# Strategic *Vantage* Selection for Learning Viewpoint-Agnostic Manipulation Policies

**Sreevishakh Vasudevan**[1], **Som Sagar**[1] **and Ransalu Senanayake**[1]
[1]Arizona State University

## Abstract

Vision-based manipulation has shown remarkable success, achieving promising performance across a range of tasks. However, these manipulation policies often fail to generalize beyond their training viewpoints, which is a persistent challenge in achieving perspective-agnostic manipulation, especially in settings where the camera is expected to move at runtime. Although collecting data from many angles seems a natural solution, such a naive approach is both resource-intensive and degrades manipulation policy performance due to excessive and unstructured visual diversity. This paper proposes **Vantage**, a framework that systematically identifies and integrates data from optimal perspectives to train robust, viewpoint-agnostic policies. By formulating viewpoint selection as a continuous optimization problem, we iteratively fine-tune policies on a few vantage points. Since we leverage Bayesian optimization to efficiently navigate the infinite space of potential camera configurations, we are able to balance exploration of novel views and exploitation of high-performing ones, thereby ensuring data collection from a minimal number of effective viewpoints. We empirically evaluate this framework on diverse standard manipulation tasks using multiple policy learning methods, demonstrating that fine-tuning with data from strategic camera placements yields substantial performance gains, achieving average improvements of up to 46.19% when compared to fixed, random, or heuristic-based strategies.

## 1 Introduction

Vision-based robot learning critically depends on the quality, consistency, and comprehensiveness of visual input, making camera placement a decisive yet frequently overlooked factor in training robust manipulation policies [1, 2]. The choice of camera viewpoint directly influences feature extraction, state estimation, and ultimately, policy performance. For instance, consider a robot tasked with picking objects from a cluttered table: a top-down camera initially offers a clear overhead view, but as the robot's arm moves to grasp an object, it may obstruct the camera, complicating precise manipulation. Conversely, a side-view camera can continuously track the robot's motion without obstruction but may fail to clearly represent essential object attributes such as length or orientation, crucial for successful grasping [3].

Despite the importance of viewpoints, most manipulation policies become highly specialized to their training viewpoints, performing reliably only when the camera is the same as in training. While suitable for static research settings, these policies fail in humanoid robots with constantly moving heads, mobile manipulators with the camera mounted on pan-tilt head or mobile bases, robots on moving assembly lines [4, 5], etc. One might assume that simply training on data from many random viewpoints enhances robustness, but in practice, excessive variation or over-augmentation increases the sample complexity and can lead to poor convergence. From the perspective of bias–variance tradeoff [6], high-capacity models exposed to overly diverse data may overfit noise across viewpoints and struggle to learn consistent features. This phenomenon parallels findings in domain generalization,

Submitted to the First Exploration in AI Today Workshop at ICML (EXAIT at ICML 2025). Do not distribute.

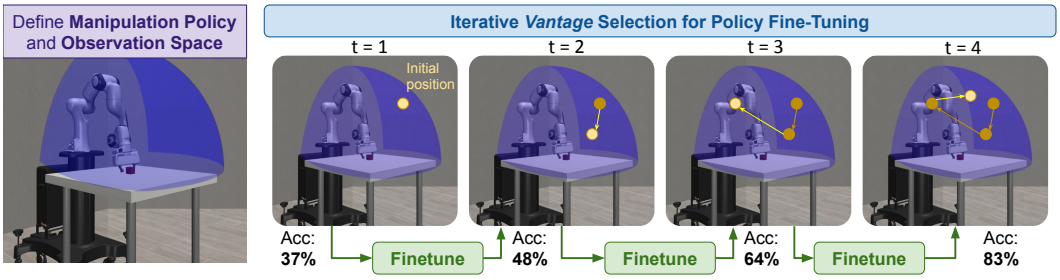


Figure 1: Starting from an initial camera viewpoint, we first define the model and search space, then sequentially fine-tune the policy at three additional selected viewpoints. Each fine-tuning step yields a measurable gain in task success rate rising from 37% at the first refine to 48 %, 64 %, and finally 83 % accuracy, illustrating how strategic viewpoint selection progressively enhances

where large shifts between training distributions hinder model generalization [7]. Therefore, it is important to learn the manipulation policy from a few meaningful viewpoints.

Instead of training a viewpoint-agnostic policy from scratch, we find it more stable to fine-tune a pretrained policy with some additional strategic views. This process also aligns with the prevailing trend of fine-tuning, as models continue to grow in size and complexity [8]. For example, a warehouse robot trained to grasp items from a fixed viewpoint pre-deployment may later need to adapt to dynamic environments such as variable shelf heights and mobile platforms at deployment [9]. This shift necessitates structured methods capable of systematically identifying a few informative camera perspectives, thereby balancing diversity in viewpoints with the stability required for effective learning.

Since our goal is to find the optimal viewpoint from the space of all viewpoints, which increases the model's performance across diverse camera perspectives, using the minimum number of samples as possible, we propose a Bayesian Optimization (BO)-driven framework for systematically identifying and combining training viewpoints to enhance policy performance. Instead of exposing the robot to all possible viewpoints at once, risking instability and poor convergence, we adopt a structured, iterative approach. This method systematically finds the most informative camera angles over multiple iterations, balancing exploration of new viewpoints with the stability required for effective learning. By doing so, our approach ensures that the robot learns from perspectives that enhance task performance while avoiding the pitfalls of excessive viewpoint variation. *Our main contribution is a framework that iteratively finds additional camera viewpoints to progressively finetune an arbitrary manipulation policy to ensure the manipulation policy is agnostic to viewpoints.*

## 2 Related Work

**Active vision in robotics:** The paradigm of active vision, wherein an agent dynamically controls its viewpoint to enhance perceptual efficiency, has long been a foundational pursuit in robotics and computer vision [10, 11]. Recent advances have leveraged deep learning to address the next best view (NBV) problem, where the agent selects optimal views to maximize information gain. Learning based methods have emerged, using reinforcement learning or uncertainty modeling [12, 13] to guide viewpoint selection for tasks such as object recognition and 3D reconstruction [14, 15, 16]. Notably, recent works have substantially raised the bar. For instance, GenNBV [2] introduced a generalizable NBV policy that learns in a 5D action space, while SUGARL [4] learns intrinsic sensorimotor rewards to guide view selection under partial observability. Active Neural Mapping [17] and VIN-NBV [18] further refined the NBV task using receding horizon planning and view quality introspection networks. In multi-view tasks, Hou et al. [19] developed methods to select informative views for efficient 3D understanding [20]. Additionally, Bayesian methods for NBV planning [16] are gaining traction in robotics, such as in safe contact-based exploration [21] and dynamic view planning for robotic manipulation [5]. Recent studies have also explored attention-driven NBV strategies for targeted perception in complex environments [22] and affordance-driven NBV planning for robotic grasping in cluttered scenes [23]. While active vision aims to solve the problem of "where to look" at inference time, complementary to them, we address the problem of how to collect optimal data at the training stage to develop a perspective agnostic policy.

