2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Learning Optimal Strategies for Needle Handover in Surgical Suturing

Cholin Kim[1*], Jeonghyeon Yoon[1*], Hakyun Lee[2], Sihyeoung Park[1], Hyojae Park[1], Seungjun Lee[3], Michael Yip[4], and Minho Hwang[1†]

***Abstract*— Automation of suturing subtasks, such as needle handover, has the potential to reduce surgeons' fatigue and improve surgical efficiency. Needle handover is particularly challenging due to the combinatorial nature of grasping and handover strategies, uncertainties in needle pose estimation, and inaccuracies inherent in cable-driven surgical robots such as the da Vinci system. In this work, we present a reinforcement learning framework for needle handover, spanning the process from initial pickup to a desired grasping state. We formulate the task as a goal-oriented planning problem and design a state–action representation that captures grasping and handover configurations. A DQN-based policy is trained with disturbances that reflect real-world kinematic errors to ensure robustness. The learned policy was validated on the da Vinci Research Kit (dVRK) and quantitatively compared with human teleoperation. Results demonstrate that our approach achieves human-level efficiency in terms of handover attempts (1.65 ± 0.50 vs. 1.62 ± 0.55), while improving consistency and joint-limit avoidance. The proposed framework demonstrates the potential of reinforcement learning for safe and reliable automation of surgical handover and points to opportunities for extending autonomy to more complex handover scenarios.**

I. INTRODUCTION

Robotics integrated with AI is expanding the possibilities of automation across industrial and everyday contexts. Alongside these advances, surgical automation has gained increasing attention. Nevertheless, robotic surgery remains in its infancy with respect to automation. Even small errors directly affect patient safety, while the acquisition and annotation of surgical data are limited by privacy constraints. These challenges hinder the development of robust and reliable learning-based automation systems.

Automating repetitive subtasks such as suturing and knot tying can reduce surgeon fatigue and enable surgeons to focus on high-level decision making [1]. Motivated by this need, prior work has explored automation across a variety of subtasks including suturing [2], [3], peg transfer [4], [5], and knot tying [6]. Among them, suturing is particularly important yet depends on the successful resolution of challenging subtasks such as needle and tool pose estimation [7]–[9], needle reorientation [10], thread manipulation [11], and deformable tissue perception [12]. These components have been extensively investigated, underscoring their importance in enabling robust suturing automation.

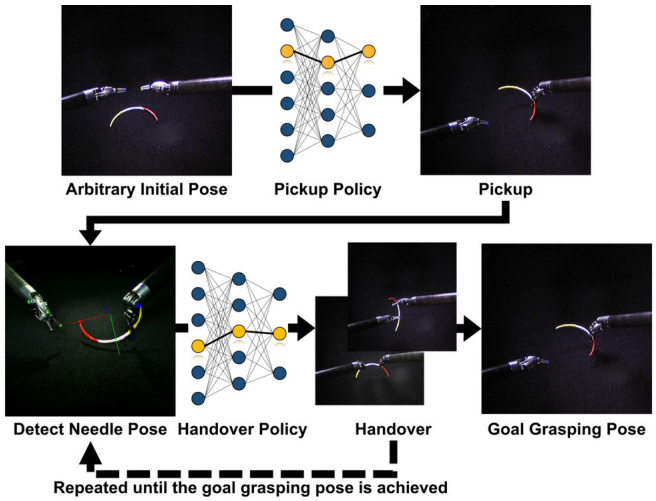

Fig. 1. **Overview of the proposed framework.** The algorithm observes arbitrary initial needle poses and computes an optimal sequence of pickup and handover actions to reach the goal grasping state. The needle pose is updated after each handover, and the process is repeated until the goal grasping state is achieved.

Handover is a critical step in suturing, as it aligns the needle with the desired stitching trajectory. Since most suture throws require reorientation [13], repeated handovers are inevitable. Establishing efficient handover strategies is therefore a key challenge in surgical automation. Proper grasping of the needle improves both efficiency and accuracy [14], highlighting the need for methods that minimize unnecessary handovers while maintaining reliability.

However, automating handover is difficult for two main reasons. First, the number of possible strategies grows exponentially with the initial needle pose and target grasping state, making high-level decision making complex. Second, handover requires precise bimanual coordination under uncertainties such as needle slippage and pose estimation errors. These challenges necessitate methods that can robustly reduce handover counts while preserving stability. Reducing unnecessary handovers shortens procedure time and improves reliability by decreasing the frequency of potentially error-prone manipulations.

