IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Towards Human-Robot Collaborative Surgery: Trajectory and Strategy Learning in Bimanual Peg Transfer

Zhaoyang Jacopo Hu[1], Ziwei Wang[2] , Yanpei Huang[2] , Aran Sena[2] , Ferdinando Rodriguez y Baena[1], and Etienne Burdet[2]

***Abstract*—While the traditional control of surgical robots relies on fully manual teleoperations, human-robot collaborative systems promise to address issues such as workspace constrains and laborious tasks. In particular, shared control can reduce the surgeon's workload and improve the overall surgical performance by supporting the surgeon effort during movements while keeping them in charge of complex control phases. In this letter, we propose a segmentation of the bimanual peg transfer task that alternates manual and autonomous control correspondingly. The authority allocation in this shared control framework considers both the limitation of learning-based methods and the higher dexterity of humans during physical interaction. The motion and strategies are transferred from an expert human to a da Vinci Research Kit (dVRK) using an epsilon-greedy on a maximum entropy inverse reinforcement learning algorithm. The model generated enables to train an intelligent agent that can skillfully collaborate with the human operator during the surgical task. The proposed shared control framework is verified both on a virtual platform and then on a real dVRK to assess its usability and robustness. The results show that, compared to traditional teleoperation, our method can achieve faster and more consistent peg transfers. An analysis of the participants' effort also reveals a significantly lower perception of the workload.**

***Index Terms*—Human-Robot Interaction, Autonomous Agents, Shared Control, Imitation Learning, Medical Robots.**

I. Introduction

MEDICAL robots such as the da Vinci Surgical System [1], [2] offer surgeons increased dexterity and visual feedback during minimally invasive surgery. The study in [3] showed that, compared to traditional manual laparoscopic surgery, the da Vinci allows operators to perform quicker and more efficiently standardised exercises, making complex surgical procedures more accessible for people without surgical experience. However, the operator's performance can decrease over time when carrying out long surgical operations, which can lead to errors [4], [5]. This is exacerbated by repetitive or intensive actions during the procedure. For example, the frequent need to clutch (i.e. pressing the pedal to reposition the hand controllers in an ergonomic position without affecting the robot manipulators) is generally regarded as one source leading to longer completion times which can affect the surgical workflow and operator's cognitive load [6]. Recent work with fully automated agents [7] has also directed the attention towards the possibility of using automation to improve surgical performance. These factors have motivated research in human-robot interaction to allow the surgeon to collaborate with an autonomous agent during the completion of a task [4]. In [8], different human-robot collaboration models were presented: supervised, shared, and full autonomy. The study reveals that a shared control, where the human and the robot sequentially alternate their control authority on the manipulators, leads to results with higher specificity and minimal effort to the participants. The works on shared control in surgery [9]–[11] corroborated the idea of a collaborative shared framework between surgeon and autonomous robot. However, while the segmentation of a shared control has been widely explored, the interchange of the control of the surgical tool between operator and autonomous agent still needs an effective adaptive component to ensure a natural relationship that does not interrupt the surgical workflow [12]. The agent needs to fit in the surgeon's technique and strategy for completing a task with intuitive movements to the human operator and smooth transitions between the operator and agent [13]. For example

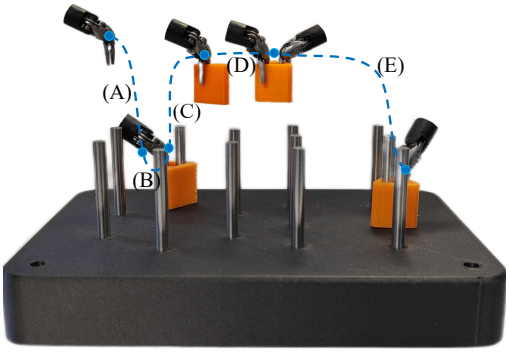

Fig. 1. Segmentation of the bimanual peg transfer task using shared control. The segments A and C are controlled by the autonomous agent. B and D are carried out by the human operator and E by the assistant.

Manuscript received: January, 2023; Revised: April, 2023; Accepted: May 31, 2023. This work was supported by the EU NIMA (FETOPEN 899626) grants, EPSRC and Intuitive Surgical in the form of an industrial CASE studentship. *(Corresponding authors: Zhaoyang Jacopo Hu, Ziwei Wang, Etienne Burdet)*

[1]Zhaoyang Jacopo Hu and Ferdinando Rodriguez y Baena are with the Department of Mechanical Engineering, Imperial College London, SW7 2AZ, (e-mail: jacopo.hu20@imperial.ac.uk).

[2]Ziwei Wang, Yanpei Huang, Aran Sena, and Etienne Burdet are with the Department of Bioengineering, Imperial College London, W12 0BZ, United Kingdom. Ziwei Wang is also with the School of Engineering, Lancaster University, Lancaster LA1 4YW (e-mail: ziwei.wang@ieee.org, e.burdet@imperial.ac.uk).

Digital Object Identifier 10.1109/LRA.2023.3285478.

in the past, methods to automate surgical tasks have been investigated through waypoints strategy [14], image-guided algorithms [15], and learning from demonstration [16]. With these methods the agent learns the trajectory necessary to complete a task, but not the strategy that ensures the natural surgical workflow and ergonomics of the surgeon. Furthermore, it is still problematic to correctly define all the different costs associated with the elements of the surgical environment. Maximum entropy inverse reinforcement learning (MaxEnt IRL), a method used to obtain the utility function of the environment, had previously been applied in [10], [17]–[19]. In [10], the authors implement MaxEnt IRL to generate the trajectory of a suturing motion, however without proposing a method to increase the robustness of the policy generated or analysing its effectiveness with human operators.

