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

Overcoming Imperfect Kinematics in Surgical Robotics Through Sim-to-Real Visuomotor Learning

Zhaoxuan Yan[1,2], Kaizhong Deng[1,2], Zhaoyang Jacopo Hu[1,3],
George P. Mylonas[1,2]*Member, IEEE*, Daniel S. Elson*[1,2]

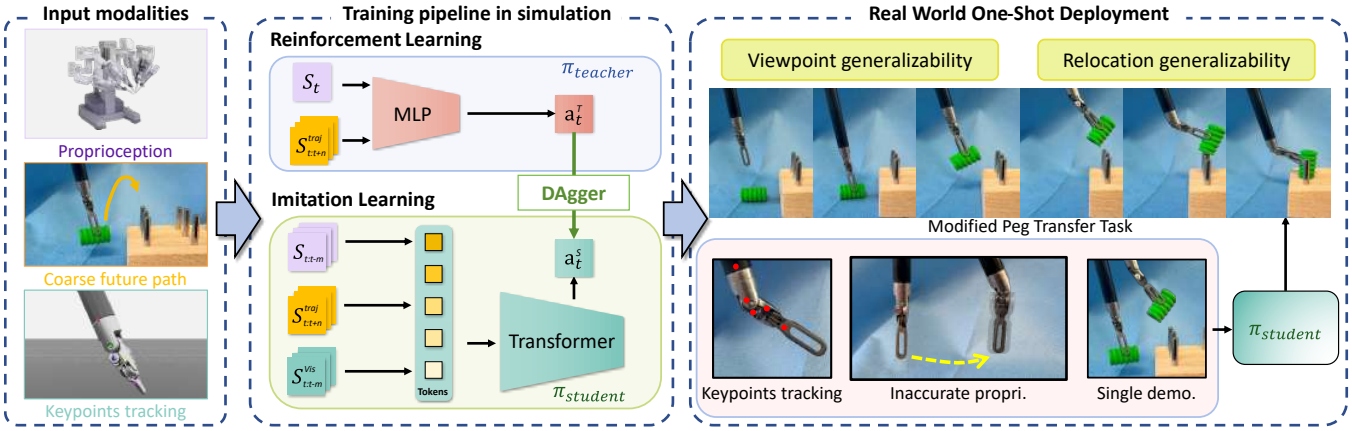

Fig. 1: We proposed a learning framework that actively compensates for a surgical robot's kinematic inaccuracies by training a visuomotor controller within a teacher-student paradigm. The policy learns to fuse unreliable proprioceptive data with reliable visual feedback, enabling robust generalization to variations in camera viewpoint and workspace relocation upon sim-to-real deployment to a physical dVRK.

***Abstract*— Robot-Assisted Surgery is integral to modern minimally invasive procedures, with automation emerging as the next frontier to enhance precision and reduce surgeon fatigue. This evolution is largely impeded by the inherent kinematic inaccuracies of surgical robots, where unreliable internal sensors lead to significant control errors. While previous methods attempted to mitigate these issues through complex model-based calibration, they often suffer from high cost and limited effectiveness. This work utilises a learning-policy to actively compensate for hardware inaccuracies using closed-loop visual feedback that was trained from a teacher-student learning framework. The policy can fuse unreliable internal readings with precise external visual data, allowing it to correct for kinematic errors in real time without needing a perfect physical model. The learned policy was successfully deployed on the da Vinci Research Kit, where experiments validated the fundamental feasibility of using external vision to overcome internal sensor deficits. This research provides a foundational and reliable control methodology, paving the way for more advanced and robust surgical automation.**

I. INTRODUCTION

Robot-Assisted Surgery has gained widespread application in the field of minimally invasive surgery, with systems like the da Vinci surgical system becoming a benchmark of commercial and clinical success [1]. Building on this foundation, the field is advancing from teleoperation towards task automation, aiming to develop intelligent systems capable of autonomously executing surgical procedures that transforms the robot from a passive tool into an active, intelligent agent [2].

Previous studies with a focus of surgical automation have attempted peg transfer [3], [4], [5], shunt insertion [6], [7], endoscopic control [8], instrument manipulation [9], tissue manipulation [10], tissue retraction [11], and tissue resection [12]. It is worth noting that some of the methodology designs are shifting from task-specific solutions towards being general-purpose task-agnostic agents [13].

The foundation of such a task-agnostic approach is the ability to reliably execute any given trajectory [14]. However, it is fundamentally challenged by the surgical system's inherent kinematic inaccuracy [15]. This inaccuracy — arising from hardware factors such as hysteresis, mechanical flexibility, and joint clearance — yields unreliable joint angle measurements [16]. Consequently, a discrepancy emerges between the robot's commanded trajectory and its actual executed motion, affecting the robot's ability to perform delicate surgical tasks. Addressing this challenge remains essential for reliable surgical automation.

Classical approaches typically rely on model-based calibration to improve kinematic accuracy. Geometric calibration adjusts Denavit-Hartenberg (DH) parameters by aligning measured and modelled end-effector poses [17]; gravity compensation identifies dynamic parameters to offset gravitational and cable-induced disturbances [18]; and vision-based calibration employs RGB-D sensing to correct cable-

[1]Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London.
*Corresponding author.
[2]Department of Surgery and Cancer, Imperial College London.
[3]Department of Mechanical Engineering, Imperial College London, Exhibition Road, London, SW7 2AZ, UK.
Corresponding email: daniel.elson@imperial.ac.uk

driven inaccuracies [19]. However, such methods generalise poorly across surgical environments and offer limited overall effectiveness [20]. More recently, learning-based approaches have shown promise in estimating and compensating for kinematic errors [21], [22], yet typically require precise end-effector pose measurements, which are difficult to obtain in practice.

