IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

SurgIRL: Towards Life-Long Learning for Surgical Automation by Incremental Reinforcement Learning

Yun-Jie Ho[1], Zih-Yun Chiu[1], Yuheng Zhi[1], and Michael C. Yip[1] *Senior Member, IEEE*

***Abstract*—Surgical automation holds immense potential to improve the outcome and accessibility of surgery. Recent studies use reinforcement learning to automate various surgical tasks. However, these policies are developed independently, and their reusability is limited when applied to other scenarios, making it more time-consuming for robots to incrementally solve tasks. Inspired by how human surgeons build their expertise, we propose Surgical Incremental Reinforcement Learning (SurgIRL). SurgIRL aims to (1) acquire new skills by referring to external policies (knowledge) and (2) build an expandable knowledge base and reuse it to solve multiple unseen tasks incrementally (incremental learning). Our SurgIRL framework includes three major components. We first define an expandable knowledge set containing heterogeneous policies that can be helpful for surgical tasks. Then, we propose Knowledge Inclusive Attention Network with mAximum Coverage Exploration (KIAN-ACE), which enhances learning performance through extensive navigation of the knowledge base. Finally, we develop incremental learning pipelines to expand and reuse a knowledge base and solve multiple surgical tasks sequentially. Our simulation experiments show that SurgIRL efficiently learns to automate ten surgical tasks separately or incrementally. We also demonstrate successful sim-to-real transfers of SurgIRL's policies on the da Vinci Research Kit (dVRK). The results represent an initial step towards lifelong robot learning for surgical automation.**

***Index Terms*—Surgical Robotics: Planning, Medical Robots and Systems, Reinforcement Learning, Incremental Learning**

I. Introduction

Autonomous robot-assisted surgery has recently attracted more attention due to its potential to improve the efficacy and accessibility of surgery. Automation techniques can alleviate the challenges of minimally invasive surgery, including the physical and ergonomic challenges of instrumentation, the training required by practitioners to be competent, as well as the observed increase in procedural time. However, automating surgery has yet to be fully realized due to its high-stakes and complex nature [1], thus requiring further research effort.

Prior studies focus on developing techniques to automate different surgical tasks, such as suturing [2]–[6], blood suction [7]–[9], tissue dissection and retraction [10], [11], and endoscopic camera control [12]–[14]. These techniques include optimization, visual servoing, imitation learning, and reinforcement learning (RL). RL, due to its less dependence on specific objective functions and high-quality expert demonstrations, has become popular in recent years.

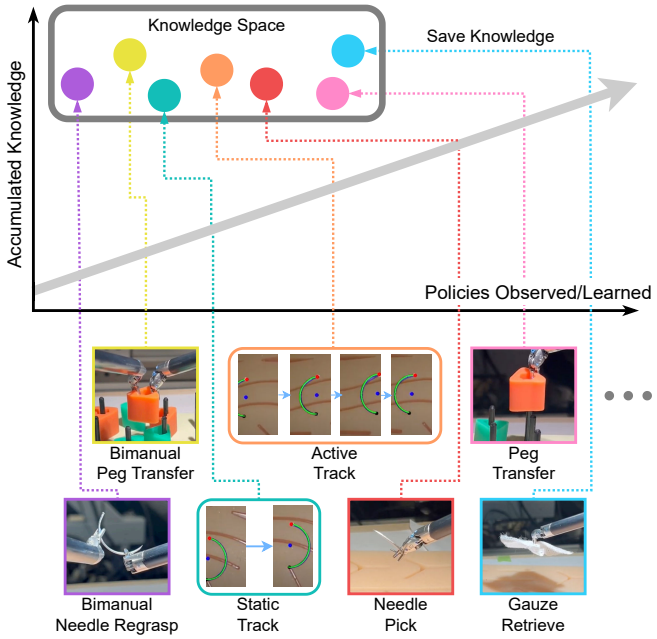


Fig. 1: A robot incrementally automates various surgical tasks, including endoscopic camera control and surgical manipulation. We propose SurgIRL that enables surgical robots to learn multiple tasks by building a knowledge base with policies from different sources. The diverse policies learned and successfully transferred to the real world demonstrate the flexibility and effectiveness of our framework.

Current efforts to automate surgical tasks by RL solve each task independently from scratch. While the trained policies can solve their assigned tasks, they are difficult to collect and form a knowledge base that helps surgical robots learn unseen tasks. Without collecting diverse policies and organizing them for future use, it takes a substantially longer time for surgical robots to automate more complex tasks.

Human surgeons, however, undergo a more efficient process to gradually develop their expertise. They often observe how others perform operations, and these external policies, together with their past experiences, become their knowledge when they perform surgery on their own [15]. This learning process is named incremental learning in prior literature [16], [17] since the knowledge will keep accumulating over time. Incremental learning is considered a key to efficient learning due to its flexibility [16], i.e., there is no constraint of when and what policies can be used to update the knowledge.

Manuscript received: April 16, 2025; Revised: August 17, 2025; Accepted: October 7, 2025.

This paper was recommended for publication by Editor Pietro Valdastri upon evaluation of the Associate Editor and Reviewers' comments.

This research was supported by the Telemedicine and Advanced Technology Research Center and NSF #2045803.

[1]Yun-Jie Ho, Zih-Yun Chiu, Yuheng Zhi, and Michael C. Yip are with the Electrical and Computer Engineering Dept., University of California San Diego, La Jolla, CA 92093 USA. {y8ho, zchiu, yzhi, yip}@ucsd.edu

Digital Object Identifier (DOI): see top of this page.

This work explores how a robot incrementally solves multiple surgical tasks by referring to an expandable knowledge set containing external policies and previously trained skills (Fig. 1). We propose an incremental learning framework, Surgical Incremental RL (SurgIRL), to achieve surgeon-level learning efficiency and target capabilities such as sample-efficient and incremental learning. SurgIRL follows the Knowledge-Grounded RL (KGRL) [17] formulation and includes three components. The first component is a knowledge set containing heterogeneous policies that may be helpful for surgical tasks. This knowledge set will be incorporated into the learning process and can be arbitrarily expanded with other policies during training. Next, we propose Knowledge Inclusive Attention Network with mAximum Coverage Exploration (KIAN-ACE) that maximizes the usage of the knowledge set during learning, leading to better training efficiency, performance, and convergence. Finally, SurgIRL includes incremental learning pipelines that can incrementally learn surgical tasks regardless of their similarity. SurgIRL's flexibility to integrate an expandable knowledge base enables surgical agents to efficiently learn a sequence of surgical tasks that require great precision and high dexterity.

