Mesh-based Dynamics with Occlusion Reasoning for Cloth Manipulation

RSS 2022
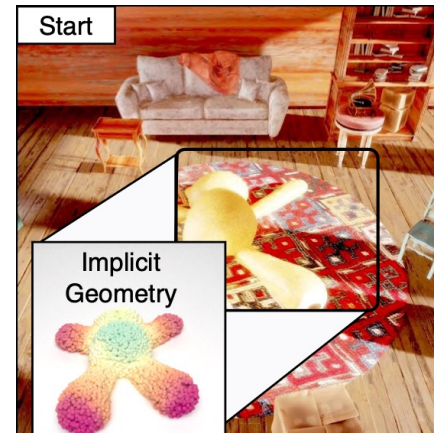
Pipeline
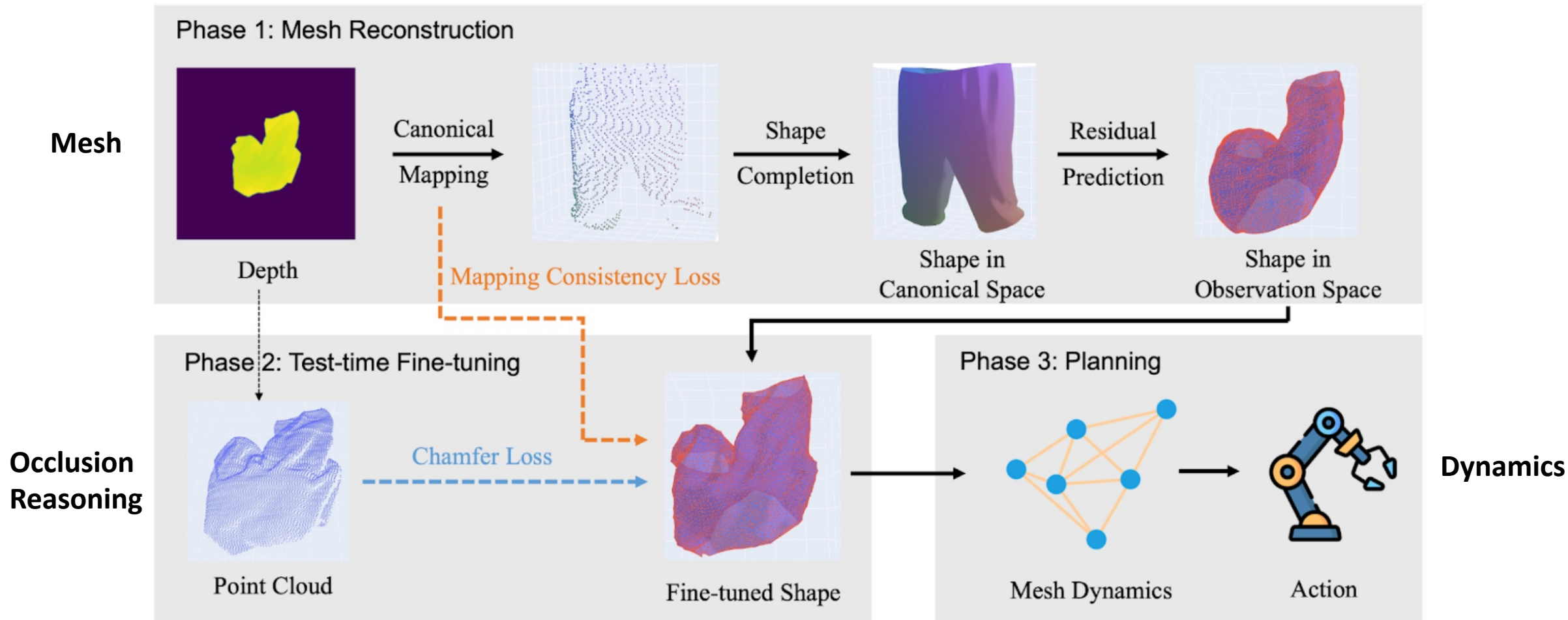
Model construction