Recent work [23], [24] decouples the policy architecture into a high-level planner and a low-level controller. This separation enables task-level reasoning independent of complex robot dynamics [25]. Therefore, the low-level controller can focus on a task-agnostic trajectory tracking [26]. Our work targets this low-level controller, as reliable trajectory execution is a prerequisite for the success of any high-level planner.

While the hierarchical structure simplifies the controller's objective, training a low-level policy is challenging due to the cable-drive surgical robot's unreliable proprioceptive data. The teacher-student paradigm is an effective methodology across various complex robots, including humanoids [27], quadrupeds [28], and manipulators [29]. In this framework, a teacher policy is first trained in simulation with access to privileged information, such as accurate physical state. It is then distilled into a student policy that relies solely on the imperfect sensory data available on the physical system. This naturally frames kinematic inaccuracy as a discrepancy between an ideal teacher with a perfect kinematic model and a realistic student that must operate with the faulty kinematics

This paper proposes a learning-based teacher-student framework that leverages closed-loop visual feedback to actively compensate for the system's inherent kinematic errors. The approach has been validated in high-fidelity simulation [30] and subsequently deployed onto a physical robotic platform, da Vinci Research Kit (dVRK) [31], via sim-to-real transfer. The main contributions of this work are:

- A low-level visual–motor controller, trained in a teacher–student framework, that learns to compensate for internal kinematic errors using external visual feedback.
- A sim-to-real transfer evaluation on a physical dVRK platform in a serial setup, demonstrating competitive performance compared to both classical control methods and learning-based policies.
- Real-world experiments that highlight the robustness of the policy to the dVRK's kinematic inaccuracies, establishing a foundation for future research in surgical automation.

II. METHOD

A. Overview of Teacher Student Framework

As illustrated in Fig. 2, our methodology is centered on a teacher-student learning framework designed for sim-to-real transfer. The teacher-student framework comprises two distinct training phases designed to address the challenges of precise trajectory tracking in robotic surgery. Initially, the teacher policy is trained using Reinforcement Learning, specifically the Proximal Policy Optimization (PPO) algorithm [32], to learn an optimal trajectory tracking strategy. This policy serves as an expert supervisor, providing high-quality demonstration data for the subsequent student training phase.

The student policy then aims to replicate the teacher's expert performance whilst operating under realistic conditions, learning to fuse unreliable, non-privileged internal state information from the physical dVRK with reliable external visual feedback to achieve precise control.

However, naive Imitation Learning approaches for this policy distillation suffer from distribution shift. To address this challenge, we employ the Data Aggregation (DAgger) algorithm [33], an interactive imitation learning method that enhances the student's robustness by iteratively collecting expert corrections. It enables the policy to learn effective recovery strategies from its own errors.

B. "Ideal" Teacher Policy

Training Pipeline. The teacher policy is trained using privileged information, which is available exclusively in the simulator. Specifically, its input state is composed of the robot's true proprioceptive data S_t^{Jnt}, consisting of joint positions, joint velocities, and the last executed action. Access to this complete and accurate state information is what significantly enhances the efficiency and final performance of the RL training process. The input also contains a preview of 30 future waypoints from the target trajectory $S_{t:t+n}^{Tgt}$, which allows the policy to anticipate upcoming movements for smoother control. The policy itself is formed by a Multi-Layer Perceptron (MLP), and its output is a 6-dimensional vector representing the relative position increment for each joint at the next step.

Reward Design. The policy is guided by a reward function designed to encourage both high tracking precision and smooth motion. The specific components and their weights are detailed in Table I. As position and orientation errors are often difficult to optimize simultaneously, we employ a multiplicative coupling of their respective reward terms. This design compels the policy to be balanced, as it must reduce both errors simultaneously to achieve a high reward. Furthermore, for a high-precision demand such as a surgical task, balancing initial exploration with final fine-grained control is

TABLE I
REWARD FUNCTION TERMS AND WEIGHTS

Reward Terms	Expressions	Weights
Position error	$\exp\left(-\frac{\|\mathbf{p}-\mathbf{p}^{\text{tg}}\|^2}{\sigma_{\text{pos}}^2}\right)$	0.6
Orientation error	$\exp\left(-\frac{\|\mathbf{q}-\mathbf{q}^{\text{tg}}\|^2}{\sigma_{\text{orn}}^2}\right)$	0.1
Action rate	$-\|\dot{\mathbf{a}}\|^2$	0.35
Joint velocities	$-\|\dot{\theta}\|^2$	0.01
Balanced error	$\exp\left(-\left(\frac{\|\mathbf{p}-\mathbf{p}^{\text{tg}}\|^2}{\sigma_{\text{pos}}^2}+\frac{\|\mathbf{q}-\mathbf{q}^{\text{tg}}\|^2}{\sigma_{\text{orn}}^2}\right)\right)$	8

