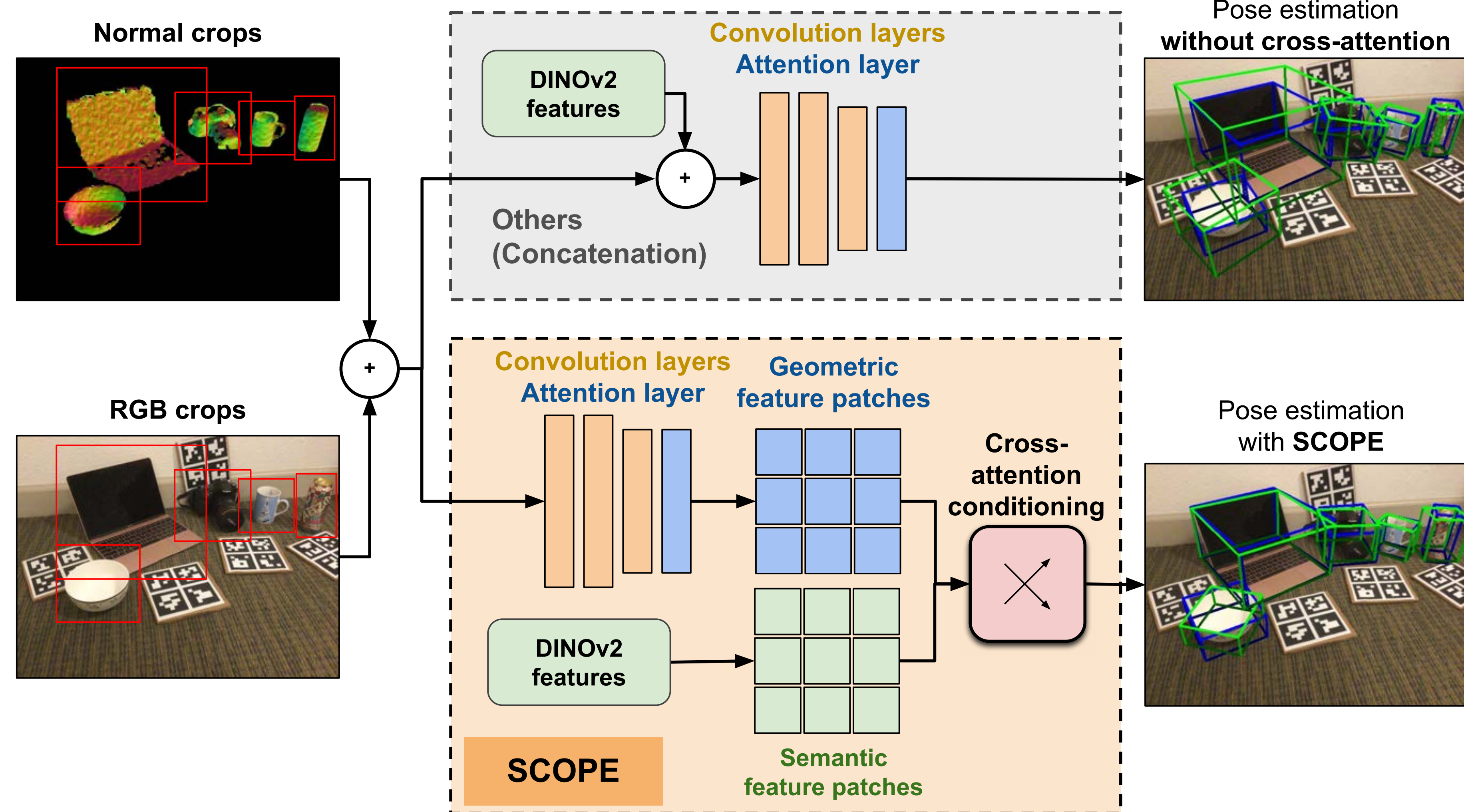


Peter Hönig, Stefan Thalhammer, Jean-Baptiste Weibel, Matthias Hirschmanner, and Markus Vincze

Motivation

- Robots must manipulate **unseen objects** in **open environments**.
- Category-level pose estimation often depends on **discrete labels** → **limited**.
- Key insight: **Replace category labels with continuous semantic priors** (DINOv2 features).
- Use **cross-attention conditioning** in a **diffusion model** → **better generalization & Sim2Real transfer**.