■ Lift up the cloth

Input Point Cloud   Canonical Coordinate   Shape Completion   Pose in Task Space
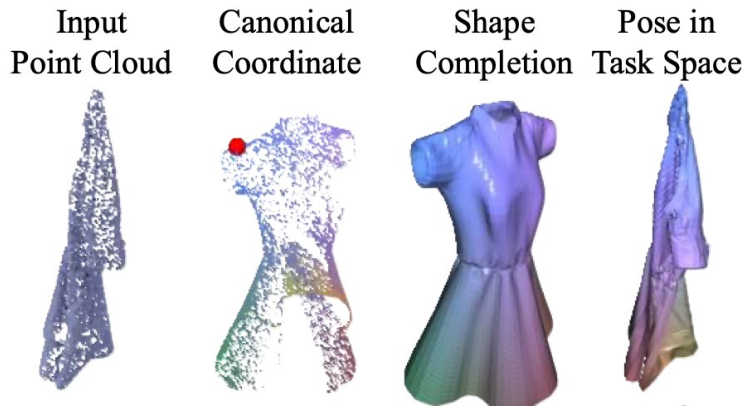
GarmentNets: Category-Level Pose Estimation
for Garments via Canonical Space Shape Completion
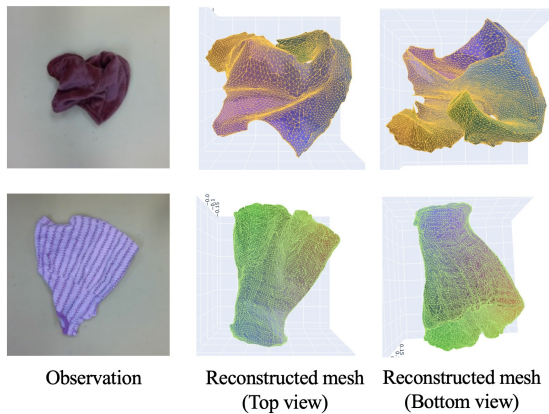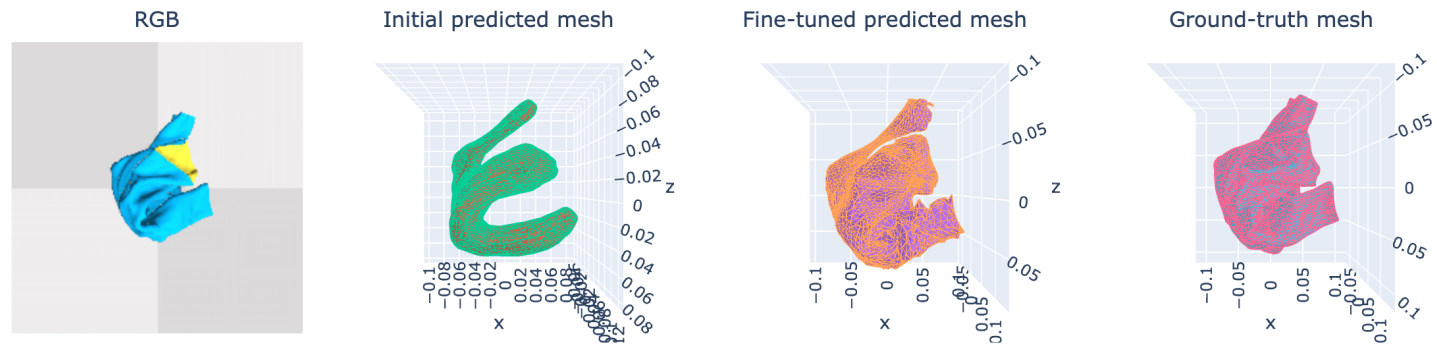**ICCV 2021**

