Towards Data-Efficient VLA Post-Training: a case study of an industrial task

Anna Schooneveld, Geet Jose, Cuong Kasperzyk, Surabhi Rammurthy, IJ Smallman, Riccardo Secoli
Cambridge Consultants Ltd.
United Kingdom

***Abstract*—Warehouse and logistics environments expose the limitations of rigid automation, where infrequent but costly edge cases, such as misaligned or damaged packages, remain unresolved. Humanoid robots are increasingly viewed as a flexible solution, motivating box picking as a practical benchmark for real-world deployment. In this work, we explore whether a Vision-Language-Action (VLA) model can provide a reliable, specialist humanoid policy in the low-data regime characteristic of early industrial deployment. We post-train NVIDIA GR00T N1.5 on a box-picking task using only 60 teleoperated demonstrations. We introduce a pragmatic data collection strategy focused on clear behaviour decomposition and sufficient per-behaviour coverage, emphasising that careful dataset design, rather than size alone, is critical for reliable low-data deployment. Despite the limited dataset, the resulting policy achieves a 97.0% success rate and generalises to unseen box orientations and substantially different lighting conditions. We compare our approach to the Improved 3D Diffusion Policy (iDP3), a from-scratch diffusion-based humanoid model, trained on the same dataset. Unlike the VLA, iDP3 fails to reliably learn key behaviours, highlighting the advantage of large-scale VLA pre-training when adapting to new humanoid embodiments for bounded industrial tasks.**

***Index Terms*—Data-Efficient Learning, Low-data regime, Real-world evaluation, Humanoid Robots, Vision-Language-Action (VLA) Models, Industrial manipulation, New embodiment adaptation, Diffusion-based policies**

I. Introduction

Humanoid robots are rapidly approaching industrial viability, with the warehouse and logistics sector emerging as the primary early adopter. Because warehouses are built to human scale, deploying humanoids for picking, palletizing, and transport requires minimal infrastructure modification [1]. Their adoption is driven by severe labour constraints: order picking accounts for over 55% of operating costs [2], and manual handling roles are chronically understaffed [3]. Traditional automation struggles with case picking, frequently failing at edge cases like misaligned or damaged boxes. Analysts expect humanoids to resolve these limitations by deploying human-like flexibility where rigid automation breaks down [1], [4]. Therefore, box picking represents a natural, practically motivated benchmark for evaluating real-world humanoid performance.

To achieve flexible, generalising behaviour, policy architectures must move beyond rigid hand-crafted controllers. Vision-Language-Action (VLA) models [5]–[10] represent a highly promising frontier. These transformer-based architectures combine joint-state information with vision and language understanding via a Vision-Language Model (VLM) [11]–[14]. Because VLMs are typically pre-trained on web-scale data, and the full VLA is then further pre-trained on vast cross-embodiment robotics datasets, VLAs function as foundation models that can be post-trained for specific downstream applications. However, post-training a VLA on a new robot or task often still requires hundreds or thousands of demonstrations [9], [10]. This is a major bottleneck for industrial deployment, since data collection is expensive and slow. Many VLAs report strong zero-shot or few-shot performance on new tasks [5], [6], [8], [9], but results on unseen embodiments are less common. There are a few exceptions [6], [7], [10], although none of these discuss humanoids. The low-data regime is specifically discussed in [7] and [6], but success rates (63.3% and 70% respectively) remain moderate, which is promising for research but still below the $>$ 90% reliability required for many industrial applications. In industrial settings, the objective is a rapidly trainable specialist policy that can generalise robustly within a bounded task. As well as being small, the dataset must therefore also enable the model to learn the required behaviours reliably and to distinguish when each one should be triggered. How to achieve this in a low-data setting remains under-explored in the VLA literature. Studies such as [15] and [16] discuss practical suggestions for data collection, but only at the scale of many thousands of demonstrations. A few studies do discuss practical low-data design pointers [17], [18], but these focus on diffusion policies rather than post-training a pre-trained VLA.

In this paper we address the gap in low-data VLA post-training by presenting a case study in which we post-train the NVIDIA GR00T [9] for an *unseen* humanoid platform on a box-picking task. Using this task, we investigate how to structure a small training set so that all desired behaviours are reliably expressed. We also train the Improved 3D Diffusion Policy (iDP3) [19], a leading diffusion-based model for humanoids [20], [21], from scratch on the same dataset, and compare its performance to the VLA. Much current VLA research focuses on simulation benchmarks [22], [23], but because of the Sim2Real gap [24]–[26], strong performance in simulation does not always translate to a real robot. For this reason, all our training and evaluation is conducted on real robotic hardware.

