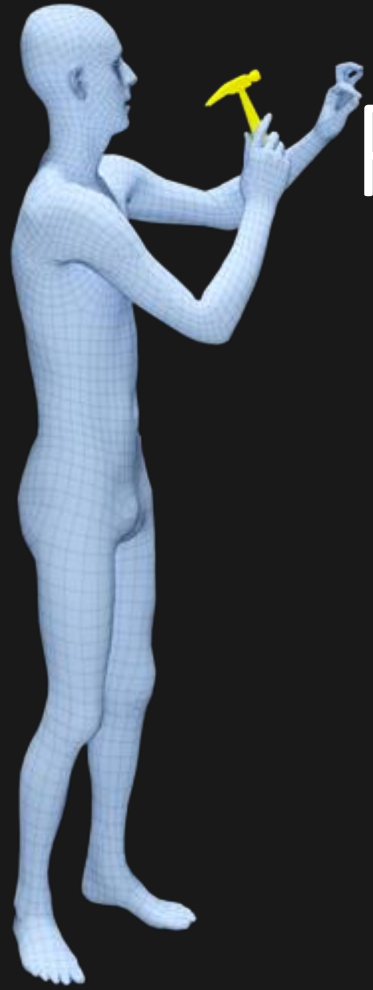
Towards… 3D Human-centric Perception and Synthesis



Dimitris Tzionas

University of Amsterdam

Research Direction

Understand how real **people:**

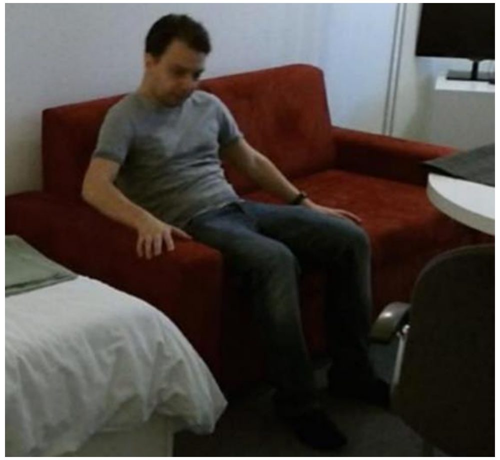
- **Look**
- **Move**
- **Interact**

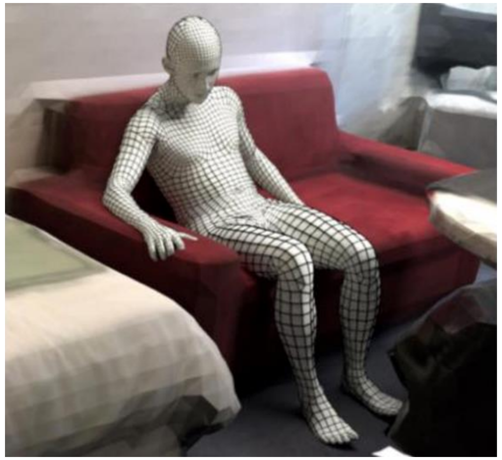
Long term goal:

- Develop ***human-centered AI***
- Assistive AI that:
    - **Perceive** humans in their environment
    - Understand their **behavior**
    - Help them achieve their **goals**

Research Direction

Holistic 3D Scene Understanding:
- Modelling how people, objects, spaces look
- Estimating their 3D shape and pose
- Inferring their semantics and spatial relationships
- Employing all above information to reason about:
  - how people act
  - how they interact with objects & people
  - how they perform tasks





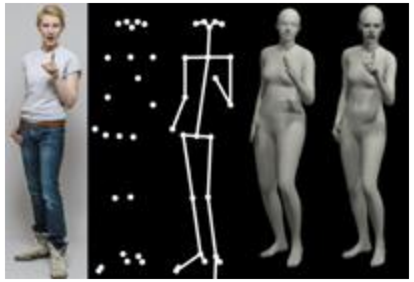
Effortless for ***humans*** and ***animals***

Challenging for ***computers*** → Challenges exist at ***all*** levels of abstraction
- Ill-posed 3D inference from a 2D image
- Semantic interpretation

Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17



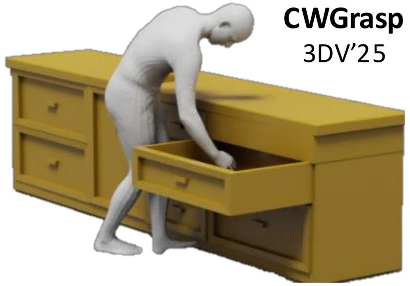
SMPL-X & SMPLify-X
CVPR’19

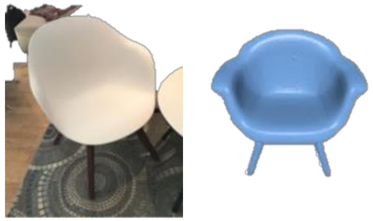
PROX
ICCV’19

GRAB ECCV’20

ARCTIC CVPR’23

CWGrasp
3DV’25





SDFit
arXiv 2024

3D Human Understanding

- in ‘isolation’
- in interaction …

… using their

- hands
- whole body
  (body + face + hands)

SMPL Body Model

SMPL specs:

$$\underset{\text{model}}{M}(\underset{\text{pose}}{\theta}, \underset{\text{shape}}{\beta}) =$$

Look like real people
Move like real people
Small number of parameters
Easy to animate
Easy to fit to data

SMPL Model

[1] - Loper et al. *“SMPL: A Skinned Multi-Person Linear Model”*. ToG/SIGA 2015

MANO Hand Model

3dMD Scanner

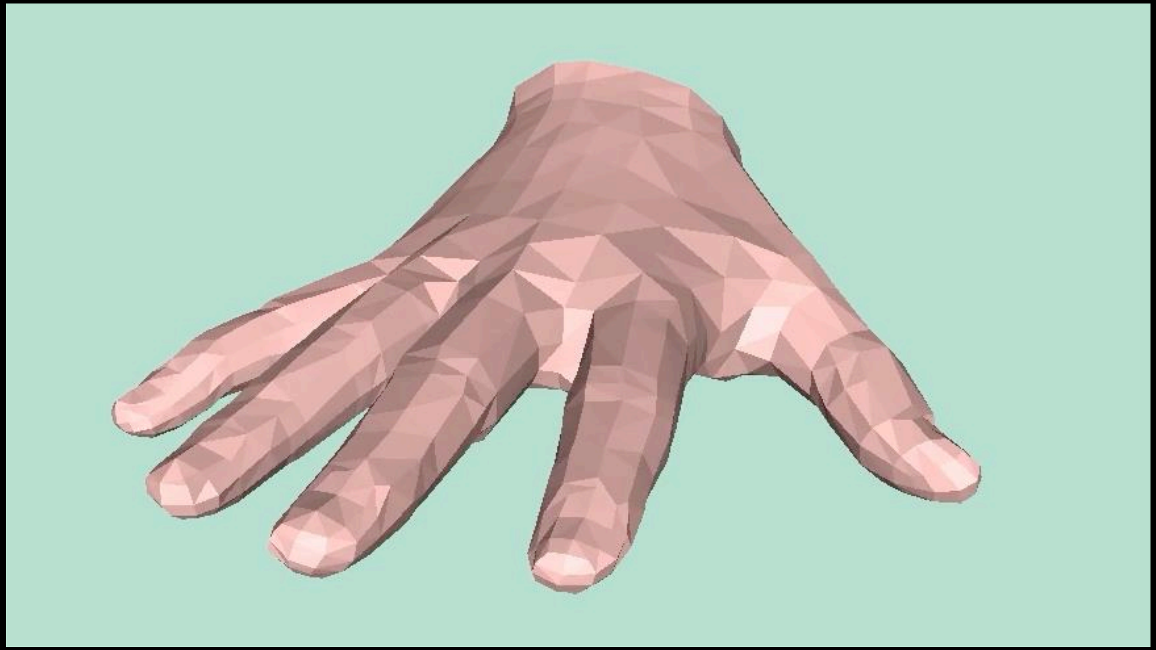
Raw images (5 views) → 3D mesh

MANO Hand Model

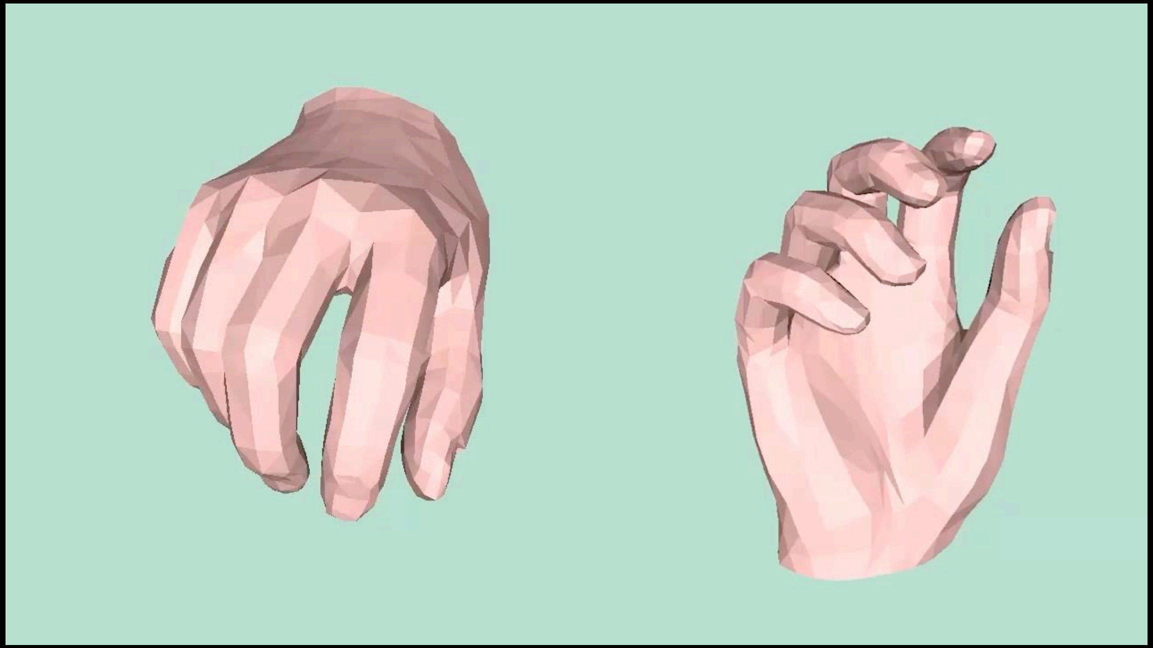
Scans

Registrations

MANO Hand Model



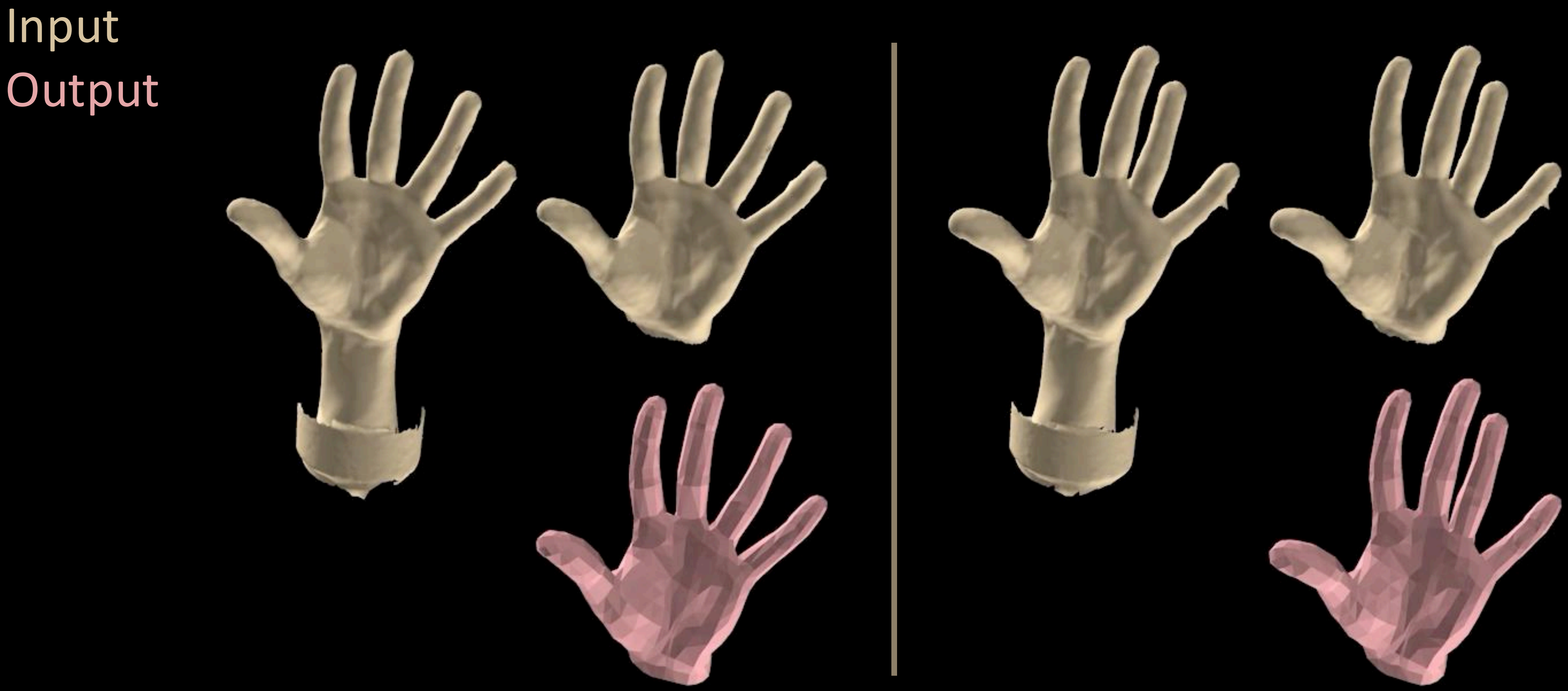
Shape Space (PCA)



Pose Space (PCA)

Diff. Background → Diff. PCA Component

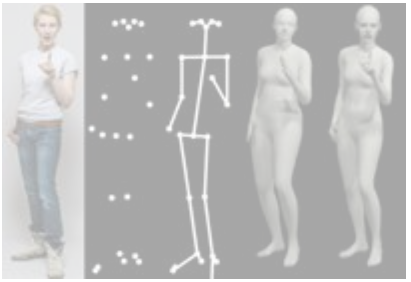
MANO Hand Model



Research Map

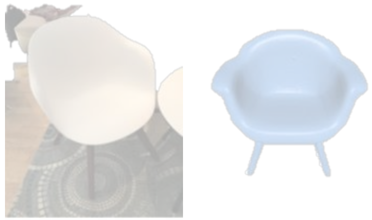
MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17



