Feedback Models for Robotic Cloth Manipulation

ELLIIT@Lund - Visiting scholar seminar series

Oriol Barbany - 12 November 2025

Who am I?

Oriol (/uRiɔl/) or Uri for short :)

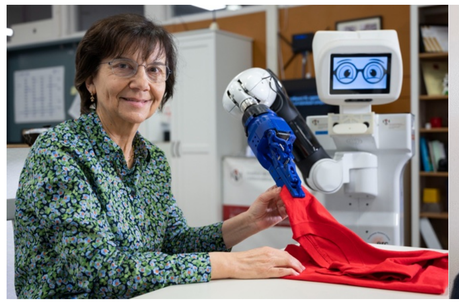
Born and raised in **Barcelona**

BSc Telecom. Engineering

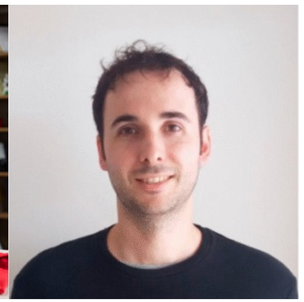
MSc Data Science

PhD Perception for Robotics

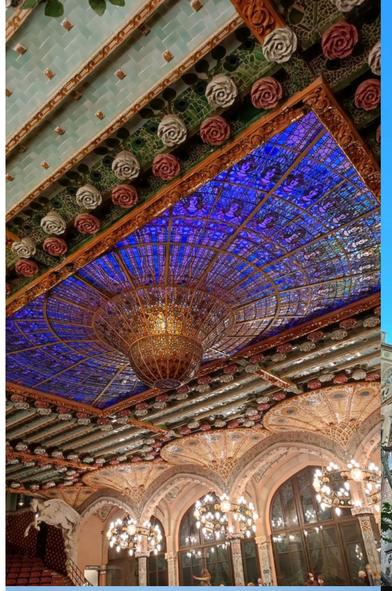
Carme Torras

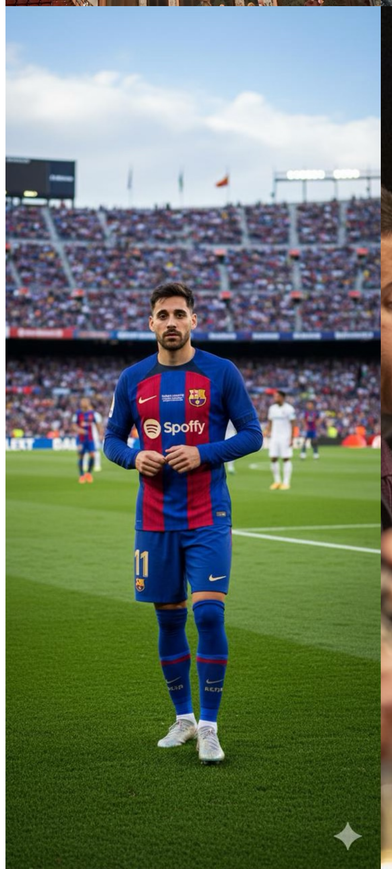


Adrià Colomé

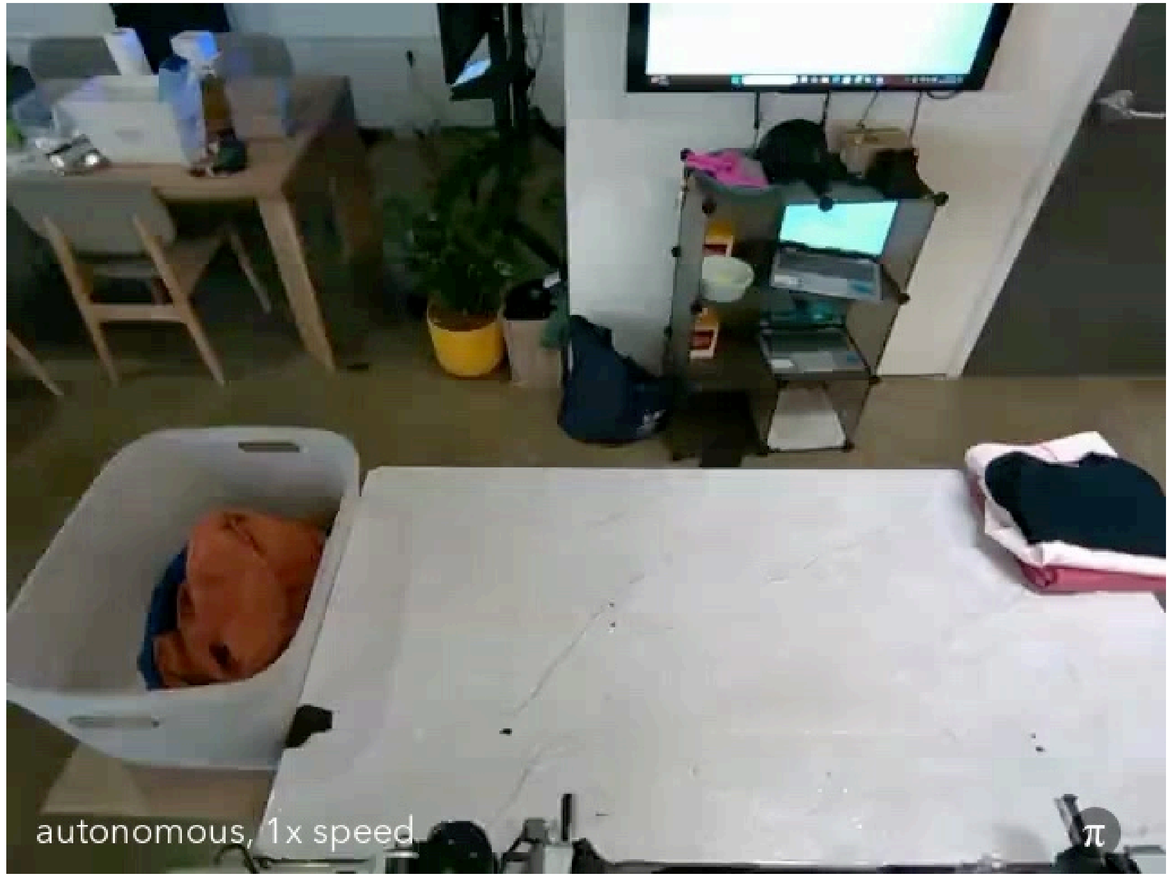


Amir Zamir



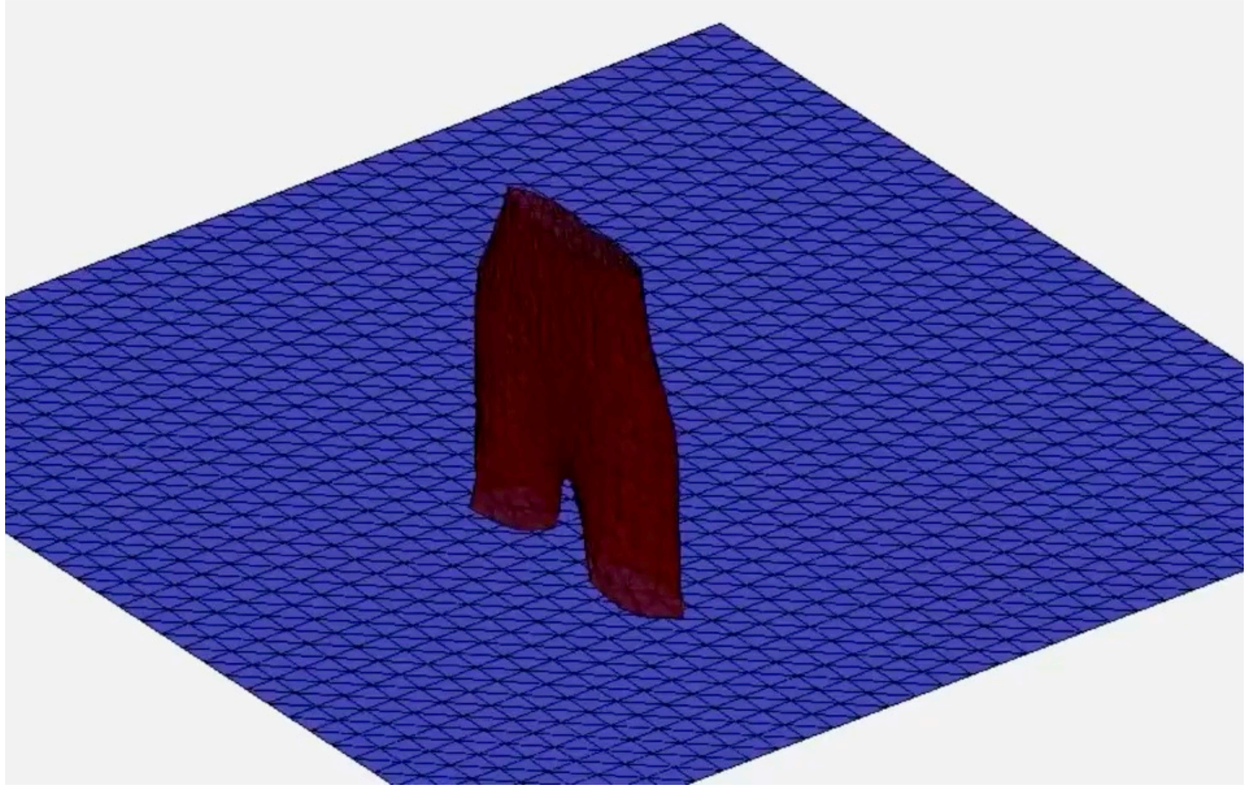


Cloth perception is challenging



Video taken from π_0 (https://www.physicalintelligence.company/blog/pi0)

Simplify cloth modeling by assuming inextensibility [1]

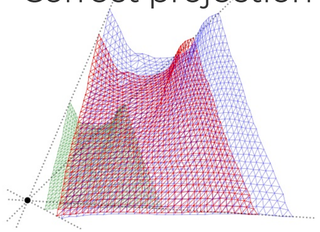


[1] Coltraro, F., et al. (2022). An inextensible model for the robotic manipulation of textiles. Applied Mathematical Modelling

Cloth perception is challenging

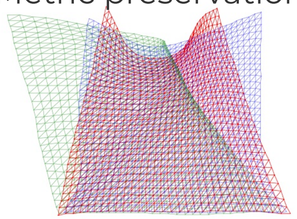
Does *inextensibility* [1] simplify perception?