**Viewpoint selection for generalization:** Beyond selecting views for immediate perceptual gain, recent research has explored how actively varying viewpoints can improve the robustness and generalization of learned visual representations. Jayaraman et al.[24] demonstrated that agents can learn viewpoint-invariant features by controlling the camera and predicting changes across views. Follow up work from Wu et al. [25] used neural predictors to estimate utility in active 3D reconstruction, while Lin et al. [26] leveraged neural implicit representations for efficient scene understanding. In parallel, multi-agent or collaborative view selection strategies like MAP-NBV [27] have been developed to jointly optimize multi-view acquisition under budgeted exploration, enabling robust understanding of complex scenes. Additionally, approaches such as Pred-NBV [28] have utilized prediction-guided strategies to enhance the efficiency of 3D object reconstruction. Parallel work in domain generalization has highlighted that excessive diversity between training and test distributions can degrade model performance, a phenomenon analyzed through invariant feature learning and distributional robustness frameworks [7, 29]. Distinct from prior work, Vantage targets the fine-tuning of pre-trained policies using data collected from a strategically chosen set of camera viewpoints via Bayesian Optimization. Rather than optimizing a single next view, it emphasizes maximizing the downstream task performance and generalization of learned policies across unseen and dynamic camera settings. This policy-centric use of viewpoint planning represents a unique contribution that bridges active vision and policy adaptation in real-world environments. Vantage leverages existing policies and enhances their robustness through informed viewpoint selection, addressing the challenges of dynamic and unpredictable operational conditions.

**Data diversity and generalization:** Although increasing training set diversity often aims to improve robustness, unchecked heterogeneity can instead degrade model performance. Recht et al. [30] report 11–14 percent drops in ImageNet top-1 accuracy on a harder test split, indicating that naively adding variation without matching distributional priors harms generalization. This effect mirrors the classical bias–variance tradeoff: with fixed model capacity, excess data heterogeneity inflates variance and thus test error [31]. In robotic manipulation, the DROID dataset [32] further illustrates that policies trained on overly diverse, in-the-wild scenes under-performs unless coupled with targeted adaptation or fine-tuning mechanisms, underscoring the need for our selective, BO-driven viewpoint strategy.

## 3 Methodology

**Formulating the optimization problem**. We define a viewpoint, $\theta$, as a 3D camera placement around the robot, where the camera is oriented toward the center of the manipulator's workspace. Given a pre-trained manipulation policy, $\pi$, our objective is to identify the optimal camera viewpoint, $\theta^{\text{vantage}}$, that when used for fine-tuning the manipulation policy, maximizes task performance, $J(\cdot)$, across the test space of viewpoints, $\Theta_{\text{test}}$. To find this vantage, we search across the large, continuous space of candidate training viewpoints, $\Theta_{\text{train}}$:

$$\theta^{\text{vantage}} = \underset{\theta_{\text{train}} \in \Theta_{\text{train}}}{\arg\max}\, J(\pi_{\theta_{\text{train}}}(\theta_{\text{test}})), \quad \forall \theta_{\text{test}} \in \Theta_{\text{test}}. \tag{1}$$

Here, $\pi_{\theta_{\text{train}}}(\theta_{\text{test}})$ indicates rolling out the policy, which was trained on data obtained from the viewpoint $\theta_{\text{train}}$, at test angles $\theta_{\text{test}} \in \Theta_{\text{test}}$. In practice, we consider $\Theta_{\text{train}}$ and $\Theta_{\text{test}}$ to have the same boundaries. Without loss of generality, we consider the performance metric, $J$, as the average manipulation success rate calculated over several rollouts on a discrete grid of test viewpoints. Note that training is much more expensive than testing because training requires collecting trajectories and updating neural network parameters whereas testing only requires policy rollouts. Considering these challenges, we iteratively optimize this expensive blackbox function, $J$, using a variant of Bayesian optimization (BO) [33] to obtain optimal viewpoints to train the neural network in such a way that policy performs well for test viewpoints.

**Modeling the performance metric**. We model the relationship between $J$ and $\theta_{\text{train}}$ as a Gaussian process,

$$J(\pi_{\theta_{\text{train}}}) \sim \mathcal{GP}(\mu(\theta_{\text{train}}), k(\theta_{\text{train}}, \theta'_{\text{train}})), \tag{2}$$

with predicted mean success rate, $\mu(\theta_{\text{train}})$, and a similarity metric between viewpoints, $k(\theta_{\text{train}}, \theta'_{\text{train}})$, computed using a squared-exponential kernel [34]. This probabilistic surrogate model iteratively guides the BO in selecting informative viewpoints—vantage points—in a sample-efficient manner.