SMPL-X & SMPLify-X
CVPR'19

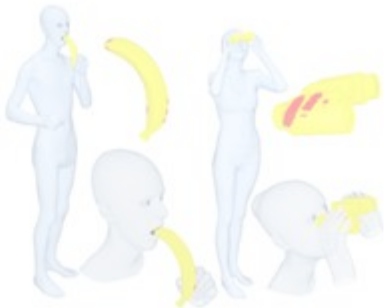
PROX
ICCV'19



SDFit
arXiv 2024

GRAB
ECCV'20

ARCTIC
CVPR'23



CWGrasp
3DV'25

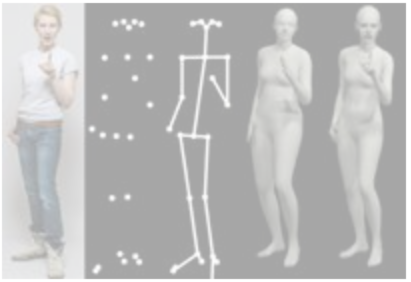
Personalized
3D Hand Mesh

Automatically

Research Map

MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17

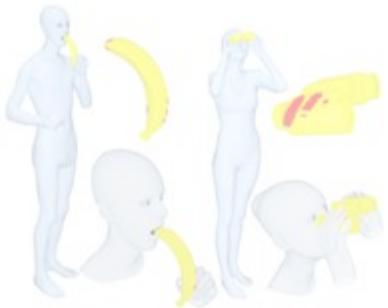


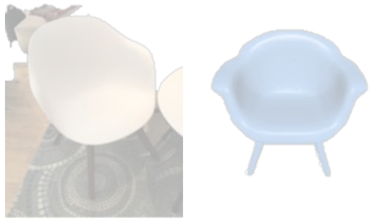
SMPL-X & SMPLify-X
CVPR'19

PROX
ICCV'19

GRAB
ECCV'20

ARCTIC
CVPR'23





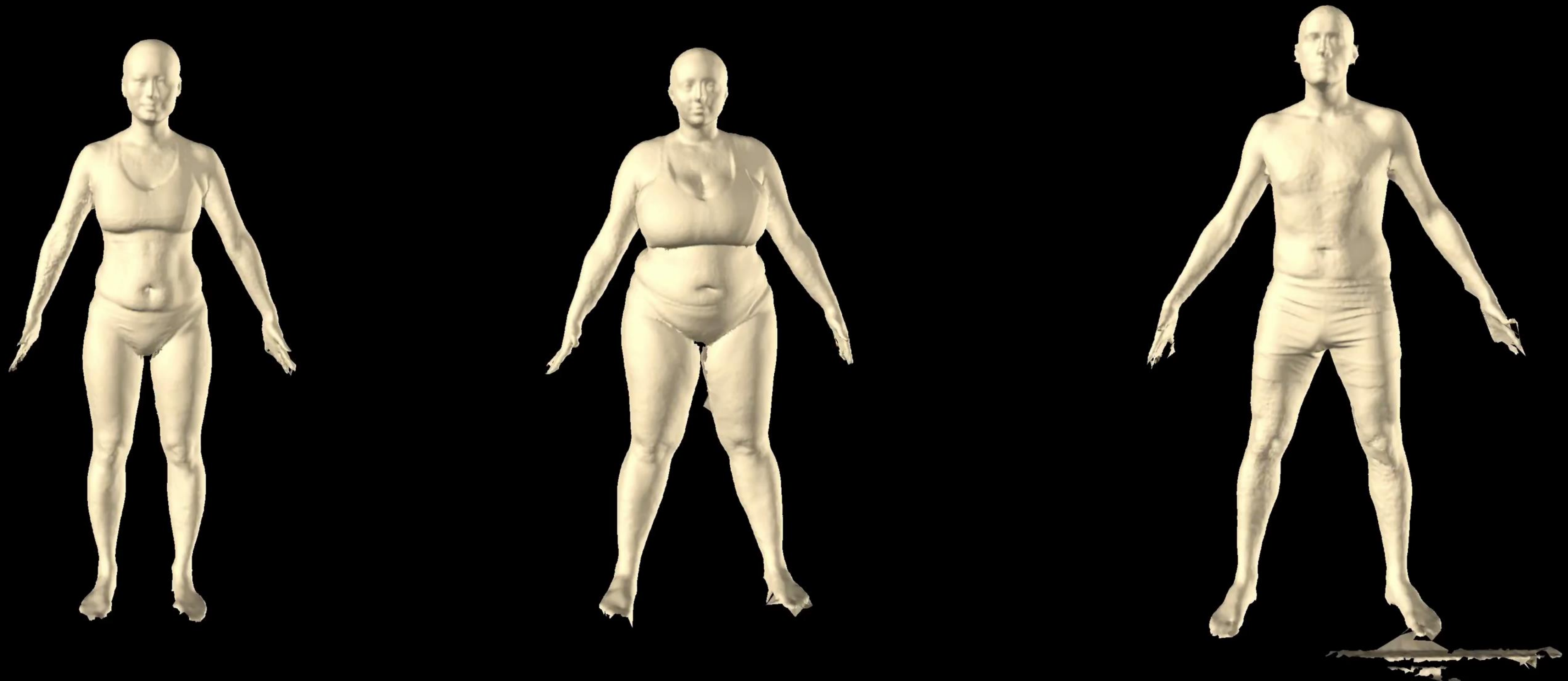
SDFit
arXiv 2024

CWGrasp
3DV'25

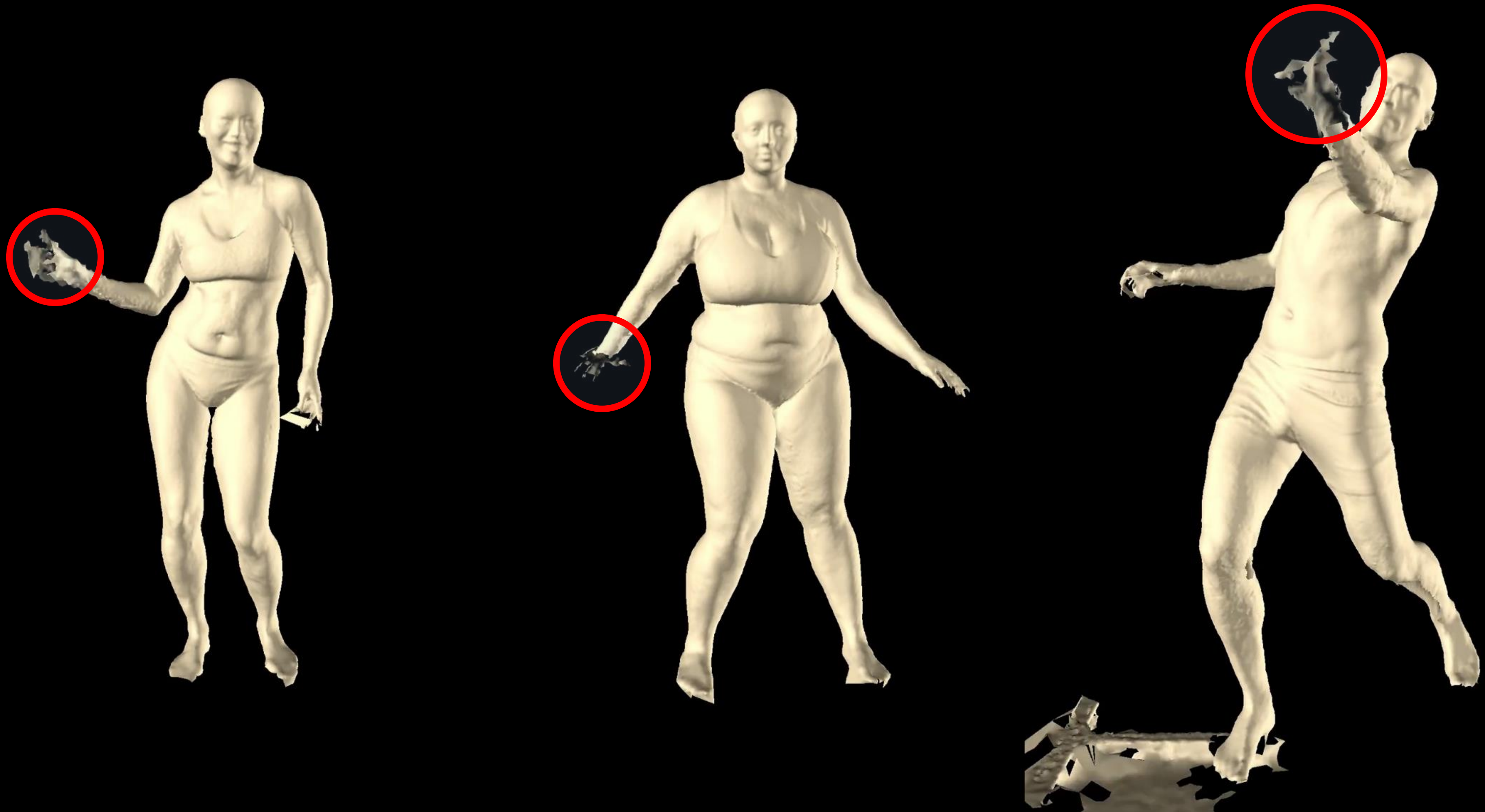
Hands:
Only part of the story

SMPL+H: Body + Hands

SMPL+H: Body + Hands



SMPL+H: Body + Hands



SMPL+H: Body + Hands

SMPL+H: Body + Hands

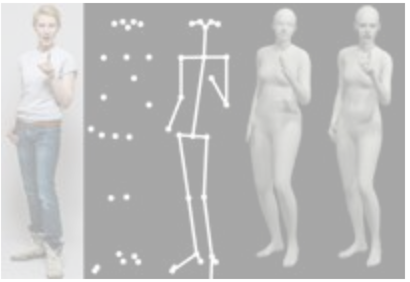
Input

Output

Research Map

MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17

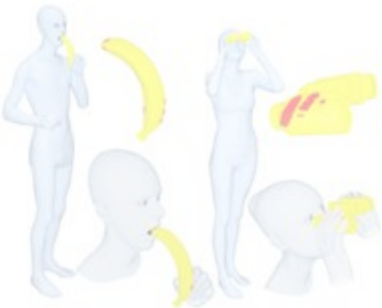


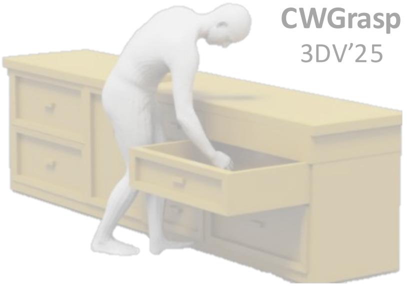
SMPL-X & SMPLify-X
CVPR'19

PROX
ICCV'19

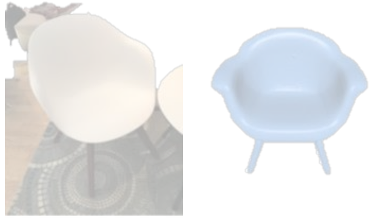
GRAB
ECCV'20

ARCTIC
CVPR'23





CWGrasp
3DV'25



SDFit
arXiv 2024

Personalized
Body + Hands

from
3D/4D Scans

Research Map

MANO
SIGGRAPH-Asia'17

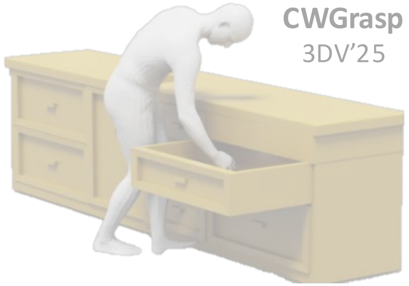
SMPL+H
SIGGRAPH-Asia'17

SMPL-X & SMPLify-X
CVPR'19

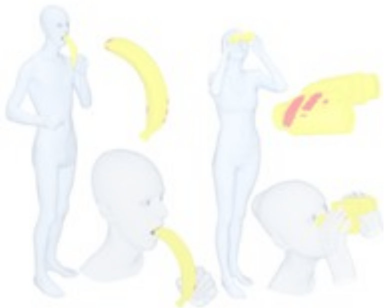
PROX
ICCV'19

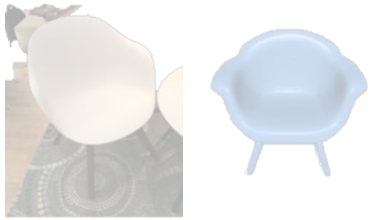
GRAB
ECCV'20

ARCTIC
CVPR'23



CWGrasp
3DV'25





SDFit
arXiv 2024

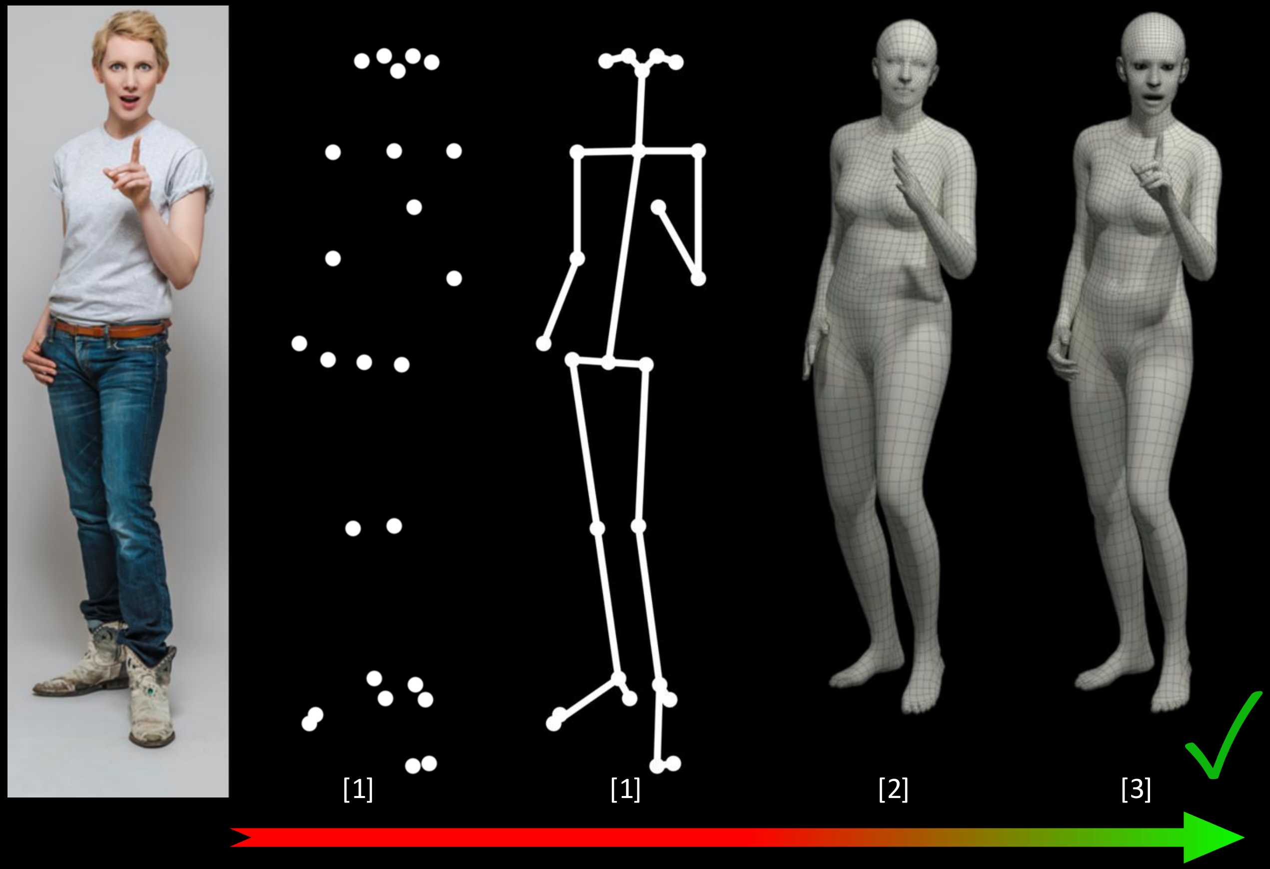
Personalized
Body + Hands (+ Face)

from
a Single RGB Image

SMPL-X: Body + Face + Hands

HOMUNCULUS

SMPL-X: Body + Face + Hands

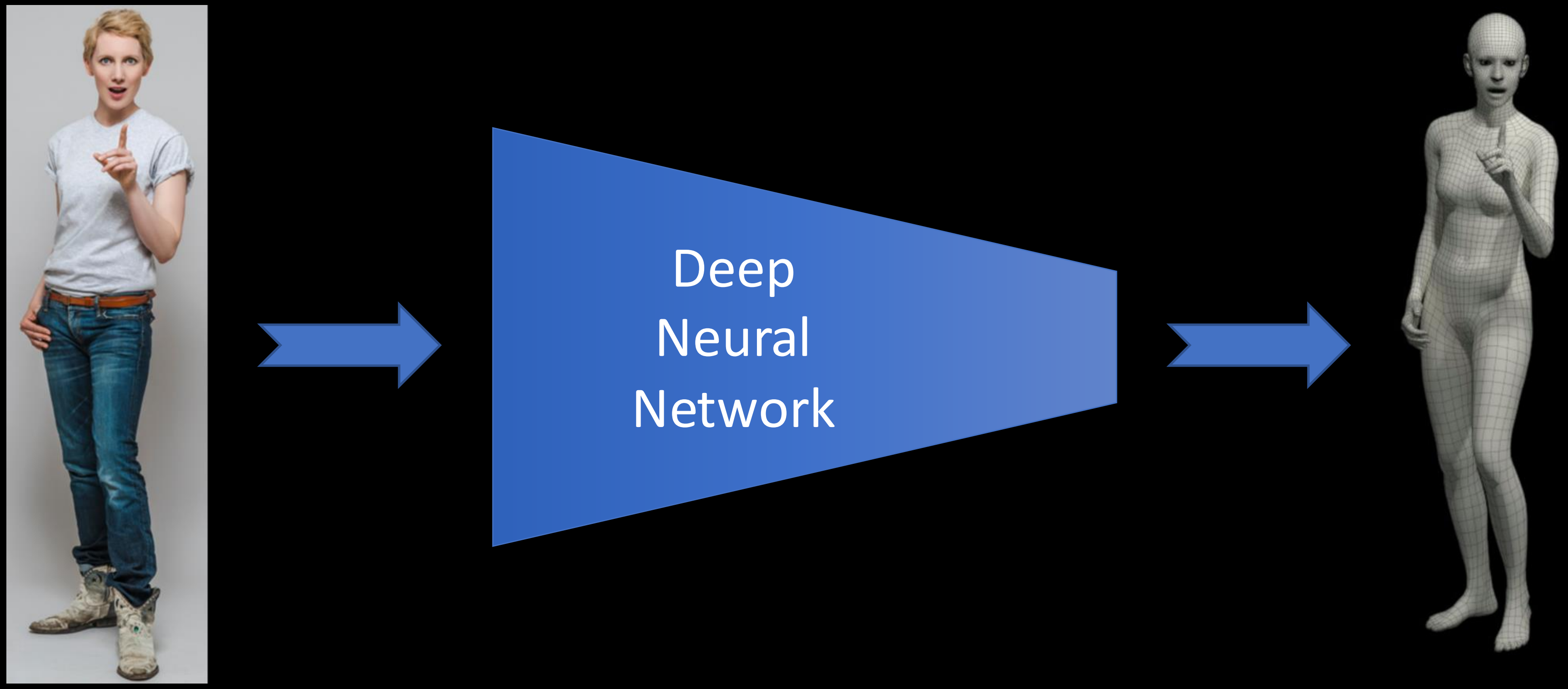


SMPL-X

SMPL eXpressive

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh *Real-time multi-person 2D pose estimation using part affinity fields* CVPR 2017
[2] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black *SMPL: A skinned multi-person linear model* SIGA 2015
[3] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas and M. J. Black *Expressive Body Capture: 3D Hands, Face, and Body from a Single Image* CVPR 2019
[4] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, M.J. Black *Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image* CVPR 2019

SMPL-X: Body + Face + Hands



Goal: Train a DNN to directly map RGB pixels to SMPL-X

Problem: No existing training data !

Impossible to manually annotate full-3D bodies !

SMPL-X: Body + Face + Hands

J_{2D-est}

Bottom-Up [1]

Match !

Projection Camera K

$\Pi_K(J_{3D})$

Top-Down

$$E_J = \| J_{2D-est} - \Pi_K(J_{3D}) \|_2^2$$

Solution: Optimization methods need no training

Proxy 2D joints are easy to annotate / detect

[1] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. *Real-time multi-person 2D pose estimation using part affinity fields*. CVPR, 2017

SMPL-X: Body + Face + Hands

Objective Function to Optimize – SMPLify-X

Whole-Body Pose

Whole-Body Shape

Facial Express.

Estim. Camera

Estim. 2D Joints

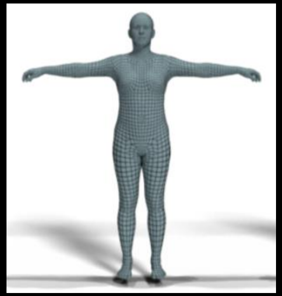
$$E(\theta, \beta, \psi) =$$

$$E_J(\theta, \beta, K, J_{est}) +$$ 2D Joints data term

$$\lambda_{\theta_b} E_{\theta_b}(\theta_b) + \lambda_{\theta_f} E_{\theta_f}(\theta_f) + \lambda_{\theta_h} E_{\theta_h}(\theta_h) + \lambda_{\theta_\alpha} E_{\theta_\alpha}(\theta_b) +$$ "Standard" Priors

$$\lambda_\beta E_\beta(\beta) + \lambda_\varepsilon \mathrm{E}_\varepsilon(\psi) +$$ Shape & Expression priors

$$\lambda_P \mathrm{E}_P(\theta, \beta, \psi)$$ Self-penetration penalty



$M(\beta, \theta, \psi) \in \mathbb{R}^{10475\text{x}3}$

ψ Facial Expression

Body Pose: $\theta = [\theta_b, \theta_f, \theta_h]$

θ_f Jaw Pose

θ_h *Hand Pose*

θ_b Main Body Pose

β Body Shape

SMPL-X: Body + Face + Hands

Optimization-based Fitting

OpenPose [1]

SMPLify-X [2]
Convergence Visualization

[1] Cao et al. TPAMI'19
[2] Pavlakos et al. CVPR'19

SMPL-X: Body + Face + Hands

Research Map

MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17

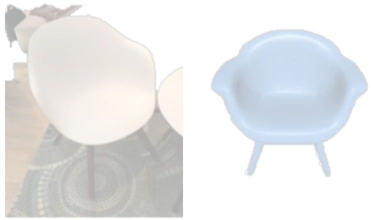
SMPL-X & SMPLify-X
CVPR'19

PROX
ICCV'19

Humans in 'Isolation'

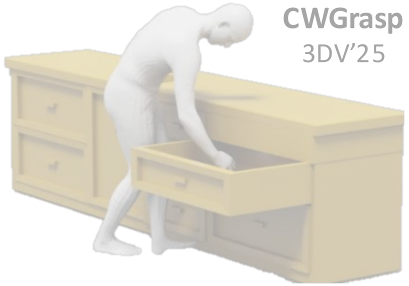
GRAB
ECCV'20

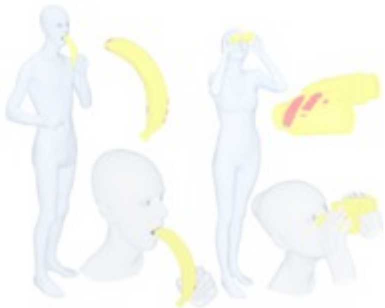
ARCTIC
CVPR'23



SDFit
arXiv 2024

CWGrasp
3DV'25

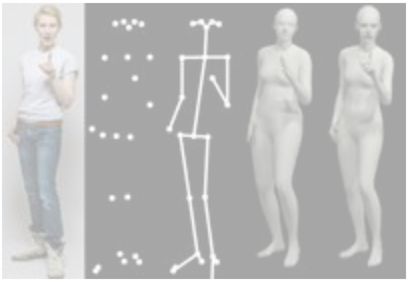




Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17



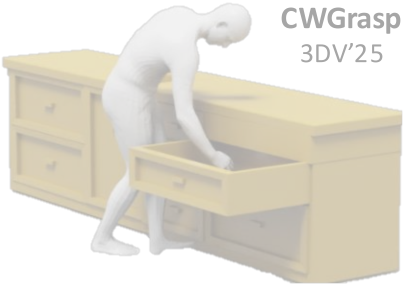
SMPL-X & SMPLify-X
CVPR’19

PROX
ICCV’19

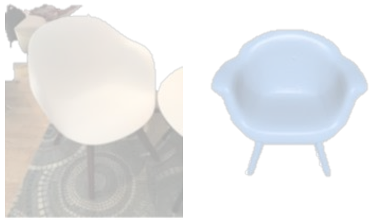
Humans in ‘Interaction’

GRAB
ECCV’20

ARCTIC
CVPR’23



CWGrasp
3DV’25

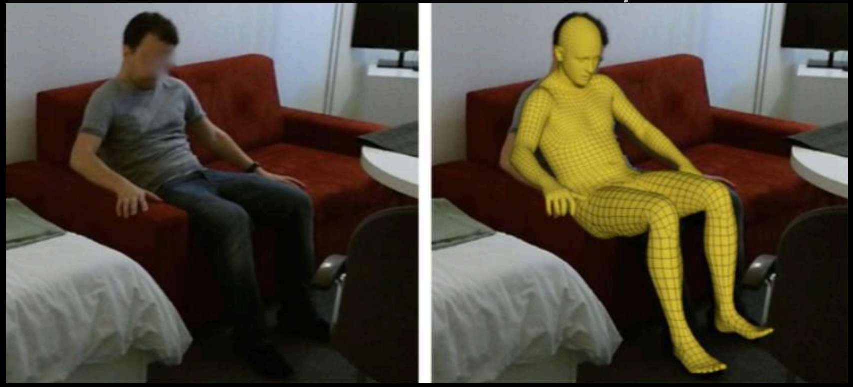


SDFit
arXiv 2024

PROX: Human Scene Interaction

Reference RGB

SMPLify-X
Overlay on RGB



PROX: Human Scene Interaction

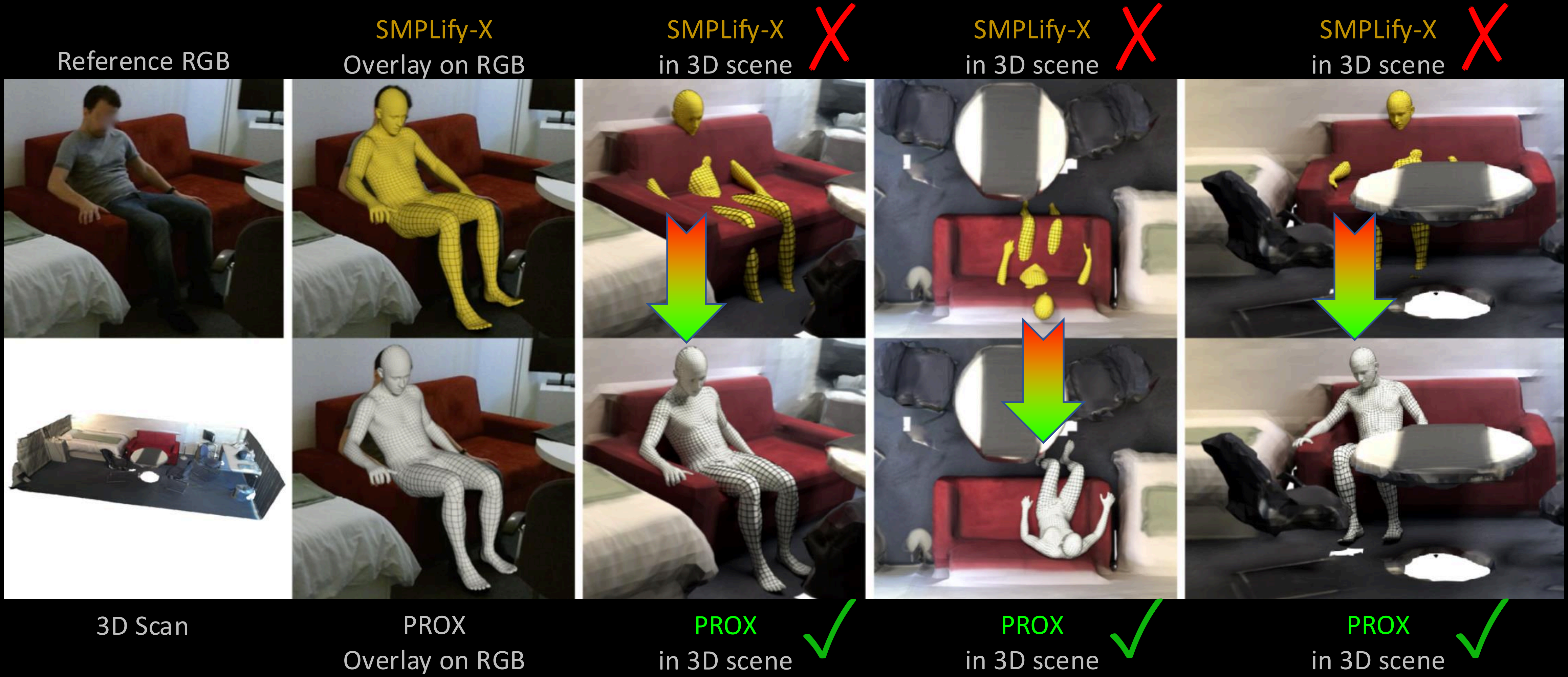
Reference RGB | SMPLify-X Overlay on RGB | SMPLify-X in 3D scene | SMPLify-X in 3D scene | SMPLify-X in 3D scene

3D Scan

Contact constraints:

- Object cannot *inter-penetrate*
- Interaction means *proximity*

PROX: Human Scene Interaction



PROX: Human Scene Interaction

Penetrations

- 3D Distance Field @ room level

Contact:

- Manually annotate likely contact vertices V_P
- Encourage close proximity between: V_P & *Scene* M_S
- If vertices V_P are "close by" in distance & orientation

[1] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas and M. J. Black. *Expressive Body Capture: 3D Hands, Face, and Body from a Single Image*. CVPR 2019

PROX: Human Scene Interaction

Whole-Body Pose — Whole-Body Shape — Facial Express. — Estim. Camera — Estim. 2D Joints — 3D Scene Mesh

$$E(\theta, \beta, \psi) =$$

$$E_J(\theta, \beta, K, J_{est}) +$$ 2D Joints data term

$$\lambda_{\theta_b} E_{\theta_b}(\theta_b) + \lambda_{\theta_f} E_{\theta_f}(\theta_f) + \lambda_{\theta_h} E_{\theta_h}(\theta_h) + \lambda_{\theta_\alpha} E_{\theta_\alpha}(\theta_b) +$$ "Standard" Priors

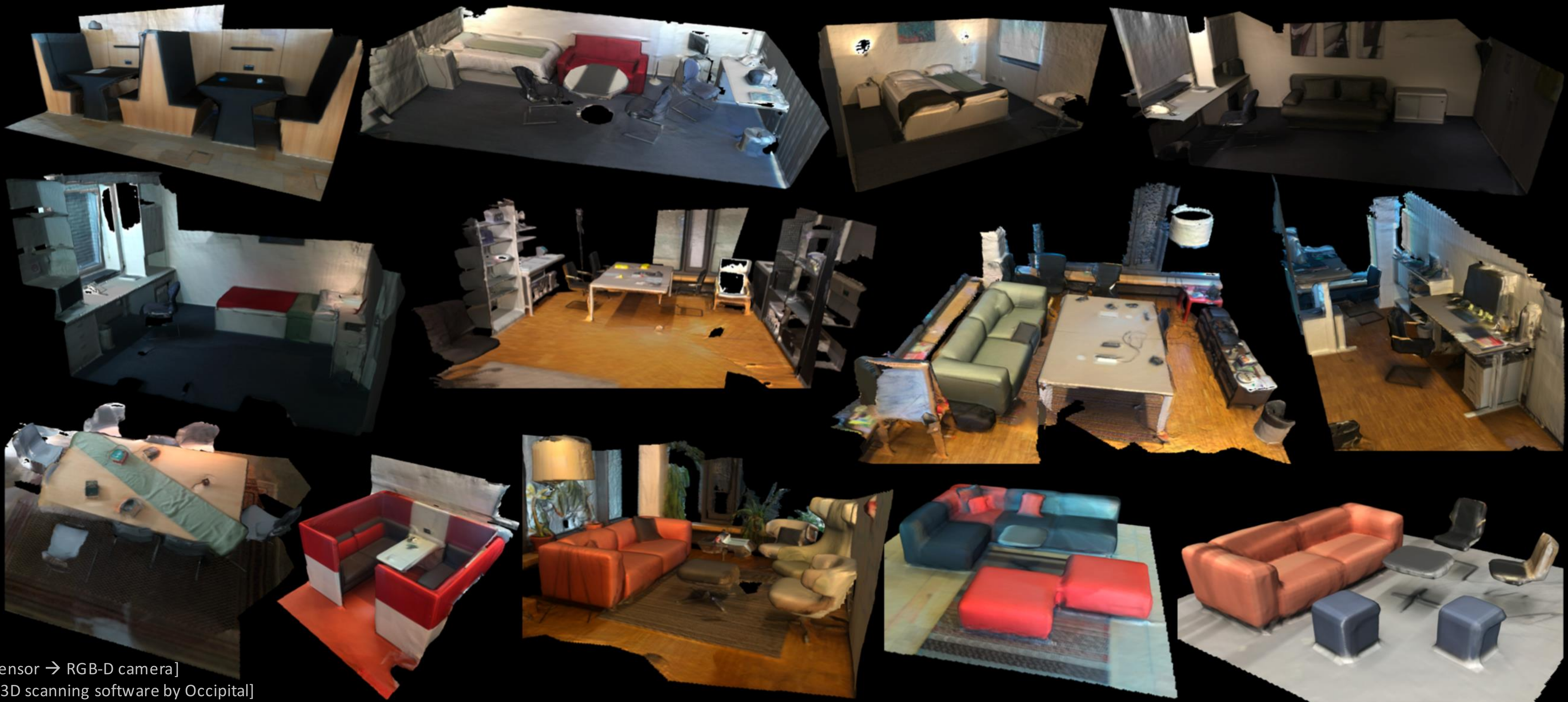
$$\lambda_\beta E_\beta(\beta) + \lambda_\mathcal{E}\, \mathrm{E}_\mathcal{E}(\psi) +$$ Shape & Expression priors

$$\lambda_P\, \mathrm{E}_P(\theta, \beta, \psi) +$$ self-Penetration (penalty)

$$\lambda_{\hat{P}}\, \mathrm{E}_{\hat{P}}(\theta, \beta, \psi, \mathrm{M}_S) +$$ Human-Scene Penetration (penalty)

$$\lambda_C\, \mathrm{E}_C(\theta, \beta, \psi, \mathrm{M}_S)$$ Contact (encourage)