In this letter, we develop a collaborative control framework of the peg transfer task with task segmentation and sequential human-robot collaboration scheme. Using this framework, we implement an epsilon-greedy in Maximum Entropy Inverse Reinforcement Learning (EG MaxEnt IRL) to learn both the trajectories and strategies from an expert user, maximising the generalisation of the trajectories with the minimum number of expert demonstrations.

We then perform a user study with ten participants to evaluate the benefits of our method by comparing the bimanual peg transfer task completed by $manual\ teleoperation$ (MT) against $shared\ control$ (SC). To validate and analyse our proposed setup design, learning algorithm and shared control framework, we implement the system to perform a bimanual peg transfer task, which is a fundamental skill of laparoscopic surgery [7], [20]. Several virtual platforms with standardised tasks for robotic surgery are available to execute the testing and validation of an agent, however simulators of the dVRK are designed either to simulate the actual surgical experience [21], [22] or to develop fully autonomous agents for surgical procedures [23]. Therefore, in this letter we modify the SurRoL environment [23] to support a real-time simulator, which only requires Python. This modification will enable the development of a framework that integrates interfaces to control and interact with the agent in the same environment in which the autonomous programme is trained.

The letter is organised as follows: Section II describes the proposed segmentation of the bimanual peg transfer and the framework used for shared control. Section III describes adaptation of the simulated environment and the training algorithm of the agent. Section IV details the experimental procedure on the simulated and real robotic platform, including protocols and evaluation methods. Section V analyses the results, and Section VI discusses them.

II. Working Principle

A. Shared Control during Peg Transfer Task

Previous works [9], [10] have shown that the surgical motion can be effectively segmented to accommodate a shared control between human operator and agent. To ensure safety and take advantage of the robot dexterity, the segmentation of a task should account for surgical motions that require higher attention of the human operator, such as when in proximity of objects and physical interactions, in contrast to repetitive and laborious spatial relocation of the surgical tools. Physical interaction are particularly complicated as they require the use of an adaptive control to solve the instability of the system. Furthermore, physical interactions require to account for a larger domain of physical dimensions which can be challenging for reinforcement learning training algorithms. Therefore, to complete the peg transfer task, the control authority of the arm controlled by the participant is divided as illustrated in Fig. 1, where each subtask is appropriately allocated to either the human or the agent skills and capacity. The proposed process begins by triggering the trained agent to autonomously position the surgical hand in proximity of the target block (A). Once the surgical tool is in a strategic position (i.e. the block can be successfully grabbed if approached from above), which is learned during training from expert demonstrations, the human operator is in control again and can continue the movement and easily grab the block (B). The task is followed by the autonomous motion of the agent that brings the block closer to the other arm (C). The operator then performs a hand-off by physically interacting with the human assistant (D) who places it into the final peg position (E). This segmentation of the bimanual task is controlled by the human operator through a pedal and it allows to preserve the natural coordination between the two arms during hand-off while facilitating the control and workload of the human operator. In this work, we intend to represent the motions that a surgeon would perform alongside an assistant to clear the surgical space when removing debris or objects.

B. Control Framework

The proposed control framework for training and human-robot interaction (HRI) testing of the agent is shown in Fig. 2. The demonstration of the task is provided by an expert user. In this work, an operator is considered expert if they have more than 300 hours of experience with the hand controllers.

To train the agent, the SurRoL environment is modified to integrate haptic controller interfaces (sigma.7 and omega.7 Force Dimension). The setup developed in Fig. 2 allows to train the agent in the SurRoL platform and then subsequently quickly test it with a human assistant in the same environment. This allows to rapidly train and test the agent and it permits to assess the model generated without alteration caused by the transition from sim-to-sim before implementing sim-to-real [24]. The features of the expert demonstrations are extracted and matched against the motions of the agent in the simulation using a gradient-based optimisation, which then updates the reward values of the environment. Once the training is completed, the model of the expert can be tested in either the same surgical simulator or the physical dVRK platform with a human operator by using a shared control framework.

III. Autonomous Agent

A. EG MaxEnt IRL Algorithm

The EG MaxEnt IRL allows to retrieve the reward function of an unknown environment through expert demonstrations.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

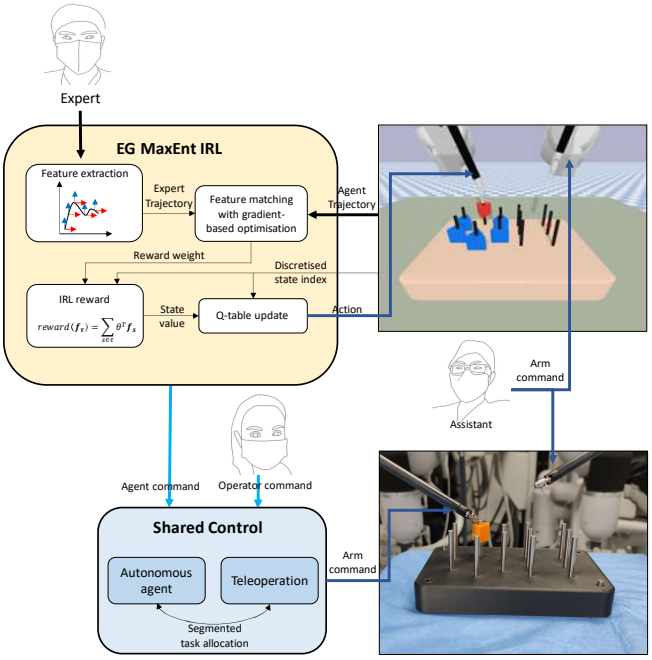

Fig. 2. Process diagram for the training and testing of the bimanual peg transfer using shared control.

