

Learning Interactive Viewpoint Planning for Foresighted Object Discovery Considering Human Preferences

Nils Dengler Sicong Pan Vamsi Kalagaturu Rohit Menon Murad Dawood Maren Bennewitz

Abstract—Viewpoint planning is an important task in any application where scenes need to be viewed from different angles to achieve sufficient coverage and discover objects. The mapping of confined spaces such as shelves is especially challenging since objects occlude each other and the scene can only be observed from the front, posing limitations on the possible viewpoints. Therefore, manipulation actions such as moving objects aside to unveil space behind them will often be necessary to discover hidden objects. In this paper, we propose a deep reinforcement learning framework that generates promising perception and manipulation actions aiming at reducing the entropy of the scene representation. In particular, our pipeline extends standard viewpoint planning by predicting adequate minimally invasive push actions to uncover occluded objects and increase the visible space. These actions are learned from human input and consider their preferences in scene changes. As simulated and real-world experimental results with a UR5 arm equipped with a Robotiq 2f85 gripper and an RGB-D camera show, our system is able to significantly increase the covered space compared to different baselines, while the executed push actions highly benefit the viewpoint planner with only minor changes to the object configuration in the scene.

I. INTRODUCTION

Viewpoint planning (VPP) is an important part of various robotic applications to gain an understanding of the environment. It enables a robot to identify occluded objects on a crowded table [1] or to estimate the amount of harvestable fruits in horticulture [2]. For manipulation purposes, VPP can also output the best grasping position by viewing and mapping the object from different angles [3]. However, the performance of VPP is highly dependent on the given scenario. While the possibility of good viewpoints, and thus scene coverage, increases when the robot is able to act freely in the environment, the environment itself can constrain actions and reduce the possible coverage, e.g., in a confined indoor space such as a shelf. If the robot’s task is to map the current state of a shelf, the difficulty increases with the number of objects present and their closeness.

In particular, the limited top-down view from above the scene makes it difficult to identify and map objects that are hidden behind other objects or closer to the back of



Fig. 1: Motivation of our approach. The robot’s task is to map the shelf board as completely as possible in a 2.5D occupancy height map. Without manipulative actions it is not possible to plan viewpoints necessary to uncover spaces behind big objects. However, with just a short push of the right box that was learned from human preferences, the information value of the same viewpoint increases because of the newly uncovered object (white circle), while maintaining the overall layout of the scene.

the shelf. An obvious solution to this problem is to introduce manipulation actions to uncover previously occluded space. However, human preferences play an important role in the arrangement of indoor environments [4]. Therefore, not all manipulation actions are equally suitable, as they could lead to undesired large changes in the layout of the scene. To overcome this problem, minimally invasive pushing actions should instead be used to uncover previously occluded areas. Furthermore, in cluttered scenes it is easier to identify promising non-prehensile push positions than, for example, grasping poses on objects to rearrange them, since the constrained environment limits the number of possible actions.

In this paper, we therefore present a novel framework called viewpoint push planning. We apply deep reinforcement learning (DRL) to perform VPP in confined scenarios, while predicting suitable minimally invasive, non-prehensile push actions to improve shelf coverage. In particular, for the immense number of possible object configurations within the shelf, DRL can learn the best viewpoints from experience and generalize them to different configurations. Furthermore, we use a push prediction network trained in a supervised manner that outputs the best pushing action to free views and aid the mapping process.

As long as the VPP finds a new promising viewpoint that increases the map coverage, it is not necessary to perform a push action, as it would unnecessarily alter the environment. Therefore, we use an action selection method to decide

All authors are with the Humanoid Robots Lab, University of Bonn, Germany. Murad Dawood and Maren Bennewitz are additionally with the Lamarr Institute for Machine Learning and Artificial Intelligence, Germany. This work has partially been funded by the European Commission under grant agreement number 964854 –RePAIR – H2020-FETOPEN-2018-2020 and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy, EXC-2070 – 390732324 – Phenorob and under the grant number BE 4420/4-1 (FOR 5351: KI-FOR Automation and Artificial Intelligence for Monitoring and Decision Making in Horticultural Crops, AID4Crops).

whether a VPP or a pushing action is best given the current situation. Fig. 1 shows an example scenario where a packed shelf board needs to be mapped. As can be seen, a short minimally invasive push is sufficient to move one of the larger boxes, exposing the space behind it and uncovering the previously occluded area.