PROX: Human Scene Interaction



[Structure Sensor → RGB-D camera]
[Skanect → 3D scanning software by Occipital]

PROX: Human Scene Interaction

PROX: Human Scene Interaction

PROX Dataset – Key for training ML

PROX: Human Scene Interaction

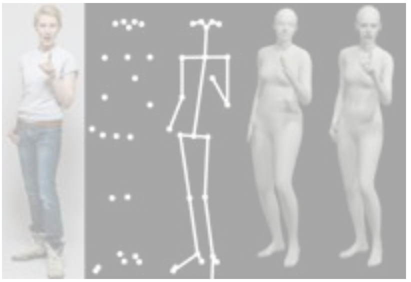
Example Application:
Learn to Populate a 3D Scene

"Populating 3D Scenes by Learning Human-Scene Interaction"
M. Hassan, P. Ghosh, J. Tesch, D. Tzionas, M. Black
CVPR 2021

Research Map

MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17



SMPL-X & SMPLify-X
CVPR'19

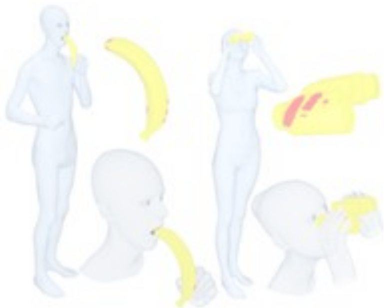
PROX
ICCV'19

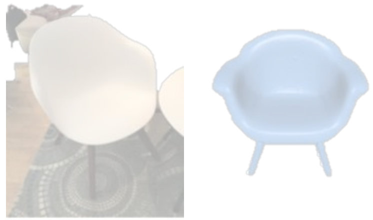
Pre-scanned objects

CWGrasp
3DV'25

GRAB
ECCV'20

ARCTIC
CVPR'23



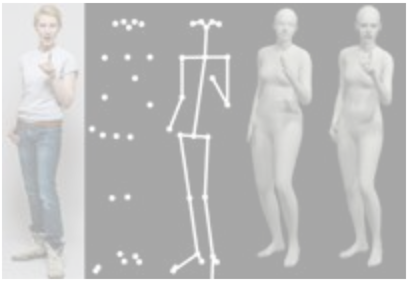


SDFit
arXiv 2024

Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17



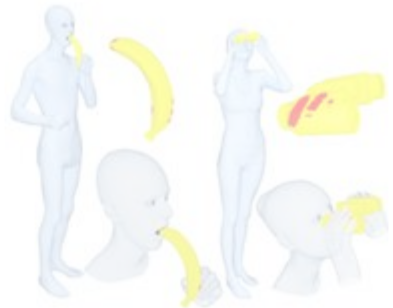
SMPL-X & SMPLify-X
CVPR’19

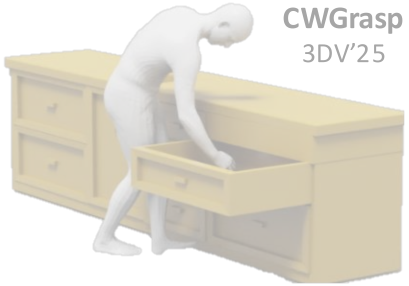
PROX
ICCV’19

SDFit
arXiv 2024

GRAB
ECCV’20

ARCTIC
CVPR’23

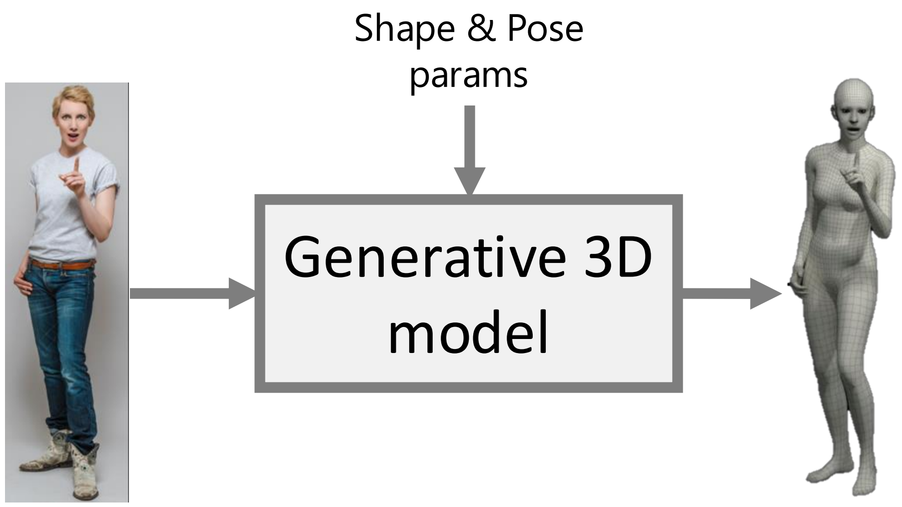




CWGrasp
3DV’25

3D Objects from
Single 2D Image 😊

3D Objects from Single Images

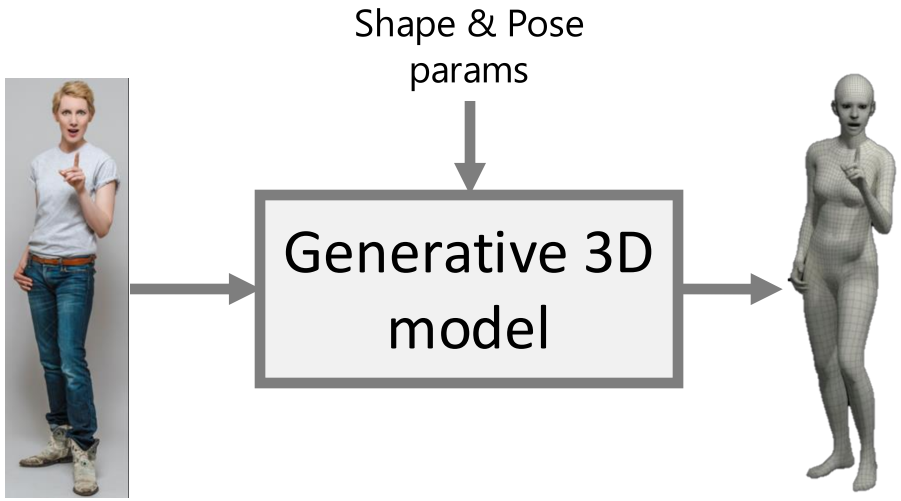


What is a SMPL-like generative model for Objects?

What is a SMPLify-like reconstr. method for Objects?

Bodies	VS	Objects
😊 Single class (human-body class)		Many classes (chairs, sofas, cars, etc) 😓
😊 Intra-class shape variance		Intra- & Inter-class shape variance 😓
😊 'Fixed' Topology		Arbitrary Topology 😓
😊 OpenPose-like keypoint detection & 2D-3D correspondences		Unsolved keypoint detection & 2D-3D correspondences 😓

3D Objects from Single Images



What is a SMPL-like generative model for Objects?

What is a SMPLify-like reconstr. method for Objects?

Bodies	VS	Objects
Single class (human-body class)		Many classes (chairs, sofas, cars, etc)
Intra-class shape variance		Intra- & Inter-class shape variance
'Fixed' Topology		Arbitrary Topology
OpenPose-like keypoint detection & 2D-3D correspondences		Unsolved keypoint detection & 2D-3D correspondences

Answering is **non-trivial**

3D Objects from Single Images

Rich 3D databases [1, 2]

Robust detectors & classifiers [3, 4]

Category-level models

Objects

Many classes
(chairs, sofas, cars, etc)

Intra- & Inter-class
shape variance

Arbitrary
Topology

Unsolved
keypoint detection &
2D-3D corresp.

[1] **Objaverse-XL**, Deitke et al., NeurIPS 2023

[2] **ShapeNet**, Chang et al., arXiv 2015

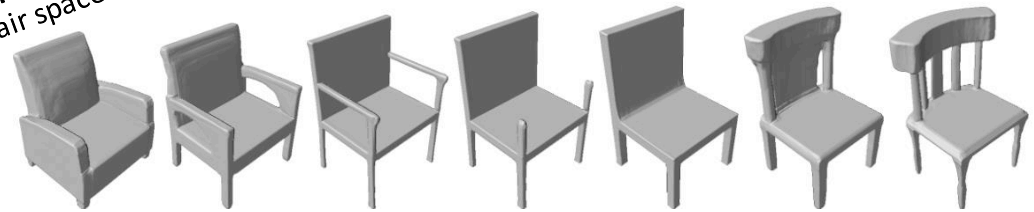
[3] **Mask R-CNN**, He et al., ICCV 2017

[4] **SAM 2**, Ravi et al., arXiv 2024

3D Objects from Single Images

DeepSDF [2]
chair space



Deep implicit shape models [1, 2]

Traverse latent space → Morphable

Category-level models

Morphable SDF (mSDF) model

Objects

Many classes
(chairs, sofas, cars, etc)

Intra- & Inter-class
shape variance

Arbitrary
Topology

Unsolved
keypoint detection & 2D-3D corresp.

[1] **Occupancy Networks**, Mescheder et al., CVPR 2019

[2] **DeepSDF**, Park et al., CVPR 2019


3D Objects from Single Images

Foundational features:

- 2D input image [1, 2]
- 3D mesh [3] from mSDF

2D-3D Corresp. → RANSAC & PnP

Category-level models

Morphable SDF (mSDF) model

Deep Features +PnP Corresp.

Objects

Many classes (chairs, sofas, cars, etc)

Intra- & Inter-class shape variance

Arbitrary Topology

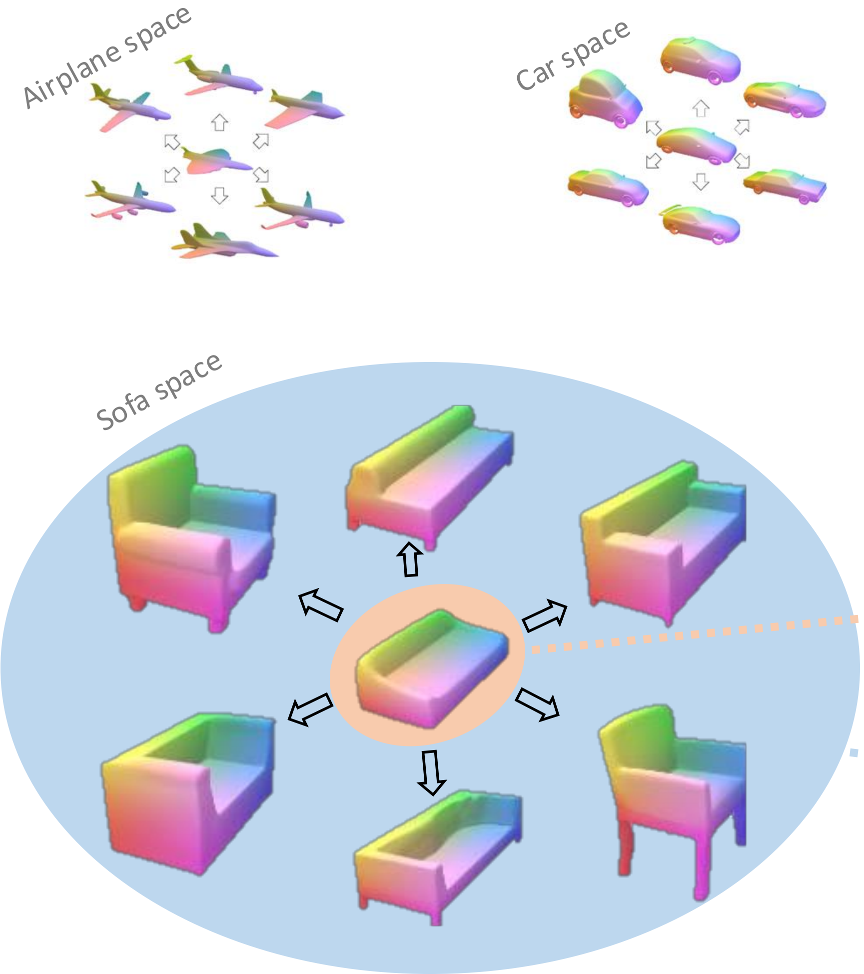
Unsolved keypoint detection & 2D-3D corresp.

[1] **DINOv2**, Oquab et al., TMLR 2024

[2] **ControlNet**, Zhang et al., ICCV 2023

[3] **Diff3F**, Dutt et al., CVPR 2024

3D Objects from Single Images



mSDF choice:
DIT [1] model

Template

Morphed Shapes
compact latent space

Cross-shape
Corresp!

Category-level models 😊

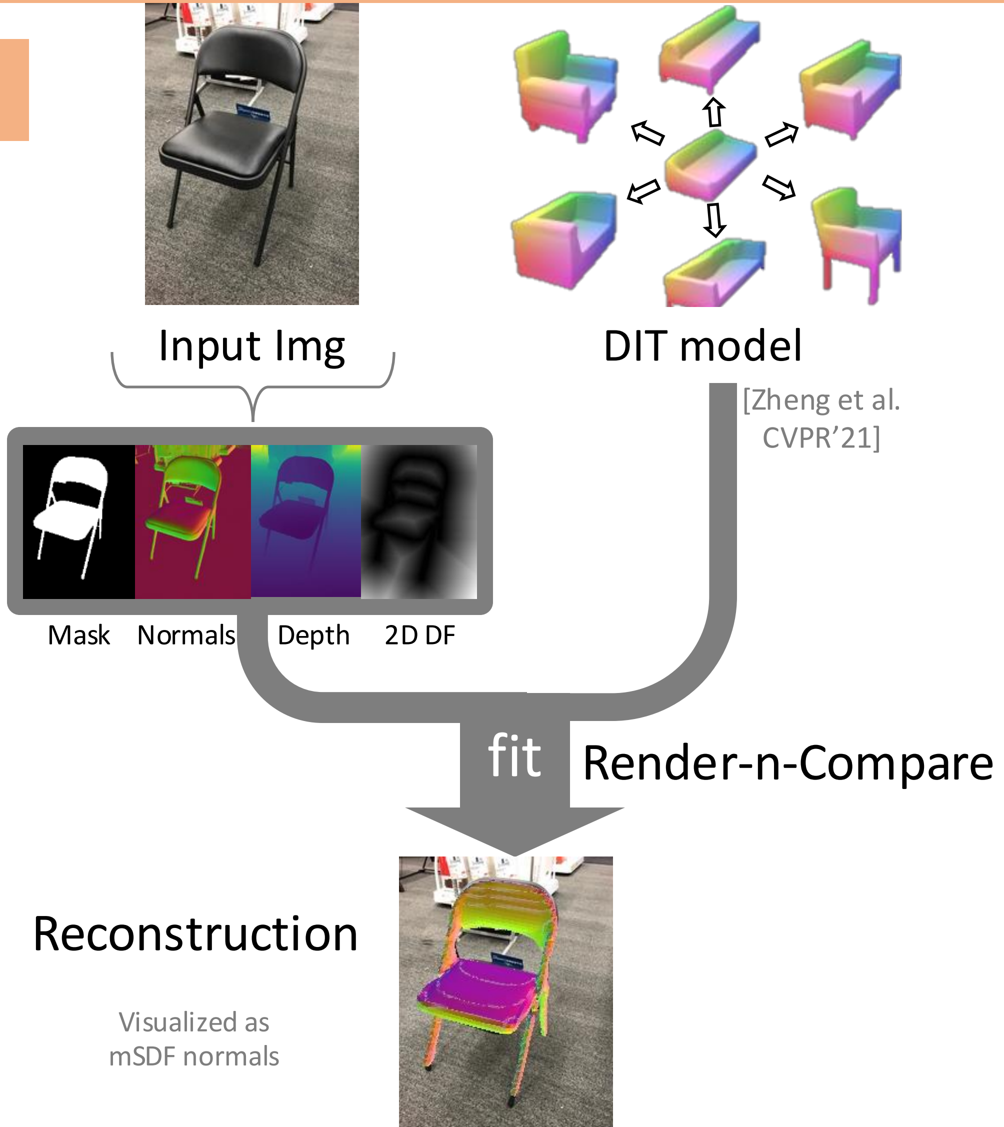
Morphable SDF (mSDF) model 😊

Deep Features +PnP Corresp. 😊

[1] **Deep implicit templates** for 3D shape representation, Zheng et al., CVPR 2021

