# PickSim: A dynamically configurable Gazebo pipeline for robotic manipulation

Guillaume Duret, Nicolas Cazin, Mahmoud Ali, Florence Zara, Emmanuel Dellandréa, Jan Peters, Liming Chen

# PickSim: A dynamically configurable Gazebo pipeline for robotic manipulation learning*

Guillaume Duret[1,3]
guillaume.duret@ec-lyon.fr

Nicolas Cazin[1]
nicolas.cazin@ec-lyon.fr

Mahmoud Ali[1]
mahmoud.ali@ec-lyon.fr

Florence Zara[2]
florence.zara@liris.cnrs.fr

Emmanuel Dellandrea[1]
emmanuel.dellandrea@ec-lyon.fr

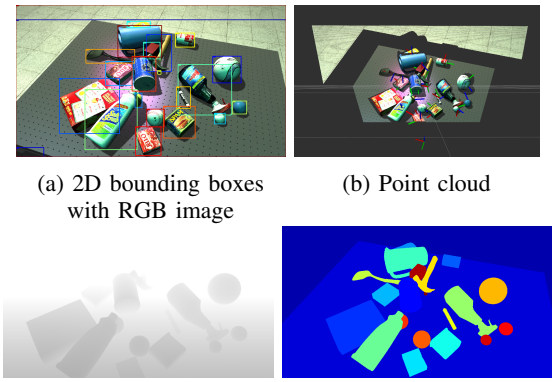
Jan Peters[3]
jan.peters@tu-darmstadt.de

Liming Chen[1]
liming.chen@ec-lyon.fr

[1]Univ Lyon, Centrale Lyon, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69130 Ecully, France
[2]Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, F-69622 Villeurbanne, France
[3]Intelligent Autonomous Systems Lab, Technical University of Darmstadt, 64289 Darmstadt, Germany

***Abstract*—State-of-the art robot learning approaches are data-driven and require a large amount of diverse robot data which are prohibitively expensive to acquire. In this paper, we present PickSim, an open source dynamically configurable Gazebo-based pipeline for the purpose of robot manipulation learning. Specifically, we propose a new plugin and pipeline, based on the well known open source robot software Gazebo. We showcase the potential of PickSim through the generation of 6D pose estimation dataset, potentially randomized over a number of factors (e.g., lighting, texture, shape, etc.). Rich and perfect annotations (e.g., object masks, poses, occlusions, etc.) are automatically generated and enable the learning and evaluation of computer vision models over a number of varying factors. PickSim unlocks multi-task robot manipulation learning when a robot model and physics engine are integrated into it, and thereby paves the way to the generation of large-scale robot manipulation data as required in general-purpose robot manipulation learning.**

***Index Terms*—6D pose estimation, Occlusion, Gazebo, Benchmark.**



(a) 2D bounding boxes with RGB image (b) Point cloud

(c) Depth map (d) Semantic map

Fig. 1: Examples of annotations proposed for the dataset generation thanks to our Gazebo's plugin called PickSim.

## I. Introduction

Robotic learning currently relies on data-driven methods. It is therefore essential to acquire high quality annotated training data. This issue is all the more important for robotic manipulation, which requires the learning of interactions between robots and objects [5]. The use of real data requires a significant amount of time to generate them and becomes a costly solution [15], [16]. In this context, the generation of synthetic data [22], [23], [28], which is very successful in the fields of computer vision and robotic simulation, is an interesting solution for the creation of various datasets related to manipulation tasks.

The challenge then lies in the ability to generate relevant datasets for learning robotic manipulation. Indeed, this data vary according to many parameters such as the robotic tasks themselves, the number of cameras present in the scene, their different viewpoints, etc. Taking an example of object tracking while grasping, training 6D pose tracking on non robotic dataset could face difficulty in front of robot unique environment and specific gripper occlusions. Specific manipulation task dataset could be used to developed or fine-tune vision model for the purpose of vision-based robot learning [8]. Secondly, the ability to generate multiple manipulation dataset is crucial for the ambitious goal of global robotics manipulation skills. This branch of research would require large scale datasets of multiple manipulation task [5], [17], [21].

