Extended Abstract

Motivation As one of the frontiers in healthcare, surgical robotics promises enhanced precision and improved patient outcomes through intelligent robotic automation. Surgical tasks often involve multiple identical or visually similar objects where the target cannot be identified from environmental context alone. This necessitates goal-conditioned approaches that explicitly specify task objectives.

To address this need, we investigate how different goal-conditioning methods affect learning in ambiguous robotic manipulation using a goal-conditioned peg transfer task where a robot must move a specified block among multiple colored blocks on a board.

Method We create an ambiguous peg transfer task, with two block and four block variations, to test goal-conditioning surgical robotic scenarios. The robot must move a goal block from its initial position to a target peg. Randomization of goal block position and color encourages the policy to utilize provided goal-conditioning features to determine task specification.

We generate heuristic-based demonstrations for training, including robot state and either (1) all block positions or (2) all block positions and colors as observations. We post-process demonstrations with different goal encodings, covering spatial and semantic types with varying complexity. These include goal block position, target peg position, one hot encoding of the goal block, and RGBA color encoding of the goal block.

We establish baseline performance using Behavior Cloning (BC), revealing limitations that motivate reinforcement learning. We then train using Deep Deterministic Policy Gradient (DDPG) with Hindsight Experience Replay (HER) and heuristic demonstrations.

Implementation We compare our goal-conditioning encoding across our ambiguous peg transfer task with two block and four blocks, and across observations with and without block colors. Specifically, we test five goal encoding combinations: one hot encoding of the goal block, one hot encoding of the goal block and target peg position, RGBA color of the goal block, goal block position, and goal block and target peg positions.

Results Our behavior cloning experiments reveal challenges in goal-conditioned imitation learning, demonstrating that imitation learning can require extensive data yet fail to generalize across task variations. BC achieved only 10% evaluation success with high run-to-run variability, indicating instability in learning goal-conditioned policies from demonstrations alone. In contrast, we see higher evaluation success across all goal encodings when trained with DDPG with HER and demonstrations.

Spatial conditioning with DDPG with HER consistently outperforms semantic approaches, with performance gaps increasing as task complexity grows. In a two-block peg transfer environment, spatial methods achieve 60% success using DDPG with HER while semantic methods reach 20-40%. In a four-block environment, spatial methods maintain 40-50% success while semantic methods fail completely. Qualitative video analysis reveals semantic methods fail primarily due to manipulation control problems, and in the more complex four-block environment, may also fail due to goal identification issues.

Discussion Our experiments reveal key insights: (1) Goal-conditioned BC suffers from overfitting with substantial training-evaluation gaps, demonstrating the value of active learning approaches like DDPG with HER for robust surgical robotic manipulation. (2) Spatial representations are more effective than semantic representations. Spatial coordinates provide actionable information, whereas semantic methods create representation gaps that agents struggle to bridge in complex environments. (3) Task complexity fundamentally affects conditioning method viability. Failure modes transition from manipulation execution issues to goal identification breakdown as complexity increases. (4) Color information improves semantic method performance in simple environments but degrades in complex ones. (5) Robot behavior degrades with increased task complexity.

Conclusion Our findings highlight the importance of representation choice in goal-conditioned reinforcement learning and provide insights for designing effective goal specification mechanisms in autonomous surgical robotic systems.

Goal-Conditioned Reinforcement Learning for Surgical Robotic Manipulation

Daphne Barretto
Department of Computer Science
Stanford University
daphnegb@stanford.edu

Alycia Lee
Department of Computer Science
Stanford University
alylee15@stanford.edu

Elsa Bismuth
Institute for Computational and Mathematical Engineering
Stanford University
elsabis@stanford.edu

Abstract

Surgical robotic systems must be capable of handling ambiguous manipulation tasks where multiple similar objects require explicit goal specification. We investigate how different goal-conditioning methods affect learning in ambiguous robotic manipulation using a goal-conditioned peg transfer task where a robot must move a specified block among multiple colored blocks on a board. We compare conditioning on spatial goal representations, i.e. 3D coordinates, with semantic representations, such as one-hot encodings and color specifications, across two peg transfer tasks of differing complexity using DDPG with HER. Behavior cloning experiments establish baseline performance limitations, with pure imitation learning achieving only 10% success despite extensive demonstration data, motivating our focus on reinforcement learning approaches. Spatial conditioning consistently outperforms semantic approaches, with performance gaps increasing as task complexity grows. In a two-block peg transfer environment, spatial methods achieve 60% success while semantic methods reach 20-40%. In a four-block environment, spatial methods maintain 40-50% success while semantic methods fail completely. Qualitative video analysis reveals semantic methods fail primarily due to manipulation control problems, and in the more complex four-block environment, may also fail due to goal identification issues. These results demonstrate that spatial coordinates offer direct, actionable guidance for robot control and task execution, whereas semantic representations may impose abstraction challenges that worsen with greater task complexity. Our findings offer important insights for the design of goal-conditioned surgical robotic systems.

1 Introduction

As one of the frontiers in healthcare, surgical robotics promises enhanced precision and improved patient outcomes through intelligent robotic automation. While currently deployed surgical robots rely on surgeons controlling the robotic arm from a console, future generations must possess the ability to not only autonomously carry out tasks, but also to handle ambiguity inherent in real-world surgical environments, and thereby generalize across varying scenarios (Lee et al. (2024); Liu et al. (2024); Power (2024)). Surgical tasks often involve multiple identical or visually similar objects where the target cannot be identified from environmental context alone. This necessitates goal-conditioned approaches that explicitly specify task objectives.

Goal-conditioned reinforcement learning addresses this challenge by conditioning policies on goal representations, enabling agents to pursue different objectives within the same environment (Liu et al. (2022)). How this conditioning affects learning and performance in ambiguous manipulation tasks could depend on different methods of goal specification, such as spatial coordinates and semantic representations. The selection of goal representation has implications for not only robot performance and learning efficiency, but also real-world applicability, particularly for more complex tasks.

Spatial representations, such as 3D coordinates, provide actionable information that directly corresponds to robot control commands. In contrast, semantic representations, such as categorical one-hot encodings or color specifications, require the agent to bridge an abstraction gap between high-level task descriptions and low-level motor control. In this work, we systematically investigate how different goal conditioning methods affect learning and performance in ambiguous robotic manipulation tasks representative of surgical scenarios. We create an ambiguous peg transfer task by introducing multiple potential goal objects to move to a designated target peg.

