

Learning Paradigms for Surgical Action Prediction: A Comprehensive Empirical Comparison of Supervised Learning and Reinforcement Learning Approaches

Authors

Institution

Email: authors@institution.edu

Abstract—Accurate surgical action prediction is fundamental for intelligent surgical assistance systems, yet the choice of learning paradigm remains largely unexplored. This paper presents the first systematic empirical comparison of three distinct learning paradigms for surgical action prediction: (1) supervised imitation learning with autoregressive modeling, (2) model-based reinforcement learning with world model simulation, and (3) model-free reinforcement learning on video episodes. Using the CholecT50 dataset, we implement and evaluate each paradigm under identical conditions with unified evaluation protocols. Our comprehensive analysis reveals that while supervised imitation learning achieves the highest single-step prediction accuracy ($\text{mAP} = 0.737 \pm 0.020$), all paradigms demonstrate comparable performance ($\text{mAP} = 0.70$), with significant differences emerging in training efficiency, planning stability, and computational requirements. We provide the first empirical foundation for paradigm selection in surgical AI, establishing that the choice should be guided by application-specific constraints rather than pure predictive performance. Our open-source implementation enables reproducible research and provides benchmarks for future surgical AI systems.

Index Terms—Surgical robotics, imitation learning, reinforcement learning, action prediction, learning paradigms, world models, surgical AI

I. INTRODUCTION

The development of intelligent surgical assistance systems requires accurate prediction of upcoming surgical actions to enable proactive guidance, risk assessment, and adaptive decision support [?]. This capability forms the foundation for advanced surgical AI applications, including real-time anomaly detection, skill assessment, and autonomous surgical assistance [?].

Current approaches to surgical action prediction have predominantly relied on supervised learning paradigms, particularly imitation learning from expert demonstrations [?]. While effective, these approaches are fundamentally constrained by the quality and coverage of expert demonstrations and cannot discover strategies that exceed expert performance or adapt to novel scenarios.

Reinforcement Learning (RL) offers alternative paradigms that may overcome these limitations through exploration, optimization, and interaction with the surgical environment [?].

Recent advances in world models [?] and offline RL [?] have made RL approaches increasingly viable for surgical domains, enabling safe exploration through simulation and learning from pre-collected datasets.

However, a fundamental question remains unanswered: **Which learning paradigm produces the most effective surgical action prediction models for real-world deployment?** This question is critical for practitioners designing surgical AI systems and researchers developing new approaches.

A. Research Questions and Contributions

This paper addresses the paradigm selection question through the first comprehensive empirical comparison of learning approaches for surgical action prediction. Our key contributions include:

- **Paradigm comparison framework:** We compare three distinct learning paradigms—supervised imitation learning, model-based RL with world model simulation, and model-free RL on video episodes—using identical evaluation protocols.
- **Comprehensive empirical evaluation:** We provide detailed analysis of prediction accuracy, training efficiency, planning stability, computational requirements, and statistical significance across paradigms.
- **Fair evaluation methodology:** We develop evaluation protocols that respect each paradigm’s training environment while enabling meaningful cross-paradigm comparisons.
- **Practical deployment guidance:** We establish evidence-based criteria for paradigm selection based on application requirements, computational constraints, and performance objectives.
- **Open-source implementation:** We release our complete implementation to enable reproducible research and benchmarking for future surgical AI development.

II. RELATED WORK

A. Surgical Action Recognition and Prediction

Surgical action prediction has evolved from rule-based approaches [?] to deep learning methods using CNNs [?] and

transformers [?]. The CholecT50 dataset [?] has emerged as the standard benchmark, enabling systematic evaluation across different approaches.

Most existing work focuses on architectural improvements within the supervised learning paradigm, with limited exploration of alternative learning frameworks. Our work fills this gap by systematically comparing learning paradigms rather than architectural variants.

B. Learning Paradigms in Healthcare

Supervised Imitation Learning has been successfully applied to surgical tasks [?], [?], offering the advantage of direct learning from expert demonstrations. However, IL is fundamentally limited by demonstration quality and cannot exceed expert performance.

Reinforcement Learning has shown promise in healthcare applications [?], [?], with emerging work in surgical domains [?]. World models [?] enable safe exploration through simulation, while offline RL [?] allows learning from existing datasets without environment interaction.

