

DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness

Supplementary Materials

A. Overview

In the main text, we introduce DexGrasp Anything, a physics-aware diffusion generator that incorporates three tailored physical constraints for generating dexterous grasps. Along with this, we present the largest and most diverse dataset for dexterous grasp generation to date. To further demonstrate the improvements brought by our method and dataset, this supplementary material provides more comprehensive experimental results (Sec. B) and details the filtering process (Sec. C) used for dataset construction. Additionally, we have included a **demo video** in the supplementary files that showcases our **zero-shot real-world experiments on unseen objects**, which we highly recommend reviewing for a deeper understanding of our method’s practical applications and performance.

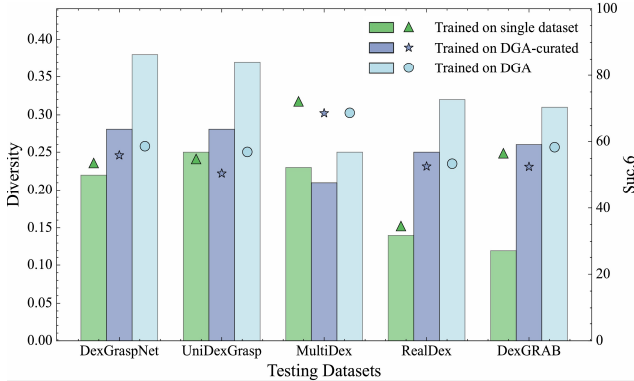


Figure 9. Cross-dataset evaluation. Comparison of diversity (bars) and all-direction grasp success rates (triangles/stars/circles) across models trained on different datasets. Trained on single dataset indicates models were trained on the same dataset they are tested on.

B. Evaluation Results

B.1. Results for Cross-dataset Evaluation

We present the comprehensive cross-dataset evaluation result for the DexGrasp Anything diffusion generator on five

existing datasets and our dataset, as shown in Table 6 and Figure 9. Qualitative results are presented in Figure 12. The results demonstrate that training on our large-scale, diverse dataset significantly enhances generation diversity while achieving comparable or higher grasping success rates compared to training on the respective original datasets.

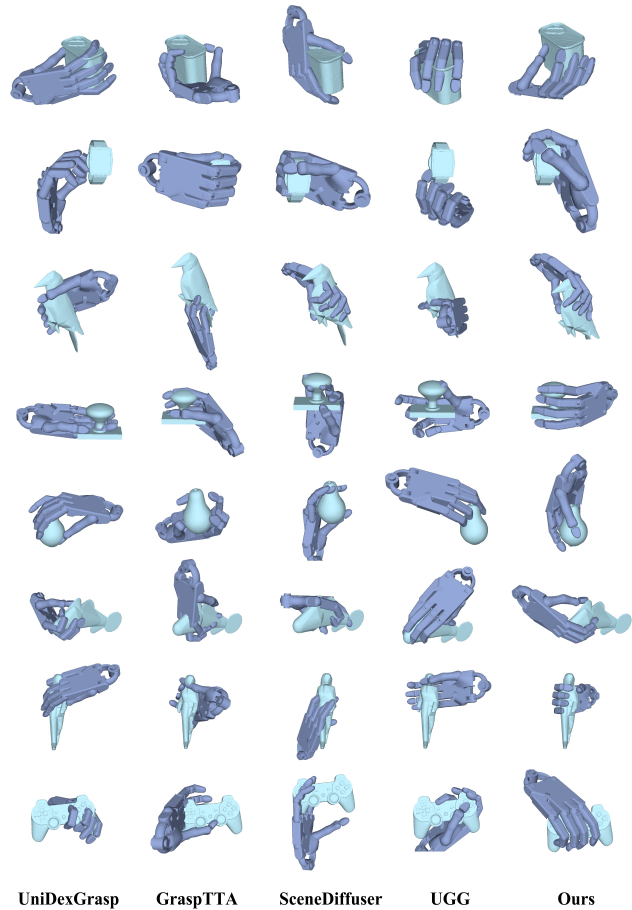


Figure 10. Qualitative visualization of comparisons on grasping poses.

B.2. Qualitative Results for Comparisons

We provide additional qualitative results comparing Dex-Grasp Anything with existing state-of-the-art methods in Figure 10.

