# Robot Skill Generalization via Keypoint Integrated Soft Actor-Critic Gaussian Mixture Models

Iman Nematollahi<sup>\*1</sup>, Kirill Yankov<sup>\*1</sup>, Wolfram Burgard<sup>2</sup>, and Tim Welschehold<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Freiburg, Germany
<sup>2</sup> Department of Engineering, University of Technology Nuremberg, Germany
\* Equal contribution

**Abstract.** A long-standing challenge for a robotic manipulation system operating in real-world scenarios is adapting and generalizing its acquired motor skills to unseen environments. We tackle this challenge employing hybrid skill models that integrate imitation and reinforcement paradigms, to explore how the learning and adaptation of a skill, along with its core grounding in the scene through a learned keypoint, can facilitate such generalization. To that end, we develop Keypoint Integrated Soft Actor-Critic Gaussian Mixture Models (KIS-GMM) approach that learns to predict the reference of a dynamical system within the scene as a 3D keypoint, leveraging visual observations obtained by the robot's physical interactions during skill learning. Through conducting comprehensive evaluations in both simulated and real-world environments, we show that our method enables a robot to gain a significant zero-shot generalization to novel environments and to refine skills in the target environments faster than learning from scratch. Importantly, this is achieved without the need for new ground truth data. Moreover, our method effectively copes with scene displacements.

# 1 Introduction

Robot skill generalization remains a long-standing challenge in the field of robotic manipulation, mainly due to the stochastic nature and inherent variability of unstructured real-world environments. For effective and autonomous operation, a robot must excel not only in acquiring new motor skills, but also in adapting and generalizing these skills to unseen and evolving contexts. However, this necessity often stands in contrast with the common practice of learning robot skills in imitation and reinforcement learning settings, which typically confine a robot to operate optimally within the rigid boundaries of its training environment. Such methods frequently stumble when faced with novel environments where the distribution of observations diverges from the one encountered during training. Moreover, the lack of adaptability intrinsic to most robot skill models imposes a significant constraint, hindering the robot's capacity to refine its learned skills within the target environment.

Numerous efforts have been made to address the need for generalizable and adaptive robot skills, yet predominantly addressed them as separate challenges. On the one hand, attempts to construct generalizable robot skills have primarily focused on enabling robots to learn object-centric and task-specific visual representations, fostering a higher readiness for their transferability to diverse yet structurally similar environments [1–3].

http://kis-gmm.cs.uni-freiburg.de

#### Iman Nematollahi\*, Kirill Yankov\*, Wolfram Burgard, Tim Welschehold



**Fig. 1.** KIS-GMM employs a hybrid model to acquire adaptive and generalizable robot skills. Initiating in a source environment (illustrated on the left), where the robot has access to few skill demonstrations and a pivotal keypoint ( $\xi^*$ ) of the skill in the scene, the Gaussian Mixture Model (GMM) agent learns a dynamical system from these demonstrations and controls the robot at high-frequency *F*. The Soft Actor-Critic (SAC) and Keypoint (KEY) agents, operating at lower frequencies, respectively learn to refine skills for adaptability and predict a 3D keypoint for generalization to novel environments. In transitioning to a target environment (illustrated on the right), where neither demonstrations nor ground-truth keypoints are available, KIS-GMM effectively predicts the essential keypoint ( $\hat{\xi^*}$ ) for grounding the skill in the novel scene. This framework allows the seamless execution of skills, initially learned in a source environment, within unseen target settings.

However, the lack of adaptability in these methods hinders performance refinement when initial generalizations fail. On the other hand, endeavors to create adaptive robot skills adopt a strategy that confines the refinement of a learned skill policy exclusively to its initial training environment [4–6]. Nonetheless, while both pursuits are crucial, there has been limited exploration into the potential synergy of these two key aspects.