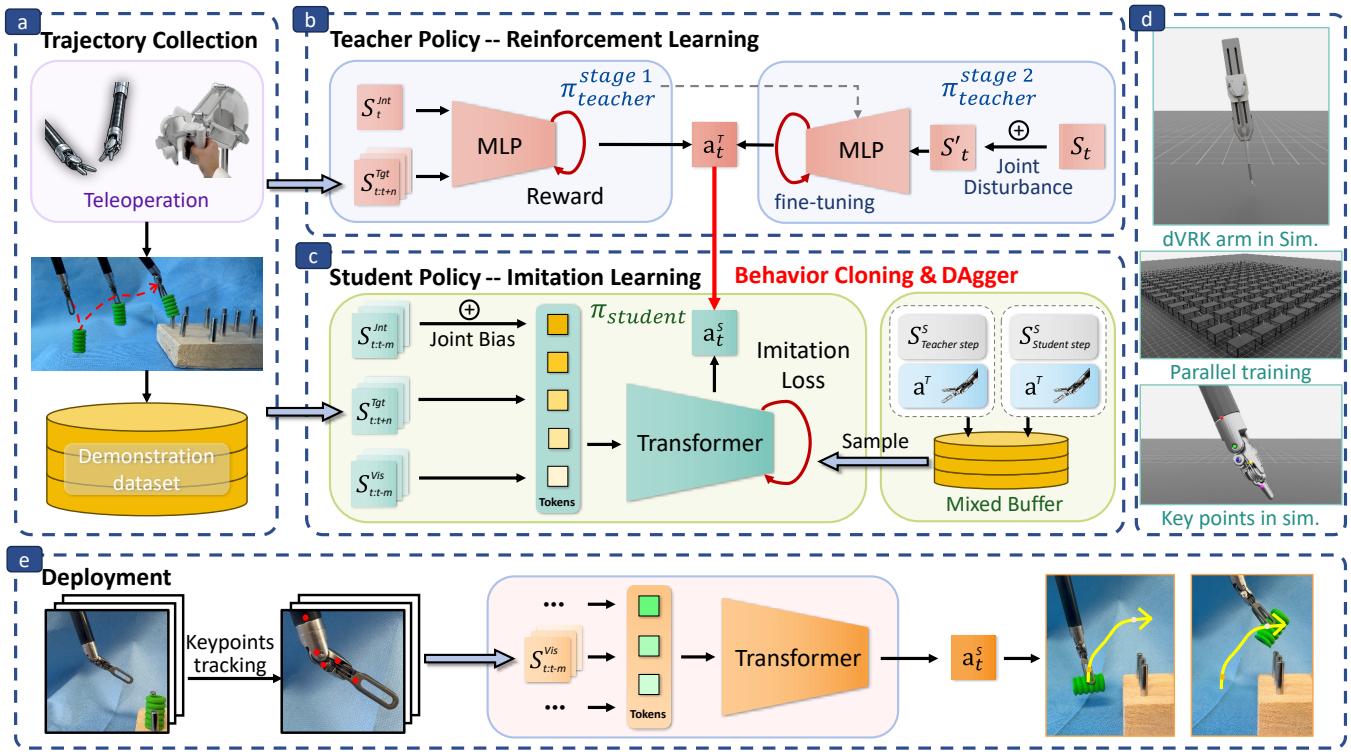

Fig. 2: **The overview of the Training framework**: **(a)**: The framework begins with collecting expert trajectories via teleoperation. **(b)**: A teacher policy is then trained in simulation using reinforcement learning and fine-tuned with joint disturbances to develop robust recovery behaviour. **(c)**: Subsequently, a student policy uses Imitation Learning (DAgger) to distil the teacher's expertise, learning to fuse biased proprioceptive data with external visual keypoints to compensate for kinematic inaccuracies introduced in simulation. **(d)**: The simulation comprises multiple parallel training environments, each containing one dVRK patient-side arm. Keypoints tracked in the simulation are indicated with colour labels. **(e)**: Finally, the trained student policy is deployed directly on the physical robot for real-world task execution.

critical. We therefore introduce a curriculum learning strategy that progressively tightens the reward function's error tolerances, effectively guiding the policy from initial, large-scale movements to the delicate, high-precision adjustments required in the final stages. The detailed curriculum schedule is provided in Table II.

Fine-Tuning for Robustness. A teacher that is only an expert on the ideal trajectory (stage 1) is insufficient for guiding an imperfect imitator. During the subsequent imitation learning, the student policy will inevitably explore states that deviate from the expert's distribution. To proactively address this, the teacher undergoes a second stage (stage 2) of fine-tuning with the goal of transforming it from a perfect executor into a "recovery expert". This is achieved by continuing the training while introducing procedural joint perturbations of gradually increasing magnitude and noise rate. This process forces the policy to learn how to recover from significant state deviations, resulting in a robust teacher that can provide effective, corrective supervision throughout the student's learning process.

C. "Realistic" Student Policy

State Representation. The student's observations consist of potentially inaccurate proprioceptive data from dVRK and visual information needed for the policy to learn how to compensate. Similar to the teacher policy, this uses joint positions, joint velocities, and the last executed action. In addition, our method relies on a more lightweight and robust representation of visual information: a set of five 2D keypoints projected from 3D keypoints on the instrument to represent the end-effector's pose as $S^{Vis}_{t-m:t}$. Finally, a history of observations is included in the state to allow the policy to implicitly infer system dynamics for smoother control actions.

Domain randomisation. Domain randomisation is employed during training to improve the policy's robustness. We note that real dVRK kinematic errors are known to be configuration-dependent, correlated across joints, and

TABLE II
CURRICULUM LEARNING PARAMETERS

Position Error Curriculum		Orientation Error Curriculum	
ϵ_{pos}	σ^2_{pos}	ϵ_{orn}	σ^2_{orn}
0.3	0.13	1.0	1.44
0.15	0.032	0.6	0.52
0.08	0.0092	0.15	0.032
0.01	0.00014	0.08	0.0092
0.005	3.6×10^{-5}	0.05	3.6×10^{-3}
0.003	1.3×10^{-5}	0.03	1.3×10^{-3}
0.002	5.76×10^{-6}	0.02	5.76×10^{-4}
0.001	1.44×10^{-6}	0.01	1.44×10^{-4}

For position error curriculum, ϵ_{pos} (in m) and σ^2_{pos} represent the threshold and sigma in the reward term. For orientation error curriculum, ϵ_{orn} (in rad) and σ^2_{orn} represent the threshold and sigma in the reward term.

history-dependent [21], [22], [16]. Rather than faithfully replicating this error structure, our approach follows the established domain randomisation principle: training under an unstructured distribution that envelops the real-world variation produces robust transferable policies. The magnitude of each randomisation parameter was selected based on estimated real-system error bounds and the convergence of the learned policy. Therefore, independent random biases are sampled uniformly from ± 0.05 rad for each joint and held fixed throughout the episode, simulating a consistent but random kinematic offset. Similarly, to make the policy less sensitive to camera placement, the camera's pose is randomly initialised with a positional offset of up to $\pm 2\,cm$ and a rotational offset of up to $\pm 10\,^{\circ}$. This encourages the policy to learn view-invariant features and allows for a more flexible real-world setup without requiring precise calibration. Finally, in addition to these per-episode randomisations, a small amount of Gaussian noise is applied to all observation inputs at each time step to further enhance the policy's robustness.

