Scene Understanding using Part-Based Object Affordances

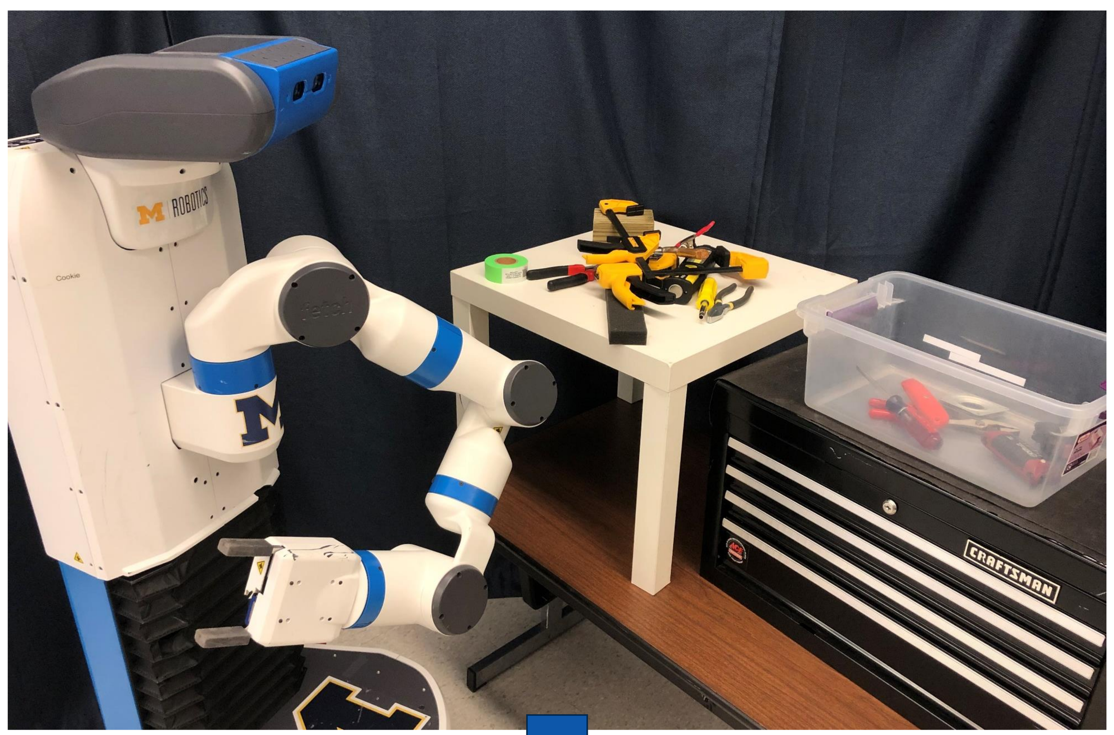
Jana Pavlasek, Karthik Desingh, Odest Chadwicke Jenkins

