FLEX: Full-Body Grasping Without Full-Body Grasps

Purva Tendulkar

Dídac Surís

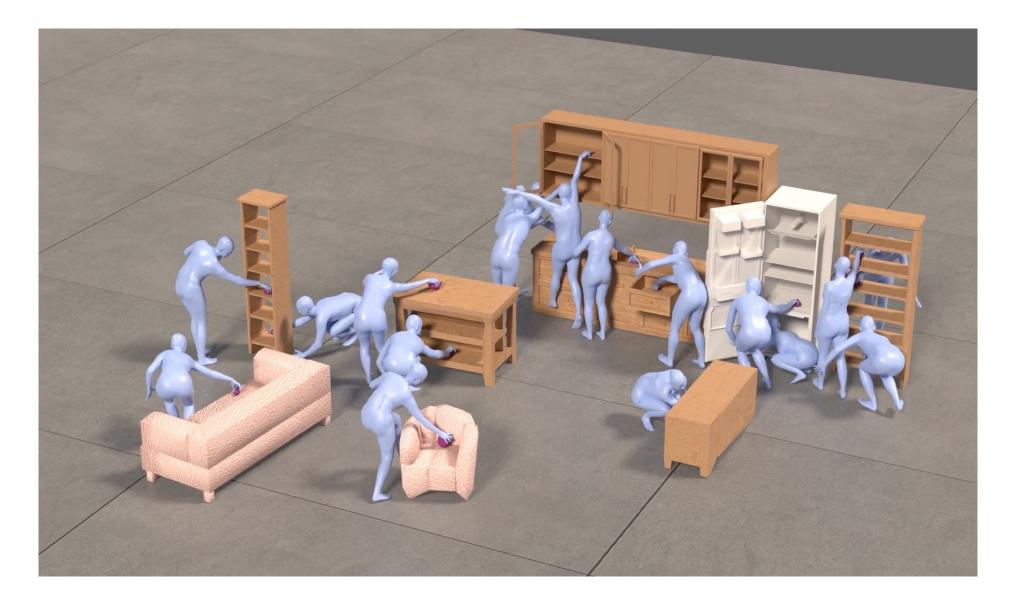Carl Vondrick
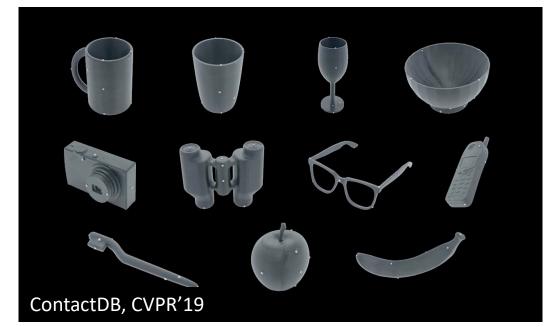
CVPR 2023

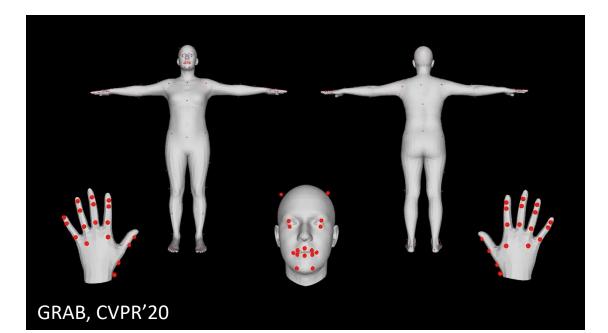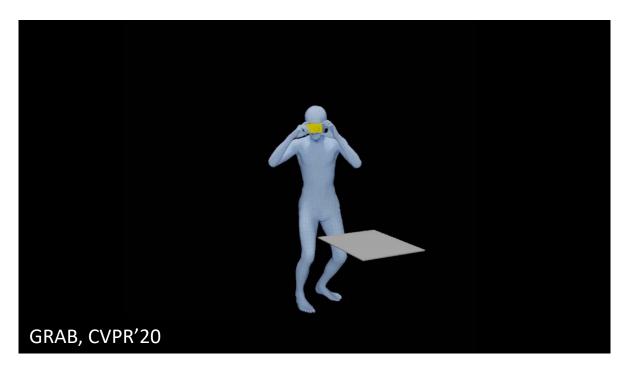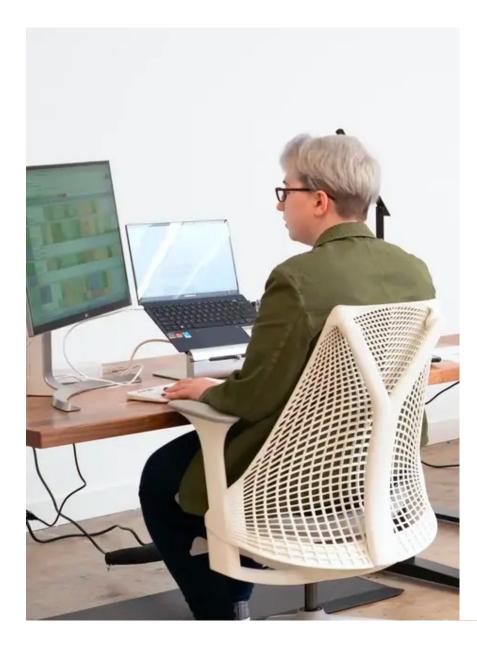COLUMBIA UNIVERSITY

Task: Full-Body Grasping in Scenes

ContactDB, CVPR'19

GRAB, CVPR'20

GRAB, CVPR'20

Failures of Existing Approaches

Failures of Existing Approaches

GOAL (CVPR'22)

Failures of Existing Approaches

GOAL (CVPR'22)   SAGA (ECCV'22)

Failures of Existing Approaches

GOAL (CVPR'22)   SAGA (ECCV'22)

