McGill

AI-Assisted Robotic Surgery for Gynecologic Cancer

Yoav Brezinov MD

Experimental Surgery

McGill University, Montreal

June 2024

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Intelligence
AP	Average Precision
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CT	Computed Tomography
CV	Computer Vision
DVIS++	Decoupled Video Segmentation
EC	Endometrial Cancer
IOU	Intersection over Union
IP	Infundibulopelvic (ligament)
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
NACT	Neo-Adjuvant Chemotherapy
OC	Ovarian Cancer
PET	Positron Emission Tomography
PointRend	Point-based Rendering
RH	Radical Hysterectomy
SL	Supervised Learning
SSL	Semi-supervised Learning
TL	Transferred Learning
YTVIS	YouTube Video Instance Segmentation

Acknowledgments

This work wouldn't be possible without the innovative ideas of my supervisor Dr. Walter Gotlieb. With his guidance and support, the aspiration to bring surgery closer to perfection has taken a considerable step toward reality.

I want to thank the research committee members, Dr. Melica Nourmoussavi Brodeur, Dr. Shannon Salvador and Dr. Susie Lau for advising me on the preparation of this project.

I am extremely grateful to Dr. Jeremie Abitbol for his insights and inspiration. His contribution to the success of this work is invaluable.

I sincerely thank Maxim Rusakov for creating the model architecture and David Bullock for building the cloud-based AI infrastructure. Their professionalism and dedication were essential for the development of this project.

I extend my appreciation to the entire team of the Gynecologic Oncology department at the Jewish General Hospital and the Lady Davis Institute staff for providing me with a comfortable and productive research environment. Special thanks to Dr. Angela Tatar and Dr. Amber Yasmeen.

I want to conclude this by expressing my sincere gratitude to my wife, Olga. Thank you for your unconditional support and approval of all my crazy ideas!

Contribution of Authors

This thesis was written by me, Yoav Brezinov, a Master's student at McGill University. I authored every section of this thesis, including the introduction, methods, results, and discussion of Chapters 1 and 2.

Chapter 1:
Computer Vision in Surgery – A Systematic Review

Abstract (English)

Objectives: This systematic review aims to map existing surgical computer vision models, identify the leading disciplines in medical CV technology and pinpoint existing gaps as well as to study the segmentation tasks, open-source databases, types of neural networks, and available AI resources.

Methods: This systematic review was performed according to the PRISMA 2020 statement guidelines. PubMed® database website was used to identify eligible studies. The search criteria included original works on surgical CV models, published between 2014 and 2024, written in English and conducted with data from humans.

The extracted variables included the segmentation task, the surgical/medical discipline, the dataset type and source, the neural network type and the anatomical region in focus.

Results: The search criteria identified 3,355 articles. After excluding non-eligible articles (3,266), abstracts of the remaining 89 were reviewed and 71 models were included in the final analysis.

The stomach was the most popular organ in surgical CV models (12.7%) and the two most common disciplines were general surgery and gastroenterology. Most models concentrated on identifying pathological features (tumors, inflammation, etc.). Endoscopies were the leading source of data (26.7%), followed by laparoscopies (22.7%). Open-source datasets were used in 22.5% of the studies.

Nearly all models used CNNs with ResNet as the leading network architecture. Ten percent of the models incorporated pre-training methods, and optimization methods were employed in 12% of the studies.

Conclusion: This work provides a detailed analysis of surgical segmentation models from a surgeon's perspective. It offers recommendations for advancing surgical CV technology important for transitioning from proof-of-concept to practical application stage.

Abstract (French)

Objectifs: Cette revue systématique vise à cartographier les modèles existants de vision par ordinateur en chirurgie, à identifier les disciplines dominantes dans la technologie de la vision par ordinateur médicale et à repérer les lacunes existantes. Elle étudie les tâches de segmentation, les bases de données open source, les types de réseaux neuronaux et les ressources en IA disponibles.

Méthodes: Cette revue systématique a été réalisée selon les lignes directrices de la déclaration PRISMA 2020. La base de données PubMed® a été utilisée pour identifier les études éligibles. Les critères de recherche incluaient des travaux originaux sur les modèles de vision par ordinateur en chirurgie, publiés entre 2014 et 2024, rédigés en anglais et réalisés sur des données humaines.

Les variables extraites comprenaient la tâche de segmentation, la discipline chirurgicale/médicale, le type et la source de données, le type de réseau neuronal et la région anatomique ciblée.

Résultats: Les critères de recherche ont identifié 3,355 articles. Après exclusion des articles non éligibles (3,266), les résumés des 89 articles restants ont été examinés et 71 modèles ont été inclus dans l'analyse finale.

L'estomac était l'organe le plus populaire dans les modèles de vision par ordinateur en chirurgie (12.7%) et les deux disciplines les plus courantes étaient la chirurgie générale et la gastroentérologie. La plupart des modèles se concentraient sur l'identification des caractéristiques pathologiques (tumeurs, inflammations, etc.). Les endoscopies étaient la principale source de données (26.7%), suivies par les laparoscopies (22.7%). Les bases de données open source étaient utilisées dans 22.5% des études.

Presque tous les modèles utilisaient des CNN avec ResNet comme architecture de réseau principale. Dix pour cent des modèles incorporaient des méthodes de pré-formation et des méthodes d'optimisation étaient employées dans 12% des études.

Conclusion: Ce travail fournit une analyse détaillée des modèles de segmentation chirurgicale du point de vue d'un chirurgien. Il offre des recommandations pour faire progresser la technologie de la vision par ordinateur en chirurgie, essentielle pour passer de l'étape de preuve de concept à celle de l'application pratique.

Introduction

Computer vision (CV) involves the automated interpretation of images and videos by computer systems to understand the environment. It draws inspiration from the human vision (1). A digital image is a two-dimensional array of numbers captured by modern cameras or created by imaging software. A video is a sequence of digital images. CV provides meaning to digital images. Numbers transform into objects, and by learning the relationships among these objects within an image, as well as tracking the transitions across consecutive images, the computer develops a sense of vision.

Modern surgery increasingly relies on imaging, endoscopy, and minimally invasive methods. Endoscopic procedures allow surgeons to access internal organs through natural orifices or small incisions. Concurrently, sophisticated imaging techniques such as MRI and CT scans provide detailed internal views that guide surgical decisions and procedures. These minimally invasive approaches improve the accuracy of surgeries, significantly shorten hospital stays, and lessen postoperative pain (2–5). This shift towards technologically advanced surgical methods opens numerous opportunities for CV applications.

The integration of CV technology into surgical practice holds great potential to advance modern medicine. By enabling automated visual data analysis, CV technologies can enhance surgical precision, reduce errors, and improve patient outcomes. Recent years have seen an explosion of surgical CV models with various applications, including pre-operative planning, intra-operative navigation, and postoperative assessment. These models are concentrated on several key segmentation tasks: detection of anatomical structures, identification of surgical instruments and phases, integration with robotic systems, tissue discrimination and surgical skill assessment.

Despite the rapid development and promising results, the application of CV technologies varies significantly across different surgical disciplines. Some areas, such as ophthalmology and gastroenterology are benefiting from multiple CV research projects, while in other medical fields CV research is minimal (6,7). Furthermore, the clinical implementation of these technologies varies widely, with significant discrepancies in the development and real-world application of CV models.

Rationale and Objectives of the Review

This systematic review aims to map the existing landscape of surgical CV models. The goal is to identify which medical disciplines are leading CV technology development and where gaps exist that require further research and new models. This includes studying the milieu of the segmentation tasks, the use of open-source databases, the types of neural networks, and the AI resources utilized for building and training these models.

Mapping of CV models to surgical disciplines and body areas

This review will catalogue the various CV models according to the specific surgical disciplines and body areas. Analyzing the disciplines and organs used in current models will help to identify new research opportunities.

Resource for new researchers

The review will gather information about key components of CV research, including open-source databases and popular neural network architectures. This will help new researchers in the field with planning and conducting their research.

Methods

This systematic review was performed according to the PRISMA 2020 statement guidelines (8). The gathering and analysis of data, as well as results reporting and discussion, followed the checklist provided in the PRISMA 2020 publication.

PubMed® database website was used to identify eligible studies. The last search was on April 15, 2024. All the eligible studies were arranged in one group for analysis.

Inclusion criteria

- Original work on surgical CV models
- Published in the last decade (2014 - 2024)
- Written and published in English
- Conducted with data from human subjects

Exclusion criteria

- Non-original articles (review, meta-analysis)
- Retracted studies and studies with no abstract available
- Non-CV studies and non-surgical or medical models
- Non-human data (e.g. mice)

Data collection process

The following keywords were used to generate the search criteria: surgery, minimally invasive surgery (MIS), robotic surgery, laparoscopy, endoscopy, computer vision, object recognition, surgical scene segmentation, and anatomical structure identification.

Using the advanced search option and the designated keywords, a search string was generated:

“(Surgery OR "Minimally Invasive Surgery" OR Laparoscopy OR Endoscopy OR "Robotic Surgery") AND (Computer Vision OR "Object Recognition" OR "Surgical Scene Segmentation" OR "Anatomical Structure Identification")”.

Initial refining to exclude ineligible results was done by applying filters for publication date (10 years), article type (not review or systematic review), article language (English), species (Humans) and text availability (Abstract).

After initial refining, the abstracts of eligible studies were manually reviewed to exclude non-surgical or medical models and retracted articles.

The remaining articles were included in the analysis and reviewed by a single reviewer (the author of this thesis). The inclusion of all the eligible studies prevented selection bias and risk of bias assessment was not performed.

Data items description

For each article, the following variables were extracted, categorized and organized for analysis:

1. Segmentation task: Identifies the specific computer vision task (e.g. surgical phase/step detection). Similar tasks were grouped by category. If a model performed multiple segmentation tasks, each task was counted independently.
2. Surgical/medical discipline: Specifies the surgical or medical specialty in the study. In studies encompassing more than one discipline, each discipline was counted independently. Sub-specialties were categorized into the main disciplines. For instance, bariatric surgery, upper GI surgery and colorectal surgery were categorized as general surgery. Models involving imaging were categorized into radiology.
3. Dataset type and source: Classifies the dataset used for CV model training (e.g. laparoscopy videos). Open-source datasets were also noted.

4. Neural network type and backbone/framework: Categorizes the types of neural networks used in the model's architecture as well as the network's backbone and pre-training methods.
5. Anatomical region: Groups the models by organs or anatomical areas.

All the relevant data was systematically collected to eliminate any possible biases.

Ethical considerations

This systematic review did not require ethical approval as it did not involve any patient data, and thus there were no concerns regarding privacy or confidentiality. The review solely involved the collection and analysis of publicly available academic literature and published research.

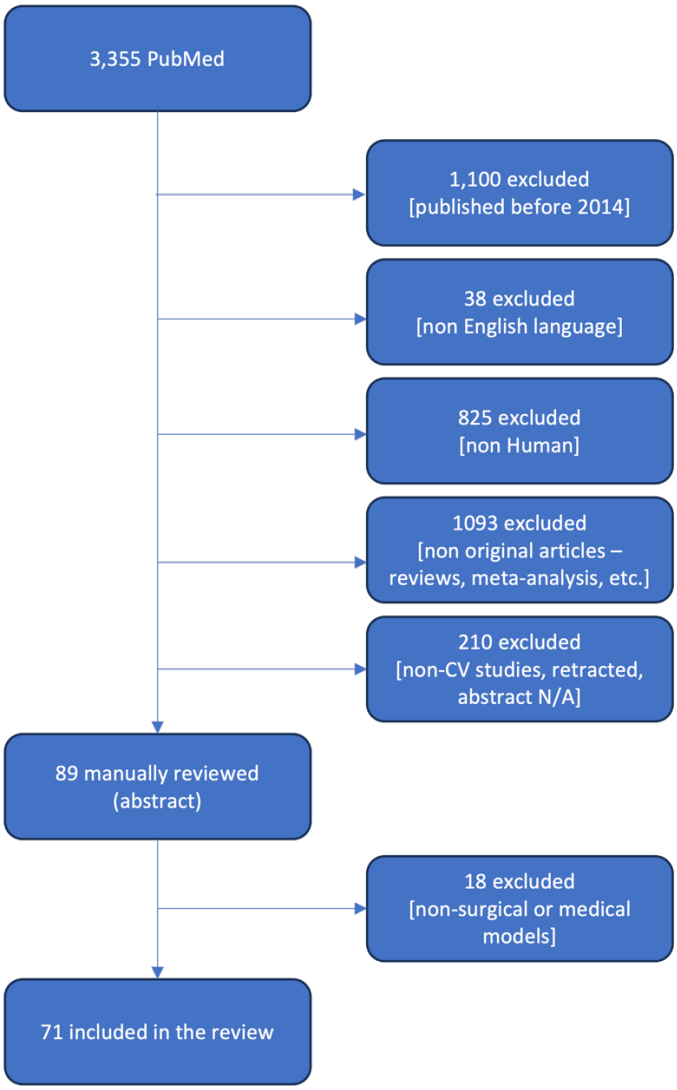
The protocol of the review ensures that no copyright violations occur.

Results

The search criteria identified 3,355 articles. 1,100 articles were published before 2014 and were excluded. 38 excluded for non-English language and 825 were omitted for data from non-human subjects. 1,093 were non-original articles (e.g. reviews or meta-analyses) and 210 were without an abstract, retracted or found to be non-CV studies.

The remaining 89 abstracts were manually reviewed and 18 were excluded for non-surgical or medical models. The final analysis included 71 studies (Figure 1.1).

Figure 1.1: Data selection flow chart (PRISMA diagram)



The stomach was the most prevalent organ in segmentation models (12.7%), followed by the eye (9.9%) and the esophagus (8.4%). Overall, the gastrointestinal (GI) system was the most segmented system of the human body. CV models, focusing on organs of the alimentary tract (from the esophagus to the rectum), comprised 36.6%. After adding digestive and endocrine organs (liver, gallbladder and pancreas) this ratio increased to 43.7%. Figure 1.2 shows a body map of all the CV projects included in the review.

The two most common medical disciplines in CV models were general surgery and gastroenterology. Data from these two disciplines was the focus of 23.6% of models, each, and together they cover roughly 50% of all the models. Figure 2.3 illustrates the distribution of CV models by medical discipline.

Overall, 13 different segmentation tasks were identified. Recognition of pathological anatomy, such as tumors, fractures and inflammation, was the most popular category, seen in 36.8% of the models. The second most common task was the detection of surgical instruments, found in 18.4% of studies and surgical step/phase identification was the third (15.8%). Figure 2.4 visualizes the spreading of segmentation tasks among models.

Most data were derived from endoscopies and laparoscopic surgeries, 26.7% and 22.7% respectively. Open surgeries comprised 18.6%, mostly cataract and other eye surgery (50% of the open surgeries). Most data were derived from internal or restricted access sources, with only 22.5% of the studies using open-source datasets. Some of the most notable open-source datasets are:

1. **CATARACTS** – Consists of 50 high-definition videos of cataract surgeries.

   URL: https://ieee-dataport.org/open-access/cataracts

2. **Cholec80** – Contains 80 video recordings of cholecystectomy surgeries.

   URL: https://paperswithcode.com/dataset/cholec80

3. **Bypass40** – Includes 40 videos of gastric bypass surgeries.

   URL: https://arxiv.org/abs/2102.12218v1

4. **KVASIR** - Features gastrointestinal tract images from endoscopies.

   URL: https://datasets.simula.no/kvasir/

5. **MISAW** - Includes 27 sequences of micro-surgical anastomosis.

   URL: https://paperswithcode.com/dataset/misaw

6. **PH2** – A dermoscopic image database of skin lesions.

   URL: https://github.com/dataset-ninja/ph2

7. **ISIC** – A collection of dermoscopic images for melanoma diagnosis.

   URL: https://www.isic-archive.com

8. **EndoVis** – includes surgical videos and images for various tasks.

   URL: https://www.endovis.org

9. **PTB ECG** – Contains ECG recordings with clinical summaries.

   URL: https://physionet.org/content/ptbdb/1.0.0/

10. **MICCAI 2017** – Various datasets from the MICCAI 2017 challenge, including liver tumor segmentation and robotic instrument segmentation.

    URL: https://github.com/ternaus/robot-surgery-segmentation

11. **BCCD** – A small-scale dataset for blood cell detection, that includes annotated images of white and red blood cells.

    URL: https://github.com/Shenggan/BCCD_Dataset

12. **LISC** – For leukocyte segmentation and classification, includes hematological images of peripheral blood.

    URL: https://universe.roboflow.com/wbcs/wbc-lisc

13. **LUNA16** – For pulmonary nodule detection in CT scans, annotated by experienced radiologists.

    URL: https://luna16. grand-challenge.org

14. **Ali Tianchi** – Includes CT scans for pulmonary nodule prediction.

    URL: https://tianchi.aliyun.com

15. **HAM10000** – 10,015 dermoscopic images of pigmented skin lesions.