**Batched optimization**. While (2) can be used in a standard BO setting [33], we further enhance the efficiency by evaluating multiple tests viewpoints simultaneously during exploration. To this end, we

**Algorithm 1** *Vantage*

**Step 1: Gather Initial Data**
Sample $q$ random viewpoints $\{\theta_{\text{train}}^{(j)}\}_{j=1}^{q}$, where each $\theta_{\text{train}}^{(j)} \in \Theta_{\text{train}}$
Generate manipulation datasets $\{\mathcal{D}^{(j)}\}_{j=1}^{q}$, from robot trials or simulation at viewpoints $\{\theta_{\text{train}}^{(j)}\}_{j=1}^{q}$
Fine-tune the original policy on $\mathcal{D}_j$ independently
Evaluate the fine-tuned models across $\Theta$ to obtain success rates $\{J_i\}_{i=1}^{q}$
Initialize a historical dataset $\mathcal{D}_{gp} \leftarrow \{(\theta_j, J_i)\}_{i=1}^{q}$
**Step 2: Bayesian Optimization Loop**
**for** $i = 1$ to $N$ **do**
  Use $\mathcal{D}$ and Bayesian optimization to select q new angles $\{\theta_{\text{new},j}\}_{j=1}^{q} = (\theta_h^{\text{new},j}, \theta_v^{\text{new},j}) \in \Theta$
  Generate datasets $\mathcal{D}_{\text{new},j}$ at $\theta_{\text{new},j}$
  Fine-tune the original model separately on $\mathcal{D}_{\text{new},j}$
  Evaluate the model fine-tuned at $\theta_{\text{new}}$ to obtain $J_{\text{new}}$
  Update $\mathcal{D}_{gp} \leftarrow \mathcal{D}_{gp} \cup \{(\theta_{\text{new},j}, J_{\text{new},j})\}$
**end for**
**Step 3: Final Selection**
$\theta^* \leftarrow \arg\max_{\theta \in \Theta} J(\pi_\theta)$

use the q-Upper Confidence Bound (q-UCB) [35] as our acquisition function. Consider a batch of viewpoints, $\Theta_q = (\theta_1, \ldots, \theta_q)$, where $\theta \in \Theta_{\text{train}}$, for fine-tuning the manipulation policy. Following [35], we obtain the next batch of viewpoints to train by maximizing the joint acquisition score,

$$\Theta_q^{\text{next}} = \arg\max_{\Theta_q \subset \Theta_{\text{train}}} \alpha_{\text{qUCB}}(\Theta_q) \tag{3}$$

$$= \arg\max_{\Theta_q \subset \Theta_{\text{train}}} \mathbb{E}_{\tilde{\Theta} \sim \mathcal{N}\left(\mu(\Theta_q),\, \frac{\beta\pi}{2}\Sigma(\Theta_q)\right)} \left[\max_{i=1,\ldots,q} \left(\mu(\theta_i) + |\tilde{\Theta}_i - \mu(\theta_i)|\right)\right], \tag{4}$$

where $\mu(\Theta_q) \in \mathbb{R}^q$ and $\Sigma(\Theta_q) \in \mathbb{R}^{q \times q}$ are the GP posterior mean and covariance on the batch, respectively. The scalar $\beta > 0$ controls exploration vs. exploitation: larger $\beta$ spreads the reparameterized posterior, encouraging sampling of high-uncertainty viewpoints, while smaller $\beta$ focuses on high-mean regions. In practice, the expectation is approximated by quasi-Monte Carlo sampling of $\tilde{\Theta}$, yielding an easy to compute acquisition score for selecting the next batch of viewpoints.

Once new $q$ viewpoints are selected, they are mapped from the BO space to real-world coordinates, and the policy is fine-tuned using datasets generated from these viewpoints. The updated policy is then evaluated and the corresponding success rates are used to update the prior. This is done for a fixed number of iterations. Upon completion of all iterations, we select the policy fine-tuned with the highest observed success rate as our final model. This process is described in Algorithm 1.

## 4 Experiments

We use *Vantage* to improve manipulation policies trained in RoboSuite [36] using datasets from RoboMimic [37]. To represent a range of common manipulation challenges with varying levels of difficulty, we benchmark across the standard tasks: Lift, Square, and Pick & Place. To have models of different architectures and learning capacities, we use BC [38], BCQ [39], Diffusion Policy [40], and BC Transformer [37] models.

### 4.1 Defining the space of allowed camera placements

Let $b \in \mathbb{R}^3$ denote the robot's base position and let $r > 0$ be the fixed radial distance at which the camera is placed. Let the sphere of radius $r$ around $b$ is defined by $S = \{x \in \mathbb{R}^3 : \|x - b\| = r\}$. For experimental purposes, we restrict our space of viewpoints to the *spherical quadrant* in front of the robot and above the table (the blue area in Fig. 1):

$$\Theta = \Theta_{\text{train}} = \Theta_{\text{test}} = \{x \in S : (x - b) \cdot u \geq 0 \quad \text{and} \quad (x - b) \cdot n \geq 0\},$$

where $u$ is the unit vector pointing forward from the robot's base and $n$ is the upward normal to the table plane. Any viewpoint $\theta \in \Theta$ can be uniquely parameterized by the horizontal and vertical

angles $\theta_h$ and $\theta_v$, respectively:

$$x(\theta_h, \theta_v) = b + r \begin{bmatrix} \cos\theta_v \, \cos\theta_h \\ \cos\theta_v \, \sin\theta_h \\ \sin\theta_v \end{bmatrix}, \quad \theta_h \in \left[-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right], \quad \theta_v \in \left[-\tfrac{\pi}{4}, \tfrac{\pi}{4}\right]$$

Any $\theta = (\theta_h, \theta_v) \in \Theta$ is considered a valid camera placement. Robot policies are initially trained from the viewpoint $\theta = (0, 0)$. The resulting model is then fine-tuned using data collected from one additional viewpoint. To quantify the variability of our method, in all experiments, this process is repeated four times, each with a different randomly selected initialization point for the GP. Next, we evaluate the model's performance on a uniform grid over the viewing space $\Theta$ and fit a GP surrogate to these measurements. By applying the q-UCB acquisition function to the GP, we identify the next eight most informative viewpoints for further data collection and fine-tuning.

We parameterize each viewpoint by normalized coordinates $\nu_h, \nu_v \in [0, 1]$. The GP is trained on these normalized inputs and thus only *sees* values in the unit square. To convert a normalized sample $(\nu_h, \nu_v)$ into actual camera angles, we apply the affine transformation:

$$\theta_{h/v} = \left(\nu_{h/v} - 0.5\right)\left(\theta_{h/v}^{\max} - \theta_{h/v}^{\min}\right) + \frac{\theta_{h/v}^{\max} + \theta_{h/v}^{\min}}{2}$$

All methods (grid search, random search, and *Vantage*) were allocated identical compute resources and used the same hyperparameters (see Appendix and code). Each task–policy combination underwent eight fine-tuning steps per iteration per method, and 4 iterations were done leading to 32 models fine-tuned for each combination.

