SAGES 2026
March 25-28
Tampa Convention Center | Tampa, Florida

Surgical Video "Highlights Reel" Generation Based on AI-based Video Content Inference

Marzie Lafouti, Jonathan Hubermann, Liane S. Feldman, Amir Hooshiar
Department of Surgery, McGill University, Montreal, Canada

Introduction

Clinical Need
(Manual review bottleneck)

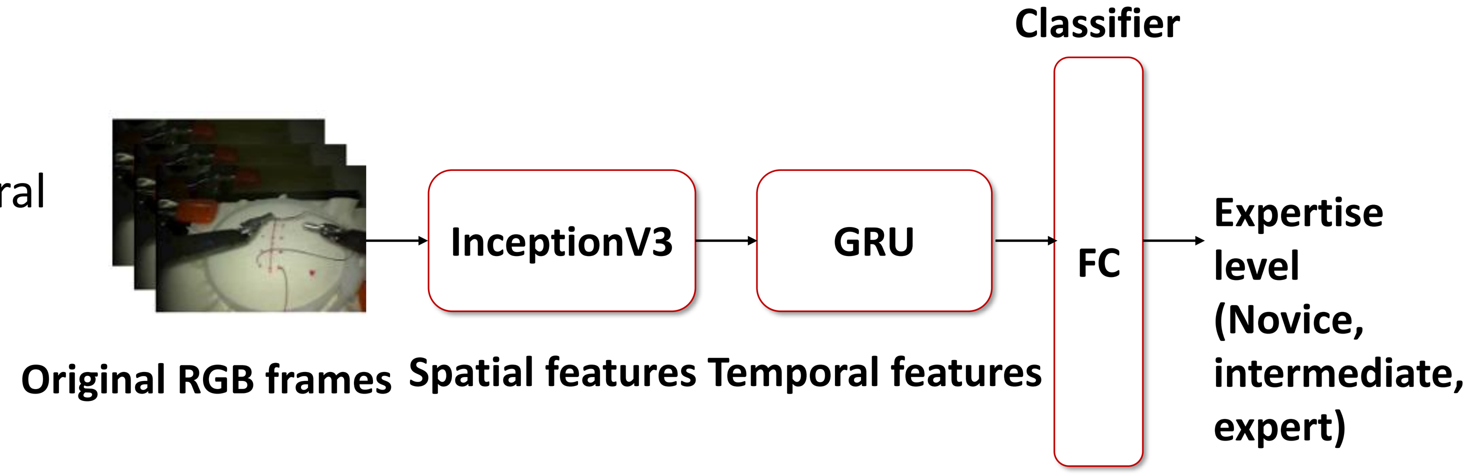
Importance
(Objective skill assessment & scalability)

Study Overview & Novelty
(AI-driven automated "highlight reels")

