MagicArticulate: Make Your 3D Models Articulation-Ready

Chaoyue Song, Jianfeng Zhang, Xiu Li, Fan Yang, Yiwen Chen, Zhongcong Xu,
Jun Hao Liew, Xiaoyang Guo, Fayao Liu, Jiashi Feng, Guosheng Lin

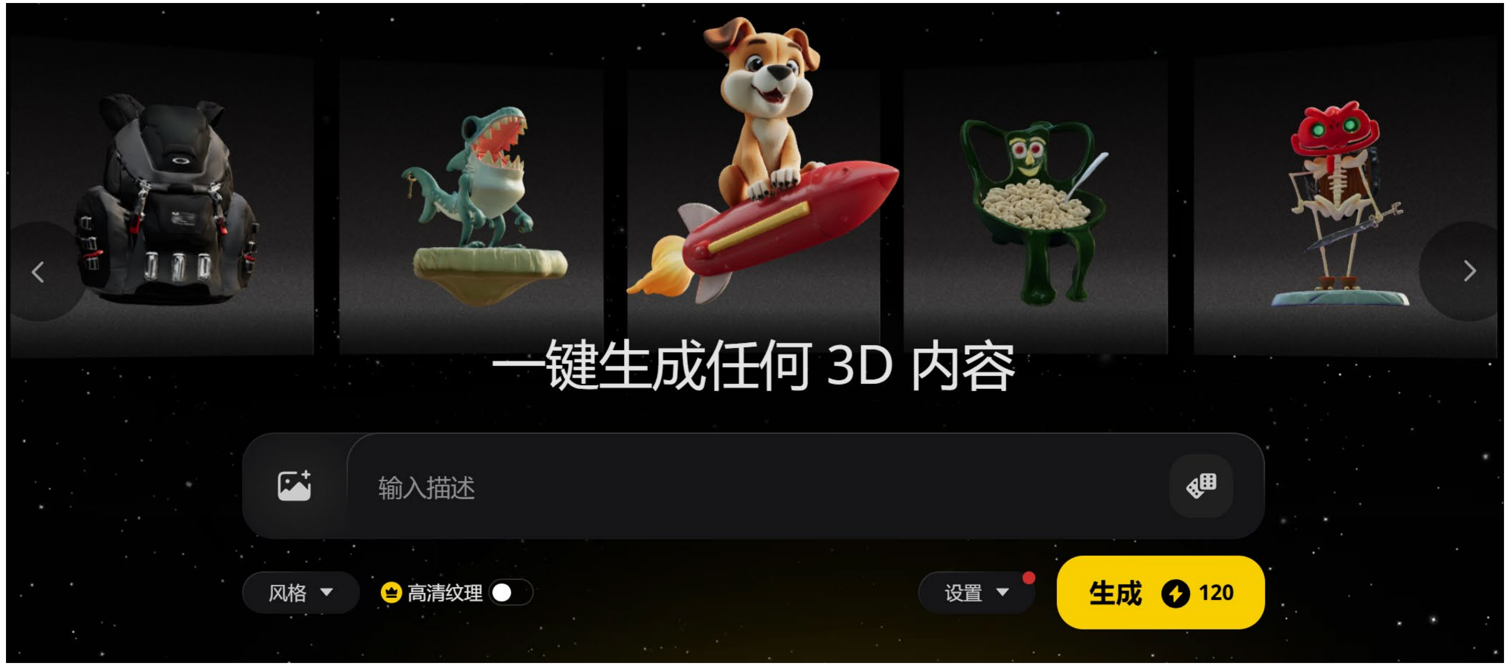
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

ByteDance | Seed

Agency for Science, Technology and Research SINGAPORE

Why “Articulation-Ready”?

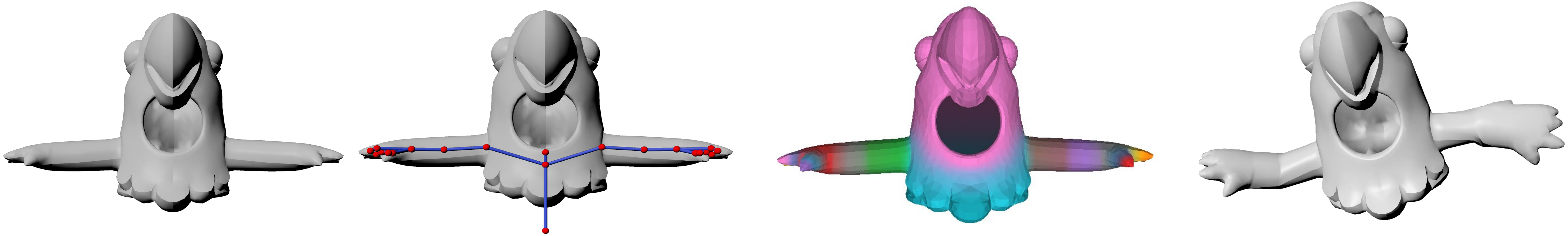
Clay



Tripo

Impressive geometry, texture, but… **Static**

What is “Articulation-Ready” (Rigging)?



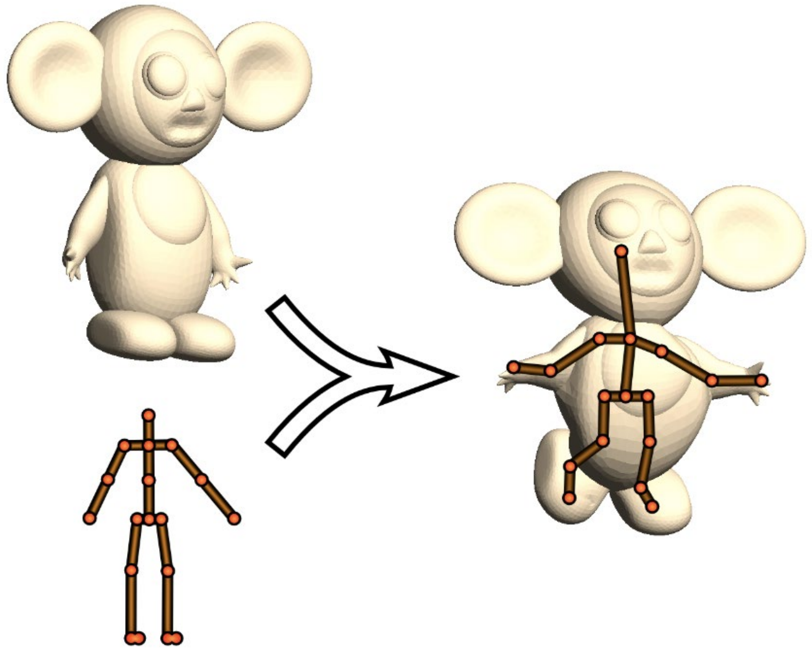
Input mesh | Skeleton | Skinning weights | Animation

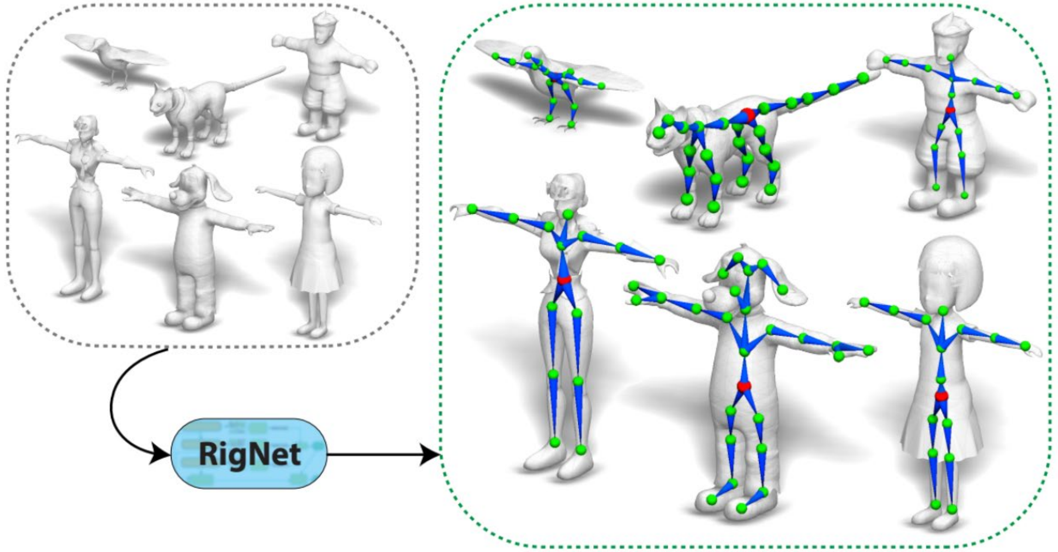
LBS: $\mathbf{v}' = (\sum_{i=1}^{n} w_i T_i)\mathbf{v}$

Previous solutions

Manual rigging:

Manual rigging is time-consuming and requires significant expertise.



