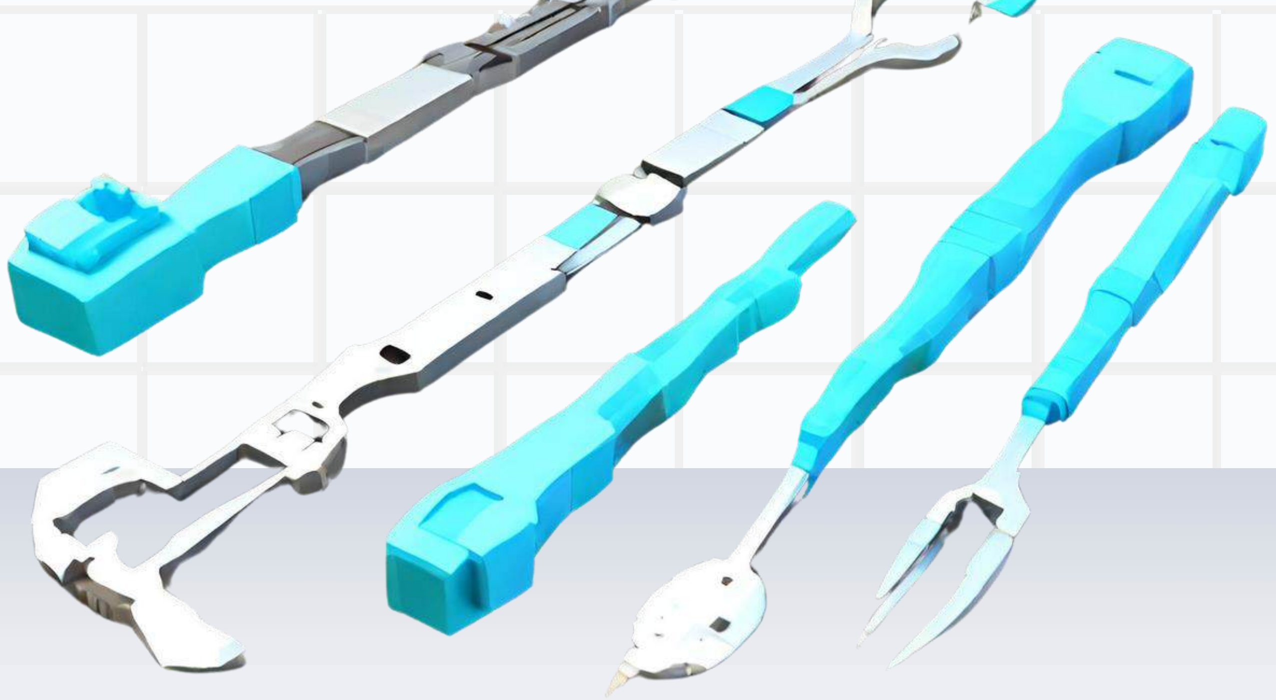


Surgeons's Eye

UIP Y15 PRESENTATION

TEAM

Surgeon's Eye

DHANASHEKAR K
MSRIT, CSE, SEM 8

SAHIL KULKARNI
MSRIT, CSE, SEM 6

JAMUNA S MURTHY
Assistant Professor (PhD)
MSRIT , CSE

PRATEEK M
MSRIT, CSE, SEM 6

UDAY CHOWDARY
MSRIT, CSE, SEM 6

OVERVIEW

Computer-assisted minimally invasive surgery has great potential in benefiting modern operating theatres. The video data streamed from the endoscope provides rich information to support context-awareness for next-generation intelligent surgical systems.

To achieve accurate perception and automatic manipulation during the procedure, learning based technique is a promising way, which enables advanced image analysis and scene understanding in recent years. However, learning such models highly relies on large-scale, high-quality, and multi-task labelled data. This is currently a bottleneck for the topic, as available public dataset is still extremely limited in the field of CAI.

MOTIVATION

Limited Understanding of Visual Content: ChatGPT primarily processes text data and lacks direct capabilities to understand or interpret visual content, such as videos. This limitation makes it less effective in analyzing and responding to information presented in video format, including medical procedures or imaging data.

Difficulty in Processing Complex Medical Terminology: Medical questions often involve specialized terminology and concepts that may be challenging for ChatGPT to understand accurately. Without specific training or knowledge in medical terminology, ChatGPT may struggle to provide relevant or accurate responses to medical queries.

CONTRIBUTION

In this research, we present A Surgical Assistance and Guidance based system for Designing Surgical-Visual Question Answering with the release of the first integrated dataset (named Surgeons's Eye) with multiple image-based perception tasks to facilitate learning-based automation in Ophthalmology(Eye), Cholecystectomy(Gall Bladder), Hysterectomy(Uterus) surgery.

Our Surgeons's Eye dataset is developed based on full-length videos of entire Ophthalmology, Cholecystectomy and Hysterectomy procedures. Specifically, two different yet highly correlated tasks are formulated in the dataset, including surgical workflow recognition, and instrument and key anatomy segmentation.

In addition, we provide experimental results with state-of-the-art models as reference benchmarks for further model developments and evaluations on this dataset.

Background

New technologies in robotics and computer-assisted intervention (CAI) are widely developed for surgeries, to release the burden of surgeons on manipulating instruments, identifying anatomical structures, and operating in confined spaces.

The advancements of these technologies will thereby facilitate the development of semi and fully-automatic robotic systems which can understand surgical situations and even make proper decisions in certain tasks. Nowadays, surgeries conducted in a minimally invasive way, e.g. laparoscopy, have become popular, and videos recorded through the laparoscope are valuable for image- based studies.

To enhance surgical scene understanding for image-guided automation, one promising solution is to rely on learning-based methods.

Background

As deep learning methods are data-driven, a large amount of surgical data is required to train and obtain reliable models for task executions. Especially in surgical field, due to various types of surgical procedures, corresponding image and video data vary in surgical scenes, surgical workflows, and the instruments used .

For laparoscopic minimally invasive surgery (MIS), some datasets have been established and released to improve the learning-based algorithm for different surgical tasks, such as action recognition, workflow recognition, instrument detection and segmentation. However, these datasets are not collected from real-surgery nor large enough to develop applicable and robust models in practice.

Recently, several research teams have worked on developing dataset at large scales, but most are only designed and annotated for one certain task. In terms of clinical applicability, data from different modalities are needed to better understand the whole scenario, make proper decisions, as well as enrich perception with multi-task learning strategy.