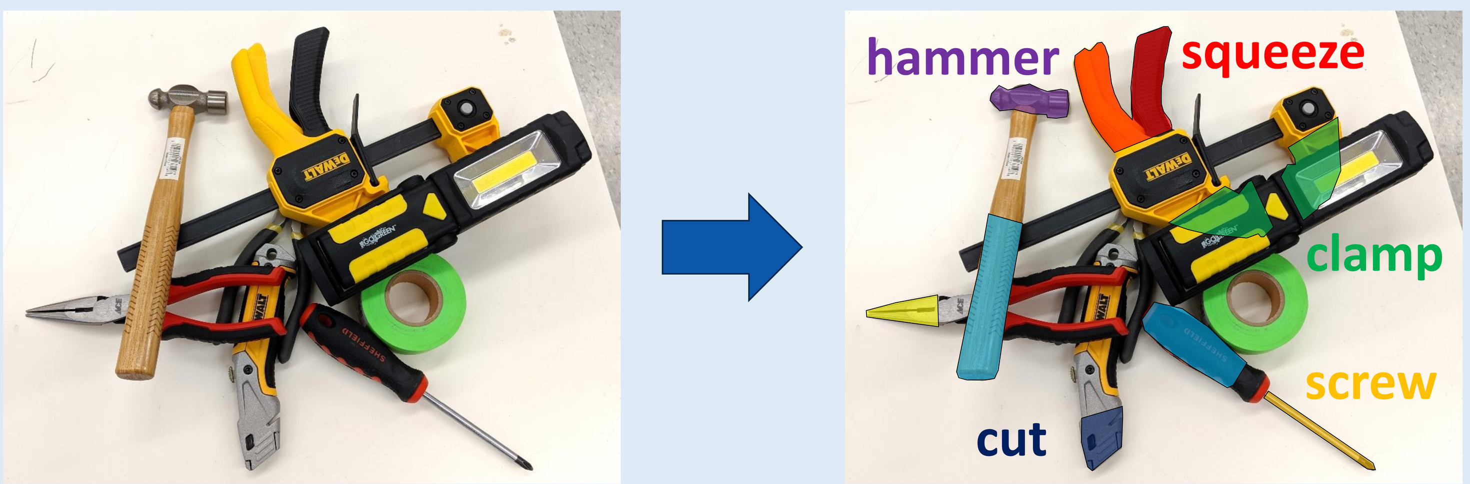
Laboratory for Progress

Robotics Institute, University of Michigan, Ann Arbor, USA

Perceiving a scene as a collection of actions available to the robot is more effective for task execution.

Introduction





Robot observation | Affordances in the scene

The ability to use tools is critical to enable robots to perform assembly, repairs and maintenance of equipment in remote locations. To effectively use tools, a robot needs to identify and localize articulated objects in unstructured, cluttered environments. Rather than representing the scene in terms of object classes, it can be represented as a collection of *object affordances* [1], or actions available to the robot. Since affordances tend to be associated with object parts, localizing objects by parts is useful. A parts-based method is robust to occlusions because the relationships between parts can be used to infer the location of occluded parts. We propose a parts-based pose estimation method which combines domain knowledge about the object model with a deep learned detector using generative inference.

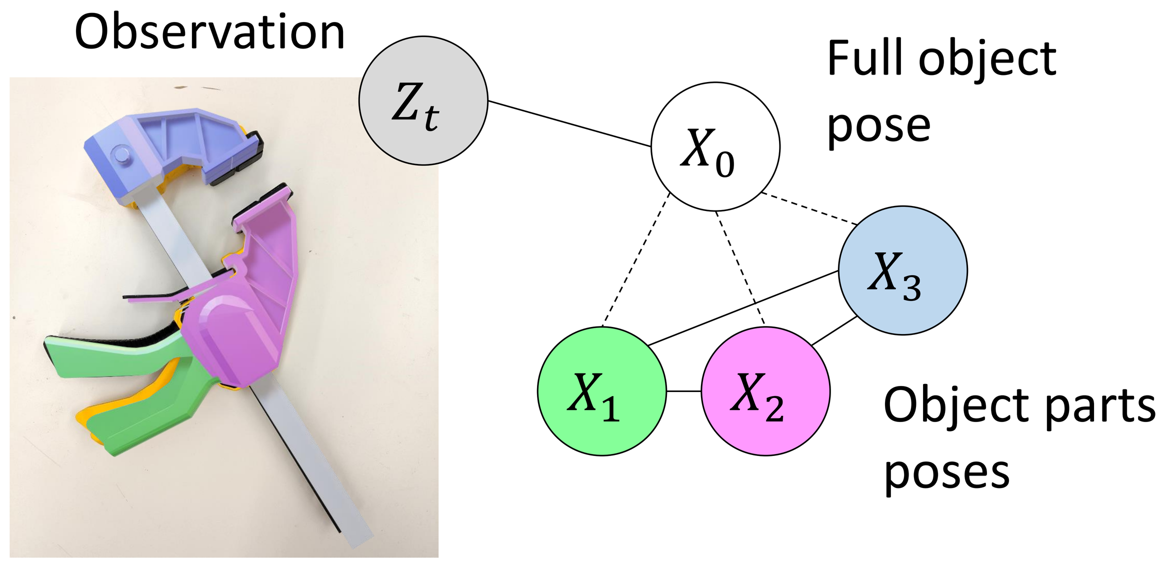
Object Localization in Clutter

Detection is robust when objects are isolated. | Under occlusion, detections might fail, but some parts can be detected. | Parts-based generative inference can correct failures.

Parts-Based Object Localization

Objects are modelled as a Markov Random Field (MRF) where each node is the pose of an object part. We maintain a belief over object pose as a set of particles and perform inference using belief propagation [2]. We assume mesh models and part models are given, as well as part affordances, as affordance templates [3].

