ArtiBoost: Boosting Articulated 3D Hand-Object Pose Estimation via Online Exploration and Synthesis

[1,2]Lixin Yang⋆, [1]Kailin Li⋆, [1]Xinyu Zhan, [1]Jun Lv, [1,2]Wenqiang Xu, [1]Jiefeng Li, [1,2]Cewu Lu†
[1]Shanghai Jiao Tong University, China [2]Shanghai Qi Zhi Institute, China
{siriusyang, kailinli, kelvin34501, LyuJune_SJTU, vinjohn, ljf_likit, lucewu}@sjtu.edu.cn

Appendices

A. The Training Details

The backbones of classification-based (*Clas*) and regression-based (*Reg*) baseline networks are initialized with ImageNet [2] pretrained model. In *Clas*, the output resolution of 3D-heatmaps is $28 \times 28 \times 28$. The MLP branch that predicts object rotation adopts three fully-connected layers with 512, 256 and 128 neurons for each, and a final layer of 6 neurons that predict the continuity representation [4] of object rotation: $\mathbf{r}_o \in \mathfrak{so}(3)$. We train the network 100 epochs with Adam optimizer and learning rate of 5×10^{-5}. The training batch size across all the following experiments is 64 per GPU and 2 GPUs in total. The framework is implemented in PyTorch. All the object models and textures are provided by the original dataset. For all the training batches, the blended rate of original real-world data and ArtiBoost synthetic data is approximately $1:1$. We empirically find that this real-synthetic blended rate achieves the best performance.

B. Objects' Symmetry Axes

In the hand-object interaction dataset, it is far more challenging to predict the pose of an object than in the dataset only contains objects, since the objects are often severely occluded by the hand. Therefore, we relax the restrictions of the objects' symmetry axes following the practices in [1,3]. Supposing the set $\mathcal{S}$ contains all the valid rotation matrices based on the object's predefined symmetry axes, we calculate $\mathcal{S}$ with the following step:

1) Firstly, as shown in Fig 1, we align the object to its principal axis of inertia.

2) Secondly, we define the axis $\mathbf{n}$ and angle θ of symmetry in Tab 1 under the aligned coordinate system, where the object's geometry does not change when rotate this object by an angle of θ around $\mathbf{n}$. Here we get the predefined rotation matrix $\mathbf{R}_{def} = \exp(\theta \mathbf{n})$.

3) To get a more accurate rotation matrix $\mathbf{R}$, we use the Iterative Closest Point (ICP) algorithm to fit a $\Delta\mathbf{R}$. The ICP minimizes the difference between $\Delta\mathbf{R} * \mathbf{R}_{def} * \mathbf{V}_o$ and $\mathbf{V}_o$, where $\mathbf{V}_o$ is the point clouds on object surface. Finally, we have $\mathbf{R} = \Delta\mathbf{R} * \mathbf{R}_{def}, \mathbf{R} \in \mathcal{S}$.

Figure 1. **YCB objects' principal axis of inertia.** The x, y and z axis are colored in red, green and blue, respectively.

Objects	Axes: $\mathbf{n}$	Angle: θ
002_master_chef_can	x, y, z	180°, 180°, ∞
003_cracker_box	x, y, z	180°, 180°, 180°
004_sugar_box	x, y, z	180°, 180°, 180°
005_tomato_soup_can	x, y, z	180°, 180°, ∞
006_mustard_bottle	z	180°
007_tuna_fish_can	x, y, z	180°, 180°, ∞
008_pudding_box	x, y, z	180°, 180°, 180°
009_gelatin_box	x, y, z	180°, 180°, 180°
010_potted_meat_can	x, y, z	180°, 180°, 180°
024_bowl	z	∞
036_wood_block	x, y, z	180°, 180°, 90°
037_scissors	z	180°
040_large_marker	x, y, z	180°, ∞, 180°
052_extra_large_clamp	x	180°
061_foam_brick	x, y, z	180°, 90°, 180°

Table 1. **YCB objects' axes of symmetry**. ∞ indicates the object is revolutionary by the axis.

Objects	Our *Clas* sym	Our *Clas* sym + **Arti**	Objects	Our *Clas* sym	Our *Clas* sym + **Arti**
002_master_chef_can	27.62	**25.59**	003_cracker_box	63.68	**46.13**
004_sugar_box	48.42	**39.20**	005_tomato_soup_can	33.31	**31.90**
006_mustard_bottle	35.16	**32.01**	007_tuna_fish_can	24.54	**23.81**
008_pudding_box	39.92	**35.04**	009_gelatin_box	45.99	**37.81**
010_potted_meat_can	41.44	**36.47**	011_banana	98.69	**79.87**
019_pitcher_base	105.66	**84.82**	021_bleach_cleanser	91.66	**72.31**
024_bowl	**31.74**	32.37	025_mug	65.46	**54.28**
035_power_drill	74.95	**52.70**	036_wood_block	51.24	**50.69**
037_scissors	88.10	**66.52**	040_large_marker	30.76	**29.33**
052_extra_large_clamp	78.87	**55.87**	061_foam_brick	34.23	**31.53**

Table 2. Full MSSD results (mm) on **DexYCB** testing set.

