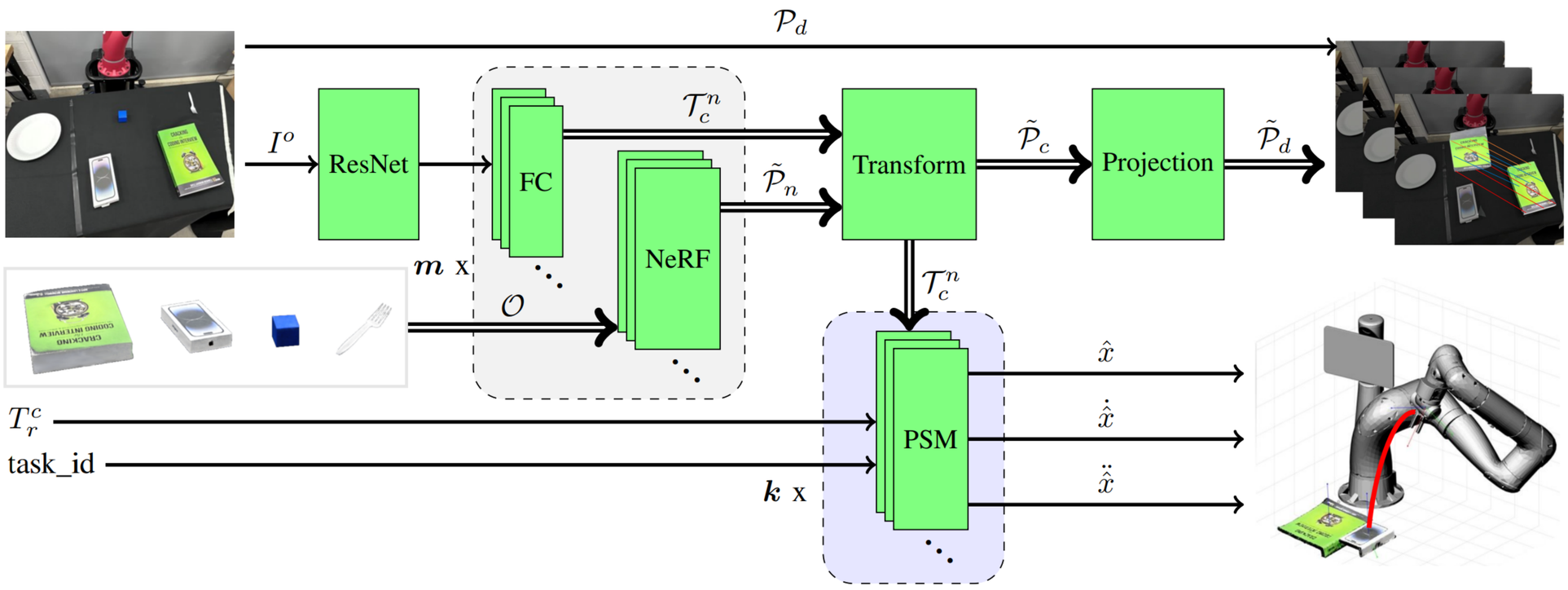
Self-Supervised Visual Motor Skills via Neural Radiance Fields

Paul Gesel, Noushad Sojib, and Momotaz Begum
paul.gesel@unh.edu, noushad.sojib@unh.edu, mbegum@cs.unh.edu

University of New Hampshire

CARL Cognitive Assistive Robotics Lab

Network Architecture

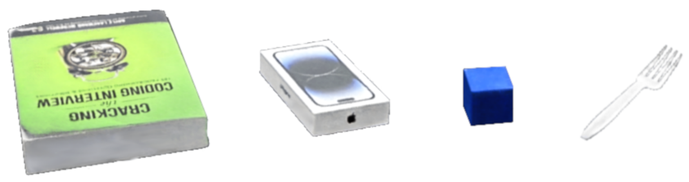


Objectives

- Visual imitation learning that is robust to significant clutter, occlusions, lighting conditions changes, and spatial variations in goal configurations
- Capable of learning from a single demonstration
- Stable trajectory prediction with respect to the robot's end effector

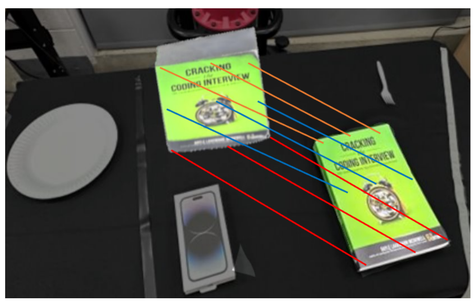
Neural Object Models

- Use neural networks to model objects
- Fast and accurate models learned via Instant-NGP
- Use object models to extract labeled training data