We first establish baseline performance using behavior cloning to understand the fundamental challenges in goal-conditioned imitation learning. Initial behavior cloning experiments achieve only 10% success despite extensive demonstration data. This critical finding motivates our transition to reinforcement learning approaches and informs our understanding of representation effectiveness. Our experimental framework then compares spatial goal representations against semantic representations across varying levels of task complexity using Deep Deterministic Policy Gradient (DDPG) with Hindsight Experience Replay (HER).

Our findings using DDPG with HER reveal that while spatial conditioning methods consistently outperform semantic approaches, the performance gap varies with task complexity and observation richness. We demonstrate how task complexity affects the viability of different conditioning approaches and provide insights into the underlying causes of policy failures via qualitative analysis, distinguishing between high-level task understanding deficits and low-level motor control limitations. We identify several failure modes and patterns across different conditioning methods and task complexities, such as goal identification and block manipulation failures. Our findings have immediate implications for the design of goal-conditioned surgical robotic systems and highlight important research directions for designing task specifications for effective robotic control policies.[1]

2 Related Work

Peg transfer is a task listed in the Fundamentals of Laparoscopic Surgery where an agent—human or machine—controls a robot arm to move objects from one peg to another peg, demonstrating precise and dexterous technical skills in surgical manipulation.[2] Automated surgical robotic manipulation for this task has shown success utilizing DDPG with HER and demonstrations, where DDPG provides continuous control for robotic manipulation, HER assists learning in sparse reward tasks, and demonstrations guide exploration during learning by using Q-filtered behavior cloning (Xu et al. (2021); Long et al. (2023); Chiu et al. (2021); Richter et al. (2019)). These reinforcement learning techniques build upon standard imitation learning for surgical robotic manipulation (Huang et al. (2023a,b)).

However, these existing approaches all assume the agent's goals can be uniquely identified from environmental context alone. This breaks down in surgical scenarios with multiple identical or similar objects where current methods cannot distinguish which specific object to manipulate without additional goal-conditioning from another source. Preliminary work has shown success in hierarchical reinforcement learning (Huang et al. (2023b)) and transformer-based methods (Fu et al. (2024)), which both use intermediate goal-conditioned tasks to solve a larger specific goal. Additionally, Lingaraju (2023) discusses goal-conditioning for overall task goals via concatenating the agent's observation with its desired goal, but does not explore this concept further. In other domains, goal-conditioning has been explored in this direct concatenation method as well as indirectly via language and video encoders (Jang et al. (2022); Kim et al. (2024); Intelligence et al. (2025)).

[1] Code is provided at `github.com/daphne-barretto/SurRoL/tree/ddpg_with_her` and `github.com/daphne-barretto/SurRoL/tree/bc` for DDPG with HER and BC experiments respectively.

[2] `sages.org/wiki/fundamentals-laparoscopic-surgery/`

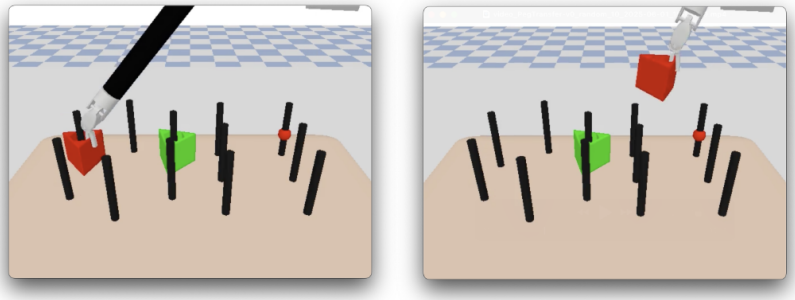

Figure 1: Sample task execution for the two-block environment, where goal-conditioning indicates the red block should be moved to the target peg (labeled with red sphere). (Left) The robot arm grasps the red block. (Right) The robot arm moves the red block towards the target peg.

Building upon these ideas, we investigate goal-conditioned reinforcement learning for surgical robotic manipulation on ambiguous peg transfer tasks, where the goal must be specified from outside the environment.

3 Method

3.1 Goal-Conditioned Peg Transfer Task

We create an ambiguous peg transfer task to test goal-conditioning in surgical robotic scenarios, building upon the Surgical Robot Learning (SurRoL) simulator and its peg transfer environment (Xu et al. (2021)). In the task, the robot must move a goal block from its initial position to a target peg. We create two variations of our goal-conditioned peg transfer task with differing levels of complexity (Fig. 1, 2):

1. Two-block environment: contains two blocks colored red and green
2. Four-block environment: contains four blocks colored red, green, blue, and yellow

Each episode begins with the specified blocks randomly distributed across six pegs on the left side of the board and with a randomly selected target peg from the six pegs on the right side of the board. The robot must correctly identify the designated goal block, pick it up with its robot arm, and transfer it to a target peg on the right side in order for the episode to be considered a success. Both the goal block and target peg are randomized in each episode. While the target peg is provided to the agent, the randomization of the goal block encourages the robot to utilize the goal-conditioning features to determine the complete task specification.

From the environment information alone, this task is ambiguous such that while the robot can observe all block positions and colors (depending on the experiment), it cannot determine which specific block needs to be moved to the target peg without additional goal conditioning information. This is reflective of real-world surgical scenarios where multiple identical or visually similar objects are present and the system must rely on external specifications to determine which to manipulate for a given circumstance.

3.2 Demonstration Data Generation for Observations

We generate heuristic-based demonstrations for training our agent policies. To do this, we run our environments with heuristic processes and save successful demonstrations with information on the target peg and varying observations until the specified number of demonstrations are successful and saved. Unlike the target peg, information specifying the goal peg is only provided indirectly through the observations.

