SutureFormer: Learning Surgical Trajectories via Goal-conditioned Offline RL in Pixel Space

Huanrong Liu[1,2,*], Chunlin Tian[1,*], Tongyu Jia[3,*], Tailai Zhou[3,*], Qin Liu[1], Yu Gao[3], Yutong Ban[7], Yun Gu[5,6], Guy Rosman[4,†], Xin Ma[3,†], and Qingbiao Li[1,2,†]

[1] Faculty of Science and Technology, University of Macau, Macau, China
[2] University of Macau Advanced Research Institute in Hengqin, Zhuhai, China
[3] Senior Department of Urology, The Chinese PLA General Hospital, Beijing, China
[4] School of Medicine, Duke University, Durham, North Carolina, USA
[5] Shanghai Key Laboratory of Flexible Medical Robotics, Tongren Hospital, Institute of Medical Robotics, Shanghai Jiao Tong University, Shanghai, China
[6] School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, Shanghai, China
[7] Global College, Shanghai Jiao Tong University, Shanghai, China
qingbiaoli@um.edu.mo
[*] Equal contribution. [†] Corresponding author.

Abstract. Predicting surgical needle trajectories from endoscopic video is critical for robot-assisted suturing, enabling anticipatory planning, real-time guidance, and safer motion execution. However, most surgical recordings lack synchronized robot kinematics, and dense frame-level annotations are prohibitively expensive to obtain from clinical experts. To address these challenges, we propose SutureFormer, a goal-conditioned offline reinforcement learning framework that reformulates trajectory prediction as visual navigation in pixel space. SutureFormer operates solely on raw image sequences under sparse supervision, eliminating the need for kinematic signals or dense annotations. It encodes variable-length video observations using a Spatial CNN and Transformer architecture to capture both local spatial cues and long-range temporal dependencies. It then autoregressively predicts future waypoints within an action space comprising discrete directions and continuous magnitudes. A guidance channel constructed from nine sparse keyframe waypoints provides goal conditioning. To enable stable offline training from expert demonstrations, we adopt Conservative Q-Learning with confidence-weighted rewards and behavioral cloning regularization. We evaluate SutureFormer on a new kidney wound suturing dataset comprising 1,158 trajectories from 50 patients, where it reduces Average Displacement Error by 56.8% compared to the strongest baseline, demonstrating the effectiveness of pixel-level sequential action modeling and the offline RL formulation for surgical trajectory prediction. The code will be available in https://github.com/HaroldHuanrongLIU/MICCAI2026_SutureFormer.

Keywords: Surgical Trajectory Prediction · Offline Reinforcement Learning · Robot-Assisted Surgery · Endoscopic Video Analysis

1 Introduction

Robot-assisted surgery is evolving from tele-operation toward task-level autonomy, where intelligent systems anticipate surgical intent and provide proactive assistance [21,1]. Within this paradigm, the capacity to predict instrument trajectories, particularly during suturing, one of the most technically demanding and outcome-sensitive maneuvers in minimally invasive procedures.

From the perspective of bounded rationality [18], surgeons operate under inherent cognitive constraints: limited attention span, finite working memory, and time pressure collectively bound the optimality of intraoperative decisions. A learning-based trajectory prediction system that distills expert demonstrations into anticipatory guidance can therefore serve as a decision-support mechanism, helping less experienced surgeons approximate expert-level performance and ultimately improving access to high-quality surgical care.

Despite growing interest in deep learning for surgical scene analysis [12], including workflow recognition [9] and scene understanding [13], research on precise procedural assistance for trajectory prediction remains nascent. Current methods face two critical practical barriers. First, most approaches depend on robot kinematic signals, including joint angles, end-effector poses, and gripper states [15,17,20]. Such data are available only on platforms with accessible kinematic interfaces (e.g., the da Vinci Research Kit) and often require direct collaboration with device manufacturers. This dependency fundamentally limits transferability: a policy trained on one robotic platform cannot generalize to another, let alone to the vast archive of conventional laparoscopic video where no kinematic readout exists. In addition, many methods require dense temporal annotations, which are prohibitively expensive to obtain.