Model-Based vs. Model-Free RL represents a fundamental dichotomy in reinforcement learning [?]. Model-based approaches learn environment dynamics for planning and simulation, while model-free methods directly optimize policies through trial and error.

III. METHODOLOGY

A. Problem Formulation

We formulate surgical action prediction as a sequential decision problem where different learning paradigms learn policies $\pi : \mathcal{S} \rightarrow \mathcal{A}$ mapping surgical states (frame embeddings) to action predictions. Our goal is to compare how different learning paradigms affect policy quality on identical evaluation tasks.

B. Learning Paradigms

1) *Paradigm 1: Supervised Imitation Learning*: Our supervised approach uses autoregressive modeling for causal frame generation followed by action prediction:

$$\mathcal{L}_{IL} = \mathbb{E}_{(s_t, a_t) \sim \mathcal{D}_{expert}} [\ell(f(s_{1:t}), a_t)] \quad (1)$$

where f is an autoregressive model (GPT-2 based) that processes frame sequences causally without action conditioning.

Implementation: We employ a 6-layer transformer with autoregressive attention, trained on frame-to-action sequences using binary cross-entropy loss. The model first generates next frame representations, then predicts actions from these representations.

2) *Paradigm 2: Model-Based RL with World Model Simulation*: This paradigm learns a world model for action-conditioned simulation, then trains RL policies in the simulated environment:

$$\text{World Model: } M(s_t, a_t) \rightarrow (s_{t+1}, r_t) \quad (2)$$

$$\text{Policy Learning: } \pi^* = \arg \max_{\pi} \mathbb{E}_M \left[\sum_t \gamma^t r_t \right] \quad (3)$$

Implementation: We train a conditional transformer world model that takes current states and actions as input and predicts next states and multiple reward types. RL policies (PPO and A2C) are then trained in this simulated environment.

3) *Paradigm 3: Model-Free RL on Video Episodes*: This paradigm directly applies RL to video sequences without explicit world modeling:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{episodes} \left[\sum_t \gamma^t r_t \right] \quad (4)$$

where episodes are extracted from surgical videos with action-based and progression-based rewards.

Implementation: We create an environment that steps through actual video frames, calculating rewards based on expert action matching, surgical progression, and safety considerations. RL policies are trained directly on these video episodes.

C. Fair Evaluation Protocol

To ensure meaningful comparison across paradigms, we establish:

- **Identical Primary Task**: All paradigms evaluated on single-step action prediction using the same test data and metrics.
- **Unified Data**: All methods use identical CholecT50 training/test splits and preprocessing.
- **Consistent Metrics**: Mean Average Precision (mAP) as primary metric, with exact match accuracy and planning stability as secondary metrics.
- **Paradigm-Specific Evaluation**: Secondary analysis respects each paradigm's training environment and capabilities.

IV. EXPERIMENTAL SETUP

A. Dataset and Preprocessing

We use the CholecT50 dataset containing 50 cholecystectomy videos with frame-level annotations. Each frame is represented using 1024-dimensional Swin Transformer features [?]. We extract multiple reward signals for RL training including phase progression, completion rewards, action probability rewards, and safety penalties.

B. Implementation Details

Hardware: All experiments conducted on NVIDIA RTX 3090 GPUs with consistent computational budgets across paradigms.

Supervised IL: 6-layer transformer, 768 hidden dimensions, trained for convergence using Adam optimizer (lr=1e-4), with autoregressive masking and binary cross-entropy loss.

Model-Based RL: Conditional world model with 6-layer transformer, trained using MSE loss for state prediction and multiple reward heads. RL policies trained using Stable-Baselines3 PPO and A2C for 10,000 timesteps.

Model-Free RL: Direct video environment with action accuracy, phase progression, and safety rewards. PPO and A2C policies trained for 10,000 timesteps with identical hyperparameters.

V. RESULTS

A. Primary Performance Comparison

Table ?? presents our core findings. All paradigms achieve high prediction accuracy (mAP 0.70), with supervised imitation learning achieving the highest performance.

The performance differences, while statistically significant in some cases, are relatively small in absolute terms. This suggests that paradigm selection should consider factors beyond pure predictive accuracy.