Parts based model as an MRF:



$$p(X,Z) = \prod_{(i,j)\in \text{edges}} \underbrace{\boxed{\psi_{i,j}(X_i,X_j)}}_{\text{Pairwise}} \prod_{i\in \text{nodes}} \underbrace{\boxed{\phi_i(X_i,Z_t)}}_{\text{Unary}}$$

Unary potential: Learned model. Describes correspondence of the observation to a given hypothesis

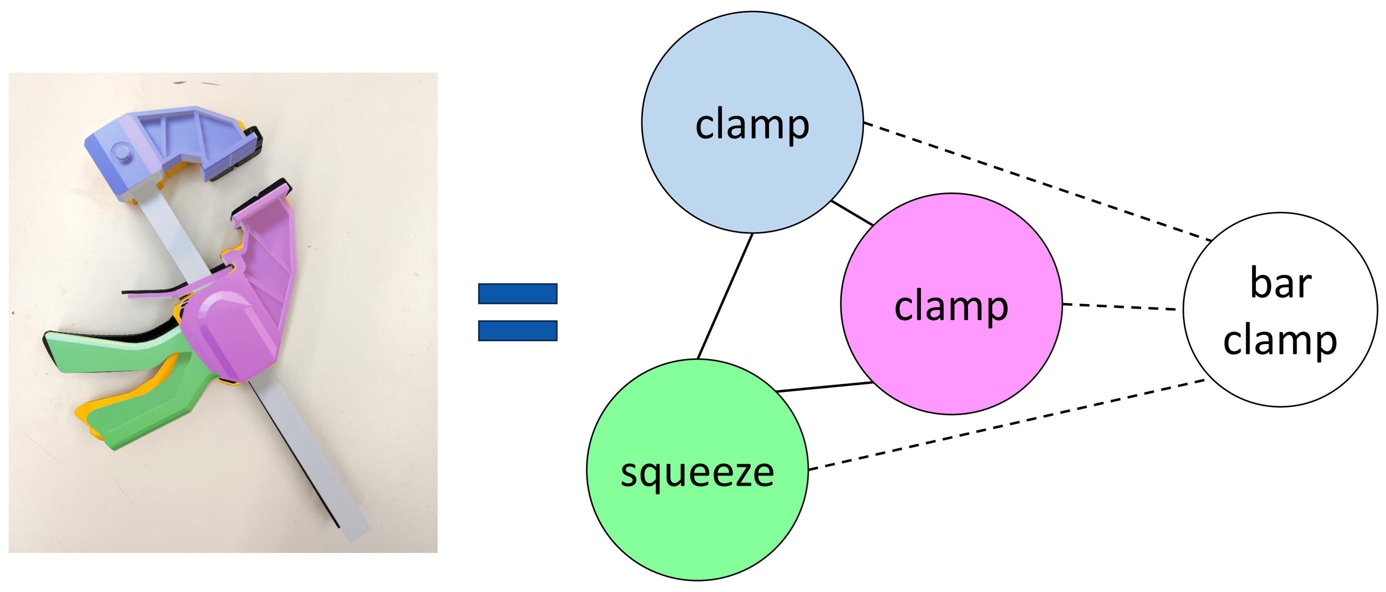
Pairwise potential: Describes agreement of the hypothesis to the geometric parts-based model.

Robot Observation | **Unary** $\phi_i(X_i,Z_t)$ | **Pairwise** $\psi_{i,j}(X_i,X_j)$

Localized object pose in 3D

Future Work

To loosen our reliance on mesh models, we can directly detect the affordances in the scene, as in [4,5]. The objects can be represented in terms of their affordances and geometric constraints, creating a generalizable object representation.



References

[1] J. J. Gibson, "The theory of affordances," in Perceiving, acting and knowing: toward an ecological psychology. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, 1977, pp. 67–82.
[2] K. Desingh, S. Lu, A. Opipari, and O. C. Jenkins, "Efficient nonparametric belief propagation for pose estimation and manipulation of articulated objects," Science Robotics, vol. 4, no. 30, 2019.
[3] S. Hart, P. Dinh, and K. Hambuchen. "The affordance template ROS package for robot task programming," in 2015 IEEE International Conference on Robotics and Automation , pages 6227–6234. IEEE, May 2015
[4] M. Kokic, J. A. Stork, J. A. Haustein, and D. Kragic, "Affordance detection for task-specific grasping using deep learning," in 2017 IEEE-RAS 17th International Conference on Humanoid Robotics.
[5] T.-T. Do, A. Nguyen, and I. Reid, "AffordanceNet: An end-to-end deep learning approach for object affordance detection," in 2018 IEEE International Conference on Robotics and Automation.