Our key findings are as follows: (1) with just 60 training episodes, our best VLA model achieves a success rate of 97.0% on our box-picking task. (2) less training is more effective, although until the tails of the loss curve, the model is reasonably resistant to performance degradation from overfitting. (3) reliable learning of behaviours depends on both sufficient representation and sufficiently distinct triggering conditions; in our setting, six demonstrations were often not enough, and overlapping state contexts could cause behavioural interference. (4) Our model can generalise to box angles unseen in training and is fairly robust to lighting variation: performance drops only slightly (5%) under a very different lighting setup. (5) a dataset half the size does not suffice: with only 20–30 episodes, performance degrades by 33.3%–66.7% due to overfitting. (6) background randomisation does not prevent overfitting in our low-data setting. (7) the iDP3 trained on the same 60-episode dataset fails to learn most of the required behaviours.

II. Methods

A. Data Collection Setup

All experiments were conducted on a modified Unitree G1 humanoid robot. To improve grasp reliability, the factory end-effectors were replaced with custom 3D-printed hands. These provide passive compliance via sponge-like rubber contact points, which compensates for the lack of native impedance control in the position-commanded VLA. Visual data (RGB images for the VLA and point clouds for iDP3) were captured using an ORBBEC Gemini 215 camera on a custom head mount, as the standard RealSense D435 on the G1 does not provide point clouds of sufficient quality for iDP3 [19].

Demonstrations were collected via teleoperation using a motion capture system (Vicon Motion Systems Ltd., Oxford). Data acquisition and synchronisation were handled through a custom ROS-based software stack, which ensured temporally aligned recording of proprioceptive states, camera stream, and robot control commands.

B. Task Definition and Dataset

The box-picking task in this study imposes the following behavioural requirements:

- **Grasping:** the robot must successfully lift a box placed at any position within grasping reach, grasping the sides of the box with both hands.
- **Rotation invariance:** the robot must adapt its hand orientation to the box angle to grasp boxes across the full range of rotations encountered in deployment.
- **Absence handling:** the robot must not attempt to pick up in the absence of a box.

As an optional extra behaviour, we also considered the ability to pick up boxes from the edge of the table, outside of immediate grasping reach.

The post-training dataset contains 60 episodes in total, with its composition determined by the behavioural requirements of the task and structured to provide sufficient coverage of the required behaviours. For the core pickup task, 48 episodes cover two box positions (close to the robot and at the middle of the table), four nominal box orientations ($0°$ [landscape], $45°$, $90°$ [portrait], and $135°$; see Fig. 1), and three starting hand configurations (A, B, and C) positioned at varying distances

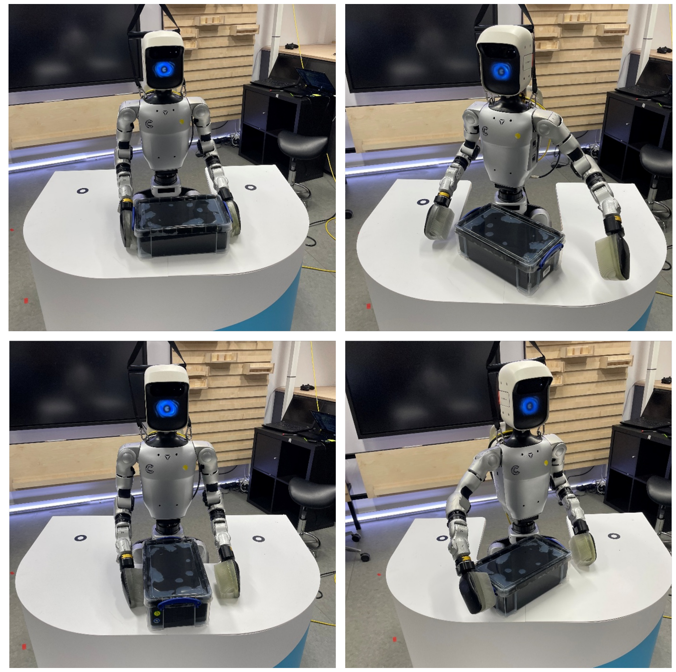

Fig. 1. The four box orientations and hand positions in the training set. Clockwise from top left: $0°$, $45°$, $90°$, $135°$.

from the box, with each combination repeated twice. An additional six episodes capture absence handling, explicitly demonstrating inaction to prevent the model from grasping if there is no box. Finally, the remaining six episodes target the optional edge-of-table behaviour, where the robot must first drag a distant box to within grasping reach before lifting. This behaviour is most distinct from the others, and was included to test whether it could be learned reliably with the same limited representation.

