# EIGENDEXPLORE: STRUCTURED EXPLORATION FOR DEXTEROUS MANIPULATION WITH HUMAN PRIORS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Dexterous manipulation poses a challenging high-dimensional optimization problem, as useful behaviors require coordinated motion across many hand joints. In reinforcement learning (RL) and sampling-based trajectory optimization, exploration commonly relies on independent robot joint perturbations, making coordinated behaviors difficult to discover. Prior work reduces this search space for grasp learning using low-dimensional spaces of coordinated joint motions learned from human hand data, but this restricts the expressivity required for general manipulation. Some combine learned and joint-space actions to restore expressivity, but this increases dimensionality and introduces redundancy. We study these effects across diverse manipulation settings, varying action dimensionality, exploration strategy, and the source of human data. Our experiments suggest that human-motion priors are most effective when used to structure exploration rather than change the action representation. Motivated by this finding, we propose EigenDEXplore, which induces correlated exploration by adding perturbations along human-derived eigenvectors to independent joint-space noise, leaving the action space unchanged. Across multiple dexterous hands, EigenDEXplore consistently outperforms joint-space and learned action-space baselines in grasping, in-hand reorientation, and contact-rich manipulation. These gains span unstructured and reference-guided RL, trajectory optimization, and sim-to-real deployment, and are largest in settings with less reward shaping and curriculum design. Project page: `sites.google.com/view/eigendexplore`.

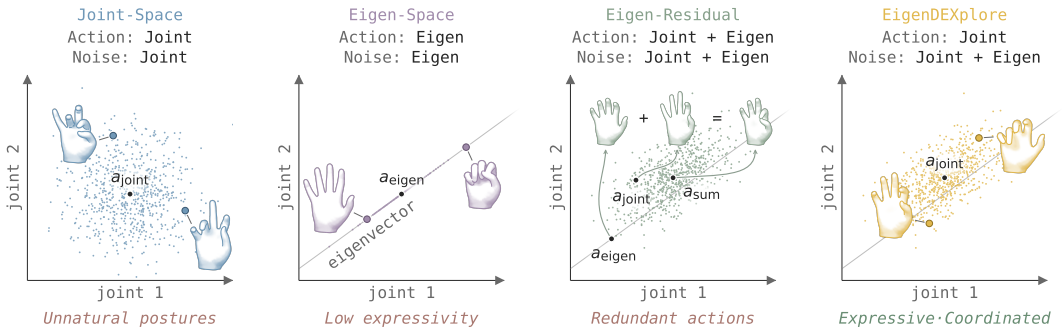


Figure 1: **Action spaces and exploration noise.** Joint-Space perturbs joints independently, which can produce unnatural hand configurations. Eigen-Space encourages coordinated motion but constrains actions to a learned PCA subspace. Eigen-Residual restores full joint-space expressivity by combining joint and eigen actions, but it introduces redundancy in the action parameterization. EigenDEXplore combines joint and eigen noise while retaining a joint-space action representation.

## 1 INTRODUCTION

Dexterous hands enable versatile grasping, in-hand manipulation, and precise interaction with tools and articulated objects. Discovering these contact-rich behaviors, however, is challenging because effective manipulation relies on highly coordinated finger motions that occupy only a small fraction of the hand's high-dimensional action space. Reinforcement learning (RL) and sampling-based trajectory optimization commonly explore this space using independent Gaussian perturbations across



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

joints, making it difficult to discover such coordinated behaviors efficiently. As a result, these methods often require substantial computation and carefully engineered rewards, costs, or curricula.

Learning from human hand motion offers a way to capture useful coordination patterns, providing guidance that would otherwise need to be designed for each task. Prior work captures these patterns in low-dimensional action spaces, including linear eigen-spaces learned using PCA (Ciocarlie et al., 2007; Agarwal et al., 2023) and nonlinear latent spaces learned using VAEs (Dimou et al., 2021; 2024). These spaces let the policy directly command coordinated movements such as grasping and twisting (see Fig. 2). Sampling actions corresponding to these modes during RL exploration can also provide more informative feedback per sample for policy updates and accelerate learning (Agarwal et al., 2023; Lum et al., 2024). However, these reduced action spaces cannot express the independent joint control required for general manipulation (Gupta et al., 2026a). Another approach combines these learned and independent joint-space actions to restore expressivity (Gupta et al., 2026b), but increases the action dimensionality and introduces redundancy. It remains unclear whether the benefits come from directly controlling coordinated motions, the noise-induced exploration, or changes to action-space dimensionality and expressivity.

To disentangle these effects, we systematically compare action representations and exploration distributions across diverse manipulation tasks and robot hands. *These comparisons suggest that the primary benefit of human-motion priors comes from correlated exploration, while action constraints or redundancy can hinder performance.* Joint-space actions already allow the policy to express both coordinated and independent joint movements. The difficulty is discovering useful joint combinations, which correlated noise makes easier to sample. Motivated by this finding, we introduce **EigenDEXplore**, a simple method that incorporates this human-motion prior into the sampling distribution while retaining the original joint-space actions. It combines two noise components: perturbations along the learned PCA directions encourage coordinated motion, while independent joint-space noise maintains exploration coverage. The prior is derived from hand motion recovered from large-scale human video data (Lightwheel, 2026), without requiring task-specific trajectories. After RL training, the eigenbasis can be discarded and the policy deployed deterministically.

We compare EigenDEXplore with joint-space and learned eigen-space baselines across unstructured RL with SimToolReal (Kedia et al., 2026) and DeXtreme (Handa et al., 2023), reference-guided bimanual RL with DexMachina (Mandi et al., 2026), and trajectory optimization with SPIDER (Pan et al., 2025), spanning six dexterous hands. EigenDEXplore consistently outperforms them, with 34% more consecutive successes in early SimToolReal training, larger gains in extended runs, and 40% more in DeXtreme. These gains transfer to hardware, raising SimToolReal average progress across 4 tasks from 33% to 82% and DeXtreme consecutive successes by 41%. In DexMachina, it improves average object-tracking accuracy by 28% without contact reward shaping or curricula, with larger gains on higher-DoF hands, enabling learning where joint-space sampling struggles or fails. In SPIDER, it reduces retargeting cost by 16% on average across six hands, again most on higher-DoF hands. We also find that large-scale egocentric video (Lightwheel, 2026) yields stronger human-motion priors than smaller, task-focused motion-capture datasets (Taheri et al., 2020; Fan et al., 2023), and validate our choice of retained PCA dimensionality through exploration-distribution experiments.

In summary, our main contribution is EigenDEXplore, a method that guides exploration with human-derived coordination patterns while preserving joint-space actions. We identify correlated exploration as the primary source of improvement and demonstrate the method's effectiveness across reinforcement learning, trajectory optimization, and sim-to-real deployment. We release benchmarking code, a plug-and-play implementation, and pretrained PCA models for six dexterous hands.