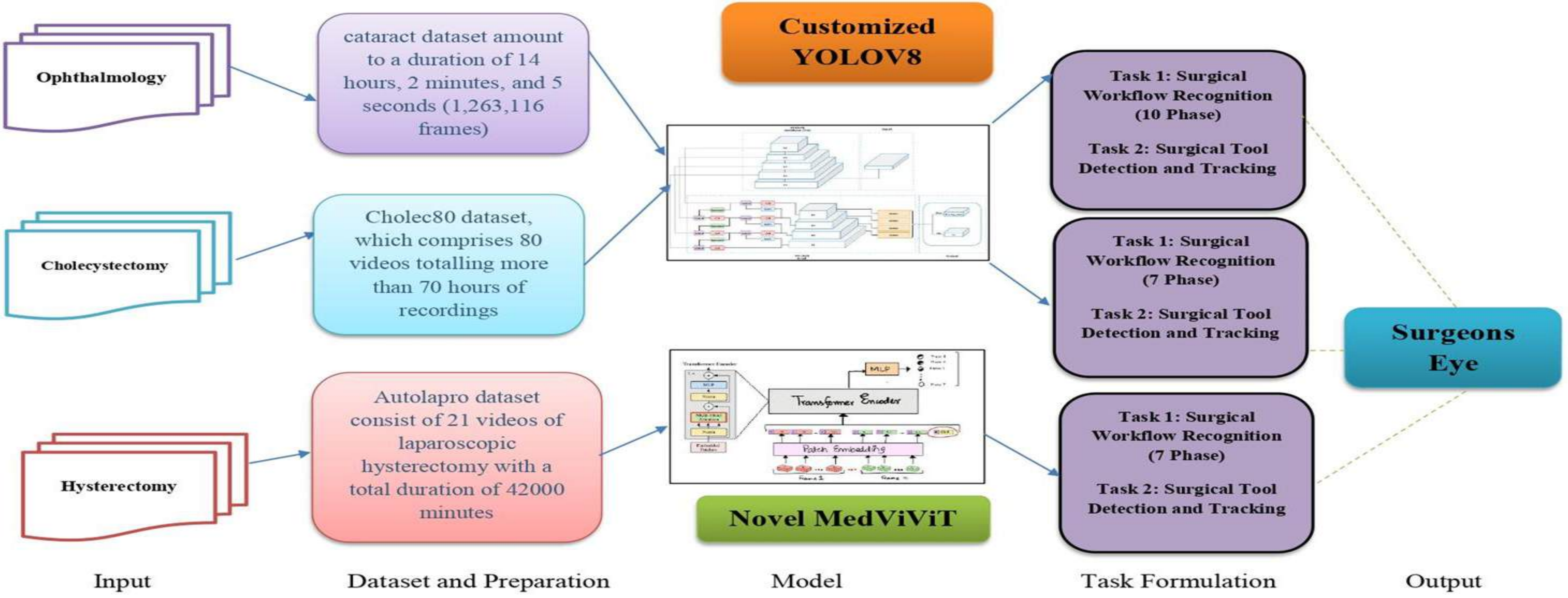
Background

Besides, there are few datasets designed for automation tasks in surgical application, among which the automatic laparoscopic field-of-view (FoV) control is a popular topic as it can liberate the assistant from such tedious manipulations with the help from surgical robots.

Therefore, surgical tasks with multiple modalities of data should be formulated and proposed for image-guided surgical automation.

For proposed application, we target :

- Ophthalmology
- Laparoscopic Cholecystectomy
- Laparoscopic Hysterectomy


Ophthalmology
cataract dataset amount to a duration of 14 hours, 2 minutes, and 5 seconds (1,263,116 frames)
Cholecystectomy
Cholec80 dataset, which comprises 80 videos totalling more than 70 hours of recordings
Hysterectomy
Autolapro dataset consist of 21 videos of laparoscopic hysterectomy with a total duration of 42000 minutes
Customized YOLOV8
Novel MedViViT
Transformer Encoder
MLP
Patch Embedding
Task 1: Surgical Workflow Recognition (10 Phase)
Task 2: Surgical Tool Detection and Tracking
Task 1: Surgical Workflow Recognition (7 Phase)
Task 2: Surgical Tool Detection and Tracking
Task 1: Surgical Workflow Recognition (7 Phase)
Task 2: Surgical Tool Detection and Tracking
Surgeons Eye
Input
Dataset and Preparation
Model
Task Formulation
Output

Ophthalmology

- Cataract surgery is one of the most frequently performed microscopic surgeries in the field of ophthalmology. The goal behind this kind of surgery is to replace the human eye lense with an artificial one, an intervention that is often required due to aging.
- The entire surgery is performed under microscopy, but co-mounted cameras allow to record and archive the procedure.
- The videos collected in this dataset were recorded at Klinikum Klagenfurt (Austria) and contain annotations of the 10 different phases of Cataract surgery:

Base Paper :

- Schoeffmann, K., Taschwer, M., Sarny, S., Münzer, B., Primus, M. J., & Putzgruber, D. (2018, June). Cataract-101: video dataset of 101 cataract surgeries. In Proceedings of the 9th ACM multimedia systems conference (pp. 421-425).

Dataset Design and Task Formulation

Task 1:

In total, video recording of the cataract dataset amount to a duration of 14 hours, 2 minutes, and 5 seconds (1,263,116 frames). All videos have PAL resolution (720x540 pixels) and are encoded as MP4 files, using H.264/AVC with profile High as video codec (25 fps, about 1.25 MBit/s birate), resulting in a total storage requirement of 8.4 GiB.

Task 2:

After extracting the frames, we had to curate the obtained frames. Following that, since there were 10 phases, we annotated the images to create the dataset. We annotated over 3000 images so that we could train it for instance segmentation.

Task 1,2: Customized YOLOv8 for Classification and Segmentation

- In YOLOv8, the segmentation heads consist of convolutional layers followed by upsampling layers (such as transposed convolution or nearest-neighbor upsampling). The output of each segmentation head is a set of feature maps with the same spatial resolution as the input image, where each pixel represents the probability or confidence of belonging to a particular class or object category.