### 4.2 Results

Our experiments demonstrate that *Vantage* yields significant improvements in policy performance across a variety of tasks and architectures while requiring only a small number of fine-tuning steps. Table 1 summarizes the success rates of BC and Table 2 the success rates of Diffusion policies under the baseline (default), $\Theta$, and dynamic camera settings, benchmarking *Vantage* against grid and random search. We observe that our method consistently outperforms or matches alternative strategies, with the largest gain seen for the Diffusion policy on Pick & Place—an increase from 37.01% to 83.20%, a 46.19% improvement. Moreover, as detailed in Appendix, by leveraging the q-UCB acquisition function within a GP framework, we present theoretical guarantees of asymptotic optimality, ensuring that each new viewpoint selection is near-optimal under standard regularity conditions. Together, these findings demonstrate that Vantage combines strong theoretical foundations with practical efficiency, delivering faster convergence and higher success rates than exhaustive or naïve sampling strategies.

Figure 4 illustrates the progression of best success rates over BO iterations for each model class. In all cases, Vantage converges more rapidly than both grid and random search, reaching near-optimal performance within 10–12 steps. Furthermore, Fig. 2 shows the detailed improvement of each policy for each task over $\Theta$ when using Vantage, with consistent gains observed across all tasks. Figure 3 shows the success rate of a policy before and after fine-tuning. Figure 5 shows each iteration slowly leads towards the global maxima.

Table 1: Success Rate of BC policy

| | Camera Placement | Base Model | Grid Search | Random Search | Vantage |
|---|---|---|---|---|---|
| Lift | Default | 100.00% | 100% | 100% | 100% |
| | $\Theta$ | 6.90% | 9.18% | 23.15% | 23.5% |
| | Dynamic | 91.66% | 93.33% | 100.00% | 100.00% |
| Pick Place | Default | 70.00% | 70.00% | 70.00% | 70.00% |
| | $\Theta$ | 0.80% | 1.04% | 2.53% | 3.90% |
| | Dynamic | 3.00% | 4.23% | 7.66% | 9.88% |
| Square | Default | 30.00% | 20.00% | 30.00% | 20.00% |
| | $\Theta$ | 0.24% | 0.61% | 0.74% | 1.00% |
| | Dynamic | 0.00% | 0.00% | 0.00% | 0.00% |

Table 2: Success Rate of Diffusion policy

| | Camera Placement | Base Model | Grid Search | Random Search | Vantage |
|---|---|---|---|---|---|
| Lift | Default | 100.00% | 100% | 100% | 100% |
| | $\Theta$ | 50.40% | 58.57% | 65.82% | 66.10% |
| | Dynamic | 98.33% | 100% | 100% | 100% |
| Pick Place | Default | 90.00% | 90.00% | 90.00% | 90.00% |
| | $\Theta$ | 37.01% | 22.09% | 66.11% | 83.20% |
| | Dynamic | 88.33% | 71.54% | 92.33% | 97.16% |
| Square | Default | 60.00% | 60.00% | 60.00% | 70.00% |
| | $\Theta$ | 2.38% | 2.38% | 12.08% | 14.80% |
| | Dynamic | 8.33% | 8.33% | 40.00% | 54.66% |

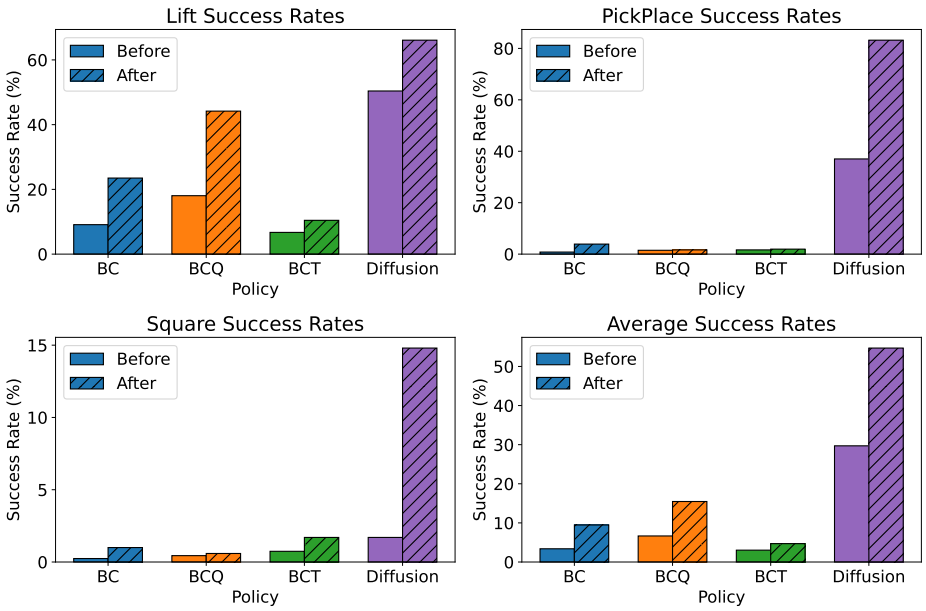


Figure 2: Success rates for each model across Lift, PickPlace, Square, and Average metrics. Solid bars indicate default model and hatched bars indicate after applying Vantage

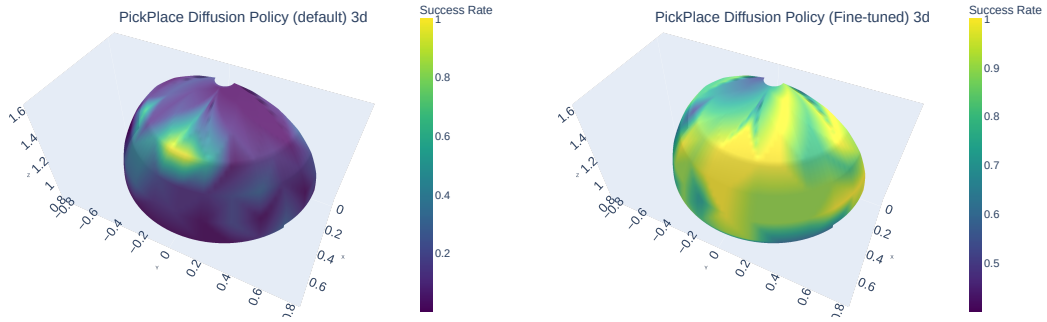


Figure 3: Comparison of default fine-tuned diffusion policies for pick place task across a uniform grid in $\Theta$

## 5 Conclusions and Future Work

We presented *Vantage*, a Bayesian optimization–driven framework for systematically identifying and combining camera viewpoints to enhance the robustness and generalization of vision-based robot policies. By modeling the success rate over the space of possible camera placements with a Gaussian Process and employing an Upper Confidence Bound acquisition strategy, our method efficiently balances exploration of new viewpoints and exploitation of high-performing ones, avoiding the instability caused by indiscriminate diversity in training data.