Learning control from pixels is a challenging but transformative idea, one that has seen remarkable success in adjacent fields. For instance, purely vision-based imitation learning agents have mastered complex control tasks like driving, learning to navigate and plan by directly mapping image inputs to steering commands [14,3]. Cai et al. [4] demonstrated that cost functions for reinforcement learning (RL) can be learned directly from images, bypassing the need for explicit state estimation. Similarly, Tamar et al. proposed Value Iteration Networks (VIN) [19] that a neural network can learn to perform goal-directed reasoning, generalizing its policy to novel environments by embedding a computational process akin to planning within its architecture. These insights suggest that the "kinematic bottleneck" in surgical trajectory prediction is not a technological inevitability, but a design choice.

To address these limitations, we propose SutureFormer, which reformulates surgical trajectory prediction as goal-conditioned visual navigation in pixel space using offline expert demonstrations. A goal-conditioned policy iteratively predicts future needle waypoints based solely on local visual context and sparse keyframe guidance. The observation encoder combines a Spatial CNN and Transformer to capture spatial cues and long-range temporal dependencies. A discrete-action Conservative Q-Learning (CQL) agent then autoregressively generates trajectory waypoints using a 9 direction action space with continuous step mag-

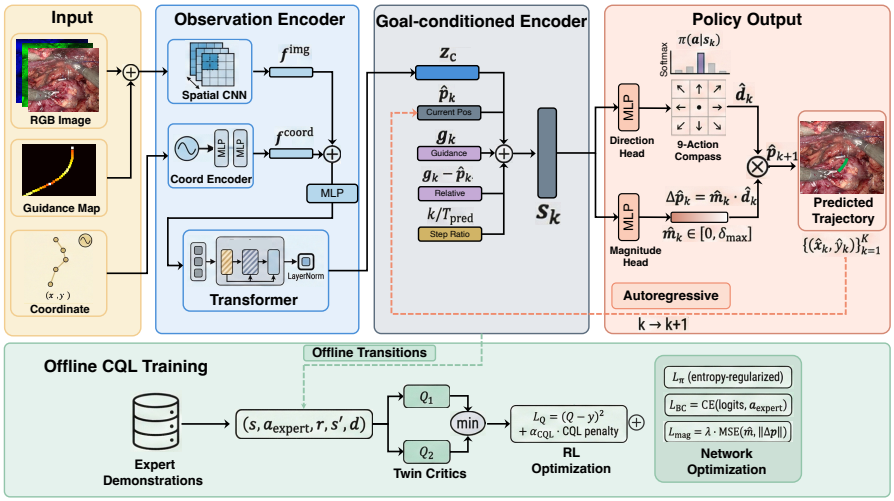

Fig. 1. Overview of the proposed framework. The Observation Encoder fuses visual and coordinate features and aggregates them via a Transformer to produce $\mathbf{z}_c$. The Goal-conditioned Encoder combines $\mathbf{z}_c$ with the current position, guidance target, relative displacement, and step ratio into a unified state $\mathbf{s}_k$. The Policy Output predicts direction and magnitude to autoregressively generate the trajectory. The model is trained entirely offline via Conservative Q-Learning with twin critics and supplementary behavior cloning and magnitude supervision losses.

nitudes, conditioned on keyframe goals during training and polynomial extrapolation at inference. In practice, SutureFormer requires only endoscopic video frames and 9 sparse keyframes annotations per trajectory. To our knowledge, this is the first framework to formulate surgical trajectory prediction as visual navigation task with offline RL in pixel space. Our method outperforms current state-of-the-art methods, demonstrating the viability of offline RL for surgical trajectory prediction.