    URL:

    https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

16. **VocalFolds** - Dataset of 536 laryngeal images.

    URL: https://github.com/imesluh/vocalfolds

Figure 1.2: Map of computer vision models

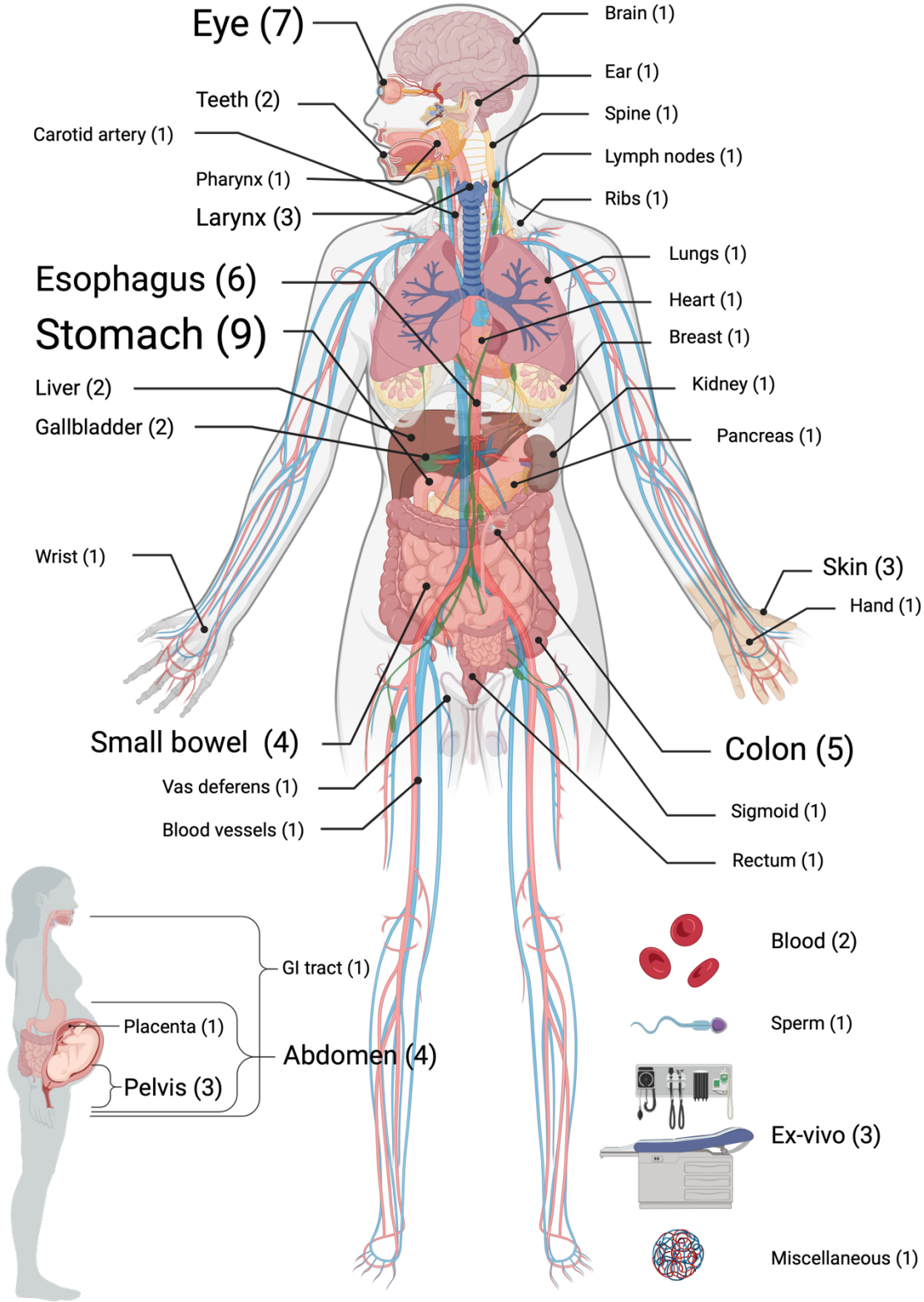
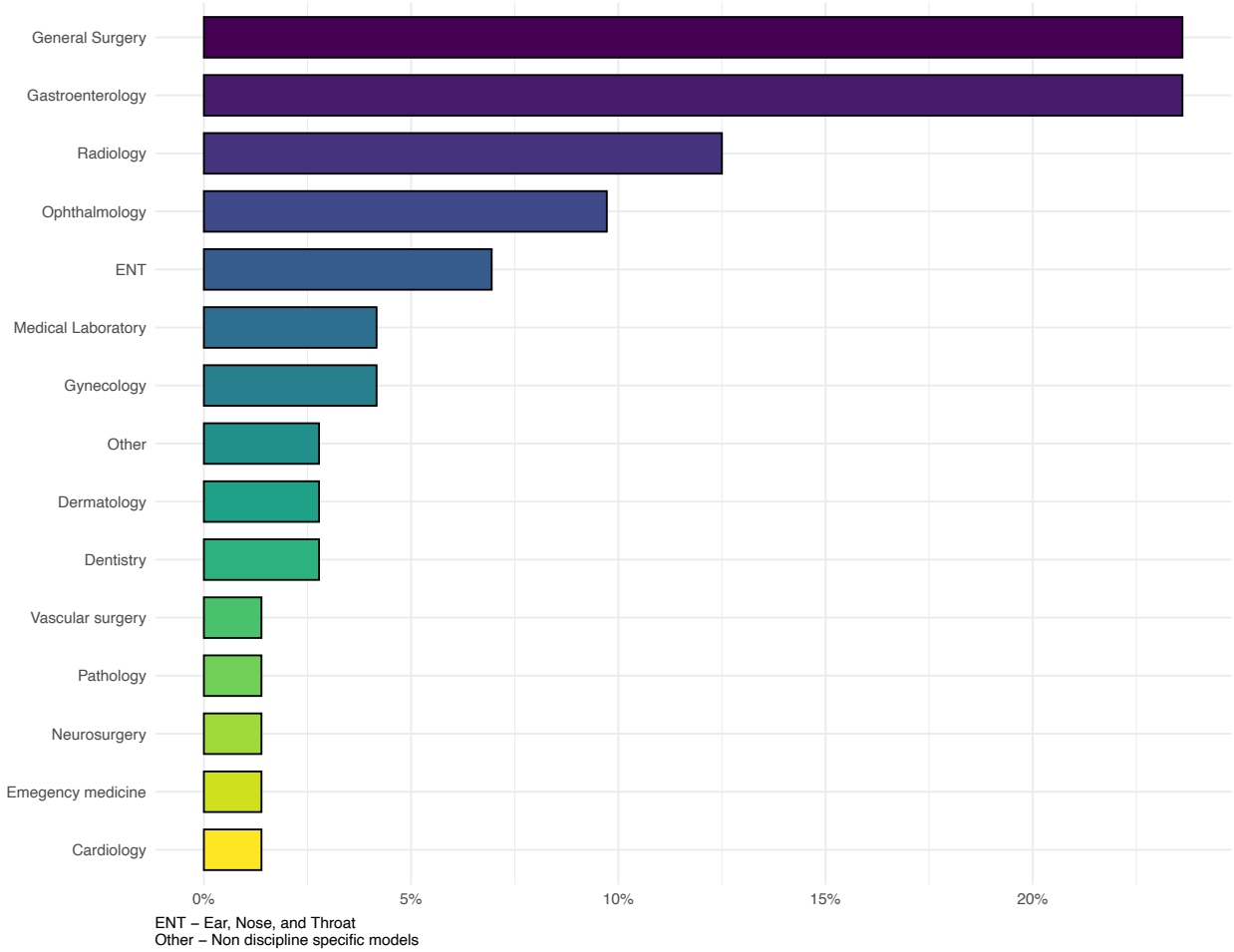


Figure 1.3: Distribution of medical disciplines in computer vision models



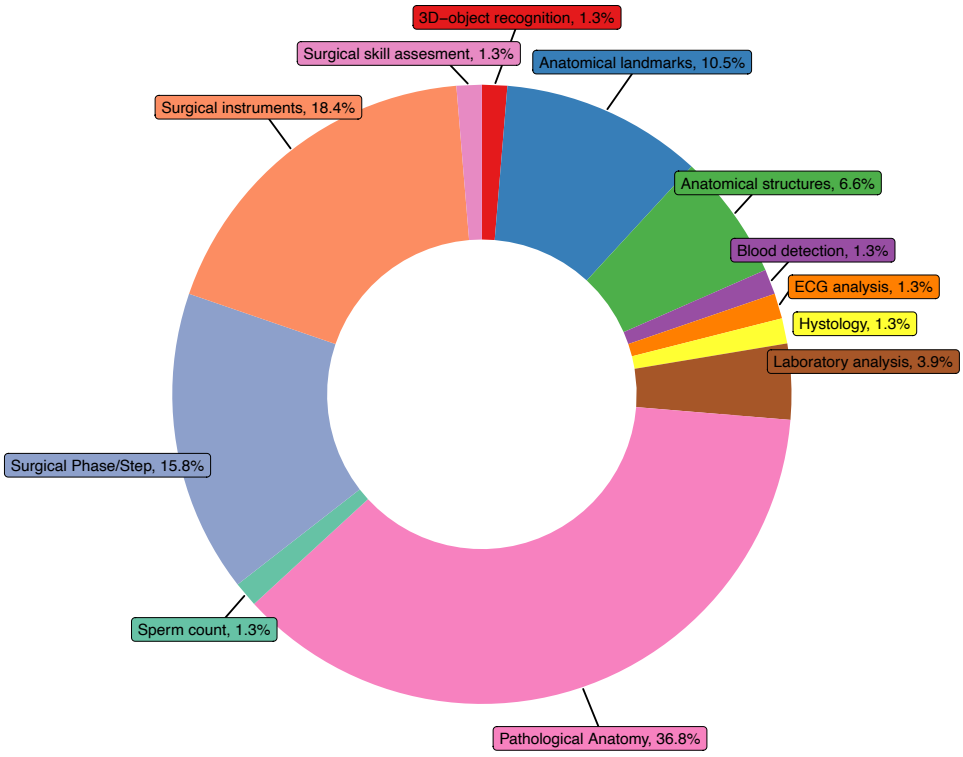
Nearly all models utilized convolutional neural networks (CNNs). ResNet was the leading architecture of CNNs, used in 18.6% of the models. Earlier architectures, AlexNet and VGG (Visual Geometry Group) were used in 3% and 1% of the models respectively. MobileNet was seen in 3% of the models and GoogleNet (Inception) in 5%.

Ten percent of the models incorporated pre-training methods using visual data from ImageNet and Microsoft COCO (Common Objects in Context) datasets.

Optimization methods such as SENet (Squeeze-and-Excitation Network), YOLO (You Only Look Once) and YOLACT (You Only Look At CoefficienTs) were employed in 12% of the studies.

Detectron (Facebook), Tensorflow (Google) and IBM Visual Insights (IBM) were the most common object detection and segmentation platforms.

Figure 1.4: Spreading of segmentation tasks in computer vision models



Discussion

This review presents an analysis of surgical segmentation models from the surgeon's perspective. Of the 3,355 academic articles found in the original search, 71 were found to be eligible according to our inclusion/exclusion criteria and were systematically reviewed. The majority of the studies were from general surgery or gastroenterology, and the GI tract was the most segmented anatomical system (Figures 1.2, 1.3). While there is a clear correlation between the anatomical site and the medical discipline it is unclear why other medical disciplines are underrepresented in current segmentation models. This may be partially explained by the availability of open-source and limited-access surgical datasets. Of the 16 open-source datasets and one limited-access database of the Japanese Society of Endoscopic Surgery (JSES), six (35.3%) contain images or videos of the GI tract. CV models of ophthalmic surgeries, a discipline with good representation (9.7%), relied on open-source data (CATARACTS dataset) in 33.3% of the studies. These findings likely explain the low number of models in some disciplines, such as neurosurgery (1.4%), with few or no available open-source data. Therefore, to increase the spread of CV technology to various medical fields, it is imperative to develop high-quality open-source datasets. This data can be derived from volunteers, anonymized real medical or surgical data or new data created with generative AI (9).

The most common segmentation tasks included the detection of pathological changes in normal anatomy, identification of surgical instruments and recognition of surgical steps or phases (Figure 2.4). Most models (71%) focused on one of these three categories. There was a redundancy in each of these categories despite existing models with good accuracy. For powerful pathology detection models, the use of multiple sources may be logical due to the extreme variability of this type of data and the high cost of overlooking. However, the detection of

surgical instruments requires less data and many instruments are shared between different surgical disciplines. Moreover, robust algorithms for this purpose were previously published (10). Few of the analyzed models concentrated on the detection of surgical anatomy. It was mostly anatomical landmarks detection (10.5%), rather than specific anatomical structures (6.6%). Upgrading surgical CV technology from the proof-of-concept stage to the next phase with practical applications in real surgery requires a coordinated approach. Similarly to The Human Genome Project (11), it is time for complete segmentation of the human anatomy. This can be achieved only by organized and systematic collection of labelled surgical data, system by system and organ by organ. While this task is undoubtedly challenging and requires significant time and resources, the benefits it yields are even greater and well worth the effort. The surgical robot will evolve from a device that passively mimics the surgeon's movements to an advanced system that enhances surgical precision and assists with operating decision-making. This will pave the way for autonomous surgical machines.

This review found that CNNs were the backbone of most CV models, which is not surprising. By exploring the neural network architectures employed in the studies we can learn the fundamentals of CV models and analyze common optimization methods. ResNet (Residual Network), the most common type of CNN in this review was published in 2015 by Microsoft Research (12,13). Earlier neural network designs, like AlexNet and VGGNet, faced a challenge; as more layers were added during training, accuracy began to decline. ResNet architecture solved this problem by enabling the use of extremely deep neural networks while maintaining very high accuracy. The original publication introduced a 152-layer residual network that achieved a staggering 3.57% error rate, surpassing the average 5% human error rate (13).

In machine learning (ML), the “learning” is either by supervised or unsupervised methods or in some cases a mix of the two (semi-supervised). In supervised learning, the model is trained on labelled data. With high-quality labels and data, this is the most accurate training method, but also the most resource-intensive. When labelled data is unavailable or resources are scarce unsupervised or semi-supervised methods are used. Unsupervised learning can be used for exploring unknown patterns or categories, such as the study of genes or diseases (14,15). For medical CV projects, the supervised approach is the most common and associated with the best results (16). Semi-supervised learning is based on partially labelled data. Reinforcement learning, another common type of ML is based on rewards and penalties. It can be used for robotic motion, decision-making and drug dosing. For training a fully automated robotic system, the combination of multiple ML methods is prudent.

In this review, analyzing surgical CV systems, it was found that supervised ML was used consistently across all the models. Ten percent of the models incorporated transferred learning. This method reuses models trained on related data at the starting point of model training. Alternatively, labelled data (not necessarily medical) is used for pre-training. This improves model performance on the target task, especially when the new dataset is small (17). ImageNet and Microsoft COCO were the most common sources for pre-training.

1. **ImageNet** - Large visual database with millions of annotated images organized according to the WordNet hierarchy, where each node is depicted by hundreds or thousands of images.

   URL: http://www.image-net.org/

2. **COCO (Common Objects in Context)** - Contains over 330,000 images, with more than 200,000 labelled images and over 1.5 million object instances.

   URL: https://cocodataset.org/

3. **YouTube-8M** - Labelled video dataset that consists of millions of YouTube videos with high-quality annotations of 3,800 visual entities.

   URL: https://research.google.com/youtube8m/

Hosseinzadeh et al. (17) tested transferred learning from general objects to medical images. They found that for image segmentation, it is more important to have detailed, specific features rather than broad, general ones and self-supervised learning models on ImageNet outperform supervised ones. They conclude that ImageNet models can be improved with training on data specific to the medical field. This statement underscores the significance of developing extensive, open-source databases of medical images.

Twelve percent of the studies employed other optimization methods such as SENet, YOLO or YOLACAT.

1. **SENet** (Squeeze-and-Excitation Network) - Enhances feature representation, best for detailed classification tasks.
2. **YOLO** (You Only Look Once) - Fast object detection, optimal for real-time applications.
3. **YOLACT** (You Only Look At CoefficienTs) - Combines fast detection with instance segmentation, suitable for applications needing both speed and detail.

By combining all the information from this review, we can summarize the basics of a surgical CV project. First, it's essential to define the objects of interest and the segmentation task. Since data from existing models can be transferred to new models, it's important to focus on areas with

low CV coverage to avoid redundancy. The next step is to select the data source. Some open-source datasets were previously listed. If a new dataset is created, it is advisable to make it publicly available whenever possible. The quality of the labels should be addressed, and specialized personnel should be involved when labelling difficult anatomical structures. Since there are effective models for delineating the borders of a structure by drawing a bounding box or clicking on the structure (e.g., Facebook's Segment Anything Model), the use of polygons to create manual masks is unnecessary (18). Once the labelled dataset is created, the model can be built and trained. The appropriate neural network architectures can be selected such as ResNet and optimization techniques like YOLACT can be utilized. Supervised learning is the most popular training method. Semi-supervised techniques are employed when the number of labels is limited. Transferred learning is important for improving performance. Optimal pre-training involves using medical or surgical data, which is scarce. Nonetheless, pre-training with general objects (e.g., ImageNet dataset) is also effective. Although not covered by this review, it is equally crucial to consider the content of the testing set and model performance metrics. Finally, making the trained model available on code-sharing platforms like GitHub can facilitate knowledge transfer to future projects.

