Dexterous Grasping Pose Generation for Multi-Finger Hand

Lei Zhang

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme

Nov.2023

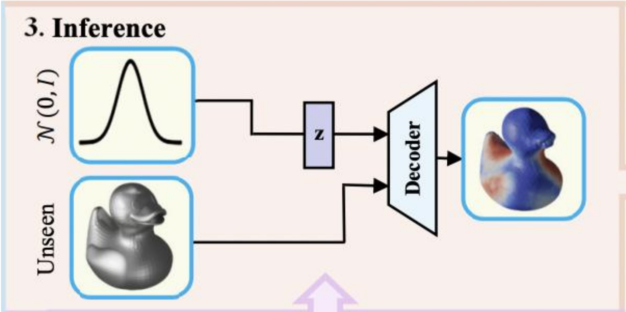
Five Finger Hand grasping

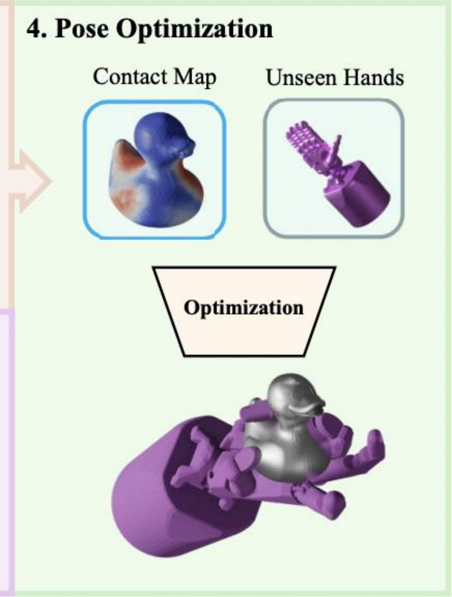
Shadow Hand

DLR-HIT II Hand

Reference: GenDexGrasp-->Grasp Generation for FFH

Generalized Contact Map --> Grasp Generation using optimization method





Hand-object representation
Contact map: distance

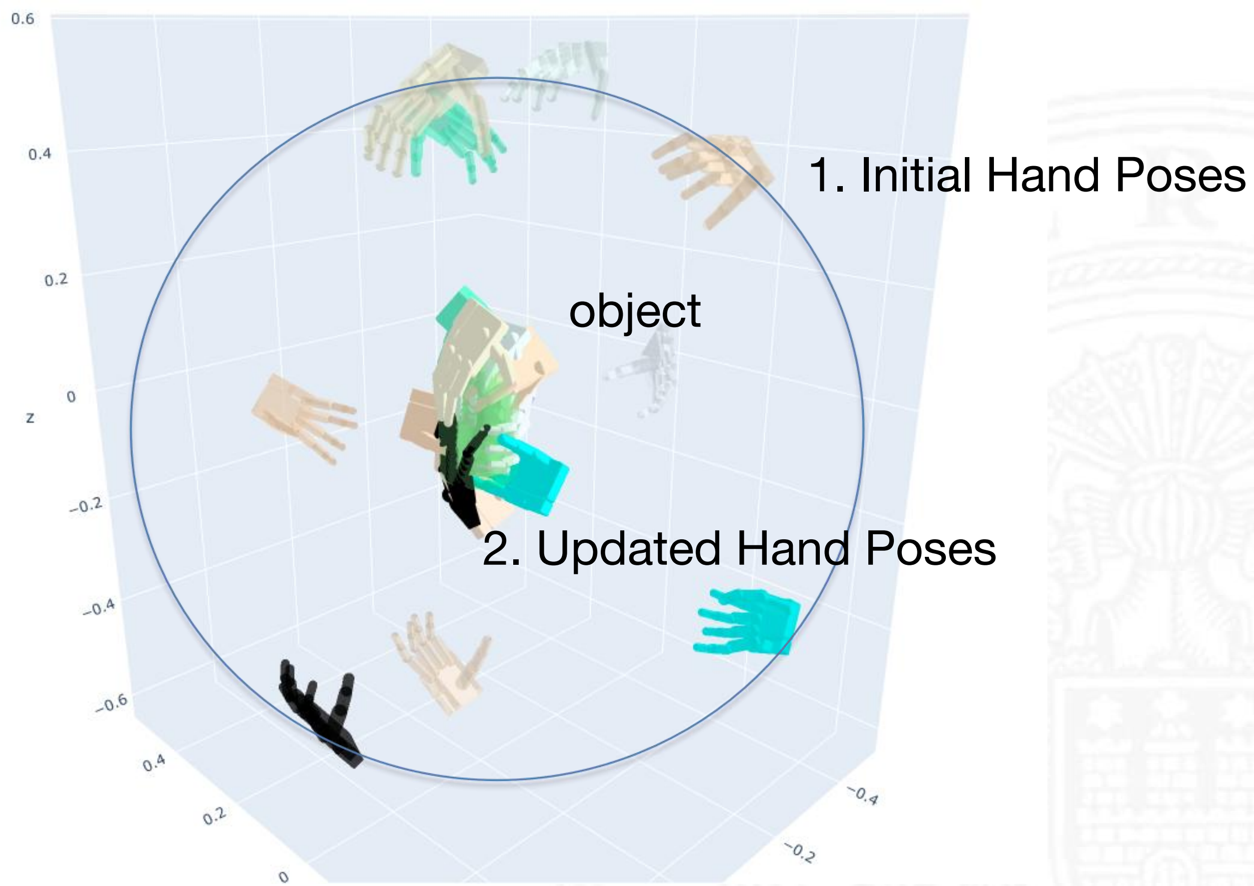
GenDexGrasp: Generalizable Dexterous Grasping
DexGraspNet: A Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation

1. Data Generation of FFH Grasping Dataset

Data Generation: Object Dataset

over 1,700 objects from 3dNet, the Yale-CMU-Berkeley (YCB) Dataset, Princeton ModelNet, Dex-Net, and the MVTec Industrial 3D Object Detection Dataset (MVTec ITODD)

Data Generation: Generate Grasping Poses for all objects



Loss: Energy Function

$$E = E_{\text{fc}} + w_{\text{dis}} E_{\text{dis}} + w_{\text{pen}} E_{\text{pen}} + w_{\text{prior}} E_{\text{prior}}$$

GenDexGrasp: Generalizable Dexterous Grasping
DexGraspNet: A Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation

Grasping Validation: Isaac gym

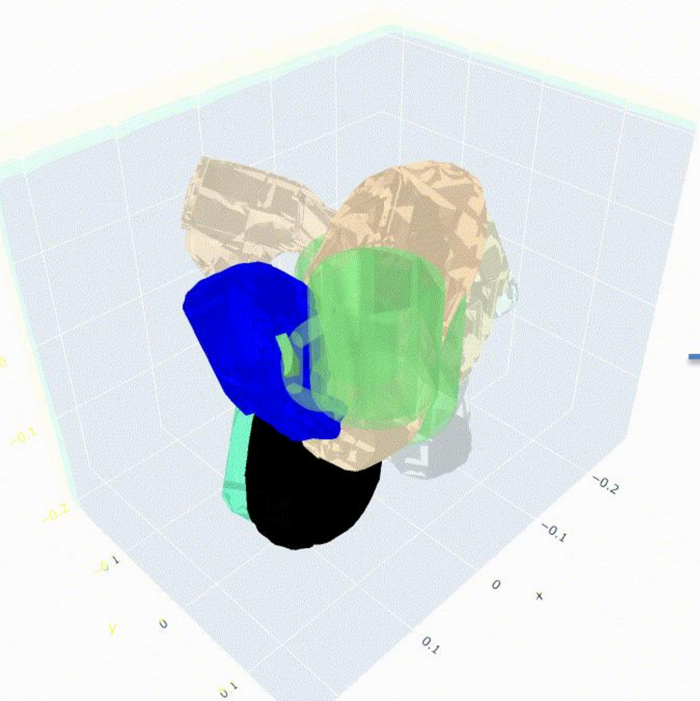
Place hand with initial pose with small offset related to the goal pose