■ Implicit geometry

ACID: Action-Conditional Implicit Visual
Dynamics for Deformable Object Manipulation
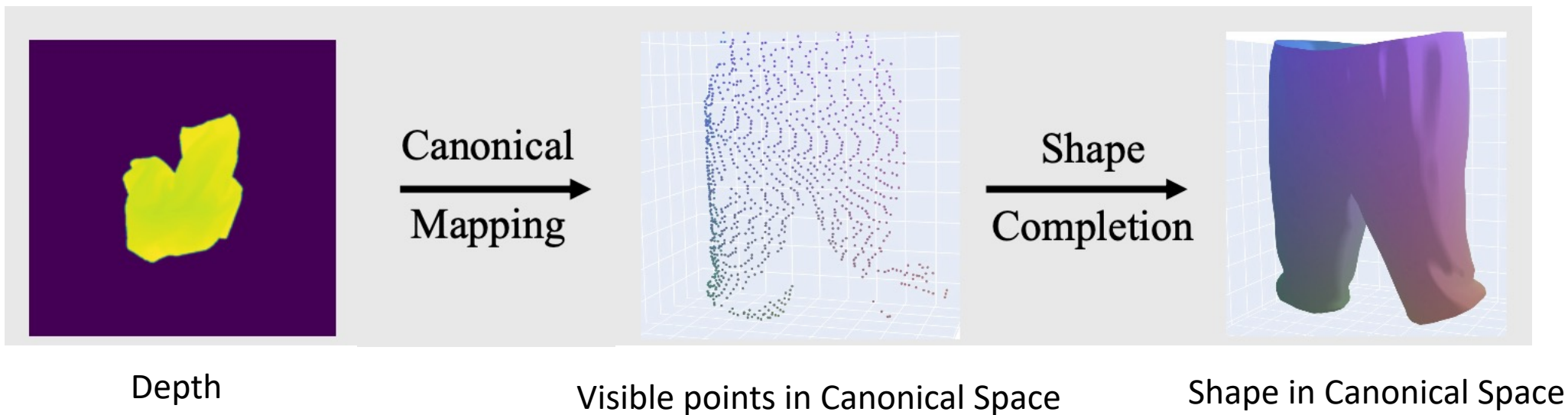**RSS 2022 Best Student Paper**

Model construction

■ Crumpled states

Given only a partial observation of a crumpled cloth, generate a complete mesh of the cloth.

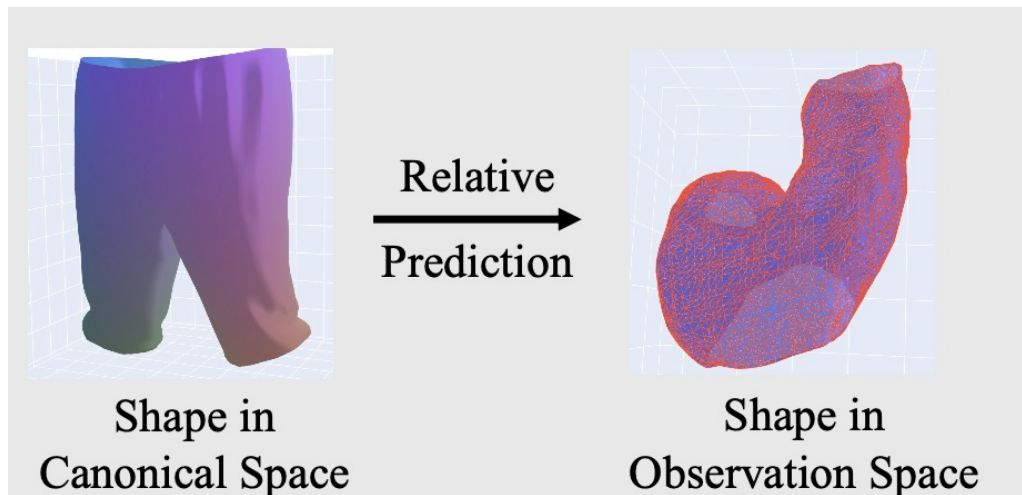# Model construction

■ Mapping to canonical state

Depth   Visible points in Canonical Space   Shape in Canonical Space

Trained in a category-dependent manner

Model construction

■ Relative predictions

predict a 3-dimensional residual vector f for each point x_i^c.

