FieldGen: From Teleoperated Pre-Manipulation Trajectories to Field-Guided Data Generation

Wenhao Wang[*2], Kehe Ye[*2,4], Xinyu Zhou[*2], Tianxing Chen[*3,4], Cao Min,
Qiaoming Zhu, Xiaokang Yang, Yongjian Shen, Yang Yang, Maoqing Yao, Yao Mu

[1]MoE Key Lab of Artificial Intelligence, AI Institute, SJTU, [2]AgiBot, [3]HKU MMLab, [4]Lumina Group

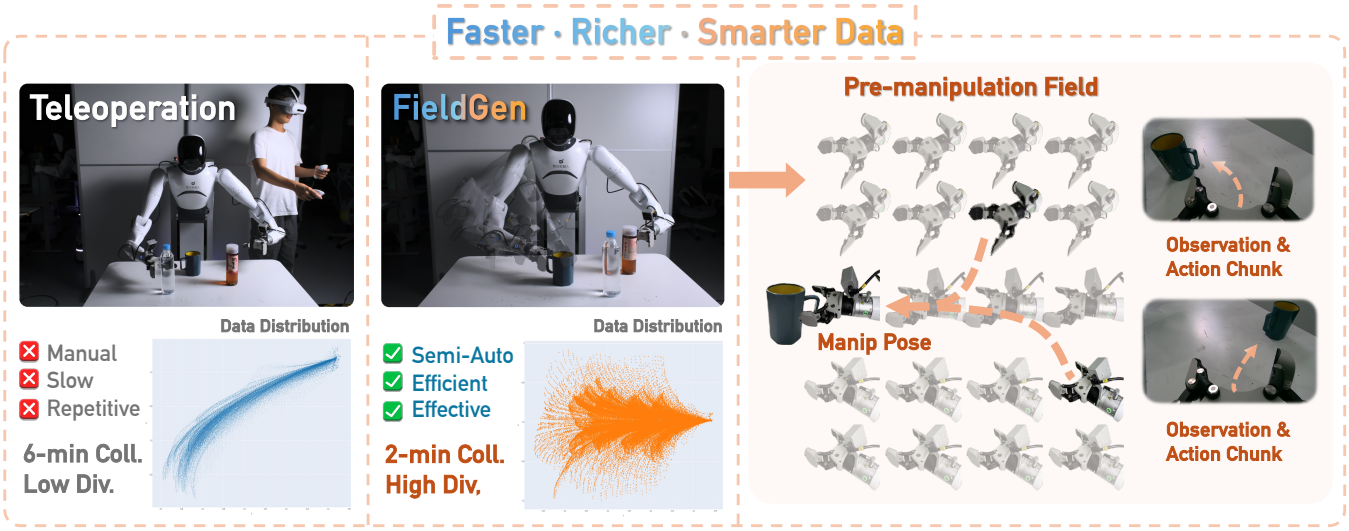


Fig. 1: **FieldGen** is a semi-auto data generation framework that enables scalable collection of diverse, high-quality real-world manipulation data with minimal human involvement.

***Abstract*— Large-scale, diverse datasets are essential for training robust robotic manipulation policies, yet existing data collection approaches face fundamental trade-offs between scale, diversity, and quality. Simulation-based methods achieve scale but suffer from sim-to-real gaps, while teleoperation produces high-quality demonstrations but is prohibitively expensive and limited in behavioral diversity. We propose FieldGen, a novel field-guided data generation framework that enables scalable collection of diverse, high-quality real-world manipulation data with minimal human involvement. Our key insight is that manipulation tasks can be decomposed into two phases: a pre-manipulation phase, where trajectory variation is acceptable, and a fine manipulation phase, which requires precise expert demonstration. FieldGen leverages this decomposition through a semi-automated pipeline. In the first stage, focused human demonstrations capture critical manipulation poses and contact interactions. In the second stage, an attraction field is constructed to automatically generate diverse pre-manipulation trajectories that converge to successful manipulation configurations. By decoupling these phases, FieldGen combines high-quality supervision for fine manipulation with scalable automated generation of behaviorally diverse reach data. Experiments show that policies trained on FieldGen-collected data outperform those trained on traditional teleoperation, achieving higher success rates with substantially less human effort and enabling stable long-duration data collection.**

I. INTRODUCTION

Recent end-to-end embodied intelligence models have achieved promising progress in robotic manipulation, demonstrating strong generalization capabilities and few-shot performance [2], [17], [21], [6], [5], [23]. However, these models critically depend on large-scale, diverse datasets to achieve effective performance. The collection of real-world robotic data requires substantial human effort through teleoperation and manual demonstration [3], [27], [22], making the creation of large-scale datasets prohibitively expensive.

Existing data collection approaches face a fundamental trade-off between scale, diversity, and data quality. Simulation-based methods can generate vast amounts of data with spatial randomization, but suffer from persistent sim-to-real gaps and limited behavioral diversity [25], [4], [19], [24]. Conversely, teleoperation produces high-quality demonstrations but faces severe scalability constraints: operators control only one robot at a time, experience cognitive fatigue during extended sessions, and converge to stereotypical motion patterns despite explicit instructions for diversity. Our analysis of large-scale teleoperation datasets reveals pronounced multimodal distributions that burden policy learning compared to more uniform behavioral diversity. Semi-automated approaches like

PATO [10] and GCENT [30] attempt to bridge this gap but still rely on pre-trained policies and may not escape human behavioral constraints.