Failures of Existing Approaches

Failures of Existing Approaches

GOAL (CVPR'22)   SAGA (ECCV'22)
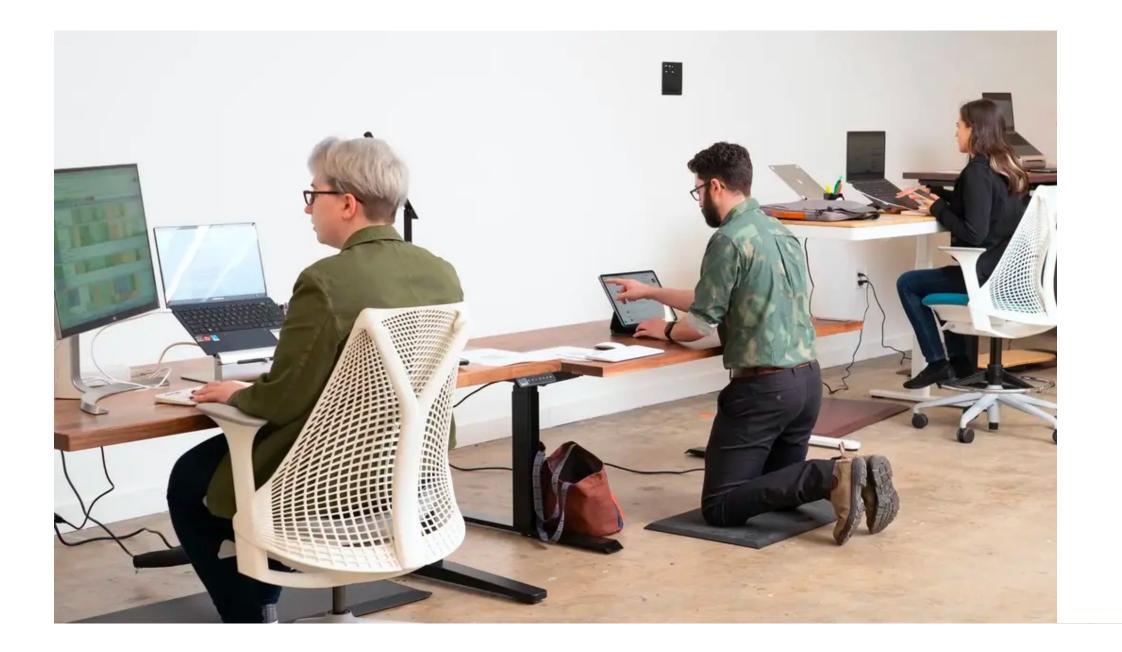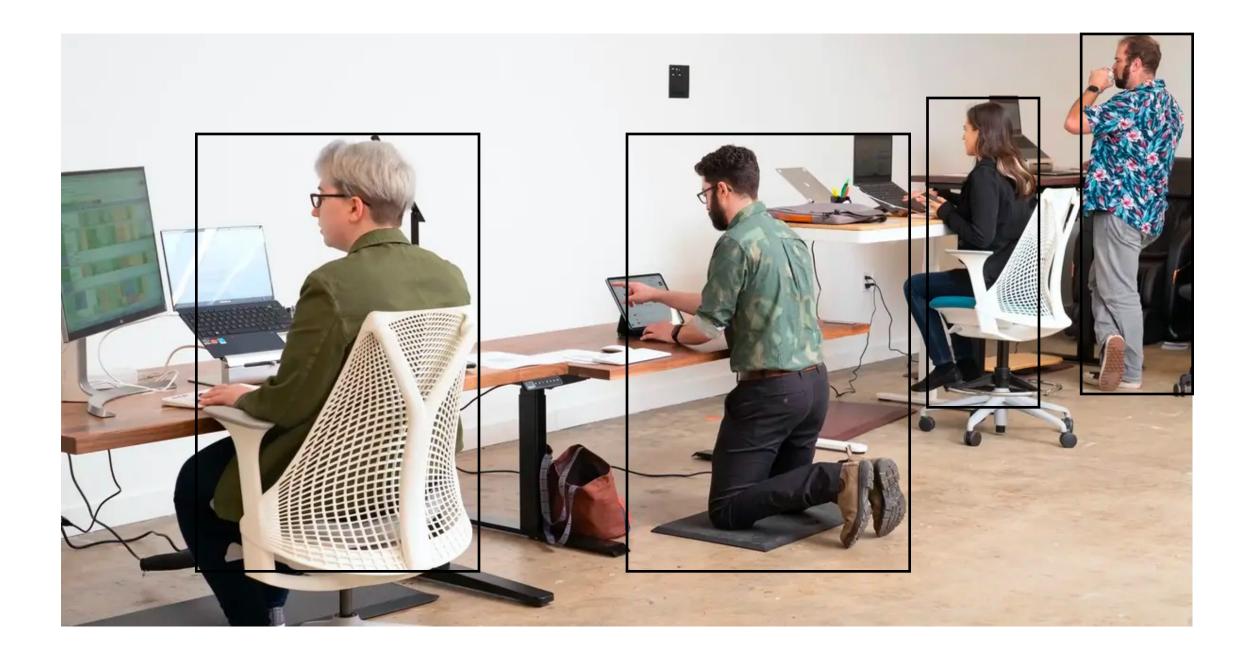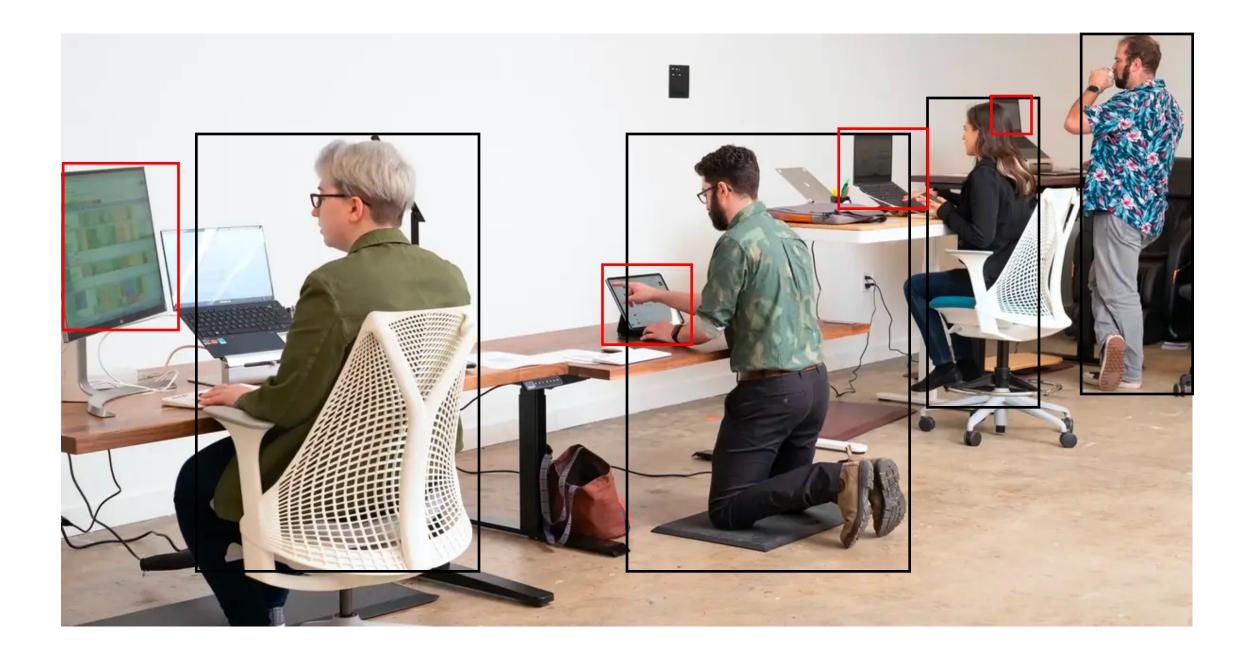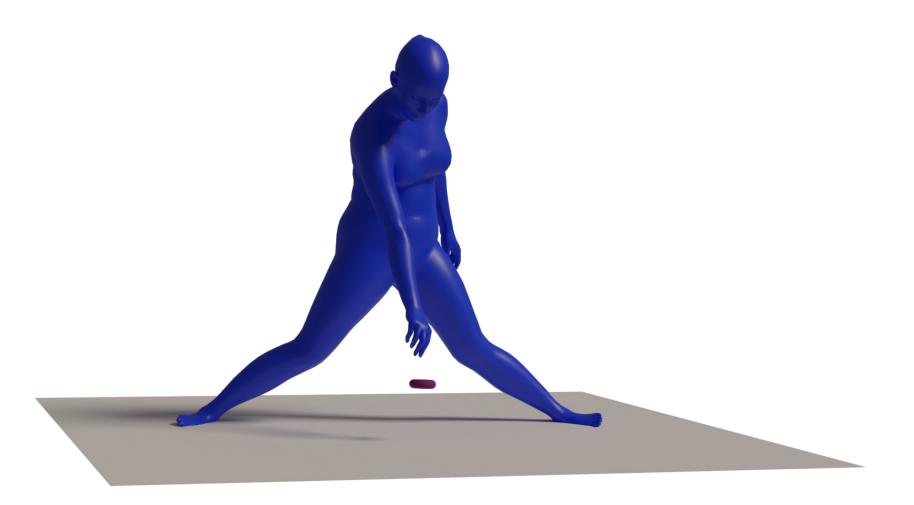