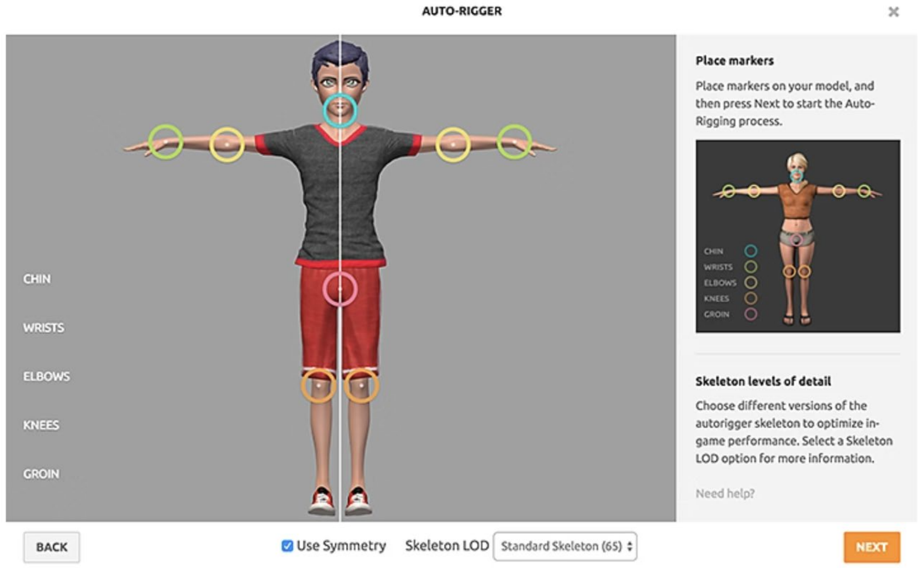


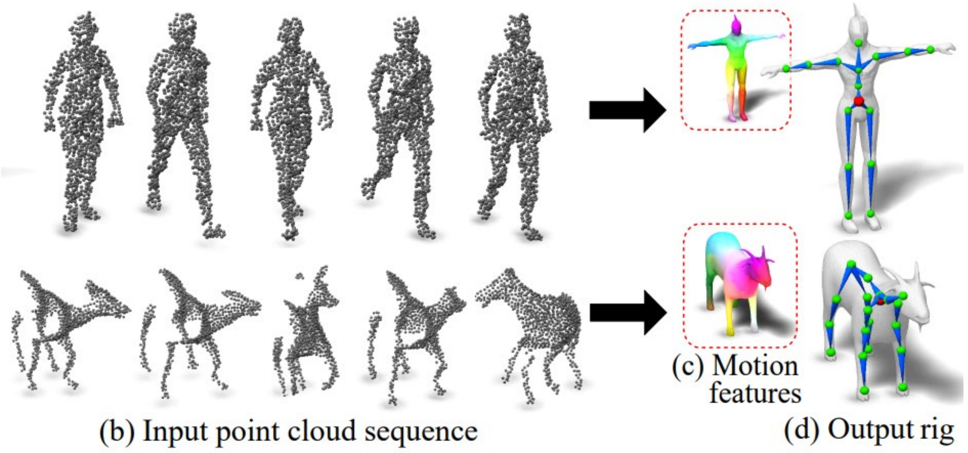
Automatic rigging:

1. Template-based

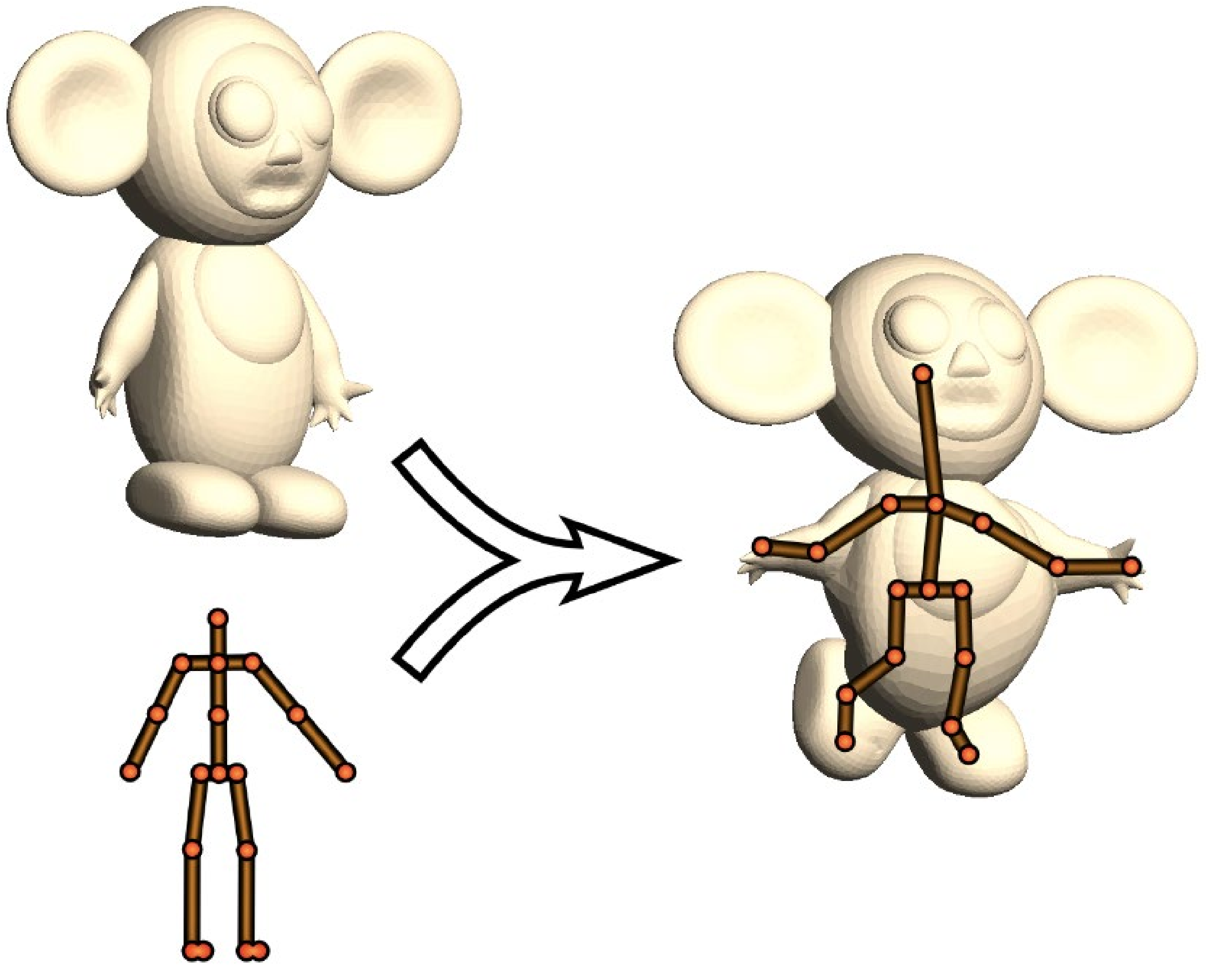
2. Template-free

3. Rely on additional inputs





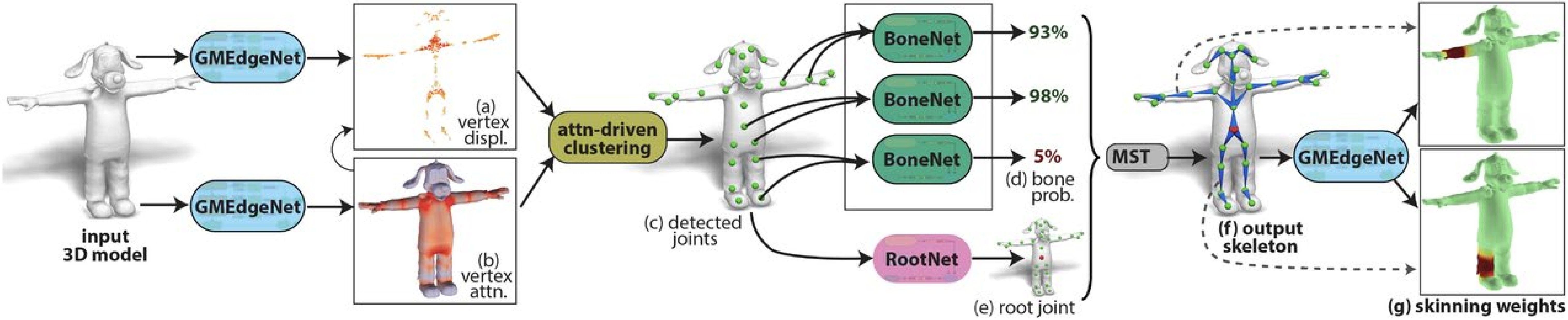
Previous solutions: template-based



- Rely on predefined templates.
- Fit a predefined skeleton template to the 3D model with the least fitting cost.
- Difficult to generalize to diverse categories.