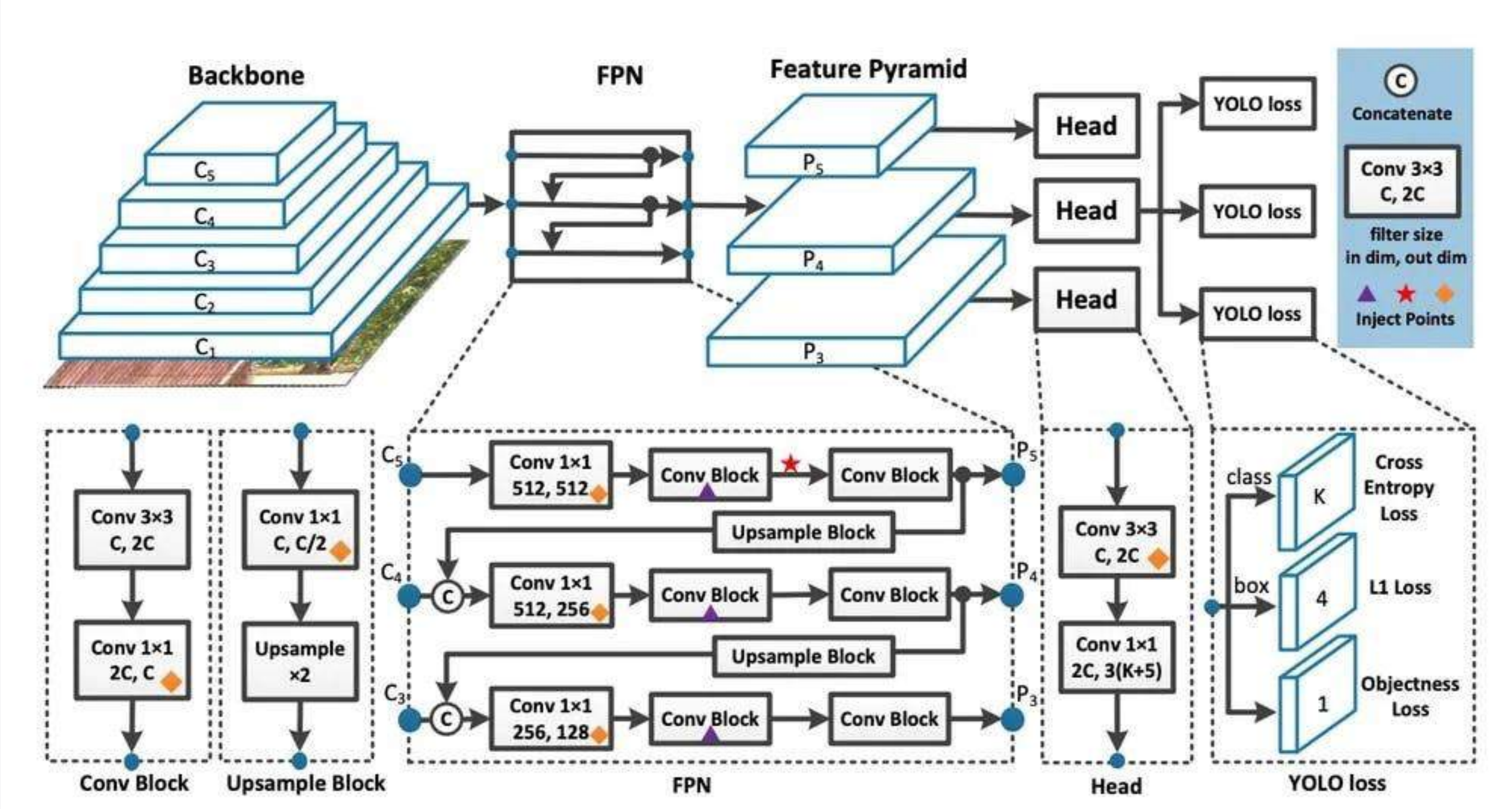
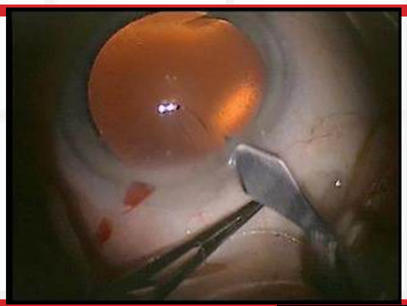


Figure : Architecture diagram of YOLOv8 Model

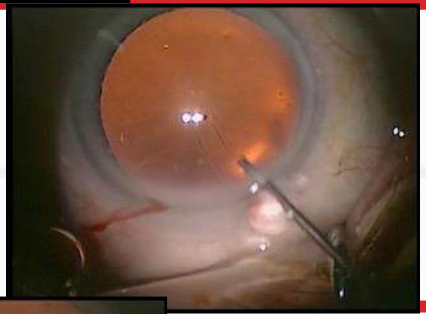
PHASE 01

Incision



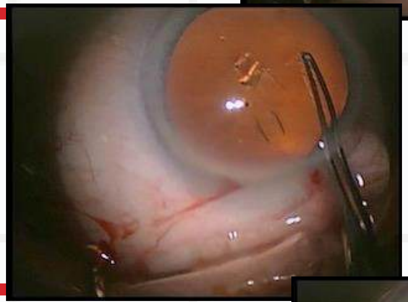
PHASE 02

Viscous Agent Injection



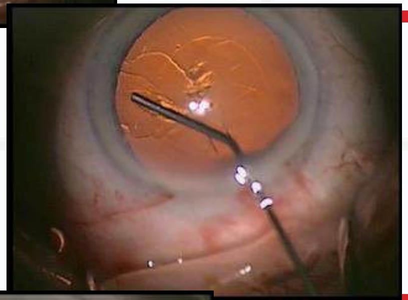
PHASE 03

Rhexis



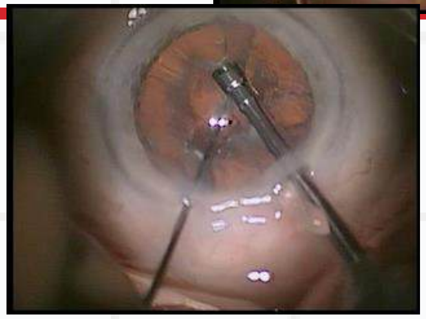
PHASE 04

Hydrodissection



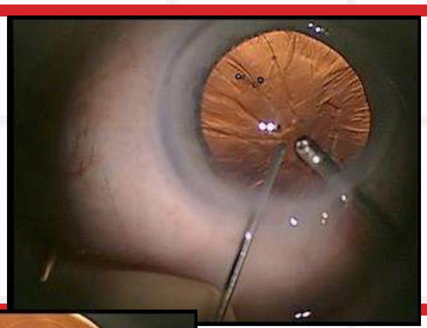
PHASE 05

Phacoemulsification



PHASE 06

Irrigation/Aspiration



PHASE 07

Capsule Polishing

PHASE 08

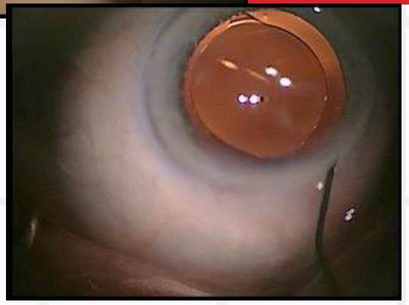
Lens Implant

PHASE 09

Viscous Agent Removal

PHASE 10

Tonifying and Antibiotics



CATARACT SURGERY TOOLS (11 tools)

i1: primary knife
i2: secondary knife
i3: bonn forceps
i7:rycroft cannula
i8: capsulorhexis forceps
i9: capsulorhexis cystotome
i10: phaco handpiece
i11: AI handpiece
i12: I/A handpiece
i16: Hydro cannula
i20: eye retractors

- i21: Lens injector.