## 2 Related Work

**Human priors for manipulation.** Human data can guide dexterous manipulation by providing examples of useful behavior and desired outcomes. Demonstrations and pose priors support learning through policy supervision (Rajeswaran et al., 2018; Qin et al., 2022; Guzey et al., 2025), initialization (Bauza et al., 2025; Dasari et al., 2023), and reward design (Mandikal & Grauman, 2021). Hand and object trajectories provide task objectives for RL (Lum et al., 2025; Gupta et al., 2026a; Mandi et al., 2026; Xu et al., 2025; Li et al., 2025; Liu et al., 2025; Zhao et al., 2026; Chen et al., 2025) and trajectory optimization (Pan et al., 2025; Paliwal et al., 2026; Liu et al., 2024), allowing

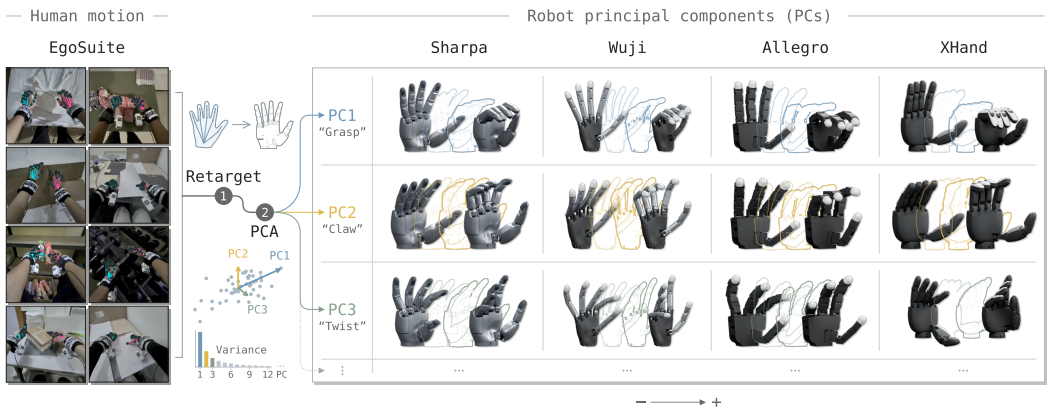


Figure 2: **Constructing robot principal components from human motion.** We retarget EgoSuite human hand motion to each robot hand and apply PCA in its joint space. We visualize the first three principal components for Sharpa, Wuji, Allegro, and XHand.

robot motions to adapt to embodiment differences and contact dynamics. Procedurally trained controllers can also execute human-derived goals at deployment (Yin et al., 2025; Kedia et al., 2026; Gupta et al., 2026b; Kuang et al., 2026). The success of task guidance from even a single video or demonstration motivates examining how larger human datasets can support learning across tasks.

**Action representations for dexterous manipulation.** Early studies showed that a few coordinated patterns capture much of the variation in human grasp postures (Santello et al., 1998), motivating eigen-space grasp planning (Ciocarlie et al., 2007; Ciocarlie & Allen, 2009). Encoding joint coordination in control representations supports RL and transfer in dexterous manipulation (Agarwal et al., 2023; Lum et al., 2024; Dimou et al., 2024; Yang et al., 2026; Yuan et al., 2025; Puang et al., 2025; Berg et al., 2024) and lowers trajectory-optimization cost for a fixed number of sampled trajectories (Yang et al., 2024). However, low-dimensional eigen-space control can restrict dexterity (Lum et al., 2024; Lee et al., 2026), including the fine adjustments needed to establish sufficient contact during tool manipulation (Gupta et al., 2026a). Adding eigen-space residuals to joint-space actions preserves full expressivity and can improve learning over joint-only control despite introducing redundant action coordinates (Gupta et al., 2026b). Reduced dimensionality therefore cannot explain all reported gains, raising the question of how these interfaces shape exploration.

**Structure in exploration.** These interfaces can turn independent action noise into correlated joint perturbations, confounding the benefits of coordinated commands and coordinated sampling. Temporal correlations (Raffin et al., 2021; Eberhard et al., 2023; Hollenstein et al., 2024) and policy-derived correlations across actuators (Chiappa et al., 2023) show how structuring exploration noise can improve learning efficiency. Human motion offers such structure, with the source data and learned representation determining which movements are encouraged. Existing priors often use grasp-focused recordings or small task-specific demonstration sets, with learned representations that vary in implementation and dimensionality (Lum et al., 2024; Dimou et al., 2024; Yuan et al., 2025; Yang et al., 2026). Our study separates the effects of human-motion priors on exploration from their use in the action representation, and examines how source data and prior dimensionality influence manipulation performance.

## 3 Method

EigenDEXplore uses human hand coordination to guide exploration while preserving native joint-space control. We retarget human motion to each robot hand and use PCA to extract a reusable basis of coordinated joint movements (Sec. 3.1). We then combine perturbations along these directions with independent joint-space noise (Sec. 3.2), sampling around the policy mean in reinforcement learning (Sec. 3.3) and the nominal control sequence in trajectory optimization (Sec. 3.4).

### 3.1 Human-Derived Coordination Prior

We use a dataset of human hand postures to construct a coordination prior for each robot hand. We pool postures from both hands by mirroring one hand into the other's coordinate frame, forming

a shared canonical representation. We adapt the kinematic retargeting procedure from (Li et al., 2025) to map human hand motion to each robot, with hand-specific alignment, scaling, and finger correspondences. We apply PCA to the retargeted joint angles to identify common patterns of joint coordination, with eigenvalue $\lambda_i$ measuring the variance captured by component $i$. To retain a chosen fraction of the human dataset variance, we select the first $k$ principal components needed to reach that threshold. The resulting $k$ depends on both the hand and dataset. For example, retaining 90% of the variance gives $k = 9$ for Sharpa (22 joints) and $k = 6$ for XHand (12 joints). We express the first $k$ directions in the policy's normalized joint-action coordinates (each joint scaled to $[-1, 1]$ by its limits) as the rows of the eigenbasis $E \in \mathbb{R}^{k\times d}$, where $d$ is the number of hand joints. Fig. 2 illustrates the leading directions across hands.

### 3.2 Action Spaces and Exploration Noise

We compare EigenDEXplore with three action formulations (Fig. 1): Joint-Space, which does not use $E$, as well as Eigen-Space and Eigen-Residual, which build $E$ into the action representation. These comparisons help disentangle the effect of incorporating human coordination into the action representation versus using it only to structure exploration. They also characterize the associated tradeoffs in action-space dimensionality, expressivity, and exploration structure (Sec. 4.1). Let $a_q \in \mathbb{R}^d$ and $a_e \in \mathbb{R}^k$ denote the joint- and eigen-space coordinates of a command before exploration. Let $q^\star \in \mathbb{R}^d$ denote the sampled absolute joint-position target in these normalized coordinates.

$$\begin{aligned}
&\text{Joint-Space } (q) && q^\star = a_q + \sigma_q \odot \eta_q && \text{action dim} = d \\
&\text{Eigen-Space } (e) && q^\star = E^\top(a_e + \sigma_e \odot \eta_e) && \text{action dim} = k \\
&\text{Eigen-Residual} && q^\star = a_q + \sigma_q \odot \eta_q + E^\top(a_e + \sigma_e \odot \eta_e) && \text{action dim} = d + k \\
&\text{EigenDEXplore} && q^\star = a_q + \sigma_q \odot \eta_q + E^\top(\sigma_e \odot \eta_e) && \text{action dim} = d
\end{aligned} \tag{1}$$

Here, $\eta_q \sim \mathcal{N}(0, I_d)$ and $\eta_e \sim \mathcal{N}(0, I_k)$ are independent standard Gaussian vectors, $\sigma_q \in \mathbb{R}^d$ and $\sigma_e \in \mathbb{R}^k$ are perturbation standard deviations, and $\odot$ denotes elementwise multiplication. We initialize $\sigma_{e,i} \propto \sqrt{\lambda_i}$, the standard deviation of the human data along component $i$, so eigen noise initially follows the spread of human postures. For EigenDEXplore, we rescale $\sigma_q$ and $\sigma_e$ so that, summed over joints, the initial noise variance matches Joint-Space, split equally between the two.

### 3.3 EigenDEXplore for Reinforcement Learning

Given an observation $o$, the policy network predicts $\mu_q = \pi_\theta(o)$, which we use as the joint-space command $a_q$. PPO (Schulman et al., 2017) learns the policy parameters $\theta$ and the noise standard deviations $\sigma_q$ and $\sigma_e$, while $E$ remains fixed. Conditioned on the observation $o$, the EigenDEXplore sampling rule in Eq. (1) is equivalently written as

$$q^\star \sim \mathcal{N}(\mu_q, \Sigma), \qquad \Sigma = \operatorname{diag}(\sigma_q^2) + E^\top \operatorname{diag}(\sigma_e^2) E. \tag{2}$$

The covariance $\Sigma$ adds the independent joint-noise covariance to the covariance of the projected eigen noise. PPO evaluates action log probabilities and entropy under this distribution. At deployment, we execute $\mu_q$ deterministically and discard $E$ and the exploration parameters.