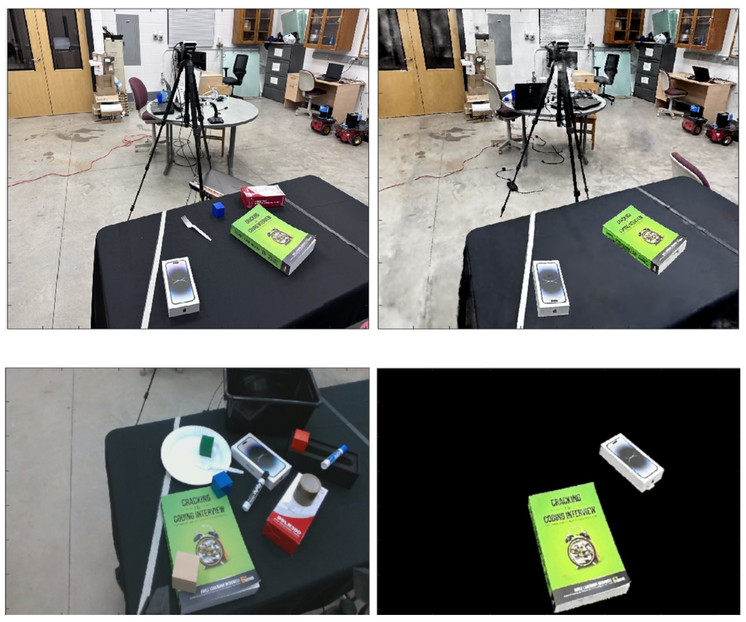


Robust Keypoint Correspondence

- Estimate object rigid body transform via keypoint correspondence
- Robust matching via transformer-based LoFTR network



- Discover object rigid body transforms via self-supervised loss



Goal Conditioned DS-Policy

- PSM is a dynamic system-based trajectory learning method that models motion with a time-invariant differential equation
- Network output constrained to produce stable motions with respect to the robot's end effector
- Goal transform and velocity profile learned from demonstration

Experimental Results

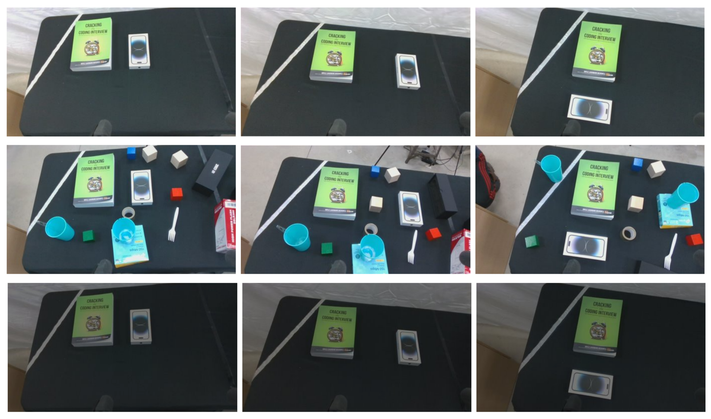


Figure 1: first row: sample training images for object stacking task, second row: test images with added clutter, third row: test images with artificial lighting variations

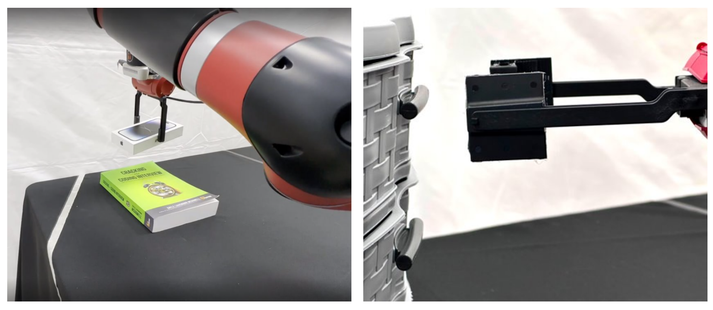


Figure 2: Object stacking task (left) and drawer opening task (right)

Conclusion

- We propose a novel network architecture for visual imitation learning that takes advantage of recent advances in image-to-image correspondence and NeRFs
- In combination with a DS-based output layer, the learned visual motor policy becomes robust to clutter, occlusions, lighting conditions, and spatial variations in goal configurations

Acknowledgements

Research support
IIS 1830597

NSF