TECHNISCHE UNIVERSITÄT DRESDEN

MICCAI 2025
28th International Conference on Medical Image Computing and Computer Assisted Intervention
23-27 September 2025
Daejeon REPUBLIC OF KOREA

Universitätsklinikum Carl Gustav Carus
DIE DRESDNER.

VISION LANGUAGE MODELS FOR AUTOMATED VIDEO ANALYSIS AND DOCUMENTATION IN LAPAROSCOPIC SURGERY

Hannah S. Muti[1,2] | Esther H. Stueker[1,2] | Fiona R. Kolbinger[2,3] | Florian Oehme [2] | Jakob N. Kather[1,4]
Else Kröner Fresenius Center for Digital Health, Dresden University of Technology, Germany[1]
Department for Visceral, Thoracic and Vascular Surgery, University Hospital Carl Gustav Carus Dresden, Germany[2]
Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA[3]
Medical Department 1, University Hospital Carl Gustav Carus Dresden, Germany[4]

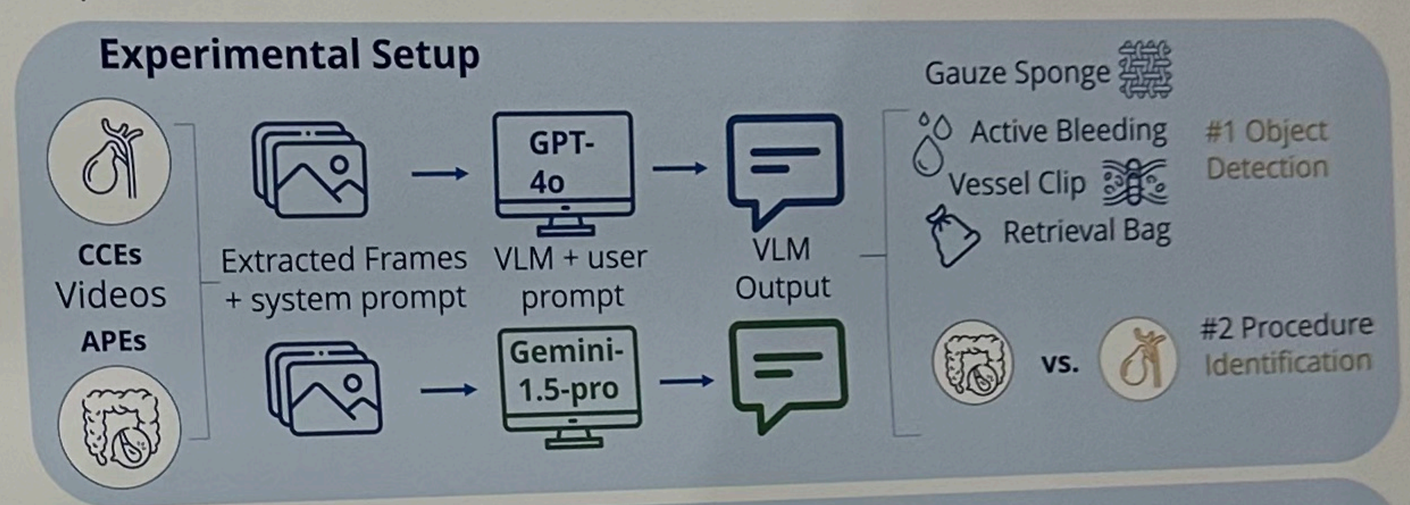
Research Question

VLMs have inherent surgical video analysis capabilities – but to what extent? Can VLMs perform surgical video analysis and write accurate surgical reports?

Experimental Design

Assessment of VLMs for increasingly complex surgical vision analysis tasks:

- Object identification
- Procedure identification
- Generating a Surgical Report



Models

Gemini 1.5 pro vs. *GPT4-o*

Data

- Laparoscopic Appendectomies: subset of the LapApp dataset[1]
- Laparoscopic Cholecystectomies: subset of the CholecT45 dataset[2]
- n=30

[1] Laparoscopic appendectomy dataset, multi-institutional, Germany. Kolbinger et al.,*unpublished*
[2]Laparoscopic cholecystectomy dataset, University Hospital Strasbourg, France. Nwoye et al., 2022