We evaluate our SurgIRL agents on ten surgical tasks in simulation and real-world environments. The simulation results demonstrate that KIAN-ACE outperforms the state-of-the-art KGRL method in learning all tasks separately. Moreover, we show that our incremental learning framework effectively accumulates knowledge to solve multiple unseen tasks. Our SurgIRL agents can be continuously deployed to learn more surgical tasks beyond this work. Lastly, we run all the trained policies on the real da Vinci Research Kit (dVRK) and show successful sim-to-real transfers.

Our main contributions are as follows:

1) We introduce SurgIRL, a KGRL algorithm that enables surgical robots to build and reuse an expandable knowledge base when learning multiple tasks incrementally.
2) We propose KIAN-ACE in SurgIRL to update policies and show that it outperforms the SoTA KGRL algorithm due to its enhanced knowledge base navigation strategy.
3) We propose incremental learning pipelines in SurgIRL to sequentially learn surgical tasks with different attributes.
4) We apply SurgIRL's policies to the real dVRK and show successful sim-to-real transfer on various surgical tasks.

II. Related Work

Prior work has developed techniques to automate different surgical procedures. One line of research focuses on optimizing robot trajectories. For example, [3], [4], [8], [14], [18], [19] proposed optimization approaches for surgical tasks such as suturing, needle manipulation, blood suction, endoscope control, and tissue manipulation. Another line of research studies how visual information can effectively guide robots to complete those tasks [2], [6], [7], [11], [20]–[22].

While optimization and visual servoing methods effectively automate various surgical tasks, they require specific optimization objectives or vision modules. To circumvent the need, prior studies have looked into automating surgical tasks by learning from demonstrations (LfD). [12] leveraged demonstrations to learn features of interest in endoscopic images and guide camera positioning. [23] proposed an LfD-based model predictive control framework to manipulate soft tissues. [5] learned surgical action primitives from human demonstrations and composed them to perform suturing. [10], [24], [25] used imitation learning to learn fundamental tasks such as tissue retraction, knot tying, and peg transfer from expert demonstrations. LfD enables surgical robots to perform human-like behaviors but requires sufficient high-quality demonstrations, which are difficult to collect.

RL has recently gained popularity due to its less dependence on demonstrations. Prior work has proved that RL leads to success in automating surgical tasks such as tissue manipulation [26]–[31], needle manipulation [32]–[36], camera control [13], blood suction [9], and rope cutting [37]. In addition, [38] presented a context-aware RL framework for object pick-and-place in surgical environments. [39] showed how multi-agent RL enables surgical robots to cooperate with humans. [40] studied how to incorporate safety constraints into RL for surgical tasks. [41], [42] used expert demonstrations to improve the sample efficiency when training surgical agents with RL. Nevertheless, these works focus on using RL to solve different tasks separately without considering how the previously developed skills can be accumulated and reused to build the expertise of surgical agents incrementally. SurgIRL takes into account this essential learning behavior, aiming for efficient learning over multiple surgical tasks.

III. Methods

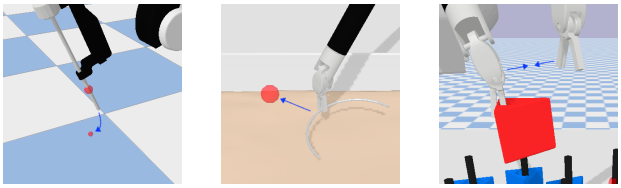
Our goal in this work is two-fold:

1) improve autonomous surgical agents' learning efficiency by flexibly observing various external policies and
2) build and organize an expandable knowledge base and reuse it to learn unseen surgical tasks incrementally.

We present SurgIRL, a framework that follows the KGRL formulation, to achieve this goal. We first review the formulation of KGRL and introduce the major components of SurgIRL: an external surgical knowledge set, a policy-learning algorithm that incorporates the knowledge set (KIAN-ACE), and incremental learning pipelines based on KIAN-ACE.

A. Knowledge-Grounded Reinforcement Learning (KGRL)

KGRL [17] is an RL paradigm studying how agents can efficiently learn by referring to an arbitrary set of external policies. Formally, a KGRL problem is mathematically formulated as a Knowledge-Grounded Markov Decision Process (KGMDP). KGMDP is defined by a tuple $\mathcal{M}_k = (\mathcal{S}, \mathcal{A}, \mathcal{T}, R, \rho, \gamma, \mathcal{G})$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the transition probability distribution, R is the reward function, ρ is the initial state distribution, γ is the discount factor, and $\mathcal{G}$ is the set of external knowledge policies. KGRL aims to find an optimal policy $\pi^*(\cdot|\cdot;\mathcal{G}) : \mathcal{S} \to \mathcal{A}$ that maximize the accumulative expected return $\mathbb{E}_{\mathbf{s}_0 \sim \rho, \mathcal{T}, \pi^*} \left[\sum_{t=0}^{T} \gamma^t R_t \right]$.



(a) First policy (b) Second policy (c) Third policy

Fig. 2: Visualization of three external knowledge policies for surgical tasks. The first policy guides a surgical manipulator to approach an object or move to a point. The second policy moves an arm with an object in hand toward a target. The third policy involves two arms trying to hand over an object.

The external knowledge set $\mathcal{G}$ contains n knowledge policies, $\{\pi_{g_1}, \ldots, \pi_{g_n}\}$. A knowledge policy can be any state-action mapping, guiding an agent to explore the environment. The external knowledge set should be flexibly expandable and easily shared among different tasks, allowing an agent to leverage the knowledge with the utmost efficiency.

A KGRL agent should be knowledge-acquirable, sample-efficient, generalizable, compositional, and incremental [16], [17]. Our SurgIRL framework follows the KGRL formulation and aims to achieve incremental learning for surgical agents.

B. Surgical Knowledge Set

SurgIRL includes an expandable external knowledge set, $\mathcal{G}$, to guide surgical policy learning. We define the initial $\mathcal{G}$ to contain three policies that can be helpful for surgical manipulation. The first policy guides a surgical manipulator to approach an object or move to a point. The second policy moves an arm with an object in hand toward a target. The third policy involves two arms approaching each other and trying to hand over an object. Fig. 2 visualizes these policies.

In this work, all tasks share the same number of external policies initially. These external policies can be irrelevant to the surgical tasks we considered. For example, the second and third external policies might not lead to success in endoscopic camera control when there is no object involved. Irreverent knowledge policies help an agent explore the environments, and unifying the external knowledge sets makes it easier to share and accumulate knowledge over tasks.