To overcome these challenges, we propose FieldGen, a novel field-guided data generation framework that enables scalable collection of diverse, high-quality real-world manipulation data with minimal human involvement. FieldGen is built on a key observation that manipulation tasks can be naturally decomposed into two phases with fundamentally different requirements: a pre-manipulation phase involving reaching and approaching the target object, where trajectory variation is tolerable as long as paths converge to appropriate manipulation configurations, and a fine manipulation phase requiring precise contact-rich interactions where expert demonstrations provide the most value. Our method operates as a semi-automated pipeline that leverages this decomposition by first collecting a small set of high-quality human demonstrations focused on the fine manipulation phase, particularly critical manipulation poses and contact interactions. We then construct an attraction field for the pre-manipulation phase that guides trajectories toward successful manipulation configurations identified from these human demonstrations. Using automated scripts, we generate large numbers of randomized initial observations and compute trajectories that asymptotically converge to this attraction field, producing extensive observation-action pairs for the pre-manipulation phase. By decoupling these phases, FieldGen achieves three critical benefits: high-quality supervision for fine manipulation through focused human demonstration, scalable automated generation of diverse pre-manipulation trajectories, and efficient large-scale data collection that maintains both behavioral diversity and data quality while dramatically reducing human involvement compared to traditional teleoperation approaches.

We summarize our contributions as follows: (1) **Pre-manipulation field (PMF)**. We introduce the pre-manipulation field as a new concept, defined as an abstract representation that models and extrapolates from a small set of demonstration trajectories. PMF provides a unified framework for large-scale, semi-automated, real-robot data collection. (2) **Semi-automated collection via phase decoupling**. Building on PMF, we propose an efficient, semi-automated pipeline that decouples manipulation into pre-manipulation phases and fine manipulation phases, utilizing teleoperated manipulation poses to generate large-scale datasets, thereby substantially improving collection efficiency automatically. (3) **Empirical validation**. Experiments show that our method produces larger, higher-quality datasets in less time; models trained on our data are more robust; and operator workload is markedly reduced, enabling long-duration, stable data collection.

II. RELATED WORKS

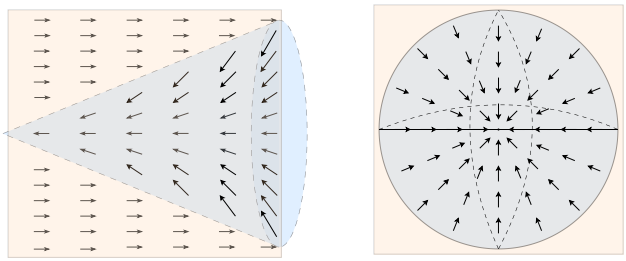
A. Teleoperation for Robotic Data Collection

Teleoperation with VR and custom hardware has been widely adopted to collect high-fidelity real-robot demonstrations. Systems leverage head-mounted displays [7], [34], cameras [26], [15], hand controllers or gloves [9], [18], [28], and low-cost bimanual platforms [31], [11], [12], [1], [13] to improve precision, reduce latency, and raise operator throughput. However, teleoperation suffers from fundamental scalability and quality limitations as human operators experience cognitive fatigue and unconsciously converge to stereotypical motion patterns, resulting in expensive data collection with insufficient diversity for training robust robotic policies. Recent efforts have attempted to improve teleoperation efficiency through assistance mechanisms. Recent efforts have attempted to address these limitations through assistance mechanisms. PATO [10] employs learned policies to automate repetitive subtasks while preserving human oversight for critical decisions, effectively reducing operator cognitive load, and GCENT [30] expanded it to real-world long-horizon tasks. However, such approaches still rely on pre-trained policies and may inherit the diversity limitations inherent in human demonstrations, as they automate existing human-demonstrated subtasks rather than fundamentally changing the data generation paradigm. In contrast, our approach differs by decomposing manipulation tasks at a more fundamental level and automatically generating diverse pre-manipulation trajectories through field-guided planning, enabling scalable diversity generation.

B. Automated and Simulation-Based Data Generation

To generate large-scale data with minimal human input, leveraging simulation engines to assist in data synthesis has become an emerging trend [32], [16], [14], [33], [29]. Large-scale simulation platforms like RoboTwin [25], MimicGen [24], and DexMimicGen [19] can generate millions of diverse manipulation scenarios through procedural generation and domain randomization. These systems achieve unprecedented scale by parallelizing data collection across multiple simulated environments and automatically varying object properties, lighting conditions, and scene configurations. Recent advances have pushed simulation-based generation to new scales. RoboTwin 2.0 [4] provides a comprehensive benchmark with diverse manipulation tasks, while HumanoidGen [20] focuses specifically on bimanual dexterous manipulation. These platforms demonstrate the potential for generating vast quantities of training data that would be impossible to collect through teleoperation alone. However, simulation-based approaches face several fundamental limitations that constrain their effectiveness. First, the persistent sim-to-real gap remains a critical challenge, as policies trained exclusively on synthetic data often fail to transfer robustly to physical systems due to discrepancies in contact dynamics, material properties, and sensor characteristics. Second, purely simulation-generated data suffers from limited behavioral diversity, as it primarily relies on procedurally-generated trajectories that lack the rich variability of contact interactions and manipulation strategies observed in human behaviors. These constraints fundamentally limit the ability of simulation-only approaches to capture the full complexity of dexterous manipulation tasks.

Our approach sidesteps the sim-to-real problem by generating data directly in the real world. By automating the pre-manipulation phase while relying on human demonstration



(a) (b)

Fig. 2: (a) Cone Field for Position. (b) Spherical Field for Orientation.

for contact-rich interactions, FieldGen combines automated scalability with real-world fidelity.