2 Methods

Problem Statement. Given a variable-length observed video sequence $V_{0:T_P} = \{I_t\}_{t=0}^{T_P}$ with corresponding needle tip coordinates $\{(x_t, y_t)\}_{t=0}^{T_P}$, our goal is to predict the remaining trajectory $\{(\hat{x}_t, \hat{y}_t)\}_{t=T_P+1}^{T_P+T_F}$, where T_P and T_F denote the number of observed and predicted frames, respectively. We formulate this task as a goal-conditioned Markov Decision Process and solve it using offline CQL. Fig. 1 illustrates the overall architecture, which employs an encoder-decoder framework where the decoder functions as a pixel-space reinforcement learning trajectory generator conditioned on the past frames.

2.1 Observation Encoder

Each frame is represented by a 128×128 local crop extracted from the full image and centered at the needle tip. The crop is augmented with a guidance channel, a single-channel heatmap encoding the trajectory path confidence at each pixel, and concatenated with the RGB data to form a 4-channel input. The observation encoder comprises: (i) a SpatialCNN extracts features from all observed input frames $\{I_t\}_{t=0}^{T_P}$; (ii) a sinusoidal coordinate encoder that maps all 2D coordinate $\{(x_t, y_t)\}_{t=0}^{T_P}$, normalized to the range $[0, 1]$, then subsequently transformed by a two-layer MLP; and (iii) a Transformer that aggregates the variable-length observation sequence.

Specifically, for each trajectory, an image feature $\mathbf{f}^{\text{img}}$ and a coordinate feature $\mathbf{f}^{\text{coord}}$ are extracted from historical observation, concatenated, and projected into the Transformer dimension of d_{model}. The Transformer employs a causal attention mask and a padding mask to support variable observation lengths T_{obs}. Finally, LayerNorm is applied to the last valid output token to yield the contextual representation $\mathbf{z}_{\text{c}}$.

2.2 Goal-conditioned Encoder

At each prediction step k, a state representation is constructed by combining the observation context with the current navigation state:

$$\mathbf{s}_k = \phi\big(\mathbf{z}_{\text{c}},\ \hat{\mathbf{p}}_k,\ \mathbf{g}_k,\ \mathbf{g}_k - \hat{\mathbf{p}}_k,\ k/T_{\text{pred}}\big) \tag{1}$$

where $\mathbf{g}_k$ denotes the guidance coordinate at step k. Three independent sinusoidal coordinate encoders are utilized to map the current position $\hat{\mathbf{p}}_k$, the guidance target $\mathbf{g}_k$, and the relative displacement $\mathbf{g}_k - \hat{\mathbf{p}}_k$. Additionally, a linear layer maps the scalar step ratio k/T_{pred} to a feature space. The concatenation ϕ of these encoded features with the contextual representation $\mathbf{z}_{\text{c}}$ yields the final state vector $\mathbf{s}_k$.

The proposed method employs a discrete 9-direction action space. A direction head outputs the logits over these 9 actions, whereas a magnitude head predicts a continuous scalar step magnitude $\hat{m}_k \in [0, \delta_{\max}]$. During inference, the predicted displacement is formulated as:

$$\hat{\mathbf{d}}_k = \sum_{a=1}^{9} \pi(a|\mathbf{s}_k)\, \mathbf{u}_a \Delta\hat{\mathbf{p}}_k = \hat{m}_k\, \hat{\mathbf{d}}_k, \qquad \hat{\mathbf{p}}_{k+1} = \text{clip}(\hat{\mathbf{p}}_k + \Delta\hat{\mathbf{p}}_k, 0, 1) \tag{2}$$

where $\hat{\mathbf{d}}_k = \sum_{a=1}^{9} \pi(a|\mathbf{s}_k)\, \mathbf{u}_a$ represents the softmax-weighted direction vector that produces smooth motion, and $\mathbf{u}_a$ denotes the unit vector corresponding to action a.