C. Knowledge Inclusive Attention Network with mAximum Coverage Exploration (KIAN-ACE)

We propose KIAN-ACE in SurgIRL that solves tasks by leveraging an external knowledge set. KIAN-ACE builds upon Knowledge-Inclusive Attention Network (KIAN) [17] and improves its efficiency in navigating the knowledge base.

KIAN-ACE adopts the same policy architecture as KIAN, as illustrated in Fig. 3. A policy has four components: an external knowledge set, an inner actor, knowledge keys, and a query. The external knowledge set contains surgical policies introduced in Section III-B and is expandable with other policies. The inner actor, $\pi_{in}(\cdot|\cdot;\theta) : \mathcal{S} \to \mathcal{A}$, is a learnable function approximator that enables an agent to develop skills different from the external ones. Each external or internal policy is

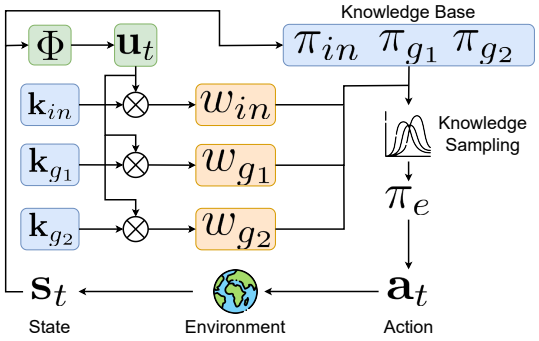


Fig. 3: The policy architecture of KIAN-ACE [17]. Given a state $\mathbf{s}_t$, the query Φ outputs a vector $\mathbf{u}_t$, which *attends* each knowledge key, $\mathbf{k}$, to calculate the weight of each policy, w. These weights are used to perform knowledge sampling. Finally, an action is generated from the sampled policy, π_e.

paired with a state-independent knowledge key, $\mathbf{k}_{in} \in \mathbb{R}^{d_k}$ or $\mathbf{k}_{g_j} \in \mathbb{R}^{d_k}$, which is a d_k-dimensional learnable embedding that represents the policy. Knowledge keys encode all policies in a unified representation space, allowing the knowledge set to consist of policies in diverse forms. Finally, the query, $\Phi(\cdot;\phi) : \mathcal{S} \to \mathbb{R}^{d_k}$, is a learnable function approximator that outputs a d_k-dimensional vector, $\mathbf{u}_t$, given a state. This output and the knowledge keys will be used to perform *knowledge sampling* during training.

KIAN-ACE generates actions through embedding-based attention [17], predicting the weight of each policy as:

$$\mathbf{u}_t = \Phi(\mathbf{s}_t;\phi),\ w_{t,in} = (\mathbf{u}_t \cdot \mathbf{k}_{in})/c_{t,in}, \tag{1}$$

$$w_{t,g_j} = (\mathbf{u}_t \cdot \mathbf{k}_{g_j})/c_{t,g_j},\ \forall j \in \{1,\ldots,n\}, \tag{2}$$

where $c_{t,in} \in \mathbb{R}$ and $c_{t,g_j} \in \mathbb{R}$ are normalization factors. Then, an agent performs *knowledge sampling* to select a policy based on these weights using Gumbel softmax [43]:

$$e \sim \texttt{gumbel\_softmax}([w_{t,in}, w_{t,g_1}, \ldots, w_{t,g_n}]^\top). \tag{3}$$

Finally, an action is sampled from $\pi_e(\cdot|\mathbf{s}_t)$. Note that all learnable components, including the inner actor, knowledge keys, and the query, can be updated by any policy-gradient algorithm. External policies are not updated to prevent impacting previously solved tasks and to limit knowledge base growth. This work uses maximum entropy algorithms such as Soft Actor-Critic (SAC) [44] to encourage knowledge usage.

KIAN-ACE and KIAN differ in how they update policies and navigate knowledge sets. KIAN's learning objective is

$$\pi^* = \arg\max_\pi \sum_t \mathbb{E}_\pi \left[R_t + \alpha H(\pi(\cdot|\mathbf{s}_t))\right], \tag{4}$$

$$H(\pi(\cdot|\mathbf{s}_t)) \approx \sum_i \pi(\mathbf{a}_i|\mathbf{s}_t) \log \pi(\mathbf{a}_i|\mathbf{s}_t), \tag{5}$$

where $\alpha \in \mathbb{R}$ is a hyperparameter, and $H(\pi(\cdot|\mathbf{s}_t))$ is the entropy of the policy approximated by multiple action samples. In maximum-entropy exploration, KIAN purely relies on maximizing the randomness of the sampled actions, which can lead to the weights, $[w_{t,in}\ w_{t,g_1} \ldots\ w_{t,g_n}]$, being highly biased [17]. Biased weights result in an agent not leveraging the knowledge set enough to achieve efficient learning. Therefore, we propose also to maximize *the randomness of knowledge*

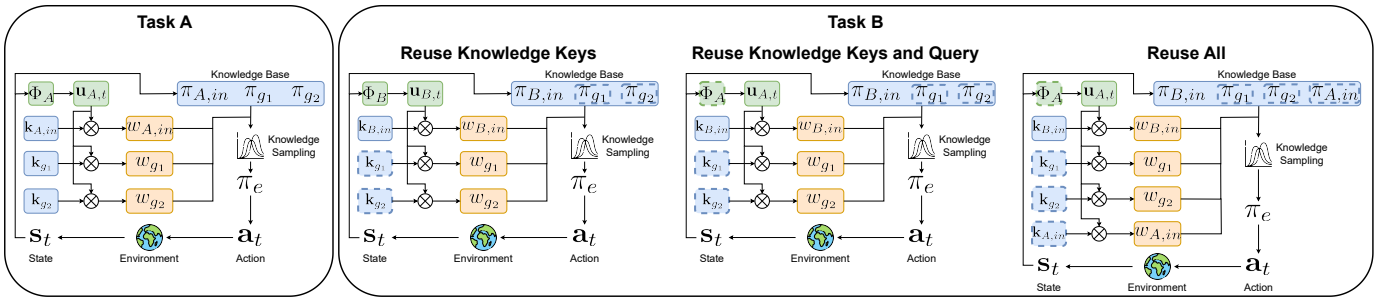


Fig. 4: SurgIRL's three pipelines to learn Task A and Task B incrementally. The dotted blocks indicate the components reused from one task to another. A robot can switch between multiple pipelines based on tasks' state/action spaces and similarity. *Reuse Knowledge Keys* suits tasks with different state/action spaces. *Reuse Knowledge Keys and Query* suits tasks with the same state/action spaces but with different environmental dynamics. *Reuse All* suits tasks with greater similarity.

