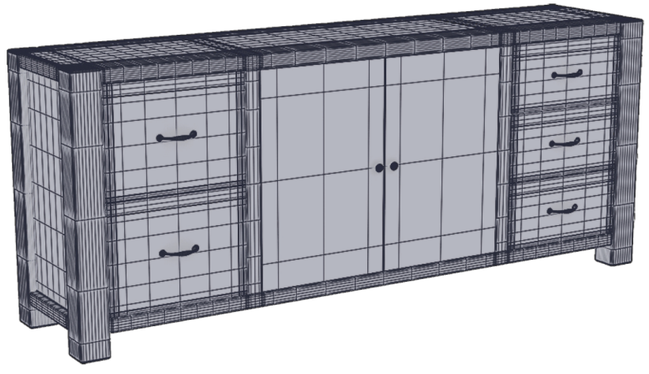
3D Geometry → **Textures** → **Motion controller & attributes**

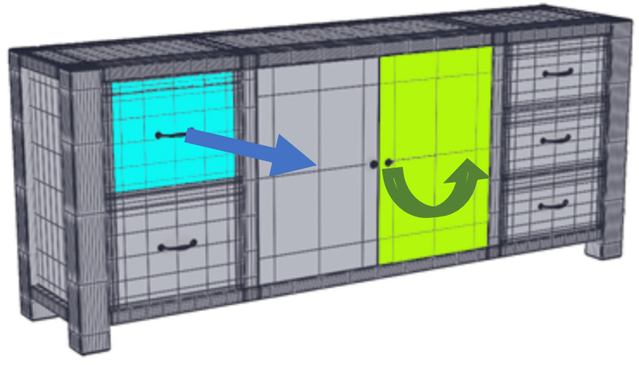




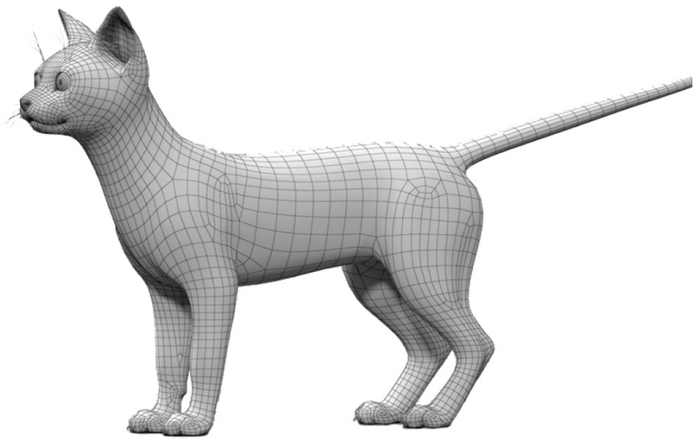


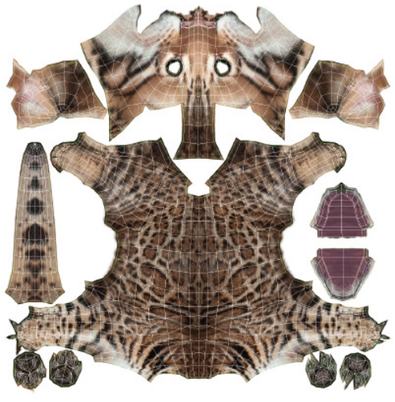
Crafting **interactive** 3D models

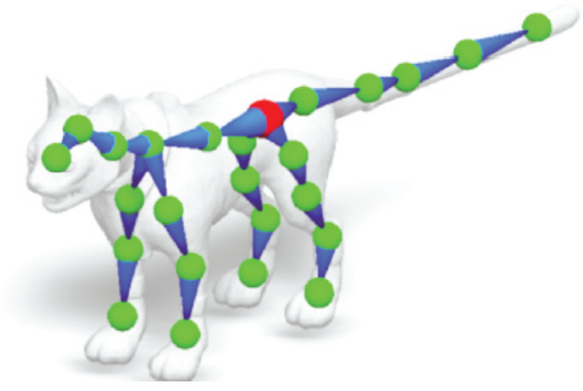




3D Geometry → **Textures** → **Motion controller & attributes** → **Interactive Model**

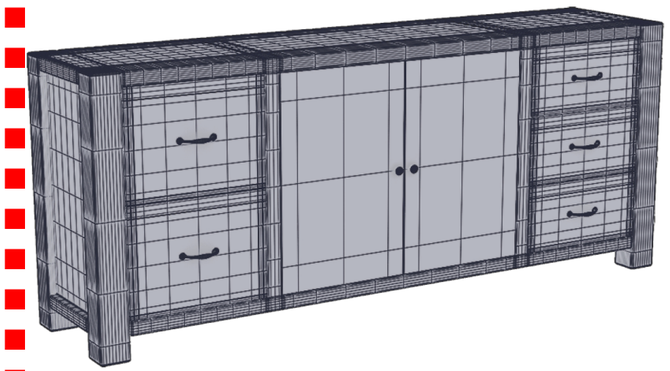


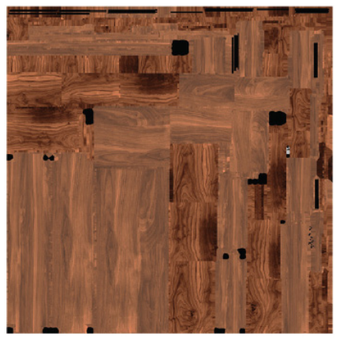


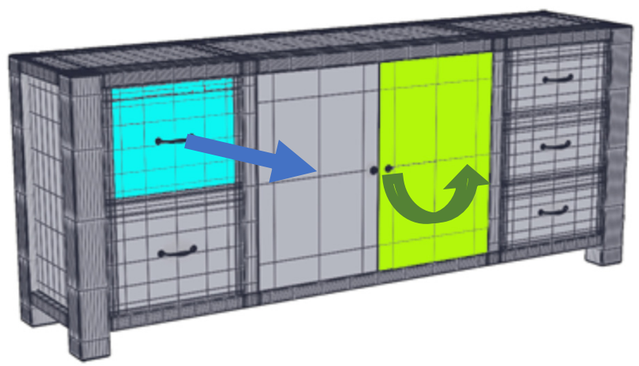


[models from turbosquid]

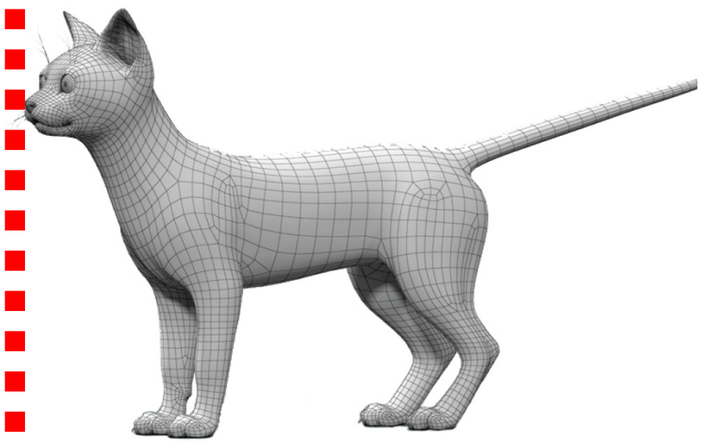
Lots of AI progress here!

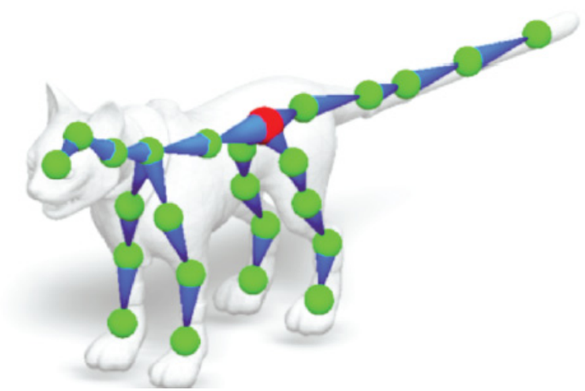






3D Geometry → Textures → Motion controller & attributes → Interactive Model





[models from turbosquid]

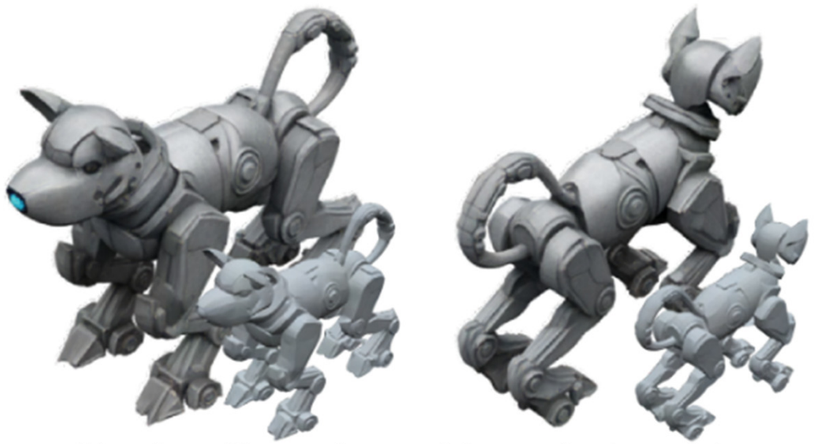
Neural 3D Reconstruction & Generation

Portable transistor radio, dark cover, speaker grille, brand logo on front.

Metallic dog-like robot with articulated legs and futuristic design elements.

Neural 3D Reconstruction & Generation

Portable transistor radio, dark cover, speaker grille, brand logo on front.

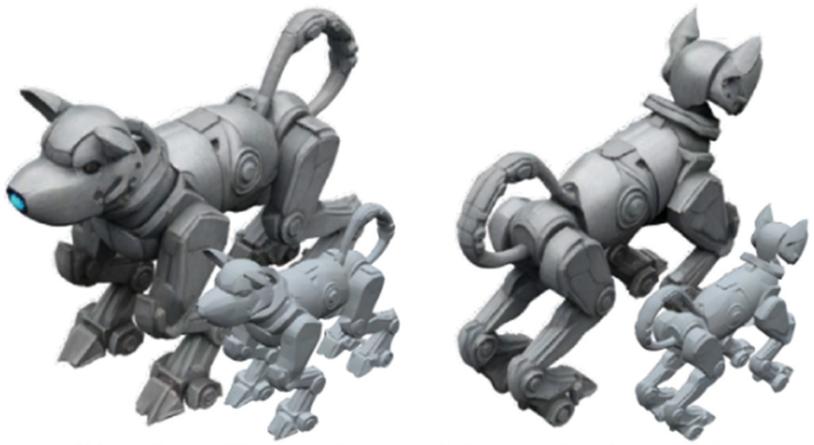


Metallic dog-like robot with articulated legs and futuristic design elements.

[TRELLIS, CVPR 2025]

Neural 3D Reconstruction & Generation

Portable transistor radio, dark cover, speaker grille, brand logo on front.

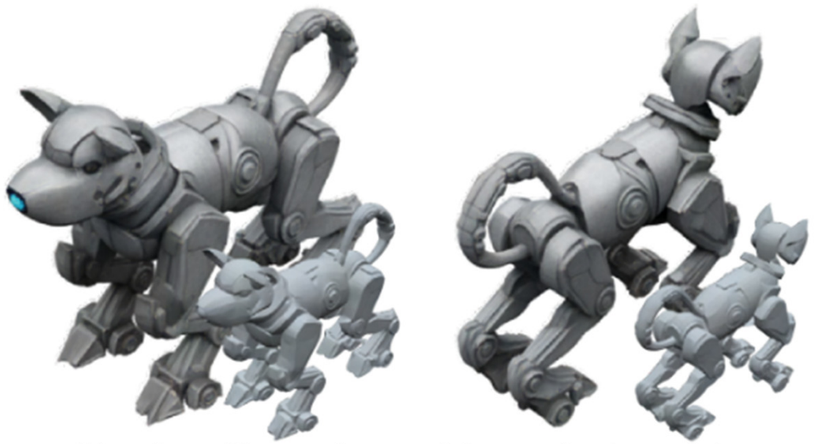


Metallic dog-like robot with articulated legs and futuristic design elements.

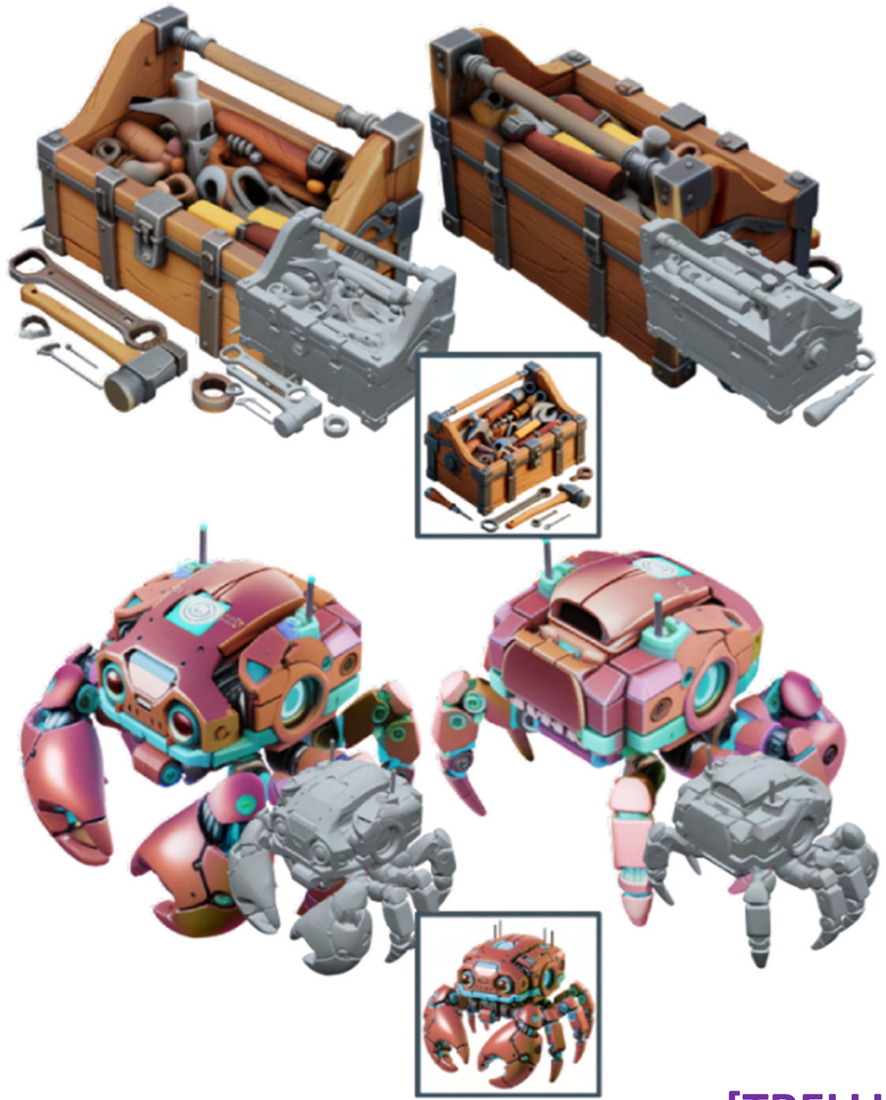
[TRELLIS, CVPR 2025]

Neural 3D Reconstruction & Generation

Portable transistor radio, dark cover, speaker grille, brand logo on front.

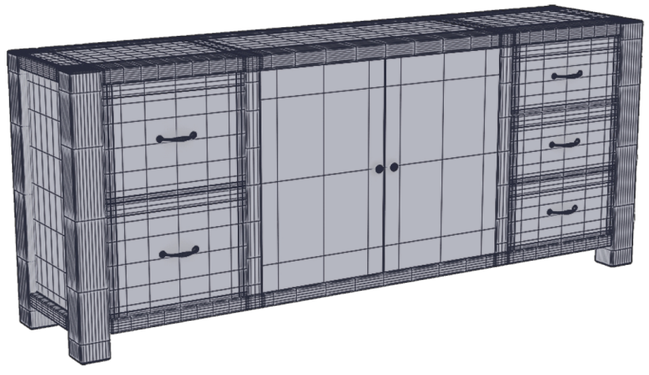


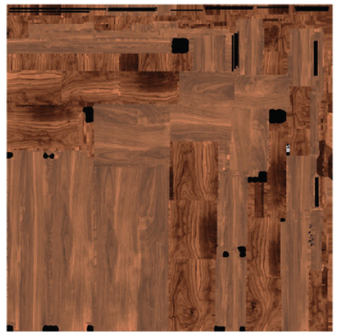
Metallic dog-like robot with articulated legs and futuristic design elements.

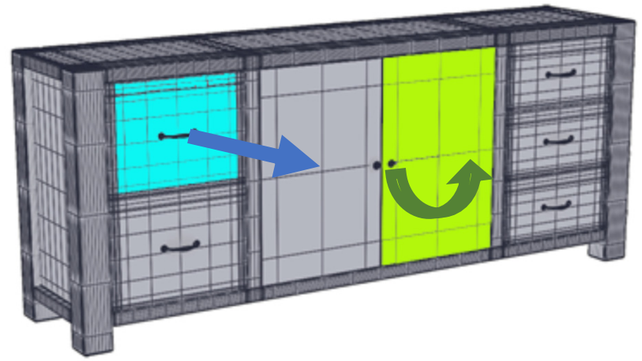


[TRELLIS, CVPR 2025]

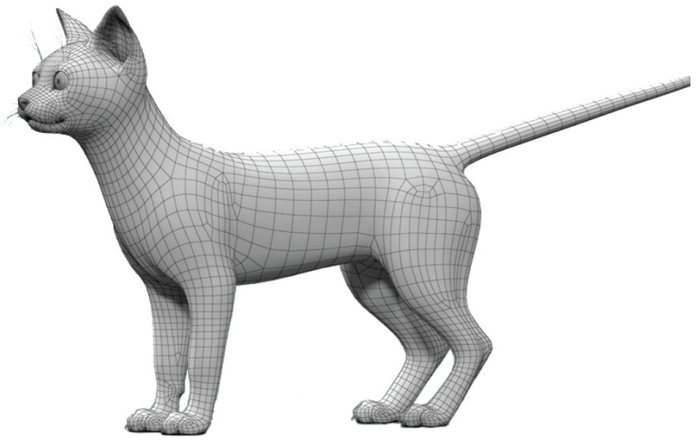
Need more AI progress here!

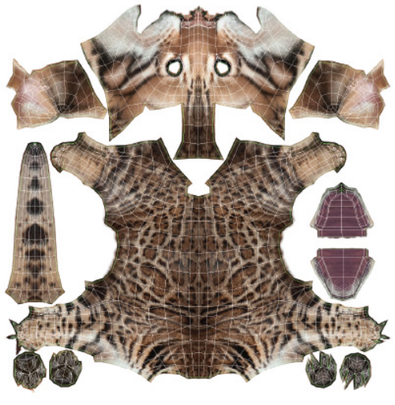


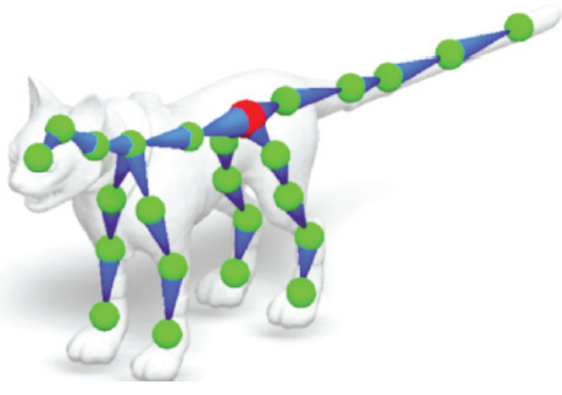




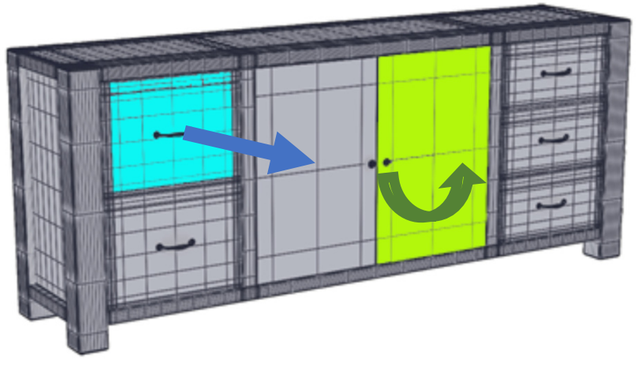
3D Geometry → **Textures** → **Motion controller & attributes** → **Interactive Model**







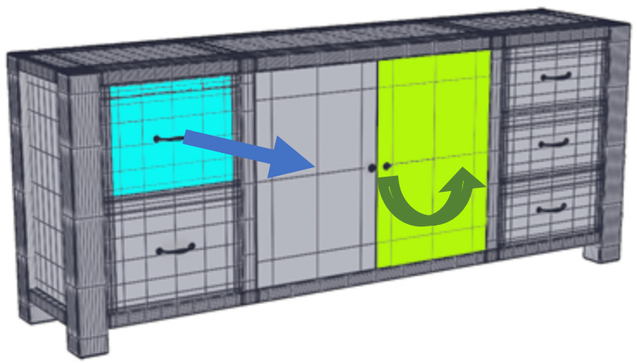
Learn motion controllers & interaction attributes



Given a 3D shape, infer:

- moving parts
- motion types
- motion axes

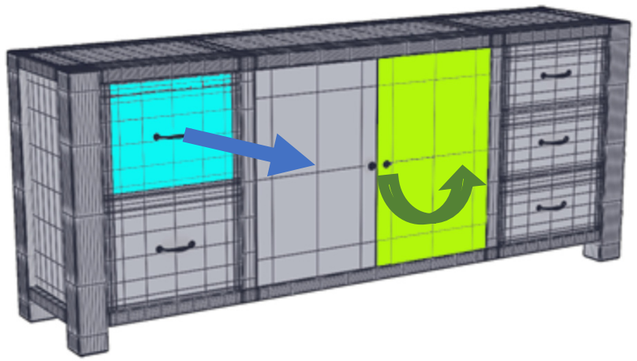
Learn motion controllers & interaction attributes

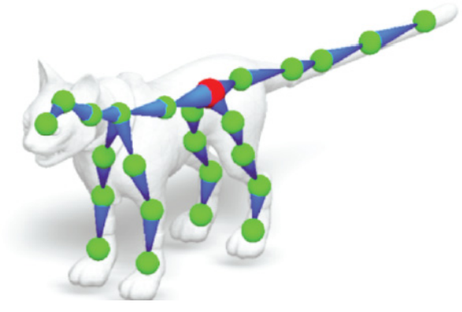


Given a 3D shape, infer:

- moving parts
- motion types
- motion axes
- material types
- material properties

Learn motion controllers & interaction attributes

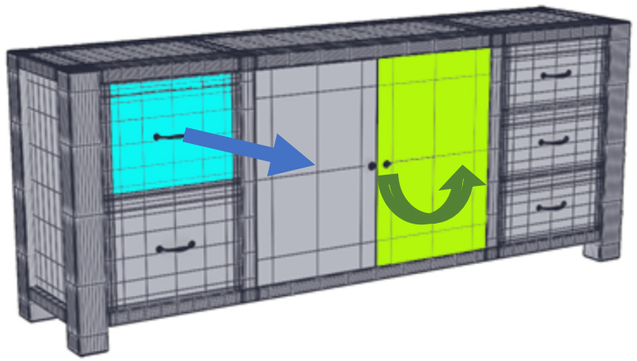


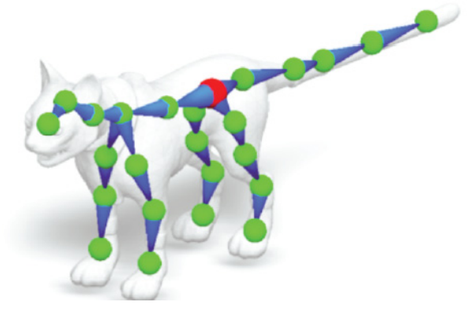


Given a 3D shape, infer:

- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

Learn motion controllers & interaction attributes





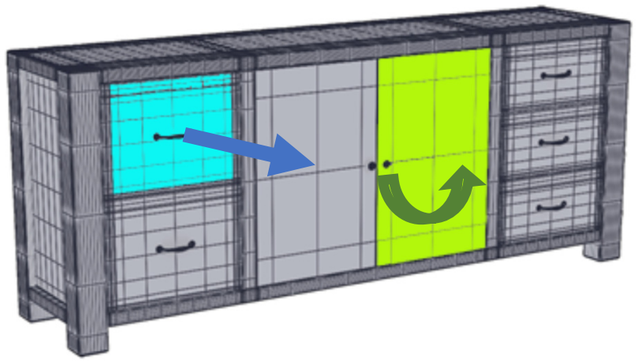
Given a 3D shape, infer:

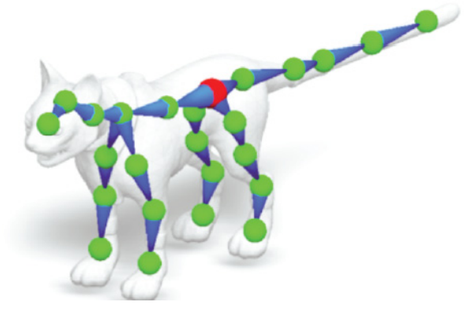
- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

Challenges:

- Limited datasets/annotations
- Generalization

Learn motion controllers & interaction attributes



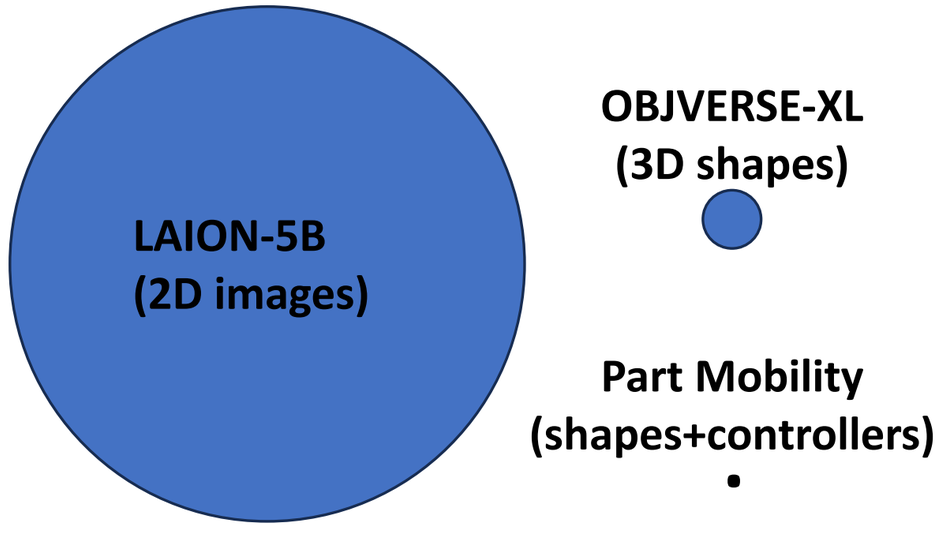


Given a 3D shape, infer:

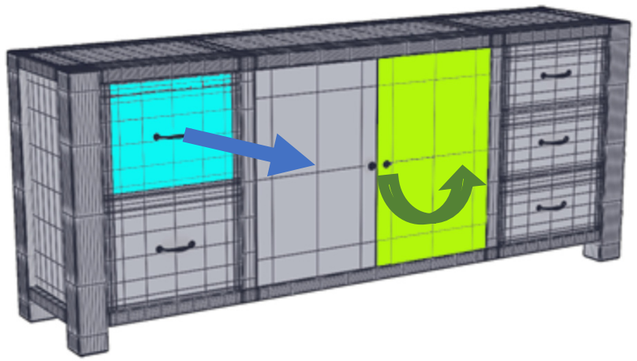
- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

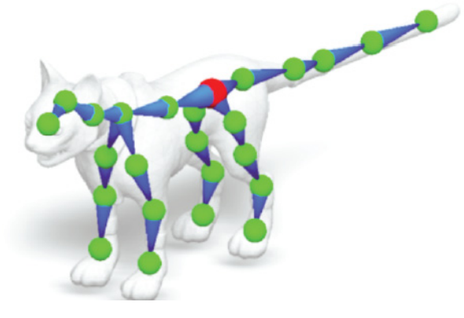
Challenges:

- Limited datasets/annotations
- Generalization



Learn motion controllers & interaction attributes





Given a 3D shape, infer:

- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

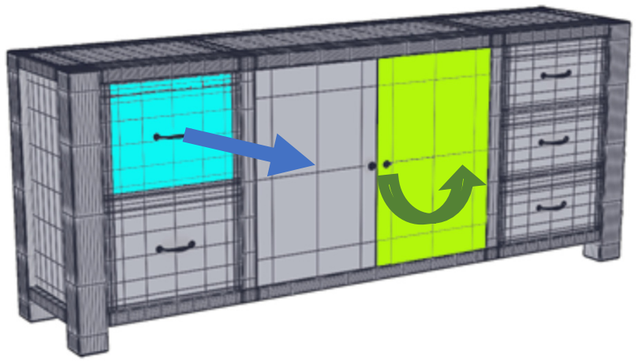
Challenges:

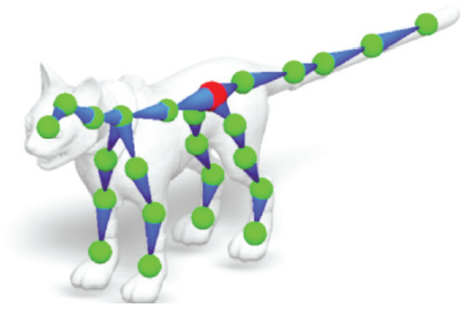
- Limited datasets/annotations
- Generalization