The expert trajectory used in our model is recorded in the same SurRoL environment that will be used for training. The proposed algorithm uses Q-Learning to populate the Q-table with the expected rewards for a specific action at a certain state. To encourage exploration for a more robust policy, we use an epsilon-greedy method [25]. With this modification, the agent will randomly choose an action with ϵ probability instead of the action with maximum reward. Notice that epsilon-greedy is normally used as a tool to find a more optimal solution. However, the proposed EG MaxEnt IRL uses it to update the reward function on a larger state space. This allows the agent to generalise more the trajectory to the goal, making it more robust to states that have not been visited by the expert.

Using EG MaxEnt IRL, we train the agent to generate a reward function of the environment from one expert trajectory. During the training episode, the reward function is calculated according to the MaxEnt reward equation [17]:

$$\text{reward}(\boldsymbol{f_\tau}) = \sum_{\text{s}\in\tau} \theta^{\text{T}} \boldsymbol{f_s} \qquad \theta \leftarrow \theta + \beta \nabla L(\theta) \tag{1}$$

where $\boldsymbol{f_\tau}$ is the feature count of the trajectory τ, $\boldsymbol{f_s}$ is the state feature count, and θ is the weight vector that is optimised at the end of each training episode using the Lagrangian gradient with β learning rate. The value function is then used to populate the Q-table and the policy $\pi(s,a)$ is generated using the Bellman equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \text{max}_{a'} \left[Q\left(s', a'\right) \right] - Q(s,a) \right] \tag{2}$$

where s is the current state, a is the current action, r is the reward value calculated from (1), α is the learning rate, γ is the discount factor, and s' and a' is the next state and action. The

Algorithm 1 EG MaxEnt IRL

Import expert trajectories dataset $D = [\tau_1, \ldots, \tau_T]$ and feature vectors $f = [f_1, \ldots, f_T]^T$
Initialize random reward weights $\theta = [\theta_1, \ldots, \theta_S]$
function AGENT TRAINING
  for `1 to N` **do**
    while *Episode not done* **do**
      if $n_{rand} < \epsilon$ **then**
        $a \leftarrow random\ action$
      else
        $a \leftarrow max[Q(s,a)]$
      $\pi(a \mid s) \leftarrow Q\text{-}Learning$
      $\boldsymbol{f_\tau} \leftarrow \sum_{s\in\tau} \boldsymbol{f_s}$
    $\nabla L(\theta) \leftarrow \tilde{\boldsymbol{f}} - \sum_{s\in S} D_s \boldsymbol{f_s}$
    $\theta \leftarrow \theta + \beta \nabla L(\theta)$
  return EG MaxEnt IRL model
while Simulation not done **do**
  EG MaxEnt IRL model$\leftarrow state$
  Update Simulation $\leftarrow action$

TABLE I
EG MAXENT IRL HYPERPARAMETERS.

Hyperparameter	Value
Weight θ learning rate	0.3
Input dimensions	x, y, z
Bellman equation learning rate	0.3
Total number of states	10,648
γ Discount factor	0.99
Action step size	$2.5mm$
ϵ Epsilon-Greedy	0-0.1

algorithm implementation is described in Algorithm 1 using the hyperparameters in Table I.

B. Agent Training and HRI Environment

The SurRoL environment was modified to enable both training and HRI in an unified platform. The models in the simulation consist of the dVRK robotic arms and the peg board with the blocks. Since the EG MaxEnt IRL algorithm requires state dimensions in discretised form, the state space of the SurRoL workspace is divided in a 22×22×22 grid with 2.5 mm sensitivity in the x, y and z dimensions. The action space is simplified to movements along the three Cartesian dimensions, discretised to a 2.5 mm step per action. Notice that the agent step length coincides with the sensitivity, which allows us to assume a deterministic transition probability of the movements. With this approximation, the agent is represented by the dVRK end-effector which moves between the 2.5×2.5×2.5 mm states. Since the action space for the expert demonstrator is continuous, we discretise its movements in the simulator. This ensures to both preserve the trajectory features and enable its implementation in SurRoL. It is worth noting that algorithms like the MaxEnt IRL method suffer from the curse of dimensionality. This is because the state of the agent along the trajectory in the surgical simulation is uniquely specified by flattening the 3-dimensional coordinates into a 1-dimensional index-array. This process implies that the