Automatic Rigging and Animation of 3D Characters (Pinocchio), Baran et al., Siggraph 2007

Previous solutions: template-free

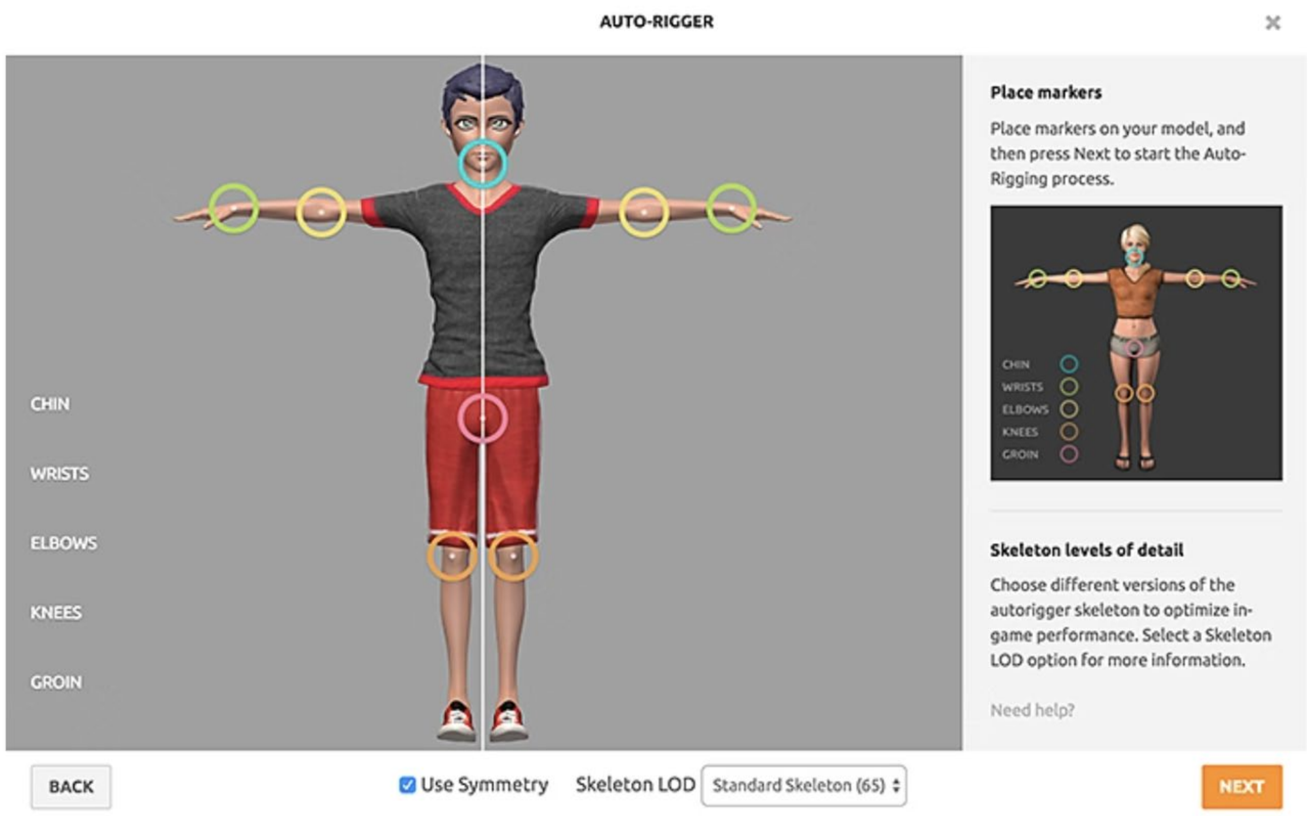


- Strong assumption that input shapes maintain a consistent upright and front-facing orientation.

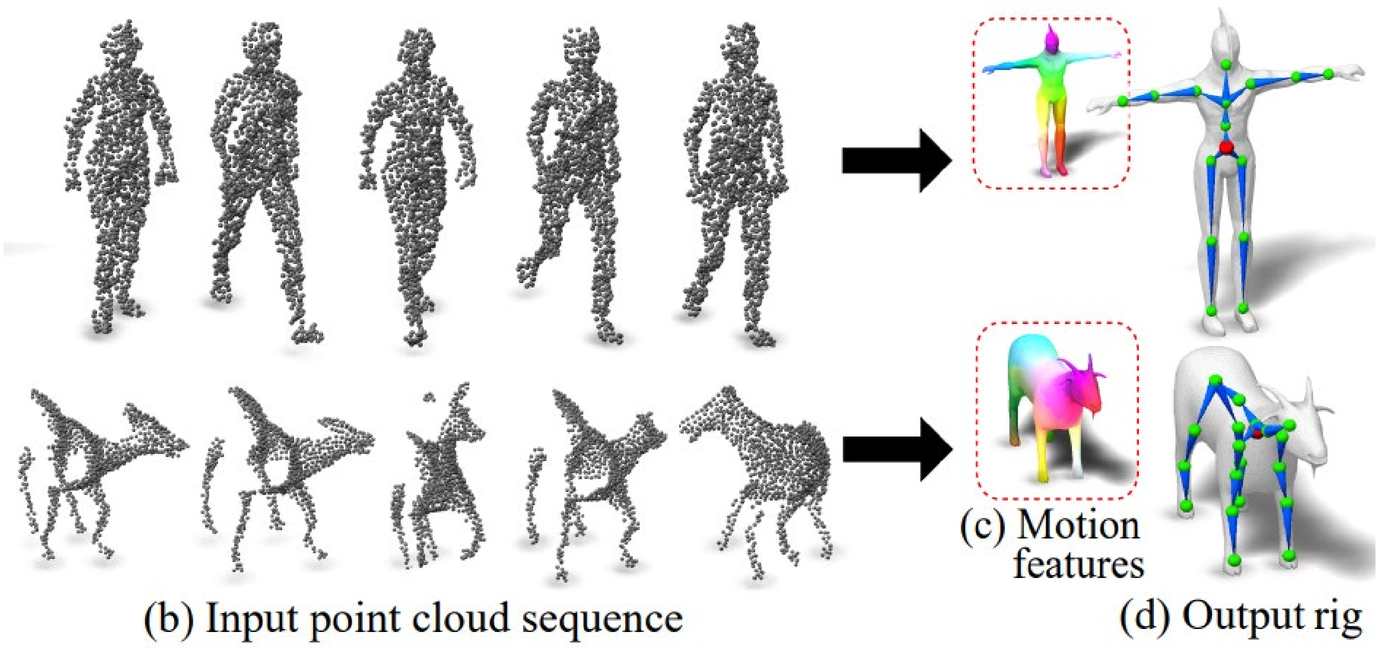
- Difficult to scale up.

RigNet: Neural Rigging for Articulated Characters, Xu et al., Siggraph 2020

Previous solutions: additional inputs



Mixamo: Rely on manual annotations



Require mesh or point cloud sequences

MoRig: Motion-aware rigging of character meshes from point clouds, Xu et al., Siggraph Asia 2022

Previous solutions: Summary

- the lack of a **large-scale**, **diverse** dataset for training generalizable models.

- the need for an effective framework capable of handling **complex mesh topologies**, accommodating **varying skeleton structures.**

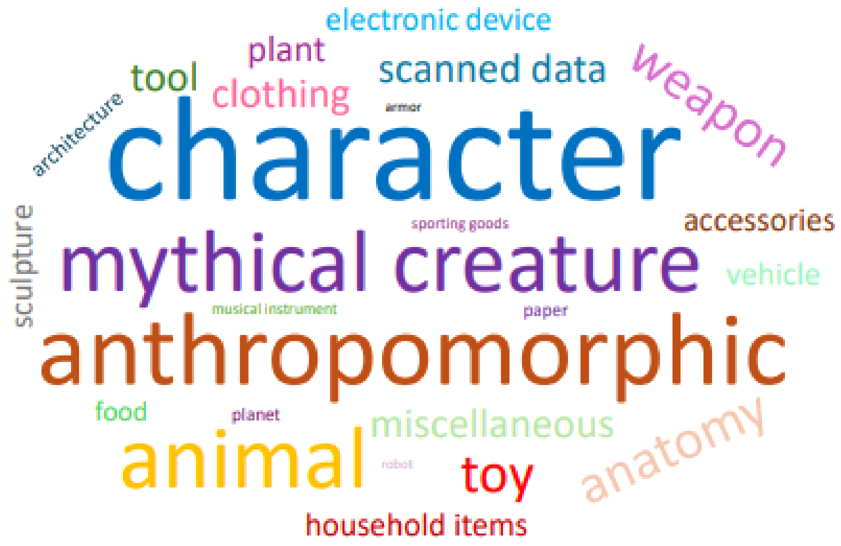
Our solution: MagicArticulate

- Introduce **Articulation-XL**, a large-scale dataset containing over 33k 3D models with high-quality articulation annotations.

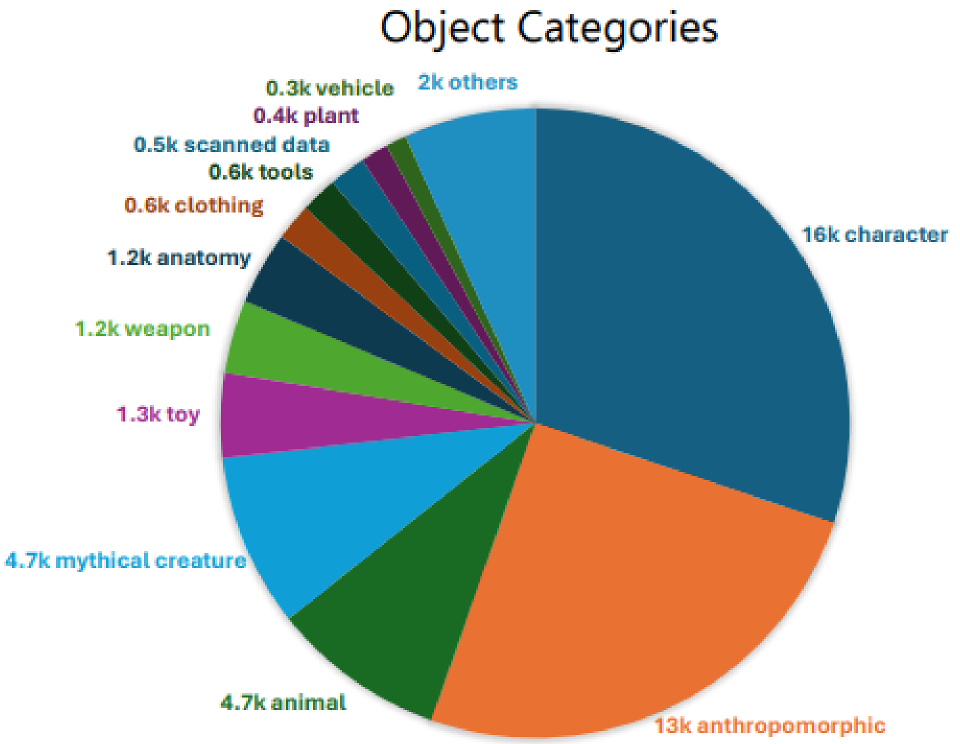
- Formulate skeleton generation as a **sequence modeling problem**.