* Equal Contribution. † Corresponding author.
[1]Cholin Kim, Jeonghyeon Yoon, Sihyeoung Park, Hyojae Park, and Minho Hwang are with the Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Republic of Korea {cholin, yjh1434, psh120, hyojae, minho}@dgist.ac.kr
[2]Hakyun Lee is with MedInTech, Inc., Seoul, Republic of Korea hylee@medintech.co.kr
[3]Seungjun Lee is with the School of Undergraduate Studies, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 42988, Republic of Korea lsg3486@dgist.ac.kr
[4]Michael Yip is with the Electrical and Computer Engineering Department, University of California San Diego, La Jolla, CA 92093 USA yip@ucsd.edu

Prior research has generally treated handover and pickup planning separately [10], [15]–[17], with most studies focusing on how to hand over the needle [10], [16]. Previous methods typically optimized the handover sequence assuming ideal pickup conditions, which limits applicability in real surgical scenarios where initial grasping is often suboptimal. In contrast, our work is motivated by the observation that handovers inherently arise when the target grasping cannot be achieved in a single pickup. We therefore propose an integrated approach that considers pickup and handover jointly when planning regrasping strategies.

Rule-based approaches often result in redundant handovers and inefficiencies, particularly in teleoperation where physical constraints such as joint limits are difficult to anticipate. Moreover, these methods require distinct rule sets for different initial poses, limiting scalability. To overcome these challenges, we formulate the needle regrasping problem as a reinforcement learning (RL) task, enabling an agent to minimize the number of handovers required to reach the goal grasping state from arbitrary initial poses. Similar to how human operators refine their strategies through experience, the RL agent learns optimal policies via trial-and-error interactions with the environment and reward feedback. RL is particularly suitable for this problem, as it can efficiently explore high-dimensional, combinatorial state–action spaces where traditional dynamic programming or sampling-based approaches struggle to optimize complex reward functions involving factors such as joint margins.

In this work, we design a reward function that incorporates these elements and train a DQN agent to acquire robust handover strategies. Furthermore, we address the sim-to-real gap, a persistent challenge in surgical automation due to nonlinear hysteresis and structural inaccuracies in cable-driven robots such as the da Vinci system [18]. These discrepancies are especially critical when manipulating small needles, representing a major bottleneck in automation [4], [5]. To mitigate this issue, disturbances are modeled within the RL environment, enabling the agent to acquire strategies that remain robust under real-world uncertainties. This integrated approach allows the simultaneous consideration of combinatorial handover sequences and physical robot inaccuracies, which is difficult to achieve with conventional optimization methods.

The main contributions of this work are

- To the best of our knowledge, we present the first automation framework that unifies pickup and handover planning to achieve a target grasping state.
- We formulate handover sequence optimization as a discrete RL problem that explicitly accounts for joint margin constraints and uncertainties during the handover, and train a DQN agent to minimize the number of handovers while maintaining stable and successful grasping.
- We validate the proposed policy on the da Vinci Research Kit (dVRK) [19], demonstrating human-level performance and robustness in teleoperation benchmarks.

II. Related Work

Surgical automation with robotic systems has been actively studied for decades, with particular focus on repetitive subtasks such as suturing [20], [21], knot tying [22], and tissue manipulation [23]. Among these, suturing has received the most attention, as it represents a core surgical procedure that cannot be performed without precise needle handling.

A. Suturing Automation

Learning-based control has recently gained momentum in robotics [22], [23], and this trend has extended to surgical automation. For example, the Surgical Robot Transformer, a transformer-based imitation learning model, demonstrated high generalization performance and success rates across multiple tasks including tissue lifting, needle pickup and handover, and knot tying [22]. Incremental reinforcement learning has also been proposed to enable task transfer between surgical subtasks [23]. More recently, beyond these approaches, vision–language–action models have been leveraged to achieve long-horizon suturing, encompassing needle pickup, needle throw, and knot tying [24]. While these studies have contributed significant progress toward learning-based control for suturing automation, they did not address needle handover from the perspective of long-horizon sequence optimization and were often limited to constrained conditions. As a result, prior approaches have not simultaneously considered factors such as the initial needle pose, grasping and handover states, and robot configurations.

B. Needle Manipulation

Due to the high precision required in suturing, needle manipulation strategies have also been extensively investigated [10], [15], [16]. Prior works combined stereo vision-based pose estimation with motion planning to implement automated needle pickup [15], while others integrated stereo vision and deep learning to detect unmodified needle poses and perform handover via visual servoing [16]. In particular, this approach improved needle visibility during handover and achieved a 96.7% success rate. Reinforcement learning has further been applied to formalize the bimanual regrasping problem, demonstrating high success rates and robustness in both simulation and real-world experiments [10].