In this paper, we step towards closing this gap by proposing the Keypoint-Integrated Soft Actor-Critic Gaussian Mixture Models (KIS-GMM, see Figure 1). KIS-GMMs employ a hybrid skill model that begins by learning a trajectory-based Gaussian mixture dynamical system from demonstrations. This learned skill is subsequently refined through physical interactions of a soft actor-critic reinforcement learning agent within its environment. During this refinement phase, KIS-GMMs harness the robot's observations to predict a 3D keypoint corresponding to the dynamical system's reference point in the scene, empowering it to generalize to structurally similar but previously unseen environments. This is achieved by transforming the motion model to the reference frame given by the predicted keypoint in new settings. Through comprehensive evaluations, we demonstrate that our KIS-GMM equips a robot with the ability for remarkable zero-shot generalization in unseen environments. It expedites skill refinement in the target environment, outpacing the learning-from-scratch approach, and does so without the need for the new environment's ground truth data. Additionally, our model effectively handles scene displacements. We showcase KIS-GMM's capability to successfully open four visually distinct doors and drawers in the real world, having been initially trained on just one of each, respectively.

### 2 Related Work

A fundamental challenge for autonomous agents in unstructured real-world scenarios is to adapt and generalize their skill set to manipulate objects in dynamic, unfamiliar environments [7]. For skill generalization, the focus is on *where* to apply a previously learned skill in a new, unseen environment. In contrast, skill adaptation emphasizes *how* to refine the robot's skill for optimal performance. Extensive research in robotics and computer vision has delved deeply into addressing either adaptability or generalization challenges.

**Robot Skill Generalization:** A core objective of machine learning is to distill abstract information from limited training data that can be generalized to unseen observations, often leveraging similarities in the training data to formulate an inductive bias [8]. In recent years, deep learning has facilitated the zero-shot and few-shot generalization of robotic skills within an imitation learning paradigm [1, 9–14]. In robot skill learning, a common approach for incorporating inductive bias is through learning the visual representation of data. Various methods exist to derive such generalizable visual representations. This includes generative models like Variational Autoencoders (VAEs) [15] and Generative Adversarial Networks (GANs) [16] or object-centric models such as video prediction [17, 18], image or point cloud segmentation [19–23], and keypoint detection [2, 3, 24–26] networks. While robot skills using these generalization techniques can manipulate objects in novel environments, they often lack an adaptation mechanism for failed generalizations. In contrast, our skill model refines its keypoint prediction in response to sub-optimal generalizations in new scenes.

**Robot Skill Adaptation:** While robot skill adaptation techniques have traditionally leaned on end-to-end deep reinforcement learning [27] and policy search [28], there is a growing interest toward hybrid skill models aiming for better sample efficiency [4, 5]. These models begin by learning robot motion policies from supervised data and later refine them in the environment through reinforcement learning. While some approaches overlook the intrinsic policy structure during adaptation, either by optimizing policy parameters [29, 30] or adding residual action at each time step [31–33], others preserve the geometric structure of the initial policy [23, 6]. In our previous study [4], we introduced a framework to learn robot skills using Gaussian Mixture Models [34–36] and subsequently adapt them in their trajectory distribution space using a Soft Actor-Critic agent [37] that interacts with the environment. While these approaches can adapt robot policies and enhance skill performance in the initial training environment, they struggle to generalize in new settings with different visual appearances, even if the skill goals remain similar.

### **3** Problem Formulation

We consider the motion of a robot to be driven by a dynamical system regulated by a set of first-order differential equations, conditioning the robot's velocity on the current and target states [34]. Through this framework, we define a robot skill as a trajectory map, indicating the requisite motion to reach a target state from any starting state. Our goal is first to master a skill in one environment and subsequently generalize its application to unseen environments. This generalization is achieved by transferring the dynamical system's reference point, while offering the flexibility to adapt the skill in the new domain if needed.