For our goal-conditioned peg transfer task, the heuristic process moves the robot arm above the goal block (which it directly has information for), lowers onto the goal block, grasps the goal block, lifts

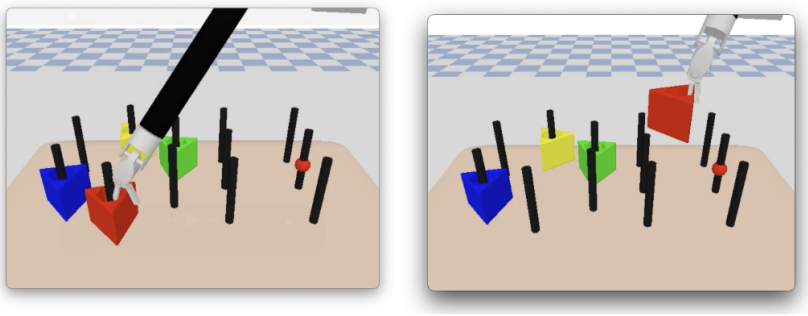

Figure 2: Sample task execution for the four-block environment, where goal-conditioning indicates the red block should be moved to the target peg (labeled with red sphere). (Left) The robot arm grasps the red block. (Right) The robot arm moves the red block towards the target peg.

up the goal block, moves the robot arm with the goal block above the target peg with an alignment offset, and releases the goal block onto the target peg.

In our demonstration data, the default provided observations are robot state, which captures the complete pose (position and orientation) of the robot's end-effector tip as well as the gripper's jaw opening state (jaw angle). We then concatenate the target block position, all block positions in (x, y, z) format, or all block positions in (x, y, z) format and colors in RGBA [R, G, B, A] format—depending on the experiment. There is still ambiguity about which block is the goal block when provided all block positions and all block positions and colors, which is resolved via our goal encodings from post-processing. Separate from the observations for the agent, we save a one hot encoding of the goal block for each demonstration in the file for our own post-processing.

3.3 Post-Processing for Goal Encodings

We post-process our generated data demonstrations with different goal encodings, covering both spatial and semantic encoding types. These goal encodings have a range of complexity and abstraction.

1. Spatial information:
   (a) Goal block position: the (x, y, z) coordinates of the goal block
   (b) Target peg position: the (x, y, z) coordinates of the target peg
2. Semantic information:
   (a) A one hot encoding of the goal block, e.g. [0, 1, 0, 0] if there are four blocks
   (b) The RGBA color encoding of the goal block, e.g. [1, 0, 0, 1] for red

The goal block position is the most direct information we could encode for goal-conditioning. The positional information itself is redundant in that the observations already provide block positions, but in an arbitrary order such that it is not specified which position is associated with the goal block. The goal block position then serves as an identifiable encoded location for the goal block, which could result in the block position observations themselves being ignored.

The target peg position provides direct information about the target position, but does not assist in disambiguating the goal block. This is used in combination with the other goal encodings.

The one hot encoding identifies the goal block from the given observations. For instance, if the observations contain the position for four blocks and the goal block is the second one, the one hot encoding would be [0, 1, 0, 0]. This encoding still requires the agent to use the spatial information from the observation itself, and instead identifies the goal block from this existing spatial information.

The RGBA color encoding indirectly identifies the goal block from the given observations. The RBGA color itself is redundant in that the observations containing colors already provide RGBA color for all blocks. In this case, the agent would have to learn to match the goal block color to the list of observed block positions and colors to find the goal block position.

3.4 Learning algorithms

3.4.1 Behavior Cloning

We use Behavior Cloning (BC) with goal-conditioning utilizing our successful demonstrations to train policies to imitate these behaviors. By testing BC, we can establish baseline BC performance and validate the necessity of reinforcement learning approaches.

3.4.2 DDPG with HER and Demonstrations

We use Deep Deterministic Policy Gradient (DDPG) with Hindsight Experience Replay (HER) and demonstrations to learn continuous control problems with sparse rewards. DDPG is an actor-critic algorithm designed for continuous action spaces, using separate actor-critic networks to learn both a policy and a value function respectively. HER addresses the challenge of learning from sparse rewards by relabeling failed experiences as successful ones. For instance, during training, when an agent fails to reach the intended goal, HER retrospectively treats the state it actually reached as if it were the true goal, thereby creating additional training data from every episode. This approach is particularly valuable in goal-conditioned tasks like the one we evaluate, where the agent might have some success with low-level control and manipulation of objects, but fail to achieve the specific goal. In this case, HER enables the agent to learn useful control skills even from its failed attempts.[3] Building on our BC work, we use our heuristic-based demonstrations to guide exploration during learning with Q-filtered behavior cloning (Nair et al. (2018)).[4]

4 Experimental Setup

4.1 Behavior Cloning

We ran experiments to establish baseline BC performance and validate the necessity of reinforcement learning approaches. We trained three runs of each experiment and then conduct 100-episode evaluations, and we report the training and evaluation success. Each run is trained to 100 epochs with batch size 64 and 5,000 demonstrations.

We trained BC for our ambiguous peg transfer task with two blocks. We tested two observation variants:

1. Robot state and all block positions
2. Robot state, all block positions, and all block colors

We also tested two goal encoding combinations:

1. Goal block and target peg positions, labeled "Spatial"
2. One hot encoding of the goal block and target peg position, labeled "Semantic"

4.2 DDPG with HER and Demonstrations

We ran experiments to investigate goal-conditioned reinforcement learning across observation variants and goal encodings on our ambiguous peg transfer task. We trained three runs of each experiment, and report the evaluation success averages and standard deviations. Each run of our goal-conditioning experiments is trained to 100 epochs.

4.2.1 Preliminary Work: Demonstration Count

In preliminary work, we determined the number of demonstrations to use to train DDPG with HER. Figure 3 shows that DDPG with HER performance is comparable when trained on either 1,000

[3] In our implementation, the target peg position is provided through a labeled "desired goal" which is differentiated from the relabeled "achieved goal" that HER uses for training on failures. Both the achieved and desired goals are passed as inputs to the DDPG networks during training.

[4] In later sections, we refer to this algorithm as "DDPG with HER" without explicitly mentioning the demonstrations used in training. Demonstrations were used for guided exploration unless otherwise noted.

