# MegaPose: 6D pose estimation of novel objects via render&compare

**Anonymous Author(s)**
Affiliation
Address
email

## Response to the reviewers

## 1 Robustness to illumination conditions

In Figure 1, we show qualitative predictions of our approach for the watering can on the TUD-L dataset. Please notice the high accuracy of our approach despite challenging illumination conditions.

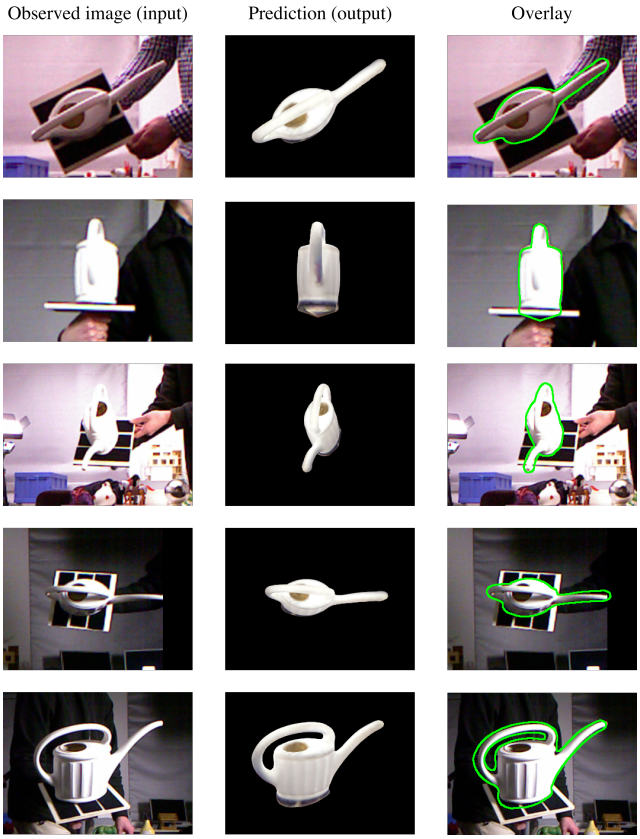


**Figure 1:** Qualitative examples on the TUD-L dataset. Each row presents one example prediction on a real image. The first column is the real observed image, the second column is the prediction of our approach here illustrated using a rendering of the object's CAD model in the predicted pose, and an overlay of the prediction and output is shown on the right.

Submitted to the 6th Conference on Robot Learning (CoRL 2022). Do not distribute.

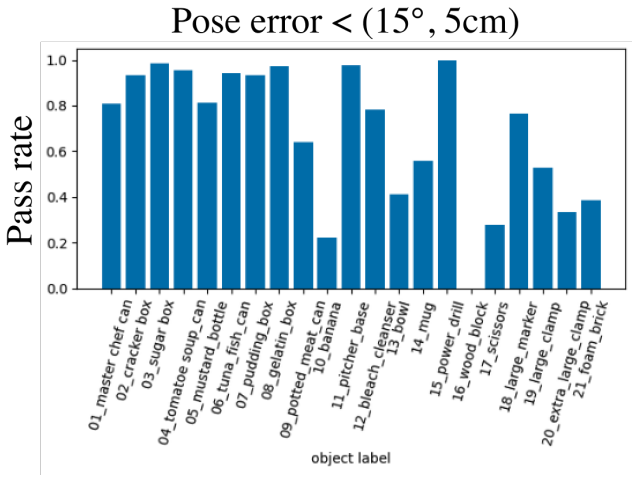


**Figure 2:** Per-object analysis on the YCB-V dataset. For each object, we report the percentage of estimates for which the error between our pose prediction and the ground truth is within 5 centimeters in translation and 15 degrees in rotation.

## 2 Failure modes and performance on specific types of objects

We carry out a per-object analysis of the performance of our approach on the YCB-V dataset. For each of the 21 objects of the dataset, we report the percentage of predictions for which the error with the ground truth is within a threshold of $15^\circ$ in rotation and 5cm in translation. Results are reported in Figure 2.

Next, we illustrate the main failure modes of our approach using a set of objects which have a performance below average on this dataset. Examples of failure cases are presented in Figure 4. We observed three main failure modes to our approach. First, we observe the orientation of a novel object may be incorrectly predicted if the object has a similar visual appearance under different viewpoint. We observed this failure mode in particular for textureless objects such as a red bowl that appears similar whether it is standing upside or it is flipped. Second, we observe that our approach may fail to disambiguate the pose of objects that are asymmetric but for which it is necessary to look at fine details on the objects to disambiguate multiple possible poses. An example is a pair of scissors which have left and right handles with slightly different dimensions. In both of these failure modes, we observed that our refiner gets stuck into a local minimal due to an inaccurate coarse estimate outside of the basin of attraction of our refiner model. Finally, using a CAD model with incorrect scale leads to an incorrect estimation of the depth of the object due to the object scale/depth ambiguity in RGB images. We observe for example that the translation estimates of the wooden block of YCB-V have systematically large error despite the rendering of our prediction correctly matching the contours of the object in the observed image. This is because the scale of the CAD model of the wooden block publicly available does not match the correct dimensions of the real object which was used for annotating the ground truth.