The predicted coordinates in the observation space x_i are obtained by $x_i^o = x_i^c + f$

Shape in Canonical Space → Relative Prediction → Shape in Observation Space

Trained in a category-dependent manner
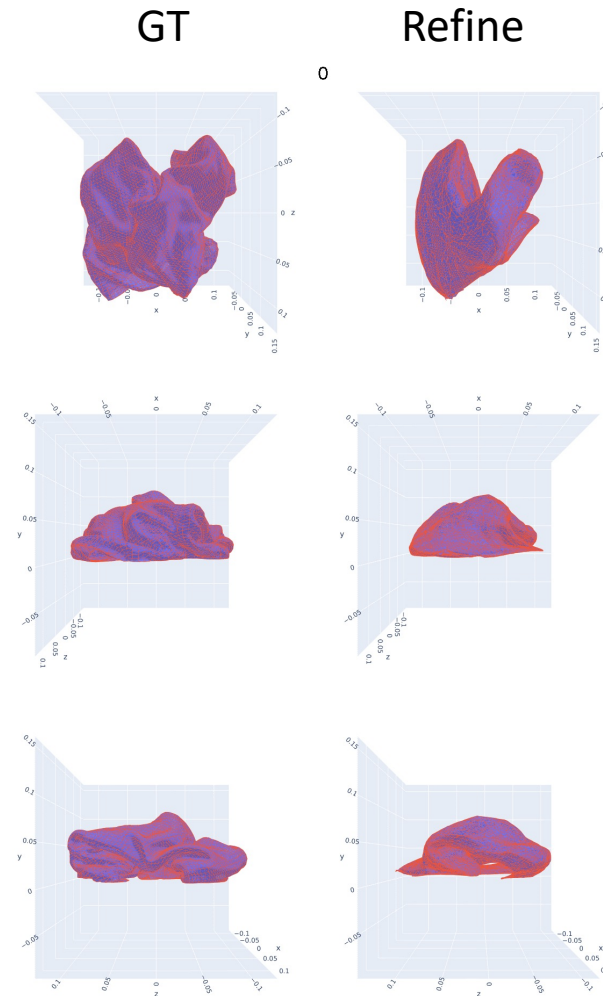
Model construction

- Test-time finetuning

Unidirectional Chamfer loss

$$\mathcal{L}_{\mathcal{C}}(\tilde{V}^t; P^t) = \frac{1}{|P^t|} \sum_{p_i \in P^t} \min_{\tilde{v}_j \in \tilde{V}^t} d(p_i, \tilde{v}_j)$$
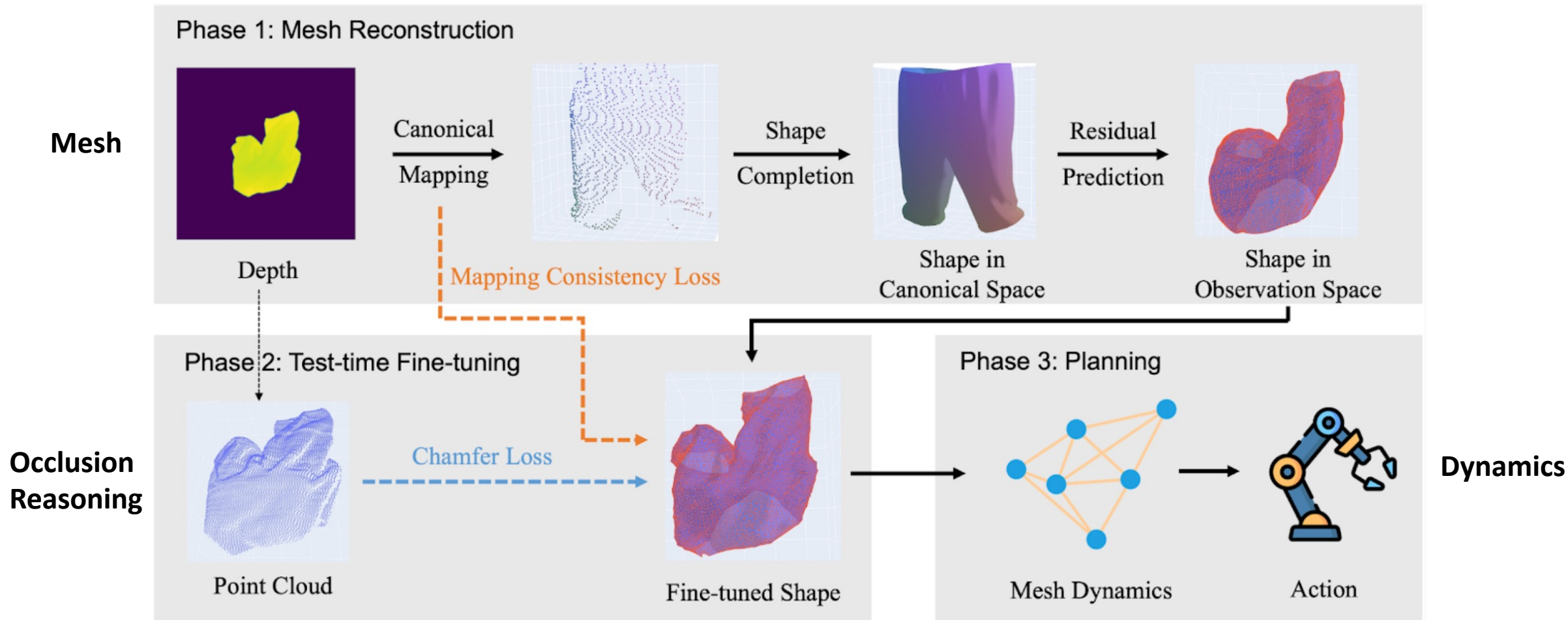
observation of the cloth is P^t,
predicted vertex in observation space is V^t.