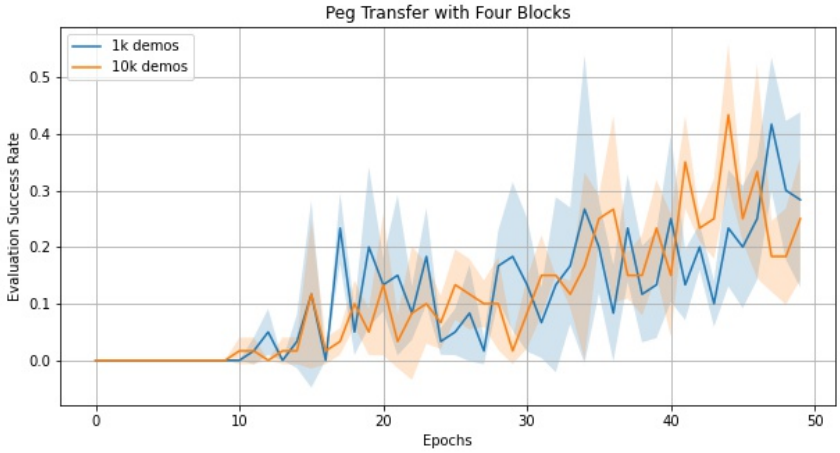

Figure 3: Evaluation success rates of DDPG with HER trained on 1,000 and 10,000 demonstrations on our goal-conditioned peg transfer task with four blocks. The final performance is comparable, so we used 1,000 demonstrations in goal-conditioned DDPG with HER experiments.

or 10,000 demonstrations, so we used 1,000 demonstrations for all following DDPG with HER experiments. We selected these values based on past behavior cloning experiments on the Needle Reach environment from SurRoL, which is similar to peg transfer tasks in that it requires the robot to move itself to a target position.

4.2.2 Baseline

We train a baseline on both the two block environment and the four block environment, with observations that include only the robot state and the goal block position, with no information about other block positions or any block colors. This baseline is used to validate the feasibility of the task when the goal is directly provided and no ambiguity is introduced, as well as provide success rate comparisons across learning.

4.2.3 Goal-Conditioning Experiments

We trained DDPG with HER and heuristic demonstrations for our ambiguous peg transfer task with two and four blocks.

For both environments, we tested two observation variants:

1. Robot state and all block positions
2. Robot state, all block positions, and all block colors

For both environments, we also tested five goal encoding combinations:

1. One hot encoding of the goal block
2. One hot encoding of the goal block and target peg position
3. RGBA color of the goal block
4. Goal block position
5. Goal block and target peg positions

By combining one hot block encoding with target peg position, we evaluate whether adding spatial context helps ground semantic information. By combining goal block and target peg positions, we test whether providing complete spatial context (both current and destination positions) offers improved conditioning by giving the agent full spatial awareness of the task at hand. Testing these different goal-conditioning methods across environment complexities and observation types enables us to investigate which conditioning approaches are most robust and generalizable.

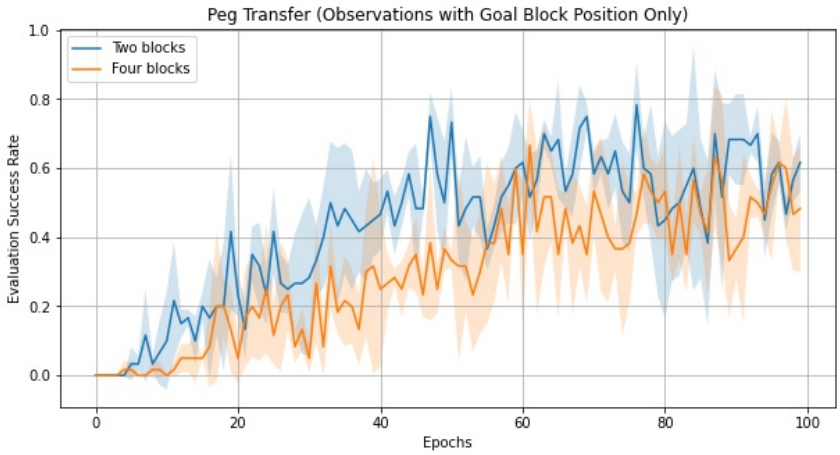

Figure 4: Evaluation success rates of DDPG with HER vs. training epochs on the ambiguous peg transfer task where observations include robot state and target block position. No goal-conditioning was performed.

5 Results

5.1 Behavior Cloning

Our BC experiments achieved maximum success rates of only 25% in training, with severe generalization challenges in evaluation (best evaluation performance: 10%). These initial results show that while BC provides useful baseline comparisons, reinforcement learning approaches may be more robust and reliable for completing the goal-conditioned peg transfer tasks.[5]

5.2 DDPG with HER Baseline

Our DDPG with HER baseline utilizes the most direct spatial information. The results clearly demonstrate task solvability for both two block and four block complexity levels. Figure 4 shows that on the two-block environment, DDPG with HER achieves a success rate of about 60% by epoch 100. On the four-block environment, it reaches a comparable success rate of around 50% with notably slower learning efficiency. This performance difference aligns with our expectations, since the four-block environment introduces additional complexity, even when the goal block position is explicitly provided. This result also establishes an approximate upper bound for spatial conditioning methods, since directly providing the goal block position represents the most actionable spatial information.

5.3 DDPG with HER Goal Conditioning Experiments: Two Blocks

The ambiguous peg transfer task with two blocks reveals important patterns in the effectiveness of goal-conditioning approaches. As shown in Figure 5, conditioning on spatial information generally outperforms conditioning on semantic information. Specifically, conditioning on goal block position achieves the highest success rates, reaching approximately 60% by epoch 100 across both observation variants.

Semantic methods show limited performance in the two-block environment. Both one-hot block encoding and RGBA color encoding achieve modest success rates of around 20% by the end of training when no color information is included in the observation, and 30-40% when color information is included. This suggests that with fewer objects in the environment, semantic representations can be somewhat learnable, though they still underperform spatial methods. Importantly, we see that when color information is added, these semantic methods learn faster and achieve higher performance. This makes sense as these semantic conditioning features directly correspond to the color information in

[5]Additional details about BC experimental results and analysis are provided in Appendix A.

