2D Object Detection and Segmentation for Surgical Instruments

Bachelor Thesis Proposal by Kerim Gregory Birgi
Supervisor: Jonas Hein

I. Introduction

Surgeries are executed by highly specialized surgeons with years of practice. Due to the nature of surgeries and the need for highly precise movements, there still remain some risks in any kind of surgery. The field of Computer Vision and Mixed Reality aims to minimize these risks by providing surgical guidance systems and safety measurements for surgeons. One example of such a system is an alert prior to wrongful behaviors conducted by the surgeon such as an inaccurate screw placement before insertion into the patient.
One of the foundations of these surgical aid systems is the detection of tools used during surgery and the calculation of their exact positions and poses. It is crucial that this is done both fast and highly accurate.
The goal of this thesis is to provide an overview of the state of the art models for the detection of bounding boxes of surgical instruments and their respective segmentation.
There are options to continue the thesis by incorporating synthetic data into the model's training pipeline. Training on real data is expensive because of the collection and annotation, therefore we would like to see if it's feasable to train the model using synthetic data and furthermore what an ideal ratio of synthetic data to real data would be

II. Related Work

The release of the ROBUST-MIS challenge and dataset [7] has allowed for easier development of models by providing high quality data of surgical scenarios. The challenge had multiple teams attempt to build solutions for binary segmentation, multi-instance detection and multi-instance segmentation.
Following, there has been a more steady release of papers with a focus on object detection in the surgical environment.
[6] develops a model for real-time bounding box calculations with high accuracy but does not perform any segmenting.
[2] develop two new models with Mask R-CNN and Mask2former for instance segmentation of surgical instruments.
[1] uses Yolact architecture and builds on top of it with an attention mechanism and multi-scale feature fusion to create a model for real time instance segmentation of surgical instruments during non invasive surgeries utilizing the same benchmark dataset as [7].
Besides [7], which is from 2019, none of these works compare multiple methods of instance segmentation. Furthermore none of them incorporate the utilization of synthetic data into the model training pipeline.

III. Methodology

The goal of the thesis is to create a model for the detection of both modal and amodal masks of a set of known tools during a surgical scenario as accurately and efficiently as possible. For the development of the model we assume that we have video footage of a spinal surgery from a calibrated camera set around the surgery table, including a depth range camera, and that we know all the intrinsic and extrinsic parameters. During the surgery, tools used will include a drill and screwdriver with markers on them. By utilizing marker tracking, we have ground truth information from which we can calculate ground truth masks.
The surgical environment means that the background of all footages are similiar and thus the data differentiates itself from typical benchmarks such as COCO.
In this thesis we show an overview of state of the art models and compare them using our own dataset. As a baseline we compare a Yolact architecture [3], which is based on convolutional neural nets, and a DETR architecture [4], which is based on transformers. Further architectures could be included if determined feasible and as time permits.
With the possibility of incorporating synthetic data into the pipeline, we face two options for continuance. One

option is to use a mix of both synthetic and real data or only synthetic data to train the model and examine how well it fares in comparison to the model trained on real data.
Alternatively, the model could be trained on synthetic data and then Segment Anything [5] can be used to refine the model output.

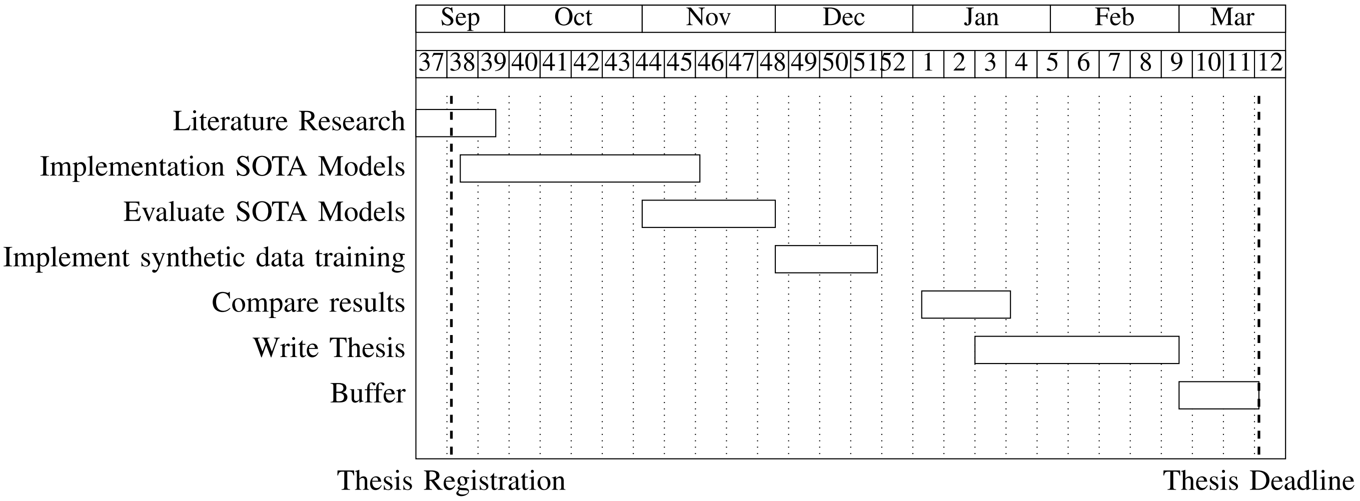

Fig. 1. Proposed time plan on the basis of calendar weeks.

References

[1] Juan Carlos Angeles-Ceron, Gilberto Ochoa-Ruiz, Leonardo Chang, and Sharib Ali. Real-time instance segmentation of surgical instruments using attention and multi-scale feature fusion, 2021.

[2] Britty Baby, Daksh Thapar, Mustafa Chasmai, Tamajit Banerjee, Kunal Dargan, Ashish Suri, Subhashis Banerjee, and Chetan Arora. From forks to forceps: A new framework for instance segmentation of surgical instruments, 2023.

[3] Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. Yolact: Real-time instance segmentation, 2019.

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers, 2020.

[5] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[6] Yuying Liu, Zijian Zhao, Faliang Chang, and Sanyuan Hu. An anchor-free convolutional neural network for real-time surgical tool detection in robot-assisted surgery. *IEEE Access*, 2020.

[7] Tobias Ross, Annika Reinke, Peter M. Full, Martin Wagner, Hannes Kenngott, Martin Apitz, Hellena Hempe, Diana Mindroc Filimon, Patrick Scholz, Thuy Nuong Tran, Pierangela Bruno, Pablo Arbeláez, Gui-Bin Bian, Sebastian Bodenstedt, Jon Lindström Bolmgren, Laura Bravo-Sánchez, Hua-Bin Chen, Cristina González, Dong Guo, Pål Halvorsen, Pheng-Ann Heng, Enes Hosgor, Zeng-Guang Hou, Fabian Isensee, Debesh Jha, Tingting Jiang, Yueming Jin, Kadir Kirtac, Sabrina Kletz, Stefan Leger, Zhixuan Li, Klaus H. Maier-Hein, Zhen-Liang Ni, Michael A. Riegler, Klaus Schoeffmann, Ruohua Shi, Stefanie Speidel, Michael Stenzel, Isabell Twick, Gutai Wang, Jiacheng Wang, Liansheng Wang, Lu Wang, Yujie Zhang, Yan-Jie Zhou, Lei Zhu, Manuel Wiesenfarth, Annette Kopp-Schneider, Beat P. Müller-Stich, and Lena Maier-Hein. Robust medical instrument segmentation challenge 2019, 2020.