Correct projection



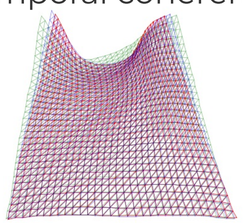
Tracked points on the surface project to image

Metric preservation



Constant geodesics no extension or shrinkage)

Temporal coherence

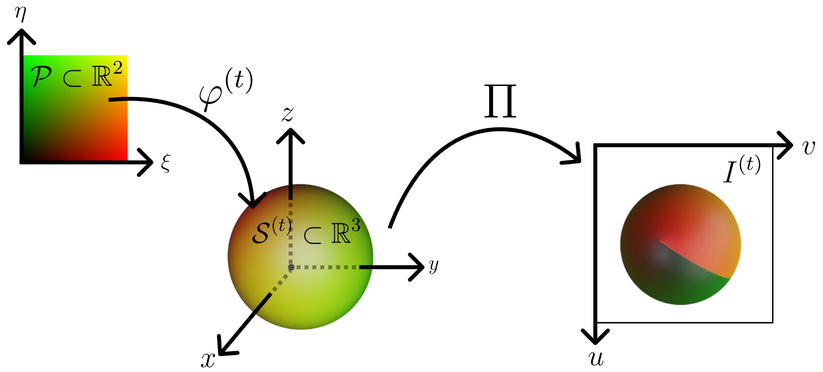


Reduce frame-to-frame flickering

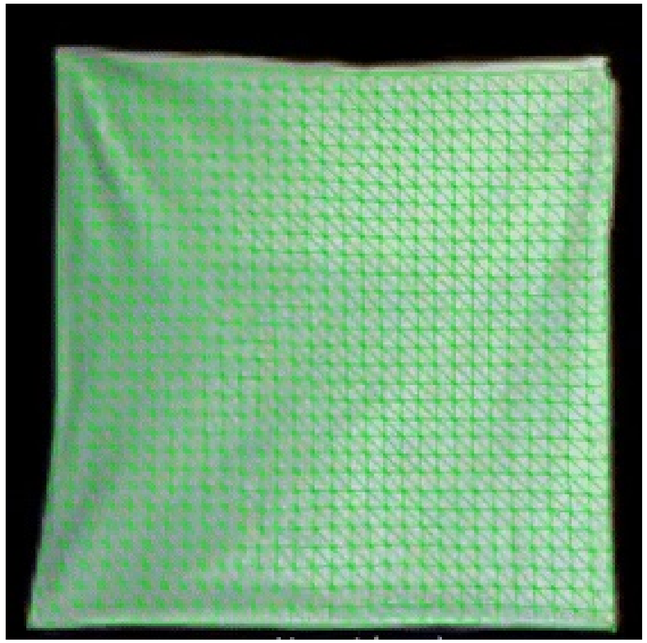
Yes, but:

- Approximate constraint: QP with linear constraints
- Slow to compute

Use neural parametric representation (φ: neural net) of the surface [2]



- Analytical constraint
- Parallelizable
- Continuous surface

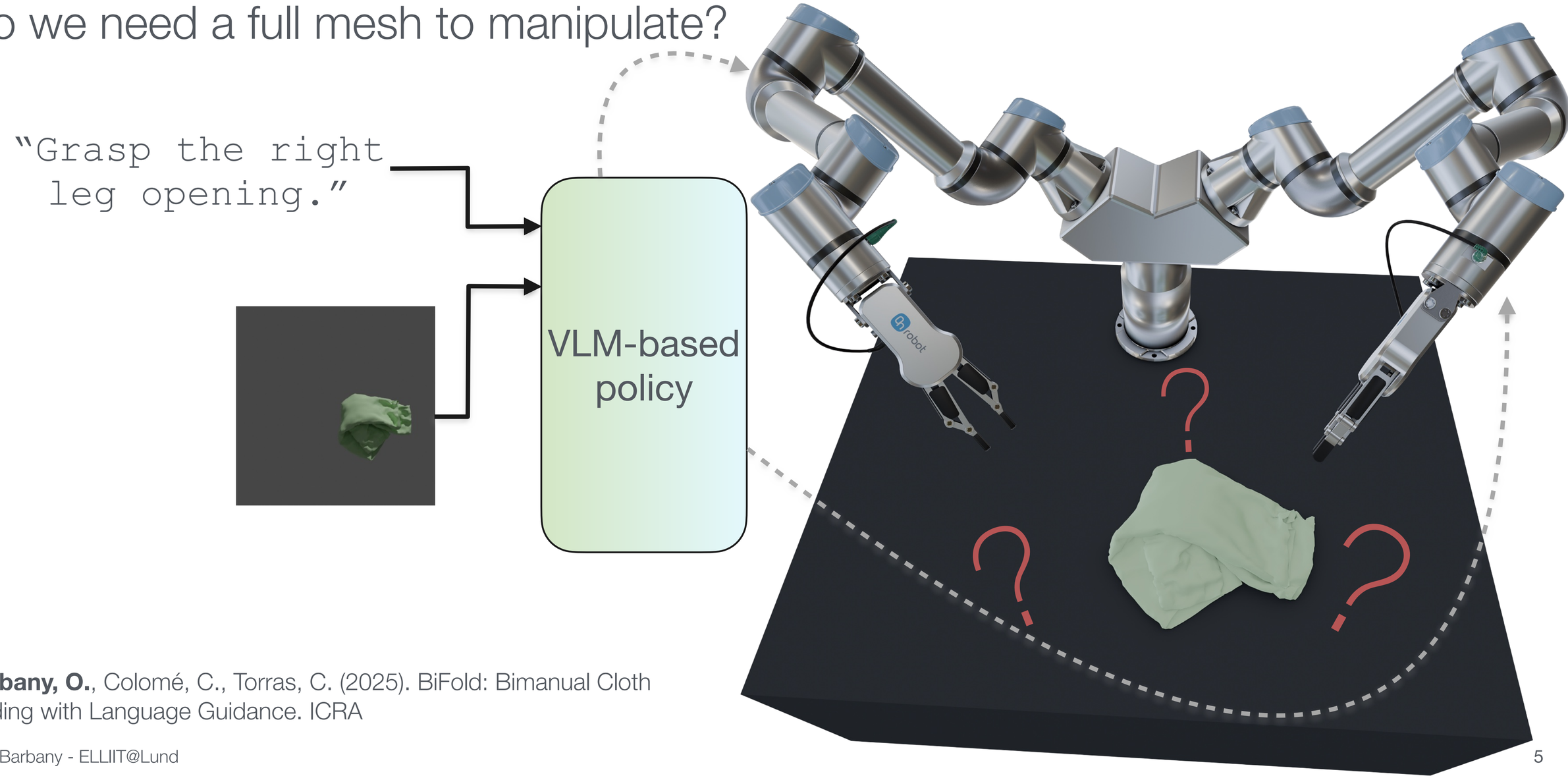


[1] Coltraro, F., et al. (2022). An inextensible model for the robotic manipulation of textiles. Applied Mathematical Modelling

[2] **Barbany, O.**, Colomé, C., Torras, C. (2024). Deformable surface reconstruction via Riemannian metric preservation. CVIU

Cloth perception is challenging

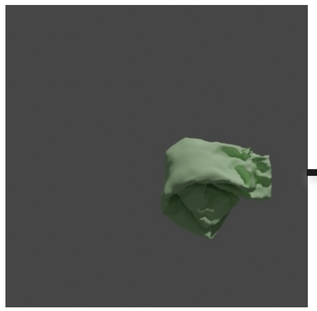
Do we need a full mesh to manipulate?

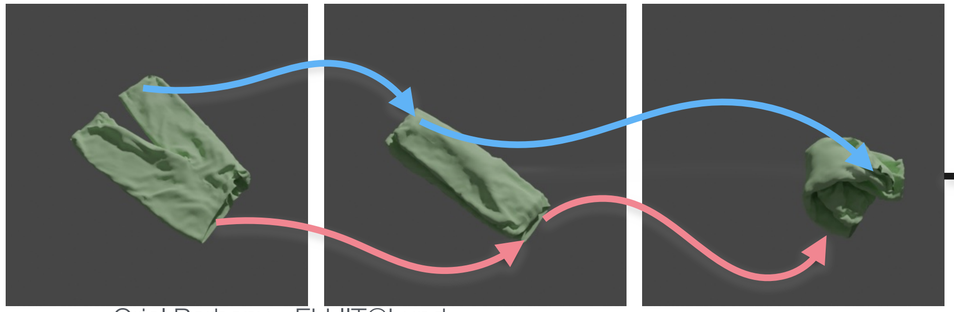


Barbany, O., Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

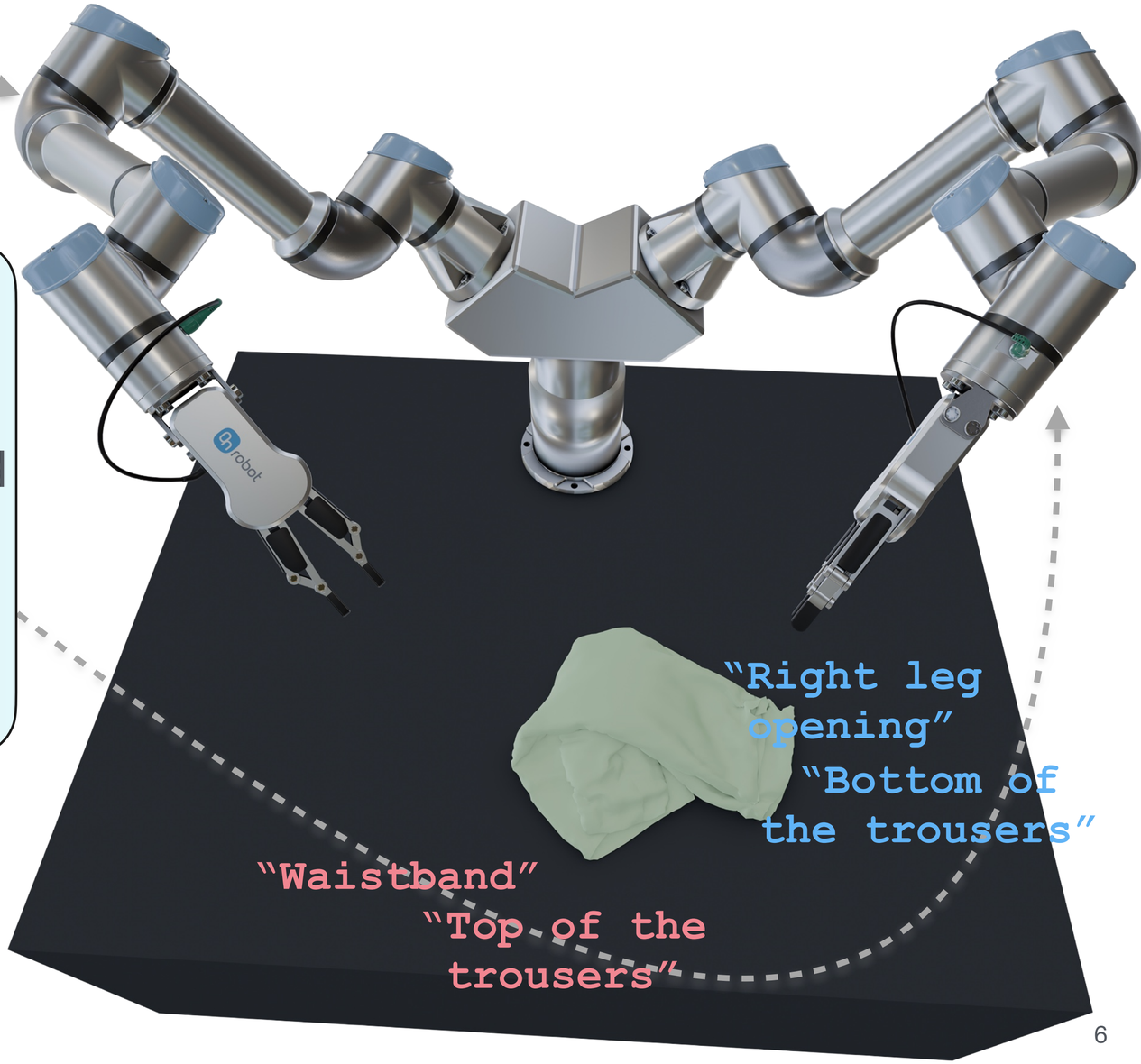
Static Cloth perception is challenging

"Grasp the right leg opening."