Close hand

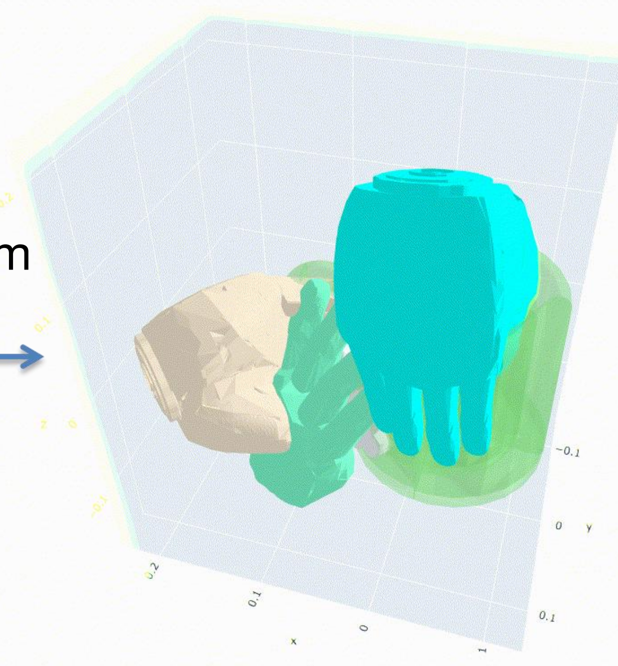
Check whether grasping success

GenDexGrasp: Generalizable Dexterous Grasping
DexGraspNet: A Large-Scale Robotic Dexterous Grasp Dataset for General Objects Based on Simulation