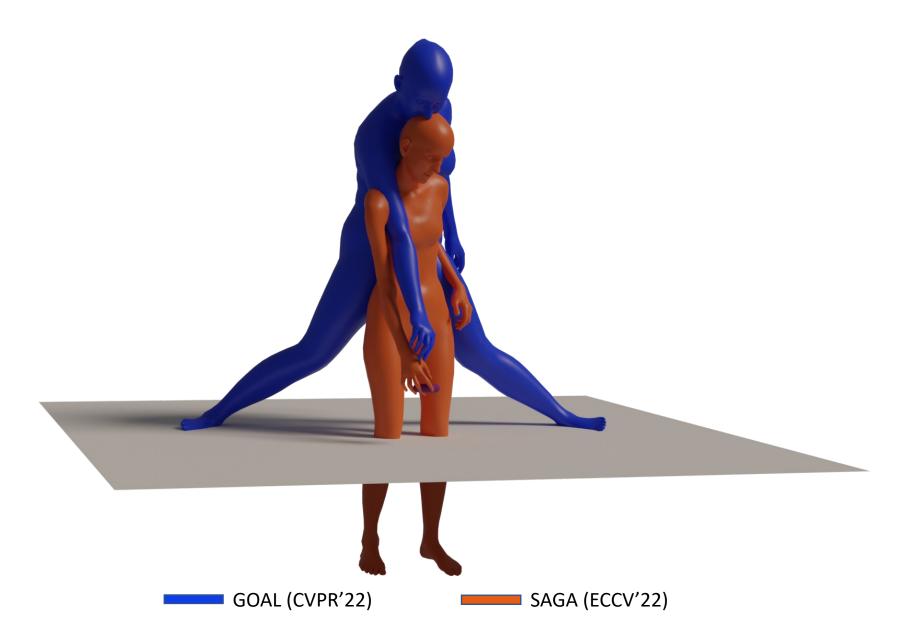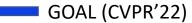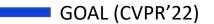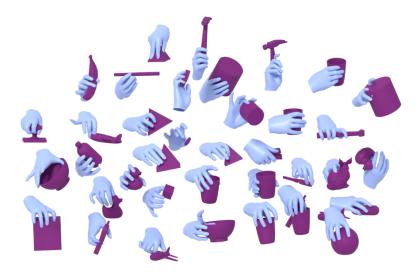
Do we *really* need full-body grasping data to solve full-body grasping?

