

A EXAMPLE OF IMAGES FROM THE DATASET

Figures 8 and 9 showcase the diverse range of scenes captured from two distinct points of view, emphasizing the breadth of scene variations contained within the dataset. These figures serve as visual examples that highlight the dataset’s richness and variety.

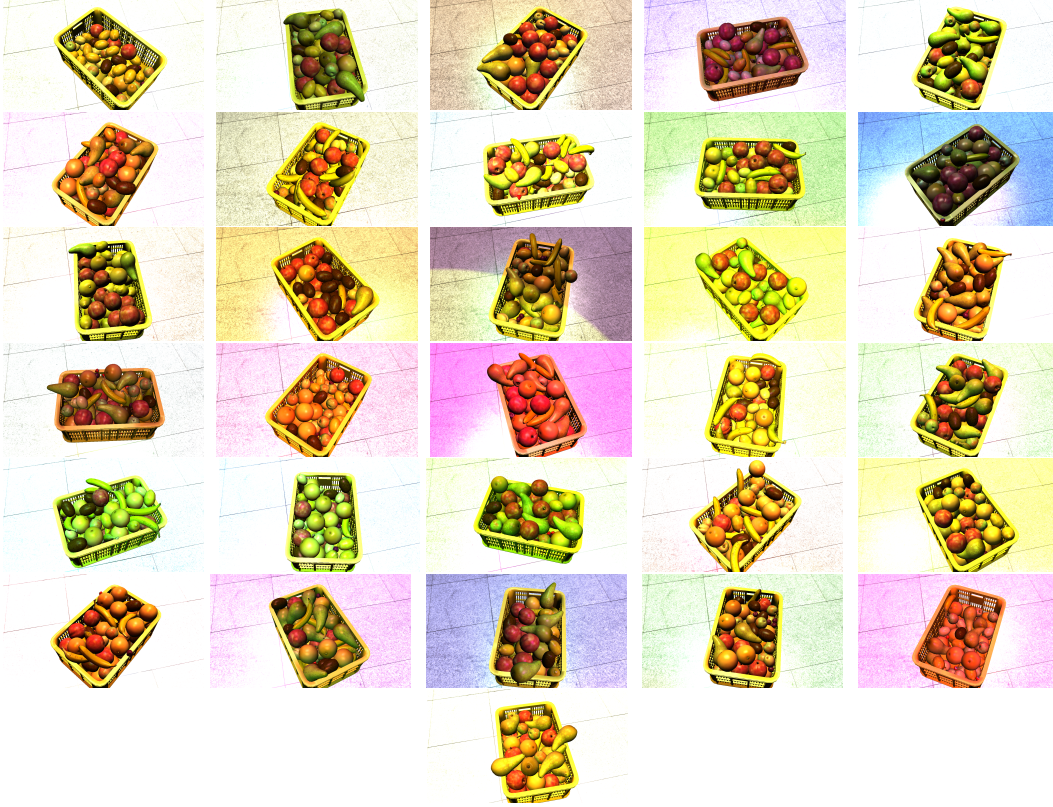


Figure 8: 31 first scenes for a unique point of view

The remaining supplementary material will adhere to the datasheet for dataset guidelines (Gebru et al., 2021) to ensure the provision of comprehensive and practical information about the dataset. It will include a wealth of details and documentation to assist users in effectively utilizing the dataset and gaining a deeper understanding of its characteristics and applications.

B MOTIVATION

- **For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.**

We have generated this dataset as there is currently no known fruit bin picking dataset available. Its purpose is to introduce the demanding new task of fruit bin picking. This dataset effectively establishes benchmark scenarios for 6D pose estimations, encompassing scene generalization, camera point of view generalization, and occlusion robustness.

C COMPOSITION

- **What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)? Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.**

The dataset comprises exclusively of synthetic images depicting fruit bin-picking scenarios in the fields of computer vision and robotics. Each scene consists of a varied assortment of fruits, such as apples, apricots, bananas, kiwis, lemons, oranges, peaches, pears, along with a floor and a bin. A total of 15 images are captured from 15 distinct cameras for each scene.

- **How many instances are there in total (of each type, if appropriate)?**

The dataset consists of an extensive collection of over 1 million images, accompanied by a remarkable total of 40 million 6D pose annotations for all the fruits included.

- **Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).**

The Fruitbin dataset is generated on a large scale using the PickSim pipeline, ensuring that no data is lost throughout the process. For more detailed technical information, please refer to the code folder "fruitbin". It is important to note that while sampling is performed to create benchmarks for 6D pose estimation, this dataset does not aim to replace the existing FruitBin dataset. Both datasets and benchmarks have been developed and made available to the community.

- **What data does each instance consist of? "Raw" data (e.g., unprocessed text or images) or features? In either case, please provide a description.**

Each dataset instance comprises images with a rich number of annotations and metadata.