Methodology:

- Self-supervision

Learn motion controllers & interaction attributes





Given a 3D shape, infer:

- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

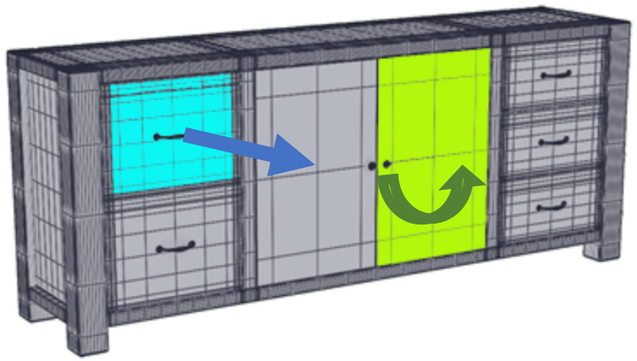
Challenges:

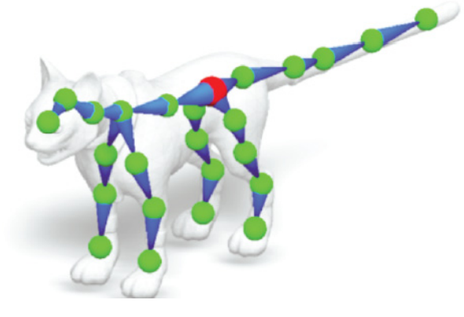
- Limited datasets/annotations
- Generalization

Methodology:

- Self-supervision
- Distillation from 2D models

Learn motion controllers & interaction attributes





Generate objects with:

- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

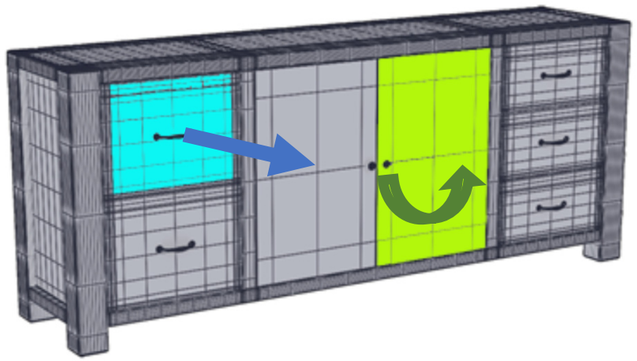
Challenges:

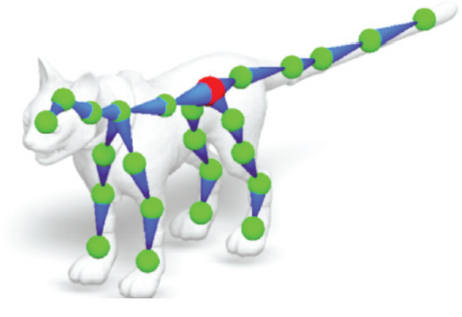
- Limited datasets/annotations
- Generalization
- Generate motion controllers compatible with geometry

Methodology:

- Self-supervision
- Distillation from 2D models

Learn motion controllers & interaction attributes





Generate objects with:

- moving parts
- motion types
- motion axes
- material types
- material properties
- articulation hierarchy
- degrees of freedom
- deformations

Challenges:

- Limited datasets/annotations
- Generalization
- Generate motion controllers compatible with geometry

Methodology:

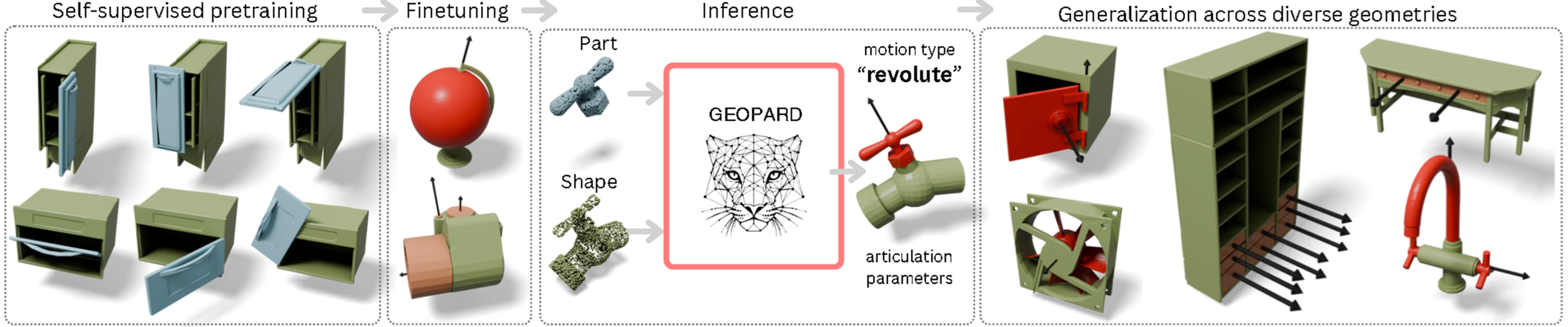
- Self-supervision
- Distillation from 2D models
- Latent space encoding motion controllers & shape

Part I:

GEOPARD: Geometric Pretraining for Articulation Prediction in 3D Shapes

Pradyumn Goyal[1] Dmitry Petrov[1] Sheldon Andrews[2] Yizhak Ben-Shabat[3]
Hsueh-Ti Derek Liu[3] Evangelos Kalogerakis[4]

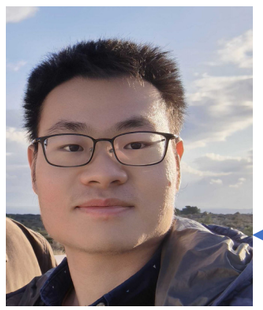
[1]UMass Amherst [2]École de technologie supérieure [3]Roblox [4]TU Crete



https://arxiv.org/pdf/2504.02747

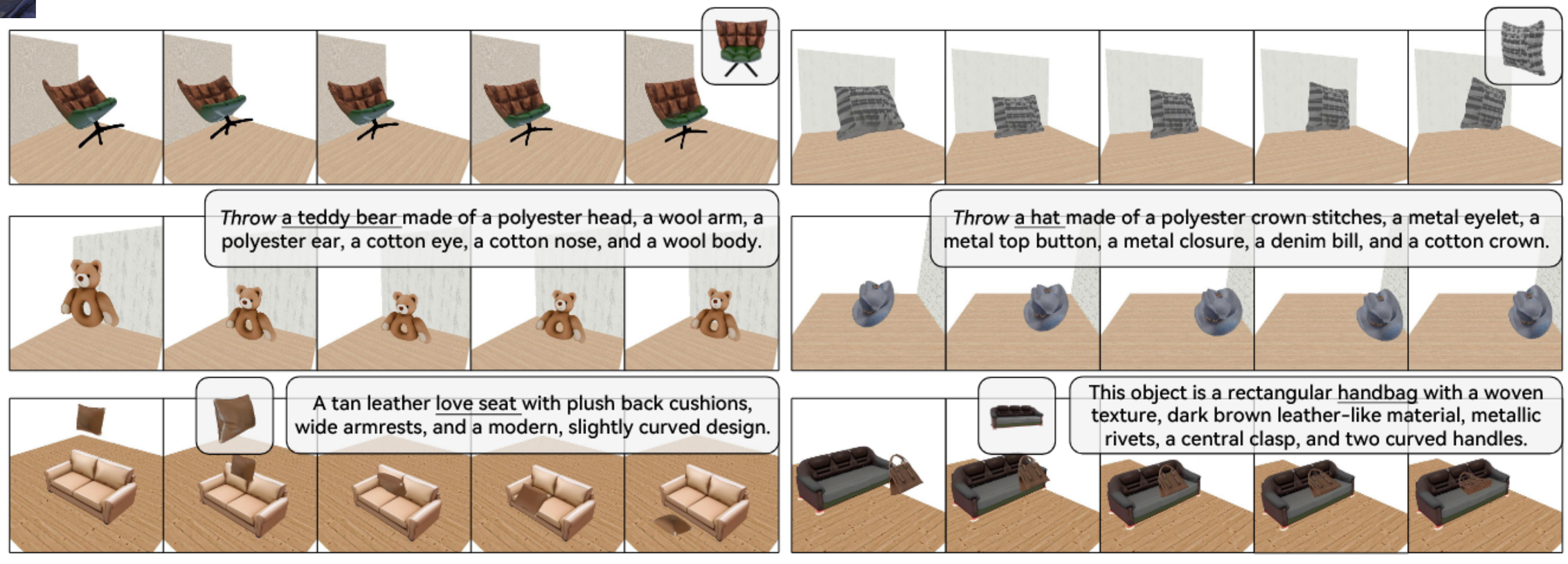
(accepted at ICCV 25)

Part II:



SOPHY: Learning to Generate Simulation-Ready Objects with PHYsical Materials

Junyi Cao Evangelos Kalogerakis



https://arxiv.org/pdf/2504.12684

(accepted at WACV 26)

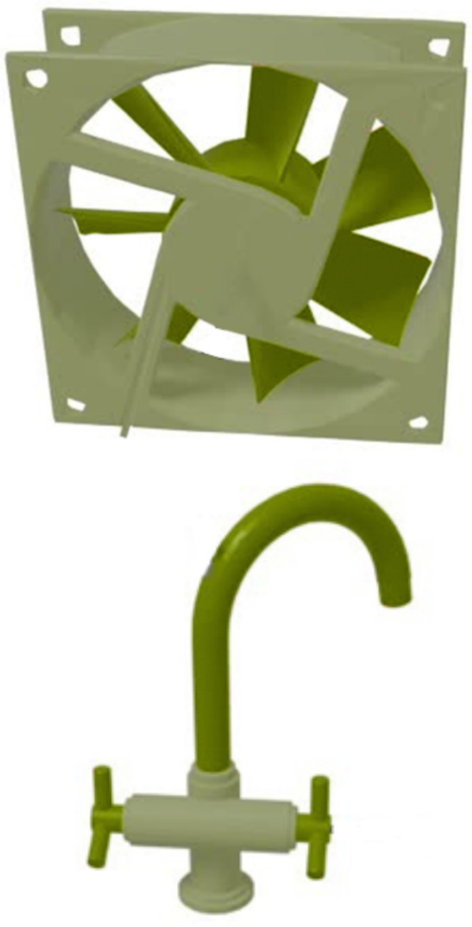
Outline

1. GEOPARD: Geo**metric** P**retraining for** Ar**ticulation** Pre**diction in 3D Shapes**

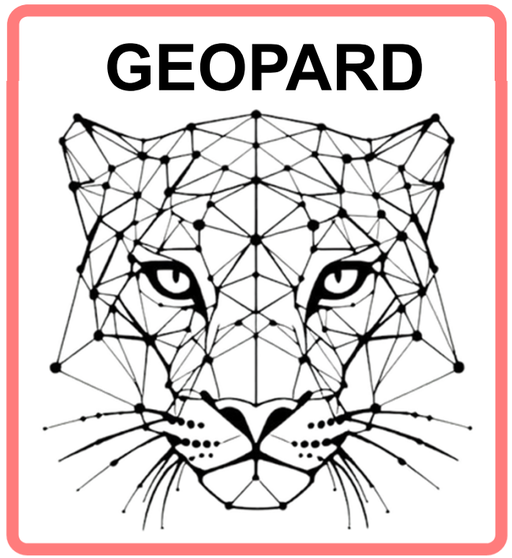
2. SOPHY: **Generating** S**imulation-Ready** O**bjects with** Phy**sical Materials**

3. Future directions

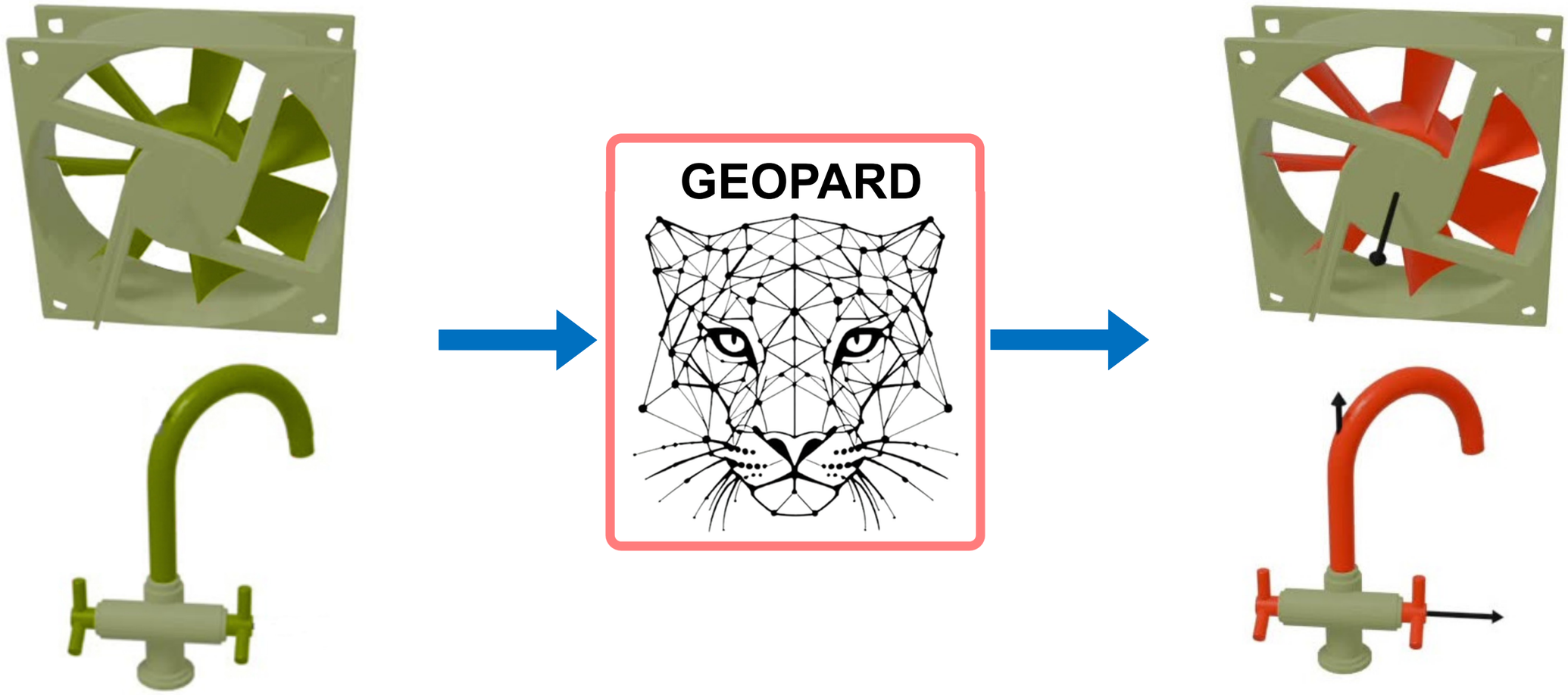
Goal: estimate articulation from a static shape



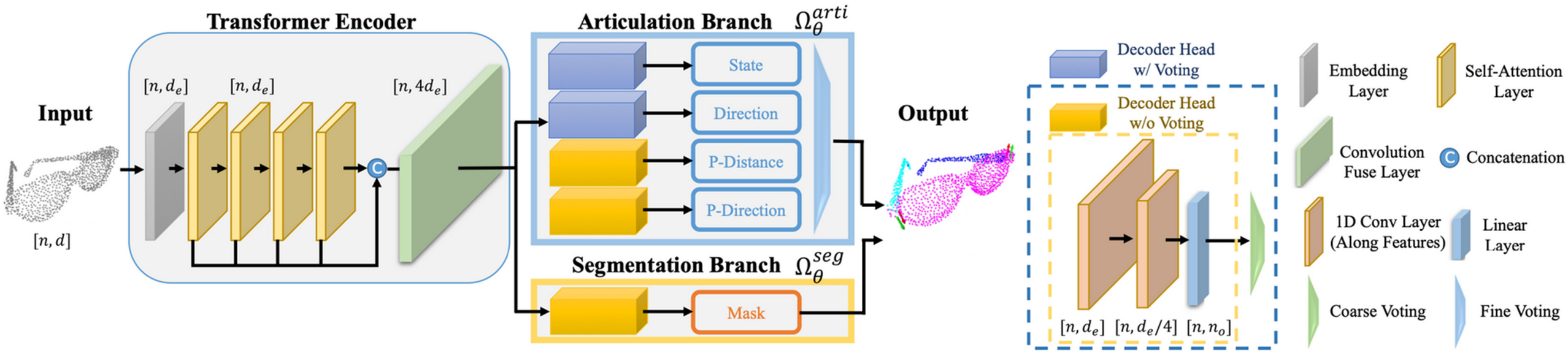
Goal: estimate articulation from a static shape