Our experimental evaluation demonstrates that the learned VPP is able to significantly reduce the entropy of the map compared to two baselines and two VPP systems from the literature [5], [6], while the complete framework, including the execution of push predictions, is able to reduce the entropy even further. The main contributions of our work are as follows:

- A 2.5D heightmap representation for mapping of confined spaces.
- A viewpoint planning framework including human guided pushing actions based on deep reinforcement learning.
- A qualitative and quantitative evaluation in simulation as well as on a real robot including the usability of our map representation for object reasoning and retrieval tasks.

II. RELATED WORK

Viewpoint planning has been used in applications ranging from the reconstruction of single objects to large-scale applications such as the mapping of crops in green houses. While in the past, sample-based approaches were mostly used to perform viewpoint planning [2], [7]–[9], recently learned planners have become more popular [10]–[12]. However, reinforcement learning, as in this paper, has not often been used for VPP, although it can highly benefit VPP approaches by reducing the planning time and learn a generalizable planning behavior for multiple objects and scenes. Peralta *et al.* [14] applied reinforcement learning to generate best views for the reconstruction of 3D houses and reduced the number of needed planning steps, while increasing the mapping accuracy wrt. a uniformly sampled approach. Zeng *et al.* [15] used reinforcement learning to map sweet pepper crops with a robotic arm and an in-hand camera in a greenhouse environment. The authors applied an octomap and ray casting to determine the next observation for the agent. However, this increases the time needed to plan an action compared to our approach, which relies on a 2.5D representation.

Previous work in the area of interaction with objects in confined spaces has concentrated on reasoning about individual objects [16]–[18] or their retrieval [19]–[23]. In contrast to our work, most of these approaches use fixed cameras, no long-term updated map representation, or assume that a lot of object information is known a priori. Furthermore, none of these approaches consider the invasiveness of the actions and perform each action regardless of the resulting scene change. For example, Bejjani *et al.* [19] used a single-shot representation and reinforcement learning for object retrieval with initially unknown object positions. Due to the incomplete map representation, the agent might investigate irrelevant areas first, while possibly changing the layout

of objects completely. Using a complete map for action planning would counteract this behavior. Huang *et al.* [16] rely on a fixed camera position for object reasoning with Gaussian distributions for likely positions of an occluded object. By using different camera positions and VPP, as in this paper, some objects would become visible even though they are occluded from the front, which supports object reasoning.

There has not been much research regarding the long-term mapping of confined spaces. Miao *et al.* [24] recently proposed a system to reconstruct objects placed in shelves. The authors create an occlusion dependency graph to reason about the relations between the objects and to calculate the best retrieval and rearrangement strategy. However, the authors assume that initial object models are given, which is a strong assumption. Furthermore, the complete 3D reconstruction of objects is time-consuming and not the focus of this work. While the reconstructed object knowledge can also be used to reduce planning time for object reasoning and retrieval, our approach is faster, since not all objects need to be processed individually.

III. OUR APPROACH

In the following, we first give an overview of our system and introduce our framework in detail afterwards.

A. Overview

We consider the following problem. In the confined environment of a shelf, a stationary robotic arm needs to map a shelf board as completely as possible. The shelf board is packed with a number of different objects, which may occlude each other. A schematic overview of our approach is shown in Fig. 2. Our system receives a depth map of the environment from an RGB-D camera mounted on the end-effector of the arm and orthographically back-projects it from the corresponding point cloud into a 2.5D occupancy height map. In order to decouple the feature extraction from the learning process and to promote faster convergence of the reinforcement learning agent for viewpoint planning, we use a variational auto-encoder [25] to encode the height map into a latent space of size 32. Since object occlusion can lead to areas that cannot be fully mapped by viewpoint planning alone, we sample push proposals on the height map to predict the best push candidate that will uncover additional space in a minimally invasive manner. Depending on the change in mapped space, an action selection method decides whether the best action in this situation is a push, moving to a new viewpoint, or if the mapping process has finished.

B. Environment Representation

We use a 2.5D representation of the environment since it can be efficiently updated. From a depth image, we compute the corresponding point cloud and orthographically back-project it into the top-down view of the scene to calculate a 2.5D occupancy height map as shown in Fig. 3. To update the map, we implemented log-odds for occupancy updates [26] and update only the cells in the current field of view. For