Although both planning- and learning-based methods have advanced suturing and needle manipulation, needle handover itself has not been sufficiently addressed. In particular, prior approaches were not designed to jointly consider the initial needle pose, grasping and handover states, robot kinematic constraints, and pose uncertainty during the handover process. Moreover, many existing systems execute handover according to predefined procedures rather than optimizing the sequence of handover decisions [25]. Motivated by these limitations, we formulate surgical needle handover as a goal-oriented task planning problem and employ reinforcement learning to learn handover strategies that minimize handovers while achieving the optimal grasping state. Furthermore, we model pose variations occurring during handover as noise,

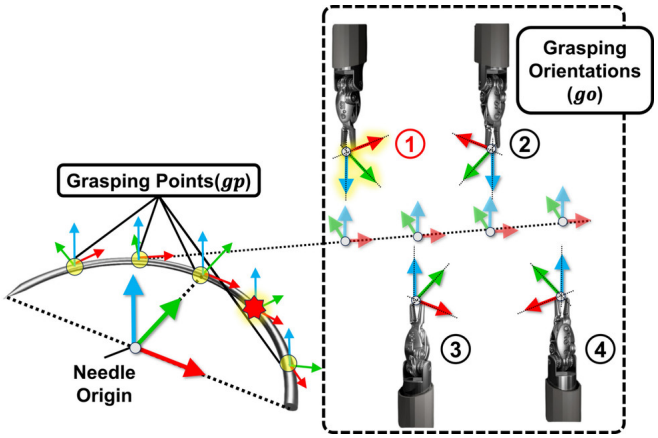

Fig. 2. **Grasping state representation.** The state is characterized by a grasping point and a grasping orientation. Each grasping point is associated with a local frame whose x-axis is aligned tangentially to the needle, while the grasping orientation specifies the relative rotation between this local frame and the robot end-effector.

enabling the learned policy to robustly handle uncertainties in real surgical environments.

III. Method

A. State and Action Space

The objective of this work is to learn a handover strategy that reaches a goal grasping pose from an arbitrary initial needle pose with a minimal number of handovers. To achieve this, we design state and action representations that explicitly capture how the needle is grasped and how handovers are executed. Specifically, four elements are defined: grasping point (gp), grasping orientation (go), grasping hand (gh), and handover orientation (ho). These elements are used to construct both the state and action spaces. The reward function incorporates joint margins (safety buffers from joint limits), enabling the policy to learn not only strategies that minimize handovers but also kinematically favorable configurations.

Table I summarizes the state and action elements and their domains. Here, gp_i^m (grasping point), go_i^m (grasping orientation), and gh_i^m (grasping hand) denote the needle grasping location, grasping angle, and the hand used, respectively. The needle pose $\mathbf{x}_n$ represents the 3D translation and quaternion orientation of the needle, while ho denotes a discrete handover orientation. Fig. 2 illustrates the grasping elements, and Fig. 3 shows examples of handover orientations and disturbances.

In this work, we consider a semicircular suture needle with a known radius r, as is widely used in robotic-assisted minimally invasive surgery (RAMIS). The definition of the needle frame is shown in Fig. 2. Grasping points (gp_i^m) are defined as discrete points obtained by uniformly dividing the arc, excluding the endpoints, yielding N_{gp} candidate points. Each point has an orientation rotated about the z-axis of the needle frame.

Grasping orientations (go_i^m) define the direction in which the needle is grasped at each gp_i^m and are grouped into two sets: the first set is obtained by rotating the gp_i^m frame around its y-axis at fixed intervals, and the second set is generated by flipping the first set 180° about the z-axis. This construction ensures uniform coverage of possible grasping directions. The grasping hand (gh_i^m) specifies which robotic arm is used. When the needle size changes, adaptation can be achieved simply by modifying the needle radius parameter during training. Together, gp, go, and gh uniquely define the robot's grasping state.

TABLE I

Elements for State and Action

Element	Symbol	Domain
Grasping elements	$gp_i^m,\ go_i^m,\ gh_i^m$	Discrete
Needle pose (n)	$\mathbf{x}_n = \{\mathbf{t}_n \in \mathbb{R}^3, \mathbf{q}_n \in \mathbb{R}^4\}$	Continuous
Handover orientation	ho_i	Discrete

$$m \in \{s, a\}, \quad i = 0, \ldots, N_{gp/go/gh/ho} - 1$$

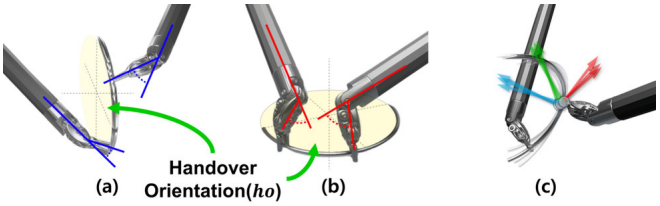

Fig. 3. **Handover orientation and disturbances.** (a, b) The action specifies the handover orientation by placing the needle on a transparent yellow plane, referred to as the handover plane. Within this plane, the needle can rotate to generate different handover orientations. (c) Disturbances introduced during training simulate uncertainties that may occur during grasping, improving the robustness of the learned policy.