Goal: estimate articulation from a static shape

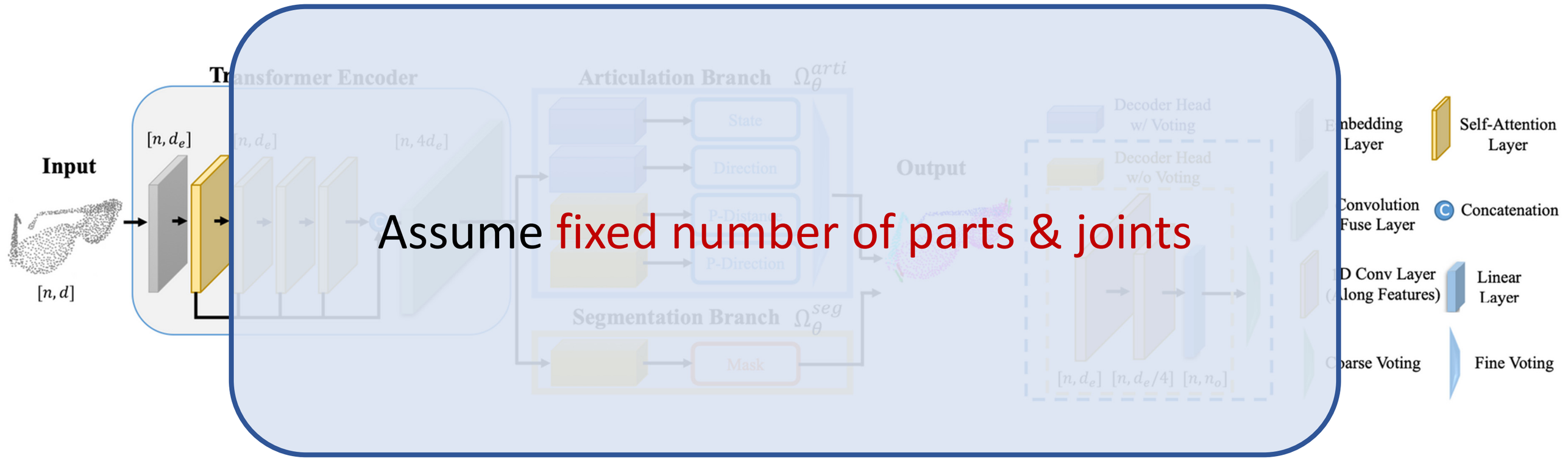


Prior Work



CAPT, ICRA 2024

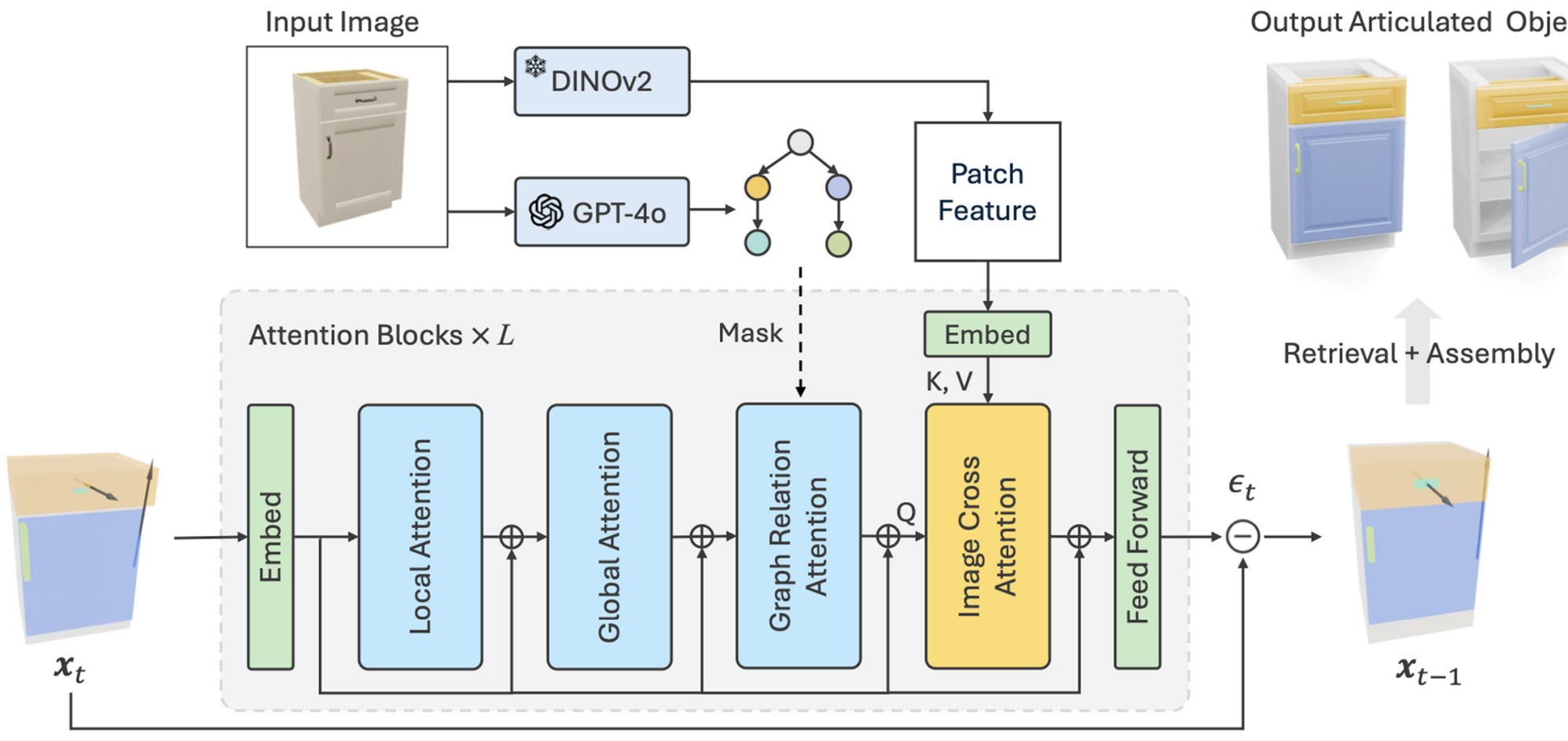
Prior Work



Assume fixed number of parts & joints

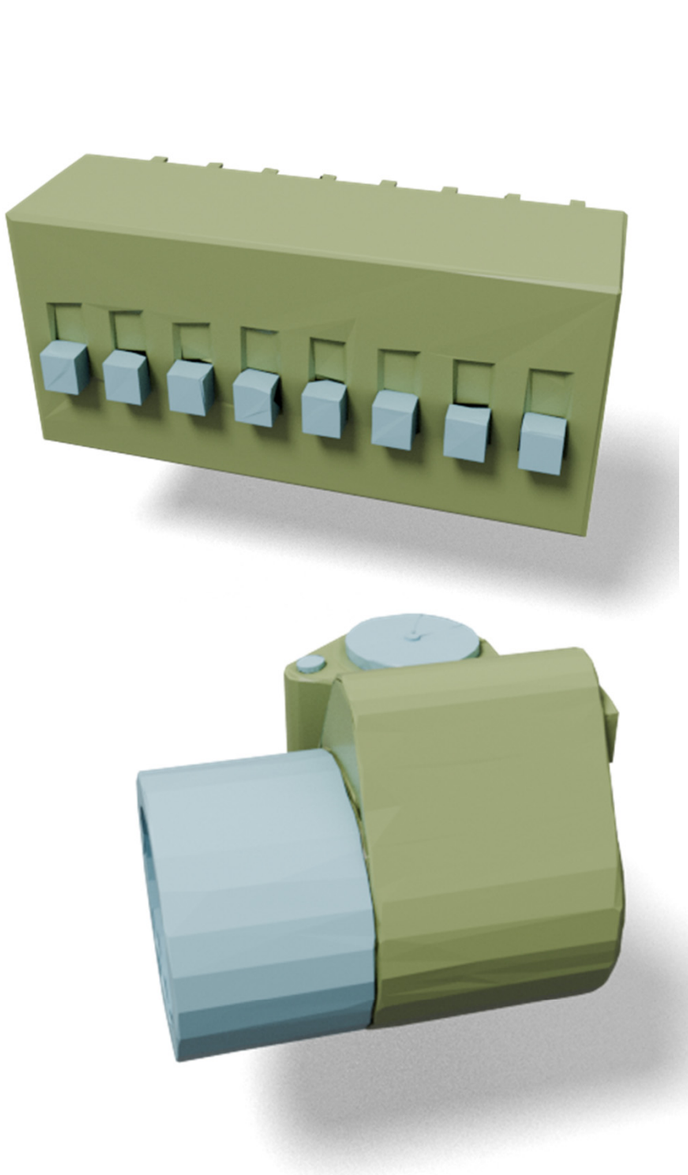
CAPT, ICRA 2024

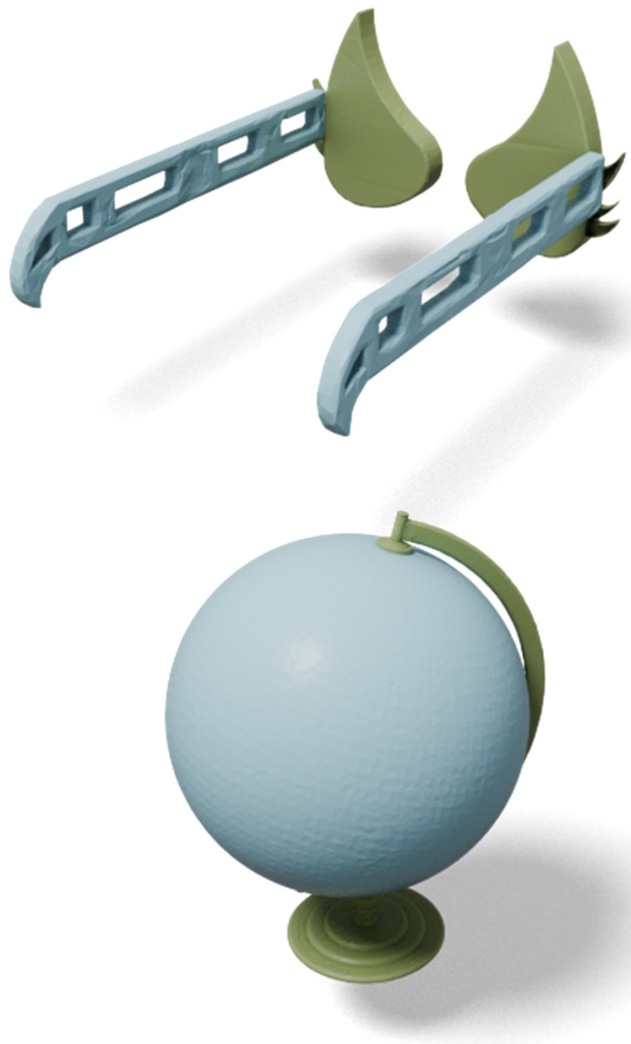
Prior Work



SINGAPO, ICLR 2025

Input assumption: 3D segmented models

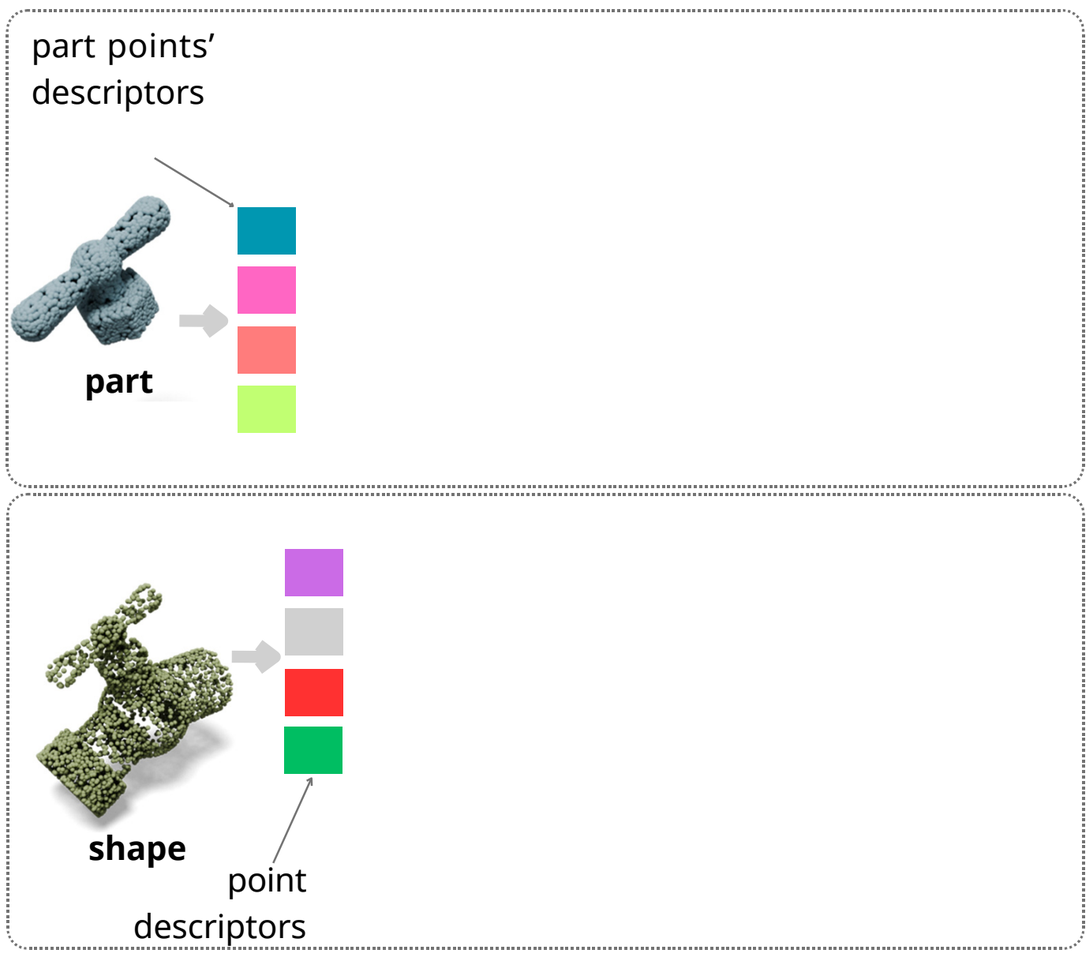




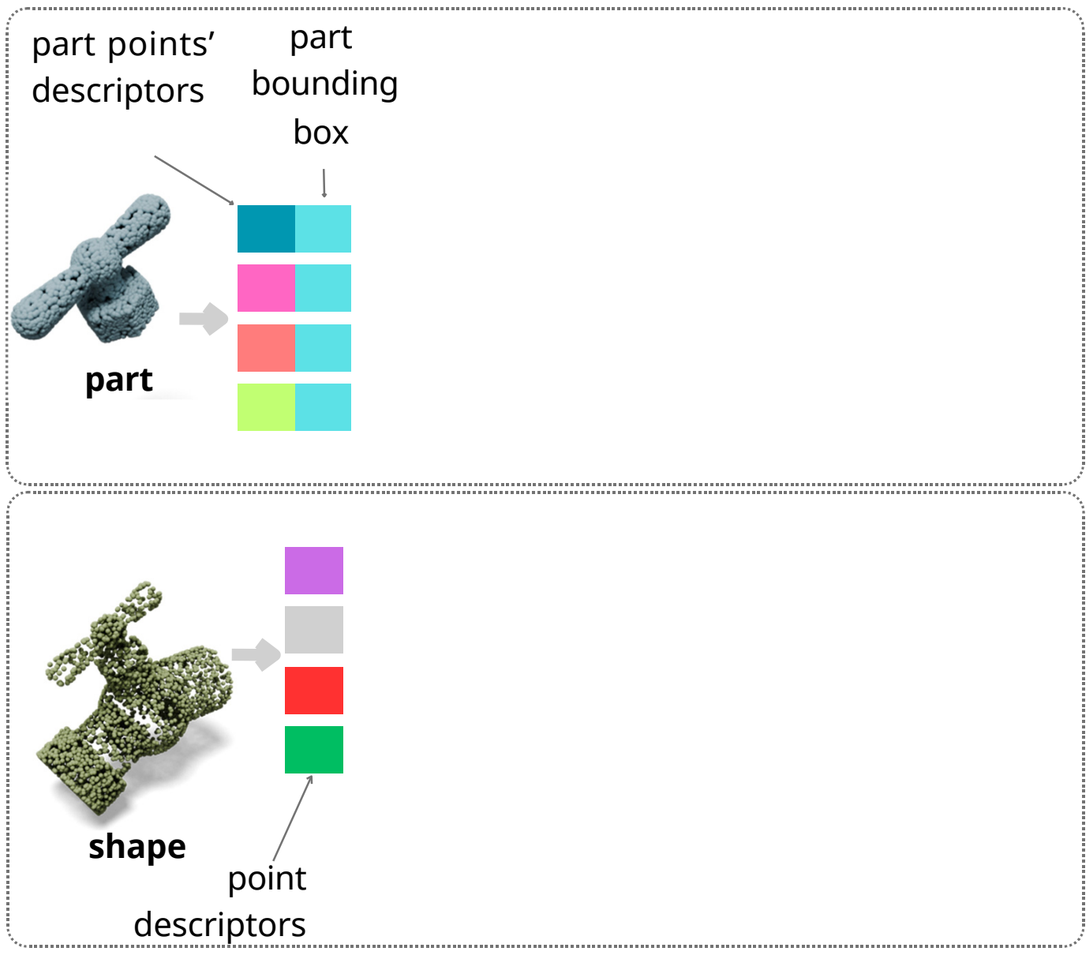
3D parts from 2D foundation model adaptation