III. METHOD

A. Problem Formulation

We formulate the problem as collecting or generating pairs of observations o_t and action sequences a_t. To this end, we divide the data collection process into two phases: (1) Fine manipulation phase. In this stage, we use teleoperation to annotate the manipulation pose. This pose are further used to construct the pre-manipulation field $\mathcal{F}_{Gen}$, an abstract representation of the demonstrated trajectories. (2) FieldGen-based reach phase. In this stage, we employ FieldGen to conduct large-scale, semi-automated data collection. Randomized observations are sampled from diverse end-effector configurations, and corresponding action sequences are generated by asymptotically rolling out toward the pre-manipulation field.

This formulation ensures that fine manipulation is supervised by human expertise, while large-scale reach trajectories are automatically synthesized, providing broad coverage and scalable data generation.

B. $\mathcal{F}_{Gen}$ Field Construction

In this part, we use the previously collected fine manipulation trajectories to construct the pre-manipulation field $\mathcal{F}_{Gen}$. Based on this field, automated scripts are employed to randomly drive the robot arm, during which large amounts of end-effector poses and corresponding RGB observations are recorded. For each sampled pose, we compute a reach action sequence by evaluating its spatial relationship to the field and solving the corresponding inverse kinematics (IK). In this way, we generate observation–action pairs that serve as training data.

We decompose the generated field $\mathcal{F}_{Gen}$ into two components: a cone field for position, $\mathcal{F}_{pos}$, and a spherical field for orientation, $\mathcal{F}_{ori}$. The position field captures how the end-effector should approach the manipulation point, while the orientation field ensures alignment of the gripper with the desired manipulation orientation.

1) Cone Field for Position: Intuitively, the cone field constrains the end-effector to approach the manipulation position in alignment with the gripper axis. When the end-effector lies within the cone, the motion smoothly converges toward the target. When outside, the motion is first redirected into the cone and then guided to the manipulation point, mimicking a natural reach-and-align strategy.

Formally, let the manipulation goal position be $p_G \in \mathbb{R}^3$, with axis direction $\hat{u}$ (opposite to the gripper closing axis) and cone half-angle θ. For a sampled point Q with position p_Q, we decompose

$$a = \hat{u}^\top (p_Q - p_G), \qquad r = \big\| (p_Q - p_G) - a\hat{u} \big\|. \quad (1)$$

The cone surface is given by

$$r = \tan\theta\, a, \quad a \geq 0. \quad (2)$$

As illustrated in Figure 2a, the cone field is structured such that G acts as a zero-gravity sink, with field lines converging smoothly toward G within the cone and aligning parallel to $\hat{u}$ outside the cone.

Under the influence of the cone field, if Q lies inside the cone ($r \leq \tan\theta\, a$), the attraction vector is defined by a smooth half-cycloid curve in the plane spanned by $(G, Q, \hat{u})$:

$$x(t) = \mu(t - \sin t), \qquad y(t) = \nu(1 - \cos t), \qquad t \in [0, \pi], \quad (3)$$

with μ, ν chosen so that the curve starts at Q and ends at G. If Q lies outside the cone ($r > \tan\theta\, a$), it is first projected along $\hat{u}$ onto the cone surface at P, then follows the inner cycloid curve $P \to G$.

Thus the translational delta action is

$$\Delta p = \mathrm{Curve}(p_Q \to p_G \mid \hat{u}, \theta). \quad (4)$$

2) Spherical Field for Orientation: The spherical field ensures direct alignment of the gripper with the target manipulation orientation. The three axes of the spherical field correspond to roll, pitch, and yaw, with all field lines converging toward the center of the sphere as shown in Figure 2b. It provides a smooth corrective rotation that gradually drives the end-effector toward the goal orientation R_G.

Let $R_G \in SO(3)$ denote the goal orientation and R_Q the sampled orientation. The relative rotation is

$$R_\Delta = R_G^\top R_Q. \quad (5)$$

The axis–angle representation is

$$\omega = \log(R_\Delta) \in \mathbb{R}^3, \quad (6)$$

and the corrective angular velocity is defined as

$$\Delta R = -K_R\, \omega. \quad (7)$$

After obtaining the generated trajectory based on these fields, the length of action sequence is determined by dividing the trajectory length by the parameter β. Finally, we extract the action sequence of one chunk-sized segment; if the sequence length is shorter than the chunk size, we pad it with the last point of the trajectory.

IV. EXPERIMENT

A. Objective and Setup

We design experiments to evaluate FieldGen along three questions:

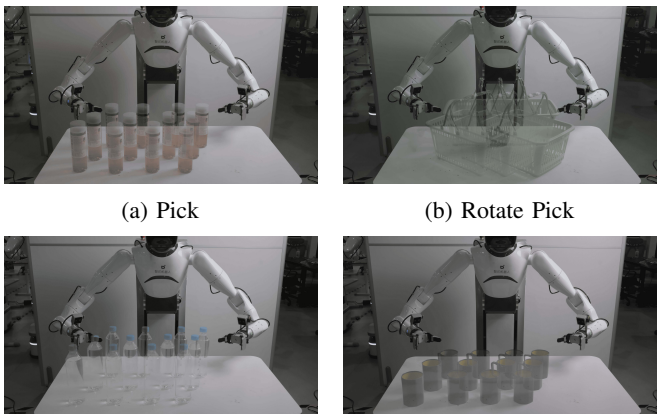
(1) Under the same collection time budgets, whether FieldGen produces training data that yields stronger policy performance than purely teleoperated data?

(2) At a fixed dataset size, whether FieldGen provides higher per-sample data quality, assessed by success rate and generalization across tasks?

(3) Through ablation studies, whether each component of FieldGen contributes essentially to overall effectiveness.

