AffordPose: A Large-scale Dataset of Hand-Object Interactions with Affordance-driven Hand Pose: Supplementary Material

Juntao Jian[1] Xiuping Liu[1] Manyi Li[2⋆] Ruizhen Hu[3] Jian Liu[4⋆]
[1]Dalian University of Technology [2]Shandong University
[3]Shenzhen University [4]Tsinghua University

1. Detailed Statistics of AffordPose Dataset

Table 1 provides more detailed statistics about our AffordPose dataset. For each object category, we select tens of objects with different shapes for the dataset construction. We list the number of collected hand-object interactions per object category corresponding to each affordance label. Note that although one object may have multiple affordances, not all the objects of one category share the same affordances. For example, it's common for most mugs to enable the hand-object interactions for ”handle-grasp” and ”wrap-grasp”, but only a few of them have a plate at the bottom to provide the ”support” affordances.

Our collected hand-object interactions exhibit the many-to-many correspondence between the affordances and the object categories. One object category may enable multiple affordances, e.g. the bags support both ”lift” and ”pull” affordances, while one affordance may be related to different object categories, e.g. the ”lift” is marked on the bags, earphones, handle-bottles, pots, to enable the interactions on different functional parts and shapes. The large variation of our dataset enables comparisons in the data analysis and learning experiments for a better understanding of affordance-driven hand-object interactions.

2. Affordance Selection

We select the hand-centered affordance labels to build our AffordPose dataset. The hand-centered affordances are the ones that are highly related to specific actions of human hands, rather than the ambiguous and complicated tasks. For example, the affordance ”pour” is often implemented with human hands. But the fact is that human often grasps and tilts objects with their hands to implement the ”pour” affordance. In other words, ”pour” indicates the goal of a series of actions, rather than a specific hand-object interaction. Therefore, we consider ”handle-grasp” and ”wrap-grasp” as hand-centered affordances, while ”pour” is discarded.

As mentioned in the paper, we organized a panel discussion to select 8 hand-centered affordances from the related works [1, 2, 3]. Table 2 lists these affordances with their definitions to describe the features of the corresponding object parts of these affordances. The related object parts, annotated by the volunteers, are also listed in Table 2.

3. Image-based Interaction Classification

We report the detailed statistics of the performance of the image-based hand-object classification application in Table 3. As said in our main paper, we found a high corre-

Table 1: The detailed statistics of our dataset, including the object number (#Object) of each category and the number of annotated hand-object interactions (#Interaction) per affordance.

		Bag	Bottle	Dispenser	Earphone	Faucet	Handle-bottle	Jar	Keyboard	Knife	Laptop	Mug	Pot	Scissors
	#Object	53	52	34	50	55	32	45	53	57	50	55	48	57
#Interaction	Handle-grasp						756			1596		1456	140	1596
	Press			952					1484		1400			
	Lift	1484			1400		168						28	
	Wrap-grasp		1456	952			896	1260				1540	1344	
	Twist		1428	868		392	560	1176						
	Support						28	280				56	728	
	Pull	140												
	Lever					1148								

Table 2: The affordance labels and their definitions used to construct our dataset, as well as the corresponding object parts annotated by the volunteers.

Hand-object Affordance	**Definition [1, 2, 3]**	**Corresponding Parts (Object/part)**
Handle-grasp	An object extension while affords the ability to easily operate the object.	Handle-bottle/handle Knife/handle Mug/handle
Lever	Any handle which can rotate up to a point. For example, knobs rotate but are not levers because they do not provide handles. Levers must be treated differently from twistable objects or handles because if they are twisted too much they will break.	Faucet/switch
Lift	A part that almost helping people to lift the rest of the entire object.	Bag/handle Earphone/top band
Press	A mechanical feature of objects while either have buttons or can interact with a finger.	Keyboard/key Dispenser/pressing lid
Pull	An object part that affords the ability to promote the object or the part to move easily.	Pull/zipper
Support	A part can be supported with the palm to safely support the entire body of the object	Pot/bottom Mug/support plate
Twist	These objects can either be detached or provide special functionality by twisting them in a clockwise or counterclockwise motion	Jar/lid Faucet/switch
Wrap-grasp	The 'wrap-grasp' trait is afforded by parts which are explicitly meant to be grasped in a hand-wrapping motion. Just because a hand can wrap around an object does not mean it affords wrap-grasp. It must be useful to grip the part in this way.	Bottle/body Mug/body Pot/body

Table 3: Detail quantitative results of hand-object interaction classification. $precision/recall$ is reported for each affordance and each object category.

	Bag	Bottle	Dispenser	Earphone	Faucet	Handle-bottle	Jar	Keyboard	Knife	Laptop	Mug	Pot	Scissors	Mean
Handle-grasp						92.55%/ 88.74%			100%/ 100%		98.21%/ 98.65%	80.00%/ 78.79%	100%/ 100%	97.95%/ 97.79%
Lever					99.14%/ 98.30%									99.00%/ 98.30%
Lift	99.45%/ 99.78%			100%/ 99.88%		88.10%/ 94.87%						100%/ 44.44%		98.97%/ 99.35%
Press			95.63%/ 84.68%					100%/ 100%		100%/ 100%				98.15%/ 96.09%
Pull	96.67%/ 92.06%													96.67%/ 92.06%
Support						71.43%/ 95.24%	92.35%/ 95.48%				100%/ 77.78%	93.44%/ 97.49%		92.38%/ 96.33%
Twist		99.03%/ 98.20%	83.53%/ 94.80%		97.21%/ 95.43%	93.06%/ 96.26%	99.39%/ 97.60%							94.53%/ 96.87%
Wrap-grasp		99.78%/ 99.00%	98.90%/ 98.18%			93.19%/ 91.81%	97.22%/ 97.71%				98.36%/ 97.83%	98.04%/ 96.39%		97.77%/ 97.10%
Mean	99.28% 99.27%	99.42% 98.61%	92.93% 92.30%	100% 99.88%	98.68% 97.62%	92.36% 92.12%	97.58% 97.29%	100% 100%	100% 100%	100% 100%	98.31% 98.03%	95.67% 95.54%	100% 99.90%	97.31%/ 97.29%

lation between classification performance and object functionality. The fewer affordances the object category has, the better classification it obtains. For example, the categories earphone, keyboard, knife, and laptop each have only one affordance type and gain the highest interaction classification results.

References

[1] Sheng Deng, Xun Xu, Chaozheng Wu, Ke Chen, and Kui Jia. 3d affordancenet: A benchmark for visual object affordance understanding. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1778–1787, 2021. 1, 2

[2] Mohammed Hassanin, Salman Khan, and Murat Tahtali. Visual affordance and function understanding: A survey. *ACM Comput. Surv.*, 54(3), 2021. 1, 2

[3] Chao Xu, Yixin Chen, He Wang, Song-Chun Zhu, Yixin Zhu, and Siyuan Huang. Partafford: Part-level affordance discovery from 3d objects. *ArXiv preprint arXiv:22022.13519*, 2022. 1, 2