PartSTAD, ECCV 2024

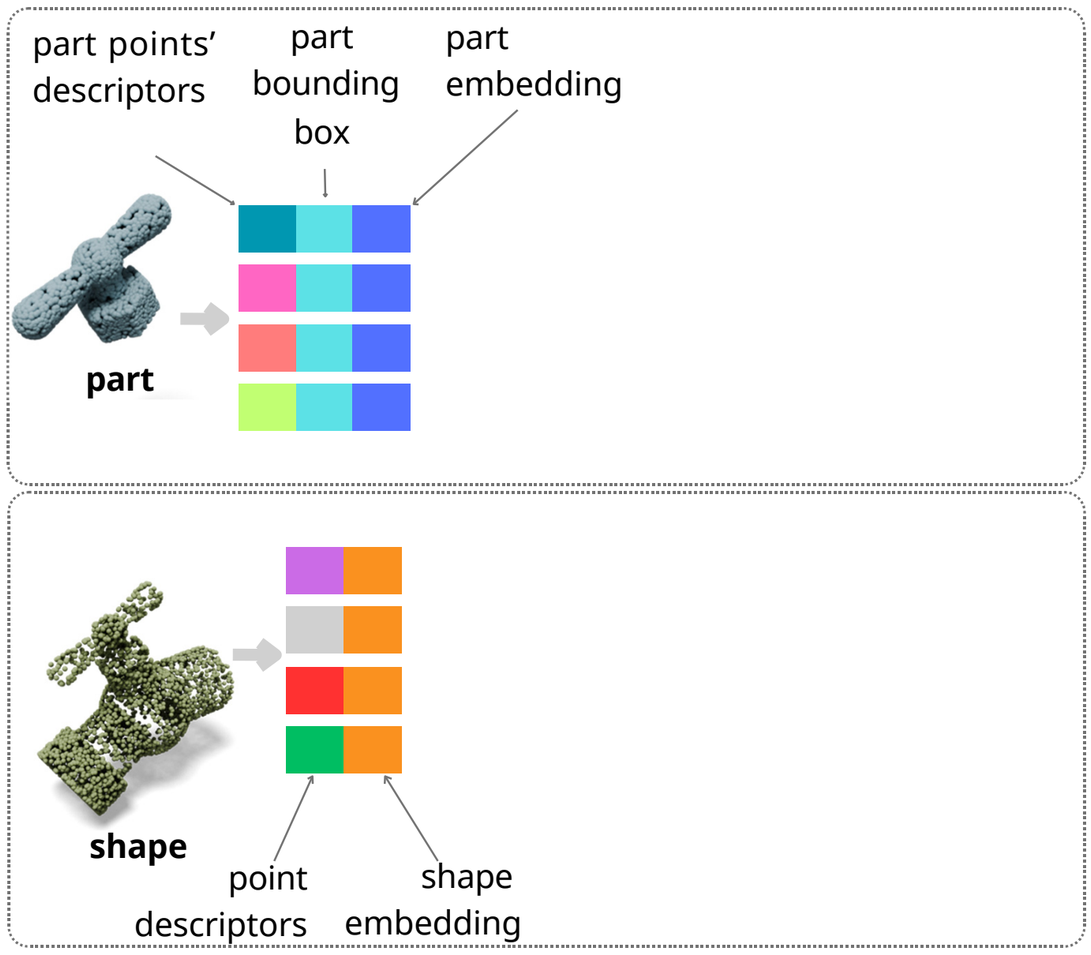
Network Architecture



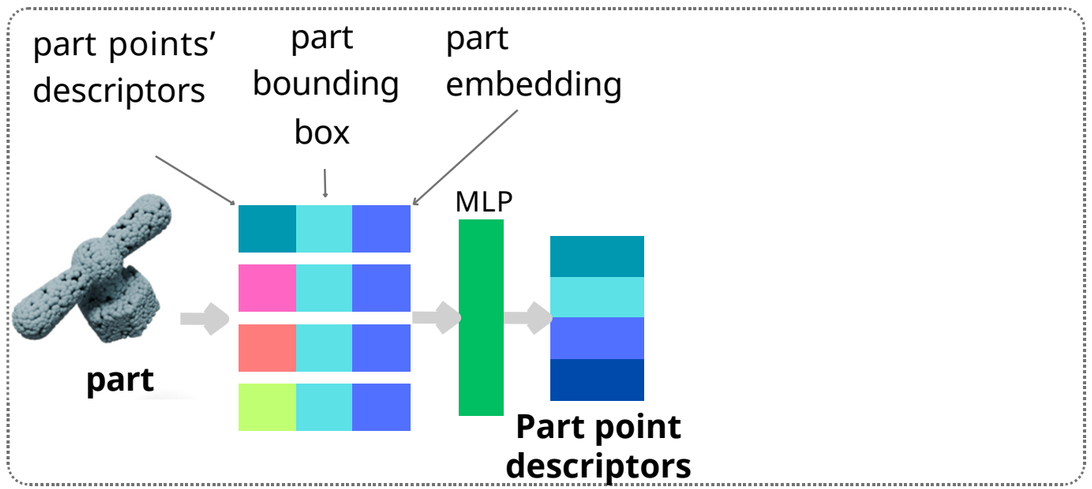
Network Architecture

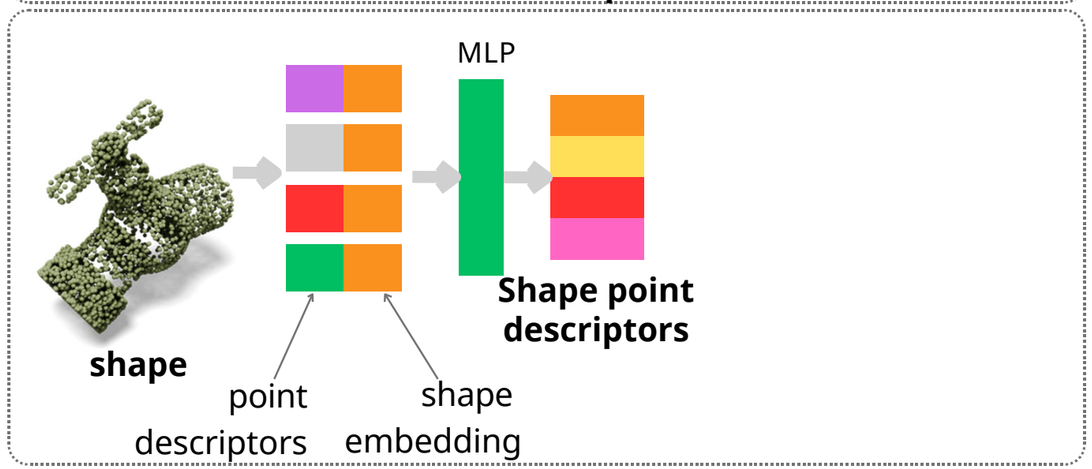


Network Architecture

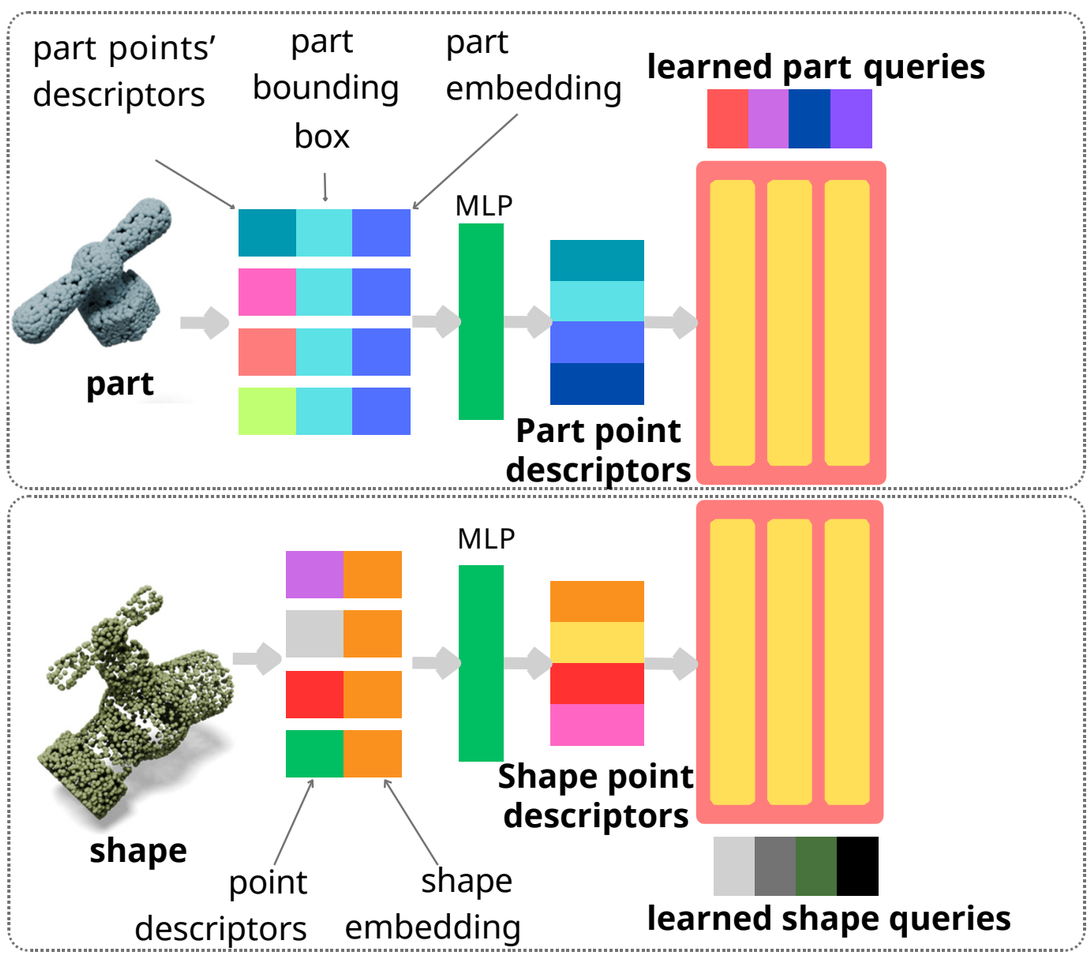


Network Architecture

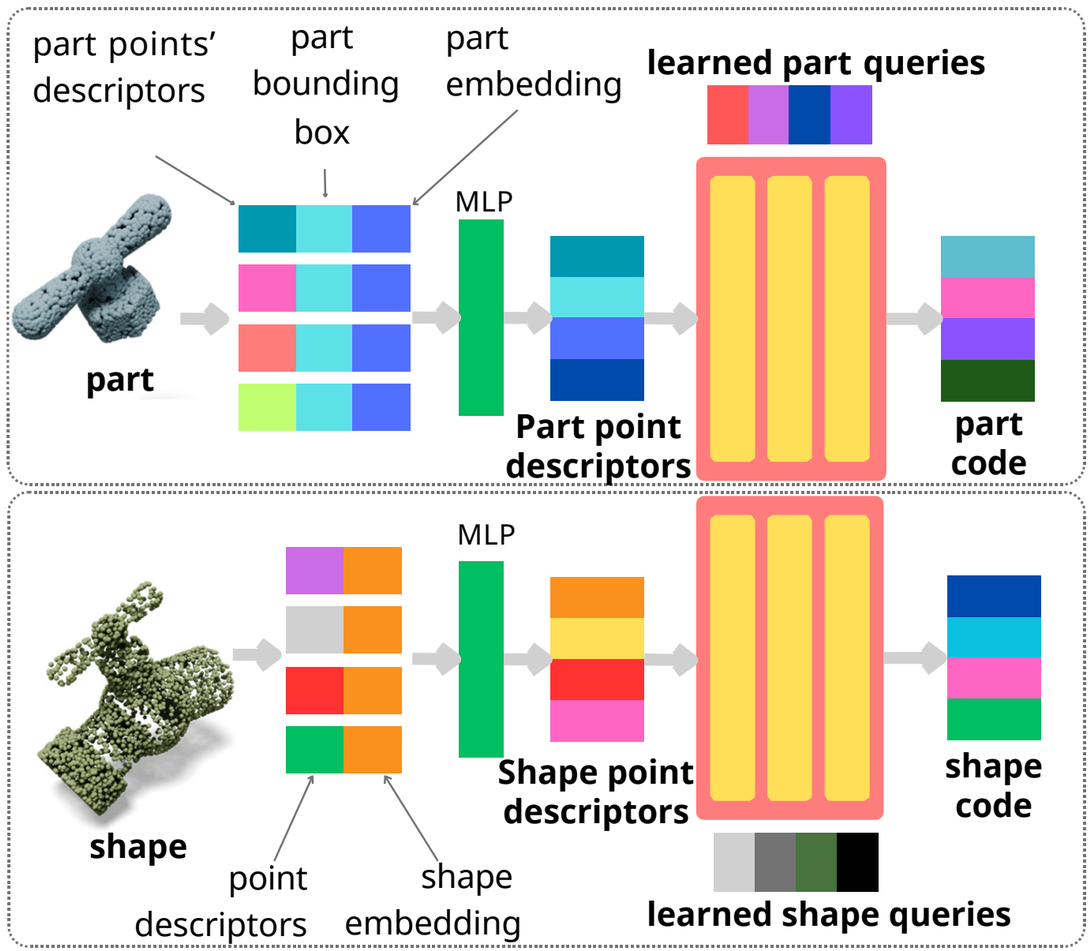




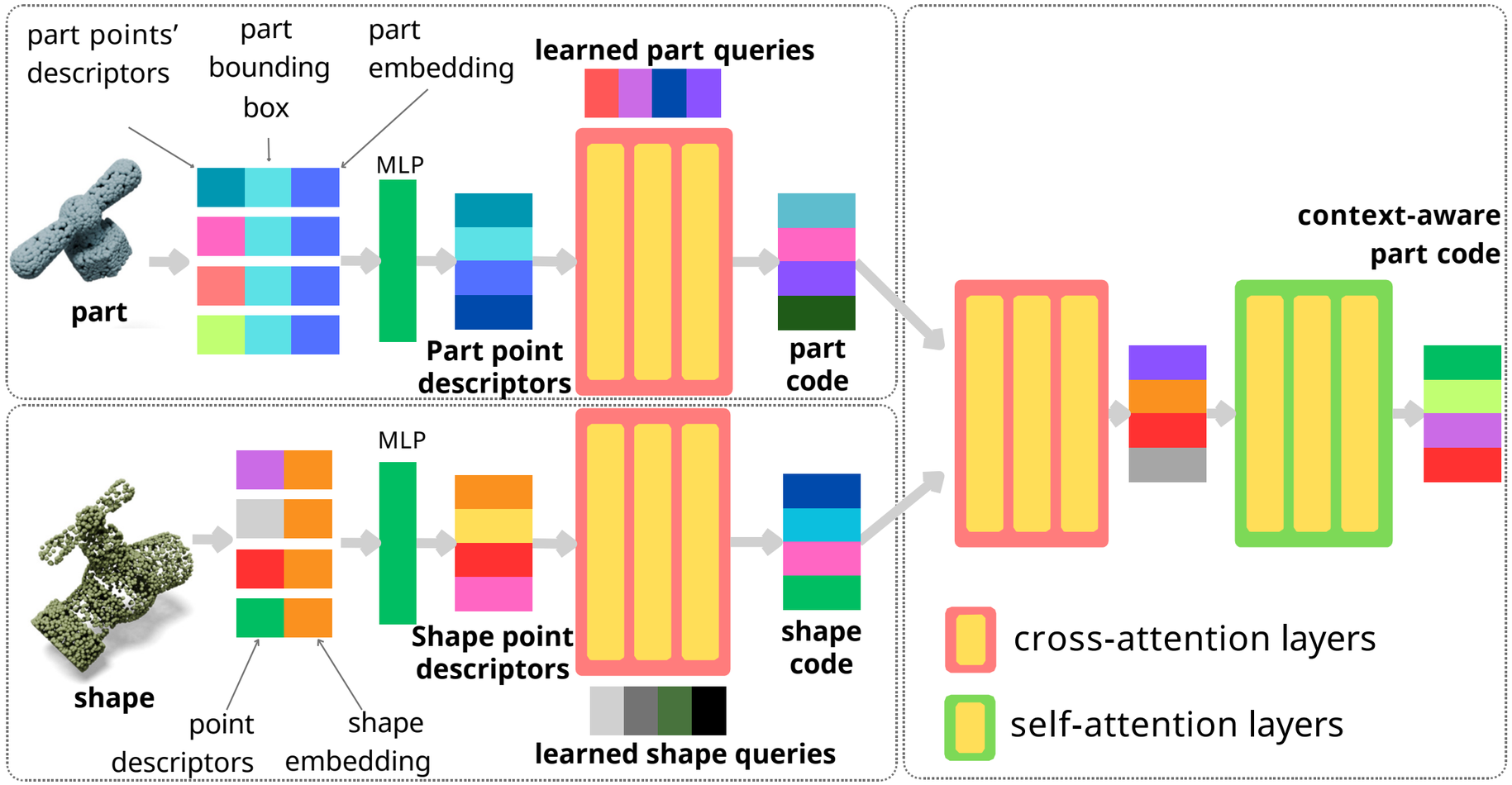
Network Architecture



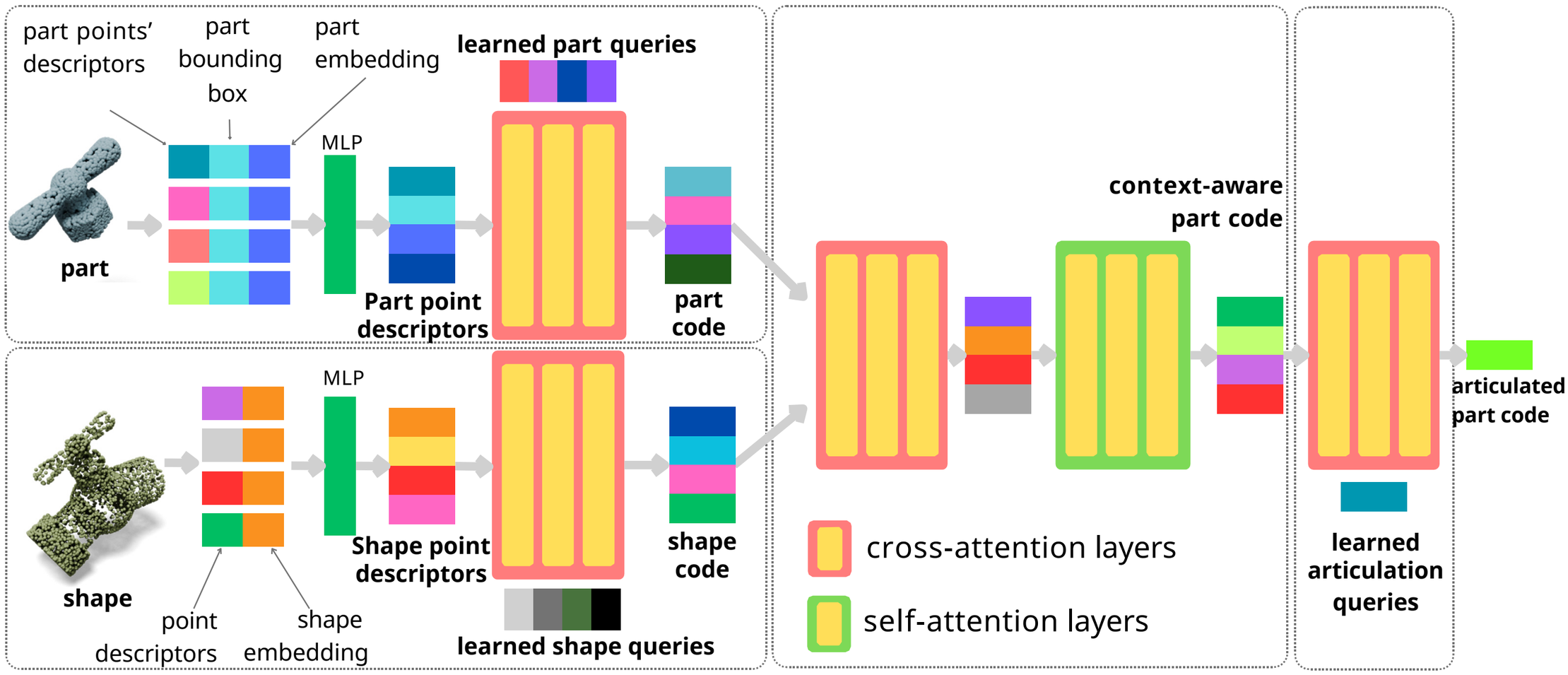
Network Architecture



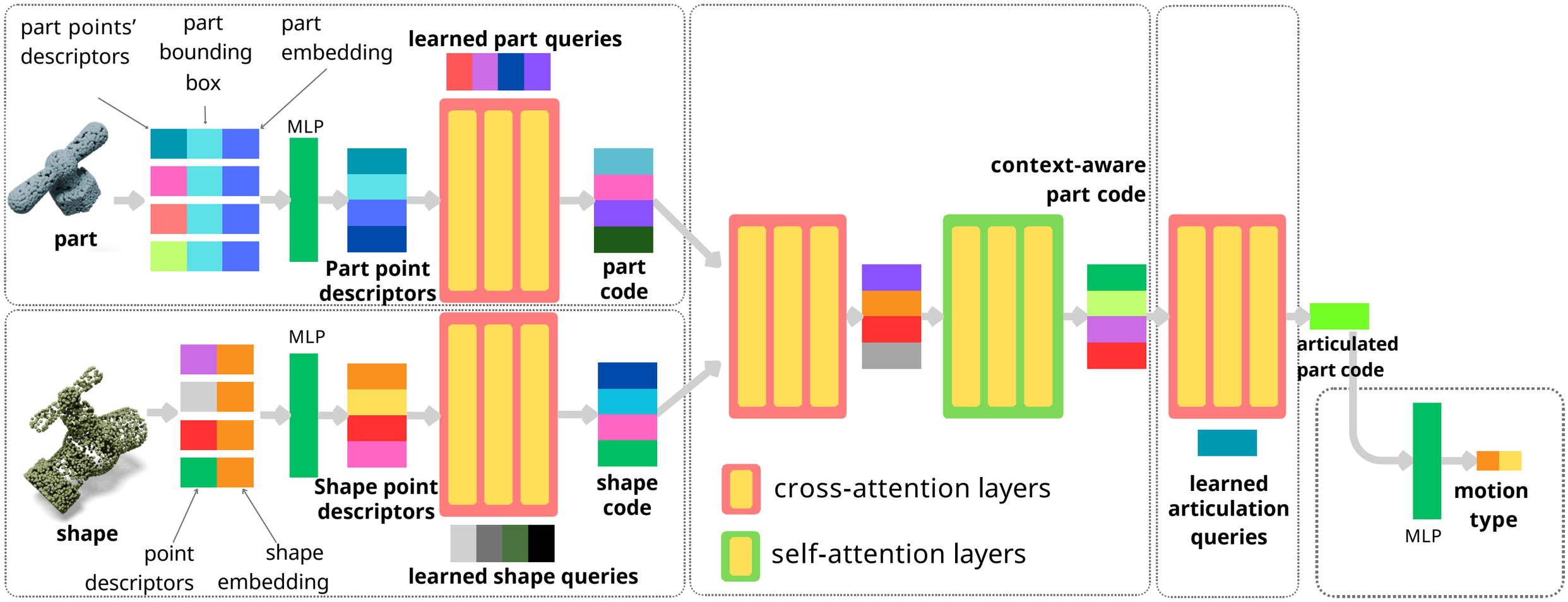
Network Architecture



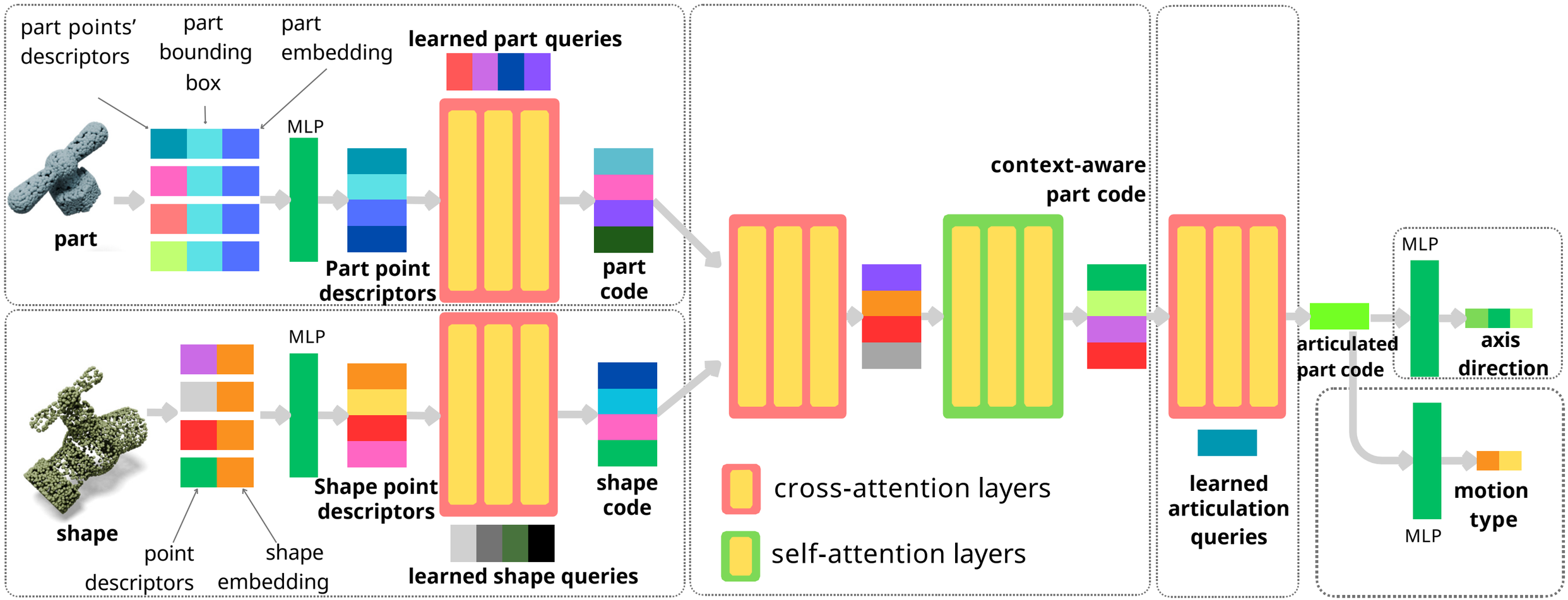
Network Architecture



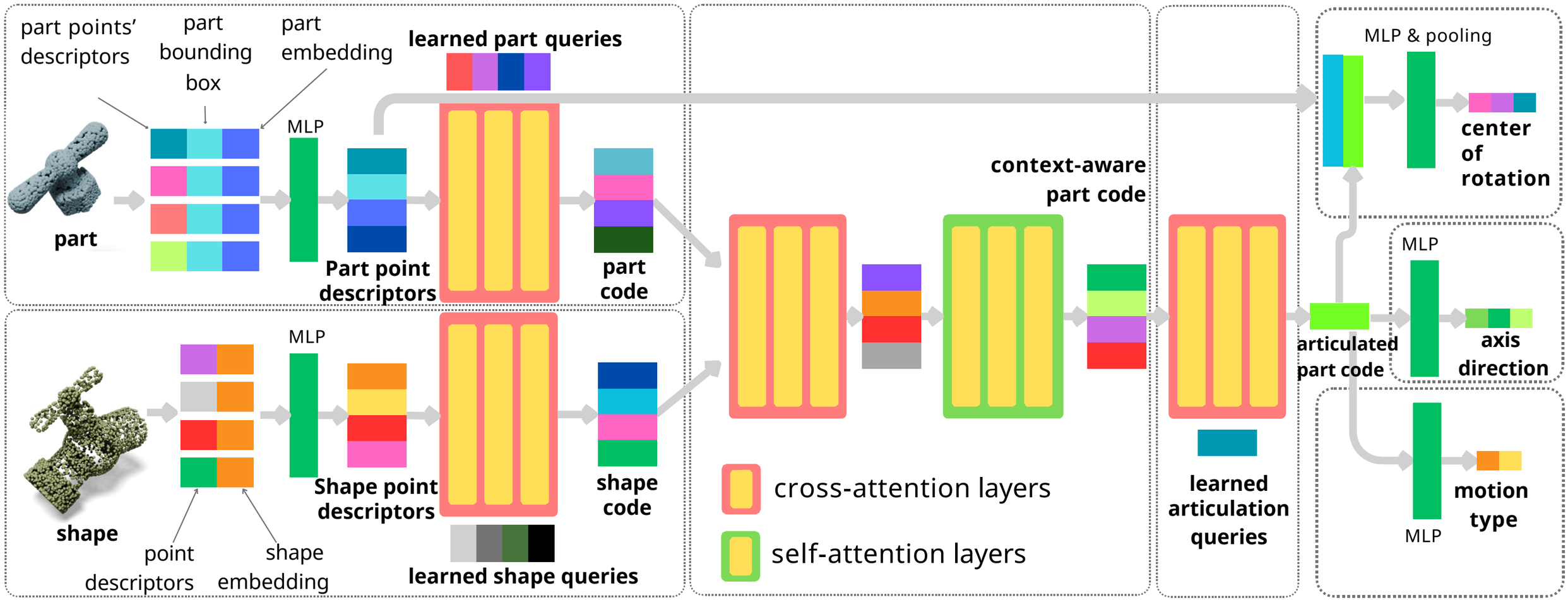
Network Architecture



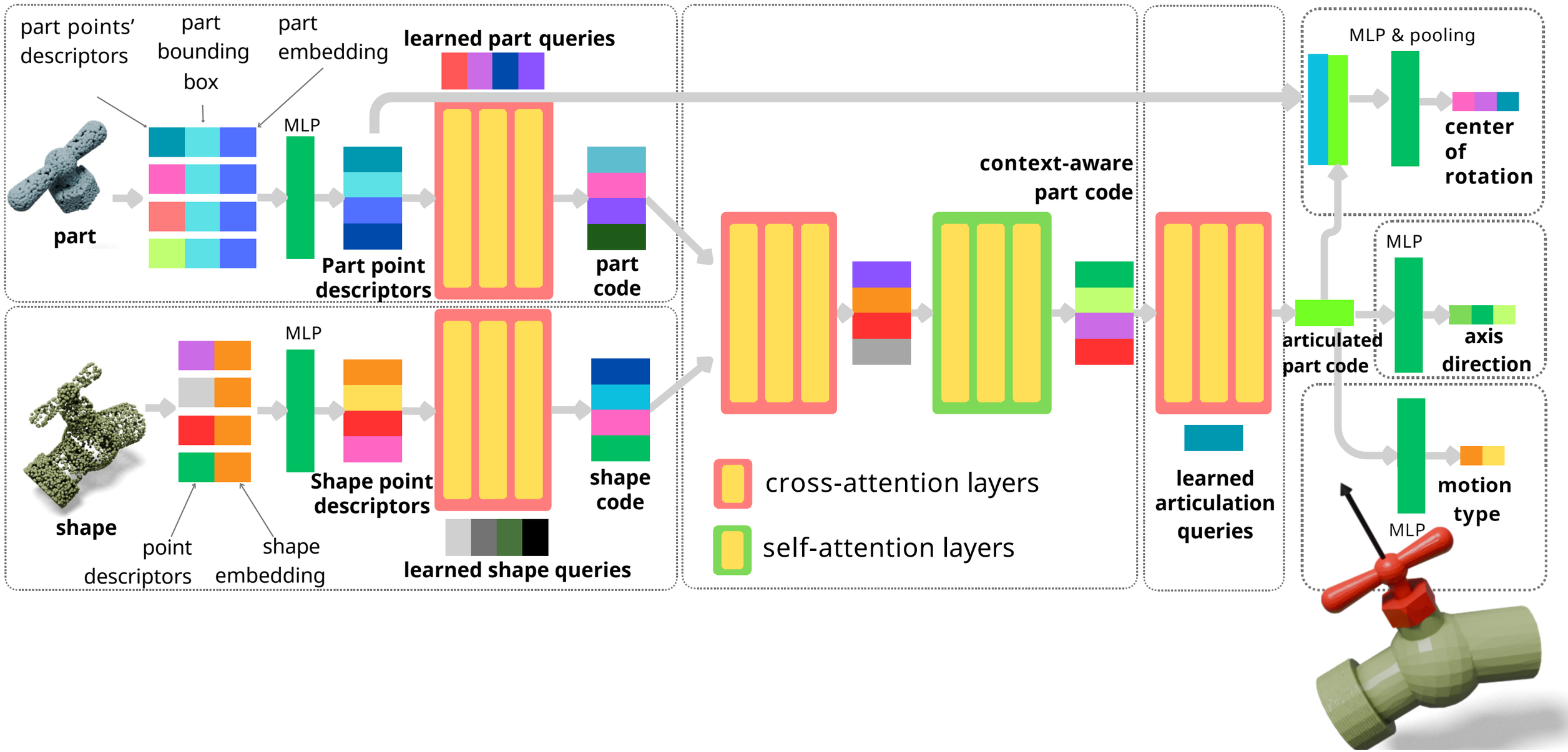
Network Architecture



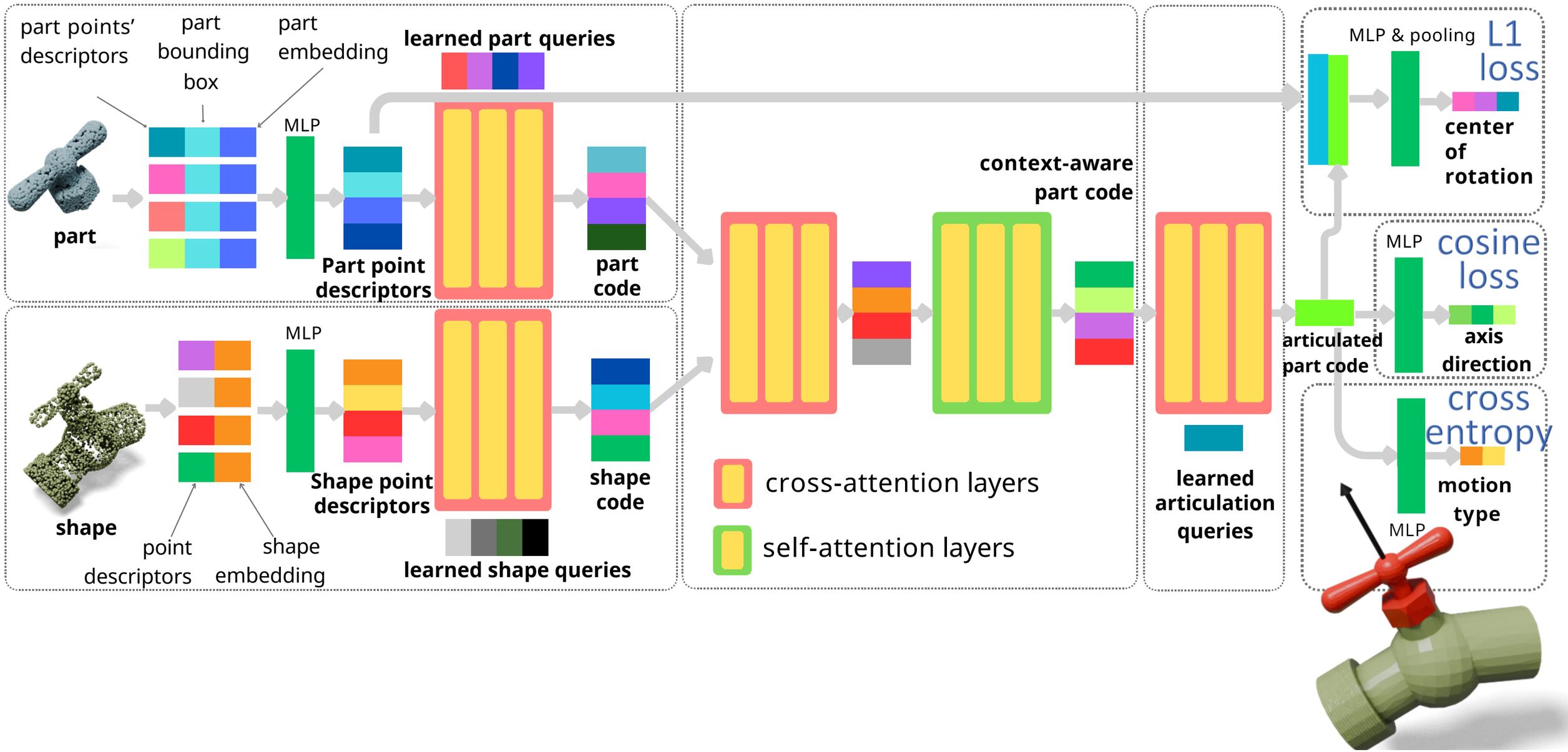
Network Architecture



Network Architecture



Training


part points' descriptors
part bounding box
part embedding
learned part queries
MLP
part
Part point descriptors
part code
context-aware part code
MLP & pooling
L1 loss
center of rotation
MLP
cosine loss
axis direction
articulated part code
cross entropy
motion type
MLP
shape
point descriptors
shape embedding
MLP
Shape point descriptors
shape code
learned shape queries
cross-attention layers
self-attention layers
learned articulation queries

Datasets with annotated articulation are small!