Our empirical evaluation on a suite of manipulation tasks (Lift, Nut Assembly, Pick & Place) and across multiple policy architectures (BC, BCQ, BCT, Diffusion) demonstrates that fine-tuning on viewpoints selected by Vantage yields consistent and significant performance gains over fixed or heuristic placements. Notably, we observe average improvements ranging from 6% for BC to 46.19% for Diffusion policies on the hyperplane evaluation, with marked robustness in dynamic camera settings. As future work, we will move beyond simulation to validate our approach on real robot platforms, assessing sim-to-real transfer and evaluating performance under real-world sensor noise and lighting variations.

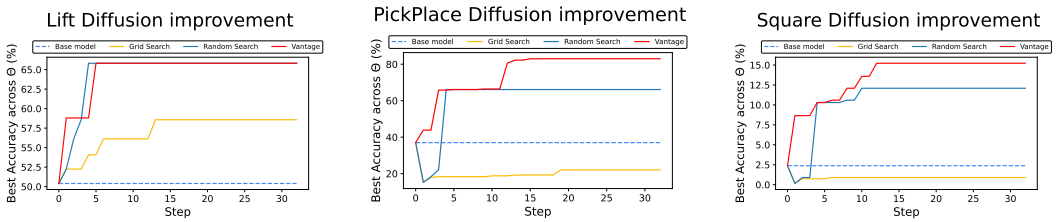


Figure 4: Success rate across Θ of the best model per each model fine-tuned

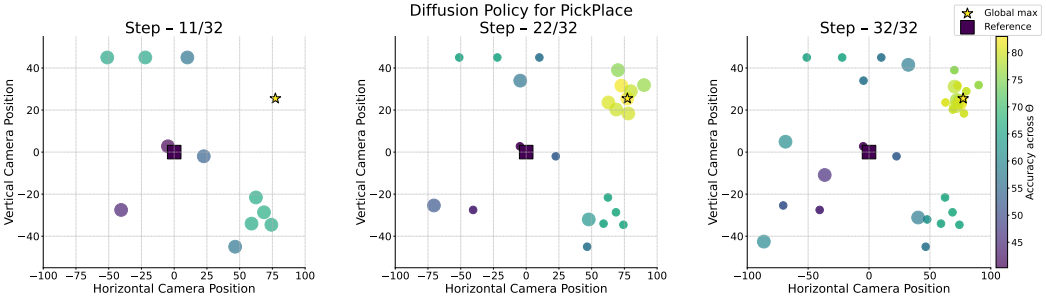


Figure 5: Rollout progression across camera viewpoints and training iterations for each task. Each column illustrate sequential outputs of the Diffusion policy. The plots visualize how task performance evolves as the policy is fine-tuned with optimized viewpoints - bigger circles indicate points from current iteration while smaller ones indicate older iteration

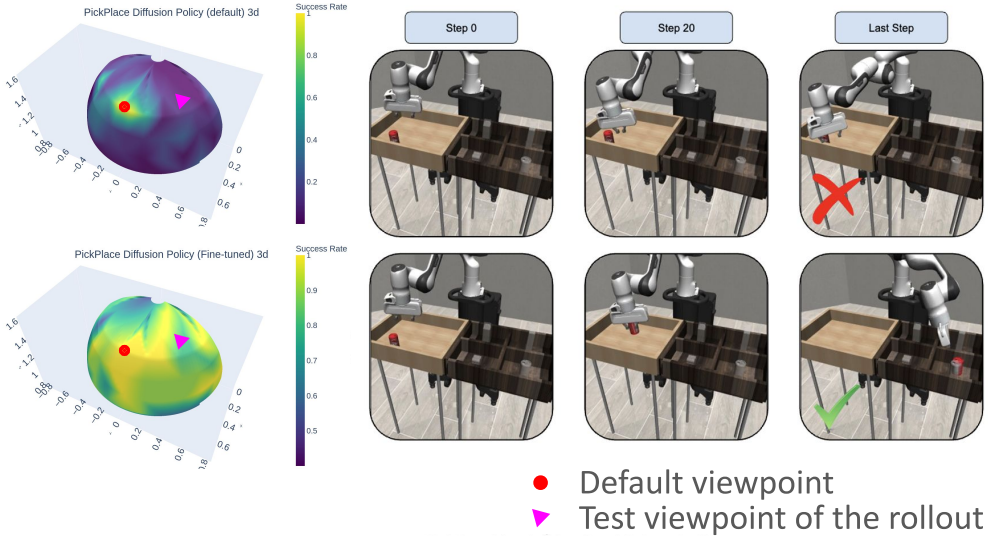


Figure 6: Comparison of the base model and model fine-tuned with Vantage, performing PickPlace task from a different viewpoint.

## References


[1] Soomin Lee, Le Chen, Jiahao Wang, Alexander Liniger, Suryansh Kumar, and Fisher Yu. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[2] Yi-Ting Chen, Cheng-Wei Hsieh, Yu-Ting Hsu, and Yu-Chiang Frank Wang. Gennbv: Generalizable next-best-view policy for active 3d reconstruction. In *CVPR*, 2024.

[3] Xuechao Zhang, Dong Wang, Sun Han, Weichuang Li, Bin Zhao, Zhigang Wang, Xiaoming Duan, Chongrong Fang, Xuelong Li, and Jianping He. Affordance-driven next-best-view planning for robotic grasping. In *Conference on Robot Learning (CoRL)*, 2023.

[4] Yikai Shang and Michael S. Ryoo. Sugarl: Sensorimotor unsupervised goal-aware active reinforcement learning. In *NeurIPS*, 2023.

[5] Yifan Shi, Yuanpei Chen, Yifei Xu, Lianjun Li, Lin Shao, Liyuan Zheng, and Qifeng Chen. Viso-grasp: Vision-language informed spatial object-centric 6-dof active view planning and grasping. *arXiv preprint arXiv:2503.12609*, 2025.