Visual Q&A

Question 01

What is the tool moving from bottom-right to bottom - center in Phase 4 ?

.

Answer 01

The too is moving from bottom-center in Phase 4 is the *Hydro Cannula*.

The Hydro Cannula is a specialized cannula that is used to inject a stream of balanced salt solution (BSS) into the anterior chamber of the eye.

Question 02

What is happening in 22nd second ?

.

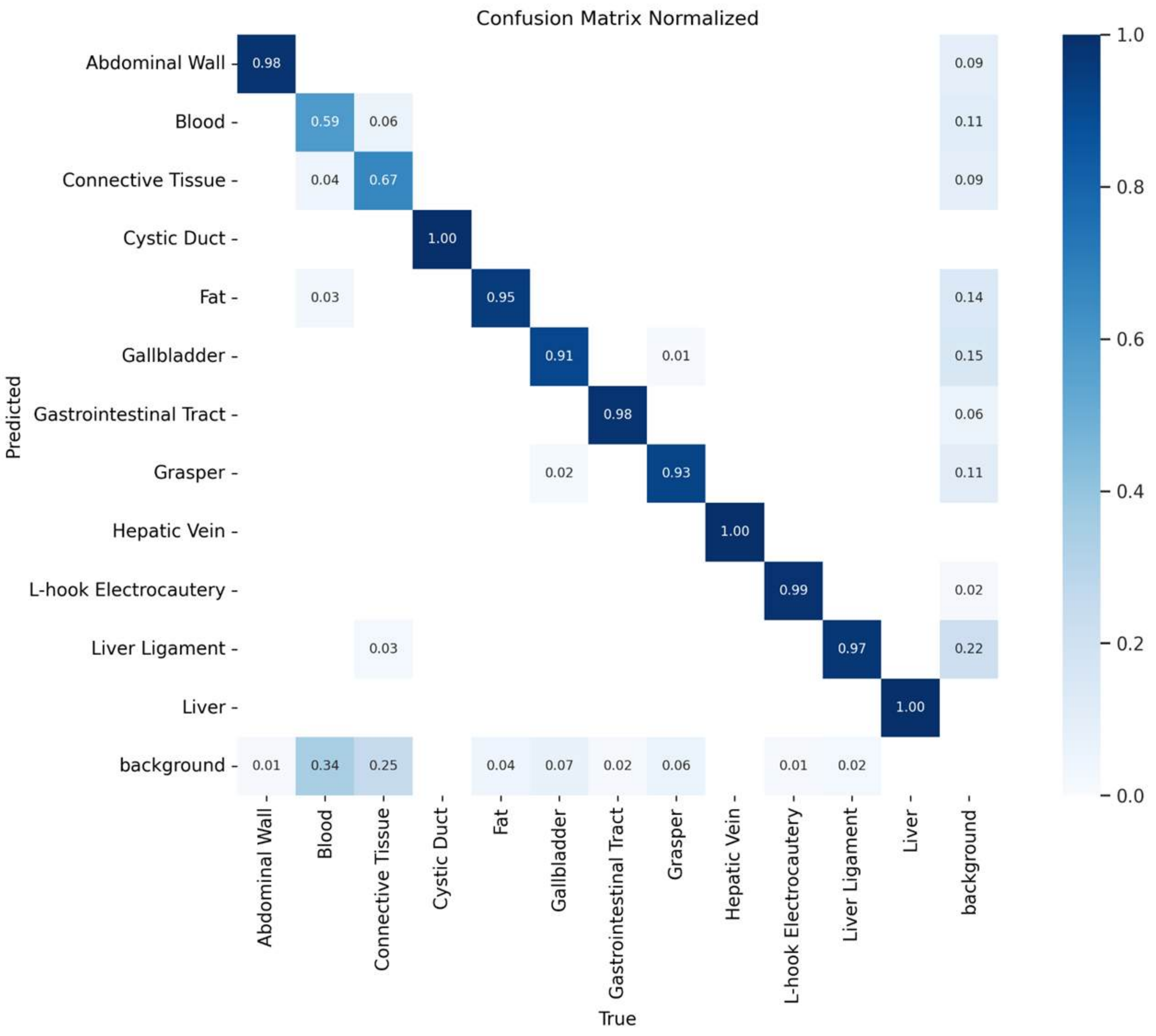
Answer 02

In the 22nd second of the surgical video, the surgeon is performing the *Incision Phase*. This phase involves the creation of corneal incisions to access the anterior chamber of the eye.

.