3D Objects from Single Images

SDFit - Framework



Category-level models

Morphable SDF (mSDF) model

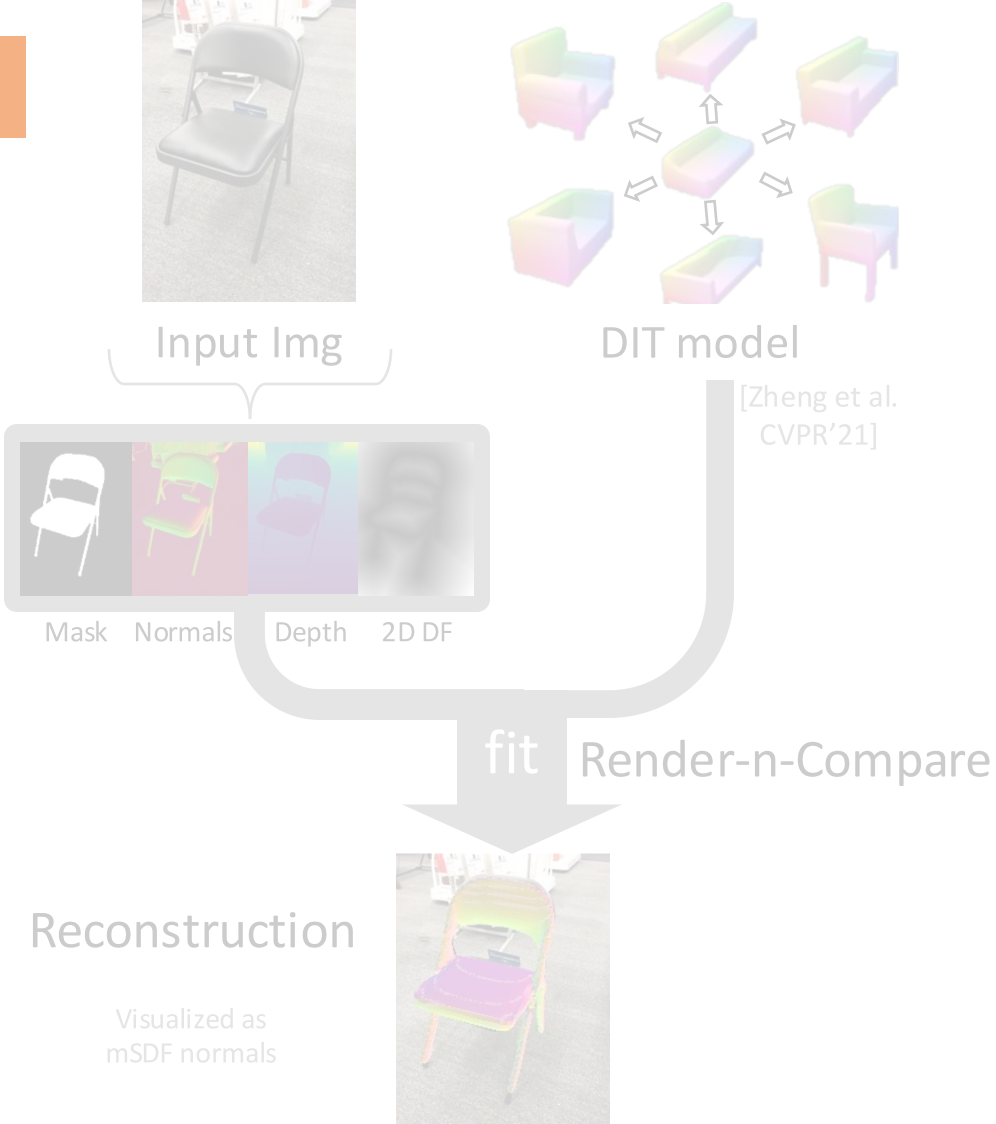
Deep Features +PnP Corresp

3D Objects from Single Images

SDFit - Framework

Needs good **initialization:**

- **Shape**
- **Pose**



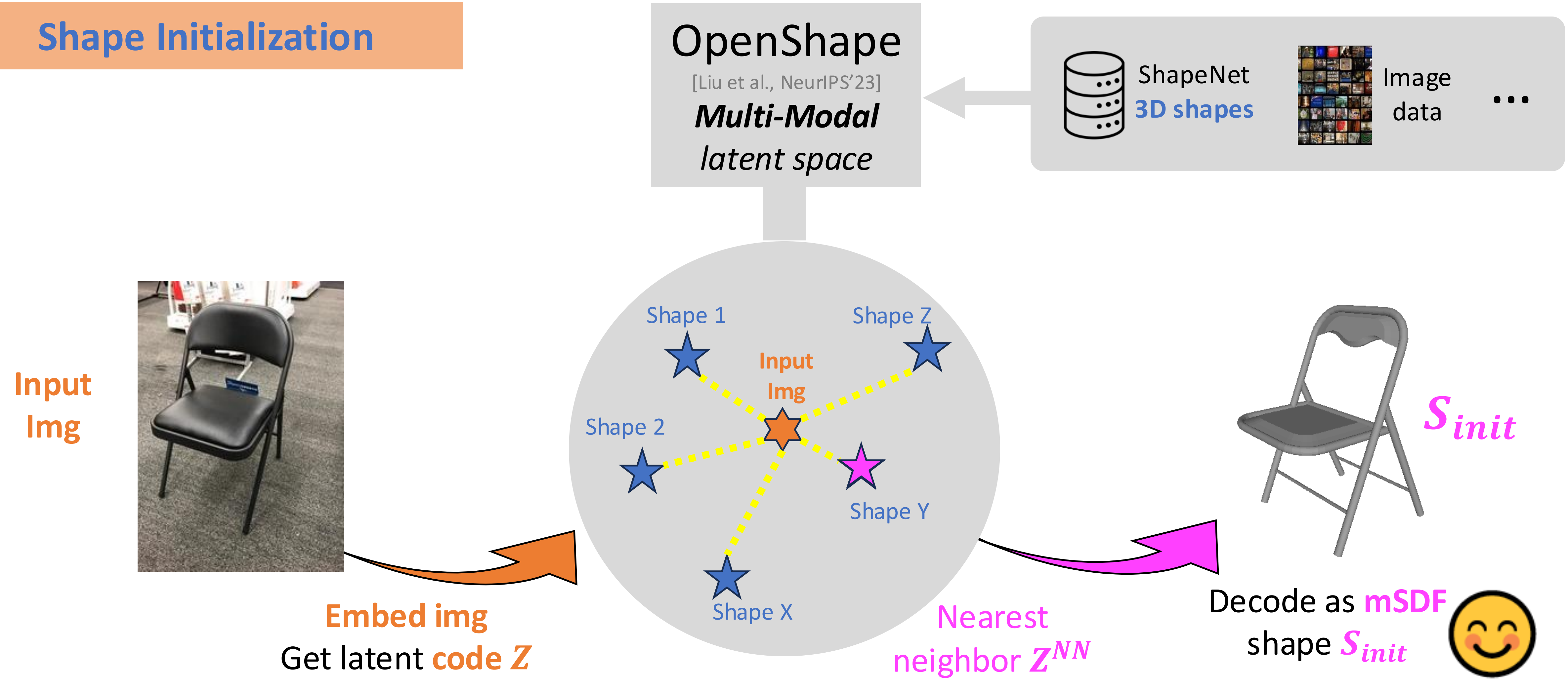
Category-level models

Morphable SDF (mSDF) model

Deep Features +PnP Corresp

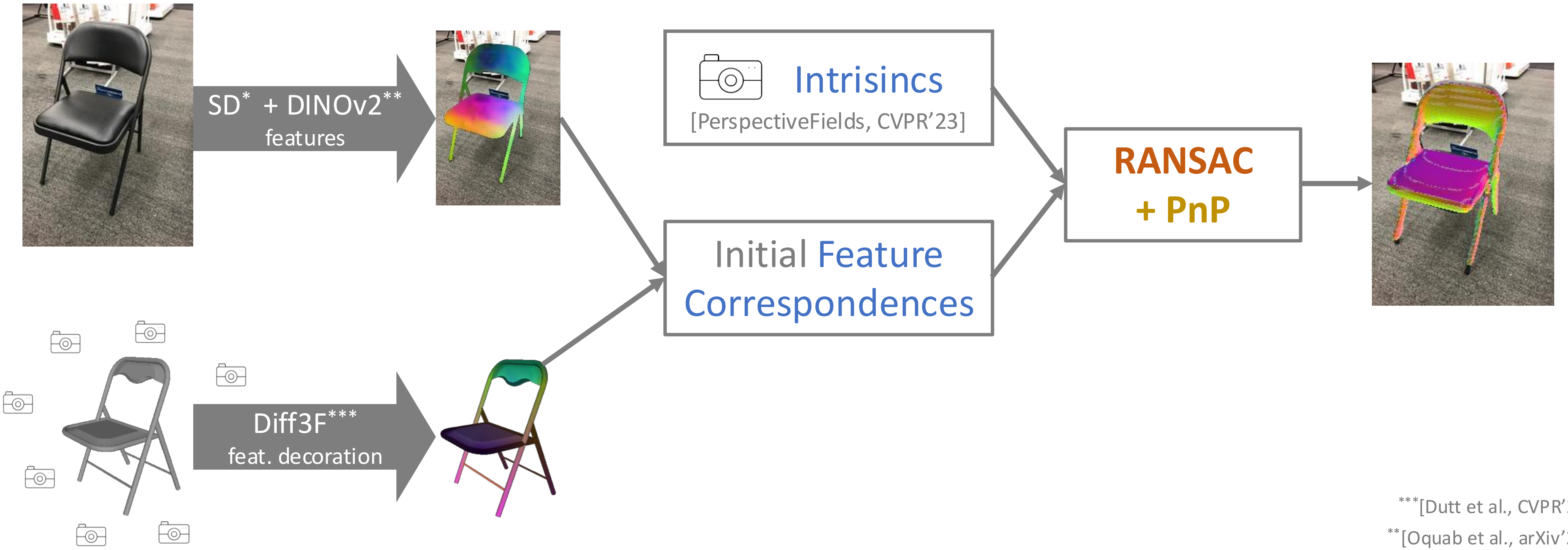
3D Objects from Single Images

Shape Initialization



3D Objects from Single Images

Pose Initialization



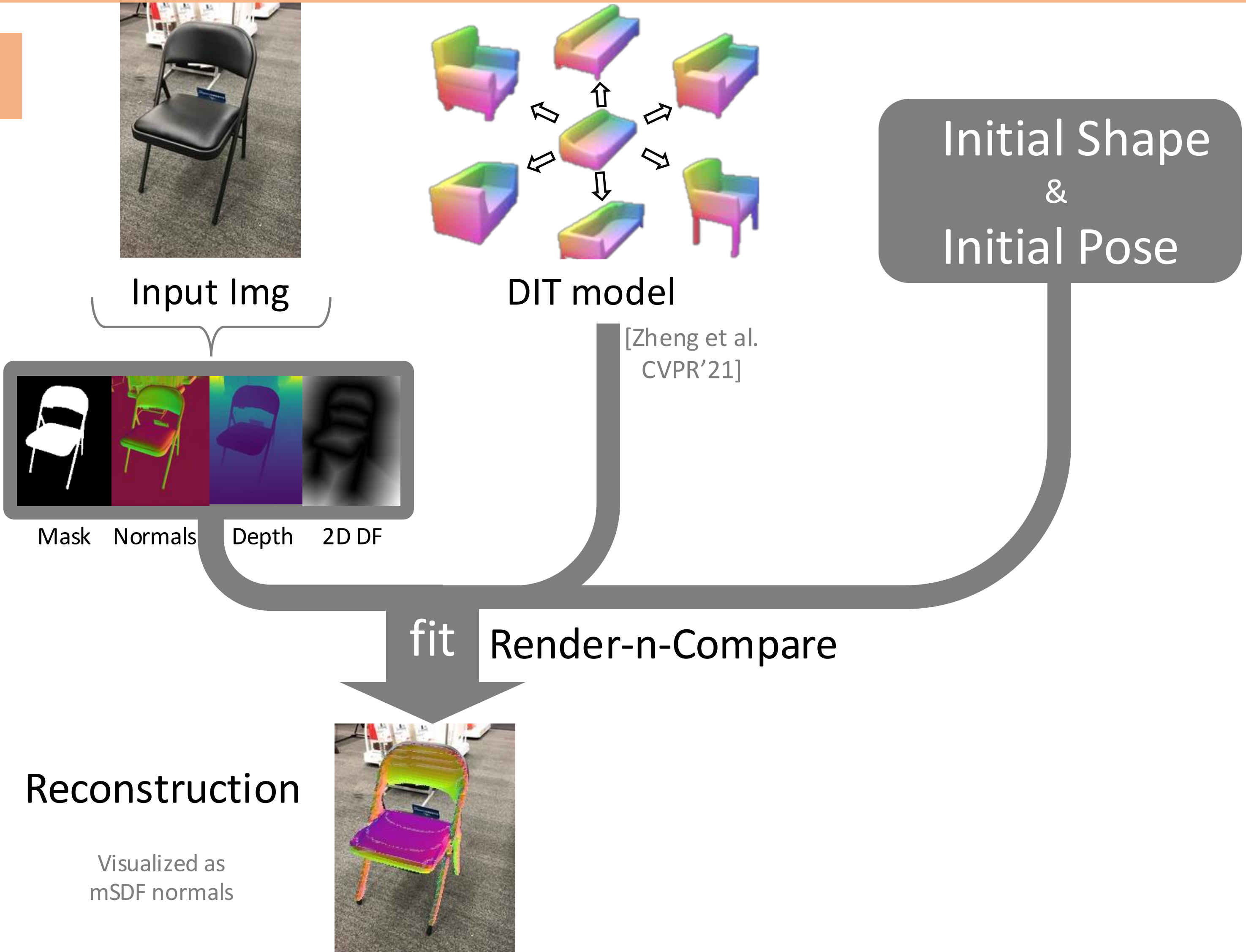
***[Dutt et al., CVPR'24]

**[Oquab et al., arXiv'24]

*[Rombach et al., CVPR'22]

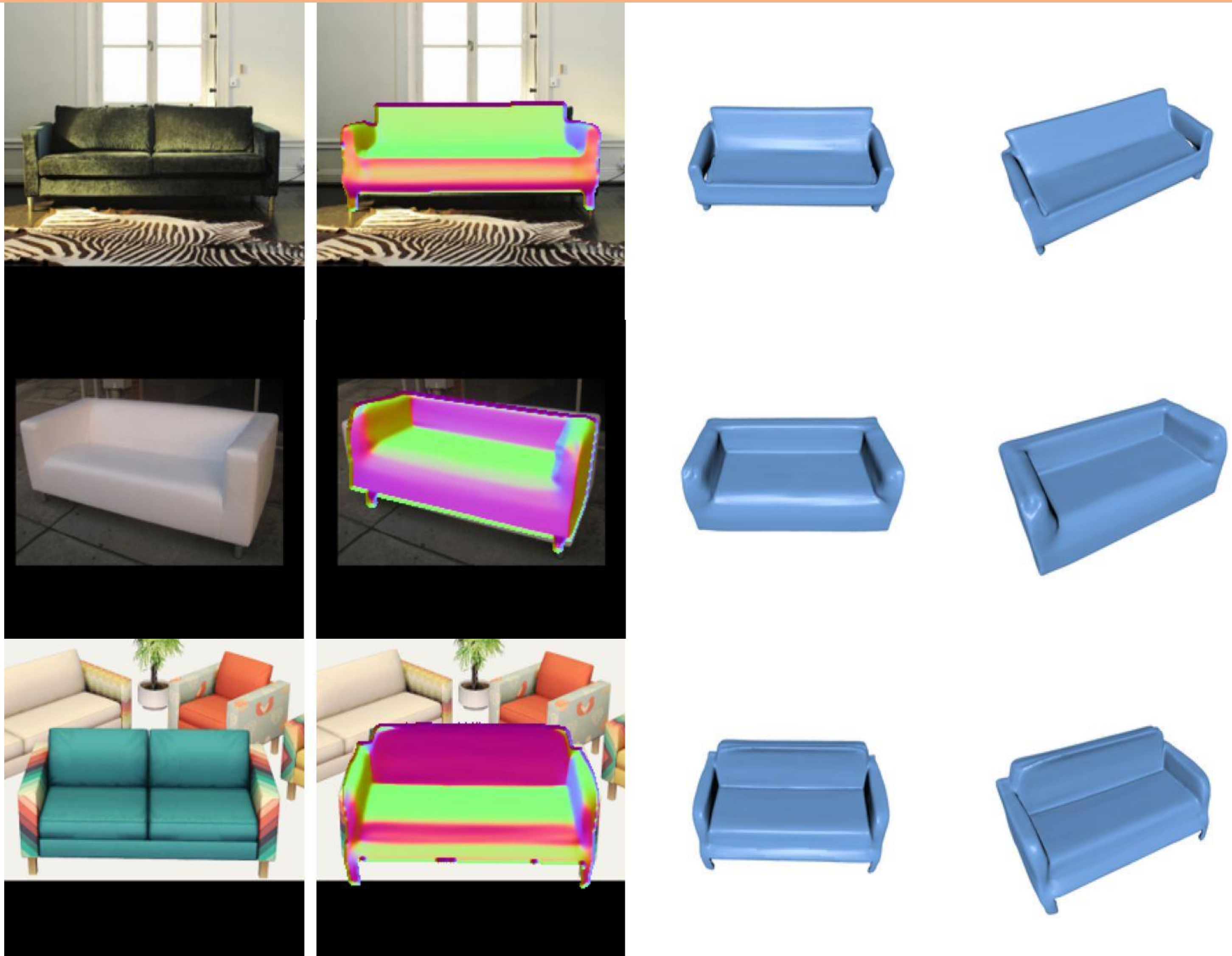
3D Objects from Single Images

SDFit - Framework

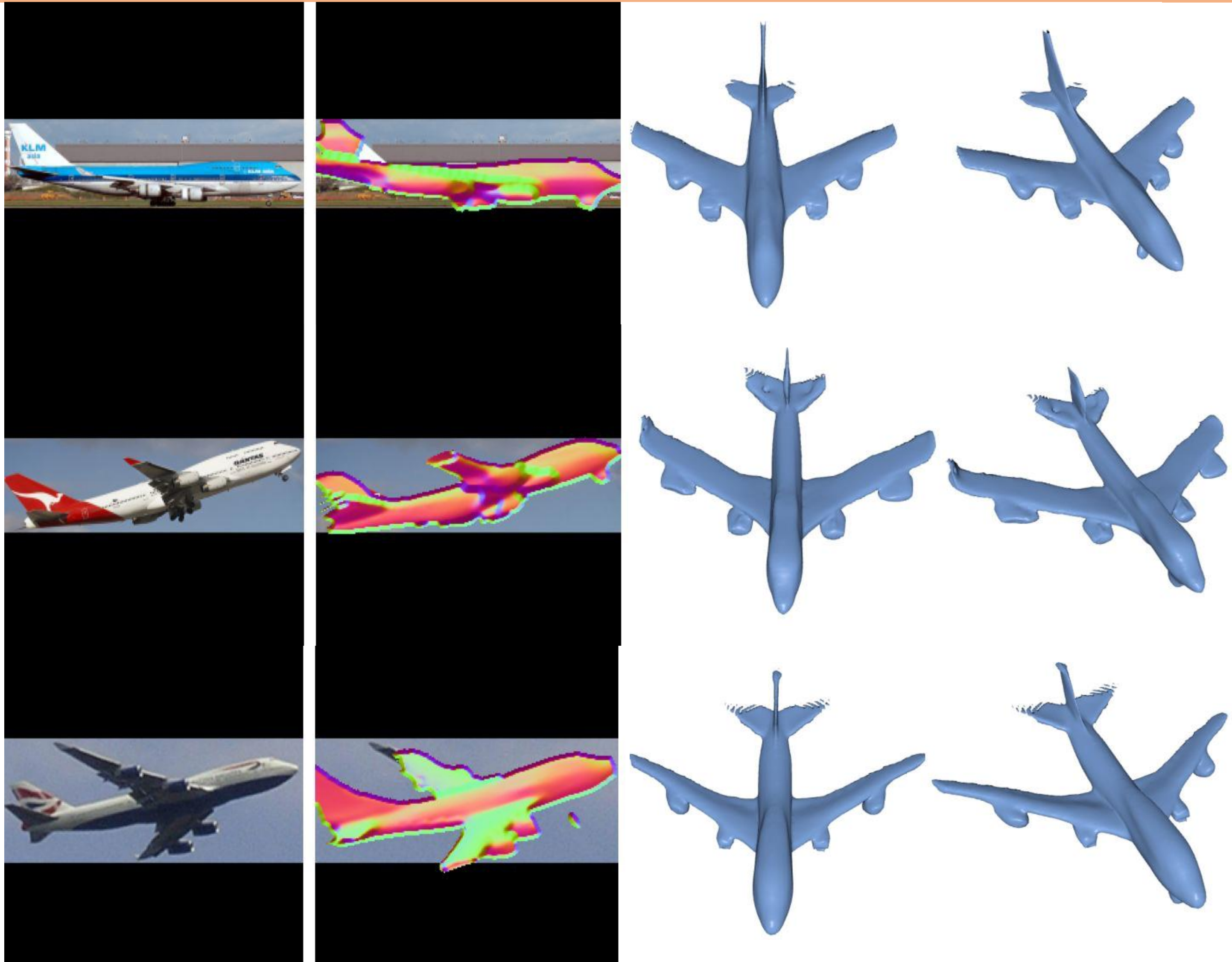


3D Objects from Single Images

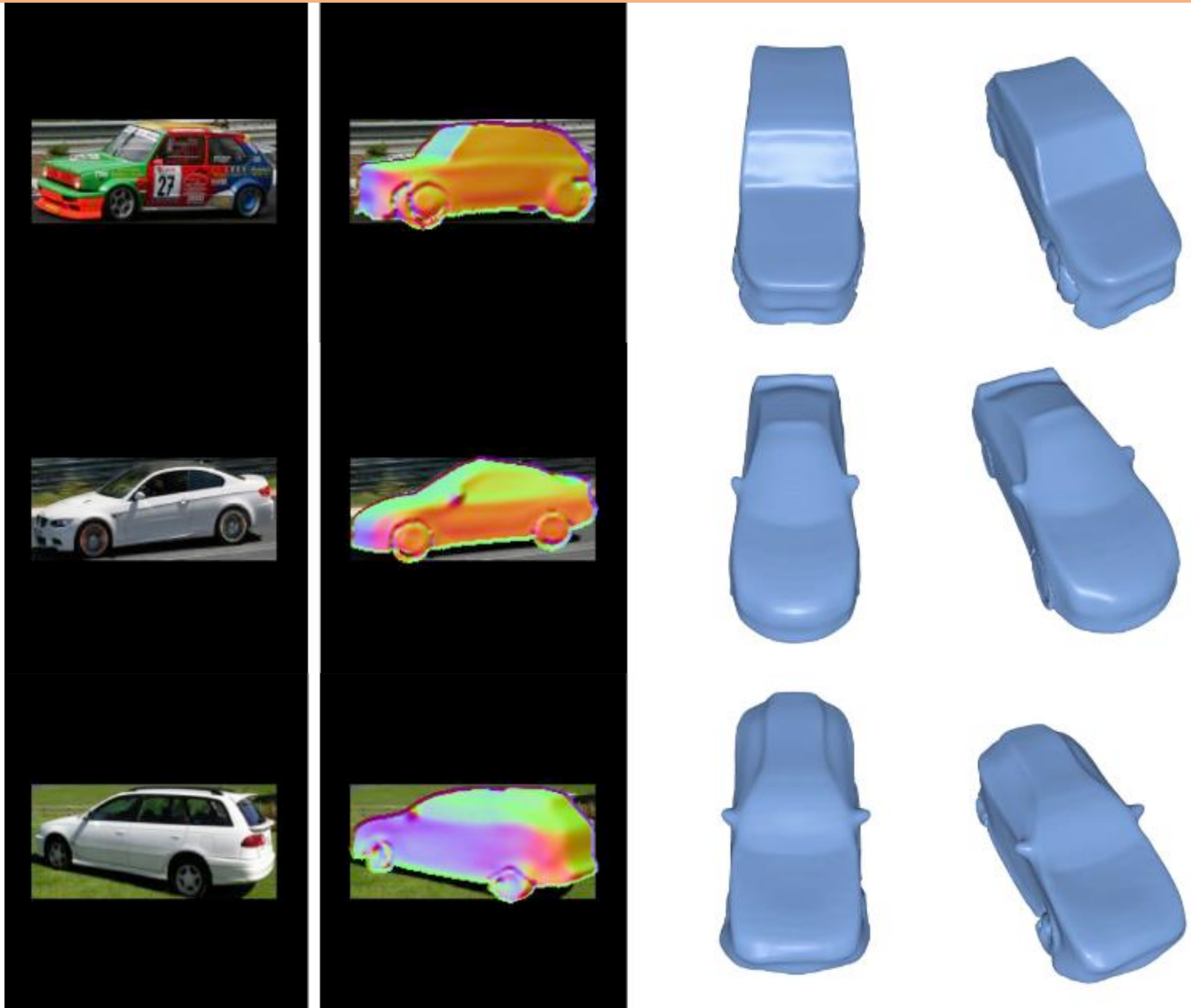
3D Objects from Single Images



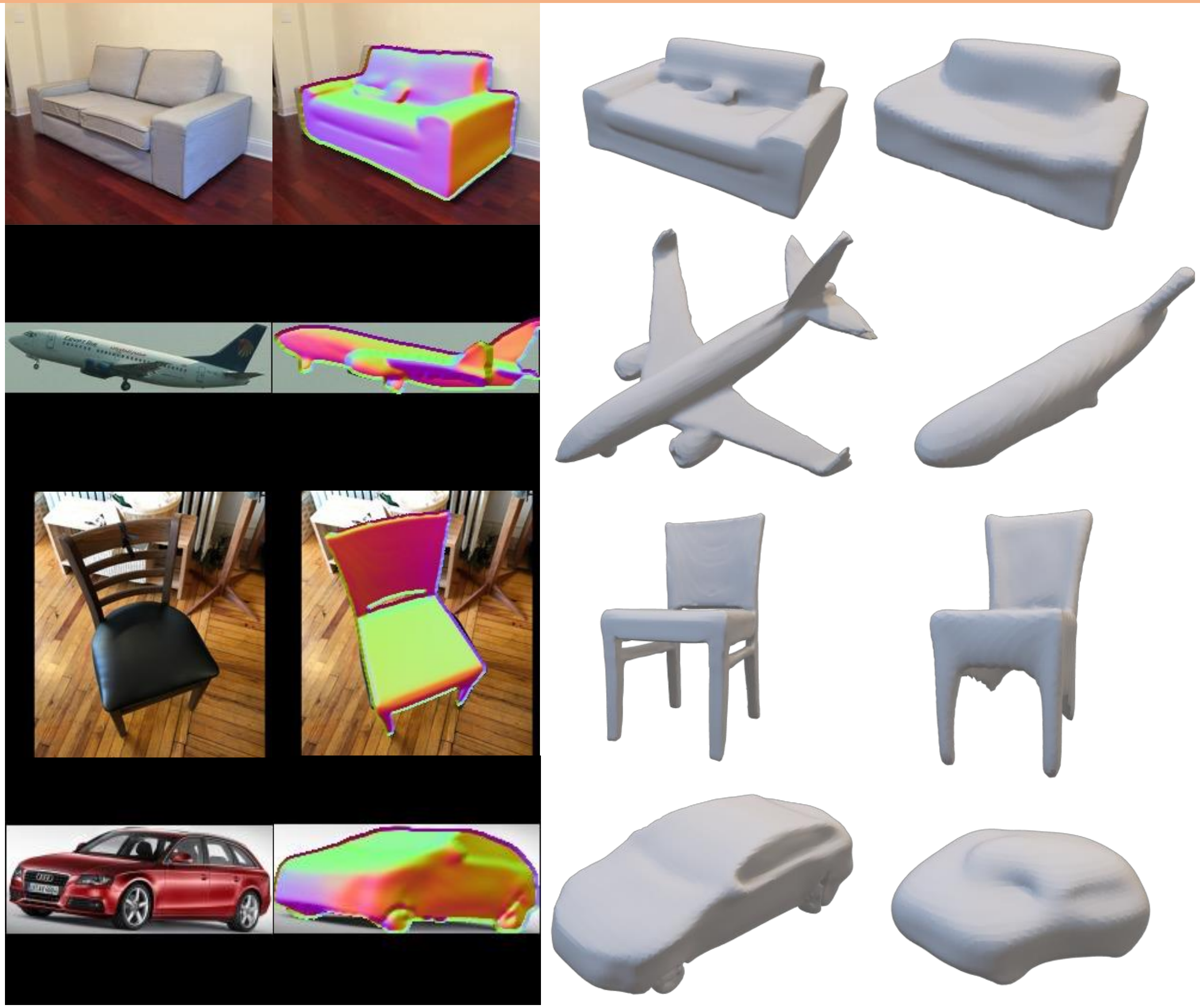
3D Objects from Single Images



3D Objects from Single Images



3D Objects from Single Images



Input Image **SDFit** (Ours) **SDFit** (Ours) **ZeroShape** [1]

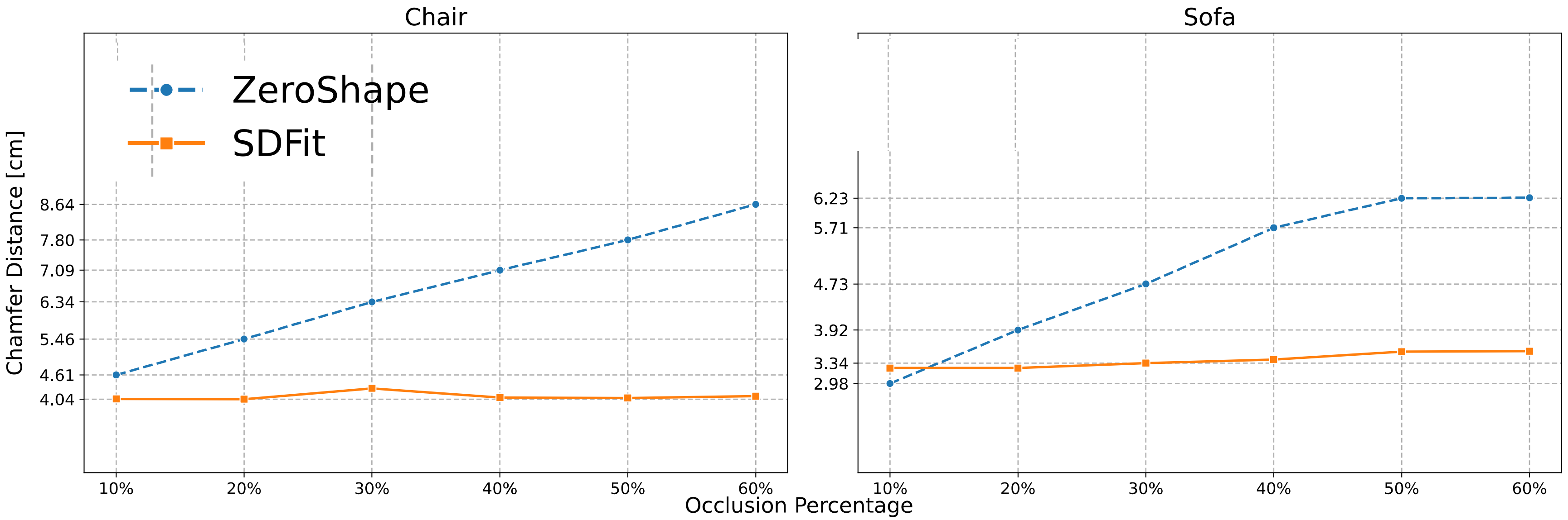
SDFit:

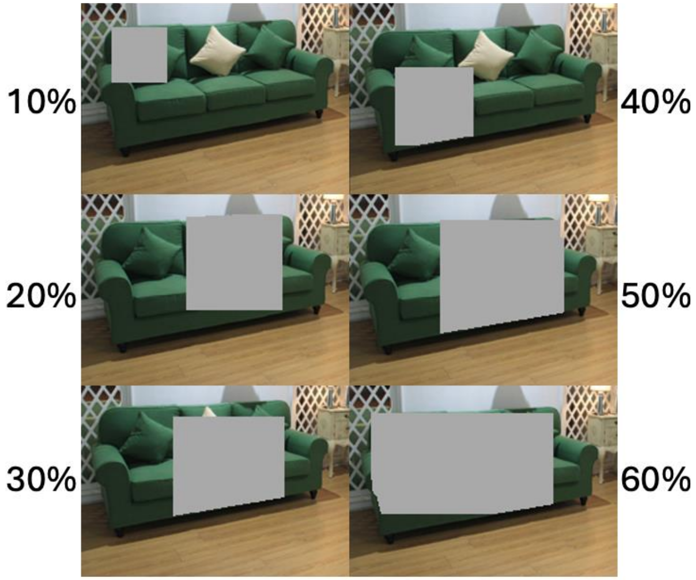
😊 Recovers Self-Occluded parts via its inherent Shape Prior