Nevertheless, developing such a framework for easy creation of synthetic datasets of manipulation tasks remains a challenge. Indeed, current data generation tools are either purely computer vision based (and thus they are not defined within robotic learning software), or lack the maturity to easily record fully annotated random datasets. The difficulty resides in developing a framework to generate this data while providing the following components: (i) fully automatic registration of all data with a good range of annotations; (ii) a complete and flexible robotic environment; and (iii) an integrated and easily configurable scene domain randomization.

*This work was in part supported by the French Research Agency, l'Agence Nationale de Recherche (ANR), through the projects Learn Real (ANR-18-CHR3-0002-01), Chiron (ANR-20-IADJ-0001-01), Aristotle (ANR-21-FAI1-0009-01). It was granted access to the HPC resources of IDRIS under the allocation 2022-[AD011012172R1] made by GENCI.

In this context, we propose to directly improve the well-known and open source robotic Gazebo software [14]. In particular, because scene understanding is a key step to learn robot manipulation skills, we highlight in this paper its ability to seamlessly generate simulated scenes of objects with the creation of fully widely annotated vision dataset. Moreover, as 6D pose estimation is a key task for robotic manipulation [13], we will demonstrate its utility in the creation of 6D pose estimation dataset.

Consequently, we propose PickSim an integrated ROS automatic pipeline creating customized annotated datasets from simple meshes. Our open source contributions are:

- a newly added gazebo plugin to add main annotations for 6D pose estimation, including 2D bounding boxes, instance, segmentation masks and occlusion rates;
- a pipeline architecture to automatically generate datasets for 6D pose estimation;
- a short review of current softwares for generating 6D pose estimations datasets;
- a unique post-processing step taking into account occlusion rates in order to generate ready to train customized benchmark datasets.

In the following, Section II presents previous work in computer vision annotations and generation of 6D pose dataset. Section III presents our pipeline to generate synthetic data suitable for 6D pose estimation. Section IV presents some limitations and Section V concludes this article with some perspectives.

## II. Related work

Deeply dependent on deep learning and data, computer vision has developed a bunch of tools with the purpose of reducing the barrier of data generation [6], [11]. These frameworks, containing a good list of annotation, unlock the generation of high quality dataset for multiple computer vision task. However, robotics software don't have this maturity of configurable and easy generating of datasets.

### A. Computer vision annotations tools

Computer vision community relied on video game software in order to compute synthetic datasets and train vision models even when the final application is robotics [20], [27]. For example, Unity perception [1] and UnrealCV [25] support the computing of some annotations like 2D and 3D bounding boxes and segmentation masks.

With the need of data continually increasing with the development of deeper and deeper models, multiple tools have been developed over the years proposing advanced rendering quality with ray tracing and focusing on facilitating the process of creating synthetic data for computer vision. In this idea, BlenderProc [6], [7], based on the open source Blender software, has been one of the most popular tools for computing computer vision data. Recently, Kubric [11] proposes a rich annotation tool using Blender for rendering and Pybullet for the simulation. The authors highlight that the process of generating datasets is as significant as model architectures and demonstrate the flexibility of their generation by realising multiple datasets for different computer vision tasks. Let's note that Kubric hasn't generate 6D pose estimation dataset.

Table I presents an overview of the current generation tool for 6D pose with their corresponding annotation. It highlights our ability to generate annotation for 6D pose with even more annotation like occlusion rate.

### B. Robotic oriented computer vision tools

More robotics oriented, similarly to Kubric [11] and BlenderProc [6], Nvsii [20] proposed an easy to install API with ray tracing rendering capabilities with OptiX and randomization tools. Stillebelen [26] introduced some category based assets to work on synthetic dataset generation through domain randomisation. Both, leveraging domain randomisation, demonstrated similar model performance compared to real data trained models. Unfortunately, they did not offer a complete robotics environment, such as path planning.

