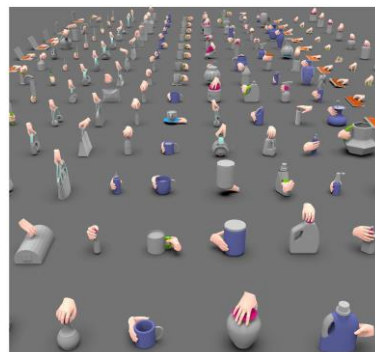


## Introduction

- We present the AffordPose dataset, a large-scale dataset of fine-grained hand-object interactions with affordance-driven hand pose.
- We provide comprehensive data analysis to understand how affordance affects the detailed arrangement of hand poses to complete the appropriate interaction.
- We conduct experiments on three tasks, i.e. hand-object affordance understanding, affordance-oriented hand-object interaction generation, image-based applications to validate the effectiveness of our dataset in learning the fine-grained hand-object interactions.



| dataset      | mod. | syn/real | #obj | #hand pose | intent            |
|--------------|------|----------|------|------------|-------------------|
| HO3D         | RGBD | real     | 10   | 68         | -                 |
| DexYCB       | RGBD | real     | 20   | 1k         | -                 |
| YCBAfford    | RGB  | syn      | 68   | 367        | -                 |
| Obman        | RGBD | syn      | 2.7k | 21k        | -                 |
| FPHAB        | RGBD | real     | 26   | 273        | object affordance |
| ContactPose  | RGBD | real     | 25   | 2.3k       | human objective   |
| GRAB         | Mesh | real     | 51   | 1.3k       | human objective   |
| OakInk-image | RGBD | real     | 100  | 1k         | human objective   |
| OakInk-shape | Mesh |          | 1.7k | 49k        |                   |
| Ours         | Mesh | syn      | 641  | 26k        | part affordance   |

## Dataset Statistics

| Statistics | All   | Bag  | Bottle | Dispenser | Earphone | Faucet | Handle bottle | Jar  | Keyboard | Knife | Laptop | Mug  | Pot  | Scissors |
|------------|-------|------|--------|-----------|----------|--------|---------------|------|----------|-------|--------|------|------|----------|
| #Object    | 641   | 53   | 52     | 34        | 50       | 55     | 32            | 45   | 53       | 57    | 50     | 55   | 48   | 57       |
| #Afford    | 8     | 2    | 2      | 3         | 1        | 2      | 5             | 3    | 1        | 1     | 1      | 3    | 4    | 1        |
| #Hand      | 26712 | 1624 | 2884   | 2772      | 1400     | 1540   | 2408          | 2716 | 1484     | 1596  | 1400   | 3052 | 2240 | 1596     |

## Dataset Analysis

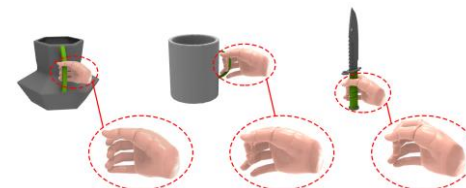
### The Representative Hand Poses



### Different affordances on Same object



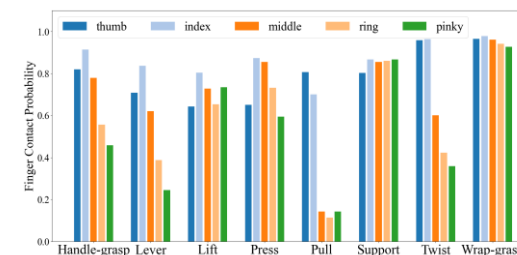
### Same Affordance on Different Object Categories



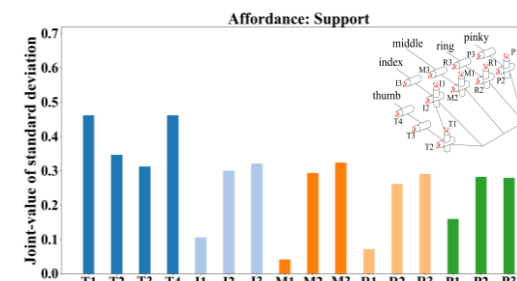
### Diversities Hand Pose on Same Affordance



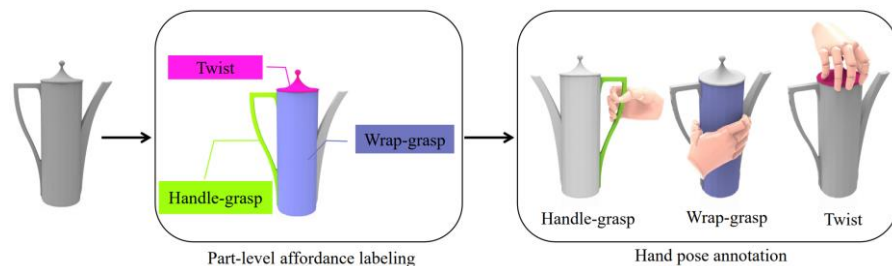
### Fingers Contacting Probabilities for Each Affordance