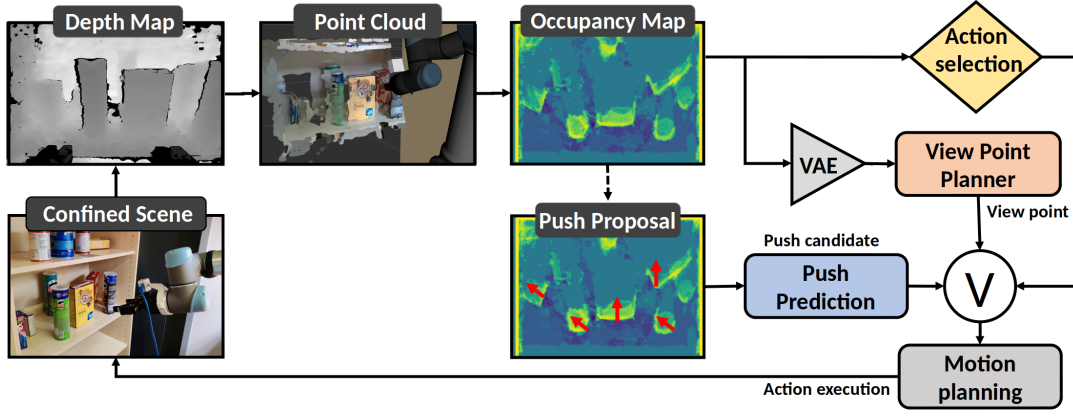


Fig. 2: Overview of our viewpoint pushing framework. Our system receives a depth map from an RGB-D camera on the end-effector of the robotic arm and converts it into the corresponding point cloud. We then orthographically backproject the point cloud into a 2.5D occupancy height map and feed it into a variational auto-encoder (VAE) to obtain the latent space. An action selection method decides whether to map or to perform a predicted push or the next viewpoint action. The latent space is used for the viewpoint planner to select the next best viewpoint, while push proposals sampled from the height map are used to predict the best push candidate.

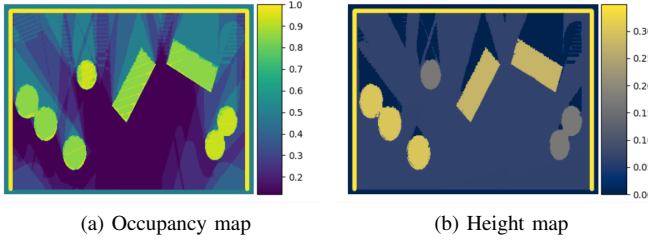


Fig. 3: Example visualization of the 2.5D state representation after mapping. The left image shows occupancy probability values of the map, while the right image shows the maximum height value of each cell. The brighter the color of the map, the higher its value.

the height component of the map, we use the maximum measured height value for each cell.

Note that in case grasping requires a more accurate representation of the objects, our map can be used to plan a path to the object, while the actual grasping action can be performed on a local higher-dimensional representation. We provide an application example of our map for object reasoning and grasping in the experimental section. As previous approaches have shown, 2.5D representations can be used to compute grasping points and manipulate objects [27]–[29].

C. Viewpoint Planning (VPP)

For VPP, we use reinforcement learning (RL) to generate the next best view. The overall goal of our VPP agent is to map the environment as completely as possible. As in previous approaches [30], we generate three fixed viewpoints at the beginning of each episode, which the agent uses to generate an initial map. This reduces the need for the agent to learn good initial viewpoints that are similar for each state configuration. For the three fixed viewpoints, we use a camera pose facing towards the center of the shelf and two poses at the left and right corners of the shelf, pointing diagonally inside the shelf to the opposite corner. By using these three viewpoints, we usually achieve an initial map coverage of around 50%.

To evaluate the quality of the agent during testing and for training, we use the entropy $h(M)$, information gain $IG(M)$, and motion cost $c(p_t)$ as defined below:

$$h(M) = \frac{m_u^t}{cells_M} \quad (1)$$

$$IG(M) = \frac{m_u^{t-1} - m_u^t}{m_u^{t-1}} \quad (2)$$

$$c(p_t) = d(p_{t-1}, p_t) \quad (3)$$

$h(M)$ states the ratio of the unknown cells m_u^t to the total number of cells of map M at time t . The lower the entropy, the higher the information value of the map. Note that we consider each cell with an occupancy probability value $occ > 0.5 + \tau_{unknown}$ as occupied, $occ < 0.5 - \tau_{unknown}$ as free, and unknown otherwise, with $\tau_{unknown}$ as uncertainty threshold. $IG(M)$ indicates the percentage decrease of unknown cells m_u from the previous time step $t-1$ to time step t . Finally, we use the Euclidean distance $c(p_t)$, between two viewpoints p_{t-1} and p_t to approximate the move cost.

During training, an episode ends if the change of $h(M)$ for the last three viewpoints is below a certain threshold τ_{single} and the sum of changes is below the threshold τ_{sum} . Furthermore, we terminate the episode if any part of the arm has contact with an object or an object drops in or out of the shelf.