Validation Performance with Isaac Gym

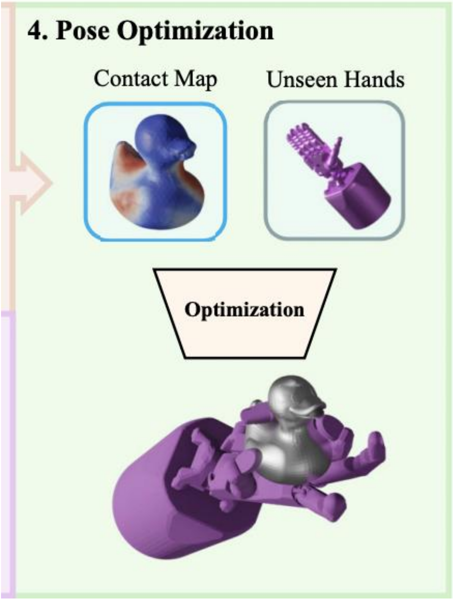


validation in isaac gym

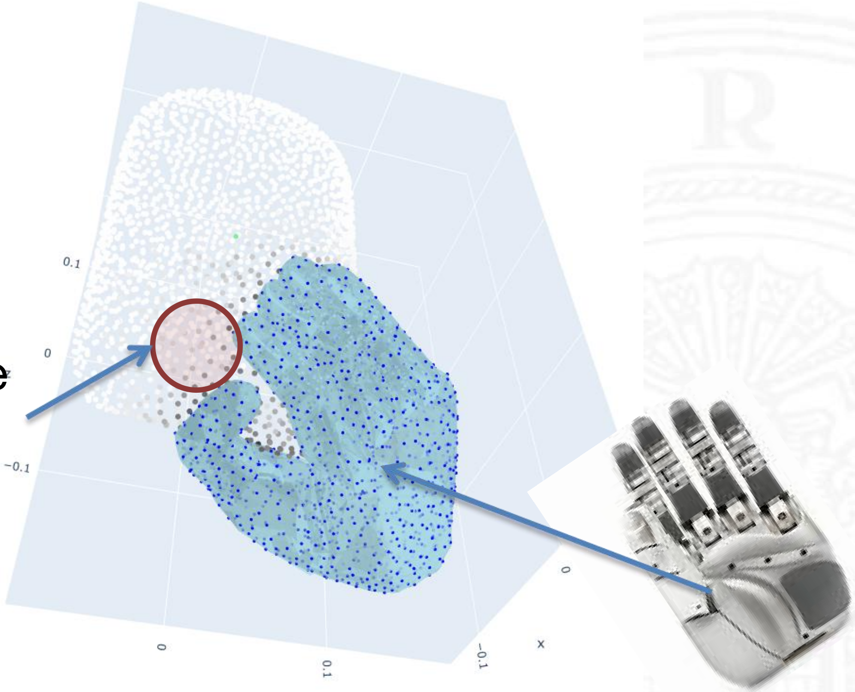


Result based on DLR-HIT hand

GenDexGrasp-->Contact Map as prior information



Distance value
Black points:
close to hand

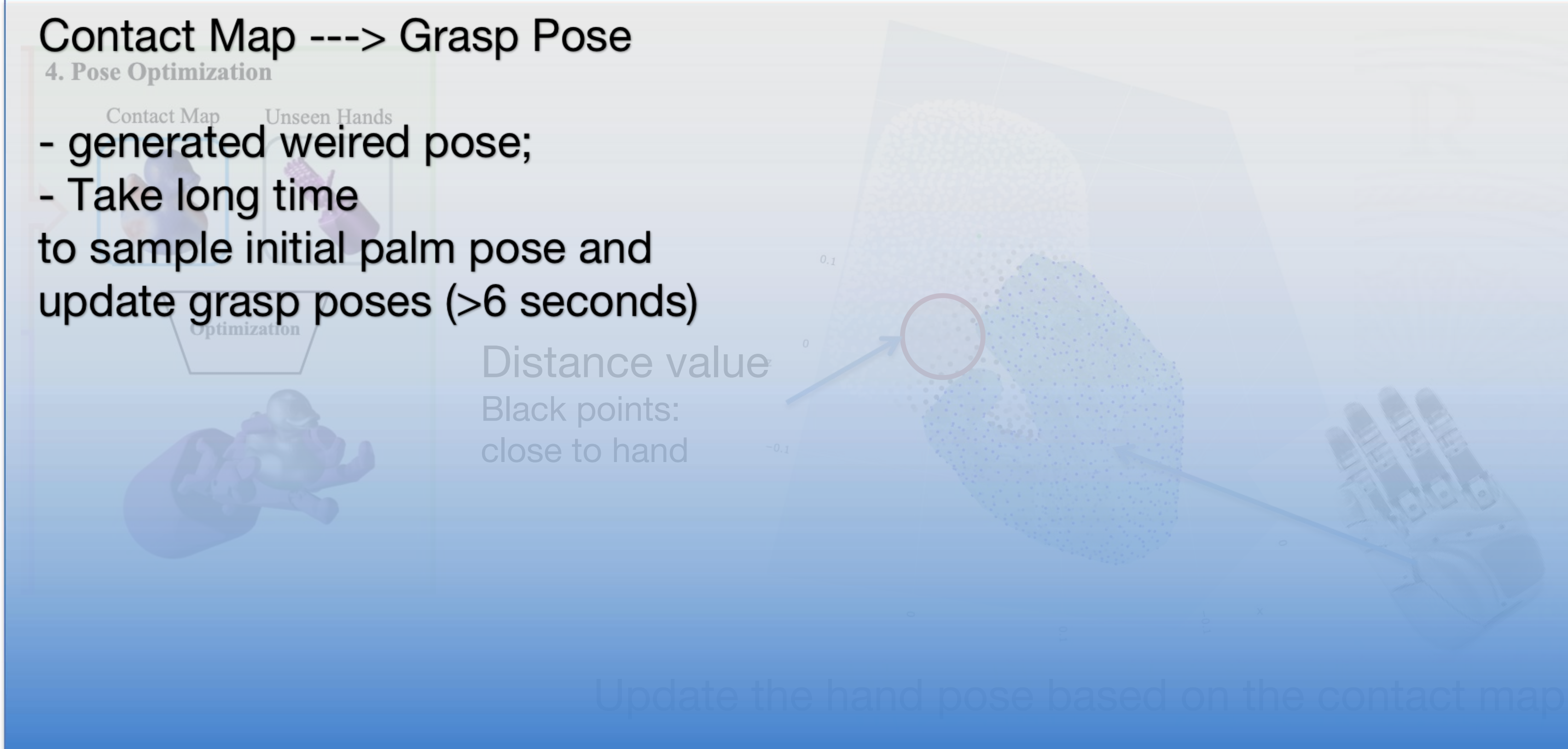


Update the hand pose based on the contact map

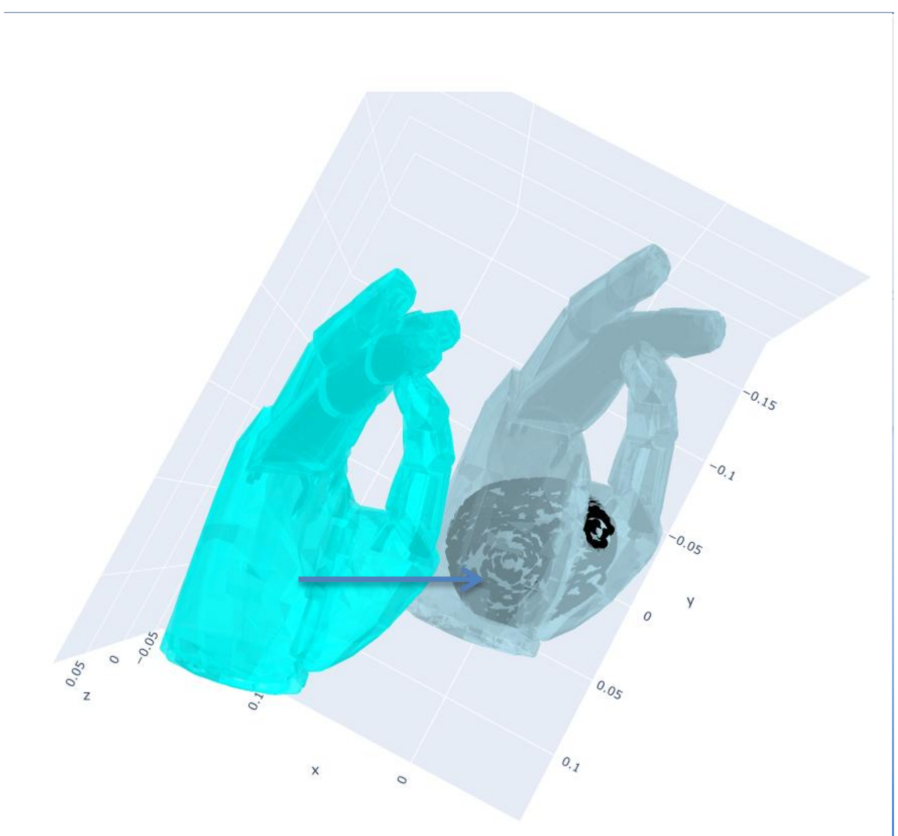
Contact Map: Disadvantages

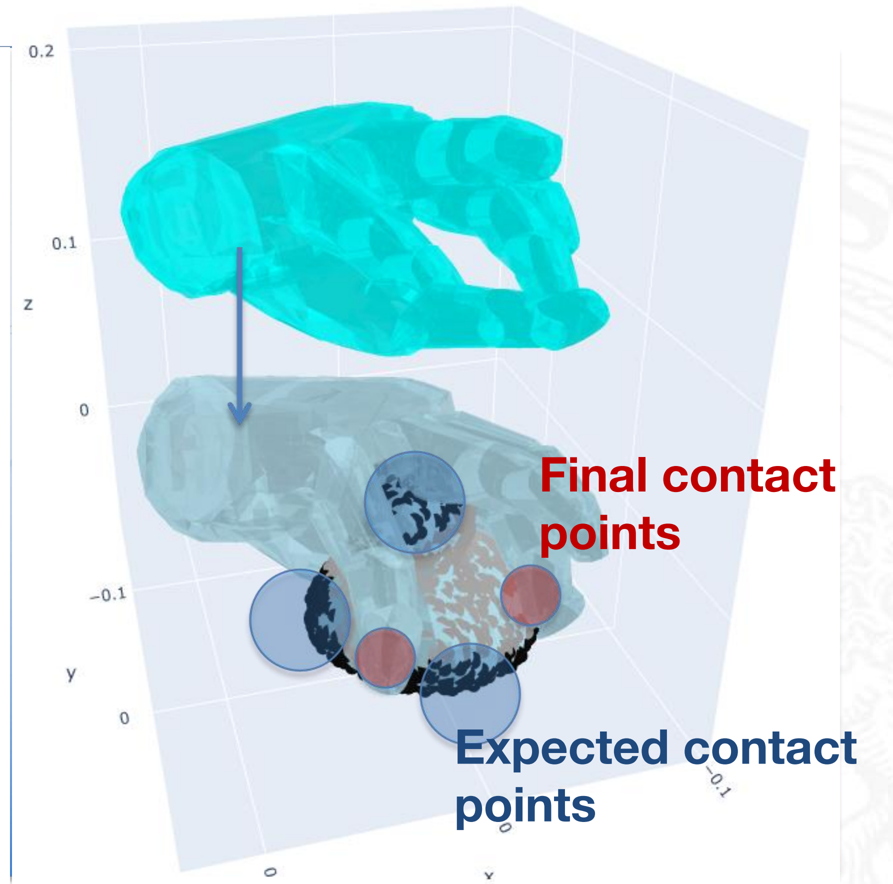
Contact Map ---> Grasp Pose

- generated weired pose;
- Take long time
to sample initial palm pose and
update grasp poses (>6 seconds)



Disadvantages of Grasping Optimization



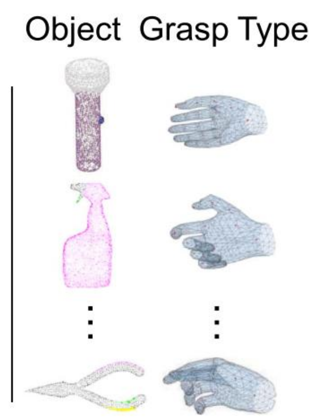


2. Different hand object representations

Reference: Different Hand-Object Representations: Touch Code

Touch Code

Thumb	Index	Middle	Ring	Little	Palm	Point Color
① ② ③	④ ⑤ ⑥	⑦ ⑧ ⑨	⑩ ⑪⑫	⑬⑭⑮	⑯	
$\boxed{1}$ 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0	Blue
0 0 0	1 0 0	1 0 0	1 0 0	1 0 0	0	Purple



(a) (b) (c) (d) (e)

Relationship between
functional grasping and **touch code**

Reference: FunctionalGrasp: Learning Functional Grasp for Robots via Semantic Hand-Object Representation, Dalian University of China