sampling during maximum-entropy exploration. The objective of KIAN-ACE becomes

$$\pi^* = \arg\max_{\pi} \sum_t \mathbb{E}_\pi \left[R_t + \alpha H(\pi(\cdot|\mathbf{s}_t)) + \beta H(\hat{\mathbf{w}}_t) \right], \quad (6)$$

$$\hat{\mathbf{w}}_t = \texttt{softmax}([w_{t,in}, w_{t,g_1}, \ldots, w_{t,g_n}]^T) \quad (7)$$

$$H(\hat{\mathbf{w}}_t) = -\hat{w}_{t,in} \cdot \log \hat{w}_{t,in} - \sum_{j=1}^{n} \hat{w}_{t,g_j} \cdot \log \hat{w}_{t,g_j}, \quad (8)$$

where $\beta \in \mathbb{R}$ is a hyperparameter. By adding the term $H(\hat{\mathbf{w}}_t)$, the weights of all knowledge policies become more uniform. Uniform weights ensure that an agent *maximizes the usage (coverage) of the knowledge set* during exploration.

The value of β controls the degree of knowledge set coverage. Higher β encourages navigating the knowledge set.

D. Decay of β

As an agent gets better at a task, it explores the knowledge set less frequently and focuses more on maximizing the expected return, mirroring the exploration-exploitation trade-off observed in general RL. Prior literature has shown that this trade-off can be task-dependent, and the learning process can be sensitive to the speed of decaying exploration [45].

We propose different scheduling functions to decay β in (6) throughout the training process. We aim to study their effectiveness on tasks with different levels of difficulty.

$$\beta_{linear}(t) = \max(1 - \frac{\gamma t}{T},\ c_e), \quad (9)$$

$$\beta_{cosine}(t) = \max\left(\frac{1}{2}\left(1 + \cos\left(\tfrac{\gamma \pi t}{T}\right)\right), c_e\right), \quad (10)$$

$$\beta_{exp}(t) = e^{-\gamma t} + c_e, \quad (11)$$

where t is the current timestep, T is the total number of timesteps, $\gamma \in \mathbb{R}$ is a constant that determines the decay rate, and $c_e \in \mathbb{R}$ is a constant that ensures some exploration throughout the training process, leaving room for further improved performance. Equation (9) is a linear decay function that follows a fixed decay rate; (10) is a cosine annealing function, which leads to a slower decay rate in the early training stage and a faster rate in the late training stage; (11) is an exponential decay function, which leads to a faster decay rate in the early training stage and a slower rate in the late training stage. The characteristics of these decay functions are also discussed in prior literature [46].

We investigate through experiments the effectiveness of these decay functions when learning with KIAN-ACE. We expect that for simple tasks, a function decaying faster in the early training stage will lead to better performance; for challenging tasks, a decay function with a slower decay rate is more likely to guide an agent to find an optimal policy.

E. Incremental Learning for Surgical Tasks

In SurgIRL, we present incremental learning pipelines based on KIAN-ACE's model architecture to expand and organize a knowledge set. This knowledge set can further benefit multiple surgical tasks by reusing (1) the previously observed or learned policies, (2) their internal relationships, and (3) the knowledge set navigation strategies.

In KIAN-ACE, the reusable *knowledge* includes the external/inner policies, the knowledge keys, and the query. External and internal policies are previously developed skills that can help explore new environments. A knowledge key encodes a policy to a representation space shared by all policies, enabling an agent to identify their internal relationships. A query informs an agent which policies are more helpful for a task. Knowledge keys and queries together guide an agent to navigate a knowledge set.

We propose three incremental learning pipelines that reuse different components of KIAN-ACE to achieve knowledge accumulation. A robot can adopt any of them or switch between multiple pipelines when learning a sequence of tasks. Here, we describe potential scenarios of using each pipeline based on the state/action and task similarity.

1) For tasks with different observation/action spaces, an agent can **only reuse knowledge keys**, which are state-independent, to acquire the relationships between policies.
2) For tasks with the same space but with different environmental dynamics, an agent can **reuse knowledge keys and queries** to efficiently navigate a knowledge set.
3) For tasks with greater similarity, an agent can **reuse all components** to take full advantage of prior knowledge.

Fig. 4 visualizes the incremental learning pipelines. To further streamline the incremental learning process, task similarity measures, such as [47], can be integrated to determine task relationships and select appropriate pipelines. We leave this integration as future work. Note that there is no limitation

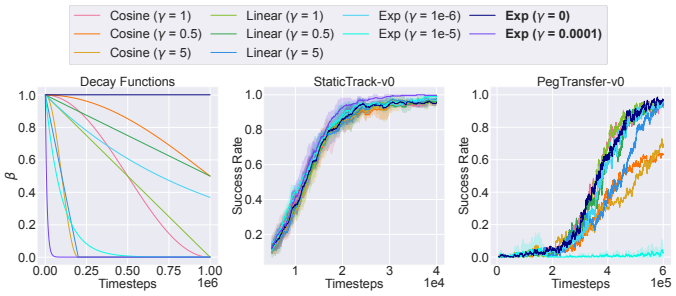


Fig. 5: Visualization of the decay functions with different values of γ and the corresponding learning curves. In StaticTrack (simple), the exponential decay function with a larger γ results in better training efficiency. In PegTransfer (challenging), a constant β, i.e., the exponential decay function with $\gamma = 0$, slightly outperforms other decay functions.

on the number of tasks learned incrementally or their order. The freedom to learn any sequence of tasks highlights the flexibility of our SurgIRL framework.

IV. Experiments and Results

We evaluate our SurgIRL framework on ten surgical tasks [48]. These tasks include (1) controlling surgical manipulators, such as a patient side manipulator (PSM) and an endoscopic camera manipulator (ECM), (2) object manipulation, such as gauze retrieval, needle pickup, peg transfer, bimanual needle regrasping, and bimanual peg transfer, and (3) camera viewpoint control and object tracking.