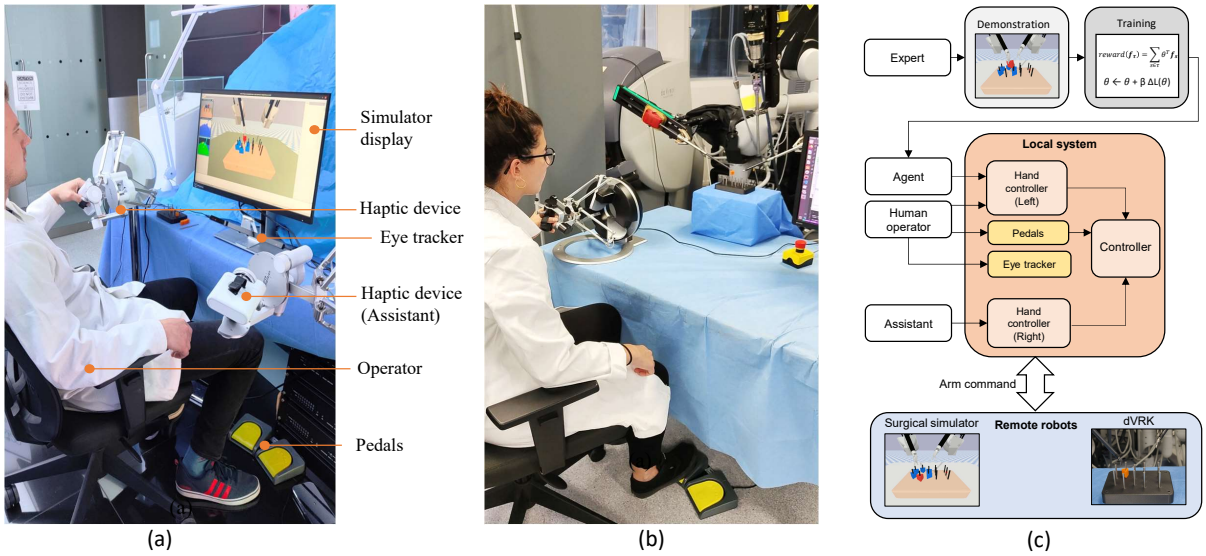

Fig. 3. Experimental setup using surgical simulation and physical dVRK platform. (a) Participant performing the task on the same surgical simulation environment on which the autonomous agent was trained. (b) Users with experience on the surgical simulator validate the controllers on the physical dVRK platform. (c) Bimanual framework for the unified real-time simulator and RL environment in a local-remote setup.

numbers of possible states in which the robot can be increases exponentially according to $(gridsize)^3$ [10]. Therefore, the number of states created in the agent training is 10,648.

IV. Experiment

Two experiments were conducted to verify the proposed method with both the surgical simulator and the dVRK surgical robot. Ten subjects without motor impairment (6 males, mean age = 26±5 years, all right-handed and without previous surgical experience) participated in the study. All participants were informed about the experiment's purpose and protocol, and signed a consent form before starting. Ten subjects joined the simulation experiment and eight of them (4 males, mean age = 25±2 years) joined the physical robot test on the dVRK platform. The experiments were approved by the Research Ethics committee of Imperial College London (21IC7042).

A. Experiment Setup

1) Simulation Platform: The first experiment was conducted with the surgical simulation platform, which is the same environment on which the agent was trained. The teleoperation system consists of local console and remote virtual laparoscopic instruments. The local console includes two hand controllers (sigma.7 Force Dimension [13]) and two foot pedals. To ensure safety, we monitor operator presence by only relinquishing control authority to the agent while the left pedal is being pressed – with control released back to the user when the pedal is released. The right pedal is used as clutch foot to uncouple the controller from the robotic arms to allow repositioning and better ergonomics [2]. Additionally, we add a constraint to Algorithm 1 to ensure that the agent movements are in the safe state space that was learned from the expert trajectory. The local console communicates with the remote virtual robot via TCP/IP to simulate the teleoperation delay between them. The hand controller can remotely control the virtual robot in seven DoF (position, orientation and grasping). The simulator monitor is equipped with Tobii Pro Fusion [26] to track the participant's eyes and observe eye related features which will be used to calculate concentration and stress.

2) dVRK Platform: The experiment was conducted with the dVRK robotic platform as illustrated in Fig. 3(b). Five participants performed the task with direct vision of the peg board, while three additional subjects were seated at the 3D dVRK interface console and tasked to perform the experiment with the 3D visual image coming from the endoscope. The experiment conducted with these two visual systems show that the condition of direct vision, which replicates and validates the setup used with the simulation platform, does not hinder the users during the task. We then formed a teleoperation with the hand controllers (omega.7 Force Dimension [13]) as local console and patient side manipulators of the dVRK system as the remote robots, without using the dVRK master tool manipulators. The hand controller could move the end-effector of the Endo-wrist in 6 DoFs motion and grasping. The pedal interface is still used to allocate agent authority and activate the clutching function. The control of the robot works with the same modes and safety functions as in the virtual platform.

B. Task and Control Modes

The participants were seated at about 1 m distance from the peg board (about 0.5 m eyeline distance) and they could easily reach the hand controller and the pedal. During the peg transfer task, the handover requires the block to reach the robotic arm controlled by the assistant. The surgical scenario that we intend to represent is the manoeuvre that the surgeon performs to pass objects to an assistant to remove and clear the space of debris or gauzes. Two control modes were used in the experiments. In the $MT\ mode$, the participants use their left hand to control the remote virtual/physical left-side surgical instrument fully

by themselves. In the *SC mode*, the intelligent agent performs subtasks A and C (detailed in Section II) and the human operator cooperates with the agent to perform subtasks B and D. In both control modes, the right-side remote robot is controlled by an expert assistant.

C. Experimental Protocol

1) Surgical Simulation: The participants were instructed on how to use the interfaces and how to complete the bimanual peg transfer task by controlling the left arm of the robot using both types of control. The experiment consisted of a total of 6 blocks. First, 2 blocks (MT and SC) of 5 minutes each to allow the participants to get familiar with the task and the interface. This is followed by 4 blocks made of 2 consecutive MT blocks and 2 consecutive SC blocks of 5 minutes each and the participants were instructed to complete as many peg transfers as possible. The order between MT and SC was randomised to disregard the learning during the experiment.

2) dVRK Platform: The experiment asked 8 participants that had previous experience from using the simulated platform to use again the two types of controllers in different order. However, unlike the surgical simulator, the real robot cannot be rapidly reset to its original position. Therefore, the experiment metrics were limited to measuring a total of 10 successful peg transfers for each control mode.