To this end, we regard robot skills as 3D trajectories. We represent a trajectory by  $\Xi = \langle \xi_1, \ldots, \xi_n \rangle$ , where  $\xi_i \in \mathbb{R}^d$  is the robot's geometric pose. Importantly,

each  $\xi_n$  can exemplify a dynamical system's reference point, represented by  $\xi^*$ . Furthermore, each rollout of a robot skill prompts a collection of observations, denoted as  $\mathcal{O} = \langle \mathbf{o}_1, \dots, \mathbf{o}_n \rangle$ , captured from the environment through the robot's wrist camera. To learn a generalizable adaptive robot skill, we first need a series of demonstrations taken from one single environment, represented as  $\mathbf{X} = \langle \Xi_1, \dots, \Xi_j \rangle$ , with each demonstration exemplifying the geometric course of the respective skill. Our initial goal is to learn a parametric skill model, denoted by  $f_{\theta}(\xi - \xi^*) = \dot{\xi}$ , which, conditioned on the dynamical system's reference point, derives the robot's next desired velocity given the current pose. Subsequently, we aim to master the learned skill in the source environment by adapting its trajectory parameters  $\theta$  through a refinement policy  $\pi_{\phi}(\xi, \mathbf{o}) = \Delta\theta$ , considering the robot's poses and observations in a sparse task completion reward setup.

Seeking to generalize our skill model to structurally similar but unseen environments, our approach involves training a 3D keypoint detector, represented as  $K_{\omega}(\mathbf{0}) = \hat{\xi}^*$ , during the skill refinement phase in the source environment. The keypoint detector leverages the wrist camera observations to predict the dynamical system's reference point. By transferring this representation to unseen environments, the robot can operate in novel settings. If the skill's performance in the target environment proves unsatisfactory, we retain the capability to refine both the skill and the keypoint detector specifically for that environment, utilizing sparse task completion rewards garnered within this novel context.

# 4 KIS-GMM

4

We introduce the Keypoint Integrated Soft Actor-Critic Gaussian Mixture Models (KIS-GMM), a robot learning framework specifically crafted for fostering adaptive and generalizable robot skills. This framework is segmented into two distinct phases. In the first phase, a robot skill is acquired by learning a parameterized dynamical system from a few demonstrations within a single source environment. This stage masters the skill not only by adapting dynamical system parameters through physical interactions, high-dimensional sensory data, and sparse rewards but also by deriving a pivotal keypoint crucial for grounding the skill in the scene. In the second phase, the dynamical system is generalized to unseen environments that share structural similarities to the source environment by transferring the learned keypoint, conducting any necessary refinements to either the skill policy or keypoint detection if required.

#### 4.1 Mastery of a Robot Skill in a Source Environment

We employ the SAC-GMM hybrid algorithm [4], a fusion of dynamical systems and deep reinforcement learning, for an adaptive formulation of robot skills. Initial learning of a parametrized dynamical system stems from a mixture of Gaussians, derived from a handful of demonstrations (GMM agent). The Soft Actor-Critic framework (SAC agent) then exploits robot interactions with the source environment to refine this dynamical system further. Provided with the current state - including the latent observation and the robot state - and a reward for the previous action, the SAC agent adapts the original GMM for the following interactions. For detailed insights into SAC-GMM, refer to our prior work [4].

We develop a 3D keypoint detector (KEY agent), that builds upon the SAC-GMM algorithm. Our keypoint detection approach draws inspiration from and resembles the



**Fig. 2.** KIS-GMM learns to adapt a robot skill and generalize it to unseen environments. Adaptation is achieved through parameter refinements  $(\Delta \theta)$  utilizing physical interactions and sparse rewards. Generalization is facilitated by transferring a learned keypoint  $(\hat{\xi}^*)$  to novel environments.

