Files description

The dataset includes two splits, `seen` and `unseen.` `seen` contains images of objects used in training, while `unseen` comprises images of objects reserved for testing.

In each split “seen” and “unseen”, there are three folders:

- `image`: contains *.jpg* images of each scene. Each scene is identified by a SHA-256 string, for instance, `0a5bd779e492513880bef534543ff031b169a045ed7ac809c5600c3268038f4d`. The size of each image is 416×416.
- `grasp_instructions`: contains *.pkl* grasp instructions. Each grasp instruction corresponds to an image, represented by the prefix of the file name. For example, if the grasp instruction file name is `0a5bd779e492513880bef534543ff031b169a045ed7ac809c5600c3268038f4d_0_0.pkl`, it corresponds to the image `0a5bd779e492513880bef534543ff031b169a045ed7ac809c5600c3268038f4d.jpg`. You can load the grasp instruction by using pickle, it should be a sentence like this: “*Grab hairbrush on its bristles.*"
- `grasp_label`: contains *.pt* grasp labels. The file name convention is the same as `grasp_instructions`. Each grasp label file contains a `torch.Tensor` shape of $M \times 6$, where M represents the varying number of grasp poses across files. The grasp pose is detailed by six parameters:

$$(F, x, y, w, h, \theta)$$

Here, F indicates the quality of the grasp poses, which may be ignored. $\{x, y, w, h, \theta\}$ are five parameters defined in the introduction, describing the position and rotation of the grasp pose.