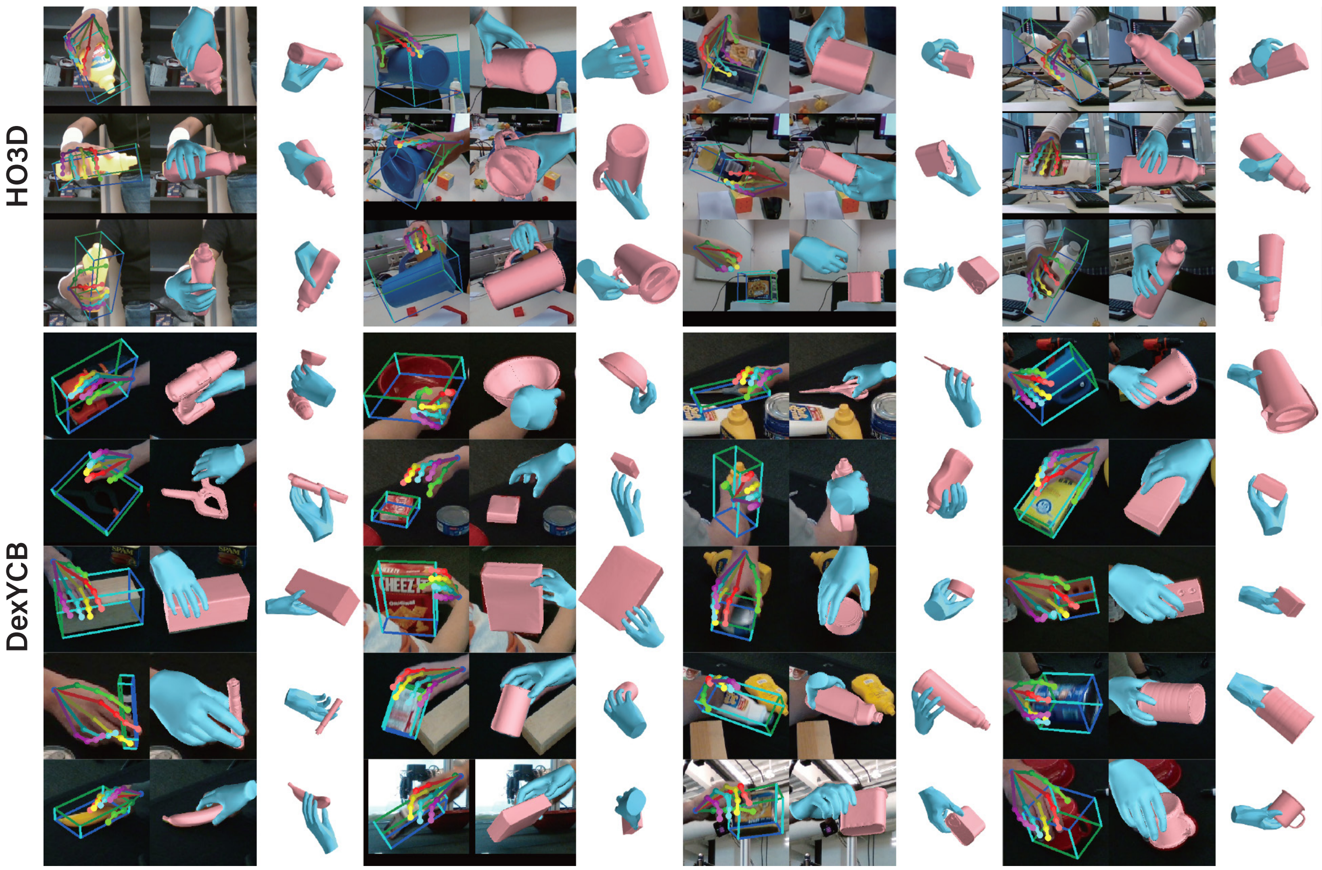

Figure 2. (**Best view in color**) More qualitative results on **HO3D** ($1^{st} \sim 3^{rd}$ rows) and **DexYCB** ($4^{th} \sim 8^{th}$ rows) datasets.

C. Additional Results

We demonstrate 20 YCB objects' MSSD on DexYCB in Tab. 2. With ArtiBoost, our network can predict a more accurate pose for almost every object. More qualitative results on HO3D and DexYCB testing set are shown in Fig. 2.

References

[1] Yu-Wei Chao, Wei Yang, Yu Xiang, Pavlo Molchanov, Ankur Handa, Jonathan Tremblay, Yashraj S. Narang, Karl Van Wyk, Umar Iqbal, Stan Birchfield, Jan Kautz, and Dieter Fox. DexYCB: A benchmark for capturing hand grasping of objects. In *CVPR*, 2021. 1

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 1

[3] Shreyas Hampali, Sayan Deb Sarkar, Mahdi Rad, and Vincent Lepetit. HandsFormer: Keypoint transformer for monocular 3d pose estimation ofhands and object in interaction. In *arXiv preprint arXiv:2104.14639*, 2021. 1

[4] Yi Zhou, Connelly Barnes, Lu Jingwan, Yang Jimei, and Li Hao. On the continuity of rotation representations in neural networks. In *CVPR*, 2019. 1