On the other hand, complete robotic simulators usually lack in vision annotations features and in their ability to easily generate datasets although scene understanding and perception is a key task for robotics application. In terms of open source robotics simulation software, Gazebo [14], PyBullet [4], and Mujoco [29] are some of the most used [3] in robotics manipulation but they only consider basic annotations and the rendering quality is limited. Recently, omniverse replicator using Isaac Sim [19] is getting more and more success to generate synthetic computer vision. It is a robotics simulator that uses the physics engine PhysX5 with state-of-the-art ray tracing rendering. This simulator, developed by NVIDIA, is very complete and offers lots of annotations including occlusion. Unfortunately, hardware restriction (Nvidia RTX GPUs) and the fact that this is not an open source project limit its use and potential community contributions. Another promising example is SAPIEN [33] simulator, which integrates OpenGL and OptiX for rendering. Unfortunately, this software only have basic annotations and have not been aimed at generating occlusion rates.

Table I presents an overview of the current generation tool for 6D pose with their corresponding annotation. It highlights our ability to generate annotation for 6D pose estimation while being in a friendly robotic environment. It could be noted that current vision model development is outside of robot learning software although vision and control are highly linked. Our work benefit of the ability of vision and control to be developed in the same simulator avoiding domain gaps.

### C. 6D pose synthetic dataset generation pipelines

For the annotated data of previous synthetic 6D pose datasets, we illustrate the disparity between current pipelines for generating 6D pose estimation. Answering different specificity, different pipeline have been developed to generate synthetic 6D pose estimation. Minghao Gou1 and Haolin Pan [10] used BlenderProc to generate synthetic data in order to get more objects variety for unseen object 6D pose estimation task.

| | Replicator | Stillleben | BlenderProc | NVISII | Ignition | Kubric | Our plugin |
|---|---|---|---|---|---|---|---|
| Instance segmentation | yes | yes | yes | yes | yes | yes | yes |
| Semantic segmentation | yes | yes | yes | yes | yes | yes | yes |
| 2D bounding boxes | yes | no | yes | yes | no | yes | yes |
| 3D bounding boxes | yes | no | no | no | no | yes | yes |
| Bounding Box 2d Loose | yes | no | no | no | no | no | yes |
| Occlusion rate | yes | no | no | yes | no | no | yes |
| 6D pose | yes | yes | yes | yes | no | yes | yes |
| Depth | yes | yes | yes | yes | yes | yes | yes |
| Point cloud | yes | no | no | no | yes | no | yes |
| Normals | yes | yes | yes | yes | yes | yes | yes |
| Physics | Isaac Sim | PhysX | Bullet | Pybullet | Multiple | Pybullet | Multiple |
| Rendering engine | Multiple | OpenGL | Blender | OptiX | Multiple | Blender | Ogre |
| Rendering capability | ***** | ** | **** | **** | **** | **** | *** |
| Robotics environment | **** | *** | ** | **** | ***** | *** | ***** |
| Open source | no | yes | yes | yes | yes | yes | yes |

TABLE I: Comparison of computer vision softwares developed for the generation of 6D pose estimation annotated datasets with their respective rendering capability and affinity with robotics environment.

Bowen Wen [32] generates his own dataset using Blender for 6D pose tracking, and Jonathan Tremblay [30] generates stereo data using Unreal Engine computer vision. These pipelines don't have direct connections with robotics environment.

Highlighting the need of robotic oriented dataset and showing interest on real robotic dataset, Synpick [24] focuses on combining the physical simulator PhysX and Stilleben [26] in order to compute a dataset on 6D pose tracking while grasping by suction. This is the only 6D pose dataset to our knowledge including robotic manipulation : suction. Unfortunately, PhysX4 stays a low level physics engine and suffer of a lack of high level robotic capabilities as loading robotic arms or even path planning. MetagraspNet [9] had been the first to generate datasets with different levels of difficulty and occlusion rates for robotic bin picking. However this dataset with Isaac sim only work on a specific digital twin scene with YCB objects without domain randomization.