Strengths and Limitations

The strength of this review is the systematic analysis of a large number of studies, providing a comprehensive overview of current surgical CV models. It offers insights into areas with low CV coverage, provides sources of surgical data, and outlines a general framework for model structure and training.

This review has limitations. It mainly covers general surgery and gastroenterology and may not represent the full scope of surgical CV models. Some studies may be overlooked when scanning

a single archive with strict search criteria. This review does not cover commercial CV models that have not been published in the academic literature. Since conducted by a single reviewer some selection biases and subjective conclusions may persist.

Conclusion

This work provides a detailed analysis of surgical segmentation models from a surgeon's perspective, identifying trends and gaps in the current research. It offers valuable recommendations for advancing surgical CV technology. These efforts are important for transitioning CV models from proof-of-concept to practical applications, upgrading MIS and paving the way for autonomous surgical systems.

References

1. Kenneth D-H. Introduction. A Practical Introduction to Computer Vision with OpenCV. 1st ed. John Wiley & Sons, Incorporated; 2014. p. 29–37.
2. Abitbol J, Cohn R, Hunter S, Rombaldi M, Cohen E, Kessous R, et al. Minimizing pain medication use and its associated costs following robotic surgery. Gynecol Oncol. 2017 Jan;144(1):187–92.
3. Lavoué V, Gotlieb W. Benefits of Minimal Access Surgery in Elderly Patients with Pelvic Cancer. Cancers (Basel). 2016 Jan 12;8(1).
4. Caroff DA, Chan C, Kleinman K, Calderwood MS, Wolf R, Wick EC, et al. Association of open approach vs laparoscopic approach with risk of surgical site infection after colon surgery. JAMA Netw Open. 2019 Oct 2;2(10):e1913570.
5. Spaner SJ, Warnock GL. A brief history of endoscopy, laparoscopy, and laparoscopic surgery. J Laparoendosc Adv Surg Tech A. 1997 Dec;7(6):369–73.
6. Tong Y, Lu W, Yu Y, Shen Y. Application of machine learning in ophthalmic imaging modalities. Eye Vis (Lond). 2020 Apr 16;7:22.
7. Chiurillo I, Sha RM, Robertson FC, Liu J, Li J, Le Mau H, et al. High-Accuracy Neuro-Navigation with Computer Vision for Frameless Registration and Real-Time Tracking. Bioengineering (Basel). 2023 Dec 7;10(12).
8. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev. 2021 Mar 29;10(1):89.
9. Musalamadugu TS, Kannan H. Generative AI for medical imaging analysis and applications. Future Medicine AI. 2023 Sep 6;

10. Nema S, Vachhani L. Surgical instrument detection and tracking technologies: Automating dataset labeling for surgical skill assessment. Front Robot AI. 2022 Nov 4;9:1030846.
11. Randal J. The human genome project. Lancet. 1989 Dec 30;2(8678–8679):1535–6.
12. Elhassouny A, Smarandache F. Trends in deep convolutional neural Networks architectures: a review. 2019 International Conference of Computer Science and Renewable Energies (ICCSRE). IEEE; 2019. p. 1–8.
13. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016. p. 770–8.
14. Yun T, Cosentino J, Behsaz B, McCaw ZR, Hill D, Luben R, et al. Unsupervised representation learning improves genomic discovery and risk prediction for respiratory and circulatory functions and diseases. medRxiv. 2023 Aug 29;
15. Zamani N, Russell P, Lantz H, Hoeppner MP, Meadows JR, Vijay N, et al. Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics. 2013 May 24;14:347.
16. Chen X, Wang X, Zhang K, Fung K-M, Thai TC, Moore K, et al. Recent advances and clinical applications of deep learning in medical image analysis. Med Image Anal. 2022 Jul;79:102444.
17. Hosseinzadeh Taher MR, Haghighi F, Feng R, Gotway MB, Liang J. A systematic benchmarking analysis of transfer learning for medical image analysis. Domain Adapt Represent Transf Afford Healthc AI Resour Divers Glob Health (2021). 2021 Sep 21;12968:3–13.
18. Segment Anything | Meta AI [Internet]. [cited 2024 May 24]. Available from: https://segment-anything.com/

Chapter 2:
AI-Assisted Robotic Surgery for Gynecologic Cancer

Abstract (English)

Objectives: This work aims to develop and validate an AI algorithm for tissue delineation in robotic gynecologic cancer surgery.

Methods: Videos of robotic surgeries for gynecologic cancer were collected and catalogued. The timestamps for each surgical phase were recorded. Instances of anatomical structures such as the ureter, obturator nerve, IP ligament, and internal and external iliac arteries were also documented. Each frame was labelled by a surgeon using the polygonal mask method. With this data, a surgical computer vision algorithm was created employing supervised training and transferred learning methods.

Results: From 2013 to 2022, 897 videos were recorded. 4,526 frames were extracted and 6,040 masks were created. The AP in the training set was 42.94%. AP at 50% IOU was 87.6% and 37.5% with a 75% IOU threshold. Smaller objects had lower AP, 23.92% compared to medium and large objects, 53.14% and 57.34%, respectively.

Conclusion: This is the first CV model focusing on the automatic recognition of key pelvic structures with high segmentation complexity. Future versions with improved performance coupled with existing algorithms for surgical tool recognition could be implemented in robotic surgeries assisting in surgical navigation, increasing surgical precision and reducing intra-operative complications.

The proof of concept established here sets the stage for future research and implementation of CV in robotic surgeries.

Abstract (French)

Objectifs: Ce travail vise à développer et valider un algorithme d'IA pour la délimitation des tissus en chirurgie gynécologique robotisée du cancer.

Méthodes: Des vidéos de chirurgies robotiques pour le cancer gynécologique ont été collectées et cataloguées. Les horodatages de chaque phase chirurgicale ont été enregistrés. Des cas de structures anatomiques telles que l'uretère, le nerf obturateur, le ligament IP et les artères iliaques internes et externes ont également été documentés. Chaque image a été étiquetée par un chirurgien selon la méthode du masque polygonal. Avec ces données, un algorithme chirurgical de vision par ordinateur a été créé utilisant une formation supervisée et des méthodes d'apprentissage transférées.

Résultats: De 2013 à 2022, 897 vidéos ont été enregistrées. 4,526 images ont été extraites et 6,040 masques ont été créés. L'AP dans l'ensemble de formation était de 42.94%. AP à 50% d'IOU était de 87.6% et 37.5% avec un seuil de 75% d'IOU. Les objets plus petits avaient un AP inférieur, 23.92%, par rapport aux objets moyens et grands, 53.14% et 57.34%, respectivement.

Conclusion: Il s'agit du premier modèle CV axé sur la reconnaissance automatique des structures pelviennes clés présentant une complexité de segmentation élevée. Les futures versions offrant des performances améliorées, associées aux algorithmes existants de reconnaissance des outils chirurgicaux, pourraient être mises en œuvre dans les chirurgies robotisées facilitant la navigation chirurgicale, augmentant ainsi la précision chirurgicale et réduisant les complications peropératoires.

La preuve de concept établie ici ouvre la voie à de futures recherches et à la mise en œuvre du CV dans les chirurgies robotiques.

Introduction

The rationale for this work is the integration of computer vision (CV) into minimally invasive surgery (MIS). Surgery has evolved from large incisions and inert tools to operations performed through keyhole openings with sophisticated devices. While effective, this technology has integral challenges related to abdominal access, intra-abdominal navigation and access to deep tissues. The lack of haptic feedback in robotic surgery (RS) demands advanced surgical skills and complicates tumor palpation to verify its complete removal. CV technology holds the potential to overcome all these challenges and beyond. By leveraging its capabilities, surgical precision can be increased, reducing complications and improving oncological outcomes. The integration of CV with RS is the first step towards autonomous surgical systems.

The objective of this project is to develop a CV model for the automatic identification of complex pelvic structures encountered in gynecologic oncology surgeries.

Literature review

This literature review will explore three fundamental domains: surgical treatment of gynecological cancer, minimally invasive surgery (MIS), and computer vision. The final part of the review will concentrate on integrating computer vision with MIS methods to improve the surgical treatment of gynecological cancer.

Surgical treatment of gynecological cancer

Surgery is a key factor in the diagnosis and management of gynecological cancers. Worldwide, more than one million surgeries are performed each year for the three most common gynecological cancers: endometrial, ovarian and cervical. One-tenth of these surgeries are performed in northern America (Table 2.1).

Table 2.1: Gynecological cancer incidence, mortality and surgical burden

	Endometrial Ca.	Ovarian Ca.	Cervical Ca.	Total
Incidence				
World	420,368	324,603	662,301	1,407,272
Northern America	73,977	24,484	15,654	114,115
Mortality				
World	97,723	206,956	348,874	653,553
Northern America	13,543	15,554	6,692	35,789
Surgeries *(est.)*				
World	~407,000	~292,000	~331,000	1,030,000
Northern America	~71,000	~22,000	~7,000	100,000

- *Source: International Agency for Research on Cancer, WHO* (1).
- *Surgeries are estimated by assuming that most Stage I-III endometrial (97%) and ovarian (90%) cancers are treated surgically. For cervical cancer surgery is performed mostly for localized disease and only for a small portion of regional and distant cases – 50% of all cases.*

Most cases of endometrial cancer (EC), the most common gynecological cancer in northern America, are confined to the uterus at diagnosis (2). Surgery is the main treatment, with adjuvant radiotherapy prescribed according to patients' age and histological risk factors. Chemotherapy is

reserved for patients with advanced-stage or high-grade cancers (3). The standard surgery includes total hysterectomy (removal of the uterus body and cervix), bilateral adnexectomy (removal of both ovaries and fallopian tubes) and bilateral lymphadenectomy (removal of pelvic and para-aortic lymph nodes). The lymphadenectomy can be either complete (removal of all the lymph nodes in the area) or by mapping and removing sentinel lymph nodes only. In some very low-risk patients, the lymphadenectomy may be omitted, on the other hand, any suspicious nodes must be removed (4). Additional procedures may be performed for staging or disease control, such as omentectomy and tumor debulking.

The surgery can be performed either by laparotomy or by minimally invasive methods (next section). MIS is widely used and recommended for the surgical treatment of EC (4). Although hysterectomy can be performed vaginally, this method is not recommended for oncological cases and will not be discussed.

Most patients (80%) undergo these surgeries without or with only mild complications (5). Nausea, ileus and pain are the most frequent complications. Blood transfusions and infections are also common. More severe, infrequent complications, include wound dehiscence, thromboembolic events and injuries to vital organs (5). Some complications can be mitigated with MIS, but the rate of intra-operative organ injuries is similar or even higher in some reports (6). Bladder, bowel, ureteric and vascular injuries are most prominent.

Ovarian cancer (OC), the second most common gynecological cancer in North America, is usually in advanced stage when diagnosed (7). Surgery remains the main component of treatment with adjuvant or neo-adjuvant chemotherapy (NACT). The surgical goal is the complete removal of macroscopic tumor tissue (debulking) and the amount of residual disease has a direct impact on patients' survival (7). In addition to tumor debulking, hysterectomy with bilateral

adnexectomy and omentectomy are routinely performed. Lymphadenectomy is reserved for bulky lymph nodes and staging. Tumor debulking may involve bowel resection, diaphragmatic stripping, splenectomy, peritonectomy and other radical procedures. In early-stage disease, the surgery is limited to the removal of the uterus with the adnexa, infracolic omentectomy, multiple biopsies of the peritoneum and any suspicious lesions, and selected pelvic and para-aortic lymphadenectomy.

Today, most of these surgeries are performed by minimally invasive methods. MIS is also possible in advanced OC, especially in cases of interval debulking (surgery after NACT) with a good response to NACT (8). Studies comparing oncological outcomes of interval debulking with primary debulking have shown similar results (9).

OC surgeries are associated with a higher rate of complications compared to EC, due to a higher ratio of advanced cases and more radical resections. Nevertheless, the types of complications are similar and MIS has not been shown to reduce the rate of organ injuries.

Cervical cancer is the most common gynecological cancer worldwide. However, it is the least common of the three cancers in Western countries, as shown in Table 2.1. This decrease is due to the universal implementation of HPV immunization and screening for pre-invasive cervical changes.

Like other cancers, the type of surgical treatment depends on the disease stage. Small tumors confined to the cervix with limited depth of invasion can be treated by conization (resection of the distal part of the cervix) or by simple hysterectomy with or without adnexectomy (10,11). Deeply invasive tumors limited to the upper two-thirds of the vagina can be treated by radical hysterectomy (RH) or by chemoradiation based on clinical factors. RH involves the removal of the uterus, the proximal part of the vagina, the uterosacral and uterovesical ligaments, and the

parametrium bilaterally with pelvic lymphadenectomy. Some advanced cases and recurrences are treated by pelvic exenteration – complete resection of the pelvic organs, including the urinary bladder and the distal part of the intestine (rectum and sigmoid).

While MIS is routinely used for simple hysterectomies and in selected pelvic exenterations, its use in RH has been restricted following the LACC study. This study found higher mortality and recurrence rates in patients after MIS compared to laparotomy (12).

The core of onco-gynecological surgeries is the excision of the uterus with the uterine cervix and the adnexa (ovaries and fallopian tubes) along with pelvic lymph node dissection. All these procedures can be performed by MIS.

Minimally Invasive Surgery (MIS)

MIS employs endoscopic techniques and specialized instrumentation to conduct surgical interventions through small (1-2cm) incisions, markedly reducing the physical trauma traditionally associated with open surgical procedures. To visualize the abdominal content and create operating space the abdomen is inflated with CO_2 (inert and transparent gas). Then, a camera with a light source is introduced into the abdominal cavity and the image is displayed on a screen. Next, surgical tools, specially designed for laparoscopy, are inserted. The surgeon looks at the screen and operates with the tools inside the abdomen (Figure 2.1).

Robotic surgery (RS) is an advanced evolution of laparoscopy. In robotic surgery, the surgeon operates a console that controls robotic arms docked to surgical instruments and a 3D camera. This setup provides enhanced dexterity, a greater range of motion, and a magnified, three-dimensional view of the surgical site. The robotic system translates the surgeon's hand movements into smaller, more precise movements. The surgeon is capable of performing

complex procedures with increased precision and reduced fatigue, given the ergonomic design of the console.

Figure 2.1: Image of laparoscopic surgery

- *Royalty-free image. Downloaded from https://pixabay.com/*

Although experiments with laparoscopic devices started more than a century ago, real laparoscopies began only in the 70s after the development of fiber optic technology (13). MIS was initially adopted by gynecologists for tubal ligation. Ten years later, in 1988, the first laparoscopic cholecystectomy (excision of the gall bladder) was performed and after a relatively short adaptation period, it became the standard of care. The evident advantages of MIS, such as less post-operative pain, faster recovery and minimal scaring have contributed to its widespread

recognition and acceptance (14,15). Reducing post-operative pain, not only improves patients' satisfaction but also decreases the need for pain medication and the associated side effects. Furthermore, improved pain control after surgeries is associated with lower rates of other surgical complications, such as atelectasis (collapsed lung sections) and thromboembolic events because less abdominal pain enables deeper inspiration and earlier mobilization. This results in faster recovery and shorter hospital stay, which increase surgical service capacities and reduce operative costs. Oncological patients also benefit from shorter intervals between surgery and adjuvant chemotherapy.

The introduction of MIS into surgical oncology was more gradual. Although diagnostic laparoscopic procedures had started in 1971, therapeutic MIS began twenty years later with the development of advanced coagulation and tissue dissection tools (13). The first laparoscopic hysterectomy was reported in 1989 and pelvic lymphadenectomy in 1993 (16). In the following years, minimally invasive gynecologic oncology was continuously developed and today this is the standard method for EC (4). Obesity is common among patients with endometrial cancer and MIS has proven to be beneficial for these cases (17,18).

Robotic surgery debuted 25 years ago with the first report of robotically assisted tubal re-anastomosis (19,20). Due to improved ergonomics and visualization, this MIS method is gaining popularity in gynecologic oncology enabling to perform the most sophisticated surgeries with great precision. *Casarin et al*. explored the modes of endometrial cancer surgery from 2008 to 2015 and found an increase from 9% to 57% of these surgeries performed robotically (21). They showed that RS had better outcomes and comparable costs. The feasibility of RS in ovarian cancer after NACT was affirmed by *Abitbol et al*. and others (22–24).