kPAM framework [24]. During the skill refinement phase within the source environment, the Key agent utilizes the wrist camera's observations to predict the 3D position of the dynamical system's reference point in the frame of the wrist camera, represented as  $\xi^* = (x_c, y_c, z_c)$ . For clarity, within the source environment, the ground truth keypoint is designated as a specific point located on the surface of the object that is integral to the relevant skill. For example, in the case of door and drawer opening skills, this keypoint is strategically positioned on the handles. Our KEY agent employs image encoding to generate a 2D heatmap, facilitating the determination of the keypoint's coordinates within the camera's frame. Concurrently, a depth map is produced to capture the depth value of the corresponding keypoint. Together, these processes result in the determination of  $\hat{\xi^*} = (\hat{x_c}, \hat{y_c}, \hat{z_c})$ . Moreover, a keypoint presence probability evaluates the uncertainty regarding the keypoint's presence within the camera's frame. Using the ground truth keypoint of the dynamical system, we train the Keypoint detector with average Euclidean distance ( $\mathcal{L}_{ED}$ ) and the keypoint presence measure via binary cross entropy loss ( $\mathcal{L}_{KP}$ ):

$$\mathcal{L}_{ED}(\omega) = \frac{1}{N} \sum_{i=0}^{N} \sqrt{(\hat{\xi}_i^* - \xi^*)^2}, \quad \mathcal{L}_{KP}(\omega) = -\frac{1}{N} \sum_{i=0}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)].$$

To ensure that the keypoint detector effectively learns the crucial features associated with the ground truth keypoint, we employ image augmentation techniques during the training process. Specifically, we randomly obscure portions of the image pixels where the keypoint pixel is absent, enhancing the model's recognition of essential features contributing to the keypoint. During inference, we maintain a running average of the predicted keypoints for which the keypoint presence measure exceeds a predetermined threshold.

#### 4.2 Generalization and Adaptation of a Robot Skill to Unseen Environments

The ability of our skill model to learn the dynamical system's reference keypoint enables the transfer of this task-specific representation to structurally similar, yet previously unseen environments. Consequently, when faced with a new environment, our model is capable of grounding the skill in the new scene using the predicted keypoint. If performance in a target environment falls short of expectations, our model leverages sparse task completion rewards from that environment to fine-tune both the SAC and KEY agents. Since access to the ground truth reference point is unavailable in new environments, we adopt the running average of the predicted keypoints from the last three successful trajectories as a pseudo-label, enabling further refinement of the KEY agent. Iman Nematollahi<sup>\*</sup>, Kirill Yankov<sup>\*</sup>, Wolfram Burgard, Tim Welschehold



**Fig. 3.** The top row displays four simulated CALVIN environments, while the second and third rows present four real-world door and drawer setups respectively. In all cases, robot skills are trained in source environment *A* and then generalized to environments *B*, *C*, and *D*.

# **5** Experimental Evaluation

We evaluate the effectiveness of KIS-GMM in acquiring adaptable and generalizable robot skills within simulated and real-world environments. Our experiments aim to investigate: (i) the generalization ability of our robot skill model to structurally similar yet unseen environments, (ii) the time efficiency of refining a skill in a target environment versus learning it from scratch, and (iii) our model's capacity to handle scene displacements.

#### 5.1 Experimental Setup

6

We assess our approach in both simulated and real-world settings. In the simulation, we conduct our experiments within the CALVIN manipulation environments [38]. CALVIN provides four distinct environments, engineered to test the generalization of robot skills in structurally similar contexts. Each environment varies in texture and the positioning of static elements like drawers and buttons, see the first row of Figure 3. In CALVIN, We pursue the mastery of two distinct robot skills - *Drawer Opening* and *Button Pushing*.

In the real-world experiment, we examine *Door Opening* and *Drawer Opening* skills. We utilize our 7-DOF Franka Emika Panda robot arm, equipped with a parallel gripper, to interact with four distinct doors and drawers. These vary in attributes like color, texture, handle design, and placement, as depicted in the second and third rows of Figure 3. We obtain the robot's wrist camera observations using a FRAMOS Industrial Depth Camera D435e mounted on the gripper. We affix ArUco markers to doors and drawers, allowing us to identify when they are opened successfully and provide a corresponding reward. Additionally, to facilitate autonomous robot operation without human intervention, our doors and drawers are fitted with closing mechanisms. As a result, the doors and drawers close automatically when the robot lets go of the handle, initiating a new interaction episode.