### 3.4 EigenDEXplore for Trajectory Optimization

In trajectory optimization, exploration refines a sequence of joint targets rather than learning a policy. In SPIDER (Pan et al., 2025), these sequences are optimized to track reference hand and object motion. EigenDEXplore leaves SPIDER's joint-space trajectory representation and objective unchanged, and modifies only the perturbations used to sample candidate trajectories. At each iteration, the optimizer generates $N$ candidate sequences by perturbing the current joint target $q_t$ at each time $t$ with the EigenDEXplore noise in Eq. (1), using $q_t$ in place of $a_q$. It evaluates each candidate trajectory in simulation, keeps the lowest-cost 10% of trajectories, indexed by $\mathcal{S}$, and updates each target by adding a weighted average of the selected perturbations,

$$q_t \leftarrow q_t + \sum_{j\in\mathcal{S}} w_j \left[\sigma_q \odot \eta_{q,t}^{(j)} + E^\top\left(\sigma_e \odot \eta_{e,t}^{(j)}\right)\right]. \tag{3}$$

Here, $\eta_{q,t}^{(j)}$ and $\eta_{e,t}^{(j)}$ are the Gaussian noise vectors of candidate $j$ at time $t$, scaled by SPIDER's noise schedule, while $\sigma_q$ and $\sigma_e$ remain fixed. The weights $w_j$ sum to one, with larger weights for lower-cost candidates.

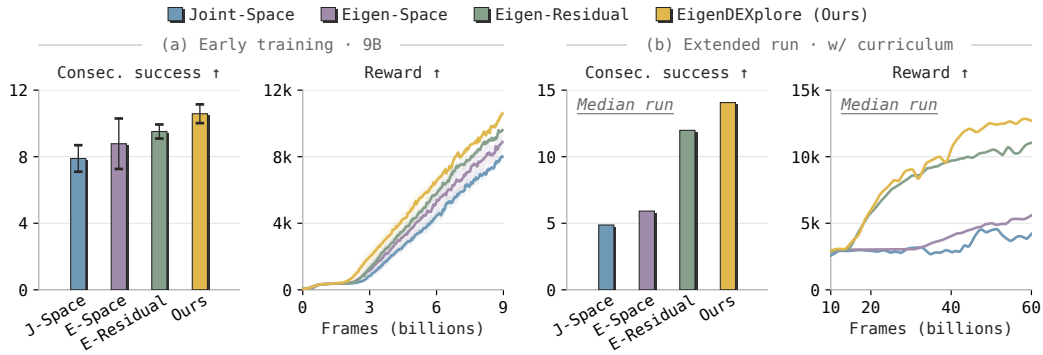


Figure 3: **SimToolReal.** EigenDEXplore achieves the highest mean consecutive successes and reward in dexterous tool manipulation. (a) Fixed-budget performance at 9B frames. Error bars show 95% CIs and shading shows $\pm 1$ SE across five seeds. (b) Extended training with a success-tolerance curriculum using one median run per method.

# 4 Experiments

## 4.1 Action spaces and exploration

We compare all four formulations (Sec. 3.2) in SimToolReal and DeXtreme over five seeds under each benchmark's protocol, reporting consecutive successes (number of object-pose goals reached).

### 4.1.1 SimToolReal

SimToolReal (Kedia et al., 2026) trains a Sharpa hand-arm policy to manipulate procedurally generated tools toward target poses. Following its original ablation budget, we first train for 9B frames with a fixed success tolerance of 6 cm and $50°$, emphasizing grasping and coarse reorientation. Eigen-Space and Eigen-Residual improve over Joint-Space, while EigenDEXplore achieves the highest reward and a mean of 10.6 consecutive successes vs. 7.9 for Joint-Space (Fig. 3a).

We then select each method's median run by consecutive successes at 9B frames and continue it to 60 billion total frames with a curriculum that tightens the tolerance toward 1 cm, $8°$, requiring finer manipulation. We use one run per method because each extended run takes $\sim$1 week. Under these tighter requirements, Eigen-Space's early advantage fades, while Eigen-Residual and EigenDEXplore sustain stronger learning, with EigenDEXplore reaching 14.1 consecutive successes by the end of training, compared with 12.0 for Eigen-Residual and 4.9 for Joint-Space (Fig. 3b).

### 4.1.2 DeXtreme

DeXtreme (Handa et al., 2023) trains an Allegro hand to repeatedly reorient a large cube toward randomly sampled target orientations. We train each variant for 2 billion environment frames with automatic domain randomization (ADR), which widens the ranges of observation noise, action delays, and physical properties as the policy meets a success threshold. The domain randomization (DR) level is measured in nats per dimension (NPD), the average log width of the randomized ranges. Higher values indicate further curriculum progression. Negative values correspond to widths below one.

Reorienting a large cube demands finger configurations that may be rare in human postures. Eigen-Space performs poorly, as it is confined to a 10-D subspace of Allegro's 16 joints. EigenDEXplore achieves the highest

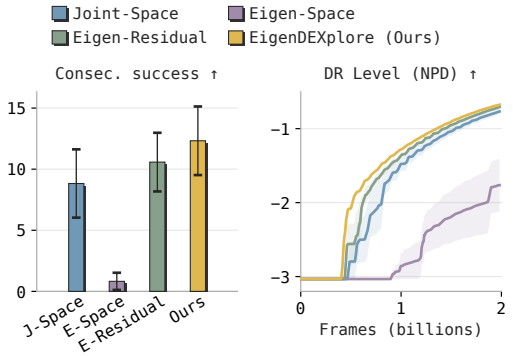


Figure 4: **DeXtreme.** EigenDEXplore achieves the highest mean consecutive successes for in-hand reorientation. **Left:** performance at 2B frames with DR Level fixed at $-0.5$. **Right:** DR Level during training. Error bars show 95% CIs and shading shows $\pm 1$ SE across five seeds.