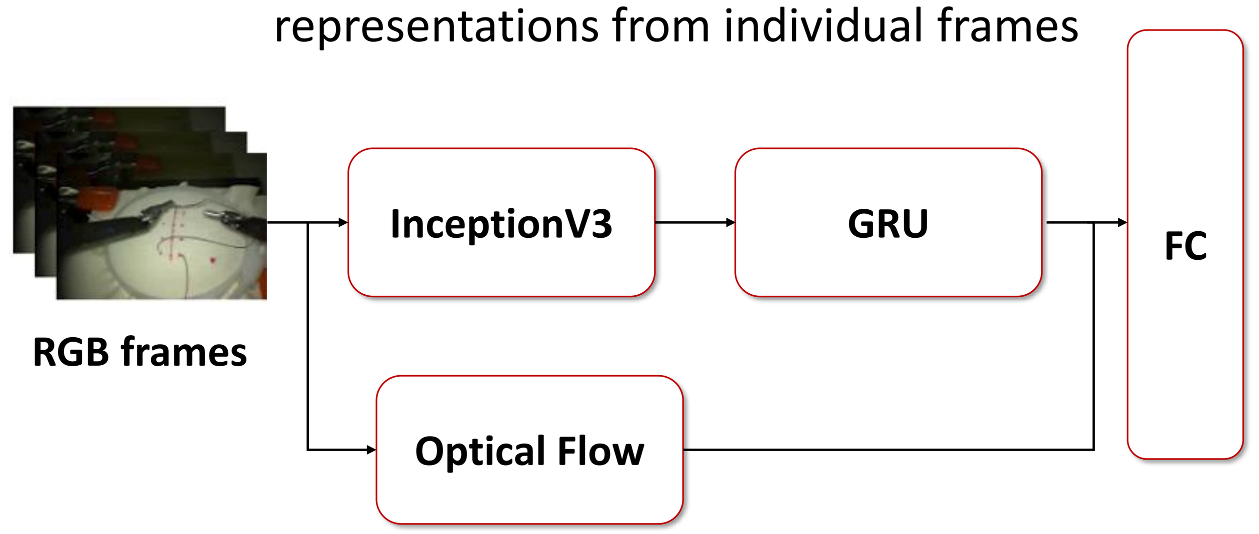
Materials & Methods

Hypothesis

Integrating spatial frame representations with temporal sequence modeling can accurately classify surgical expertise and automatically generate interpretable "highlight reels" of critical tool-tissue interactions.



Spatial features:
A pre-trained InceptionV3 extracted spatial representations from individual frames



Validation: On JIGSAWS dataset with established expertise labels (suturing 39 videos)

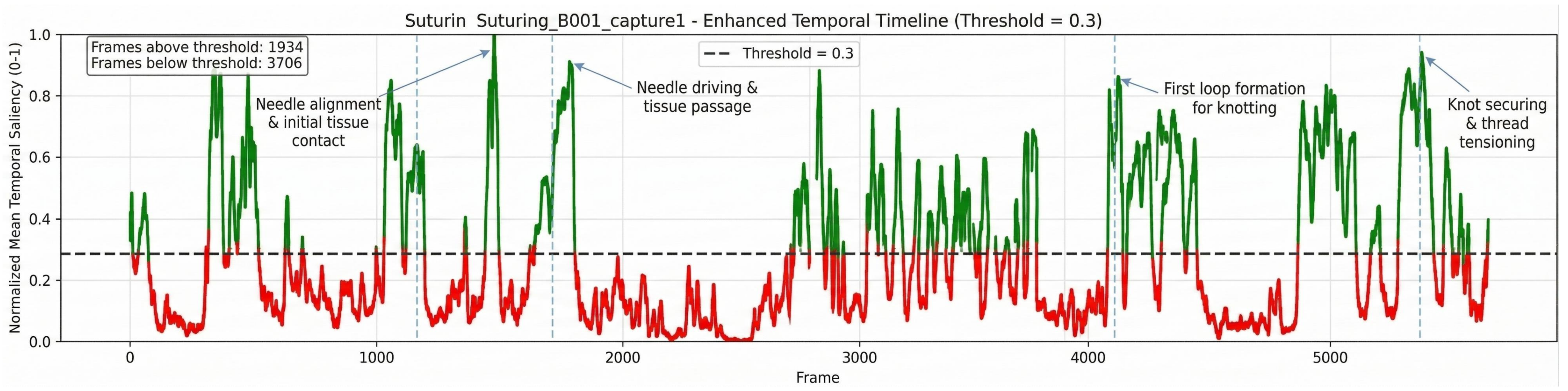
Temporal features:
A Gated Recurrent Unit (GRU) captured temporal dependencies, technique consistency and motion patterns across the video sequence

