GRASP3R: Unposed and Uncalibrated RGB Robotic Object Grasping

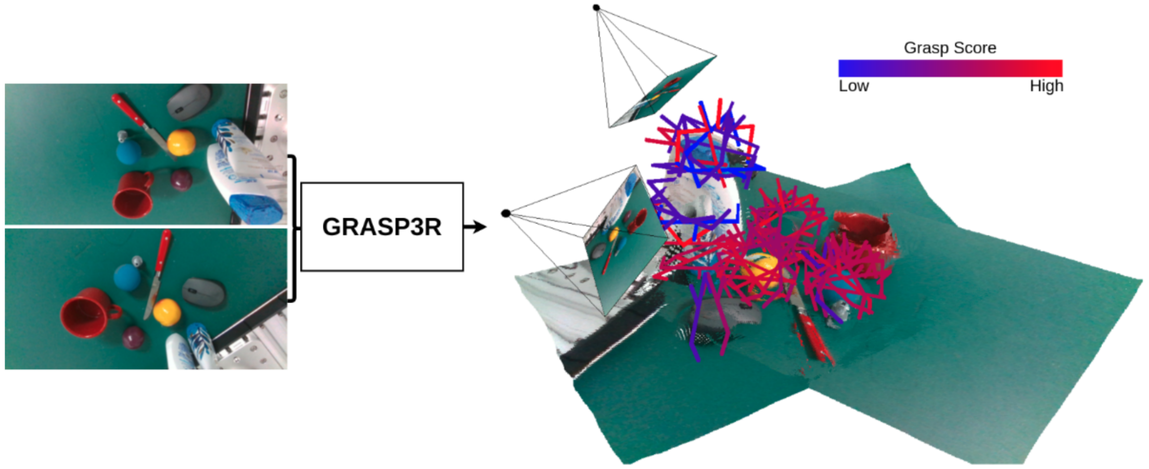
Loughborough University

Aston University
BIRMINGHAM UK

AURRIGO

Overview

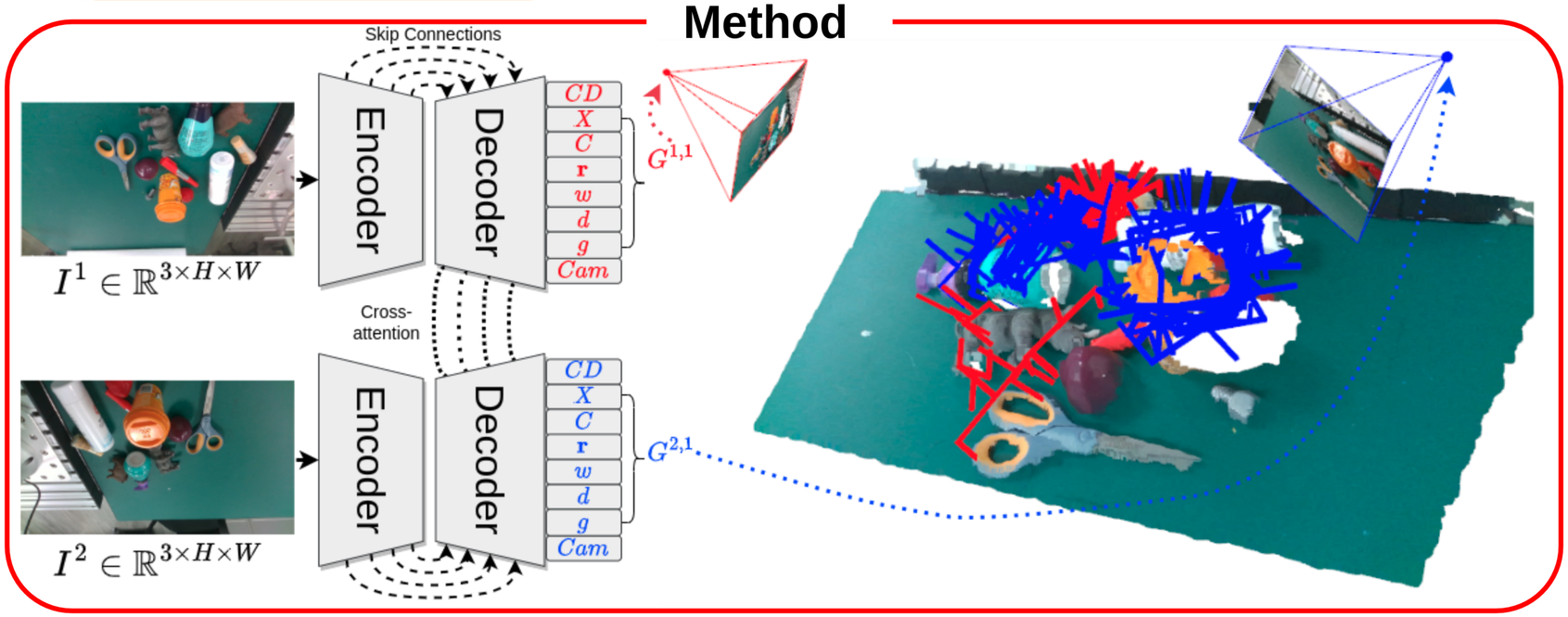
- **Challenge:** Depth-based grasping systems are costly and struggle with transparent or reflective objects. They usually require camera-specific calibration and retraining, making real-world use difficult.
- **Solution:** An end-to-end vision method that predicts dense 7-DoF grasp poses from just two uncalibrated RGB images.
- **Key Innovation: GRASP3R** fuses 3D scene reconstruction with grasp prediction. By learning geometry from CAD models, it enables effective collision avoidance without the need for a depth sensor.