MPI Limits

Large-scale MoCap Datasets
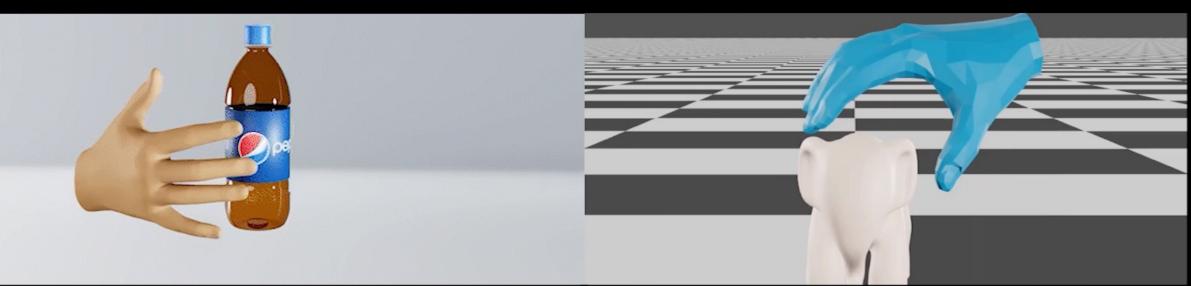
GRAB CVPR '20

GraspTTA ICCV '21

Grasp'D ECCV '22

TOCH ECCV '22

Large-scale MoCap Datasets
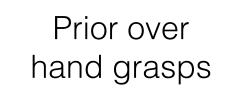
Hand Grasping
Datasets & Methods

Large-scale MoCap Datasets
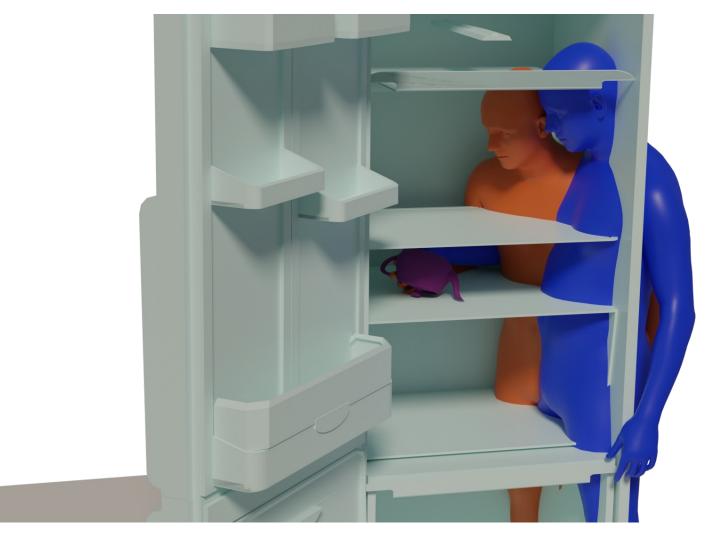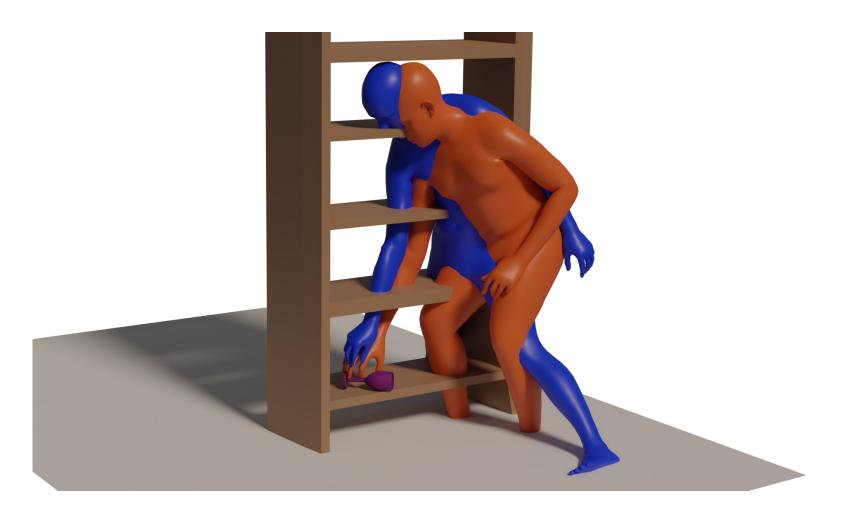
Hand Grasping
Datasets & Methods