- Predict skinning weights using a **functional diffusion process**.

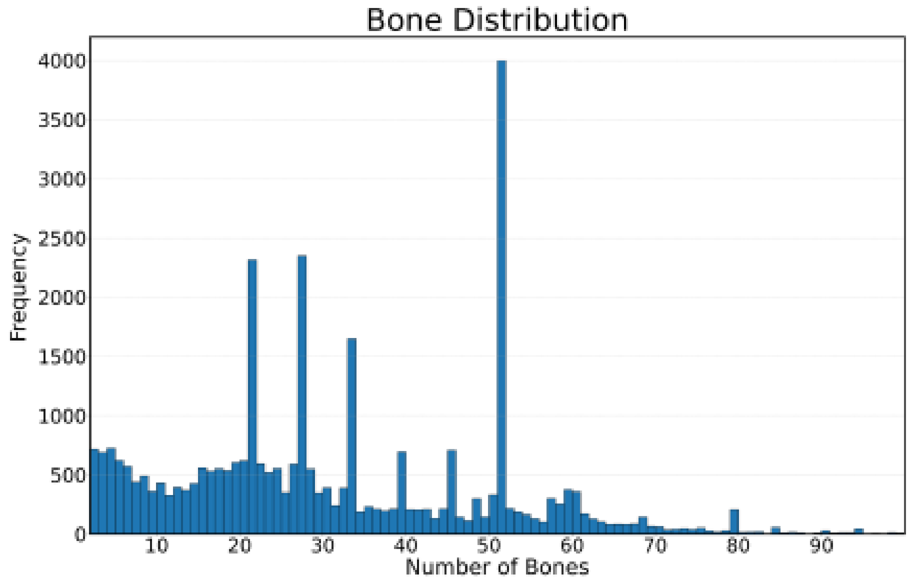
Dataset: Articulation-XL



(a) **Word cloud of Articulation-XL categories.**



(b) **Breakdown of Articulation-XL categories.**



(c) **Bone number distributions of Articulation-XL.**

Articulation-XL2.0 with over 48K data has been open sourced.

Dataset: Articulation-XL

1. Initial data collection.

2. VLM-based filtering.

3. Category label annotation.

Table 1. **Data statistics.**

Source	All 3D data	with rigging	high quality rigging	low quality rigging
GitHub	2.08M	64K	42K	22K
Objaverse1.0	0.89M	10K	6K	4K
Sum	2.97M	74K	48K	26K

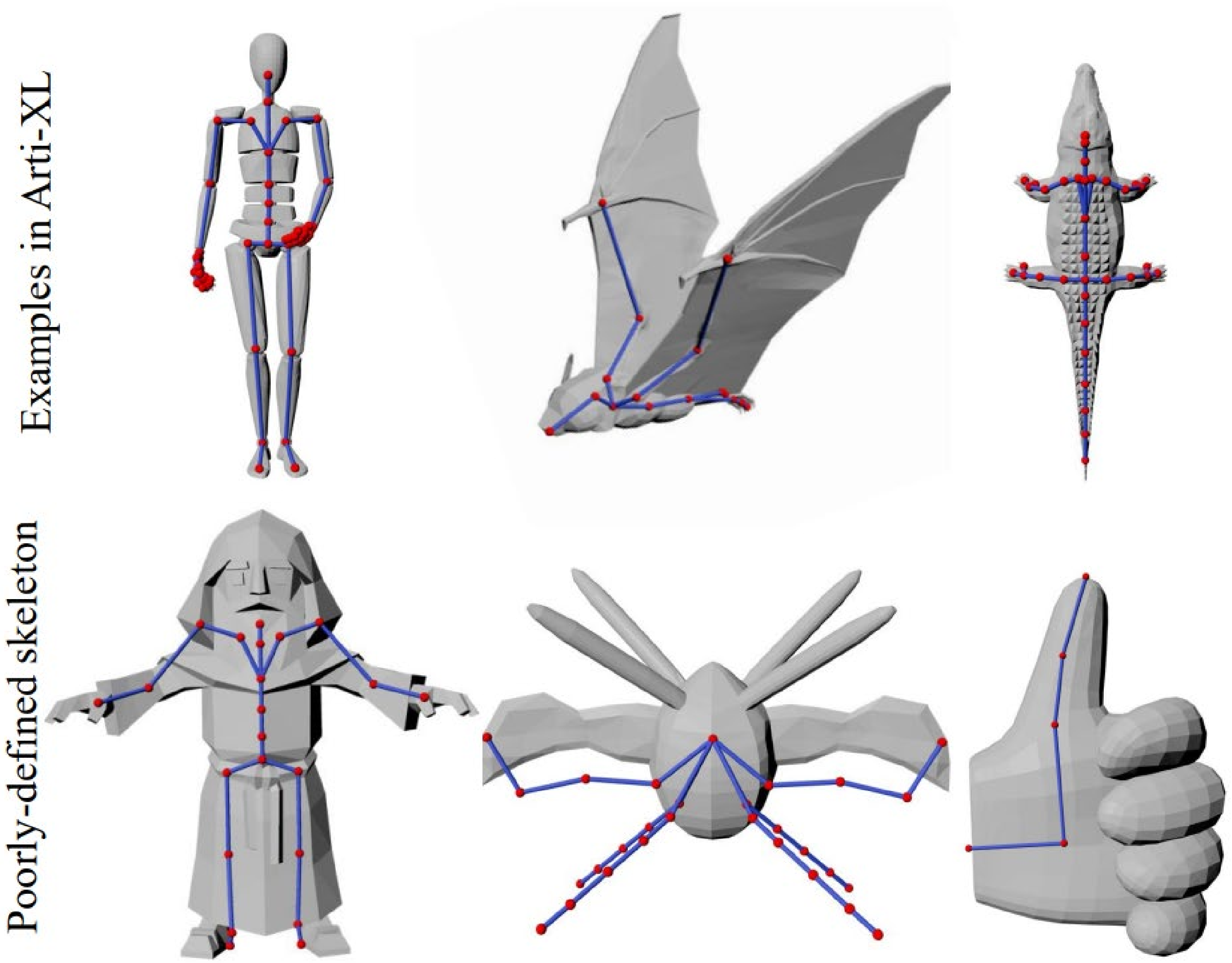
Articulation-XL2.0, the data with rigging has been deduplicated (over 150K).

Dataset: Articulation-XL

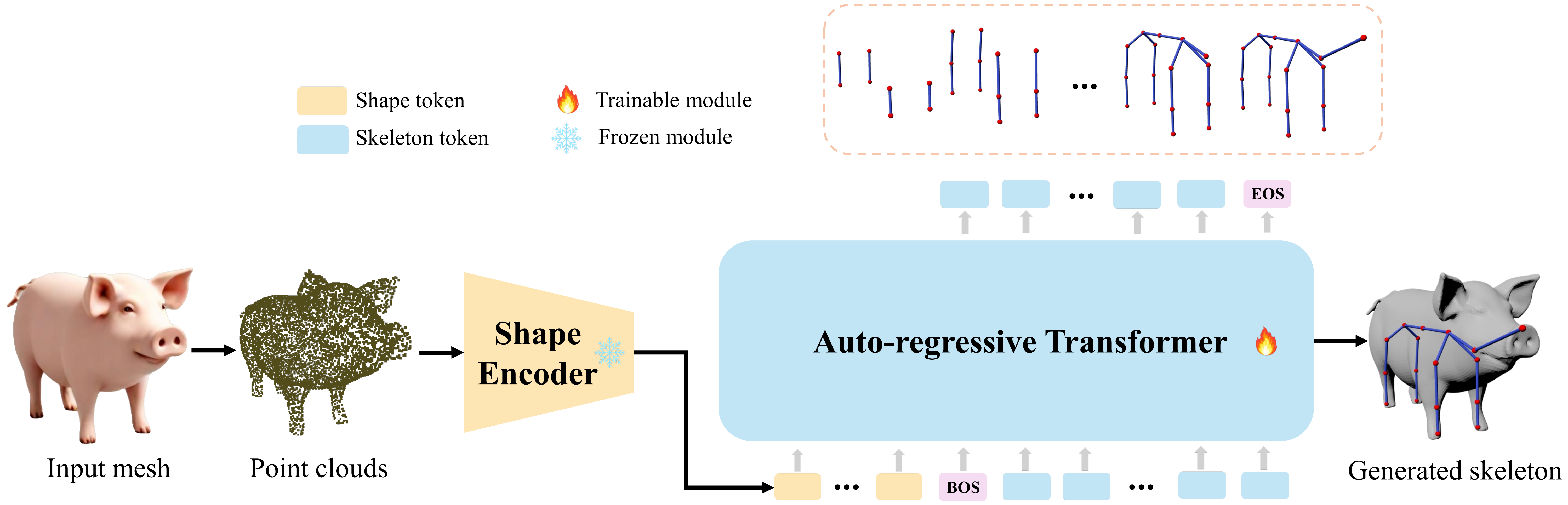
1. Initial data collection.