Sample questions

EXPERIMENTS



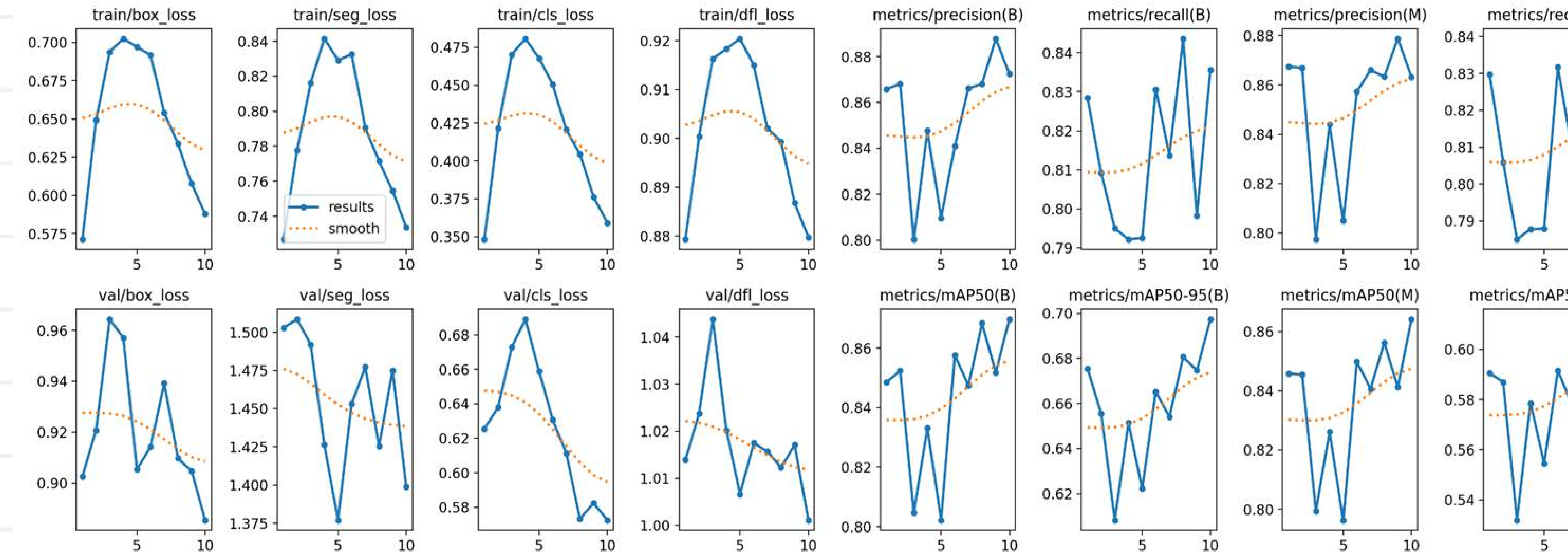
Phase Detection model was trained for image classification using yolov8 with dataset of more than 10000 images .

Phase Detection Accuracy: 95 %

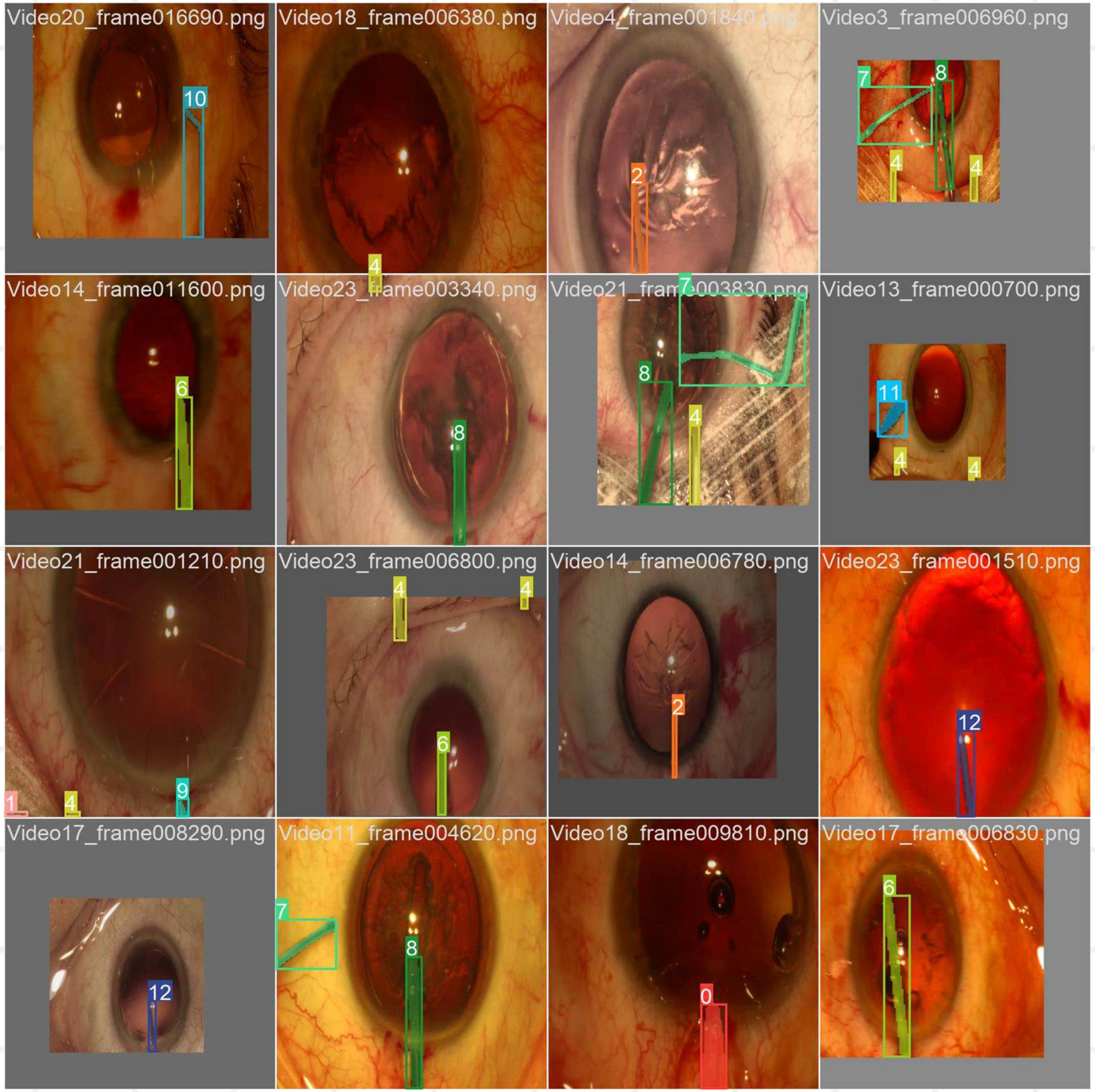
Tool detection Accuracy: 80%

Tool detection model was trained for instance segmentation using yolov8 with dataset more than 3000 images.