### Standard Deviations of Joint Values



## Dataset construction process



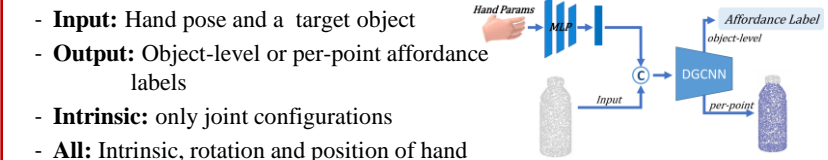
## Acknowledgements

This work was supported by NSFC programs (61976040), Emerging Interdisciplinary Cultivation Project of Jiangxi Academy of Sciences (2022YXXJC0101), Shenzhen Collaborative Innovation Program (CJGJZD2021048092601003), GD Natural Science Foundation (2021B1515020085), and Shenzhen Science and Technology Program (RCYX20210609103121030). Manyi Li is also supported by the Excellent Young Scientists Fund Program (Overseas) of Shandong Province (Grant No.2023HWYQ034).

## Experiments

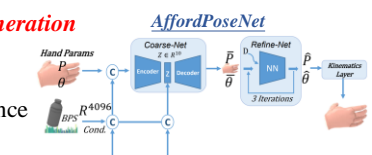
### Task1: Hand-object Affordance Understanding

| Methods                    | Inputs    | Handle-grasp | Lever | Lift  | Press | Pull  | Support | Twist | Wrap-grasp | Mean  |
|----------------------------|-----------|--------------|-------|-------|-------|-------|---------|-------|------------|-------|
| Classification (Accuracy)% | Intrinsic | 92.79        | 99.88 | 98.43 | 98.60 | 99.10 | 91.58   | 90.13 | 93.72      | 94.40 |
|                            | All       | 99.50        | 100   | 99.16 | 95.00 | 91.00 | 99.65   | 98.35 | 98.73      | 98.39 |
| Localization (IoU)%        | Intrinsic | 95.62        | 96.44 | 97.94 | 94.89 | 77.78 | 88.96   | 90.80 | 97.78      | 95.36 |
|                            | All       | 95.05        | 96.99 | 97.90 | 94.42 | 77.78 | 96.59   | 95.17 | 98.91      | 96.29 |



### Task2: Affordance-oriented Interaction Generation

- Input:** Object and a affordance label
- Output:** Hand pose related to the affordance
- Test Set:** Expanded by 200+ objects



| Metrics                              | GrabNet | AffordPoseNet |       |       |       |      |         |       |            |       | Mean |
|--------------------------------------|---------|---------------|-------|-------|-------|------|---------|-------|------------|-------|------|
|                                      |         | Handle-grasp  | Lever | Lift  | Press | Pull | Support | Twist | Wrap-grasp | Mean  |      |
| Penet.Depth(cm) ↓                    | 0.87    | 1.01          | 0.60  | 1.02  | 0.09  | 0.94 | 1.85    | 0.97  | 0.94       | 0.89  |      |
| Solid.Intsec.Vol(cm <sup>3</sup> ) ↓ | 3.20    | 5.32          | 2.44  | 3.37  | 0.97  | 2.82 | 22.93   | 1.83  | 6.04       | 4.57  |      |
| Contact Ratio(%) ↑                   | 96.06   | 100           | 100   | 92.50 | 92.86 | 75   | 100     | 96.88 | 97.26      | 96.06 |      |
| Affordance accuracy(%) ↑             | -       | 80            | 72.73 | 92.50 | 95.24 | 0    | 87.50   | 53.13 | 98.63      | 83.51 |      |



### Task3: Image-based Applications

- Render the RGB images of the hand-object interactions from AffordPose
- Hand-Object Interaction Classification**
  - Input:** RGB Image
  - Output:** hand pose
  - Network:** ResNet-18
- Hand Mesh Recovery**
  - Input:** RGB Image and affordance
  - Output:** Hand pose
  - Network:** Modified from I2L-MeshNet

|               |       | Handle-grasp | Lever  | Lift   | Press  | Pull   | Support | Twist  | Wrap-grasp | Mean   |
|---------------|-------|--------------|--------|--------|--------|--------|---------|--------|------------|--------|
| Mesh Recovery | MPVPE | 12.2         | 18.44  | 45.36  | 11.82  | 24.3   | 26.36   | 14.02  | 9.58       | 16.4   |
|               | MPJRE | 0.2516       | 0.2455 | 0.2278 | 0.1478 | 0.3796 | 0.1962  | 0.2043 | 0.1279     | 0.1892 |