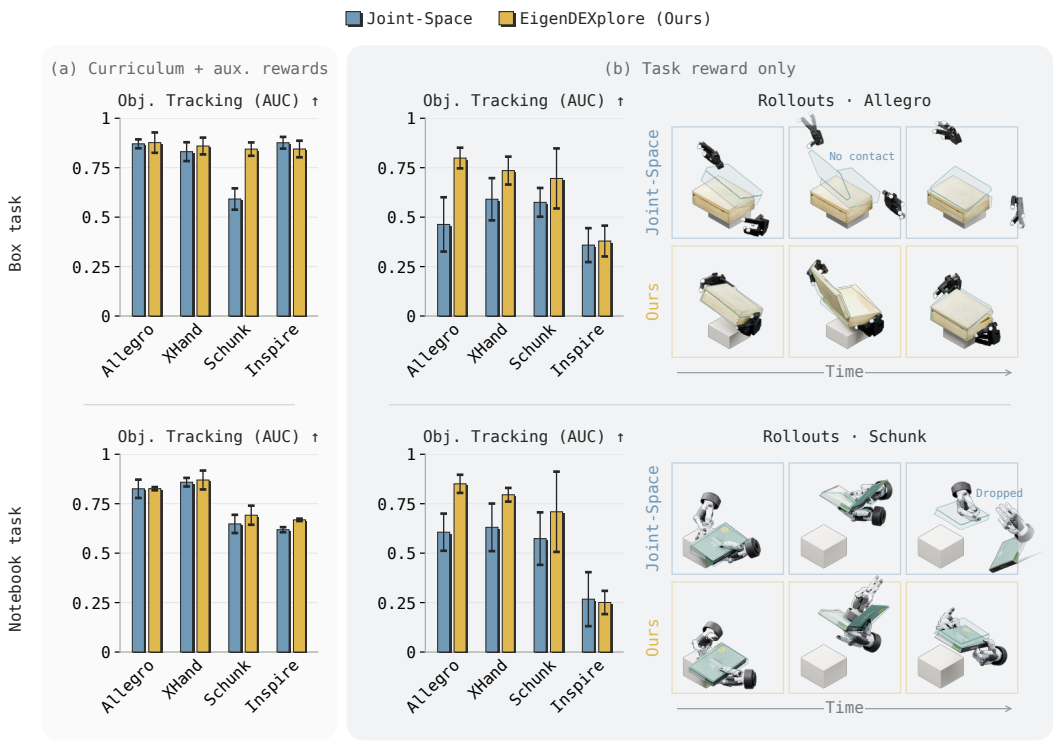


Figure 5: **DexMachina.** With curriculum and auxiliary rewards, EigenDEXplore performs similarly to Joint-Space in most settings (a). However, with task reward alone, EigenDEXplore substantially improves object tracking on higher-DoF Allegro and XHand (b, left). Representative task-reward-only rollouts (b, right). Bars show mean ADD-AUC with 95% CIs across five seeds.

mean consecutive successes under shared randomization (12.3 vs. 8.8 for Joint-Space) and progresses earlier through the DR curriculum (Fig. 4).

Across both benchmarks, EigenDEXplore improves over Joint-Space without changing the action representation, Eigen-Space struggles with more demanding manipulation, and Eigen-Residual trails EigenDEXplore slightly, consistent with a cost to expanding the action space. Additional delta-action variants show similar trends and are reported in App. A.

### 4.2 DexMachina

DexMachina (Mandi et al., 2026) trains bimanual policies to track object motions from ARCTIC demonstrations. To study how the benefit of EigenDEXplore varies with training assistance and hand embodiment, we compare it with Joint-Space on box and notebook manipulation using Allegro (16 DoF), XHand (12), Schunk (9), and Inspire (6), with five seeds and about 2B frames per configuration. We report ADD-AUC, which summarizes object-tracking accuracy across error thresholds. We evaluate two levels of training assistance: the full setup, whose auxiliary motion and contact rewards and virtual-controller curriculum initially guide the object along the demonstration before withdrawing, and the object-tracking task reward alone.

With curriculum and auxiliary rewards, EigenDEXplore performs similarly to Joint-Space in most settings, with a notable improvement for Schunk on the box task (0.59 to 0.84 ADD-AUC, Fig. 5a). This guidance already addresses much of the exploration difficulty, leaving less room for structured exploration to improve performance. With task reward alone, EigenDEXplore achieves substantially higher object-tracking accuracy than Joint-Space, with the largest gains on the higher-DoF Allegro and XHand, for example from 0.46 to 0.80 ADD-AUC on the Allegro box task, and little change on Inspire (Fig. 5b). In the representative rollouts, EigenDEXplore maintains stable contact while opening and closing both objects, whereas Joint-Space never touches the box with Allegro and drops the notebook mid-task with Schunk. The larger gains on higher-DoF hands are consistent with structured exploration helping most when coordinated actions must be discovered in a large

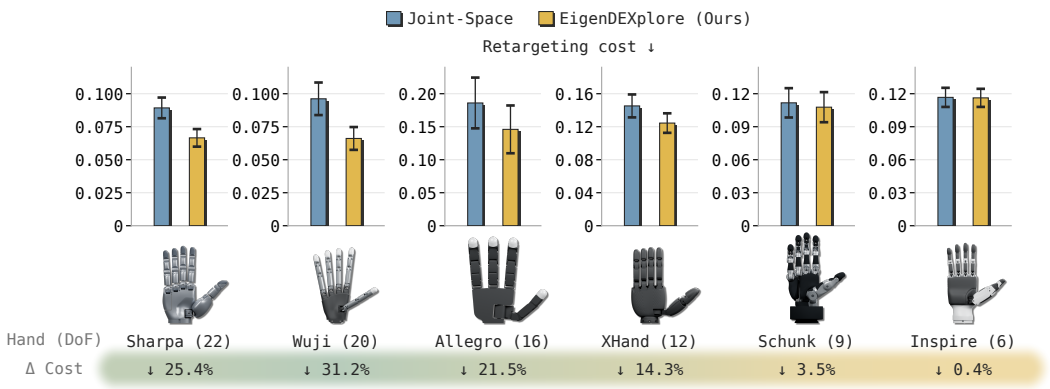


Figure 6: **SPIDER across hands.** EigenDEXplore reduces mean retargeting cost across all hands, with the largest reductions on higher-DoF hands and the smallest on lower-DoF hands. Δ cost (%) is relative to the Joint-Space baseline. Error bars show 95% CIs over ten seed-wise task averages.

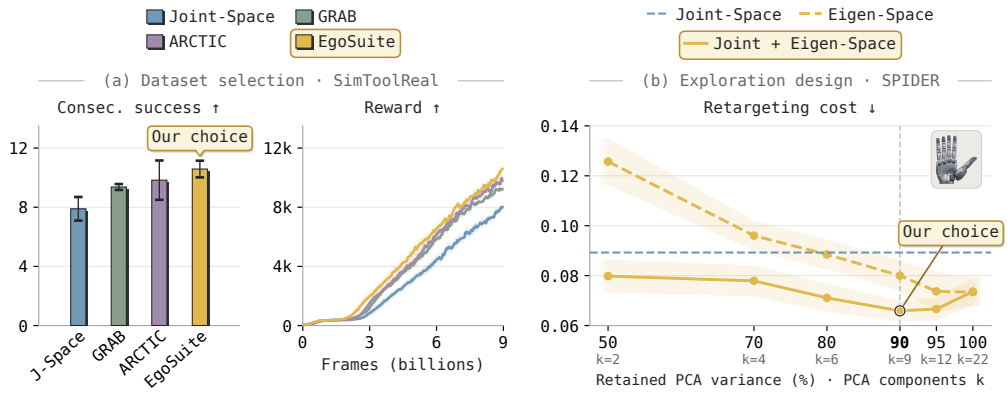


Figure 7: **Dataset and exploration analysis.** (a) Dataset selection: EgoSuite achieves the highest mean consecutive success and training reward in SimToolReal at 9B frames. (b) Exploration design: Sharpa SPIDER ablations vary the retained PCA variance (and thus the number of PCA components $k$) with and without joint-space noise. The selected 90% setting with joint-space noise achieves the lowest mean retargeting cost. Shading shows $\pm 1$ SE and error bars show 95% CIs.

search space without guidance. Lower-DoF hands such as Inspire are already highly coordinated, with three components capturing 90% of the variance, leaving less room for improvement.

### 4.3 SPIDER

We use SPIDER (Pan et al., 2025) to evaluate EigenDEXplore in sampling-based trajectory optimization across hand embodiments. We optimize ten object-motion tracking tasks from OakInk2 (Zhan et al., 2024) and compare Joint-Space with EigenDEXplore on six hands: Sharpa (22 DoF), Wuji (20), Allegro (16), XHand (12), Schunk (9), and Inspire (6). Both methods use the same tasks, a cold start from the same fixed initial hand pose, and a fixed budget of 16 optimizer updates per commit, by which most runs have plateaued, with ten seeds per task. We report the retargeting cost at the end of optimization, where lower is better. This cost is SPIDER's optimization objective for tracking the reference hand and object trajectory.

EigenDEXplore reduces mean retargeting cost on every hand, and the reduction grows with DoF (Fig. 6). The largest reductions are on the high-DoF Wuji (31%) and Sharpa (25%), followed by Allegro (22%) and XHand (14%), while costs on Schunk and Inspire are nearly unchanged (4% and 0.4%), mirroring the trend in DexMachina (Sec. 4.2). Because SPIDER keeps noise scales fixed rather than learned (Sec. 3.4), these gains come from the sampling distribution alone.