During training, the guidance coordinates $\mathbf{g}_k$ are constructed by placing confidence-weighted annotated trajectory points within the 128×128 crop, applying dilation via a 3×3 max filter, and concatenating the result with the RGB data to form a 4-channel input. During testing, due to the unavailability

of future trajectory points, pseudo-guidance is generated via polynomial extrapolation from the observed points. All the positions are clamped to the range of $[0, 1]^2$ to avoid exceeding the normalized image boundary.

2.3 Offline Conservative Q-Learning

SutureFormer is trained entirely offline using a static dataset of expert demonstrations. During the training phase, the ground-truth trajectories are traversed to extract expert transitions. At each step k, the spatial displacement $\mathbf{p}_{k+1} - \mathbf{p}_k$ is discretized into the nearest compass direction via cosine similarity (assigned an idle action if $\|\mathbf{p}_{k+1} - \mathbf{p}_k\| < 10^{-6}$), and the corresponding expert magnitude is defined as $m_k^* = \|\mathbf{p}_{k+1} - \mathbf{p}_k\|_2$.

The proposed framework adopts Conservative Q-Learning (CQL) [10], which incorporates a penalty into the standard Bellman backup to mitigate the overestimation of the Q-function for out-of-distribution actions. By penalizing the Q-values associated with unobserved actions, this approach ensures that the learned policy remains proximate to the demonstrated expert behavior. The critic loss for a given transition $(\mathbf{s}, a_{\text{expert}}, r, \mathbf{s}', d)$ is formulated as

$$\mathcal{L}_Q = (Q(\mathbf{s}, a_{\text{expert}}) - y)^2 + \alpha_{\text{CQL}} \cdot \left(\log \sum_{a'} \exp Q(\mathbf{s}, a') - Q(\mathbf{s}, a_{\text{expert}}) \right) \quad (3)$$

where $y = r + \gamma^n (1 - d) \cdot V(\mathbf{s}')$ represents the n-step target with a discount factor $\gamma = 0.95$, whereas the parameter $\alpha_{\text{CQL}} = 0.01$ controls the strength of the conservative regularization. The maximum-entropy soft value target $V(\mathbf{s}')$ is computed as:

$$V(\mathbf{s}') = \sum_{a'} \pi(a'|\mathbf{s}') \left[\min\big(Q_1^{\text{tgt}}(\mathbf{s}', a'), Q_2^{\text{tgt}}(\mathbf{s}', a')\big) - \alpha \log \pi(a'|\mathbf{s}') \right] \quad (4)$$

where $\alpha = 0.2$ denotes a fixed entropy temperature. Furthermore, the actor is optimized via an entropy-regularized objective function utilizing the estimated Q-values:

$$\mathcal{L}_\pi = \mathbb{E}_{\mathbf{s}} \left[\sum_a \pi(a \mid \mathbf{s}) \left(\alpha \log \pi(a \mid \mathbf{s}) - \min(Q_1(\mathbf{s}, a),\, Q_2(\mathbf{s}, a)) \right) \right] \quad (5)$$

This objective is supplemented by a behavior cloning loss $\mathcal{L}_{\text{BC}} = \text{CE}(\text{logits}, a_{\text{expert}})$ with a weighting coefficient $\lambda_{\text{BC}} = 1.0$, and a supervised magnitude loss $\mathcal{L}_{\text{mag}} = \lambda_{\text{mag}} \cdot \text{MSE}(\hat{m}, \|\mathbf{p}_{\text{target}} - \mathbf{p}_{\text{current}}\|_2)$ with $\lambda_{\text{mag}} = 100$.

We note a critical design choice of separating gradient flows. Specifically, the Q-networks receive states with detached gradients to prevent the propagation of gradients into the observation encoder. Conversely, the losses associated with the actor and the step magnitude allow gradients to propagate through the state encoder and into the contextual representation $\mathbf{z}_c$. This configuration ensures that the objectives of trajectory prediction, rather than those of value estimation, dictate the formation of the visual representations.