While the benefits of MIS are clear, some concerns regarding the use of this technique in oncology exist. First, the method requires special training and expensive equipment not readily available. Second, it requires larger operating rooms and additional operating time. Palpation is not possible and there is a concern about missing tumor tissue or inadequate tumor resection. There were reports on port-site metastasis and the negative effect of pneumoperitoneum on tumor spread (25–27). While concerning, many of these claims were disputed. Still, no prospective studies comparing MIS with laparotomy for ovarian cancer are available, and the retrospective data is inconclusive, mandating judicious use.

The advanced capabilities of robotic systems enable us to perform the most complex oncological procedures – debulking for ovarian cancer and pelvic exenteration for recurrent cervical cancer. The challenge is to develop MIS further to make it safer, faster and easier to perform. This, potentially, will translate to cost reduction and improved surgical and oncological outcomes. One compelling method to achieve that is by incorporating computer vision into MIS systems.

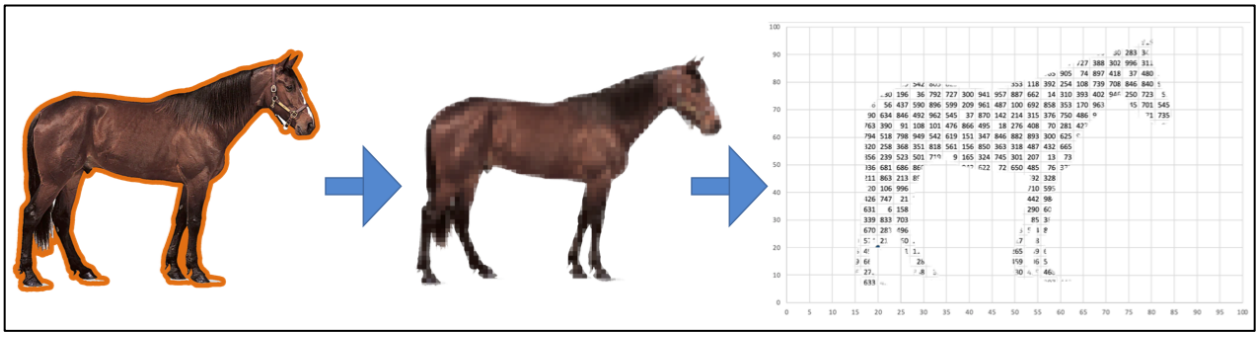
Computer Vision (CV)

Computer vision (CV) is a group of artificial intelligence (AI) algorithms that process visual data and make decisions or provide instructions. The core of CV models is the ability to recognize patterns in a digital image. Training a computer to "understand" the scene requires a substantial amount of data and sophisticated machine-learning methods.

Similar to a baby learning its environment, a computer needs to learn from multiple different images. The most common method is supervised learning, where the objects in the image are outlined and labelled. The machine "sees" a digital image as a matrix of numbers. Each cell in this matrix is a pixel (colored dot). The location of the cell in the matrix is the location of the pixel in the image, and the number stored in this cell is the color of this pixel (Figure 2.2). By

outlining an object in the image, the computer learns its shape and colors. When a new image is presented, it will try to identify familiar patterns and classify the objects (object detection). For example, an algorithm to identify a tumor on a snapshot from surgery requires multiple images of surgical scenes with and without tumors for training. Each image has to be labelled by marking the areas of the tumor and the normal tissue. A portion (usually 20-25%) of these images is used for testing the algorithm's accuracy (internal validation). The accuracy can be improved by choosing a different training model and by adding labelled data.

Figure 2.2: The digital image (illustration)

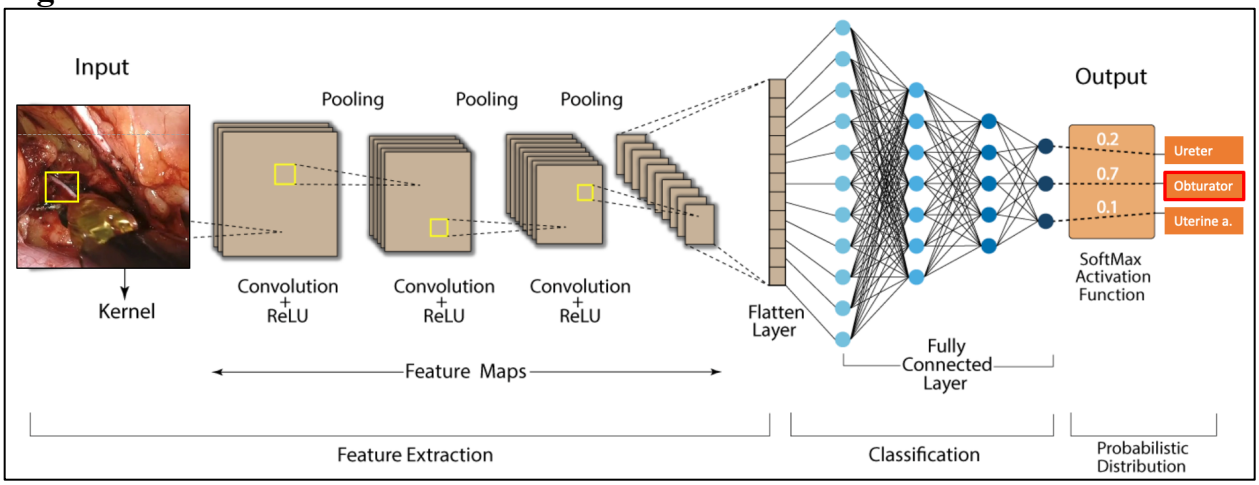


The introduction of neural networks revolutionized computer vision and many models available today surpass human vision capabilities. Deep learning is a subset of machine learning that involves neural networks with many layers. Convolutional neural networks (CNNs) are a specific type of deep learning model that incorporates convolutional layers. CNNs are used with great success for object detection, image segmentation, facial recognition, and medical image analysis (28). Their convolutional part involves a filter (kernel) applied to an image to produce a feature map. This operation effectively captures patterns in an image, such as edges, textures, and shapes. The number of layers (network depth) is equal to the number of filters applied. By stacking multiple convolutional layers, with each layer capable of recognizing more complex

patterns, CNNs can learn increasingly complex features (29,30). This makes them particularly effective for image and video recognition (Figure 2.3). U-Net, DeepUnet, ResNet and DenseNet are the most popular CNN models used for medical image analysis (30).

Performance metrics are used to compare and assess machine learning models. For CV models, the interception over union (IOU) is used to evaluate the accuracy of object detection (Figure 2.4).

Figure 2.3: Convolutional neural network architecture

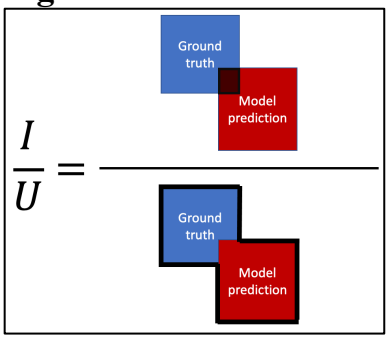


Interception is the area of overlap between the ground truth (manual mask of the object) and the model prediction. Union is the sum of these areas. The best possible IOU is 100% when the masks are identical. In reality, the prediction cannot be perfect and a threshold is used according to task requirements. Mean average precision (mAP) is an important metric to estimate the detection of multiple objects within an image. It is a combination of precision ($\frac{TP}{TP+FP}$) and recall ($\frac{TP}{TP+FN}$). While precision is the ratio of correct predictions from all the predictions, recall is the ratio of correct predictions from all the possible objects. Average precision (AP) is calculated

individually for each class of objects. In a model designed to detect surgical tools, tumors and blood vessels in an image of a surgical scene, the AP is calculated separately for each type (class) of objects. The AP is computed by plotting several precision-recall curves with different IOU thresholds. This method avoids the problem of choosing a single IOU threshold to evaluate the model. Finally, the mAP is the average of APs of all the classes. Today, mAP is the standard metric used by computer vision researchers to evaluate object detection models.

Existing CV models were able to recognize the surgical phases of cholecystectomy with 90% accuracy (31,32) and evaluate surgical skills with high precision (33). A robust model for detecting polyps during colonoscopy was developed (34) and strong models for surgical tools detection exist (35). The existing data and technology can be developed to enhance many minimally invasive surgical procedures, including gynecologic oncology.

Figure 2.4: Intersection over Union (IOU)



Integrating CV with MIS to improve surgical and oncological outcomes

Combining CV with MIS is like mixing two powerful ingredients to create a super medicine. The unique set-up of laparoscopic and robotic surgery creates a solid foundation for the application of computer vision technologies (32). This fusion can make surgeries more precise and efficient and set the stage for the development of autonomous surgery in the future.

Complete tumor removal is a key component of oncologic surgery and is strongly associated with progression-free and overall survival (36). Existing methods to improve intra-surgical tumor detection include pre-operative imaging, palpation of abdominal organs and infusion of tumor-specific fluorescence dyes (37). Since RS systems lack haptic feedback, palpation is not possible and the imaging techniques are invasive and associated with adverse effects. Tumor detection by CV is a non-invasive and potentially very effective method. Furthermore, it may facilitate the detection of microscopic and "hidden" tumor tissue by incorporating hyperspectral imaging and fusion with computed tomography (CT) scans (38,39).

Surgical navigation is another domain that can be enhanced with CV. Some tumors are very challenging to locate, due to their size and location. Although detected pre-operatively by positron emission tomography (PET) scanning, it takes a great deal of time to find and resect them. Usually, they are located retroperitoneally, or deeply embedded in abdominal tissue, impossible to see with the naked eye. CV algorithm can assist in locating those tumors by fusion of PET and CT scans and surgical image registration. Every modern smartphone contains a gyroscope and accelerometer and these sensors are small enough to be installed in a dissection tool. Then, by using an accessible anatomical point as a reference (e.g. aortic bifurcation) the AI model can navigate the surgeon from this point to the tumor.

Lymph node dissection, a routine procedure of many oncological surgeries, can cause significant injuries to major blood vessels, the obturator nerve and the ureters. The resection of tumor tissue from healthy organs and dissection of the bladder from the uterine cervix during hysterectomy are challenging and associated with significant morbidity. Some of these complications may be avoided by CV methods. First, the computer can assist visually by delineating important anatomical structures. Second, it can use audio alerts to warn the surgeon when a surgical tool is

dangerously close to a vital organ. Lastly, the tools themselves can be manipulated by the computer when in proximity to certain structures. For example, thermal injuries of the ureter are caused by electrocautery tools during dissection in the area. CV algorithm trained to detect ureters can manipulate the energy of the tool according to its distance to the ureter, potentially avoiding those injuries.

CV models for tumor detection, surgical navigation or dissection require training on labelled surgical data. Similar to surgeons, comprehensive anatomical knowledge is an essential component of any surgical CV algorithm. In this project, a large database of surgical images was created and labelled to produce a working prototype of a surgical CV model.

Methods

Data collection

Video data was collected from consecutive robotic surgeries in the Department of Gynecologic Oncology at the Jewish General Hospital (JGH) from 2013 to 2022. The videos were captured by Intuitive da Vinci surgical system proprietary endoscopic camera and recorder. The recording started with the activation of the camera and concluded following its retraction from the abdominal cavity at the end of the surgery. Segments of the procedure before introducing the camera and after its removal were not captured on video. In total 897 videos were collected and organized into a database. Most videos are between 2-4 hours in length.

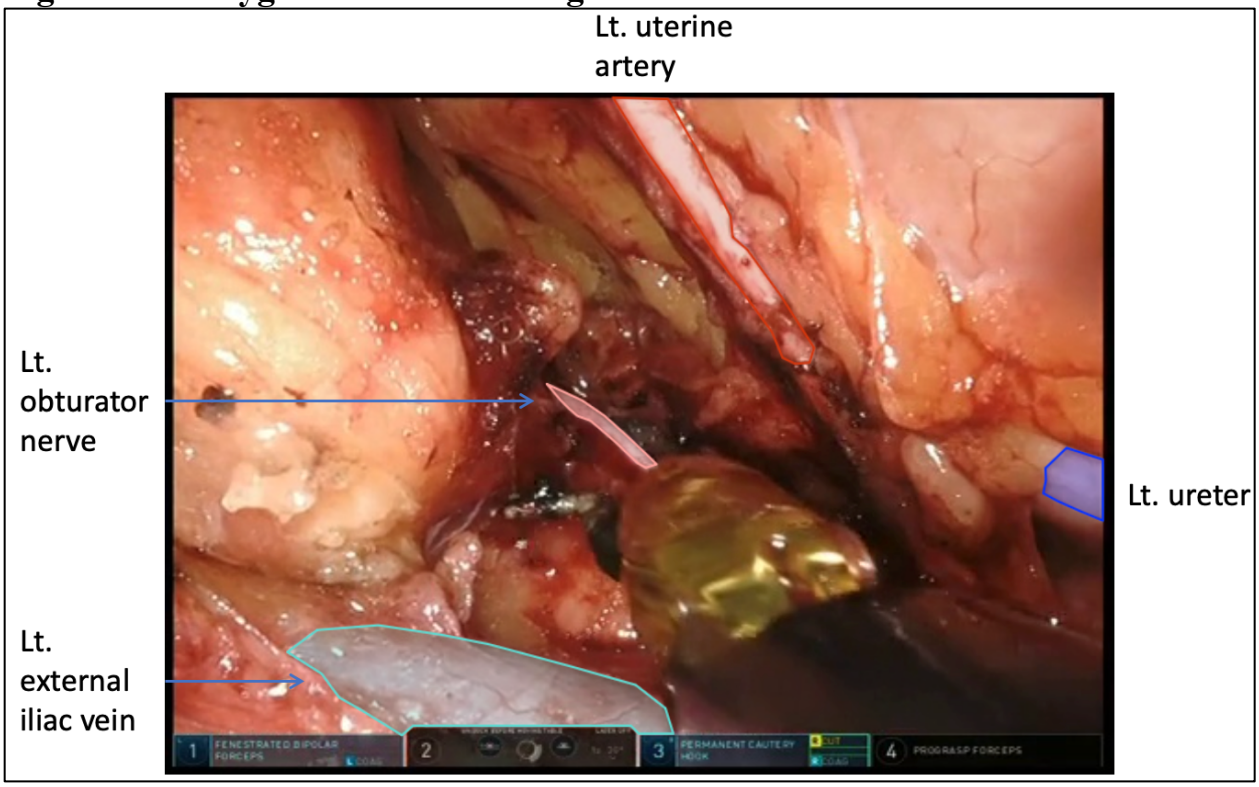
Data labelling

The supervised machine-learning method was used for algorithm training. The video content was catalogued by surgical phases and anatomical structures with the corresponding timestamps (Table 2.2). Clips with ureters and obturator nerves were used for frame extraction. Label Studio ver. 1.11.0, an open-source data labelling platform, was used for labelling. The frames were uploaded to the labelling platform and annotated by a surgeon (the author of this thesis) utilizing both semantic and polygonal segmentation methods (Figure 2.5). In addition to labelling the structures of interest (ureters and obturator nerves), adjacent anatomical structures were also segmented to increase the algorithm's accuracy. Frames of indocyanine green (ICG) imaging, used for sentinel lymph node mapping, were also labelled.

Table 2.2: Surgical recordings catalogue (sample)

	Surgery ID	Clip name	Surgical Phase	Anatomical Structure	Side	Start Time	End Time	Comments
1	***YO12345***	*Vid001*	*Retroperitoneal Dissection*	*Ureter*	*Right*	*1:10:15*	*1:12:29*	*Bleeding*
2	---	*Vid002*	*Retroperitoneal Dissection*	*Ureter*	*Left*	*1:42:12*	*1:53:41*	---
3	---	*Vid003*	*Lymphadenectomy*	*Obturator n.*	*Right*	*2:12:23*	*2:17:56*	*EndoBag*
4	***AV67890***	*Vid001*	*Uretrolysis*	*Ureter*	*Right*	*0:55:43*	*1:06:12*	
5	---	*Vid002*	*Uretrolysis*	*Ureter*	*Left*	*1:31:01*	*1:35:23*	
6	---	*Vid003*	*Vaginal vault closure*	*Ureter*	*Both*	*2:36:12*	*2:45:10*	*Gauze*
7	***BR54321***	*Vid001*	*Ligation of the IP ligament*	*Ureter*	*Right*	*1:14:32*	*1:17:33*	
8	---	*Vid002*	*Lymphadenectomy*	*Obturator n.*	*Left*	*1:30:12*	*1:34:55*	*ICG*
9	---	*Vid003*	*Ligation of the uterine a.*	*Ureter*	*Left*	*1:50:55*	*1:53:33*	
10	***EZ09876***	*Vid001*	*Ligation of the IP ligament*	*IP ligament*	*Right*	*1:33:18*	*1:36:20*	
11	---	---	---	*Ureter*	*Right*	*1:34:22*	*1:34:50*	
12	---	*Vid002*	*Retroperitoneal Dissection*	*Ureter*	*Left*	*2:18:22*	*2:23:14*	
13		---	*Retroperitoneal Dissection*	*Internal iliac a.*	*Left*	*2:19:20*	*2:22:12*	*Bleeding*
14	***IN16505***	*Vid001*	*Uretrolysis*	*Ureter*	*Left*	*0:45:14*	*0:49:23*	
15	---	*Vid002*	*Uretrolysis*	*Ureter*	*Left*	*0:50:12*	*0:55:22*	

Figure 2.5: Polygonal mask labelling method with Label Studio



Data splits

The training and testing sets were designed to include frames from different patients. In real life, the technique of each surgery is slightly different and some anatomical variations between patients exist. When frames are extracted sequentially from a video of the same surgery, similar and even identical frames are common, especially during periods of minimal surgical activity. Splitting such images randomly between the training and the testing sets can artificially increase the model performance. To overcome this phenomenon, different clips from the same surgery were allocated to the training set and one surgery was chosen to serve exclusively as the testing set. This method allows more credible model testing and simulates a more realistic situation.