Policy Architecture. Inspired by ACT [34], a Transformer architecture comprising 4 encoder layers and 1 decoder layer is employed to capture complex temporal and multi-modal dependencies. The self-attention mechanism can explicitly model these dependencies, which is crucial for achieving precise and smooth control. Each proprioceptive state and the keypoint state are tokenised via linear projections, with modality-specific embeddings added to each token. Learnable positional embeddings are added to the token representations of historical observations within a fixed-length temporal window.

Training Pipeline. The student policy is trained to minimize an imitation MSE loss between its predicted action and the teacher's expert action:

$$\mathcal{L} = \mathbb{E}_{s_t}\left[\|\hat{a}_t - a_t^*\|_2^2\right], \tag{1}$$

where $\hat{a}_t$ is the predicted action and a_t^* is the optimal action produced by the teacher policy.

Training exclusively on new interactive data, however, can cause the policy to catastrophically forget the initial expert behaviour. This is addressed by a dual-buffer strategy that samples from both a static Expert Buffer of ideal trajectories and a rolling Mixed Buffer of data from the DAgger phase [33]. This combination of a structured curriculum and balanced data sampling allows the policy to learn stable error correction while retaining the high precision of the original expert behaviour.

D. Sim-to-Real Transfer

A successful sim-to-real transfer requires a real-world perception module that can provide 2D keypoint coordinates with high consistency and low latency. We adopt the pyramidal Lucas-Kanade (LK) optical flow algorithm to track the keypoints. In our workflow, an operator first manually initializes keypoints on the initial frame, which are then automatically tracked in real-time by the LK algorithm. This method allows fast and accurate tracking, and provides the geometric consistency over other more granular tracking methods [35].

III. EXPERIMENTS

In this section, we present a series of experiments to evaluate the effectiveness of our proposed policy. We first introduce the experimental setup and simulation performance, followed by a detailed analysis addressing the following key research questions:

Q1. How effectively does our policy compensate for the dVRK's kinematic inaccuracies when the target trajectory is relocated to different regions of the workspace?

Q2. How well does our policy generalise across different camera viewpoints?

Q3. To what extent does the policy's success depend on the availability and quality of keypoint observations?

A. Environment Setup

Robotic platform. Our experimental setup is a distributed system designed for real-time, closed-loop control, with the overall architecture illustrated in Fig. 3. Teleoperation is performed using a Sigma 7 controller to command the Cartesian motion of the dVRK instrument. The Realsense D405 Camera video stream and the dVRK trajectory information are recorded. All communication between these components is managed by the Robot Operating System (ROS), enabling the entire loop to operate at 30Hz. The resulting action command is executed by the dVRK with a PD controller.

Task Descriptions. A modified peg transfer task was adopted as a benchmark to validate the proposed framework. The standard peg transfer task is widely used as an evaluation benchmark on the dVRK [3], [4], [5], owing to its clear correspondence to clinical manipulation skills. To increase task complexity and introduce orientation variability, the task objective is defined as placing a peg onto a vertical stick on the pegboard, whilst the peg is initialised on its side adjacent to the board. This setup encourages greater wrist rotation of the instrument, making the task more challenging and representative of dexterous surgical manipulation.

Data Collection. A total of 80 successful demonstrations were collected via teleoperation. Domain randomisation was applied throughout data collection by varying the position and orientation of both the peg and the pegboard on the workspace surface, initialising the robot arm from diverse configurations, and perturbing the camera pose.

Baseline models. To comprehensively evaluate our proposed policy, its performance is compared against two distinct baselines:

- **IK Replay:** This replays the recorded Cartesian trajectory from teleoperation in joint space relying on the inverse kinematics (IK) of the dVRK. It is a non-learning, open-loop method that is affected by the dVRK's kinematic inaccuracies.
- **ACT:** The Action Chunking Transformer [34] is a state-of-the-art imitation-learning-based policy for robotic manipulation. It provides a strong benchmark for comparison with other task-level policy learning methods.

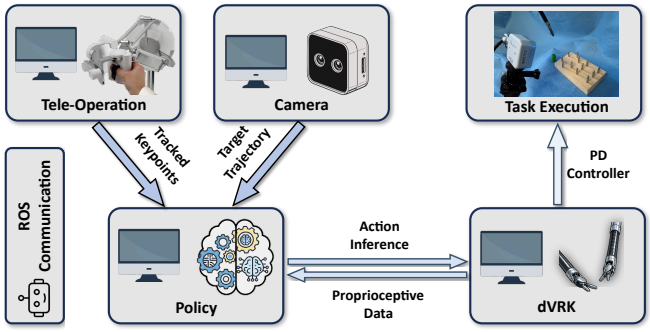

Fig. 3: **Overview of the robotic control system**: The central Policy module integrates a target trajectory from the teleoperation system, tracked keypoints from a camera, and proprioceptive data from the dVRK to infer the action. This inferred action is then sent back to the dVRK, whose internal PD Controller performs the final task execution. The entire distributed system is connected and communicates via ROS.

The primary evaluation metric for all real-world experiments is the Success Rate. A trial on the peg-transfer task is considered successful if the peg is transferred from its start position to the target pin in a smooth motion, without getting stuck.