Reference: Hand Object Representations: Touch Code

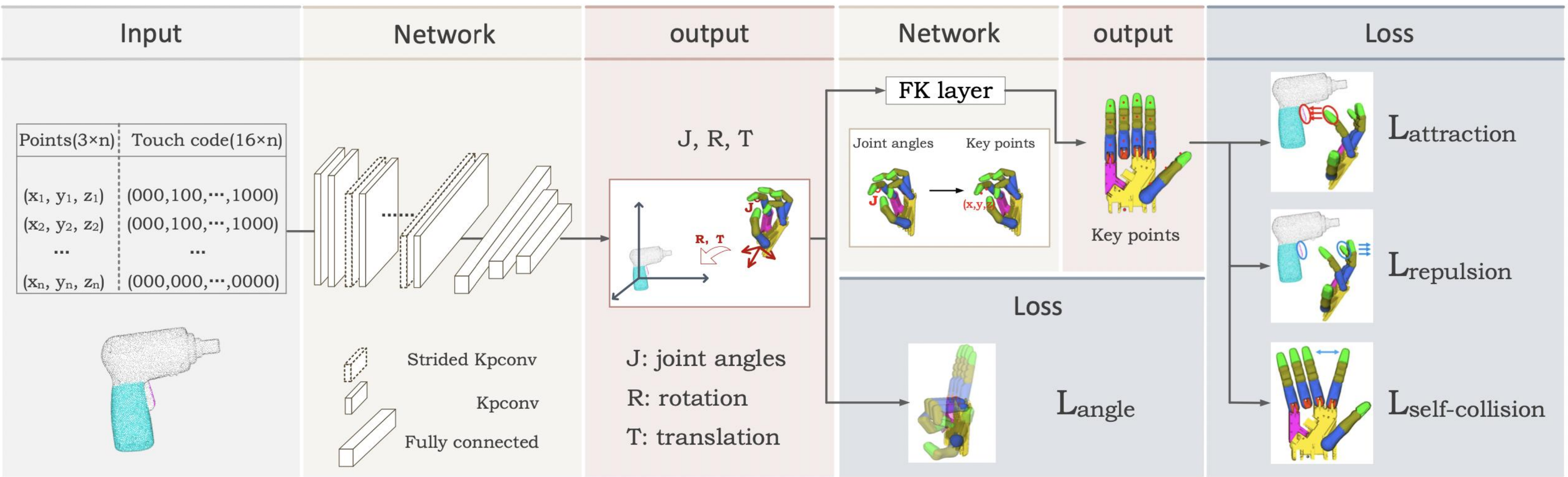
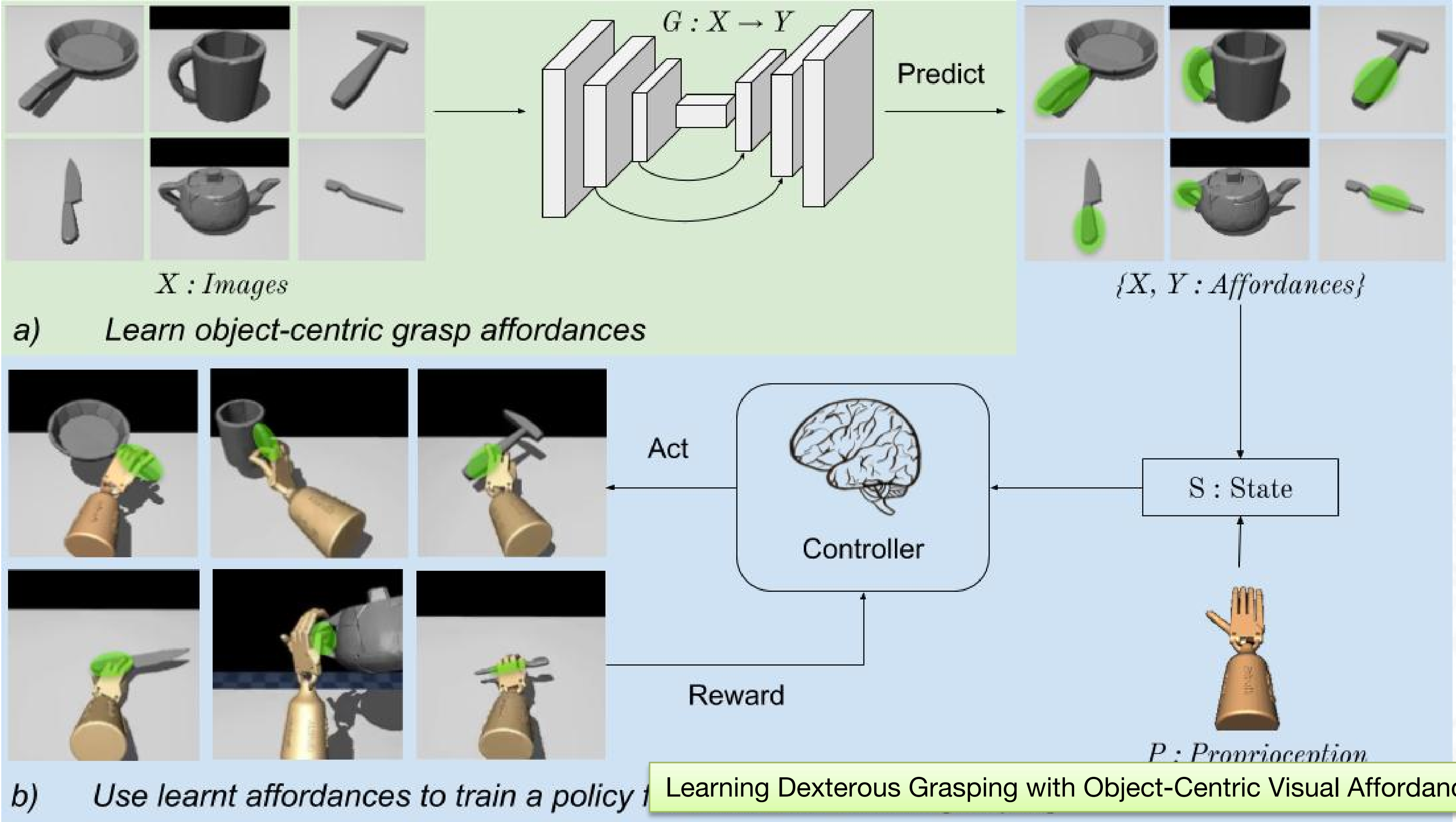


Figure 4: The overall architecture of our functional grasp synthesis framework. The original point cloud of the object with the 16-bit 'touch code' is fed into the network, which generates the configurations of the hand that conform to the functional grasp under the guidance of four loss functions.

Toward Human-Like Grasp: Dexterous Grasping via Semantic Representation of Object-Hand

