highly dexterous manipulation capabilities that could not be simply programmed or teleoperated by humans [2, 17, 31]. However, these approaches are often tailored to a single manipulation skill, limiting their broad applicability.

What prevents RL from being more generally applicable to vision-based dexterous manipulation? We first investigate this by identifying the inherent properties of dexterous manipulation that differentiate this application domain from others. Then, we examine how these properties contribute to challenges in applying RL algorithms and develop a collection of novel techniques to address the challenges. Putting together our experiences and techniques, we outline a recipe for applying sim-to-real RL to vision-based humanoid manipulation tasks and show promising results. Below, we articulate the key challenges and our strategies to tackle them.

Challenge in environment modeling. The first challenge in applying RL to dexterous manipulation lies in the difficulty (or impossibility) of matching a simulated environment with the real environment. While real-world RL circumvents this problem, training with physical hardware is highly demanding regarding hardware quality, maintenance support, controller robustness, and safety. With a system as high-dimensional as a humanoid with multi-fingered hands, real-world exploration becomes even less tractable. In contrast, simulations offer unlimited chances of trial and error in a virtual sandbox, motivating the development of sim-to-real RL approaches. While previous successes in RL-based locomotion [16, 21, 26, 43] are encouraging, we observe that previous successes in dexterous manipulation involve much more laborious real-to-sim engineering efforts that are task-specific or hardware-specific [2, 17, 31]. To better align simulation with the real world, we propose an automated real-to-sim tuning module that substantially reduces the engineering effort required for the environment modeling gap.

Challenge in reward design. While the reward function is commonly used as a general interface for specifying a task to train RL policies, it is notoriously hard to design generalizable rewards for manipulation tasks, especially for those that are contact-rich or long-horizon. Prior work often resorts to hand engineering based on the knowledge of human experts [40, 61], which has limited scalability in the long run. This challenge differentiates manipulation from locomotion, where many tasks of interest can be specified with variations of the reward for a single "walking" task. We propose a general principle to design rewards for dexterous manipulation tasks: disentangle a full task into intermediate "contact goals" and "object goals". We use a novel keypoint-based state representation to specify contact goals. Following our reward design techniques, a task as long-horizon and contact-rich as bimanual handover can be learned with RL *tabula rasa*.

Challenge in policy learning. A well-defined reward function does not guarantee the successful learning of RL policies due to the sample complexity and reward sparsity of exploring in a high-dimensional space. The variety and complexity of contact patterns in dexterous manipulation with multi-fingered hands further exacerbate the problem. Although unsupervised methods [7, 30, 39] have been proposed to encourage exploration by favoring novel state visitations, they do not fundamentally reduce the difficulty of hard-exploration problems. We tackle this challenge by introducing two practical techniques: (1) initializing tasks with task-aware hand poses; (2) breaking down hard exploration problems into sub-tasks with much-reduced dimensionality, training expert policies on the sub-tasks, then distilling them into a generalist policy for the full task. We experimentally verify that these techniques improve sample efficiency of learning and study how different divide-and-conquer schemes vary in effectiveness.

Challenge in object perception. Compared to other robotic tasks, object perception is particularly important for manipulation because the task is inevitably coupled with interaction with objects. Object perception is a long-standing challenge because the variety of objects is uncountable in shapes, sizes, masses, colors, textures, and many other properties. Research in applying sim-to-real RL to dexterous manipulation is bottlenecked by this dilemma — while object representations that are more expressive and information-dense can improve dexterity and capability of the learned policy, they also present a larger sim-to-real gap. To overcome this challenge, we propose to use a mixture of low-dimensional and high-dimensional object representations, with modality-specific data augmentation on the high-dimensional features to reduce the sim-to-real perceptual gap. We systematically study how this combination could help achieve a good balance between learning dexterous manipulation policy and reliably transferring the policy onto real robot hardware.

The strategies we outline above form a complete recipe of sim-to-real RL for vision-based dexterous manipulation. We show successful results of learning a collection of three dexterous manipulation tasks on humanoids and conduct systematic ablation studies.

II. Background

A. Deep Reinforcement Learning Applications to Robotics

The successes of deep RL across a wide range of applications [1, 6, 14, 21, 24, 26, 48, 56] have sparked lots of excitement in recent years. However, works over the years have identified brittleness with this paradigm, most notably the sensitivity to hyperparameters [19] and questionable reproducibility [23] due to the high variance intrinsic to RL algorithms.

Among the open problems in RL, the most important and long-standing is exploration. In supervised learning, it is often assumed that data is given. In RL, however, agents need to collect their own data and update their policy based on the collected data. The problem of *how* data is collected is known as the exploration problem. Real-world robotics, with high-dimensional observations and dynamics and often sparse rewards, present a particularly challenging set of hard exploration problems for RL. While there have been works that algorithmically scale exploration to high-dimensional inputs by encouraging visitation to novel states [4, 7, 30, 38, 39, 50, 53],