Prior over
body poses

Prior over
hand grasps

Large-scale MoCap Datasets

Hand Grasping
Datasets & Methods

Prior over
body poses

Prior over
hand grasps

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

α, w → $\mathcal{G}$ grasp generator → $\widetilde{\theta_h, t_h, R_h}$ → MANO

↑ Object

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration
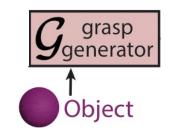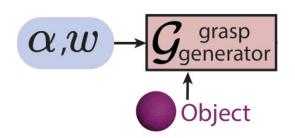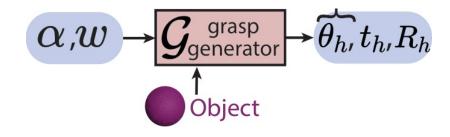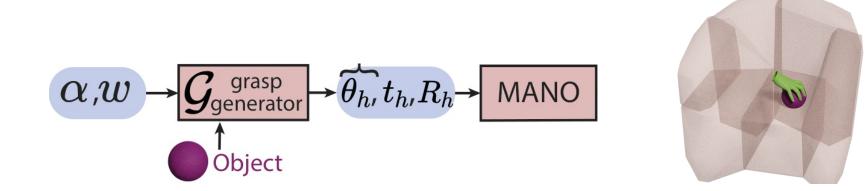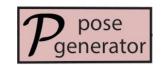