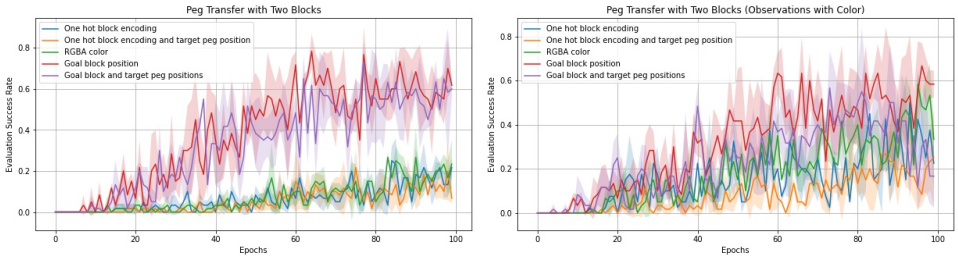

Figure 5: Evaluation success rates of DDPG with HER vs. training epochs on on the goal-conditioned peg transfer task with two blocks. Observations include robot state and all block positions (left), and also include all block colors (right).

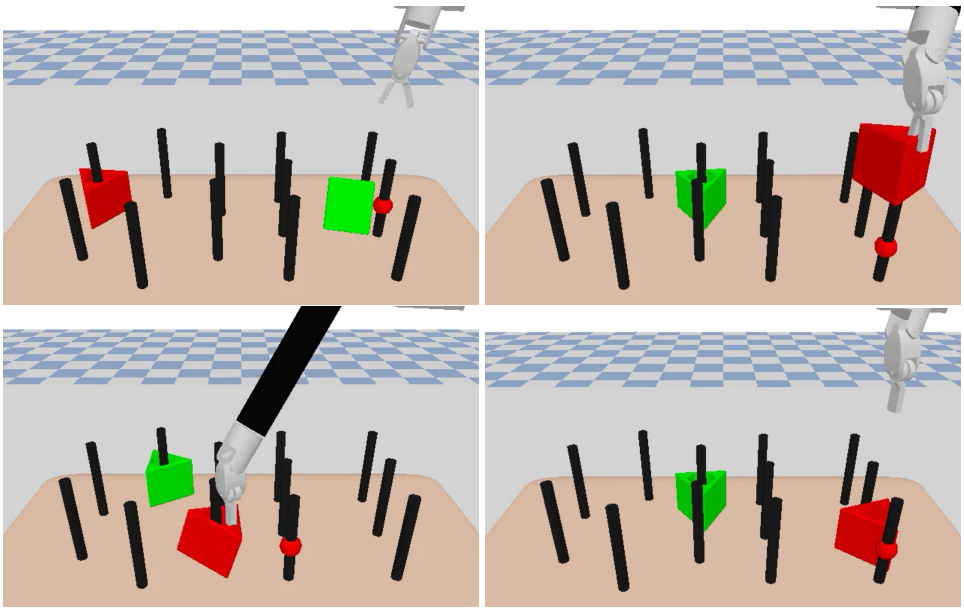

Figure 6: Failure modes of evaluation episodes on the goal-conditioned peg transfer task with two blocks, which are largely manipulation failures. (top left) Goal block is dropped near the target peg. (top right) Robot hovers above target peg without dropping block onto the peg before the episode terminates. (bottom left) Robot struggles to grasp the goal block. (bottom right) Robot drops the goal block near but not on the target peg.

the observation space, and the policy learns to connect these features with the colors of all blocks on the board to determine which block is the goal. However, even with this correspondence advantage, these semantic methods are still outperformed by spatial conditioning.

Combining one hot block encoding with target peg position achieves the worst performance across both observation variants, likely due to the additional complexity introduced when semantic and spatial representations are concatenated in the observation, which may lead to confusion during learning.

Video analysis of evaluation episodes reveals distinct behavioral patterns between conditioning methods and observation conditions. When observations include block color information, both one-hot block encoding and RGBA color conditioning occasionally achieve successful task completion (link to video). However, analysis of failed episodes shows that while goal identification is typically successful, both methods face consistent manipulation challenges. Figure 6 shows depictions of failure modes described below.

Conditioning on the one hot block encoding typically succeeds in identifying and grasping the correct goal block but struggles with precise placement, frequently dropping blocks near rather than on the

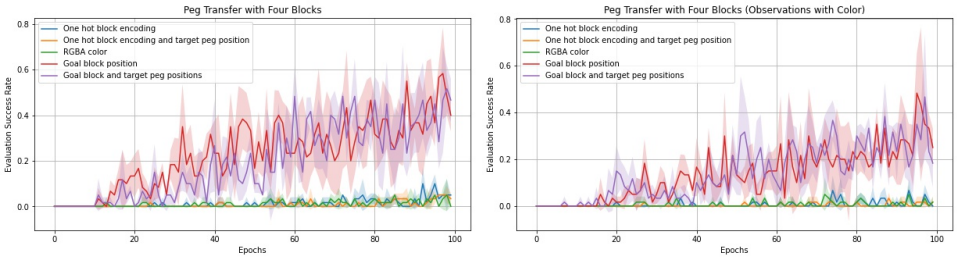

Figure 7: Evaluation success rates of DDPG with HER vs. training epochs on the goal-conditioned peg transfer task with four blocks. Observations include robot state and all block positions (left), and also include all block colors (right).

target peg (link to video). The robot sometimes attempts to recover by trying to re-grasp a dropped goal block, although episodes usually terminate before successful completion of the task (link to video).

Conditioning on RGBA color shows similar manipulation difficulties as well as somewhat more hesitant behavior. The robot is also able to successfully identify and grasp goal blocks but often hovers above target pegs without dropping the block before the episode terminates (link to video). Both methods occasionally struggle with the initial step of grasping the goal block (link to video). Both methods demonstrate that goal identification is largely successful in the simplified two-block setting, with failures primarily due to imprecise manipulation control rather than incorrect block selection.

Comparing across observation variants shows how visual color information of all blocks on the board can influence robot confidence. Without colors of all blocks in the observations, the robot's behavior becomes more tentative. When conditioning on RGBA color, the robot successfully identifies and grasps goal blocks but shows more instances of hesitation to place the block on the target peg (link to video, often dropping blocks prematurely before reaching the peg (link to video). This suggests that while the robot can learn where the goal block is without direct color information, the policy appears to be less confident in its execution timing.

These qualitative results from the two-block environment demonstrate that goal identification is largely successful for semantic methods in simpler environments as failures are primarily due to imprecise manipulation and control rather than incorrect goal block selection.