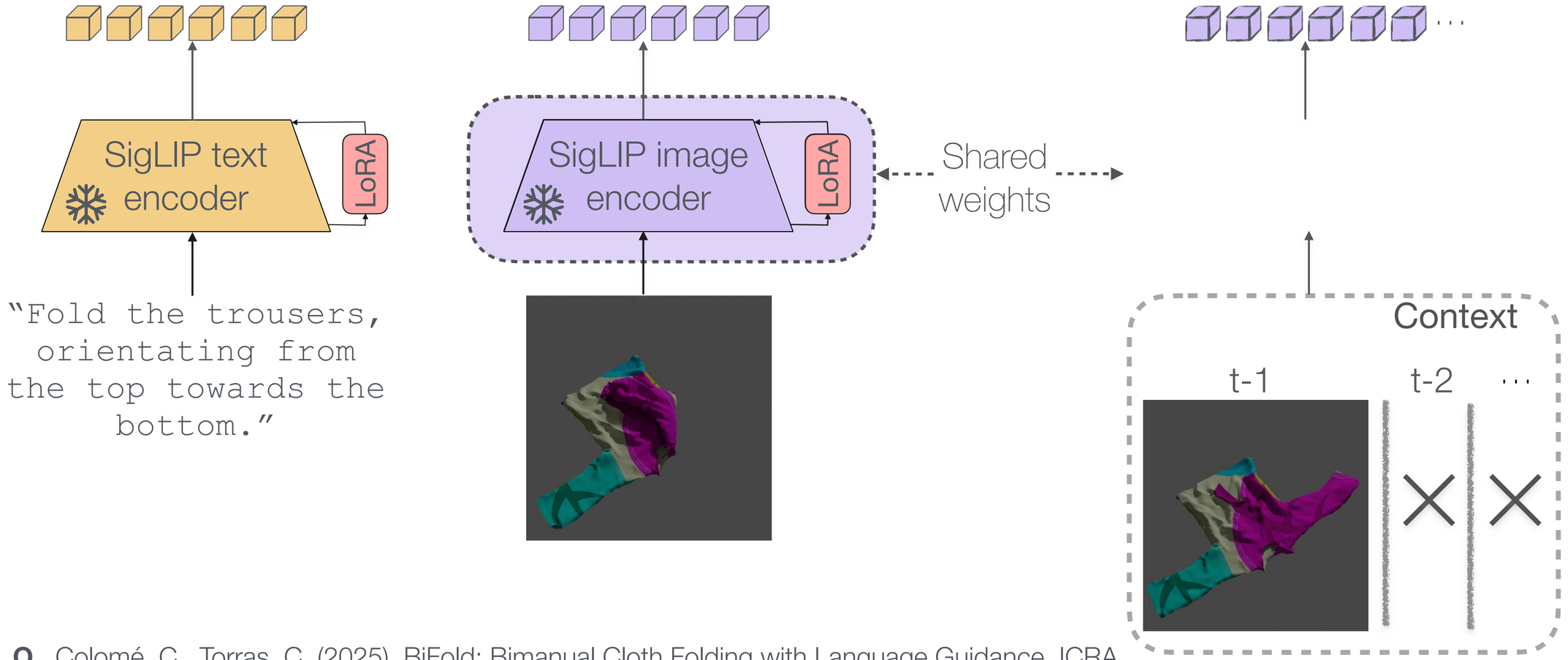




VLM-based policy

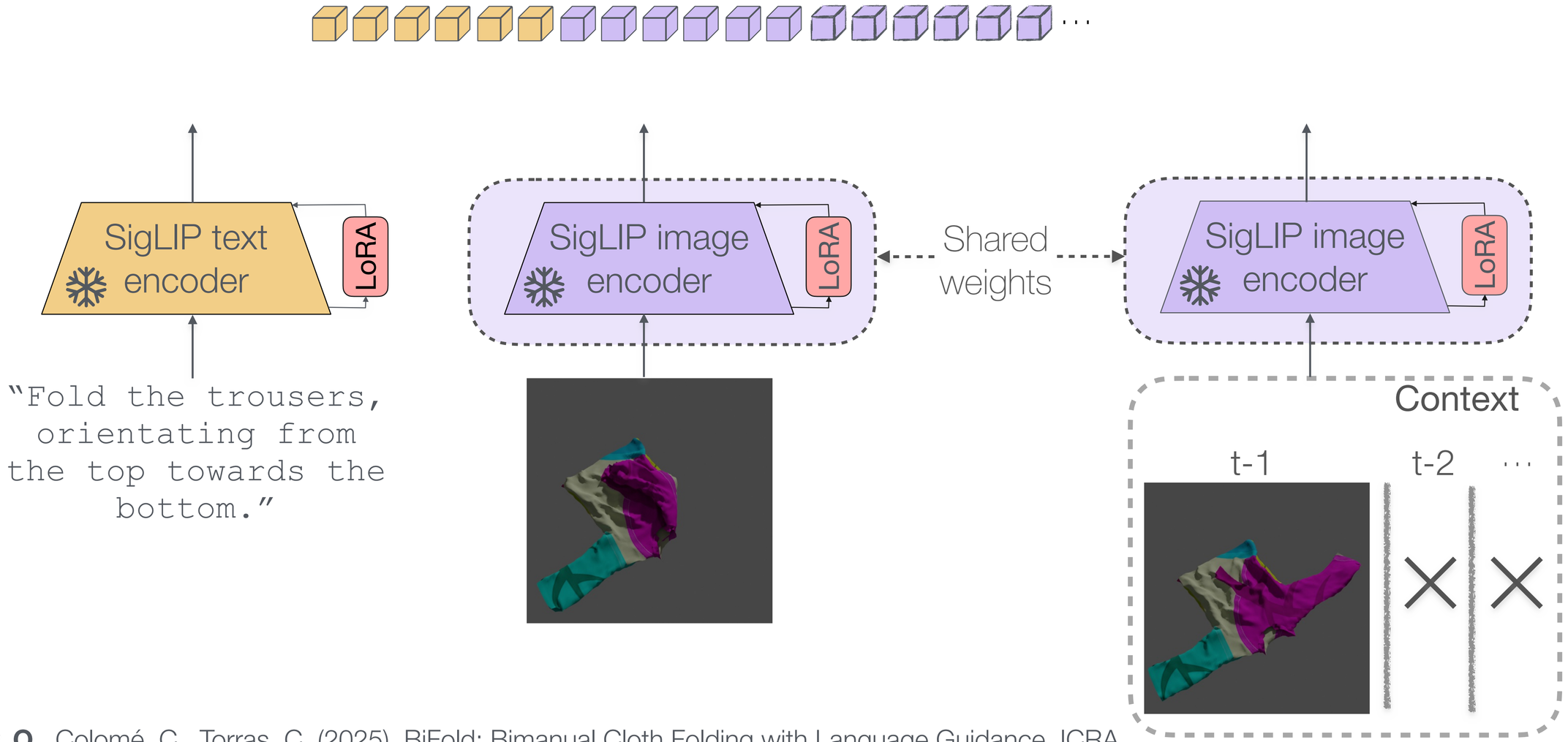


BiFold architecture



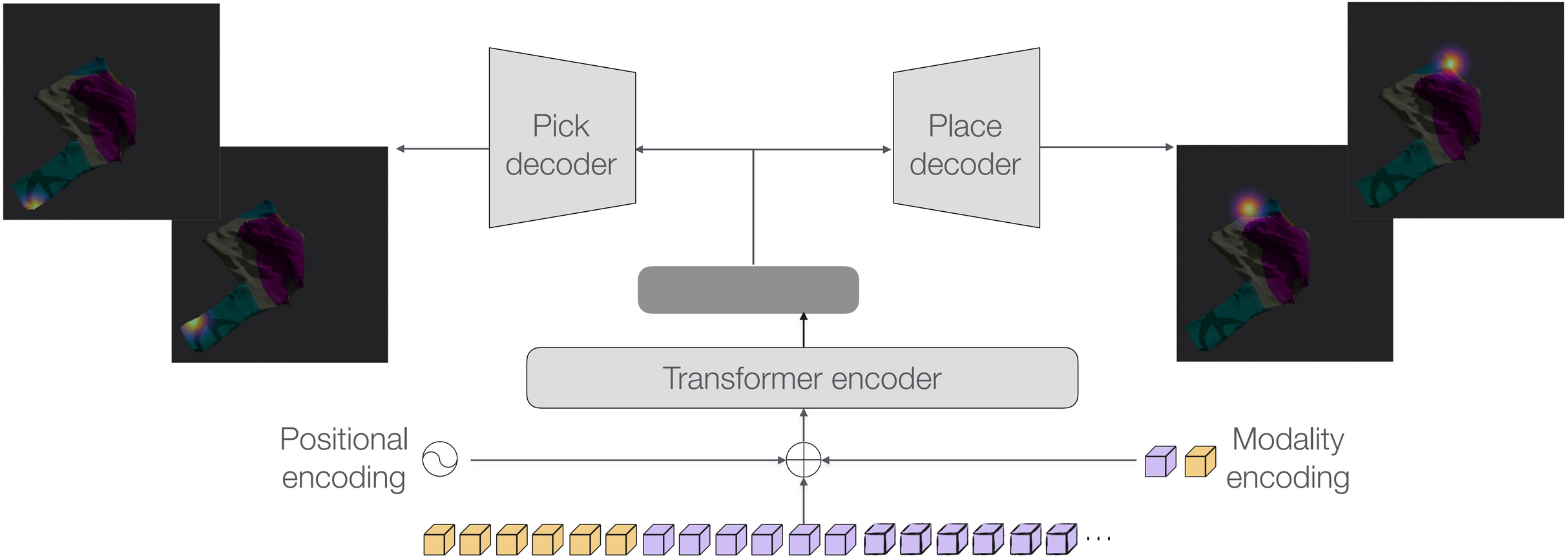
Barbany, O., Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