Table II presents an overview of the generation pipeline used in the creation of 6D pose dataset. It can be highlight that, the existing 6d pose datasets are using numerous pipelines to address multiple needs and specificity (rendering, robotics...) and no universal software is used. We propose the closest open source pipeline unifying robotics capabilities and effort in terms of domain randomisation. In the table II, very few are even generating them in robotics environments, while the final application of all of them is robotic grasping.

## III. Generation of synthetic data

The purpose of creating robotic task specific dataset requires task steps as vision and manipulation to be made and trained in the same environment. Therefore, we propose to directly improve a robotics software to facilitate the generation of robotic environment dataset. Indeed, the direct use of a robotic software such as Gazebo in order to generate computer vision synthetic data, offers numerous advantages since it: (i) directly integrates physical engines to get physically realistic results; (ii) enables the easy loading of robots and sensors with native control of them; (iii) unlocks the possibility of creating robotic task oriented datasets. In this section, we present our Gazebo's plugin and complete pipeline PickSim developed to generate dataset suitable for 6D pose estimation for robotics.

### A. A plugin for Gazebo

The original Gazebo software, thanks to its connection with the ROS community, is very suitable to create scenes including several 3D cameras and robot urdf files. Additionally the control of these robots is facilitated by the use of ROS. It enables the connection to different open-sourced libraries, such as MoveIt [2] for path planning. However, it still lacks annotation for computer vision. In this context, we propose a new plugin for Gazebo with new features, developed using the common Ogre [12] and OpenGL libraries. On Fig. 1, we can see the proposed annotations for each frame of dataset: (a) 2D and 3D bounding boxes; (b) 6D poses and occlusion rate of all objects; (c) depth map of the scene ; (d) semantic segmentation. These features are then sent through ROS messages to be easily accessed via the classical ROS architecture.

### B. Our pipeline called PickSim

We now present our complete pipeline PickSim which provides a whole ROS architecture for domain randomization, recording and generating datasets. In addition, scripts are also provided to automatically split the generated datasets for deep learning training, evaluation and benchmark. Our pipeline is composed of four steps.

*a) Pre-processing:* The software allows the user to download some classical computer vision datasets like COCO, ade20k or to download assets. This makes it compatible with any existing open source mesh asset. These meshes are then being processed to automatically generate SDF gazebo compatible models. Finally, our pre-processing step enables to generate several variations of these models in order to build randomized scenes by adjusting textures, objects' properties, poses, etc.

*b) Scene randomization:* In simulation and synthetic data, overcoming the Sim2Real Gap remains a major challenge. For this reason, numerous methods are used, such as fine tuning [**?**], domain adaptation [**?**], and domain randomization [**?**]. To train for models with generalization skills,

| | MetagraspNet | SinPick | Unseen6D | Track Net | Falling things | PickSim |
|---|---|---|---|---|---|---|
| Physic Simulation | Isaac Sim | PhysX 4.1 | BlenderProc | Blender | Unreal | Gazebo |
| Rendering engine | Isaac | Stilleben | BlenderProc | Blender | Unreal | Gazebo |
| Domain randomization | no | no | no | yes | yes | yes |
| Meshes randomization | no | no | no | no | no | yes |
| Robot environment | yes | yes | yes | no | no | yes |
| Robot control | yes | no | no | no | no | yes |
| Sensor based cameras | no | no | no | no | yes | yes |

TABLE II: Comparison of proposed pipeline for generating 6D pose estimation with our Gazebo's plugin called PickSim.

we propose to automatically randomize domains with a wide range of different scenes. For this purpose, configuration files are proposed to design the number of objects, the number of cameras, lighting conditions or texture randomization, without making any line of code. This approach makes it possible to easily generate gazebo world files.