Image format selection

Frames, i.e., snapshots of surgical scenes extracted from surgical videos, were stored and processed in Portable Network Graphics (PNG) format. In comparison to the more commonly used Joint Photographic Experts Group (JPEG) format, PNG holds the following key advantages:

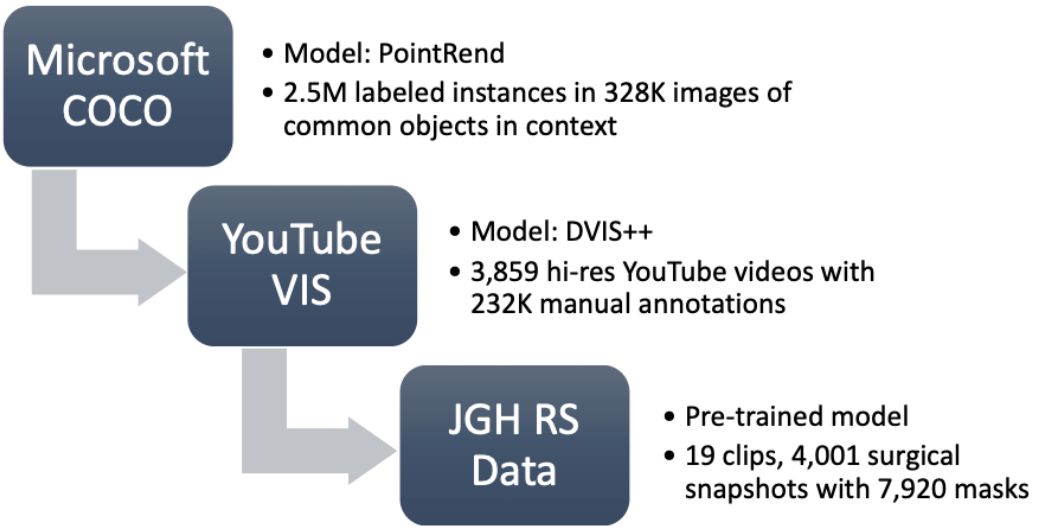
- **Lossless Compression** - The image data is preserved exactly as it was originally captured, without any loss of detail or quality.
- **Bit Depth** - PNG supports a higher bit depth than standard JPEG images, allowing for a broader range of colors and finer details.
- **Absence of Artifacts** – Images in JPEG format are compressed to reduce file size. This compression can introduce artifacts such as blurring and loss of details.
- **Consistency** – The lossless and artifact-free nature of the PNG format provides reproducibility across different computer vision experiments and ensures that the algorithm's performance is not influenced by variability in image quality.

Given these advantages, PNG was chosen as the preferred image format for this study.

Model selection and training

For optimal results, a combination of frame-based and video segmentation models was used incorporating transfer learning techniques (Figure 2.6).

Figure 2.6: Transferred learning method for model training

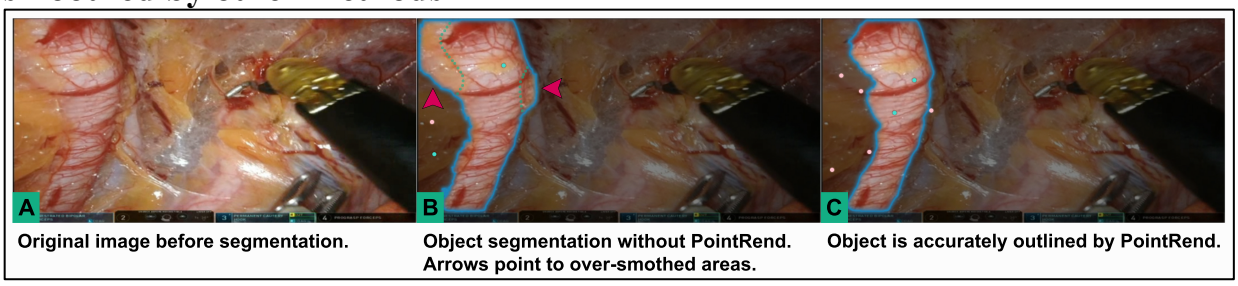


Transfer learning method for model training was used to improve the model performance on surgical data. The model was pre-trained on large scale common objects and videos repositories and than applied on the Jewish General Hospital (JGH) Robotic Surgeries (RS) dataset.

PointRend (Point-based Rendering) (40) is the image segmentation model that was chosen and video segmentation was performed with DVIS++ (Decoupled Video Segmentation) (41). PointRend was pre-trained on the Microsoft COCO (Common Objects in Context) dataset (42) and the YTVIS (YouTube Video Instance Segmentation) dataset (43) was used for DVIS++ pre-training. PoindRend enables smooth and accurate object delineation (Figure 2.7). This is an important attribute, especially for oncological surgeries when fine dissection of crucial anatomical structures is performed. Pre-training on the Microsoft COCO dataset, containing 2,500,000 labeled instances of 328,000 images utilizing over 70,000 worker hours provides the model solid capabilities for general world image segmentation. This technique of knowledge transfer between models effectively compensates for classification challenges imposed by training sets of limited size (44). The same principle was used for pre-training the DVIS++-based model utilizing 3,859 high-resolution YouTube videos – 2,985 training videos, 412 validation

videos and 453 test videos. In total, the YTVIS 2021 version repository contains 232,000 high-quality manual annotations (https://youtube-vos.org/dataset/vis/).

Figure 2.7: PointRend model outputs crisp object boundaries in areas over-smoothed by other methods

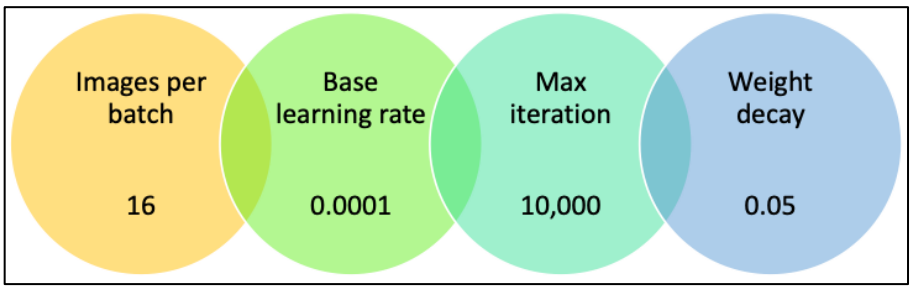


After pre-training, the final model was trained on the Jewish General Hospital dataset containing labelled data of gyne-oncological robotic surgeries.

Model parameters

The following hyperparameters were used for model tuning (Figure 2.8):

Figure 2.8: Video Classification Model Hyperparameters



- **Images per batch** – The number of images processed together in one pass (one batch) during the training of the model. A batch size of 16 means that the model will update its weights after learning from 16 images. This is a balanced number that can provide a stable estimate while being computationally manageable.

- **Base learning rate** – The initial rate at which the model learns. It determines the size of the steps taken during optimization. A learning rate of 0.0001 is relatively small, this cautious approach to learning where the model makes minor adjustments to weights can be more beneficial.
- **Max iteration** – The total number of training cycles the model will pass during training. Setting the max iterations to 10,000 ensures the model will have sufficient full dataset passes until convergence or satisfactory performance is achieved.
- **Weight decay** – A regularization technique used to prevent overfitting by penalizing large weights. With a weight decay of 0.05, the model applies a moderate penalty to the weights, which helps in generalizing the learning and avoiding overfitting the training data.

Performance metrics

The model was evaluated by traditional performance metrics for video classification tasks.

- **Precision** – The ratio of correct predictions from all the predictions made by the model ($\frac{TP}{TP+FP}$). High precision is especially important in tasks sensitive to false positive results including surgical scene segmentation. This metric is equivalent to the positive predictive value used in evaluating medical diagnostics.
- **Recall** – The ratio of correct predictions from all the existing objects ($\frac{TP}{TP+FN}$). Contrary to precision, recall estimates the model coverage or the ratio of objects missed by the model. This metric is identical to sensitivity in medical studies.
- **Intersection over union (IoU)** – Measures the overlap between the ground truth (defined manually by the labeller) and the model prediction $\left(\frac{Area\ of\ Intersection}{Area\ of\ Union}\right)$ on a scale

between 0 (no overlap) and 1 (perfect overlap). In this project, an IoU of 40% was set as the threshold for correct prediction due to the complexity of predicting anatomical structures.

- **Average precision (AP)** – Usually a tradeoff between precision and recall exists. For tasks requiring high performance in both metrics (most cases), AP is used to balance between these parameters. It provides a single metric to assess model quality across different precision and recall levels. In this study, the area under the precision-recall curve was used to calculate the AP.

$$AP = \sum_{k=1}^{N} P(k)\Delta r(k)$$

Ethical considerations

This study was approved by the Jewish General Hospital's ethics board and was conducted with a firm commitment to upholding ethical standards, particularly in terms of patient confidentiality. The video content was edited to eliminate any visuals that could potentially identify patients or surgical team members. All utilized patient data underwent irreversible de-identification, stripping away personal identifiers like names, birthdates, IDs, medical record numbers and Régie de l'assurance maladie du Québec (RAMQ) codes.

Results

In total, 4,526 frames from 19 clips were extracted with 6,040 masks. The majority of the data is from two main surgical phases: retroperitoneal dissection and lymph node extraction. The training set contained 5,465 (90%) labelled structures and the testing set 575 (10%). Table 2.3 shows the anatomical structures in each set.

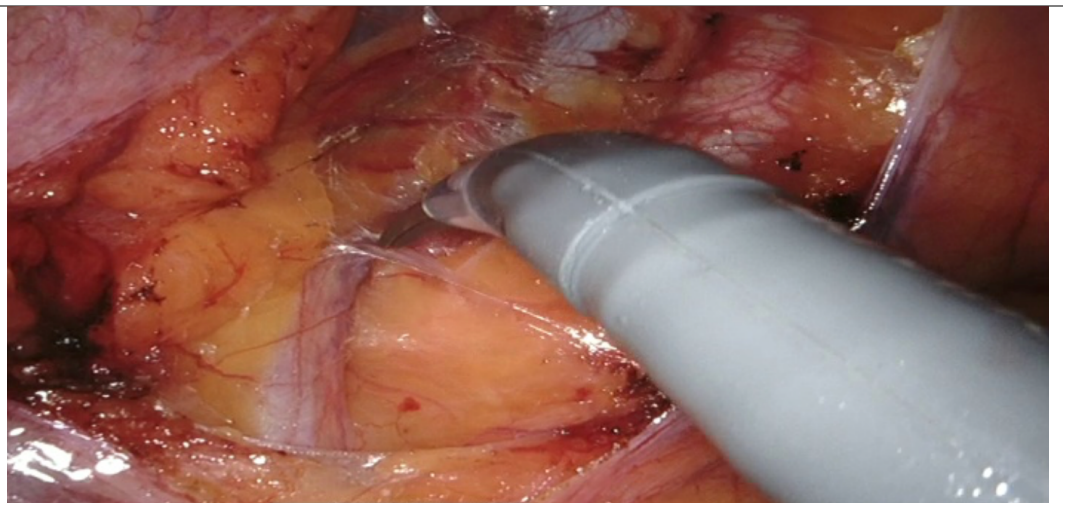
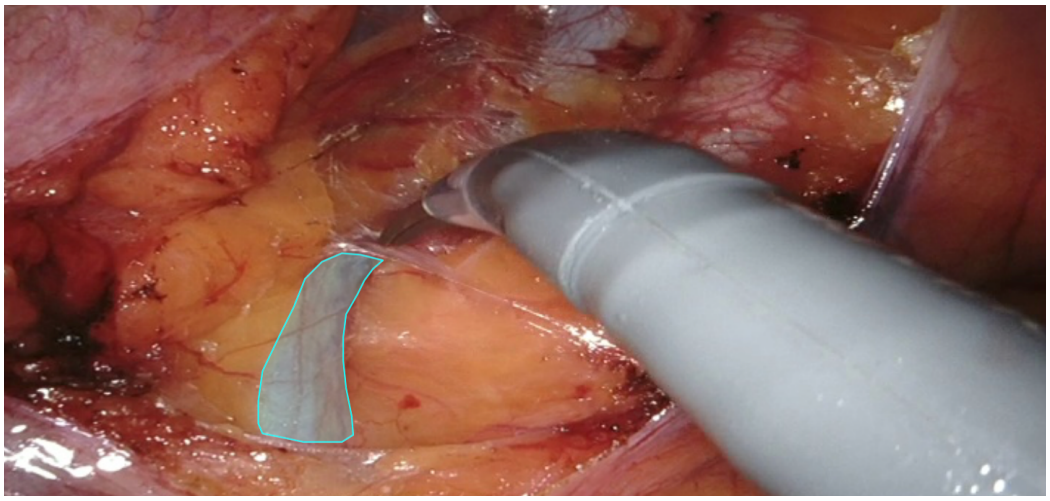
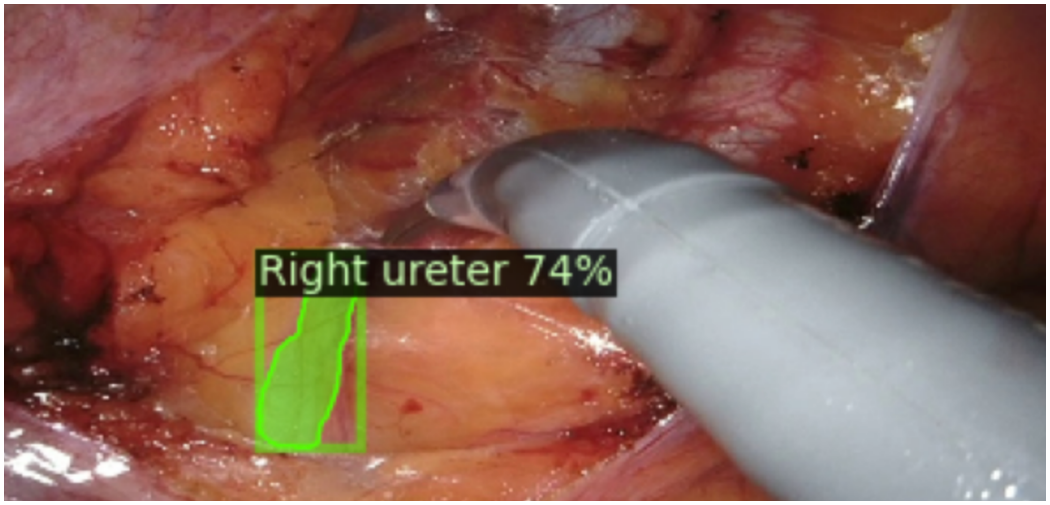
Table 2.3: Training and testing sets

	Anatomical Structure	Training Set	Testing Set	Total	Average Precision
1	IP ligament	133	63	196	---
2	External iliac artery	212	---	212	63.94%
3	Internal iliac vein	1	---	1	---
4	Internal iliac artery	399	---	399	32.89%
5	Obturator nerve	1742	---	1742	45.78%
6	Ureter	2978	512	3490	36.05%
		5465	**575**	**6040**	

The AP in the training set was 42.94%. AP at 50% IOU was 87.6% and 37.5% with a 75% IOU threshold. Smaller objects had lower AP, 23.92% compared to medium and large objects, 53.14% and 57.34%, respectively. Per structure, the external iliac artery had the highest AP - 63.94%, followed by the obturator nerve with 45.78%, the ureter with 36.05% and the internal iliac artery with 32.89% average precision (Table 2.3). In the external validation set, the mean average precision (mAP) was 10.4%.