2. VLM-based filtering.

3. Category label annotation.



Auto-regressive skeleton generation



Skeleton tokenization: sequence of bones

$$p(\mathcal{S}|\mathcal{M}) = p(\mathbf{J}, \mathbf{B}|\mathcal{M})$$

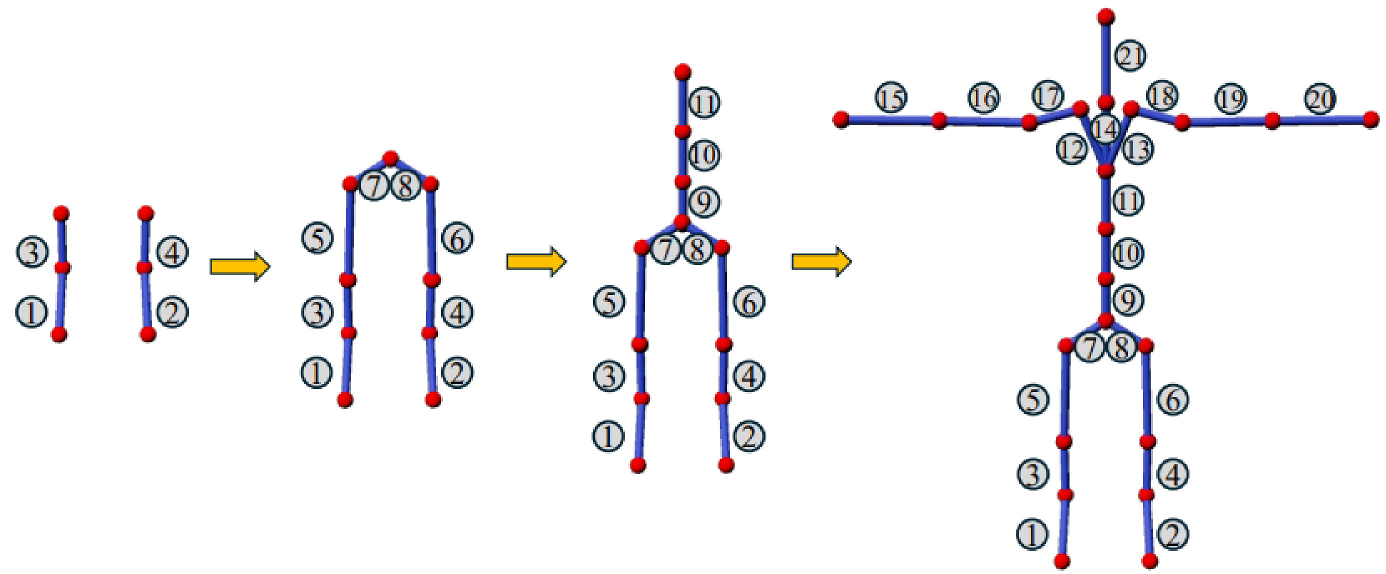
B1 = (x1, y1, z1, x2, y2, z2)

B2 = (x2, y2, z2, x3, y3, z3)

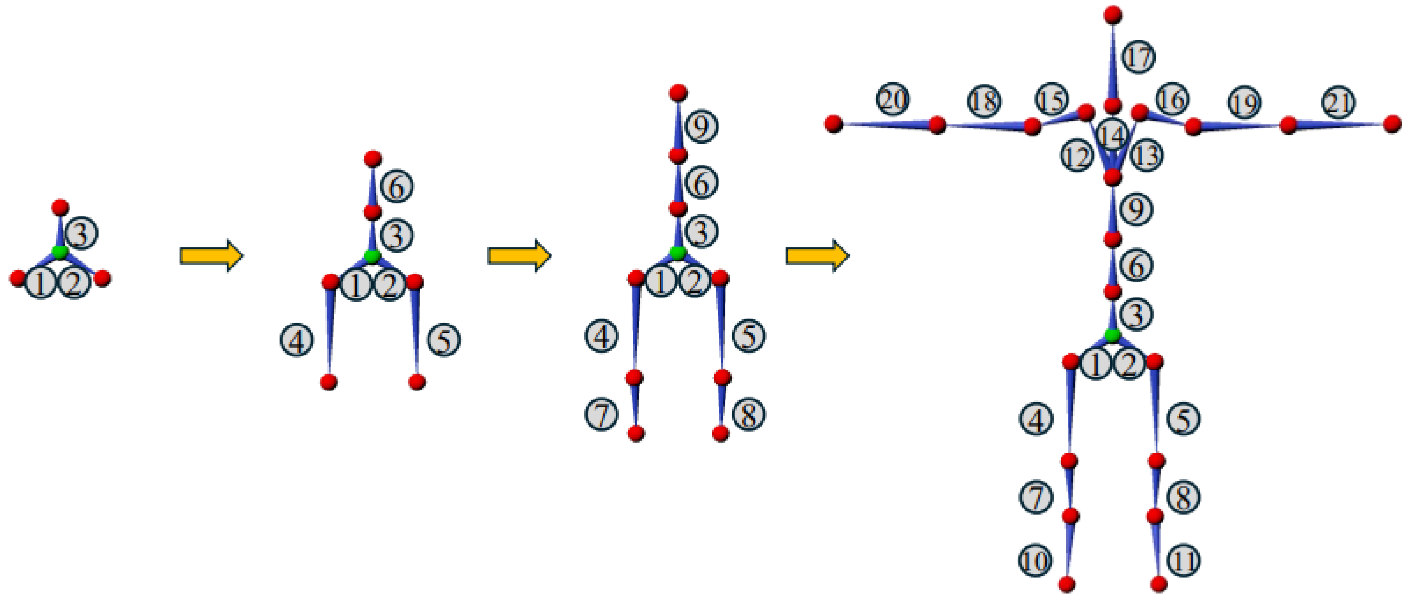
normalization --> discretization --> 6b sequence

How to sort this sequence?