The difference between state and action lies in the last element. In the action space, ho denotes the orientation of the needle during handover. At a fixed handover position, one of N_{ho} predefined orientations is selected. Each ho is defined as a discrete orientation on a given plane, where the needle is rotated within the plane and the plane itself is oriented in different directions. Fig. 3 illustrates how different ho choices can lead to distinct robot configurations even under the same grasping state, thereby affecting joint margins.

The final state element, $\mathbf{x}_n$, is modeled as a continuous variable to account for errors arising during manipulation. Needle manipulation requires high precision, but in tendon-driven robots (e.g., da Vinci systems), nonlinear hysteresis and visual occlusion can introduce deviations. To simulate these uncertainties, Gaussian rotational noise of up to $\pm 30°$ was applied around the x- and y-axes of the desired grasping frame. This disturbance is applied to ho and incorporated into $\mathbf{x}_n$, thereby enhancing the robustness of the learned policy. An example of disturbances is shown in Fig. 3.

In our experiments, we set $N_{gp} = 5$, $N_{go} = 4$, $N_{gh} = 2$, and $N_{ho} = 8$. This compact representation allows efficient definition and computation of pickup and handover scenarios, while also enabling straightforward expansion of the state–action space if needed.

B. Training Flow

The algorithm begins by observing a needle randomly placed within the workspace such that at least one arm can

reach it. The first action determines the pickup strategy, while all subsequent actions are executed at a predefined handover position. The agent receives the state $\mathbf{s} = [gp^s, go^s, gh^s, \mathbf{x}_n]$ as input and outputs the next action $\mathbf{a} = [gp^a, go^a, gh^a, ho]$. After the initial pickup, the grasping hand gh^s alternates between the two arms regardless of the chosen gh^a.

For suturing, the optimal grasping state is typically defined as grasping the needle at one-third of its arc with the gripper perpendicular to the needle axis detailed in Section 2B of the Supplement [26]. One such state is designated as the goal state (highlighted in Fig. 2). The chosen grasping state is transformed from the needle frame to the desired grasping frame and mapped to executable Cartesian coordinates.

At each state update, the handover count is incremented by one. The process repeats until the agent reaches the goal state. The objective is to minimize handover count while maximizing the cumulative reward, thereby learning an optimal handover sequence policy.

1) Reward: For each state–action pair, the end-effector poses are computed and used to evaluate the reward. The reward function is designed to encourage minimal handovers and kinematically favorable strategies. It is defined as:

$$r = \begin{cases} -5, & \text{if truncation occurs,} \\ +10 + \delta, & \text{if termination occurs,} \\ \alpha + \delta, & \text{otherwise,} \end{cases}$$

where:

- **Truncation conditions** include: (i) reaching a joint limit, (ii) exceeding the episode step limit, (iii) reselecting the same grasping point, and (iv) collision with the ground.
- **Termination condition**: reaching the goal grasping state.
- **Transition reward** α is applied otherwise; we set $\alpha = -1$.
- **Reward scaling**: the values $(-5, +10, -1)$ were heuristically chosen to balance exploration and discourage excessive handovers.
- **Joint margin term** $\delta \in [0, 1]$ encourages neutral joint configurations, thereby reducing the risk of singularities and improving dexterity. Inspired by [14], it is computed as

$$f_i(\theta_i) = \frac{4\,|\theta_i - \theta_{i,\text{mid}}|^2}{|\theta_{i,\max} - \theta_{i,\min}|^2}, \quad \theta_{i,\text{mid}} = \frac{\theta_{i,\max} + \theta_{i,\min}}{2} \tag{1}$$

and

$$\delta = 1 - \frac{1}{T} \sum_{t=0}^{T-1} \sum_{i=1}^{N} f_i(\theta_i) \tag{2}$$

where $\theta_{i,\max}$ and $\theta_{i,\min}$ are the joint limits and N is the number of joints considered.

For physical truncations (e.g., joint limits, collisions), rewards are evaluated at two points: during pickup (at grasping and pickup completion) and during handover (at preparation and transfer completion).

