Datasheet for Imitrob dataset

J. Sedlar[1,*]**, K. Stepanova**[1,*]**, M. Tuna**[2]**, R. Skoviera**[1]**, J. Behrens**[1]**,**
G. Sejnova[1]**, J. Sivic**[1]**, R. Babuska**[1,3]
[1]CIIRC, CTU in Prague, [2]MFF, CUNI Bratislava, [3]Cognitive Robotics, 3mE, TU Deflt

1 Datasheet for dataset Imitrob

Questions from the Datasheets for Datasets (https://arxiv.org/abs/1803.09010) paper, v7.

1.1 Motivation

For what purpose was the dataset created?

The *Imitrob* dataset was created with the aim to enable imitation learning of manipulation tasks purely from visual observations. This includes the ability to recognize 6D pose of the hand-held objects. Current methods are typically trained and tested in different conditions than this kind of tasks, so it is very difficult to estimate how they will perform in manipulation tasks with hand-held tools. As expected, the tested methods showed quite low accuracy in the case of the mainipulation with hand-held tools, especially when generalization to new users, camera viewpoints, or tasks was needed. This motivated the creation of a new dataset, which would enable benchmarking of these methods.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by Jiri Sedlar, Karla Stepanova, Radoslav Skoviera, Gabriela Sejnova, Jan K. Behrens, and Josef Sivic within CIIRC CTU in Prague (Imitation learning centre http:\imitrob.ciirc.cvut.cz) in collaboration with Matus Tuna from Comenius University in Bratislava and Robert Babuska from TU Delft.

Who funded the creation of the dataset?

---

*Both authors contributed equally. E-mail: jiri.sedlar@cvut.cz, karla.stepanova@cvut.cz

1.2   Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

The dataset consists of RGB-D images extracted from 104 video sequences, accompanied by 6D annotation. The videos capture simple manipulation tasks with 3 hand-held tools (glue gun, grout float, and roller) such as applying glue along a given trajectory, polishing a surface, or flattening a cloth.

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

The *Imitrob* dataset contains images extracted from 104 video sequences (56 in the *ImitrobTest* dataset and 48 in the *ImitrobTrain* dataset) of hand-held tool manipulations. The *ImitrobTest* component of the dataset contains 61 660 images and the *ImitrobTrain* component contains 39 326 images.

Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

The dataset contains all the possible instances.

What data does each instance consist of?

Each video frame contains the following data:

- 6D pose of the recorded tool (`6DOF/*.json`)
- 2D image coordinates of the tool 3D bounding box vertices and centroid (`BBox/*.json`)
- depth image (`Depth/*.png`)
- RGB image (`Image/*.png`)

In addition, each frame of the *ImitrobTrain* dataset also contains:

- binary mask of the segmented tool and hand (`Mask_thresholding/*.png`)
- RGB image with the segmented tool and hand opaque and the background transparent (`Mask/*.png`)

Each video sequence in the *Imitrob* dataset contains:

- 3D coordinates of the tool bounding box vertices and centroid with respect to the HTC Vive Tracker (`BB_in_tracker`) and intrinsic camera matrices for cameras C1 (`K_C1`) and C2 (`K_C2`) (`parameters.json`)

The training/test component, tool, task, subject, camera, left/right hand or presence/absence of clutter are identified in the name of the video sequence folder.

Is there a label or target associated with each instance?

Yes, each image is annotated with the 6D pose of the tool as well as the video sequence labels, including the identifier of the training/test component, tool, task, subject, camera viewpoint, left/right hand, or presence/absence of clutter.

Is any information missing from individual instances?

The 6D pose for individual data frames was interpolated. When the time difference between consecutive HTC Vive frames was longer than 100 ms, the corresponding camera images were discarded to ensure sufficient accuracy of the ground truth data.

Otherwise no information is missing and the data is complete.

Are relationships between individual instances made explicit (e.g., users' movie ratings, social network links)?

Yes, the relationships are fully identified by the video sequence labels (see above) and the position of the frame in the sequence.

Are there recommended data splits (e.g., training, development/validation, testing)?

We explicitly state the data splits used for training and testing of the 6D pose estimator. The training set is an (augmented) subset of the *ImitrobTrain* dataset, and the test set is a subset of the *ImitrobTest* dataset.

Are there any errors, sources of noise, or redundancies in the dataset?