Figure 2.9: Ground truth and model predictions on distinct images

A. Frame #4431 (right ureter)

Original frame	
Ground truth	
Model prediction	

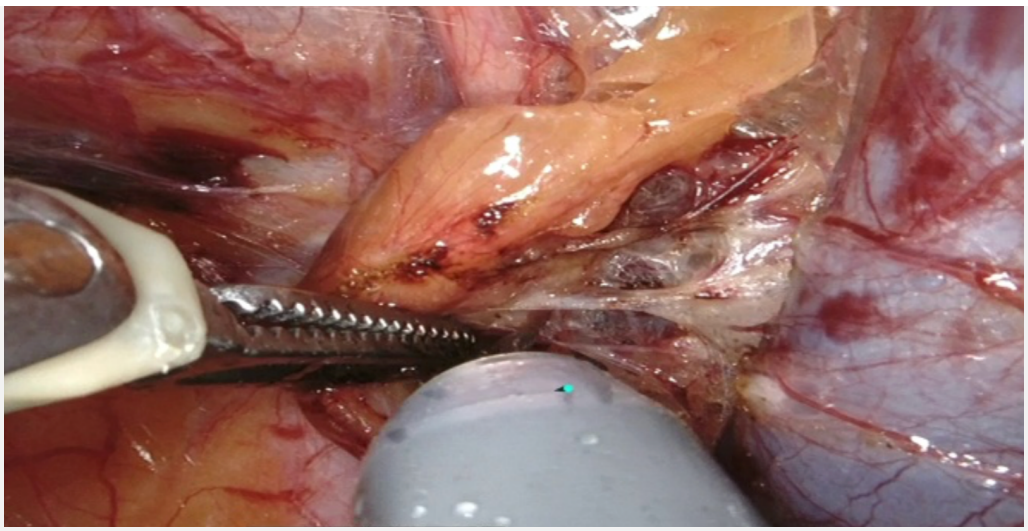
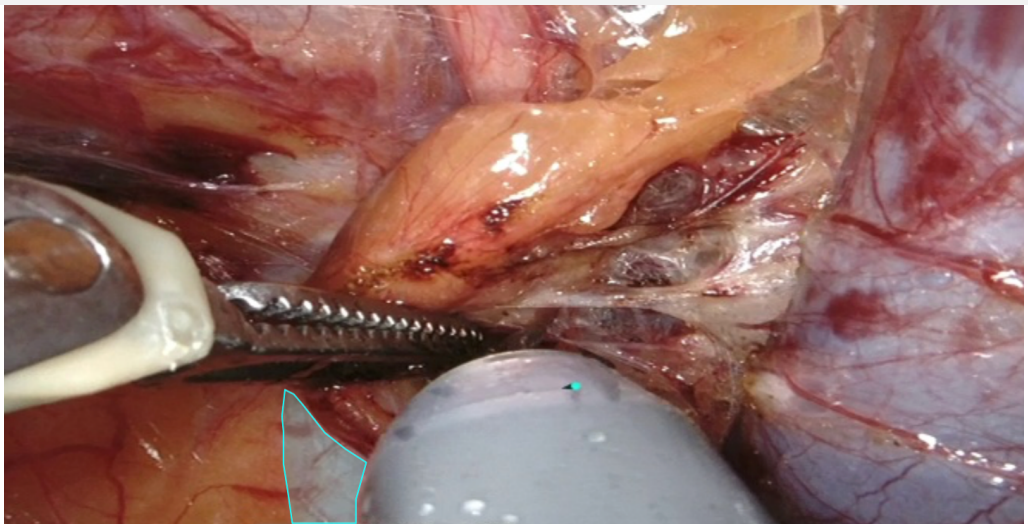
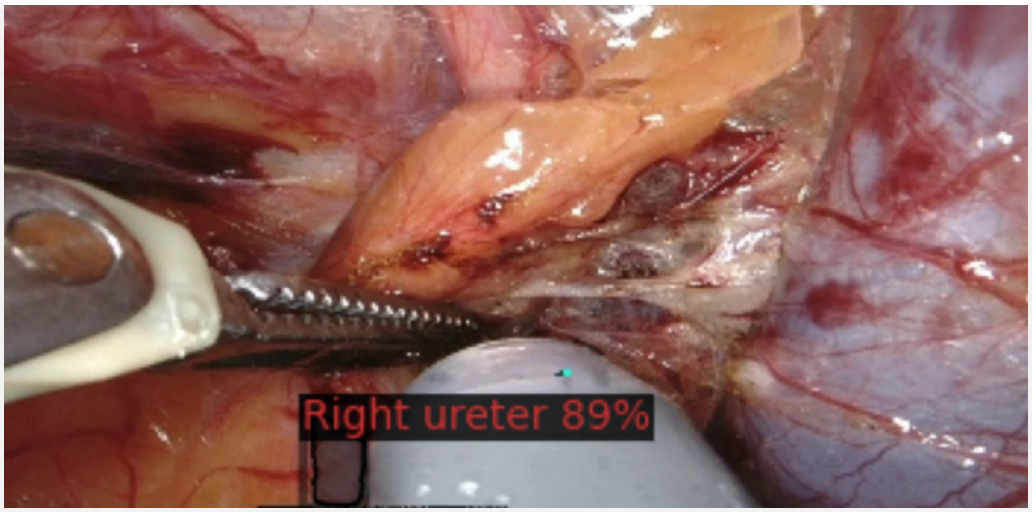
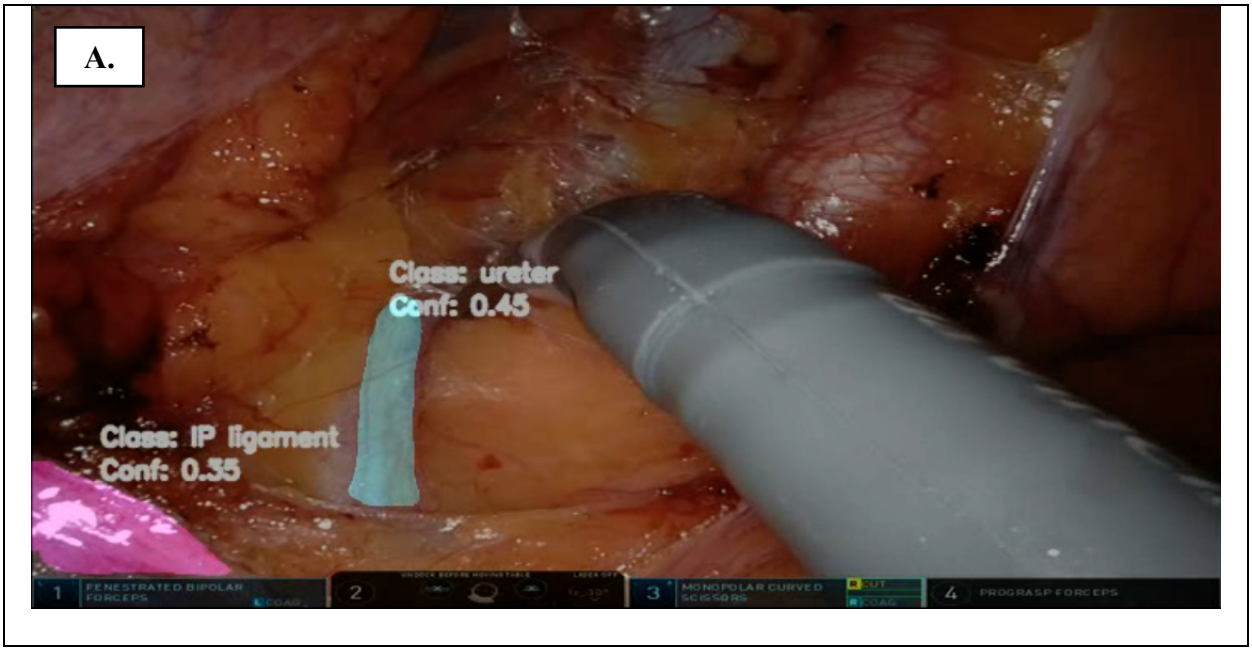
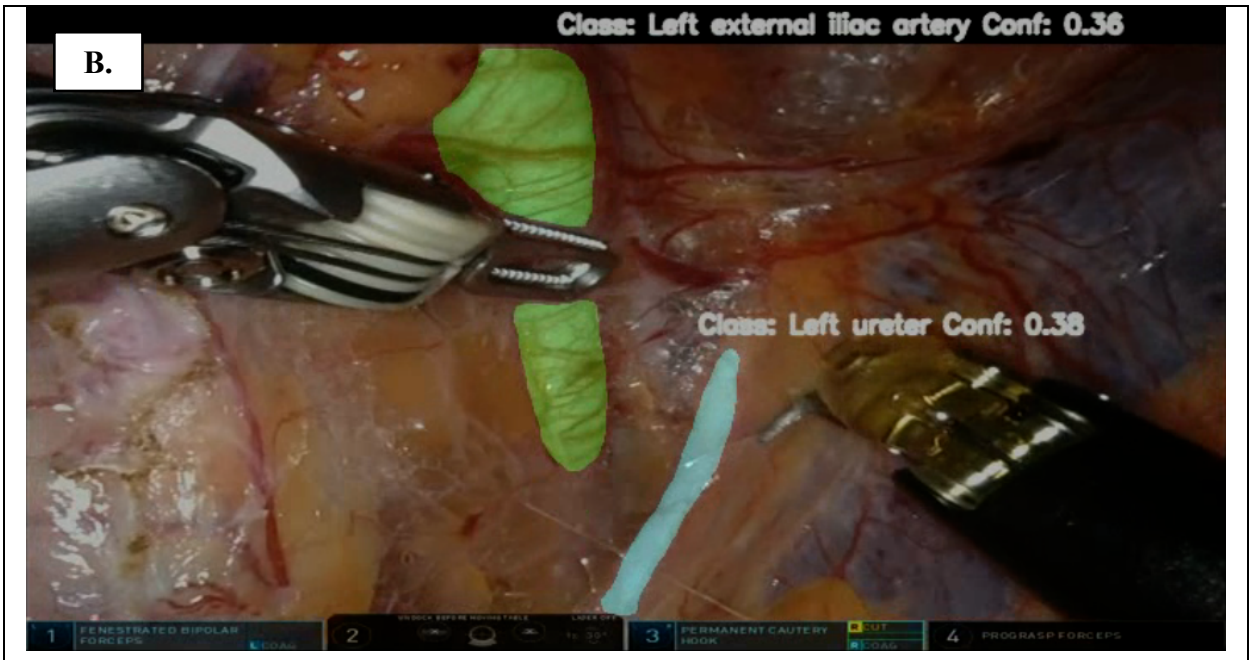
B.	*Frame #4639 (right ureter)*
Original frame	
Ground truth	
Model prediction	

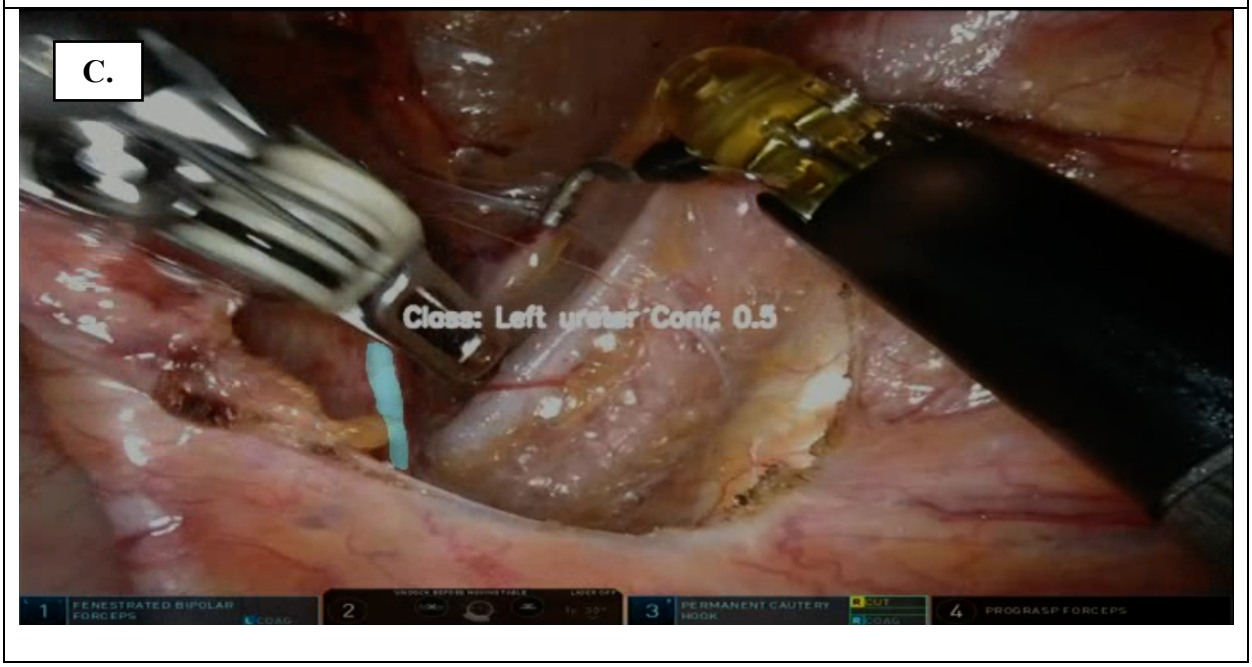
Figure 2.9 demonstrates two examples of model predictions. The original frame, extracted from a surgical clip can be seen, followed by the manual label, in this case, of the right ureter. The third image in Figures 2.9A and 2.9B shows the label created by the model. In both instances, the correct location of the right ureter was identified, but there was not a complete overlap between the ground truth and the predicted mask. Although the alignment is not perfect, the actual discrepancy is only a few millimeters, with questionable practical significance.

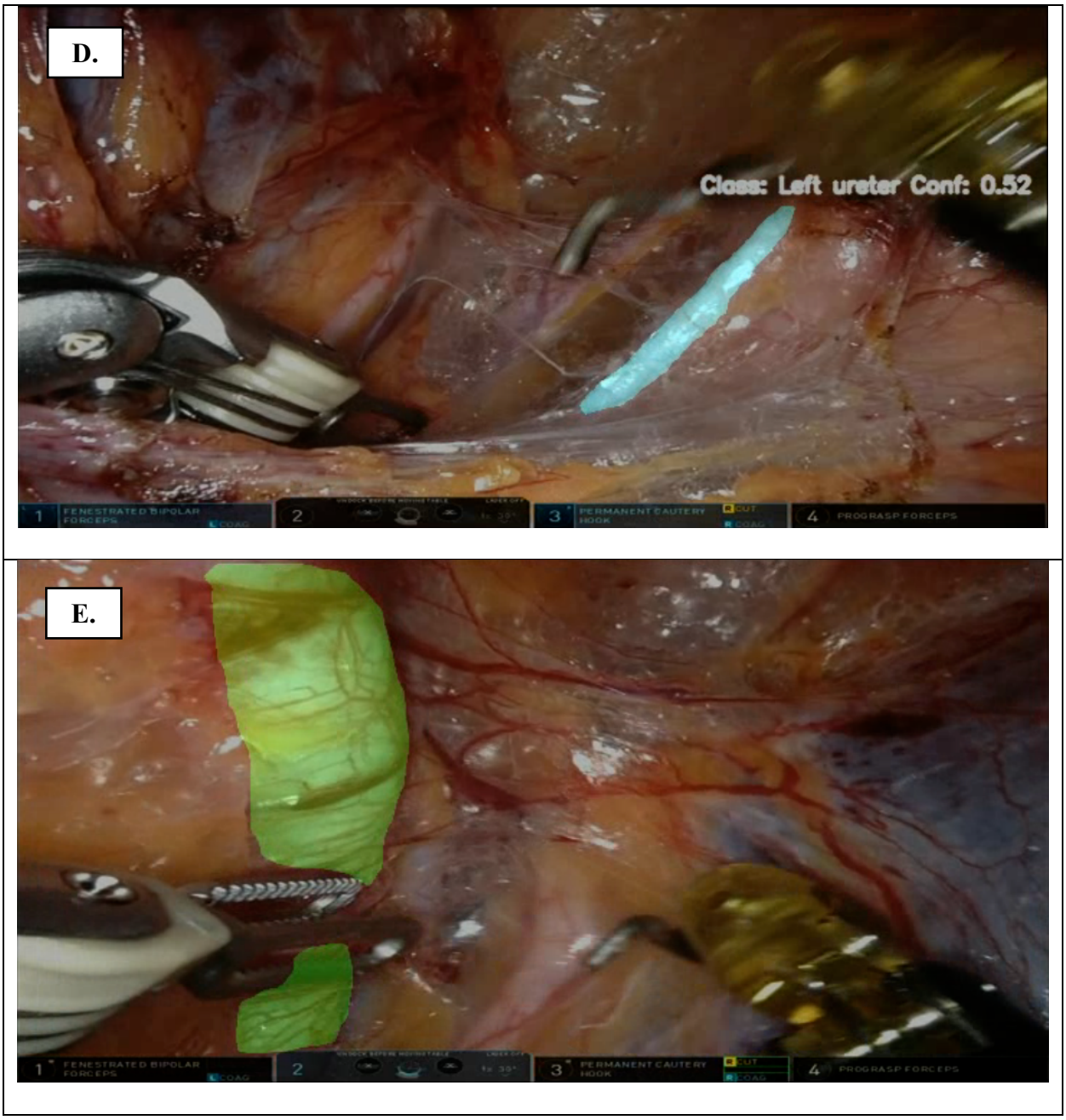
After testing the model on distinct frames, the model was tested on an unseen surgical video from a "new" patient. This simulates a more realistic and practical usage of the model. Figure 2.10 shows some snapshots taken during real-time model implementation. The ureter, the IP ligament and the external iliac artery are accurately delineated and classified correctly in Figures 2.10A-C. In Figures 2.10D and 2.10E the ureter is not identified. However, the external iliac artery was properly identified in the same image. Figure 2.10F demonstrates the opposite situation where the ureter is outlined perfectly but the external iliac artery is overlooked by the model. This is probably related to the surgical tool crossing it.