*c) Record data:* Thanks to these generated world files, simulations can then be easily launched in Gazebo. They generate datasets which include the recorded annotations (presented in Table I, with some ones shown in Fig. 1): instance and semantic segmentation, 2D and 3D bounding boxes, bounding box 2d loose, occlusion rate, 6D pose estimation, depth map, point cloud, and norms.

*d) Dataset architecture, post processing and training:* The generated datasets are naturally organised world by world, with N frames taken by N cameras. To avoid losing any information about the current scenes, the dataset automatically generates meta data about the recordings.

This amount of meta data allows us to process and filter the data about the world or camera as needed. One of the advantages of our pipeline is that it makes it possible to count the number of instances on the images and even filter the data according to the different occlusion rates of instances. This allow for the generation of ready to train datasets (training, evaluation and testing) targeting different scenarios or needs.

Moreover, a considerable effort has also been made in the post processing step of the dataset in order to easily split the data for training, evaluation and testing. This will enable the precise study of the generalization of the models for different points of view, different worlds or different occlusion rates. It is hoped that this will be useful when it comes to generating benchmarks for robotics.

## IV. Limitation and future work

In our contributions of Gazebo, some limitations can be highlighted. Firstly Gazebo doesn't have headless. It also has low flexibility in its rendering capabilities and its maintaining end is 2025 as it is for ROS. Thus, future work could be to migrate these contributions to the revised gazebo Ignition or even other open source robotic software, such as SAPIEN [33] both adding ray tracing rendering with OptiX. This could also offer better rendering and annotation computing speed.

Multiple direction of future work are possible : It could be focus on the development of new vision feature making our framework even more global (optical flow, NERF [18], NOCS [31] ...). Further evaluation of the current domain randomization pipeline for real world applications need to be studied. Finally, robotics oriented dataset, including robotic manipulation can be generated for the community.

## V. Conclusion and discussion

In this paper, we proposed: (i) a plugin for Gazebo software; (ii) a complete pipeline for domain randomization; and (iii) a 6D pose benchmark dataset generation pipeline, which could be useful for the computer vision and robotics community. Our development could easily be used for different applications, including 3D reconstruction, 6D pose estimation, 2D instance/semantic segmentation, 2D/3D bounding boxes, object tracking and robot learning. Moreover, our new plugin gives Gazebo the ability to compute state of the art annotations for 6D pose estimation and provides a pipeline to automatically generate datasets on 6D pose estimation and implicitly cover other dataset tasks, including object detection and segmentation or multi-view application for robotics application.

Let's note that the ability to easily generate training data is crucial, but it is equally imperative that the different annotations can directly be used by the robotics communities for their own use and implicitly contribute to other communities and tasks. Thus, the fact that our dataset is generated in a robotics software itself unlock the creation of specific robotic task datasets for a manipulation challenges without using biased pre-trained vision models directly.

Moreover, the current challenge is to use or create a common open source software for the different computer graphics (rendering, efficient simulation, deformation objects, etc.), computer vision (deep scene understanding, etc.) or robotics (robotic control, sensors, etc.) communities. This is not an easy, but a long term objective. We believe that it could only be achieved in a highly flexible software with control over the different modalities (physique engine, rendering, annotations, etc.). Gazebo has been one of the first software offering a good flexibility like multiple physics engine. Gazebo ignition is going even further in adding some annotations such as segmentation masks and multiple rendering engines. Only this kind of flexible software will allow people to add specific task-oriented features. We also hope to convince researchers that Open Source software is the only way of gathering whole communities. Specifically, improving robotics simulation softwares will increase the collaborations between communities and is a major factor for the purpose of robotic manipulation learning.

## References

[1] Steve Borkman, Adam Crespi, Saurav Dhakad, Sujoy Ganguly, Jonathan Hogins, You-Cyuan Jhang, Mohsen Kamalzadeh, Bowen Li, Steven Leal, Pete Parisi, Cesar Romero, Wesley Smith, Alex Thaman, Samuel

Warren, and Nupur Yadav. Unity Perception: Generate Synthetic Data for Computer Vision. jul 2021.