During every training and test episode, the robot's end effector's starting position is randomly sampled from a Gaussian distribution. Additionally, we add a uniform noise distribution to the skill's reference point in the scene. For all simulated and real-world setups, robot skills are trained within the source environment *A*. Subsequently, we assess the generalization and adaptability of these learned skills within environments *B*, *C*, and *D*.

#### 5.2 Evaluation Protocol

We compare our KIS-GMM against the following baselines:

- GMM: This baseline utilizes the same dynamical system learned from demonstrations in our approach. However, it lacks adaptability and awareness of the dynamical system's reference point in the target environments.
- SAC-GMM (w/ GT) [4]: This skill model utilizes robot interactions and observations for adaptability and refinement of the GMM baseline. Additionally, it benefits from access to ground-truth reference points in the target environments, which aids its performance in these contexts.
- SAC-GMM: This baseline is similar to the previous one but lacks access to ground truth in the target environments, making it a closer representation of real-world scenarios.
- **KEY-GMM**: Building on the GMM baseline, this model integrates a keypoint detector to predict the dynamical system's reference point in target environments. However, it still lacks adaptability.

All quantitative results represent the mean success rate of each skill model over 100 trials, computed across three different random seeds.

#### 5.3 Experiments in Simulation

We begin by assessing our approach with the *Drawer Opening* and *Button Pushing* skills in CALVIN environments. Table 1 reports the quantitative results representing the accuracy with which each skill model generalizes to unseen target environments. Our KIS-GMM is not only effective in achieving notable zero-shot generalization to unseen environments, but it is also capable of refining its performance in target environments. Remarkably, it accomplishes this refinement without requiring access to target environments' ground truth data. As a result, KIS-GMM achieves an accuracy exceeding 90% across all environments for both assessed skills. Upon investigating the performance of the SAC-GMM baselines, we find that while they efficiently refine the initial GMM model in source environments and can generalize to unseen environments given access to ground truth data, they fail in the more realistic scenario where access to such data is unavailable. Furthermore, while the KEY-GMM models exhibit results that come close

Robot	Drawer Opening			Button Pushing				
Skill	Source	Target			Source	Target		
Model	A	В	С	D	A	В	С	D
GMM	73%	5%	4%	5%	88%	5%	6%	6%
SAC-GMM (w/GT)	96%	60%	96%	93%	95%	39%	66%	84%
SAC-GMM	96%	6%	8%	9%	95%	9%	14%	16%
KEY-GMM	67%	42%	60%	61%	94%	40%	58%	79%
KIS-GMM (Zero-Shot)	95%	52%	94%	81%	96%	33%	57%	80%
KIS-GMM (Refined)	х	93%	94%	93%	x	<b>97</b> %	94%	95%

**Table 1.** Superior performance of KIS-GMM over baseline models in generalizing and adapting *Drawer Opening* and *Button Pushing* skills, initially mastered in environment *A*, to target environments *B*, *C* and *D*.



(a) Drawer Opening in Environment B(b) Button Pushing in Environment BFig. 4. KIS-GMM swiftly refines both robot skills in the target environment B, outpacing learning the skills from scratch, without even requiring ground truth reference points.

to matching the zero-shot variations of KIS-GMM, they lack the capability to refine their performance. This often culminates in them achieving sub-optimal accuracy rates.

As Table 1 illustrates, the zero-shot generalization of skills to environment *B* can lead to subpar results, largely due to significant table texture changes. In contrast to KEY-GMM, our adaptive skill model demonstrates the capacity to refine its performance in the target environment when zero-shot generalization doesn't meet expectations. As depicted in Figure 2, the comparative timelines for skill refinement in a new domain versus mastering the skill from the ground up reveal a distinct advantage for our model. For both robot skills assessed, KIS-GMM refines the skill significantly faster than the time it would take to learn it from scratch (approximately four times quicker), importantly, without requiring the ground truth reference point of the dynamical system in the target environment.