BiFold architecture



Barbany, O., Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

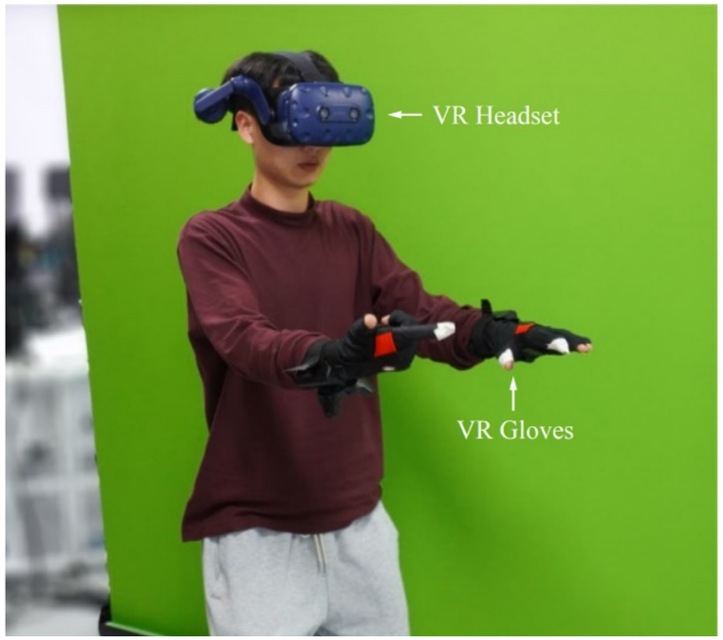
BiFold architecture

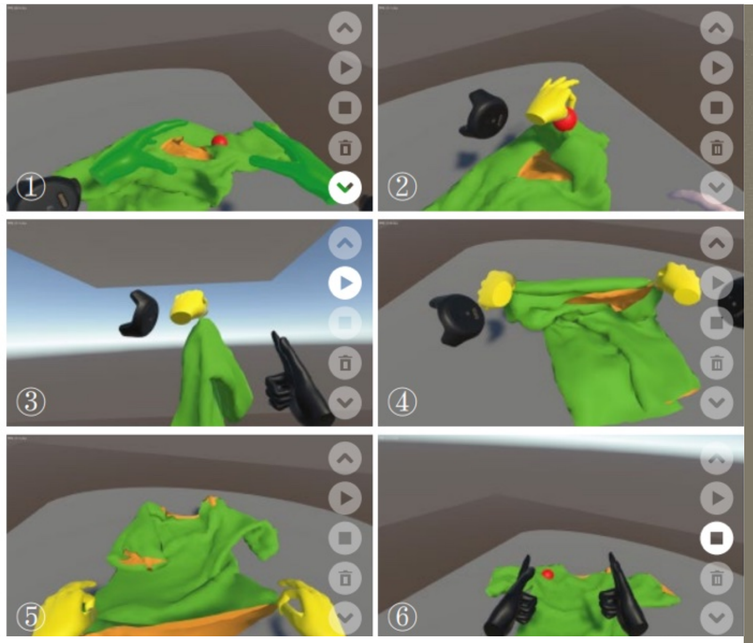


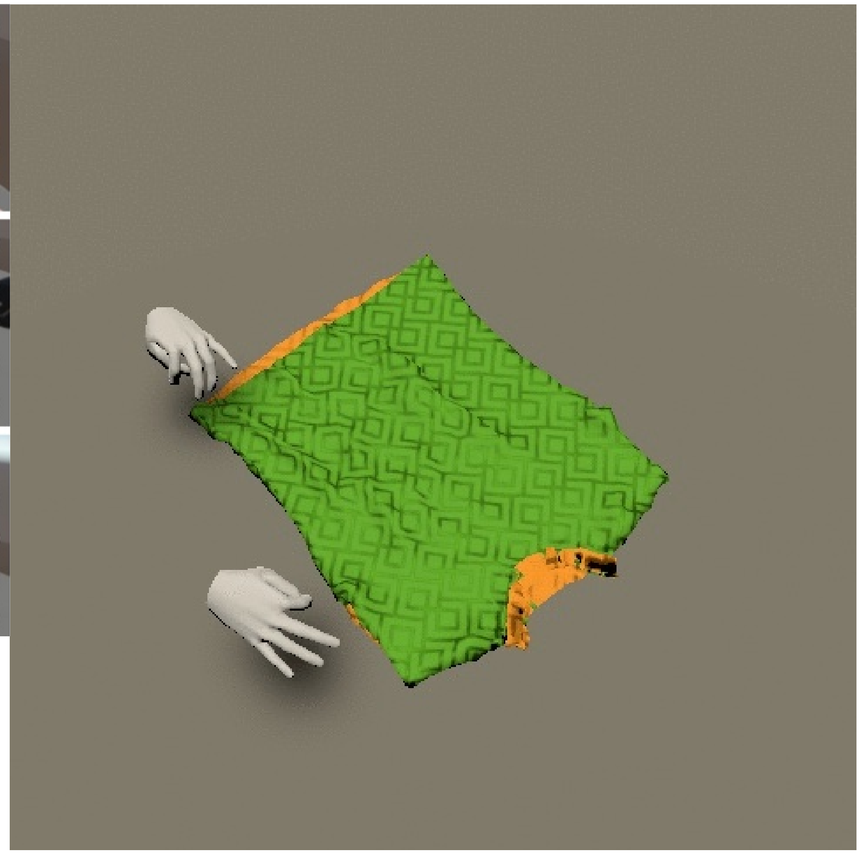
Barbany, O., Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

Folding dataset

Learn from human demos in VR [1]







Images taken from https://huggingface.co/datasets/robotflow/vr-folding

[1] Xue, H. et al. (2023). GarmentTracking: Category-Level Garment Pose Tracking. CVPR

Folding dataset

Assign language instruction using NOCS [1]

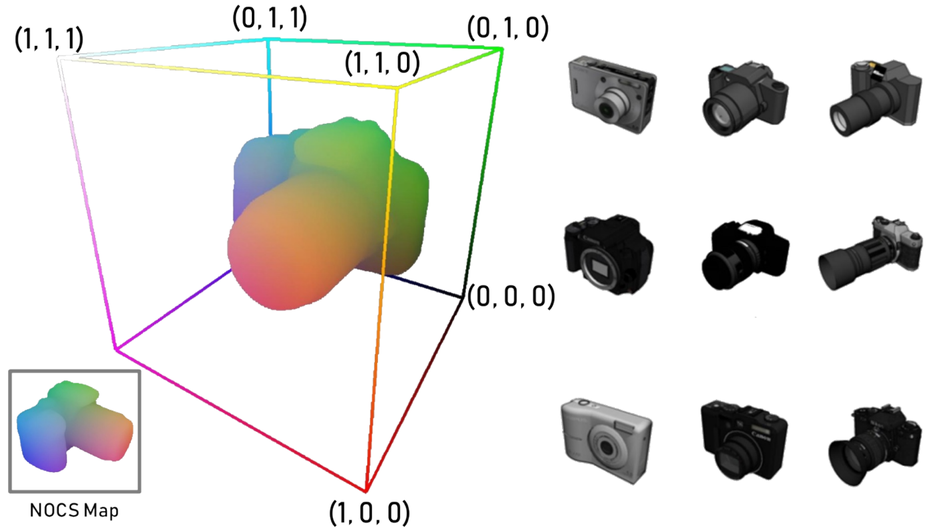
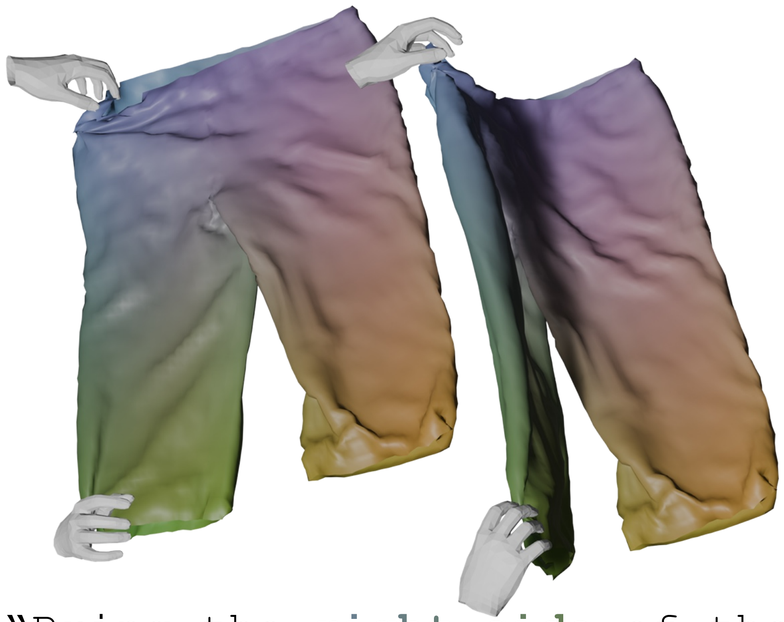
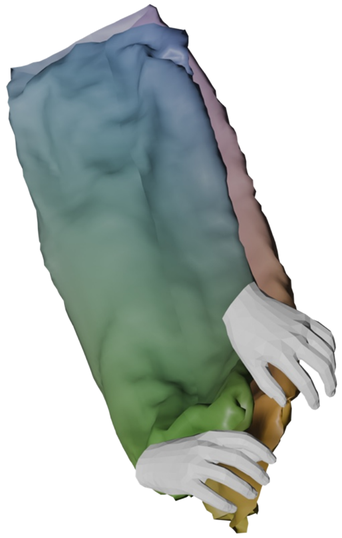


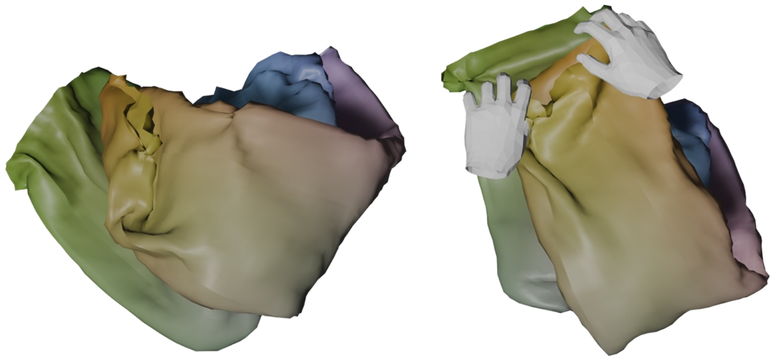
Image taken from [1]

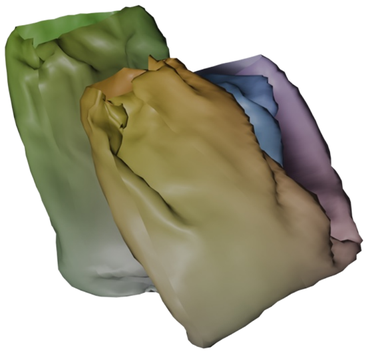


"Bring the **right side** of the trousers towards the **left side** and fold them in half."



"Bend the trousers in half, from **bottom** to **top**."