D. User Performance Evaluation

While conducting the experiments with each participant, the success rate and temporal performance of only the user controlled arm were measured. The evaluation of the participants' subjective workload was performed using a custom NASA TLX [27] to score the perceived Control, Mental, Temporal, Physical effort. The participants' mental load were also analytically calculated by processing the gaze data obtained from the Tobii Pro Fusion eye tracker.

V. Results

A. Peg Transfer Trajectory

A comparison with the previous method applied in [10], which can be obtained by setting the epsilon-greedy variable to zero, determined that the addition of the epsilon-greedy component in the algorithm proposed allows to update a larger volume of the state space. The counting of the fully explored states (i.e. the reward values associated with all possible actions in that state have been updated) shows that the use of epsilon-greedy increases the fully explored states by 45% while it decreases the fully unexplored by 34%. The increased exploration of the states enables the agent to complete the task with initial starting states that are different from those provided by the expert. We provide evidence of this in Fig. 4(a), which illustrates the agent reaching movement towards the target using deviated initial states. We then perform user demonstrations of the entire human-robot collaborative bimanual peg transfer using MT and SC to validate control framework developed. Typical trajectories of a trained user carried out at the end of the experiment are shown in Fig. 4(b).

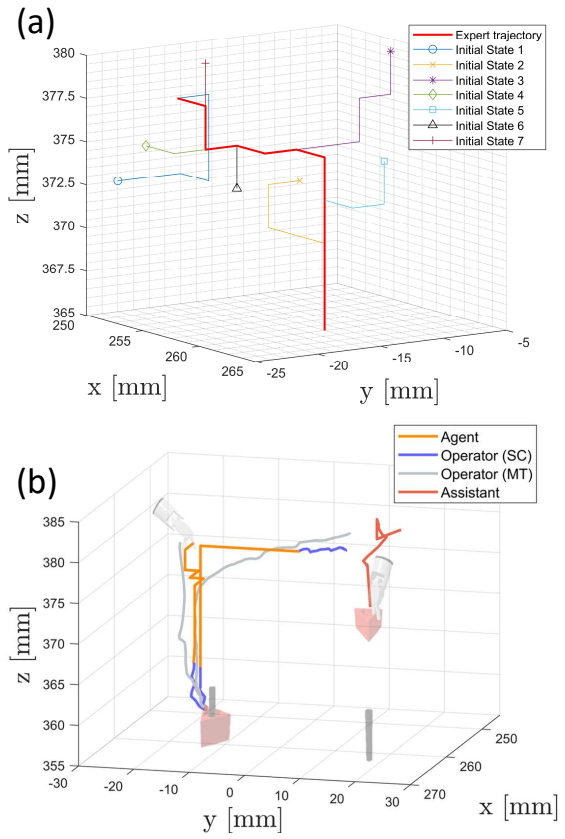

Fig. 4. (a) Agent trajectory with different starting states during target approach. (b) Bimanual peg transfer user trajectories achieved after the participants became familiar with the task.

B. Performance Comparison between MT and SC

The successful peg transfer results of MT and SC are shown in Fig. 5(a) and were analysed using statistical comparison tests [28] that determine significant differences between the two controllers. A peg transfer is defined as successful if the block is not dropped after the participant attempts to lift it. The data obtained was normally distributed, as assessed by Shapiro-Wilk's test [28] (all $p > 0.2$). A paired t-test [28] was conducted to determine the effect of different control schemes. When using MT mode, the participants could perform an average of 2 successful transfers within 5 minutes; while when robotic assistance is added, SC mode achieved about 6 successful transfers in the same time period, which is significantly higher than the MT mode ($p < 0.001$). As shown in Fig. 5(b), the average completion time has the similar tendency, with the average completion time decreasing from using MT (134.9 ± 89.9 s) to SC (37.3 ± 5.69 s), (Wilcoxon Signed-Rank test, $p = 0.005$). Fig. 5(c) illustrates the average responses and the standard deviation of the ten participants to the questionnaire. The rating results of mental effort and physical effort were normally distributed for both control modes ($p > 0.06$) and a paired t-test was used for normally distributed data and the Wilcoxon Signed-Rank test [28] for non-normally distributed data. It is interesting to observe that participants felt they had more *control* on the virtual environment when cooperating with intelligent agent rather than when they had

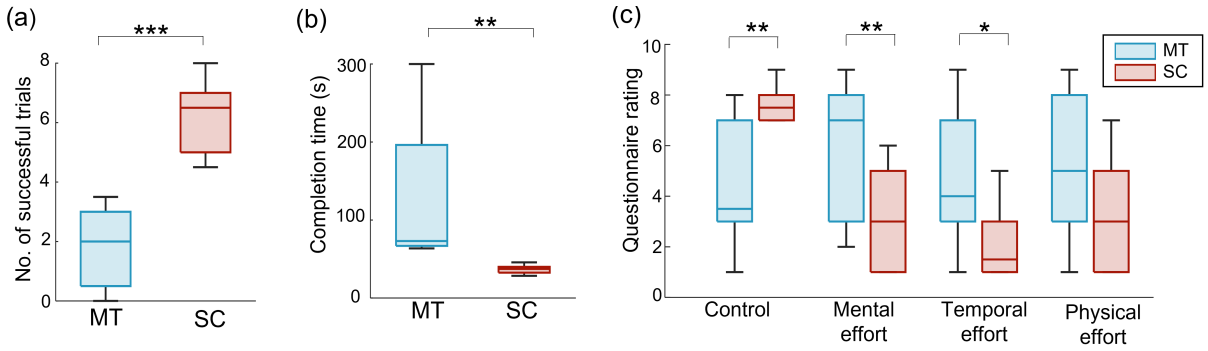

Fig. 5. Experiment results with the laparoscopic simulator platform. (a) Average number of success trials. (b) Average completion time. (c) Subjective rating of questionnaire. Asterisks denote significant effects at *p<0.05, ** p<0.01, *** p<0.001.

full control of the virtual robot ($p = 0.008$). In addition, the participants felt SC also had less *Mental effort* ($p = 0.004$) and *Temporal effort* ($p = 0.02$) with respect to MT, but they did not felt significantly difference in *Physical effort* ($p = 0.06$) when using both control modes.

