

Supplementary for Reconstruct Locally, Localize Globally: A Model Free Method for Object Pose Estimation



The results of our method on YCB-Video Dataset. We train our network using 20% images of each sequence, and visualize the results on samples of the rest. These result images show that our method also works for multi-object pose estimation. Thanks to the nature of Mask R-CNN, the extended model is able to detect, segment and estimate the pose for multiple objects at the same time, in a single forward. The visualization contains the results of detection(in bounding boxes), segmentation(in contours), and pose estimation. We projection the 3D points of the ground truth object model using the predicted poses. The shape of the predicted projections match up with the silhouettes of the objects.