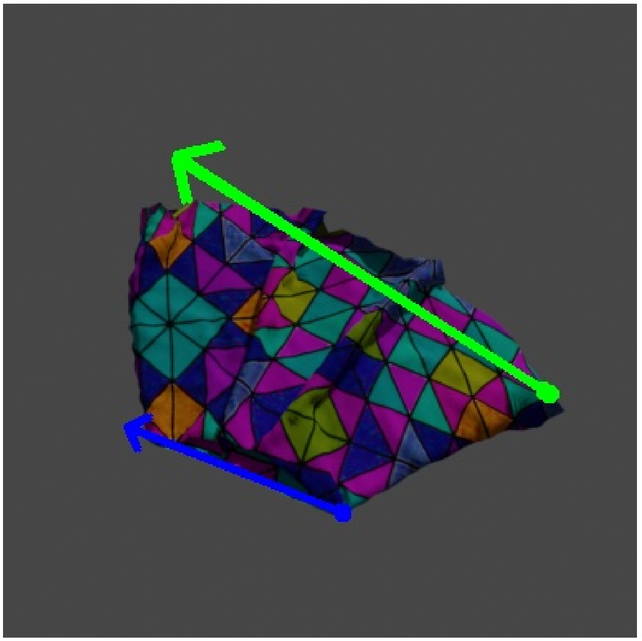




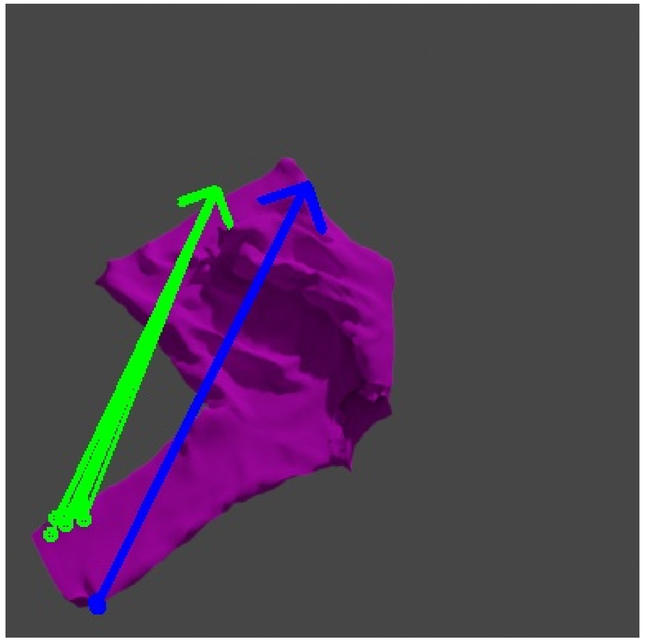
"Straighten out the **bottom** part of the trousers."

[1] Wang, H., et al (2019). Normalized object coordinate space for category-level 6d object pose and size estimation. CVPR

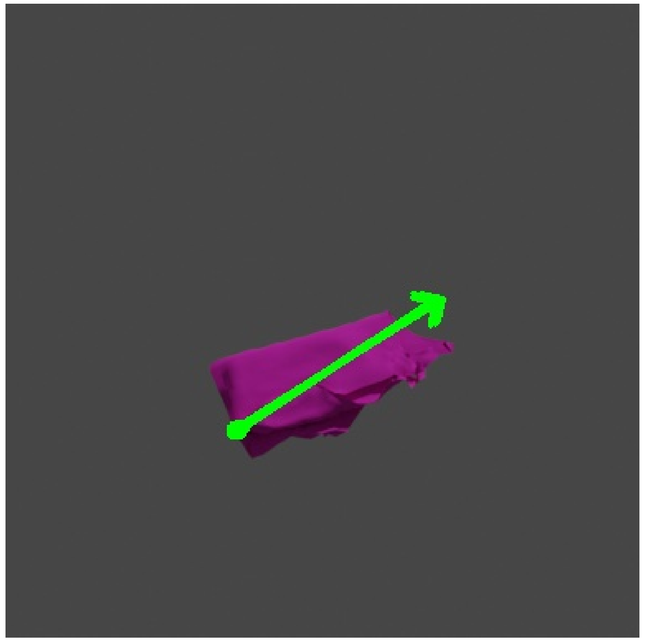
BiFold folding dataset



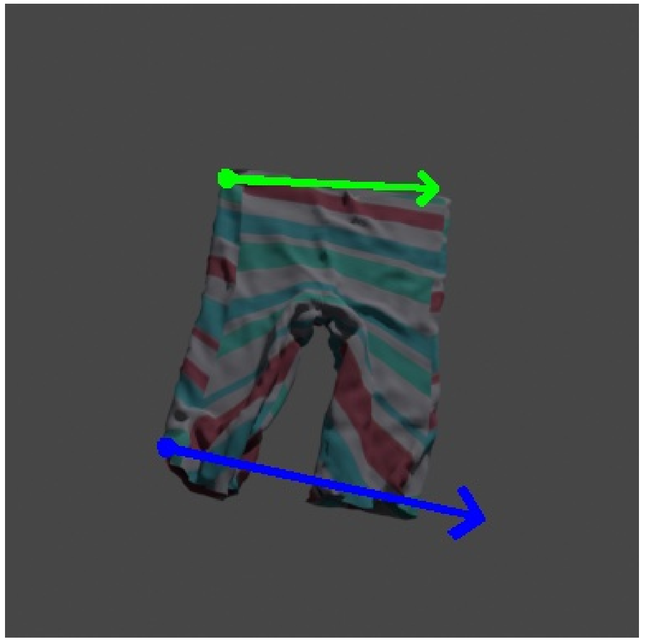
"Fold the skirt in half horizontally, right to left."



"Bring the left sleeve to the center."



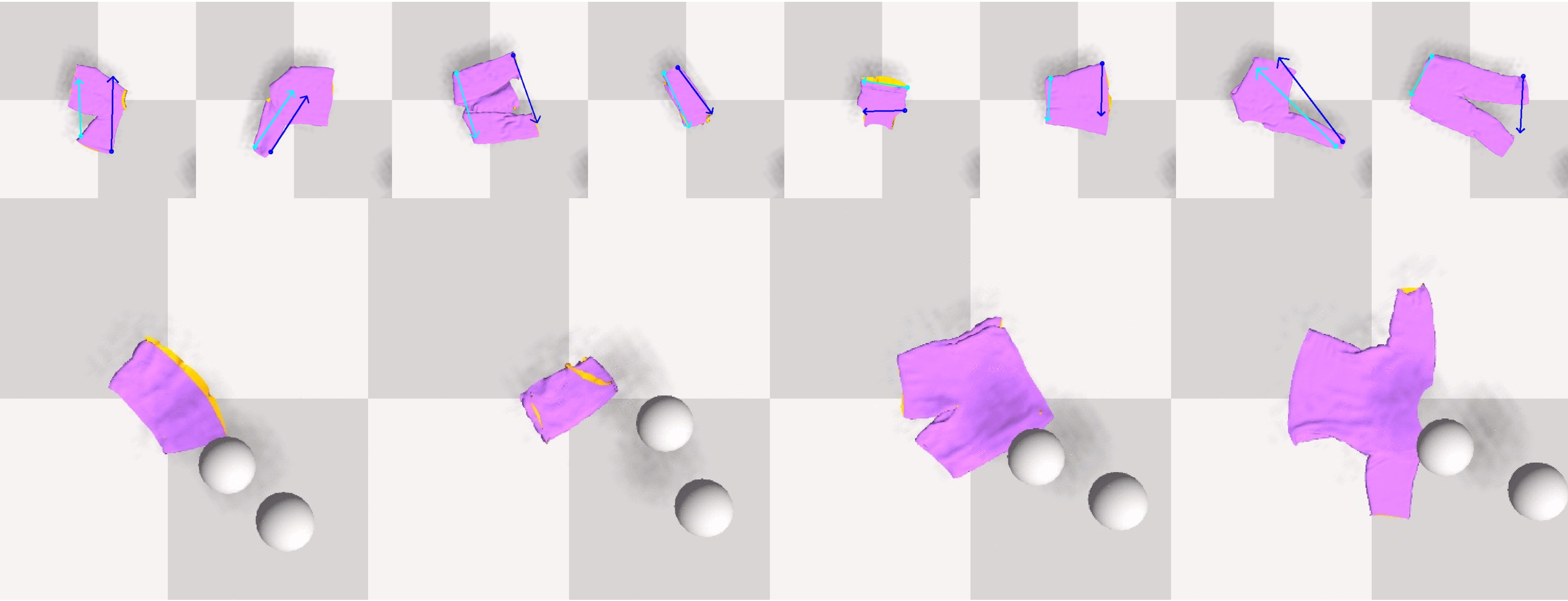
"Fold the top, making sure the bottom right side touches the top right side only using the right arm."



"Create a fold in the trousers, going from right to left."

Barbany, O., Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

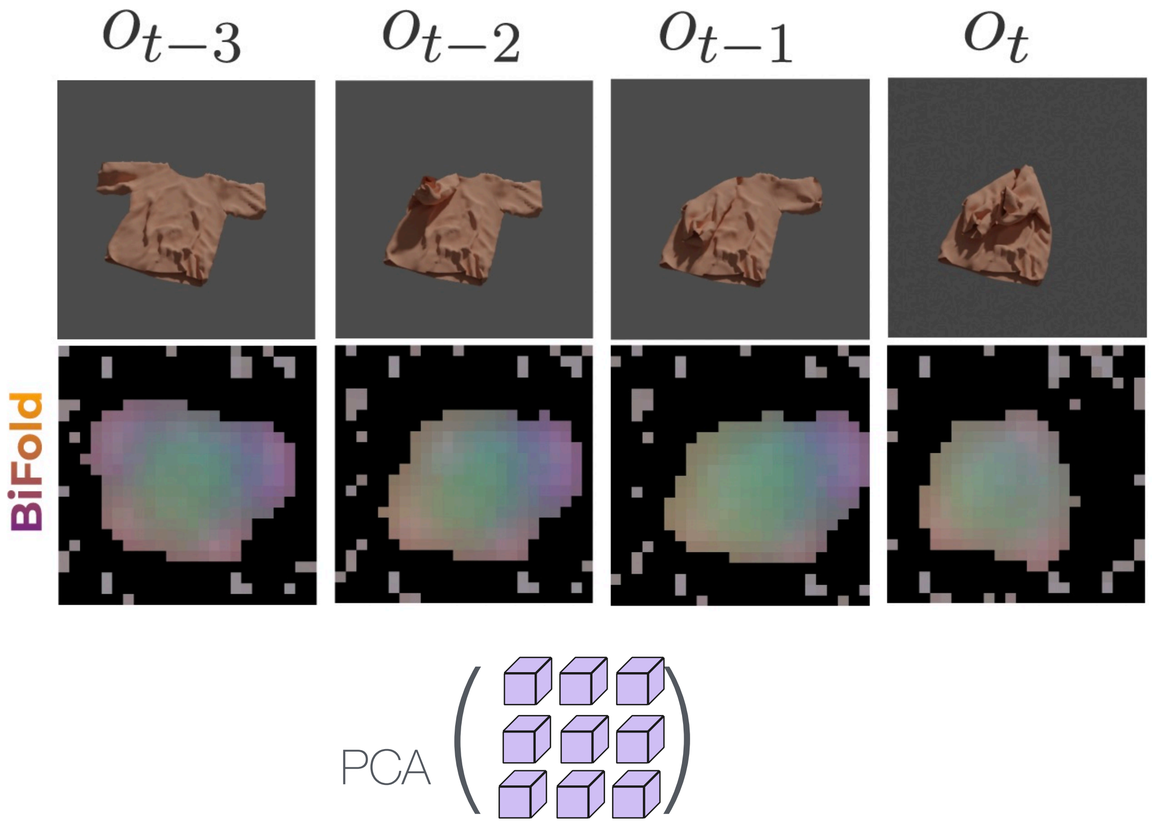
Zero-shot sim2sim on SoftGym [1]