A. Training Setup

We train SurgIRL agents in the simulation platform SurRoL [48]. We modify some environments in SurRoL for simulation and training stability. In GauzeRetrieve, NeedlePick, PegTransfer, NeedleRegrasp, and BiPegTransfer, a policy does not control whether a gripper is closed/opened. A gripper is automatically closed whenever its distance to an object is smaller than a threshold, which ranges from 0.01 to 0.1 in different tasks. Similarly, a gripper is automatically opened whenever the distance between an object and the goal is smaller than 0.01, or the other gripper holds the object. Additionally, we design a dense reward function as follows:

$$R_t = -c_{og} \cdot d_{t,og} - c_{ro} \cdot d_{t,ro} - c_{rg} \cdot d_{t,rg} - p + 20 \cdot success, \quad (12)$$

where $d_{t,og}, d_{t,ro}$, and $d_{t,rg} \in \mathbb{R}$ are the object-goal distance, robot-object distance, and robot-goal distance, respectively, c_{og}, c_{ro}, and $c_{rg} \in \mathbb{R}$ are the coefficients of these distances, $p \in \mathbb{R}_{\geq 0}$ is the penalty when collision happens, and $success$ indicates whether the task is successfully completed. The values of c_{og}, c_{ro}, and c_{rg} lies within $\{0, 1\}$, and $p \in \{0, 2\}$.

B. Simulation Experiments

1) Experimental Setup: We compare the performance of four different algorithms: external policies as introduced in Section III-B, RL (SAC [44]), KIAN [17], and our KIAN-ACE. KIAN and KIAN-ACE use SAC to perform actor-critic policy learning. The actor architecture of RL and the inner actor architectures of KIAN and KIAN-ACE are multi-layer perceptrons (MLPs) with three hidden layers and a hidden size of 512 units. The dimension of each knowledge key $d_k = 4$. The query network is an MLP with three hidden layers and a hidden size of 64 units. The actor learning rate ranges from 5×10^{-5} to 3×10^{-4}. The entropy coefficient α lies within $\{10^{-1}, 10^{-5}\}$ or can be automatically tuned [44].

2) Analyses of Exploration Decay: We first study how the speed of decaying exploration affects KIAN-ACE's learning performance. We test the linear, cosine, and exponential decay functions proposed in Section III-D on a simpler environment, StaticTrack, and a more complex one, PegTransfer.

Fig. 5 visualizes each decay function with different values of γ and their corresponding learning curves. Each error band in Fig. 5 represents the 95% confidence interval. In StaticTrack, the exponential decay function, $\beta_{exp}(t)$, with a larger γ outperforms others. In PegTransfer, the decay functions that decay slower in the early training stage perform better. Additionally, we found that a constant β without decay, i.e., the exponential decay function with $\gamma = 0$, leads to a slightly better result. These observations align with our expectation that for simple tasks, β should decay faster in the early training stage to achieve better training efficiency, and a non-decaying β ensures sufficient exploration to master more complicated tasks. Since the exponential decay function, $\beta_{exp}(t)$, generally outperforms others, we use it for the subsequent experiments with γ lying within $\{0, 2 \times 10^{-4}\}$.

3) Single-Task Learning: Fig. 6 shows the learning curves of each task trained separately. Each error band in Fig. 6 represents the 95% confidence interval. KIAN-ACE achieves the best performance and sample efficiency in all tasks. The improvement is more evident in challenging tasks such as NeedleRegrasp and BiPegTransfer. In addition, the error bands of KIAN-ACE are smaller than those of other methods, indicating that its training results are more consistent.

4) Incremental Learning: We evaluate SurgIRL's incremental learning pipelines, as introduced in Section III-E, by dividing the ten surgical tasks into four groups. Table I shows each group's task sequence, task similarity, and the incremental learning pipeline used. The policy of the first task in each group comes from Fig. 6. We perform the following modifications to the incremental learning pipelines or environments based on our understanding of the surgical tasks: For tasks in Group 2, we only reuse $\mathbf{k}_{g_1}$ and $\mathbf{k}_{g_2}$ instead of all external knowledge keys due to the last external policies of these tasks being inconsistent. For GauzeRetrieve in Group 3, we do not reuse the inner actors of previous tasks since the object dynamics of a gauze is very different from that of a needle or a cube. For tasks in Group 4, we unify the state space of StaticTrack and ActiveTrack so that the query and inner actor can be reused between them.

Fig. 7 shows the learning curves of each task trained incrementally. Our SurgIRL framework successfully learns multiple surgical tasks in sequence and achieves better or similar results compared to other single-task agents. For tasks with greater similarity, such as NeedlePick and PegTransfer in Group 3 and StaticTrack and ActiveTrack in Group 4, incremental learning can effectively improve the training efficiency. Otherwise,

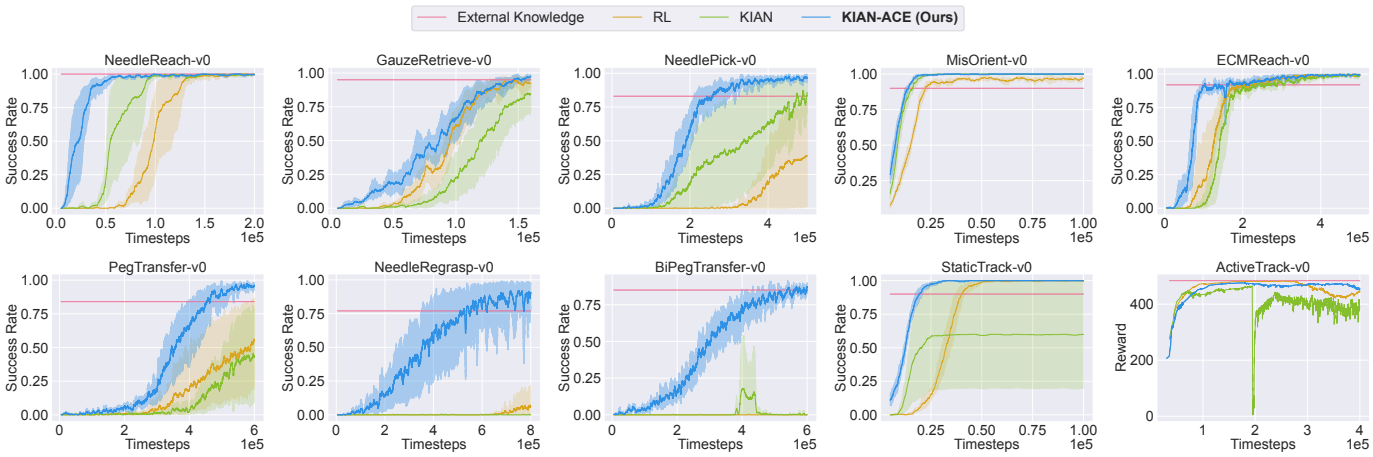


Fig. 6: The performance of external policies, RL [44], KIAN [17], and KIAN-ACE in single-task learning. KIAN-ACE outperforms other methods and has more consistent training results, demonstrating the effectiveness of its exploration strategy.