2.3K shapes, 46 object categories

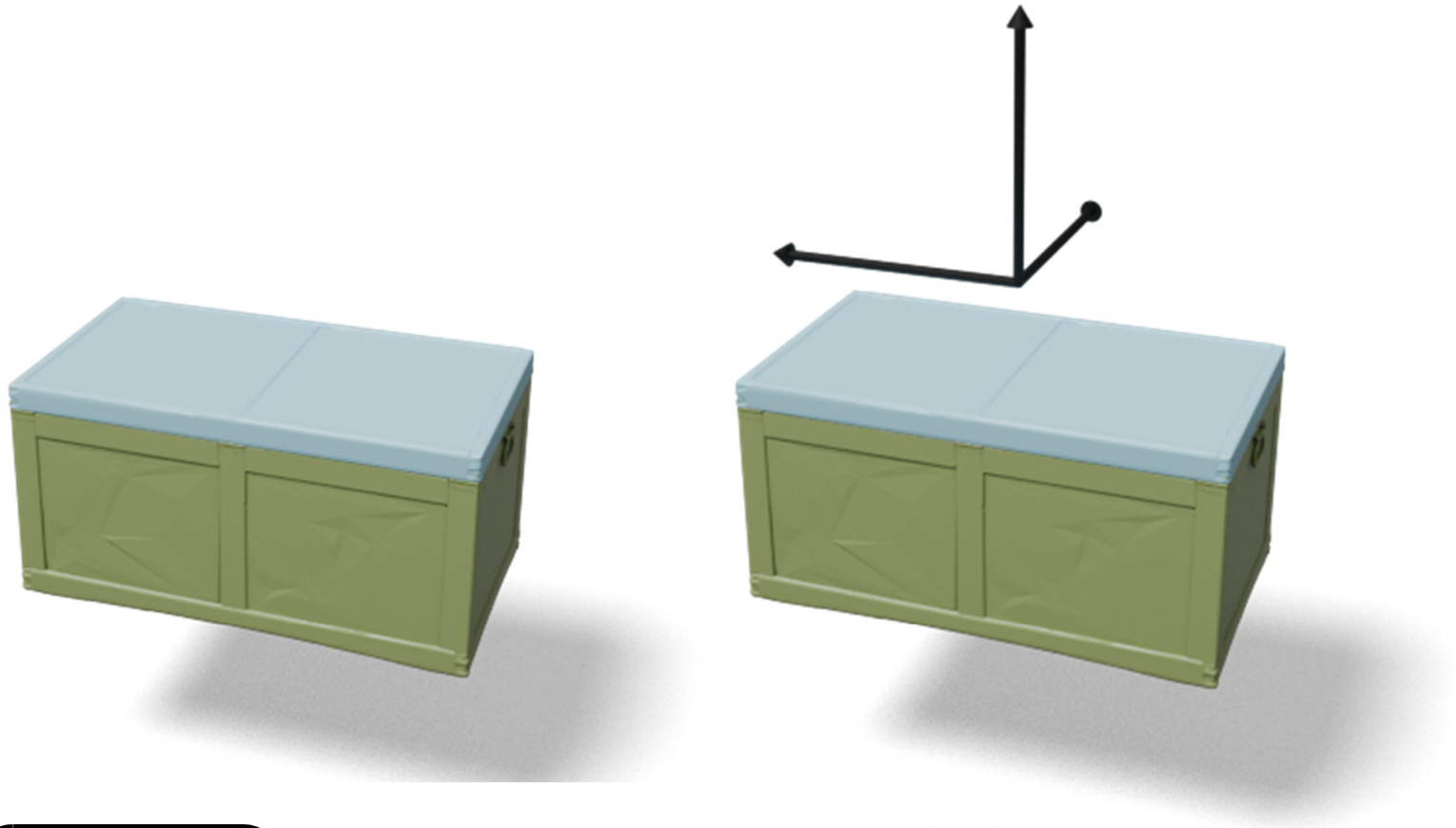
PartNet-Mobility, CVPR 2020

Self-supervision: find **possible** motions

Part

Shape

Self-supervision: find **possible** motions

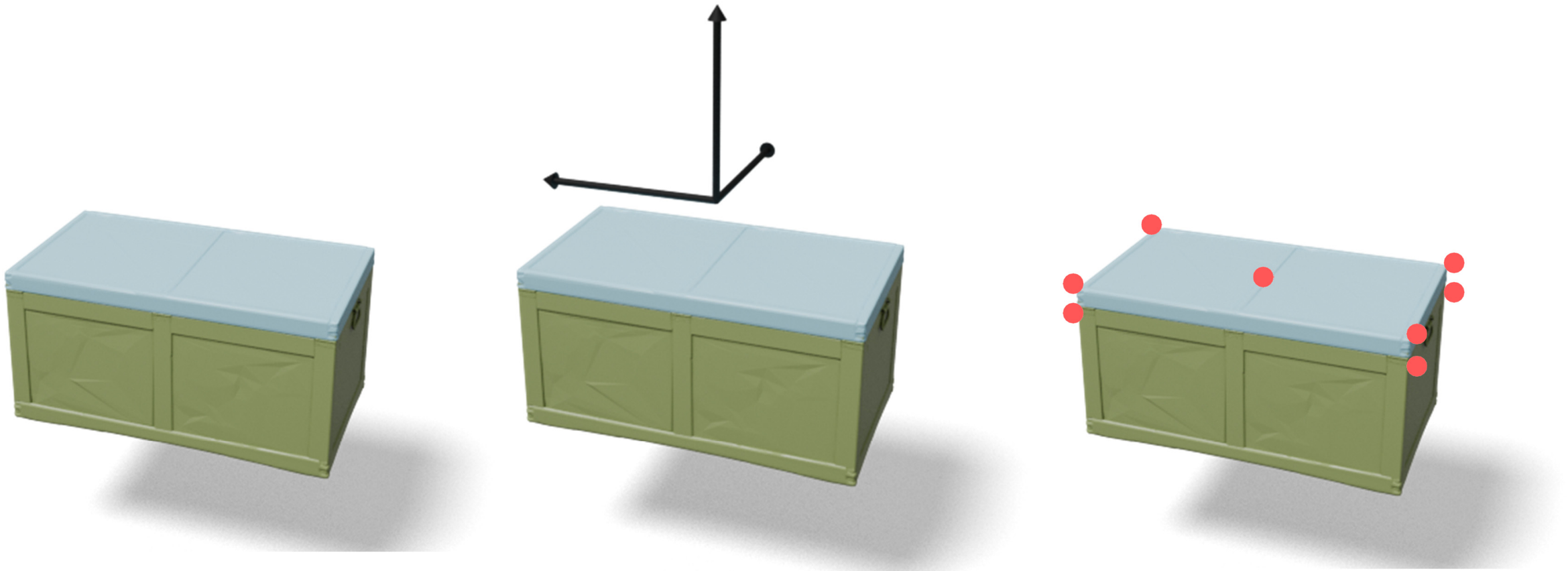


Part

Shape

Candidate motion axes

Self-supervision: find **possible** motions



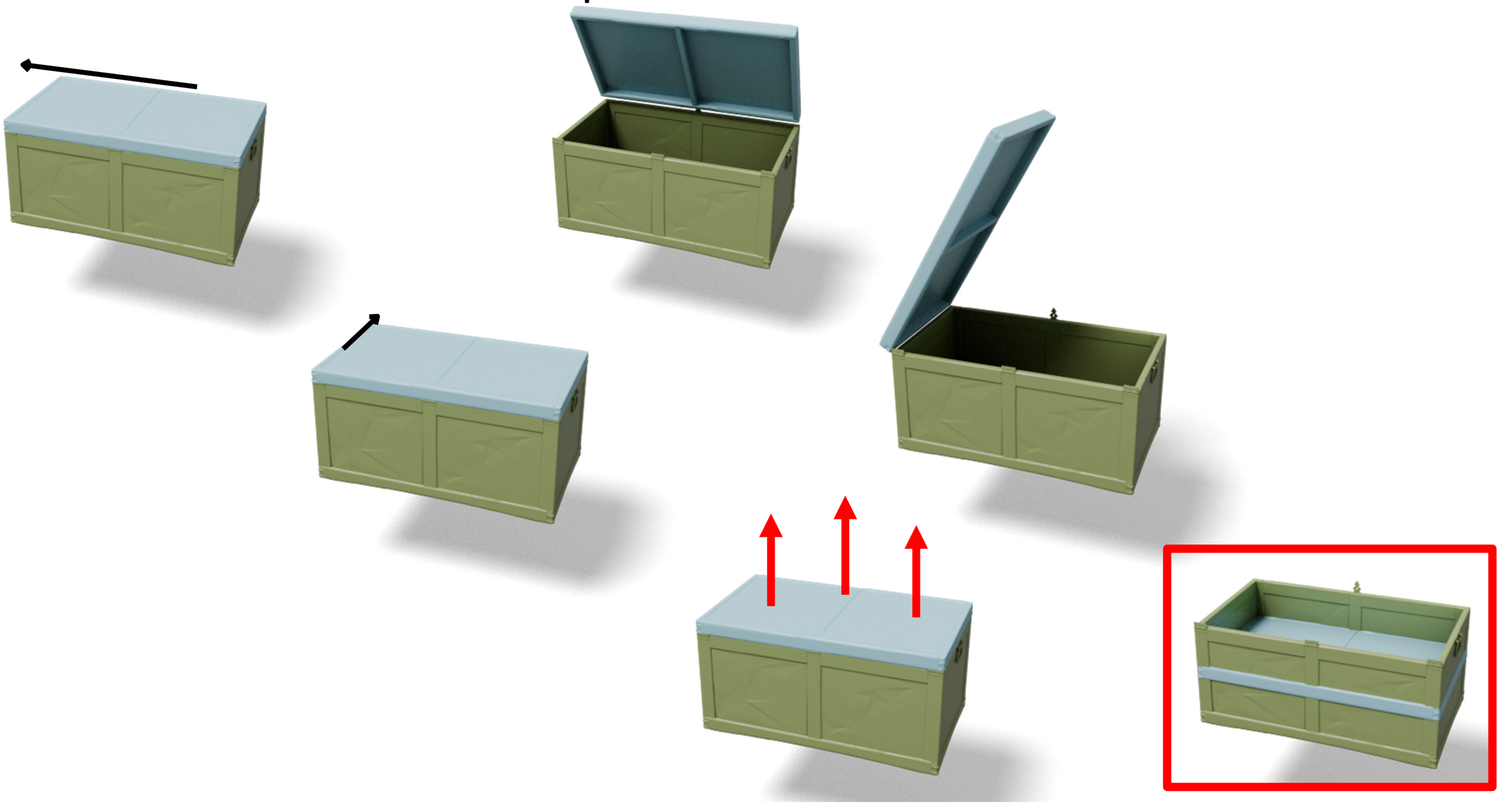
Part

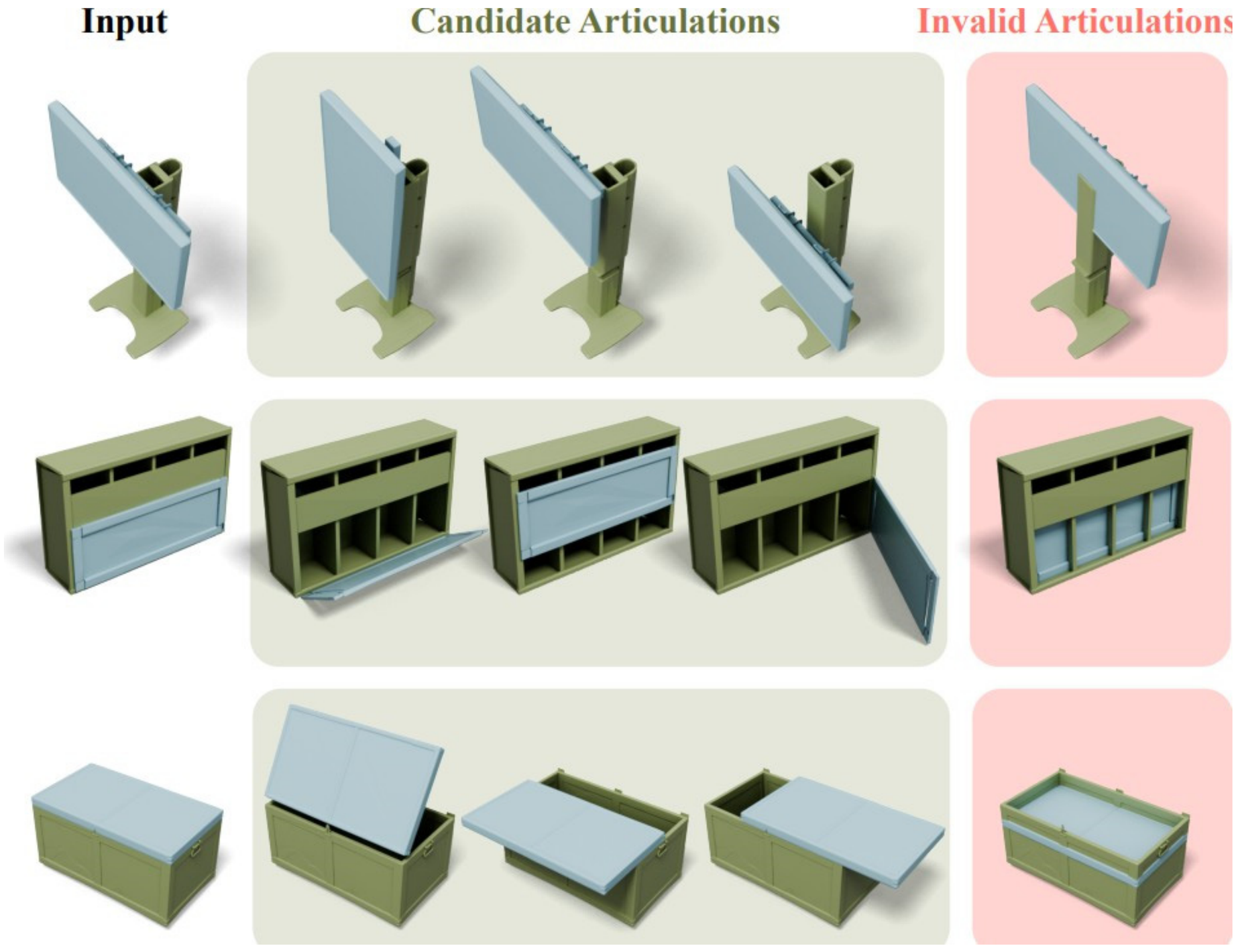
Shape

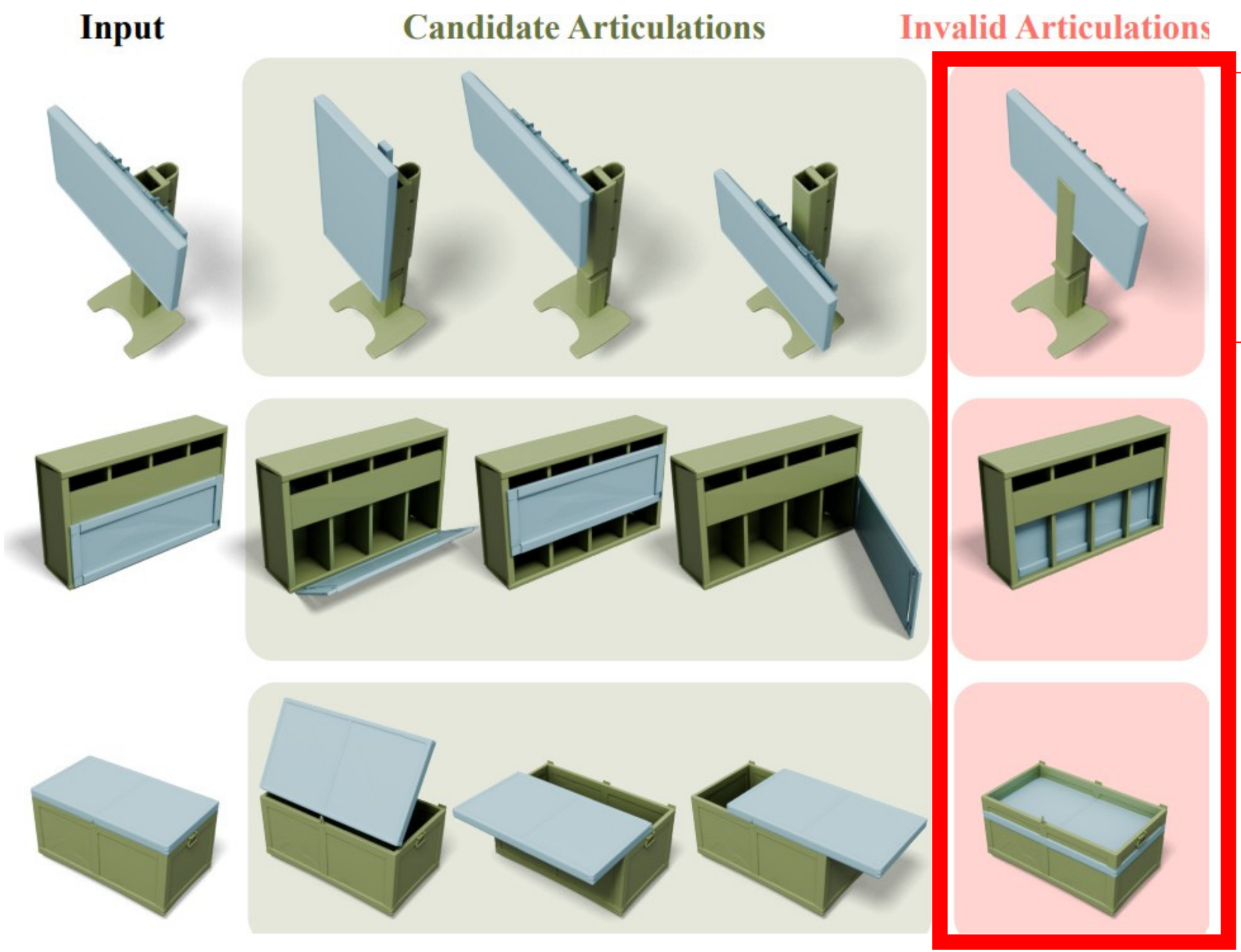
Candidate motion axes

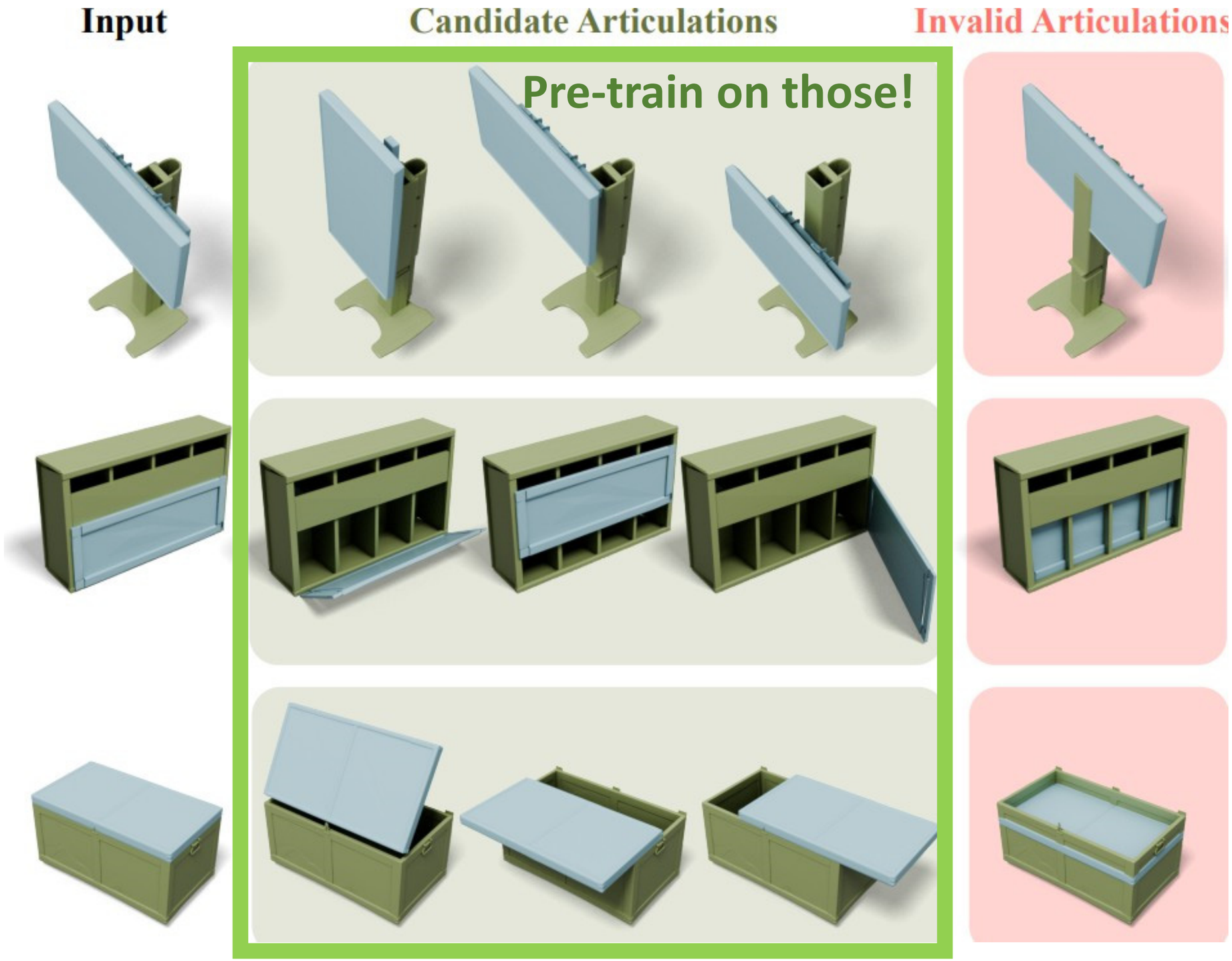
Candidate rotation centers

Self-supervision: simulator

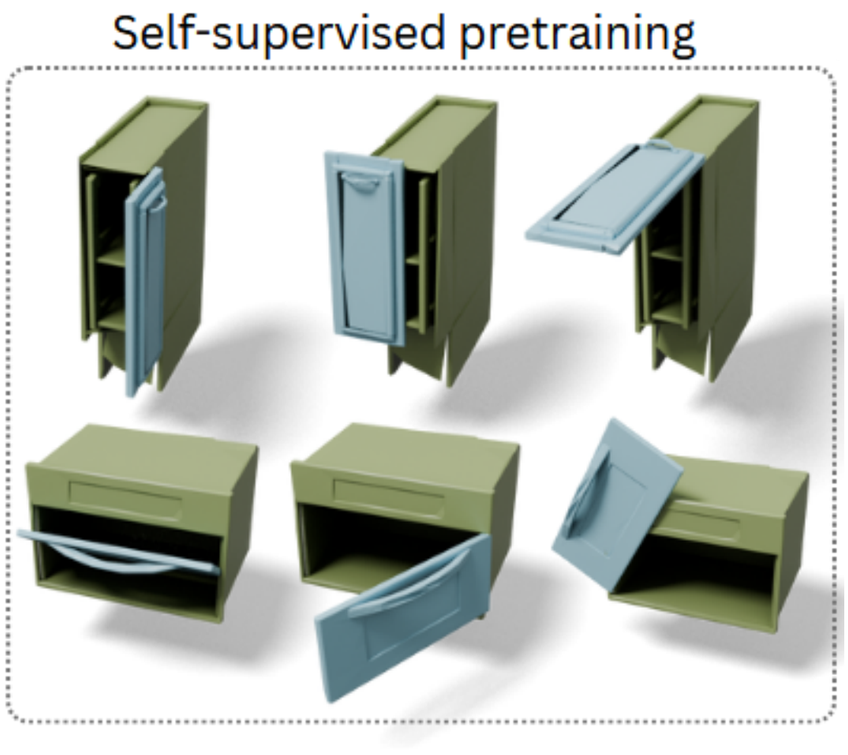



Input
Candidate Articulations
Invalid Articulations

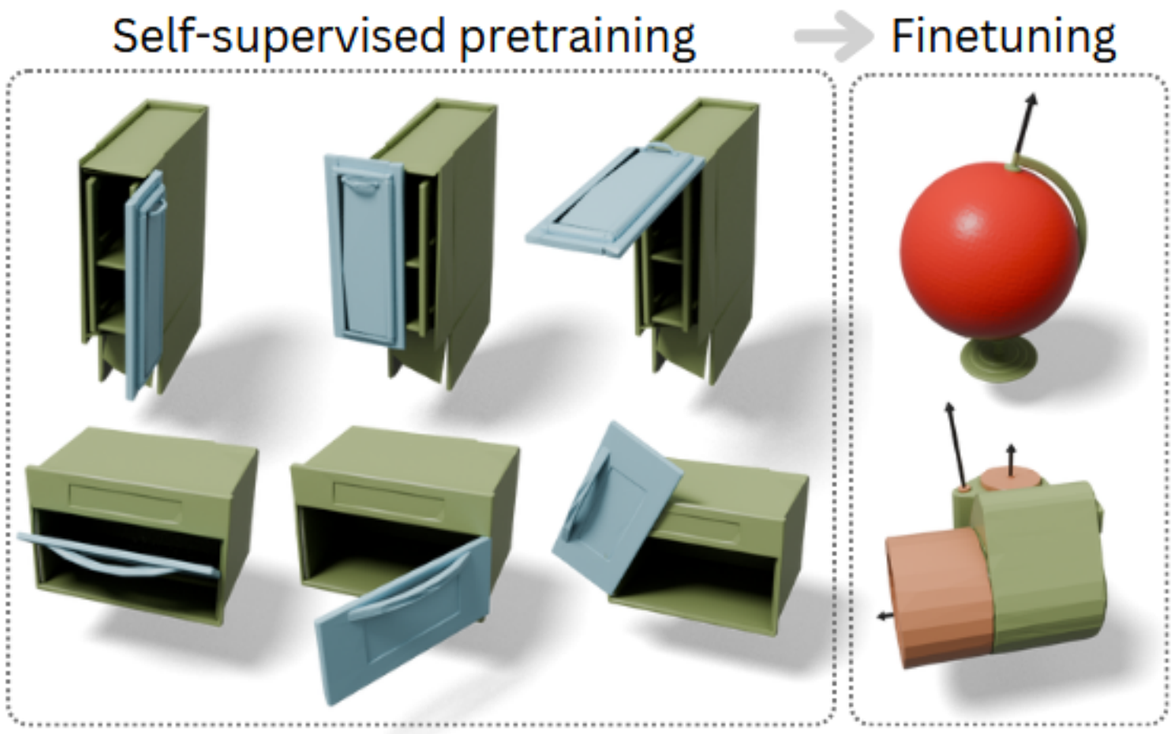

Input
Candidate Articulations
Invalid Articulations
reject articulations causing detachment & self-collisions


Input
Candidate Articulations
Invalid Articulations
Pre-train on those!

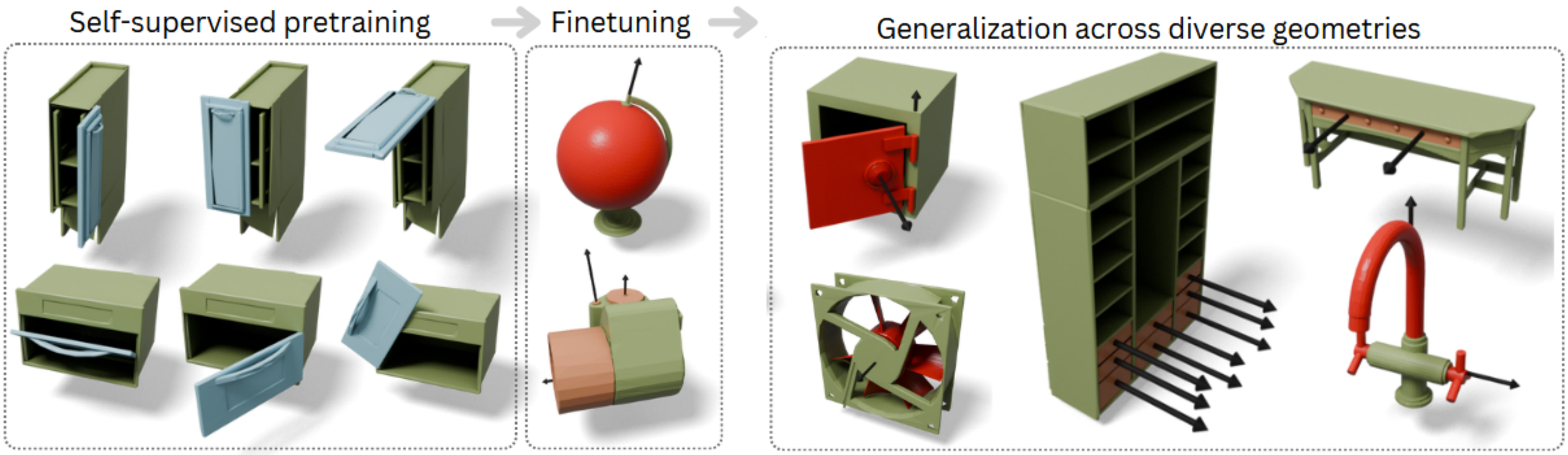
GEOPARD in a nutshell



GEOPARD in a nutshell



GEOPARD in a nutshell



Comparisons

Method	Axis Error ↓
CAGE	12.83
SINGAPO	11.20
GEOPARD	**9.18**

CAGE: Controllable Articulation Generation, Li et al. CVPR 24
SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects, Liu et al, ICLR 25

Comparisons

Method	Axis Error ↓	Rot. Center Error ↓
CAGE	12.83	0.11
SINGAPO	11.20	0.12
GEOPARD	**9.18**	**0.05**

CAGE: Controllable Articulation Generation, Li et al. CVPR 24
SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects, Liu et al, ICLR 25

Comparisons

Method	Axis Error ↓	Rot. Center Error ↓	Motion Type Precision ↑
CAGE	12.83	0.11	0.89
SINGAPO	11.20	0.12	0.90
GEOPARD	**9.18**	**0.05**	**0.92**

CAGE: Controllable Articulation Generation, Li et al. CVPR 24
SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects, Liu et al, ICLR 25

Comparisons

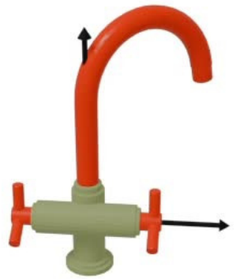
Method	Axis Error ↓	Rot. Center Error ↓	Motion Type Precision ↑	Motion Type Recall ↑
CAGE	12.83	0.11	0.89	0.89
SINGAPO	11.20	0.12	0.90	0.91
GEOPARD	**9.18**	**0.05**	**0.92**	**0.93**

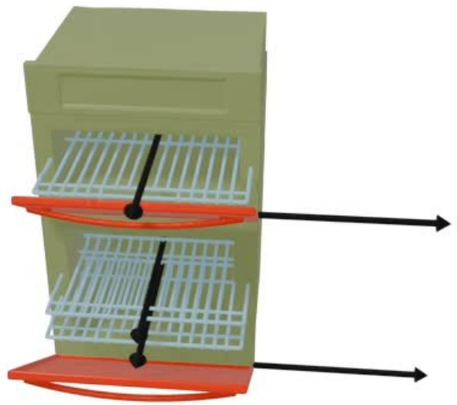
CAGE: Controllable Articulation Generation, Li et al. CVPR 24
SINGAPO: Single Image Controlled Generation of Articulated Parts in Objects, Liu et al, ICLR 25

Ablation

GEOPARD variant	Axis Error ↓	Rot. Center Error ↓	Motion Type Precision ↑	Motion Type Recall ↑
No pretrain.	10.67	0.06	0.87	0.88
With pretrain.	**9.18**	**0.05**	**0.92**	**0.93**

Generalization across diverse categories





Outline

1. GEOPARD: Geometric Pretraining for Articulation Prediction in 3D Shapes

2. SOPHY: **Generating Simulation-Ready Objects with Physical Materials**

3. Future directions

Goal: generate **simulation-ready** 3D objects

Input:
image or/and
text prompt

Goal: generate **simulation-ready** 3D objects

Input:
image or/and text prompt

Output:
object geometry+texture+physical materials

Background

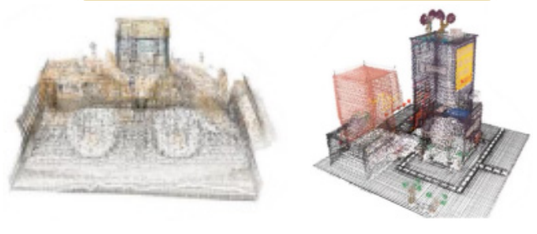
Triangular Mesh

MeshAnything V2
Chen *et al*, arXiv 24

NeRF

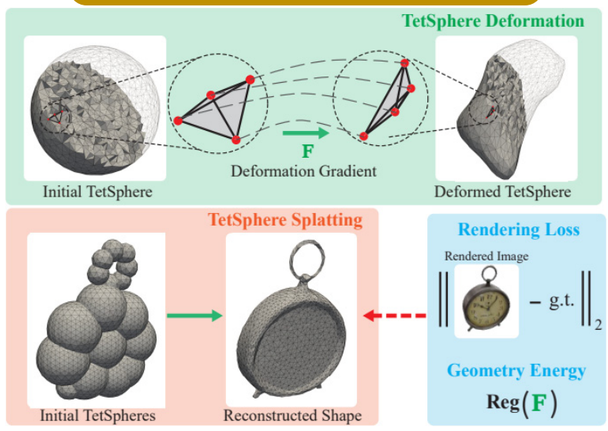
Latent-NeRF
Metzer *et al*, CVPR 23

3D Gaussians



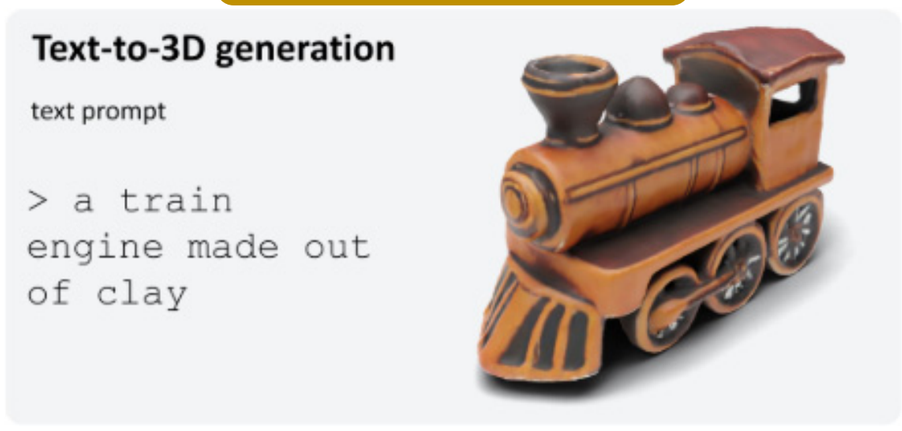
GRM
Xu *et al*, ECCV 24

Tetrahedral Mesh



TetSphere Splatting
Guo *et al*, ICLR 25

Textured Mesh



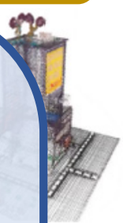
Meta 3D AssetGen
Siddiqui *et al*, NeurIPS 24

Background

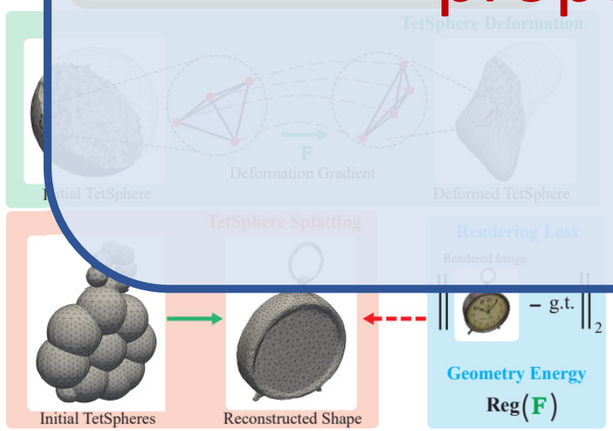
Triangular Mesh

NeRF

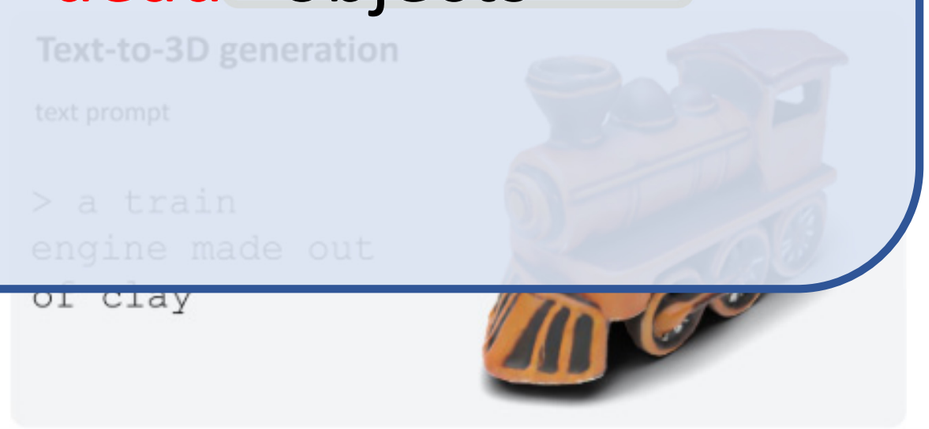
3D Gaussians



Completely agnostic to physics & material properties i.e., "dead" objects



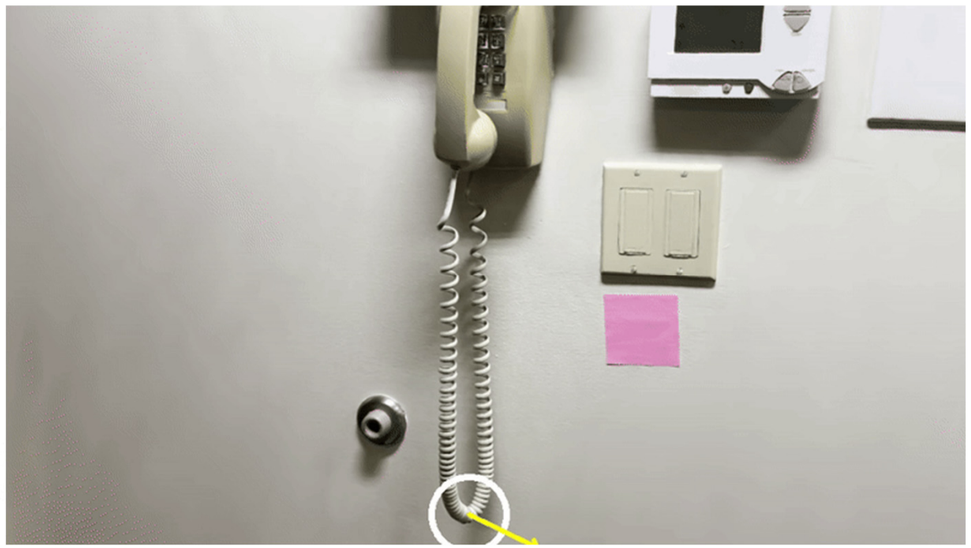
TetSphere Splatting
Guo *et al*, ICLR 25



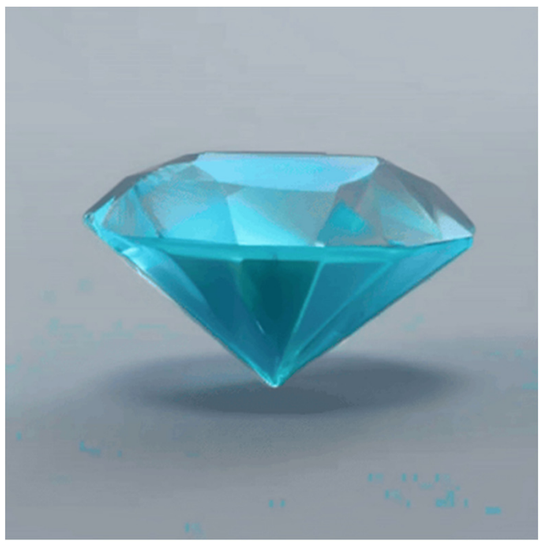
Meta 3D AssetGen
Siddiqui *et al*, NeurIPS 24

Prior Work

PhysGaussian



PhysDreamer



PhysMotion

PhysGaussian: Physics-integrated 3D Gaussians for Generative Dynamics. CVPR 2024.
PhysDreamer: Physics-Based Interaction with 3D Objects via Video Generation. ECCV 2024.
PhysMotion: Physics-Grounded Dynamics From a Single Image. arXiv 2024.

Prior Work

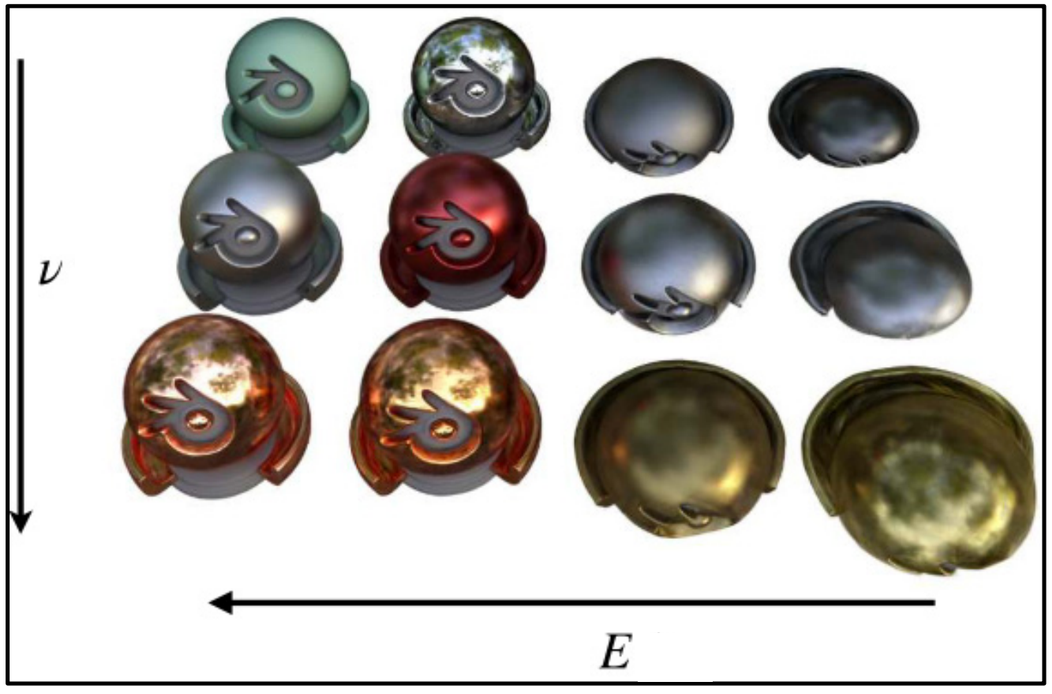
PhysGaussian: Physics-integrated 3D Gaussians for Generative Dynamics. CVPR 2024.
PhysDreamer: Physics-Based Interaction with 3D Objects via Video Generation. ECCV 2024.
PhysMotion: Physics-Grounded Dynamics From a Single Image. arXiv 2024.