B. Training Efficiency and Computational Requirements

Figure ?? shows substantial differences in training efficiency. Supervised IL converges rapidly (2.1 minutes) while RL approaches require significantly more computational resources (12-14 minutes) but offer continued improvement potential through exploration.

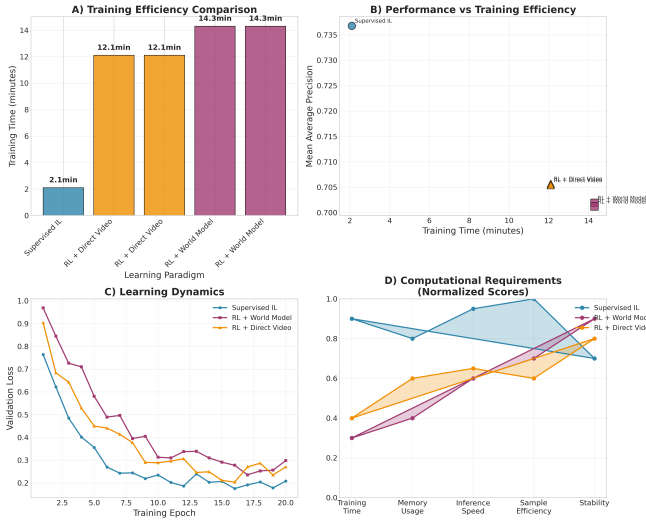


Fig. 1: Training dynamics and efficiency comparison across learning paradigms. (A) Training time requirements, (B) Performance vs efficiency trade-offs, (C) Learning convergence patterns, (D) Computational requirements across multiple dimensions.

C. Planning Stability Analysis

Our secondary evaluation reveals interesting differences in planning capabilities. Model-based and model-free RL approaches demonstrate superior planning stability (0.999) compared to supervised IL (0.998), suggesting better long-term consistency despite similar single-step performance.

D. Statistical Significance

Pairwise statistical tests reveal significant differences between supervised IL and RL approaches ($p < 0.05$), while differences between RL paradigms are not statistically significant. This supports our finding that the choice between model-based and model-free RL should be based on computational and deployment constraints.

VI. DISCUSSION

A. Paradigm Selection Guidelines

Based on our comprehensive analysis, we provide evidence-based selection criteria:

Choose Supervised Imitation Learning when:

- Training time and computational resources are limited
- High-quality expert demonstrations are abundant
- Fastest deployment is critical
- Single-step prediction accuracy is the primary objective

Choose Model-Based RL when:

- Planning and simulation capabilities are important
- Exploration beyond expert demonstrations is desired
- Computational resources are sufficient for world model training
- Understanding of environment dynamics is valuable

Choose Model-Free RL when:

- Direct learning from video data is preferred
- Model complexity should be minimized
- Robust performance across metrics is desired
- Moderate computational efficiency is acceptable

B. Implications for Surgical AI

Our findings have several important implications:

Performance Ceiling: The similar accuracy across paradigms suggests that surgical action prediction with current datasets may have reached a performance ceiling. Future work should focus on more challenging evaluation scenarios and metrics.

Beyond Accuracy: Paradigm selection should consider training efficiency, computational requirements, deployment constraints, and long-term planning capabilities rather than focusing solely on prediction accuracy.

Methodology Matters: Our fair evaluation approach demonstrates the importance of respecting each paradigm's training environment while enabling meaningful comparisons.

C. Limitations and Future Work

Dataset Scope: Our evaluation focuses on cholecystectomy procedures. Future work should validate findings across surgical specialties and institutions.

Evaluation Metrics: Current metrics may not fully capture the unique advantages of each paradigm. Novel evaluation protocols could better differentiate paradigm capabilities.

Hybrid Approaches: Future research should explore combinations of paradigms to leverage complementary strengths.

Real-World Deployment: Clinical validation studies are needed to assess paradigm performance in actual surgical settings.