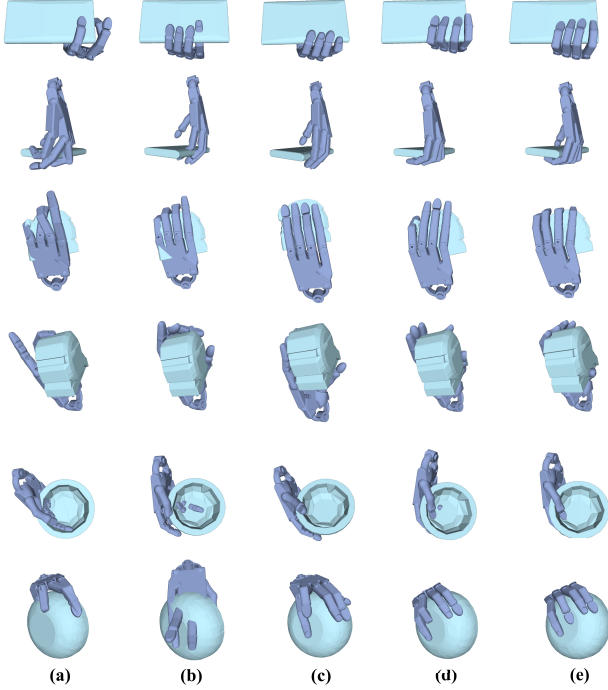


Figure 11. Visualizations of the ablation study. Each pair (row 1-2, 3-4, 5-6) of rows corresponds to the different views of the same grasp for the same object.

B.3. More Visualizations for Ablation Studies

We provide additional visualizations for the ablation studies in Figure 11, where we progressively incorporate three physical constraints during the training and sampling process of our diffusion generator. Column (a) represents the baseline, while columns (b), (c), and (d) illustrate the results after incrementally adding the SRF, ERF, and SPF constraints, respectively, to the baseline. Finally, column (e) shows the results after incorporating the LLM-enhancement into the representation extraction module.

B.4. More Visualizations of Generated Poses

We present more visualizations of the generated grasping poses by our methods on various challenging objects from [1, 2] in Figure 14 and Figure 15. Our models produce reasonable and stable grasping poses for complex and irregular objects such as a robot model (3rd row, 2nd col in Figure 15) and a loong head (7th row, 3rd col in Figure 15).

C. Implementation details

We rigorously evaluated each grasp pose in our dataset to ensure that the object is held firmly without significant penetration. For hand-object penetration computation, we employ two approaches. The first approach, adopted by [3] and also used in the External-penetration Repulsion Force, calculates the Euclidean distance between each hand point and its nearest neighbor on the object surface. The second approach, introduced by [4], transforms the object and each robot hand link into the local hand coordinate system based on the robot’s configuration. For the palm, penetration is measured as the signed distance between the sampled object points and the mesh surface of the palm, represented by a signed distance field. For each phalange link, it is approximated as cylinders, and object points are projected onto the cylinders’ bounding volumes to compute signed distances, adjusted using a mask to differentiate internal and external points. We combine both methods to enforce strict filtering conditions for our dataset.

D. Limitations and Future Works

As shown in Figure 13, we notice that our method produces sub-optimal poses with obvious penetration for objects with extremely thin shapes (e.g. masks, plates etc.). To address these challenges, enhancing affordance modeling or integrating tactile feedback into the robotic grasping system would be promising directions for future works.

References

- [1] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023. 2
- [2] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. Objaverse-xl: A universe of 10m+ 3d objects. *Advances in Neural Information Processing Systems*, 36, 2024. 2
- [3] Siyuan Huang, Zan Wang, Puhao Li, Baoxiong Jia, Tengyu Liu, Yixin Zhu, Wei Liang, and Song-Chun Zhu. Diffusion-based generation, optimization, and planning in 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16750–16761, 2023. 2
- [4] Yinzhen Xu, Weikang Wan, Jialiang Zhang, Haoran Liu, Zikang Shan, Hao Shen, Ruicheng Wang, Haoran Geng, Yijia Weng, Jiayi Chen, et al. Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4737–4746, 2023. 2

Table 6. Cross-dataset evaluation results. The **bold** values indicate the best performance, and the underlined values indicate the second-best performance.