C. Mental Effort Evaluation of MT and SC using Gaze Data

In addition to the questionnaire in Fig. 5(c), the gaze data collected from the eye tracking unit on the simulator platform can be analysed to assess mental load. Table II summarises the features extracted from the raw gaze data, which have been studied in prior works investigating mental load during teleoperation [29]–[34]. These features can be divided into *eye movement*, *blink*, and *pupillary response*. The two main eye movement features are fixations (periods where the subject's gaze ceases moving and remains fixed on a point) and saccades (rapid movements of the eye between distant focus points). It can be seen that several of the eye movement features exhibit a shift as a result of changing from the MT to SC condition in a significant manner. The mean fixation duration and max fixation duration both are higher during MT, while fixation and saccade frequencies are also lower in MT. Within the blink features, it can be observed that the maximum blink duration increased for MT. Finally for pupillary response, the index of pupillary activity (IPA) is used to track the diameter of the pupil, and it has been linked to mental load responses, where higher values indicated higher task difficulty [30].

TABLE II
MEAN VALUES ACROSS PARTICIPANTS FOR GAZE FEATURES ASSOCIATED WITH MENTAL LOAD DURING THE PEG TRANSFERS.

Eye Movement Features	**Ratio MT/SC**	**p**
Fixation Frequency	0.54	**0.0001**
Mean Fixation Duration	2.02	**0.001**
Max Fixation Duration	1.64	**0.001**
Saccade Frequency	0.52	**0.0001**
Mean Saccade Duration	1.06	0.72
Max Saccade Duration	0.73	0.51
Mean Saccade Speed	0.94	0.20
Max Saccade Speed	0.81	0.14
Blink Features		
Blink Frequency	0.60	**0.0004**
Mean Blink Duration	0.90	0.14
Max Blink Duration	0.74	**0.01**
Pupillary Response Features		
IPA Left	0.90	0.19
Mean Left Pupil Diameter	1.02	0.16
Left Pupil Deviation	0.93	0.29
Max Left Pupil Diameter	0.98	0.26
IPA Right	0.75	**0.01**
Mean Right Pupil Diameter	1.01	0.23
Right Pupil Deviation	0.91	0.23
Max Right Pupil Diameter	0.98	0.13

D. dVRK Platform Results

Eight subjects were randomly selected to test the controllers with the physical dVRK. Their temporal performance with the system is shown in Fig. 6. The similar means and variations in the results indicate that the performances between subjects 1-5 and 6-8 are comparable despite the different visual system used. The participants required an average completion time of 29.77 ± 4.05 s using MT and 24.57 ± 1.43 s via SC without significant difference ($p = 0.15$), which may due to the practice with the simulator. However, we found that the performance between subjects varied considerably especially under the MT control mode. In addition, as shown in Fig. 6(b), the control mode may have different effect to different subjects. To subject 2, 3, 6 and 8, SC control mode provides obvious assistance to complete the task efficiently than MT, but this tendency did not appear on subject 1, 5 or 7.

VI. DISCUSSION

In this letter, EG MaxEnt IRL was used to create a model of the surgeon strategy and generate trajectories for the bimanual peg transfer task. The algorithm was successfully implemented in SurRoL for validation. The simulation environment was modified to directly integrate both a RL training environment and a hand controller for remote teleoperation. Since epsilon-greedy promotes both exploration and exploitation, the reward

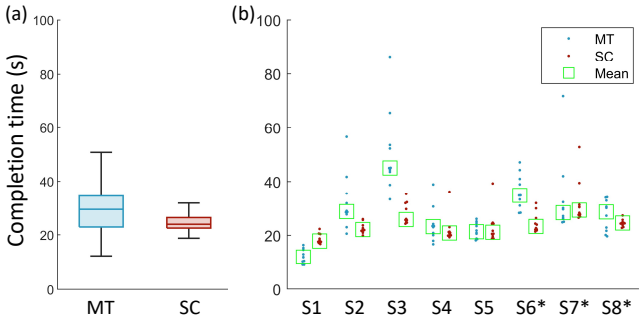

Fig. 6. Experiment results with the dVRK system. (a) Average completion time of SC and MT. (b) Completion time with 10 trials for each control mode per subject. *Subjects that used console and endoscope during the experiment.

values in the Q-table are not only updated near the states visited by the expert and the agent explores a larger volume of the workspace, allowing to obtain a more robust policy.

Noticeably from Fig. 4, the agent learned the expert strategy to approach the block from above and not other directions. This shows the importance of using EG MaxEnt IRL: if the agent had only learned to end at the target, it would have reached it from a non-specific direction, interrupting the surgical workflow as the surgeon would have to manually reposition the end-effector above the block. Instead, since the agent learned the strategy, the surgeon can just resume the directional movement of the previous autonomous segment. This is clear in Fig. 4(b) as the operator and the agent continue each others trajectory without abrupt change in direction.