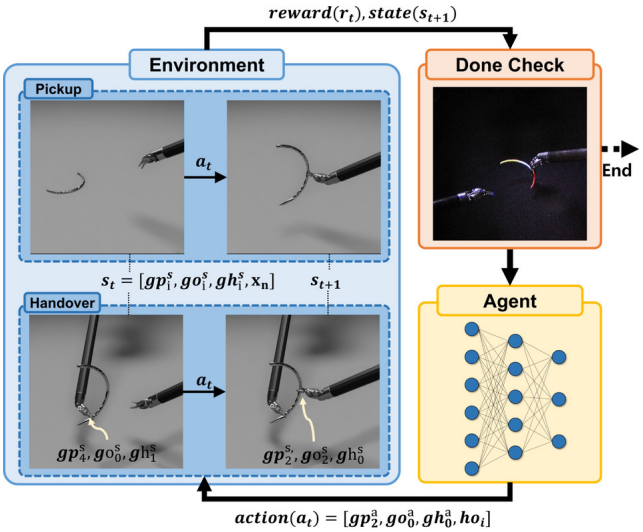

Fig. 4. **Training flow.** The agent takes the state as input, generates an action, and updates the state through its interaction with the environment. The upper box in the environment illustrates the state transition during the initial pickup at the beginning of an episode, while the lower box depicts state updates through handovers in subsequent steps. Training continues until a termination or truncation condition is met, with rewards guiding the policy learning process. Abbreviations are defined in the Method section.

When computing δ, the kinematics of the da Vinci robot were considered. The first three joints mainly contribute to positioning, while the last three dominate orientation. Thus, δ was computed only for the last two joints, excluding the wide-range roll joint ($-270°$ to $270°$).

IV. Experiment

We validated the proposed reinforcement learning (DQN) based automated handover policy on a surgical robot platform and compared it against human teleoperation. The objective of the experiment was to evaluate whether the algorithm could reliably perform the entire process from needle pickup to handover, and to quantitatively analyze differences in execution strategies.

A. Experimental Setup

Experiments were conducted on a da Vinci Research Kit (dVRK) [19] equipped with two large needle drivers. Human teleoperation was performed through the master console while observing the workspace via the endoscopic camera (ECM). For automation, needle perception was provided by an external Basler RGB camera placed opposite the ECM. Both cameras were tilted to cover the 10 cm × 10 cm workspace, and extrinsic calibration was performed to establish the spatial relationship between the cameras and the robot poses (Fig. 5).

At the beginning of each trial, the needle was randomly placed within the workspace. The algorithm observed the needle, generated the corresponding state, and selected an action. The selected pose was executed through visual servoing, allowing the robot to grasp the needle and transfer it to the predefined handover location. The goal state was defined as PSM1 grasping the needle at the specified goal point and orientation (Fig. 5).

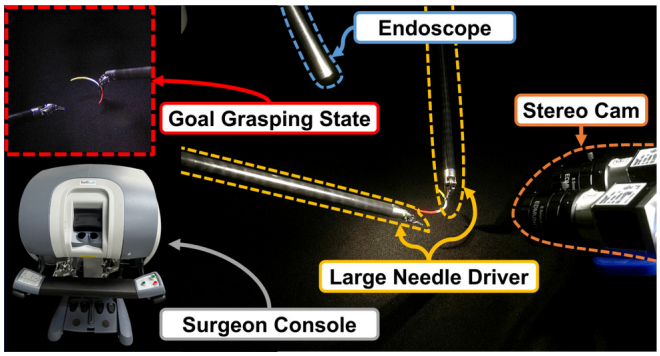

Fig. 5. **Experimental setup.** The dVRK was equipped with large needle drivers, an endoscope for teleoperation, and an external stereo camera for perception. Teleoperation was carried out via the surgeon console (bottom left), and both teleoperation and the automated policy were evaluated in achieving the goal grasping state (top left).

B. Perception and Visual Servoing

Accurate perception and control are essential for reliable handovers, especially given the cable-driven hysteresis of the dVRK. To mitigate perception uncertainty, visual servoing was employed.

Needle Pose Estimation: We extended the framework of HOUSTON [16] by combining SAM2-based [27] segmentation with color-based segmentation for robust perception under occlusions and illumination changes. Needle contours were extracted from stereo RGB images, followed by ellipse fitting and point sampling. Candidate points were matched using the Hungarian algorithm, and 3D triangulation was performed. Outliers were removed by RANSAC plane fitting (inlier radius 3 mm, 300 iterations). Keypoints (needle endpoints, color crossing points) were then used to compute the needle frame.

Tool Pose Estimation: Tool pose was estimated with a particle filter [9], modeling the base-to-camera transformation as a lumped error. By minimizing the difference between kinematic projections and observed projections, estimation errors were reduced.

Visual Servoing: Motions toward the grasp pose suggested by the policy were executed using waypoint-based visual servoing. An intermediate waypoint was defined 2 cm behind the target pose along the approach direction of the end-effector. Based on the estimated needle and tool poses, pose refinement was applied before the final approach. The trigger was generated manually by the operator, after which the end-effector translated 2 cm along the negative z-axis to complete the grasp.