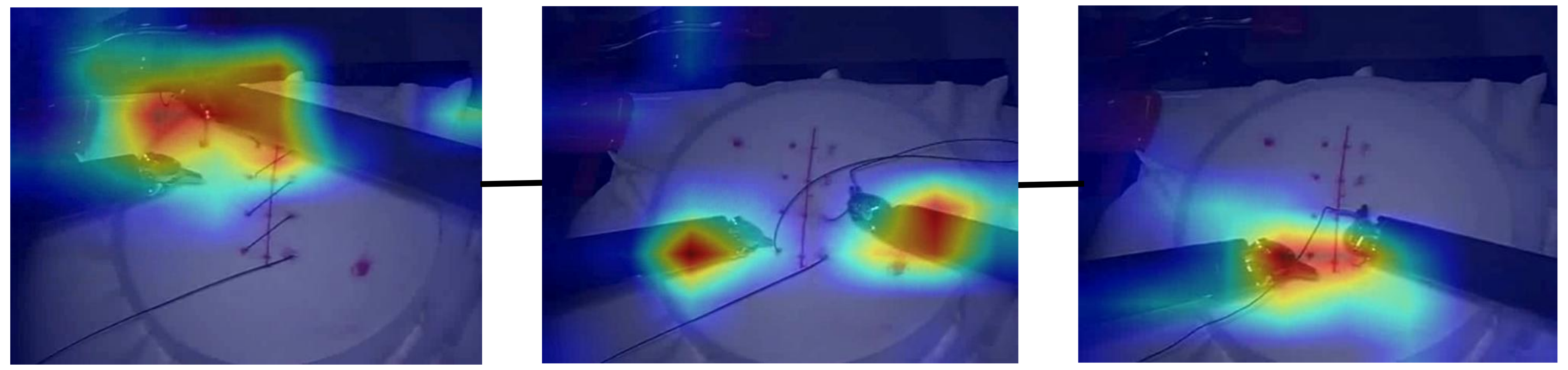
Classification: Fused spatiotemporal features are passed to a fully-connected classifier to output the expertise level.

Alternative Variant

- Evaluated a motion-augmented variant using late feature fusion of Optical Flow and RGB to capture bimanual coordination and motion smoothness.
- Underperformance

Results





Interpretability: Visualizing localized tool-tissue interactions

Table 1: Performance comparison.

	Accuracy	Macro Precision	Macro Recall	Macro F1	Class Expert	Class Intermediate	Class Novice
RGB	**0.9524**	**0.9420**	**0.9175**	**0.9234**	P: 0.861 R: 0.989 F1: 0.921	P: 0.990 R: 0.764 F1: 0.863	P: 0.975 R: 0.999 F1: 0.987
RGB+Optical-Flow	0.3803	0.6403	0.4911	0.3958	P: 0.244 R: 0.945 F1: 0.388	P: 0.851 R: 0.312 F1: 0.456	P: 0.826 R: 0.217 F1: 0.343

Model Performance

- The primary RGB pipeline achieved **95% classification accuracy**.
- Strong discrimination between novice and expert levels
- The motion-augmented variant underperformed (38% accuracy), heavily diluted by noisy Optical Flow from camera drift, glare, and phantom movement.

Conclusion

Findings & Impact

Our spatiotemporal pipeline accurately classifies surgical expertise and generates intuitive high-interpretability "highlight reels".

* This transforms clinician involvement by reducing manual review, maintaining capability and accelerating assessment.

Future Vision

Expansion to specific task-level metrics and longitudinal tracking of trainees

References

[1] Balvardi S, Kaneva P, Semsar-Kazerooni K, et al. Effect of video-based self-reflection on intraoperative skills: A pilot randomized controlled trial. *Surgery*. 2024.

[2] Balvardi S, Semsar-Kazerooni K, Kaneva P, et al. Validity of video-based general and procedure-specific self-assessment tools for surgical trainees in laparoscopic cholecystectomy. *Surgical Endoscopy*. 2023.

[3] Power D, Burke C, Madden MG, et al. Automated assessment of simulated laparoscopic surgical skill performance using deep learning. *Sci Rep*. 2025.

[4] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision. *IEEE Conference on Computer Vision and Pattern Recognition*. 2016:2818-2826.

[5] Gao Y, Vedula SS, Reiley CE, et al. Jigsaws: A surgical activity dataset for human motion modeling. *Modeling and Monitoring of Computer Assisted Interventions*. 2014:3-10.

Surgical Performance Enhancement and Robotics Centre (SuPER)