Testing Dataset	DexGraspNet				UniDexGrasp				MultiDex				RealDex				DexGRAB			
Training Dataset	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑	Suc.6 ↑	Suc.1 ↑	Pen. ↓	Div ↑
DexGraspNet	53.6	90.4	21.5	0.22	49.3	82.4	14.9	0.19	55.6	90.1	9.1	0.17	38.4	77.5	19.2	0.17	48.1	84.0	19.7	0.18
UniDexGrasp	45.4	82.4	16.4	0.23	<u>54.8</u>	90.8	<u>18.9</u>	0.25	52.8	90.3	9.4	0.18	38.4	79.3	20.7	0.19	37.5	79.6	<u>20.1</u>	0.19
MultiDex	46.8	83.1	18.1	0.20	43.9	81.3	14.5	0.19	72.2	<u>96.3</u>	9.6	<u>0.23</u>	29.6	69.1	<u>20.1</u>	0.23	52.9	87.9	21.0	0.15
RealDex	47.3	79.5	18.7	0.05	43.8	81.3	15.8	0.04	57.5	89.2	11.6	0.06	34.6	71.2	23.1	0.14	38.5	79.8	22.7	0.08
DexGRAB	41.0	75.8	18.7	0.12	43.9	81.4	14.1	0.10	62.1	90.9	<u>9.3</u>	0.11	35.2	71.5	24.0	0.11	<u>56.5</u>	91.8	28.6	0.12
DGA-curated(ours)	<u>55.9</u>	87.3	20.9	<u>0.28</u>	50.5	84.6	14.0	<u>0.28</u>	68.7	95.9	12.5	0.21	<u>52.6</u>	85.7	21.5	<u>0.25</u>	52.5	89.0	22.9	<u>0.26</u>
DGA(ours)	58.6	<u>88.5</u>	<u>17.8</u>	0.38	56.9	<u>86.7</u>	16.7	0.37	<u>68.8</u>	96.9	9.5	0.25	53.4	<u>84.4</u>	22.4	0.32	58.3	<u>90.2</u>	23.2	0.31

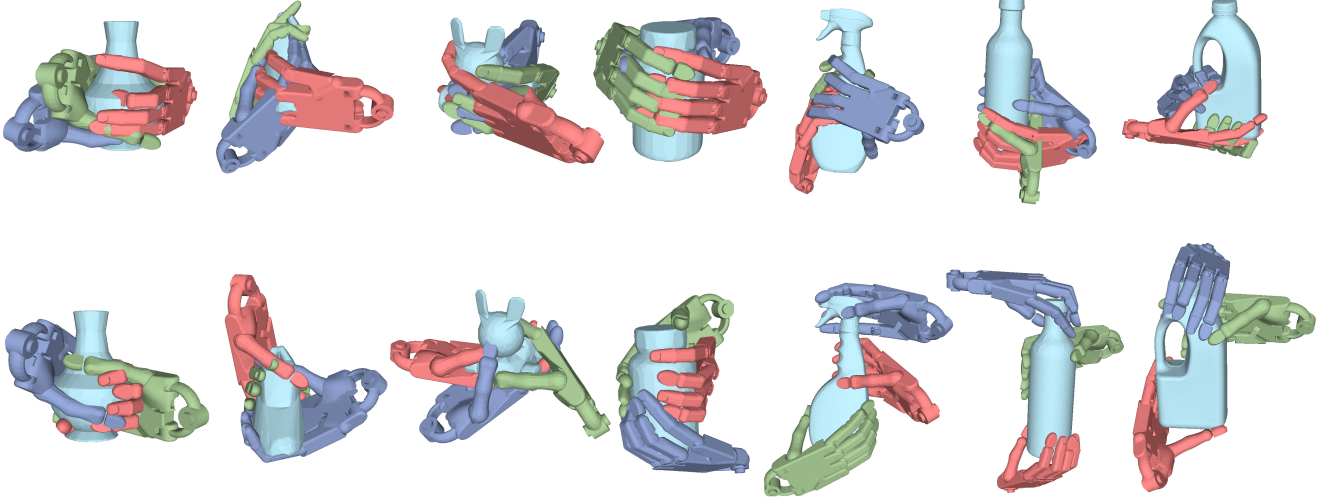


Figure 12. Visualization of cross-dataset evaluation results. The top row shows models trained on single dataset, while the bottom row displays models trained on our dataset.