[6] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning requires rethinking generalization. In *International Conference on Learning Representations (ICLR)*, 2017.

[7] Kunal Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. *Journal of Machine Learning Research*, 17(57):1–35, 2013.

[8] 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.

[9] Jingdong Hou, Han Zhang, and Xiaoming Liu. Learning to select views for efficient multi-view understanding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[10] Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

[11] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1990.

[12] Ransalu Senanayake. The role of predictive uncertainty and diversity in embodied ai and robot learning. *arXiv preprint arXiv:2405.03164*, 2024.

[13] Ransalu Senanayake and Fabio Ramos. Bayesian hilbert maps for dynamic continuous occupancy mapping. In *Conference on Robot Learning*, pages 458–471, 2017.

[14] Emily Dunn, Cristian Rodriguez, and Thomas Funkhouser. Learning viewpoint invariant visual representations by predicting views from novel viewpoints. In *ICRA*, 2017.

[15] Soomin Lee, Le Chen, Jiahao Wang, Alexander Liniger, Suryansh Kumar, and Fisher Yu. Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *arXiv preprint arXiv:2209.08409*, 2022.

[16] Herbert Wright, Weiming Zhi, Matthew Johnson-Roberson, and Tucker Hermans. Robust bayesian scene reconstruction by leveraging retrieval-augmented priors. *arXiv preprint arXiv:2411.19461*, 2024.

[17] Ruoyu Yan, Lin Wang, Yuxin Zhang, Yinghuan Shi, and Yang Gao. Active neural mapping. In *ICCV*, 2023.

[18] Martin Frahm and et al. Vin-nbv: A view introspection network for next-best-view planning. *arXiv preprint arXiv:2505.06219*, 2025.

[19] Jingdong Hou, Han Zhang, and Xiaoming Liu. Learning to select views for efficient multi-view understanding. In *CVPR*, 2024.

[20] Xinyi Liu, Tianyi Zhang, Matthew Johnson-Roberson, and Weiming Zhi. Splatraj: Camera trajectory generation with semantic gaussian splatting. *arXiv preprint arXiv:2410.06014*, 2024.

[21] Frederik Vinter-Hviid, Christoffer Sloth, Thiusius Rajeeth Savarimuthu, and Iñigo Iturrate. Safe contact-based robot active search using bayesian optimization and control barrier functions. *Frontiers in Robotics and AI*, 11:1344367, 2024.

[22] Akshay K. Burusa, Eldert J. van Henten, and Gert Kootstra. Attention-driven next-best-view planning for efficient reconstruction of plants and targeted plant parts. In *Biosystems Engineering*, volume 248, pages 248–262, 2024.

[23] Xuechao Zhang, Dong Wang, Sun Han, Weichuang Li, Bin Zhao, Zhigang Wang, Xiaoming Duan, Chongrong Fang, Xuelong Li, and Jianping He. Affordance-driven next-best-view planning for robotic grasping. In *Conference on Robot Learning (CoRL)*, 2023.

[24] Dinesh Jayaraman and Kristen Grauman. Learning viewpoint-invariant visual representations by predicting views from novel viewpoints. *CVPR*, 2018.

[25] Xiaoyu Wu and colleagues. Neural next-best-view planning for active 3d reconstruction. In *CVPR*, 2023.

[26] Ying Lin and colleagues. Neural implicit active vision for scene understanding. In *ICCV*, 2023.

[27] Harnaik Dhami, Vishnu D. Sharma, and Pratap Tokekar. Map-nbv: Multi-agent prediction-guided next-best-view planning for active 3d object reconstruction. *arXiv preprint arXiv:2307.04004*, 2023.

[28] Harnaik Dhami, Vishnu D. Sharma, and Pratap Tokekar. Pred-nbv: Prediction-guided next-best-view planning for 3d object reconstruction. *arXiv preprint arXiv:2307.04004*, 2023.

[29] Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *International Conference on Learning Representations (ICLR)*, 2021.

[30] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? In *Proceedings of the 36th International Conference on Machine Learning*, pages 5389–5400, 2019.

[31] Stuart Geman, Elad Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4(1):1–58, 1992.

[32] Alexander Khazatsky, Karl Pertsch, Suraj Nair, and *et al.* DROID: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

[33] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, 25, 2012.

[34] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[35] James T. Wilson, Riccardo Moriconi, Frank Hutter, and Marc Peter Deisenroth. The reparameterization trick for acquisition functions, 2017.

[36] Yuke Zhu, Josiah Wong, Ajay Mandlekar, Roberto Martín-Martín, Abhishek Joshi, Kevin Lin, Abhiram Maddukuri, Soroush Nasiriany, and Yifeng Zhu. robosuite: A modular simulation framework and benchmark for robot learning, 2025.

[37] Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. In *arXiv preprint arXiv:2108.03298*, 2021.

[38] Michael Bain and Claude Sammut. A framework for behavioural cloning. In *Machine intelligence 15*, pages 103–129, 1995.

[39] Scott Fujimoto, David Meger, and Doina Precup. Off-policy deep reinforcement learning without exploration. In *International conference on machine learning*, pages 2052–2062. PMLR, 2019.

[40] 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, 2024.

[41] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias W. Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, May 2012.

# 6 Appendix

## 6.1 Theoretical Guarantees

Imagine a *black-box* function which, for each fine-tuning viewpoint $\theta$, returns the robot's success rate when evaluated across a uniform grid spanning the entire allowable camera placement space $\Theta$. Initially, we know almost nothing about this function, so we sample a few random angles and observe the outcomes. Each time we fine-tune the policy from a new angle, we learn more about which regions of the viewing space are most informative and which angles yield high success rates. By fitting a smooth probabilistic model (a Gaussian Process) over these observations, we can predict both the expected success rate at untested angles and how uncertain those predictions are. The GP-UCB rule then lets us pick the next angles that strike the best balance between exploring uncertain regions and exploiting angles already known to work well. Over successive rounds, this strategy systematically steers us toward the optimal viewpoints, guaranteeing that we improve faster than if we had naively tried every angle or picked random ones.

