Where2Act: From Pixels to Actions for Articulated 3D Objects Summary

Kaichun Mo, Leonidas Guibas, Mustafa Mukadam,
Abhinav Gupta, Shubham Tulsiani

1 Motivation

Robots need a perception system that can understand affordances, identify where certain actions affect the environment, and predict what that effect will be. Humans can innately understand these concepts and bringing these capabilities to robots would be invaluable for task planning. Existing methods either predict what action can be done, but now how it affects the environment or vice versa. In addition, passive observations of other agents interacting with the world can be useful, but may not provide the data distribution needed compared to on-policy exploration.

2 Approach

The authors train three different networks. The actionability scoring module generates a heatmap to indicate which pixels are likely actionable and could be moved (though not how they could be moved). The action proposal module is a generative network that predicts an action to execute at each pixel. Finally, the action scoring module estimates how likely this action with these parameters is to if executed at this pixel. To train these modules, they first sample random interactions to give 10,000 positive trajectories where an interaction successfully moved the object. They then intelligently sample around previous positive trajectories to refine their estimates.

3 Results

The highlighted interaction points on objects seem to correspond to common sense interaction points for pushing and pulling. This is reflected in their high F-scores and accuracy in their quantitative results. However, their success rate for the final action is fairly low across the board. This is somewhat understandable given the complicated nature of the task, but also indicates a failure in their action proposal and action scoring module or their sample and reject methodology. They also show qualitative results on real-world objects that look promising.

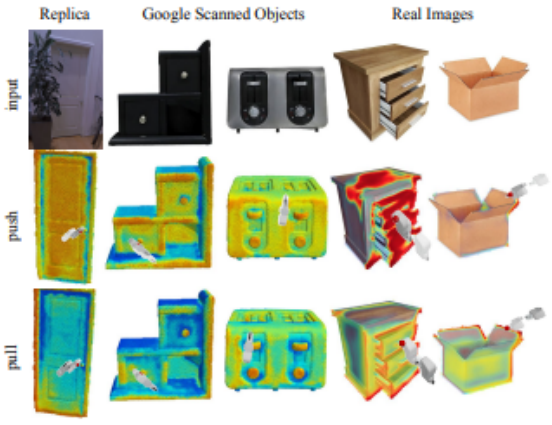

Figure 1: Visualized results over various real-world objects