In the following, we define the actions, observations, and the reward function of our agent.

1) **Action Space:** As action we use the 5D Cartesian pose $(x, y, z, pitch, yaw)$ of the camera on the end effector. We ignore the roll, as it does not have much impact on the information gain of a viewpoint. To speed up training and reduce the likelihood of damage to the arm, we restrict the (x, y, z) space of the actions to a reasonable area in front of the shelf with a small overlap inside the shelf. For pitch and yaw, we experimentally evaluated the angle of rotation so that the arm cannot damage the camera during its movements.



Fig. 4: Two height map object configurations with example sampled push candidates (red arrow). The brighter the color, the higher the object. Unknown space is marked dark blue. The labels indicate how the human experts evaluated the quality of the push candidate in terms of minimally invasiveness and additional uncovered space.

2) **Observation Space:** The primary component of the observation space is the latent space of the state representation, which is of size 32. Additionally, we include the last action in 5D Cartesian coordinates, the information gain the agent receives from this viewpoint, and the motion cost to move there. Furthermore, we incorporate a Boolean collision indication. Finally, we calculate the center point of the largest unknown area from our map and provide it to the agent as 3D Cartesian coordinates, offering the agent insight into a promising region of interest.

3) **Reward:** We keep the reward function as simple as possible to aid the training. The reward consists of two parts:

$$r_{coll} = \begin{cases} -\eta, & \text{if collision or drop from shelf} \\ 0, & \text{else} \end{cases} \quad (4)$$

$$r_{gain} = \alpha * -c(p_t) + \beta * IG(M) \quad (5)$$

r_{coll} penalizes each contact with the object according to a predefined high negative value η . Since the goal is to map the environment in as few time steps as possible, we penalize each step in r_{gain} according to the move cost $c(p_t)$. The higher $IG(M)$, the more space has been uncovered during one step, which is positively rewarded. α and β are used to weight closer viewpoints with the effectiveness of the viewpoint. Finally, the total reward can be expressed as $r = r_{coll} + r_{gain}$.

D. Human Guided Push Prediction

Standard viewpoint planning aims to minimize entropy without altering the environment. In contrast, interactive perception [31], as proposed in this paper, incorporates additional actions to modify the current object configuration to enhance perception and coverage of the environment. However, any manipulation action involves a trade-off between reducing the entropy of the environment and disrupting the current scene configuration, which may not be desirable for human users. Therefore, we emphasize minimally invasive actions to strike a balance between these competing requirements. Specifically, we utilize non-prehensile push actions since they are straightforward to compute and execute in confined spaces, while other types of manipulation actions would increase computational efforts due to environmental constraints.

1) **Learning of the Push Prediction Network:** To predict the push that best supports the VPP, we employ a similar approach as Eitel et al. [32] to train a prediction network that outputs the optimal push for the current scene. Our goal is to predict the minimally invasive push action that increases the area that can be mapped by the VPP agent while balancing the mentioned trade-off. To achieve this, three human experts manually labeled 5,000 simulated map configurations and an additional 850 configurations of the real-world scene according to their personal preferences for altering the scene, considering the trade-offs involved.

The experts were tasked with labeling images annotated with the starting point of the push and its direction, assigning binary labels "good push" (= 1) and "bad push" (= 0) based on their intrinsic needs, preferring pushes that minimally change the object configuration while achieving a significant entropy reduction. The maximum considered push length was 10 cm.

Example images, including the labels, are illustrated in Fig. 4. In the left image, the push candidate was labeled "positive" since the push would only alter the position of one object while uncovering the previously hidden space marked with the white circle. In contrast, the second example was labeled "negative" because the push involves the displacement of up to four objects, potentially causing damage as they are pushed against the edges of the shelf.

It's worth noting that we opted for manual labels rather than automatically generated ones because, without computationally expensive physics simulations, it's challenging to accurately predict the effect of pushing an object on itself as well as its neighbors. Therefore, we rely on human expertise.

2) **Generation of Push Proposals:** To generate push proposals, we sample a number of push candidates for each map configuration by ray casting from three fixed positions outside the shelf to the back of the shelf. The 3D Cartesian coordinate of the first occupied cell along the ray is then considered a push candidate, along with eight push angles and a predefined push length. In our experiments we sampled 100 to 150 push proposals on average. As input to the network, we generate push maps by translating the starting point of the push to the center of the map and rotating the image according to the push direction.

