

DEXTERITYGEN: Foundation Controller for Unprecedented Dexterity

Zhao-Heng Yin^{1,2}, Changhao Wang², Luis Pineda², Francois Hogan², Krishna Bodduluri², Akash Sharma²,
Patrick Lancaster², Ishita Prasad², Mrinal Kalakrishnan², Jitendra Malik², Mike Lambeta²,
Tingfan Wu², Pieter Abbeel¹, Mustafa Mukadam²

¹BAIR, UC Berkeley ²FAIR at Meta
zhaohengyin.github.io/dexteritygen

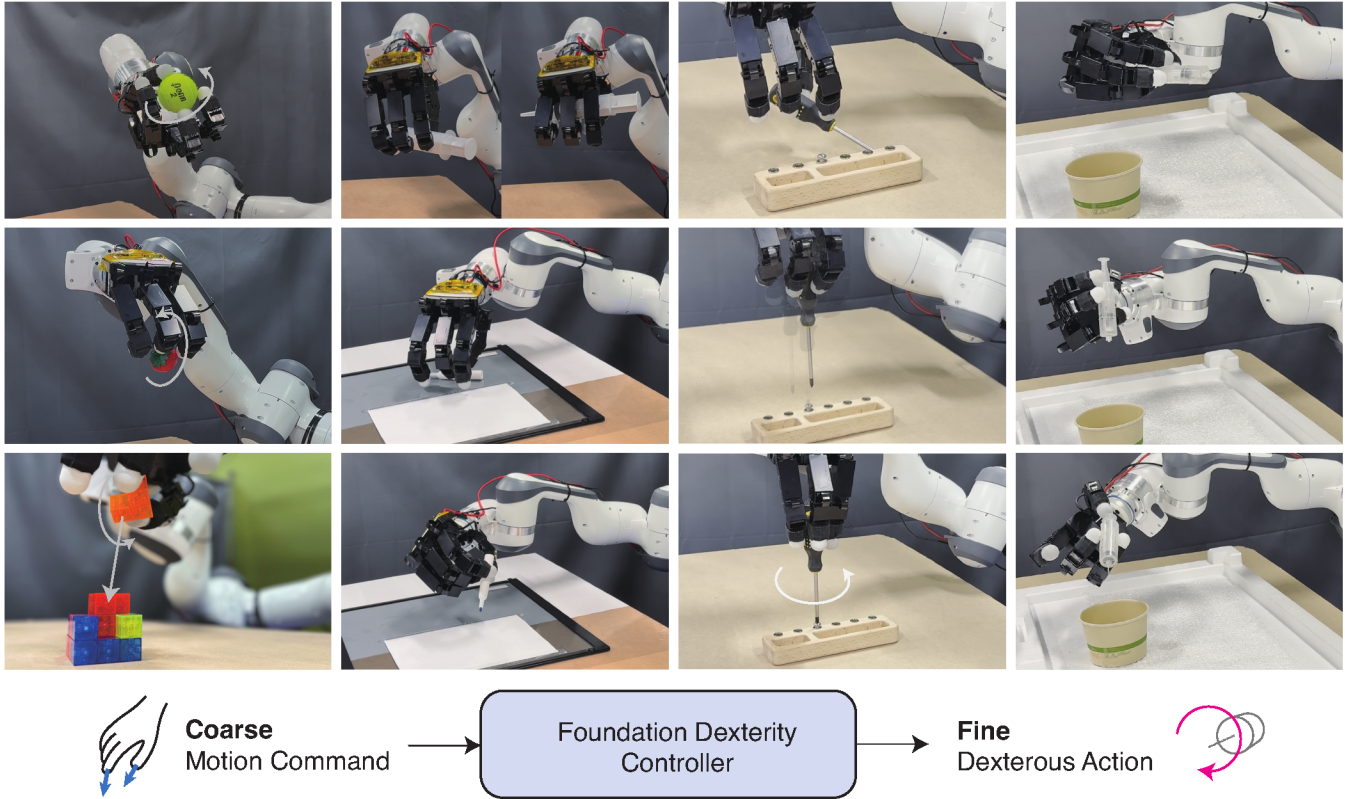


Fig. 1: We introduce DexterityGen (DexGen) as a foundation controller that achieves unprecedented dexterous manipulation behavior with teleoperation. DexGen is a generative model that can translate an unsafe, coarse motion command produced by external policy to safe and fine actions. With human teleoperation as a high-level policy, DexGen exhibits unprecedented dexterity from diverse object rotation and regrasping to using pen, syringe, and screwdriver.

Abstract—Teaching robots dexterous manipulation skills, such as tool use, presents a significant challenge. Current approaches can be broadly categorized into two strategies: human teleoperation (for imitation learning) and sim-to-real reinforcement learning. The first approach is difficult as it is hard for humans to produce safe and dexterous motions on a different embodiment without touch feedback. The second RL-based approach struggles with the domain gap and involves highly task-specific reward engineering on complex tasks. Our key insight is that RL is effective at learning low-level motion primitives, while humans excel at providing coarse motion commands for complex, long-horizon tasks. Therefore, the optimal solution might be a combination of both approaches. In this paper, we introduce DexterityGen (Dex-

Gen), which uses RL to pretrain large-scale dexterous motion primitives, such as in-hand rotation or translation. We then leverage this learned dataset to train a dexterous foundational controller. In the real world, we use human teleoperation as a prompt to the controller to produce highly dexterous behavior. We evaluate the effectiveness of DexGen in both simulation and real world, demonstrating that it is a general-purpose controller that can realize input dexterous manipulation commands and significantly improves stability by 10-100x measured as duration of holding objects across diverse tasks. Notably, with DexGen we demonstrate unprecedented dexterous skills including diverse object reorientation and dexterous tool use such as pen, syringe, and screwdriver for the first time.

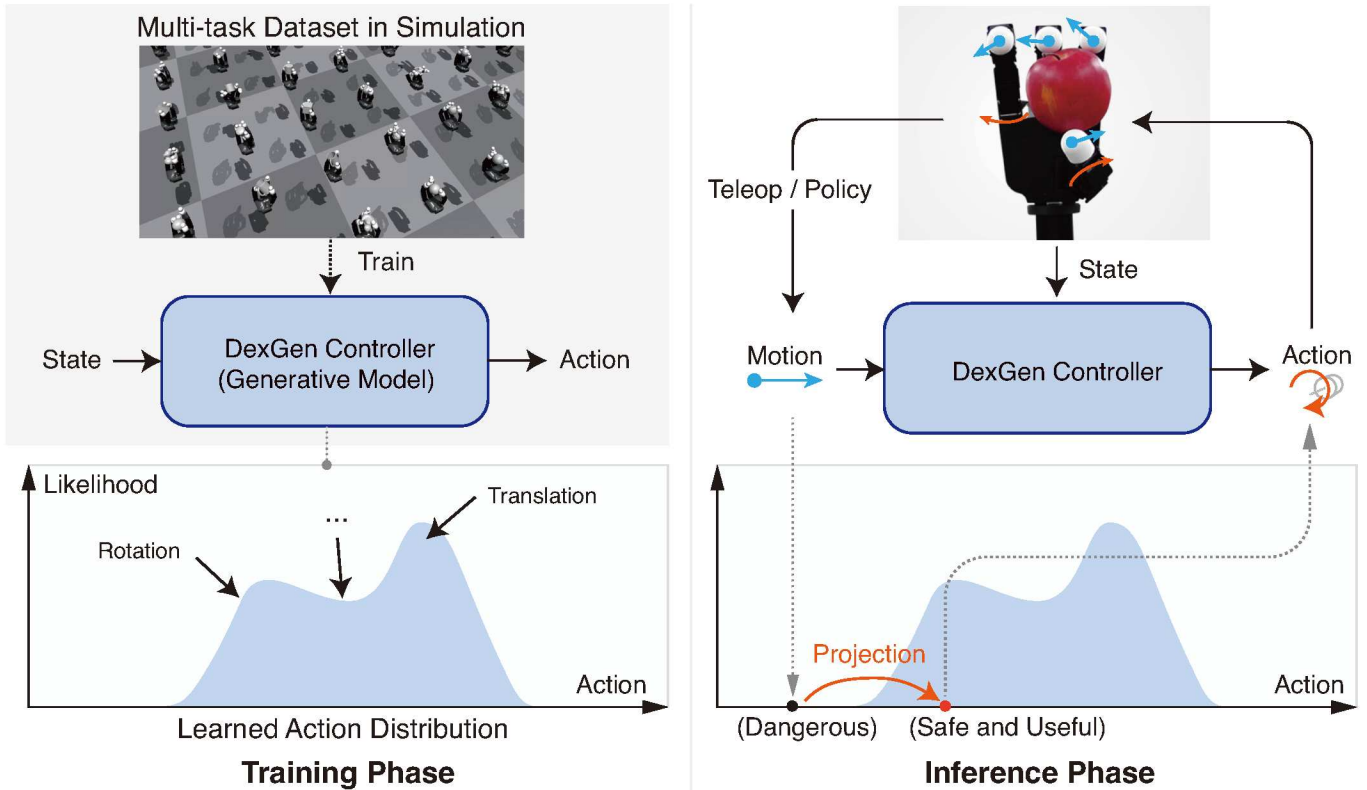


Fig. 2: Overview of proposed framework. **Left (Training):** We collect a large multi-task dexterous in-hand manipulation dataset in simulation to pretrain a generative model that can generate diverse actions conditioned on the current state. The pretrained generative model can produce useful actions including rotation, translation, and more intricate behaviors. **Right (Inference):** During inference, we can project dangerous motion produced by teleoperation or policy back to a high-likelihood action with guided sampling. This makes DexGen capable of assisting a coarse high-level policy to perform complex object manipulations.