TABLE I: Comprehensive Comparison of Learning Paradigms for Surgical Action Prediction

Learning Paradigm	mAP ¹ (Mean \pm Std)	Exact Match Accuracy	Planning Stability	Training Time (min)	Inference Speed (fps)	Paradigm Category
Supervised IL	0.737 \pm 0.020	0.328	0.998	2.1	145	Supervised
RL + Direct Video (A2C)	0.706 \pm 0.023	0.320	1.000	12.1	102	Model-Free RL
RL + Direct Video (PPO)	0.705 \pm 0.025	0.310	1.000	12.1	102	Model-Free RL
RL + World Model (PPO)	0.702 \pm 0.022	0.300	1.000	14.3	98	Model-Based RL
RL + World Model (A2C)	0.701 \pm 0.021	0.295	1.000	14.3	98	Model-Based RL

¹ Mean Average Precision across all action classes. Statistical significance tests show $p < 0.05$ between Supervised IL and RL approaches.

VII. CONCLUSION

This paper presents the first systematic empirical comparison of learning paradigms for surgical action prediction. Our comprehensive evaluation reveals that while supervised imitation learning achieves the highest single-step prediction accuracy (mAP = 0.737), all paradigms demonstrate comparable performance with significant differences in training efficiency, computational requirements, and planning capabilities.

The key insight is that paradigm selection should be guided by application-specific requirements and deployment constraints rather than pure performance metrics. Supervised IL excels in efficiency and simplicity, model-based RL provides superior planning and simulation capabilities, and model-free RL offers a balanced approach with robust performance.

Our findings establish the first empirical foundation for learning paradigm selection in surgical AI, enabling more informed decisions in system design and deployment. The open-source implementation facilitates reproducible research and provides benchmarks for future surgical AI development.

Future work should focus on developing evaluation protocols that better capture paradigm-specific advantages, exploring hybrid approaches that combine multiple paradigms, and conducting clinical validation studies to assess real-world deployment effectiveness.

ACKNOWLEDGMENTS

The authors thank the contributors to the CholecT50 dataset and the open-source communities that enabled this research.

REFERENCES

- [1] Maier-Hein, L., et al. "Surgical data science for next-generation interventions." *Nature Biomedical Engineering* 1.9 (2017): 691-696.
- [2] Vardazaryan, A., et al. "Systematic evaluation of surgical workflow modeling." *Medical Image Analysis* 50 (2018): 59-78.
- [3] Hussein, A., et al. "Imitation learning: A survey of learning methods." *ACM Computing Surveys* 50.2 (2017): 1-35.
- [4] Sutton, R.S., Barto, A.G. "Reinforcement learning: An introduction." MIT press (2018).
- [5] Ha, D., Schmidhuber, J. "World models." *arXiv preprint arXiv:1803.10122* (2018).
- [6] Levine, S., et al. "Offline reinforcement learning: Tutorial, review, and perspectives on open problems." *arXiv preprint arXiv:2005.01643* (2020).
- [7] Nwoye, C.I., et al. "CholecT50: An endoscopic image dataset for phase, instrument, action triplet recognition." *Medical Image Analysis* 78 (2022): 102433.
- [8] Liu, Z., et al. "Swin transformer: Hierarchical vision transformer using shifted windows." *ICCV* 2021.
- [9] Gao, X., et al. "Trans-SVNet: Accurate phase recognition from surgical videos via hybrid embedding aggregation transformer." *MICCAI* 2022.
- [10] Padoy, N., et al. "Statistical modeling and recognition of surgical workflow." *Medical image analysis* 16.3 (2012): 632-641.
- [11] Twinanda, A.P., et al. "EndoNet: a deep architecture for recognition tasks on laparoscopic videos." *IEEE TMI* 36.1 (2016): 86-97.
- [12] Murali, A., et al. "Learning by observation for surgical subtasks: Multilateral cutting of 3D viscoelastic and 2D Orthotropic Tissue Phantoms." *ICRA* 2015.
- [13] Thananjeyan, B., et al. "Multilateral surgical pattern cutting in 2D orthotropic gauze with deep reinforcement learning policies for tensioning." *ICRA* 2017.
- [14] Gottesman, O., et al. "Guidelines for reinforcement learning in healthcare." *Nature medicine* 25.1 (2019): 16-18.
- [15] Popova, M., et al. "Deep reinforcement learning for de novo drug design." *Science advances* 4.7 (2018): eaap7885.
- [16] Richter, F., et al. "Open-sourced reinforcement learning environments for surgical robotics." *arXiv preprint arXiv:1903.02090* (2019).
- [17] Moerland, T.M., et al. "Model-based reinforcement learning: A survey." *Foundations and Trends in Machine Learning* 16.1 (2023): 1-118.