Our prediction network consists of five convolution layers, each followed by a max-pooling layer, and the rectified linear unit activation function. With the labeled push maps, we followed the training structure of [32], with the Adam optimizer and a learning rate of $1e-4$. For inference, we feed the push proposal maps into the network and execute the push with the highest output of the network.

E. Action Selection

Since a push is not required after each executed viewpoint, and pushing can lead to several new viewpoints that increase the observable space, viewpoint planning is always performed until the VPP RL agent terminates. To decide afterwards whether another push is needed or the mapping process is finished, we proceed as follows. We assume that

Simulation	Entropy Reduction
Ours	30.0% \pm 0.9%
Random	15.5% \pm 0.6%
RSE [6]	16.5% \pm 0.5%
GMC [5]	21.8% \pm 0.8%
Real World	Entropy Reduction
Ours	26.1% \pm 1.0%
Random	20.3% \pm 1.0%

TABLE I: Evaluation of the viewpoint planner: Reduced entropy compared to the initial entropy of the map computed with 3P (three fixed viewpoints). Each metric is shown with its standard error. The experiments were carried out in 100 simulated and 15 real-world scenarios. As can be seen, our approach lead to the highest reduction in entropy in comparison to all baselines. The reduction is significant according to the paired t-test with a p-value of 0.05.

the VPP agent is trained to an optimal policy and only terminates according to the criteria defined in III-C. Furthermore, we assume that the push prediction network always outputs the best possible push action. Since we assume the push prediction to be optimal, if the previous push action did not lead to an entropy reduction of $\Delta h(M) > \tau_{push}$, we do not perform any further push action and terminate the mapping process. In this way, we deal with the trade-off between entropy reduction and change in the current configuration of the scene and also avoid infinite pushing loops.

IV. EXPERIMENTAL EVALUATION

We evaluated the entropy reduction, the number of steps needed, the planning time, and the object displacement of our complete pipeline and compared the performance to different baselines in both simulation and real-world experiments. Furthermore, to demonstrate that our 2.5D map is a sufficient state representation of the environment, we showcased its application to object reasoning and grasping.

All experiments were conducted on a UR5 robot equipped with a Robotiq 2f85 gripper and a Realsense D405 for close-range depth images mounted on top of the end effector. We trained and evaluated our approach on a computer with an i7-6800k CPU and an RTX2070 GPU. For the reinforcement learning agent, we utilized the stable-baselines3 framework [35] with the truncated quantile critics algorithm (TQC) [36]. Furthermore, we employed the same actor-critic architecture with a small three-layered network of size 256 each. For our experiments, we utilized a shelf with dimensions of 40 cm \times 80 cm \times 40 cm.

Throughout the experiments, we used the following thresholds: $\tau_{unknown} = 0.2$, $\tau_{single} = 0.01$, $\tau_{sum} = 0.05$, and $\tau_{push} = 0.01$. Additionally, for the reward function of the VPP agent, we set $\alpha = 1$ and $\beta = 10$.

A. Evaluation of the Viewpoint Planner

To evaluate our viewpoint planner individually, we executed the 3-point fixed planner (3P) used for training and evaluated how much our learned planner improves the resulting map. Furthermore, we compared it to a random

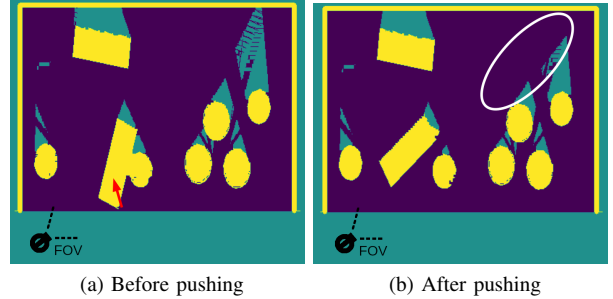


Fig. 5: Qualitative example of an occupancy map update in simulation after a minimally invasive push. (a) shows the map before the push (green indicates unknown space), with the arrow indicating the planned push action. (b) shows the map after the push, where the cells in the marked area became visible due to the push.

method and reimplemented two sample-based state-of-the-art next-best view planners: a greedy planner by Delmerico et al. [6] (RSE) and a global planner by Pan et al. [5] (GMC). We compared against RSE to show the limitations of greedy planners in confined spaces and against GMC to demonstrate the overall benefit of a learned approach over a sample-based method. We adapted the methods to work in our confined-space multi-object scenario by sampling only reachable poses a priori and removing the region of interest constraint. For comparability, the two state-of-the-art systems and the random sampling method used the same workspace as our system to generate the viewpoints, the map generated by 3P as the initial representation, and the same number of view poses as our agent in the environment.

