



OCTOBER 1 - 5, 2023

IEEE/RSJ International Conference on Intelligent Robots and Systems

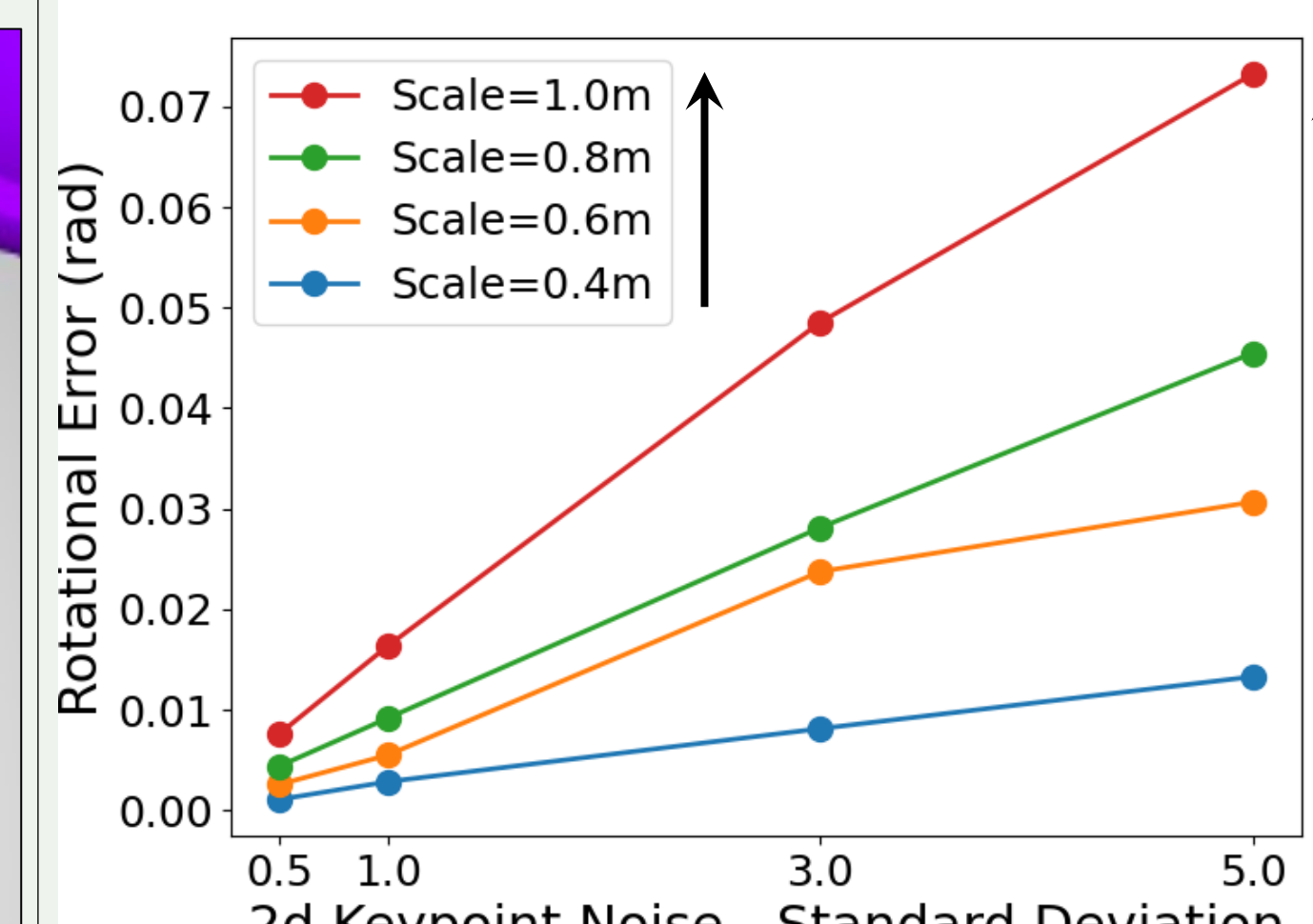