Training was carried with the ORBIT-Surgical framework [30], which is built upon NVIDIA Isaac Sim, using a single NVIDIA RTX 5090 GPU. The two-stage teacher policy training took approximately 40 hours in total, utilizing 8192 parallel environments for data collection, with a total batch size of 8192×64. The student policy was subsequently trained for approximately 20 hours using 200 parallel environments and a batch size of 200×256. The ACT was trained on the demonstration dataset for 200k gradient steps with its original hyper-parameters. The proprioceptive modalities are tokenised identical to the proposed policy while visual inputs are tokenised respective to the original implementation.

B. Simulation Performance

Before deployment on the physical robot, the trained policies were first validated in the simulation environment. This evaluation serves two main purposes: to verify the effectiveness of our teacher-student learning framework and to establish an upper-bound performance benchmark for the subsequent real-world experiments.

While the teacher policy achieves near-perfect trajectory tracking with privileged information, the student policy exhibits slight but expected deviations. To precisely quantify these observations, we present a summary of the average performance in Table III. The visualisation of the error through a rollout episode is shown in Fig. 4. This shows that the teacher policies maintain a position error and orientation of approximately $0.6\,mm$ and $0.015\,rad$ throughout the trajectory. The student policy error, while slightly higher at around $1.2\,mm$ and $0.03\,rad$, remains low and stable, following the teacher's performance profile. Notably, the Stage 2 teacher's performance is nearly identical to that of Stage 1, demonstrating that the robustness fine-tuning did not sacrifice precision.

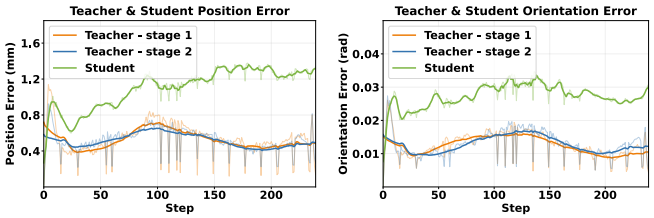

Fig. 4: **Tracking Performance of Teacher and Student Policies in Simulation.** The plots compare the position error (**left**) and orientation error (**right**) for the teacher and student policies over a rollout trajectory. Both Stage 1 and Stage 2 teacher policies demonstrate nearly identical, high-precision performance with low, stable errors. The student policy successfully tracks the trajectory but exhibits a consistently higher tracking error.

The graph in Fig. 4 indicates a slight upward trend in student error, which is an acceptable outcome for DAgger. This is because while DAgger mitigates the cumulative error arising from distribution shifts during imitation learning, it cannot entirely eliminate it. These results quantitatively confirm the success of our proposed learning framework. The student policy successfully distils the precision tracking ability from the expert teacher, which learns to achieve high-precision tracking by leveraging external visual feedback to compensate for its inaccurate proprioceptive information.

C. Real-World Deployment

1) Relocation generalizability: Relocation generalizability experiments were conducted to evaluate the policy's effectiveness across different workspace configurations. In this experimental protocol, the workspace was horizontally shifted relative to the robot whilst maintaining the internal relative positioning of objects. The policy was required to complete the task by replicating the demonstration trajectory within the shifted workspace configuration.

This experiment could test policy's ability to compensate for the dVRK's kinematic inaccuracies across different regions of the workspace. The underlying hypothesis is that these inaccuracies are workspace-dependent, as different joint configurations can lead to varying error characteristics. A truly robust policy must therefore generalize its corrective behaviour rather than overfit to a single region.

To test this, we evaluate our policy against a traditional open-loop inverse kinematic baseline, named as inverse kinematic (IK) Replay, which is supposed to achieve the task with high success rate when there are no kinematic issues. The expert trajectories for the peg-transfer task were systematically relocated horizontally (along the x-axis) from $-4\,cm$ to $+6\,cm$, forcing the robot to operate in different

TABLE III
TRACKING ERROR IN SIMULATION.

Metric	Teacher Policy		Student Policy
	Stage 1	Stage 2	
Position error $[mm]$	0.6	0.6	1.2
Orientation error $[rad]$	0.015	0.015	0.03

TABLE IV
RELOCATION ROBUSTNESS

Method	↑Success Rate after x_{shift} $[cm]$					
	-4	**-2**	**0**	**+2**	**+4**	**+6**
Ours	**15/20**	17/20	18/20	19/20	**15/20**	**11/20**
IK Replay	0/20	**20/20**	**20/20**	**20/20**	0/20	0/20

The trajectory was relocated by shifting x_{shift} it horizontally on the platform in a range of $-4\,cm$ to $6\,cm$, where $0\,cm$ represents the original location. **Ours** represents our proposed method.

TABLE V
VIEWPOINT ROBUSTNESS

Method	↑Success Rate under Camera Pose Variation				
	Original	**±1 cm**	**±3 cm**	**±5°**	**±10°**
Ours	19/20	**17/20**	**13/20**	**14/20**	**8/20**
ACT	19/20	15/20	7/20	10/20	3/20

During the evaluation, the camera pose was varied to assess its sensitivity to pose changes. Two types of variation were considered: horizontal translation (in cm) and yaw rotation (in °). **Original** indicated the original camera pose identical to the training setup.

joint configurations within its workspace.

The results of this experiment are summarized in Table IV. The IK Replay baseline performs well in a very narrow range, achieving 100% success at the original location (0 cm) and with a small ±2 cm offset. However, its performance collapses to 0% success with any larger offset. In contrast, our policy demonstrates significantly more robust performance. It maintains a high success rate of over 80% across a wide range of offsets (±2 cm). Even at the largest tested offset of +6 cm, it still achieves a remarkable success rate of 55% compared to IK Replay.

The failure of IK Replay under large workspace offsets can be attributed to the dVRK's kinematic inaccuracies. The initial success at the original position is deceptive; it occurs not because the system is inherently accurate, but because the teleoperated trajectory was collected in this exact configuration and has implicitly encoded the specific, configuration-dependent kinematic errors of that workspace region. This open-loop method is fragile because the robot's joint configuration and its corresponding error profile change when the trajectory is relocated to different workspace positions. Furthermore, the consistent failure pattern highlights that the underlying issue represents a repeatable, deterministic kinematic error rather than random control noise.