Yolov8 Performance Analysis



Inference


Video20_frame016690.png
Video18_frame006380.png
Video4_frame001840.png
Video3_frame006960.png
Video14_frame011600.png
Video23_frame003340.png
Video21_frame003830.png
Video13_frame000700.png
Video21_frame001210.png
Video23_frame006800.png
Video14_frame006780.png
Video23_frame001510.png
Video17_frame008290.png
Video11_frame004620.png
Video18_frame009810.png
Video17_frame006830.png

Laparoscopic Cholecystectomy

- Laparoscopic cholecystectomy is a surgical procedure used to remove the gallbladder. The gallbladder is a small organ located beneath the liver, and it stores bile produced by the liver. Bile is essential for the digestion of fats in the small intestine.
- The Cholec80 dataset contains 80 videos of cholecystectomy surgeries performed by 13 surgeons. The videos are captured at 25 fps. The dataset is labeled with the phase (at 25 fps) and tool presence and contain annotations of the 7 different phases of Gallbladder surgery

Base Paper :

- Seenivasan, L., Islam, M., Krishna, A. K., & Ren, H. (2022, September). Surgical-vqa: Visual question answering in surgical scenes using transformer. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 33-43). Cham: Springer Nature Switzerland.

Dataset Design and Task Formulation

Task 1:

In cholec, we utilized the Cholec80 dataset, which comprises 80 videos totalling more than 70 hours of recordings. Frame extraction was conducted every second, and all videos were stored in MP4 format, utilizing the High profile as the video codec (25 fps), with a total storage requirement of 70GB.

Task 2:

The next task was to select the appropriate frames/images for training, ensuring a diverse range of images to achieve the highest accuracy. We needed to curate images from various sources to create the proper dataset.

Task 1,2: Customized YOLOv8 for Classification and Segmentation

- In YOLOv8, the segmentation heads consist of convolutional layers followed by upsampling layers (such as transposed convolution or nearest-neighbor upsampling). The output of each segmentation head is a set of feature maps with the same spatial resolution as the input image, where each pixel represents the probability or confidence of belonging to a particular class or object category.

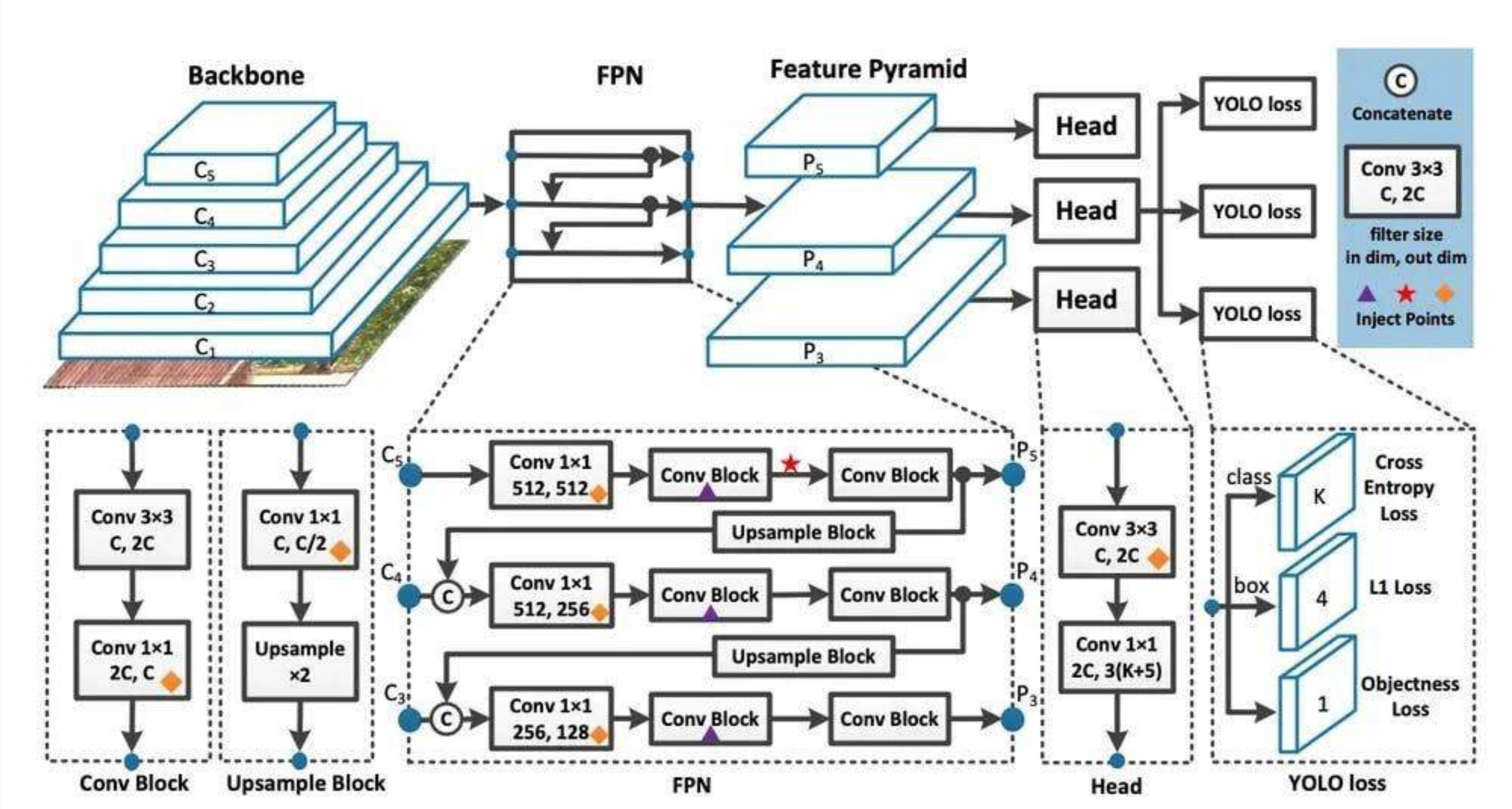


Figure : Architecture diagram of YOLOv8 Model

PHASE 01

Preparation

PHASE 02

Calot Triangle Dissection

PHASE 03

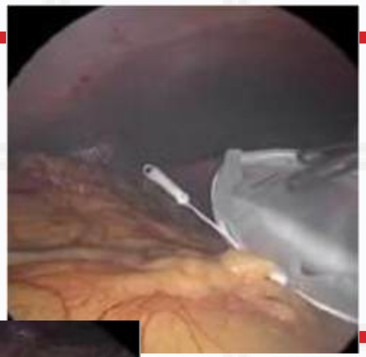
Clipping Cutting

PHASE 04

Gallbladder Dissection

PHASE 05

Gallbladder Retraction



PHASE 06

Cleaning Coagulation

PHASE 07

Gallbladder Packaging

GALL BLADDER SURGERY WORKFLOW

GALLBLADDER SURGERY TOOLS (2 tools)

Grasper, L-Hook
Electrocautry

Visual Q&A

Question 01
What is happening at 58th second ?