For the experiments, we sampled 100 different shelf configurations in simulation and 15 in the real world, each with 8 to 10 objects. As shown in Tab. I, our agent outperformed all baselines in simulation, significantly according to the paired t-test ($p=0.05$), with an average 30% reduced entropy compared to 3P, while requiring 4.94 viewpoints. Since RSE is a greedy sampling-based method, it sometimes performed worse or equal to the random agent and therefore shows no significant difference since it is designed for single object reconstruction and adheres to the local optimum. In contrast, the performance of GMC is better due to its global sampling; however, our approach achieves a 37% higher entropy reduction relative to GMC. Furthermore, GMC took 7.29s per planning step, whereas our approach planned the next viewpoint in 0.04s. As evident from the lower part of Tab. I, the simulation results are corroborated by the real-world experiments, wherein our system performed nearly as effectively as in simulation.

B. Evaluation of the Viewpoint Push Planner

To evaluate our complete pipeline and to show the influence of the human guided push actions to reduce the entropy after VPP, we calculated the entropy before and after the push actions. Furthermore, to show the benefits of our trained push prediction network, we applied a method that randomly selects a push candidate generated by our push sampling method (see Sec. III-D.2). Each combination of

Metrics	Learned Push	Random Push
entropy reduction	27.62% \pm 2.2	37.05% \pm 2.6
displacement	*0.24 m \pm 0.03	0.34 m \pm 0.03
#iterations	3.5 \pm 0.24	3.7 \pm 0.21
object drop rate	0.04	0.21

TABLE II: Evaluation of the complete viewpoint push planner pipeline: Reduced entropy compared to the map before executing push actions, object displacement, and number of pipeline iterations as well as object drops. Each metric is shown with its standard error. The experiments were carried out in 100 simulated scenarios. As can be seen, our complete pipeline was able to reduce the entropy of the map generated by our learned VPP even further. In comparison to a random push selection method, our trained push proposal network lead to reduced object displacements and a reduced object drop rate. The object displacements and entropy reduction are significant according to the paired t-test with a p-value of 0.05.

VPP followed by a push action is called a pipeline iteration and is terminated when the criteria defined in Sec. III-E is met. We evaluated the pipeline on the same 100 simulated environments as used in IV-A. As Tab. II shows, both pushing methods were clearly able to reduce the map entropy resulting from VPP alone by 27.62% and 37.05%, with an average of 3.5 and 3.7 pipeline iterations. While the random push selection method outperformed our trained method in terms of entropy reduction, our method shows the improvement wrt. the learned minimally invasive actions by reducing the likelihood of object drops and too large object displacements. An object is considered dropped if it is tilted by more than 45° or falls off the shelf. For evaluating the displacement, we compared the object configuration before and after the full pipeline execution, using the Euclidean distance of the center point of each object as the evaluation metrics. Out of the 100 simulated trials, our learned pipeline reduced the number of object drops after the push by 80% and was able to reduce the displacement by 29.4% compared to the random method, reflecting the human preferences that were introduced during training. To test the influence of the push length, we performed each trial with a push length of 10 cm, 7 cm, 5 cm, and 2 cm. The longer the push length, the greater is the reduction in entropy. However, also the displacements increase significantly according to the trade off between entropy reduction and configuration disruption. Empirically, we evaluated 5 cm to be the best in terms of this trade off and used it for the results in Tab. II. Fig. 5 shows a qualitative comparison of a map before and after a push action. As can be seen, the VPP is able to detect more, previously hidden space, while the push itself changed the object configuration only slightly. On average, our method takes 0.31 s to sample the push candidates, while the network takes 0.1 s for inference.

To summarize, we demonstrated that using a predicted non-prehensile push aids the VPP while maintaining the overall structure of the scene. The complete pipeline is easily transferable to the real-robot system without loss of

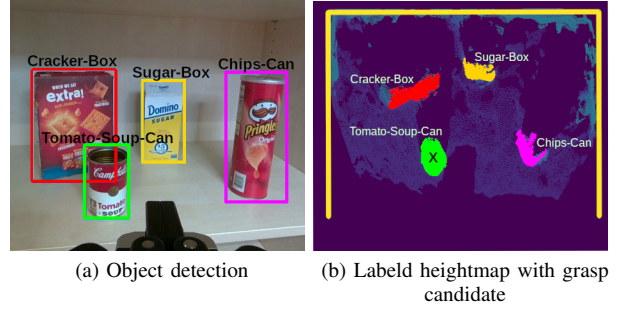


Fig. 6: Visualization of the capabilities of our map representation. The left image shows the current scene with the corresponding bounding boxes resulting from the object detection. On the right, our map representation is shown with the labels of each object added as a new layer and the center point of the tomato soup can for object retrieval.

performance as can be seen in the accompanying video¹.

