Reinforcement Learning vs Imitation Learning for Surgical Action Prediction: A Comprehensive Evaluation Using Actual Trained Models

Author Name
Institution
Address
Email: author@institution.edu

Abstract—This paper presents a comprehensive comparison between reinforcement learning (RL) and imitation learning (IL) approaches for surgical action prediction using actual trained models. We evaluate trajectory-level performance using cumulative mean Average Precision (mAP) analysis on the CholecT50 dataset, comparing Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC), and supervised imitation learning. Our analysis uses a world model trained on the same data as a latent simulator, eliminating sim-to-real transfer issues. Results show that Imitation Learning achieves the best performance with 0.979 mAP. Statistical analysis reveals significant differences between methods, providing insights for surgical AI development.

I. Introduction

Surgical action prediction is crucial for computer-assisted surgery systems. While imitation learning from expert demonstrations has been the standard approach, reinforcement learning offers potential advantages in temporal modeling and optimization. This work provides a rigorous comparison using actual trained models.

II. Methods

A. Experimental Setup

We compare three approaches using actual trained models:

- **Imitation Learning (IL)**: World model trained via supervised learning on expert demonstrations
- **Proximal Policy Optimization (PPO)**: On-policy RL using the world model as environment
- **Soft Actor-Critic (SAC)**: Off-policy RL with continuous action spaces

All methods use the same world model as the underlying simulator, trained on identical CholecT50 data, ensuring fair comparison without sim-to-real gaps.

B. Evaluation Protocol

We employ single-step inference where each method predicts the next action given the current visual state. Evaluation uses cumulative mAP trajectory analysis, computing mAP at each timestep using predictions from the start to that timestep.

III. Results

A. Main Findings

Overall Performance: $best_method[0].replace(',_'').title() achieves the hi$
RL vs IL Comparison:
Overall Performance: Imitation Learning achieves the highest mean mAP of 0.979, demonstrating competitive performance among all methods. **RL vs IL Comparison:**

- PPO underperforms IL: -0.678 mAP (-69.3%)
- SAC underperforms IL: -0.647 mAP (-66.1%)

Trajectory Stability: Methods show varying degradation patterns over time. Imitation Learning demonstrates the best trajectory stability. **Statistical Significance:** Pairwise comparisons reveal:

- Sac vs Ppo: p = 0.000, Cohen's d = 0.28
- Sac vs Imitation Learning: p = 0.000, Cohen's d = -8.26
- Ppo vs Imitation Learning: p = 0.000, Cohen's d = -8.83

IV. Discussion

A. Key Insights

Our evaluation using actual trained models reveals important insights:

1. **Model Architecture Matters**: The competitive performance of Imitation Learning suggests that direct imitation of expert behavior remains highly effective for surgical action prediction.

2. **Training Methodology**: Using the same underlying world model eliminates confounding factors from sim-to-real transfer, providing a fair comparison of learning paradigms.

3. **Trajectory Analysis**: Cumulative mAP evaluation reveals how prediction quality evolves over surgical procedures, crucial for understanding clinical applicability.

B. Clinical Implications

IL remains competitive for surgical applications, particularly where direct mimicry of expert behavior is desired.

V. Conclusion

This comprehensive evaluation using actual trained models provides rigorous comparison of RL and IL for surgical action prediction. Imitation Learning demonstrates superior performance, confirming the effectiveness of imitation learning.

TABLE I

COMPREHENSIVE COMPARISON: REINFORCEMENT LEARNING VS IMITATION LEARNING FOR SURGICAL ACTION PREDICTION USING ACTUAL TRAINED MODELS

Method	Mean mAP	Std mAP	Initial mAP	Final mAP	Degradation	Stability	Videos
Imitation Learning	0.979	0.010	0.998	0.961	0.037	-0.037	5
Sac	0.331	0.052	0.602	0.242	0.360	-0.360	5
Ppo	0.301	0.070	0.516	0.228	0.288	-0.288	5

Degradation = Initial mAP - Final mAP; Stability = -Degradation (higher is better)
All methods use actual trained models, not simulations

TABLE II
STATISTICAL SIGNIFICANCE TESTS BETWEEN METHODS

Comparison	Mean Difference	t-statistic	p-value	Effect Size
Sac vs Ppo	0.031	4.472	0.000***	0.285
Sac vs Imitation Learning	-0.647	-129.740	0.000***	-8.255
Ppo vs Imitation Learning	-0.678	-138.778	0.000***	-8.830

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

TABLE III
DETAILED PERFORMANCE ANALYSIS BY METHOD

Method	Min mAP	Max mAP	Median mAP	Range
Imitation Learning	0.961	0.990	0.983	0.029
Sac	0.269	0.404	0.331	0.135
Ppo	0.215	0.378	0.308	0.163

Future work should explore hybrid approaches combining both paradigms.

REFERENCES

[1] Add your references here...