5.4 DDPG with HER Goal Conditioning Experiments: Four Blocks

The results from the four-block environment reinforces the patterns observed in the two-block setting, with spatial conditioning methods maintaining outperformance over semantic ones. As shown in Figure 7, conditioning on goal block position, and on goal block plus target peg positions achieve success rates of around 40-50% by epoch 100 when trained on demonstrations without color information, and 20-30% when trained on demonstration that include color. These spatial methods demonstrate consistent learning progress throughout training, demonstrating that the agent can effectively leverage direct spatial coordinates even in more complex environments.

In contrast, conditioning on semantic methods—including block encoding, block encoding plus target peg position, and color encoding—remain near zero success rates throughout training. This complete failure to perform the task suggests that one-hot block encodings and color features provide virtually no useful signal in the more complex four-block environment. Even augmenting the block encoding with the target peg position fails to provide sufficient signal for successful task performance, likely for the same reason that combining the two methods underperformed in the two-block environment.

The comparison between observation variants, i.e. between Figure 7 left and right, shows that interestingly, the spatial conditioning methods perform slightly worse when demonstrations include color information, while semantic conditioning methods remain completely ineffective. This finding contradicts our intuition that color information should benefit semantic methods by providing clearer indication for identifying the goal block. Instead, the results suggest that color information may introduce noise to the observation space that interferes with conditioning on spatial information,

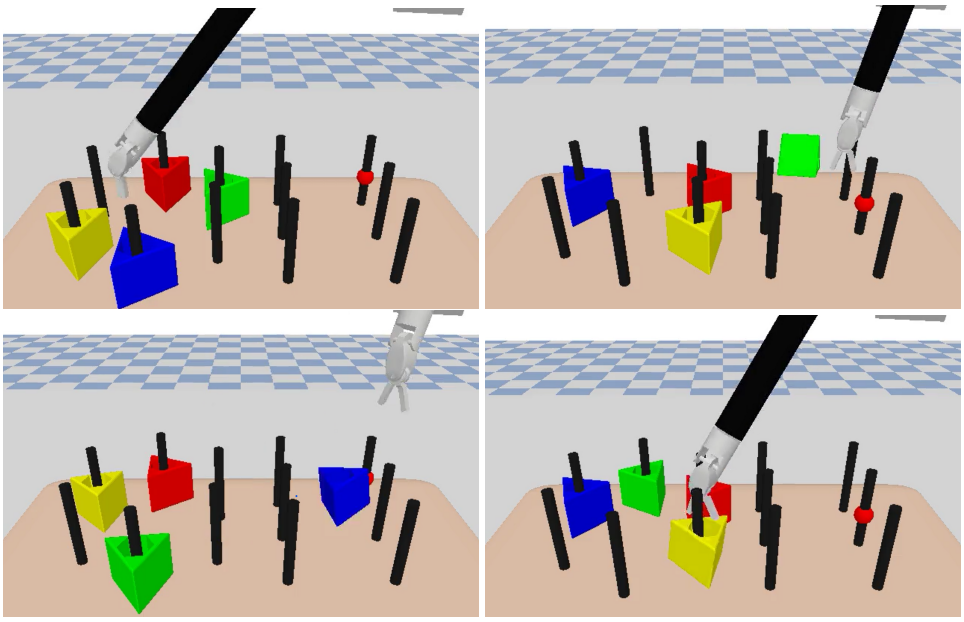

Figure 8: Failure modes of evaluation episodes on the goal-conditioned peg transfer task with four blocks, which include both goal block identification and manipulation failures. (top left) Robot fails to identify the goal block and opens/closes the gripper in mid-air. (top right) Robot fails to place goal block on target peg, by releasing it on the board instead of over or near the target peg. (bottom left) Robot fails to place goal block on target peg, releasing it prematurely before the gripper has reached the peg. (bottom right) Robot struggles to grasp the goal block.

while failing to provide any meaningful advantage to semantic approaches in this more complex environment.

Analysis of evaluation episodes shows more noticeable differences between conditioning approaches on the four-block environment. Figure 8 shows depictions of failure modes described below. Even with increased complexity of the task, spatial conditioning, specifically conditioning on the target block, is able to maintain goal identification. Successful examples show its effectiveness (link to video), while failures primarily occur during manipulation execution, where the agent correctly identifies and attempts to grasp goal blocks but fails during pick-up or placement on the target peg (link to video).

Semantic conditioning methods show inability to perform the task whatsoever in the four-block setting, with goal identification failures becoming prominent—along with challenges with manipulation. When observations include colors of all blocks, conditioning on RGBA color shows various failure modes including goal identification failures, in which the gripper opens and closes in mid-air without approaching any block (link to video); unsuccessful grasping attempts (link to video); and manipulation failures during attempted placement on the target peg (link to video). Some episodes show the agent attempting to pick up multiple blocks in sequence (link to video), suggesting confusion about goal specification.

Conditioning on one-hot block encoding shows similar issues in its failed episodes, such as frequent goal identification failures, where the robot fails to approach any block and instead opens and closes the gripper in mid-air. When the robot is able to successfully identify the goal, there are still manipulation failures, indicating that increased environmental complexity affects both high-level goal understanding and low-level execution capabilities for semantic methods.

This qualitative analysis supports the quantitative findings, highlighting that semantic conditioning methods fail to scale effectively to more complex settings.