C. Training and Evaluation

To investigate the best way of post-training a VLA in a low-data setting, we post-trained various configurations of the NVIDIA GR00T N1.5 model on (subsets of) our dataset. For comparison, we also trained iDP3 from scratch. Training followed similar approach as in [9], [19] with the exception of epoch number for GR00T, which is discussed in Section III-A1.

We then evaluated these models in real world experiments using two evaluation types: basic and generalisation. All models underwent a basic evaluation, which assessed their ability to reproduce the behaviours from training. Some models were also assessed via a generalisation evaluation, which measures adaptation to unseen box orientations. Table I lists the test positions and orientations for each evaluation. We performed 10 repeats per test. Our evaluation pipeline included an automatic stop to detect successful picks based on joint effort and arm retraction. Visually correct picks outside these conditions were manually scored as successful. Picks that demonstrated the desired behaviours were scored as 1, whereas picks that did not exhibit all required behaviours (e.g., due to incorrect hand placement) were scored as 0.5. For 60° and 120°, we assigned a score of 1 to both a 90°-style pick and an "arms rotated" pick. Taking 60° as an example, this is because the training set does not include specific instructions for 60°, which is visually roughly halfway between 45° and 90°, so the model could interpret it as either.

III. Experiments and Results

A. How do we train the best model?

1) Less is more when it comes to epochs: To find the optimal training duration, we evaluated three model checkpoints with different numbers of epochs selected based on their position in the loss curve: *Short* (23.01 epochs, ~ 0.15 loss, just after the steep descent), *Middle* (32.75 epochs, ~ 0.1 loss, between descent and plateau), and *Long* (45.39 epochs, ~ 0.05 loss, far into the plateau). Results are shown under "Epoch tests" in Table I. The Short model performed best in the basic evaluation (98.3%), followed by the Middle (96.7%) and Long (84.2%) models. To rule out that the difference between Short and Middle is due to random fluctuations, we performed a generalisation evaluation, which again favoured the Short model (95.0% vs 90.0%). We therefore selected the Short model as the baseline. The results indicate that shorter post-training is preferable in our low-data regime. We hypothesise that shorter runs retain more of the pre-trained GR00T abilities, which are overwritten by overfitting during post-training. The small difference between the Short and Middle models suggests some robustness to modest overtraining, but performance clearly degrades for the Long model. This policy displayed unwanted behaviours, including failures at the no-box test, which indicate overfitting to joint states instead of the visual absence of a box.

2) Behaviours must be covered by a sufficient number of episodes: Neither of the three models could successfully pick up boxes in the far position. The robot did not extend its hands for the pulling motion shown in training, grabbed too close, and missed the box. This suggests that the six demonstrations allocated to this behaviour were not sufficient for the model to learn it reliably. The edge-of-table case required a more distinct motion pattern, therefore it likely benefited less from shared structure elsewhere in the dataset.

3) Behaviours must be sufficiently distinct: At an early stage of the study, we explored a "hand-over" behaviour by collecting six episodes starting from the terminal grasp pose (i.e. holding the box at an elevated position), in which the robot released the box when a person placed a hand over it. However, when these episodes were incorporated into the training set, the robot released the box at the end of every pick trajectory, regardless of the presence of a hand. Consequently, we excluded the hand-over requirement from the task. The failure mode likely arises because the terminal joint configuration of a standard pick trajectory is nearly indistinguishable from the initial state of a hand-over episode. In the low-data regime, such overlap can cause behavioural interference, underscoring the need to ensure that behaviours occupy sufficiently distinct regions of the state space.

B. Can our model generalise?

1) Our model can adapt to unseen angles: the generalisation evaluation in Table I shows that the baseline model can adapt to box angles unseen in the training dataset with a success rate of 95.0%, indicating that the policy is not merely memorising the four trained grasps, but can interpolate between them.