GRASP3R: Seeing is Grasping

7-DoF Pose from Uncalibrated RGB Only

Method



Results

- **New State-of-the-Art for RGB-Only Grasping:** On the GraspNet-1Billion benchmark, **GRASP3R** is competitive with state-of-the-art depth-based systems, while remaining depth-free (DF) and enabling collision detection (CD) through its pointmaps.

Method	DF	CD	Seen ↑	Similar ↑	Novel ↑	Average ↑	Params ↓
GraspNet [8]	✗	✗	27.56 / 29.88	26.11 / 27.84	10.55 / 11.51	21.40 / 23.08	1.0M
RGB Matters [12]	✗	✗	27.98 / 32.08	27.23 / 30.40	12.25 / 13.08	22.49 / 25.19	55.1M
GSNet [32]	✗	✓	67.12 / 63.50	54.81 / 49.18	24.31 / 19.78	48.75 / 44.15	∼15.3M†
SBGrasp [20]	✗	✓	63.83 / -	58.46 / -	24.63 / -	48.97 / -	1.6M
HGGD [38]	✗	✓	64.45 / 61.17	53.59 / 47.02	24.59 / 19.37	47.54 / 42.52	3.4M
EconomicGrasp [37]	✗	✗	68.21 / 62.59	61.19 / 51.73	25.48 / 19.54	51.63 / 44.62	15.75M
ZeroGrasp [15]	✗	✓	72.43 / -	65.45 / -	28.49 / -	55.46 / -	-
GRASP3R	✓	✓	46.22 / 43.28	29.29 / 27.86	9.32 / 8.33	28.28 / 26.49	32.1M

- **Real-Time Capability:** The model operates at **5 FPS** at test time, making it suitable for real-world robotics applications.

Conclusion

- **GRASP3R** is the first method to demonstrate high-quality, collision-aware 7-DoF grasp prediction on GraspNet-1Billion using only unposed, uncalibrated RGB images.
- This work paves the way for more accessible, robust, and cost-effective robotic systems that can be deployed "in the wild" with standard cameras.
- **Future Work: GRASP3R** opens promising avenues for scaling to larger datasets and more diverse multi-image scenarios to further enhance robotic grasping capabilities.