6 Discussion

Based on these findings, our work has important implications for how we design goal-conditioning mechanisms for RL in ambiguous robotic tasks. Our experiments revealed several key insights:

1. **Improved performance with active learning:** Goal-conditioned BC suffers from overfitting with the best method in training performing worst in evaluation. This demonstrates that goal-conditioned imitation learning struggles to generalize beyond demonstration patterns, a critical limitation for surgical robotics where adaptability is essential. The substantial performance gap between BC (10% maximum evaluation) and DDPG with HER (60% success) represents a 6× improvement, demonstrating that success in completing the peg transfer tasks can be aided with active learning rather than pure demonstration imitation.
2. **Spatial versus semantic representation effectiveness for goal-conditioning:** We find significant differences in how conditioning on spatial versus semantic representations impact DDPG with HER performance. Spatial coordinates provide actionable information, whereas semantic methods create a representation gap that agents struggle to bridge, particularly in complex environments. While the robot seemed to be able to map semantic representations to physical actions in the simple two-block environment, this capability breaks down entirely in the four-block environment where semantic methods fail to provide any useful learning signal.
3. **Task complexity as a factor on performance of goal-conditioned methods:** Our analysis demonstrates that task complexity fundamentally affects which conditioning methods are viable. While semantic methods achieve moderate success rates in two-block environments, they completely fail in four-block scenarios, suggesting a task complexity threshold beyond which semantic representations become unusable. Qualitative analysis reveals that failure modes transition from manipulation execution issues in the two-block environment to a breakdown in goal identification for semantic methods as complexity increases.
4. **Effect on learning of including color information in environment observations:** When color information is included in observations, semantic methods show improved performance in the simple two-block environment, as the policy can establish clearer correspondences between goal specifications and visual features. However, this advantage disappears in the more complex four-block environment, and color information degrades spatial conditioning performance, suggesting it may introduce noise.
5. **Robot behavior degradation with increased task complexity:** Qualitative video analysis reveals how the robot's behaviors change with increasing task complexity. In the two-block environment, semantic methods show hesitant but often successful goal identification followed by manipulation struggles. In four-block environments, these methods completely fail, producing erratic behaviors like mid-air gripper activation and random block grasping that suggest total goal confusion.

Our work has several limitations:

1. We focused exclusively on a relatively simple task with static objects, and it remains unclear whether our findings generalize to more complex manipulation scenarios involving dynamic environments where object positions change during an episode. The advantage of using spatial conditioning methods that we observed might diminish in such settings.
2. We only tested two RL algorithms—Behavioral Cloning and DDPG with HER. Different learning algorithms might show different patterns in how they utilize various conditioning representations.
3. Our semantic representations were relatively simple. More sophisticated semantic representations might better bridge the gap between abstract specifications and desired physical actions.

7 Conclusion

As advancements are made in surgical robotics towards fully autonomous operations, effective methods for goal specification become critical for handling ambiguous scenarios with multiple similar

objects. Our comprehensive analysis demonstrates that both algorithm choice and representation design significantly impact goal-conditioned surgical robotics performance. Our behavior cloning analysis reveals that training metrics can be misleading, with spatial representations achieving superior generalization despite lower training performance. For reinforcement learning approaches, this work demonstrates that goal-conditioned RL using spatial goal conditioning outperforms semantic approaches for peg transfer robotic manipulation tasks. While spatial coordinates provide actionable information for the robot to accomplish its task, semantic representations create a representation gap that makes it difficult to learn on as task complexity increases, with failures transitioning from manipulation issues to a breakdown in goal identification. These findings highlight the importance of representation choice, proper evaluation protocols, and algorithm selection in goal-conditioned RL, providing essential insights for designing effective goal specification mechanisms in autonomous surgical robotic systems.

Future research directions include exploring methods to improve learning with more complex goal-conditioned encodings, such as human-language goal encodings (e.g., "red") where the mapping of the encoding to the RGBA observation value must be learned before the relevant spatial observation can be used. One promising approach draws inspiration from BC-Z and VLA architectures that separate goal-conditioned encoding from learning the policy, using systems like language encoders to transform the goal conditioning into spaces easier for the policy to learn. Additionally, investigating how our findings extend to dynamic environments and more complex manipulation tasks would provide valuable insights into the broader applicability of spatial versus semantic conditioning approaches. Future work should also explore whether different learning algorithms exhibit different downstream performance and behaviors depending on goal representation choices, as our findings may be specific to DDPG with HER.

8 Team Contributions

All team members contributed to experiment implementation and training, as well as all deliverables.

- **Daphne Barretto:** Initial simulation environment and data generation setup. Encoding implementation. DDPG with HER experimentation and analysis.
- **Alycia Lee:** Initial AWS setup for goal-conditioning and encoding implementation. DDPG with HER experimentation and analysis. Video policy analysis setup.
- **Elsa Bismuth:** Behavioral cloning implementation, experimentation, and analysis. Goal-conditioned behavioral cloning experimentation and analysis.

Changes from Proposal We changed our original project proposal (to fine-tune VLA models for surgical robotic manipulation) to focus on topics and algorithms more directly implementing what we learned from the course. In our current project, we worked with behavioral cloning, deep deterministic policy gradient and hindsight experience replay, and goal-conditioning methods. Due to these changes, some previous work from this project was cut for the final paper, including data generation and training for image-based observations and behavioral cloning ablation studies.

References

Zih-Yun Chiu, Florian Richter, Emily K Funk, Ryan K Orosco, and Michael C Yip. 2021. Bimanual regrasping for suture needles using reinforcement learning for rapid motion planning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 7737–7743.

Jiawei Fu, Yonghao Long, Kai Chen, Wang Wei, and Qi Dou. 2024. Multi-objective Cross-task Learning via Goal-conditioned GPT-based Decision Transformers for Surgical Robot Task Automation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 13362–13368.

Tao Huang, Kai Chen, Bin Li, Yun-Hui Liu, and Qi Dou. 2023a. Guided reinforcement learning with efficient exploration for task automation of surgical robot. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 4640–4647.

Tao Huang, Kai Chen, Wang Wei, Jianan Li, Yonghao Long, and Qi Dou. 2023b. Value-informed skill chaining for policy learning of long-horizon tasks with surgical robot. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 8495–8501.

Physical Intelligence, Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, et al. 2025. $pi\_\{0.5\}$: a Vision-Language-Action Model with Open-World Generalization. *arXiv preprint arXiv:2504.16054* (2025).

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

Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, Quan Vuong, Thomas Kollar, Benjamin Burchfiel, Russ Tedrake, Dorsa Sadigh, Sergey Levine, Percy Liang, and Chelsea Finn. 2024. OpenVLA: An Open-Source Vision-Language-Action Model. arXiv:2406.09246 [cs.RO] `https://arxiv.org/abs/2406.09246`

Audrey Lee, Turner S Baker, Joshua B Bederson, and Benjamin I Rapoport. 2024. Levels of autonomy in FDA-cleared surgical robots: a systematic review. *NPJ Digital Medicine* 7, 1 (2024), 103.