Our subsequent experiment investigates the adaptability of various skill models to scene displacements occurring between the training and evaluation phases. Specifically, we manipulate the relative positioning of the robot and tables and delineate three tiers of displacement difficulty for adaptation: "*Easy*" signifies displacements from 0 to 10 centimeters, "*Medium*" represents displacements from 10 to 20 centimeters, and "*Hard*" encompasses displacements from 20 to 30 centimeters. All displacements are randomly sampled from a uniform distribution. As indicated in Table 2, the adaptive SAC-GMM model, while outperforming the GMM baseline in "*Easy*" displacement scenarios, falls short in managing the more challenging "*Medium*" and "*Hard*" scenarios. This limitation primarily stems from these models' inability to ground their motion models to the displaced scene by predicting reference keypoints. Both KEY-GMM and KIS-GMM skill models demonstrate a capability to handle displacement noise effectively due to their training in detecting the 3D keypoint corresponding to the dynamical system's reference point within the environment. Consequently, their performance remains robust despite considerable displacements between the training and evaluation stages.

Robot	Drawer Opening / Environment A				Button Pushing / Environment A				
Skill	Noise		Noisy		Noise	Noisy			
Model	Free	Easy	Medium	Hard	Free	Easy	Medium	Hard	
GMM	73%	23%	5%	5%	88%	35%	5%	1%	
SAC-GMM	<b>96</b> %	64%	3%	2%	95%	92%	25%	3%	
KEY-GMM	67%	72%	65%	65%	89%	93%	90%	92%	
KIS-GMM	95%	<b>97</b> %	94%	<b>97</b> %	96%	96%	92%	95%	

**Table 2.** KIS-GMM is effective in coping with scene displacements between training and evaluation stages, by employing a 3D keypoint detector for grounding its dynamical system.

Robot Skill Generalization via KIS-GMMs

$\sim$	Robot	Real World Door Opening				Real World Drawer Opening			
	Skill	Source	Target			Source	Target		
Model		A	В	С	D	A	В	С	D
GMM		56%	11%	9%	5%	67%	6%	17%	8%
SAC-GMM		93%	5%	13%	7%	96%	6%	19%	15%
KEY-GMM		49%	44%	55%	23%	65%	59%	52%	34%
KIS-GMM (Zer	o-Shot)	93%	<b>84</b> %	<b>88</b> %	41%	94%	<b>82</b> %	<b>89</b> %	86%
KIS-GMM (Ref	ined)	x	х	х	83%	x	х	х	х

**Table 3.** KIS-GMM effectively masters the task of opening four distinct doors and drawers in realworld settings, after being trained initially on just a single door and drawer (source environment A).

#### 5.4 Experiments in Real-World

Table 3 presents the accuracy results of the Door Opening and Drawer Opening skill models as they generalize to previously unseen target settings. For both of these skills, the refinement phase spanned roughly five hours of physical interactions, equating to approximately 1,000 episodes, during which the SAC and KEY agents were actively trained. These results echo the outcomes observed in the simulated environments, underscoring the robustness and general applicability of our findings. KIS-GMM demonstrates remarkable zero-shot generalization and adaptability in the real-world setting, outperforming all baseline models. GMM and SAC-GMM models, lacking awareness of the dynamical system's reference point in target environments, struggle to anchor the skill to these new contexts, leading to an inability to generalize the learned skill effectively. Additionally, though the KEY-GMM models can generalize robot skills to unfamiliar settings, their inability to adapt these skills often leads to subpar performance. Our KIS-GMM model consistently achieves an accuracy rate of over 80% across all environments for both evaluated skills, with the sole exception of the Door Opening skill in environment D. Notably, while our KEY agent adeptly identifies the door handle and anchors the skill within the scene, the door handle in environment D is designed with a slippery surface, inherently altering its point of maximum leverage compared to other environments. Recognizing this nuance, we undertook further refinement of the door opening skill model specifically for environment D. After approximately an hour of refinement (~200 episodes), the accuracy of our model increased to 83%.