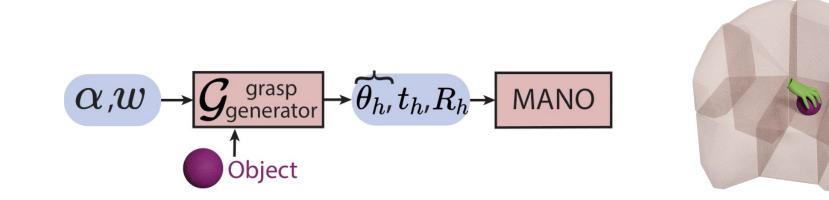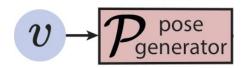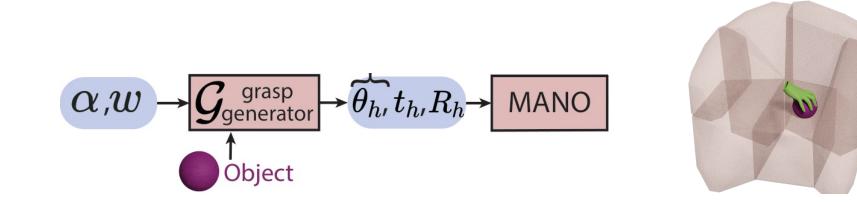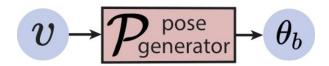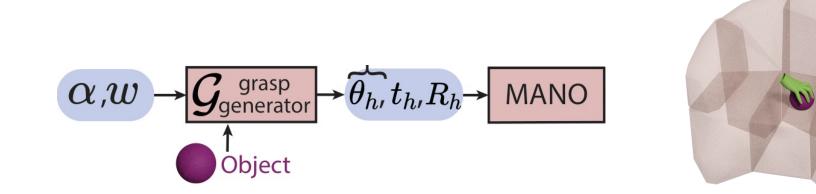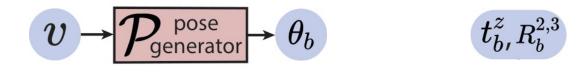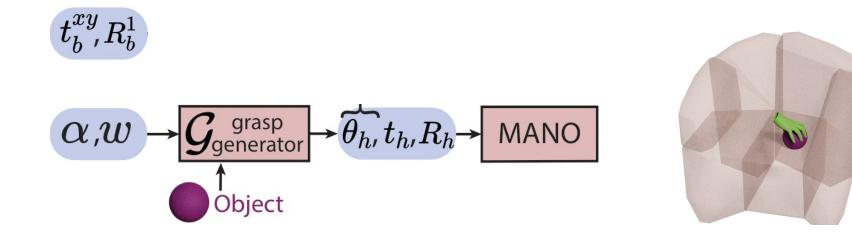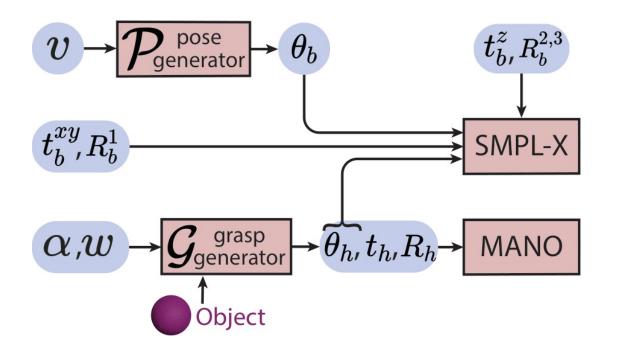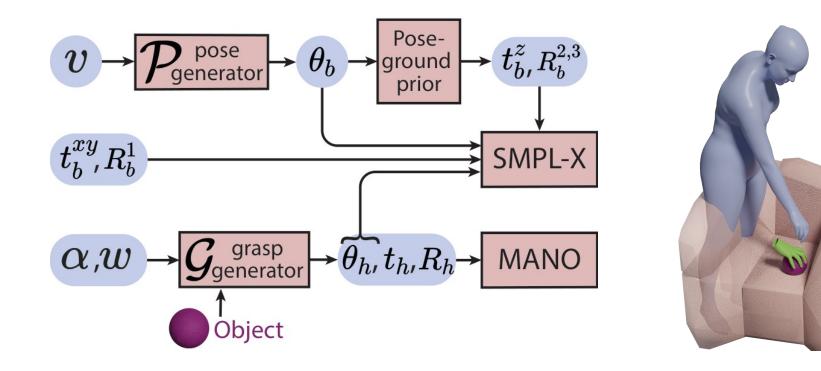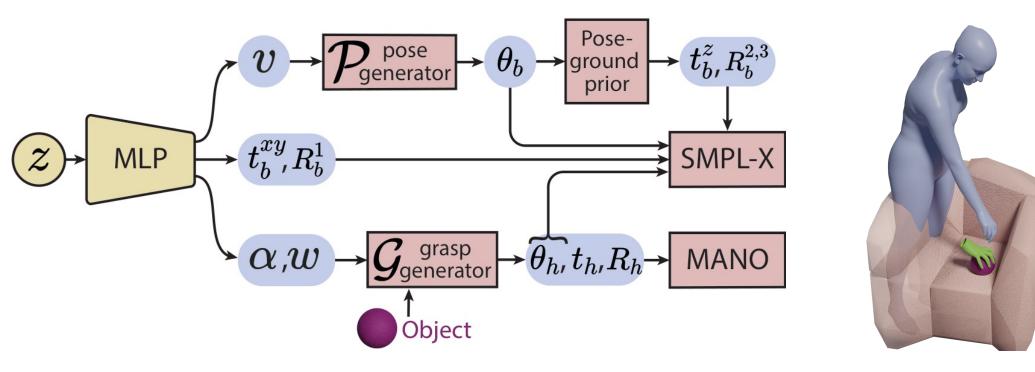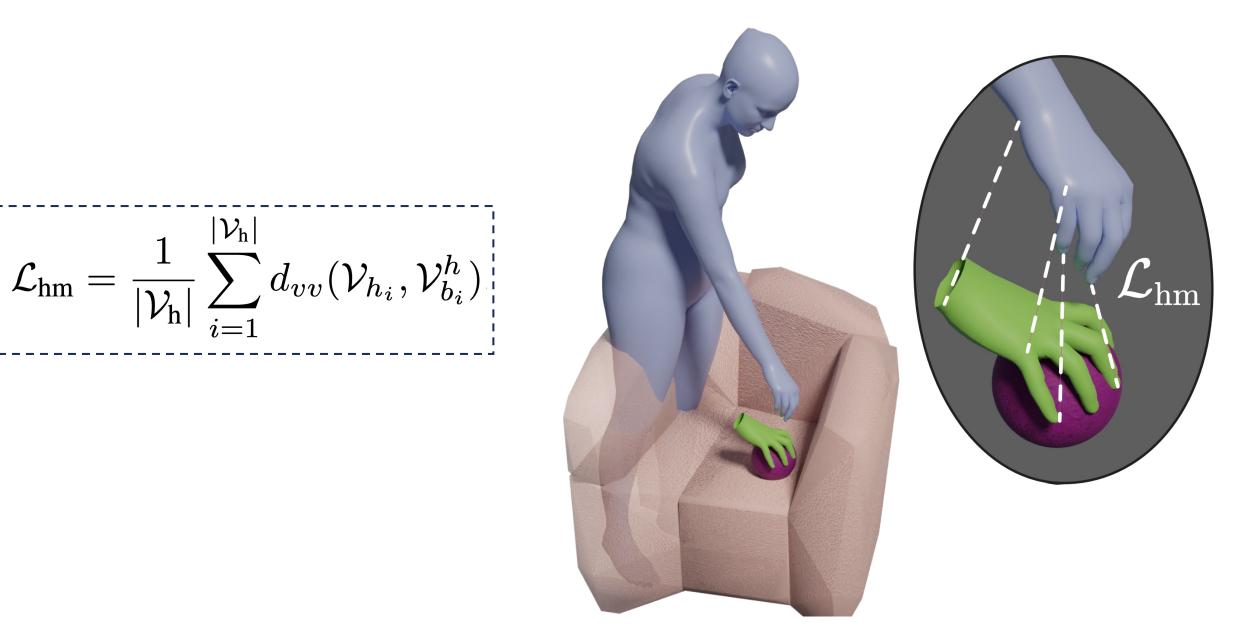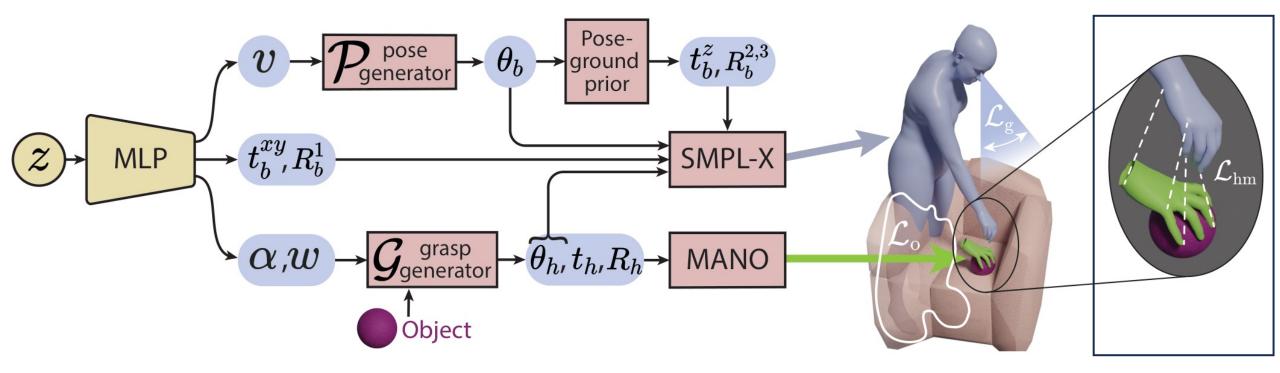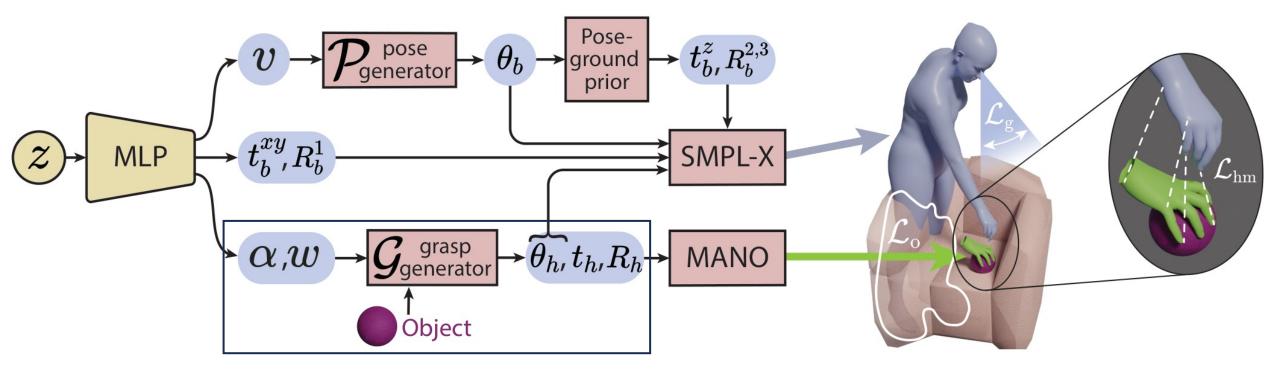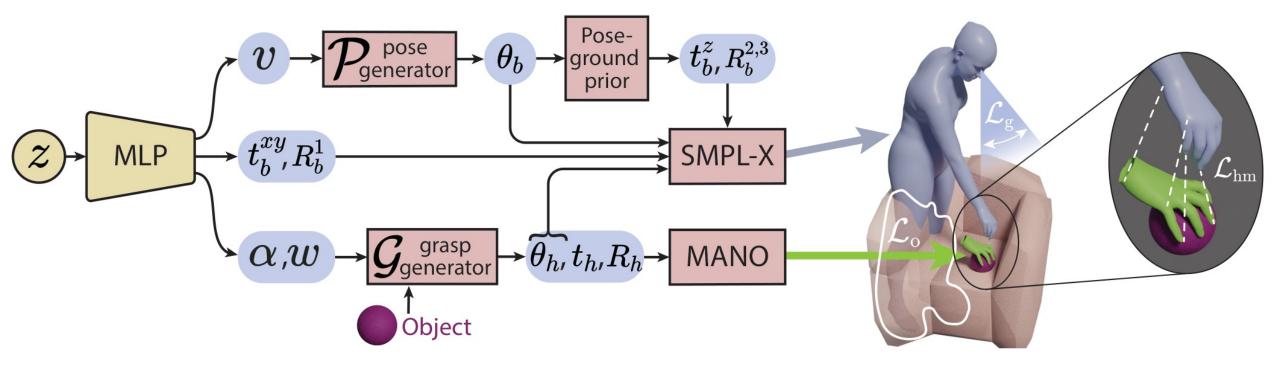
Hand Matching Loss