[2] D.M. Coleman, Ioan Alexandru Sucan, Sachin Chitta, and Nikolaus Correll. Reducing the barrier to entry of complex robotic software: a moveit! case study. *ArXiv*, abs/1404.3785, 2014.

[3] Jack Collins, Shelvin Chand, Anthony Vanderkop, and David Howard. A review of physics simulators for robotic applications. *IEEE Access*, 9:51416–51431, 2021.

[4] Erwin Coumans and Yunfei Bai. Pybullet, a python module for physics simulation for games, robotics and machine learning. 2016.

[5] Sudeep Dasari, Frederik Ebert, Stephen Tian, Suraj Nair, Bernadette Bucher, Karl Schmeckpeper, Siddharth Singh, Sergey Levine, and Chelsea Finn. Robonet: Large-scale multi-robot learning. *CoRR*, abs/1910.11215, 2019.

[6] Maximilian Denninger, Martin Sundermeyer, Dominik Winkelbauer, Youssef Zidan, Dmitry Olefir, Mohamad Elbadrawy, Ahsan Lodhi, and Harinandan Katam. BlenderProc. oct 2019.

[7] Maximilian Denninger, Dominik Winkelbauer, Martin Sundermeyer, Wout Boerdijk, Markus Knauer, Klaus H. Strobl, Matthias Humt, and Rudolph Triebel. Blenderproc2: A procedural pipeline for photorealistic rendering. *Journal of Open Source Software*, 8(82):4901, 2023.

[8] Chelsea Finn and Sergey Levine. Deep Visual Foresight for Planning Robot Motion. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 2786–2793, oct 2016.

[9] Maximilian Gilles, Yuhao Chen, Tim Robin Winter, E. Zhixuan Zeng, and Alexander Wong. MetaGraspNet: A Large-Scale Benchmark Dataset for Scene-Aware Ambidextrous Bin Picking via Physics-based Metaverse Synthesis. In *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, volume 2022-Augus, pages 220–227. IEEE, aug 2022.

[10] Minghao Gou, Haolin Pan, Hao-Shu Fang, Ziyuan Liu, Cewu Lu, and Ping Tan. Unseen object 6d pose estimation: A benchmark and baselines, 2022.

[11] Klaus Greff, Francois Belletti, Lucas Beyer, Carl Doersch, Yilun Du, Daniel Duckworth, David J Fleet, Dan Gnanapragasam, Florian Golemo, Charles Herrmann, Thomas Kipf, Abhijit Kundu, Dmitry Lagun, Issam Laradji, Hsueh Ti Liu, Henning Meyer, Yishu Miao, Derek Nowrouzezahrai, Cengiz Oztireli, Etienne Pot, Noha Radwan, Daniel Rebain, Sara Sabour, Mehdi S.M. Sajjadi, Matan Sela, Vincent Sitzmann, Austin Stone, Deqing Sun, Suhani Vora, Ziyu Wang, Tianhao Wu, Kwang Moo Yi, Fangcheng Zhong, and Andrea Tagliasacchi. Kubric: A scalable dataset generator. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2022-June, pages 3739–3751. IEEE, jun 2022.

[12] Gregory Junker. *Pro OGRE 3D programming*. Apress, 2007.

[13] Kilian Kleeberger, Richard Bormann, Werner Kraus, and Marco F Huber. A Survey on Learning-Based Robotic Grasping. *Current Robotics Reports*, 1(4):239–249, 2020.

[14] Nathan Koenig and Andrew Howard. Design and use paradigms for Gazebo, an open-source multi-robot simulator. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3:2149–2154, 2004.

[15] Sergey Levine, Peter Pastor, Alex Krizhevsky, and Deirdre Quillen. Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection.