In contrast to the brittle IK Replay, our policy's success demonstrates the key advantage of its learning-based design. Instead of relying on a flawed internal kinematic model, it has learned a generalizable mapping from visual error to corrective action. This allows it to handle the robot's configuration-dependent inaccuracies, allowing it to succeed where the open-loop method fails.

Despite its overall performance, we also observe that our policy's success rate presents a slight downward trend as the trajectory offset increases. We attribute this to a combination of three potential factors. Firstly, the expert dataset may not fully cover the entire workspace, causing the policy to encounter out-of-distribution states at large offsets. Secondly, a sim-to-real gap in perception may arise. While keypoints are derived from ideal projections in simulation, the physical camera's lens distortion can reduce tracking accuracy near the periphery of the image. Thirdly, the largest offsets may push the robot towards its physical workspace boundaries, where its mechanical control and precision can be inherently less reliable.

2) Viewpoint Generalizability: A key objective of our work is to develop a policy that is robust to variations in camera placement, relaxing the strict requirement for a fixed, precisely calibrated camera common in many SOTA methods. To quantitatively evaluate this generalizability, we compare our policy against another baseline, the Action Chunking Transformer [34]. The camera pose was systematically varied from its initial calibrated position, with perturbations in both translation (±1 cm and ±3 cm) and rotation (±5° and ±10°).

The results of this comparison are presented in Table V. When operating at original calibrated camera pose, both our policy and the ACT baseline achieve a high success rate of 95%. As the camera pose is varied, the performance of both methods decreases, but at different rates. Under the largest translational shift of ±3 cm, our policy's success rate is 65%, while ACT's is 35%. Similarly, under the largest rotational variation of ±10°, our policy achieves a 40% success rate, compared to 15% for the ACT baseline. The data shows that our policy maintains a consistently higher success rate than the ACT baseline under the camera pose perturbations.

The results reveal that the superior generalisability of our policy is a direct outcome of our training methodology. The camera pose randomisation employed during training forced our policy to learn a more view-invariant control strategy, focusing on geometric relationships that remain consistent across different viewpoints. In contrast, the ACT collapses when the visual input shifts into an out-of-distribution domain. However, the results also show a degradation in our policy's performance as the perturbations become larger.

3) Visual Features Robustness: Reliable keypoint tracking in real surgical environments remains challenging; hence, robustness to corrupted or incomplete observations is a critical property for any clinically viable policy. To evaluate this, we conduct an ablation study examining two complementary aspects of the policy's dependence on external keypoint inputs: (1) its tolerance to observation corruption, and (2) the necessity of keypoint feedback beyond proprioceptive information alone. Results are presented in Table VI.

Robustness to corrupted keypoint observations. In **Noisy**, Gaussian noise with a standard deviation of 5 pixels was added to all keypoint estimates at inference time. The policy maintained robust performance, with the success rate declining only marginally from 90% to 80%, demonstrating tolerance to moderate localisation errors. In the **Drop 2** condition, two of the five tracked keypoints were randomly masked during each episode of inference, causing the success

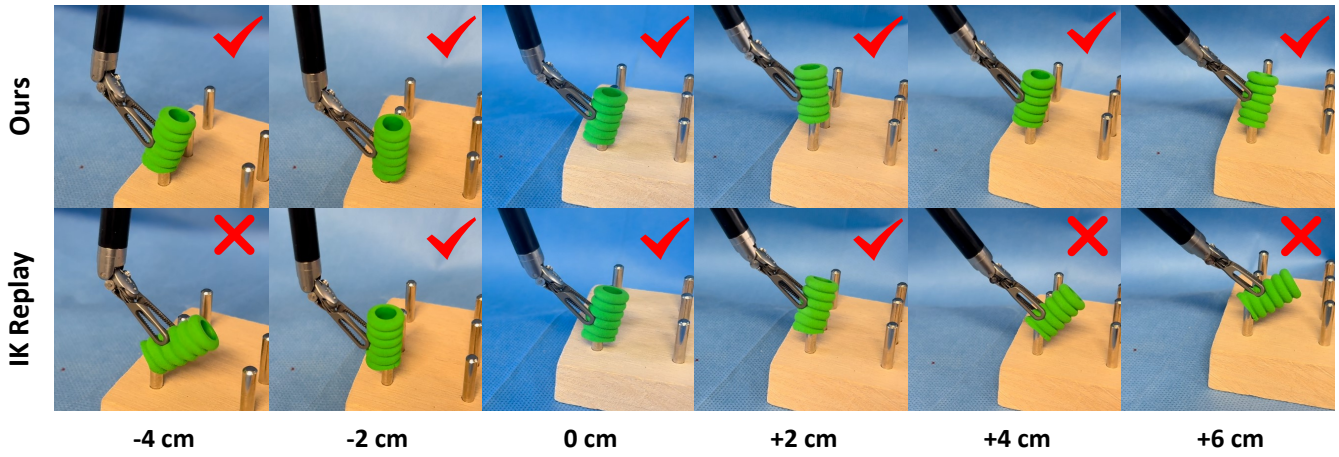

Fig. 5: **Example frames from the Relocation Generalizability experiment.** This figure presents a qualitative comparison between our policy (**Ours**, top row) and the **IK Replay** baseline (bottom row). The peg transfer task was systematically shifted horizontally from -4 cm to +6 cm. The results show that our policy successfully completes the task across a wide range of locations (indicated by ✓). In contrast, the IK Replay baseline succeeds only at or near the original location (0 cm) but fails at larger offsets (indicated by ✗). **Ours** represents our proposed method.

rate to drop substantially to 30%. These results indicate its inherent tolerance to minor tracking jitter commonly encountered in real-world keypoint tracking systems.

