FailGen: Frontier-Aware Semantic Data Synthesis for Embodied Manipulation

Zhirui Hu[1*] Sen Cui[2*] Jialin Tang[1] Jingheng Ma[2] Changshui Zhang[2] Shiji Zhou[1,3†] Xiao Zhang[1]
[1]School of Artificial Intelligence, Beihang University
[2]School of Automation, Tsinghua University
[3]Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing, Beihang University
hky0323@buaa.edu.cn, zhoushiji25@buaa.edu.cn

1. Introduction

Embodied agents require training environments that are both semantically diverse and physically valid, especially for long-horizon manipulation [5, 10, 11]. Existing data-generation pipelines have improved the scalability of robot learning data, but they often treat generation as either one-shot augmentation or open-ended task synthesis. As a result, generated tasks may be physically infeasible, already trivial, or weakly aligned with the current policy's failure modes. We present FailGen in Figure 1, a frontier-aware iterative data-generation framework for embodied manipulation. Unlike prior trajectory-centric augmentation methods [12] and open-ended generative simulation systems [19], FailGen explicitly couples data generation to the learner's evolving competence frontier. At each iteration, FailGen estimates the frontier from rollout statistics, prompts a VLM to propose semantic variations near this frontier, validates them with a hybrid solver combining motion planning and learned control, and consolidates successful demonstrations through replay-based behavioral cloning. Our main contribution is a policy-conditioned evolutionary data-generation paradigm that turns data synthesis from static dataset expansion into closed-loop capability expansion. Our preliminary results suggest that FailGen expands the policy's effective capability range and improves performance stability across challenging manipulation settings involving geometric obstruction, contact-rich interaction, and long-horizon logic.

2. Method

FailGen is a closed-loop frontier-aware iterative data-generation framework. Given a current policy, FailGen first identifies task variations near the policy's current competence boundary, then attempts to solve these tasks with a hybrid demonstration generator, and finally consolidates the generated demonstrations into the policy through replay-based behavioral cloning. This Propose–Solve–Learn loop

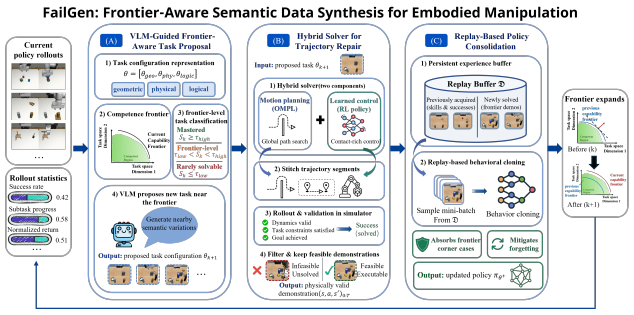


Figure 1. Overview of FailGen: a self-evolving loop of frontier-aware task generation, hybrid solver, and policy refinement for hard embodied manipulation.

is repeated to adapt the training distribution as the policy improves. We consider challenging embodied manipulation tasks whose difficulty arises from a combination of geometric constraints, physical interaction, and sequential structure [4, 11, 13]. We denote a task configuration by $\theta = [\theta_{geo}, \theta_{phy}, \theta_{logic}]$, where θ_{geo} captures geometric variations, θ_{phy} captures physical interaction requirements, and θ_{logic} captures long-horizon dependencies. In FailGen, θ serves as a compact interface between semantic task proposal and physical trajectory generation [2, 14, 15].

A. VLM-Guided Frontier-Aware Task Proposal: The central principle of FailGen is that curriculum generation should be conditioned on the evolving competence of the policy. Instead of decomposing a target task into a fixed sequence of subtasks, FailGen uses policy performance as feedback for proposing the next task configuration. At iteration k, the VLM is provided with the current task context and rollout statistics, and proposes a new configuration θ_{k+1} near the current failure boundary. We estimate the policy frontier with thresholded rollout statistics. Given the success rate or subtask-completion score s_k for configuration θ_k, FailGen applies three update rules: if $s_k \geq \tau_{\text{high}}$, the task is treated as mastered and difficulty is increased along one or more semantic dimensions; if $\tau_{\text{low}} < s_k <$

*Equal contribution.
†Corresponding author.

Table 1. Policy success rate across three difficulty levels.