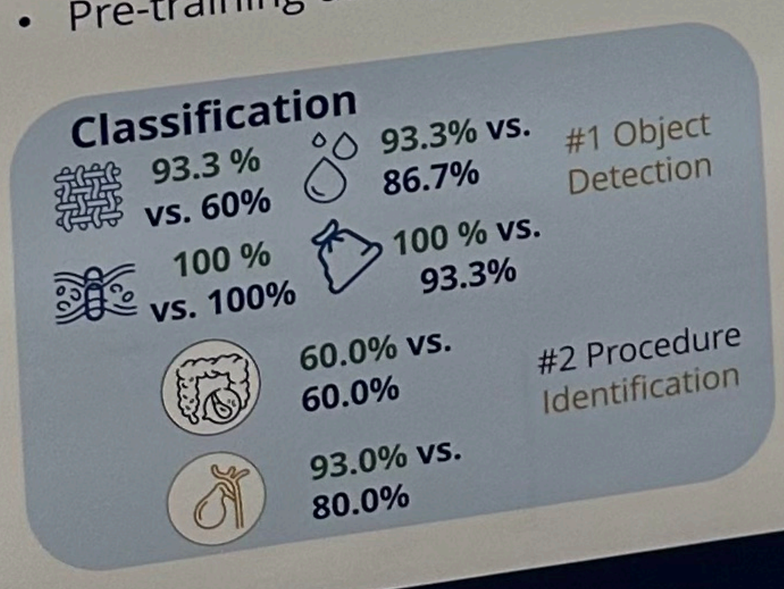
Results

Object detection

- High accuracy
- Gemini 1.5 pro outperforms GPT4-o

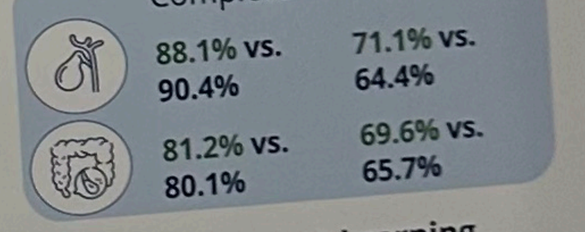
Procedure identification

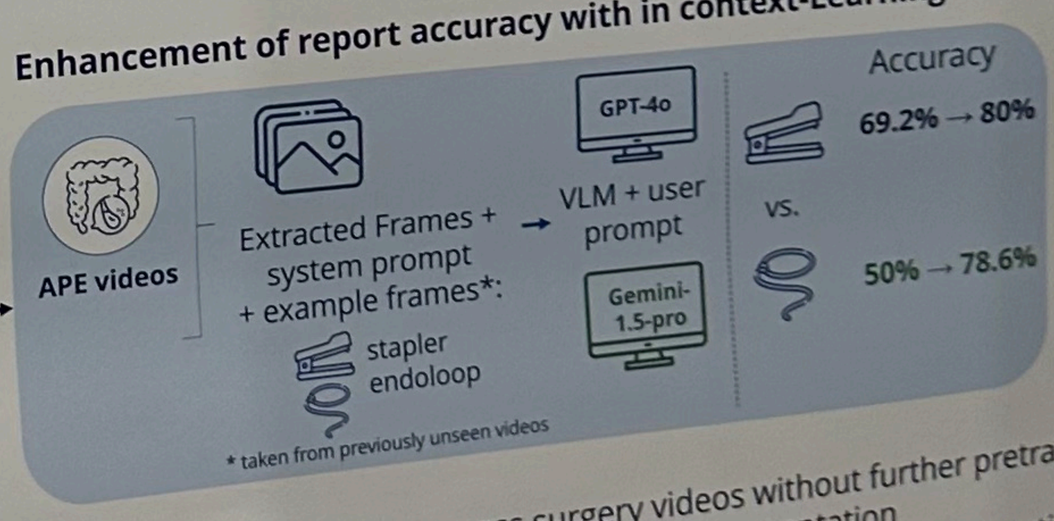
- Moderate accuracy
- Gemini 1.5 pro outperforms GPT4-o
- Pre-training data: potential bias



Report Generation

- Moderate-high accuracy
- **Highest inaccuracies in method of organ separation**





Conclusion

- VLMs show inherent capabilities to assess surgery videos without further pretraining
- In the future, VLMs could perform real-time surgical documentation
- After authorization by a trained surgeon, this could streamline the documentation process, reducing the documentational burden on surgeons

Gefördert durch:
Else Kröner Fresenius Stiftung

Check out our publication!
Stueker, EH; Kolbinger, FR; Saldanha, OL; Digomann, D; Pistorius, S; Oehme, F; Van Treeck, M; Ferber, D; CML; Weitz, J; Distler, M; Kather, JN; Muti, HS. Vision-language models for automated video analysis and documentation in laparoscopic surgery: a proof-of-concept study. **International Journal of Surgery**. July 17, 2025. | DOI: 10.1097/JS9.0000000000003069

Get in touch!

EKFZ Digital Health