## 3 CAD model quality

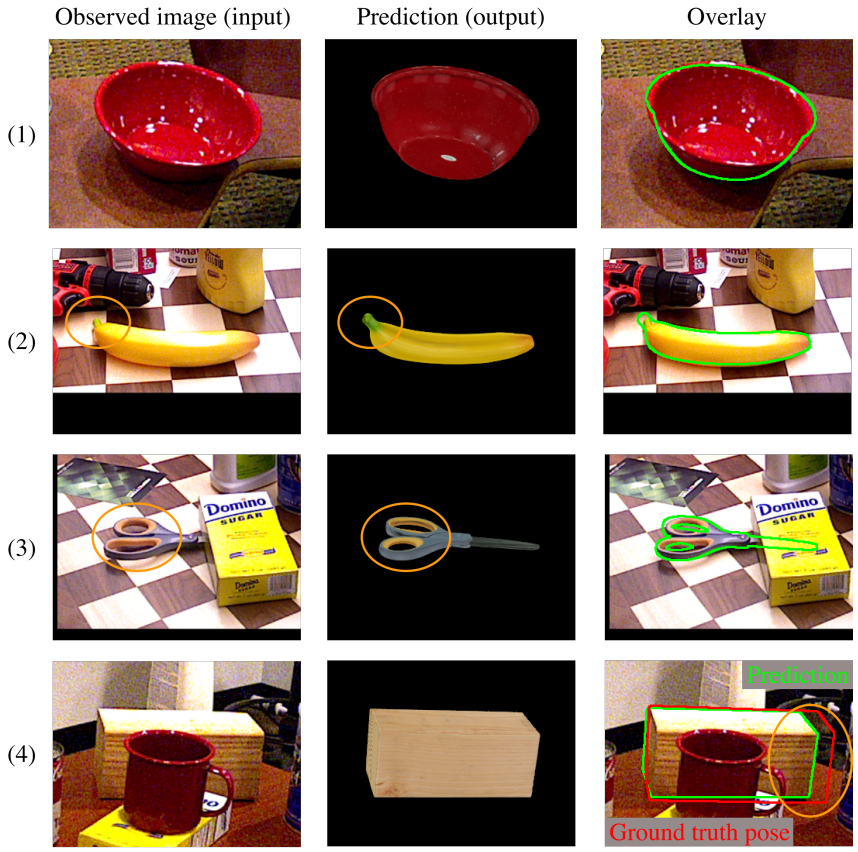


**Figure 3:** Illustration of the main failure modes of our approach. In (1) and (2), the contours of the object in the predicted poses correctly overlay the observed image, but the pose is incorrect because these objects have a similar appearance under different viewpoints. In (3), our approach fails to correctly distinguish the left and right handles with different dimensions in order to disambiguate the orientation of the asymmetric pair of scissors. In (4), our pose prediction does not match the ground truth annotation, because the CAD model of the wooden block we use for pose estimation has different dimensions that do not match the dimensions of the real objects which was used for annotating the ground truth. Please notice in all examples how the contours of the object in the predicted pose are closely aligned with the contours of the object in the input image.

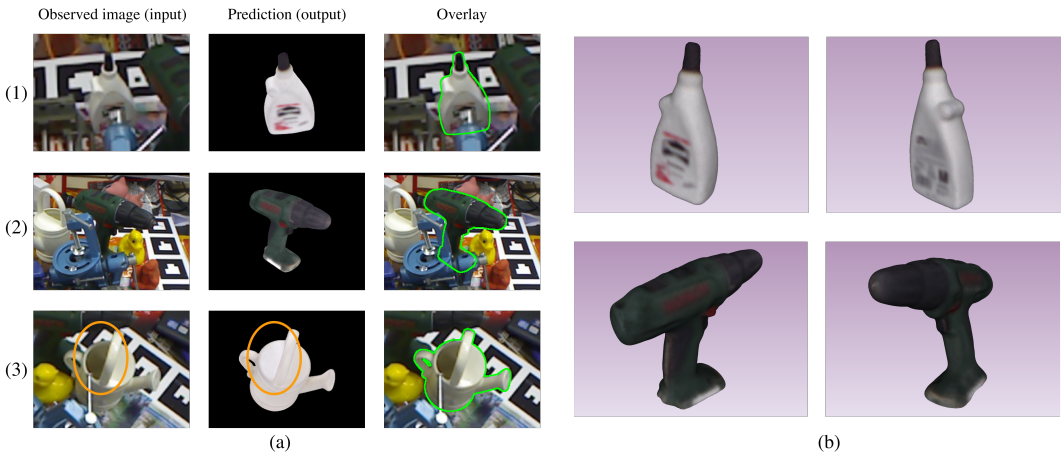


**Figure 4:** Predictions using low-fidelity CAD models. In (a) we show the result of our approach on LineMOD Occlusion for three different objects which have only low-fidelity CAD models available. In (1) and (2), the quality of the mesh and textures is poor as illustrated in (b). Notice for example how the annotations on the glue box or the brand of the drill are not readable on the CAD models. In (3), the hole of the watering can does not appear in the CAD model. Despite these discrepancies between the real object and the CAD model, our approach correctly estimates the pose of each object.