[1] **ZeroShape:** Regression-based Zero-shot Shape Reconstruction, Huang et al., CVPR 2024

3D Objects from Single Images

3D Shape Reconstruction





Occlude observed masks & feature maps

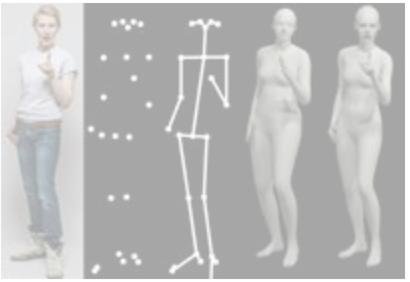
SDFit: Robust to occlusions

[1] **ZeroShape:** Regression-based Zero-shot Shape Reconstruction, Huang et al., CVPR 2024

Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17

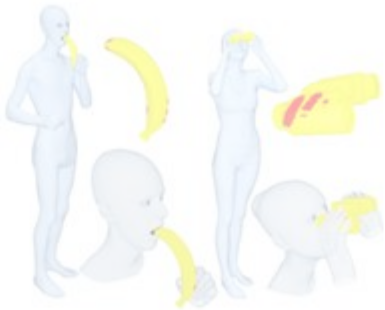


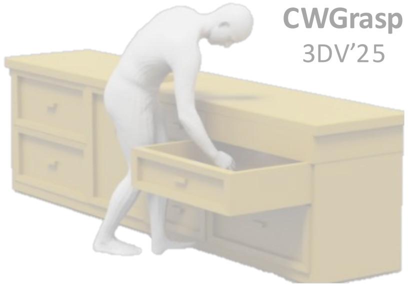
SMPL-X & SMPLify-X
CVPR’19

PROX
ICCV’19

GRAB
ECCV’20

ARCTIC
CVPR’23





CWGrasp
3DV’25

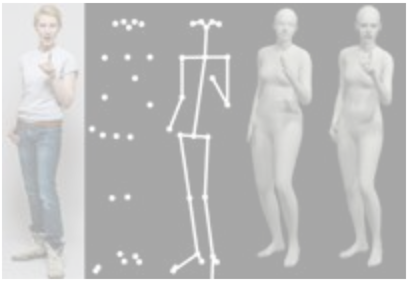
SDFit
arXiv 2024

Objects alone
not interesting

Research Map

MANO
SIGGRAPH-Asia'17

SMPL+H
SIGGRAPH-Asia'17



SMPL-X & SMPLify-X
CVPR'19

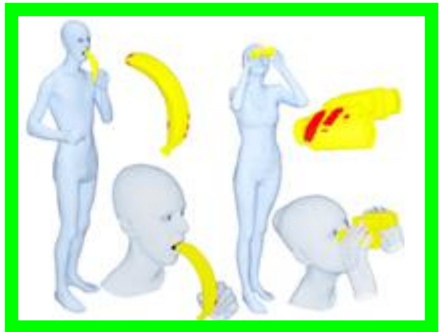
PROX
ICCV'19

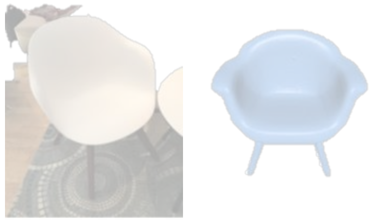
Human-Object Interaction

with Whole Bodies

GRAB
ECCV'20

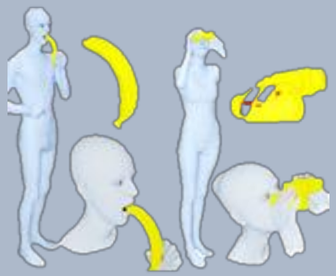
ARCTIC
CVPR'23



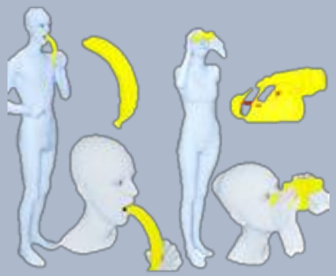


SDFit
arXiv 2024

CWGrasp
3DV'25

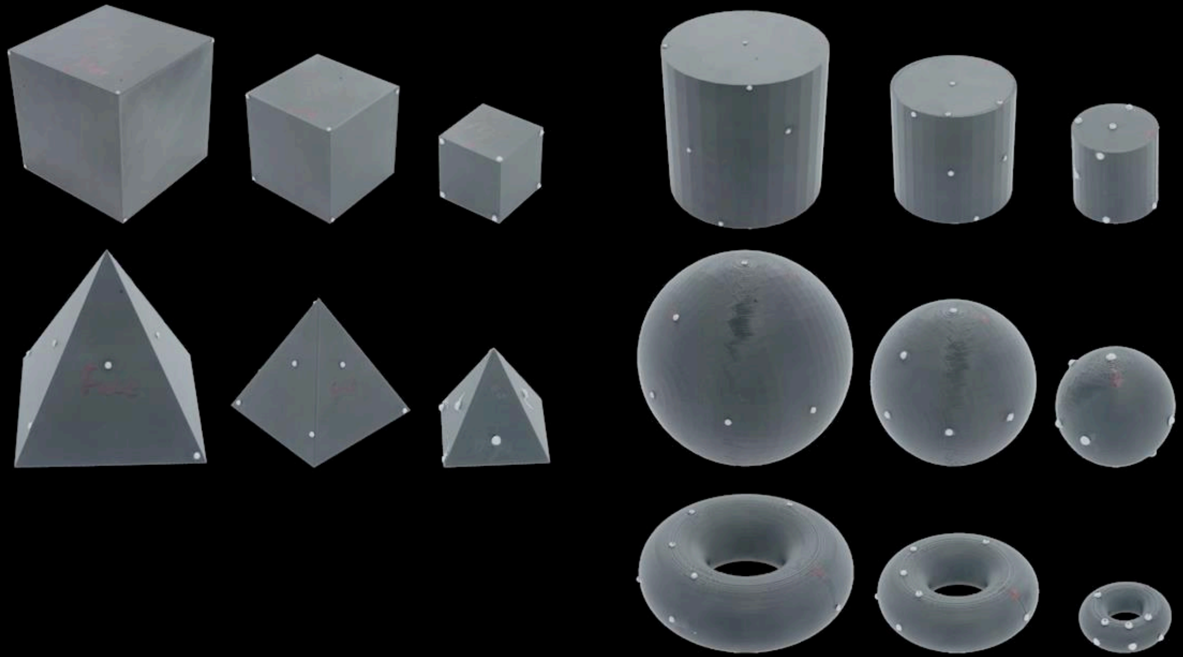


GRAB: Whole-body Grasps

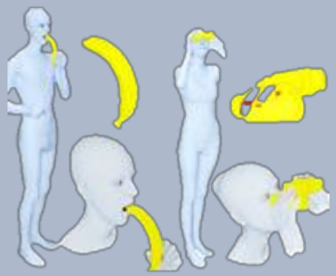


GRAB: Whole-body Grasps

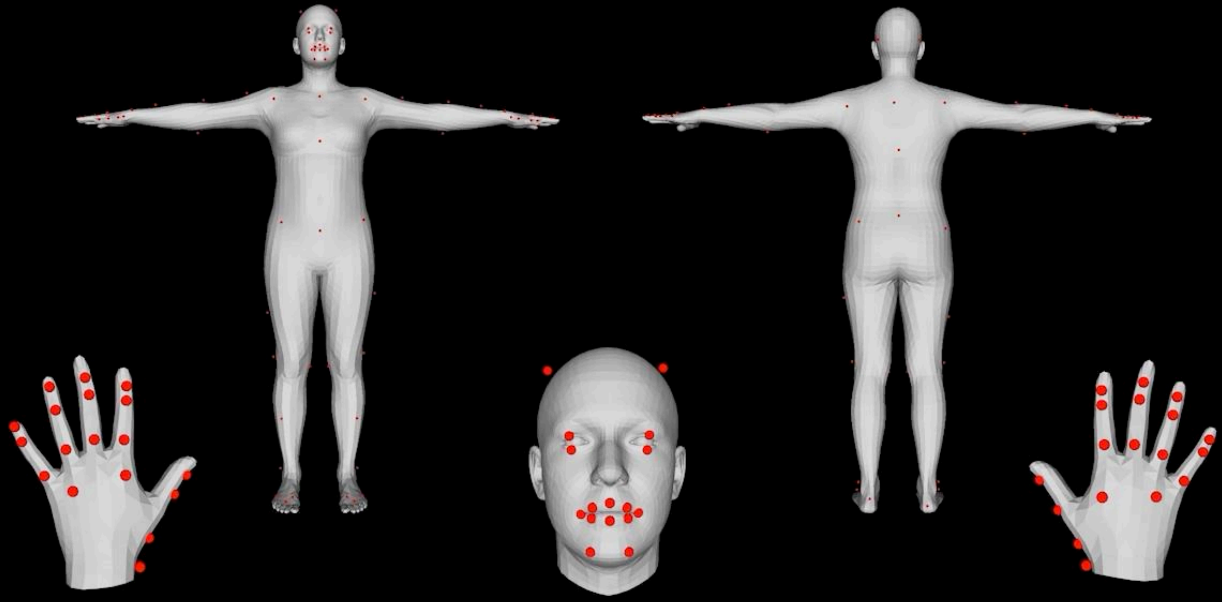
High-res & 54-camera Vicon MoCap system
1.5mm radius hemi-spherical markers
Semi-automatic MoCap data cleaning & labeling



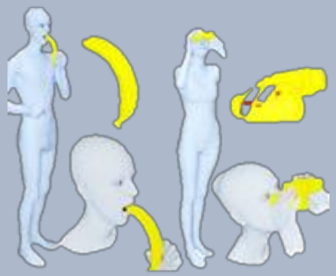
[ContactDB, CVPR'19]



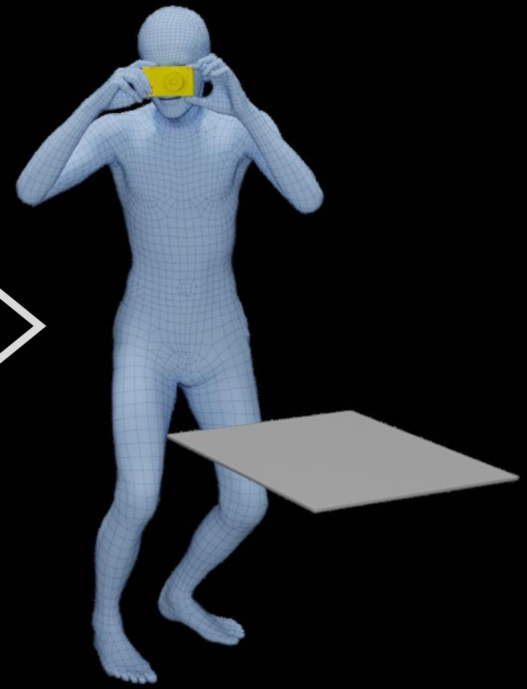
GRAB: Whole-body Grasps



High-res & 54-camera Vicon MoCap system
1.5mm radius hemi-spherical markers
Semi-automatic MoCap data cleaning & labeling



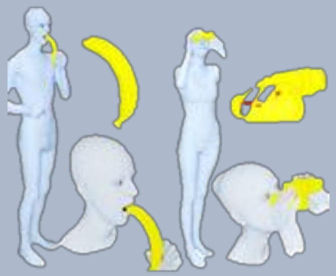
GRAB: Whole-body Grasps



High-res & 54-camera Vicon MoCap system
1.5mm radius hemi-spherical markers
Semi-automatic MoCap data cleaning & labeling

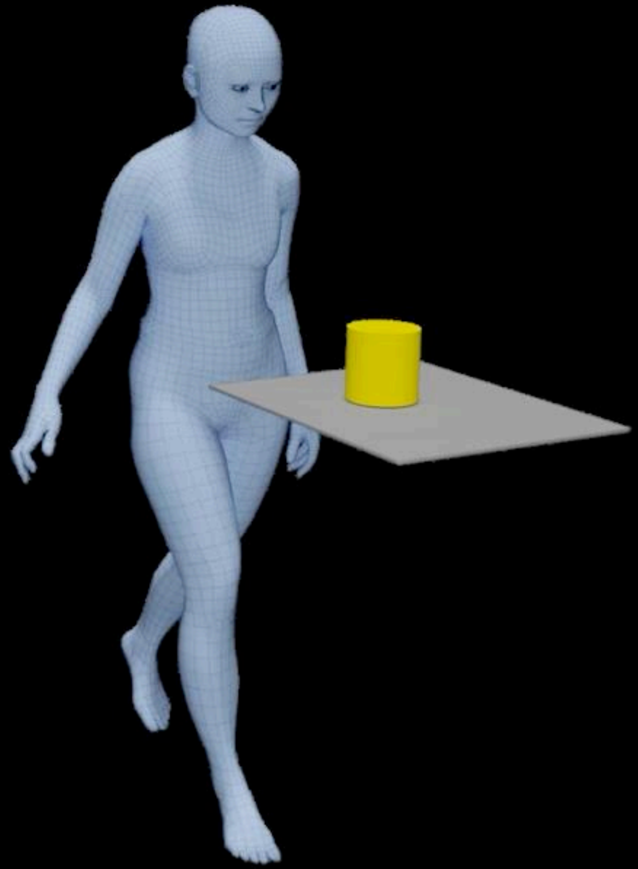
Adapt MoSh++ [1] for SMPL-X (body + face + hands)
Rigid fitting for object meshes to markers

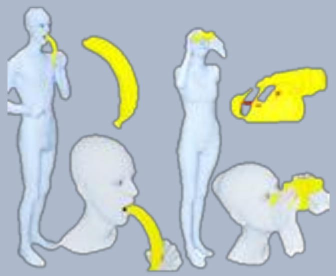
[1] N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, M. J. Black. *AMASS: Archive of Motion Capture as Surface Shapes*. ICCV 2019