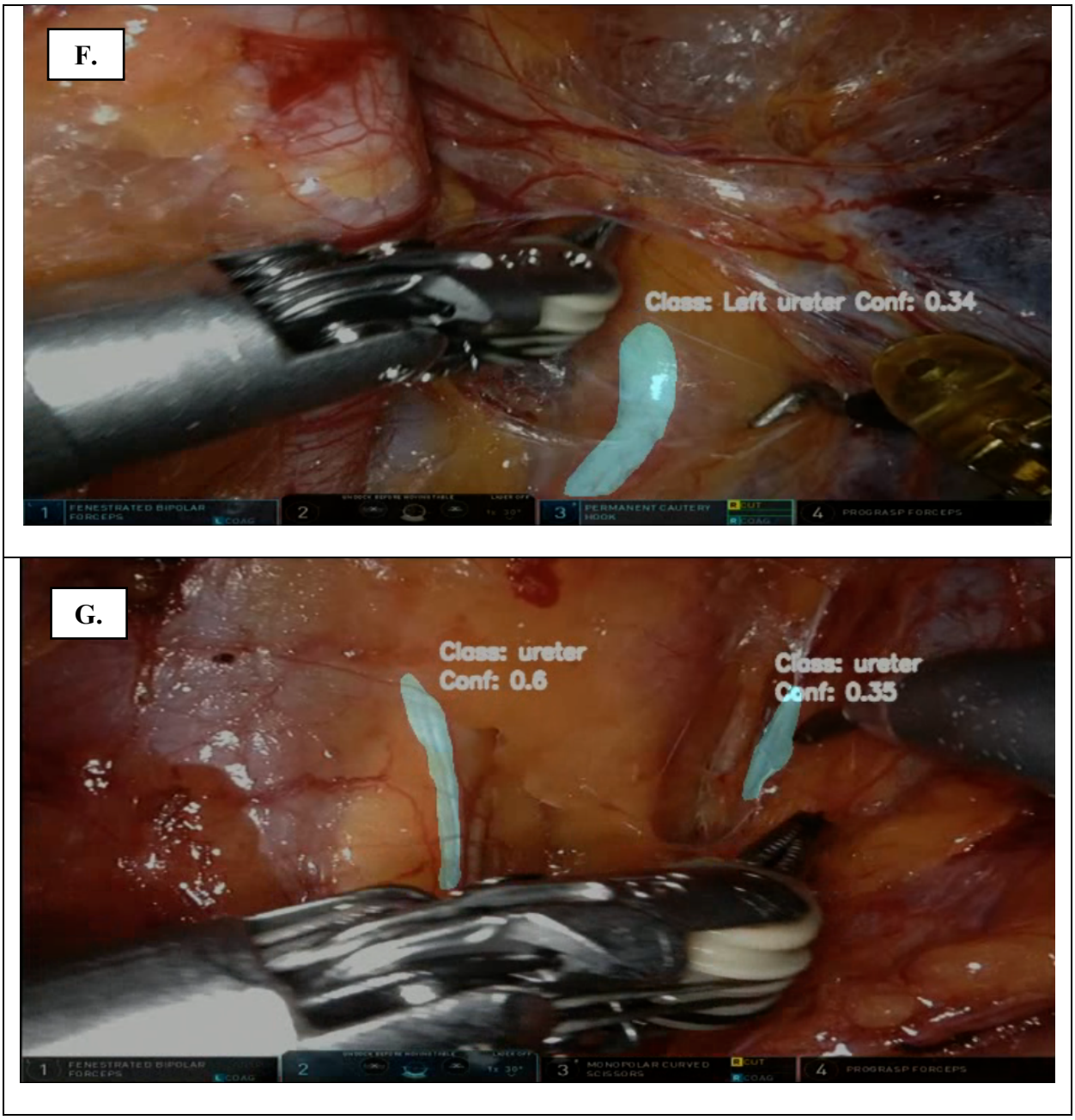
Figure 2.10: Real-time classification of an unseen surgical clip

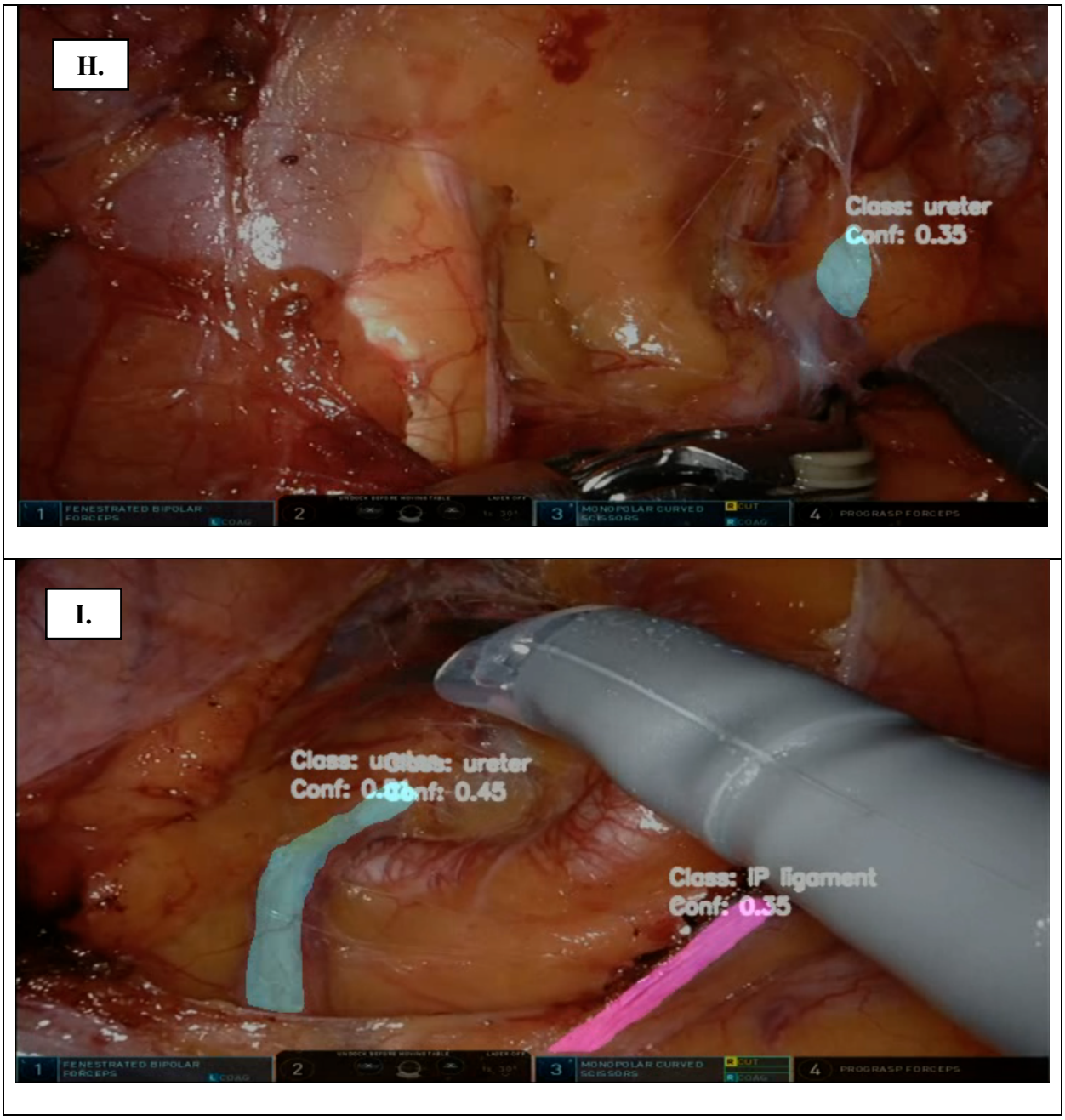












In Figure 2.10G the model mistakenly identified two ureters. One (on the left) is correct, while the second structure (on the right) is the genitofemoral nerve. This structure wasn't labelled and usually ignored by the model as can be seen in Figure 2.10H. The final snapshot, Figure 2.10I, shows two errors made by the model. While the ureter is outlined properly, the external iliac artery is missed and the IP ligament is incorrectly masked.

Discussion

This work is a prototype system for the automatic recognition of important anatomical structures encountered in gyne-oncologic surgery. Incorporating MIS in gynecologic oncology offers significant benefits to patients and surgeons and provides an opportunity to advance oncologic surgery to the next level. Incorporating CV in these surgeries will increase precision, reduce complications and potentially improve oncological outcomes.

In traditional abdominal surgeries (i.e. laparotomies), the surgeons rely on their eyes and experience to identify critical anatomical structures. To increase precision, some surgeons use magnifying lenses. Laparoscopy, using sophisticated optical systems, further improves visualization. Robotic surgery, by canceling tremor, combines advanced vision with precise motion. CV can delineate anatomical structures and tumors with great accuracy. Adding this to robotic systems will significantly increase the surgical precision. This will have a direct effect on patients' outcomes by reducing the rate of intra-operative complications and enabling the complete removal of tumor tissue. For instance, ureteral injuries can be significantly reduced by AI-assisted delineation of the ureter. This will aid surgeons to safely dissect tumor and other tissues from the ureter. Moreover, AI and CV can manipulate the surgical tools, in proximity to the ureter, to avoid injury. Finally, the complete removal of tumor will increase the rates of progression free and overall survival.

While accurate models for surgical tool detection exist (45), models for the identification of pelvic anatomy are scarce and focus on simple structures (e.g. uterus) rather than more challenging objects (e.g. ureter). During surgery, great effort is made to locate these structures and carefully dissect the tissue surrounding them to avoid complications. Thus, a model capable of precisely identifying them will have a significant impact on gyne-oncologic surgery.

In this project, a large video dataset of robotic surgeries for gynecologic cancer was created and organized. Videos were divided into surgical phases and the beginning, the end, and the anatomical structures of each phase were recorded. A surgeon labelled nineteen clips from this data. While labelling all the clips in the dataset is not feasible, it establishes a strong basis for the ongoing development of the model and new surgical CV systems.

Data annotation was done in three steps. First, watching the entire surgery and documenting timestamps of surgical phases and anatomical structures. Second, creating clips of specific phases with important anatomical structures. Third, extracting frames from the clips and labeling the anatomical structures in each frame. The main challenge was to accurately identify and consistently label the structure, as it was often only partially visible or obscured by surgical tools, blood, or smoke.

The polygonal mask method was used for labeling. In this method, a series of connected points is drawn around an object to form a polygon that outlines the object's shape. This polygon is then used to create a binary mask, with the interior of the polygon representing the object and the exterior representing the background. This method is particularly useful for objects with complex or irregular shapes such as tumors and some anatomical structures. Polygonal masks closely follow the contours of an object, making them ideal for tasks that require detailed segmentation, such as medical imaging. Other existing methods are bounding boxes and pixel-wise annotations. Polygonal masks are a good balance between annotation efficiency and accuracy, making them more practical for large datasets (46).

The labelled structures included the ureters, obturator nerves, iliac arteries, and infundibulopelvic (IP) ligament. These are the main structures in pelvic surgery. Successful and uncomplicated surgery relies on a clear understanding of their location and pathway. Other pelvic organs, such

as the uterus or ovaries were omitted since models for their identification exist (Chapter 1). Moreover, the results of this work show that large or superficial structures require considerably fewer labels to achieve high AP, compared to small or deep organs. For example, the AP of the external iliac artery is 63.94% with only 212 labels, while the AP of the obturator nerve is 45.78% with 1742 labels.

PointRend and DVIS++ were chosen for medical data labeling, specifically for segmenting anatomical structures, due to their advanced capabilities in handling complex image segmentation tasks. PointRend excels at refining the boundaries of objects, which is crucial for accurately delineating anatomical structures that often have intricate and irregular shapes. DVIS++, on the other hand, is tailored for dynamic video segmentation, allowing it to maintain consistent tracking of anatomical structures across surgical procedures where the field of view is constantly changing.

The selection of these models over alternatives like Mask R-CNN or U-Net is based on their superior performance in previous studies, particularly in tasks requiring precise boundary refinement and temporal consistency. These features are vital for medical applications, where the accurate identification of anatomical boundaries can directly impact clinical outcomes. While other models may be easier to implement or require less computational power, they often lack the precision necessary for medical environments.

The model was trained by the supervised learning (SL) method. Although this is the most resource-intensive technique, it yields the highest accuracy. In the next stage of model development, labels will be created automatically (by the model) and then evaluated and adjusted manually. This approach resembles semi-supervised learning (SSL). Works directly comparing supervised and semi-supervised learning for training medical models are inconsistent. Some

argue that SSL achieves comparable accuracy with considerably fewer labelled data (47), while others disagree and find no benefit in adding unlabeled data (48). Probably SL is the optimal method for medical and surgical models, as accuracy is crucial, and the cost of errors is high.

To optimize the model, it was pre-trained on general objects in images and videos from COCO and DVIS datasets. These two databases are common sources of labelled data for transferred learning (TL), however they contain no medical or surgical data. Since no other sources of relevant surgical data were available, this was the best option for improving the model. The use of TL is widespread in medical CV models (49), however, its ability to improve medical or surgical object segmentation is questioned and not without challenges (50,51). While all acknowledge the necessity of developing public archives of medical and surgical data, using TL from general objects is the next best option until then.

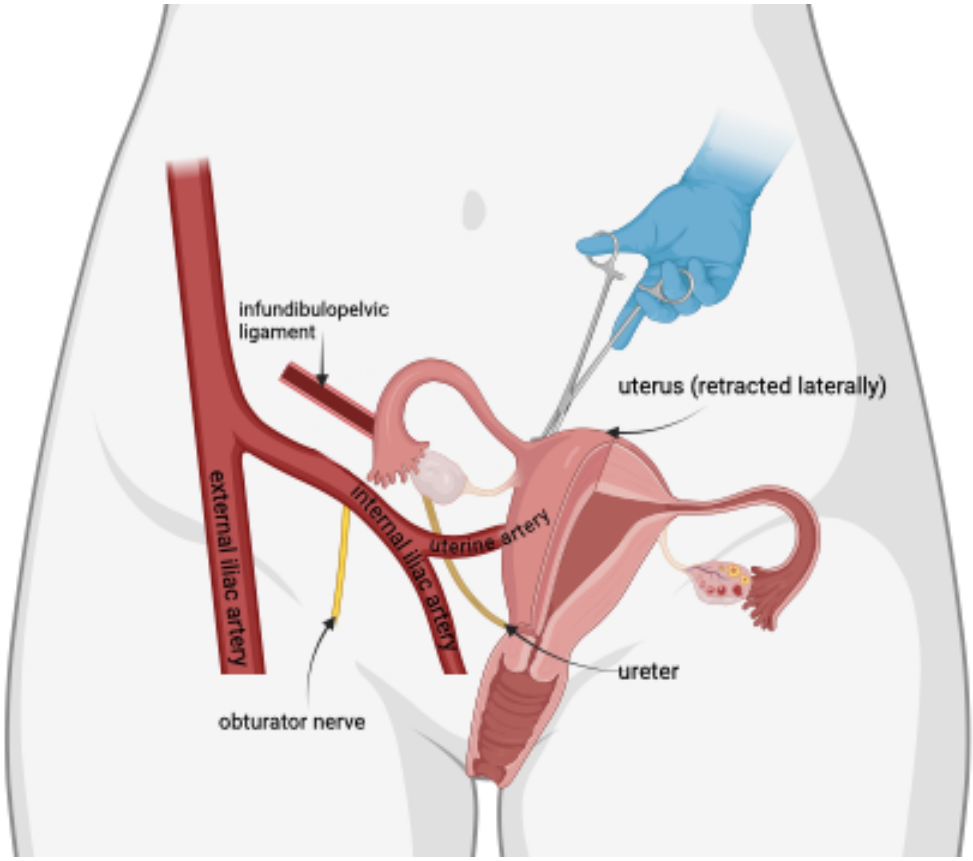
The hyperparameters are crucial for the performance and efficiency of CV models. In this work, the hyperparameters setting attempts to balance between computational effort and performance. Setting images per batch to 16 provides stable gradient estimation, fitting well within GPU memory limits. A learning rate of 0.0001 allows for cautious, incremental weight updates, improving convergence and model performance. Max iterations at 10,000, enable sufficient space for thorough learning, ensuring satisfactory performance without overloading resources. A weight decay of 0.05 helps prevent overfitting by penalizing large weights and aiding in model generalization. Together, these hyperparameters optimize training effectiveness and resource management, leading to a robust and efficient model.