Mapping consistency loss

$$\mathcal{L}_{\mathcal{M}}(P^t) = \frac{1}{|P^t|} \sum_{p_i \in P^t} d(g(f(p_i)), p_i)$$

GT   Refine

Pipeline

Mesh

Occlusion Reasoning

Dynamics

Mesh-based Dynamics

■   Dynamic model

Given a down-sampled mesh from reconstruction.

A GNN encodes the input feature on the nodes and on the edges and then conducts multiple message passing steps between the nodes and edges.
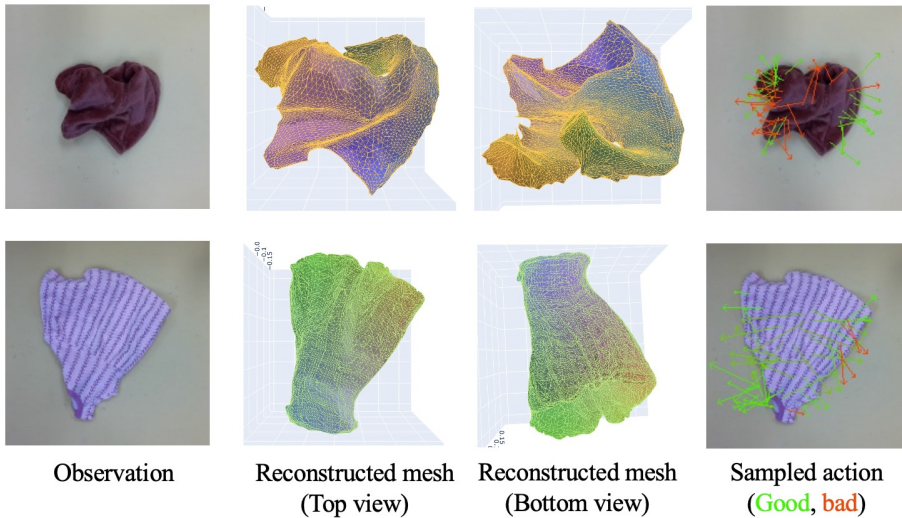
The decoder will decode the latent features of each node into the predicted acceleration for that node.