[16] Zhihao Liu, Quan Liu, Wenjun Xu, Lihui Wang, and Zude Zhou. Robot learning towards smart robotic manufacturing: A review. *Robotics and Computer-Integrated Manufacturing*, 77:102360, oct 2022.

[17] Zhihao Liu, Quan Liu, Wenjun Xu, Lihui Wang, and Zude Zhou. Robot learning towards smart robotic manufacturing: A review. *Robotics and Computer-Integrated Manufacturing*, 77:102360, oct 2022.

[18] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[19] Filipe F. Monteiro, Andre Luiz Buarque Vieira e Silva, João Marcelo Xavier Natário Teixeira, and Veronica Teichrieb. Simulating real robots in virtual environments using nvidia's isaac sdk. *Anais Estendidos do Simpósio de Realidade Virtual e Aumentada (SVR)*, 2019.

[20] Nathan Morrical, Jonathan Tremblay, Yunzhi Lin, Stephen Tyree, Stan Birchfield, Valerio Pascucci, and Ingo Wald. NViSII: A Scriptable Tool for Photorealistic Image Generation. may 2021.

[21] Iman Nematollahi, Erick Rosete-Beas, Seyed Mahdi B. Azad, Raghu Rajan, Frank Hutter, and Wolfram Burgard. T3VIP: Transformation-based 3D Video Prediction. *IEEE International Conference on Intelligent Robots and Systems*, 2022-October:4174–4181, sep 2022.

[22] OpenAI, Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, Jonas Schneider, Nikolas Tezak, Jerry Tworek, Peter Welinder, Lilian Weng, Qiming Yuan, Wojciech Zaremba, and Lei Zhang. Solving rubik's cube with a robot hand, 2019.

[23] Xue Bin Peng, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. In *Proceedings - IEEE International Conference on Robotics and Automation*, pages 3803–3810, 2018.

[24] Arul Selvam Periyasamy, Max Schwarz, and Sven Behnke. SynPick: A Dataset for Dynamic Bin Picking Scene Understanding. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, volume 2021-Augus, pages 488–493. IEEE, aug 2021.

[25] Weichao Qiu and Alan Yuille. UnrealCV: Connecting computer vision to unreal engine. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9915 LNCS, pages 909–916. Springer Verlag, 2016.

[26] Max Schwarz and Sven Behnke. Stillleben: Realistic Scene Synthesis for Deep Learning in Robotics. *Proceedings - IEEE International Conference on Robotics and Automation*, (May):10502–10508, 2020.

[27] Thang To, Jonathan Tremblay, Duncan McKay, Yukie Yamaguchi, Kirby Leung, Adrian Balanon, Jia Cheng, William Hodge, and Stan Birchfield. NDDS: NVIDIA deep learning dataset synthesizer, 2018. https://github.com/NVIDIA/Dataset$_{synthesizer}$.

[28] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. pages 23–30, 09 2017.

[29] Emanuel Todorov, Tom Erez, and Yuval Tassa. MuJoCo: A physics engine for model-based control. *IEEE International Conference on Intelligent Robots and Systems*, pages 5026–5033, 2012.

[30] Jonathan Tremblay, Thang To, and Stan Birchfield. Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, volume 2018-June, pages 2119–21193. IEEE, jun 2018.

[31] He Wang, Srinath Sridhar, Jingwei Huang, Julien Valentin, Shuran Song, and Leonidas J. Guibas. Normalized object coordinate space for category-level 6d object pose and size estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[32] Bowen Wen, Chaitanya Mitash, Baozhang Ren, and Kostas E Bekris. Se(3)-TrackNet: Data-driven 6D pose tracking by calibrating image residuals in synthetic domains. In *IEEE International Conference on Intelligent Robots and Systems*, pages 10367–10373. IEEE, oct 2020.

[33] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, Li Yi, Angel X. Chang, Leonidas J. Guibas, and Hao Su. SAPIEN: A SimulAted Part-Based Interactive ENvironment. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11094–11104. IEEE, jun 2020.