

Abstract

- No parameter update; no auxiliary recovery module
- diagnostic analysis on 12 diverse RL Bench tasks
- 1,334 natural VLA failures, not synthetic data
- Studies failure-driven in-context robot learning

- We ask whether a VLA can use its own failed attempt as in-context evidence for a subsequent retry.
- Failure-conditioned context improves cumulative retry success over matched blind retry.
- The gains come from both a structural retry cue and content-dependent corrective language.
- Semantic re-grounding and contact-precision recovery remain difficult.

Methods and Materials

- RoboRetry**
 - We introduce ROBORETRY, a controlled probe.
$$a_{t:t+H-1} = \pi(o_t, s_t, l)$$

$$a_{t:t+H-1} = \pi(o_t, s_t, l, \mathcal{F}_{0:k-1})$$
- FailureSlot**
 - We introduce FAILURESLOT, a failure dataset with human corrective annotations with an 8-category taxonomy.
- Evaluation**
 - Instantiated on Pi0-FAST, evaluated on 12 RL Bench tasks. Paired retry protocol: Blind Retry v.s. Failure-Conditioned. Metric: SR@k = success within first k attempts.

Case Study

- The robotic arm closes too high. Continue lowering the height, and then align with the edge of the red cup. Rough spatial corrections ✓
- Move the arm to the right, aim at the left cross-shaped faucet handle, align it, and then close the gripper. Precision-sensitive ✗
- The gripper tried to open the middle drawer, which is the wrong target. Move up a layer to open the top drawer. Wrong target ✗

Conclusions

- Failure as context is useful for retry, improvements come from both **format following** and **description quality**.
- High-quality error-correcting languages are helpful, with **full-modal** contexts yielding the greatest benefit.
- Language modeling recovery is **limited** and still not adapt at complex recovery.

Does failure context improve retry?

Failure slots improve retry beyond blind retry.

How does it improve retry performance?

Retry lift combines a slot-format cue and content-dependent correction.

When is retry still insufficient?

- Works well: spatial corrections
- Still hard: discrete target, subgoal choice, precision-sensitive

A more native and actionable representation is needed for future

Contact / Resources

- Github: github.com/jedward225/RoboRetry
- Emails: jedward.jiajun@gmail.com, xumd@tsinghua.edu.cn

References

- James, S., Ma, Z., Arrojo, D. R., & Davison, A. J. (2020). Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2), 3019-3026.
- Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., & Zettlemoyer, L. (2022, December). Rethinking the role of demonstrations: What makes in-context learning work?. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 11048-11064).