Sequence ordering



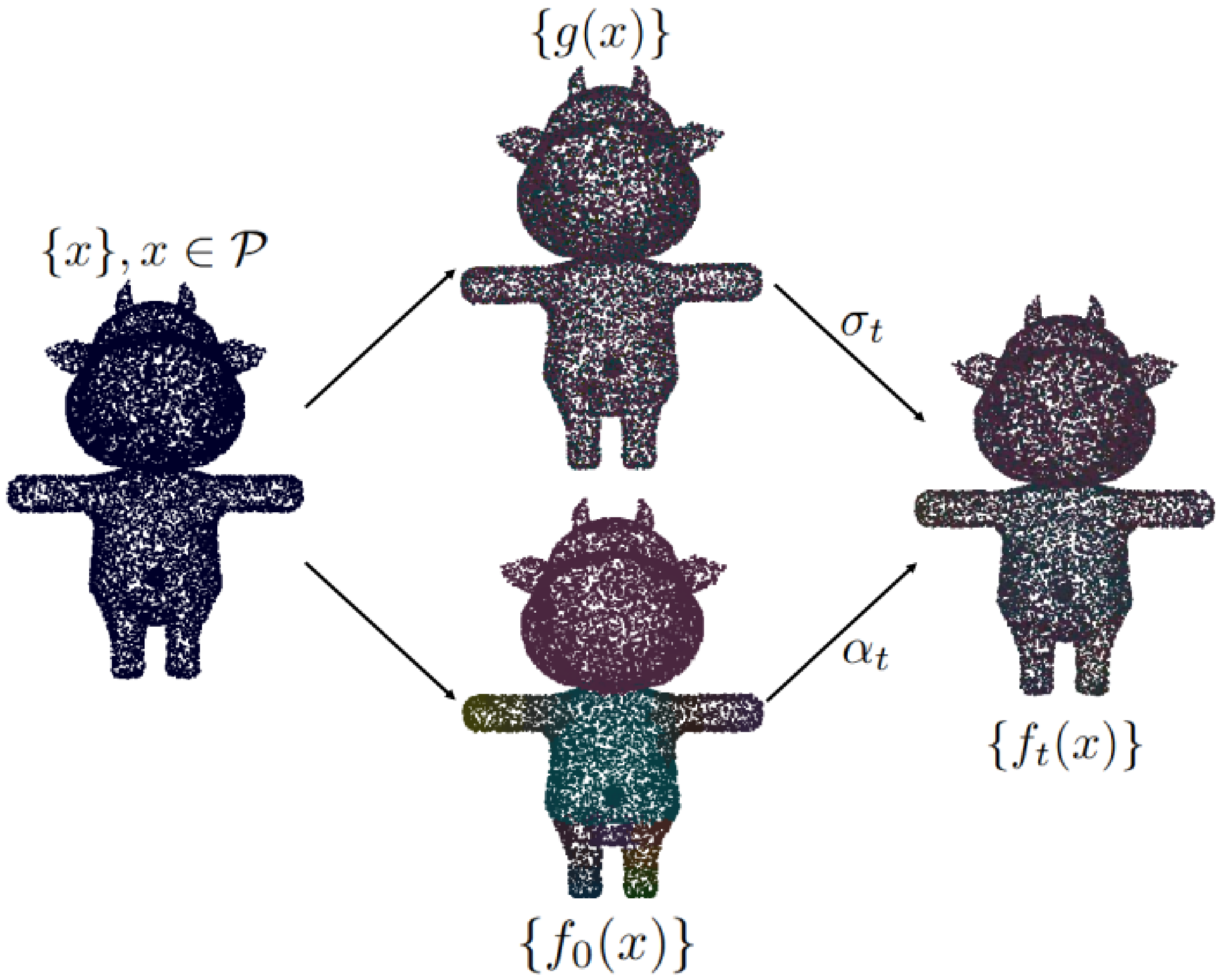
Spatial sequence ordering



Hierarchical sequence ordering

$$\mathcal{L}_{pred} = \mathrm{CE}(\mathbf{T}, \hat{\mathbf{T}})$$

Skinning weight prediction: functional diffusion

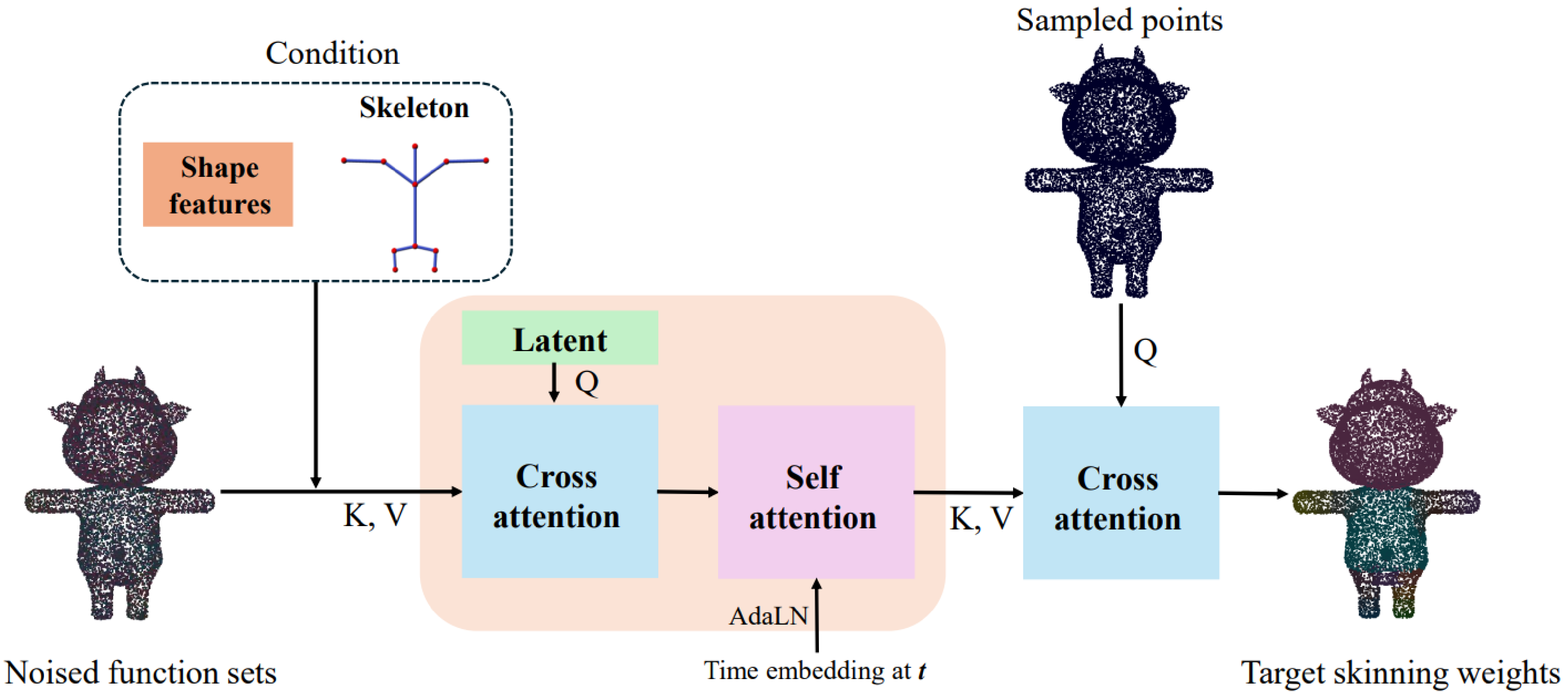


$$f_0 : \mathcal{X} \to \mathcal{Y}.$$

$$f_t(x) = \alpha_t \cdot f_0(x) + \sigma_t \cdot g(x), \quad t \in [0, 1]$$

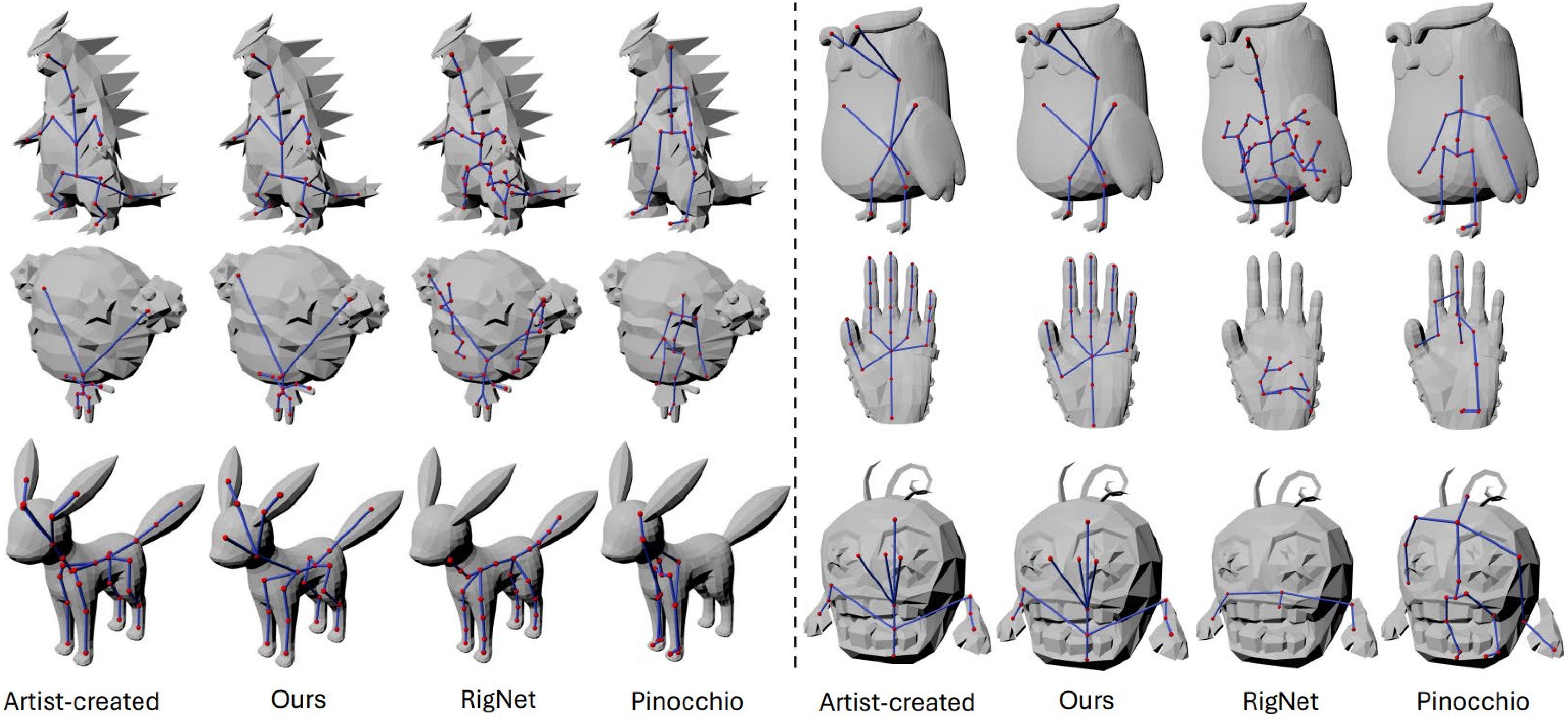
$$D_\theta[f_t, t](x) \approx f_0(x).$$

Skinning weight prediction

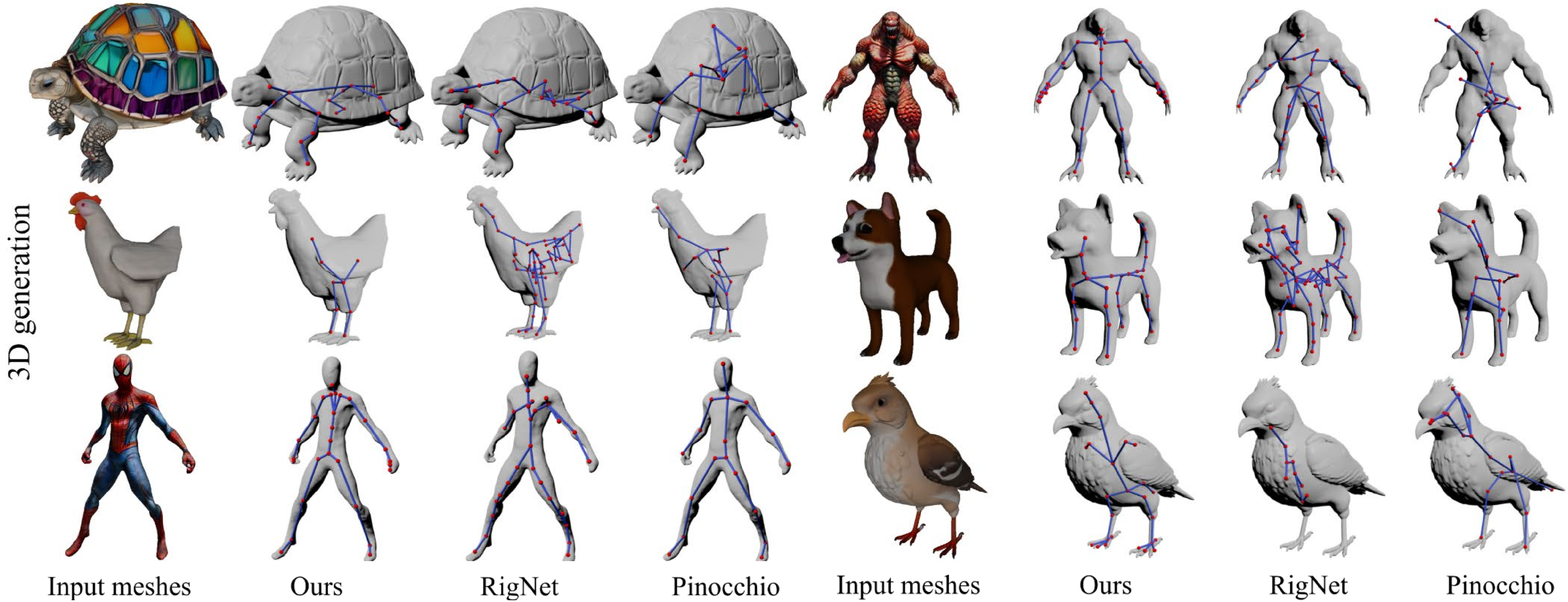


$$f : \mathcal{P} \rightarrow (\mathcal{W} - \mathcal{G}) \qquad \mathcal{L}_{denoise} = \left\| D_\theta \left(\{x, f_t(x)\}, t \right) - f_0(x) \right\|_2^2, \quad x \in \mathcal{P}.$$