Challenges

1. Lack of available datasets that include detailed material parameters

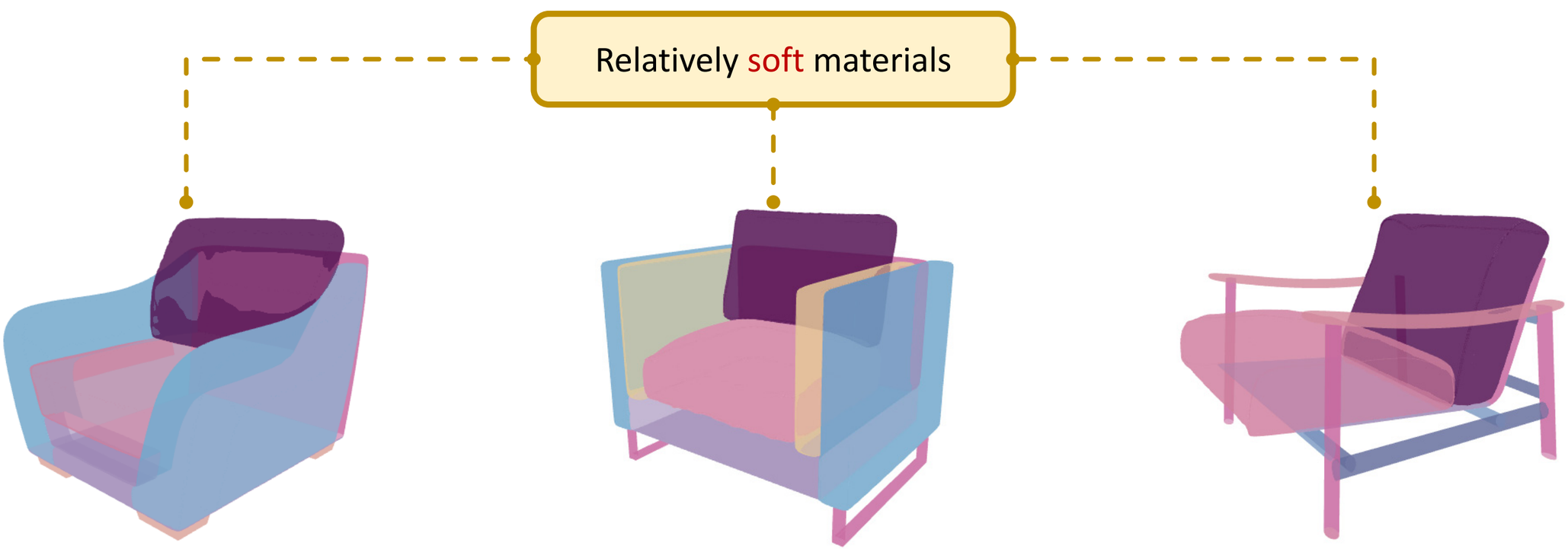
Material Parameters

Young's modulus E
Poisson's ratio ν
Yield stress σ
Friction angle ϕ
…



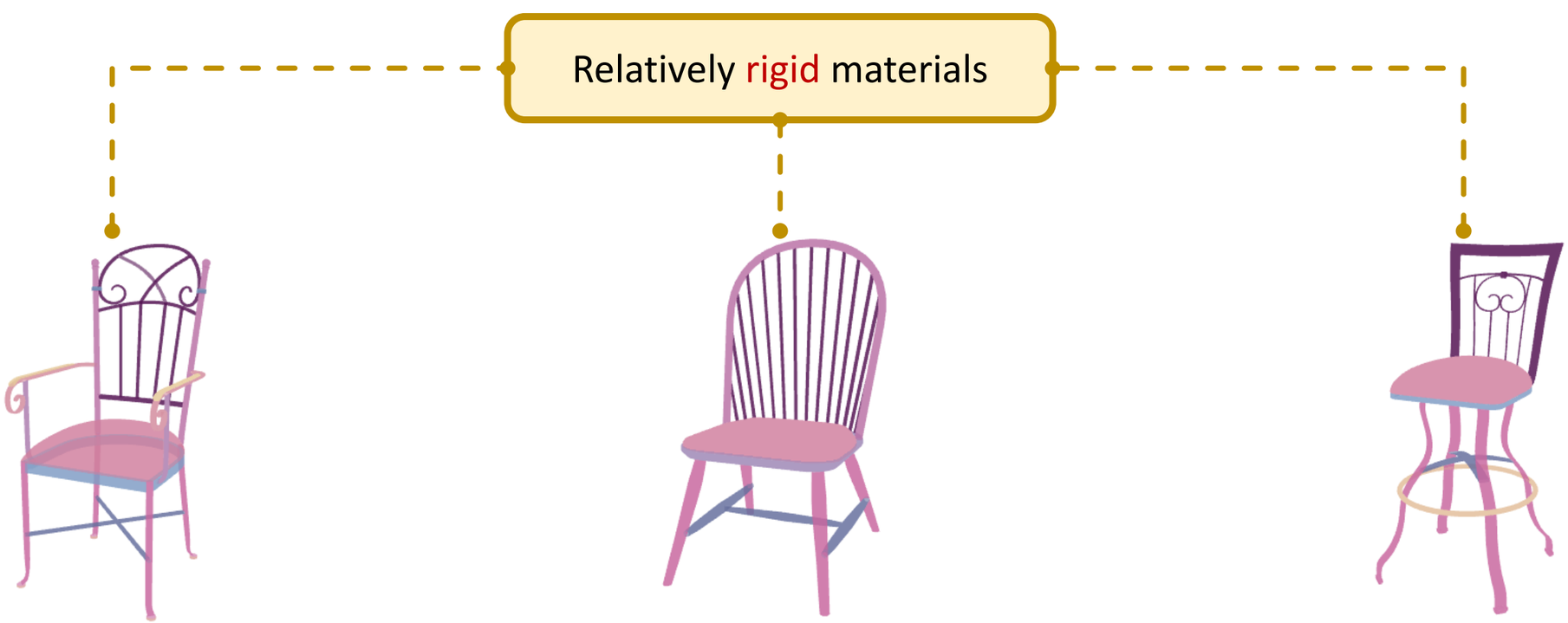
Challenges

2. How to jointly model the interplay of shape and material?



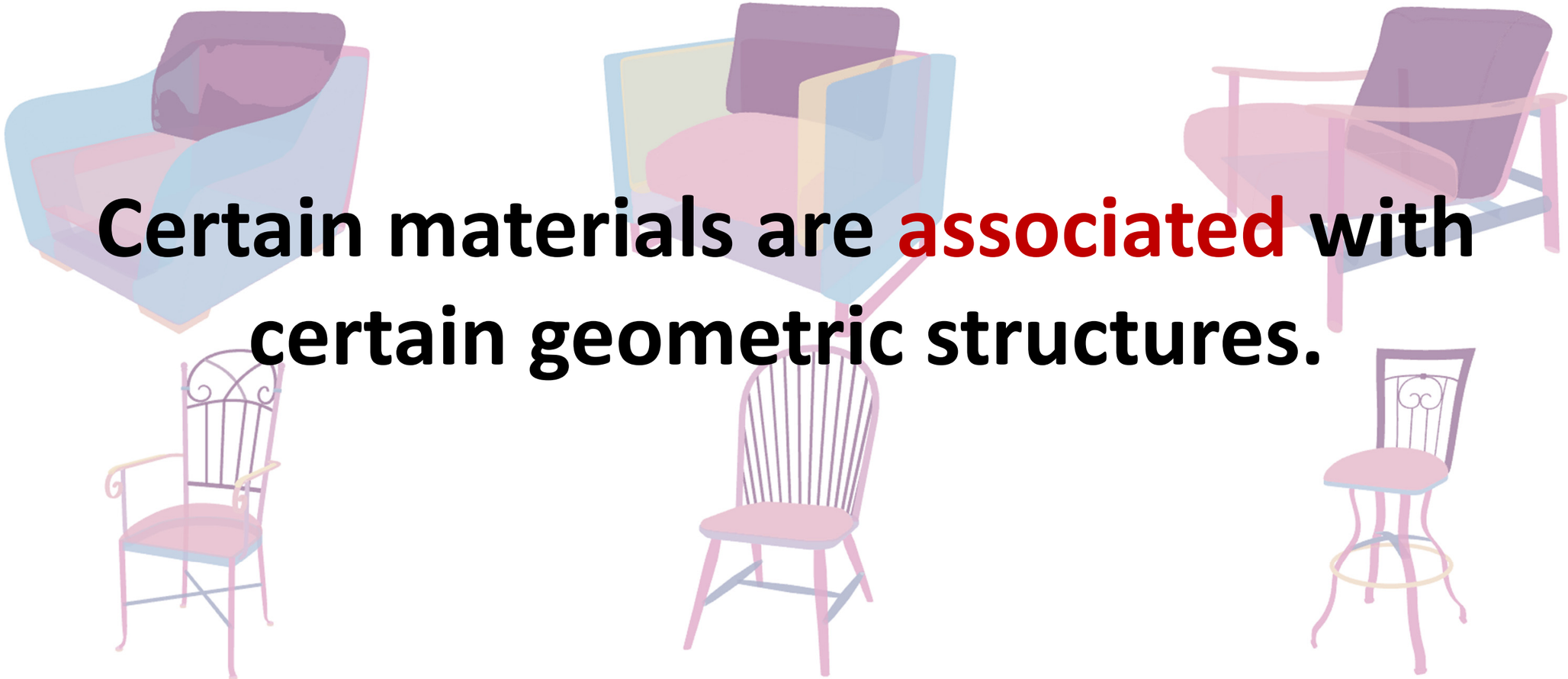
Challenges

2. How to jointly model the interplay of shape and material?



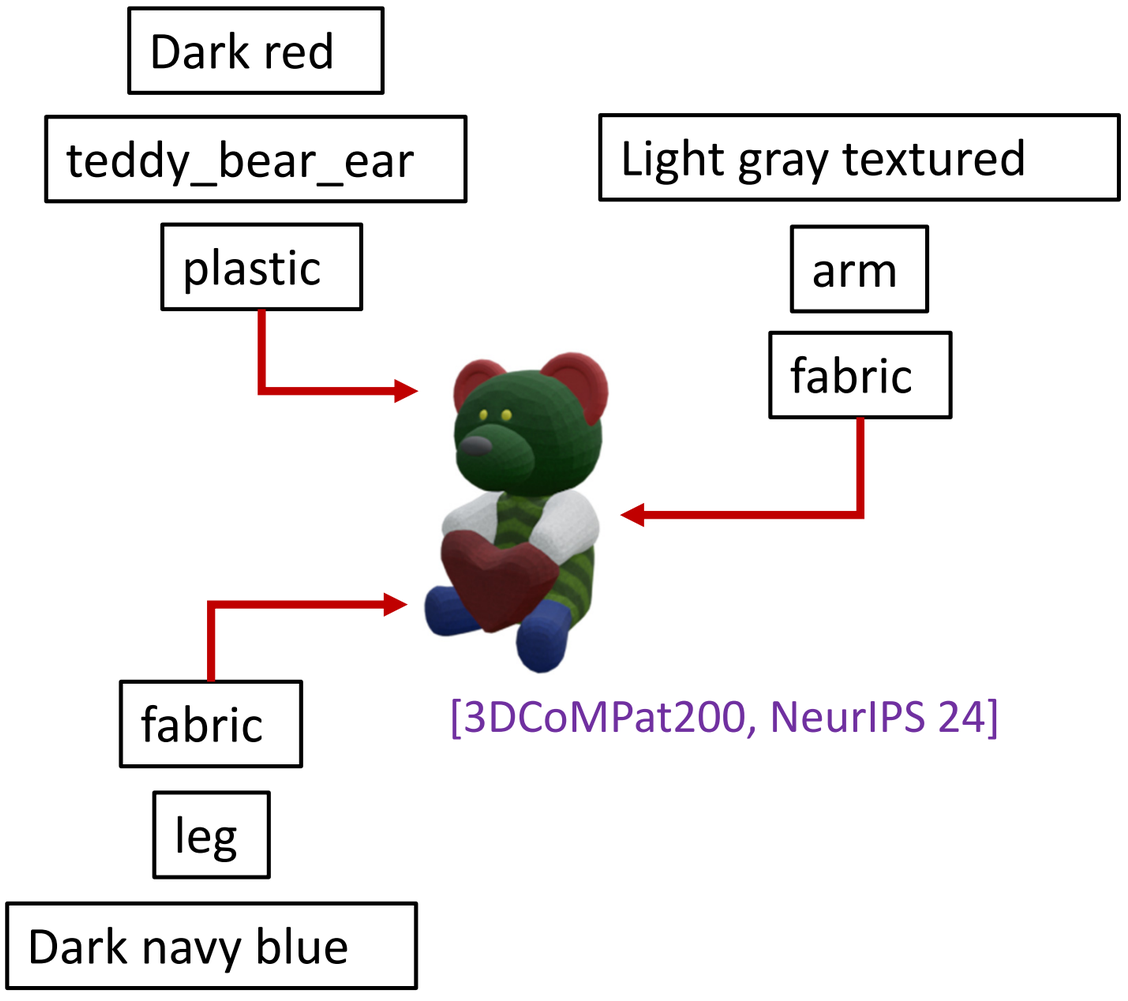
Challenges

2. How to jointly model the interplay of shape and material?



Certain materials are associated with certain geometric structures.

Material-annotated Dataset



[3DCoMPat200, NeurIPS 24]

12 shape categories, 3K objects, 15K parts

bag	bed	chair	crib	hat	headband

love_seat	pillow	planter	sofa	teddy_bear	vase

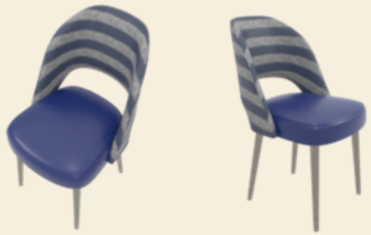
Automated Assembled Text

The teddy bear is made of *a light gray fabric arm*, *a dark navy-blue fabric leg*, *a dark red fabric ear*, …

Material-annotated Dataset

Assistance from Vision Language Models (VLMs)!

Rendered Images



Text Descriptions

A chair made of 3 parts: (1) a pewter gray **metal** *leg*, (2) a dark navy blue **fabric** *seat*, and (3) a blue, gray stripes **fabric** *backrest*.

Query Prompts

Please choose the most suitable **fine-grained material type** for the part(s): [*seat*, *backrest*] . . . Available choices for **fabric**: [cotton, wool, . . .] . . .

VLM Output

```
{
  "seat": "polyester",
  "backrest": "cotton"
}
```

Material-annotated Dataset

Assistance from Vision Language Models (VLMs)!

Rendered Images

Text Descriptions

A chair made of 3 parts: (1) a pewter gray **metal** *leg*, (2) a dark navy blue **fabric** *seat*, and (3) a blue, gray stripes **fabric** *backrest*.

Query Prompts

Please choose the most suitable **fine-grained material type** for the part(s): [*seat*, *backrest*] . . . Available choices for **fabric**: [cotton, wool, . . .] . . .

VLM Output

```
{
  "seat": "polyester",
  "backrest": "cotton"
}
```

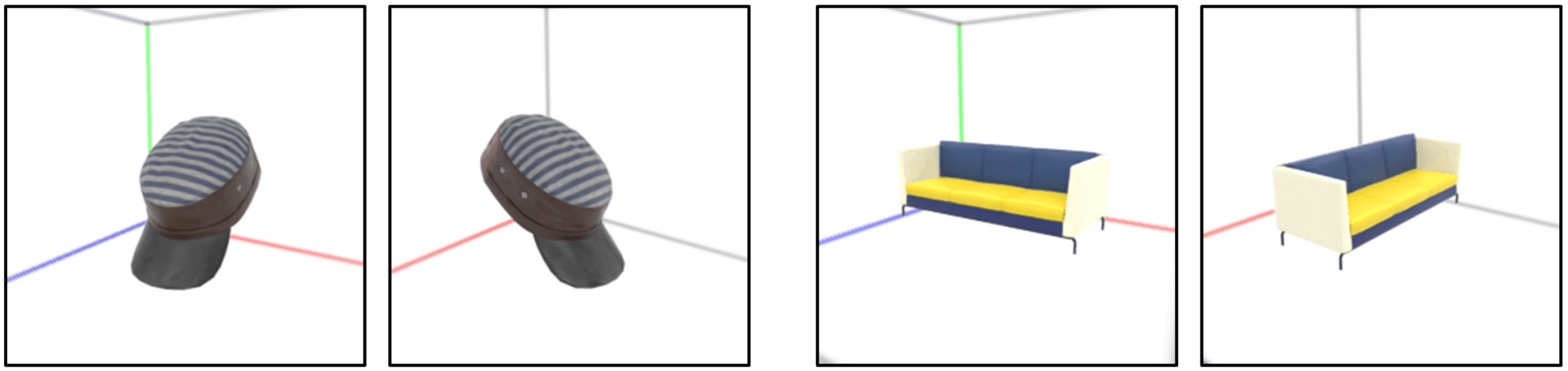
A chair made of 3 parts: (1) a pewter gray **metal** *leg*, (2) a dark navy blue (polyester) **fabric** *seat*, and (3) a blue, gray stripes (cotton) **fabric** *backrest*.

Please specify the **material models**, along with **material parameters** to simulate the motion of the object . . .

```
{
  "leg": {
    "E": 200000000000,
    "ν": 0.3,
    . . .
  }
  "seat": . . . ,
  . . .
}
```

Simulations

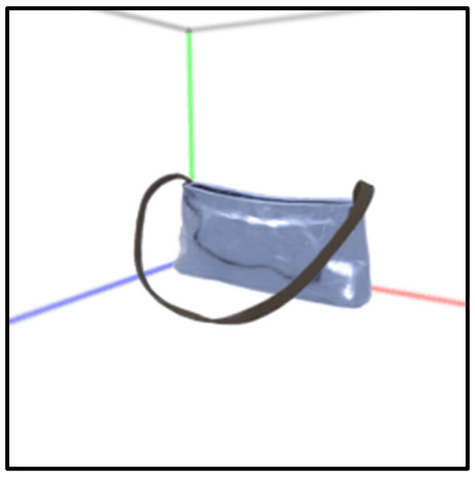
Test case: Drop

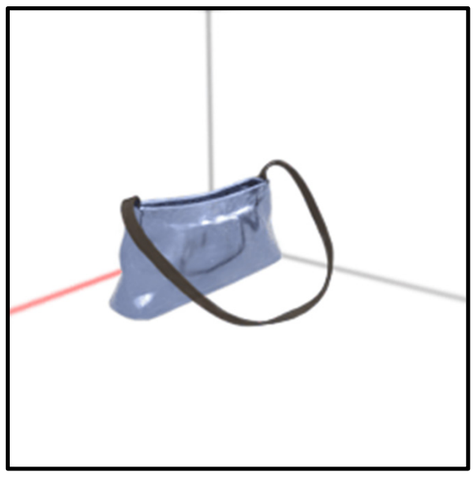


Hat **Sofa**

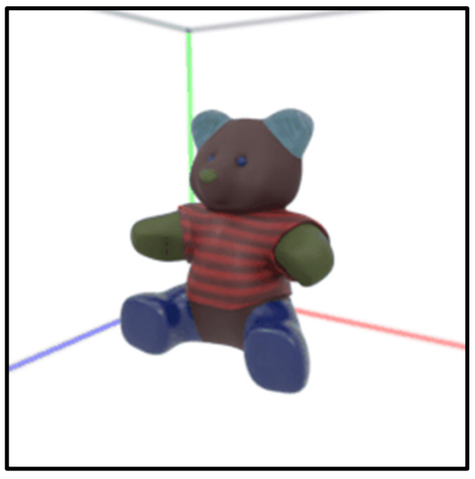
Simulations

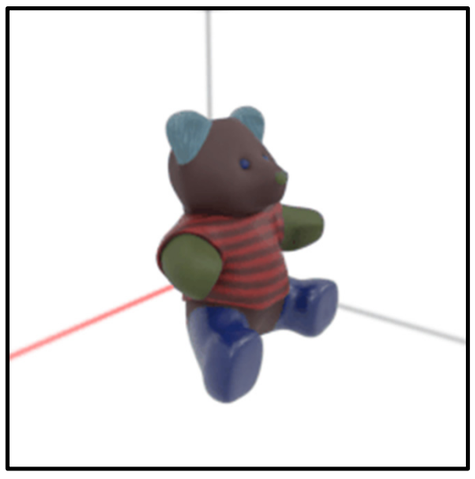
Test case: Throw





Bag

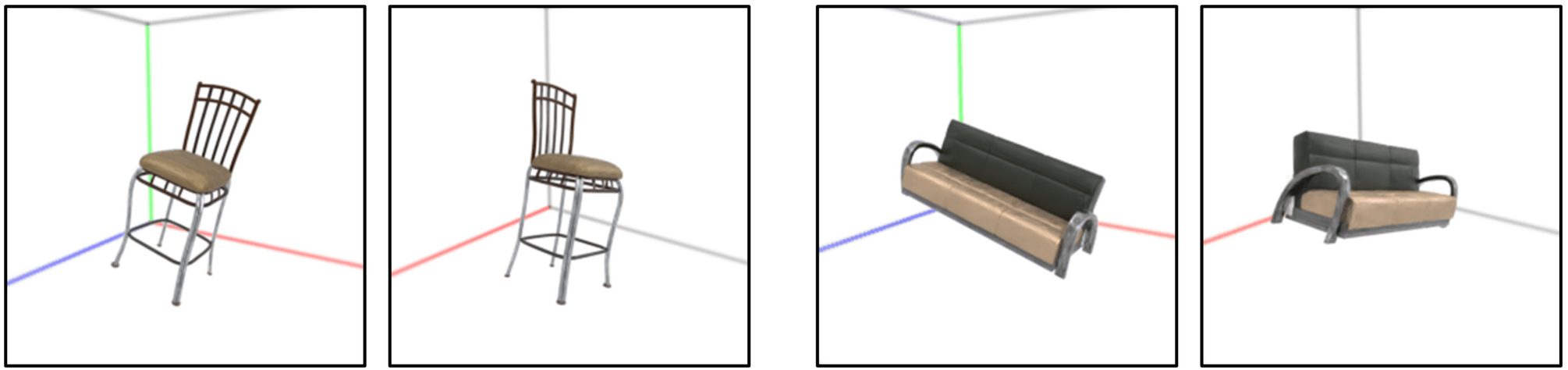




Teddy Bear

Simulations

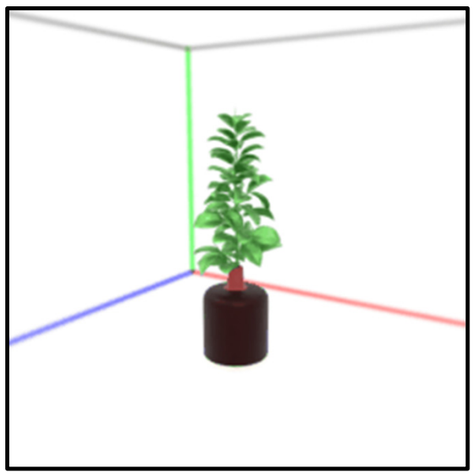
Test case: Tilt



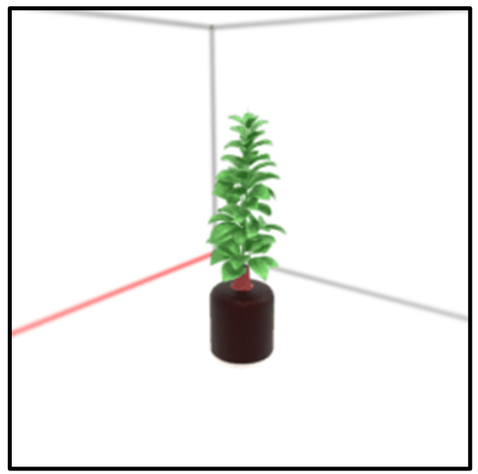
Chair **Sofa**

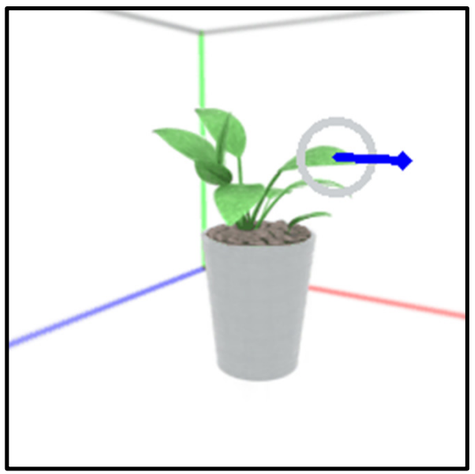
Simulations

Test case: Drag

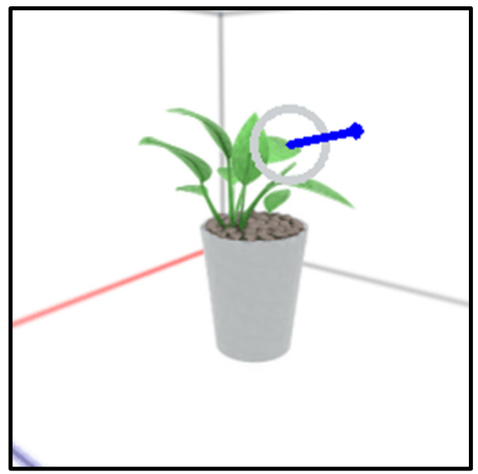


Planter



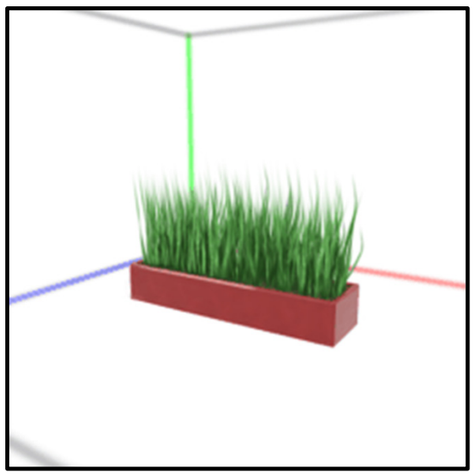


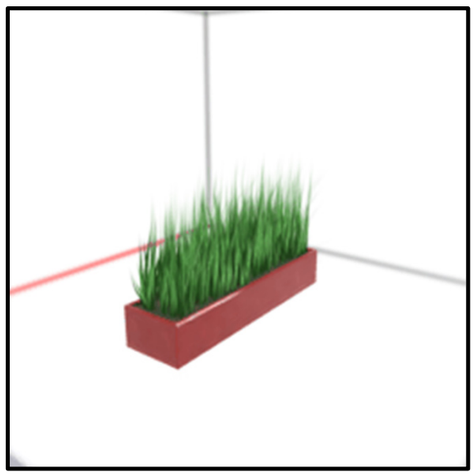
Planter



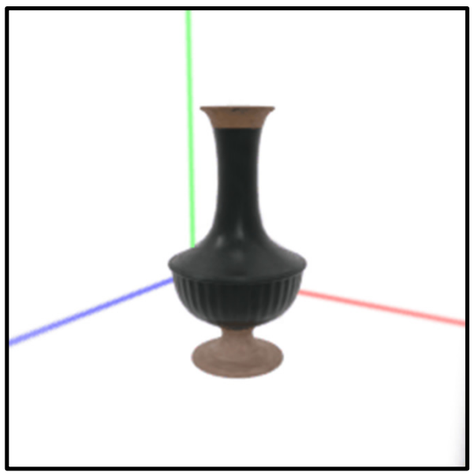
Simulations

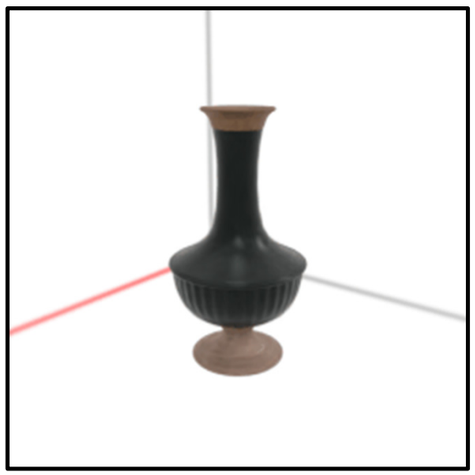
Test case: Wind





Planter

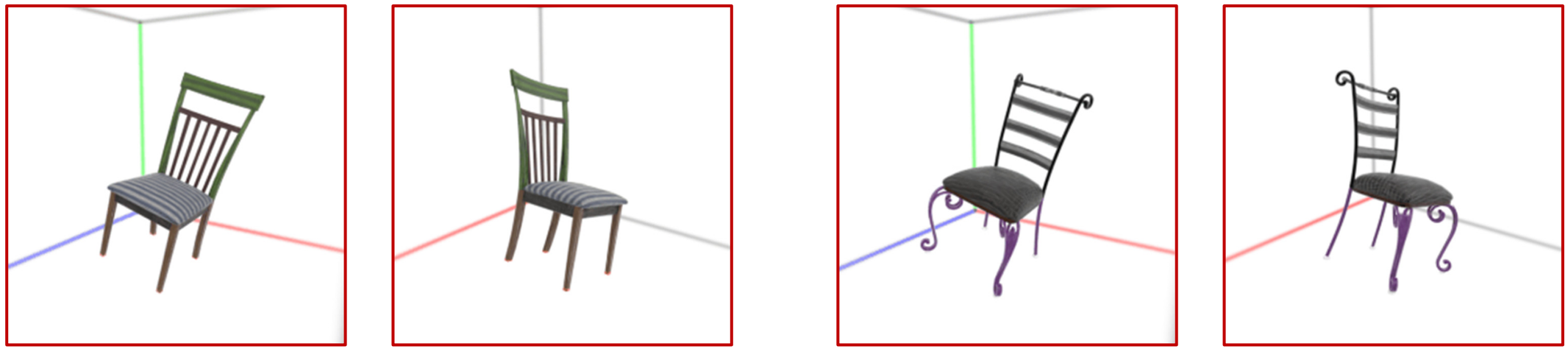




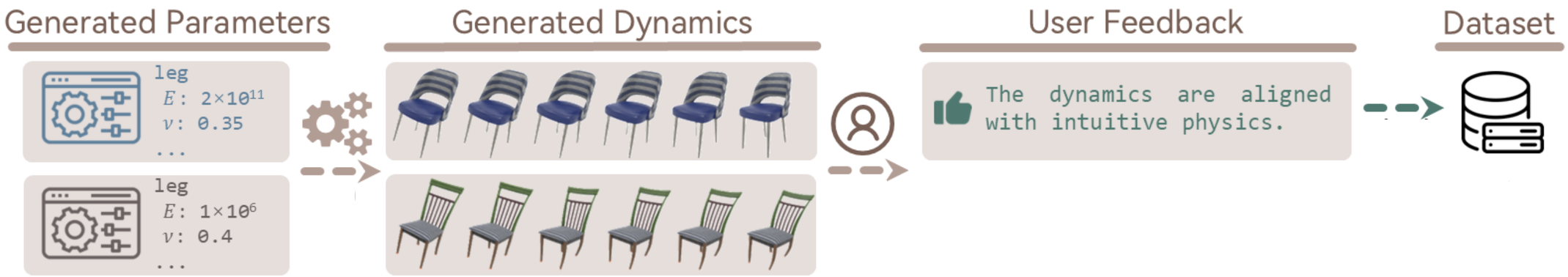
Vase

Material-annotated Dataset

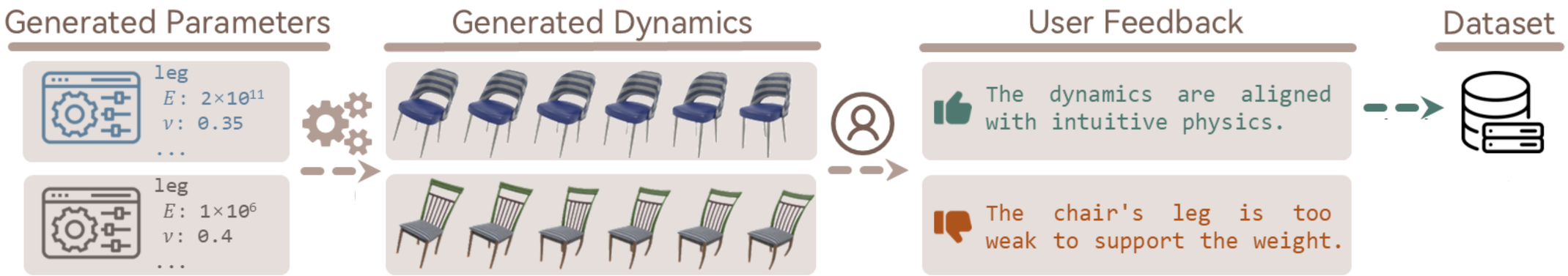
Expert vertification for checking simulation results from the VLM parameters!