The results obtained from the surgical simulator (Fig. 5) show that participants without surgical experience can achieve a higher success rate for the bimanual peg transfer task when using SC. In accordance to the results of [7], which exhibited consistently higher success rates with a fully autonomous agent than through direct human operation, our results find a similar performance improvement while offering the possibility for the human to retain a certain level of control while still using an autonomous agent to improve their performance. The results also suggest that the learning curve for the participants when using SC is smaller than MT. The temporal performance indicates that the users can achieve a higher consistency in the speed, with small time difference between each successful attempt. The ratings in the questionnaire show that overall the users perceive a lower workload when using SC instead of MT, in the categories analysed. A common comment provided by the participants was that "the firm grip of the block was the most significant challenge" when completing the task. Interestingly, although the SC mode only assists in the trajectory and does not provide direct help in the grasping, the participants felt significantly more in control during SC as they can simply continue the motion set by the agent.

The algorithm is then successfully implemented on the real dVRK platform with eight trained participants. The results recorded in Fig. 6 are comparable with those obtained on the simulation platform. When using SC, the participants still maintained a higher time consistency. When using MT, they required a shorted average time to complete the task if compared to the simulator experiments. This can be attributed to both the previous training with the simulator and to the more rich sensory information of the real surgical environment.

From the gaze results in Table II, we observed higher fixation duration and max fixation duration in MT, along with lower fixation and saccade frequencies, when compared to SC. The increase in saccades for SC may be a result of the experiment fixed time-limit and the increased number of successful trials during this time limit. More successful grasps over the time span will result in the operator performing more action sequences, thus requiring them to shift their attention more times. Conversely, for MT, as fewer successful trials were accomplished within the time limit, the higher fixation duration may indicate that the operator spent much more time focusing on specific steps in their action sequences during a trial. While some prior works have shown that increased fixation duration has a negative correlation with mental load [29], [31], a recent study has highlighted differing fixation behaviours depending on whether the subject is experiencing a *perceptual* or *cognitive load* [32], where the fixation duration will either decrease or increase respectively. Given the visual scene does not rapidly change during the experiment, it can be argued the subjects are primarily experiencing a cognitive load and so the observed result is consistent with prior works. Additionally, the maximum blink duration is higher during MT, which is again consistent with prior studies that show a positive relationship between blink duration and mental load [29], [33], [34]. Furthermore, the right IPA is significantly higher during SC. While this indicates a higher load for SC, it is possible that, similar to the fixation and saccade frequency feature, the higher value may be linked to the greater number of successful trials the participant completes. The participants perform more work, and so their eye must engage in more movement. It is not clear why the IPA for the right eye indicates significance, while the left eye does not. While it is possible this is due to head movements, alternatively it may be linked to dominant eye effects recently observed during mental load studies considering pupil diameter sizes in response to challenging surgical tasks, where in many cases the right pupil was observed to be larger than the left [35]. This discrepancy in pupil behaviours may be used in future research for the effective assessment of mental load where left or right eye dominance varies between individuals. Overall, these gaze features agree with the subjective ratings, where the MT condition exhibited a higher mental workload.

VII. Conclusion

The system demonstrated the benefits of training agents for shared control of surgical robots during the completion of fundamental movements in robotic surgery using EG MaxEnt IRL. It showed how shared control impacts mental load if compared to traditional manual teleoperation. We developed the algorithm in an unified simulator platform that can accelerate training and testing of HRI control frameworks and successfully implement it on both a virtual and real dVRK. We showed that our control framework can allow users to experience significant reduction of effort and mental load while improving performance. Future works will explore tasks

needle manipulation and suturing. As explained in [10], [19], a significant limitation of MaxEnt IRL is its susceptibility to large state space, which limits the number of dimensions that can be used during training. Therefore a further improvement of the algorithm is to investigate methods that can be adapted to directly handle continuous state space environments, such as the deep deterministic policy gradient [36]–[38].

ACKNOWLEDGMENT

This work was carried out in the Hamlyn Centre at Imperial College London. We thank Kiran Bhattacharyya and Stephen Laws for the help provided during the completion of this work.

REFERENCES

[1] 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*, 2014, pp. 6434–6439.

[2] C. Freschi, V. Ferrari, F. Melfi, M. Ferrari, F. Mosca, and A. Cuschieri, "Technical review of the da vinci surgical telemanipulator," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 9, no. 4, pp. 396–406, 2013.

[3] G. Hubens, H. Coveliers, L. Balliu, M. Ruppert, and W. Vaneerdeweg, "A performance study comparing manual and robotically assisted laparoscopic surgery using the da vinci system," *Surgical Endoscopy and other interventional techniques*, vol. 17, no. 10, pp. 1595–1599, 2003.

[4] Y. Kassahun *et al.*, "Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 11, no. 4, pp. 553–568, 2016.

[5] J. Van Den Berg, S. Miller, D. Duckworth, H. Hu, A. Wan, X.-Y. Fu, K. Goldberg, and P. Abbeel, "Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations," in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 2074–2081.

[6] T. Da Col, A. Mariani, A. Deguet, A. Menciassi, P. Kazanzides, and E. De Momi, "Scan: System for camera autonomous navigation in robotic-assisted surgery," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020, pp. 2996–3002.

[7] M. Hwang *et al.*, "Automating surgical peg transfer: Calibration with deep learning can exceed speed, accuracy, and consistency of humans," *IEEE Transactions on Automation Science and Engineering*, 2022.

[8] K. E. Kaplan, K. A. Nichols, and A. M. Okamura, "Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task," in *2016 IEEE International Conference on Robotics and Automation*, 2016, pp. 723–729.

[9] N. Padoy and G. D. Hager, "Human-machine collaborative surgery using learned models," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 5285–5292.