Necessity of keypoint feedback over proprioception alone. When four keypoints were masked (**Drop 4**), the policy failed entirely, yielding a 0% success rate. In the **No Access** condition, the keypoint modality was entirely removed during both training and inference, yielding a 0% success rate. This confirms that the policy does not succeed by relying solely on inaccurate proprioceptive signals. Instead, it utilised keypoint-based visual feedback to perceive and adapt to the current pose.

TABLE VI
ROBUSTNESS TO KEYPOINTS TRACKING

	↑Success Rate				
Method	**Original**	**Noisy**	**Drop 2**	**Drop 4**	**No access**
Ours	9/10	8/10	3/10	0/10	0/10

In this ablation study of visual keypoints tracking input, the success rate of five cases are reported. **Original** is identical to the proposed policy setup. **Noisy** represents adding Gaussian noise to the tracking input. **Drop 2** and **Drop 4** represents masking out two or four out of totally five tracked keypoints. **No access** means removing this input modality during training and further evaluation

IV. DISCUSSION AND CONCLUSION

This paper presents a novel visuomotor controller developed within a teacher-student learning framework, that actively compensates for the inherent kinematic inaccuracies of surgical robots by leveraging closed-loop visual feedback. An expert teacher, trained in an ideal simulation via reinforcement learning with privileged information, provides supervision for a student policy that learns through an interactive process to fuse unreliable proprioception with reliable external 2D keypoints. Real-world deployment of the learned policy demonstrated its ability to compensate for workspace-dependent kinematic errors and its significant robustness to camera pose variations. Therefore, this work establishes a complete methodology for developing low-level controllers that can overcome physical hardware limitations for precise surgical trajectory tracking.

However, this work still faces key challenges. Firstly, the visual perception module has critical limitations: it requires manual initialization and is fragile to visual perturbations like keypoint occlusion and significant lighting changes; moreover, the projective nature of 2D keypoints can make it difficult to unambiguously represent the tool's complete 3D pose. Secondly, whilst simulated linear errors are incorporated during model training, non-linear systematic errors present in the dVRK are not modelled, potentially limiting the framework's capacity to compensate for such disturbances. Thirdly, the policy has only been exhaustively evaluated on a single exemplar task, which may not fully reflect its performance across the broader spectrum of surgical procedures.

Future work should focus on several key directions. We intend to integrate an autonomous perception system capable of learning to select task-relevant keypoints whilst generalising to novel surgical tools and environments. Additionally, non-linear dVRK system errors will be characterised and injected into the simulation to enable the policy to compensate for such disturbances more effectively. The proposed framework will further be evaluated across a broader range of surgical tasks to assess its generalisation capability and robustness. We hope this work serves as a foundational step towards overcoming the hardware limitations of current surgical robots, paving the way for more robust and intelligent surgical automation systems.

ACKNOWLEDGMENTS

This paper is independent research funded by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC), the Cancer Research UK (CRUK) Imperial Centre, the Wellcome Trust ITPA MedTechOne awards, and Imperial-CSC scholarship.

REFERENCES

[1] C. D'Ettorre, A. Mariani, A. Stilli, F. Rodriguez y Baena, P. Valdastri, A. Deguet, P. Kazanzides, R. H. Taylor, G. S. Fischer, S. P. DiMaio, A. Menciassi, and D. Stoyanov, "Accelerating Surgical Robotics Research: A Review of 10 Years With the da Vinci Research Kit," *IEEE Robotics & Automation Magazine*, vol. 28, no. 4, pp. 56–78, 2021.

[2] S. Schmidgall, J. D. Opfermann, J. W. Kim, and A. Krieger, "Will your next surgeon be a robot? Autonomy and AI in robotic surgery," *Science Robotics*, vol. 10, no. 104, p. eadt0187, 2025.

[3] H. Zheng, Z. J. Hu, Y. Huang, X. Cheng, Z. Wang, and E. Burdet, "A User-Centered Shared Control Scheme with Learning from Demonstration for Robotic Surgery," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 195–15 201.

[4] Z. J. Hu, Z. Wang, Y. Huang, A. Sena, F. Rodriguez y Baena, and E. Burdet, "Towards Human-Robot Collaborative Surgery: Trajectory and Strategy Learning in Bimanual Peg Transfer," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4553–4560, 2023.

[5] J. Chen, Z. Wang, R. Zhu, R. Zhang, W. Bai, and B. Lo, "Path Generation with Reinforcement Learning for Surgical Robot Control," in *2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 2022, pp. 1–4.

[6] K. Dharmarajan, W. Panitch, B. Shi, H. Huang, L. Y. Chen, M. Moghani, Q. Yu, K. Hari, T. Low, and D. Fer, "Robot-assisted vascular shunt insertion with the dvrk surgical robot," *Journal of Medical Robotics Research*, vol. 8, no. 03n04, p. 2340006, 2023.

[7] K. Dharmarajan, W. Panitch, B. Shi, H. Huang, L. Y. Chen, T. Low, D. Fer, and K. Goldberg, "A trimodal framework for robot-assisted vascular shunt insertion when a supervising surgeon is local, remote, or unavailable," in *2023 International Symposium on Medical Robotics (ISMR)*. IEEE, Conference Proceedings, pp. 1–8.

[8] B. Dong, J. Chen, Z. Wang, K. Deng, Y. Li, B. Lo, and G. Mylonas, "An Intelligent Robotic Endoscope Control System Based on Fusing Natural Language Processing and Vision Models," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 8180–8186.

[9] H. Zhang, K. Deng, Z. J. Hu, B. Huang, and D. S. Elson, "Hybrid Deep Reinforcement Learning for Radio Tracer Localisation in Robotic-Assisted Radioguided Surgery," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 15 465–15 471.

[10] 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)*, 2019, pp. 3875–3881.