GRAB: Whole-body Grasps

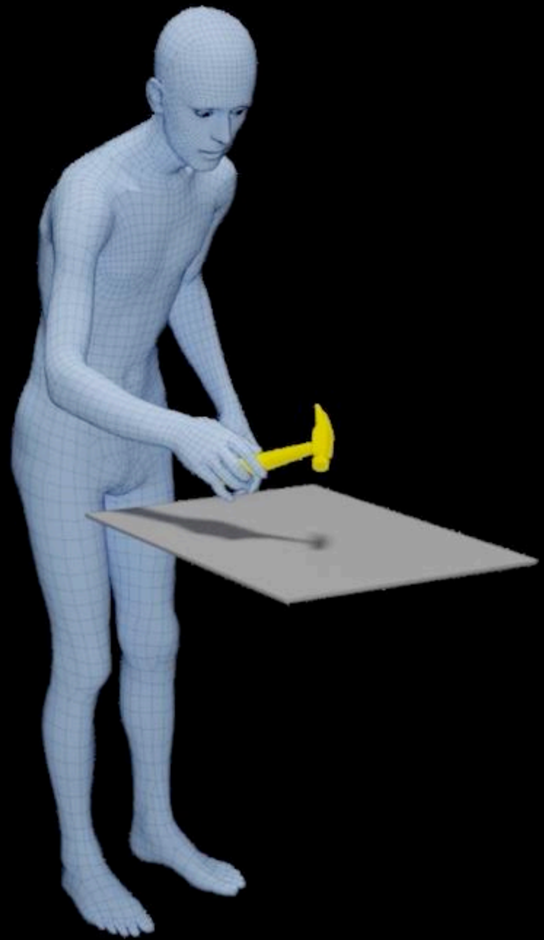
4 Interaction Intents

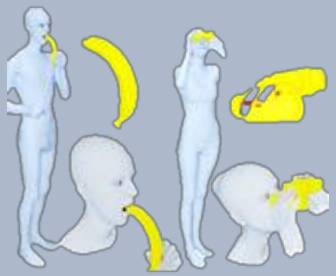




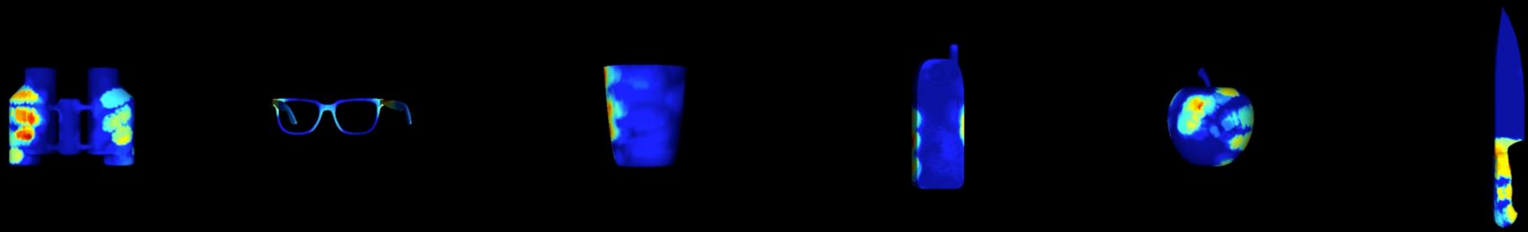
GRAB: Whole-body Grasps

Whole-Body Interaction

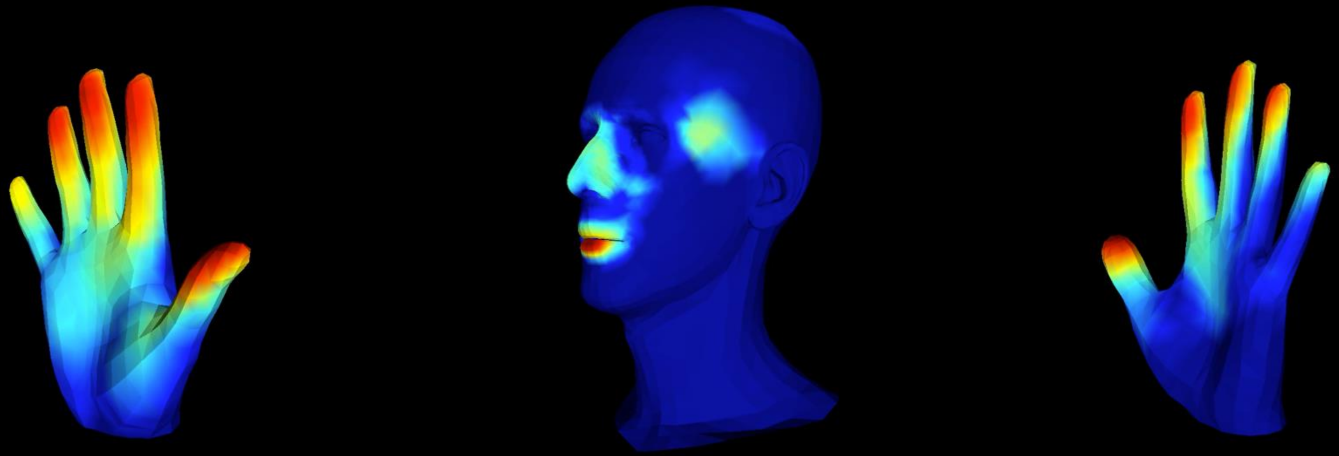


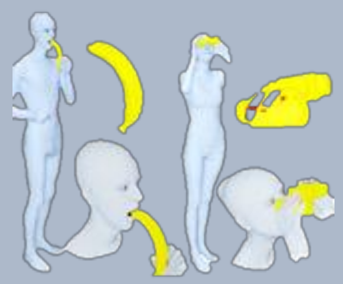


GRAB: Whole-body Grasps

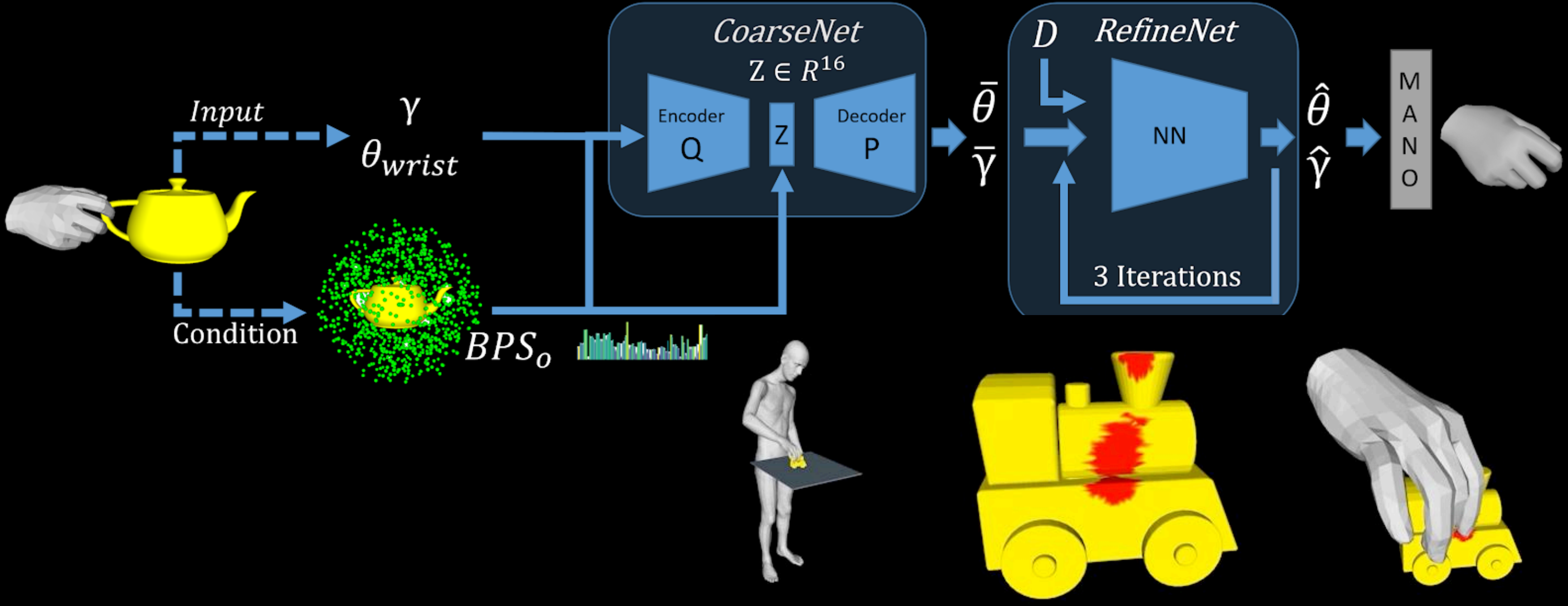


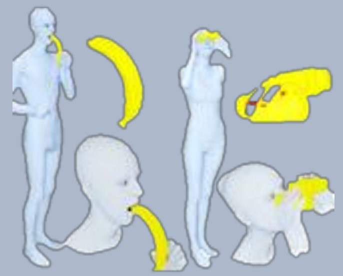
Data-driven likely contacts



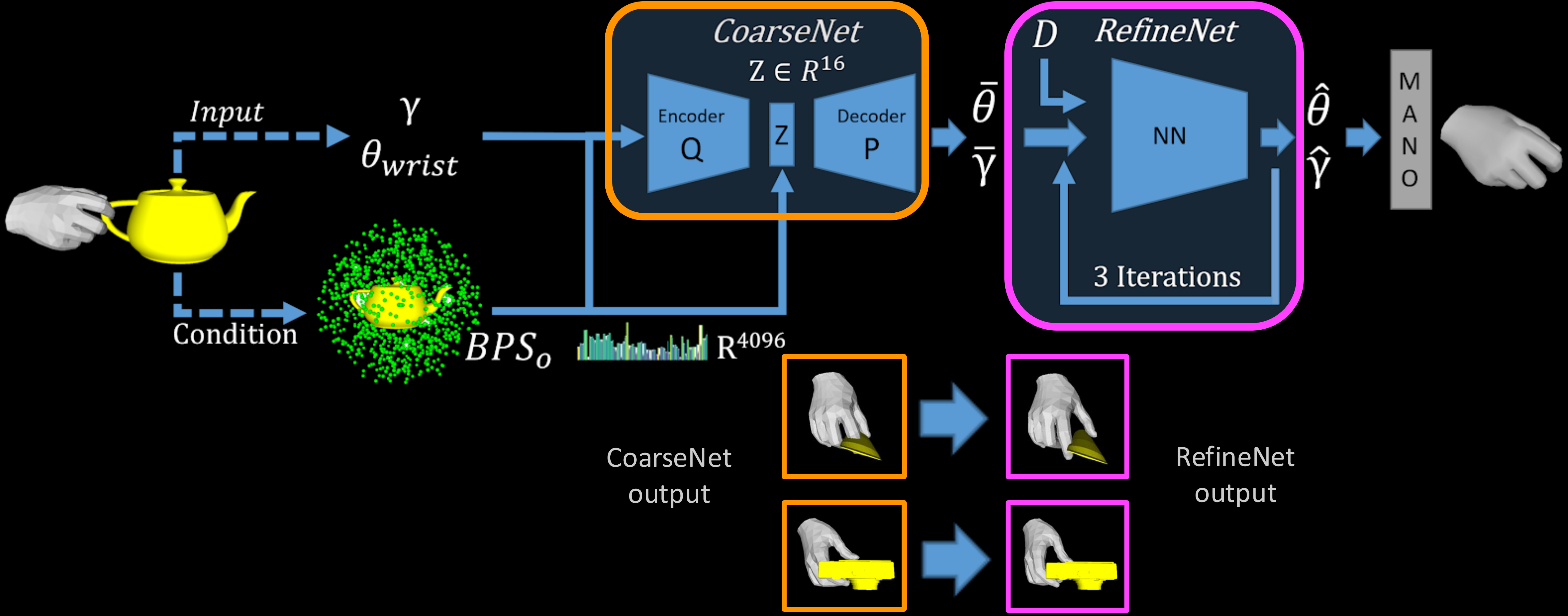


GrabNet: Grasp Synthesis

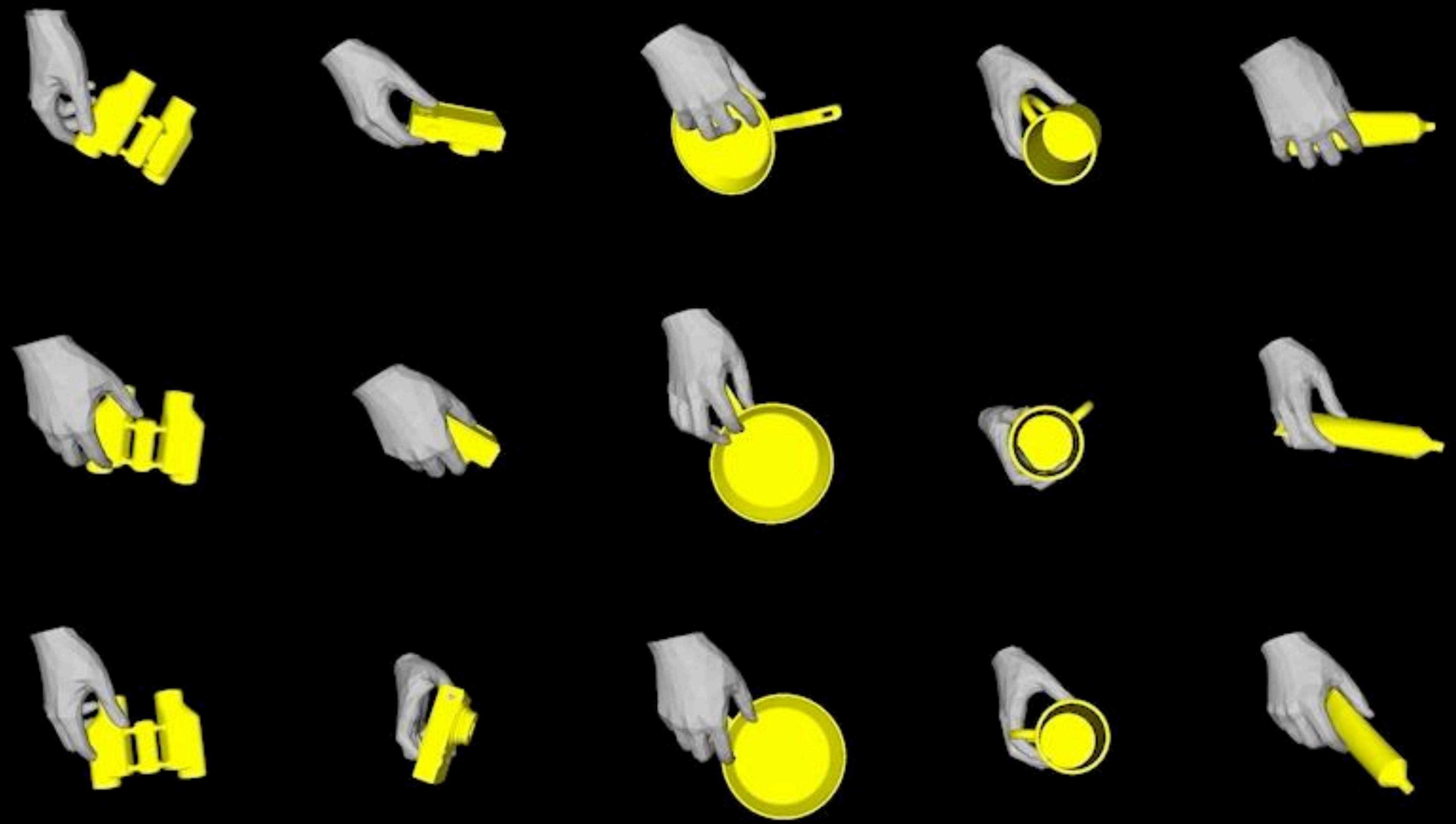




GrabNet: Grasp Synthesis



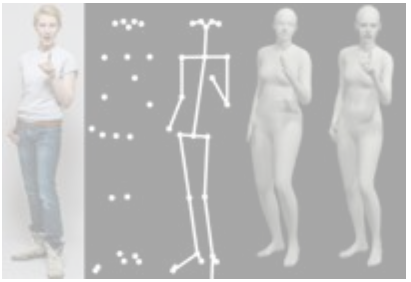
GrabNet: Grasp Synthesis



Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17



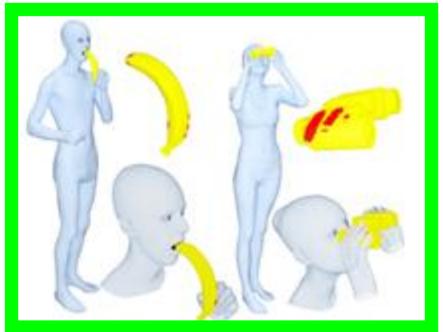
SMPL-X & SMPLify-X
CVPR’19

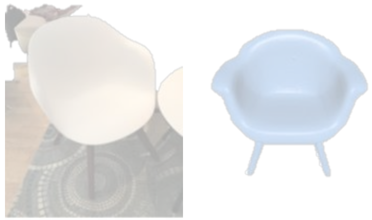
PROX
ICCV’19

CWGrasp
3DV’25

GRAB
ECCV’20

ARCTIC
CVPR’23





SDFit
arXiv 2024

Scarce data

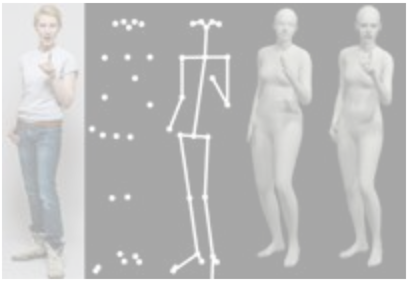
Constrained settings

Methods struggle generalizing

Research Map

MANO
SIGGRAPH-Asia’17

SMPL+H
SIGGRAPH-Asia’17

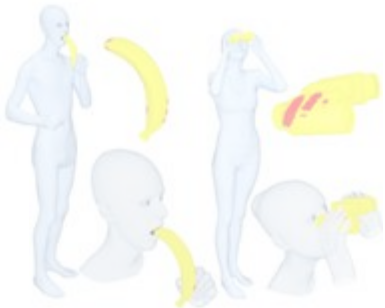


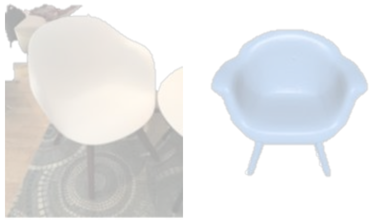
SMPL-X & SMPLify-X
CVPR’19

PROX
ICCV’19

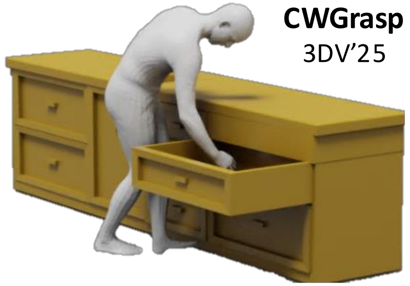
GRAB
ECCV’20

ARCTIC
CVPR’23



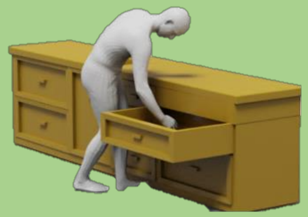


SDFit
arXiv 2024



CWGrasp
3DV’25

Go beyond data 😊

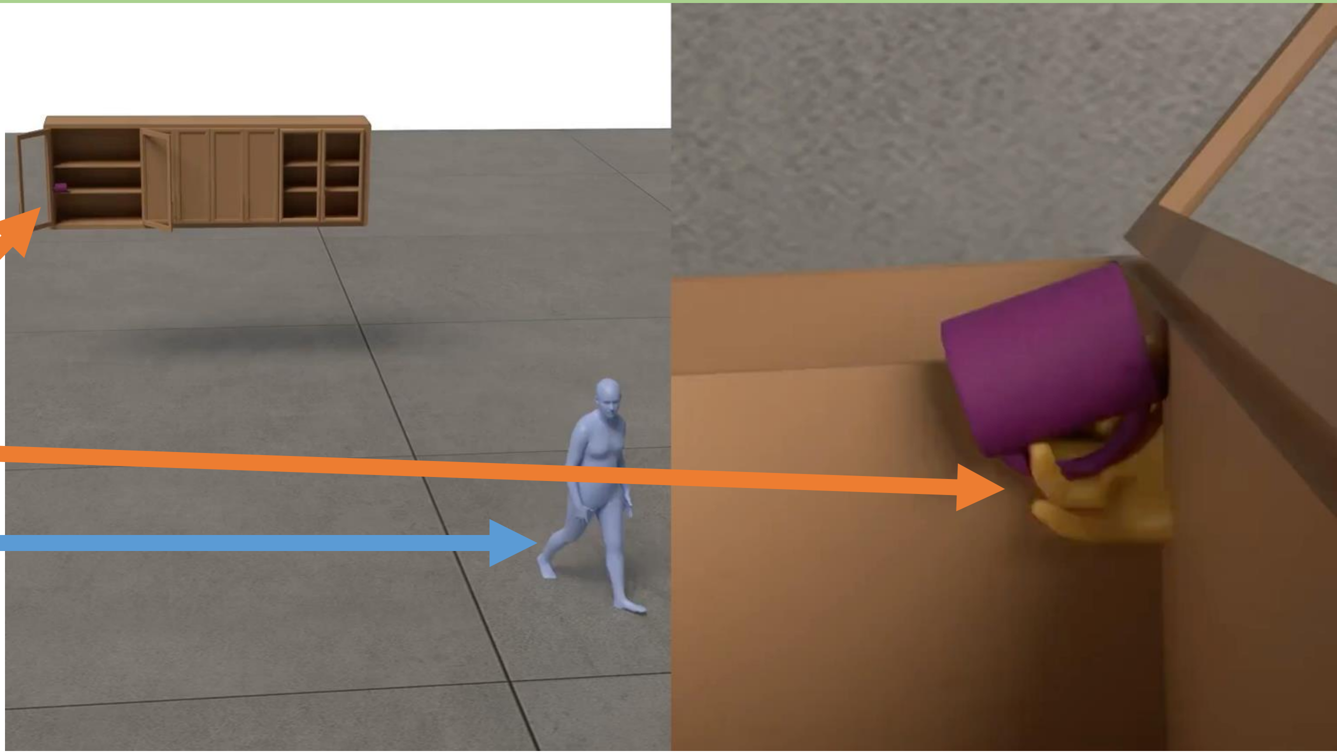


3D Grasp Synthesis

How do we go **beyond (scarce) data?**

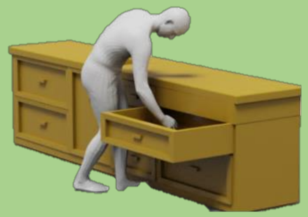
Promising → **FLEX** [1] method

- Divide & conquer!
- Generate a **hand-only** grasp [2]
- Sample random **bodies** in scene
  (500 samples, random location & pose)
- Optimize:
  - **Hand-only** grasp to match **body**
  - **Body** to match **hand-only** grasp
  - **Prune** implausible Bodies



[1] **FLEX:** Full-Body Grasping Without Full-Body Grasps. Tendulkar et al. , CVPR 2023

[2] **GrabNet**, Taheri et al., ECCV 2020



3D Grasp Synthesis

How do we go **beyond (scarce) data?**

Promising → **FLEX** [1] method

- Divide & conquer!
- Generate a **hand-only** grasp [2]
- Sample random **bodies** in scene
  (500 samples, random location & pose)
- Optimize:
  - **Hand-only** grasp to match **body**
  - **Body** to match **hand-only** grasp
  - **Prune** implausible Bodies

Great idea!

Exhaustive sampling

Intensive post-processing

Roots of limitations:

- **Non-controllable** components [2]
- **Too late** reasoning: body & scene

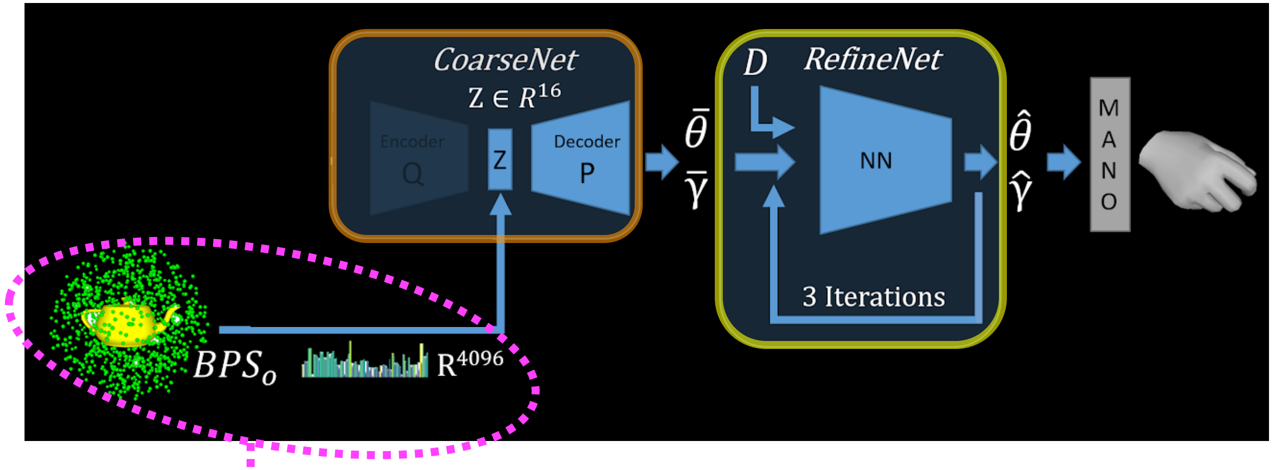
Let's solve these!

[1] **FLEX:** Full-Body Grasping Without Full-Body Grasps. Tendulkar et al. , CVPR 2023

[2] **GrabNet**, Taheri et al., ECCV 2020

3D Grasp Synthesis

Grasping hands



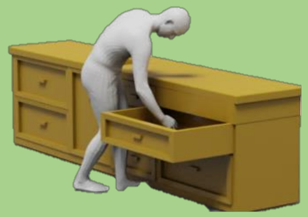
Vanilla GrabNet [1]

Only condition @ inference: Object shape

Draw 5 different samples

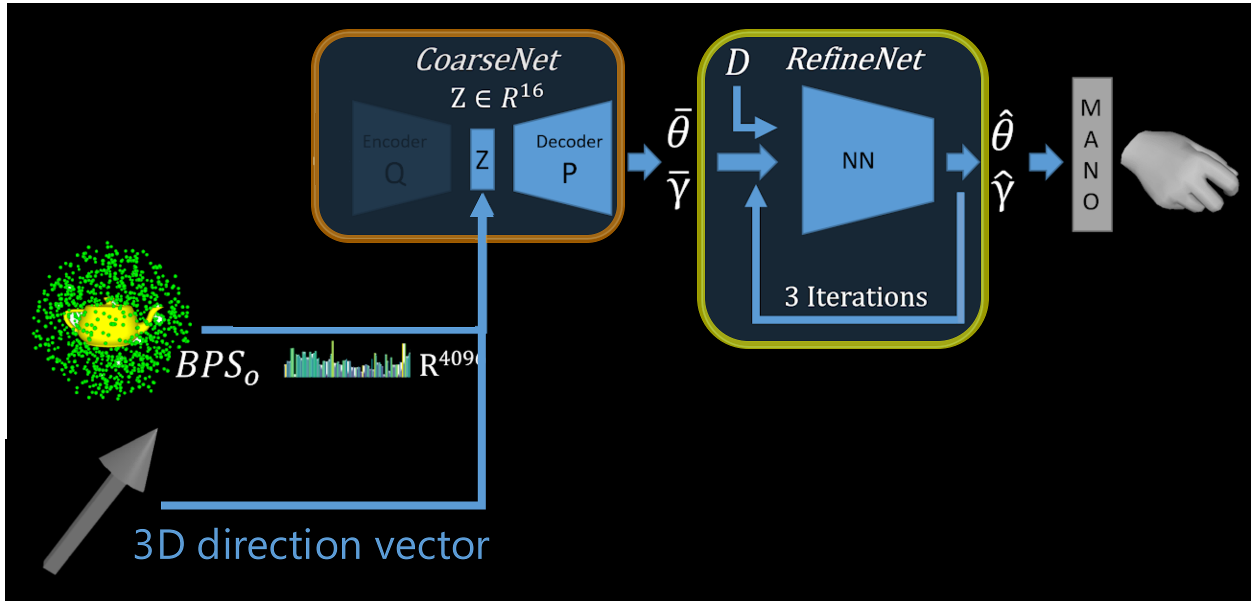
Hands with **random** palm **direction**

[1] **GRAB:** A Dataset of Whole-Body Human Grasping of Object, Taheri et al., ECCV 2020

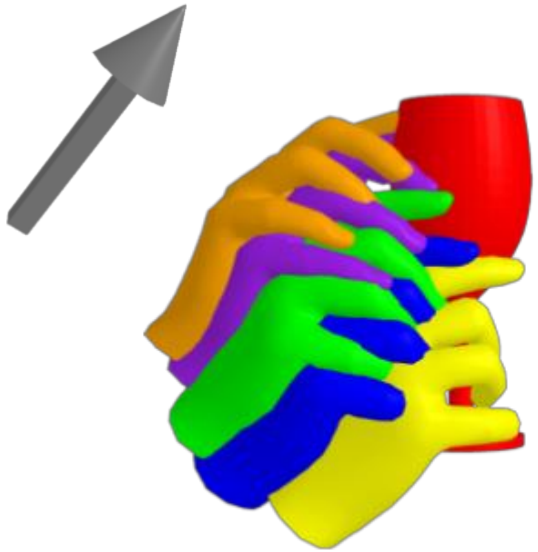


3D Grasp Synthesis

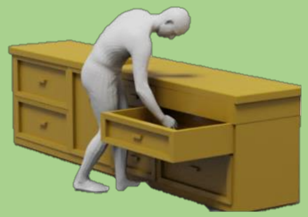
Grasping hands with palm-**direction** control



CGrasp → *'Controllable **Grasp** synthesis'*



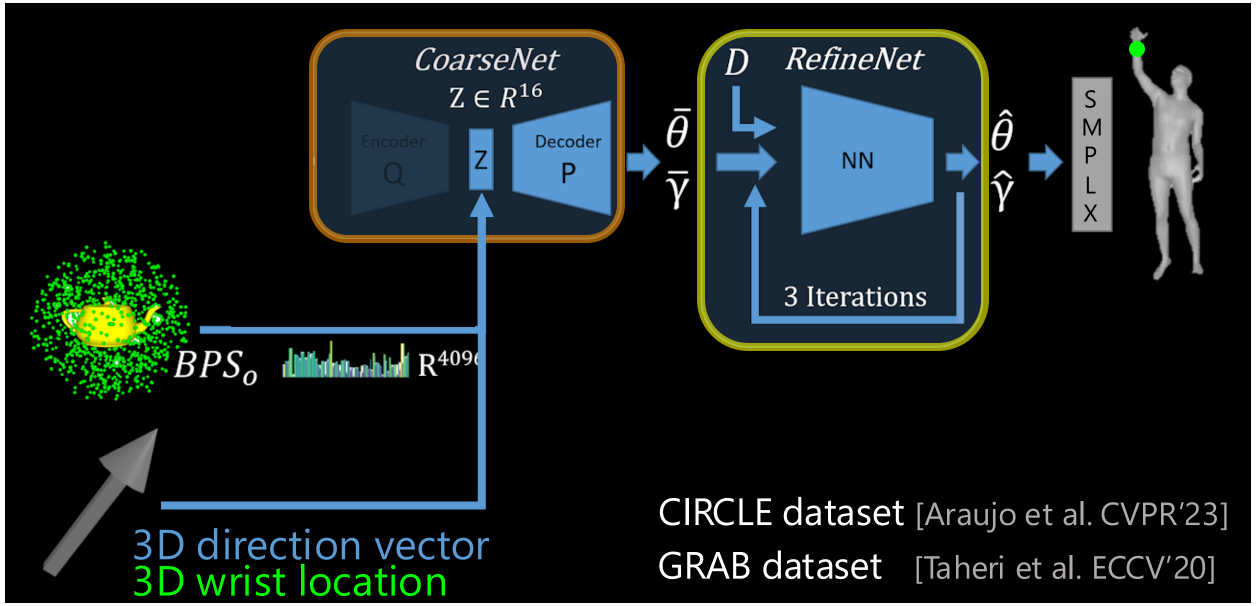
Draw 5 different samples

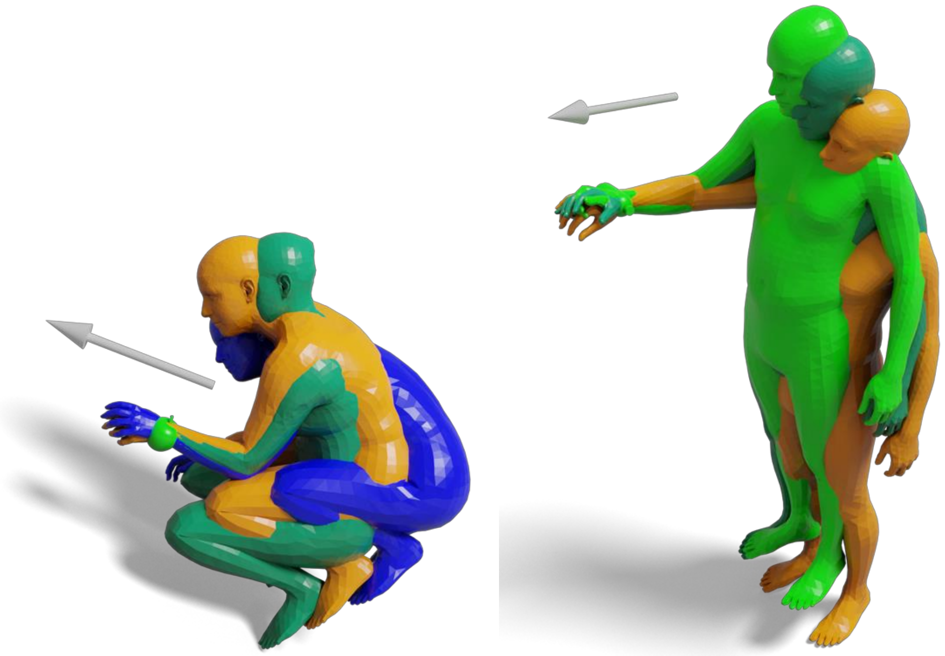


3D Grasp Synthesis

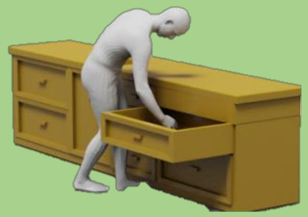
Reaching bodies

with **arm-direction** control





CReach → *'Controllable **Reach** synthesis'*

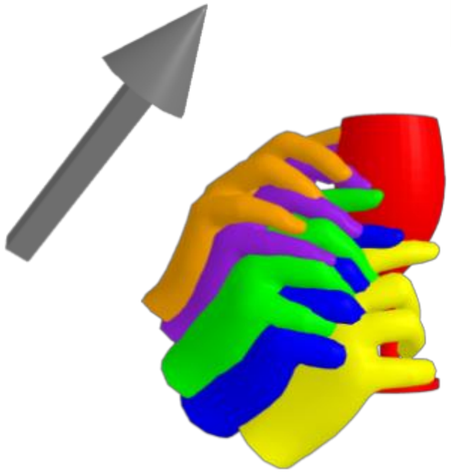


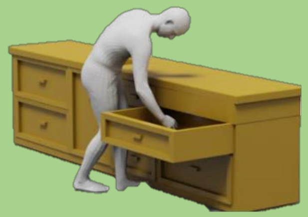
3D Grasp Synthesis

Grasping hands

Reaching bodies

Conditioned on
desired 3D direction



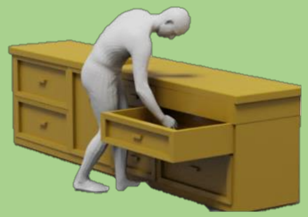


3D Grasp Synthesis

Where does the
3D direction come from ?