C. Application of the Map Representation to Grasping

To demonstrate the benefit of our map representation, we conducted two qualitative experiments on the real robot. Firstly, we augmented our pipeline with an additional object label layer. These labels were obtained using an object detection network trained on the YCB dataset [39], and then transformed onto our map representation. As illustrated in Fig. 6, this object label layer enables reasoning about individual objects within the scene. Subsequently, we utilized the map to grasp objects from the shelf. Initially, we applied 2D clustering on the labeled objects to determine the center of each object cluster. Then, by deprojecting the corresponding pixel to its 3D Cartesian world coordinate with a predefined height value of 10 cm above the shelf board height, our system successfully grasped the soup can as well as the chips can in nine out of ten cases. The single unsuccessful grasping trial was attributed to slippage of the object.

V. CONCLUSION

In this paper, we presented a pipeline for interactive viewpoint planning in confined spaces. We demonstrated the efficacy of our approach in multiple simulated and real-world scenarios with a robotic arm equipped with an RGB-D camera on the end effector for mapping of the objects on a shelf. The results show the improved performance of our system in comparison to several baselines and demonstrate that our agent is able to increase the mapped space. We showed that the executed push actions highly benefit the viewpoint planner. Furthermore, the executed pushes maintain the overall object configuration following preferred push actions of humans used to train the push prediction network. Our system runs in real time and the code of our viewpoint push planner is available on Github².