I. INTRODUCTION

Dexterous robotic hands are increasingly capturing attention due to their potential across various fields, including manufacturing, household tasks, and healthcare [38]. These robotic systems can replicate the fine motor skills of the human hand, enabling complex object manipulation [50, 4]. Their ability to perform tasks requiring human-like dexterity makes them valuable in areas where traditional automation falls short. However, effectively teaching dexterous in-hand manipulation skills to robotic hands remains a key challenge in robotics.

Recent data-driven approaches to teach robots dexterous manipulation skills can be broadly categorized into two categories: human teleoperation (for imitation learning) [20, 15, 53, 42, 12, 34, 11, 59, 51] and sim-to-real Reinforcement Learning (RL) [4, 16, 41, 63, 25, 19, 9, 24, 3, 36, 61, 60, 33, 52]. Despite their success, these methods face several limitations in practical applications. For human teleoperation, a major bottleneck is the collection of high-quality demonstrations [32, 62]. In contact-rich dexterous manipulation, it is challenging for humans to perform safe and stable object manipulation actions, often resulting in objects falling from the hand. This makes teleoperation impractical for dexterous manipulation tasks. For sim-to-real RL, challenges arise from

the significant domain gap between simulation and the real world, as well as the need for highly task-specific reward specifications when training an RL agent for complex tasks. We will discuss these challenges in more detail in Section II.

While each approach has its own set of challenges, combining their strengths offers a promising strategy to address the complexities of dexterous manipulation. Specifically, recent sim-to-real RL works [41, 63] have shown that it is possible to train simple dexterous in-hand object manipulation primitives (e.g. rotation) that can be transferred to a robot in the real world. This suggests that RL can be leveraged to generate a large-scale dataset of dexterous manipulation primitives, including in-hand object rotation, translation, and grasp transitions. Meanwhile, humans excel at composing these skills through teleoperation to address more challenging tasks. For example, Yin et al. have shown that they can perform in-hand reorientation by calling several rotation primitives sequentially [63]. However, the external inputs in these studies are limited to a few discretized commands, lacking control over low-level interactions, such as finger movements and object contact. This limitation makes it difficult to prompt existing models to generate more detailed, finger-level interaction behaviors, such as using a syringe or screwdriver.

Motivated by these observations, in this paper, we propose

a novel training framework called DexterityGen (DexGen) to address the challenges of teaching dexterous in-hand manipulation skills. Our main idea is to use a broad, multitask simulation dataset generated via RL to pretrain a generative behavior model (DexGen) that can translate a coarse motion command to safe robot actions which can maximally preserve the motion while guaranteeing safety. In real-world applications, an external policy, such as human teleoperation or an imitation policy, can be used to prompt DexGen to execute meaningful manipulation skills. Our approach effectively decouples high-level semantic motion generation from fine-grained low-level control, serving as a foundational low-level dexterity controller.

We validate our DexGen framework through both simulated and real-world experiments. In simulation, we demonstrate that DexGen significantly enhances the robustness and performance of a highly perturbed noisy policy, extending its stable operation duration by 10-100 times and enabling success even when input commands are predominantly noise. In real-world scenarios, we employ human teleoperation as a proxy for high-level motion commands and test the framework on various challenging dexterous manipulation tasks involving complex hand-object interactions across a diverse set of objects. Notably, it successfully synthesizes trajectories to solve challenging tasks, such as reorienting and using syringes and screwdrivers for the first time (with human guidance).

II. EXISTING APPROACHES: CHALLENGES AND OPPORTUNITIES

In this section, we review the challenges and opportunities with existing approaches to dexterous manipulation that motivate our work.

A. Human Teleoperation for Imitation Learning

Challenge: Dexterous manipulation via teleoperation is challenging for humans due to the following reasons:

a) Partial Observability: During in-hand manipulation, the object motion is determined by the contact dynamics between hand and object [58, 38, 23]. Successful manipulation requires perceiving and understanding contact information, such as normal force and friction, to generate appropriate torques. However, human operators face challenges in observing this information due to occlusion and limited tactile feedback. Additionally, existing discrete haptic feedback (e.g. binary vibration) alone is often inadequate for conveying complex touch interactions and contact geometries.

b) Embodiment Gap: Although human and robot hands may appear similar at first glance, they differ significantly in their kinematic structures and geometries. For example, human fingers have smooth and compliant surfaces, while the robot fingers often have rough edges. These differences result in discrepancies in contact dynamics, making it challenging to directly transfer our understanding of human finger motions for object manipulation to robotic counterparts. In our early experiments, we find the object motion very sensitive to the change of fingertip shape.

c) Motion Complexity: Dexterous in-hand manipulation involves highly complex motion. The process requires precise control of a high degree-of-freedom dynamical system. Any suboptimal teleoperation motion at any DOF can lead to failures such as breaking grasping contacts.

d) Inaccuracy of Actions (Force): Existing robot hand teleoperation systems are based on hand retargeting with position control, which lacks an intuitive force control interface to users. As a result, users can only influence force through position-control errors, making teleoperation particularly challenging in force-sensitive scenarios. Moreover, the presence of noise in real world robot system further complicates control.

Opportunity: High-level (Semantical) Motion Control While humans may find it challenging to provide fine-grained, low-level actions directly, human teleoperation or even video demonstrations can still offer valuable coarse motion-level guidance for a variety of complex real-world tasks. Humans possess intuitive knowledge, such as where a robot hand should make contact and what constitutes a good grasp. Thus, human data can be leveraged to create a high-level semantic action plan. In locomotion and whole-body control research, recent studies have proposed using teleoperation commands as high-level motion prompts [10, 17]. However, extending this approach to finegrained dexterous manipulation remains an open question.

B. Sim-to-real Reinforcement Learning

Challenge: Developing a generalized sim-to-real policy for dexterous manipulation involves two main challenges:

a) Sim-to-Real Gap: It is difficult to reproduce real-world sensor observation (mainly for vision input) and physics in simulation. This gap can make sim-to-real transfer highly challenging for complex tasks. In particular, transferring a vision-based control policy from simulation to real world for dexterous hands is a huge challenge and requires costly visual domain randomization [56]. For instance, Dextreme [16] leveraged extensive visual domain randomization with 5M rendered images to train a single object rotation policy.

b) Reward Specification: A more important issue, beyond the sim-to-real gap, is the notorious challenge of designing reward functions for long-horizon, contact-rich problems. Existing methods often involve highly engineered rewards or overly complicated learning strategies [9], which are task-specific and limit scalability.

Opportunity: Low-level (Physical) Action Control Although sim-to-real RL can be difficult, especially for those complex long-horizon or vision-based tasks, some recent works have shown that sim-to-real RL is sufficient to build diverse transferable manipulation primitives based on proprioception and touch [63]. Therefore, one opportunity for sim-to-real RL is to create rich low-level action primitives that can be combined with the high-level action plan discussed above. In this paper, we achieve this through generative pretraining.

