Pos3R: 6D Pose Estimation for Unseen Objects Made Easy

Weijian Deng[1], Dylan Campbell[1], Chunyi Sun[1], Jiahao Zhang[1], Shubham Kanitkar[2], Matthew E. Shaffer[2], Stephen Gould[1]
[1]The Australian National University [2]RIOS Intelligent Machines

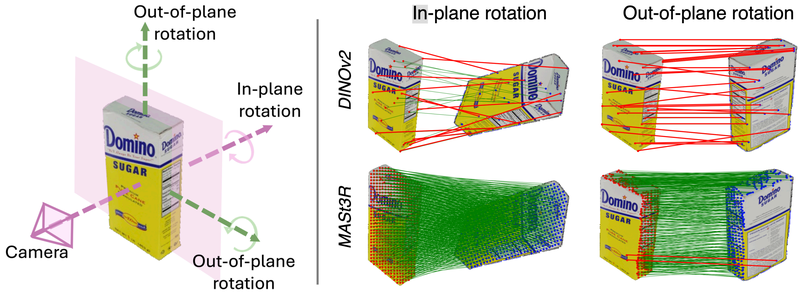
CVPR Nashville JUNE 11-15, 2025

I. Unseen Object Pose Estimation

- Task Definition: Training Free, RGB, and CAD Model
  - Training-free pipelines offer adaptability to unseen objects
  - Model-based 6D localization estimates object pose from a 3D CAD model and an RGB image

II. Why Use a 3D Foundation Model?

- Motivation
  - 2D foundation models (*e.g.*, DINOv2) have been shown to be effective at training-free pose estimation, but are not consistent under significant 3D transformations
  - 3D foundation models (*e.g.*, MASt3R) predict 3D-consistent features, which we show to be useful for pose estimation
- Correspondence Matching Quality

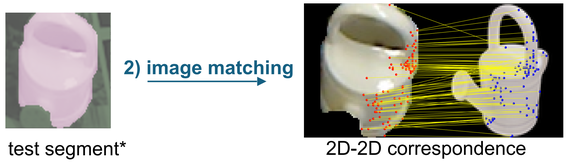


  - DINOv2 produces inconsistent matches under out-of-plane rotations since it was not trained under these transformations
  - MASt3R provides dense and stable correspondences

III. Pos3R

- Pos3R: Training-Free and Fast — Render, Match, Fit
  - Step 1: Template Rendering

    - Eight base template, covering essential orientations
    - For each, five in-plane (axial) rotations are generated around the camera's principal axis.
  - Step 2: Image Matching



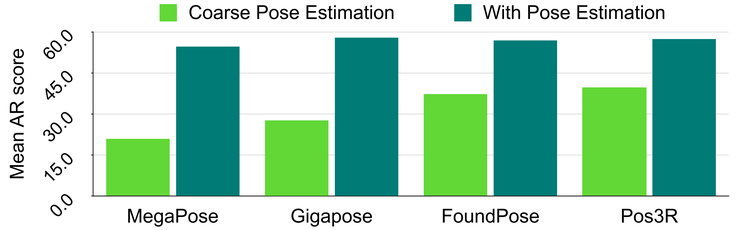
** provided by CNOS (Cnos: A strong baseline for cad-based novel object segmentation)*

    - MASt3R produces dense 2D correspondences between test segment and every template
    - Similarity is computed by summing feature similarities across correspondences
  - Step 3: Pose Fitting

    - The template with the highest score is selected
    - 3D coordinate map provides 2D-3D matches for P*n*P to estimate pose

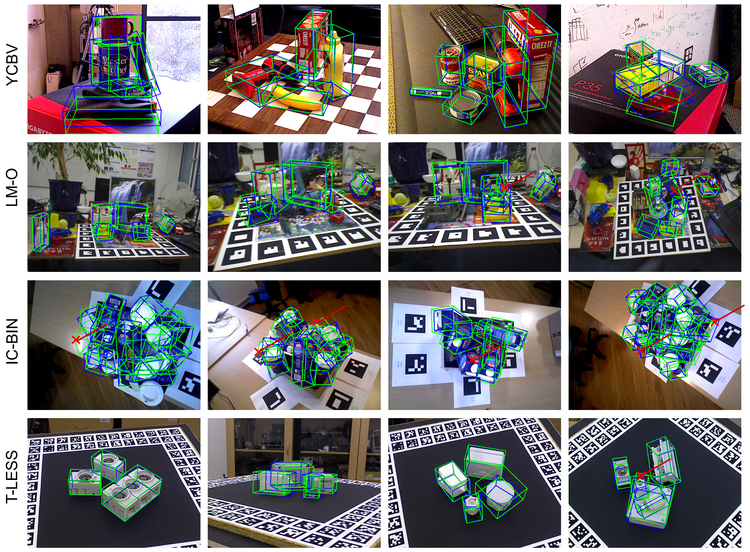
IV. Experiments

- Performance Comparison on the BOP Challenge



  - Pos3R outperforms other methods in coarse pose estimations
  - With pose refiner provided by MegaPose, Pos3R remains competitive
- Qualitative Results of 6D Pose Estimates

blue *indicates ground truth;* ***green*** *indicates the estimate*



  - Pos3R is robust to crowding, lighting changes, and texture-less objects
  - Limitation: heavy occlusion (**X**) poses a challenge