¹<https://www.youtube.com/watch?v=6C3Q2UFKq2Q>

²<https://github.com/NilsDengler/view-point-pushing>

REFERENCES

- [1] M. Li, C. Weber, M. Kerzel, J. H. Lee, Z. Zeng, Z. Liu, and S. Wermter, "Robotic occlusion reasoning for efficient object existence prediction," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.
- [2] T. Zaenker, C. Smitt, C. McCool, and M. Bennewitz, "Viewpoint planning for fruit size and position estimation," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.
- [3] M. Breyer, L. Ott, R. Siegwart, and J. J. Chung, "Closed-loop next-best-view planning for target-driven grasping," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2022.
- [4] N. Abdo, C. Stachniss, L. Spinello, and W. Burgard, "Robot, organize my shelves! tidying up objects by predicting user preferences," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*. IEEE, 2015.
- [5] S. Pan and H. Wei, "A global generalized maximum coverage-based solution to the non-model-based view planning problem for object reconstruction," *Computer Vision and Image Understanding*, 2023.
- [6] J. Delmerico, S. Isler, R. Sabzevari, and D. Scaramuzza, "A comparison of volumetric information gain metrics for active 3d object reconstruction," *Autonomous Robots*, 2018.
- [7] C. Lehnert, D. Tsai, A. Eriksson, and C. McCool, "3d move to see: Multi-perspective visual servoing towards the next best view within unstructured and occluded environments," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2019.
- [8] W. R. Scott, G. Roth, and J.-F. Rivest, "View planning for automated three-dimensional object reconstruction and inspection," *ACM Computing Surveys (CSUR)*, 2003.
- [9] D. Morrison, P. Corke, and J. Leitner, "Multi-view picking: Next-best-view reaching for improved grasping in clutter," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2019.
- [10] B. Hepp, D. Dey, S. N. Sinha, A. Kapoor, N. Joshi, and O. Hilliges, "Learn-to-score: Efficient 3d scene exploration by predicting view utility," in *Proc. of the Europ. Conf. on Computer Vision (ECCV)*, 2018.
- [11] Y. Wang, S. James, E. K. Stathopoulou, C. Beltrán-González, Y. Konishi, and A. Del Bue, "Autonomous 3-d reconstruction, mapping, and exploration of indoor environments with a robotic arm," *IEEE Robotics and Automation Letters (RA-L)*, 2019.
- [12] S. Pan, H. Hu, and H. Wei, "Scvp: Learning one-shot view planning via set covering for unknown object reconstruction," *IEEE Robotics and Automation Letters (RA-L)*, 2022.
- [13] R. Zeng, Y. Wen, W. Zhao, and Y.-J. Liu, "View planning in robot active vision: A survey of systems, algorithms, and applications," *Computational Visual Media*, 2020.
- [14] D. Peralta, J. Casimiro, A. M. Nilles, J. A. Aguilar, R. Atienza, and R. Cajote, "Next-best view policy for 3d reconstruction," in *Computer Vision-ECCV 2020 Workshops*, 2020.
- [15] X. Zeng, T. Zaenker, and M. Bennewitz, "Deep reinforcement learning for next-best-view planning in agricultural applications," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2022.
- [16] H. Huang, M. Dominguez-Kuhne, V. Satish, M. Danielczuk, K. Sanders, J. Ichnowski, A. Lee, A. Angelova, V. Vanhoucke, and K. Goldberg, "Mechanical search on shelves using lateral access x-ray," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.
- [17] H. Huang, M. Danielczuk, C. M. Kim, L. Fu, Z. Tam, J. Ichnowski, A. Angelova, B. Ichter, and K. Goldberg, "Mechanical search on shelves using a novel "bluction" tool," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2022.
- [18] H. Huang, L. Fu, M. Danielczuk, C. M. Kim, Z. Tam, J. Ichnowski, A. Angelova, B. Ichter, and K. Goldberg, "Mechanical search on shelves with efficient stacking and destacking of objects," *arXiv preprint arXiv:2207.02347*, 2022.
- [19] W. Bejjani, W. C. Agboh, M. R. Dogar, and M. Leonetti, "Occlusion-aware search for object retrieval in clutter," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2021.
- [20] M. Kang, H. Kee, J. Kim, and S. Oh, "Grasp planning for occluded objects in a confined space with lateral view using monte carlo tree search," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2022.
- [21] J. Ahn, C. Kim, and C. Nam, "Coordination of two robotic manipulators for object retrieval in clutter," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2022.
- [22] C. Nam, J. Lee, S. H. Cheong, B. Y. Cho, and C. Kim, "Fast and resilient manipulation planning for target retrieval in clutter," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 2020.
- [23] J. K. Li, D. Hsu, and W. S. Lee, "Act to see and see to act: Pomdp planning for objects search in clutter," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2016.
- [24] Y. Miao, R. Wang, and K. Bekris, "Safe, occlusion-aware manipulation for online object reconstruction in confined spaces," *Proc. of the Intl. Symposium on Robotic Research (ISRR)*, 2022.
- [25] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [26] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proc. of the IEEE Intl. Conf. on Robotics & Automation (ICRA)*, 1985.
- [27] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, "Learning synergies between pushing and grasping with self-supervised deep reinforcement learning," in *Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*.
- [28] M. Q. Mohammed, K. L. Chung, and C. S. Chyi, "Pick and place objects in a cluttered scene using deep reinforcement learning," *Intl. Journal of Mechanical Engineering and Mechatronics (IJMEM)*, 2020.
- [29] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. Funkhouser, "Tossing-bot: Learning to throw arbitrary objects with residual physics," *IEEE Transactions on Robotics*, 2020.
- [30] B. Hepp, M. Nießner, and O. Hilliges, "Plan3d: Viewpoint and trajectory optimization for aerial multi-view stereo reconstruction," *ACM Transactions on Graphics*, 2018.
- [31] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme, "Interactive perception: Leveraging action in perception and perception in action," *IEEE Transactions on Robotics*, 2017.
- [32] A. Eitel, N. Hauß, and W. Burgard, "Learning to singulate objects using a push proposal network," in *Proc. of the Intl. Symposium on Robotic Research (ISRR)*, 2020.
- [33] E. Coumans and Y. Bai, "Pybullet, a python module for physics simulation for games, robotics and machine learning," <http://pybullet.org>, 2016–2023.
- [34] Stanford Artificial Intelligence Laboratory et al., "Robotic operating system." [Online]. Available: <https://www.ros.org>
- [35] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, 2021.
- [36] A. Kuznetsov, P. Shvechikov, A. Grishin, and D. Vetrov, "Controlling overestimation bias with truncated mixture of continuous distributional quantile critics," in *International Conference on Machine Learning (ICML)*, 2020.
- [37] I. A. Şucan, M. Moll, and L. E. Kavraki, "The Open Motion Planning Library," *IEEE Robotics and Automation Magazine (RAM)*, 2012, <https://ompl.kavrakilab.org>.
- [38] D. Coleman, I. Sucan, S. Chitta, and N. Correll, "Reducing the barrier to entry of complex robotic software: a moveit! case study," *arXiv preprint arXiv:1404.3785*, 2014.
- [39] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *Proc. of the Intl. Conf. on Advanced Robotics (ICAR)*, 2015.