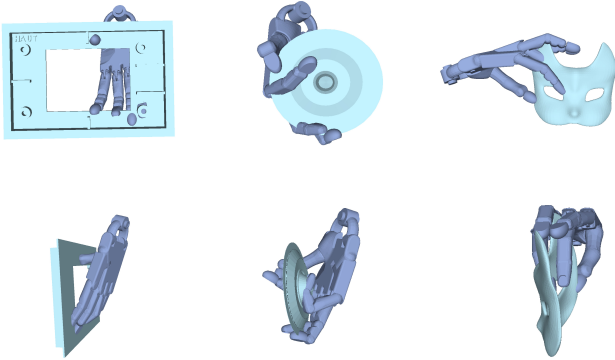


Figure 13. Visualization of failed cases.

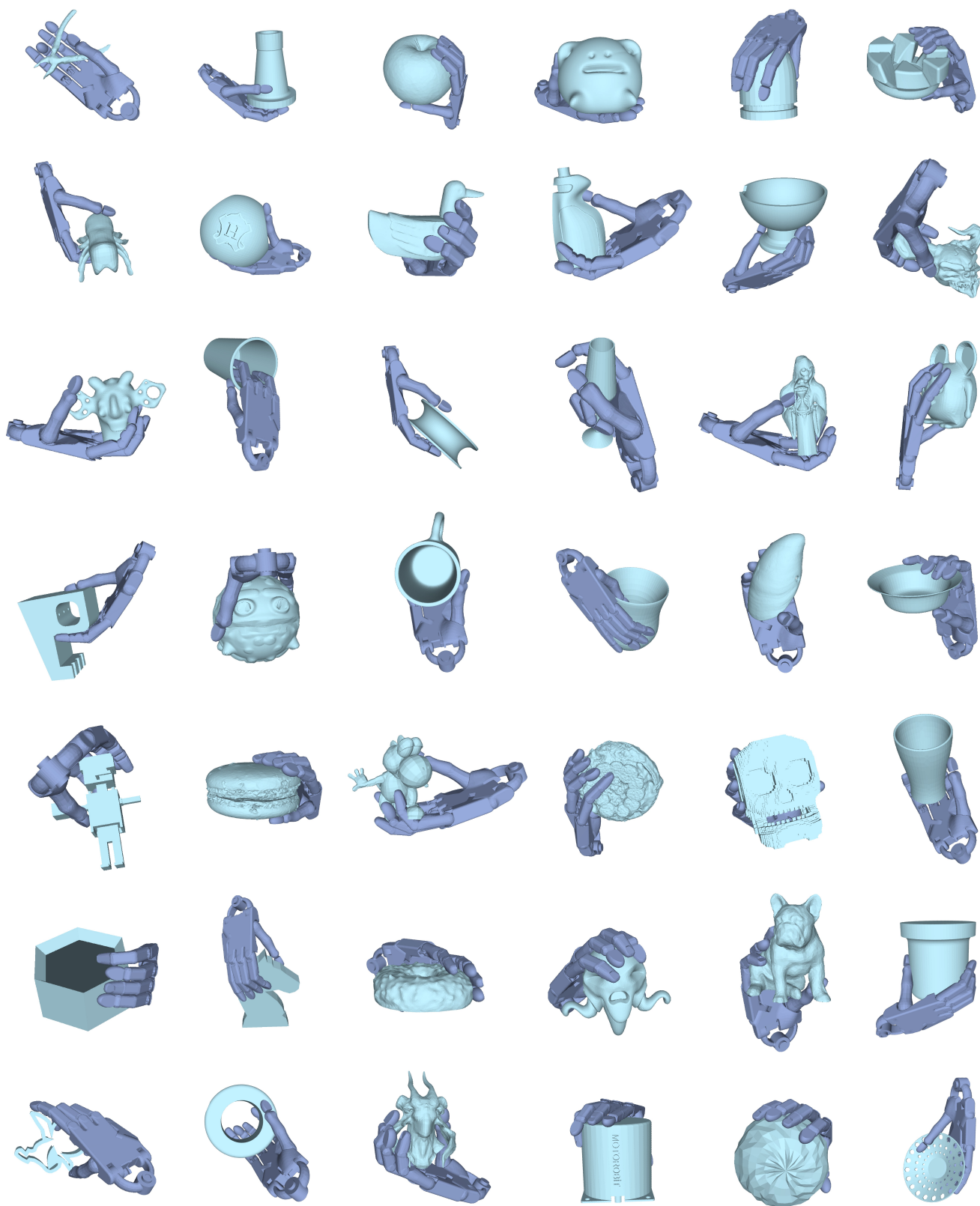


Figure 14. Visualization of our method’s results.