Hand Matching Loss

$$\mathcal{L}_{\mathrm{hm}} = \frac{1}{|\mathcal{V}_{\mathrm{h}}|} \sum_{i=1}^{|\mathcal{V}_{\mathrm{h}}|} d_{vv}(\mathcal{V}_{h_i}, \mathcal{V}_{b_i}^h)$$

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

Hand Grasp Search

Hand Grasp Search

Hand Grasp Search

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

Gaze Loss

$$\mathcal{L}_{\mathrm{g}} = \cos^{-1} \frac{V_{b \to f} \cdot V_{b \to o}}{|V_{b \to f}||V_{b \to o}|}$$

Obstacle Loss

Obstacle Loss

$$\mathcal{L}_{\mathrm{o}}^{\mathrm{inter}} = \frac{1}{|\mathcal{V}_{\mathrm{b}}|} \sum_{i=1}^{|\mathcal{V}_{\mathrm{b}}|} \left| \min \left(0, d_{\mathrm{vm}}(\mathcal{V}_{\mathrm{b}_i}, \mathcal{M}_{\mathrm{obstacle}})\right) \right|$$

Obstacle Loss

$$\mathcal{L}_{\mathrm{o}}^{\mathrm{inter}} = \frac{1}{|\mathcal{V}_{\mathrm{b}}|} \sum_{i=1}^{|\mathcal{V}_{\mathrm{b}}|} \Big| \min\left(0, d_{\mathrm{vm}}(\mathcal{V}_{\mathrm{b}_i}, \mathcal{M}_{\mathrm{obstacle}})\right) \Big|$$

Obstacle Loss

$$\mathcal{L}_{\mathrm{o}}^{\mathrm{inter}} = \frac{1}{|\mathcal{V}_{\mathrm{b}}|} \sum_{i=1}^{|\mathcal{V}_{\mathrm{b}}|} \big| \min\left(0, d_{\mathrm{vm}}(\mathcal{V}_{\mathrm{b}_i}, \mathcal{M}_{\mathrm{obstacle}})\right) \big|$$

$$\mathcal{L}_{\mathrm{o}}^{\mathrm{con}} = \frac{1}{|\mathcal{V}_{\mathrm{b}}|} \sum_{i=1}^{|\mathcal{V}_{\mathrm{b}}^{\mathrm{discon}}|} d_{\mathrm{vm}}(\mathcal{V}_{\mathrm{b}_i}^{\mathrm{discon}}, \mathcal{M}_{\mathrm{obstacle}})$$

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

Pose-Ground Prior

Pose-Ground Prior

Pose-Ground Prior

Pose-Ground Prior

- 2-Layer MLP

- MSE Regression

- Trained on 12.7 million examples of AMASS

- Frozen during inference

FLEX: Full-Body Latent EXploration

FLEX: Full-Body Latent EXploration

Dataset: ReplicaGrasp

Dataset: ReplicaGrasp

- 50 everyday objects from GRAB (e.g., wineglass, cellphone)
- 48 receptacles from ReplicaCAD (e.g., couch, drawer)
- Simulated using Habitat simulator to be in one upright and random configurations
- Total 4.8K instances

Results: Comparisons

Results: Comparisons

Results: Comparisons

Results: Comparisons

GOAL   SAGA   FLEX

Results: Diversity

Results: Diversity

Results: Diversity

Results: Diversity

Results: Diversity

Results: Quantitative

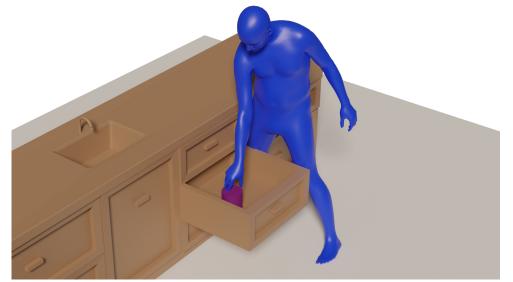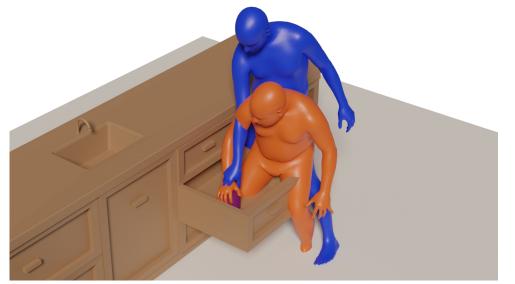| Method	ReplicaGrasp						GRAB				
	Obj Cont (%) ↑	Obj Penet (%) ↓	Obst Penet (%) ↓	Div$_{samp}$ (cm) ↑	Div$_{all}$ (cm) ↑	Ground (cm) ↓	Obj Cont (%) ↑	Obj Penet (%) ↓	Div$_{samp}$ (cm) ↑	Div$_{all}$ (cm) ↑	Ground (cm) ↓
GOAL [28]	1.14	2.14	6.87	0.20	37.89	5.31	1.62	3.50	0.11	7.87	3.56
SAGA [29]	1.21	**0.29**	7.27	1.10	43.84	9.40	**2.19**	3.43	1.06	17.35	2.39
FLEX (ours)	**2.20**	2.49	**0.53**	**10.37**	**69.78**	**0.00**	1.63	**2.73**	**24.94**	**46.46**	**0.00**