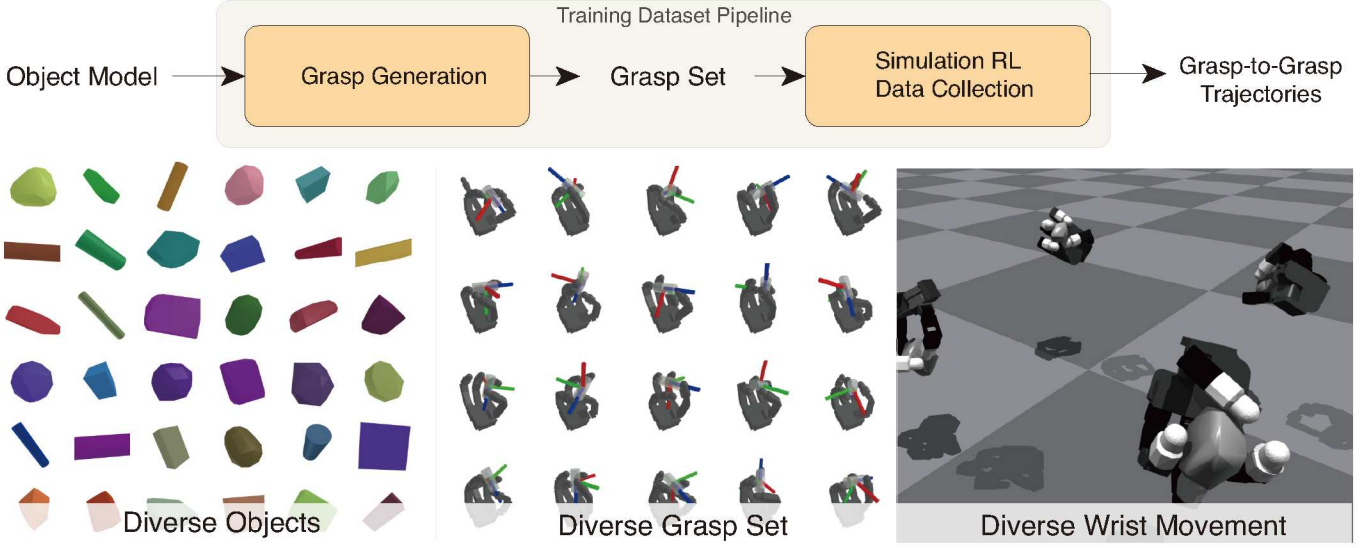


Fig. 3: **Dataset:** The Anygrasp-to-Anygrasp dataset generation pipeline is designed for the generative pretraining of DexGen. For a wide variety of objects, we extensively search for potential grasp configurations, using these as both the initial and goal states for RL policies. To ensure our diffusion model can manage diverse scenarios, we incorporate varied wrist poses, movements, and domain randomization during RL training and data collection.

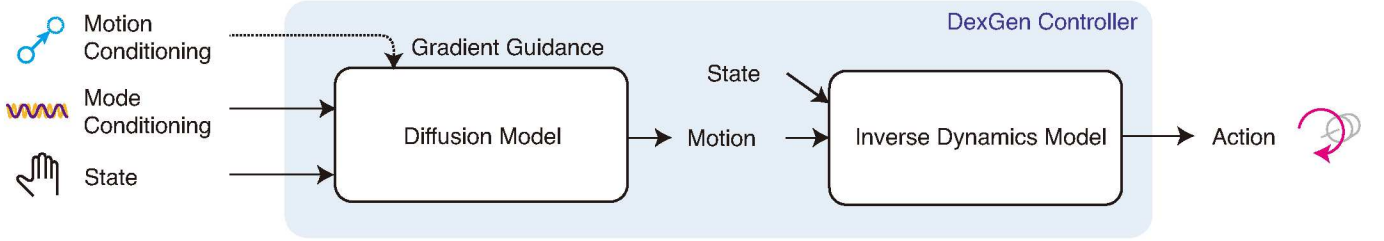


Fig. 4: **Model:** Architecture of the DexGen controller. The whole system takes robot state, external motion conditioning, and mode conditioning as input. A diffusion model first generates the motion as the intermediate action representation. The motion conditioning is not fed into the diffusion model directly but as the gradient guidance during the diffusion sampling. Then, another inverse dynamics model will translate the generated motion to executable robot action. We implement our diffusion model as a UNet in this paper. The inverse dynamics model is a residual multilayer perceptron.

III. THE DEXGEN CONTROLLER

We propose to pretrain a generative behavior model $p_\theta(a|o)$ on the simulation dataset to model prior action distribution so that it can generate stable and effective actions a conditioned on the robot state o . During inference, we can sample actions from this distribution and further aligned with external motion commands using gradient guidance. We detail the dataset used for training the model in section III-B, the model architecture in section III-C, and the inference procedure in section III-D.

A. Preliminaries

a) Diffusion Models: Diffusion Model [18] is a powerful generative model capable of capturing highly complex probabilistic distributions, which we use as our base model. The classical form of the diffusion model is the Denoising Diffusion Probabilistic Model (DDPM) [18]. DDPM defines a forward process that gradually adds noise to the data sample

$$x_0 \sim p_{data}(x):$$

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon_t, \quad (1)$$

where α_t is some noising schedule. We have $x_t \sim \mathcal{N}(\sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I)$ where $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ goes to 0 as $t \rightarrow +\infty$. DDPM trains a model $\mu_\theta(x_t, t)$ to predict denoised sample x_0 given the noised sample x_t with its timestep t . During sampling, DDPM generates the sample by removing the noise through a reverse diffusion process:

$$p(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I) \quad (2)$$

DDPM can generate high-fidelity samples in both vision and robotics applications. In addition to the power to generate data samples faithfully, diffusion models also support guided sampling [21], which turns out to be very useful in our set-up.

Specifically, when $p_{data}(x)$ is modeled by a diffusion model, we can sample from a product probability distribution

$p(x) \propto p_{data}(x)h(x)$ where $h(x)$ is given by a differentiable energy function $h(x) = \exp J(x)$. To do this, we only need to introduce a small modification to the reverse diffusion step. Given the current sample μ , we add a correction term $\alpha \Sigma \nabla J(\mu)$ to μ . Here, α is a step size hyperparameter and Σ is the variance in each diffusion step. This will guide the sample towards high-energy regions in the sample space. We can set $h(x)$ to control the style of generated samples.

b) Robot System and Notations: In this paper, we assume the robot hand is driven by a widely used PD controller. At each control timestep, we command a joint target position $\tilde{q}_t$ and the controller will use torque $\tau = K_p(\tilde{q}_t - q_t) - K_d\dot{q}_t$ to drive the joints. Here, q_t is the current joint position, and $\dot{q}_t$ is the joint velocity. K_p and K_d are two constant scalar gains. We use x_t to denote the key point positions of finger links at time t in the wrist frame. Note that our algorithm does not rely on a specific system implementation and can be extended to other robot systems. We can also specify keypoints and actions for other robots to implement our proposed algorithm.

B. Large-Scale Behavior Dataset Generation

Since human teleoperation or external policies will control the robot hand to interact with the object in diverse ways, our model should be capable of providing refinement for all these potential scenarios (states). To achieve this, we require a large-scale behavior dataset to pretrain our DexGen model, ensuring comprehensive coverage of the state space. We accomplish this by collecting object manipulation trajectories in simulation through reinforcement learning.

Anygrasp-to-Anygrasp To ensure our dataset can cover a broad range of potential states, we introduce Anygrasp-to-Anygrasp as our central pretraining task. This task captures the essential part of in-hand manipulation, which is to move the object to arbitrary configurations. For each object, we define our training task as follows. We first generate a set of object grasps using Grasp Analysis and Rapidly-exploring Random Tree (RRT) [31], similar to the Manipulation RRT procedure [25]. Each generated grasp is defined as a tuple (hand joint position, object pose). In each RL rollout, we initialize the object in the hand with a random grasp. We set the goal to be a randomly selected nearby grasp using the k Nearest-Neighbor search. After reaching the current grasp goal, we update the goal in the same way. We find it crucial to select a nearby reachable goal during the training process, as learning to reach a distant grasp directly can be difficult. After training, we use this anygrasp-to-anygrasp policy to rollout grasp transition sequences to cover all the possible hand-object interaction modes. We sample over 100K grasp for most objects during grasp generation to ensure coverage. This training procedure yields a rich repertoire of useful skills, including object translation and reorientation, which the high-level policy can leverage for solving downstream tasks (Figure 5). In addition to the Anygrasp-to-Anygrasp task, we also introduce other tasks such as free finger moving and fine-grained manipulation (e.g. fine rotation) to handle tasks that have special precision requirements.

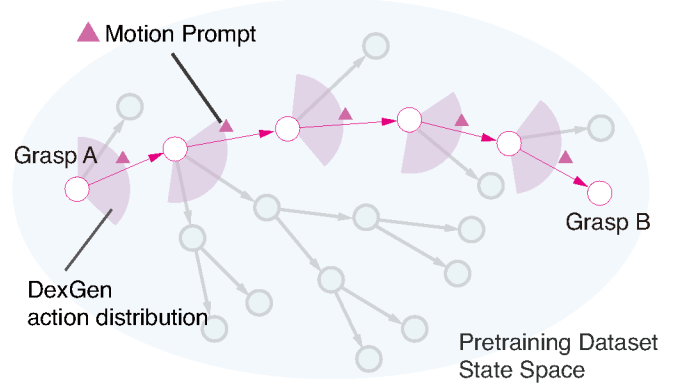


Fig. 5: Our large-scale, multi-task pretraining dataset covers diverse grasp to grasp transitions (arrows). DexGen controller learns the dataset action distribution (purple shaded area) at each state, and we can use sequential motion prompting (purple triangle) to perform a useful long-horizon skill, connecting two distance states.

During RL training, we use a diverse set of random objects and wrist poses. For each task, we include random geometrical objects with different physical properties. To enhance the robustness of our policy, we randomly adjust the wrist to different poses throughout the process, in addition to employing commonly used domain randomizations, so the policy will learn to counteract the gravity effects and exhibit prehensile manipulation behavior (Figure 3). By combining all these data, the robot hand can manipulate different kinds of objects in different wrist configurations against gravity rather than being limited to manipulating a single object at a certain pose. More details of the RL training can be found in Appendix.