2) Our model can adapt to substantially different lighting: To assess robustness to lighting variation, we repeated the basic and generalisation evaluations after replacing the diffuse overhead tube lighting with yellow spotlights. This produced a darker scene with stronger shadows and altered specular reflections on the plastic box. Results are listed under "Lighting" in Table I. Performance drops by just 5% and most tests differ from the baseline by less than one point, indicating good robustness. The main degradations occur in the no-box and 60° tests. The 60° condition was already the most challenging under normal lighting, and the altered lighting likely exacerbates this. In the no-box scenario, we suspect the cause is overfitting; due to unfamiliar lighting the model possibly cannot rely on its standard visual cues for the absence of a box, and resorts to replicating the most common trajectory in the training set (namely a pick).

C. Ablation Studies on Dataset Size

To investigate whether the task can be learnt with even fewer episodes, we performed three ablation studies by removing subsets of the training data. In the first ablation, we trained a model (referred to as *One Rep*) with just one repeat per combination of box position, rotation, and starting hand position (30 episodes). The second (*Start A*) and third (*Start C*) models were trained on episodes with starting hand positions A and C respectively (20 episodes each); chosen because they are the most distinct. Training duration was chosen to match the baseline loss of ~ 0.15.

1) Effect of reducing training dataset: The "Data ablation" results in Table I show that all three methods substantially reduce performance (33.3%–66.7%). This is likely due to overfitting. For the One Rep and Start A models, failures consisted of the robot moving its hands to the pick-done position (i.e. the endpoint of a grasp-and-lift trajectory) without picking up. We hypothesise that with insufficient data, the model defaults to this endpoint because it is the most frequent position in the dataset. The Start C model exhibited a different kind of overfitting, performing the same grasp regardless of box presence or orientation. Which overfitting behaviour occurs is likely biased by whether the starting position is closer to a pick trajectory or the pick-done position.

D. Lessons learnt from earlier dataset iterations

The data collection strategy described in Section II-B was developed iteratively. One prior iteration warrants discussion, as it provided useful insights. This dataset contained 96 episodes, with six box positions, four orientations, two starting hand positions (A+B), and two repetitions per combination. The six positions expanded the workspace beyond the central

TABLE I
EVALUATION PERFORMANCE FOR VARIOUS MODELS AND CONDITIONS

	EPOCH TESTS			LIGHTING	DATA ABLATION			IDP3
	Baseline/Short	Middle	Long	Yellow Spotlights	One Rep	Start A	Start C	
Num epochs	**23.0**	32.8	45.4	23.0 (baseline)	25.2	39.22	37.74	300
Box middle	**10**	10	9	10	2	10	10	9
Box close 0°	**10**	9.5	10	10	0	0	10	10
Box close 45°	**9**	8.5	10	9.5	0	0	5	2
Box close 90°	**10**	10	10	10	9	0	10	9
Box close 135°	**10**	10	3.5	10	10	0	5	3
No box	**10**	10	8	7	10	10	0	0
Basic total	**59/60 (98.3%)**	**58/60 (96.7%)**	**50.5/60 (84.2%)**	**56.5/60 (94.2%)**	**31/60 (51.7%)**	**20/60 (33.3%)**	**40/60 (66.7%)**	**33/60 (55.0%)**
Box close 20°	**10**	10	-	10	-	-	-	4.5
Box close 60°	**8**	6	-	6.5	-	-	-	7.5
Box close 120°	**10**	10	-	9	-	-	-	10
Box close 160°	**10**	10	-	10	-	-	-	5
Generalisation total	**38/40 (95.0%)**	**36/40 (90.0%)**	-	**35.5/40 (88.8%)**	-	-	-	**27/40 (67.5%)**
Combined total	**97/100**	**94/100**	-	**92/100**	-	-	-	**60/100**

position to include left and right locations. The task was also more complex: each box was slid to the middle, rotated to 0°, and then picked up. In addition, background randomisation was introduced by varying table placement and surrounding objects between episodes.