Standard metrics of CV model performance were used in this study. Intersection over Union (IoU) and Average Precision (AP) were selected due to their popularity, reliability and precision in evaluating computer vision systems (52,53). IoU is particularly important for medical or

surgical objects where precise detection is crucial. AP is complementary to IoU and measures the area under the precision-recall curve (PR-AUC). Since the curve is plotted by calculating the precision (correct predictions out of all predictions) and recall (correct predictions out of all objects) across different IoU thresholds, it provides a balanced estimation of the model's performance.

The AP50 (AP at 50% IoU) of the current version of the model is 87.6%, however, the AP75 is significantly lower. As discussed previously, the AP of larger and more superficial structures detection is significantly higher compared to smaller and deeper organs. Furthermore, they are often partially covered by a surgical tool or not fully dissected from the surrounding tissue, making the segmentation even more challenging. Another important factor that significantly influences the model's metrics is the allocation of images into the training set. Typically, data for prediction models is randomly split after shuffling into training and testing sets at ratios of 70-30 or 80-20. However, for CV models, especially when extracting images from videos, stratified sampling is more appropriate. Since the frames can be highly correlated (repeating frames), random splitting may achieve misleadingly high AP. Yet, the model will perform poorly on unseen data (external validation), leading to a model that is not generalizable.

Figure 2.11: Relations of key structures in retroperitoneal dissection



To avoid this problem the images in the testing set were strictly from patients and surgeries not included in the training set. This may explain the relatively lower AP compared to other surgical models. However, there are no available models segmenting the same structures and accurate comparison may not be possible. Testing the model on unseen data is more practical since the metrics resemble real-life model performance therefore increasing the clinical relevance and generalizability of the system.

The most effective way to improve performance metrics is by adding more labelled data. The current version of the system was trained on 5,465 images. While it may be sufficient for some simpler structures, such as the external iliac artery, complex structures, such as the ureter require an extensive training set for accurate segmentation. Assessing the model's behavior on new data

gives us further insights for improvement. This model is capable of outlining anatomical structures with high precision but tends to misclassify them (Figure. 2.10D, 2.10G). The two main options to increase the accuracy of the classifications are temporal information integration and incorporating anatomical landmarks and relations.

Temporal information is derived from the change in video content over time. By "understanding" the dynamics between sequential frames, video CV models can use time-related context in their segmentation tasks. For instance, the ureter has a very typical motion pattern – vermiculation. This peristaltic movement resembling a slithering worm (hence the term) or snake, can only be identified by analyzing temporal information from successive frames. Since this motion is specific to the ureter, it can assist in its identification in complex situations. Furthermore, after detecting the ureter by vermiculation, its location can be tracked by temporal information analysis. This allows the system to identify the ureter in the next frames even without the vermiculation. Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) Networks and 3D Convolutional Neural Networks (3D-CNNs) are the most common neural network types used for temporal information integration.

Incorporating anatomical relations and landmarks is another powerful option for model improvement. Figure 2.11 depicts the anatomical relations of crucial pelvic organs. These structures are key components of pelvic retroperitoneal dissection in gyne-oncological surgeries. To achieve optimal oncological results and minimize complications, it is crucial to expose them and meticulously dissect the surrounding tissues. Feeding the relations in Figure 2.11 into the model can improve its performance significantly. We previously discussed that the model can accurately identify the external iliac artery, and the use of temporal information will facilitate the detection of the ureter. The internal iliac artery location can then be inferred since it branches

from the external iliac and the uterine artery, in turn, branches from the internal iliac artery and crosses above the ureter ("water under the bridge") towards the uterus. Locating the obturator nerve is also simplified because its pathway is deep in the pelvis between the external and the internal iliac arteries (Figure 2.11).

Study limitations

Despite the encouraging results demonstrated by our surgical CV model, it has some limitations that need to be discussed. The labelling was performed by a single person and wasn't reviewed by others. In controversial cases, it may lead to subjective labelling and impact generalization. Increasing the number of labellers could have resulted in a higher volume of labels and facilitated a more even distribution of errors.

The training data is from real surgeries performed by different surgeons. The level of exposure of the anatomical structures can vary significantly. This leads to extreme variability in the training data which influences the model's convergence. This phenomenon is further aggravated by the presence of artifacts, such as debris or blood.

This model was trained exclusively on robotic surgeries from one center. Most of the surgeries were for endometrial cancer, performed by one of the three attendings. This may limit the model generalizability. While the human anatomy is similar, its visualization varies between different types of surgeries, surgeons and subjects. This model may not perform well in laparoscopic or open surgeries because of different optics and different surgical approaches. Surgeons in other centers may use different surgical techniques with distinct exposure of key anatomical structures. Patient's previous surgeries or medical conditions can distort the anatomy and impact the model metrics. The model may not be usable by other disciplines since its focus is oncologic surgeries in the pelvis.

The ratio of the training to testing sets was 90 to 10 which is not standard. This was dictated by the requirement to test the model on an unseen surgery and patient. The limited size of the testing set may reduce the accuracy of the model's performance estimation.

The model's training possibilities are also limited by the available computational power.

Study strengths

While there are some limitations, the study has notable strengths. This study represents innovative work focusing on structures in which segmentation is very complex. The created database of surgical videos is extensive with a considerable volume of labeled data.

The labels were created by a surgeon using the polygonal masks method and the model was trained only on labeled data. This is the most accurate method for training CV algorithms.

The testing set is from unseen labelled data simulating model performance in realistic conditions.

The AI infrastructure underlying this CV model is powerful and reliable.

Conclusion and summary

This is the first model focusing on key pelvic structures with high segmentation complexity. This system can be combined with existing algorithms for surgical tool recognition. Future versions with improved performance could be implemented in robotic surgeries to assist in surgical navigation, increase surgical precision and reduce intra-operative complications. The development of such systems lays the foundation for the future of autonomous surgical machines.

References

1. Cancer Today [Internet]. [cited 2024 Mar 7]. Available from: https://gco.iarc.fr/today/
2. Schilling JM, Shaker N, Shaker N, Fadare O. The 2023 FIGO staging system for endometrial carcinoma: predicted impact on stage distribution based on a retrospective analysis of 1169 cases. Am J Surg Pathol. 2024 Jan 1;48(1):123–6.
3. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer. Lancet. 2022 Apr 9;399(10333):1412–28.
4. National Comprehensive Cancer Network. Uterine Neoplasms (Version 1.2020) [Internet]. NCCN Guidelines. 2020 [cited 2020 May 16]. Available from: https://www.nccn.org/professionals/physician_gls/pdf/uterine.pdf
5. Åkesson Å, Wolmesjö N, Adok C, Milsom I, Dahm-Kähler P. Lymphadenectomy, obesity and open surgery are associated with surgical complications in endometrial cancer. Eur J Surg Oncol. 2021 Nov;47(11):2907–14.
6. Eskander RN, Miller DS, Powell M, Creasman WT. Adenocarcinoma of the Uterine Corpus and Sarcomas of the Uterus . In: DiSaia PJ, Creasman WT, Mannell RS, Eskander RN, Herzog TJ, editors. Clinical Gynecologic Oncology. 10th ed. Philadelphia, PA: Elsevier; 2022. p. 93–120.
7. Lheureux S, Gourley C, Vergote I, Oza AM. Epithelial ovarian cancer. Lancet. 2019 Mar 23;393(10177):1240–53.
8. National Comprehensive Cancer Network, Inc. 2024. Ovarian Cancer Including Fallopian Tube Cancer and Primary Peritoneal Cancer [Internet]. nccn.org. 2024 [cited 2024 Feb 28]. Available from: https://www.nccn.org/professionals/physician_gls/pdf/ovarian.pdf

9. Vergote I, Tropé CG, Amant F, Kristensen GB, Ehlen T, Johnson N, et al. Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer. N Engl J Med. 2010 Sep 2;363(10):943–53.
10. Cohen PA, Jhingran A, Oaknin A, Denny L. Cervical cancer. Lancet. 2019 Jan 12;393(10167):169–82.
11. Yang FC, Huang W, Yang W, Liu J, Ai G, Luo N, et al. Cervical cancer surgery: current state of affairs. Gynecology and Minimally Invasive Therapy. 2021 Jun;10(2):75–83.
12. Ramirez PT, Frumovitz M, Pareja R, Lopez A, Vieira M, Ribeiro R, et al. Minimally Invasive versus Abdominal Radical Hysterectomy for Cervical Cancer. N Engl J Med. 2018 Nov 15;379(20):1895–904.
13. Chang J, Rattner DW. History of minimally invasive surgical oncology. Surg Oncol Clin N Am. 2019 Jan;28(1):1–9.
14. Patil M, Gharde P, Reddy K, Nayak K. Comparative analysis of laparoscopic versus open procedures in specific general surgical interventions. Cureus. 2024 Feb 19;16(2):e54433.
15. Wattiez A, Cohen SB, Selvaggi L. Laparoscopic hysterectomy. Curr Opin Obstet Gynecol. 2002 Aug;14(4):417–22.
16. Mori KM, Neubauer NL. Minimally invasive surgery in gynecologic oncology. ISRN Obstet Gynecol. 2013 Aug 12;2013:312982.
17. Jeyaprakash P, Moussad A, Pathan S, Sivapathan S, Ellenberger K, Madronio C, et al. A systematic review of scaling left atrial size: are alternative indexation methods required for an increasingly obese population? J Am Soc Echocardiogr. 2021 Oct;34(10):1067–1076.e3.

18. Davis EF, Crousillat DR, He W, Andrews CT. Indexing left atrial volumes: alternative indexing methods better predict outcomes in overweight and obese populations. Cardiovascular. 2022;
19. Falcone T, Goldberg J, Garcia-Ruiz A, Margossian H, Stevens L. Full robotic assistance for laparoscopic tubal anastomosis: a case report. J Laparoendosc Adv Surg Tech A. 1999 Feb;9(1):107–13.
20. Leal Ghezzi T, Campos Corleta O. 30 years of robotic surgery. World J Surg. 2016 Oct;40(10):2550–7.
21. Casarin J, Song C, Multinu F, Cappuccio S, Liu E, Butler KA, et al. Implementing robotic surgery for uterine cancer in the United States: Better outcomes without increased costs. Gynecol Oncol. 2020 Feb;156(2):451–8.
22. Abitbol J, Gotlieb W, Zeng Z, Ramanakumar A, Kessous R, Kogan L, et al. Incorporating robotic surgery into the management of ovarian cancer after neoadjuvant chemotherapy. Int J Gynecol Cancer. 2019 Nov;29(9):1341–7.
23. Jorgensen K, Melamed A, Wu C-F, Nitecki R, Pareja R, Fagotti A, et al. Minimally invasive interval debulking surgery for advanced ovarian cancer after neoadjuvant chemotherapy. Gynecol Oncol. 2023 May;172:130–7.
24. Zhang Y, Grant MS, Zhang X, Paraghamian SE, Tan X, Clark LH. Comparing Laparotomy versus Robotic-Assisted Interval Debulking Surgery for Patients with Advanced Epithelial Ovarian Cancer Receiving Neoadjuvant Chemotherapy. J Minim Invasive Gynecol. 2020 Nov 25;
25. Benabou K, Khadraoui W, Khader T, Hui P, Fernandez R, Azodi M, et al. Port-Site Metastasis in Gynecological Malignancies. JSLS. 2021;25(1).

26. Abu-Rustum NR, Rhee EH, Chi DS, Sonoda Y, Gemignani M, Barakat RR. Subcutaneous tumor implantation after laparoscopic procedures in women with malignant disease. Obstet Gynecol. 2004 Mar;103(3):480–7.

27. Nagarsheth NP, Rahaman J, Cohen CJ, Gretz H, Nezhat F. The incidence of port-site metastases in gynecologic cancers. JSLS. 2004;8(2):133–9.

28. Manakitsa N, Maraslidis GS, Moysis L, Fragulis GF. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision. Technologies. 2024 Jan 23;12(2):15.

29. Guo Y, Liu Y, Georgiou T, Lew MS. A review of semantic segmentation using deep neural networks. Int J Multimed Inf Retr. 2017 Nov 24;7(2):1–7.

30. Singh R, Rani R. Semantic Segmentation using Deep Convolutional Neural Network: A Review. SSRN Journal. 2020;

31. Jin Y, Dou Q, Chen H, Yu L, Qin J, Fu C-W, et al. SV-RCNet: Workflow Recognition From Surgical Videos Using Recurrent Convolutional Network. IEEE Trans Med Imaging. 2018 May;37(5):1114–26.

32. Chadebecq F, Vasconcelos F, Mazomenos E, Stoyanov D. Computer vision in the surgical operating room. Visc Med. 2020 Dec;36(6):456–62.

33. Nguyen XA, Ljuhar D, Pacilli M, Nataraja RM, Chauhan S. Surgical skill levels: Classification and analysis using deep neural network model and motion signals. Comput Methods Programs Biomed. 2019 Aug;177:1–8.

34. Ahmad OF, Soares AS, Mazomenos E, Brandao P, Vega R, Seward E, et al. Artificial intelligence and computer-aided diagnosis in colonoscopy: current evidence and future directions. Lancet Gastroenterol Hepatol. 2019 Jan;4(1):71–80.

35. Nwoye CI, Yu T, Sharma S, Murali A, Alapatt D, Vardazaryan A, et al. CholecTriplet2022: Show me a tool and tell me the triplet - An endoscopic vision challenge for surgical action triplet detection. Med Image Anal. 2023 Oct;89:102888.
36. Tseng JH, Cowan RA, Zhou Q, Iasonos A, Byrne M, Polcino T, et al. Continuous improvement in primary Debulking surgery for advanced ovarian cancer: Do increased complete gross resection rates independently lead to increased progression-free and overall survival? Gynecol Oncol. 2018 Oct;151(1):24–31.
37. Tanyi JL, Randall LM, Chambers SK, Butler KA, Winer IS, Langstraat CL, et al. A Phase III Study of Pafolacianine Injection (OTL38) for Intraoperative Imaging of Folate Receptor-Positive Ovarian Cancer (Study 006). J Clin Oncol. 2022 Sep 7;JCO2200291.
38. van Vliet-Pérez SM, van de Berg NJ, Manni F, Lai M, Rijstenberg L, Hendriks BHW, et al. Hyperspectral Imaging for Tissue Classification after Advanced Stage Ovarian Cancer Surgery-A Pilot Study. Cancers (Basel). 2022 Mar 10;14(6).
39. Tu D-Y, Lin P-C, Chou H-H, Shen M-R, Hsieh S-Y. Slice-Fusion: Reducing False Positives in Liver Tumor Detection for Mask R-CNN. IEEE/ACM Trans Comput Biol Bioinform. 2023 Apr 7;PP.
40. Kirillov A, Wu Y, He K, Girshick R. Pointrend: image segmentation as rendering. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2020. p. 9796–805.
41. Zhang T, Tian X, Zhou Y, Ji S, Wang X, Tao X, et al. DVIS++: Improved Decoupled Framework for Universal Video Segmentation. 2023;

42. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. European Conference on Computer Vision (ECCV). Cham: Springer International Publishing; 2014. p. 740–55.
43. Yang L, Fan Y, Xu N. Video Instance Segmentation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE; 2019. p. 5187–96.
44. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010 Oct 1;22(10):1345–59.
45. Nema S, Vachhani L. Surgical instrument detection and tracking technologies: Automating dataset labeling for surgical skill assessment. Front Robot AI. 2022 Nov 4;9:1030846.
46. Mullen JF, Tanner FR, Sallee PA. Comparing the effects of annotation type on machine learning detection performance. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE; 2019. p. 855–61.
47. Al-Azzam N, Shatnawi I. Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer. Ann Med Surg (Lond). 2021 Feb;62:53–64.
48. Spiero I, Schuit E, Wijers OB, Hoebers FJP, Langendijk JA, Leeuwenberg AM. Comparing supervised and semi-supervised machine learning approaches in NTCP modeling to predict complications in head and neck cancer patients. Clinical and Translational Radiation Oncology. 2023 Nov;43:100677.
49. Kora P, Ooi CP, Faust O, Raghavendra U, Gudigar A, Chan WY, et al. Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering. 2022 Jan;42(1):79–107.

50. Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: Understanding Transfer Learning for Medical Imaging. Advances in Neural Information Processing Systems. 2019;
51. Salehi AW, Khan S, Gupta G, Alabduallah BI, Almjally A, Alsolai H, et al. A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope. Sustainability. 2023 Mar 29;15(7):5930.
52. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis. 2010 Jun;88(2):303–38.
53. Schlosser T, Friedrich M, Meyer T, Kowerko D. A Consolidated Overview of Evaluation and Performance Metrics for Machine Learning and Computer Vision. 2023 Oct 1;