Results: Quantitative

Method	ReplicaGrasp						GRAB				
	Obj Cont (%) ↑	Obj Penet (%) ↓	Obst Penet (%) ↓	Div_{samp} (cm) ↑	Div_{all} (cm) ↑	Ground (cm) ↓	Obj Cont (%) ↑	Obj Penet (%) ↓	Div_{samp} (cm) ↑	Div_{all} (cm) ↑	Ground (cm) ↓
GOAL [28]	1.14	2.14	6.87	0.20	37.89	5.31	1.62	3.50	0.11	7.87	3.56
SAGA [29]	1.21	**0.29**	7.27	1.10	43.84	9.40	**2.19**	3.43	1.06	17.35	2.39
FLEX (ours)	**2.20**	2.49	**0.53**	**10.37**	**69.78**	**0.00**	1.63	**2.73**	**24.94**	**46.46**	**0.00**

FLEX outperforms GOAL and SAGA on the dataset they were trained on, without even using it!

Results: Quantitative

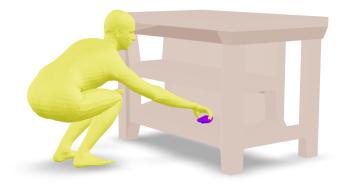| Method	ReplicaGrasp						GRAB				
	Obj Cont (%) ↑	Obj Penet (%) ↓	Obst Penet (%) ↓	Div$_{samp}$ (cm) ↑	Div$_{all}$ (cm) ↑	Ground (cm) ↓	Obj Cont (%) ↑	Obj Penet (%) ↓	Div$_{samp}$ (cm) ↑	Div$_{all}$ (cm) ↑	Ground (cm) ↓
GOAL [28]	1.14	2.14	6.87	0.20	37.89	5.31	1.62	3.50	0.11	7.87	3.56
SAGA [29]	1.21	**0.29**	7.27	1.10	43.84	9.40	**2.19**	3.43	1.06	17.35	2.39
FLEX (ours)	**2.20**	2.49	**0.53**	**10.37**	**69.78**	**0.00**	1.63	**2.73**	**24.94**	**46.46**	**0.00**

FLEX also outperforms both baselines on ReplicaGrasp.

Results: Qualitative

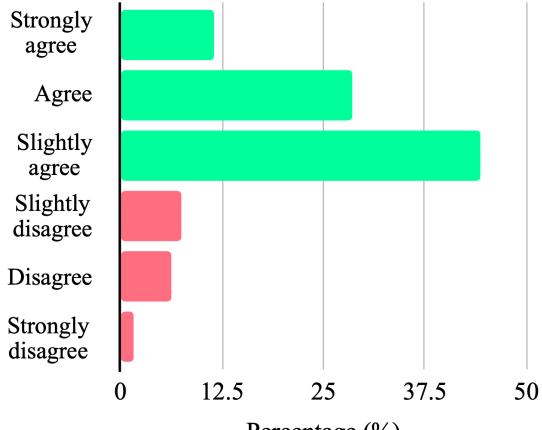Claim: **"This pose is natural and realistic for grasping the object."**.

Note: the scene may not show up in the preview, but will be visible once you accept. If it is not visible upon acceptance, please select option "7. The scene is not showing".

○ 1. Strongly disagree   ○ 2. Disagree   ○ 3. Slightly disagree   ○ 4. Slightly agree   ○ 5. Agree   ○ 6. Strongly agree   ○ 7. The scene is not showing
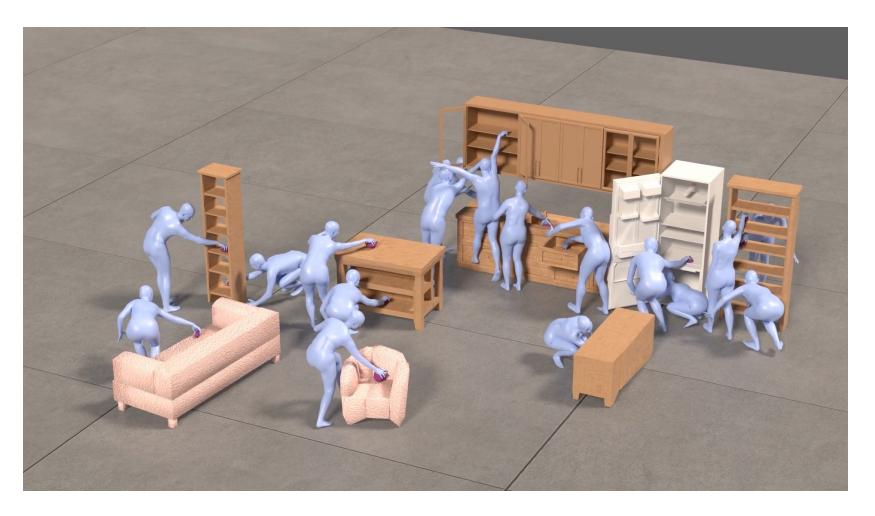
You must select one of the options above. Do not leave unselected.

(Optional) Provide a reason for your selection

Submit

Conclusion

- **FLEX** – a method to generate full-body grasps for objects in everyday scenes while avoiding obstacles.

- **ReplicaGrasp** – a dataset of 50 objects spawned in everyday scenes in a variety of configurations

- Demonstrate full-body grasping results that **outperform baselines** quantitatively and qualitatively.