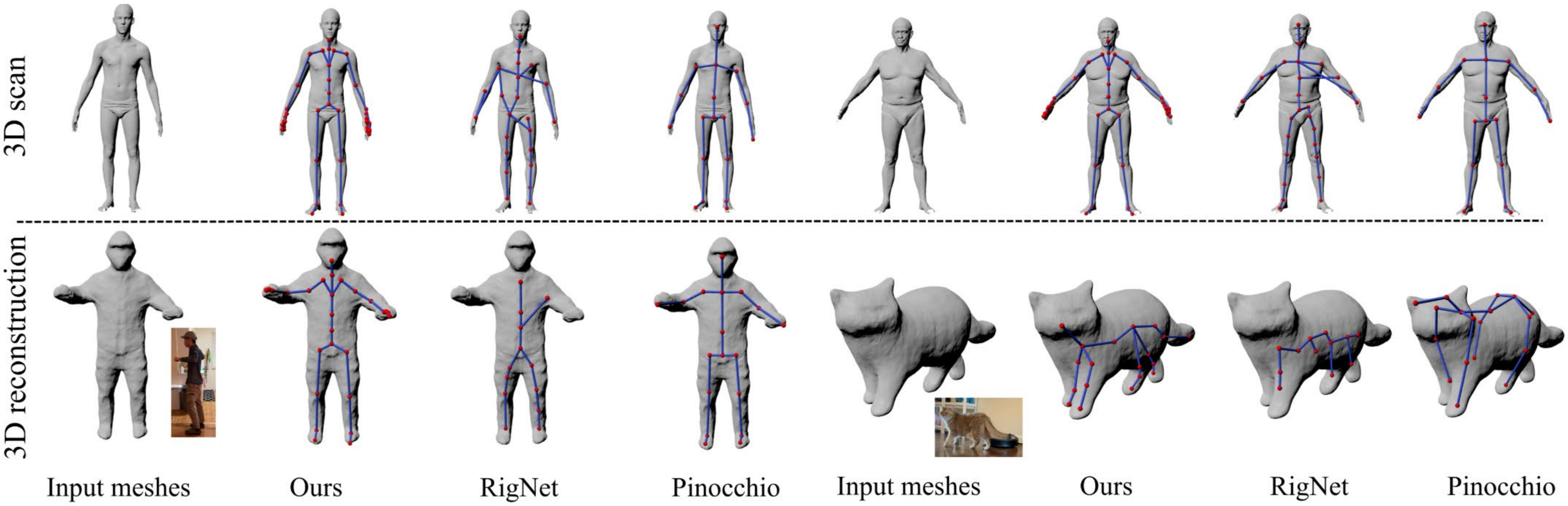
Skeleton generation results



Skeleton generation results: generalization



Skeleton generation results: generalization



Skeleton generation results

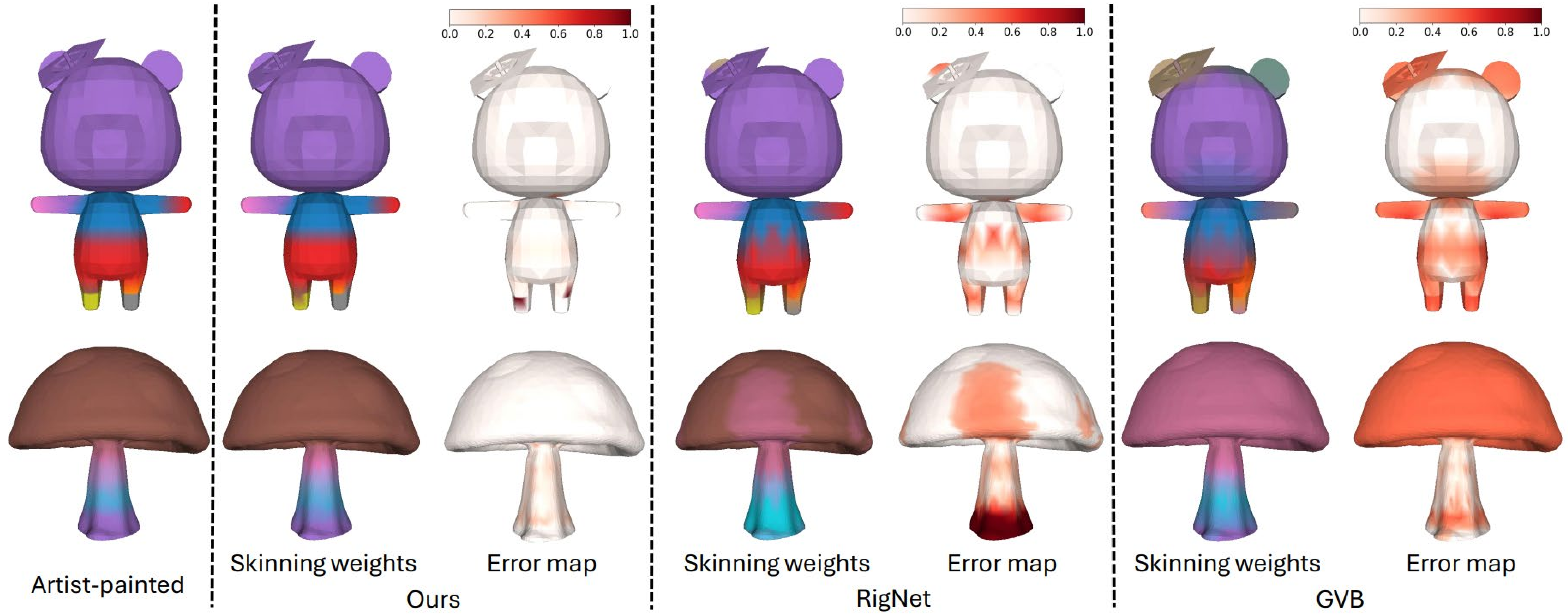
These Chamfer Distance-based metrics measure the spatial alignment between generated and ground truth skeletons. Lower is better.

	Dataset	CD-J2J	CD-J2B	CD-B2B
RigNet*	*ModelsRes.*	7.132	5.486	4.640
Pinocchio		6.852	4.824	4.089
Ours-hier*		4.451	3.454	2.998
RigNet		4.143	2.961	2.675
Ours-spatial*		4.103	3.101	2.672
Ours-hier		3.654	2.775	2.412
Ours-spatial		**3.343**	**2.455**	**2.140**
Pinocchio	*Arti-XL*	8.360	6.677	5.689
RigNet		7.478	5.892	4.932
Ours-hier		3.025	2.408	2.083
Ours-spatial		**2.586**	**1.959**	**1.661**