Experiments were conducted on the Agibot G1 robot, using an NVIDIA Orin for inference. To comprehensively evaluate the effect of data distribution and generation methods, we train all policies from scratch, removing any influence from pretrained data. We compare three policies: small RDT [21] (170M, SigLIP [35] frozen encoder), DP [8] (288M, jointly trained encoder and head) and ACT [36]. Since ACT achieved consistently low success rates across most tasks, its results are omitted from the tables for clarity. Observations are wrist-mounted RGB images; actions are end-effector delta poses. Each action chunk outputs 30 steps.

B. Equal-Time Data Effectiveness



(a) Pick (b) Rotate Pick

(c) Transparent Pick (d) Affordance Pick

Fig. 3: **Equal-Time Data Effectiveness Experiment Setups.** Rotate Pick requires the end-effector change orientation to pick objects, while Affordance Pick requires to manipulate specific position.

To verify the efficiency of FieldGen, we design experiments that compare policies trained on data collected by FieldGen with those trained on fully teleoperated data, under equal wall-clock collection budgets. Specifically, data are collected for checkpoints every 4 minutes up to 20 minutes. This setup also allows us to examine whether increasing data scale correlates with improved policy success rates. For each task, FieldGen collects a frame of the manipulation pose and then automatically gathers randomized observations until each episode reaches 1 minute in duration. As shown in Figure 3, we evaluate four manipulation tasks: Pick, Rotate Pick, Transparent Pick, and Affordance Pick. Data are obtained either through full teleoperation or through FieldGen. All policies are trained for 50 epochs until convergence, and success rates are measured across 12 randomized object placements per task.

From the results in Table I, we observe that FieldGen consistently outperforms teleoperation across all policies and all time budgets. At each checkpoint (4–20 minutes), FieldGen achieves higher success rates, exceeding teleoperation by 41.7%, 44.8%, 45.8%, 41.7%, and 35.5% respectively. This highlights the strong time efficiency of FieldGen: under equal wall-clock collection budgets, the data it generates leads to substantially stronger policy performance.

Figure 4a further shows that FieldGen scales more effectively with additional collection time. Success rates grow more sharply with time compared to teleoperation, and after 20 minutes of collection, FieldGen exceeds 80% success in all settings. In particular, diffusion-policy–based methods reach 100% success on three of the four evaluated tasks, reflecting the high quality of FieldGen data.

C. Data Quality and Generalization

To verify the quality of FieldGen data, we design experiments that compare its effect on task success rate and generalization. Specifically, we aim to assess whether policies trained on FieldGen data achieve higher success rates than those trained on teleoperation-only datasets of the same size, and whether FieldGen provides stronger generalization across different tasks and scene variations.

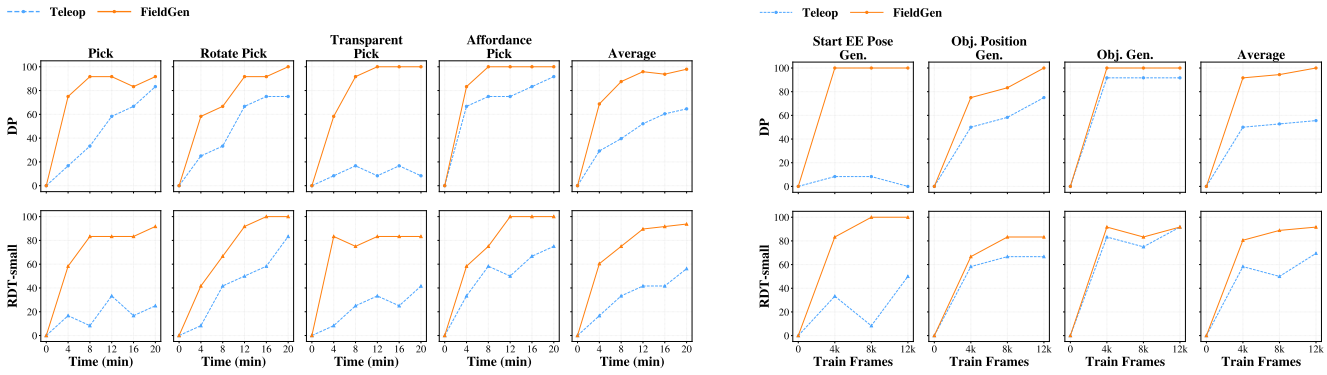
In this setup, we consider three evaluation conditions: (1) end-effector initialization, where policies are tested under diverse initial poses of the robot end-effector; (2) object placement, where policies are evaluated on varied object positions within the workspace; and (3) cross-instance generalization, where policies are transferred to unseen object instances within the same category, as shown in Figure 5. For both teleoperation and FieldGen, datasets of equal size (measured in data frames) are collected and used to train policies. All models are trained under identical schedules and hyperparameters. Success rates are measured across multiple randomized trials for each condition, enabling a direct comparison of robustness and generalization between the two data sources.

The results are summarized in Table II. Under equal dataset sizes (data frames), FieldGen consistently outperforms teleoperation across both policy backbones. Notably, for diffusion policy (DP), only 4000 FieldGen samples are sufficient to reach 100% success on both the Start EE Pose Generalization and Object Generalization tasks.

We further plot success rate against dataset size in Figure 4b. FieldGen shows a stable upward trend that follows a clear scaling curve, while teleoperation data consistently lags behind and exhibits irregular growth. This indicates that FieldGen covers a broader range of task-relevant state space, enabling more stable training and stronger performance scaling. In contrast, teleoperation trajectories are subject to

TABLE I: **Equal-Time Data Effectiveness.** Rows list tasks (DP/RDT-small); columns list Teleop vs. FieldGen across collection times.