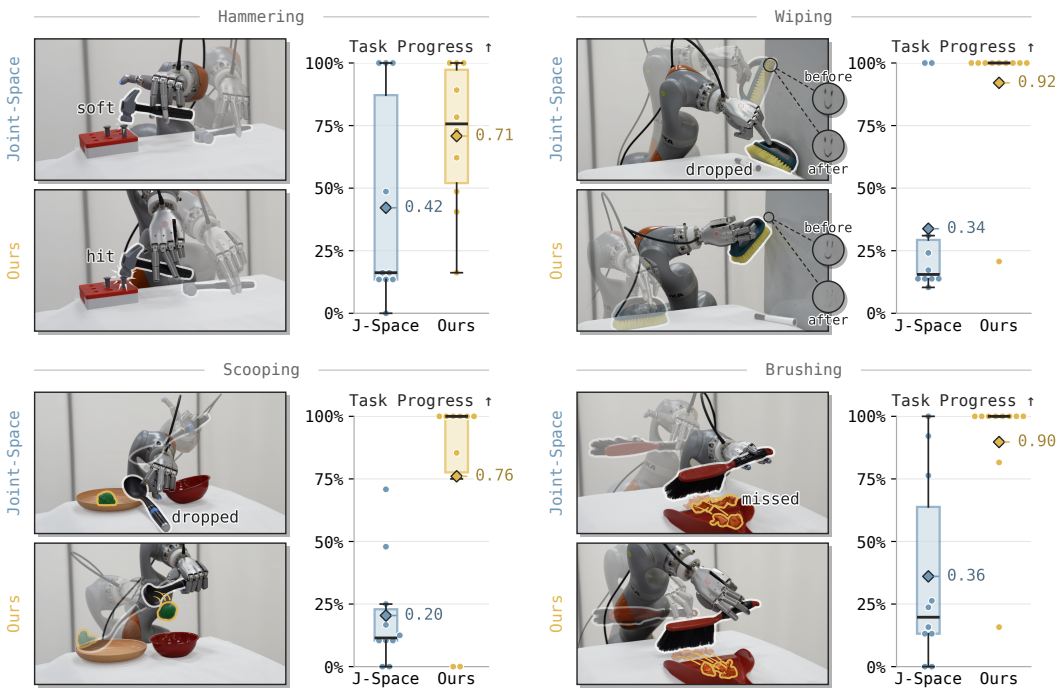


Figure 8: **Real-world SimToolReal.** EigenDEXplore achieves higher task progress on tool-use. For each task, representative rollouts (left) and task progress, the fraction of waypoints reached (right). Boxes show quartiles/medians, diamonds show means, and dots show individual trials.

### 4.4 Dataset and exploration analysis

We examine two design choices in EigenDEXplore: the source dataset used to construct the human-motion prior and the exploration configuration.

For dataset selection, we compare priors fitted to GRAB (Taheri et al., 2020), ARCTIC (Fan et al., 2023), and EgoSuite (Lightwheel, 2026) in SimToolReal, holding the exploration configuration fixed and following the protocol in Sec. 4.1.1. These datasets differ in scale and manipulation coverage: GRAB provides 1,334 motion-capture sequences involving 51 rigid objects, ARCTIC provides 339 sequences involving 11 articulated objects, and EgoSuite provides 100k hours of egocentric video spanning more than 15k tasks. All three priors improve over Joint-Space, and EgoSuite achieves the highest mean consecutive successes at 9B frames, 10.6 compared with 9.8 for ARCTIC, 9.4 for GRAB, and 7.9 for Joint-Space (Fig. 7a), supporting its use as the source dataset for EigenDEXplore.

For exploration design, we fix the prior to EgoSuite and sweep the retained-variance threshold $\rho \in \{0.50, 0.70, 0.80, 0.90, 0.95, 1.00\}$ on Sharpa, with and without independent joint-space noise, using the SPIDER setup in Sec. 4.3. This tests how strongly exploration should be concentrated in the human-motion subspace and whether exploration outside that subspace improves optimization. With eigen noise alone, retargeting cost decreases as more components are retained, and at $\rho = 0.50$, it is worse than Joint-Space (0.126 vs. 0.089), indicating that a narrow subspace restricts the search. Combining eigen noise with joint-space noise, as in EigenDEXplore, lowers cost at every threshold below $\rho = 1.00$, reaching a minimum of 0.066 at $\rho = 0.90$, 25% below Joint-Space (Fig. 7b). We therefore use this configuration in the dataset comparison and our main experiments.

These trends hold across GRAB, ARCTIC, and EgoSuite priors (App. B.1). We also find that replacing the human-derived basis with random orthonormal directions degrades performance (App. B.2).

### 4.5 Real-world evaluation

To test whether these improvements transfer to hardware, we compare EigenDEXplore with the Joint-Space baseline using the original SimToolReal and DeXtreme hardware protocols.

#### 4.5.1 SIMTOOLREAL

We evaluate the Joint-Space and EigenDEXplore policies from the extended-training comparison in Sec. 4.1.1 on four tool-use tasks: hammering, wiping, scooping, and brushing (Fig. 8). Both policies were trained for 60B frames, half of the 120B-frame budget used in the original SimToolReal training. We deploy them on the SimToolReal Sharpa hand-arm setup, using the same objects and trajectories as the original work (Kedia et al., 2026). These trajectories consist of 6D object-pose waypoints extracted from human RGB-D videos using FoundationPose (Wen et al., 2024). During execution, the policy advances to the next waypoint when the current goal pose is reached. We run ten trials per object-trajectory combination and report SimToolReal's Task Progress metric, the fraction of demonstrated waypoints reached. A trial ends at the final waypoint or on failure, such as losing the grasp or getting stuck.

EigenDEXplore achieves higher mean task progress than Joint-Space on all four tasks, averaging 82% across tasks compared with 33% for Joint-Space (Fig. 8). The Joint-Space policy typically pinches the tool between the pad of the middle finger and the backs of the index and ring fingers. It often requires several grasp attempts to lift the tool and struggles to reorient the tool in hand, instead relying on arm motion or dropping and regrasping. When the tool gets wedged deeply between the three fingers, the policy can complete the full trajectory, but the low friction on the backs of the fingers makes this grasp prone to slipping. In contrast, EigenDEXplore learns a more conventional fingertip grasp using the pads of the thumb, index, and middle fingers, with support from the side of the ring finger. This grasp is more stable and enables in-hand reorientation, allowing the policy to reach successive goal poses more smoothly. This qualitative difference is consistent with the hypothesis that eigen-structured exploration makes coordinated, human-like grasp configurations easier to discover than independent joint-space exploration. Overall, these results show that the simulation advantage observed during extended training (Sec. 4.1.1) transfers to hardware.

#### 4.5.2 DEXTREME

Each method uses one policy trained with the 2-billion-frame budget described in Sec. 4.1.2. A wrist-fixed Allegro hand repeatedly reorients a cube to successive target orientations (Fig. 9). We estimate cube pose from ArUco markers using AprilCube (Park & Agrawal, 2026). A trial ends when the cube is dropped or remains stuck in the same configuration for more than 80 seconds. Over 40 trials per method, EigenDEXplore increases mean consecutive successes from 8.45 to 11.90 (41% improvement) and the longest run from 32 to 41 (28% improvement).

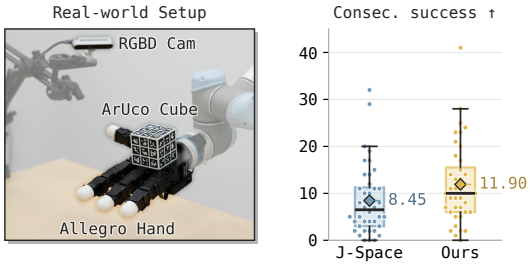


Figure 9: **Real-world DeXtreme.** EigenDEXplore increases mean consecutive successes from 8.45 to 11.90 over 40 trials per method. **Left:** hardware setup. **Right:** consecutive successes per trial. Boxes show quartiles/medians, diamonds show means, and dots show individual trials.