Reference: Hand Object Representations: Affordance Regions



Learning Dexterous Grasping with Object-Centric Visual Affordances

Reference: Hand Object Representations: Affordance Regions

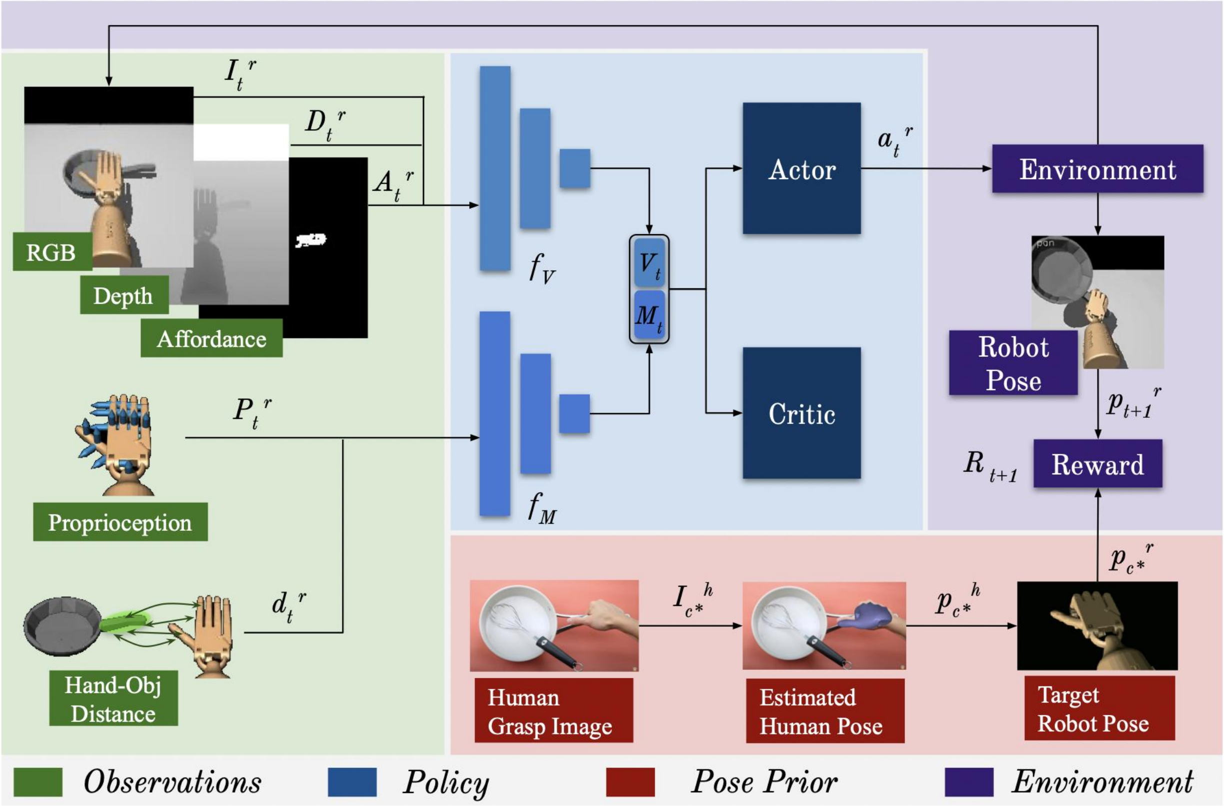


Figure 2: **Overview of DexVIP.** We use grasp poses inferred from Internet video to train a dexterous grasping policy. An actor-critic network (blue) processes sensory observations from visual and motor streams (green) to estimate agent action. Human hand pose priors derived from how-to videos (red) encourage the agent to explore worthw

DexVIP: Learning Dexterous Grasping with Human Hand Pose Priors from Video, University of Texas at Austin + Facebook AI

Reference: Hand Object Representations: Affordance Score

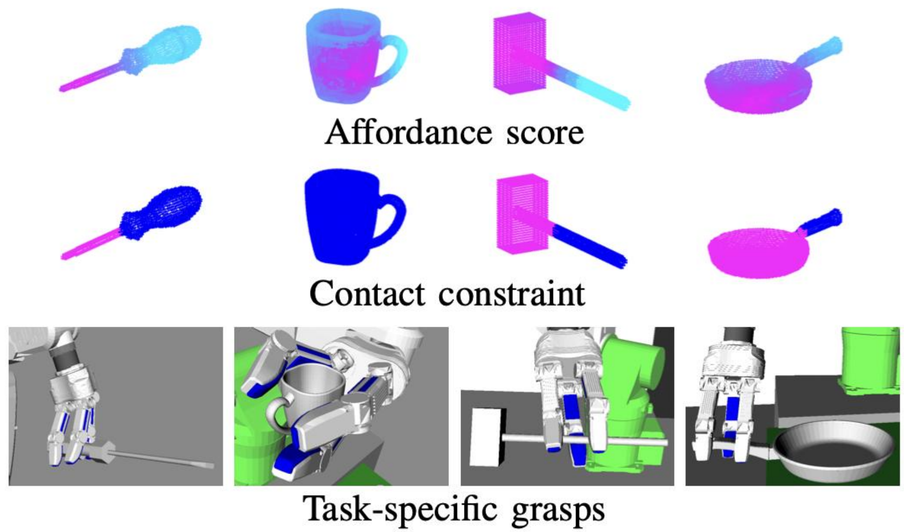


Fig. 1: Given the shape of an object and a task, we detect object part affordances. From these we formulate grasp constraints, such as a contact location constraint. These constraints are then utilized to compute task-specific grasps as shown here for example tasks poke, pour, pound and support on the objects screwdriver, mug, hammer and pan respectively. Magenta color indicates high affordance score (top) and contact avoidance constraint for grasping (middle).

Affordance Detection for Task-Specific Grasping Using Deep Learning

Reference: Hand Object Representations: Signed Distance

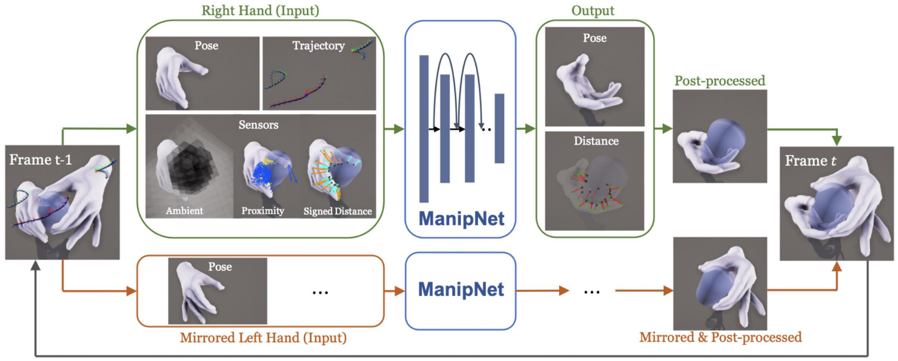


Fig. 2. The outline of our framework. Given the poses of two hands, the shapes of objects as well as the trajectories of two wrists and objects at frame $t-1$. The inputs of the two hands will be generated separately and fed into a shared neural network. Correspondingly, the poses for the two hands at frame t will be synthesized from the outputs of the neural network.

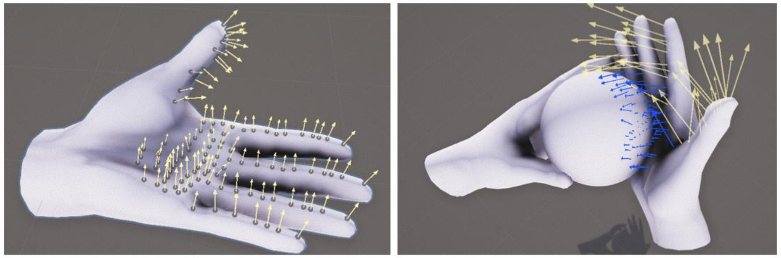


Fig. 4. Left: The 104 Proximity sensors on the hand mesh. Right: Proximity Sensors cast rays along the hand surface normal until they hit the object surface (blue arrows), or at a maximum distance (yellow arrows).

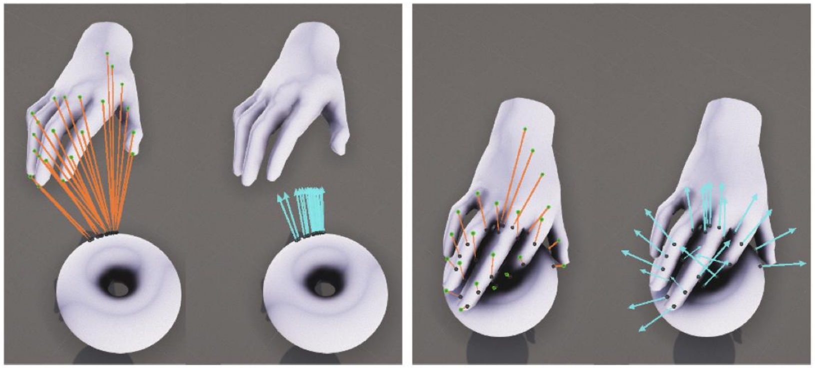


Fig. 5. Two examples of the Signed Distance sensors for the right hand. The hand joints are shown in green. Orange lines indicate the distance from the hand joints to the torus. Cyan arrows are surface normals on the torus.