Task	Model	Teleop					FieldGen				
Collect Time(min)		4	8	12	16	20	4	8	12	16	20
Pick	DP	16.7%	33.3%	58.3%	66.7%	83.3%	75.0%	91.7%	91.7%	83.3%	91.7%
	RDT-small	16.7%	8.3%	33.3%	16.7%	25.0%	58.3%	83.3%	83.3%	83.3%	91.7%
Rotate Pick	DP	25.0%	33.3%	66.7%	75.0%	75.0%	58.3%	66.7%	91.7%	91.7%	100%
	RDT-small	8.3%	41.7%	50.0%	58.3%	83.3%	41.7%	66.7%	91.7%	100%	100%
Transparent Pick	DP	8.3%	16.7%	8.3%	16.7%	8.3%	58.3%	91.7%	100%	100%	100%
	RDT-small	8.3%	25.0%	33.3%	25.0%	41.6%	83.3%	75.0%	83.3%	83.3%	83.3%
Affordance Pick	DP	66.7%	75.0%	75.0%	83.3%	91.7%	83.3%	100%	100%	100%	100%
	RDT-small	33.3%	58.3%	50.0%	66.7%	75.0%	58.3%	75.0%	100%	100%	100%
Average	DP	29.2%	39.6%	52.1%	60.4%	64.6%	68.8%	87.5%	95.8%	93.8%	97.9%
	RDT-small	16.7%	33.3%	41.7%	41.7%	56.2%	60.4%	75.0%	89.6%	91.7%	93.8%



(a) (b)

Fig. 4: (a) Equal-Time Data Effectiveness. (b) Equal-Data Generalization.

Fig. 5: **Data Quality and Generalization Experiment Setups**

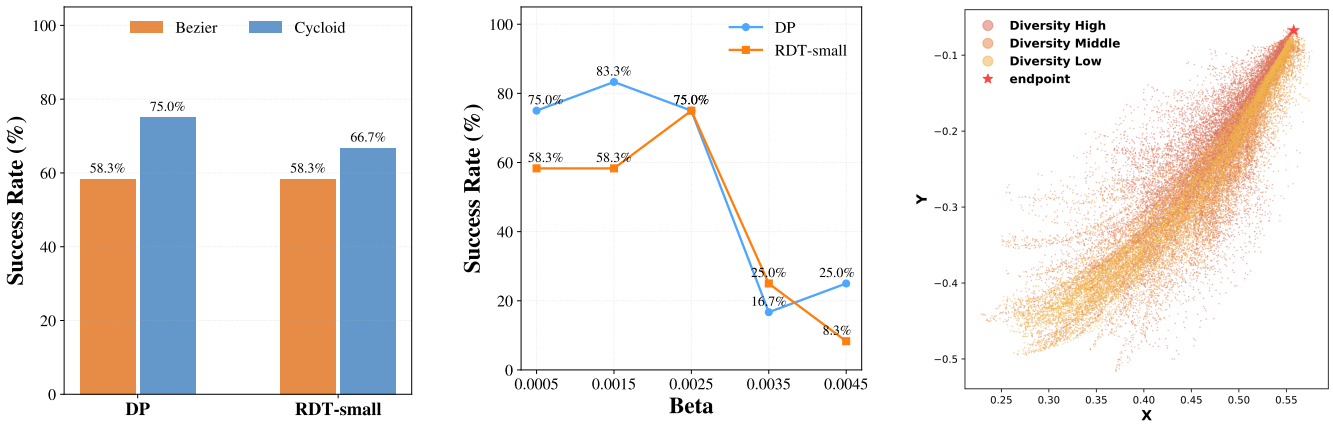
TABLE II: **Generalization Experiment Results.** We conduct controlled experiments on 3 generalization tasks: *Start EE Pose Gen*, *Obj. Position Gen*, and *Obj. Gen*, each evaluated under 3 different data volume.

Task	**Train Frames**	**Teleop**		**FieldGen**	
		DP	*RDT-small*	*DP*	*RDT-small*
Start EE Pose Gen.	4000	8.3%	33.3%	**100%**	**83.3%**
	8000	8.3%	8.3%	**100%**	**100%**
	12000	0%	50.0%	**100%**	**100%**
Obj. Position Gen.	4000	50.0%	58.3%	**75.0%**	**66.7%**
	8000	58.3%	66.7%	**83.3%**	**83.3%**
	12000	75.0%	66.7%	**100%**	**83.3%**
Obj. Gen.	4000	91.7%	83.3%	**100%**	**91.7%**
	8000	91.7%	75.0%	**100%**	**83.3%**
	12000	91.7%	91.7%	**100%**	**91.7%**
Average	4000	50.0%	58.3%	**91.7%**	**80.6%**
	8000	52.8%	50.0%	**94.4%**	**88.9%**
	12000	55.6%	69.5%	**100%**	**91.7%**

higher randomness in human motion, resulting in narrower coverage and more variable downstream performance.

D. Trajectory Diversity and Spatial Coverage

To evaluate the trajectory diversity of FieldGen, we designed experiments that compare its spatial coverage against teleoperated data. Specifically, we considered three levels of data diversity corresponding to different collection strategies. The first setting uses teleoperation with the same initial and final end-effector states, which yields relatively uniform trajectories and is referred to as low diversity. The second setting still relies on teleoperation but allows varied initial end-effector states while keeping the final object configuration fixed, thereby producing moderately diverse trajectories,



(a) (b) (c)

Fig. 6: (a) Ablation on Curve. (b) Ablation on β Parameter. (c) XY plane scatter plot of Diversity.

denoted as middle diversity. The third setting applies the FieldGen method, which automatically generates trajectories that span a wide range of spatial variations, achieving high diversity.