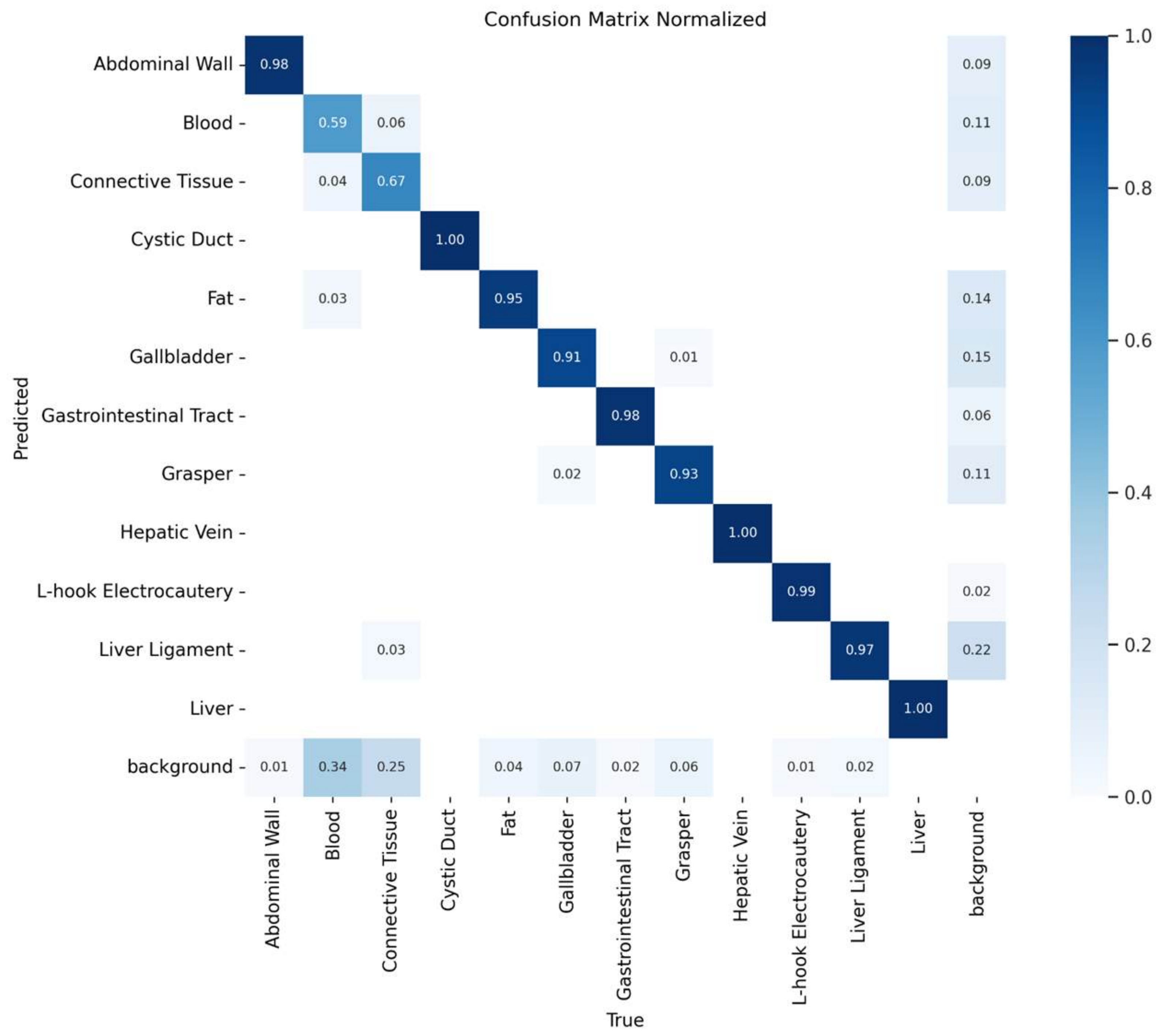
Answer 01
At the 58th second, the dissection phase of the gallbladder is taking place, and further, the dissection of the gallbladder is being performed using a hook and grasper.

Question 02
What is hook?

Answer 02
The hook is a tool used in surgery to dissect the gallbladder, primarily employed to tear tissue and remove unwanted elements.

Sample questions

EXPERIMENTS



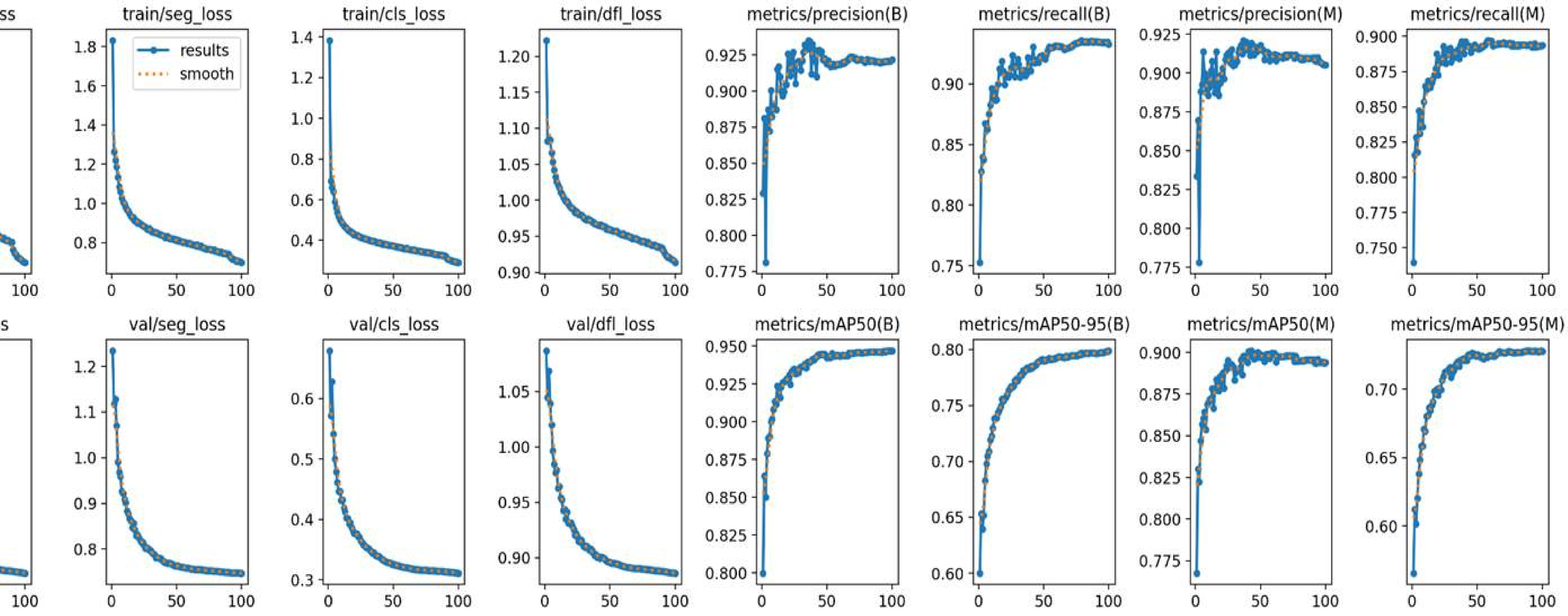
Trained the yolov8 model for instance segementation with dataset of more than 8000 images.