## 5 CONCLUSION

EigenDEXplore adds perturbations along human-derived eigenvectors to independent joint-space noise while retaining joint-space actions. Across RL, trajectory optimization, six dexterous hands, and real-world deployment, it outperforms joint-space and learned action-space baselines, with the largest gains on higher-DoF hands and under less reward shaping. These results suggest that human-motion priors help mainly by structuring exploration rather than changing the action space.

**Limitations.** EigenDEXplore uses a fixed linear basis derived from retargeted human motion, so coordination patterns absent from the data or nonlinear synergies must be explored through independent joint noise. Adapting the basis during training or using nonlinear priors could address this. The prior depends on kinematic retargeting and may be less effective for hands that differ substantially from human morphology. Improvements are also smaller for low-DoF hands. Finally, our noise is sampled independently at each timestep, and deriving temporally correlated exploration (Raffin et al., 2021; Eberhard et al., 2023) from human motion is a natural extension.

## AI USE STATEMENT

We used generative AI tools to assist with implementing our method and benchmarking code, polishing the writing, generating figures programmatically from our results, and surveying related work to complement our own literature review. The authors verified all code and experiments, reviewed all AI-assisted content, and take full responsibility for this paper, including any text, claims, or artifacts produced with AI assistance.

## REFERENCES


Ananye Agarwal, Shagun Uppal, Kenneth Shaw, and Deepak Pathak. Dexterous functional grasping. In *Conference on Robot Learning (CoRL)*, pp. 3453–3467, 2023. URL https://proceedings.mlr.press/v229/agarwal23a.html.

Maria Bauza, Jose Enrique Chen, Valentin Dalibard, Nimrod Gileadi, Roland Hafner, Murilo F. Martins, Joss Moore, Rugile Pevceviciute, Antoine Laurens, Dushyant Rao, Martina Zambelli, Martin Riedmiller, Jon Scholz, Konstantinos Bousmalis, Francesco Nori, and Nicolas Heess. DemoStart: Demonstration-led auto-curriculum applied to sim-to-real with multi-fingered robots. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6756–6763, 2025. URL https://doi.org/10.1109/ICRA55743.2025.11127813.

Cameron Berg, Vittorio Caggiano, and Vikash Kumar. SAR: Generalization of physiological agility and dexterity via synergistic action representation. *Autonomous Robots*, 48(8):28, 2024. URL https://doi.org/10.1007/s10514-024-10182-4.

Zerui Chen, Shizhe Chen, Etienne Arlaud, Ivan Laptev, and Cordelia Schmid. ViViDex: Learning vision-based dexterous manipulation from human videos. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3336–3343, 2025. URL https://doi.org/10.1109/ICRA55743.2025.11127358.

Alberto Silvio Chiappa, Alessandro Marin Vargas, Ann Huang, and Alexander Mathis. Latent exploration for reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 56508–56530, 2023. URL https://doi.org/10.52202/075280-2466.

Matei Ciocarlie, Corey Goldfeder, and Peter Allen. Dexterous grasping via eigengrasps: A low-dimensional approach to a high-complexity problem. In *Robotics: Science and Systems (RSS) Workshop on Robot Manipulation: Sensing and Adapting to the Real World*, 2007. URL https://projects.csail.mit.edu/manipulation/rss07/paper__dexterous_grasping_via_eigengrasps_a_low_dimensional_approach_to_a_high_complexity_problem__ciocarlie.pdf.

Matei T. Ciocarlie and Peter K. Allen. Hand posture subspaces for dexterous robotic grasping. *The International Journal of Robotics Research*, 28(7):851–867, 2009. URL https://doi.org/10.1177/0278364909105606.

Sudeep Dasari, Abhinav Gupta, and Vikash Kumar. Learning dexterous manipulation from exemplar object trajectories and pre-grasps. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3889–3896, 2023. URL https://doi.org/10.1109/ICRA48891.2023.10161147.

Dimitrios Dimou, José Santos-Victor, and Plinio Moreno. Learning conditional postural synergies for dexterous hands: A generative approach based on variational auto-encoders and conditioned on object size and category. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4710–4716, 2021. URL https://doi.org/10.1109/ICRA48506.2021.9560818.

Dimitris Dimou, José Santos Victor, and Plinio Moreno. Task-oriented grasping for dexterous robots using postural synergies and reinforcement learning. In *IEEE International Conference on Robotic Computing (IRC)*, pp. 65–71, 2024. URL https://doi.org/10.1109/IRC63610.2024.00016.

Onno Eberhard, Jakob Hollenstein, Cristina Pinneri, and Georg Martius. Pink noise is all you need: Colored noise exploration in deep reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2023. URL https://openreview.net/forum?id=hQ9V5QN27eS.

Zicong Fan, Omid Taheri, Dimitrios Tzionas, Muhammed Kocabas, Manuel Kaufmann, Michael J. Black, and Otmar Hilliges. ARCTIC: A dataset for dexterous bimanual hand-object manipulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12943–12954, 2023. URL https://doi.org/10.1109/CVPR52729.2023.01244.

Harsh Gupta, Mohammad Amin Mirzaee, and Wenzhen Yuan. Grasp to act: Dexterous grasping for tool use in dynamic settings. *IEEE Robotics and Automation Letters*, 11(5):6288–6295, 2026a. URL https://doi.org/10.1109/LRA.2026.3677744.

Harsh Gupta, Guanya Shi, and Wenzhen Yuan. LUCID: Learning embodiment-agnostic intent models from unstructured human videos for scalable dexterous robot skill acquisition. *arXiv preprint arXiv:2606.11628*, 2026b. URL https://arxiv.org/abs/2606.11628.

Irmak Guzey, Haozhi Qi, Julen Urain, Changhao Wang, Jessica Yin, Krishna Bodduluri, Mike Lambeta, Lerrel Pinto, Akshara Rai, Jitendra Malik, Tingfan Wu, Akash Sharma, and Homanga Bharadhwaj. Dexterity from smart lenses: Multi-fingered robot manipulation with in-the-wild human demonstrations. *arXiv preprint arXiv:2511.16661*, 2025. URL https://arxiv.org/abs/2511.16661.

Ankur Handa, Arthur Allshire, Viktor Makoviychuk, Aleksei Petrenko, Ritvik Singh, Jingzhou Liu, Denys Makoviichuk, Karl Van Wyk, Alexander Zhurkevich, Balakumar Sundaralingam, and Yashraj Narang. DeXtreme: Transfer of agile in-hand manipulation from simulation to reality. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5977–5984, 2023. URL https://doi.org/10.1109/ICRA48891.2023.10160216.

Jakob Hollenstein, Georg Martius, and Justus Piater. Colored noise in PPO: Improved exploration and performance through correlated action sampling. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 12466–12472, 2024. URL https://doi.org/10.1609/aaai.v38i11.29139.

Kushal Kedia, Tyler Ga Wei Lum, Jeannette Bohg, and Karen Liu. SimToolReal: An object-centric policy for zero-shot dexterous tool manipulation. In *Robotics: Science and Systems (RSS)*, 2026. URL https://doi.org/10.15607/RSS.2026.XXII.151.

Yuxuan Kuang, Sungjae Park, Katerina Fragkiadaki, and Shubham Tulsiani. Dex4D: Task-agnostic point track policy for sim-to-real dexterous manipulation. *arXiv preprint arXiv:2602.15828*, 2026. URL https://arxiv.org/abs/2602.15828.