Every fine-tuning step at a new viewpoint yields a measured success rate for the policy at that angle. We treat these measured success rates as noisy evaluations of an underlying true function $f(\theta)$. When we update the GP with a new observation $(\theta_t, J_t)$, the surrogate's posterior mean $\mu_t(\theta)$ and standard deviation $\sigma_t(\theta)$ reflect our current best guess of $f$ and our uncertainty. The GP-UCB acquisition function

$$\mu_t(\theta) + \beta_t \, \sigma_t(\theta)$$

scores each candidate angle by combining its predicted success rate (exploitation) with a bonus for uncertainty (exploration). By fine-tuning at the angle maximizing this score, we both collect data where we expect high success and reduce uncertainty where the model is least certain, thereby directly linking each fine-tuning experiment to measurable improvements in success rate and ensuring theoretical guarantees on cumulative performance.

In the single-point GP-UCB analysis, the information gain $\gamma_T$ quantifies the maximum reduction in uncertainty one can achieve by making $T$ sequential observations. In our batched $q$-UCB procedure, we collect $q$ observations per round for $T$ rounds, yielding a total of $qT$ queries to the underlying function. Accordingly, we must replace the sequential information gain $\gamma_T$ with the batch information gain $\gamma_{qT}$, which measures the maximum mutual information between the GP prior and $qT$ noisy evaluations. Since a GP's information gain grows sublinearly in the number of observations, this substitution preserves the same sublinear regret behavior: the cumulative batch regret is controlled by

$$qT \, \gamma_{qT} \, \beta_T,$$

just as in the single-point case but scaled to account for the larger total sample size.

**Theorem 6.1** (GP-UCB Regret Bound)**.** *Let $f : \Theta \to \mathbb{R}$ be a mapping from angles to success rates, drawn from a Gaussian process prior with kernel $k$. At each round $t = 1, \ldots, T$ choose,*

$$\theta_t \;=\; \arg\max_{\theta \in \Theta} \Big[ \mu_{t-1}(\theta) \;+\; \sqrt{\beta_t} \, \sigma_{t-1}(\theta) \Big],$$

*and observe $y_t = f(\theta_t) + \varepsilon_t$ with $\varepsilon_t$ zero-mean sub-Gaussian noise. Then, with probability at least $1 - \delta$,*

$$R(T) \;=\; \sum_{t=1}^{T} \sum_{i=1}^{q} \big[ f(\theta^*) - f(\theta_{t,j}) \big] \;=\; O\big( \sqrt{qT \, \gamma_{qT} \, \beta_T} \big),$$

*where $\theta^* = \arg\max_\Theta f$, $\gamma_T$ is the maximum information gain after $T$ steps, and $\beta_T$ is chosen as in [41].*

*Proof.* By Theorem 2 of Srinivas *et al.* (2012), with probability $1 - \delta$, for all $t$ and all $\theta$,

$$\big| f(\theta) - \mu_{t-1}(\theta) \big| \;\le\; \sqrt{\beta_t} \, \sigma_{t-1}(\theta).$$

Hence the instantaneous regret $r_t = f(\theta^*) - f(\theta_t)$ satisfies

$$r_t \;\le\; 2 \sqrt{\beta_t} \, \sigma_{t-1}(\theta_t).$$

Summing and applying Cauchy–Schwarz together with the definition of information gain gives

$$R(T) \;=\; \sum_{t=1}^{T} r_t \;\le\; 2\sum_{t=1}^{T} \sqrt{\beta_t}\,\sigma_{t-1}(\theta_t) \;\le\; 2\sqrt{\big(\sum_t \beta_t\big)\big(\sum_t \sigma_{t-1}^2(\theta_t)\big)} \;=\; O\big(\sqrt{T\,\gamma_T\,\beta_T}\big).$$

□

**Theorem 6.2** (Average Success Convergence)**.** *Under the same setting as Theorem 6.1, with probability at least* $1-\delta$*,*

$$\frac{1}{T}\sum_{t=1}^{T} J\big(\pi_{\theta_t}\big) \;\ge\; J\big(\pi_{\theta^*}\big) \;-\; O\Big(\sqrt{\tfrac{\gamma_T\,\beta_T}{T}}\Big).$$

*In particular, the mean success converges to the optimum at rate* $O(T^{-1/2})$*.*

*Proof.* From Theorem 6.1, with high probability,

$$\sum_{t=1}^{T}\big[J(\pi_{\theta^*}) - J(\pi_{\theta_t})\big] \;=\; O\big(\sqrt{T\,\gamma_T\,\beta_T}\big).$$

Divide both sides by $T$ and rearrange:

$$J(\pi_{\theta^*}) \;-\; \frac{1}{T}\sum_{t=1}^{T} J(\pi_{\theta_t}) \;=\; O\Big(\sqrt{\tfrac{\gamma_T\,\beta_T}{T}}\Big),$$

which yields the stated bound. □

**Theorem 6.3** (Rademacher Complexity Generalization)**.** *Let* $\mathcal{F}$ *be a class of indicator functions with Rademacher complexity* $\mathfrak{R}_n(\mathcal{F})$*. If* $\widehat{J}(\pi_\theta)$ *is the empirical success rate over* $n$ *i.i.d. trials at viewpoint* $\theta$*, then for any* $\delta > 0$*, with probability at least* $1-\delta$*, for all* $\theta$

$$\big|J(\pi_\theta) - \widehat{J}(\pi_\theta)\big| \;\le\; 2\,\mathfrak{R}_n(\mathcal{F}) \;+\; \sqrt{\frac{\ln(2/\delta)}{2n}}.$$

*Proof.* Apply Theorem 3.1 of Mohri *et al.* (2018) on uniform convergence: with probability $1-\delta$,

$$\sup_{f\in\mathcal{F}}\big|\mathbb{E}[f] - \widehat{\mathbb{E}}[f]\big| \;\le\; 2\,\mathfrak{R}_n(\mathcal{F}) \;+\; \sqrt{\frac{\ln(2/\delta)}{2n}}.$$

Setting $f(\cdot) = \mathbf{1}\{\text{success}\}$ for each $\pi_\theta$ gives the result. □