2.4 Reward Design with Sparse Keyframe Supervision

At each prediction step k, the reward comprises three components: a constant time penalty $r_{\text{time}} = -0.01$ encouraging efficient trajectories, a confidence-weighted proximity reward based on the distance d_k to the ground truth, and an exponential terminal bonus at the final step. The proximity reward is positive (up to $r_{\text{prox,max}} = 0.5$) when d_k falls within a threshold of 0.02 in normalized coordinates, and negative otherwise. Since clinical experts annotate only 9 keyframes per trajectory while intermediate positions are obtained via temporal interpolation, we apply confidence-based weighting: keyframe steps receive full weight ($w = 1.0$), whereas interpolated steps are down-weighted proportionally ($w = 0.5 + 0.5 \cdot confidence$, see detail in Sec.3), providing dense training signals while reflecting the lower certainty of non-keyframe positions. To handle variable-length observation and prediction sequences, we employ masked attention within the Transformer encoder to selectively attend to valid time steps while ignoring padded positions. Furthermore, we use bucketed batch sampling, which groups sequences of similar lengths into the same batch, thereby minimizing padding overhead and improving training efficiency.

3 Experiments

Datasets. We evaluate SutureFormer on real clinical dataset of robotic-assisted laparoscopic kidney wound suturing operated by expert surgeon. The dataset comprises surgical videos from 50 patients and contains a total of 1,158 trajectories, where clinical experts annotate 9 keyframes within each trajectory. The dataset is partitioned at the patient level into training, validation, and test sets, consisting of 35, 8, and 7 patients (corresponding to 861, 151, and 146 trajectories), respectively.

Interpolated Annotation. To obtain dense per-frame annotations, we independently apply cubic spline interpolation [5] to the x and y coordinate sequences as functions of the frame index. Given the 9 keyframe positions, we fit two natural cubic spline functions, $S_x(t)$ and $S_y(t)$, evaluating them at every intermediate frame within the temporal range. The interpolated coordinates are rounded to the nearest integer to yield valid pixel locations, without extrapolation beyond the first and last keyframes. Each interpolated frame receives a confidence score from 0.45 to 0.9 based on its temporal proximity to the nearest keyframe, assigning higher scores to closer frames. The expert annotated keyframes retain a confidence score of 1.0.

Implementation Details. The model is trained for 100 epochs with batch size 8 on an NVIDIA RTX 5090 GPU. We use four Adam optimizers with cosine annealing decaying to 1% of the initial learning rates: 1×10^{-4} for the observation encoder and 3×10^{-4} for the actor, critic, and magnitude heads. CQL hyperparameters are $\alpha_{\text{CQL}} = 0.01$, $\gamma = 0.95$, and $n = 3$-step returns, with soft target updates $\tau = 0.005$. Policy and magnitude updates subsample up to 2,048 transitions per batch.

Table 1. Quantitative comparison on the test set under two setting (Obs = 6, Pred = 3 and Obs = 3, Pred = 6). Specifically, Obs = 6 denotes an observed surgical trajectory comprising 6 annotated keyframes, whereas Pred = 3 represents the prediction of the future trajectory containing 3 keyframes. All metrics are evaluated in pixel space (lower is better), with the best results highlighted in **bold** and second best in **underline**.

Method	Obs = 6, Pred = 3			Obs = 3, Pred = 6		
	ADE (↓)	FDE (↓)	FD (↓)	ADE (↓)	FDE (↓)	FD (↓)
BC [2]	<u>128.15</u>	<u>146.24</u>	<u>156.83</u>	<u>137.06</u>	<u>184.69</u>	<u>195.13</u>
GAIL [8]	269.79	282.99	305.86	249.19	254.41	318.82
IBC [6]	243.97	262.12	285.60	225.74	260.49	296.82
iDiff-IL [11]	187.38	207.32	220.21	223.75	259.90	296.62
CondDiff [11]	165.20	189.47	200.31	184.15	236.84	255.62
MID [7]	151.23	165.24	192.79	172.67	229.83	246.65
Ours	**55.29**	**84.23**	**86.00**	**94.85**	**154.68**	**158.46**