C. Real-World Experiment

The proposed DQN framework was implemented with a three-layer feed-forward network. The input dimension was 10, followed by hidden layers of size 64 and 128, and an output layer of 320 discrete actions. Training was performed for 100,000 episodes with a replay buffer of 100,000 transitions, a learning rate of 1×10^{-4}, a discount factor $\gamma = 0.98$, and a batch size of 64. An ϵ-greedy strategy was adopted, where ϵ decayed linearly from 1.0 to 0.01 with $\Delta\epsilon = 2.86 \times 10^{-5}$ and remained fixed at 0.01 thereafter. The target network was synchronized with the Q-network every 1,000 episodes. Training was conducted on a workstation with an NVIDIA GeForce RTX 4070Ti Super GPU and 32 GB RAM, running Windows with PyTorch 2.4.1, CUDA 11.8, and Python 3.10.

For evaluation, the RL policy was executed in 20 trials with randomized initial needle poses. For teleoperation, four human operators each performed 20 trials (80 in total). All operators had 5–10 hours of prior teleoperation experience and were informed of the same goal grasp state and orientation as the RL policy. To ensure a fair comparison, teleoperation was restricted to the same discrete action space as the DQN algorithm. Specifically, operators were instructed to select from five predefined grasping points, grasp the needle in a near-vertical orientation, and perform handovers with orientations consistent with those used by the automated policy. This setup minimized variability and enabled a direct comparison between human and algorithmic strategies.

Performance was evaluated using four metrics: (1) number of handovers, (2) success rate, (3) number of joint limit occurrences, and (4) joint margin. These metrics allowed a quantitative comparison of the stability and consistency of the RL-based policy against human performance. To our knowledge, this is the first quantitative evaluation covering the entire process from pickup to inter-arm handover, rather than isolated transfer actions.

V. Results

Table II summarizes the overall performance under both conditions. In terms of success rate, teleoperation achieved 100% (80/80), while the DQN policy reached 85% (17/20). Although the learned policy did not achieve perfect success, its performance remained comparable to human teleoperation, demonstrating the feasibility of the proposed approach. Among the three failures observed in the DQN trials, two were caused by perception-related issues (needle detection failure and visual servoing failure), rather than the learned policy itself. The remaining failure occurred when both arms simultaneously grasped the needle in a crossing configuration, leading to a collision. These cases indicate that the evaluation reflects realistic errors that can arise in a fully autonomous system, rather than failures solely attributable to the decision-making logic of the policy.

Fig. 6 illustrates the training results of the proposed DQN framework. The network converged stably, and the learned policy achieved the goal grasping state with an average of one handover attempt. The average inference time was

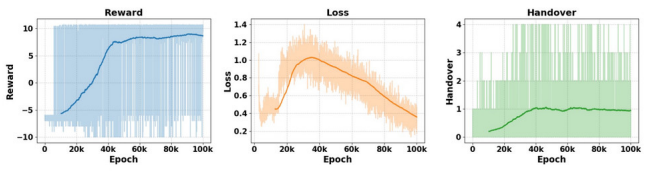

Fig. 6. **Training curves.** Rewards, losses, and the number of handovers are plotted across training epochs.

TABLE II

COMPARISON BETWEEN DQN AND TELEOPERATION IN HANDOVER PERFORMANCE AND JOINT METRICS

Method	Handover Count	Success Rate	Joint Limit Count	q1	q2	q3	q4	q5	q6
DQN Algorithm	1.65 ± 0.50	85% (17/20)	1.10 ± 0.30	0.64 ± 0.08	0.93 ± 0.04	0.50 ± 0.07	0.92 ± 0.05	0.79 ± 0.06	0.80 ± 0.05
Teleoperation	1.62 ± 0.55	100% (80/80)	1.67 ± 0.40	0.67 ± 0.07	0.99 ± 0.03	0.29 ± 0.09	0.86 ± 0.06	0.85 ± 0.05	0.75 ± 0.07

* Values are shown as mean $\pm$ standard deviation.

0.187 ms averaged over 1,000 trials, demonstrating the real-time feasibility of the method. This indicates that the learned policy can be practically deployed in surgical scenarios where low-latency responses are critical.

In terms of the number of handovers, DQN achieved an average of 1.65 ± 0.50, while teleoperation achieved 1.62 ± 0.55, showing comparable performance. Notably, DQN exhibited lower variance, indicating more consistent behavior across trials. As shown in Fig. 7, human operators exhibited greater variability in both average handovers and variance across individuals, whereas DQN reproduced stable performance at a level similar to that of human teleoperation. This demonstrates that the proposed algorithm not only achieves a competitive success rate but also provides meaningful improvements in execution consistency.