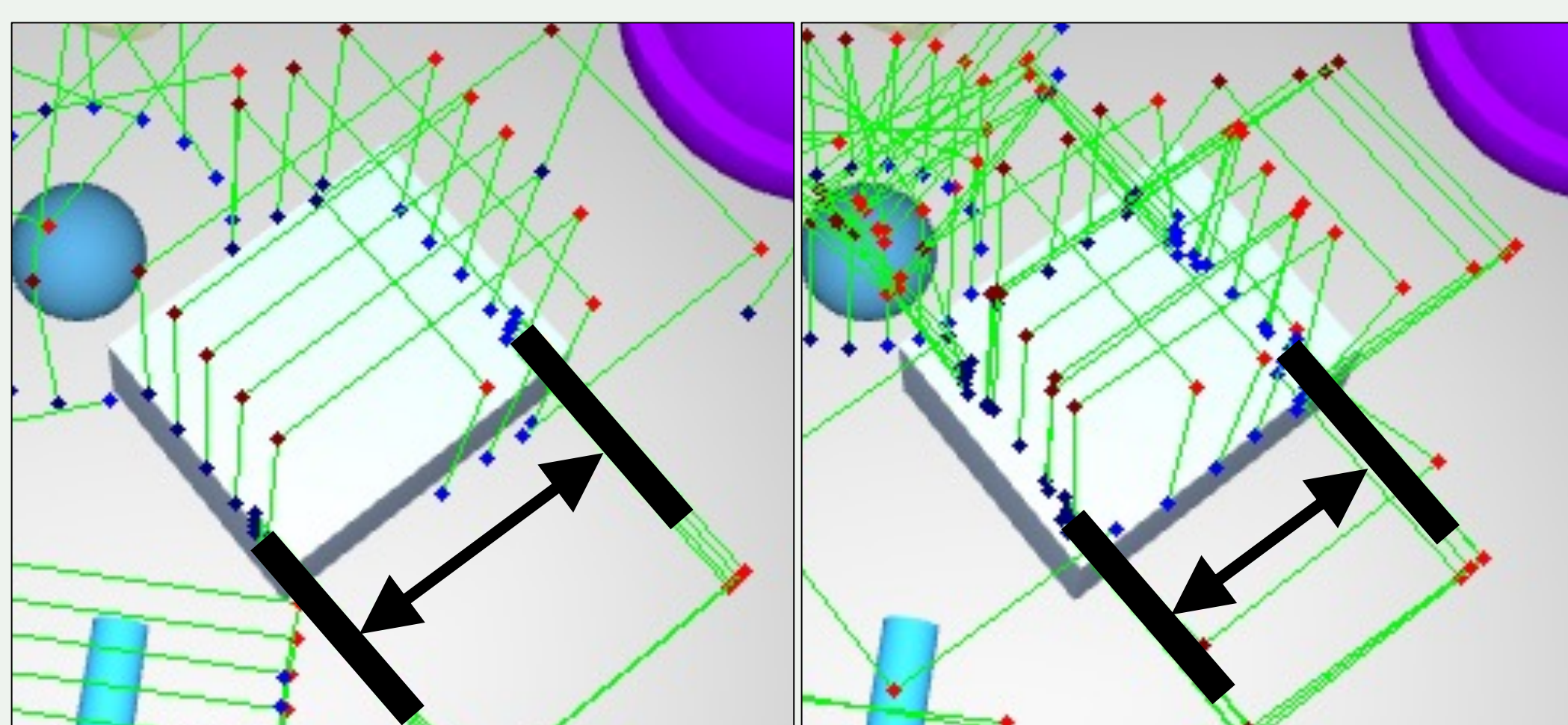
# KGNv2: Separating Scale and Pose Prediction for Keypoint-Based Grasp Synthesis on RGB-D Input