We collect a total of 1×10^{10} transitions as our simulation dataset, equivalent to 31.7 years of real world experience. Generating this dataset (by rolling out trained RL policies) requires 300 GPU hours. Although the dataset is large, we hypothesize it can still be far from sufficient as the human dexterity emerges from millions of years of evolution. Nevertheless, this simulated dataset still enables reliable dexterous behavior that have not been showed before.

C. DexGen Model Architecture

We illustrate our DexGen model architecture in Figure 4. The DexGen model has two modules. The first module is a diffusion model that characterizes the distribution of robot finger keypoint motions given current observations. Here we use 3D keypoint motions $\Delta x \in \mathbb{R}^{T \times K \times 3}$ in the robot hand frame as an intermediate action representation. This representation is particularly advantageous for integrating guidance from human teleoperation. In this context, T is the future horizon, K is the number of finger keypoints. The second module in DexGen is an inverse dynamics model, which converts the keypoint motions to executable robot actions (i.e. target joint position) $a_t = \tilde{q}_t$.

We use a UNet-based [46] diffusion model to fit the complex keypoint motion distribution of our multitask dataset. Our



Fig. 6: Real world experimental setup based on Allegro Hand with a Franka Panda Arm (Left). We use human teleoperation (Right) as a proxy for high-level policy.

model learns to generate several future finger keypoint offsets $\Delta x_i = x_{t+i} - x_t$ conditioned on the robot state at timestep t and a mode conditioning variable. The state is a stack of historical proprioception information. The mode conditioning variable is a one-hot vector to explicitly indicate the intention of the task. For instance, when placing an object we do not want the model to produce actions that will make the robot hold the object firmly. Without introducing a “release object” indicator, it is hard to prompt the hand to release the object if most of the actions in the dataset will keep the object in the palm. In our dataset, the majority of transitions are labeled with a “default” (unconditional) label, and only a small portion of them corresponding to specialized scenarios has a special mode label. We only use a specialized precision rotation mode label for screwdriver in our experiments. For releasing object, we find that disabling DexGen controller is sufficient in practice.

The inverse dynamics model is a simple residual multilayer perceptron that outputs a normal distribution to model the actions conditioned on the current robot state and motion command. We train both the diffusion model and inverse dynamics model with our generated simulation dataset using the standard diffusion model loss function and the MSE loss for regression respectively. We train these models with the AdamW optimizer [35, 29] for 15 epochs using 96 GPUs, which takes approximately 3 days. The detailed network setup can be found in the appendix.

D. Inference: Motion Conditioning with Guided Sampling

Our goal is to sample a keypoint motion that is both safe (i.e. from our learned distribution $p_\theta(\Delta x|o)$) and can maximally preserve the input reference motion. Formally, this can be written as $\Delta x \sim p_\theta(\Delta x|o) \exp(-\text{Dist}(\Delta x, \Delta x_{input}))$. Here, $\Delta x_{input} \in \mathbb{R}^{K \times 3}$ is the input commanded fingertip offset, and Dist is a distance function that quantifies the distance between the predicted sequence and the input reference. There can be many ways to instantiate this distance function. In this paper,



Fig. 7: Part of our real world testing objects, which are not present in our pretraining dataset. We include objects of different sizes, masses, and aspect ratios.

we find the following simple distance function works well empirically:

$$\text{Dist}(\Delta x, \Delta x_{input}) = \sum_{i=1}^T \|\Delta x_i - \Delta x_{input}\|^2. \quad (3)$$

The above function encourages the generated future fingertip position to closely match the commanded fingertip position. Since the action of the robot hand has a high degree of freedom (16 for the Allegro hand used in this paper), naive sampling strategies become computationally intractable. To address this, we propose using gradient guidance in the diffusion sampling process to incorporate motion conditioning. In each diffusion step, we adjust the denoised sample Δx by subtracting $\alpha \Sigma \nabla_{\Delta x} \text{Dist}(\Delta x, \Delta x_{input})$ as a guide. Here α is a parameter of the strength of the guidance to be tuned, which we will study in experiments. The generated finger keypoint movement is then converted to action by the inverse dynamics model. We use DDIM sampler [54] during inference for 10Hz control. The total sampling time is around 27ms (37Hz) on a Lambda workstation equipped with an NVIDIA RTX 4090 GPU.

IV. EXPERIMENTS

In the experiments, we first validate the effectiveness of DexGen through simulated experiments, demonstrating its ability to enhance the robustness and success rate of extremely suboptimal policies. Then, we test our system in the real world with a focus on its application in shared autonomy. Our results show that DexGen can assist a human operator in executing unprecedented dexterous manipulation skills with remarkable generalizability.

A. System Setup

In this paper, we use Allegro Hand as our manipulator and we attach the Allegro Hand to a Franka-panda robot arm. In the teleoperation experiments in real world, we use a retargeting-based system to control the robot with human hand

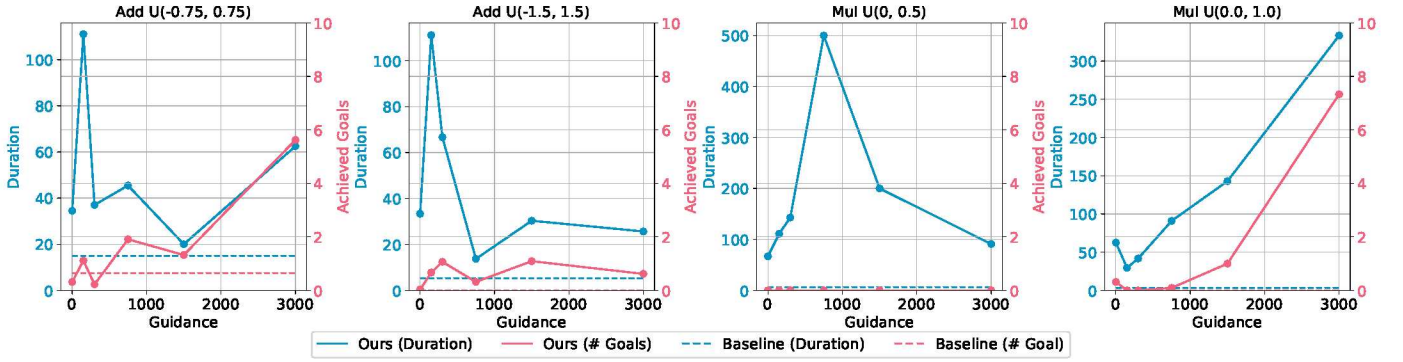


Fig. 8: Results of simulation evaluation. We use DexGen to correct several noise-corrupted expert policies. Note that each dimension of action space is bounded by $[-1, 1]$ and these noises ruin the expert action most of the time. We measure the average duration (in seconds) and number of achieved goals per trial over a 20-minute simulated experiment. As shown in the figure, DexGen can successfully improve the performance of these policies. Across the experiments, DexGen can boost the duration by 10-100x and even help an extremely perturbed policy to achieve success where the baseline fails.

gestures. The human hand pose is captured by Manus Glove and retargeted to the Allegro hand through the Geometric Retargeting algorithm [64]. We obtain the 6D human wrist pose via the Vive tracking system and use it to control the robot arm separately. Although we use this single robot setup in our experiments, we believe our method is general and can be applied to other hand setups.

B. Simulated Experiments

1) *Experimental Setup*: We first test the capability of DexGen in assisting suboptimal policies in solving the Anygrasp-to-Anygrasp task in simulation. We simulate 2 kinds of suboptimal policies with an expert RL policy π_{exp} . The first one is $\pi_{noisy}(a|s) = \pi_{exp}(a|s) + \mathcal{U}(-\alpha, \alpha)$, which simulates an expert that can perform dangerous suboptimal actions through additive uniform noise. The second is $\pi_{slow}(a|s) = \mathcal{U}(0, \alpha)\pi_{exp}(a|s)$, which is a slowdown version of expert. We compare these suboptimal experts π to their assisted counterparts $\text{DexGen} \circ \pi$. We record the average number of critical failures (drop the object) and the number of goal achievements within a certain time of different policies.

2) *Main Results*: We plot the result of different policies in Figure 8. We find that without our assistance, the noisy expert has much more frequent failures. As a result, it can only hardly achieve any goals in the evaluation. In contrast, with the assistance of DexGen, we can partially recover the performance of this noisy expert. We also find that for different policies, the optimal guidance value is also different. Fortunately, there is a common region working well for all these policies. Moreover, when the guidance is relatively small, although we can maintain the object in hand, we can not achieve the desired goal as well because DexGen does not know what the goal is. When the guidance becomes too large, the potentially suboptimal external motion command may take over DexGen guidance and lead to a lower duration in some cases.