TABLE III: **Diversity Comparison.** We compare the spatial coverage and model performance across three different diversity levels.

Diversity Level	Spatial Coverage	DP	RDT-small
Low	9.04%	0%	0%
Middle	15.44%	66.7%	41.7%
High	18.14%	83.3%	83.3%

During evaluation, the target object's spatial position is varied to test policy generalization and robustness. Figure 6c illustrates the scatter plot of the collected trajectories along the XY plane, showing that FieldGen produces a broader and more uniform coverage. To quantify this, we calculate spatial coverage by enclosing all trajectories within the minimum bounding cube, dividing the cube into N voxels of equal size, and measuring the fraction N'/N of voxels that are traversed by trajectories. As reported in Table III, FieldGen achieves the highest coverage rate, confirming its ability to explore a broader range of the workspace.

We further compare downstream policy performance using DP and RDT trained on equal-sized datasets collected under the three settings. The resulting success rates are 0%, 54.2%, and 83.3% for low, middle, and high diversity, respectively. These results demonstrate that broader spatial coverage and higher trajectory diversity not only characterize the data itself but also translate directly into more robust and capable manipulation policies. Overall, the experiments verify that FieldGen provides data of higher quality and broader coverage, enabling the training of stronger policies compared to conventional teleoperation.

E. Ablation on Curve Type: Bezier vs. Cycloid

We conduct an ablation study to analyze the effect of different curve types on task success rate. In addition to the cycloid curve used in the main approach, we evaluate Bezier curves—a commonly used alternative in trajectory generation. The results are summarized in Figure 6a. Cycloid-based trajectories achieve success rates of 75% on DP and 66.7% on RDT, corresponding to improvements of 16.7% and 8.4% over Bezier, respectively. These results highlight the clear advantage of cycloid-based trajectory generation.

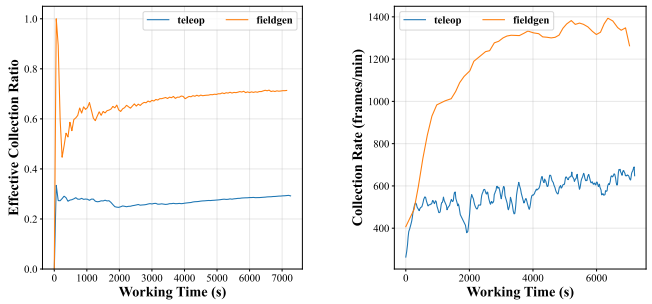
To better understand this advantage, we examine the geometric properties of the two approaches. The cycloid is generated directly under the pure geometric constraint of a rolling circle without slipping. This construction provides an explicit geometric prior and yields naturally smooth curvature variations. In contrast, the Bezier curve relies on manually defined control points to shape the trajectory. Lacking inherent geometric constraints, it often exhibits sharp curvature growth over short distances, resulting in abrupt changes that increase execution difficulty for the manipulator.

Overall, this ablation confirms that cycloid-based generation not only improves task success rates but also aligns with the physical feasibility of robot motion by producing smoother, more geometrically consistent trajectories.

F. Ablation on the β Parameter

We further conduct an ablation study on the parameter β, which controls the distance between consecutive frames in the generated trajectory. Physically, β reflects the average displacement between two successive end-effector poses in Cartesian space. The results on the manipulating task are shown in Figure 6b.

When β is set to a small value, the inter-frame distance becomes short, causing the manipulator to move only slightly within each chunk. This often leads to repeated local adjustments and oscillatory back-and-forth motions. Moreover, the overall motion speed is reduced, which significantly increases task completion time. In contrast, when β is set too large, the inter-frame distance increases, and FieldGen may generate a gripper-closing command while the manipulator is still far from the target object. In real tests, this premature closure prevents accurate manipulating and results in task failure.



(a) (b)

Fig. 7: (a) Collection time ratio. (b) Memory rate while data collecting.

Considering both efficiency and reliability, we select $\beta = 0.0025$ as a balanced choice in our experiments.

G. Less effort with long-duration data collection

To quantify the differential effort and efficiency between the teleoperation data collection pipeline and the FieldGen generation pipeline, we measured, for a teleoperation operator over a two-hour collection, the proportion of time actually spent performing data collection under different methods, and the resulting throughput, expressed as the rate at which collecting frames outputs were acquired.

We define Collection Time Ratio as the fraction of total collection time during which active data acquisition is occurring. This metric inversely captures procedural and supervisory overhead; higher values therefore indicate improved temporal efficiency. For the FieldGen pipeline, an elevated ratio further signifies diminished operator exertion, since a larger portion of the session consists of autonomous script execution requiring only low-intensity monitoring. The results are summarized in Figure 7a. In the experiment, FieldGen yielded a mean Collection Time Ratio of 66.73%, representing a 2.47× increase over the 27.07% observed with manual teleoperation, the latter demanding continuous, high-effort interaction and oversight. Frame collection rate is also as the data generation throughput (frames per minute) of the acquisition process, providing a direct measure of pipeline efficiency. The results are summarized in Figure 7b. FieldGen attained 1203.14 frames/min, exceeding manual teleoperation's 569.10 frames/min by a factor of 2.11.

Overall, FieldGen substantially outperforms manual teleoperation in both temporal utilization and data throughput. These experiment results indicate that FieldGen shifts a large fraction of the collection time cost into autonomous execution with only light supervision, thereby reducing operator cognitive and attentional load while simultaneously accelerating scalable dataset production.