Yiye Chen, Ruinian Xu, Yunzhi Lin, Hongyi Chen, Patricio A. Vela

Institute for Robotics and Intelligent Machine, Georgia Tech

## Background & Motivation

- We aim to improve upon 2.5D-based 6-DoF Grasp Detection strategy, which is more efficient compared to point cloud methods [1]
- We found two drawbacks for the state-of-the-art method KGN [1]
  - Imprecise scale prediction due to inaccurate keypoint proximity (LEFT Figure)
  - Inconsistency in sensitivity to noise with different keypoint proximity (RIGHT Figure)
- Accordingly, two solutions are designed:
  - Scale prediction branch (*sBranch*)
  - Scale-normalized keypoint offset prediction (*sKpt*)



Increased pose error with larger scale

[1] Keypoint-GraspNet: Keypoint-based 6-DoF Grasp Generation from the Monocular RGB-D input. Chen, Yiye, et al. ICRA, 2023.

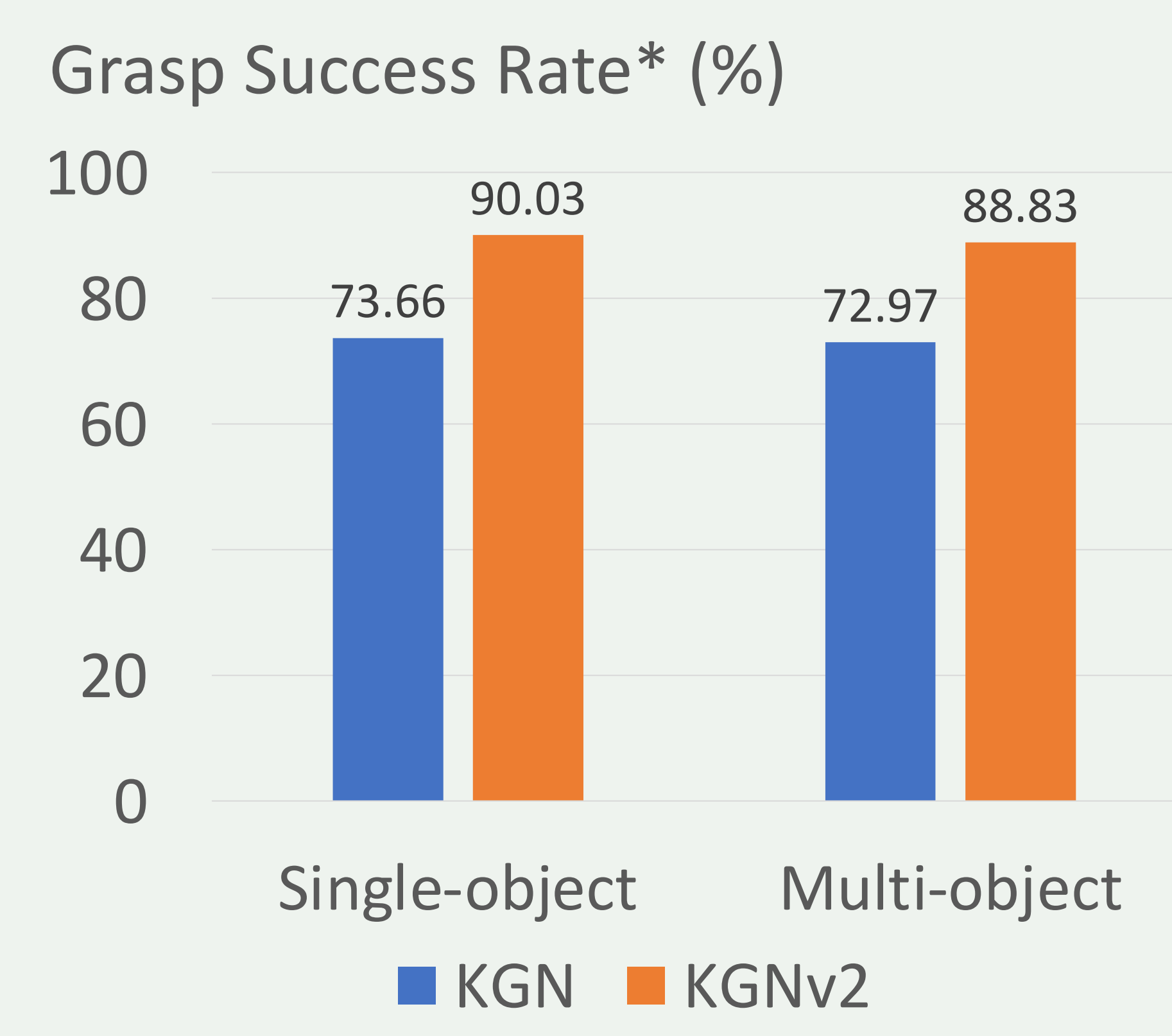
## Methodology

Scale-normalized offsets:  $O = \tilde{O}/S$

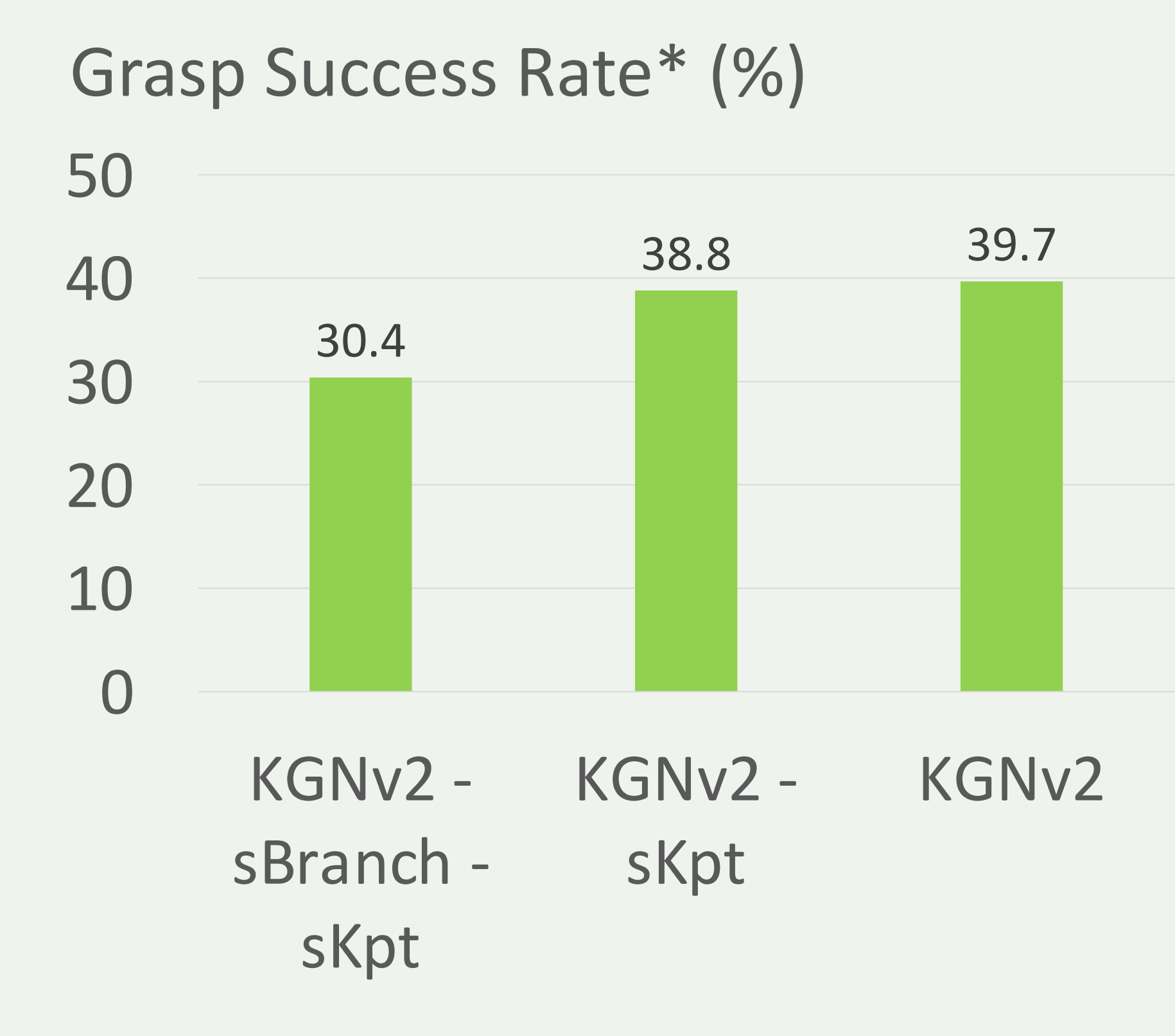
Keypoints (●) as center (●) + offsets (→):  $P_i = C + \tilde{O}_i, i = 1, 2, 3, 4$

- 6-DoF grasp poses are detected by first predicting the gripper keypoints on the image, and then recovering the pose using PnP algorithm
- The pose scale is regressed with a separate network branch, which is not subject to keypoint proximity prediction error.
- Keypoint coordinates are normalized by the pose scale to balance the sensitivity to noise.

