PickSim: A dynamically configurable Gazebo pipeline for robotic manipulation

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

To cite this version:

Guillaume Duret, Nicolas Cazin, Mahmoud Ali, Florence Zara, Emmanuel Dellandréa, et al.. PickSim: A dynamically configurable Gazebo pipeline for robotic manipulation. Advancing Robot Manipulation Through Open-Source Ecosystems - 2023 IEEE International Conference on Robotics and Automation (ICRA) Conference Workshop, Adam Norton, University of Massachusetts Lowell; Holly Yanco, University of Massachusetts Lowell; Berk Calli, Worcester Polytechnic Institute; Aaron Dollar, Yale University, May 2023, Londres, United Kingdom. hal-04074800

HAL Id: hal-04074800
https://hal.science/hal-04074800
Submitted on 3 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
PickSim: A dynamically configurable Gazebo pipeline for robotic manipulation learning

Guillaume Duret\textsuperscript{1,3} guillaume.duret@ec-lyon.fr
Nicolas Cazin\textsuperscript{1} nicolas.cazin@ec-lyon.fr
Mahmoud Ali\textsuperscript{1} mahmoud.ali@ec-lyon.fr
Florence Zara\textsuperscript{2} florence.zara@liris.cnrs.fr
Emmanuel Dellandrea\textsuperscript{1} emmanuel.dellandrea@ec-lyon.fr
Jan Peters\textsuperscript{3} jan.peters@tu-darmstadt.de
Liming Chen\textsuperscript{1} liming.chen@ec-lyon.fr

\textsuperscript{1}Univ Lyon, Centrale Lyon, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69130 Ecully, France
\textsuperscript{2}Univ Lyon, UCBL, CNRS, INSA Lyon, LIRIS, UMR5205, F-69622 Villeurbanne, France
\textsuperscript{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.

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.

\fig{fig:dataset}{Examples of annotations proposed for the dataset generation thanks to our Gazebo’s plugin called PickSim.}
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.
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 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 open source pipeline Stillleben for computer vision. In this context, we propose a new plugin for Gazebo with new features, developed using the common Stillleben for computer vision.

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,
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 Hoggins, You-Cyuan Jhang, Mohsen Kamalzadeh, Bowen Li, Steven Leal, Pete Parisi, Cesar Romero, Wesley Smith, Alex Thaman, Samuel

<table>
<thead>
<tr>
<th>Physical Simulation</th>
<th>Rendering engine</th>
<th>Domain randomization</th>
<th>Meshes randomization</th>
<th>Robot environment</th>
<th>Robot control</th>
<th>Sensor based cameras</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isaac Sim</td>
<td>Isaac</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>PhysX 4.1</td>
<td>Stilleben</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>BlenderProc</td>
<td>BlenderProc</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Blender</td>
<td>Blender</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Unreal</td>
<td>Unreal</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Gazebo</td>
<td>Gazebo</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

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