# 6 Conclusions

In conclusion, our novel KIS-GMM framework promotes robot skills that are both adaptable and generalizable, emphasizing their mutual importance for superior performance across diverse scenarios. Using the robot's observations from the skill refinement phase, our model learns to identify a 3D keypoint representation that grounds the skill's dynamical system in a novel scene. Comprehensive experiments in simulation and real-world highlight the efficacy of our proposed skill model in several key areas: 1) generalizing robot skills to unseen environments and subsequently adapting them within these new domains, 2) reducing skill refinement time in a target domain over learning from scratch, even without requiring ground truth data, and 3) handling scene displacement between training and evaluation stages. A compelling extension of our research would be to explore enhancing the sample efficiency of our skill model even further. This could be achieved by incorporating multiple keypoints that compactly define the robot's task, and by granting the refinement agent access to such keypoints.

#### References

- E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn, "Bcz: Zero-shot task generalization with robotic imitation learning," in *Conference on Robot Learning*, 2022.
- J. O. von Hartz, E. Chisari, T. Welschehold, W. Burgard, J. Boedecker, and A. Valada, "The treachery of images: Bayesian scene keypoints for deep policy learning in robotic manipulation," *arXiv preprint arXiv:2305.04718*, 2023.
- J. Gao, Z. Tao, N. Jaquier, and T. Asfour, "K-vil: Keypoints-based visual imitation learning," arXiv preprint arXiv:2209.03277, 2022.
- 4. I. Nematollahi, E. Rosete-Beas, A. Röfer, T. Welschehold, A. Valada, and W. Burgard, "Robot skill adaptation via soft actor-critic gaussian mixture models," in *Int. Conf. on Robotics and Automation*, 2022.
- A. Nair, B. Zhu, G. Narayanan, E. Solowjow, and S. Levine, "Learning on the job: Selfrewarding offline-to-online finetuning for industrial insertion of novel connectors from vision," arXiv preprint arXiv:2210.15206, 2022.
- H. Ziesche and L. Rozo, "Wasserstein gradient flows for optimizing gaussian mixture policies," arXiv preprint arXiv:2305.10411, 2023.
- J. Cui and J. Trinkle, "Toward next-generation learned robot manipulation," *Science robotics*, 2021.
- T. M. Mitchell, "The need for biases in learning generalizations," *Technical Report CBM-TR-*117, Rutgers Univ., 1980.
- 9. C. Finn, T. Yu, T. Zhang, P. Abbeel, and S. Levine, "One-shot visual imitation learning via meta-learning," in *Conference on robot learning*, PMLR, 2017.
- 10. S. James, M. Bloesch, and A. J. Davison, "Task-embedded control networks for few-shot imitation learning," in *Conference on robot learning*, PMLR, 2018.
- 11. T. Yu, C. Finn, S. Dasari, A. Xie, T. Zhang, P. Abbeel, and S. Levine, "One-shot imitation from observing humans via domain-adaptive meta-learning," *Robotics: Science and Systems*, 2018.
- 12. S. Dasari and A. Gupta, "Transformers for one-shot visual imitation," in *Conference on Robot Learning*, PMLR, 2021.
- A. Zhou, E. Jang, D. Kappler, A. Herzog, M. Khansari, P. Wohlhart, Y. Bai, M. Kalakrishnan, S. Levine, and C. Finn, "Watch, try, learn: Meta-learning from demonstrations and rewards," in *International Conference on Learning Representations*, 2019.
- D.-A. Huang, S. Nair, D. Xu, Y. Zhu, A. Garg, L. Fei-Fei, S. Savarese, and J. C. Niebles, "Neural task graphs: Generalizing to unseen tasks from a single video demonstration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.
- 15. D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014.
- 17. C. Finn, I. Goodfellow, and S. Levine, "Unsupervised learning for physical interaction through video prediction," *Advances in neural information processing systems*, 2016.
- I. Nematollahi, E. Rosete-Beas, S. M. B. Azad, R. Rajan, F. Hutter, and W. Burgard, "T3vip: Transformation-based 3d video prediction," in *Int. Conf. on Intelligent Robots and Systems*, 2022.
- M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