Figure 9: 31 first scenes for a unique point of view

- Is there a label or target associated with each instance? If so, please provide a description.**
 Each image in the dataset is accompanied by a comprehensive set of annotations and metadata. This includes essential information such as scene ID, camera ID, and detailed annotations for fruit instances, including semantic and instance segmentation masks, 2D bounding boxes (BBboxes), loose 2D BBboxes, 3D BBboxes, 6D pose estimations, and occlusion rates. Various formats are available for these features, including high or low-resolution options for training, as well as different depth and RGB sensors found in the widely used realsense camera D415. For more in-depth information, please refer to the code folder "fruitbin".
- Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.**
 Leveraging the power of synthetic data, ground truths are meticulously generated for every instance, ensuring comprehensive and accurate annotations.
- Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.**
 The relationship between images can be readily established through the available metadata, which includes information about images from the same scenes, captured from the same camera point, and data specific to each fruit category. This capability has been validated through the creation of sub-datasets designed specifically for benchmarking 6D pose estimation tasks.
- Are there recommended data splits (e.g., training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.**
 During the dataset sampling process for benchmark generation, careful splitting is conducted to align with the benchmark's objectives. For the scene generalization benchmark, 60% of the scenes are allocated for training, while 20% is reserved for evaluation with different scenes, and an additional 20% is dedicated to testing. Similarly, the camera point of view generalization follows the same splitting organization: 9 points of view are assigned as training data, 3 are allocated for evaluation, and the remaining 3 are designated for testing. The identical principles are applied to different levels of scenarios involving the occlusion rate. For the reproducibility of the benchmark the code is accessible at the code folders : "pvnet" and "DenseFusion".
- Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.**
 As the dataset is generated within a simulated environment, no known errors or sources of error exist that would undermine the training process for 6D pose estimation. To ensure optimal randomization, the setup has been carefully designed to minimize redundancies across the vast number of images, aiming for the least possible redundancy.
- Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.**
 The dataset is self-contained and not subject to any external restrictions. To ensure proper attribution and promote open sharing, the dataset can be publicly downloaded online and is licensed under Creative Commons (CC BY-NC-SA).
- Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.**
 The dataset is exclusively generated using open-source software within a simulated environment, ensuring complete transparency and openness. There is no confidential data involved in the dataset generation process.
- Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why**
 Our dataset is free of any known causes for concern or anxiety.

D COLLECTION PROCESS AND REARRANGEMENT

- **How was the data associated with each instance acquired? Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.**
The Fruitbin dataset is a large-scale dataset generated through the PickSim pipeline. Although no specific checks have been conducted on the entire dataset, visualizing samples of the different features reveals their quality. No issues or problems have been identified during this process.
- **What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?**
PickSim software is currently available with a Docker container facilitating reproducibility. This integration enables effortless deployment of the dataset generation process across thousands of CPUs in parallel, utilizing the national cluster. The estimated computation time for the generation process is approximately 40,000 CPU hours.
- **If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**
Our control over the generation of scenes is solely based on probabilistic factors, primarily achieved through randomization. The sampling strategy encompasses random variation in lighting conditions, as well as the number and initial placement of objects within the scene. This deliberate setup ensures a diverse range of scenes. The statistic figure 4 from the main paper visually demonstrates the effectiveness of our sampling approach, showcasing a well-distributed range of multi-instance images, spanning from 0 to approximately 20 instances per image.
- **Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?**
The dataset generation process was undertaken with the involvement of a dedicated PhD student.
- **Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.**
The entirety of the dataset was generated around February/March 2023. Even several months later, no recently published dataset has surpassed ours in terms of scene variability and scale for bin picking tasks according to the author’s knowledge. Notably, our dataset is also the only one specifically designed for fruit bin picking applications.
- **Were any ethical review processes conducted (e.g., by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.**
As the dataset is synthetically generated, the potential ethical concerns associated with data collection are minimal. Therefore, no ethical review has been deemed necessary for this dataset.

E PREPROCESSING/CLEANING/LABELING

- **Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.**
The dataset generation process involved three main steps. Firstly, the dataset was generated using PickSim, organized in a folder architecture consisting of world, camera, and feature directories. Secondly, the data was rearranged by features, resulting in the creation of the FruitBin dataset. Lastly, a subsampling procedure was performed on this large dataset to construct benchmark scenarios for occlusion robustness as well as scene and camera point-of-view generalization. To prepare the data for training YOLOv8, Unet, PpNet, and DenseFusion, several preprocessing steps were applied to the subsampled dataset. All the necessary details and scripts for this preprocessing are provided at the code folder "fruitbin".
- **Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)? If so, please provide a link or other access point to the “raw” data.**
Indeed, all the raw data generated using PickSim is intended to be shared with the community for potential future applications. It is important to note that not all data generated with PickSim was utilized for the paper. The generation includes more data from the camera model d415 (including infra1, infra2, and various RGB and depth sensors). Only the depth sensor data is used to get depth and RGB in the same dimension. The dataset, along with all the additional data, will be made available. Currently, only the Scenario data has been released, but the complete FruitBin dataset, including the unused data generated with PickSim, will be released. For more detailed information, please refer to the code folder "fruitbin".
- **Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.**
PickSim is readily accessible and conveniently packaged into a Docker container, simplifying its installation and facilitating testing. You can access the PickSim Docker container in the code folder PickSim.