Evaluation Metrics. All metrics are computed in pixel space by rescaling normalized coordinates to the original 1264×902 resolution. We report Average Displacement Error (ADE), the mean Euclidean distance across all prediction steps; Final Displacement Error (FDE), the Euclidean distance at the last step; and discrete Fréchet Distance (FD), measuring global trajectory shape similarity. Lower values indicate better performance for all three metrics.

4 Results

4.1 Comparison with Baselines

We compare SutureFormer against following baselines: (1) Behavioral Cloning (BC) [2], which encodes stacked frames via a U-Net [16] and regresses future coordinates with an MLP; (2) Generative Adversarial Imitation Learning (GAIL) [8], which generates trajectories conditioned on a random latent vector with a discriminator distinguishing real from generated pairs; (3) Implicit Behavioral Cloning (IBC) [6], an energy based model trained with InfoNCE loss and sampled via Langevin MCMC at inference; (4) Implicit Diffusion Policy (iDiff-IL) [11], a joint image trajectory diffusion method using a dual-head UNet to denoise both spaces simultaneously; (5) Conditional Diffusion Policy (CondDiff) [11], a variant applying DDPM denoising in trajectory space only, conditioned on the image; and (6) Motion Indeterminacy Diffusion (MID) [7], a latent-space diffusion method that encodes frames into a context vector and applies a Transformer-based DDPM denoiser for trajectory generation.

Quantitative Results. Table 1 reports results under two experimental settings. SutureFormer consistently outperforms all baselines across all metrics.

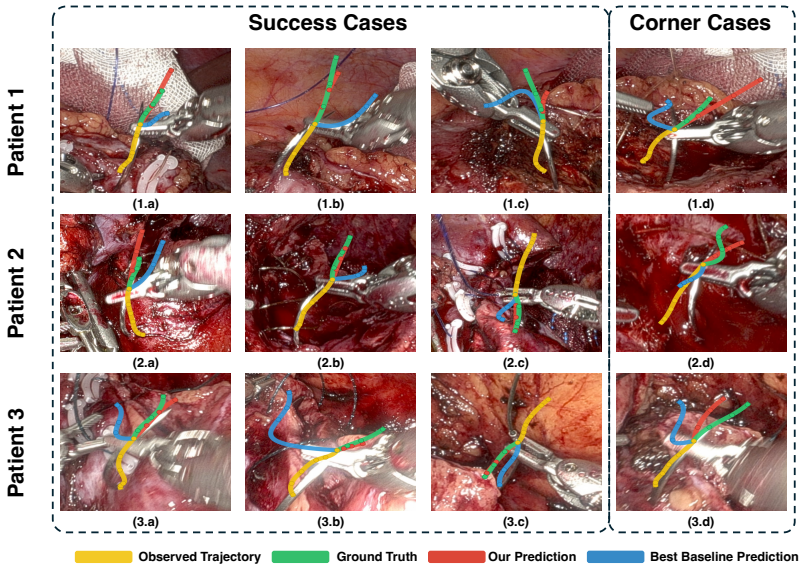

Fig. 2. Qualitative comparison of predicted trajectories on the testset. The **yellow** curve denotes the observed trajectory, the **green** curve represents the ground truth future trajectory, the **red** curve shows the prediction from our SutureFormer and the **blue** curve indicates the best baseline prediction.