- M. Käppeler, K. Petek, N. Vödisch, W. Burgard, and A. Valada, "Few-shot panoptic segmentation with foundation models," *arXiv preprint arXiv:2309.10726*, 2023.
- C. Xie, Y. Xiang, A. Mousavian, and D. Fox, "Unseen object instance segmentation for robotic environments," *IEEE Transactions on Robotics*, 2021.
- 22. A. Byravan and D. Fox, "Se3-nets: Learning rigid body motion using deep neural networks," in *Int. Conf. on Robotics and Automation*, 2017.
- 23. I. Nematollahi, O. Mees, L. Hermann, and W. Burgard, "Hindsight for foresight: Unsupervised structured dynamics models from physical interaction," in *Int. Conf. on Intelligent Robots and Systems*, 2020.
- L. Manuelli, W. Gao, P. Florence, and R. Tedrake, "kPAM: Keypoint affordances for categorylevel robotic manipulation," in *The International Symposium of Robotics Research*, Springer, 2019.
- 25. B. Chen, P. Abbeel, and D. Pathak, "Unsupervised learning of visual 3d keypoints for control," in *International Conference on Machine Learning*, PMLR, 2021.
- M. Vecerik, J.-B. Regli, O. Sushkov, D. Barker, R. Pevceviciute, T. Rothörl, R. Hadsell, L. Agapito, and J. Scholz, "S3k: Self-supervised semantic keypoints for robotic manipulation via multi-view consistency," in *Conference on Robot Learning*, PMLR, 2021.
- J. Ibarz, J. Tan, C. Finn, M. Kalakrishnan, P. Pastor, and S. Levine, "How to train your robot with deep reinforcement learning: lessons we have learned," *The International Journal of Robotics Research*, 2021.
- M. P. Deisenroth, G. Neumann, J. Peters, *et al.*, "A survey on policy search for robotics," *Foundations and Trends in Robotics*, 2013.
- 29. S. Lee, Y. Seo, K. Lee, P. Abbeel, and J. Shin, "Offline-to-online reinforcement learning via balanced replay and pessimistic q-ensemble," in *Conference on Robot Learning*, PMLR, 2022.
- A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," *Robotics: Science and Systems*, 2018.
- 31. T. Silver, K. Allen, J. Tenenbaum, and L. Kaelbling, "Residual policy learning," *arXiv preprint arXiv:1812.06298*, 2018.
- 32. T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine, "Residual reinforcement learning for robot control," in *Int. Conf. on Robotics and Automation*, IEEE, 2019.
- 33. S. Hoppe, M. Giftthaler, R. Krug, and M. Toussaint, "Sample-efficient learning for industrial assembly using qgraph-bounded ddpg," in *Int. Conf. on Intelligent Robots and Systems*, IEEE, 2020.
- 34. S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, 2011.
- P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in 2009 IEEE International Conference on Robotics and Automation, IEEE, 2009.
- N. B. Figueroa Fernandez and A. Billard, "A physically-consistent bayesian non-parametric mixture model for dynamical system learning," *Proceedings of Machine Learning Research*, 2018.
- T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on* machine learning, PMLR, 2018.
- O. Mees, L. Hermann, E. Rosete-Beas, and W. Burgard, "Calvin: A benchmark for languageconditioned policy learning for long-horizon robot manipulation tasks," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, 2022.