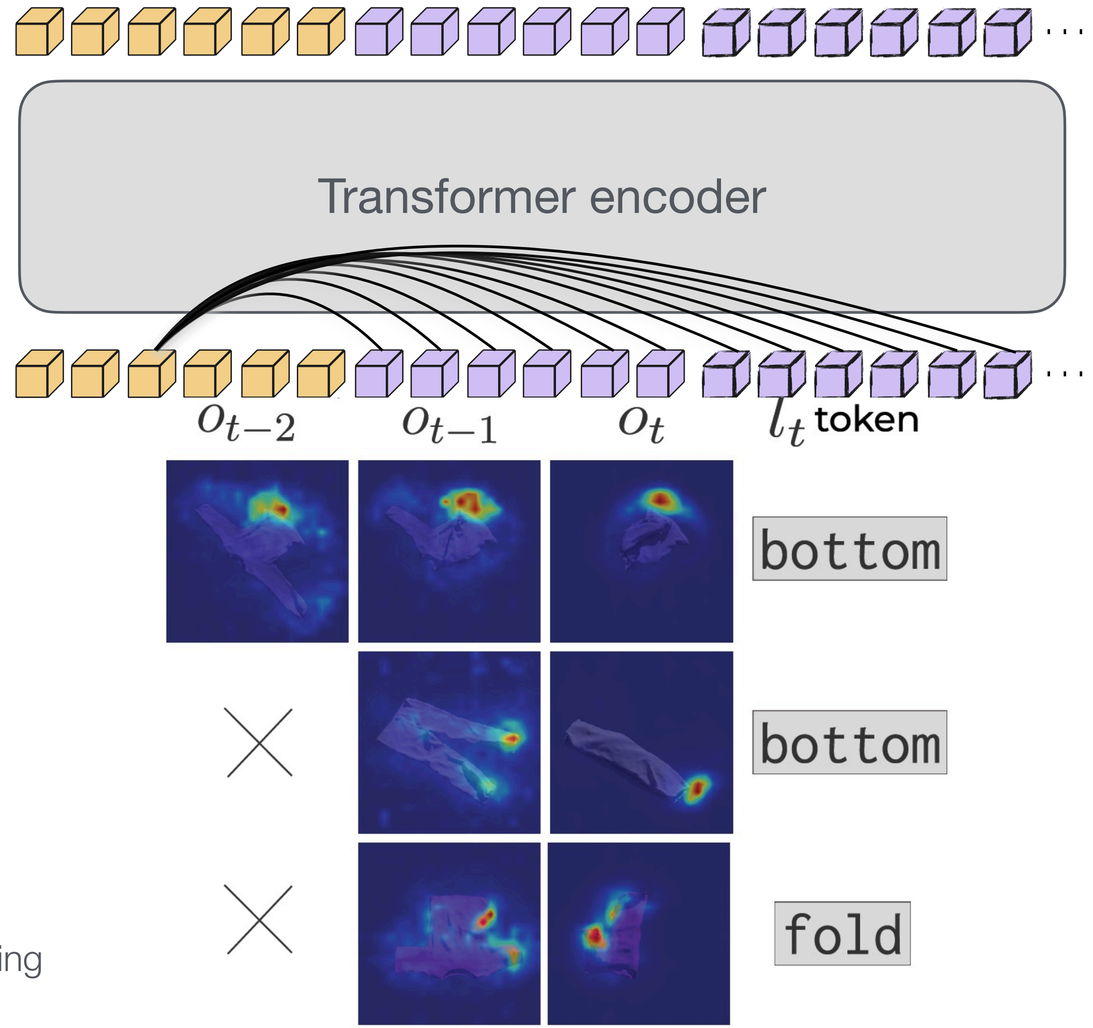


Results from **Barbany, O.**, Colomé, C., Torras, C. (2025). BiFold: Bimanual Cloth Folding with Language Guidance. ICRA

[1] Lin, X., Wang. Y., Olkin, J., Held, D. (2022). SoftGym: Benchmarking Deep Reinforcement Learning for Deformable Object Manipulation. CoRL

What does BiFold learn?

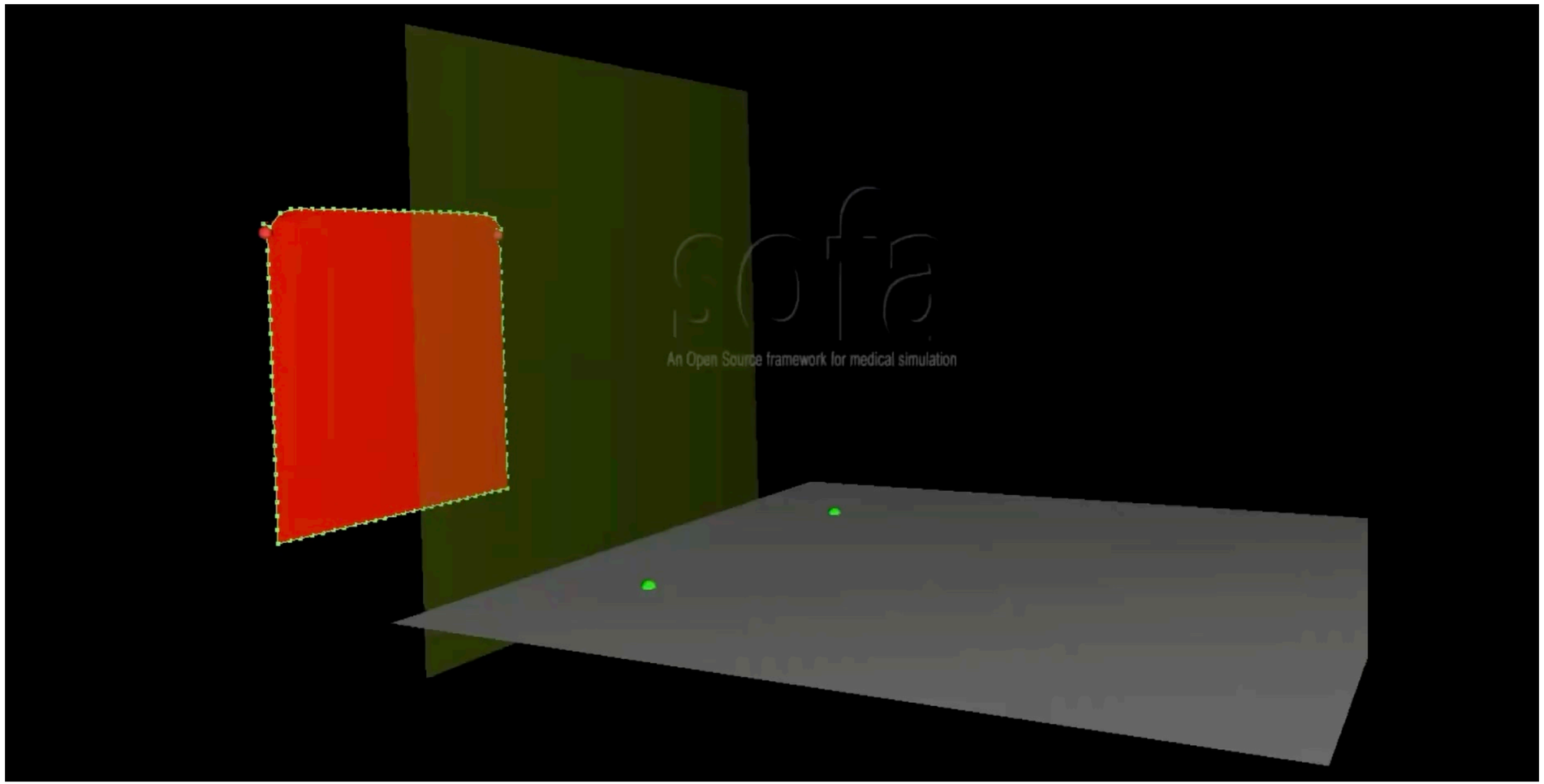




Barbany, O., Colomé, C., Torras, C. (2025). Beyond Static Perception: Integrating Temporal Context into VLMs for Cloth Folding. ICRA workshop

How reliable are (cloth) simulators?

Some of them not much…



Video taken from Jangir, R., Alenyà, G., Torras, C. (2020). Dynamic Cloth Manipulation with Deep Reinforcement Learning. ICRA

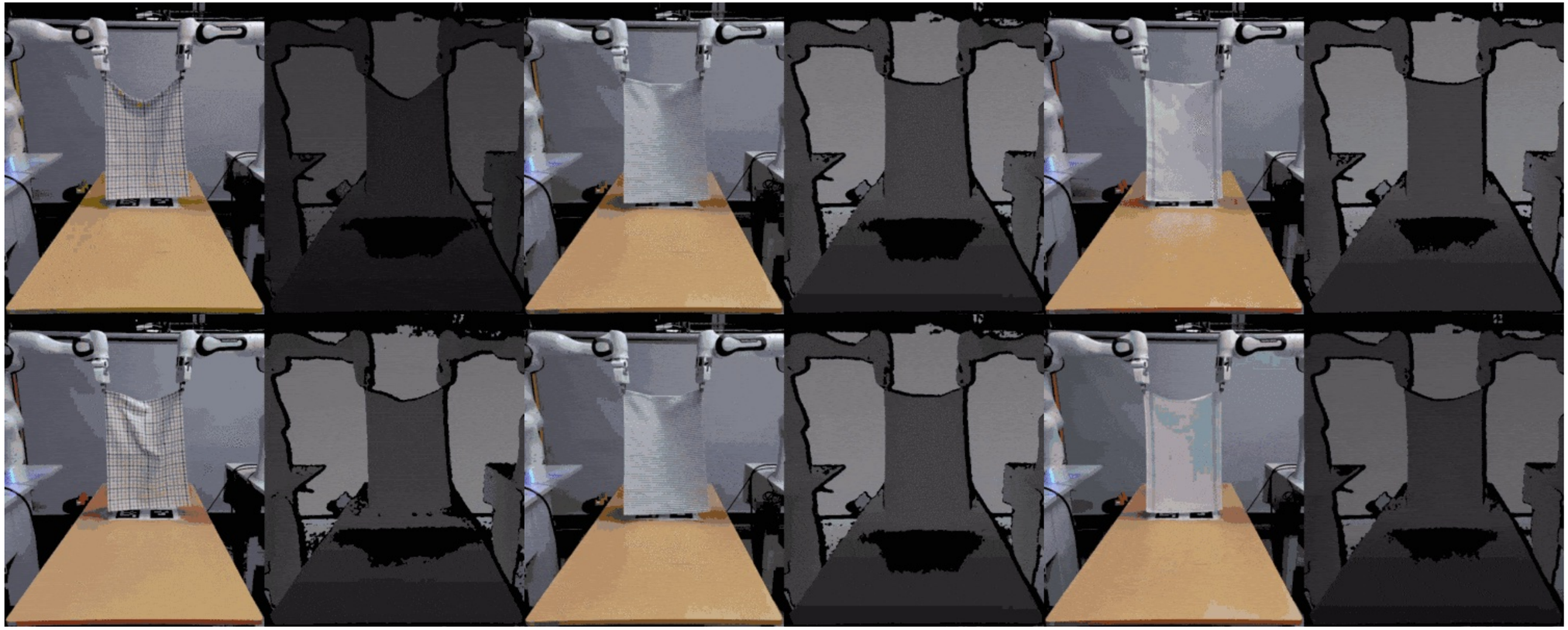
How reliable are (cloth) simulators?