F USES

- **Has the dataset been used for any tasks already? If so, please provide a description.**
The dataset has been extensively employed in the training of 6D pose estimation models, as well as for training Unet and YOLOv8 models for segmentation purposes. However, its potential applications extend beyond these domains.
- **Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.**
While there is currently no specific link available, considering the success and impact of the dataset, the future of establishing a dedicated link is quite plausible.
- **What (other) tasks could the dataset be used for?**
The dataset has been intentionally crafted with an extensive number of annotations, enabling its utilization across a wide range of tasks. These tasks encompass but are not limited to multi-view 6D pose estimation, multi-view scene reconstruction, Nerf (Neural Radiance Fields), instance and semantic segmentation, 2D and 3D bounding boxes, and even stereo 6D pose estimation (leveraging infrared data). This rich annotation coverage empowers researchers and practitioners to explore diverse applications and extract valuable insights from the dataset.
- **Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?**
The composition of the dataset has been meticulously designed to mitigate any potential unfairness, risks, or harms. Every effort has been made to ensure that the dataset’s contents and characteristics do not introduce any elements that could lead to biased outcomes or negative consequences.
- **Are there tasks for which the dataset should not be used? If so, please provide a description.**
The versatility of the dataset makes it suitable for a wide range of tasks, and there are no known tasks where its usage would be inappropriate or unsuitable. Its comprehensive nature and rich annotations make it a valuable resource for various applications and research endeavors.

G DISTRIBUTION

- **Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.**

This dataset was created with the intention of ensuring equal distribution to the entire community. Its availability is intended to benefit researchers and practitioners across various domains, fostering collaboration and enabling advancements in the field. The goal is to provide equitable access to the dataset and promote widespread utilization for the collective progress of the community.

- **How will the dataset will be distributed (e.g., tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?**
The dataset is currently accessible publicly online ensuring all data is conveniently consolidated in one place for easy access and retrieval. Ensuring anonymity, a small part of the data is available in the supplementary material.
- **When will the dataset be distributed?**
The datasets and the 8 benchmarks scenarios and are currently available in compressed zip files on our website. Considering the substantial size of the dataset, users will have the option to selectively download specific data subsets based on their requirements, such as RGB, depth, and other features, allowing for more efficient and targeted data retrieval.
- **Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.**
The license of the dataset is chosen to be creative commons: CC BY-NC-SA - This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.
- **Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.**
The dataset is readily accessible to all individuals without any constraints beyond the scope of the license. It is open and available for use by anyone interested, fostering inclusivity and facilitating widespread utilization.
- **Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.**
The dataset is made accessible to everyone without any known restrictions or regulations imposed by the authors.

H MAINTENANCE

- **Who will be supporting/hosting/maintaining the dataset?**
Our laboratory has already hosted a variety of datasets, and FruitBin is a valuable addition to this collection, now being shared with the wider community. The authors of the dataset will take responsibility for addressing any maintenance or issues that may arise, ensuring the dataset remains accessible and reliable for users.
- **How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**
To contact the dataset owners, please feel free to reach out the authors of the paper. They will be happy to assist you with any inquiries or concerns you may have regarding the dataset.
- **Is there an erratum? If so, please provide a link or other access point.**
No erratum has been identified in the dataset.
- **Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?**
Currently, the plan is to fully release the Fruit bin dataset, including the raw data. We are open to considering any improvements or corrections that may arise. Any updates or changes will be reflected on the dataset link and the online "fruitbin" code. We encourage users to stay connected to these links to access the latest information and updates regarding the dataset.
- **If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.**
The authors are not aware of any time restrictions associated with the dataset. It can be accessed and utilized without any limitations on the duration of usage.
- **Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.**
Any decision to completely cease maintenance of a previous version of the dataset would require valid justifications. In such cases, comprehensive documentation will be provided to outline the differences between the versions, ensuring transparency and assisting the community in understanding the changes made. The intention is to facilitate smooth transitions and support users in adapting to any updates or modifications to the dataset.
- **If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.**
We highly appreciate and encourage contributions from the community to enhance our dataset. If you have any improvements or suggestions, we invite you to reach out to the authors. Engaging in a discussion regarding the potential benefits of the proposed enhancements would be the most straightforward approach. In certain cases, access to our dataset hosting can be granted to facilitate collaborative efforts and ensure the successful implementation of the improvements. We look forward to collaborating with you and leveraging collective expertise to enhance the dataset for the benefit of all users.