Trained only on Trousers
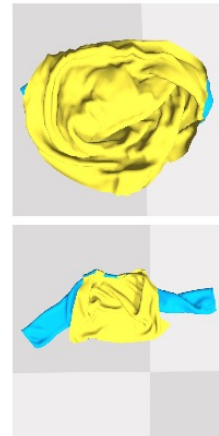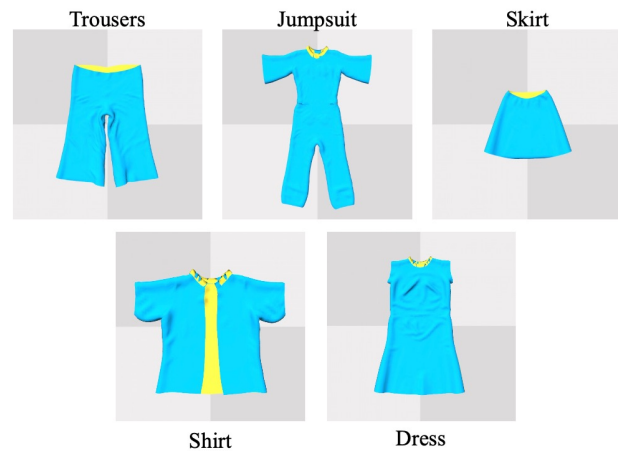
Mesh-based Dynamics

■ Planning

Randomly sample 500 pick-and-place actions and rollout each action with

the GNN dynamics model and pick the action with the highest predicted reward

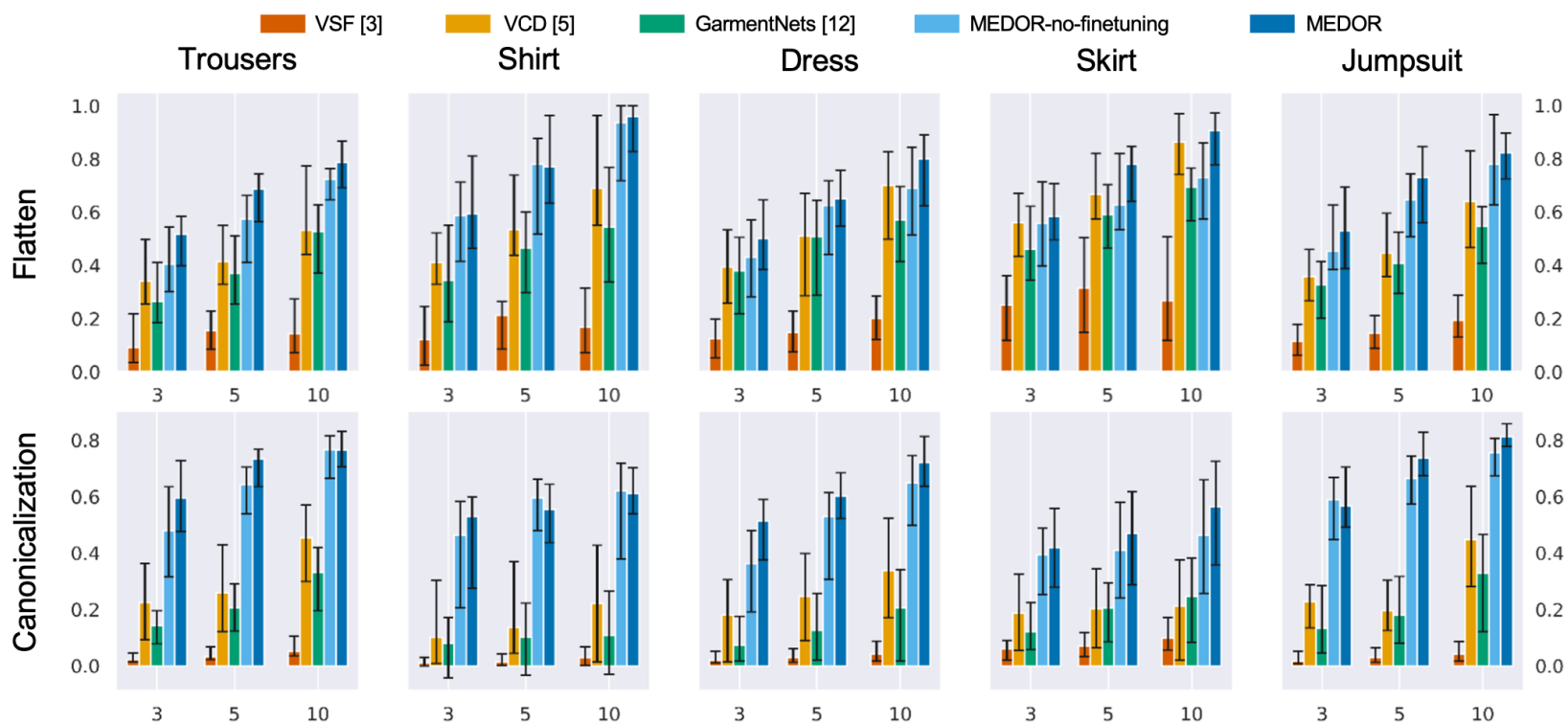| Observation | Reconstructed mesh (Top view) | Reconstructed mesh (Bottom view) | Sampled action (Good, bad) |