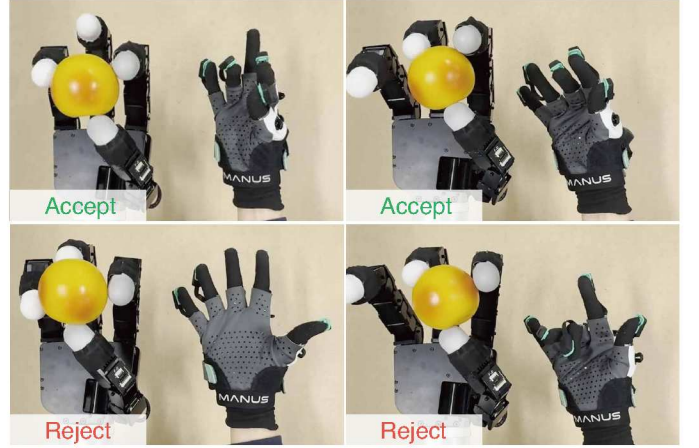


Fig. 9: DexGen can maximally preserve input action while correcting dangerous actions. DexGen can reject users' behavior (open up the palm) and keep holding the object.

C. Real World Experiments:

We have demonstrated that our system can provide effective assistance through simulated validation. Then, we further design several tasks for benchmarking in the real world. In the first set of experiments, we ask a human teleoperator to act as an external high-level policy and we evaluate whether our system can assist humans to solve diverse dexterous manipulation tasks. We introduce a set of atomic skills that covers common in-hand dexterous manipulation behavior.

- **In-hand Object Reorientation** The user is required to control the hand to rotate a given object to a specific pose. In the beginning, we initialize the object in the air over the palm, and the user needs to first teleoperate the hand to grasp the object.
- **Functional Grasping** Regrasping is a necessary step in tool manipulation. The user is asked to perform a power grasp on the tool handle placed either horizontally (normal) or vertically in the air (horizontal functional grasp).

TABLE I: Performance of evaluated methods on the real-world tasks. We report success rate (SR) and time-to-fall (TTF) / Holding Time metric which is normalized by the test episode length. The raw teleoperation baseline fails completely on those tasks, while our method can help the teleoperation policy to achieve both stability and success in diverse setups.

Task	Reorient Large (Up)		Reorient Small (Up)		Reorient Large (Down)		Reorient Small (Down)	
	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)
Teleop	0/20	<5.0%	0/20	<5.0%	0/20	<5.0%	0/20	<5.0%
Teleop + DexGen	12/20	75%	13/20	79%	10/20	63%	9/20	58%

Task	Func Grasp		Func Grasp (Horizontal)		Regrasp (Ball)		Regrasp (Cylinder)	
	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)	SR($\uparrow$)	TTF($\uparrow$)
Teleop	0/10	<5.0%	1/10	<10.0%	0/10	<5.0%	0/10	<5.0%
Teleop + DexGen	7/10	87%	6/10	80%	5/10	78%	5/10	74%

In the beginning, the user can only perform a pinch grasp or precision grasp.

- **In-hand Regrasping** We define this task as a harder version of object reorientation. In this task, the user is asked to achieve a specific grasp configuration (object pose + finger pose). In the beginning, the object is initialized with a precision grasp on the fingertip.

Besides these tasks, we demonstrate some more realistic, long-horizon tasks as well. These tasks require the user to combine the skills above. In the main text, we only study the following two tasks. We leave more examples in the demo video in the appendix.

- **Screwdriver** In this task, the user needs to pick up a screwdriver lying on the table and use it to tighten a bolt.
- **Syringe** In this task, the user needs to pick up a syringe and inject some liquid into a target region.

a) *Evaluation Protocol*: We evaluate the performance of a teleoperation system by measuring the success rate a human user can achieve when using it to solve certain tasks. Before evaluation, we let users familiarize themselves with each evaluated teleoperation system in 30 minutes. Our experiments involve 2 users in this section.

D. Real World Results

The performance of different approaches is shown in Table I. We observe that humans can hardly use the baseline teleoperation system to solve the tasks above. The user can drop the object easily during the contact-rich manipulation process. Compared to the baseline, our system can successfully help the user to solve many tasks in various challenging setups. During these experiments, we also observe the following intriguing properties of our system:

a) *Protective “Magnetic Effect”*: We find that the fingertips show some “magnetic effect” when they are in contact with the object. When the user mistakenly moves a supporting finger which may drop the object, our model can override that behavior and maintain the contact as if the fingertips are sticking to the object (Figure 9 second row). This explains why the user can achieve a much higher success rate in these dexterous tasks.

TABLE II: The breakdown success analysis of syringe and screwdriver teleoperation. These long-horizon tasks require several stages of manipulation and remain challenging.

Screwdriver	Reorient	Regrasp	Align	Use
	16/20	11/20	5/20	3/20
Syringe	Reorient	Regrasp	Use	
	15/20	9/20	4/20	

b) *Intention Following*: Although our model overrides dangerous user action, we find that in most cases our model can follow the user’s intention (action) well and move along the user-commanded moving direction. During the manipulation procedure, the user can still have a sense of agency over the robot hand and complete a complex task. This finding echoes our simulated result with noisy policies: DexGen can realize the intention in the noisy suboptimal actions.

We also present a breakdown analysis of the long-horizon tasks in Table II. For the first time, we enable such long-horizon dexterous manipulation behavior in the real world through teleoperation. Achieving tool use remains challenging as it involves several stages of complex dexterous manipulation: we can achieve a reasonable stage-wise success rate, but chaining these skills together is difficult. However, we believe that improving stage-wise policy in the future can eventually close the gap (see the conclusion part).

V. RELATED WORKS

a) Foundation Models and Pretraining for Robotics:

In recent years, the success of large foundation models in natural language processing and computer vision [2, 30, 57] has attracted much attention in building foundation models for robotics [6, 7, 13, 43, 55, 39, 66, 28, 26]. Existing works typically focus on building a large end-to-end control model by pretraining them on large real world datasets. Our framework also leverages large-scale pretraining, but it differentiates from these works in various aspects. First, we consider pretraining a controller on pure simulation datasets rather than real world datasets which require extensive human efforts in

data collection with teleoperation. Second, we study dexterous manipulation with a high DOF robotic hand and demonstrate the advantage of generative pretraining in this challenging scenario for the first time, while existing works typically consider parallel jaw gripper problems. Third, we build a low-level foundation controller that can be prompted with continuous fine-grained guidance to provide useful actions, which can be potentially integrated with high-level planning policies in the future. Most existing robotic foundation models are conditioned on discrete language prompts or task embeddings. In summary, our pretraining framework for building a foundational low-level controller presents a new perspective in the foundation model literature.

b) Shared Autonomy: Our system is also related to shared autonomy research [1, 27, 8, 47, 49], which focuses on leveraging external action guidance to produce effective actions. Some works focus on how to train RL agents with external actions (e.g. from teleoperation) [44, 14, 45]. In their setup, the external inputs are usually treated as part of observation fed to the RL policy. Compared to this line of work, our method does not involve human actions in the training phase. Another line of work assumes the existence of a few task-specific intentions and goals and reduces the shared autonomy problem to the goal or intent inference [1, 22]. A limitation of this line of work in dexterous manipulation is that they do not allow fine-grained finger control since they only provide a few options for high-dimensional action space. Our method samples fine-grained low-level behavior according to user commands in high-dimensional action space and does not suffer from this problem. The most relevant works are [5, 65], which also use some sampled distribution to correct user behavior. We use a different correction procedure and investigate a more general and challenging dexterous manipulation setup.

VI. CONCLUSION

In this paper, we have presented DexterityGen as an initial attempt towards building a foundational low-level controller for dexterous manipulation. We have demonstrated that generative pretraining on diverse multi-task simulated trajectories yield a powerful generative controller that can translate coarse motion prompts to effective low-level actions. Combined with external high-level policy, our controller exhibits unprecedented dexterity. We believe that our work opens up new possibilities in various dimensions of dexterous manipulation.

Limitations and Future Work Our exploration still has some limitations to be addressed in future works, which we discuss as follows.

- 1) **Touch Sensing** In this work, we rely on joint angle proprioception for implicit touch sensing (i.e. inferring force by reading control errors), which can be insufficient and nonrobust for fine-grained problems. In many cases, it is impossible to recover contact geometry based on joint angle error. In the future, we will add touch to pretraining, which has been shown possible for sim-to-

real transfer. We hope that this can further improve the robustness of our system.

- 2) **Vision: Hand-Eye Coordination** Our low-level controller does not involve vision. Nevertheless, we observe that vision feedback is necessary for producing accurate tool motions for many tasks such as using a screwdriver. It is unclear whether this vision processing should be in the high-level policy or part of the proposed foundation low-level controller, and we leave this to future research.
- 3) **Real-world Finetuning** In this work, we deploy our controller in a zero-shot manner. However, due to the sim-to-real gap, it can still be necessary to fine-tune our controller with some real world experience.