Jayjun Lee, Jessica Yin, Asif Rana, Nicholas Blauch, Sam Mady, Mohak Bhardwaj, Nima Fazeli, Nathan Ratliff, Karl Van Wyk, and Ankur Handa. ADEPT: Accelerating dexterity via pre-training and post-training using reinforcement learning. *arXiv preprint arXiv:2608.19182*, 2026. URL https://arxiv.org/abs/2608.19182.

Kailin Li, Puhao Li, Tengyu Liu, Yuyang Li, and Siyuan Huang. ManipTrans: Efficient dexterous bimanual manipulation transfer via residual learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6991–7003, 2025. URL https://doi.org/10.1109/CVPR52734.2025.00656.

Lightwheel. EgoSuite-Open100K: The largest fully-annotated open egocentric human dataset. Dataset release, 2026. URL https://lightwheel.ai/media/egosuite-open100k.

Xueyi Liu, Kangbo Lyu, Jieqiong Zhang, Tao Du, and Li Yi. Parameterized quasi-physical simulators for dexterous manipulations transfer. In *European Conference on Computer Vision (ECCV)*, pp. 164–182, 2024. URL https://doi.org/10.1007/978-3-031-73229-4_10.

Xueyi Liu, Jianibieke Adalibieke, Qianwei Han, Yuzhe Qin, and Li Yi. DexTrack: Towards generalizable neural tracking control for dexterous manipulation from human references. In *International Conference on Learning Representations (ICLR)*, pp. 85970–85994, 2025. URL https://openreview.net/forum?id=ajSmXqgS24.

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)*, pp. 3182–3211, 2024. URL https://proceedings.mlr.press/v270/lum25a.html.

Tyler Ga Wei Lum, Olivia Y. Lee, Karen Liu, and Jeannette Bohg. Crossing the human-robot embodiment gap with sim-to-real RL using one human demonstration. In *Conference on Robot Learning (CoRL)*, pp. 4418–4441, 2025. URL https://proceedings.mlr.press/v305/lum25a.html.

Zhao Mandi, Yifan Hou, Dieter Fox, Yashraj Narang, Ajay Mandlekar, and Shuran Song. DexMachina: Functional retargeting for bimanual dexterous manipulation. In *International Conference on Machine Learning (ICML)*, 2026. URL https://openreview.net/forum?id=YScdQhxHgS.

Priyanka Mandikal and Kristen Grauman. DexVIP: Learning dexterous grasping with human hand pose priors from video. In *Conference on Robot Learning (CoRL)*, pp. 651–661, 2021. URL https://proceedings.mlr.press/v164/mandikal22a.html.

Bhawna Paliwal, Haritheja Etukuru, William Liang, Pieter Abbeel, Nur Muhammad Mahi Shafiullah, and Jitendra Malik. Do as I do: Dexterous manipulation data from everyday human videos. *arXiv preprint arXiv:2606.19333*, 2026. URL https://arxiv.org/abs/2606.19333.

Chaoyi Pan, Changhao Wang, Haozhi Qi, Zixi Liu, Homanga Bharadhwaj, Akash Sharma, Tingfan Wu, Guanya Shi, Jitendra Malik, and Francois Hogan. SPIDER: Scalable physics-informed dexterous retargeting. *arXiv preprint arXiv:2511.09484*, 2025. URL https://arxiv.org/abs/2511.09484.

Younghyo Park and Pulkit Agrawal. AprilCube: 3D-printable fiducial targets for reliable 6-DoF pose estimation. Software, 2026. URL https://github.com/younghyopark/aprilcube.

En Yen Puang, Federico Ceola, Giulia Pasquale, and Lorenzo Natale. PCHands: PCA-based hand pose synergy representation on manipulators with $N$-DoF. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 475–482, 2025. URL https://doi.org/10.1109/Humanoids65713.2025.11203193.

Yuzhe Qin, Yueh-Hua Wu, Shaowei Liu, Hanwen Jiang, Ruihan Yang, Yang Fu, and Xiaolong Wang. DexMV: Imitation learning for dexterous manipulation from human videos. In *European Conference on Computer Vision (ECCV)*, pp. 570–587, 2022. URL https://doi.org/10.1007/978-3-031-19842-7_33.

Antonin Raffin, Jens Kober, and Freek Stulp. Smooth exploration for robotic reinforcement learning. In *Conference on Robot Learning (CoRL)*, pp. 1634–1644, 2021. URL https://proceedings.mlr.press/v164/raffin22a.html.

Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, Giulia Vezzani, John Schulman, Emanuel Todorov, and Sergey Levine. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. In *Robotics: Science and Systems (RSS)*, 2018. URL https://doi.org/10.15607/RSS.2018.XIV.049.

Marco Santello, Martha Flanders, and John F. Soechting. Postural hand synergies for tool use. *The Journal of Neuroscience*, 18(23):10105–10115, 1998. URL https://doi.org/10.1523/JNEUROSCI.18-23-10105.1998.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. URL https://arxiv.org/abs/1707.06347.

Omid Taheri, Nima Ghorbani, Michael J. Black, and Dimitrios Tzionas. GRAB: A dataset of whole-body human grasping of objects. In *European Conference on Computer Vision (ECCV)*, pp. 581–600, 2020. URL https://doi.org/10.1007/978-3-030-58548-8_34.

Bowen Wen, Wei Yang, Jan Kautz, and Stan Birchfield. FoundationPose: Unified 6D pose estimation and tracking of novel objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17868–17879, 2024. URL https://doi.org/10.1109/CVPR52733.2024.01692.

Sirui Xu, Yu-Wei Chao, Liuyu Bian, Arsalan Mousavian, Yu-Xiong Wang, Liang-Yan Gui, and Wei Yang. Dexplore: Scalable neural control for dexterous manipulation from reference-scoped exploration. In *Conference on Robot Learning (CoRL)*, pp. 2184–2199, 2025. URL https://proceedings.mlr.press/v305/xu25d.html.

Chenyu Yang, Davide Liconti, and Robert K. Katzschmann. VQ-ACE: Efficient policy search for dexterous robotic manipulation via action chunking embedding. *arXiv preprint arXiv:2411.03556*, 2024. URL https://arxiv.org/abs/2411.03556.

Xinye Yang, Zhiyuan Ma, Hongze Yu, Yuanpei Chen, Yaodong Yang, Xiaojie Chai, Xinlei Chen, and Chao Yu. LAMP: Latent motion prior-guided real-world learning for dexterous hand manipulation. *arXiv preprint arXiv:2607.06323*, 2026. URL https://arxiv.org/abs/2607.06323.

Zhao-Heng Yin, Changhao Wang, Luis Pineda, Francois Robert Hogan, Chaithanya Krishna Bodduluri, Akash Sharma, Patrick Lancaster, Ishita Prasad, Mrinal Kalakrishnan, Jitendra Malik, Mike Lambeta, Tingfan Wu, Pieter Abbeel, and Mustafa Mukadam. DexterityGen: Foundation controller for unprecedented dexterity. In *Robotics: Science and Systems (RSS)*, 2025. URL https://doi.org/10.15607/RSS.2025.XXI.103.

Haoqi Yuan, Bohan Zhou, Yuhui Fu, and Zongqing Lu. Cross-embodiment dexterous grasping with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, pp. 81413–81434, 2025. URL https://openreview.net/forum?id=twIPSx9qHn.

Xinyu Zhan, Lixin Yang, Yifei Zhao, Kangrui Mao, Hanlin Xu, Zenan Lin, Kailin Li, and Cewu Lu. OakInk2: A dataset of bimanual hands-object manipulation in complex task completion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 445–456, 2024. URL https://doi.org/10.1109/CVPR52733.2024.00050.

Hui Zhang, Julian Ferchow, Jie Song, and Mirko Meboldt. UniCross: Unified cross-skill dexterous manipulation synthesis. *arXiv preprint arXiv:2607.28198*, 2026. URL https://arxiv.org/abs/2607.28198.

