1. Hyper-parameters in the Pose Solver

For RANSAC/PnP [7], we set the threshold value for re-projection error as 2 pixels, and execute 150 iterations. For Progressive-X [1], we also set the threshold value for the reprojection error as 2 pixels, and execute 400 iterations. The additional parameters for Progressive-X are "neighborhood_ball_radius=20", "spatial_coherence_weight=0.1", "maximum_tanimoto_similarity=0.9".

2. BOP Challenge

We submitted the results on 4 datasets of the BOP challenge and will test our method on the rest 3 datasets. The results are online in BOP Leaderboards with the submission name "zebrapose".

3. YCB-V Evaluation per Object

We present a more detailed result on the YCB-V dataset [10] in Tab. 1 and Tab. 2. As the Tab. 1 shows, in the evaluation of the estimate pose w.r.t ADD(-S) metric, we show major improvement over the state of the art.

In Tab. 2, we carefully calculated the AUC with all-points interpolation algorithm with the maximum threshold of 10 cm. If we calculate the AUC with 11-points interpolation, we will reach AUC of ADD-S of 94%, and AUC of ADD(-S) of 89.8%.

4. Qualitative Results

4.1. Vertex Code Prediction LM-O

We visualized the predicted binary code of the "duck" object in LM-O dataset [2] with a few examples in Fig. 1. Due to the size limits, we only show the predicted binary code till the 11-th bits. We render the object with the predicted pose on top of the original input ROI. To make the predicted pose more visible in the figure, we set the colour of the object model as red just for this figure. So the duck appears with the orange colour (red + yellow) in the last row. We can see that the rendered object overlapped the object in the original image quite well, indicating that our predicted pose is very accurate.

4.2. Pose Prediction LM-O

Qualitative Results on LM-O [2] can be found in Fig. 2. We render the objects with estimated pose on top of the original images. The presented confidence scores are from the 2D object detection with FCOS detector [8].

4.3. Pose Prediction YCB-V

Qualitative Results on YCB-V [10] are available in Fig. 3. We render the objects with estimated pose on top of the original images. The presented confidence scores are from the 2D object detection with FCOS detector [8].
Figure 1. We visualized the predicted binary code of the "duck" in LM-O dataset [2] with a few examples. Due to the size limits, we only show the predicted binary code till the 11-th bit. We set the colour of the object model as red and render the object with the predicted pose on the top of the input ROI. We can see that the rendered object overlaps the object in the image quite well.
Figure 2. **Qualitative Results on LM-O [2]**: We render the objects with estimated pose on top of the original images. The presented confidence score are from the 2D object detection with FCOS detector [8].
Table 1. **Comparison with State of the Art on YCB-V**. We report the Average Recall of ADD(-S) in % and compare with state of the art. (*) denotes symmetric objects, (-) denotes the results missing from the original paper.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric</td>
<td>AUC of ADD-S</td>
<td>AUC of ADD(-S)</td>
<td>AUC of ADD-S</td>
<td>AUC of ADD(-S)</td>
</tr>
<tr>
<td>002_master_chef_can</td>
<td>84.0</td>
<td>50.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>003_cracker_box</td>
<td>76.9</td>
<td>51.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>004_sugar_box</td>
<td>84.3</td>
<td>68.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>005_tomato_soup_can</td>
<td>80.9</td>
<td>66.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>006_mustard_box</td>
<td>90.2</td>
<td>79.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>007_tuna_fish_can</td>
<td>87.9</td>
<td>70.4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>008_pudding_box</td>
<td>79.0</td>
<td>62.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>009_gelatin_box</td>
<td>87.1</td>
<td>75.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>010_potted_meat_can</td>
<td>78.5</td>
<td>59.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>011_banana</td>
<td>85.9</td>
<td>72.3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>019_pitcher_base</td>
<td>76.8</td>
<td>52.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>021_bleach_cleanser</td>
<td>71.9</td>
<td>50.5</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>024_bowl*</td>
<td>69.7</td>
<td>69.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>025_mug</td>
<td>78.0</td>
<td>57.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>035_power_drill</td>
<td>72.8</td>
<td>55.1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>036_wood_block*</td>
<td>65.8</td>
<td>65.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>037_scissors</td>
<td>56.2</td>
<td>35.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>040_large_marker</td>
<td>71.4</td>
<td>58.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>051_large_clamp*</td>
<td>49.9</td>
<td>49.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>052_extra_large_clamp*</td>
<td>47.0</td>
<td>47.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>061_foam_brick*</td>
<td>87.8</td>
<td>87.8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>mean</td>
<td>75.9</td>
<td>61.3</td>
<td>89.8</td>
<td>84.5</td>
</tr>
</tbody>
</table>

Table 2. **Comparison with State of the Art on YCB-V.** We report the Average Recall w.r.t AUC of ADD(-S) and AUC of ADD-S in % and compare with state of the art. (*) denotes symmetric objects, (-) denotes the results missing from the original paper.
Figure 3. Qualitative Results on YCB-V [10]: We render the objects with estimated pose on top of the original images. The presented confidence score are from the 2D object detection with FCOS detector [8].
References

[1] Daniel Barath and Jiri Matas. Progressive-x: Efficient, any-
time, multi-model fitting algorithm. In Proceedings of the
IEEE/CVF International Conference on Computer Vision,

Ying Yang, Stefan Gumhold, and others. Uncertainty-driven
6d pose estimation of objects and scenes from a single rgb
image. In Proceedings of the IEEE Conference on Computer

Single-stage 6d object pose estimation. In Proceedings of
the IEEE/CVF conference on computer vision and pattern

[4] Yinlin Hu, Joachim Hugonot, Pascal Fua, and Mathieu Salz-
mann. Segmentation-driven 6d object pose estimation. In
Proceedings of the IEEE/CVF Conference on Computer Vi-
sion and Pattern Recognition, pages 3385–3394, 2019.

[5] Shun Iwase, Xingyu Liu, Rawal Khirodkar, Rio Yokota, and
Kris M Kitani. Repose: Fast 6d object pose refinement via
deep texture rendering. In Proceedings of the IEEE/CVF
International Conference on Computer Vision, pages 3303–
3312, 2021.

[6] Yann Labbé, Justin Carpentier, Mathieu Aubry, and Josef
Sivic. Cosypose: Consistent multi-view multi-object 6d pose
estimation. In European Conference on Computer Vision,

Epnp: An accurate o (n) solution to the pnp problem. Inter-

[8] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos:
Fully convolutional one-stage object detection. In Proceed-
ings of the IEEE/CVF international conference on computer

[9] Gu Wang, Fabian Manhardt, Federico Tombari, and Xi-
angyang Ji. Gdr-net: Geometry-guided direct regression net-
work for monocular 6d object pose estimation. In Proceed-
ings of the IEEE/CVF Conference on Computer Vision and

[10] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and
Dieter Fox. Posecnn: A convolutional neural network for
6d object pose estimation in cluttered scenes. arXiv preprint