ManipNet: Neural Manipulation Synthesis with a Hand-Object Spatial Representation

Reference: Hand Object Representations: Implicit Representation

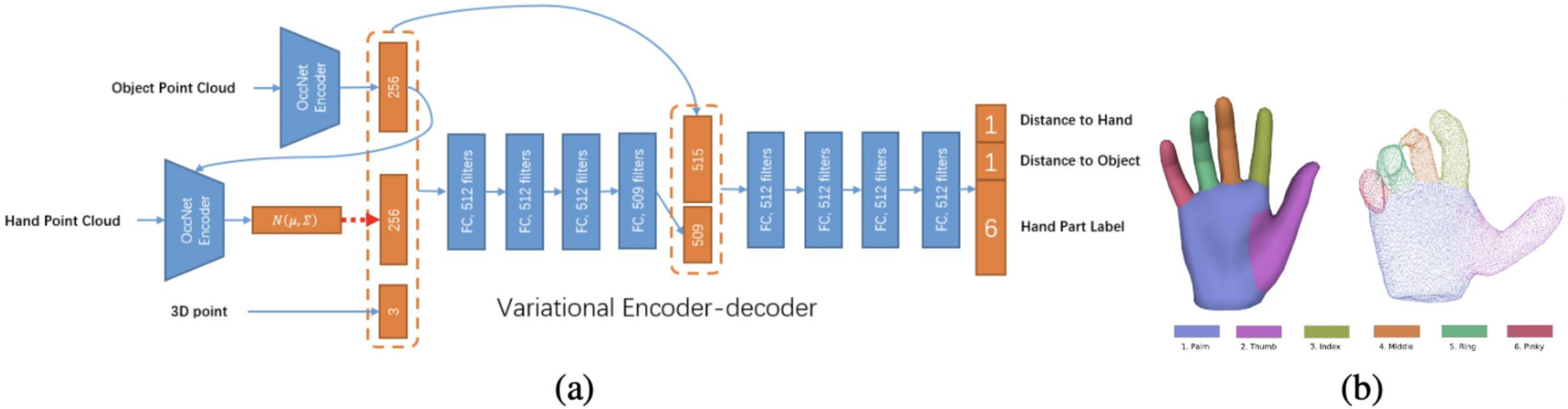


Figure 2: (a) Illustration of the generative grasping field network conditioned on the object point cloud. The red dashed arrow denotes sampling from a distribution. Architecture details are described in Appendix A. (b) Illustration of hand segmentation. Left is our hand part annotation on the MANO model. Right is an example of our *predicted* surface points with hand part labels.

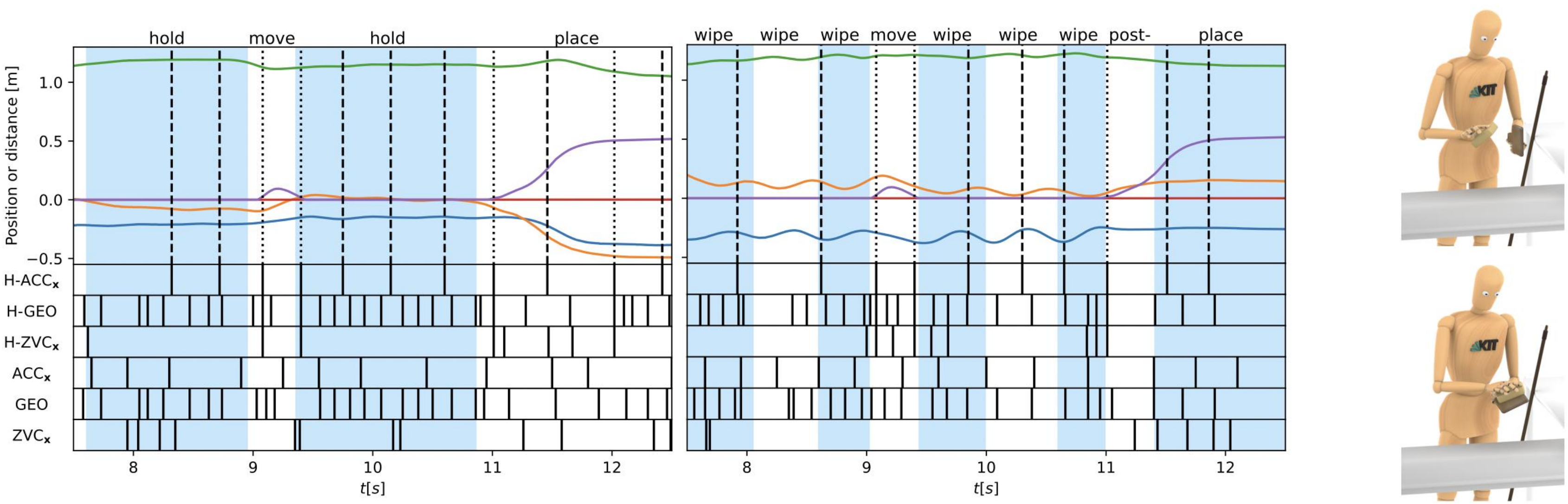
Grasping Field: Learning Implicit Representations for Human Grasps

Reference: Hand-Object Representations: Dataset

grasping
pushing
....
etc

Reference: This year's IROS: Manipulation Segmentation

Different Manipulation Affordance



(a) Left hand segmentations (b) Right hand segmentations (c) $t = \{9.5, 10\}$s

Fig. 3: Bimanual segmentation for the wipe action within the cleaning up task for the subject 1480. *(a)-(b)* Comparisons of the segmentation points (❙) obtained with different algorithms. The manual annotations of the dataset ( ) are considered as ground truth. The *top* panels differentiate between the semantic segmentation (⋮) and the trajectory subsegmentation $\text{ACC}_{\boldsymbol{x}}$ (╏) of the hierarchical segmentation $\text{H-ACC}_{\boldsymbol{x}}$. The hand trajectories (x_1 —, x_2 —, x_3 —) and the distances between the hand and *(a)* the cutting board or *(b)* the sponge (—) and between the sponge and the cutting board (—) are also depicted. *(c)* Snapshots of the task mapped onto the MMM model.

An Evaluation of Action Segmentation Algorithms on Bimanual Manipulation Datasets

Generalization between Gripper

num_link_each_finger: 3

num_finger: 5

palm: bool

num_link_each_finger: 1

num_finger: 2

palm: False

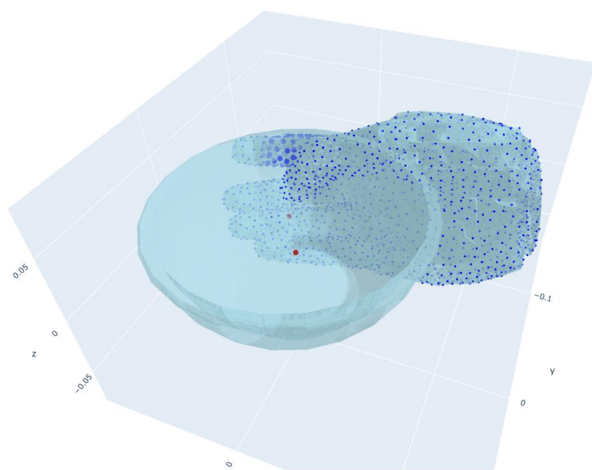
num_link_each_finger: 3

num_finger: 3

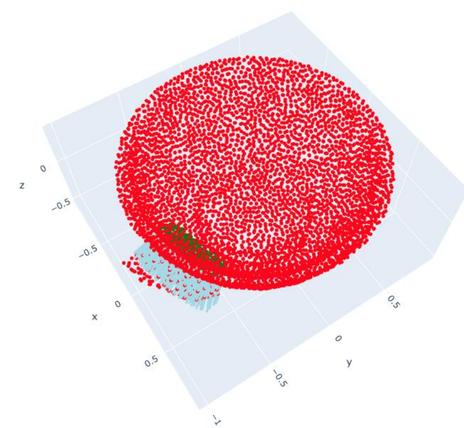
palm: bool

Calculate the distance between links and manipulated object


Shadow Hand
DLR-HIT II Hand
etc.