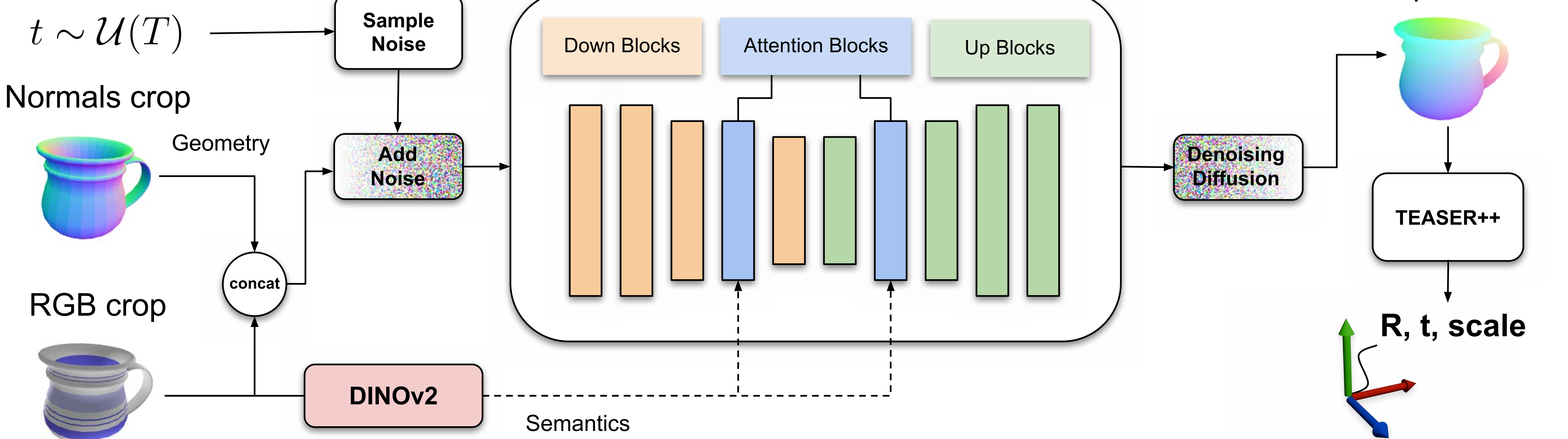


Method

Gaussian Noise

SCOPE

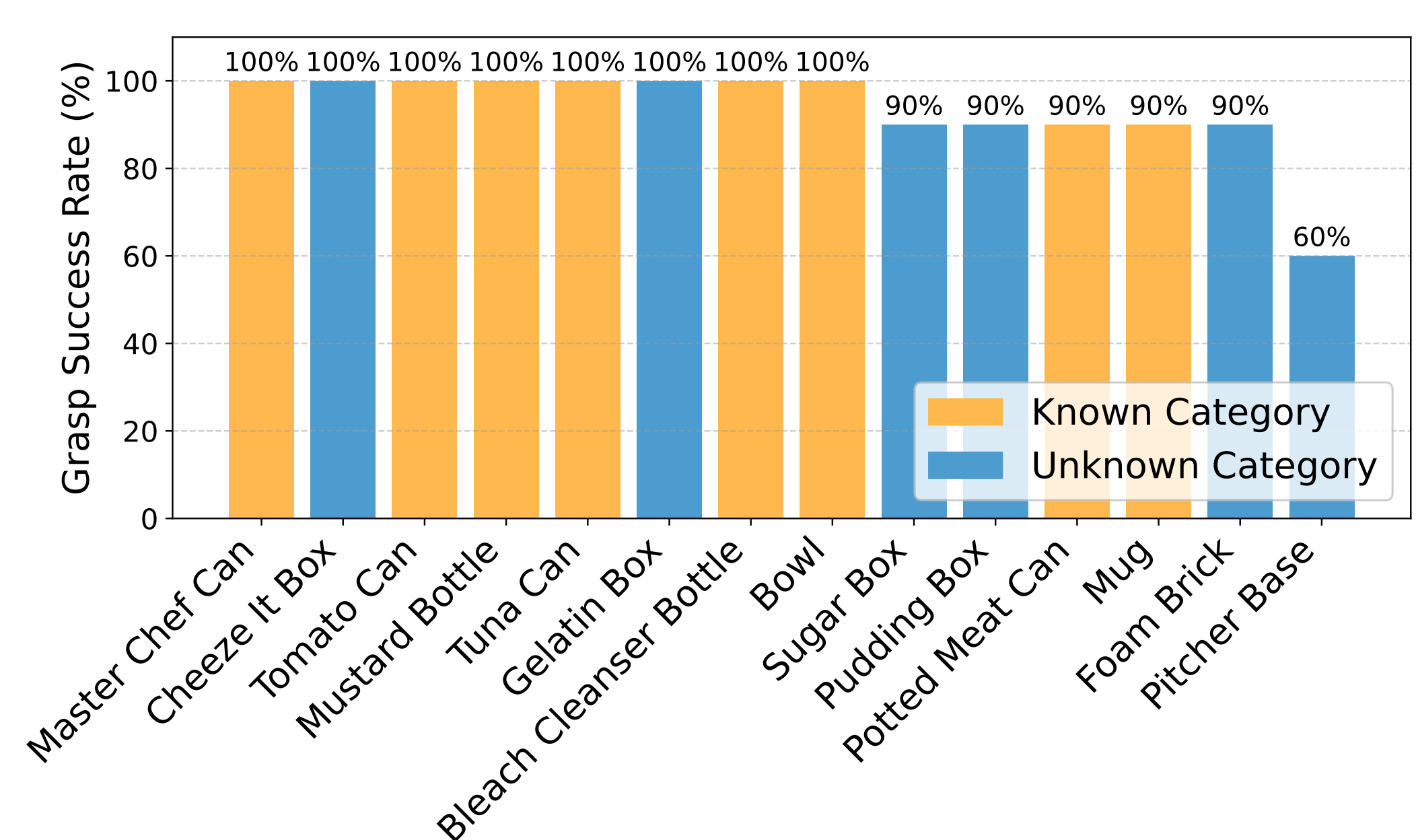