TABLE I: Incremental Learning Experimental Setup

Group	Task Sequence	Task Similarity	Incremental Learning Pipeline (Fig. 4)
1	MisOrient, ECMReach, NeedleReach	Different observation/action spaces	Reuse knowledge keys
2	NeedleRegrasp, BiPegTransfer	Same observation/action spaces but different environmental dynamics	Reuse knowledge keys and query
3	NeedlePick, PegTransfer GauzeRetrieve	Greater Similarity	Reuse all components GauzeRetrieve does not reuse inner actors
4	StaticTrack, ActiveTrack	Greater Similarity	Reuse all components

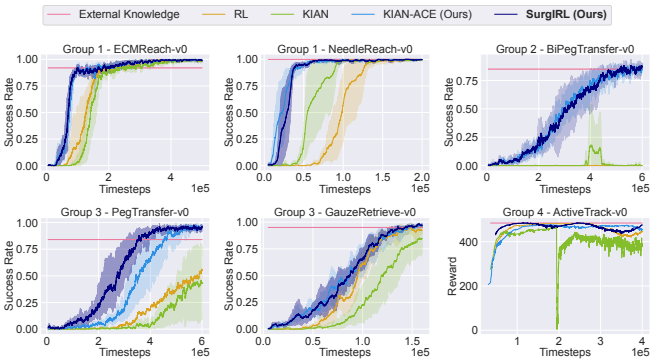


Fig. 7: Incremental learning results. Overall, SurgIRL performs better or comparable. The improvement is more evident for similar tasks. Note that our SurgIRL agents can be continuously improved by training in more surgical tasks.

an agent needs enough interaction samples to succeed in an unseen task. Moreover, if the environmental dynamics in two tasks differ, such as NeedleRegrasp and BiPegTransfer in Group 2 and PegTransfer and GauzeRetrieve in Group 3, reusing the query does not always provide helpful information on navigating the knowledge set, leading to comparable results. However, our SurgIRL agents can be continuously improved by training in more surgical tasks, opening the door to life-long surgical learning.

C. Real-Robot Experiments

1) Experimental Setup: We deploy the trained SurgIRL policies on a real dVRK [49] to demonstrate the sim-to-real transferability. A PSM attaches a Large Needle Driver that interacts with the suture needles, gauze, or blocks. An ECM holds a 1080p stereo endoscope that runs at 30 fps.

TABLE II: Performance of SurgIRL agents on the real dVRK

Success Rate	GauzeRetrieve 20 / 20	NeedlePick 19 / 20	PegTransfer 20 / 20
Success Rate	NeedleRegrasp 18 / 20		BiPegTransfer 20 / 20
	StaticTrack		ActiveTrack
Position Distance (mm)	2.22 ± 0.54		3.22 ± 0.34
Pixel Distance	29.86 ± 12.91		28.31 ± 13.82

We run prior visual tracking methods to provide an input state, $\mathbf{s}_t$, to a policy. The PSM's end-effectors and needles are tracked with [50]–[52], which run at 20fps. We use Segment Anything Model (SAM) [53] and Cutie [54] to obtain real-time needle detections. The gauze and blocks are detected using ArUco markers [55]. The ECM's end-effector is detected in its base frame using forward kinematics.

2) Experimental Results: We evaluate the best-performed models in Fig. 7 on the real dVRK. Table II shows their success rate or distance to the goal, and Fig. 8 demonstrates that our SurgIRL agents successfully automate multiple surgical tasks. We run 20 trials for each task. Note that we do not specifically evaluate MisOrient, ECMReach, and NeedleReach since they are subtasks of other tasks. Our SurgIRL policies achieve over 90% success rate on all PSM tasks, demonstrating effective sim-to-real transfers.

For ECM tasks, the goal is to move the camera such that a suture needle appears at the center of an image. After the ECM's motions are stabilized, we record the camera's distance to the goal in the 3D and image spaces. The positional errors

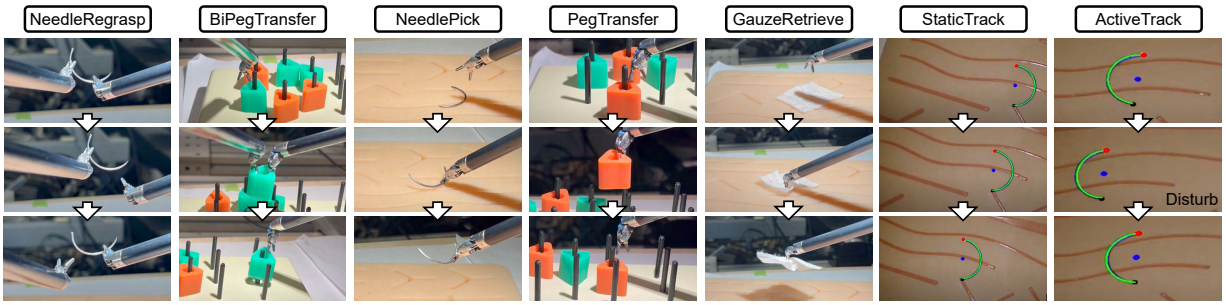


Fig. 8: Our SurgIRL agents demonstrate successful sim-to-real transfers across diverse surgical tasks using a real dVRK.

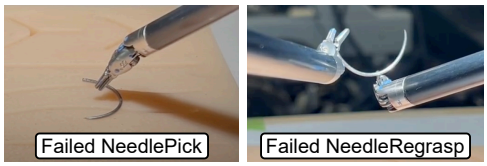


Fig. 9: Failed needle manipulation in real-world environments. The failures are due to inaccurate tool localization.

are below 4mm for all trails, and the pixel errors are below 40 for StaticTrack and below 60 for ActiveTrack. The main reasons for these errors are (1) the goal might be out of the ECM's reachable space and (2) the ECM's motors do not provide enough torque to reach the desired pose. However, these errors are not significant.

Since we do not consider environmental uncertainty when we train SurgIRL policies in this work, the failure cases in real-world environments can be attributed to inaccurate pose tracking. As shown in Table II and Fig. 9, inaccurate pose tracking affects suture-needle manipulation tasks more than others. Detecting a needle in an image is more challenging because it is highly reflective under light and very thin.