Task	Geometric Obstruction			Contact-Rich Insertion			Long-Horizon Logic		
	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3	Level 1	Level 2	Level 3
Open-Loop (Static)	55.2	30.5	0.0	45.6	20.3	0.0	38.7	12.8	0.0
FailGen w/o Buffer	80.3	75.6	68.9	78.4	70.1	63.6	74.9	31.3	0.0
FailGen	**97.5**	**92.4**	**85.8**	**95.2**	**88.6**	**81.3**	**94.6**	**65.5**	**58.1**

τ_{high}, the task is considered frontier-level and nearby variations are generated to improve robustness; if $s_k \leq \tau_{\text{low}}$, the task is treated as rarely solvable, and the system backs off to a more granular intermediate variation [3]. This process keeps synthesized tasks near the current failure boundary, reducing trivial or unusable data generation [1].

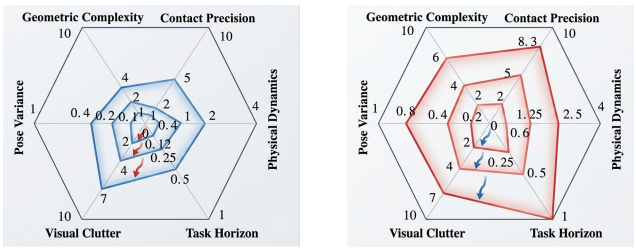
B. Hybrid Solver for Trajectory Repair: Task proposals produced by the semantic module may invalidate existing trajectories or require new interaction strategies [17]. FailGen therefore uses a hybrid solver to recover demonstrations when possible [18]. For globally constrained segments, the solver relies on RRT-Connect from OMPL to produce geometrically feasible motion skeletons [8]. For contact-rich segments, control is handed to a RL policy that adapts to physical interaction [7]. The segment-level trajectories are stitched sequentially, with each segment's terminal state serving as the next segment's initial state. The stitched trajectory is then rolled out in Robosuite/MuJoCo. This allows FailGen to filter infeasible proposals while preserving challenging but physically valid demonstrations.

C. Replay-Based Policy Consolidation: FailGen treats generated demonstrations as part of an evolving training distribution rather than as one-shot supervision. Each batch of newly solved frontier tasks is inserted into a persistent experience buffer together with demonstrations of previously acquired skills [16]. A reactive student policy is then trained with behavioral cloning over this accumulated dataset [6]. This replay-based consolidation serves two purposes. First, it allows the policy to absorb newly generated corner cases near its current failure boundary [9]. Second, it mitigates forgetting of earlier sub-skills as the curriculum shifts toward harder configurations. As the policy improves, the estimated frontier moves outward, providing the next round of task generation with a stronger base policy.

3. Results

We evaluate FailGen on manipulation tasks derived from the MimicGen dataset. All tasks are instantiated in Robosuite with the MuJoCo physics engine. For semantic curriculum generation, we use a VLM (Qwen-VL-Max) to analyze scene renderings, the current policy state, and target-task descriptions, and to generate task parameters.

Efficacy of VLM-Driven Curriculum: We examine whether VLM-guided task proposal leads to broader capability growth than non-adaptive task sampling in Fig 2. Visualizing performance across six difficulty dimensions reveals a stark contrast: while non-adaptive baselines exhibit a "spiky collapse", FailGen achieves a more isotropic expansion. These results suggest that conditioning task generation on the current policy frontier can make the synthesized curriculum more relevant to the agent's failure modes.



(a) Current Method (b) FailGen

Figure 2. The structural evolution of the capability boundary between our method (b) and current method (a).

Table 2. Ablation results on five tasks.

	Obs. Reach	Nut Assem.	Stack	Bin place	Obs. Assem.
Random Curriculum	0.18	0.23	0.45	0.33	0
Human-Designed	0.53	0.48	0.82	0.83	0.33
Planning-only	0.57	0	0	0	0
RL-only	0	0.51	0.71	0.64	0
FailGen	0.57	0.51	0.84	0.83	0.33

Solver Necessity: Ablation studies in Tab 2 confirm that both classical planning and learned control are indispensable. While planning-only ensures global geometric feasibility, it fails in contact-rich tasks. Conversely, RL-only handles local adaptation but struggles with complex geometric constraints. Only the full hybrid solver consistently succeeds across all task types, effectively bridging geometric reachability with physical interaction.

Benefits of closed-loop refinement: We compare three training paradigms to study the effect of closed-loop refinement in Tab 1. The static pipeline performs poorly on higher-difficulty configurations, suggesting that one-pass data generation is insufficient for bridging large capability gaps. The no-buffer variant degrades on long-horizon tasks, consistent with forgetting of earlier sub-skills. FailGen maintains stronger performance across difficulty levels, indicating that frontier-aware synthesis and replay-based consolidation are both important for stable capability growth.