ACKNOWLEDGMENTS

This work was partially carried out during Zhao-Heng Yin’s intern at the Meta FAIR Labs. This work is supported by the Meta FAIR Labs and ONR MURI N00014-22-1-2773. Pieter Abbeel holds concurrent appointments as a Professor at UC Berkeley and as an Amazon Scholar. This paper describes work performed at UC Berkeley and is not associated with Amazon.

REFERENCES

- [1] Daniel Aarno, Staffan Ekvall, and Danica Kragic. Adaptive virtual fixtures for machine-assisted teleoperation tasks. In *International Conference on Robotics and Automation (ICRA)*, 2005.
- [2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.
- [3] Ananye Agarwal, Shagun Uppal, Kenneth Shaw, and Deepak Pathak. Dexterous functional grasping. In *Conference on Robot Learning (CoRL)*, 2023.
- [4] OpenAI: Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39 (1):3–20, 2020.
- [5] Alexander Broad, Todd Murphey, and Brenna Argall. Highly parallelized data-driven mpc for minimal intervention shared control. In *Robotics: Science and Systems (RSS)*, 2019.
- [6] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.
- [7] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Xi Chen, Krzysztof Choromanski, Tianli Ding, Danny Driess, Avinava Dubey, Chelsea Finn, et al. Rt-2: Vision-language-action models transfer web knowledge

- to robotic control. *arXiv preprint arXiv:2307.15818*, 2023.
- [8] Tom Carlson and Yiannis Demiris. Human-wheelchair collaboration through prediction of intention and adaptive assistance. In *International Conference on Robotics and Automation (ICRA)*, 2008.
 - [9] Yuanpei Chen, Chen Wang, Li Fei-Fei, and C Karen Liu. Sequential dexterity: Chaining dexterous policies for long-horizon manipulation. In *Conference on Robot Learning (CoRL)*, 2023.
 - [10] Xuxin Cheng, Yandong Ji, Junming Chen, Ruihan Yang, Ge Yang, and Xiaolong Wang. Expressive whole-body control for humanoid robots. In *Robotics: Science and Systems (RSS)*, 2024.
 - [11] Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. In *Conference on Robot Learning (CoRL)*, 2024.
 - [12] Runyu Ding, Yuzhe Qin, Jiyue Zhu, Chengzhe Jia, Shiqi Yang, Ruihan Yang, Xiaojuan Qi, and Xiaolong Wang. Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning. *arXiv preprint arXiv:2407.03162*, 2024.
 - [13] Yilun Du, Sherry Yang, Bo Dai, Hanjun Dai, Ofir Nachum, Josh Tenenbaum, Dale Schuurmans, and Pieter Abbeel. Learning universal policies via text-guided video generation. In *Neural Information Processing Systems (NeurIPS)*, 2024.
 - [14] Yuqing Du, Stas Tiomkin, Emre Kiciman, Daniel Polani, Pieter Abbeel, and Anca Dragan. Ave: Assistance via empowerment. In *Neural Information Processing Systems (NeurIPS)*, 2020.
 - [15] Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. Dexpilott: Vision-based teleoperation of dexterous robotic hand-arm system. In *International Conference on Robotics and Automation (ICRA)*, 2020.
 - [16] Ankur Handa, Arthur Allshire, Viktor Makoviychuk, Aleksei Petrenko, Ritvik Singh, Jingzhou Liu, Denys Makoviichuk, Karl Van Wyk, Alexander Zhurkevich, Balakumar Sundaralingam, et al. Dextreme: Transfer of agile in-hand manipulation from simulation to reality. In *International Conference on Robotics and Automation (ICRA)*, 2023.
 - [17] Tairan He, Zhengyi Luo, Xialin He, Wenli Xiao, Chong Zhang, Weinan Zhang, Kris Kitani, Changliu Liu, and Guanya Shi. Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning. In *Conference on Robot Learning (CoRL)*, 2024.
 - [18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Neural Information Processing Systems (NeurIPS)*, 2020.
 - [19] Binghao Huang, Yuanpei Chen, Tianyu Wang, Yuzhe Qin, Yaodong Yang, Nikolay Atanasov, and Xiaolong Wang. Dynamic handover: Throw and catch with bimanual hands. In *Conference on Robot Learning (CoRL)*, 2023.
 - [20] Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan, and Chrisina Jayne. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.
 - [21] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In *International Conference on Machine Learning (ICML)*, 2022.
 - [22] Shervin Javdani, Siddhartha S Srinivasa, and J Andrew Bagnell. Shared autonomy via hindsight optimization. In *Robotics: Science and Systems (RSS)*, 2015.
 - [23] Xuerong Ji and Jing Xiao. Planning motions compliant to complex contact states. *The International Journal of Robotics Research*, 20(6):446–465, 2001.
 - [24] Aditya Kannan, Kenneth Shaw, Shikhar Bahl, Pragna Mannam, and Deepak Pathak. Deft: Dexterous fine-tuning for real-world hand policies. In *Conference on Robot Learning (CoRL)*, 2023.
 - [25] Gagan Khandate, Siqi Shang, Eric T Chang, Tristan Luca Saidi, Yang Liu, Seth Matthew Dennis, Johnson Adams, and Matei Ciocarlie. Sampling-based exploration for reinforcement learning of dexterous manipulation. In *Robotics: Science and Systems (RSS)*, 2023.
 - [26] Alexander Khazatsky, Karl Pertsch, Suraj Nair, Ashwin Balakrishna, Sudeep Dasari, Siddharth Karamcheti, Soroush Nasiriany, Mohan Kumar Srirama, Lawrence Yunliang Chen, Kirsty Ellis, et al. Droid: A large-scale in-the-wild robot manipulation dataset. In *Robotics: Science and Systems (RSS)*, 2024.
 - [27] Hyun K Kim, J Biggs, W Schloerb, M Carmena, Mikhail A Lebedev, Miguel AL Nicolelis, and Mandayam A Srinivasan. Continuous shared control for stabilizing reaching and grasping with brain-machine interfaces. *IEEE Transactions on Biomedical Engineering*, 53(6):1164–1173, 2006.
 - [28] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. In *Conference on Robot Learning (CoRL)*, 2024.
 - [29] Diederik P Kingma. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2014.
 - [30] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.
 - [31] Steven M LaValle and James J Kuffner. Rapidly-exploring random trees: Progress and prospects. *Algorithmic and Computational Robotics*, pages 303–307, 2001.
 - [32] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *arXiv preprint*

arXiv:2005.01643, 2020.

- [33] Toru Lin, Zhao-Heng Yin, Haozhi Qi, Pieter Abbeel, and Jitendra Malik. Twisting lids off with two hands. In *Conference on Robot Learning (CoRL)*, 2024.
- [34] Toru Lin, Yu Zhang, Qiyang Li, Haozhi Qi, Brent Yi, Sergey Levine, and Jitendra Malik. Learning visuotactile skills with two multifingered hands. In *International Conference on Robotics and Automation (ICRA)*, 2025.
- [35] I Loshchilov. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2018.
- [36] Tyler Ga Wei Lum, Martin Matak, Viktor Makoviychuk, Ankur Handa, Arthur Allshire, Tucker Hermans, Nathan D Ratliff, and Karl Van Wyk. Dextrah-g: Pixels-to-action dexterous arm-hand grasping with geometric fabrics. In *Conference on Robot Learning (CoRL)*, 2024.
- [37] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.
- [38] Allison M Okamura, Niels Smaby, and Mark R Cutkosky. An overview of dexterous manipulation. In *International Conference on Robotics and Automation (ICRA)*, 2000.
- [39] Abby O’Neill, Abdul Rehman, Abhinav Gupta, Abhiram Maddukuri, Abhishek Gupta, Abhishek Padalkar, Abraham Lee, Acorn Pooley, Agrim Gupta, Ajay Mandlekar, et al. Open x-embodiment: Robotic learning datasets and rt-x models. *arXiv preprint arXiv:2310.08864*, 2023.
- [40] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *AAAI conference on Artificial Intelligence (AAAI)*, 2018.
- [41] Haozhi Qi, Ashish Kumar, Roberto Calandra, Yi Ma, and Jitendra Malik. In-hand object rotation via rapid motor adaptation. In *Conference on Robot Learning (CoRL)*, 2023.
- [42] Yuzhe Qin, Wei Yang, Binghao Huang, Karl Van Wyk, Hao Su, Xiaolong Wang, Yu-Wei Chao, and Dieter Fox. Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. In *Robotics: Science and Systems (RSS)*, 2023.
- [43] Ilija Radosavovic, Baifeng Shi, Letian Fu, Ken Goldberg, Trevor Darrell, and Jitendra Malik. Robot learning with sensorimotor pre-training. In *Conference on Robot Learning (CoRL)*, 2023.
- [44] Siddharth Reddy, Anca D Dragan, and Sergey Levine. Shared autonomy via deep reinforcement learning. In *Robotics: Science and Systems (RSS)*, 2018.
- [45] Siddharth Reddy, Sergey Levine, and Anca Dragan. First contact: Unsupervised human-machine co-adaptation via mutual information maximization. In *Neural Information Processing Systems (NeurIPS)*, 2022.
- [46] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- [47] Sebastian Schröer, Ingo Killmann, Barbara Frank, Martin Völker, Lukas Fiederer, Tonio Ball, and Wolfram Burgard. An autonomous robotic assistant for drinking. In *International Conference on Robotics and Automation (ICRA)*, 2015.
- [48] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [49] Wilko Schwarting, Javier Alonso-Mora, Liam Pauli, Sertac Karaman, and Daniela Rus. Parallel autonomy in automated vehicles: Safe motion generation with minimal intervention. In *International Conference on Robotics and Automation (ICRA)*, 2017.
- [50] Kenneth Shaw, Ananye Agarwal, and Deepak Pathak. Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning. In *Robotics: Science and Systems (RSS)*, 2023.
- [51] Kenneth Shaw, Yulong Li, Jiahui Yang, Mohan Kumar Srirama, Ray Liu, Haoyu Xiong, Russell Mendonca, and Deepak Pathak. Bimanual dexterity for complex tasks. In *Conference on Robot Learning (CoRL)*, 2024.
- [52] Leon Sievers, Johannes Pitz, and Berthold Bäuml. Learning purely tactile in-hand manipulation with a torque-controlled hand. In *International Conference on Robotics and Automation (ICRA)*, 2022.
- [53] Aravind Sivakumar, Kenneth Shaw, and Deepak Pathak. Robotic telekinesis: Learning a robotic hand imitator by watching humans on youtube. In *Robotics: Science and Systems (RSS)*, 2022.
- [54] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations (ICLR)*, 2020.
- [55] Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. In *Robotics: Science and Systems (RSS)*, 2024.
- [56] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [57] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.
- [58] Jeffrey C. Trinkle and Richard P. Paul. Planning for dexterous manipulation with sliding contacts. *The International Journal of Robotics Research*, 9(3):24–48, 1990.
- [59] Chen Wang, Haochen Shi, Weizhuo Wang, Ruohan Zhang, Li Fei-Fei, and C Karen Liu. Dexcap: Scalable and portable mocap data collection system for dexterous

