Pos3R: 6D Pose Estimation for Unseen Objects Made Easy

Supplementary Material

I. In-Plane (Axis) Rotation Number T. In the main paper, for each template $\mathbf{I}_i$, we generate T rotations around the camera's principal axis, each rotation defined by an angle $\theta_k = \frac{2\pi k}{T}$, where $k = 0, \ldots, T-1$, covering the full 360° range. Here, we study the effect of axis rotation number T in Fig. 1. We observe that increasing T improves AR on the LM-O dataset. $T = 5$ achieves a good trade-off between performance and computational cost, with $T = 7$ offering minimal additional improvement.

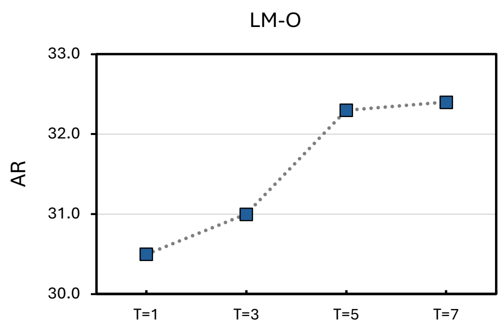

Figure 1. **The Effect of Axis Rotation Number T.** We show the impact of T on the average recall (AR) for LM-O [1]. We observe that increasing T improves AR, with $T = 5$ offering a good balance between performance and computational cost. While $T = 7$ achieves slightly higher AR, the improvement over $T = 5$ is minimal, making $T = 5$ a practical choice

II. Template Number Comparison. Table 1 compares the number of templates required by different methods. FoundPose [6] uses 600 templates, MegaPose [4] uses 576, and GigaPose [5] reduces this to 162. In contrast, Pos3R (ours) achieves competitive performance with only 40 templates, demonstrating its efficiency in significantly reducing template requirements while maintaining effectiveness.

#	Method	Template Number
1	FoundPose [6]	600
2	MegaPose [4]	576
3	GigaPose [5]	162
4	Pos3R (ours)	40

Table 1. **Comparison of template numbers used by different methods.** FoundPose [6] uses 600 templates, MegaPose [4] uses 576 templates, and GigaPose [5] uses 162 templates, while Pos3R (ours) achieves competitive results with only 40 templates.

III. Experimental Parameter Settings. Unless stated otherwise, all experiments are conducted using a single NVIDIA A100 GPU. The default subsample size of 8 is used for FastNN's subsample of MASt3R. Our method is entirely training-free, with no task-specific training involved. The size of templates and test segments is set to 320×320 pixels. For each test segment, the pose is estimated using PnP-RANSAC, which runs up to 400 iterations with an inlier threshold of 10 pixels. All masks are loaded from default CNOS mask files provided for the BOP 2023 benchmark [3]. We use the default hyperparameter of RoMa [2] for the component analysis.

References

[1] Eric Brachmann, Alexander Krull, Frank Michel, Stefan Gumhold, Jamie Shotton, and Carsten Rother. Learning 6d object pose estimation using 3d object coordinates. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pages 536–551. Springer, 2014. 1

[2] Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. RoMa: Robust Dense Feature Matching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2024. 1

[3] Tomas Hodan, Martin Sundermeyer, Yann Labbe, Van Nguyen Nguyen, Gu Wang, Eric Brachmann, Bertram Drost, Vincent Lepetit, Carsten Rother, and Jiri Matas. BOP challenge 2023 on detection segmentation and pose estimation of seen and unseen rigid objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5610–5619, 2024. 1

[4] Yann Labbé, Lucas Manuelli, Arsalan Mousavian, Stephen Tyree, Stan Birchfield, Jonathan Tremblay, Justin Carpentier, Mathieu Aubry, Dieter Fox, and Josef Sivic. Megapose: 6d pose estimation of novel objects via render & compare. *arXiv preprint arXiv:2212.06870*, 2022. 1

[5] Van Nguyen Nguyen, Thibault Groueix, Mathieu Salzmann, and Vincent Lepetit. Gigapose: Fast and robust novel object pose estimation via one correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9903–9913, 2024. 1

[6] Evin Pınar Örnek, Yann Labbé, Bugra Tekin, Lingni Ma, Cem Keskin, Christian Forster, and Tomas Hodan. Foundpose: Unseen object pose estimation with foundation features. In *European Conference on Computer Vision*, pages 163–182. Springer, 2025. 1