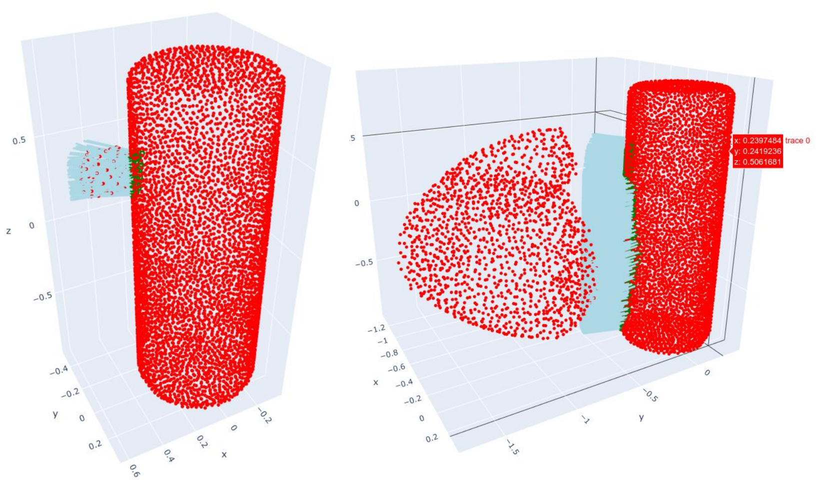


Grasp Generation

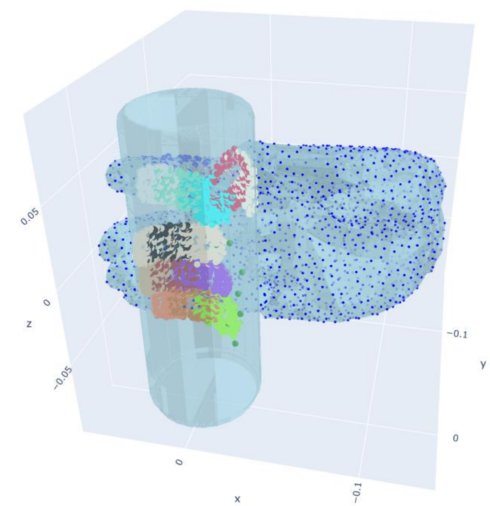


Distance between
Distal Finger and Object

Add Segmentation of links on contact map



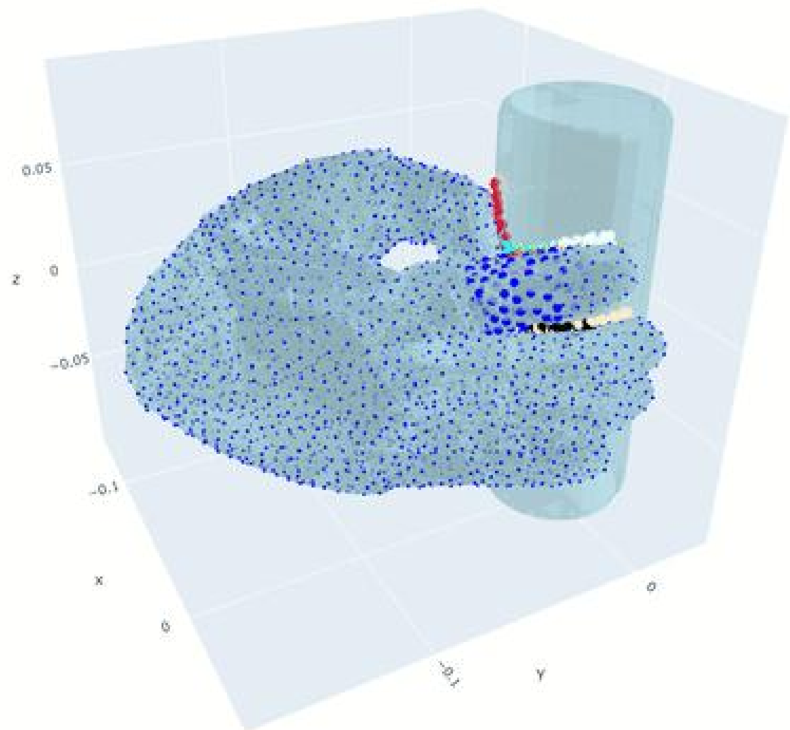
...
Distance of all LINKS



Contact Map with Segmentation

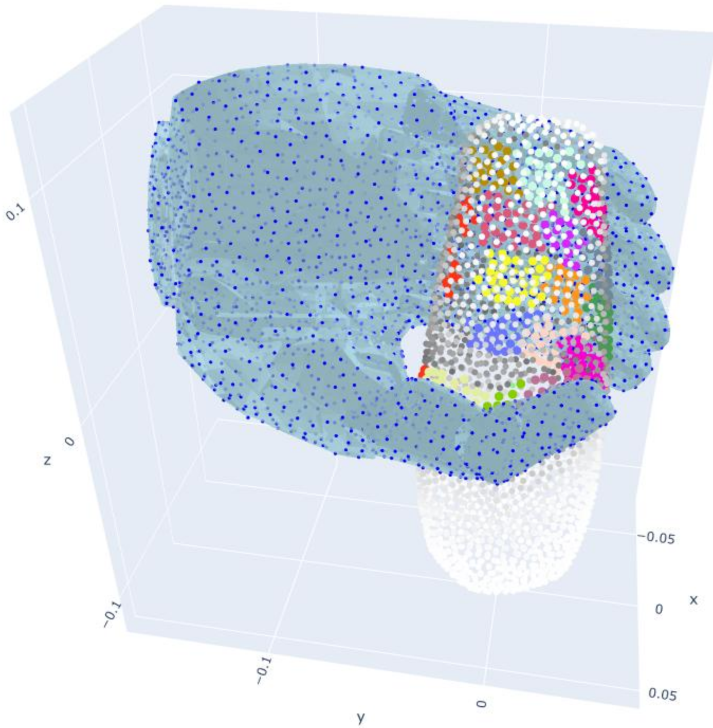
The visualized grasp pose is randomly selected from generated database

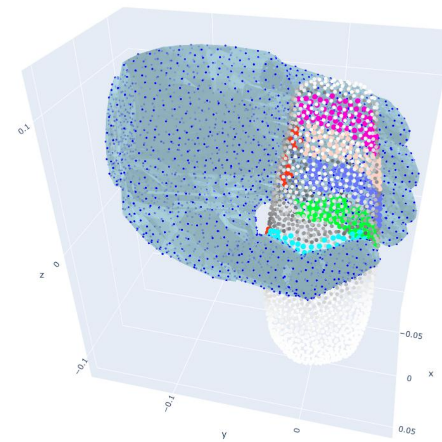
Add Segmentation of links on contact map

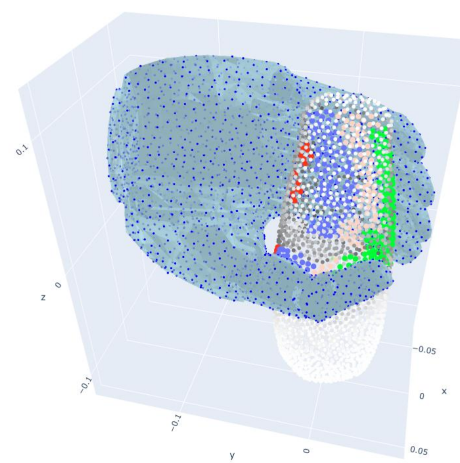


The visualized grasp pose is randomly selected from generated database

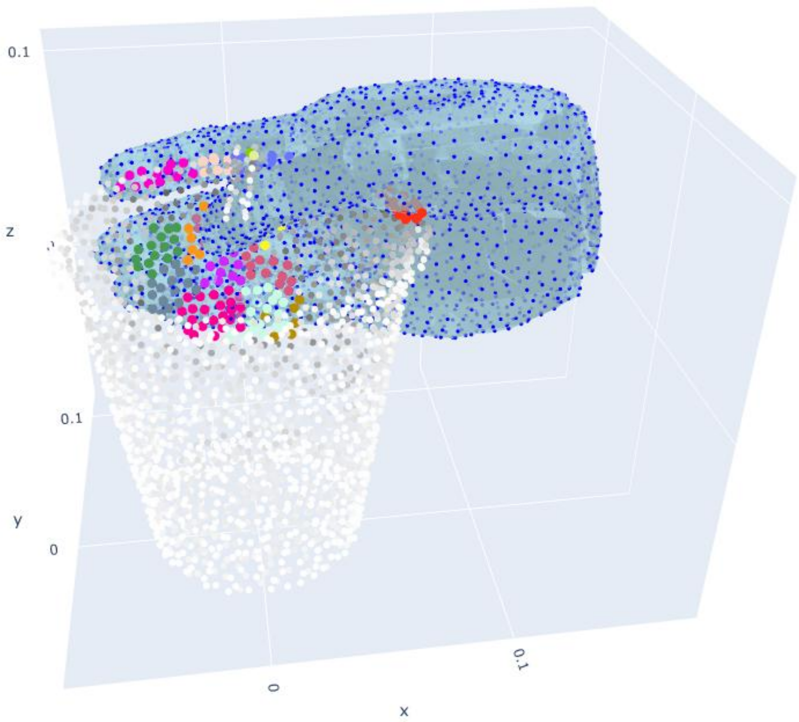
Hand-Object Representations: Group Link and Finger