manipulation. In *Robotics: Science and Systems (RSS)*, 2024.

- [60] Jun Wang, Ying Yuan, Haichuan Che, Haozhi Qi, Yi Ma, Jitendra Malik, and Xiaolong Wang. Lessons from learning to spin” pens”. In *Conference on Robot Learning (CoRL)*, 2024.
- [61] Max Yang, Chenghua Lu, Alex Church, Yijiong Lin, Chris Ford, Haoran Li, Efi Psomopoulou, David AW Barton, and Nathan F Lepora. Anyrotate: Gravity-invariant in-hand object rotation with sim-to-real touch. In *Conference on Robot Learning (CoRL)*, 2024.
- [62] Zhao-Heng Yin and Pieter Abbeel. Offline imitation learning through graph search and retrieval. In *Robotics: Science and Systems (RSS)*, 2024.
- [63] Zhao-Heng Yin, Binghao Huang, Yuzhe Qin, Qifeng Chen, and Xiaolong Wang. Rotating without seeing: Towards in-hand dexterity through touch. In *Robotics: Science and Systems (RSS)*, 2023.
- [64] Zhao-Heng Yin, Changhao Wang, Luis Pineda, Krishna Bodduluri, Tingfan Wu, Pieter Abbeel, and Mustafa Mukadam. Geometric retargeting: A principled, ultra-fast neural hand retargeting algorithm. *arXiv preprint arXiv:2503.07541*, 2025.
- [65] Takuma Yoneda, Luzhe Sun, Bradly Stadie, Matthew Walter, et al. To the noise and back: Diffusion for shared autonomy. In *Robotics: Science and Systems (RSS)*, 2023.
- [66] Tony Z Zhao, Jonathan Tompson, Danny Driess, Pete Florence, Kamyar Ghasemipour, Chelsea Finn, and Ayzaan Wahid. Aloha unleashed: A simple recipe for robot dexterity. In *Conference on Robot Learning (CoRL)*, 2024.

APPENDIX

A. DexGen Training Pipeline

We provide an overview of the full training process in Algorithm 1. The algorithm has two stages. In the first stage, we first collect manipulation trajectories with multiple RL policies. In the second stage, we distill the experience into our controller. For the dataset filtering step, we apply a very simple heuristic rule. If a rollout ends with dropping the object, then we directly discard the last 2 second transitions.

Algorithm 1 Training Procedure of DexGen Controller p_θ

Require: Manipulation tasks $\{T_i\}$ in simulation (e.g. Anygrasp to Anygrasp).

- 1: Train RL policy π_i on each $\{T_i\}$ to convergence.
 - 2: Collect training dataset $\mathcal{D} = \cup_i \text{Rollout}(\pi_i)$.
 - 3: Preprocess dataset $\mathcal{D}$ by filtering failure transitions.
 - 4: Train DexGen controller p_θ on $\mathcal{D}$.
 - 5: **return** p_θ
-

B. Implementation of Anygrasp-to-Anygrasp

The core dexterous manipulation task used by Algorithm 1 is Anygrasp-to-Anygrasp. We describe its implementation as follows.

Grasp Generation To define this task, we first need to generate the grasp set for each object with the Grasp Generation Algorithm 2. The algorithm first generates a base grasp set using heuristic sampling, and we further expand this grasp set via RRT search to ensure that it can cover as many configurations as possible. Note that there exist many approaches for synthesizing grasps. Here, we just provide one option that works well empirically.

Algorithm 2 Grasp Generation

Require: Object mesh $\mathcal{M}$. Initial Grasp Set Size N . RRT Step N_{RRT} .

- 1: Grasp Set $S \leftarrow \text{HeuristicSample}(\mathcal{M}, N)$.
 - 2: $S \leftarrow \text{GraspRRTExpand}(S, \mathcal{M}, N_{RRT})$.
 - 3: **return** S
-

Algorithm 3 HeuristicSample

Require: Object mesh $\mathcal{M}$, Num Samples N .

- 1: $S \leftarrow []$.
 - 2: **while** $\text{len}(S) < N$ **do**
 - 3: $N_{pts} = \text{random}([2, 3, 4])$. // Num grasp point.
 - 4: Point P , Normal $n \leftarrow \text{SampleSurface}(\mathcal{M}, N_{pts})$.
 - 5: **if** $\text{GraspAnalysis}(P, n)$ **then**
 - 6: Object Pose $p \leftarrow \text{RandomPose}()$.
 - 7: Finger Configuration $q \leftarrow \text{Assign}(\mathcal{M}, P, n, q)$.
 - 8: **if** $\text{NoCollision}(q, p, \mathcal{M})$ **then**
 - 9: $S \leftarrow S \cup \{(q, p)\}$
 - 10: **end if**
 - 11: **end while**
 - 12: **return** S
-

Algorithm 4 GraspAnalysis

Require: Contact Points P , Contact Normals n .

- 1: $F_{min} \leftarrow \text{Min solution to Net Force Opt. } (n)$.
 - 2: **if** $F_{min} < F_{thresh}$ **then**
 - 3: **return** TRUE
 - 4: **end if**
 - 5: **return** FALSE
-

For Algorithm 5, we originally followed the implementation proposed by [25]. However, we find that minimizing the wrench can be too strict and it is not efficient for large-scale generation. Therefore, we introduce a simplified optimization problem for grasp analysis as follows, which we find effective in practice.

Algorithm 5 GraspRRTExpand**Require:** Grasp set S , Object Mesh $\mathcal{M}$, RRT Step N_{RRT} .

```

1: for  $i = 1, 2, \dots, N_{RRT}$  do
2:    $(q, p) \leftarrow \text{RandomSample}()$ . //  $q$  finger configuration.
    $p$  object pose.
3:    $(q^*, p^*) \leftarrow \text{NearestNeighbor}((q, p), S)$ .
4:    $(q, p) \leftarrow \text{Interpolate}((q, p), (q^*, p^*))$ .
5:    $(q, p) \leftarrow \text{FixContactAndCollision}(q, p, \mathcal{M})$ .
6:    $S = S \cup \{(q, p)\}$ .
7: end for
8: return  $S$ 

```

Net Force Optimization ($\{n_i\}$)

$$\begin{aligned}
&\underset{f_i}{\text{Minimize:}} && \left\| \sum f_i n_i \right\|^2 \\
&\text{s.t.} && \forall i, f_i \geq 0, \\
&&& \exists i, f_i = 1.
\end{aligned}$$

Intuitively, we apply force f_i at each contact point along contact normal n_i and we optimize for a nontrivial force combination ($\exists f_i = 1$) that can generate a near-zero net force. If the minimizer of this problem is below a threshold, we consider this grasp as stable. Note that the second existence constraint is hard to directly parameterize as a differentiable loss function. In our implementation, we decompose this problem into several subproblems by enforcing $f_1 = 1, f_2 = 1, \dots, f_n = 1$ in each subproblem.