V. Conclusion and Future Work

We introduce SurgIRL, a surgical incremental learning framework aiming to (1) learn new surgical tasks by referring to a set of external policies and (2) build and reuse a knowledge base to solve unseen tasks incrementally. SurgIRL contains external surgical policies, a policy update algorithm, KIAN-ACE, leveraging external policies and improving knowledge base navigation efficiency, and incremental learning pipelines to expand and reuse a knowledge base for a sequence of surgical tasks. We demonstrate the effectiveness of SurgIRL in simulation and on a real dVRK.

Future work includes (1) integrating cross-embodied policy learning methods, such as [56], into SurgIRL for cross-platform generalization, and (2) filtering non-helpful knowledge policies while retraining informative ones. This work serves as a first step in surgical incremental learning, and we hope it opens the door for building lifelong surgical agents.

References

[1] B. T. Ostrander, D. Massillon, L. Meller, Z.-Y. Chiu, M. Yip, and R. K. Orosco, "The current state of autonomous suturing: a systematic review," *Surgical Endoscopy*, vol. 38, no. 5, pp. 2383–2397, 2024.

[2] S. Iyer, T. Looi, and J. Drake, "A single arm, single camera system for automated suturing," in *2013 IEEE international conference on robotics and automation*. IEEE, 2013, pp. 239–244.

[3] S. Sen, A. Garg, D. V. Gealy, S. McKinley, Y. Jen, and K. Goldberg, "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2016, pp. 4178–4185.

[4] S. A. Pedram, C. Shin, P. W. Ferguson, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous suturing framework and quantification using a cable-driven surgical robot," *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 404–417, 2020.

[5] K. L. Schwaner, I. Iturrate, J. K. Andersen, P. T. Jensen, and T. R. Savarimuthu, "Autonomous bi-manual surgical suturing based on skills learned from demonstration," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4017–4024.

[6] K. Hari, H. Kim, W. Panitch, K. Srinivas, V. Schorp, K. Dharmarajan, S. Ganti, T. Sadjadpour, and K. Goldberg, "Stitch: Augmented dexterity for suture throws including thread coordination and handoffs," *arXiv preprint arXiv:2404.05151*, 2024.

[7] F. Richter, S. Shen, F. Liu, J. Huang, E. K. Funk, R. K. Orosco, and M. C. Yip, "Autonomous robotic suction to clear the surgical field for hemostasis using image-based blood flow detection," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1383–1390, 2021.

[8] J. Huang, F. Liu, F. Richter, and M. C. Yip, "Model-predictive control of blood suction for surgical hemostasis using differentiable fluid simulations," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 380–12 386.

[9] Y. Ou, A. Soleymani, X. Li, and M. Tavakoli, "Autonomous blood suction for robot-assisted surgery: A sim-to-real reinforcement learning approach," *IEEE Robotics and Automation Letters*, vol. 9, no. 8, pp. 7246–7253, 2024.

[10] A. Pore, E. Tagliabue, M. Piccinelli, D. Dall'Alba, A. Casals, and P. Fiorini, "Learning from demonstrations for autonomous soft-tissue retraction," in *2021 international symposium on medical robotics (ISMR)*. IEEE, 2021, pp. 1–7.

[11] K.-H. Oh, L. Borgioli, M. Zefran, L. Chen, and P. C. Giulianotti, "A framework for automated dissection along tissue boundary," *arXiv preprint arXiv:2310.09669*, 2023.

[12] J. J. Ji, S. Krishnan, V. Patel, D. Fer, and K. Goldberg, "Learning 2d surgical camera motion from demonstrations," in *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2018, pp. 35–42.

[13] Y. H. Su, K. Huang, and B. Hannaford, "Multicamera 3d viewpoint adjustment for robotic surgery via deep reinforcement learning," *Journal of Medical Robotics Research*, vol. 6, no. 1-2, 2021.

[14] R. Moccia and F. Ficuciello, "Autonomous endoscope control algorithm with visibility and joint limits avoidance constraints for da vinci research kit robot," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 776–781.

[15] D. J. Harris, S. J. Vine, M. R. Wilson, J. S. McGrath, M.-E. LeBel, and G. Buckingham, "The effect of observing novice and expert performance on acquisition of surgical skills on a robotic platform," *PLoS One*, vol. 12, no. 11, p. e0188233, 2017.

[16] L. P. Kaelbling, "The foundation of efficient robot learning," *Science*, vol. 369, no. 6506, pp. 915–916, 2020.

[17] Z.-Y. Chiu, Y.-L. Tuan, W. Y. Wang, and M. Yip, "Flexible attention-based multi-policy fusion for efficient deep reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[18] X. Liang, F. Liu, Y. Zhang, Y. Li, S. Lin, and M. Yip, "Real-to-sim deformable object manipulation: Optimizing physics models with residual mappings for robotic surgery," in *2024 IEEE International*

Conference on Robotics and Automation (ICRA). IEEE, 2024, pp. 15 471–15 477.

[19] N. U. Shinde, Z.-Y. Chiu, F. Richter, J. Lim, Y. Zhi, S. Herbert, and M. C. Yip, "Surestep: An uncertainty-aware trajectory optimization framework to enhance visual tool tracking for robust surgical automation," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 6953–6960.

[20] C. D'Ettorre, G. Dwyer, X. Du, F. Chadebecq, F. Vasconcelos, E. De Momi, and D. Stoyanov, "Automated pick-up of suturing needles for robotic surgical assistance," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1370–1377.

[21] O. Özgüner, T. Shkurti, S. Lu, W. Newman, and M. C. Çavuşoğlu, "Visually guided needle driving and pull for autonomous suturing," in *2021 IEEE 17th international conference on automation science and engineering (CASE)*. IEEE, 2021, pp. 242–248.

[22] A. Wilcox, J. Kerr, B. Thananjeyan, J. Ichnowski, M. Hwang, S. Paradis, D. Fer, and K. Goldberg, "Learning to localize, grasp, and hand over unmodified surgical needles," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 9637–9643.

[23] C. Shin, P. W. Ferguson, S. A. Pedram, J. Ma, E. P. Dutson, and J. Rosen, "Autonomous tissue manipulation via surgical robot using learning based model predictive control," in *2019 International conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 3875–3881.

[24] J. W. Kim, T. Z. Zhao, S. Schmidgall, A. Deguet, M. Kobilarov, C. Finn, and A. Krieger, "Surgical robot transformer (srt): Imitation learning for surgical tasks," *arXiv preprint arXiv:2407.12998*, 2024.

[25] K. Kawaharazuka, K. Okada, and M. Inaba, "Robotic constrained imitation learning for the peg transfer task in fundamentals of laparoscopic surgery," *arXiv preprint arXiv:2405.03440*, 2024.