Answer the question:

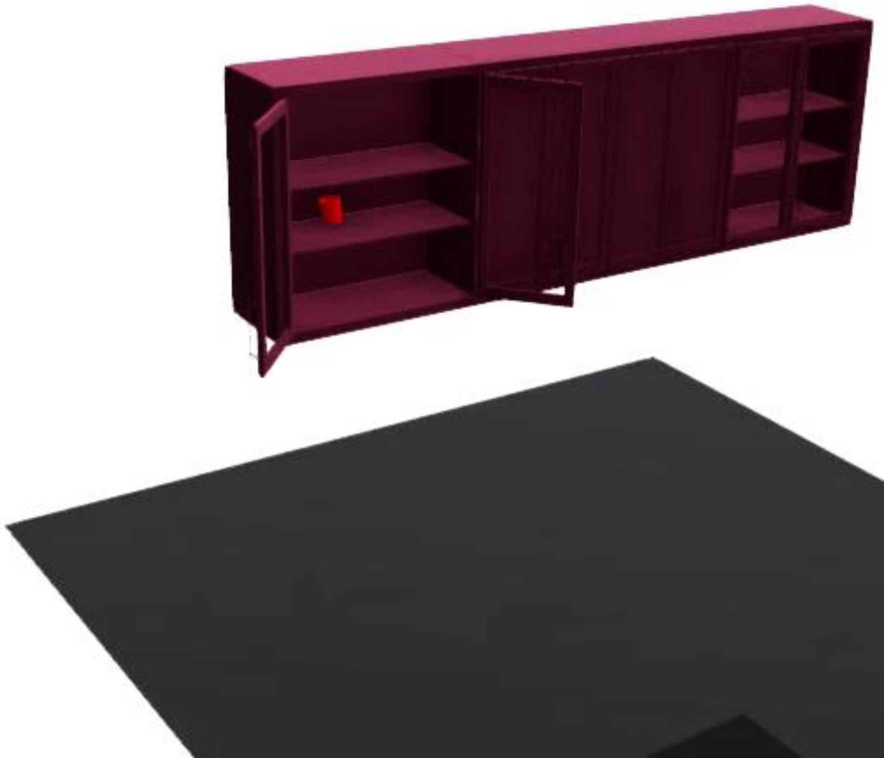
Where can an object be reached from? ?

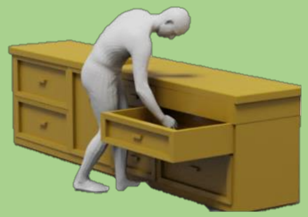


3D Grasp Synthesis

‘Local scene’ reasoning

- Object on a receptacle

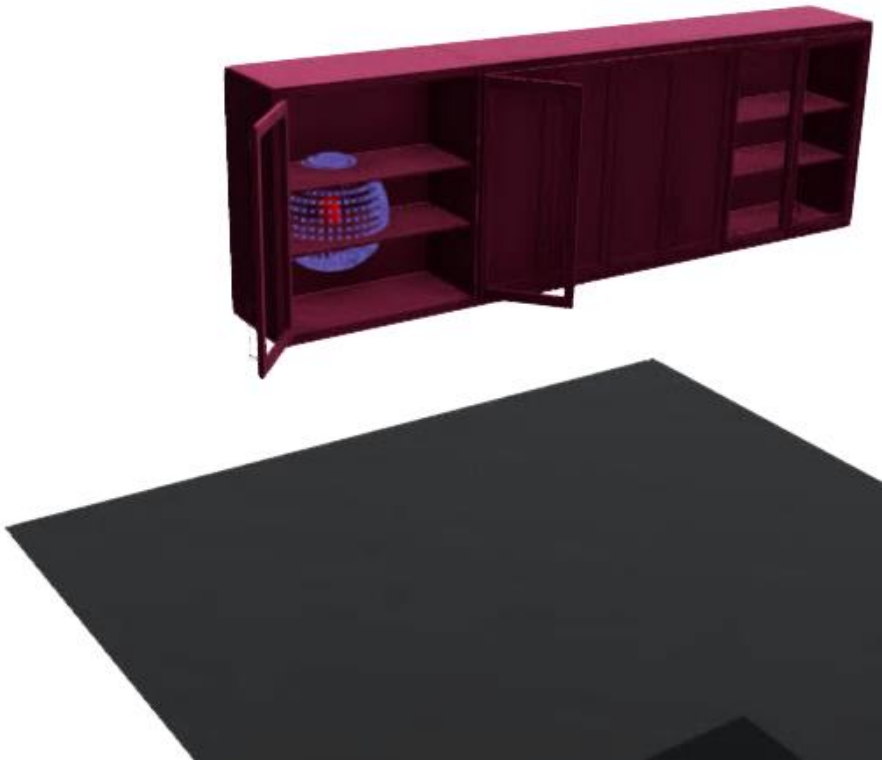


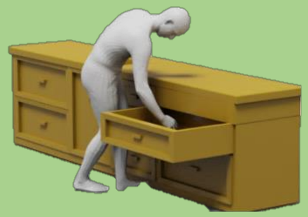


3D Grasp Synthesis

'Local scene' reasoning

- Object on a receptacle
- Sample a sphere around object

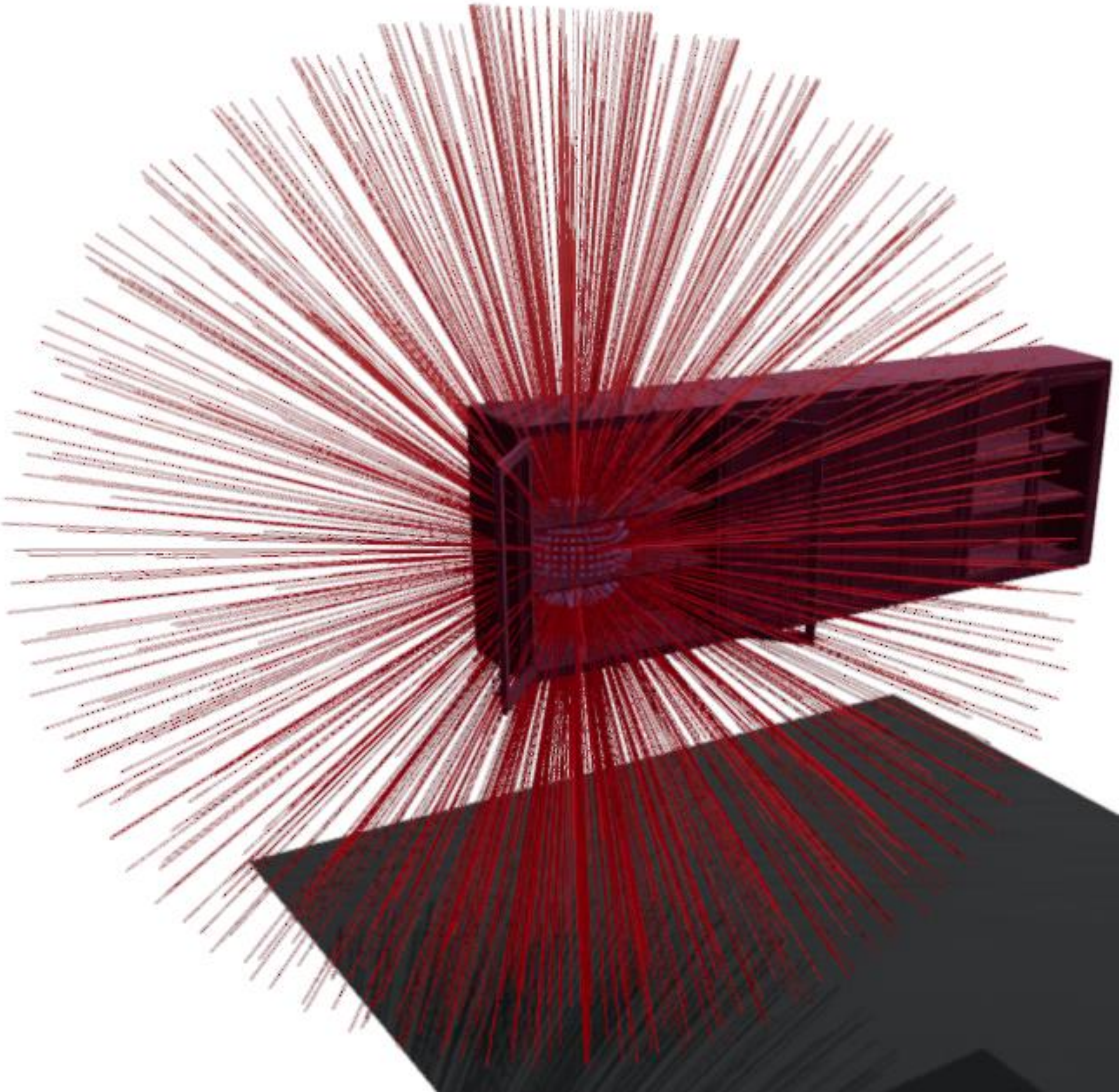


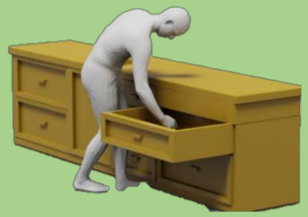


3D Grasp Synthesis

'Local scene' reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**

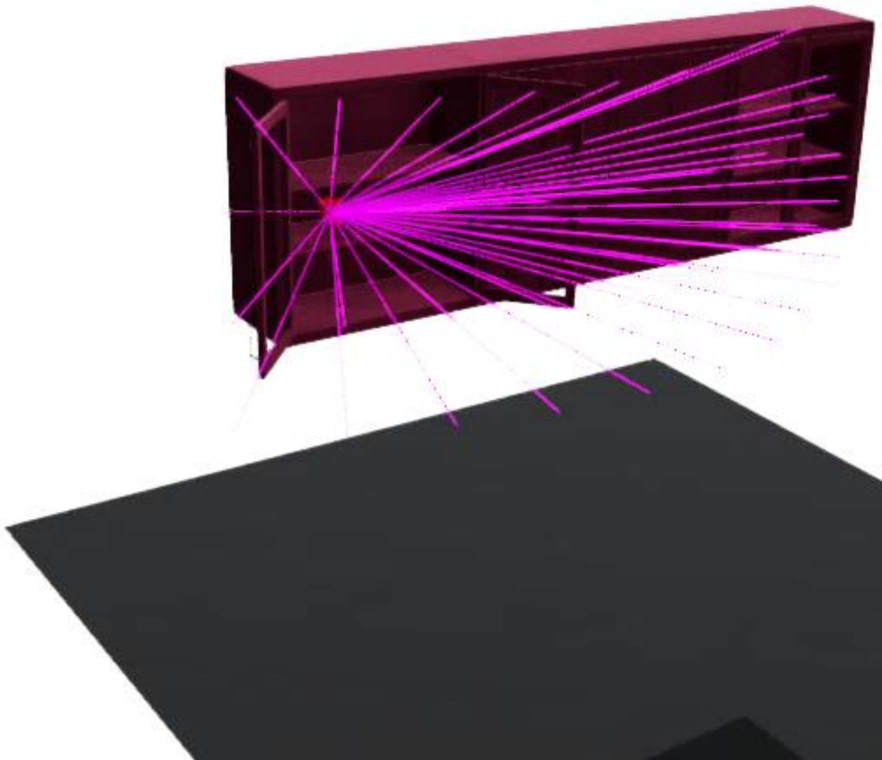


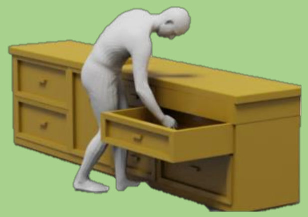


3D Grasp Synthesis

‘Local scene’ reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones

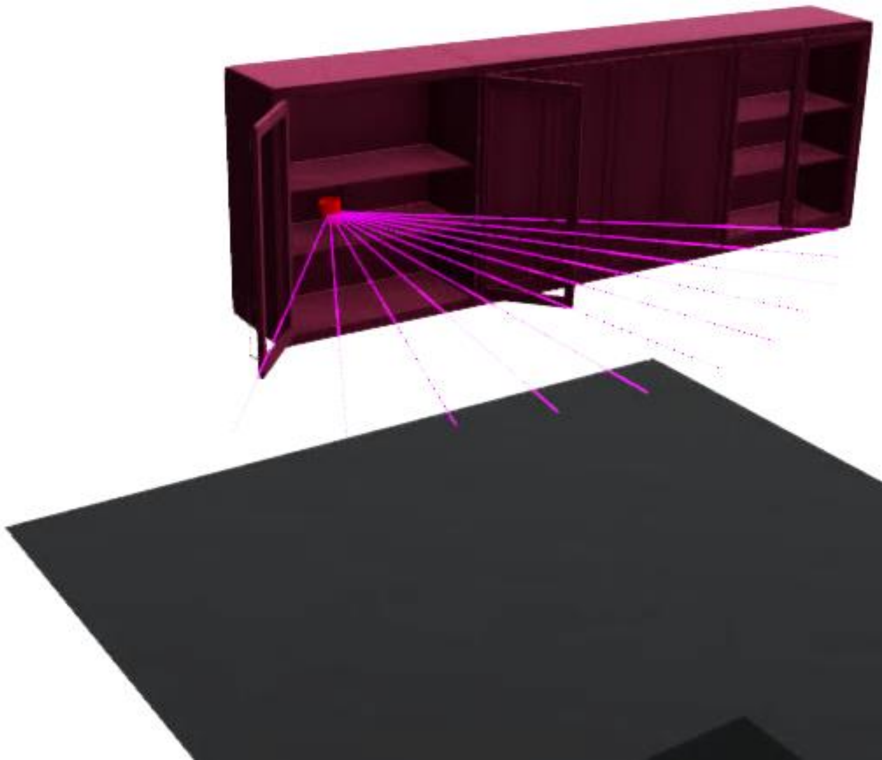


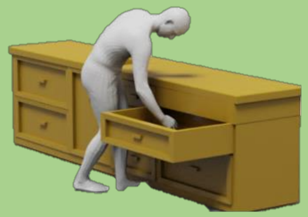


3D Grasp Synthesis

'Local scene' reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones
- Further refined **filtering**:
    - Project rays → parallel to floor

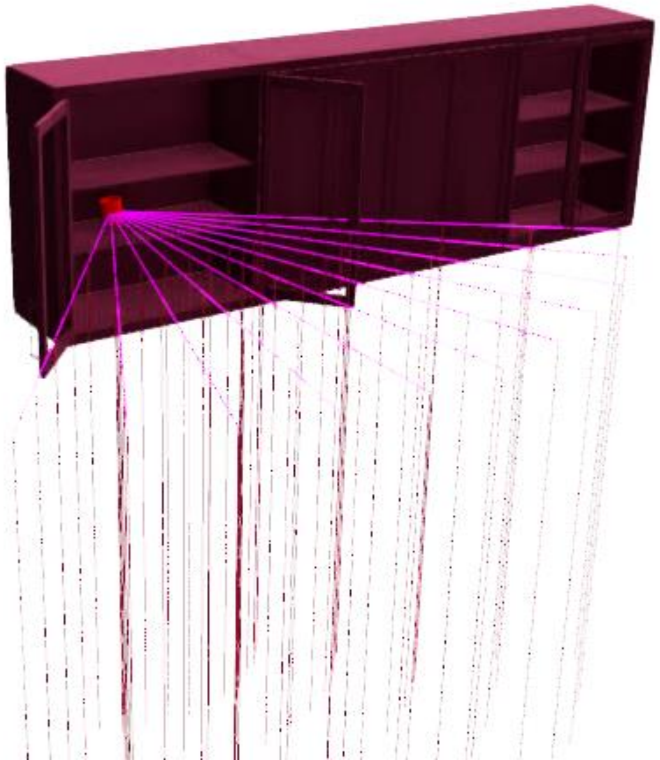


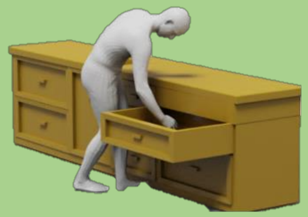


3D Grasp Synthesis

‘**Local scene**’ reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones
- Further refined **filtering**:
    - Project rays → parallel to floor
    - Shoot **rays to floor** → Check **collisions w. receptacles**

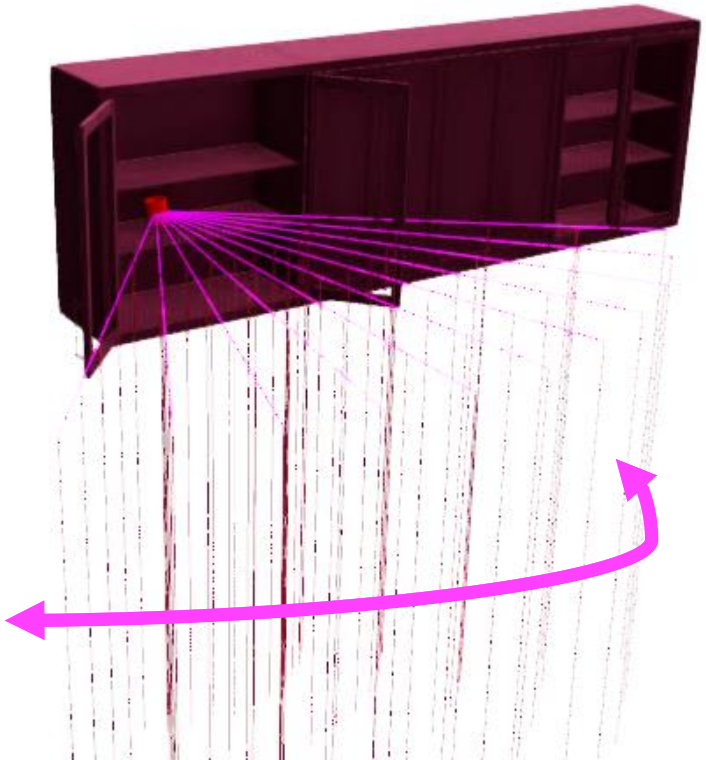


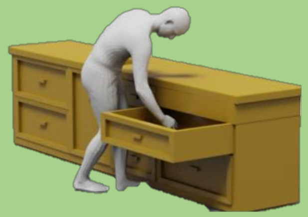


3D Grasp Synthesis

'Local scene' reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones
- Further refined **filtering**:
  - Project rays → parallel to floor
  - Shoot **rays to floor** → Check **collisions w. receptacles**
  - Also **wiggle** around a bit → Check **collisions** (arms take up volume)

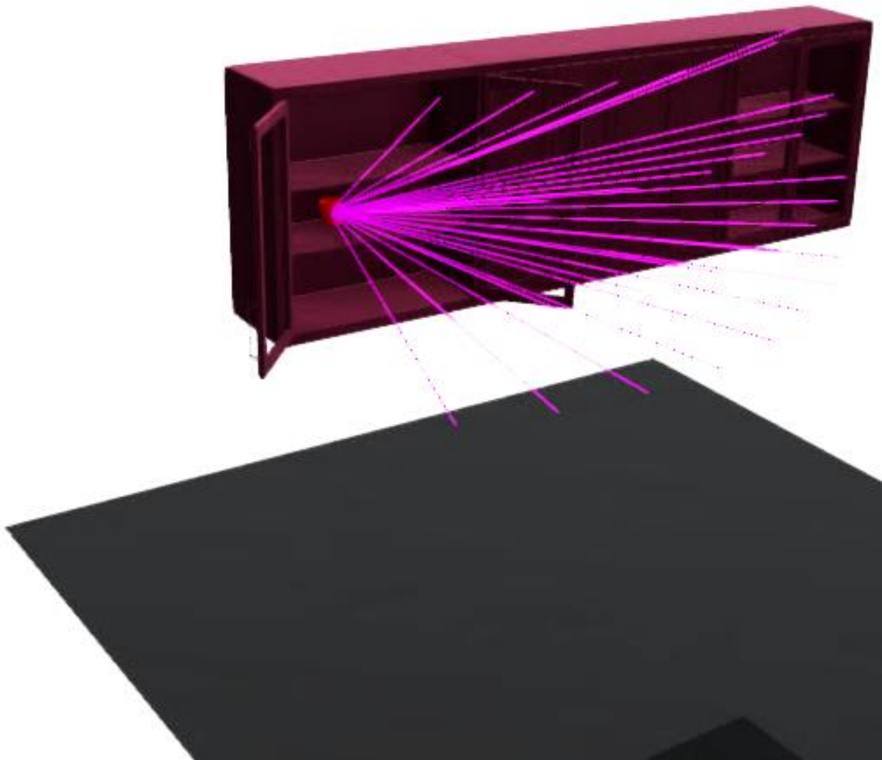


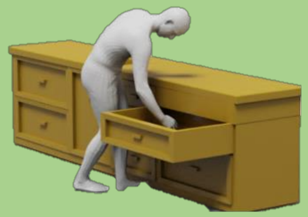


3D Grasp Synthesis

‘Local scene’ reasoning

- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones
- Further refined **filtering**:
  - Project rays → parallel to floor
  - Shoot **rays to floor** → Check **collisions w. receptacle**
  - Also **wiggle** around a bit → Check **collisions** again
- **Final** set of filtered **rays**





3D Grasp Synthesis

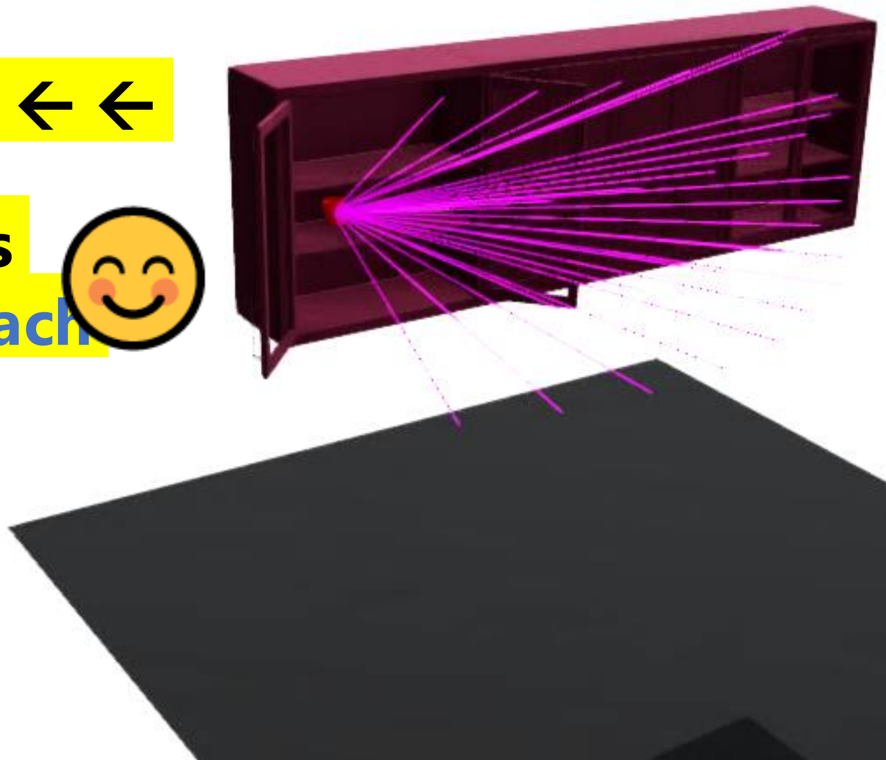
‘**Local scene**’ reasoning

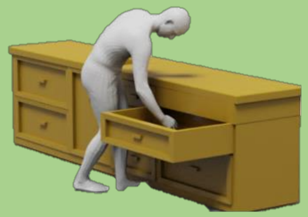
- Object on a receptacle
- Sample a sphere around object
- Shoot **rays**
- Keep only **non-intersecting** ones
- Further refined **filtering**:
    - Project rays → parallel to floor
    - Shoot **rays to floor** → Check **collisions w. receptacle**
    - Also **wiggle** around a bit → Check **collisions** again
- **Final** set of filtered **rays**