Reward Design The reward function for the Anygrasp-to-Anygrasp task is as follows. It is composed of three different terms, goal-related reward r_{goal} , style-related reward r_{style} , and regularization terms r_{reg} .

$$r = w_{goal}r_{goal} + w_{style}r_{style} + w_{reg}r_{reg}. \quad (4)$$

The goal-related reward term r_{goal} involves target object pose and finger joint positions:

$$r_{goal} = \exp(-\alpha_{pos}\|p_{obj} - p_{obj}^{target}\|^2 - \alpha_{orn}d(R_{obj}, R_{obj}^{target})) \quad (5)$$

$$- \alpha_{hand}\|q - q^{target}\|^2 \quad (6)$$

$$+ \alpha_{bonus}\mathbf{1}(\text{goal achieved}). \quad (7)$$

The regularization term includes the penalty on the action scale, applied torque, and work:

$$r_{reg} = -\alpha_{work}|\dot{q}^T||\tau| - \alpha_{action}\|a\|^2 - \alpha_{tau}\|\tau\|^2. \quad (8)$$

For the style reward, it is a penalty term on the fingertip velocity. This can elicit different manipulation styles (fast movement or slow movement). This reward term is mainly used to boost data diversity in temporal dimensions, see discussion below.

$$r_{style} = \sum_i \alpha_i \|\dot{x}_{tip}^i\|. \quad (9)$$

TABLE III: Domain Randomization Setup

Object: Mass (kg)	[0.03, 0.25]
Object: Friction	[0.5, 1.2]
Object: Shape	$\times \mathcal{U}(0.95, 1.05)$
Hand: Initial Joint Noise	[-0.05, 0.05]
Hand: Friction	[0.5, 1.2]
PD Controller: P Gain	$\times \mathcal{U}(0.8, 1.1)$
PD Controller: D Gain	$\times \mathcal{U}(0.7, 1.2)$
Random Force: Scale	1.0/2.0
Random Force: Probability	0.2
Random Force: Decay Coeff. and Interval	0.99 every 0.1s
Joint Observation Noise (white noise)	$+\mathcal{N}(0, 0.025)$
Joint Observation Noise (episode noise)	$+\mathcal{N}(0, 0.005)$
Action Noise	$+\mathcal{N}(0, 0.05)$

Goal Dynamics A crucial design in the Anygrasp-to-Anygrasp task is the goal dynamics. We find that when we set a goal very far away, the RL policy can usually fail to reach that goal and as a result, the RL learning process can plateau very early. Therefore, throughout the RL process, we set goals within a moderate distance to ensure effective RL learning. Specifically, when the current goal is achieved, we search for a grasp in our grasp cache whose object distance is within a certain range as our next goal. We achieve this through a Nearest Neighbor search. Since NN search is computationally expensive for a large grasp set, we first perform a random down-sampling at each update step before the next goal computation.

C. Boosting Dataset Diversity with Diverse Rewards

To boost the diversity of the training dataset, we use multiple reward setups to train policies of different styles and use all of these policies for data collection. In this paper, we train RL policies with different w_{style} and w_{reg} coefficients and this yields policies of both fast and slow object manipulation behavior. This ensures that real-world states, whether they are from a good policy or a suboptimal one, are effectively managed by our controller.

D. RL Training Setups

We implement all the training tasks and data collection using the IsaacGym simulator [37]. We use Proximal Policy Optimization (PPO) [48] as our RL algorithm. We use asymmetric actor-critic during training, where the actor only observes proprioception information and desired goal (represented by a relative transformation from current state to goal state), while the critic network observes all the state information such as object position and velocity etc. We use MLP to parameterize both actor and critic networks, whose hidden dimensions are both [1024, 512, 512, 256, 256]. We use a learning rate 0.0005, batch size 8192, PPO clip value 0.2, with $\gamma = 0.99$ and GAE $\tau = 0.95$. We use 8192 environments in parallel.

E. Domain Randomization

We apply extensive domain randomizations in both training and data collection. We list the randomized components in the Table III.

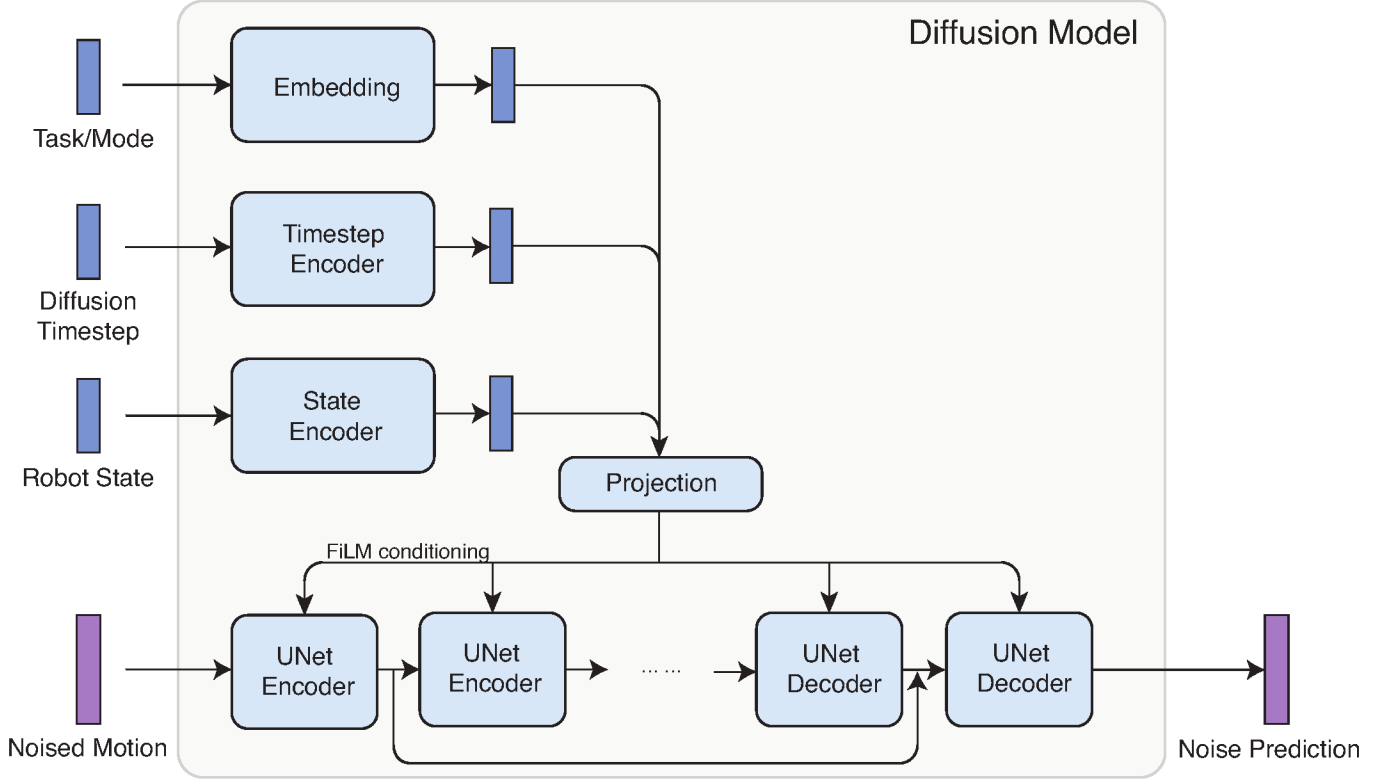


Fig. 10: Diffusion Model in DexGen Controller. We use a standard U-Net based diffusion model with FiLM conditioning.

F. Diffusion Model Architecture

We illustrate our diffusion model architecture in Figure 10. We first use a state encoder and a mode encoder to produce a compact representation for the conditional inputs. Then, we use a UNet to predict the noise added to the current sample, with FiLM-conditioning [40] in the middle layers. We use 3 blocks for both UNet encoder and decoder. For the UNet, we use a hidden dimension of 768 and replace 1D convolution layers with fully connected layers. We implement the state encoder as a 6-layer MLP with hidden dimension 1024. We also use GroupNorm after each MLP layer with group size 8. We experimented with 8 and 12 DDIM steps during the diffusion model inference. We find there is a tradeoff between sample fidelity (action accuracy) and latency, and they affect user experience in different ways.

In this paper, we use $T = 2$ as the future motion prediction horizon, which corresponds to 0.2s future. We use $K = 8$ finger keypoints (PIP of each finger and the fingertips). In early experiments, we used $K = 4$ finger keypoints (fingertips only), but this representation behaves suboptimally in the experiments and has a large inverse dynamics training loss. The simplified representation does not encode the full action information, as $K = 4$ 3D keypoints only span a 12-dimensional space but the hand is 16DOF. We stack 4 history steps of robot proprioception including fingertip position, joint position, target joint position, and control error as input to the diffusion model.