Improving VLA Workspace Randomization Robustness for Industrial Pick-and-Place via Lightweight Residual Reinforcement Learning

Siva Kailas*, Shalin Jain*, Harish Ravichandar
Georgia Institute of Technology
* denotes equal contribution

Abstract

Vision-language-action (VLA) models have recently demonstrated strong generalization across robotic manipulation tasks. However, many VLAs remain brittle to variations in initial workspace configurations, limiting their applicability in industrial settings. We investigate improving VLA robustness for industrial pick-and-place using lightweight residual reinforcement learning (RL). Instead of finetuning the base VLA, we learn a small residual multi-layer perceptron (MLP) policy that edits actions generated by a frozen VLA policy. We further investigate several design choices for stable residual learning, including curriculum-based domain randomization, network initialization, reward structure, and residual observation spaces. Using a representative industrial pipe sorting task in IsaacLab with the GR00T N1 VLA, we demonstrate that residual RL substantially improves robustness and behavioral diversity under increasing workspace randomization while requiring only a fractional increase in parameter count.

1. INTRODUCTION

Recent vision-language-action (VLA) models such as OpenVLA [6], RT-2 [14], MolmoAct [7], and GR00T [3] have demonstrated strong capabilities for general robotic manipulation. While these models exhibit strong inter-task generalization, recent studies [4, 5, 8] suggest that many VLAs remain brittle within a task when initial workspace configurations vary significantly. This intra-task brittleness presents a major challenge for deploying VLAs in industrial environments where workspace variation is unavoidable. Existing approaches often rely on collecting increasingly diverse teleoperation data, which does not scale efficiently and may still fail to adequately cover broader workspace distributions. Recent works such as [9] investigate RL finetuning of the base VLAs to improve this robustness, which can be computationally-expensive to do. Residual RL offers a parameter-efficient and training-efficient means of improving base model performance and robustness [2], but has not been characterized well for VLAs. As such, in this work, we investigate whether residual RL can improve VLA robustness without finetuning the underlying VLA itself. While residual RL has been investigated methodologically in works such as [12] to then distill the trajectories collected from trained residual RL experts back into the original VLA, our work is complementary. We are studying various design elements that can be used to instantiate a residual RL policy for a frozen VLA to determine which are useful for robustness.

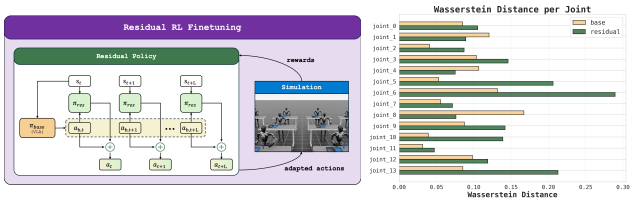

Figure 1. Left: Residual RL finetuning for action-chunked VLA policies. A lightweight residual policy edits actions from a frozen VLA policy. Right: improved VLA behavioral diversity under workspace randomization using the best studied design choices.

More specifically, we implement a lightweight residual policy that edits actions generated by a frozen VLA policy, and then to stabilize training, we additionally investigate curriculum-based domain randomization, reward shaping, residual observation design, and initialization strategies. Our experiments on an industrial pipe sorting task show that residual RL substantially improves robustness under increasing workspace randomization while adding only approximately 0.02% additional parameters.

2. METHOD

We consider a frozen base VLA policy π_{base} that generates action chunks $(a_{base,t}, \ldots, a_{base,t+L})$. A lightweight residual policy π_{res} produces corrective actions $a_t = a_{base,t} + a_{res,t}$. The residual policy is trained with PPO [11] and the base VLA remains frozen. The base VLA is queried every