1) More data is not always better if it increases task complexity: Despite the larger number of episodes, the model trained on this dataset exhibited overfitting. It failed to pick up boxes across most positions and instead performed a single grasp regardless of position (similar to the Start C model in Section III-C1). We hypothesise this is because the increased task complexity (specifically the addition of position-dependent sliding and more positions) was too difficult to learn from the limited dataset, similarly to Sections III-A2 and III-A3. This suggests that data quality and task simplicity are as important as dataset size, and that simpler, less ambiguous datasets may be more effective than increasing the number of episodes.

2) Background randomisation does not prevent overfitting in the low-data regime: Interestingly, the model could recover the correct behaviour when the background was set up to match that of a specific training episode, indicating overfitting to background cues. This shows that background randomisation alone is insufficient to prevent overfitting in low-data settings.

E. How does the VLA compare to iDP3?

To investigate the benefit of a VLA over alternative methods, we applied our 60-episode dataset to the diffusion-based iDP3 model and compared it to the VLA baseline. We trained for 300 epochs to be consistent with [19].

1) iDP3 does not learn many training behaviours: It is clear from Table I that iDP3 performs substantially worse than the VLA, and that it does not learn many behaviours in the training set, instead showing heavy overfitting. The model failed at the no-box test by attempting a pick, indicating overfitting to joint states and a lack of adaptation to visual cues. Furthermore, it did not execute the 45° and 135° grabs with the desired arm rotation. Together, these observations suggest that 60 episodes are simply not sufficient training data for iDP3 to learn the required behaviours reliably. For completeness, we also performed a generalisation evaluation. Perhaps unsurprisingly, given that it cannot perform rotations for angles in the training set, the model does not perform the correct motions for many of the unseen angles either.

IV. CONCLUSION

In this paper, we showed that a VLA can be adapted to a new humanoid embodiment for an industrially relevant box-picking task using only 60 teleoperated episodes. Our best GR00T model achieved a 97.0% success rate on the core box-picking task, and generalised to unseen box orientations and substantially different lighting conditions. Comparison with iDP3 further showed that pre-training is a major advantage in the low-data regime, as the non-pretrained iDP3 failed to learn several behaviours from the same dataset. This suggests that pre-trained VLAs provide a superior foundation for rapid deployment in industrial scenarios. The study also showed that success in the low-data regime depends strongly on dataset construction. Reliable performance requires clear separation of behaviours and sufficient representation of each one. When the dataset was reduced too far, some behaviours were not learned reliably; when it was enlarged in ways that increased complexity or irrelevant variation, learning also became less reliable. Overall, our findings indicate that reliable low-data deployment depends on balancing coverage against complexity, rather than increasing dataset size alone.

REFERENCES

[1] Supply Chain Management Review, “The value and limitations of humanoid robots in the ware-

house of the future," https://www.scmr.com/article/the-value-and-limitations-of-humanoid-robots-in-the-warehouse-of-the-future, September 2025, accessed: April 2026.

[2] R. de Koster, T. Le-Duc, and K. J. Roodbergen, *Design and control of warehouse order picking: A literature review*. Elsevier, 2007, vol. 182, no. 2.

[3] Vecna Robotics and CITE Research, "Warehouse automation statistics: The state of the market in 2023," Vecna Robotics, Tech. Rep., 2023. [Online]. Available: https://www.vecnarobotics.com/resources/warehouse-automation-statistics/

[4] Bain & Company, "Humanoid robots: From demos to deployment," Bain & Company, Tech. Rep., 2025, accessed: April 2026. [Online]. Available: https://www.bain.com/insights/humanoid-robots-from-demos-to-deployment-technology-report-2025/

[5] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," 2023. [Online]. Available: https://arxiv.org/abs/2307.15818

[6] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees *et al.*, "Octo: An open-source generalist robot policy," 2024. [Online]. Available: https://arxiv.org/abs/2405.12213

[7] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair *et al.*, "Openvla: An open-source vision-language-action model," 2024. [Online]. Available: https://arxiv.org/abs/2406.09246

[8] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn *et al.*, "π_0: A vision-language-action flow model for general robot control," 2024. [Online]. Available: https://arxiv.org/abs/2410.24164

[9] NVIDIA, J. Bjorck, F. Castañeda, N. Cherniadev, X. Da, R. Ding *et al.*, "Gr00t n1: An open foundation model for generalist humanoid robots," 2025. [Online]. Available: https://arxiv.org/abs/2503.14734