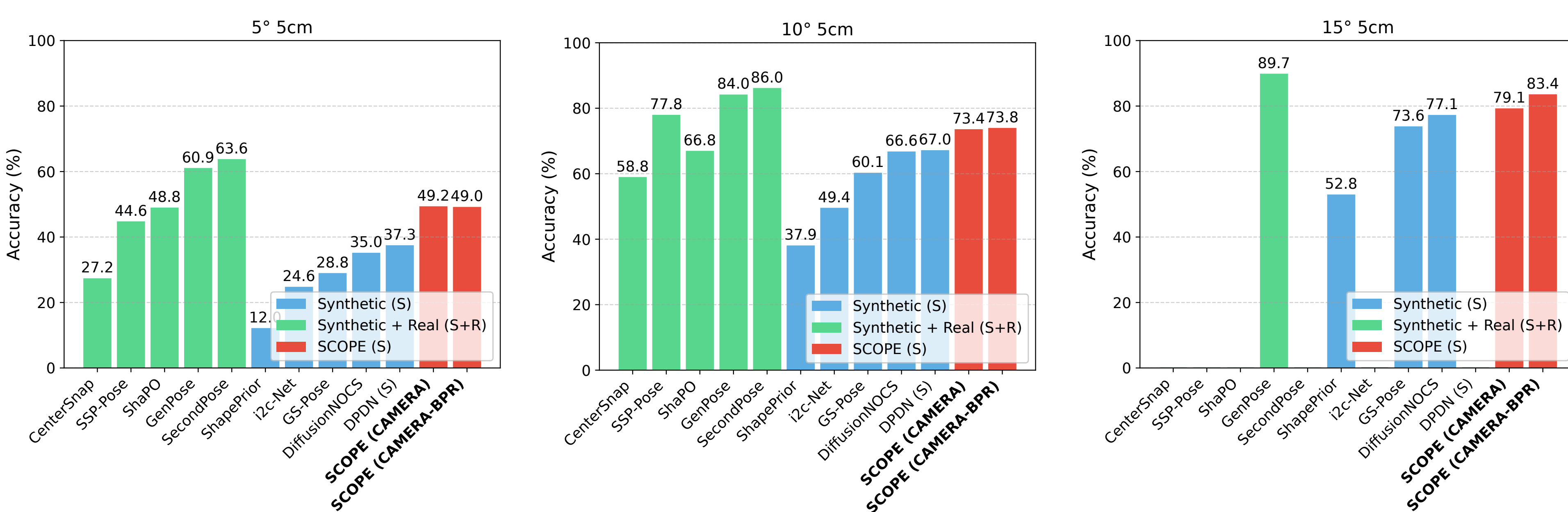
NOCS pointcloud



Results

Evaluation REAL275

Grasping (Un)Known Objects



Evaluation YCB-V