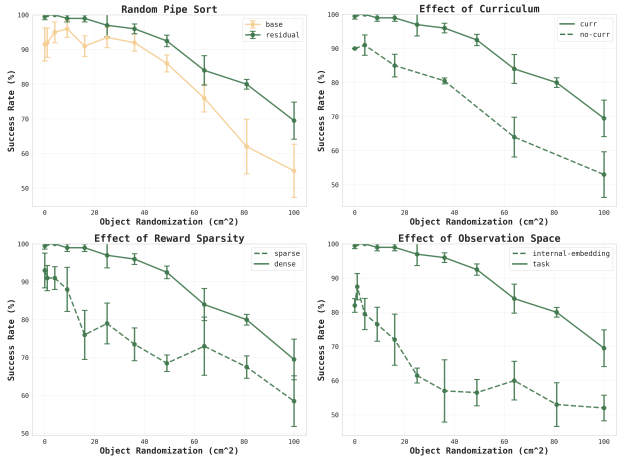

Figure 2. Left to right: residual RL under best studied design choices yields improved robustness over base VLA, curriculum learning improves robustness over no curriculum, dense rewards outperform sparse rewards for residual RL training, and distilled task observations outperform using VLA internal embeddings.

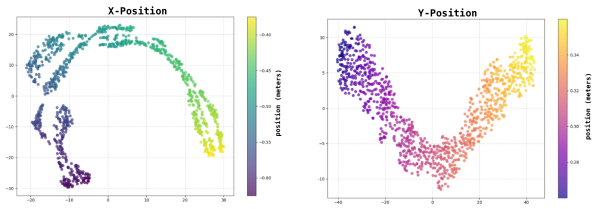

Figure 3. TSNE analysis of averaged VLM backbone embedding from GR00T N1 at varying x, y object positions in workspace. The embedding space seems to have structure despite averaging.

Network Initialization	Last Layer, Action Output Scale	Success Rate
Xavier Normal	Linear,0.1	60.00%
Kaiming Normal	TanH, 0.1	67.00%
Sparse	Linear, 0.1	88.00%
Ortho with Zero Last Layer	TanH, 0.1	85.00%

Table 1. Network Initialization Results at +/- 2cm Random Environment Initializations

L timesteps to generate an action chunk, and the residual policy edits the currently executed action at every timestep. To improve learning stability, we employ curriculum-based domain randomization inspired by [1]. The workspace randomization range is gradually increased once the policy exceeds a success threshold. Intuitively, this allows the residual policy to initially preserve the pretrained VLA behavior and progressively learn stronger corrections as task difficulty increases. We also investigate: (i) sparse vs. dense rewards, (ii) task observations vs. internal VLA embeddings, and (iii) residual network initialization strategies.

3. EXPERIMENTS

We evaluate our approach using the IsaacLab industrial `Pipe-Sort` task [10] with the GR00T N1 VLA [3]. The robot must pick up an exhaust pipe and place it into the correct bin using only first-person RGB observations and language instructions. Workspace randomization is introduced by perturbing the object spawn position across increasing $\Delta x, \Delta y$ ranges. Success is measured by whether the pipe is placed within the correct bin. Residual policies are trained using PPO with lightweight 2-layer 512-node MLPs.

4. RESULTS

Residual RL consistently improves robustness across increasing workspace randomization ranges (Figure 1). Figure 2 summarizes the primary ablation results. We additionally observe increased behavioral diversity, measured using Wasserstein distance between policy outputs under nominal and randomized workspace configurations. Curriculum learning substantially improves performance compared to directly training on large workspace randomization ranges. Dense rewards outperform sparse rewards, likely due to improved sample efficiency and more stable value estimation under PPO. We also find that residual policies conditioned on task-specific observations outperform those conditioned on internal VLA embeddings. While TSNE analysis indicates that averaged VLA embeddings preserve some workspace structure (Figure 3), the compressed representations appear insufficient for effective residual adaptation. Finally, initialization plays a critical role in training stability (see Table 1). Sparse initialization and orthogonal initialization with a zeroed final layer (both similar to [13]) significantly outperform standard Xavier and Kaiming initialization strategies. We hypothesize that initializing residual outputs near zero is important when using curriculum-based domain randomization, as early workspace configurations remain close to the pre-trained VLA distribution.