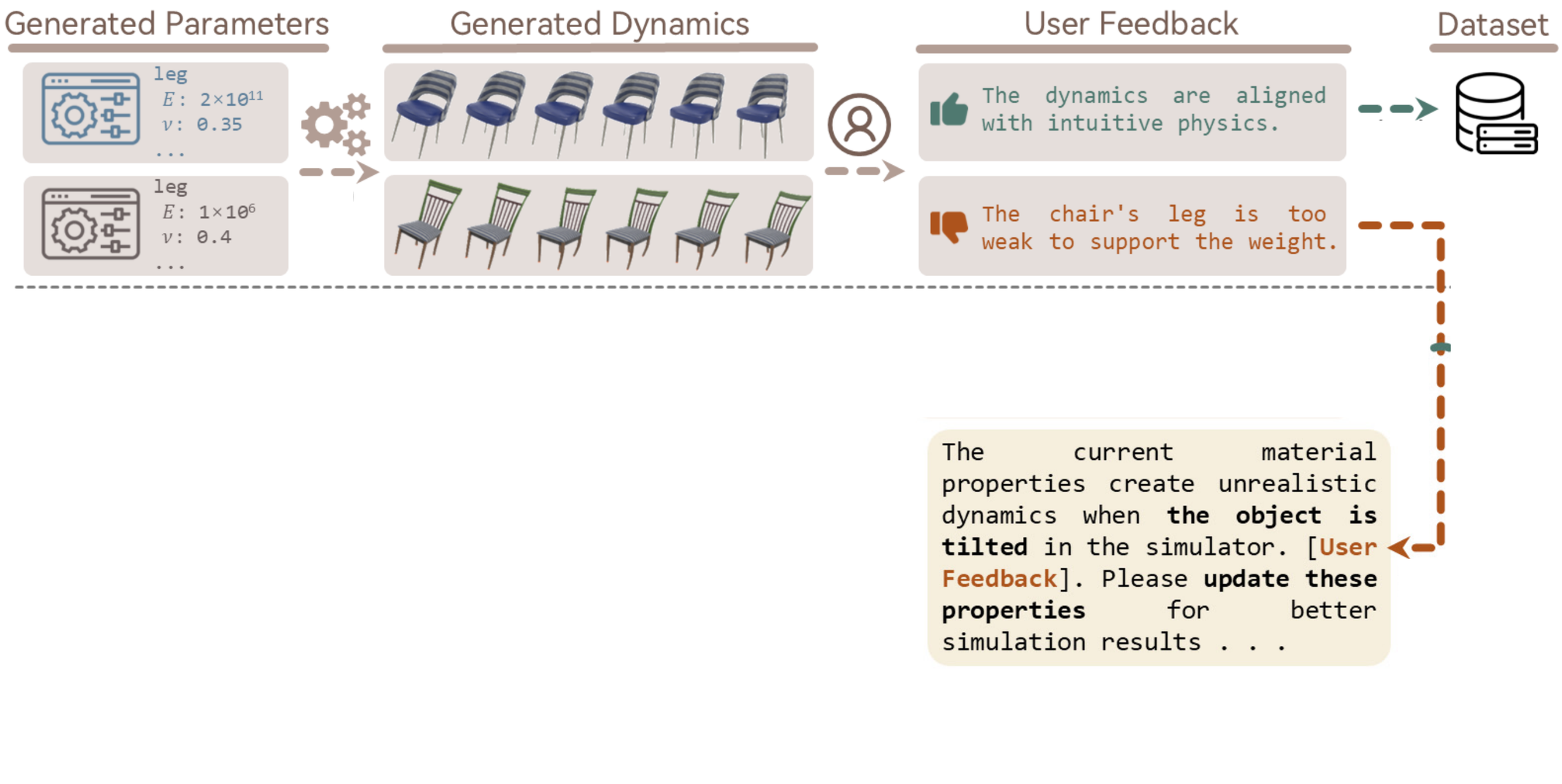
Annotation Pipeline



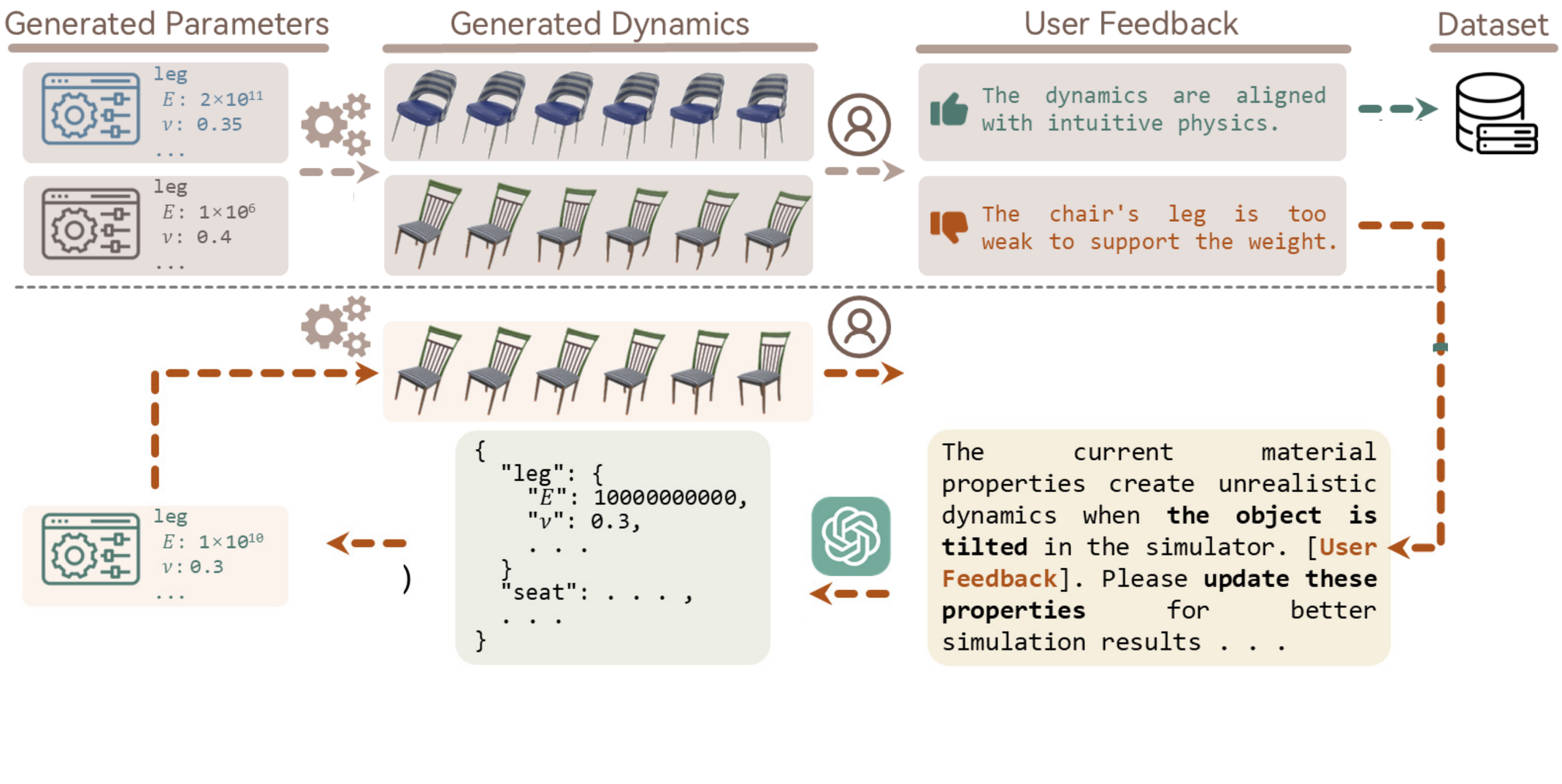
Annotation Pipeline



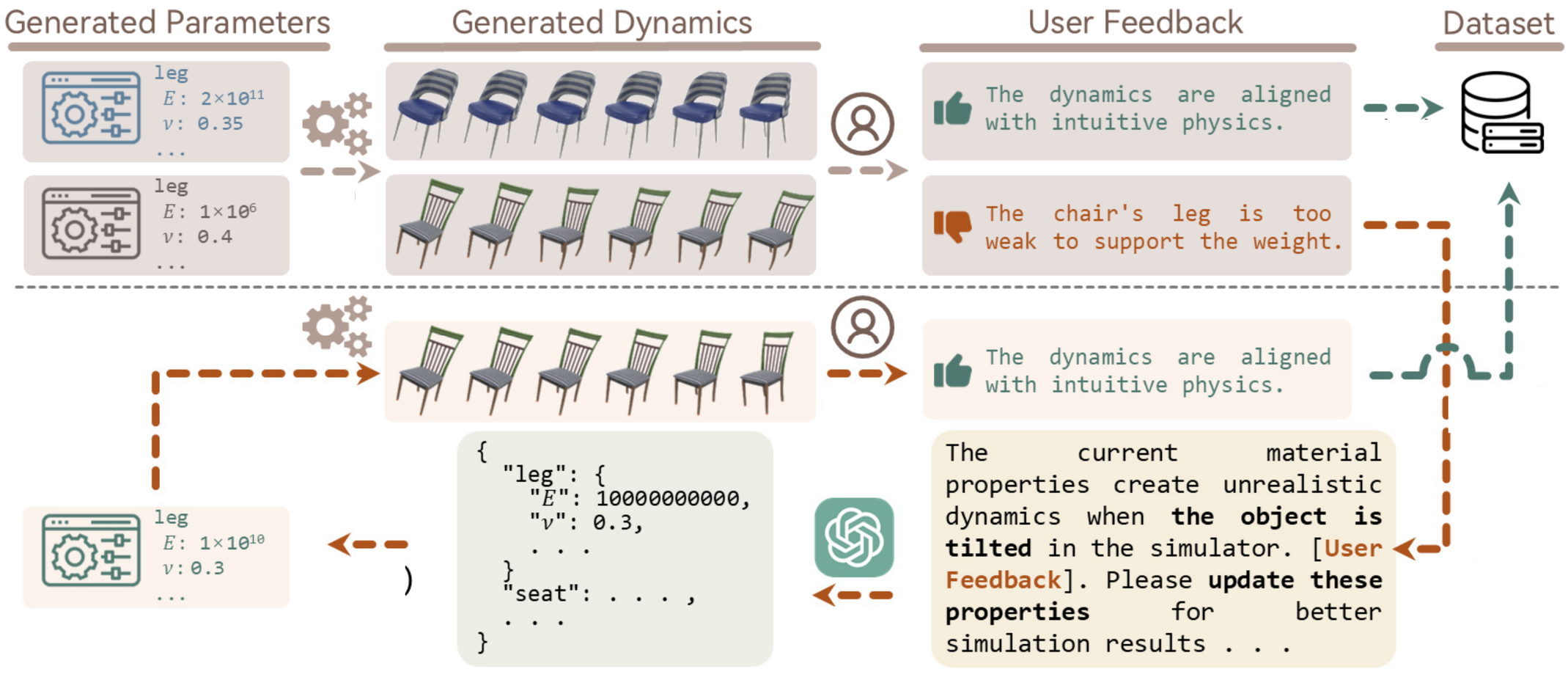
Annotation Pipeline



Annotation Pipeline



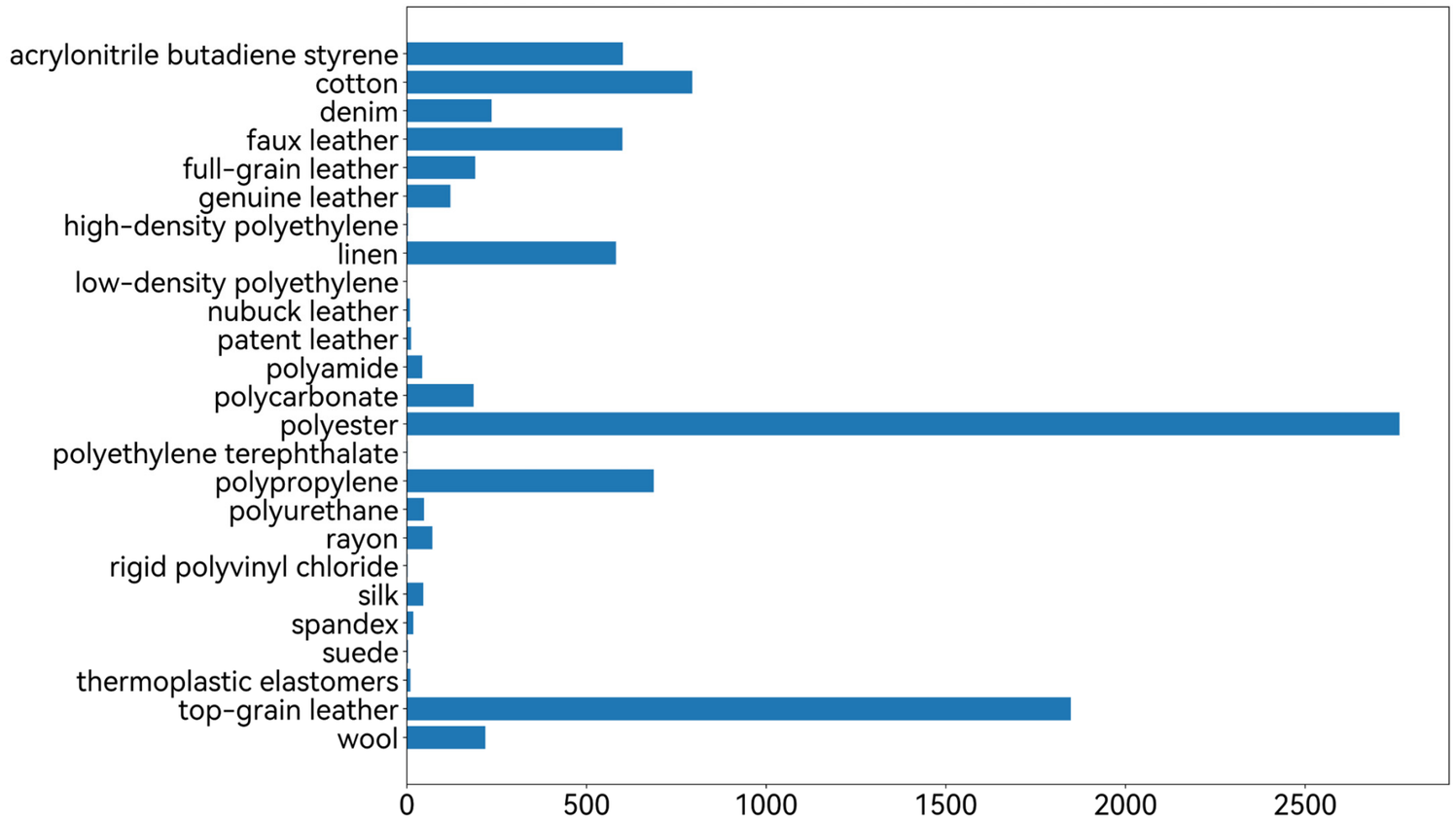
Annotation Pipeline



Dataset Statistics

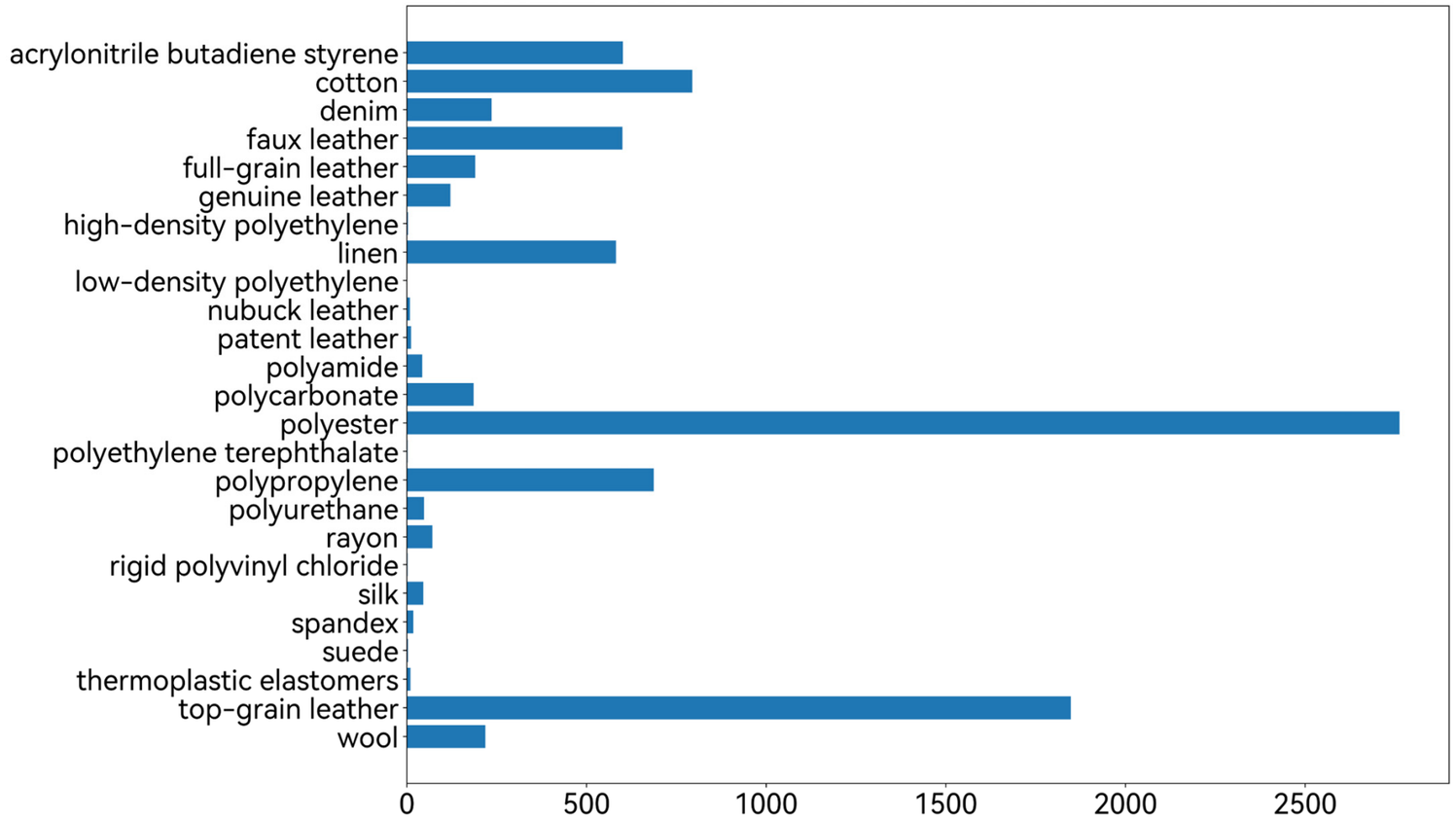
	bag	bed	chair	crib	hat	head-band	love seat	pillow	plant	sofa	teddy bear	vase	Total
Train	75	418	898	27	18	27	139	71	383	278	37	91	**2462**
Valid	10	26	52	5	2	5	21	11	21	18	6	3	**180**
Test	24	48	106	8	7	10	39	18	46	31	11	14	**362**