[26] D. Baek, M. Hwang, H. Kim, and D.-S. Kwon, "Path planning for automation of surgery robot based on probabilistic roadmap and reinforcement learning," in *2018 15th international conference on ubiquitous robots (UR)*. IEEE, 2018, pp. 342–347.

[27] N. D. Nguyen, T. Nguyen, S. Nahavandi, A. Bhatti, and G. Guest, "Manipulating soft tissues by deep reinforcement learning for autonomous robotic surgery," in *2019 IEEE International Systems Conference (SysCon)*. IEEE, 2019, pp. 1–7.

[28] A. Pore, D. Corsi, E. Marchesini, D. Dall'Alba, A. Casals, A. Farinelli, and P. Fiorini, "Safe reinforcement learning using formal verification for tissue retraction in autonomous robotic-assisted surgery," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4025–4031.

[29] P. M. Scheikl, E. Tagliabue, B. Gyenes, M. Wagner, D. Dall'Alba, P. Fiorini, and F. Mathis-Ullrich, "Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery," *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 560–567, 2022.

[30] Y. Ou and M. Tavakoli, "Sim-to-real surgical robot learning and autonomous planning for internal tissue points manipulation using reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2502–2509, 2023.

[31] A. A. Shahkoo and A. A. Abin, "Autonomous tissue manipulation via surgical robot using deep reinforcement learning and evolutionary algorithm," *IEEE Transactions on Medical Robotics and Bionics*, vol. 5, no. 1, pp. 30–41, 2023.

[32] V. M. Varier, D. K. Rajamani, N. Goldfarb, F. Tavakkolmoghaddam, A. Munawar, and G. S. Fischer, "Collaborative suturing: A reinforcement learning approach to automate hand-off task in suturing for surgical robots," in *2020 29th IEEE international conference on robot and human interactive communication (RO-MAN)*. IEEE, 2020, pp. 1380–1386.

[33] H. Su, Y. Hu, Z. Li, A. Knoll, G. Ferrigno, and E. De Momi, "Reinforcement learning based manipulation skill transferring for robot-assisted minimally invasive surgery," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2203–2208.

[34] Z.-Y. Chiu, F. Richter, E. K. Funk, R. K. Orosco, and M. C. Yip, "Bimanual regrasping for suture needles using reinforcement learning for rapid motion planning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 7737–7743.

[35] R. Bendikas, V. Modugno, D. Kanoulas, F. Vasconcelos, and D. Stoyanov, "Learning needle pick-and-place without expert demonstrations," *IEEE Robotics and Automation Letters*, vol. 8, no. 6, pp. 3326–3333, 2023.

[36] M. Caianiello, C. Iacono, A. Imperato, and F. Ficuciello, "Exploring the use of deep reinforcement learning algorithms for wound-approaching trajectories in robot-assisted minimally invasive surgery," in *2023 21st International Conference on Advanced Robotics (ICAR)*. IEEE, 2023, pp. 285–290.

[37] M. Haiderbhai, R. Gondokaryono, A. Wu, and L. A. Kahrs, "Sim2real rope cutting with a surgical robot using vision-based reinforcement learning," *IEEE Transactions on Automation Science and Engineering*, 2024.

[38] C. D'Ettorre, S. Zirino, N. N. Dei, A. Stilli, E. De Momi, and D. Stoyanov, "Learning intraoperative organ manipulation with context-based reinforcement learning," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 8, pp. 1419–1427, 2022.

[39] P. M. Scheikl, B. Gyenes, T. Davitashvili, R. Younis, A. Schulze, B. P. Müller-Stich, G. Neumann, M. Wagner, and F. Mathis-Ullrich, "Cooperative assistance in robotic surgery through multi-agent reinforcement learning," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1859–1864.

[40] K. Fan, Z. Chen, G. Ferrigno, and E. De Momi, "Learn from safe experience: Safe reinforcement learning for task automation of surgical robot," *IEEE Transactions on Artificial Intelligence*, 2024.

[41] T. Huang, K. Chen, B. Li, Y.-H. Liu, and Q. Dou, "Guided reinforcement learning with efficient exploration for task automation of surgical robot," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 4640–4647.

[42] Y. Ou and M. Tavakoli, "Towards safe and efficient reinforcement learning for surgical robots using real-time human supervision and demonstration," in *2023 International Symposium on Medical Robotics (ISMR)*. IEEE, 2023, pp. 1–7.

[43] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[44] 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, pp. 1861–1870.

[45] L. Schäfer, F. Christianos, J. Hanna, and S. V. Albrecht, "Decoupling exploration and exploitation in reinforcement learning," in *ICML 2021 Workshop on Unsupervised Reinforcement Learning*, 2021.

[46] M. Li, E. Yumer, and D. Ramanan, "Budgeted training: Rethinking deep neural network training under resource constraints," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=HyxLRTVKPH

[47] X. Liu, Y. Bai, Y. Lu, A. Soltoggio, and S. Kolouri, "Wasserstein task embedding for measuring task similarities," *Neural Networks*, vol. 181, p. 106796, 2025.

[48] J. Xu, B. Li, B. Lu, Y.-H. Liu, Q. Dou, and P.-A. Heng, "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, 2021, pp. 1821–1828.

[49] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da vinci® surgical system," in *2014 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 6434–6439.

[50] F. Richter, J. Lu, R. K. Orosco, and M. C. Yip, "Robotic tool tracking under partially visible kinematic chain: A unified approach," *IEEE Transactions on Robotics*, vol. 38, no. 3, pp. 1653–1670, 2021.

[51] Z.-Y. Chiu, A. Z. Liao, F. Richter, B. Johnson, and M. C. Yip, "Markerless suture needle 6d pose tracking with robust uncertainty estimation for autonomous minimally invasive robotic surgery," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 5286–5292.

[52] Z.-Y. Chiu, F. Richter, and M. C. Yip, "Real-time constrained 6d object-pose tracking of an in-hand suture needle for minimally invasive robotic surgery," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 4761–4767.

[53] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[54] H. K. Cheng, S. W. Oh, B. Price, J.-Y. Lee, and A. Schwing, "Putting the object back into video object segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3151–3161.

[55] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.

[56] R. Doshi, H. R. Walke, O. Mees, S. Dasari, and S. Levine, "Scaling cross-embodied learning: One policy for manipulation, navigation, locomotion and aviation," in *Conference on Robot Learning*. PMLR, 2025, pp. 496–512.