ReachingField

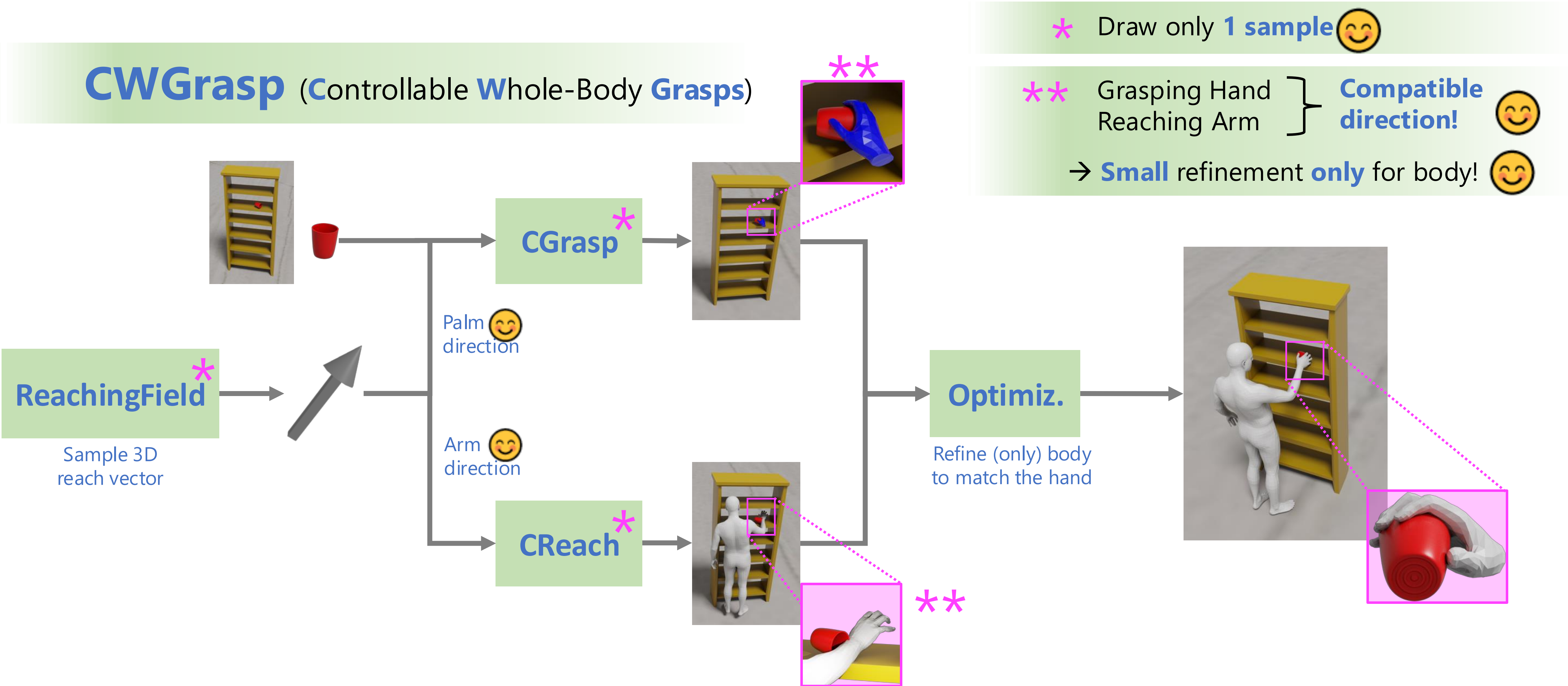
→ → Sample a 3D ray ← ←

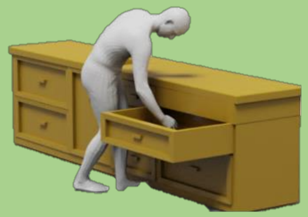
Condition on this both CGrasp & CReach 😊





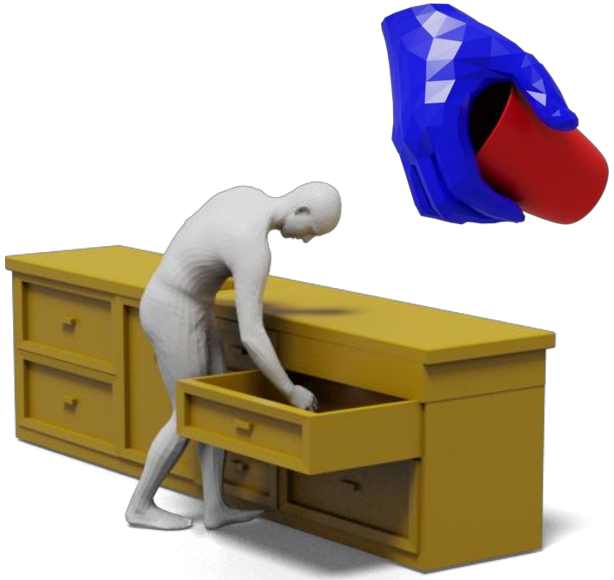
3D Grasp Synthesis

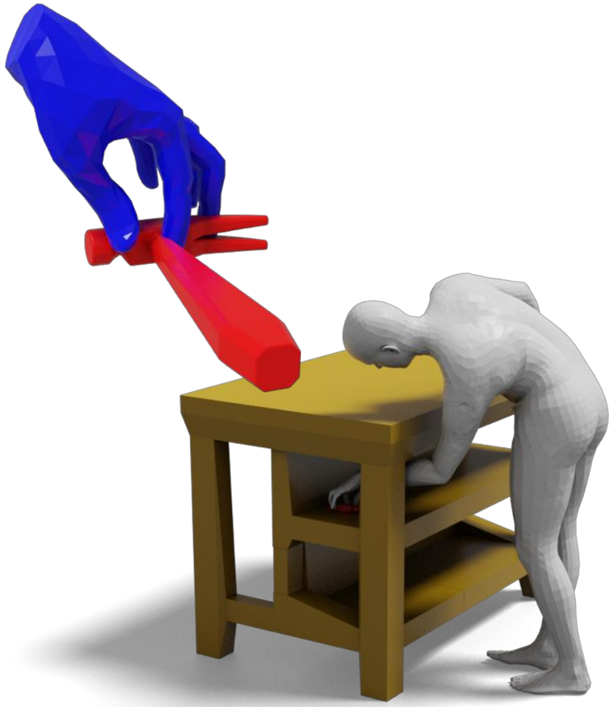


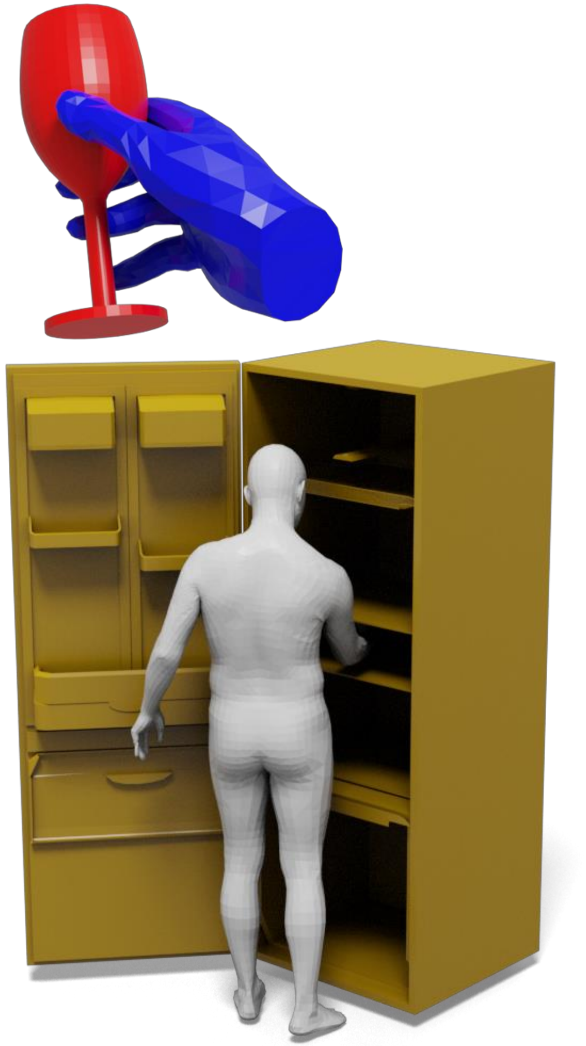


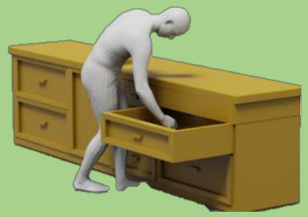
3D Grasp Synthesis

Medium-height positions 😊



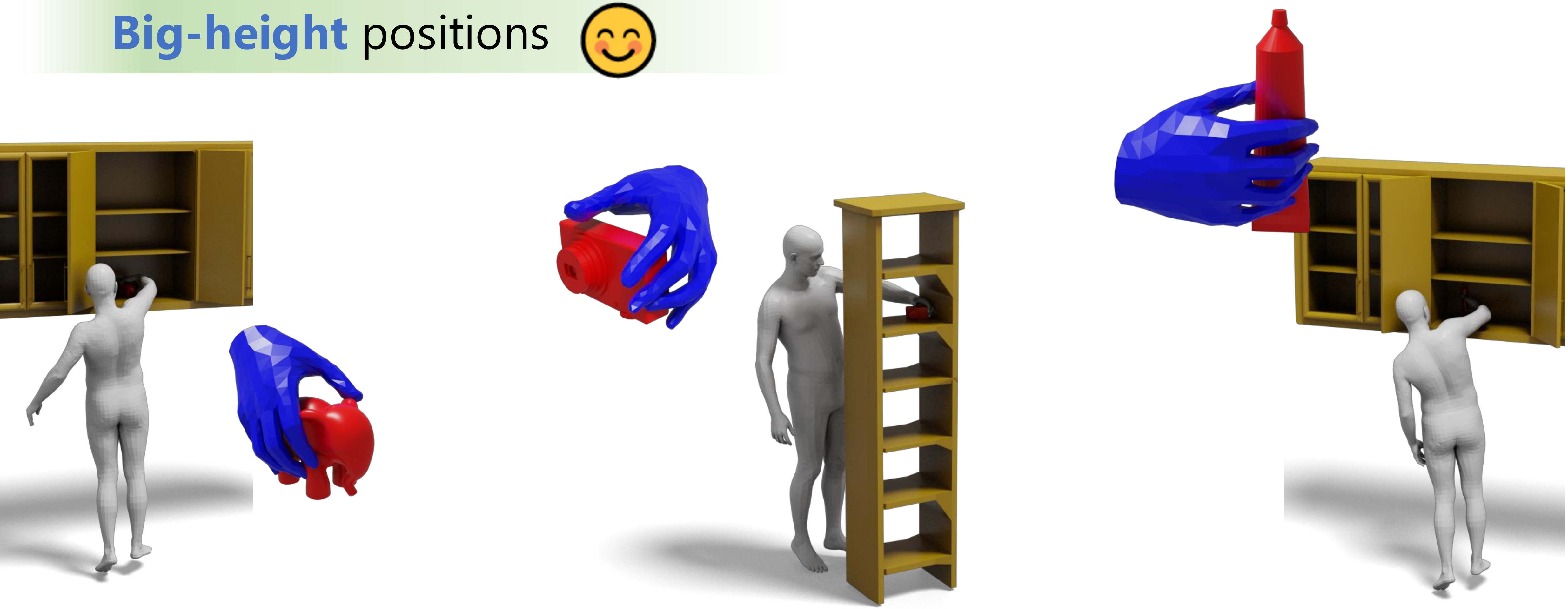


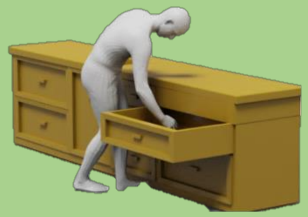




3D Grasp Synthesis

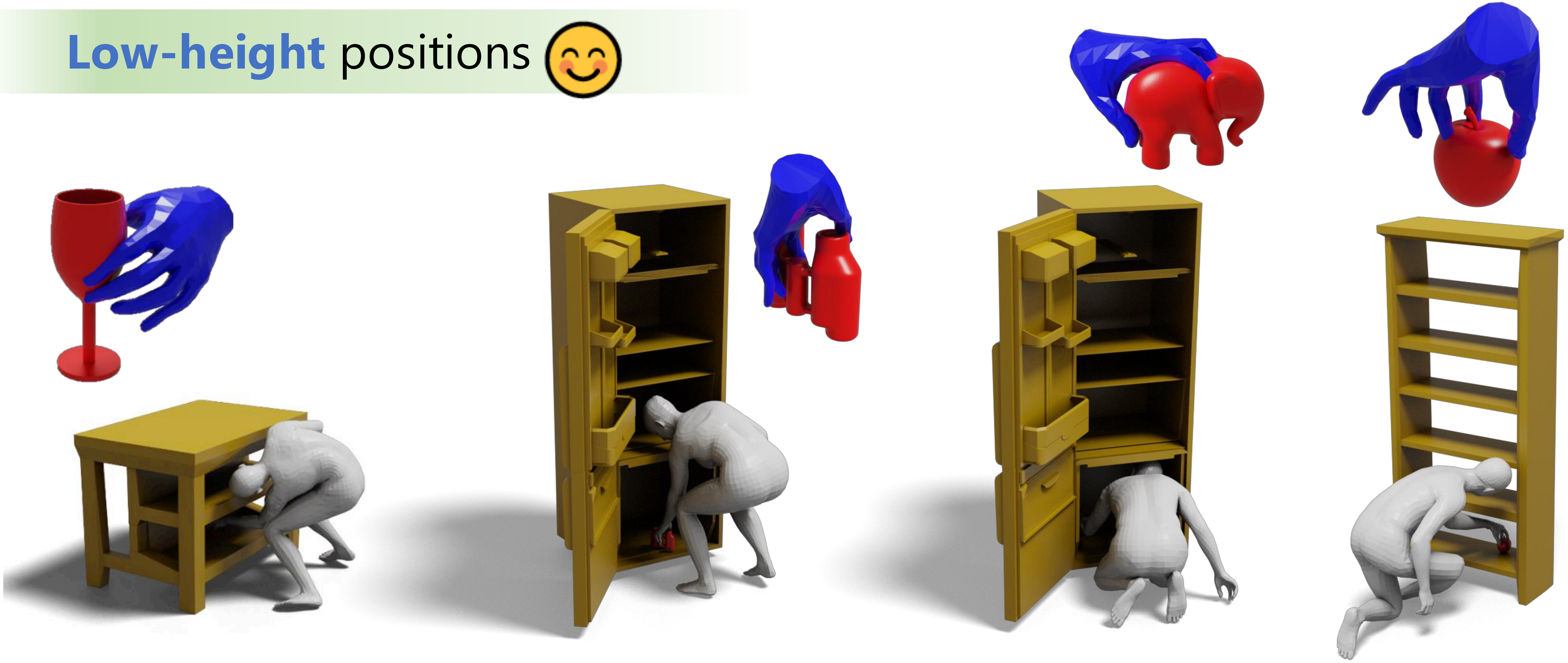
Big-height positions 😊





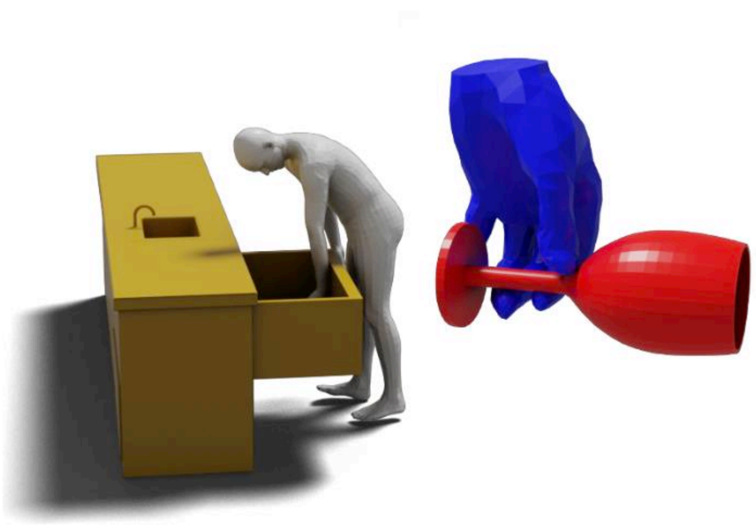
3D Grasp Synthesis

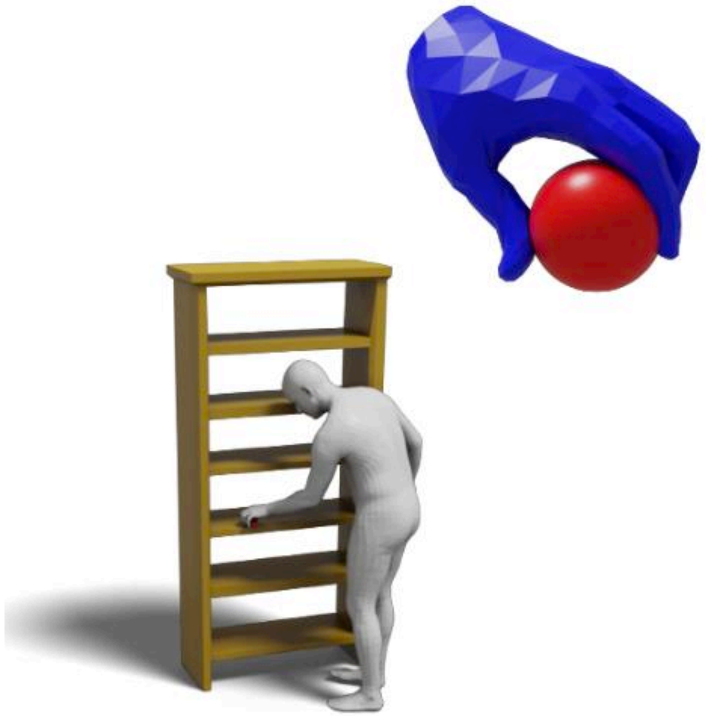
Low-height positions 😊

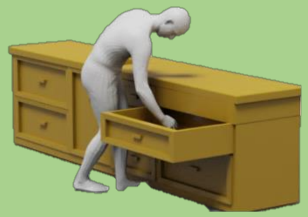


3D Grasp Synthesis

Left-hand grasps 😊







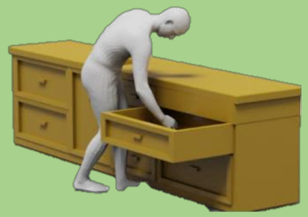
3D Grasp Synthesis

CWGrasp
(ours)

VS

FLEX [1]

[1] **FLEX:** Full-Body Grasping Without Full-Body Grasps. Tendulkar et al. , CVPR 2023



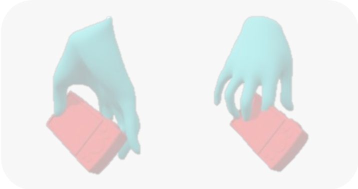
3D Grasp Synthesis

Perceptual study: **CWGrasp** vs **FLEX**

33 participants (+2 filtered out)

24 samples (+4 catch trials)

Each sample (gif): Full-body & hand-zoom view
(rated separately – random order)

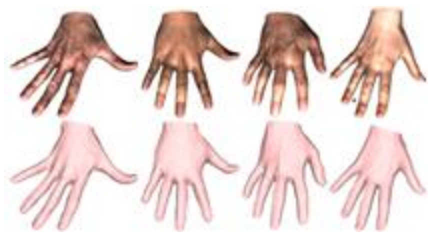


Participants preferring **CWGrasp**:

Full-body view	**70.8%**
Hand-zoomed view	**71.6%**
Both views	**71.23%**

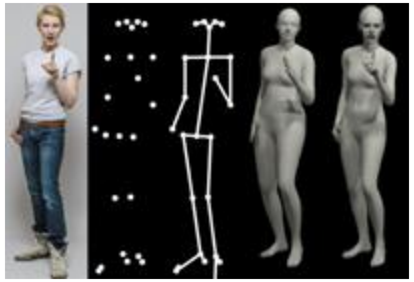
	#Body Samples ↓	Time (sec) ↓
FLEX	500	357
CWGrasp	1	23
	500x less	10x less

Research Map



MANO
SIGGRAPH-Asia'17

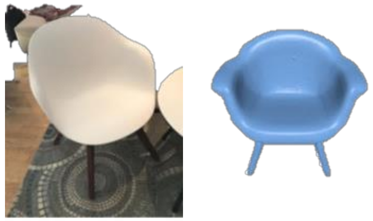
SMPL+H
SIGGRAPH-Asia'17



SMPL-X & SMPLify-X
CVPR'19

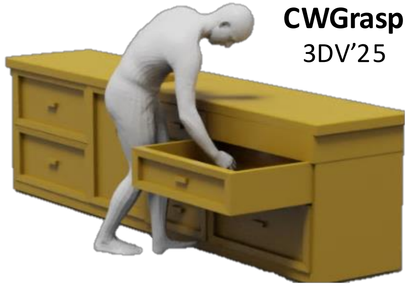
PROX
ICCV'19

Summary



SDFit
arXiv 2024

GRAB ECCV'20 **ARCTIC** CVPR'23

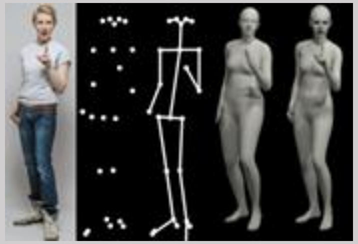


CWGrasp
3DV'25

Publications / Preprints

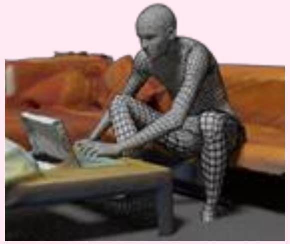
[Embodied Hands: Modeling and Capturing Hands and Bodies Together]

J. Romero* · **D. Tzionas*** · M. J. Black
SIGGRAPH-Asia 2017



[Expressive Body Capture: 3D Hands, Face, and Body from a Single Image]

G. Pavlakos* · **V. Choutas*** · N. Ghorbani · T. Bolkart · A.A.A. Osman · D. Tzionas · M. J. Black
CVPR 2019

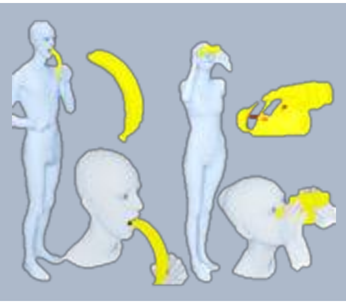


[Resolving 3D Human Pose Ambiguities with 3D Scene Constraints]

M. Hassan · V. Choutas · D. Tzionas · M. J. Black
ICCV 2019

Publications / Preprints

SDFit: 3D Object Shape and Pose by Fitting a Morphable SDF to a Single Image
Dimitrije Antić · S. K. Dwivedi · S. Tripathi · T. Gevers · D. Tzionas
arXiv, Sep. 2024



GRAB: A Dataset of Whole-Body Human Grasping of Objects
Omid Taheri · N. Ghorbani · M. J. Black · D. Tzionas
ECCV 2020

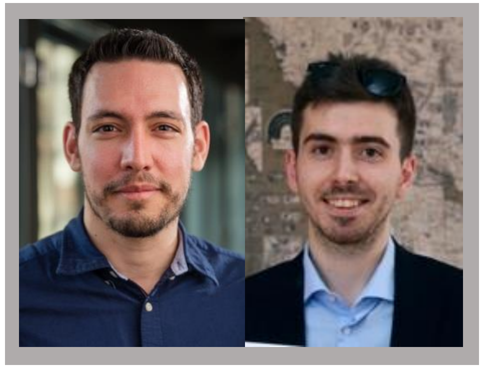
3D Whole-body Grasp Synthesis with Directional Controllability
Georgios Paschalidis · R. Wilschut · D. Antić · O. Taheri · D. Tzionas
3DV 2025

Collaborators

First Authors

Dimitrije Antić

Georgios Paschalidis



George Pavlakos

Vasileios Choutas

Javier Romero

Dimitris Tzionas

Sai K. Dwivedi

Romana Wilschut

Nima Ghorbani

Mohamed Hassan

Theo Gevers

Shashank Tripathi

Timo Bolkart

Omid Taheri

Ahmed A.A. Osman

Michael J. Black

Towards...

3D Human-centric Perception and Synthesis

Work funded by:

erc

ellis
UNIT AMSTERDAM

MAX PLANCK
GESELLSCHAFT

dtzionas.com

@dimtzionas

@dimtzionas