## Vision Dataset Experiments



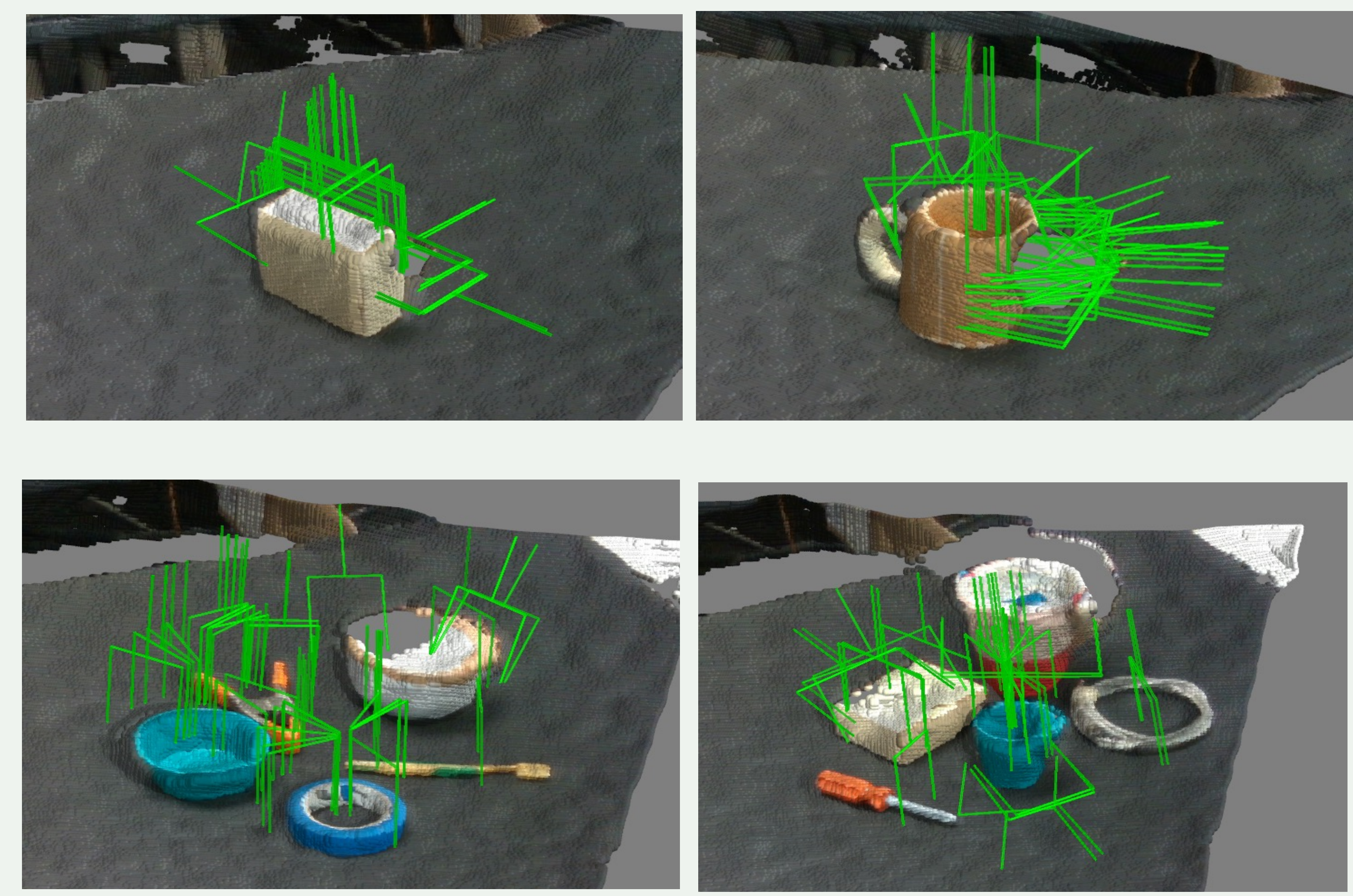
**Efficacy of KGNv2**  
(Trained and evaluated both on single- or multi-object data)



**Ablation Study**  
(Trained on single-object and evaluate on multi-object data)

- Our method outperforms the baseline by a large margin under both experiment settings (LEFT)
- Ablation study shows that both modifications contribute to the improvements (RIGHT)

## Physical Experiments



92.5% success rate on single-object grasping

80% success rate on multi-object grasping

\*Averaged across three error tolerance levels.