Shuqi Zhao, Xinghao Zhu, Yuxin Chen, Chenran Li, Yichen Xie, Xiang Zhang, Mingyu Ding, and Masayoshi Tomizuka. DexH2R: Task-oriented dexterous manipulation from human to robots. *IEEE/ASME Transactions on Mechatronics*, 31(3):3202–3213, 2026. URL https://doi.org/10.1109/TMECH.2025.3641164.

# A Additional Action Spaces

The formulations in Eq. (1) command absolute joint targets. A common alternative integrates actions as increments onto the previous target (Zhang et al., 2026; Gupta et al., 2026b). We evaluate delta variants of Joint-Space and Eigen-Residual in this form,

$$\begin{aligned} &\text{Joint-}\Delta && q_t^\star = q_{t-1}^\star + a_q + \sigma_q \odot \eta_q && \text{action dim} = d \\ &\text{Eigen-Residual-}\Delta && q_t^\star = q_{t-1}^\star + a_q + \sigma_q \odot \eta_q + E^\top(a_e + \sigma_e \odot \eta_e) && \text{action dim} = d + k \end{aligned} \tag{4}$$

Here, $a_q$ and $a_e$ act as per-step increments, scaled to a fraction of each joint range and coefficient span, and $q_t^\star$ is clipped to the joint limits. Joint-$\Delta$ integrates independent joint increments as in Zhang et al. (2026), and Eigen-Residual-$\Delta$ also integrates increments along the eigenbasis. LUCID (Gupta et al., 2026b) integrates eigen-grasp coefficients with per-joint residuals and is an instantiation of Eigen-Residual-$\Delta$.

We train both variants under the SimToolReal and DeXtreme protocols of Sec. 4.1 (Fig. 10). They perform comparably to their absolute counterparts in SimToolReal but fall well behind in DeXtreme.

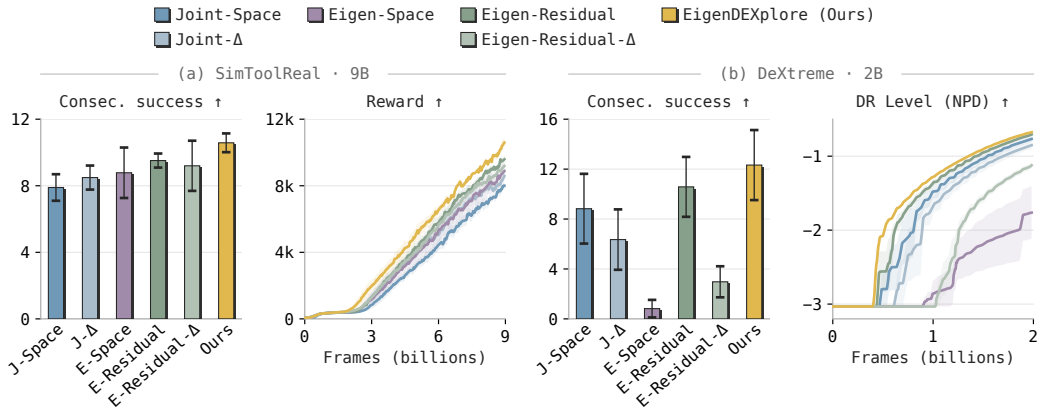


Figure 10: **Delta action spaces.** Delta variants of Joint-Space and Eigen-Residual perform comparably to their absolute counterparts in SimToolReal but fall behind them in DeXtreme. (a) SimToolReal at 9B frames, with consecutive successes (left) and reward (right). (b) DeXtreme at 2B frames, with consecutive successes at DR Level fixed at $-0.5$ (left) and DR Level during training (right). Error bars show 95% CIs and shading shows $\pm 1$ SE across five seeds.

# B Exploration Noise Analysis

## B.1 Source Dataset and Retained Variance

We repeat the exploration-design sweep of Sec. 4.4 with GRAB and ARCTIC priors in the same Sharpa SPIDER setup (Fig. 11). All three priors follow the EgoSuite trends. Eigen noise alone improves as more human variance is retained, and adding joint-space noise lowers cost across thresholds, with the lowest cost between 80% and 95% retention for all three priors. This supports retaining 90% of the PCA variance in EigenDEXplore.

## B.2 Human-Derived versus Random Directions

To test whether the benefit of EigenDEXplore comes from the human-derived directions or from correlation alone, we introduce a Random-Correlated baseline. It replaces the eigenvectors in $E$ with random orthonormal directions but keeps the EgoSuite eigenvalues at 90% retention as per-component variances. The eigenvalues are what correlate the noise, since equal variances along a full orthonormal basis would make it identical to joint-space noise. We evaluate it with eigen noise alone and combined with joint-space noise in the SPIDER setup of Sec. 4.3. In both settings, random directions perform worse than joint-space noise, while the human-derived priors match or improve on it (Fig. 12). Correlation alone therefore does not explain the benefit of EigenDEXplore, which relies on the coordination patterns in human motion.

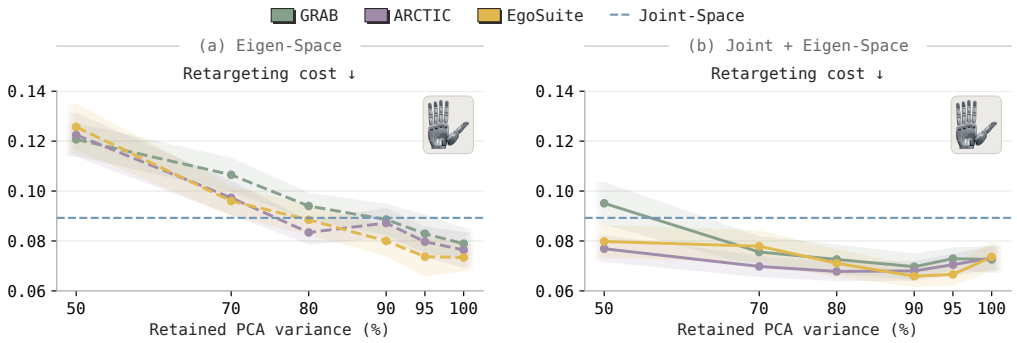


Figure 11: **Exploration design across datasets.** Priors fitted to GRAB, ARCTIC, and EgoSuite follow the same trends, so the benefit of combining eigen and joint-space noise does not depend on the source dataset. (a) With eigen noise alone, retargeting cost decreases as more components are retained. (b) Adding joint-space noise lowers cost at nearly every threshold. Dashed blue lines mark Joint-Space. Shading shows $\pm 1$ SE.

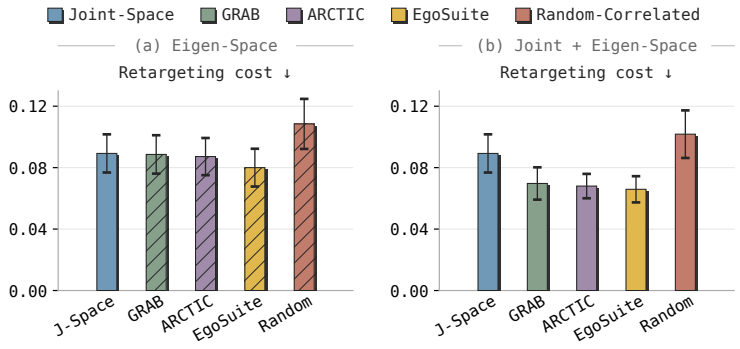


Figure 12: **Human-derived versus random directions.** The benefit of correlated exploration comes from the human-derived directions. Human-derived priors match or improve on Joint-Space, while random directions with the same eigenvalues are worse, both with eigen noise alone (a) and combined with joint-space noise (b). All eigen variants retain 90% of the variance. Error bars show 95% CIs over episodes.