Benchmark sim2real gap

Stiffer Lighter

Dynamic

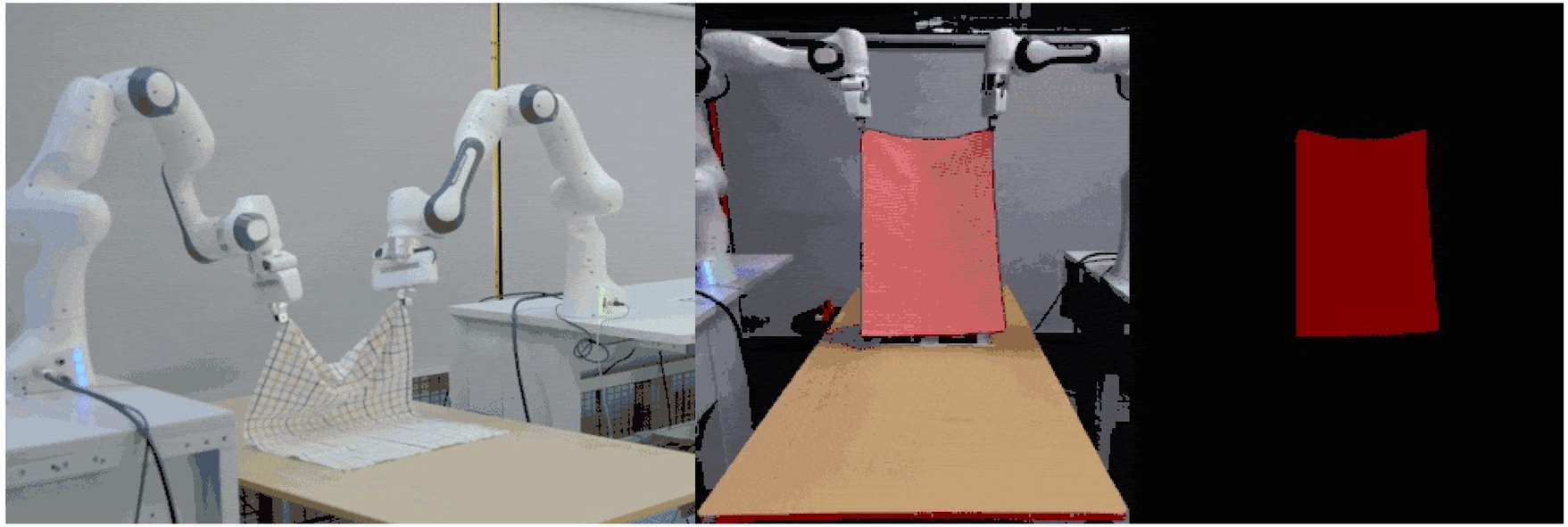
Quasi-static



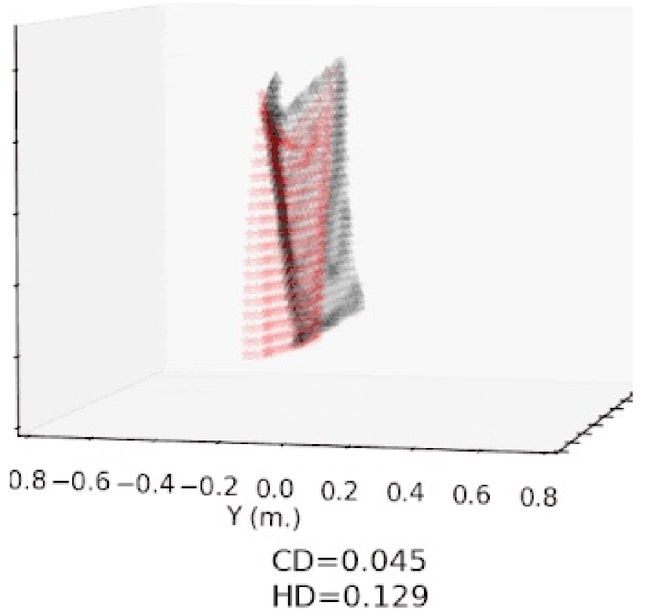
Blanco-Mulero, D., **Barbany, O.**, et al. (2024). Benchmarking the Sim-to-Real Gap in Cloth Manipulation. RA-L

How reliable are (cloth) simulators?

Obtain cloth point cloud, then compare to sim



Digital twin in 4 simulators
MuJoCo, Flex, SOFA, Bullet



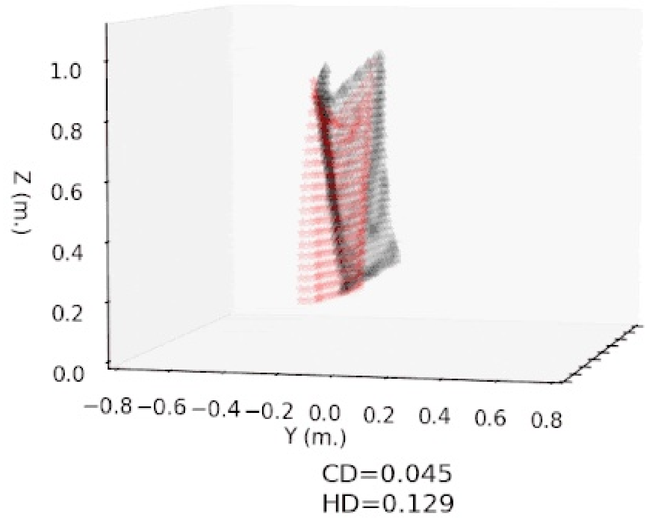
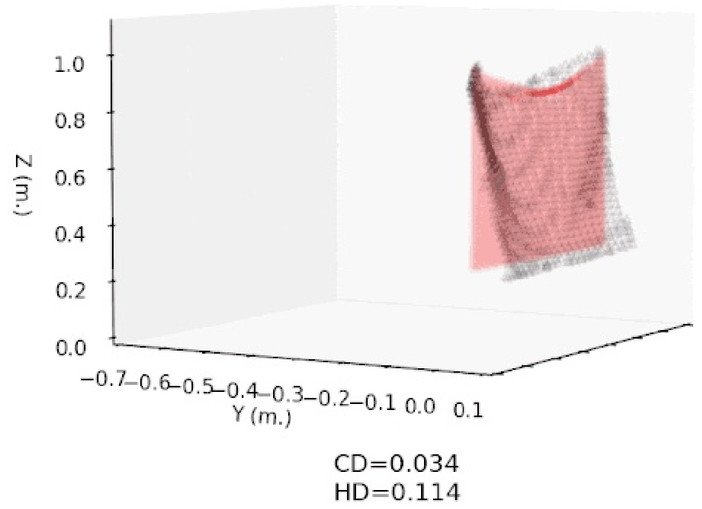
CD: Unidirectional Chamfer Distance
HD: Unidirectional Hausdorff Distance

Blanco-Mulero, D., **Barbany, O.**, et al. (2024). Benchmarking the Sim-to-Real Gap in Cloth Manipulation. RA-L

How reliable are (cloth) simulators?

Tune cloth parameters by Bayesian Opt wrt *unidirectional* CD

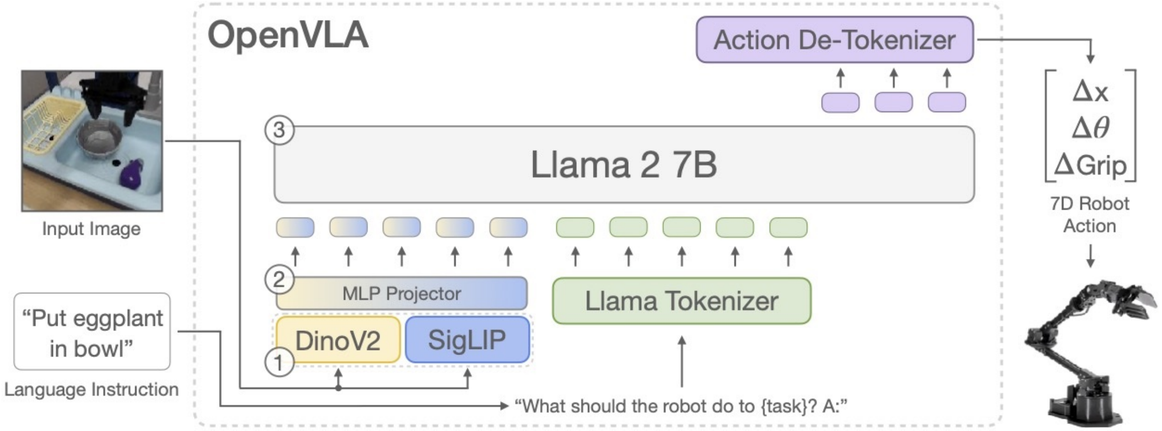
Sim
Real

	Dynamic	Quasi-static
Reality gap:	 >> 
Simulator:	**MuJoCo** > Flex ~ SOFA > Bullet	~Same order but less relevant

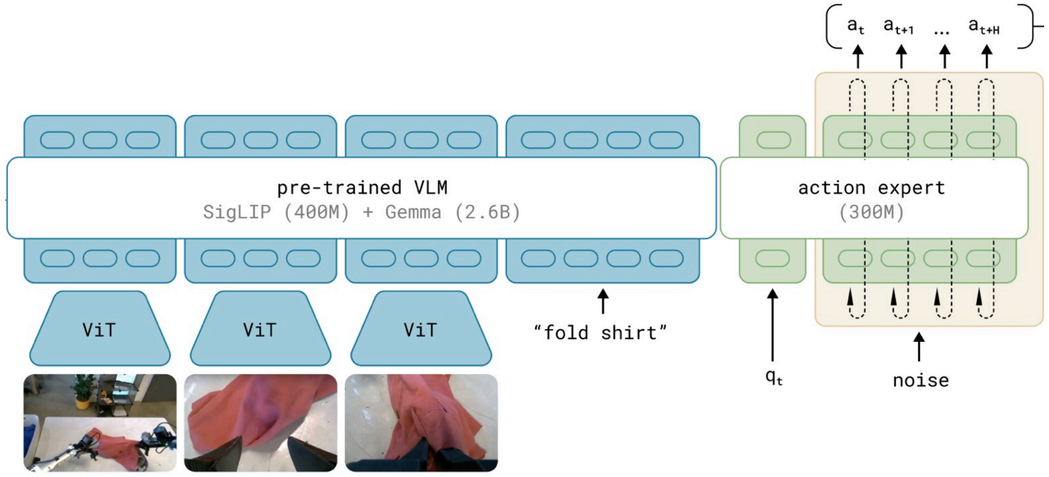
	Simulation Generic				Deformable Objects	
Physics Simulator	RGB-D	Robot Integration	Grasp	CPU/GPU	Shape	Dynamics Model
MuJoCo [13]	✓	✓	P/L	CPU & GPU	Variable[5]	Mass-Spring
Bullet [15]	✓	✓	P/L	CPU	Variable	PBD / FEM
Flex [14, 20]	✓	✓[4]	P/L	GPU	Variable	PBD
SOFA [16]	RGB	✗	P	CPU & GPU	Variable	Mass-Spring / FEM

Blanco-Mulero, D., **Barbany, O.**, et al. (2024). Benchmarking the Sim-to-Real Gap in Cloth Manipulation. RA-L

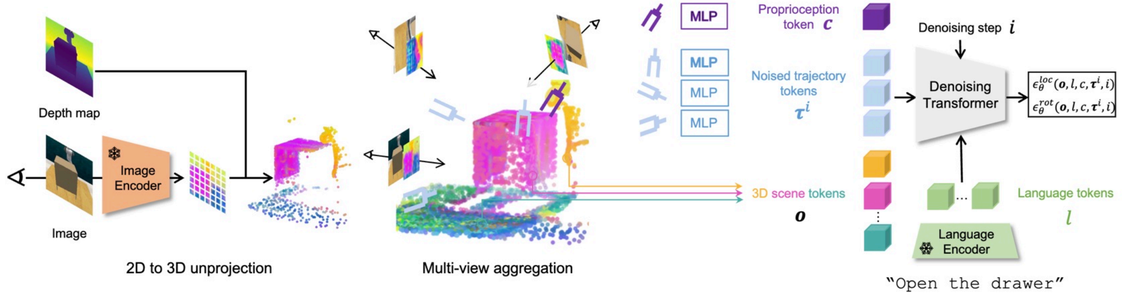
Is RGB the best we can do?



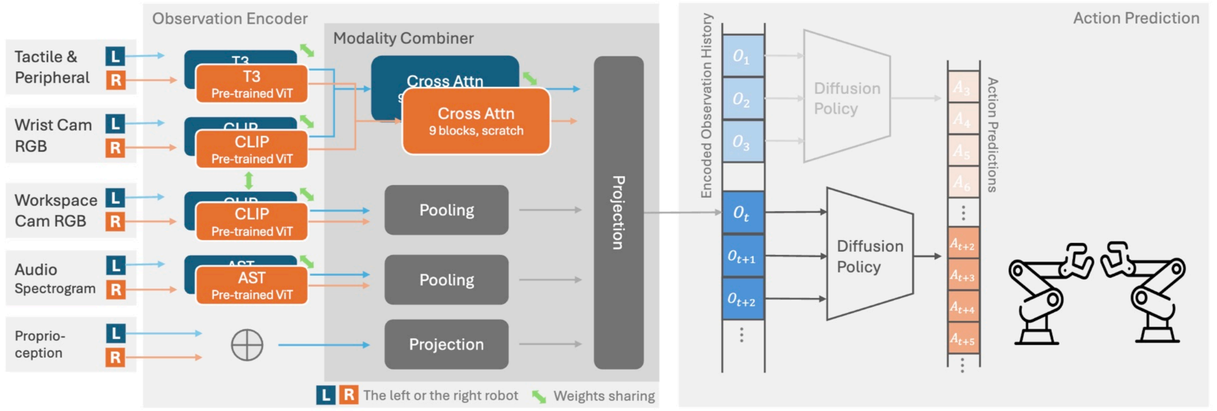
Kim, M. J. et al. (2025). OpenVLA: An Open-Source Vision-Language-Action Model. CoRL



Black, K. et al. (2024). π_0: A Vision-Language-Action Flow Model for General Robot Control. ArXiv

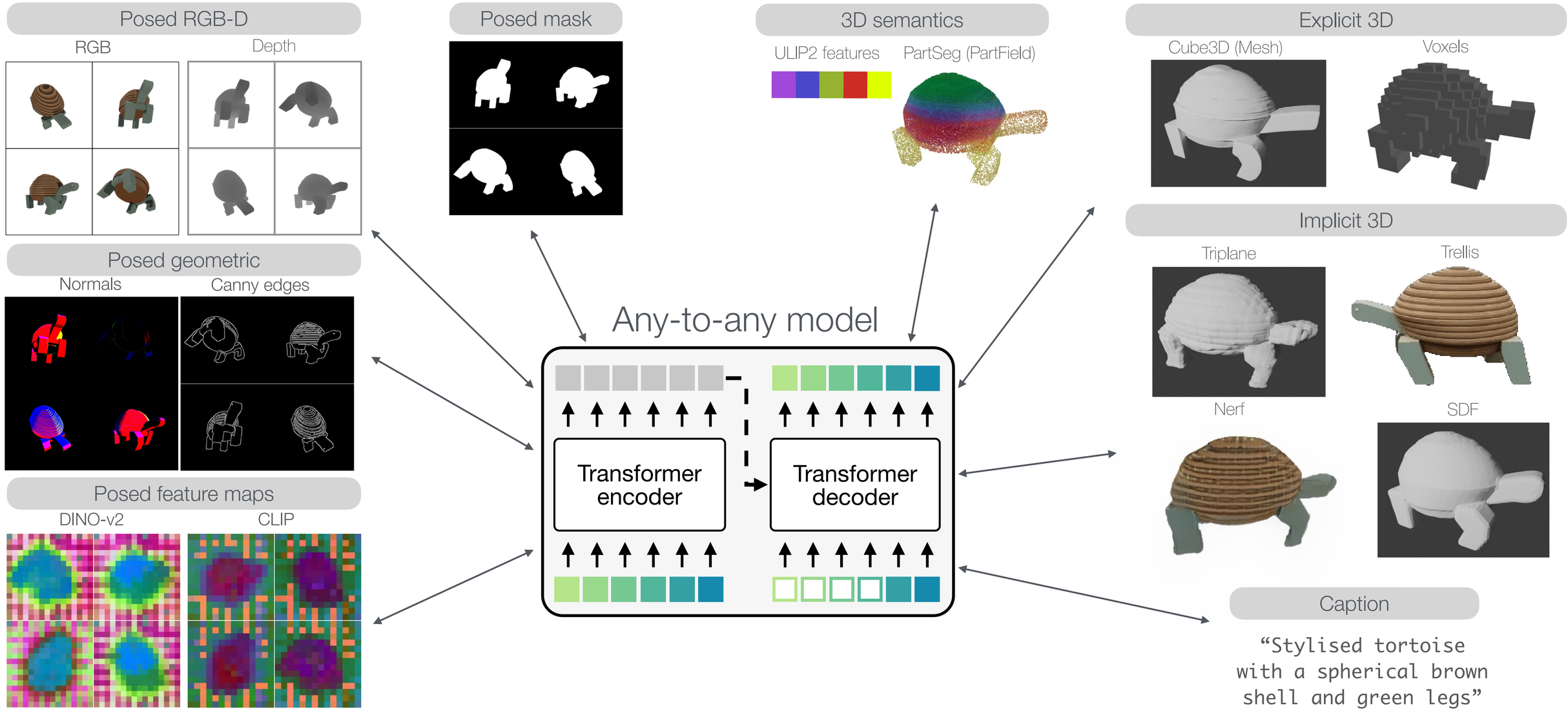


Khe, T., et al. (2024). 3D Diffuser Actor: Policy Diffusion with 3D Scene Representations. CoRL



Zhao, J. et al. (2025). PolyTouch: A Robust Multi-Modal Tactile Sensor for Contact-rich Manipulation Using Tactile-Diffusion Policies. ICRA

Foundation 2D-3D model

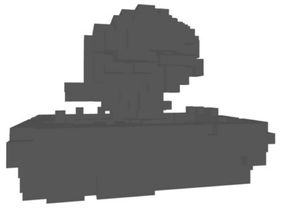


Toskov, J.*, **Barbany, O.***, et al. (2026). Bridging the gap from 2D to 3D with large-scale multimodality. In preparation

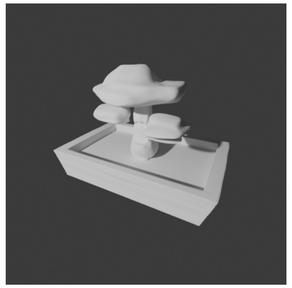
Foundation 2D-3D model

Generation:

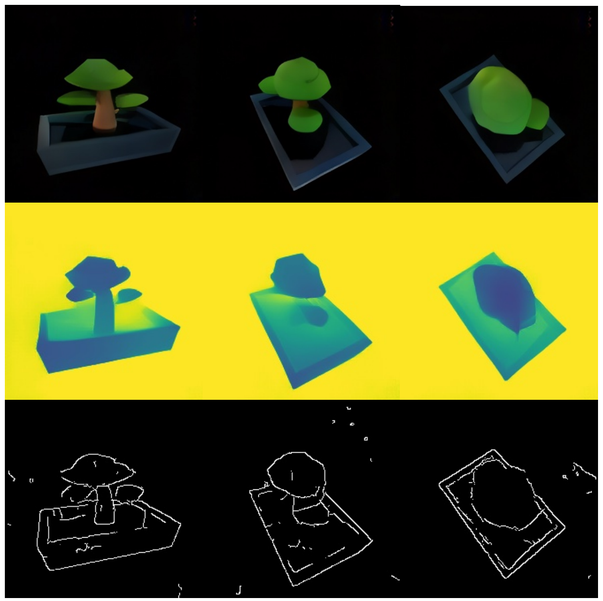
Voxels



NeRF



Mesh (Cube3D)

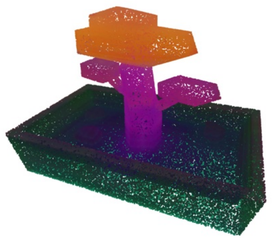


RGB

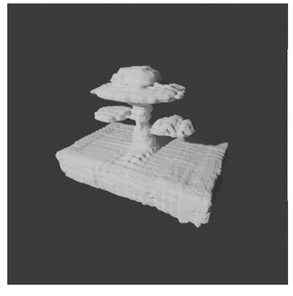
Depth

Canny edges

Mask

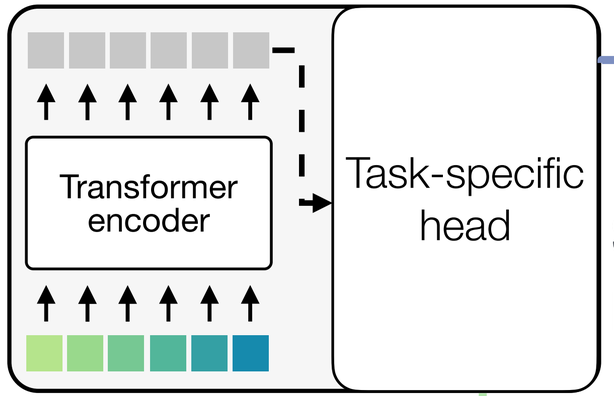


PartSeg (PartfField)



Triplane

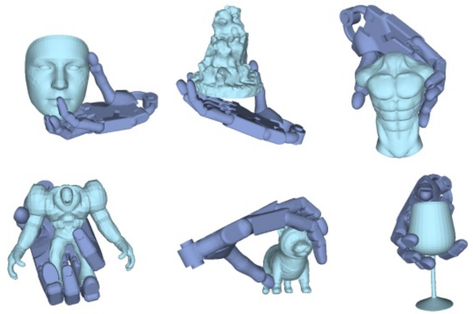
Use features for downstream tasks:



ShadowHand parameters

Class

Wu, Z. et al. (2015). 3D ShapeNets: A Deep Representation for Volumetric Shapes. CVPR



Zhong, Y. et al. (2025). DexGraspAnything: Towards Universal Robotic Dexterous Grasping with Physics Awareness. CVPR

Toskov, J.*, **Barbany, O.***, et al. (2026). Bridging the gap from 2D to 3D with large-scale multimodality. In preparation

Summary

Clothes have complicated dynamics and nearly ∞ DOF

- Perception:
  - Physical assumptions (inextensibility)
  - End-to-end solutions (more general, no GT, …)
  - Temporal information
  - Additional modalities
- Simulation:
  - Importance to identify sim parameters
  - Inaccurate sim physics

obarbany@iri.upc.edu