For joint limit analysis, a joint was counted as reaching a limit when it operated within 5% of its range boundary. On average, DQN exhibited 1.10 ± 0.30 occurrences per trial, compared to 1.67 ± 0.40 for teleoperation. This result highlights that the DQN policy more effectively avoided extreme joint configurations and maintained safer mid-range postures compared to human operators. This suggests that the DQN policy inherently favors safer motion strategies, reducing the likelihood of extreme joint usage that could compromise system reliability.

For joint margin analysis (Table II), the DQN policy exhibited overall stable trends, although differences were observed across individual joints. The joint margin ranges from [0,1], with values closer to 1 indicating more neutral joint configurations. In this study, the reward function was designed to emphasize $q5$ and $q6$, which play a primary role in handover execution. As a result, DQN achieved larger margins than teleoperation for $q6$, reflecting improved safety in joint usage, whereas lower margins were observed for $q5$. Nonetheless, this limitation may be mitigated in future work by expanding the action space or incorporating adaptive orientation strategies.

This difference can be attributed to the discrete nature of the DQN action space. Since the algorithm is restricted to a predefined set of grasping and handover strategies, it cannot easily avoid unfavorable configurations at certain joints. In contrast, human operators, despite being instructed to maintain the needle in a near-vertical orientation, acted in a continuous space with greater flexibility, allowing them to secure additional margins when necessary. Thus, the reduced performance in $q5$ reflects a structural limitation of the discrete action space. A per-joint comparison of margins is visualized in Fig. 8.

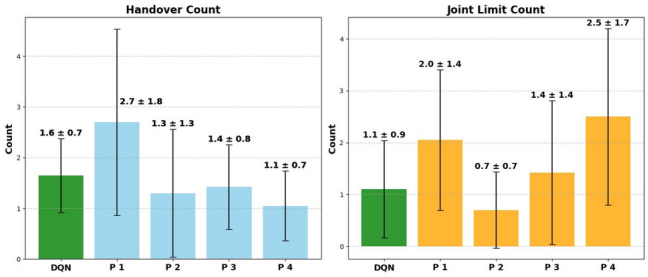

Fig. 7. **Handover count and joint limit count.** Comparison between the proposed DQN-based policy and human subjects in terms of the number of handovers and joint limit occurrences.

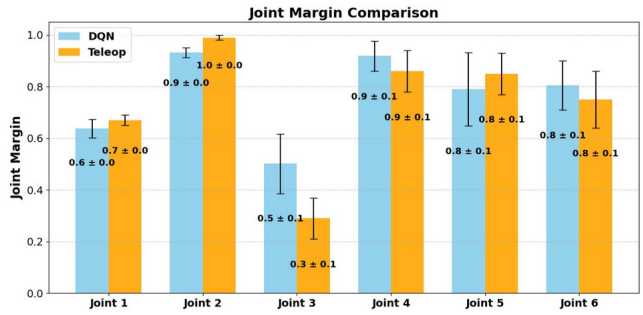

Fig. 8. **Joint margin comparison.** Joint margin scores for each joint are compared between the DQN-based policy and teleoperation.

This work presents the first integrated experimental validation of reinforcement learning-based automation covering the entire process from needle pickup to handover. The proposed DQN policy reliably executed the complete sequence, achieving efficiency and dexterity comparable to teleoperation while producing more conservative and reproducible motions. Failure analysis indicated that most errors arose from perception and low-level control modules rather than the learned policy, suggesting that performance could be further improved with enhanced sensing and control infrastructure.

VI. CONCLUSION & FUTURE WORK

This study introduced a reinforcement learning-based policy for automating surgical needle handover and, for the first time, directly compared it with human teleoperation across the full pickup-to-handover sequence. The results demonstrated that the learned policy can achieve human-comparable reliability while improving consistency and joint safety, underscoring the potential of RL for safe and reproducible surgical assistance.

Future work will address several limitations identified in this study. First, we plan to incorporate visibility-aware ser-

voing strategies to mitigate perception-related failures during handover. Second, explicit penalty terms will be introduced to discourage collision-prone cross-grasp configurations. In addition, although disturbance has currently been modeled only in terms of rotational uncertainty, translational disturbances should also be considered. To better reflect these limitations and the constraints of the current representation, we will extend the state and action definitions to a continuous space. Furthermore, while the present system relies on painted needles for reliable tracking, future work will incorporate unmodified surgical needles to better approximate realistic clinical environments. These extensions will support the development of more robust policies and provide a foundation for advancing surgical automation toward complex multi-step tasks.

ACKNOWLEDGMENT

This work was supported by the Industrial Strategic Technology Development Program (ISTDP) (RS-2024-00443054) funded by the Ministry of Trade, Industry and Energy (MOTIE, Republic of Korea); by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2025-25420118); and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT, MSIT) (RS-2025-22862972).

