Model checkpoint

Explanation of observation space

The model takes as input a RGB image from the robot workspace camera and a task string describing the task that the robot is supposed to perform.

What task the model should perform is communicated to the model purely through the task string. The image communicates to the model the current state of the world, i.e. **assuming** the model runs at three hertz, every 333 milliseconds, we feed the latest RGB image from a robot workspace camera into the model to obtain the next action to take.

Please note that the model currently does not take in additional camera images such as wrist camera images, in hand camera images, or depth.

Explanation of action space

The action dimensions we consider include seven variables for the gripper movement (x, y, z, roll, pitch, yaw, opening of the gripper). Each variable represents the absolute value, the delta change to the dimension value or the velocity of the dimension.

Link to inference colab of trained RT-1-X Tensorflow checkpoint.

Dataset Access

Dataset structure

Each data set is represented as a sequence of episodes, where each episode is represented using the RLDS episode format.

Dataset colab

We provide a self-contained colab that demonstrates how to visualize a few episodes from each data set, and how to create batches of data ready for training.

List of datasets

We provide the list of dataset that is included as part of the open-sourcing effort and their metadata in the dataset spreadsheet.

Citation

If you're using the Open X-Embodiment dataset and RT-X in your research, please cite. If you're specifically using datasets that have been contributed to the joint effort, please cite those as well. The dataset spreadsheet contains the citation for each dataset for your convenience.

Copyright Notice