Sources of noise include the calibration of the cameras and the HTC Vive controllers and the synchronization between the HTC Vive and the cameras (the HTC Vive data were interpolated to the closest camera frame and if the distance between 2 consecutive frames was longer than 100 ms, the corresponding camera images were discarded to ensure sufficiently accurate ground truth data).

Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?

Both the dataset and the supplementary code are self-contained.

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)?

N/A.

Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?

N/A.

Does the dataset relate to people?

N/A.

Does the dataset identify any subpopulations (e.g., by age, gender)?

N/A.

Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?

N/A.

Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals racial or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?

N/A.

Any other comments?

N/A.

1.3 Collection process

How was the data associated with each instance acquired?

The directly observable data (RGB-D images) were synchronized with observable HTC Vive data. The parameters of each video sequence setup (such as the tool, task, subject, camera viewpoint, left/right hand, or presence/absence of clutter) were manually annotated and associated with the corresponding data.

What mechanisms or procedures were used to collect the data (e.g., hardware apparatus or sensor, manual human curation, software program, software API)?

The visual part of the dataset was collected by two RGB-D cameras, specifically Intel RealSense D-435. The resolution of both RGB and depth images was set to 848x480 and they were recorded at 60 FPS. The 6D pose information was recorded using HTC Vive VR system in standard configuration. An HTC Vive tracker was mounted to the tools to acquire their pose. The cameras and HTC Vive system were calibrated towards a common coordinate system. The calibration of the camera and HTC Vive was validated by the average distances of associated points using a checkerboard pattern. The whole acquisition system was implemented via the Robot Operating System, using Python as the main programming language.

If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?

N/A.

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)?

Only the authors were involved in the collection process.

Over what timeframe was the data collected?

The data was collected over the timeframe of 2 months (1-2/2021).

Were any ethical review processes conducted (e.g., by an institutional review board)?

N/A.

Does the dataset relate to people?

N/A.

1.4 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)?

The data were originally recorded as ROS bag files, from which the individual data instances were extracted, synchronized, interpolated, and saved to separate folders.

For the *ImitrobTrain* dataset, the masks were created by automatic segmentation of the RGB images.

Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?

The original bag files are saved on our internal data storage, but are too big to be easily shareable.

Is the software used to preprocess/clean/label the instances available?

No, we don't provide the raw data and thus neither the code to process it. Dataset manipulation tools (for the already preprocessed and labeled data) are available on the supplementary code GitHub page: https://github.com/imitrob/imitrob_dataset_code.

1.5 Uses

Has the dataset been used for any tasks already?

This is the first use of the dataset.

Is there a repository that links to any or all papers or systems that use the dataset?

There are no papers that use our dataset, yet. Future uses will be added to the dataset/code website.

What (other) tasks could the dataset be used for?

The dataset is primarily intended for benchmarking 6D pose estimation methods in manipulation tasks with hand-held objects and evaluating their ability to generalize with respect to various conditions. It can be also used to evaluate the effect of different data augmentation methods. Another usage is the methodology for data acquisition and 6D pose estimator training for new tools and tasks and a guideline for collecting more extensive datasets and benchmarking 6D object pose estimators on various tasks with hand-held tools, e.g. in imitation learning, grasping, virtual or augmented reality, etc. In general, we hope that the presented dataset will trigger further development of 6D object pose estimation methods and their usage in various industrial tasks based on the required accuracy.

Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

N/A.

Are there tasks for which the dataset should not be used?

N/A.

1.6 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?

N/A.

How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?

The dataset is available on the dataset website: http://imitrob.ciirc.cvut.cz/imitrobdataset.php

When will the dataset be distributed?

In 2022.

Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?

The newly provided datasets and benchmarks are copyrighted by us and published under the CC BY-NC-SA 4.0 license[2].

---

[2] https://creativecommons.org/licenses/by-nc-sa/4.0/

Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

N/A.

Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

N/A.

1.7 Maintenance

Who is supporting/hosting/maintaining the dataset?

Karla Stepanova at CIIRC CTU in Prague.

How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

Contact e-mail address: karla.stepanova@cvut.cz

Is there an erratum?

Any updates to the code will be visible as commits in the GitHub repository. The dataset website will summarize all changes to the code and the dataset.

Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?

Any updates to the code will be visible as commits in the GitHub repository. The dataset website will summarize all changes to the code and 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 individuals in question told that their data would be retained for a fixed period of time and then deleted)?

N/A

Will older versions of the dataset continue to be supported/hosted/maintained?

N/A.

If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?

Yes, contributions to the dataset are welcome. Please contact the maintainer of the dataset via e-mail (see above).