Srujan Gowdru Lingaraju. 2023. *Learning from Pixels: Image-Centric State Representation Reinforcement Learning for Goal Conditioned Surgical Task Automation*. Master's thesis. University of Minnesota.

Minghuan Liu, Menghui Zhu, and Weinan Zhang. 2022. Goal-conditioned reinforcement learning: Problems and solutions. *arXiv preprint arXiv:2201.08299* (2022).

Yanzhen Liu, Xinbao Wu, Yudi Sang, Chunpeng Zhao, Yu Wang, Bojing Shi, and Yubo Fan. 2024. Evolution of surgical robot systems enhanced by artificial intelligence: A review. *Advanced Intelligent Systems* 6, 5 (2024), 2300268.

Yonghao Long, Wang Wei, Tao Huang, Yuehao Wang, and Qi Dou. 2023. Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning. *IEEE Robotics and Automation Letters* 8, 8 (2023), 4441–4448.

Ashvin Nair, Bob McGrew, Marcin Andrychowicz, Wojciech Zaremba, and Pieter Abbeel. 2018. Overcoming Exploration in Reinforcement Learning with Demonstrations. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 6292–6299. `https://doi.org/10.1109/ICRA.2018.8463162`

David Power. 2024. Ethical considerations in the era of AI, automation, and surgical robots: there are plenty of lessons from the past. *Discover Artificial Intelligence* 4, 1 (2024), 65.

Florian Richter, Ryan K Orosco, and Michael C Yip. 2019. Open-sourced reinforcement learning environments for surgical robotics. *arXiv preprint arXiv:1903.02090* (2019).

Jiaqi Xu, Bin Li, Bo Lu, Yun-Hui Liu, Qi Dou, and Pheng-Ann Heng. 2021. SurRoL: An Open-source Reinforcement Learning Centered and dVRK Compatible Platform for Surgical Robot Learning. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE.

A Behavior Cloning Results and Discussion Details

A.1 Training vs Evaluation Performance

Table 1: Behavior Cloning: Training vs Evaluation Performance

Method	Training	Evaluation	Performance Gap	Key Finding
Spatial (No Color)	18.3%	**10.0%**	+8.3%	Best generalization
Spatial (Color)	15.0%	4.3%	+10.7%	Color hurts spatial
Semantic (No Color)	18.3%	7.7%	+10.7%	Moderate performance
Semantic (Color)	**25.0%**	2.7%	+22.3%	Severe overfitting

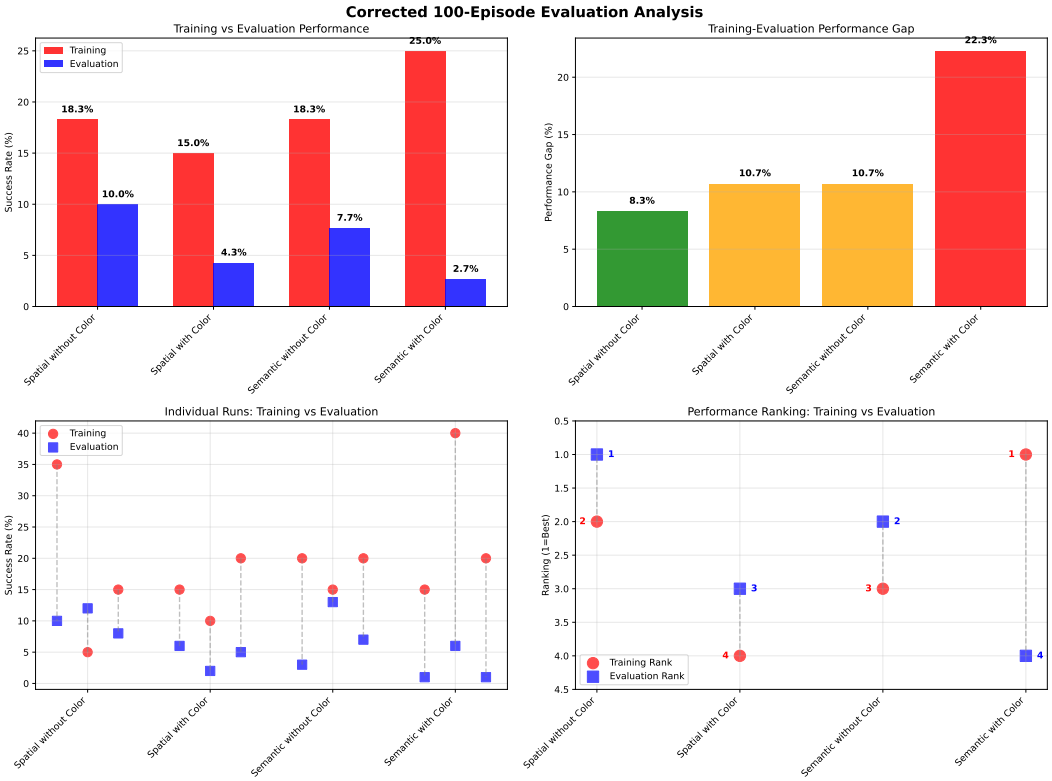

Figure 9: Behavior cloning analysis showing training vs evaluation performance. The substantial generalization gaps demonstrate fundamental challenges in goal-conditioned imitation learning.

A.2 Key Insights

Overfitting in Goal-Conditioned Learning Semantic approaches with color information showed severe overfitting: highest training performance (25.0%) but worst evaluation performance (2.7%), indicating that complex feature representations can lead to overfitting to demonstration patterns.

Spatial Conditioning Advantage Spatial conditioning without color demonstrated the best generalization with the smallest training-evaluation gap (8.3%), suggesting that direct spatial coordinates provide more robust learning signals.

Methodological Implications The inconsistency between training and evaluation metrics highlights critical methodological considerations: (1) training performance is not predictive of deployment success in goal-conditioned tasks, (2) feature complexity (color information) can impair rather

than improve generalization, and (3) rigorous evaluation protocols are essential for assessing goal-conditioned policies. These findings have immediate implications for developing reliable surgical robotic systems.

A.3 Future Directions

While these BC results demonstrate clear limitations, they suggest research directions including hybrid BC-RL approaches, advanced goal encodings, and improved evaluation protocols for surgical robotic manipulation applications.