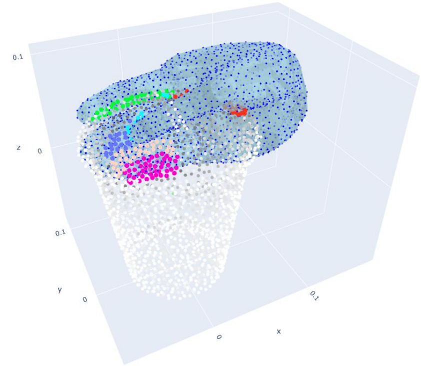


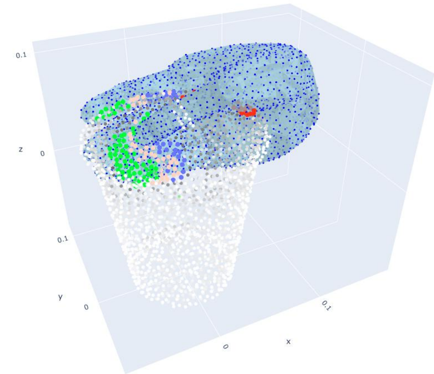




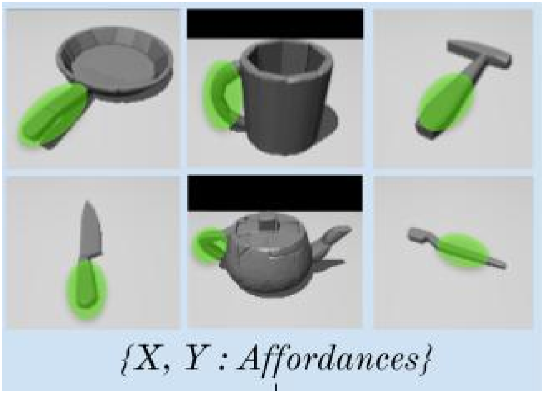
Hand-Object Representations: Group Link and Finger

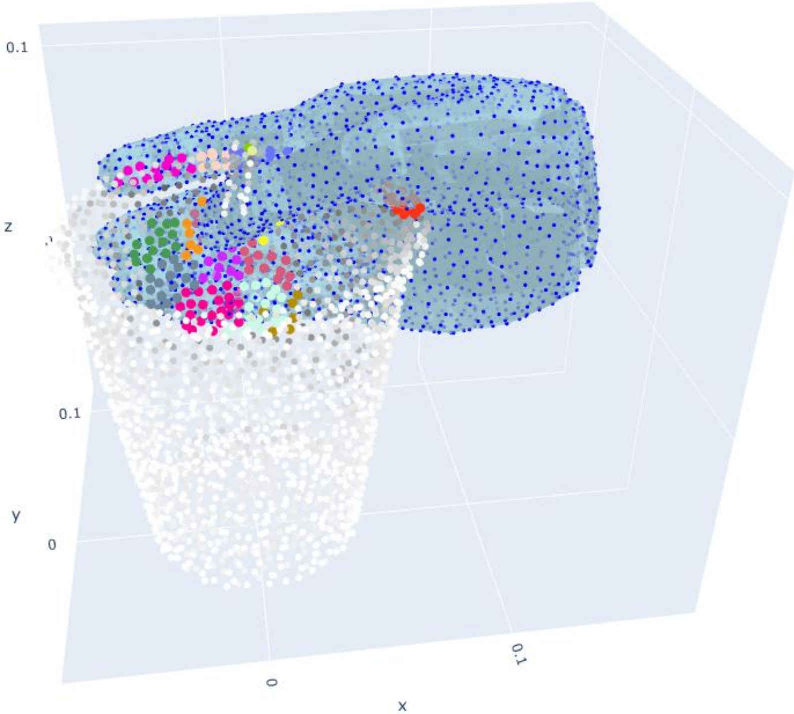






Ongoing: relationship between affordance map and contact map





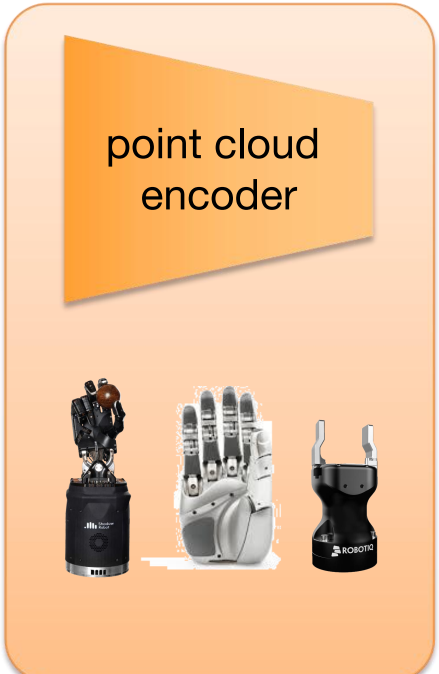
Generate reliable affordance grasping candidates based on contact map

3. How to generate grasp poses and trajectory

Ongoing:
Contact Map as Prior Information + Diffusion Model for grasp generation and trajectory generation

object pcd

FFH joints



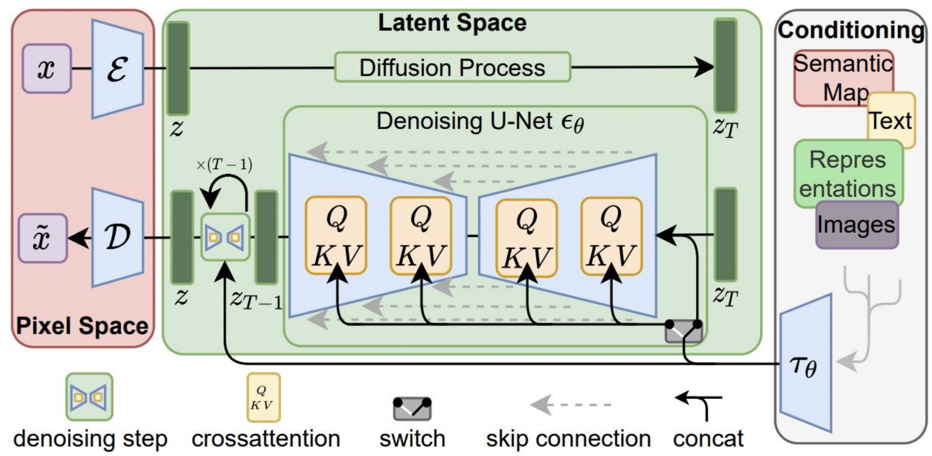


Figure 3. We condition LDMs either via concatenation or by a more general cross-attention mechanism. See Sec. 3.3

. . . Danke

Lei Zhang
lei.zhang-1@studium.uni-hamburg.de
Zhanglei.cn.de@gmail.com

Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme