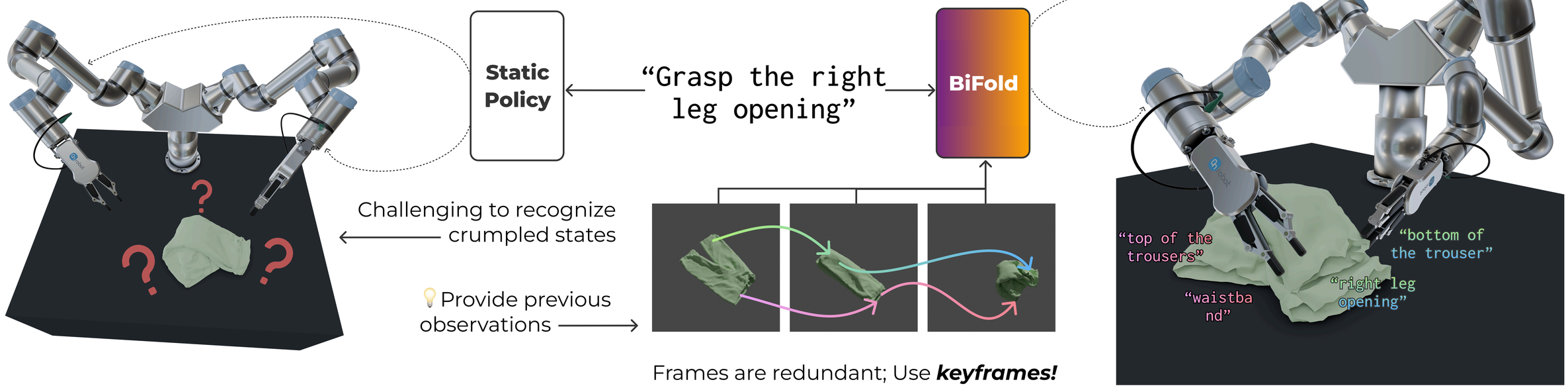
BiFold: Bimanual Cloth Folding with Language Guidance

Oriol Barbany, Adrià Colomé, Carme Torras

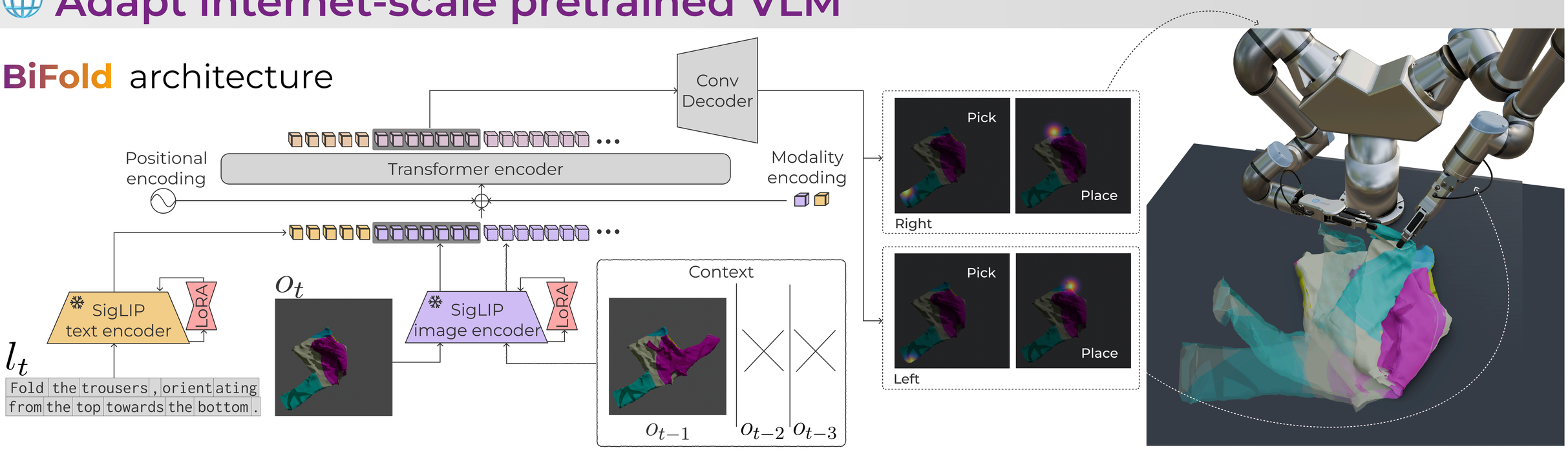
Keyframe context



Frames are redundant; Use ***keyframes!***

Adapt internet-scale pretrained VLM

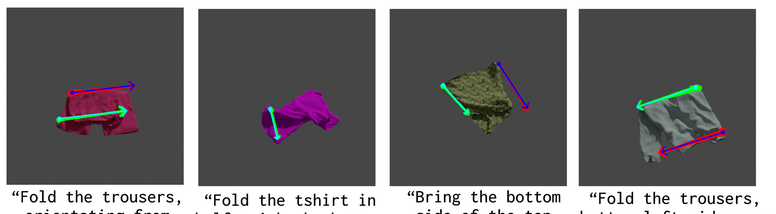
BiFold architecture



Generalizes across fabrics/textures/domains

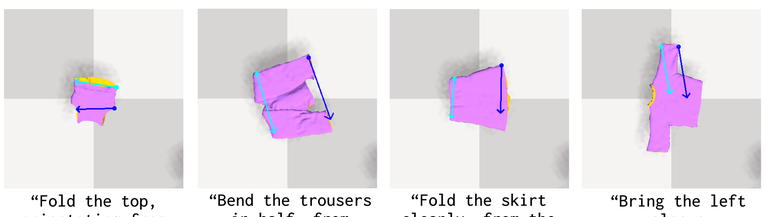
L/R Ground Truth L/R Predictions

Predictions on BiFold dataset

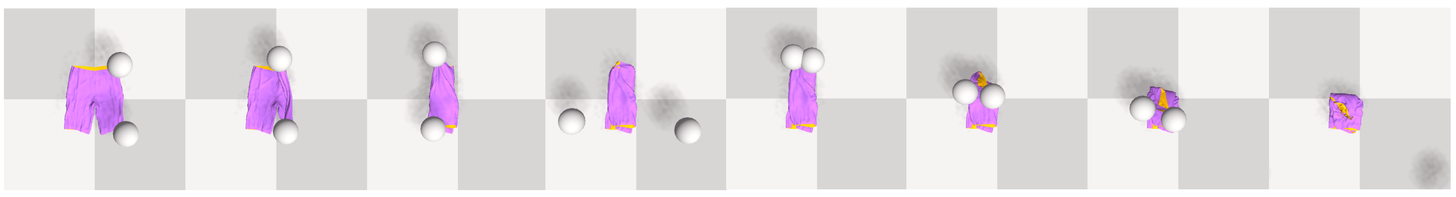


"Fold the trousers, orientating from the right towards the left." "Fold the tshirt in half, with the bottom right side overlapping the bottom left only using the right arm." "Bring the bottom side of the top towards the top side and fold them in half." "Fold the trousers, bottom left side over bottom right side."

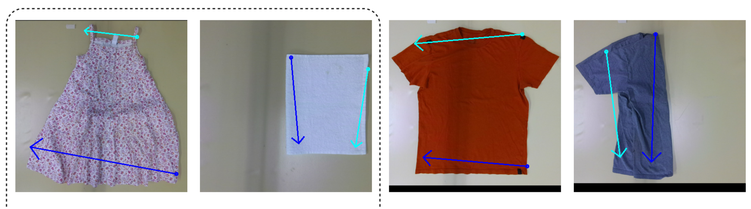
Zero-shot predictions on SoftGym



"Fold the top, orientating from the right towards the left." "Bend the trousers in half, from bottom left to bottom right." "Fold the skirt cleanly, from the left side to the right side." "Bring the left sleeve to the center."



Zero-shot predictions in real world



"Fold the waistband of the dress in half, from left to right." "Create a fold in the towel, going from top to bottom." "Bend the tshirt in half from left to right." "Fold the tshirt in half, with the top side overlapping the bottom."

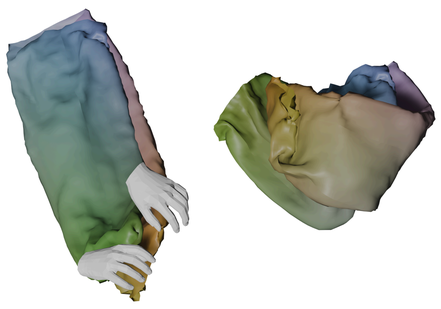
Unseen categories!

← End-to-end folding by chaining actions

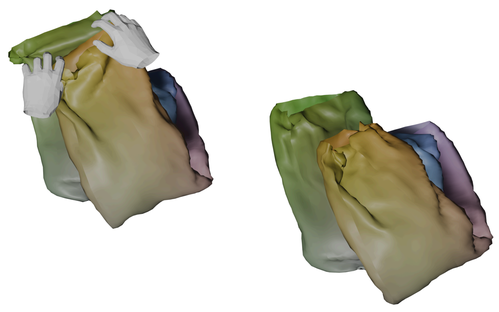
Scalable annotation pipeline

NOCS→Semantics→Language

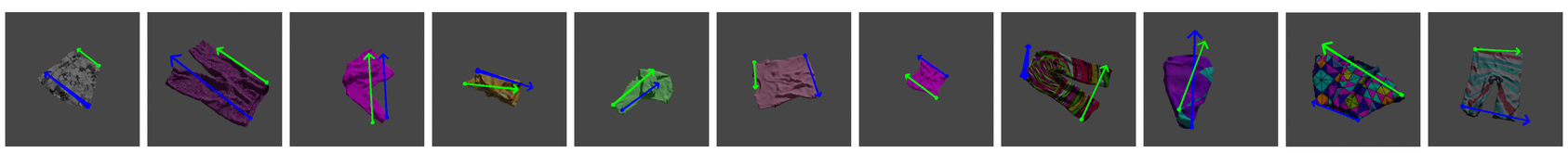
"Bring the **right side** of the trousers towards the **left side** and fold them in half."



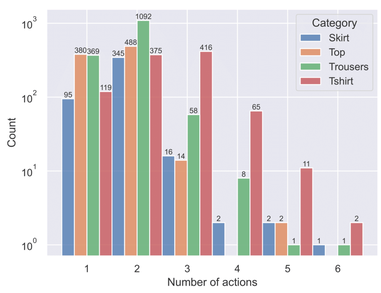
"Bend the trousers in half, from **bottom** to **top**."

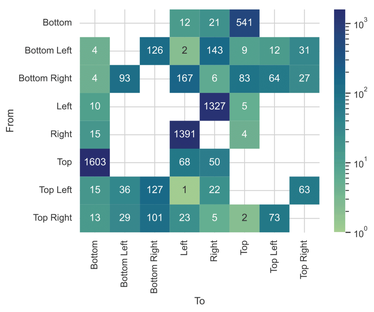


"Straighten out the **bottom** part of the trousers."



Garment	# demos	# meshes	# actions	# instructions
Skirt	464	462	857	196
Top	893	890	891	262
Trousers	1541	1539	2770	353
T-shirt	993	990	2444	254





Code, Data & Supplementary

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

ATLANTA 2025