5. CONCLUSION

We present a lightweight residual RL framework for improving robustness of frozen VLA policies under workspace randomization in industrial pick-and-place settings. Our results demonstrate that residual action editing, under the right design choices, can improve robustness and behavioral diversity while requiring minimal additional parameters. We identify important design choices for stable residual learning, including curriculum-based domain randomization and near-zero residual initialization strategies. These findings suggest that lightweight residual adaptation may provide a practical mechanism for improving VLA robustness under industrial workspace distribution shift.

References

[1] Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, et al. Solving rubik's cube with a robot hand. *arXiv preprint arXiv:1910.07113*, 2019. 2

[2] Lars Ankile, Anthony Simeonov, Idan Shenfeld, Marcel Torne, and Pulkit Agrawal. From imitation to refinement-residual rl for precise assembly. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 01–08. IEEE, 2025. 1

[3] Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, Xingye Da, Runyu Ding, Linxi Fan, Yu Fang, Dieter Fox, Fengyuan Hu, Spencer Huang, et al. Gr00t n1: An open foundation model for generalist humanoid robots. *arXiv preprint arXiv:2503.14734*, 2025. 1, 2

[4] Senyu Fei, Siyin Wang, Junhao Shi, Zihao Dai, Jikun Cai, Pengfang Qian, Li Ji, Xinzhe He, Shiduo Zhang, Zhaoye Fei, et al. Libero-plus: In-depth robustness analysis of vision-language-action models. *arXiv preprint arXiv:2510.13626*, 2025. 1

[5] Jianing Guo, Zhenhong Wu, Chang Tu, Yiyao Ma, Xiangqi Kong, Zhiqian Liu, Jiaming Ji, Shuning Zhang, Yuanpei Chen, Kai Chen, et al. On robustness of vision-language-action model against multi-modal perturbations. *arXiv preprint arXiv:2510.00037*, 2025. 1

[6] 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. *arXiv preprint arXiv:2406.09246*, 2024. 1

[7] Jason Lee, Jiafei Duan, Haoquan Fang, Yuquan Deng, Shuo Liu, Boyang Li, Bohan Fang, Jieyu Zhang, Yi Ru Wang, Sangho Lee, et al. Molmoact: Action reasoning models that can reason in space. *arXiv preprint arXiv:2508.07917*, 2025. 1

[8] Hanqing Liu, Jiahuan Long, Junqi Wu, Jiacheng Hou, Huili Tang, Tingsong Jiang, Weien Zhou, and Wen Yao. Eva-vla: Evaluating vision-language-action models' robustness under real-world physical variations. *arXiv preprint arXiv:2509.18953*, 2025. 1

[9] Guanxing Lu, Wenkai Guo, Chubin Zhang, Yuheng Zhou, Haonan Jiang, Zifeng Gao, Yansong Tang, and Ziwei Wang. Vla-rl: Towards masterful and general robotic manipulation with scalable reinforcement learning. *arXiv preprint arXiv:2505.18719*, 2025. 1

[10] Mayank Mittal, Pascal Roth, James Tigue, Antoine Richard, Octi Zhang, Peter Du, Antonio Serrano-Muñoz, Xinjie Yao, René Zurbrügg, Nikita Rudin, et al. Isaac lab: A gpu-accelerated simulation framework for multi-modal robot learning. *arXiv preprint arXiv:2511.04831*, 2025. 2

[11] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. 1

[12] Wenli Xiao, Haotian Lin, Andy Peng, Haoru Xue, Tairan He, Yuqi Xie, Fengyuan Hu, Jimmy Wu, Zhengyi Luo, Linxi Fan, et al. Self-improving vision-language-action models with data generation via residual rl. *arXiv preprint arXiv:2511.00091*, 2025. 1

[13] Hongyi Zhang, Yann N Dauphin, and Tengyu Ma. Fixup initialization: Residual learning without normalization. In *International Conference on Learning Representations*. 2

[14] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, pages 2165–2183. PMLR, 2023. 1