[11] 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)*, 2021, pp. 1–7.

[12] R. Zhang, J. Chen, Z. Wang, Z. Yang, Y. Ren, P. Shi, J. Calo, K. Lam, S. Purkayastha, and B. Lo, "A step towards conditional autonomy-robotic appendectomy," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2429–2436, 2023.

[13] J. W. B. Kim, J.-T. Chen, P. Hansen, L. X. Shi, A. Goldenberg, S. Schmidgall, P. M. Scheikl, A. Deguet, B. M. White, D. R. Tsai, R. J. Cha, J. Jopling, C. Finn, and A. Krieger, "SRT-H: A hierarchical framework for autonomous surgery via language-conditioned imitation learning," *Science Robotics*, vol. 10, no. 104, p. eadt5254, 2025.

[14] Y. Long, A. Lin, D. H. C. Kwok, L. Zhang, Z. Yang, K. Shi, L. Song, J. Fu, H. Lin, W. Wei, K. Chen, X. Chu, Y. Hu, H. C. Yip, P. W. Y. Chiu, P. Kazanzides, R. H. Taylor, Y. Liu, Z. Chen, Z. Wang, null, and Q. Dou, "Surgical embodied intelligence for generalized task autonomy in laparoscopic robot-assisted surgery," *Science Robotics*, vol. 10, no. 104, p. eadt3093, 2025.

[15] 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," 2024.

[16] Z. Cui, J. Cartucho, S. Giannarou, and F. R. y Baena, "Caveats on the First-Generation da Vinci Research Kit: Latent Technical Constraints and Essential Calibrations [Survey]," *IEEE Robotics & Automation Magazine*, vol. 32, no. 2, pp. 113–128, 2025.

[17] D. Seita, S. Krishnan, R. Fox, S. McKinley, J. Canny, and K. Goldberg, "Fast and Reliable Autonomous Surgical Debridement with Cable-Driven Robots Using a Two-Phase Calibration Procedure," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 6651–6658.

[18] H. Lin, C.-W. Vincent Hui, Y. Wang, A. Deguet, P. Kazanzides, and K. W. S. Au, "A Reliable Gravity Compensation Control Strategy for dVRK Robotic Arms With Nonlinear Disturbance Forces," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3892–3899, 2019.

[19] 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.

[20] X. Li, E. Zhang, X. Fang, and B. Zhai, "Calibration Method for Industrial Robots Based on the Principle of Perigon Error Close," *IEEE Access*, vol. 10, pp. 48 569–48 576, 2022.

[21] J. A. Barragan, H. Ishida, A. Munawar, and P. Kazanzides, "Improving the Realism of Robotic Surgery Simulation Through Injection of Learning-Based Estimated Errors," in *2024 International Symposium on Medical Robotics (ISMR)*. Atlanta, GA, USA: IEEE, June 3–5 2024, pp. 1–6, arXiv:2406.07375.

[22] J. Mahler, S. Krishnan, M. Laskey, S. Sen, A. Murali, B. Kehoe, S. Patil, J. Wang, M. Franklin, P. Abbeel, and K. Goldberg, "Learning accurate kinematic control of cable-driven surgical robots using data cleaning and Gaussian Process Regression," in *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, 2014, pp. 532–539.

[23] H. Ha, Y. Gao, Z. Fu, J. Tan, and S. Song, "UMI on Legs: Making Manipulation Policies Mobile with Manipulation-Centric Whole-body Controllers," in *Proceedings of the 2024 Conference on Robot Learning*, 2024.

[24] Q. Wu, Z. Fu, X. Cheng, X. Wang, and C. Finn, "Helpful Doggy-Bot: Open-World Object Fetching using Legged Robots and Vision-Language Models," in *arXiv*, 2024.

[25] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, and C. Finn, "HumanPlus: Humanoid Shadowing and Imitation from Humans," in *Conference on Robot Learning (CoRL)*, 2024.

[26] Z. Fu, X. Cheng, and D. Pathak, "Deep Whole-Body Control: Learning a Unified Policy for Manipulation and Locomotion," in *Conference on Robot Learning (CoRL)*, 2022.

[27] Y. Shao, B. Zhang, Q. Liao, X. Huang, Y. Gao, Y. Chi, Z. Li, S. Shao, and K. Sreenath, "LangWBC: Language-directed Humanoid Whole-Body Control via End-to-end Learning," in *Proceedings of Robotics: Science and Systems*, LosAngeles, CA, USA, June 2025.

[28] A. Mousa, N. Karavis, M. Caprio, W. Pan, and R. Allmendinger, "TAR: Teacher-Aligned Representations via Contrastive Learning for Quadrupedal Locomotion," 2025.

[29] L. Ankile, A. Simeonov, I. Shenfeld, M. Torne, and P. Agrawal, "From Imitation to Refinement – Residual RL for Precise Assembly," 2024.

[30] Q. Yu, M. Moghani, K. Dharmarajan, V. Schorp, W. C.-H. Panitch, J. Liu, K. Hari, H. Huang, M. Mittal, K. Goldberg, and A. Garg, "Orbit-Surgical: An Open-Simulation Framework for Learning Surgical Augmented Dexterity," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 15 509–15 516.

[31] 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 *IEEE Intl. Conf. on Robotics and Auto. (ICRA)*, Hong Kong, China, 2014, pp. 6434–6439.

[32] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[33] S. Ross, G. Gordon, and D. Bagnell, "A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, G. Gordon, D. Dunson, and M. Dudík, Eds., vol. 15. Fort Lauderdale, FL, USA: PMLR, 11–13 Apr 2011, pp. 627–635.

[34] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," in *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023.

[35] N. Karaev, I. Rocco, B. Graham, N. Neverova, A. Vedaldi, and C. Rupprecht, "CoTracker: It is Better to Track Together," in *Proc. ECCV*, 2024.