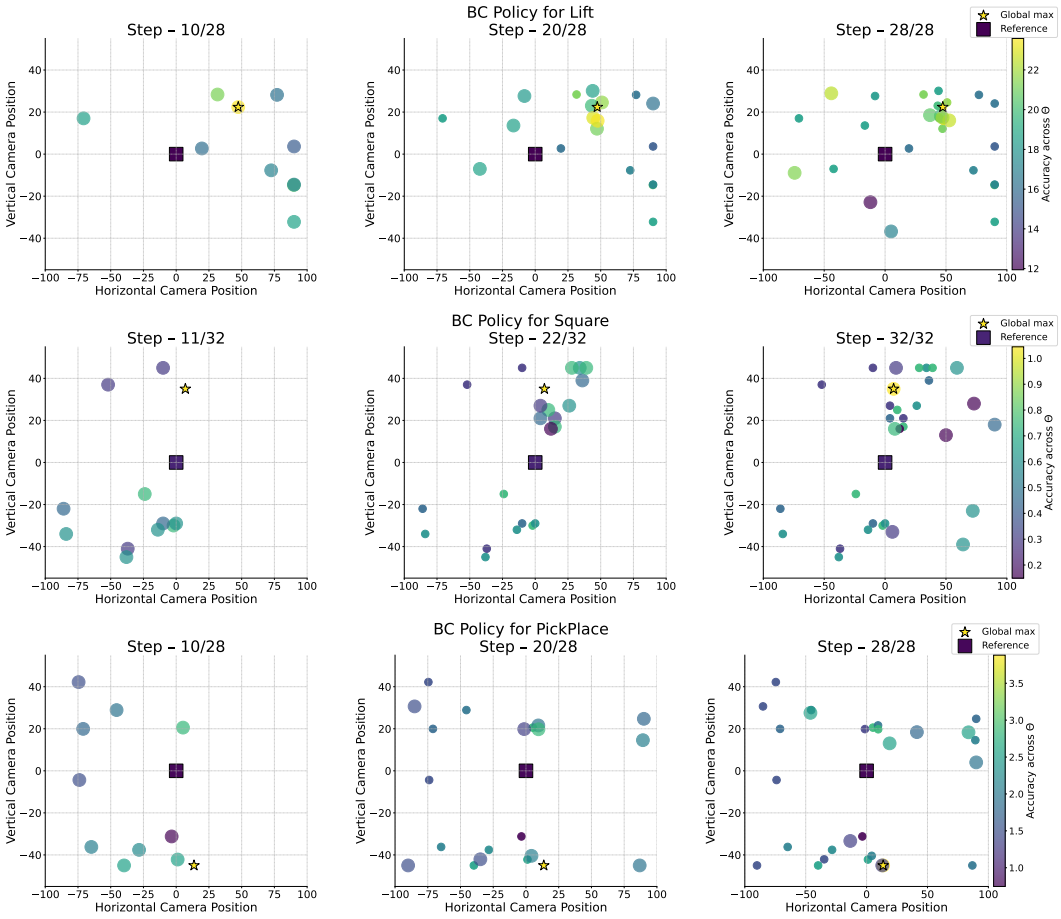


Figure 7: Rollout progression across camera viewpoints and training iterations for each task. Each row corresponds to a different manipulation task, namely Lift, Square and PickPlace while columns illustrate sequential outputs of the BC policy. The plots visualize how task performance evolves as the policy is fine-tuned with optimized viewpoints - bigger circles indicate points from current iteration while smaller ones indicate older iteration

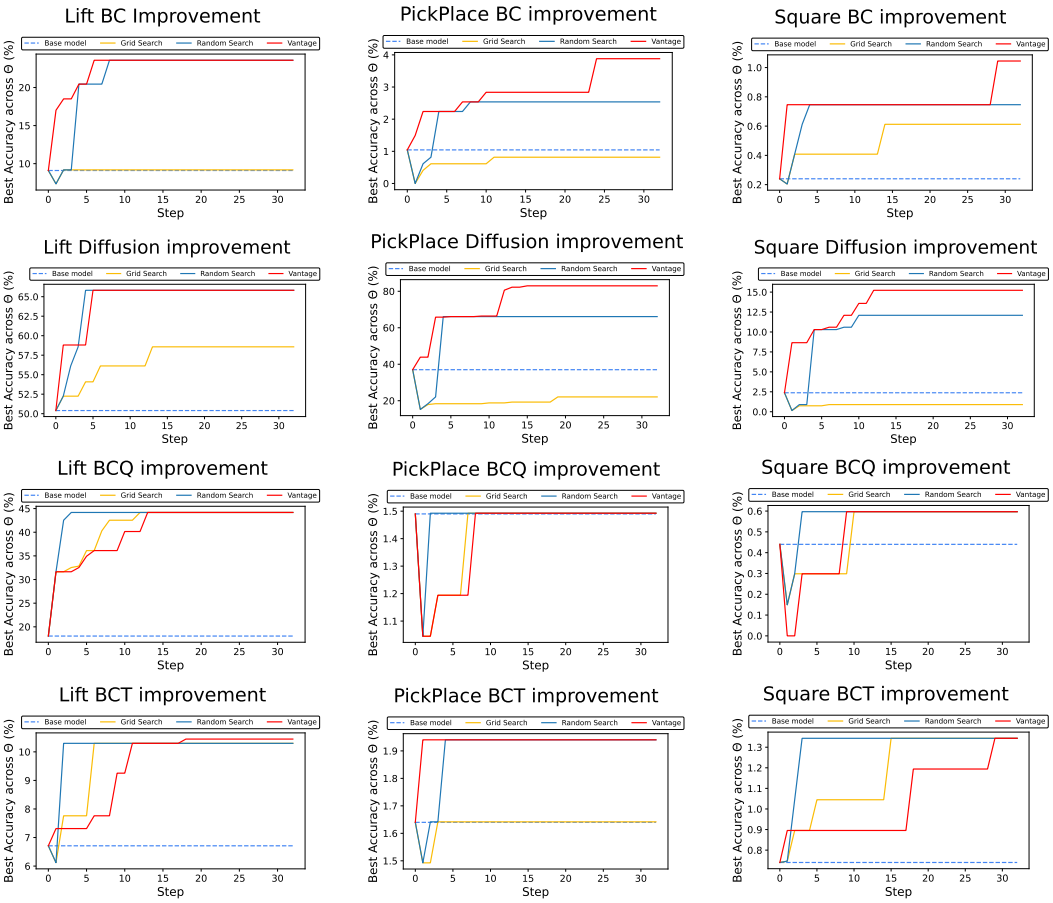


Figure 8: Success rate across Θ of the best model per each model fine-tuned (all experiments)