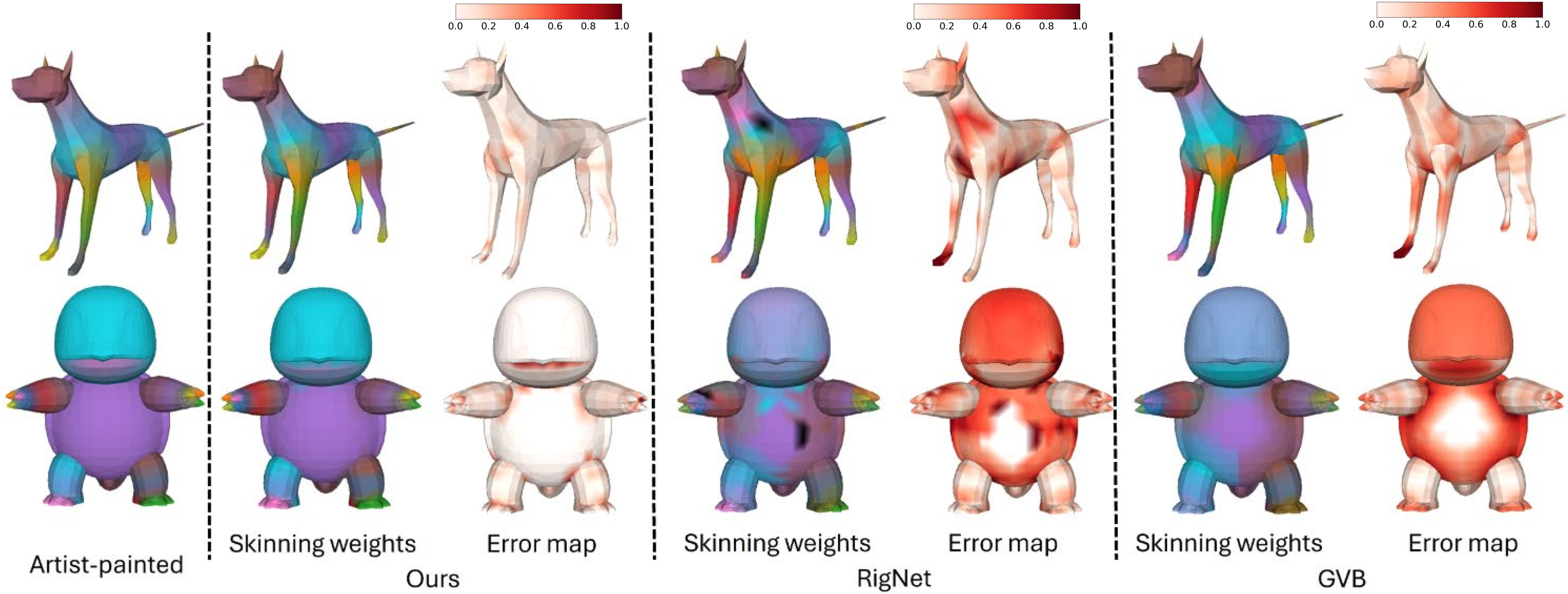
Skeleton generation results: ablation

	CD-J2J	CD-J2B	CD-B2B
w/o data filtering	2.982	2.327	2.015
4,096 points	2.635	2.024	1.727
12,288 points	2.685	2.048	1.760
Ours (8,192)	**2.586**	**1.959**	**1.661**

Skinning weight prediction results



Skinning weight prediction results



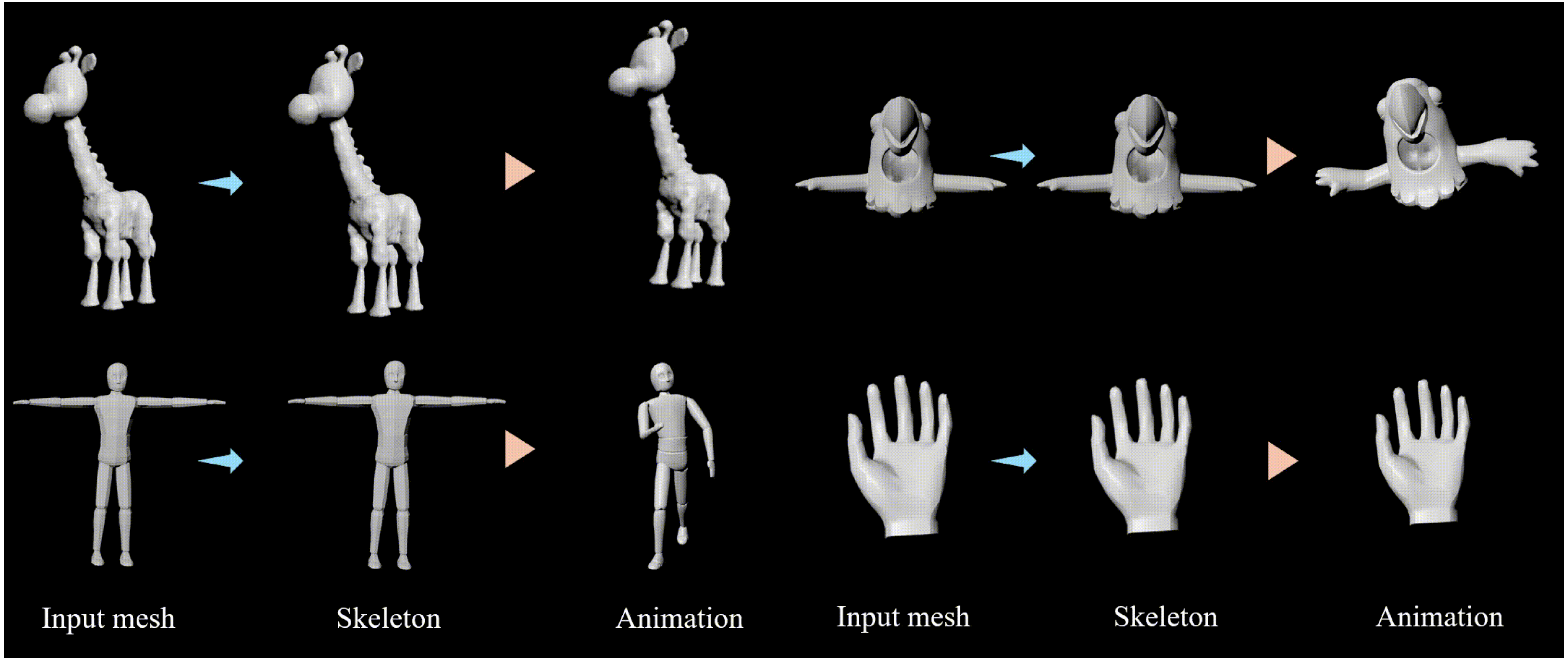
Skinning weight prediction results

	Dataset	Precision	Recall	avg L1	avg Dist.
GVB	*ModelsResource*	69.3%	79.2%	0.687	0.0067
RigNet		77.1%	**83.5%**	0.464	0.0054
Ours		**82.1%**	81.6%	**0.398**	**0.0039**
GVB	*Articulation-XL*	75.7%	68.3%	0.724	0.0095
RigNet		72.4%	71.1%	0.698	0.0091
Ours		**80.7%**	**77.2%**	**0.337**	**0.0050**

Skinning weight prediction results

	Precision	Recall	avg L1	avg Dist.
w/o geodesic dist.	81.5%	77.7%	0.444	0.0046
w/o weights norm	82.0%	77.9%	0.436	0.0045
w/o shape features	81.4%	81.3%	0.412	0.0042
Ours	**82.1%**	**81.6%**	**0.398**	**0.0039**

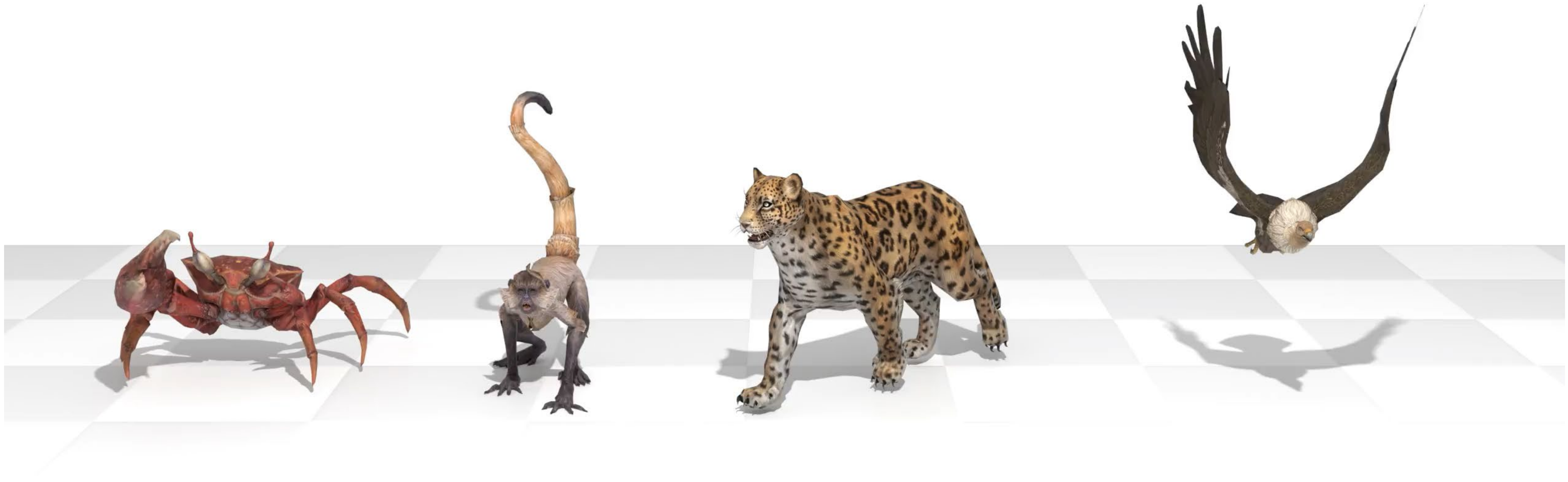
Animation results



Future directions 1: MagicArticulate V2

1. Faster inference.
    - Skeleton generation: 2s --> 1s
    - Skinning: 1.4s --> 0.06s
2. Better generalization.

Future directions 2: 3D animation



AnyTop: Character Animation Diffusion with Any Topology, Gat et al., arXiv 2025

Future directions 3: Rigid articulated objects

1. Part-rigidity

2. Articulation mode: 1 dof, 2dof, 3dof v.s. 6 dof

3. Save format: urdf v.s. glb/fbx/blend/dae…

Articulate AnyMesh: Open-Vocabulary 3D Articulated Objects Modeling, Qiu et al., arXiv 2025

Thanks!