V. CONCLUSIONS

We proposed **FieldGen**, a field-guided, semi-automated framework for real-world manipulation data generation. By decoupling reach and fine manipulation, and labeling reach trajectories via a cone (position) and spherical (orientation) attraction field, FieldGen converts free wrist-camera observations into high-quality ⟨obs, action⟩ pairs with minimal human effort. Real world experiments show that FieldGen consistently outperforms teleoperation under equal collection time, scales more favorably with additional data, and yields stronger generalization across start-pose, object-pose, and cross-instance settings. Ablations further indicate that cycloid-based trajectories improve success and motion smoothness over Bezier alternatives. Compared with teleoperation collection, FieldGen acquires data more efficiently while imposing substantially lower operator effort.

FieldGen thus offers a practical path to **faster collection** and **broader coverage** for large-scale real-robot datasets. Future work will extend the fields to multi-step tasks, incorporate uncertainty-aware control, and explore scene synthesis (e.g., 3DGS/NeRF) for fully automated observation generation.

REFERENCES

[1] Qingwei Ben, Feiyu Jia, Jia Zeng, Junting Dong, Dahua Lin, and Jiangmiao Pang. Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit. *arXiv preprint arXiv:2502.13013*, 2025.

[2] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Lucy Xiaoyang Shi, James Tanner, Quan Vuong, Anna Walling, Haohuan Wang, and Ury Zhilinsky. π_0: A vision-language-action flow model for general robot control, 2024.

[3] Qingwen Bu, Jisong Cai, Li Chen, Xiuqi Cui, Yan Ding, Siyuan Feng, Xindong He, Xu Huang, et al. Agibot world colosseo: A large-scale manipulation platform for scalable and intelligent embodied systems. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025.

[4] Tianxing Chen, Zanxin Chen, Baijun Chen, Zijian Cai, Yibin Liu, Qiwei Liang, Zixuan Li, Xianliang Lin, Yiheng Ge, Zhenyu Gu, et al. Robotwin 2.0: A scalable data generator and benchmark with strong domain randomization for robust bimanual robotic manipulation. *arXiv preprint arXiv:2506.18088*, 2025.

[5] Tianxing Chen, Yao Mu, Zhixuan Liang, Zanxin Chen, Shijia Peng, Qiangyu Chen, Mingkun Xu, Ruizhen Hu, Hongyuan Zhang, Xuelong Li, et al. G3flow: Generative 3d semantic flow for pose-aware and generalizable object manipulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1735–1744, 2025.

[6] Tianxing Chen, Kaixuan Wang, Zhaohui Yang, Yuhao Zhang, Zanxin Chen, Baijun Chen, Wanxi Dong, Ziyuan Liu, Dong Chen, Tianshuo Yang, et al. Benchmarking generalizable bimanual manipulation: Robotwin dual-arm collaboration challenge at cvpr 2025 meis workshop. *arXiv preprint arXiv:2506.23351*, 2025.

[7] Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-television: Teleoperation with immersive active visual feedback. In *Proceedings of the Conference on Robot Learning (CoRL)*, 2024.

[8] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, page 02783649241273668, 2023.

[9] Shivin Dass, Wensi Ai, Yuqian Jiang, Samik Singh, Jiaheng Hu, Ruohan Zhang, Peter Stone, Ben Abbatematteo, and Roberto Martín-Martín. Telemoma: A modular and versatile teleoperation system for mobile manipulation. *arXiv preprint arXiv:2403.07869*, 2024.

[10] Shivin Dass, Karl Pertsch, Hejia Zhang, Youngwoon Lee, Joseph J Lim, and Stefanos Nikolaidis. Pato: Policy assisted teleoperation for scalable robot data collection. *arXiv preprint arXiv:2212.04708*, 2022.

[11] Hongjie Fang, Hao-Shu Fang, Yiming Wang, Jieji Ren, Jingjing Chen, Ruo Zhang, Weiming Wang, and Cewu Lu. Airexo: Low-cost exoskeletons for learning whole-arm manipulation in the wild. *arXiv preprint arXiv:2309.14975*, 2023.

[12] Hongjie Fang, Chenxi Wang, Yiming Wang, Jingjing Chen, Shangning Xia, Jun Lv, Zihao He, Xiyan Yi, Yunhan Guo, Xinyu Zhan, Lixin Yang, Weiming Wang, Cewu Lu, and Hao-Shu Fang. Airexo-2: Scaling up generalizable robotic imitation learning with low-cost exoskeletons. *arXiv preprint arXiv:2503.03081*, 2025.

[13] Zipeng Fu, Tony Z. Zhao, and Chelsea Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. *arXiv preprint arXiv:2401.02117*, 2024.

[14] Caelan Garrett, Ajay Mandlekar, Bowen Wen, and Dieter Fox. Skillmimicgen: Automated demonstration generation for efficient skill learning and deployment. *arXiv preprint arXiv:2410.18907*, 2024.

[15] Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. Dexpilot: Vision based teleoperation of dexterous robotic hand-arm system. *arXiv preprint arXiv:1910.03135*, 2019.

[16] Ryan Hoque, Ajay Mandlekar, Caelan Garrett, Ken Goldberg, and Dieter Fox. Intervengen: Interventional data generation for robust and data-efficient robot imitation learning. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2840–2846. IEEE, 2024.

