

## Visual Imitation Made Easy

Interfaces for visual imitation require kinesthetic teaching or teleoperation which restrict large-scale data collection. Efficient collection can be facilitated using an economical and easy-to-use interface which is compatible as a robot end-effector and requires minimal setup requirements. To this end, the work presents the use of assistive tools for large-scale data collection which facilitate visual imitation. A conventional reacher-grabber tool is used to collect and learn trajectories of visual tasks using Structure from Motion (SfM) in collaboration with finger detector network. Behavioral cloning is employed to learn data augmentations collected from pixel inputs consisting of different objects and background settings. The presented setup demonstrates improved setup of 87% and 62% on pushing and stacking tasks respectively.

A suitable interface is essential for collecting real-world data and provisioning learning of robots from visual inputs. To address this challenge of large-scale data collection and learning, assistive tools are presented as suitable candidates for learning setups. Specifically, a reacher-grabber tool which is readily available, economical and compatible with robots as an end-effector is used to collect visual data. Visual inputs consist of over 1000 human trajectories captured over different objects, background conditions and lighting settings. Following demonstrations, object poses are recovered using SfM reconstruction. To extract finger locations of gripper, a finger detection network is trained on human-labeled frames. The complete framework is trained using behavioral cloning which additionally consists of augmented data learnt using an AlexNet-type architecture yielding 3D translation and 6D rotation vectors respectively.

The proposed DemoAT framework demonstrates improved convergence as a result of data augmentation. Furthermore, the overall efficacy of the setup is found suitable in comparison to behavioral cloning as a result of increased success rates on pushing and stacking tasks. While the setup is found suitable with large-scale augmentations, it presents several caveats. Firstly, the scheme dedicates its efforts to large-scale data collection which is contrary to common challenges of learning from less data. In most practical scenarios robots have access to limited data which requires an efficient learning algorithm. This is found lacking in the proposed framework. Secondly, suitability of the setup is not compared to other end-effectors such as low-cost plastic grippers or rubber surfaces. These alternatives are durable, readily-available and may be suitable for contact-rich tasks consisting of low friction between surfaces. Lastly, comparison of the proposed scheme is carried with behavioral cloning in the presence of data augmentation. The evaluation lacks other state-of-the-art visual imitation methods which depict improved results on small as well as large datasets.

The DemoAT framework presents a suitable scheme for large-scale data collection and training visual imitation learning algorithms using assistive tools. This presents a novel direction for the usage of assistive tools in robot control which can be further extended to low friction tasks and state-of-the-art visual imitation learning algorithms.

Usage of kinesthetic methods and teleoperation hinder large-scale data collection for visual imitation learning. To tackle this challenge, the work presents DemoAT framework which proposes the utilization of assistive reacher-grabber tool as a suitable end-effector for collecting and learning human demonstrations. The framework consists of pose and finger location extraction using SfM followed by training using behavioral cloning consisting of data augmentation. In comparison to conventional cloning, data augmentation improves convergence and success rate of the setup.