# Statement of Purpose

Haolun “Harry” Zhang

My research interests lie at the intersection of computer vision, deep learning, control theory, and their applications to the field of ***robot learning***. I aspire to develop algorithms and representations to help robots to better understand and interact with the world. I believe that a good visual representation is an indispensable step to help robots exhibit flexible and generalizable behaviors in downstream manipulation tasks across environments. I have been deeply interested in robotics since I took my first control theory course with the late Professor Andrew Packard, and relentlessly pursued that interest through 2 years of undergraduate research at Berkeley AI Research under Professor Ken Goldberg and Dr. Jeff Ichnowski (currently CMU faculty), and then 2 years of robotics research as an MS student at CMU in Professor David Held's lab.

My first major project in Berkeley AI Research was building a system called *Dex-Net AR* that trained robots to grasp objects from point cloud data scanned from iPhones. Instead of relying on high-resolution input depth images from an expensive depth camera, I rearchitected the deep grasp planning system, Dex-Net, to make predictions exclusively based on low-cost 3D point cloud data collected by an iPhone’s camera. I found that 3D point cloud data can be projected into depth images from arbitrary viewpoints, revealing more geometric information about the objects of interest compared to traditional top-down depth images, and helping the system to find better grasps. Moreover, with the ability to train Dex-Net using in-the-wild data collected by iPhone users, we were able to continually improve the grasp planner’s performance. Our paper on this project was accepted to **IEEE International Conference on Robotics and Automation (ICRA) 2020**. I also contributed to two other projects related to learning objects’ 3D properties such as rotation prediction and graspability prediction, which were accepted to **IEEE Conference on Automation Science and Engineering (CASE) 2020 and 2021**, respectively.

My next major project under Prof. Goldberg was on dynamic deformable object manipulation. We focused on teaching robots to dynamically control and manipulate a cable to accomplish interesting tasks such as vaulting over obstacles, knocking target objects off a base, weaving between obstacles, and jump-rope. In this project, I designed a novel algorithm to control the rope dynamics efficiently using a 3D vector of the robot’s joint angles. The key takeaway is that by using a novel and compact parameterization of the robot’s arcing trajectory, we could learn the rope’s dynamic motion efficiently, in a much smaller state space. The paper was accepted to **IEEE International Conference on Robotics and Automation (ICRA) 2021** and has contributed to a community-wide adoption of our parameterization in the field of deformable object manipulation, with over 20 citations within a year. For the work I have done on this project, I was awarded the *Warren Y. Dere Design Award* at Berkeley. The project was also featured in Bay Area Robotics Symposium, an RSS Workshop, and an ICRA Workshop.

As a Master's student in Robotics at the CMU Robotics Institute, I continued to conduct robot learning research under the supervision of Professor David Held, where I focused my efforts on perception for robot learning. The first project at CMU that I co-led focused on manipulating *articulated objects* (objects with 1-DoF such as doors, drawers, toilet lids, etc. that are common in households). The key idea is to learn a general visual representation of articulated objects and build a learning system that can generalize to novel objects. We proposed a vision-based system, *FlowBot3D*, that learns to predict the potential motions of the parts of a variety of articulated objects to guide downstream motion planning of the system to articulate the objects. To predict the object motions, I trained a PointNet++ model to output a dense vector field (*3D articulated flows*) representing the point-wise motion direction of the points in the point cloud under articulation. I then developed an analytical motion planner based on this vector field to achieve a policy that actuates the object with theoretical guarantees that it achieves maximum articulation. A single FlowBot 3D model was trained entirely in simulation across all categories of objects and generalized well to real-world unseen objects and novel categories without any fine-tuning. The takeaway here is that 3D articulated flows are a very elegant, generalizable representation of articulated objects' motions, and learning 3D articulated flows helps robots to generalize to unseen objects well. This project was accepted to **Robotics: Science and System (RSS) 2022 and was a Best Paper Finalist (1.5% selection rate).** FlowBot3D also garnered media attention; for example, MIT Tech Review China featured our paper in June 2022.

The next project I co-led at CMU also focused on generalizable 3D representations for robot manipulation policies, but instead of focusing on highly-constrained articulated objects, we focused on *free-floating objects*. We conjectured that the task-specific pose relationship between relevant parts of interacting objects is a generalizable notion of a manipulation task that can transfer to new objects in the same category. For example, the relationship between the pose of a lasagna relative to an oven or the pose of a mug relative to a mug rack. We called this task-specific pose relationship "cross-pose" and provided a mathematical definition of this concept. We proposed a vision-based system, *TAX-Pose*, that learns to estimate the cross-pose between two objects for a given manipulation task. The estimated cross-pose is then used to guide a downstream motion planner to manipulate the objects into the desired pose relationship (e.g., placing the lasagna into the oven or hanging a mug on the mug rack). My part also focused on devising and improving the TAX-Pose network architecture with Transformers and residual corrections so that the network can cross-attend to different parts of the objects and output correct correspondences. I have found that learning this fundamental 3D relation between objects yields a robust, generalizable robot manipulation policy in the real world. This paper was accepted to **Conference on Robot Learning (CoRL) 2022** and was considered "the strongest paper from the lab yet" by my advisor. With generalizable 3D representations for both articulated and free-floating objects, we are able to accomplish a variety of challenging manipulation tasks involving these objects, and I am currently working on a unified method that learns the two representations simultaneously.

During the summer of 2022, I worked on 3D vision projects at Amazon as an Applied Research Scientist for their new physical fashion stores. I created a 3D animatable virtual try-on system that synthesizes a customer's image and a catalog outfit image using StyleGAN, generates a 3D mesh from the synthesized image using a learned implicit function, and animates the 3D mesh given an input sequence. This work is currently under review at **CVPR 2023.**