Data and tasks

■ Cloth3D in SoftGym

■ Flattening and Canonicalization

■ Real robot

Results

- With baseline

Results

■ Ablations

| Method	Normalized Improvement
GarmentNets [53]	0.320 ± 0.146
No Mesh Reconstruction (VCD [5])	0.391 ± 0.174
No Finetuning and no Relative Prediction	0.560 ± 0.163
No Finetuning	0.585 ± 0.171
Joint Optimization	0.614 ± 0.157
No Consistency Loss	0.623 ± 0.148
Replace GNN by GT Dynamics	0.631 ± 0.161
Ours w/ Partial Reward	0.462 ± 0.210
Ours (full method)	$\mathbf{0.651 \pm 0.138}$
GT Mesh + Learned Dynamics	0.800 ± 0.096
GT Mesh + GT Dynamics	0.870 ± 0.076

TABLE I: Ablation experiments.

Results

■ Real robot
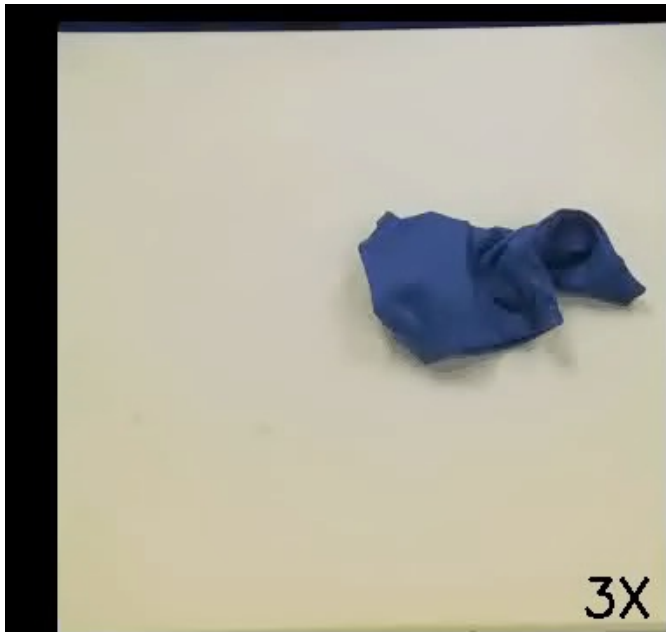
Execution 3X
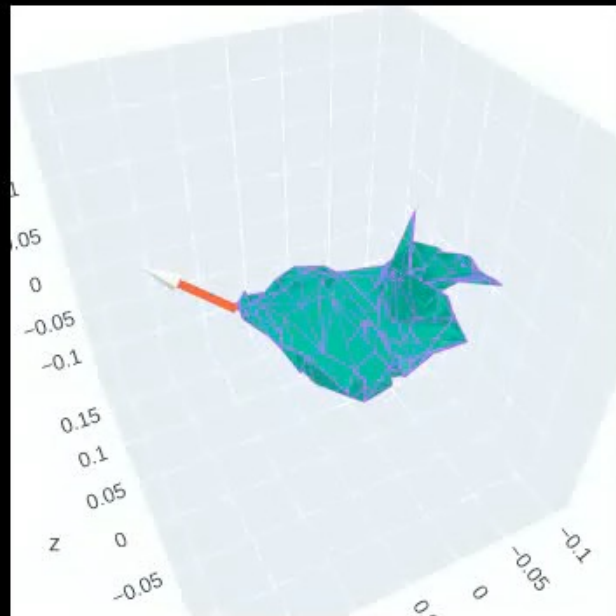
Model rollout

Conclusions

■ Model-based approach

■ Low level dynamic