[10] G. R. Team, S. Abeyruwan, J. Ainslie, J.-B. Alayrac, M. G. Arenas, T. Armstrong *et al.*, "Gemini robotics: Bringing ai into the physical world," 2025. [Online]. Available: https://arxiv.org/abs/2503.20020

[11] Z. Li, G. Chen, S. Liu, S. Wang, V. VS, Y. Ji *et al.*, "Eagle 2: Building post-training data strategies from scratch for frontier vision-language models," 2025. [Online]. Available: https://arxiv.org/abs/2501.14818

[12] S. Karamcheti, S. Nair, A. Balakrishna, P. Liang, T. Kollar, and D. Sadigh, "Prismatic vlms: Investigating the design space of visually-conditioned language models," 2024. [Online]. Available: https://arxiv.org/abs/2402.07865

[13] P. Wang, S. Bai, S. Tan, S. Wang, Z. Fan, J. Bai *et al.*, "Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution," 2024. [Online]. Available: https://arxiv.org/abs/2409.12191

[14] L. Beyer, A. Steiner, A. S. Pinto, A. Kolesnikov, X. Wang, D. Salz *et al.*, "Paligemma: A versatile 3b vlm for transfer," 2024. [Online]. Available: https://arxiv.org/abs/2407.07726

[15] A. Khazatsky, K. Pertsch, S. Nair, A. Balakrishna, S. Dasari, S. Karamcheti *et al.*, "Droid: A large-scale in-the-wild robot manipulation dataset," 2025. [Online]. Available: https://arxiv.org/abs/2403.12945

[16] T. Z. Zhao, J. Tompson, D. Driess, P. Florence, K. Ghasemipour, C. Finn *et al.*, "Aloha unleashed: A simple recipe for robot dexterity," 2024. [Online]. Available: https://arxiv.org/abs/2410.13126

[17] A. Xie, L. Lee, T. Xiao, and C. Finn, "Decomposing the generalization gap in imitation learning for visual robotic manipulation," 2023. [Online]. Available: https://arxiv.org/abs/2307.03659

[18] J. Gao, A. Xie, T. Xiao, C. Finn, and D. Sadigh, "Efficient data collection for robotic manipulation via compositional generalization," 2024. [Online]. Available: https://arxiv.org/abs/2403.05110

[19] Y. Ze, Z. Chen, W. Wang, T. Chen, X. He, Y. Yuan *et al.*, "Generalizable humanoid manipulation with 3d diffusion policies," 2025. [Online]. Available: https://arxiv.org/abs/2410.10803

[20] S. Bai, W. Song, J. Chen, Y. Ji, Z. Zhong, J. Yang *et al.*, "Towards a unified understanding of robot manipulation: A comprehensive survey," Mon Oct 13 2025 01:59:27 GMT+0000 (Coordinated Universal Time). [Online]. Available: https://arxiv.org/abs/2510.10903

[21] C. Li, J. Wen, Y. Peng, Y. Peng, and Y. Zhu, "Pointvla: Injecting the 3d world into vision-language-action models," *IEEE Robotics and Automation Letters*, vol. 11, no. 3, pp. 2506–2513, 2026.

[22] B. Liu, Y. Zhu, C. Gao, Y. Feng, Q. Liu, Y. Zhu *et al.*, "Libero: Benchmarking knowledge transfer for lifelong robot learning," *arXiv preprint arXiv:2306.03310*, 2023.

[23] W. Wang, F. Wei, Q. Li, X. Chen, Y. Liang, C. Xu *et al.*, "Mobilemanibench: Simplifying model verification for mobile manipulation," Thu Feb 05 2026 02:49:52 GMT+0000 (Coordinated Universal Time). [Online]. Available: https://arxiv.org/abs/2602.05233

[24] X. Li, K. Hsu, J. Gu, O. Mees, K. Pertsch, H. R. Walke *et al.*, "Evaluating real-world robot manipulation policies in simulation," in *Proceedings of The 8th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, P. Agrawal, O. Kroemer, and W. Burgard, Eds., vol. 270. PMLR, 06–09 Nov 2025, pp. 3705–3728. [Online]. Available: https://proceedings.mlr.press/v270/li25c.html

[25] O. M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki *et al.*, "Learning dexterous in-hand manipulation," *Int. J. Rob. Res.*, vol. 39, no. 1, pp. 3–20, Jan. 2020.

[26] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3803–3810.