[17] Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Manuel Y. Galliker, Dibya Ghosh, Lachy Groom, Karol Hausman, Brian Ichter, Szymon Jakubczak, Tim Jones, Liyiming Ke, Devin LeBlanc, Sergey Levine, Adrian Li-Bell, Mohith Mothukuri, Suraj Nair, Karl Pertsch, Allen Z. Ren, Lucy Xiaoyang Shi, Laura Smith, Jost Tobias Springenberg, Kyle Stachowicz, James Tanner, Quan Vuong, Homer Walke, Anna Walling, Haohuan Wang, Lili Yu, and Ury Zhilinsky. $\pi_{0.5}$: a vision-language-action model with open-world generalization, 2025.

[18] Aadhithya Iyer, Zhuoran Peng, Yinlong Dai, Irmak Guzey, Siddhant Haldar, Soumith Chintala, and Lerrel Pinto. Open teach: A versatile teleoperation system for robotic manipulation. *Proceedings of Machine Learning Research*, 270:2372–2395, 2024. Publisher Copyright: © 2024 Proceedings of Machine Learning Research.; 8th Conference on Robot Learning, CoRL 2024 ; Conference date: 06-11-2024 Through 09-11-2024.

[19] Zhenyu Jiang, Yuqi Xie, Kevin Lin, Zhenjia Xu, Weikang Wan, Ajay Mandlekar, Linxi Jim Fan, and Yuke Zhu. Dexmimicgen: Automated data generation for bimanual dexterous manipulation via imitation learning. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16923–16930. IEEE, 2025.

[20] Zhi Jing, Siyuan Yang, Jicong Ao, Ting Xiao, Yugang Jiang, and Chenjia Bai. Humanoidgen: Data generation for bimanual dexterous manipulation via llm reasoning, 2025.

[21] Songming Liu, Lingxuan Wu, Bangguo Li, Hengkai Tan, Huayu Chen, Zhengyi Wang, Ke Xu, Hang Su, and Jun Zhu. Rdt-1b: a diffusion foundation model for bimanual manipulation. *arXiv preprint arXiv:2410.07864*, 2024.

[22] Yushan Liu, Shilong Mu, Xintao Chao, Zizhen Li, Yao Mu, Tianxing Chen, Shoujie Li, Chuqiao Lyu, Xiao-ping Zhang, and Wenbo Ding. Avr: Active vision-driven robotic precision manipulation with viewpoint and focal length optimization. *arXiv preprint arXiv:2503.01439*, 2025.

[23] Guanxing Lu, Zifeng Gao, Tianxing Chen, Wenxun Dai, Ziwei Wang, Wenbo Ding, and Yansong Tang. Manicm: Real-time 3d diffusion policy via consistency model for robotic manipulation. *arXiv preprint arXiv:2406.01586*, 2024.

[24] 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.

[25] Yao Mu, Tianxing Chen, Shijia Peng, Zanxin Chen, Zeyu Gao, Yude Zou, Lunkai Lin, Zhiqiang Xie, and Ping Luo. Robotwin: Dual-arm robot benchmark with generative digital twins (early version). In *European Conference on Computer Vision*, pages 264–273. Springer, 2024.

[26] Yuzhe Qin, Wei Yang, Binghao Huang, Karl Van Wyk, Hao Su, Xiaolong Wang, Yu-Wei Chao, and Dieter Fox. Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[27] Quan Vuong, Sergey Levine, Homer Rich Walke, Karl Pertsch, Anikait Singh, Ria Doshi, Charles Xu, Jianlan Luo, Liam Tan, Dhruv Shah, et al. Open x-embodiment: Robotic learning datasets and rt-x models. In *Towards Generalist Robots: Learning Paradigms for Scalable Skill Acquisition@ CoRL2023*, 2023.

[28] Chen Wang, Haochen Shi, Weizhuo Wang, Ruohan Zhang, Li Fei-Fei, and C. Karen Liu. Dexcap: Scalable and portable mocap data collection system for dexterous manipulation. *arXiv preprint arXiv:2403.07788*, 2024.

[29] Lirui Wang, Yiyang Ling, Zhecheng Yuan, Mohit Shridhar, Chen Bao, Yuzhe Qin, Bailin Wang, Huazhe Xu, and Xiaolong Wang. Gensim: Generating robotic simulation tasks via large language models. *arXiv preprint arXiv:2310.01361*, 2023.

[30] Wenhao Wang, Jianheng Song, Chiming Liu, Jiayao Ma, Siyuan Feng, Jingyuan Wang, Yuxin Jiang, Kylin Chen, Sikang Zhan, Yi Wang, et al. Genie centurion: Accelerating scalable real-world robot training with human rewind-and-refine guidance. *arXiv preprint arXiv:2505.18793*, 2025.

[31] Philipp Wu, Yide Shentu, Zhongke Yi, Xingyu Lin, and Pieter Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators. *arXiv preprint arXiv:2309.13037*, 2023.

[32] Zhengrong Xue, Shuying Deng, Zhenyang Chen, Yixuan Wang, Zhecheng Yuan, and Huazhe Xu. Demogen: Synthetic demonstration generation for data-efficient visuomotor policy learning. *arXiv preprint arXiv:2502.16932*, 2025.

[33] Runyi Yu, Yinhuai Wang, Qihan Zhao, Hok Wai Tsui, Jingbo Wang, Ping Tan, and Qifeng Chen. Skillmimic-v2: Learning robust and generalizable interaction skills from sparse and noisy demonstrations. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pages 1–11, 2025.

[34] Yanjie Ze, Zixuan Chen, Wenhao Wang, Tianyi Chen, Xialin He, Ying Yuan, Xue Bin Peng, and Jiajun Wu. Generalizable humanoid manipulation with 3d diffusion policies. *arXiv preprint arXiv:2410.10803*, 2024.

[35] Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11975–11986, 2023.

[36] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.