Dataset Statistics

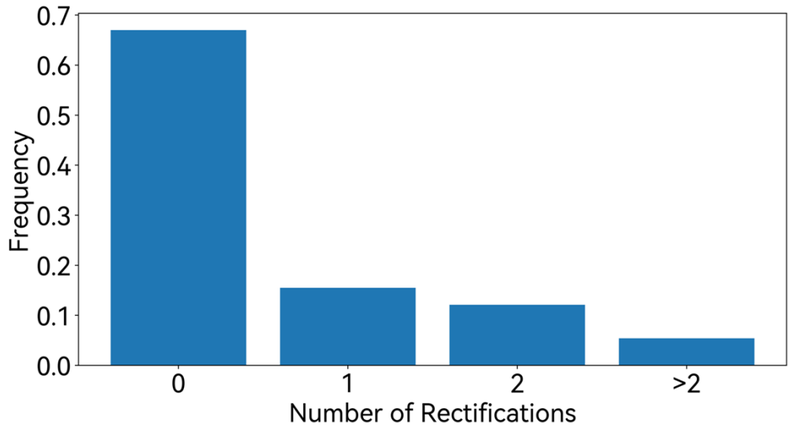


Statistics of fine-grained material types

Dataset Statistics



Statistics of fine-grained material types

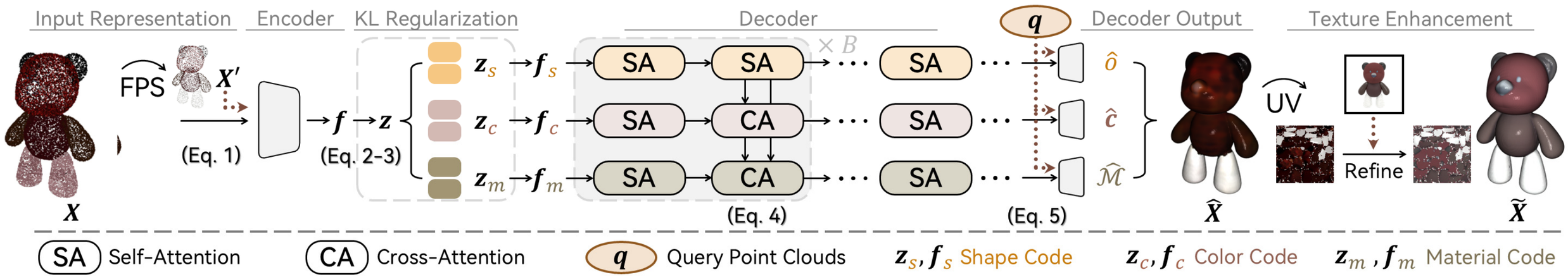


~ *1/3* of the data received at least one round of update

~ *5%* of the data received more than two rounds of updates

Statistics of expert rectification times

Learning latent spaces of geometry+texture+materials



Input Representation

For each surface point $\boldsymbol{p} \in \boldsymbol{P}$…

{

position,

rgb,

Young’s modulus,

Poisson’s ratio,

yield stress,

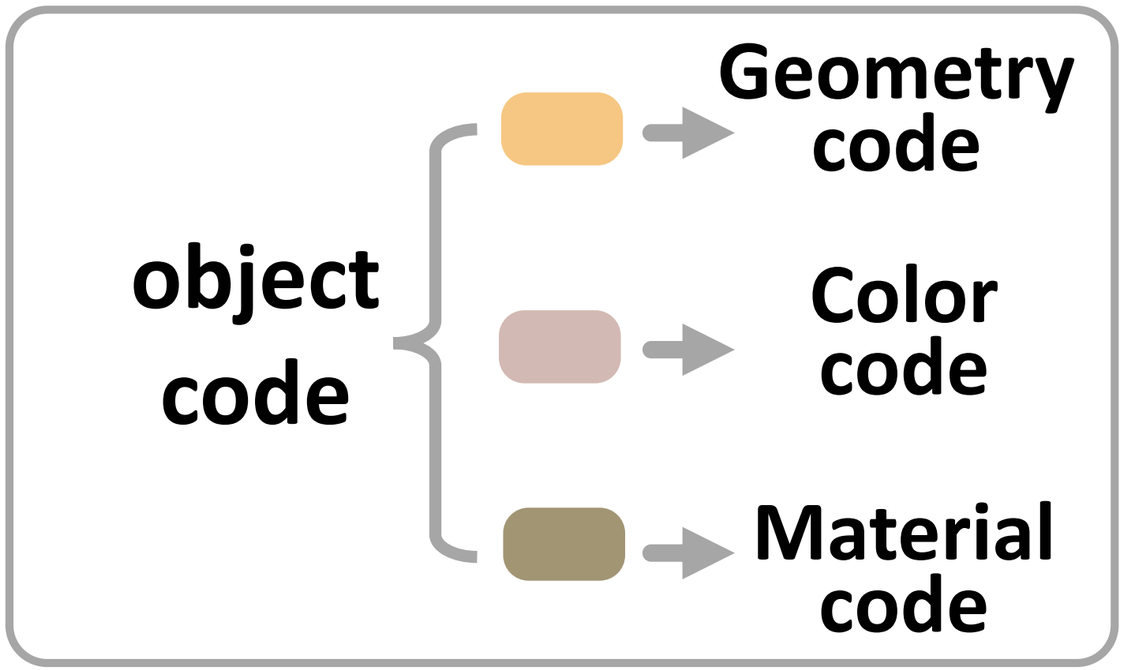
friction angle,

material type,

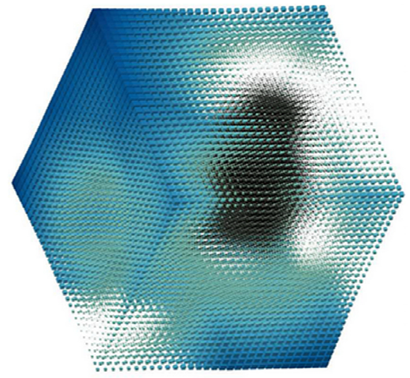
}

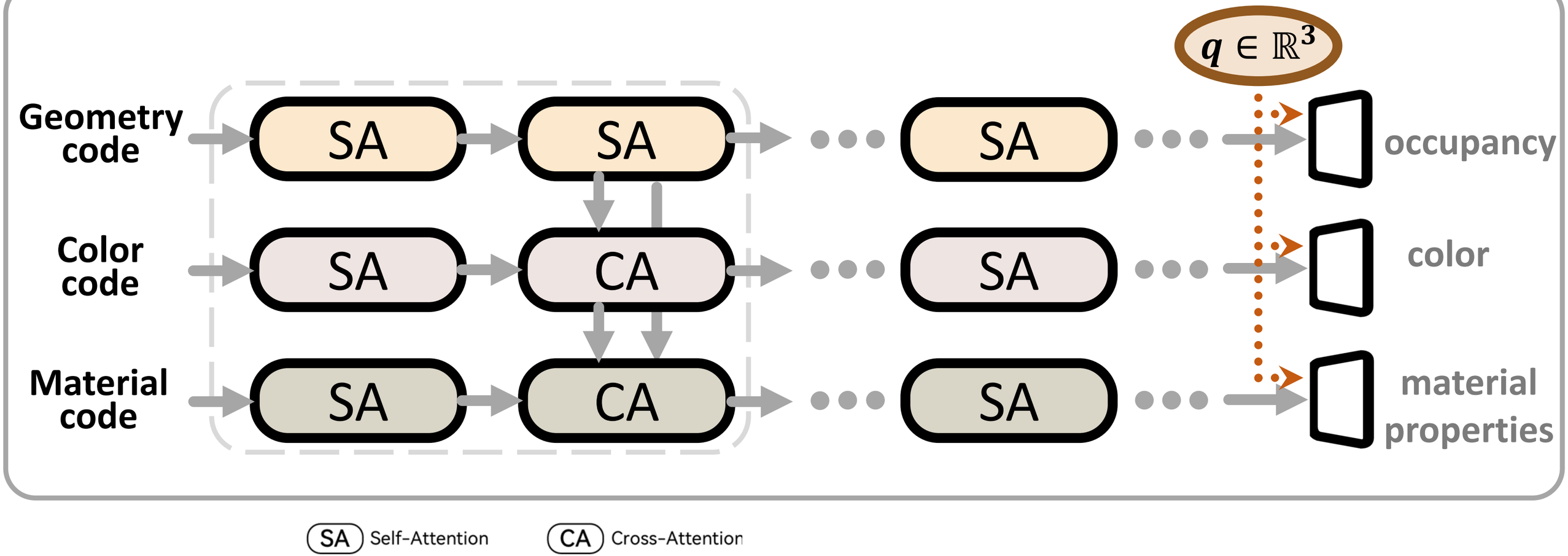
Shape & Texture & Material Encoder

Generative Model: code decomposition



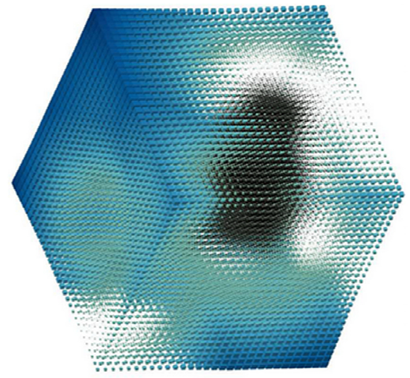
Decoder

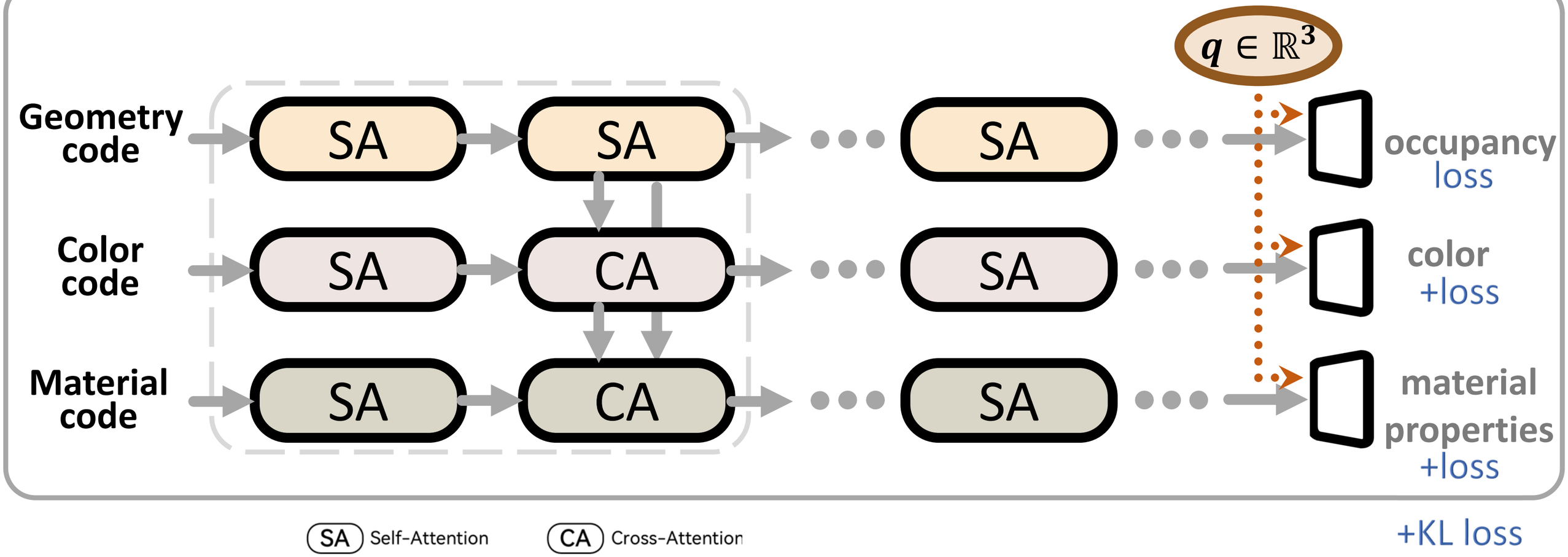




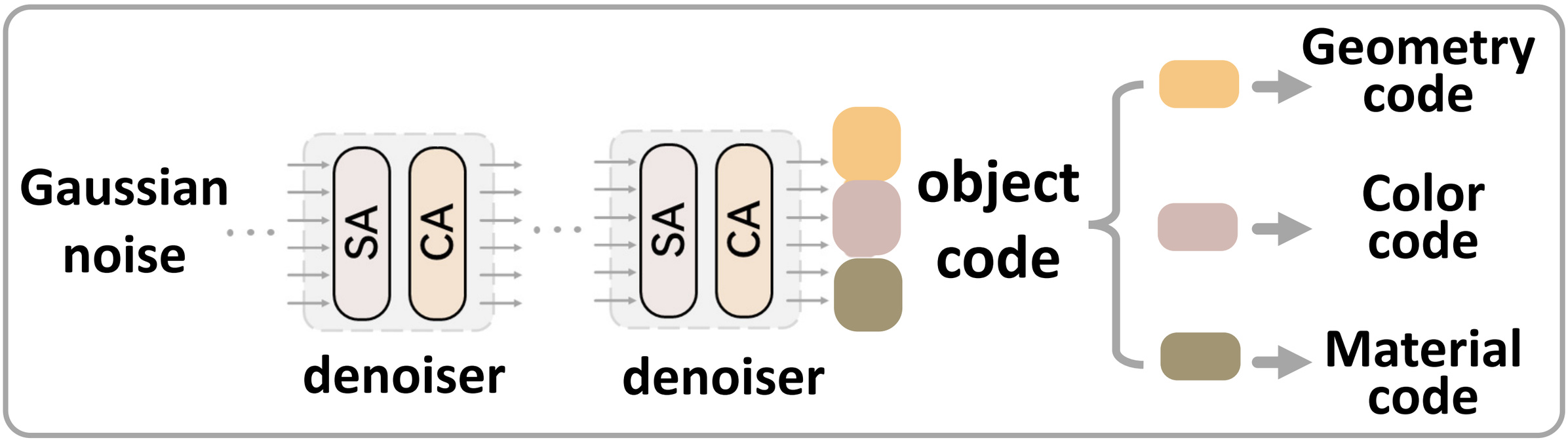
SA Self-Attention CA Cross-Attention

Losses

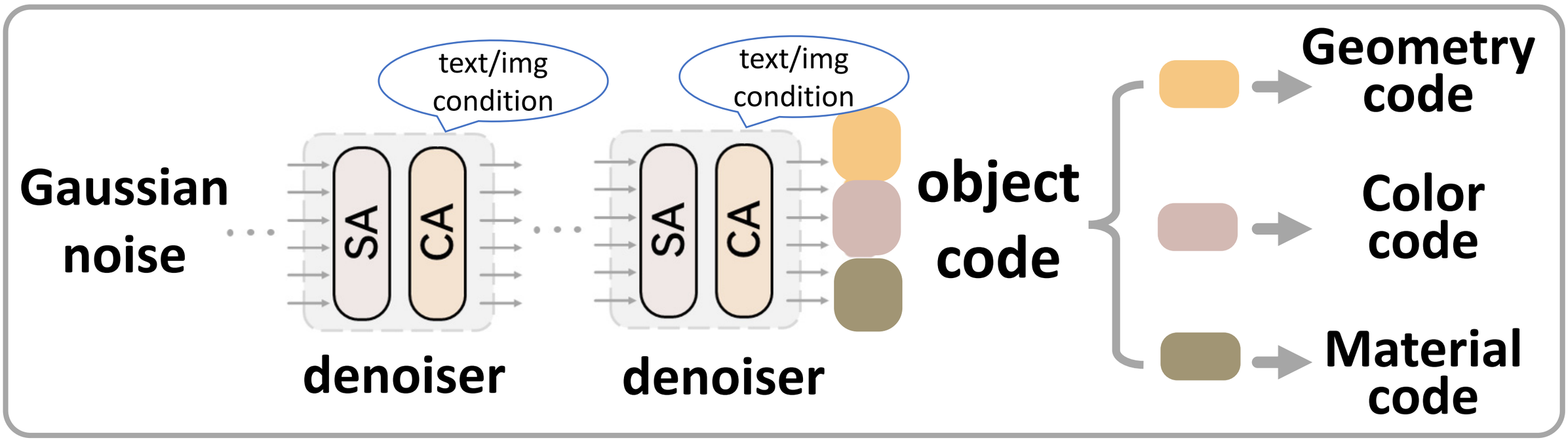



$q \in \mathbb{R}^3$
Geometry code
Color code
Material code
SA
SA
SA
SA
CA
SA
SA
CA
SA
occupancy loss
color +loss
material properties +loss
+KL loss
SA Self-Attention
CA Cross-Attention

Generative Model: diffusion



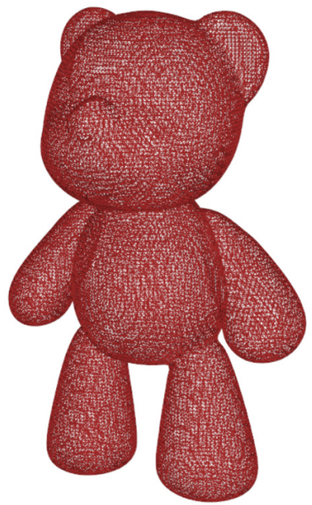
Generative Model: diffusion



Dynamics Generation

Volumetric

Sampling



MPM

Test case
Configuration

e.g. initial location
and velocity

The material point method for simulating continuum materials (MPM). SIGGRAPH 2016 courses.

Image-to-4D Experiments

Variant
“B. Dec”
“B. Perc”
SOPHY

B. Dec.: generates shape, then uses a subsequent decoder for predicting material properties

Image-to-4D Experiments

Variant
“B. Dec”
“B. Perc”
SOPHY

B. Perc.: off-the-shelf 3D generation followed by segmentation (Find3D) + ChatGPT-4o for material properties

Image-to-4D Experiments

Variant	Physical Commonsense [VideoPhy] ↑
“B. Dec”	-0.22
“B. Perc”	0.18
SOPHY	**0.40**

VideoPhy: Evaluating Physical Commonsense for Video Generation. arXiv 2024.

Image-to-4D Experiments

Variant	Physical Commonsense [VideoPhy] ↑	Text matching score [CLIP] ↑
“B. Dec”	-0.22	0.73
“B. Perc”	0.18	0.74
SOPHY	**0.40**	**0.77**

Image-to-4D Experiments

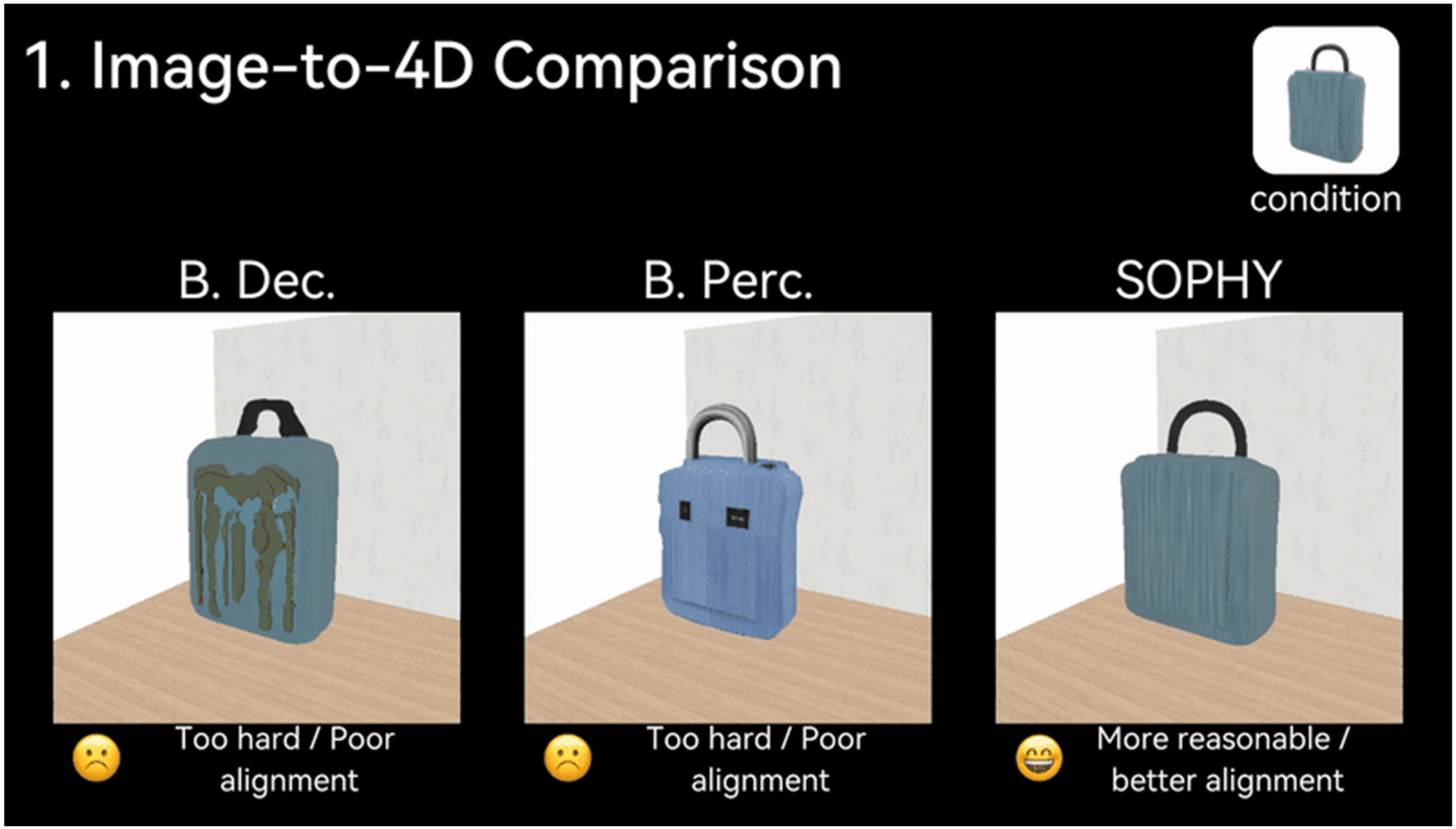
Variant	Physical Commonsense [VideoPhy] ↑	Text matching score [CLIP] ↑	Time (min) ↓
“B. Dec”	-0.22	0.73	5
“B. Perc”	0.18	0.74	7
SOPHY	**0.40**	**0.77**	**5**

Many ablations…

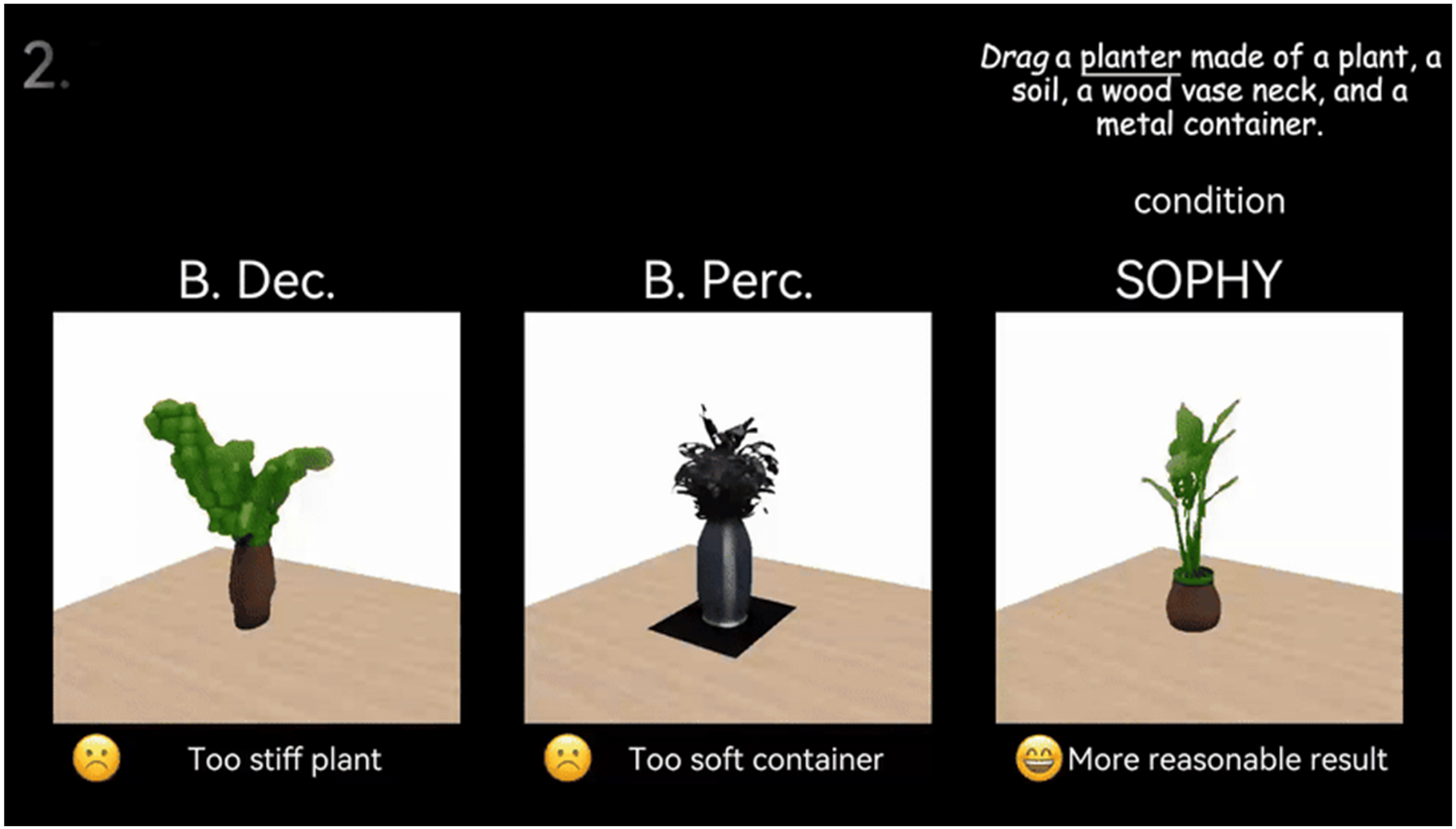
Metric	B. Dec.	w/o CA	Fused	SOPHY
M.B. Acc(%) ↑	71.04	92.77	93.23	**93.55**
MAE-log(E) ↓	1.18	0.50	0.47	**0.45**
MAE-ν($\times 10^{-2}$) ↓	4.10	3.06	**2.98**	3.06
MAE-log(σ) ↓	1.07	0.41	0.32	**0.29**
MAE-ϕ ($\times 10^{-2}$) ↓	5.45	1.89	**1.22**	1.28
MAE-ρ ↓	0.16	0.14	0.13	**0.09**
Sim-CD ($\times 10^{-3}$) ↓	53.84	17.47	10.05	**8.72**
MAE-c($\times 10^{-2}$) ↓	8.75	8.47	8.35	**8.13**
IoU(%) ↑	90.69	90.75	90.66	**90.89**
CD($\times 10^{-4}$) ↓	3.51	3.34	3.30	**3.02**
F-Score(%) ↑	93.65	93.77	93.79	**93.85**

M.B. Acc: Material Behavior Type Accuracy. MAE: Mean Absolute Error.

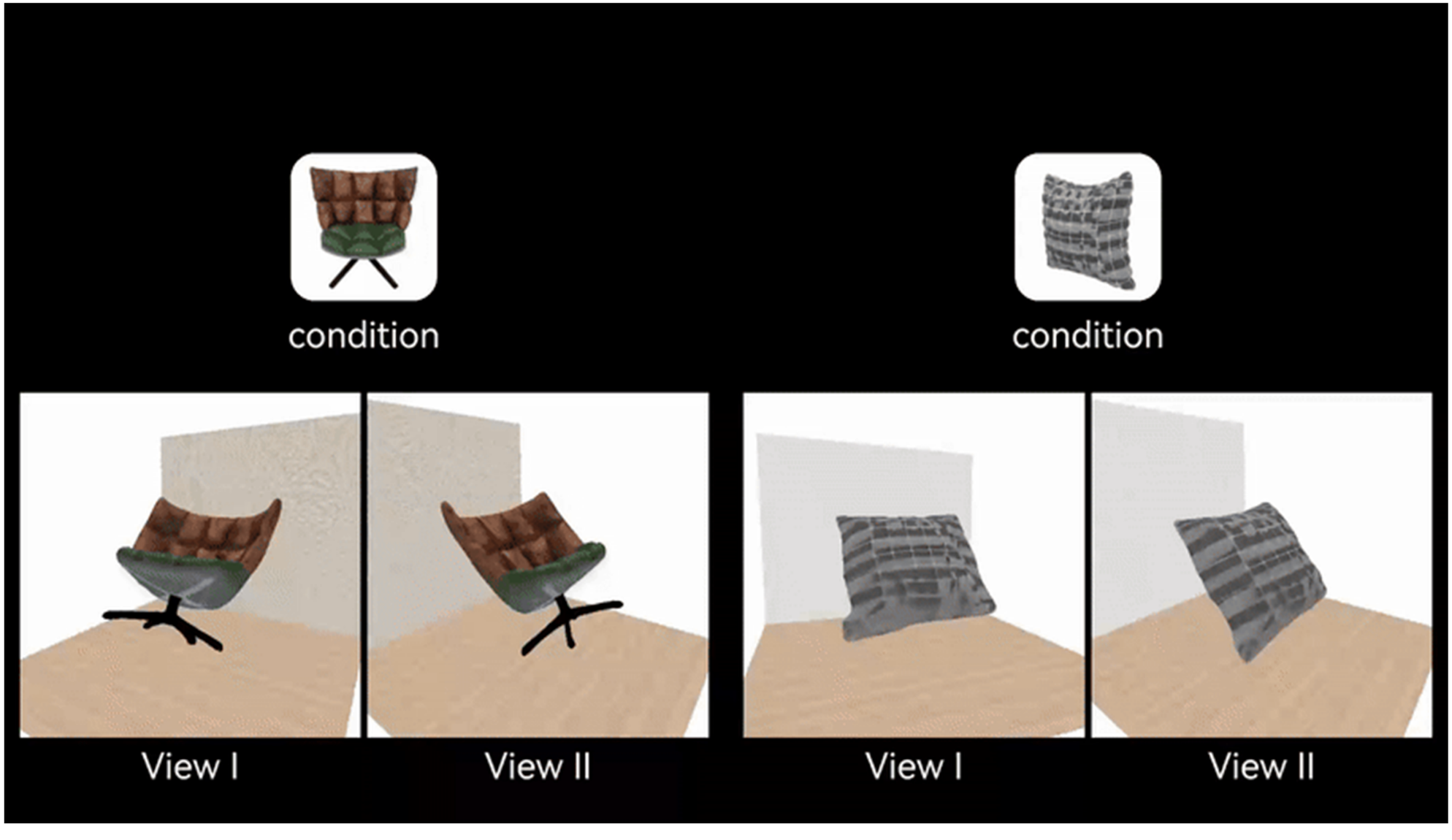
Image-to-4D Experiments



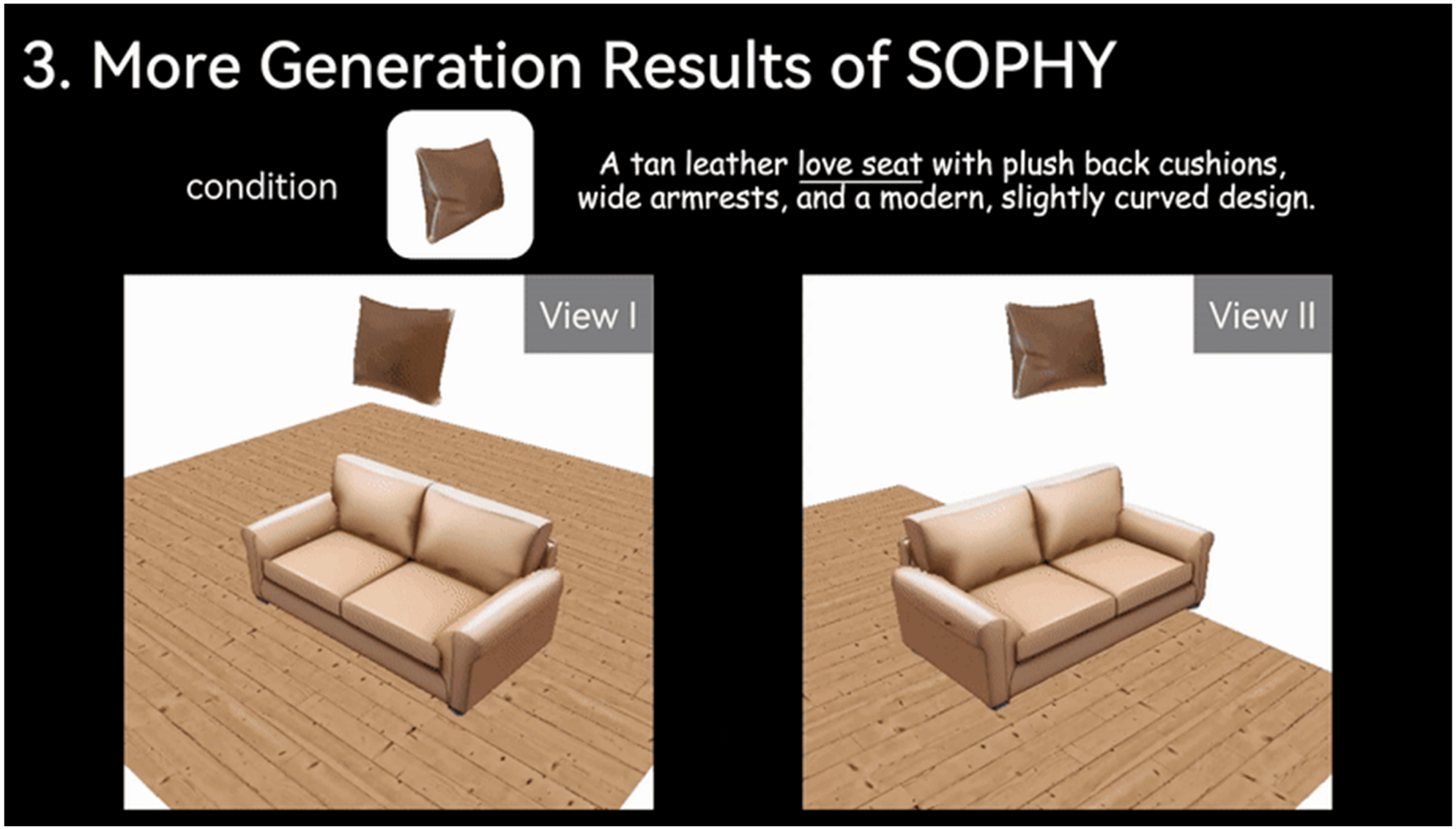
Text-to-4D Experiments



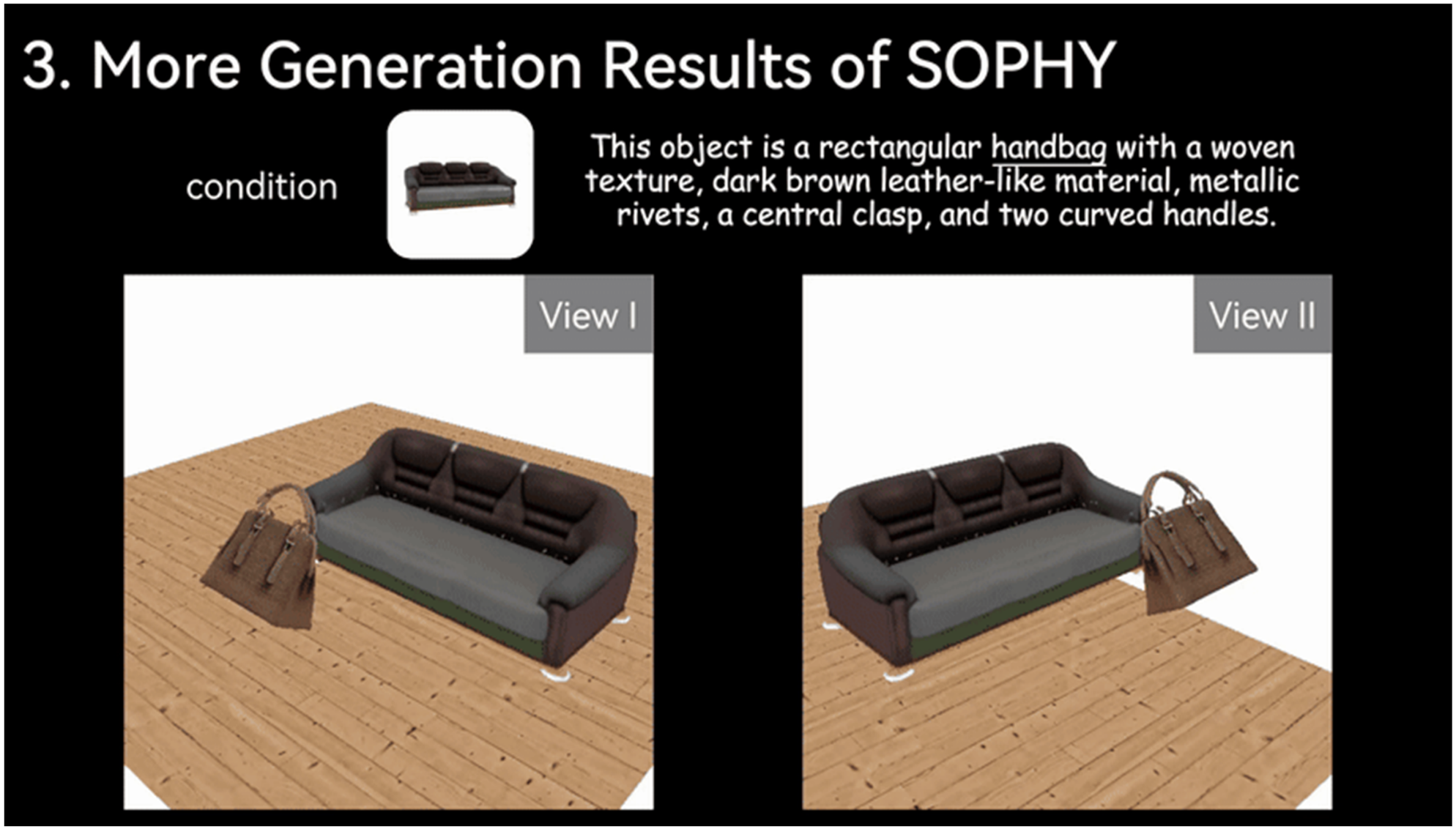
More Results



More Results



More Results



Outline

1. GEOPARD: Geometric Pretraining for Articulation Prediction in 3D Shapes

2. SOPHY: Generating Simulation-Ready Objects with Physical Materials

3. Future directions

Many more things to be done!

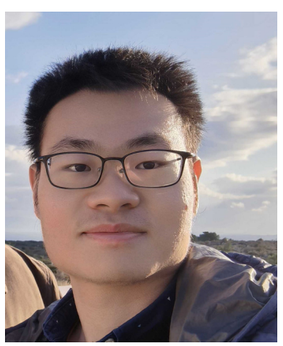
- Generative models of object geometry+articulation hierarchy+motions
- Larger-scale datasets with material annotations
- Better material verification strategies
- 4D scene generation with multiple objects+agents interacting

[models from turbosquid]

Thank you!

Student Team:

Dmitry Petrov
PhD, UMass,
(now @ Foundation)



Junyi Cao
PhD
UMass & TU Crete

Vikas Thamizharasan
PhD
UMass

Pradyumn Goyal
PhD
UMass

Tuan Ngo
PhD
UMass

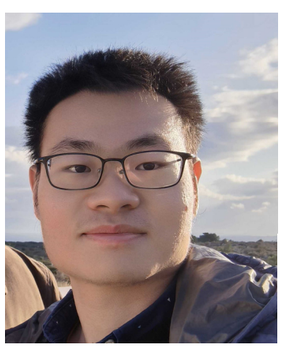
Nikitas Chatzis
MS
TU Crete & NTUA

Marios Loizou
Postdoc
TU Crete & CYENS

Thank you!

Student Team:

Dmitry Petrov
PhD, UMass,
(now @ Foundation)



Junyi Cao
PhD
UMass & TU Crete

Vikas Thamizharasan
PhD
UMass

Pradyumn Goyal
PhD
UMass

Tuan Ngo
PhD
UMass

Nikitas Chatzis
MS
TU Crete & NTUA

Marios Loizou
Postdoc
TU Crete & CYENS

Funding:

Project pages & code:

https://kalo-ai.github.io/