[10] 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*, 2020, pp. 1380–1386.

[11] M. Power, H. Rafii-Tari, C. Bergeles, V. Vitiello, and G.-Z. Yang, "A cooperative control framework for haptic guidance of bimanual surgical tasks based on learning from demonstration," in *2015 IEEE International Conference on Robotics and Automation*, 2015, pp. 5330–5337.

[12] N. Jarrasse, V. Sanguineti, and E. Burdet, "Slaves no longer: review on role assignment for human–robot joint motor action," *Adaptive Behavior*, vol. 22, no. 1, pp. 70–82, 2014.

[13] Z. Huang, Z. Wang, W. Bai, Y. Huang, L. Sun, B. Xiao, and E. M. Yeatman, "A novel training and collaboration integrated framework for humanndash;agent teleoperation," *Sensors*, vol. 21, no. 24, 2021.

[14] M. Minelli *et al.*, "Integrating model predictive control and dynamic waypoints generation for motion planning in surgical scenario," in *2020 IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2020, pp. 3157–3163.

[15] B. W. King, L. A. Reisner, A. K. Pandya, A. M. Composto, R. D. Ellis, and M. D. Klein, "Towards an autonomous robot for camera control during laparoscopic surgery," *Journal of laparoendoscopic & advanced surgical techniques*, vol. 23, no. 12, pp. 1027–1030, 2013.

[16] 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*. IEEE, 2021, pp. 1–7.

[17] B. D. Ziebart *et al.*, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[18] C. Finn, S. Levine, and P. Abbeel, "Guided cost learning: Deep inverse optimal control via policy optimization," in *International Conference on Machine Learning*. PMLR, 2016, pp. 49–58.

[19] K. Li and J. W. Burdick, "A function approximation method for model-based high-dimensional inverse reinforcement learning," *arXiv preprint arXiv:1708.07738*, 2017.

[20] H. Hur, D. Arden, L. E. Dodge, B. Zheng, and H. A. Ricciotti, "Fundamentals of laparoscopic surgery: a surgical skills assessment tool in gynecology," *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, vol. 15, no. 1, p. 21, 2011.

[21] N. Enayati, A. M. Okamura, A. Mariani, E. Pellegrini, M. M. Coad, G. Ferrigno, and E. De Momi, "Robotic assistance-as-needed for enhanced visuomotor learning in surgical robotics training: An experimental study," in *2018 IEEE International Conference on Robotics and Automation*, 2018, pp. 6631–6636.

[22] G. A. Fontanelli, M. Selvaggio, M. Ferro, F. Ficuciello, M. Vendittelli, and B. Siciliano, "A v-rep simulator for the da vinci research kit robotic platform," in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics*, 2018, pp. 1056–1061.

[23] J. Xu, B. Li, B. Lu, Y. Liu, Q. Dou, and P. 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*, 2021, pp. 1821–1828.

[24] X. B. Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation*, 2018, pp. 3803–3810.

[25] M. Tokic and G. Palm, "Value-difference based exploration: adaptive control between epsilon-greedy and softmax," in *Annual Conference on Artificial Intelligence*. Springer, 2011, pp. 335–346.

[26] S. Akshay, Y. Megha, and C. B. Shetty, "Machine learning algorithm to identify eye movement metrics using raw eye tracking data," in *2020 Third International Conference on Smart Systems and Inventive Technology*. IEEE, 2020, pp. 949–955.

[27] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[28] H.-Y. Kim, "Statistical notes for clinical researchers: Nonparametric statistical methods: 1. nonparametric methods for comparing two groups," *Restorative Dentistry and Endodontics*, vol. 39, no. 3.

[29] Y. Guo, D. Freer, F. Deligianni, and G.-Z. Yang, "Eye-tracking for performance evaluation and workload estimation in space telerobotic training," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 1, pp. 1–11, 2022.

[30] A. T. Duchowski *et al.*, "The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation," in *2018 CHI Conference on Human Factors in Computing Systems*. New York, USA: Association for Computing Machinery, 2018, p. 1–13.

[31] C. Schulz *et al.*, "Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment," *British Journal of Anaesthesia*, vol. 106, no. 1, pp. 44–50, 2011.

[32] J.-C. Liu, K.-A. Li, S.-L. Yeh, and S.-Y. Chien, "Assessing perceptual load and cognitive load by fixation-related information of eye movements," *Sensors*, vol. 22, no. 3, 2022.

[33] F. Volden, V. De Alwis Edirisinghe, and K.-I. Fostervold, "Human gaze-parameters as an indicator of mental workload," in *20th Congress of the International Ergonomics Association*. Cham: Springer, 2019.

[34] K. F. V. Orden, W. Limbert, S. Makeig, and T.-P. Jung, "Eye activity correlates of workload during a visuospatial memory task," *Human Factors*, vol. 43, no. 1, pp. 111–121, 2001.

[35] N. E. Cagiltay and G. G. M. Dalveren, "Are left-and right-eye pupil sizes always equal?" *Journal of Eye Movement Research*, vol. 12, no. 2, 2019.

[36] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[37] X.-l. Chen, L. Cao, Z.-x. Xu, J. Lai, and C.-x. Li, "A study of continuous maximum entropy deep inverse reinforcement learning," *Mathematical Problems in Engineering*, vol. 2019, 2019.

[38] Y. Long, W. Wei, T. Huang, Y. Wang, and Q. Dou, "Human-in-the-loop embodied intelligence with interactive simulation environment for surgical robot learning," *arXiv preprint arXiv:2301.00452*, 2023.