With Obs=6, Pred=3, SutureFormer achieves an ADE of 55.29 (56.8% reduction over BC at 128.15) and FD of 84.23 (vs. 146.24 for BC), indicating substantially better trajectory shape fidelity. Under the more challenging Obs=3, Pred=6 setting, all methods degrade as historical observational information decreases, yet ours remains the best, achieving ADE 94.85, FDE 154.68 and FD 158.46. SutureFormer maintains clear advantages in ADE (30.7% reduction over BC), FDE (16.2% reduction) and FD (18.7% reduction), demonstrating that the CQL-based policy produces more accurate trajectories with better shape consistency even under sparse observations. The goal-conditioned navigation formulation proves particularly beneficial when visual context is limited, as explicit target guidance compensates for reduced observational evidence. To further isolate the contribution of conservative offline reinforcement learning, we remove CQL from the training objective and observe 23.5%, 18.6%, and 18.1% degradation in ADE, FDE, and FD, respectively. This indicates that the performance gain is not merely attributable to the architecture or preprocessing, but is substantially supported by conservative offline RL.

Qualitative Results. Fig. 2 illustrates representative predicted trajectories overlaid on surgical images, alongside the corresponding ground-truth paths and prediction results from selected state-of-art baseline method. As observed across diverse suturing scenarios from multiple patients, the baseline method frequently struggles to capture the complex spatial dynamics of the surgical instruments,

resulting in predicted trajectories that significantly deviate from the true paths. In contrast, our proposed approach consistently maintains high fidelity to the ground truth. The proposed model generates smooth trajectories that conform to the general curvature of the suturing path.

5 Conclusion

This paper introduces SutureFormer, a novel framework that reformulates surgical trajectory prediction in pixel space based on goal-conditioned offline reinforcement learning using Conservative Q-Learning. By requiring only 9 keyframe annotations per trajectory and operating without robot kinematics, the framework demonstrates broad applicability to existing clinical video archives. Experimental results on 1,158 trajectories demonstrate that SutureFormer significantly outperforms best diffusion and imitation learning baselines, achieving up to a 56.8% reduction in ADE. Future work will extend SutureFormer to broader laparoscopic procedures and validate its performance through ex-vivo porcine experiments on robotic platforms, facilitating its transition toward precise, advancing its translation toward real-world cognitive surgical assistance.

Acknowledgments. This work was supported by the University of Macau under Grants SRG2024-00056-FST, 0078/2024/RIB2, and FST/SP01/2024, the Dr. Stanley Ho Medical Development Foundation under Grant SHMDF-AI/2026/002, the Natural Science Foundation of China under Grant 62373243, the Shanghai Municipal Health Commission Smart Healthcare Project under Grant 2025ZHYL021, and the Shanghai Tongren Hospital Medical-Engineering Collaboration Project under Grant lhyjzx2024-xm03. The surgical records of robot-assisted partial nephrectomy used in this work were collected by the Chinese PLA General Hospital and annotated by two expert surgeons.

Disclosure of Interests. The authors have no competing interests.

References

1. Attanasio, A., Scaglioni, B., De Momi, E., Fiorini, P., Valdastri, P.: Autonomy in surgical robotics. Annual Review of Control, Robotics, and Autonomous Systems **4**(Volume 4, 2021), 651–679 (2021). https://doi.org/https://doi.org/10.1146/annurev-control-062420-090543, https://www.annualreviews.org/content/journals/10.1146/annurev-control-062420-090543
2. Bain, M., Sammut, C.: A framework for behavioural cloning. In: Machine Intelligence 15, Intelligent Agents [St. Catherine's College, Oxford, July 1995]. p. 103–129. Oxford University, GBR (1999)
3. Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L.D., Monfort, M., Muller, U., Zhang, J., et al.: End to end learning for self-driving cars. arXiv preprint arXiv:1604.07316 (2016)

