Technische Universität München – Faculty of Informatics
Chair for Computer Aided Medical Procedures (Prof. Nassir Navab)
Practical Course: Machine Learning in Medical Imaging WS24/25

Benchmark for Surgical Video Generation

1 General Info

Project Title: Benchmark for Surgical Video Generation
Supervisors: Azade Farshad, Yousef Yeganeh
Contact Email: azade.farshad@tum.de, y.yeganeh@tum.de

2 Background and Motivation

Surgical video generation has emerged as a promising tool for surgical training, simulation, and data augmentation [1, 5]. However, accurate generation and evaluation of the quality and clinical relevance of generated surgical videos remains challenging due to the complexity of surgical scenes and the lack of standardized evaluation metrics. While traditional image generation metrics like FID and SSIM provide general quality assessments, they may not capture surgery-specific attributes that are crucial for medical applications. Additionally, identifying and proposing accurate metrics can lead to higher performing generative models that can learn the real data distribution more effectively.

3 Project Outline

This project aims to develop a comprehensive evaluation framework for assessing the quality and clinical utility of generated surgical videos to effectively assess and improve the generation quality. We aim to first generate videos of different plausible and clinically relevant surgical scenarios by (1) generating videos of the same scene with different surgical tools, (2) different organs, and (3) videos of specific surgical phases. Using the collection of real and generated surgical videos, we propose multiple evaluation metrics:

- Surgical Tool Analysis
  - Tool detection accuracy and confidence scores [3]
  - Temporal consistency of tool positioning
  - Evaluation using pre-trained surgical tool detectors on both real and generated videos
- Scene Depth Understanding
  - Comparative analysis of depth estimation performance between real and generated videos
  - Assessment of 3D structure preservation in generated sequences

- Clinical Relevance Metrics
  - Phase recognition accuracy in surgical workflows [2]
  - Assessment of surgical action continuity

The methods would be trained and evaluated on the CholecT50 [4] dataset that contains videos of surgeries and the actions performed during the surgery.

4 Technical Prerequisites

- Good background in machine learning and deep learning
- Experienced in PyTorch
- Experienced in Python
- Experience with Generative Models

5 Benefits

- Weekly supervision and discussions
- Possible novelty of the research
- The results of this work are intended to be published in a conference or journal

6 Work packages and Time-plan

	Description
WP1	Familiarizing with the literature.
WP2	Implementing the baselines
WP3	Improving the baselines and validation on relevant datasets
WP4	Implementing the model
WP5	Finalizing the results and evaluation

Table 1: Suggested Work Packages

References

[1] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22563–22575, 2023.

[2] Tobias Czempiel, Magdalini Paschali, Daniel Ostler, Seong Tae Kim, Benjamin Busam, and Nassir Navab. Opera: Attention-regularized transformers for surgical phase recognition. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, pages 604–614. Springer, 2021.

[3] Hai-Binh Le, Thai Dinh Kim, Manh-Hung Ha, Anh Long Quang Tran, Duy-Thuc Nguyen, and Xuan-Minh Dinh. Robust surgical tool detection in laparoscopic surgery using yolov8 model. In *2023 International Conference on System Science and Engineering (ICSSE)*, pages 537–542. IEEE, 2023.

[4] Chinedu Innocent Nwoye, Tong Yu, Cristians Gonzalez, Barbara Seeliger, Pietro Mascagni, Didier Mutter, Jacques Marescaux, and Nicolas Padoy. Rendezvous: Attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Medical Image Analysis*, 78:102433, 2022.

[5] Yousef Yeganeh, Rachmadio Lazuardi, Amir Shamseddin, Emine Dari, Yash Thirani, Nassir Navab, and Azade Farshad. Visage: Video synthesis using action graphs for surgery. *arXiv preprint arXiv:2410.17751*, 2024.