REFERENCES

[1] M. Yip and N. Das, "Robot autonomy for surgery," in *The Encyclopedia of Medical Robotics*, 2017, pp. 281–313.

[2] 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)*, 2016, pp. 4178–4185.

[3] 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," in *2024 International Symposium on Medical Robotics (ISMR)*, 2024, pp. 1–7.

[4] M. Hwang, J. Ichnowski, B. Thananjeyan, D. Seita, S. Paradis, D. Fer, T. Low, and K. Goldberg, "Automating surgical peg transfer: Calibration with deep learning can exceed speed, accuracy, and consistency of humans," *IEEE Transactions on Automation Science and Engineering*, vol. 20, no. 2, pp. 909–922, 2023.

[5] S. Paradis, M. Hwang, B. Thananjeyan, J. Ichnowski, D. Seita, D. Fer, T. Low, J. E. Gonzalez, and K. Goldberg, "Intermittent visual servoing: Efficiently learning policies robust to instrument changes for high-precision surgical manipulation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 7166–7173.

[6] D.-L. Chow and W. Newman, "Improved knot-tying methods for autonomous robot surgery," in *2013 IEEE International Conference on Automation Science and Engineering (CASE)*, 2013, pp. 461–465.

[7] 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)*, 2023, pp. 4761–4767.

[8] 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)*, 2022, pp. 5286–5292.

[9] 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, 2022.

[10] 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)*, 2021, pp. 7737–7743.

[11] B. Lu, H. K. Chu, K. Huang, and J. Lai, "Surgical suture thread detection and 3-d reconstruction using a model-free approach in a calibrated stereo visual system," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 2, pp. 792–803, 2020.

[12] J. Lu, A. Jayakumari, F. Richter, Y. Li, and M. C. Yip, "Super deep: A surgical perception framework for robotic tissue manipulation using deep learning for feature extraction," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 4783–4789.

[13] G. A. Fontanelli, M. Selvaggio, L. R. Buonocore, F. Ficuciello, L. Villani, and B. Siciliano, "A new laparoscopic tool with in-hand rolling capabilities for needle reorientation," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2354–2361, 2018.

[14] T. Liu and M. C. Cavusoglu, "Needle grasp and entry port selection for automatic execution of suturing tasks in robotic minimally invasive surgery," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 552–563, 2016.

[15] 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)*, 2018, pp. 1370–1377.

[16] 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)*, 2022, pp. 9637–9643.

[17] S. Lu, T. Shkurti, and M. C. Çavuşoğlu, "Dual-arm needle manipulation with the da vinci® surgical robot," in *2020 International Symposium on Medical Robotics (ISMR)*. IEEE, 2020, pp. 43–49.

[18] M. Hwang, B. Thananjeyan, S. Paradis, D. Seita, J. Ichnowski, D. Fer, T. Low, and K. Goldberg, "Efficiently calibrating cable-driven surgical robots with rgbd fiducial sensing and recurrent neural networks," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5937–5944, 2020.

[19] 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)*, 2014, pp. 6434–6439.

[20] 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)*, 2021, pp. 242–248.

[21] K. L. Schwaner, D. Dall'Alba, P. T. Jensen, P. Fiorini, and T. R. Savarimuthu, "Autonomous needle manipulation for robotic surgical suturing based on skills learned from demonstration," in *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 2021, pp. 235–241.

[22] 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," in *8th Annual Conference on Robot Learning*, 2024. [Online]. Available: https://openreview.net/forum?id=fNBbEgcfwO

[23] Y.-J. Ho, Z.-Y. Chiu, Y. Zhi, and M. C. Yip, "Surgirl: Toward lifelong learning for surgical automation by incremental reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 10, no. 12, pp. 13 145–13 152, 2025.

[24] J. Haworth, J.-T. Chen, N. Nelson, J. W. Kim, M. Moghani, C. Finn, and A. Krieger, "Suturebot: A precision framework benchmark for autonomous end-to-end suturing," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[25] K. Hari, Z. Chen, H. Kim, and K. Goldberg, "Stitch 2.0: Extending augmented suturing with ekf needle estimation and thread management," *IEEE Robotics and Automation Letters*, vol. 10, pp. 12 700–12 707, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:282459020

[26] W. J. Nelson, "Guide to suturing," *Journal of Oral and Maxillofacial Surgery*, vol. 73, no. 8 Suppl, pp. 1–62, Aug. 2015.

[27] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollar, and C. Feichtenhofer, "SAM 2: Segment anything in images and videos," in *The Thirteenth International Conference on Learning Representations*, 2025. [Online]. Available: https://openreview.net/forum?id=Ha6RTeWMd0