4. Cai, P., Wang, H., Huang, H., Liu, Y., Liu, M.: Vision-based autonomous car racing using deep imitative reinforcement learning. IEEE Robotics and Automation Letters **6**(4), 7262–7269 (2021)
5. De Boor, C., De Boor, C.: A practical guide to splines, vol. 27. springer New York (1978)
6. Florence, P., Lynch, C., Zeng, A., Ramirez, O.A., Wahid, A., Downs, L., Wong, A., Lee, J., Mordatch, I., Tompson, J.: Implicit behavioral cloning. In: Faust, A., Hsu, D., Neumann, G. (eds.) Proceedings of the 5th Conference on Robot Learning. PMLR, vol. 164, pp. 158–168. PMLR (08–11 Nov 2022), https://proceedings.mlr.press/v164/florence22a.html
7. Gu, T., Chen, G., Li, J., Lin, C., Rao, Y., Zhou, J., Lu, J.: Stochastic trajectory prediction via motion indeterminacy diffusion. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17092–17101 (2022). https://doi.org/10.1109/CVPR52688.2022.01660
8. Ho, J., Ermon, S.: Generative adversarial imitation learning. In: NIPS. p. 4572–4580. NIPS'16, Curran Associates Inc., Red Hook, NY, USA (2016)
9. Jin, Y., Long, Y., Gao, X., Stoyanov, D., Dou, Q., Heng, P.A.: Trans-svnet: hybrid embedding aggregation transformer for surgical workflow analysis. International Journal of Computer Assisted Radiology and Surgery **17**(12), 2193–2202 (2022)
10. Kumar, A., Zhou, A., Tucker, G., Levine, S.: Conservative Q-learning for offline reinforcement learning. In: NeurIPS. Curran Associates Inc., Red Hook, NY, USA (2020)
11. Li, J., Jin, Y., Chen, Y., Yip, H.C., Scheppach, M., Chiu, P.W.Y., Yam, Y., Meng, H.M.L., Dou, Q.: Imitation learning from expert video data for dissection trajectory prediction in endoscopic surgical procedure. In: MICCAI. p. 494–504. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-43996-4_47, https://doi.org/10.1007/978-3-031-43996-4_47
12. Maier-Hein, L., Vedula, S.S., Speidel, S., Navab, N., Kikinis, R., Park, A., Eisenmann, M., Feussner, H., Forestier, G., Giannarou, S., et al.: Surgical data science for next-generation interventions. Nature Biomedical Engineering **1**(9), 691–696 (2017)
13. Nwoye, C.I., Yu, T., Gonzalez, C., Seeliger, B., Mascagni, P., Mutter, D., Marescaux, J., Padoy, N.: Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. Medical Image Analysis **78**, 102433 (2022)
14. Pomerleau, D.A.: ALVINN: an autonomous land vehicle in a neural network. In: NIPS. p. 305–313. NIPS'88, MIT Press, Cambridge, MA, USA (1988)
15. Qin, Y., Feyzabadi, S., Allan, M., Burdick, J.W., Azizian, M.: davincinet: Joint prediction of motion and surgical state in robot-assisted surgery. In: IROS. pp. 2921–2928 (2020). https://doi.org/10.1109/IROS45743.2020.9340723
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
17. Shi, C., Zheng, Y., Fey, A.M.: Recognition and prediction of surgical gestures and trajectories using transformer models in robot-assisted surgery. In: IROS. pp. 8017–8024 (2022). https://doi.org/10.1109/IROS47612.2022.9981611
18. Simon, H.A., et al.: Theories of bounded rationality. Decision and organization **1**(1), 161–176 (1972)
19. Tamar, A., Wu, Y., Thomas, G., Levine, S., Abbeel, P.: Value iteration networks. Advances in neural information processing systems **29** (2016)

20. Weerasinghe, K., Reza Roodabeh, S.H., Hutchinson, K., Alemzadeh, H.: Multimodal transformers for real-time surgical activity prediction. In: ICRA. pp. 13323–13330 (2024). https://doi.org/10.1109/ICRA57147.2024.10611048
21. Yang, G.Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P.E., Hata, N., Kazanzides, P., Martel, S., Patel, R.V., et al.: Medical robotics—regulatory, ethical, and legal considerations for increasing levels of autonomy (2017)