4. Acknowledgments and Disclosure of Funding

This work is supported by the National Science Foundation of China (62401327).

References

[1] Andres Campero, Roberta Raileanu, Heinrich Küttler, Joshua B Tenenbaum, Tim Rocktäschel, and Edward Grefenstette. Learning with amigo: Adversarially motivated intrinsic goals. *arXiv preprint arXiv:2006.12122*, 2020. 2

[2] Michael Dennis, Natasha Jaques, Eugene Vinitsky, Alexandre Bayen, Stuart Russell, Andrew Critch, and Sergey Levine. Emergent complexity and zero-shot transfer via unsupervised environment design. In *Advances in Neural Information Processing Systems*, 2020. 1

[3] Sébastien Forestier, Rémy Portelas, Yoan Mollard, and Pierre-Yves Oudeyer. Intrinsically motivated goal exploration processes with automatic curriculum learning. *Journal of Machine Learning Research*, 23(152):1–41, 2022. 2

[4] Jiayuan Gu, Fanbo Xiang, Xuanlin Li, Zhan Ling, Xiqiang Liu, Tongzhou Mu, Yihe Tang, Stone Tao, Xinyue Wei, Yunchao Yao, et al. Maniskill2: A unified benchmark for generalizable manipulation skills. In *International Conference on Learning Representations*, 2023. 1

[5] Stephen James, Zicong Ma, David Rovick Arrojo, and Andrew J. Davison. Rlbench: The robot learning benchmark & learning environment. *IEEE Robotics and Automation Letters*, 5(2):3019–3026, 2020. 1

[6] Eric Jang, Alex Irpan, Mohi Khansari, Daniel Kappler, Frederik Ebert, Corey Lynch, Sergey Levine, and Chelsea Finn. Bc-z: Zero-shot task generalization with robotic imitation learning. In *conference on Robot Learning*, pages 991–1002. PMLR, 2022. 2

[7] Tobias Johannink, Shikhar Bahl, Ashvin Nair, Jianlan Luo, Avinash Kumar, Matthias Loskyll, Juan Aparicio Ojea, Eugen Solowjow, and Sergey Levine. Residual reinforcement learning for robot control. In *2019 international conference on robotics and automation (ICRA)*, pages 6023–6029. IEEE, 2019. 2

[8] Sertac Karaman and Emilio Frazzoli. Sampling-based algorithms for optimal motion planning. *The international journal of robotics research*, 30(7):846–894, 2011. 2

[9] Michael Laskey, Jonathan Lee, Roy Fox, Anca Dragan, and Ken Goldberg. Dart: Noise injection for robust imitation learning. In *Conference on robot learning*, pages 143–156. PMLR, 2017. 2

[10] Chengshu Li, Ruohan Zhang, Josiah Wong, Cem Gokmen, Sanjana Srivastava, Roberto Martín-Martín, Chen Wang, Gabrael Levine, Wensi Ai, Benjamin Martinez, et al. Behavior-1k: A human-centered, embodied ai benchmark with 1,000 everyday activities and realistic simulation. In *Conference on Robot Learning*, 2023. 1

[11] Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. In *Advances in Neural Information Processing Systems*, 2023. 1

[12] Ajay Mandlekar, Soroush Nasiriany, Bowen Wen, Iretiayo Akinola, Yashraj Narang, Linxi Fan, Yuke Zhu, and Dieter Fox. Mimicgen: A data generation system for scalable robot learning using human demonstrations. *arXiv preprint arXiv:2310.17596*, 2023. 1

[13] Oier Mees, Lukas Hermann, Erick Rosete-Beas, and Wolfram Burgard. Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks. *IEEE Robotics and Automation Letters*, 7(3): 7327–7334, 2022. 1

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

[15] Sharath Chandra Raparthy, Naomi Saphra, Grace Su, Mariano Phielipp, and Joelle Pineau. Generating automatic curricula via self-supervised active domain randomization. *arXiv preprint arXiv:2002.07911*, 2020. 1

[16] David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019. 2

[17] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011. 2

[18] David Silver, Satinder Singh, Doina Precup, and Richard S Sutton. Reward is enough. *Artificial intelligence*, 299: 103535, 2021. 2

[19] Yufei Wang, Zhou Xian, Feng Chen, Tsun-Hsuan Wang, Yian Wang, Katerina Fragkiadaki, Zackory Erickson, David Held, and Chuang Gan. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. *arXiv preprint arXiv:2311.01455*, 2023. 1