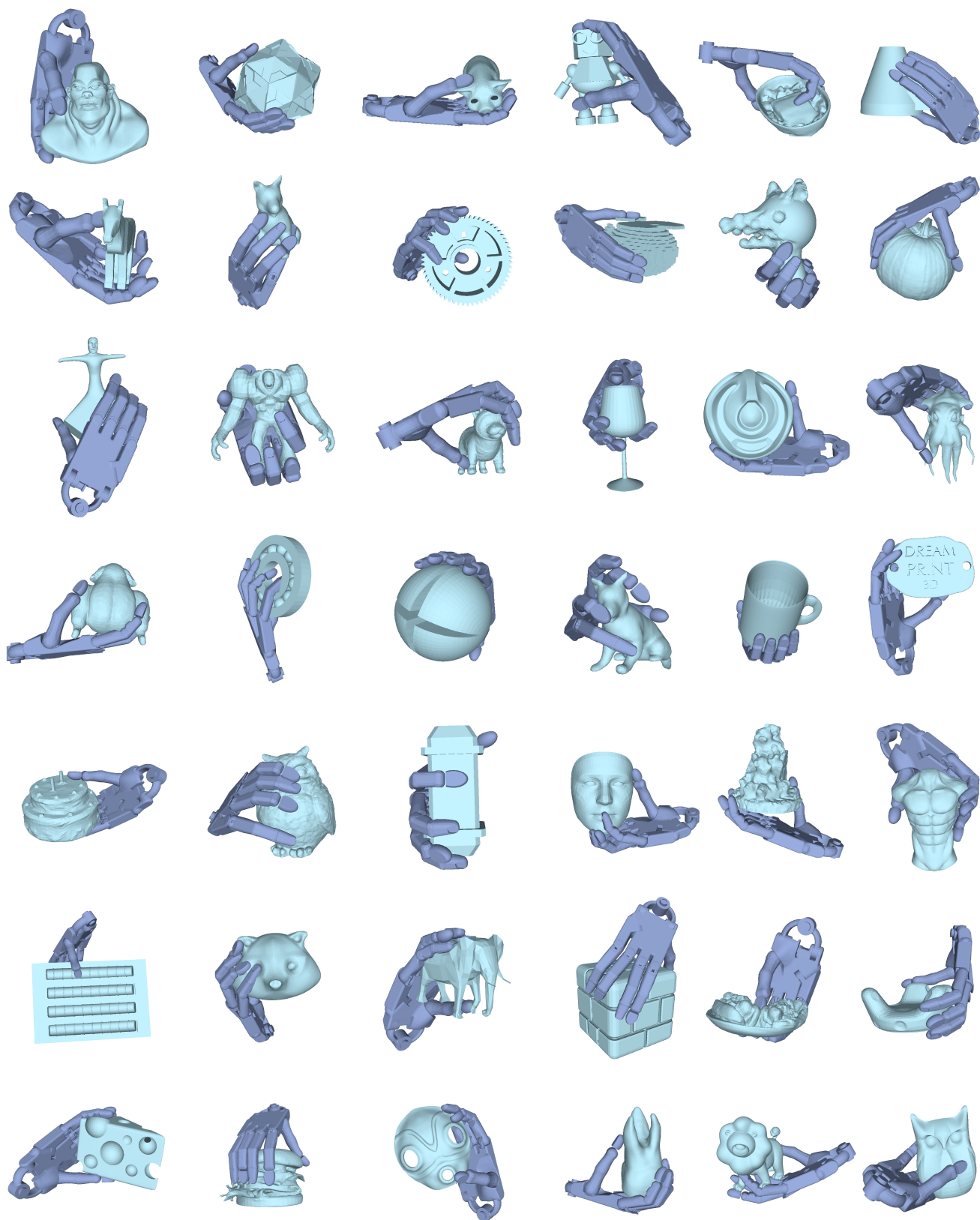


Figure 15. Visualization of our method's results.