Accuracy for Phase Detection : 92%

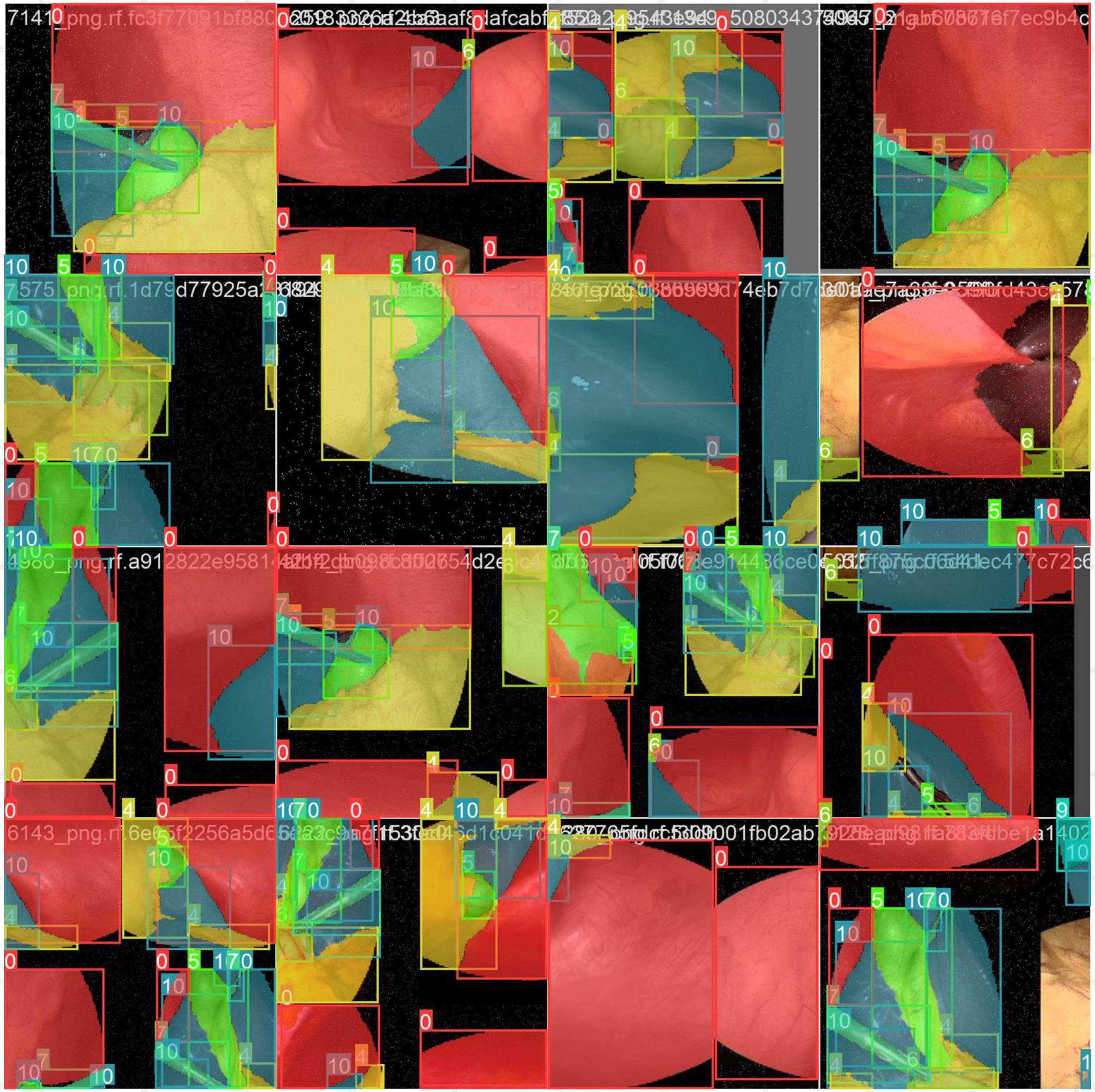
Accuracy for Tool Detection : 91.8%

Trained the yolov8 model for image classification with dataset if 7 phases with each phase containing 1000+ images

Yolov8 Performance Analysis



Inference



Laparoscopic Hysterectomy

- Laparoscopic hysterectomy is a minimally invasive surgical procedure used to remove the uterus. During a laparoscopic hysterectomy, the surgeon makes small incisions in the abdomen and inserts specialized surgical instruments, along with a laparoscope, to perform the procedure.
- The dataset contains 80 videos of laparoscopic hysterectomy collected from Prince of Wales Hospital, Hong Kong, during October to December, 2018.

Base Paper :

- Wang, Z., Lu, B., Long, Y., Zhong, F., Cheung, T. H., Dou, Q., & Liu, Y. (2022, September). Autolaparo: A new dataset of integrated multi-tasks for image-guided surgical automation in laparoscopic hysterectomy. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 486-496). Cham: Springer Nature Switzerland.

Dataset Design and Task Formulation

Task 1 :

Surgical workflow recognition, where Autolapro dataset consist of 21 videos of complete procedures of laparoscopic hysterectomy with a total duration of 42000 minutes are collected and annotated into 7 phases, proposed as the first dataset dedicating workflow analysis in this surgery

Task 2:

Instrument and key anatomy segmentation, where 1800 frames are sampled from the above clips and annotated with 5936 pixel-wise masks of 9 types, providing rich information for scene understanding. It is worth noting that these tasks, data and annotations are highly correlated to support multi-task and multi-modality learning for advanced surgical perception towards vision-based automation.

Task 1: Proposed MedViViT Model Design

- Our model extracts spatio-temporal tokens from the input video, which are then encoded by a series of transformer layers for Surgical workflow recognition.

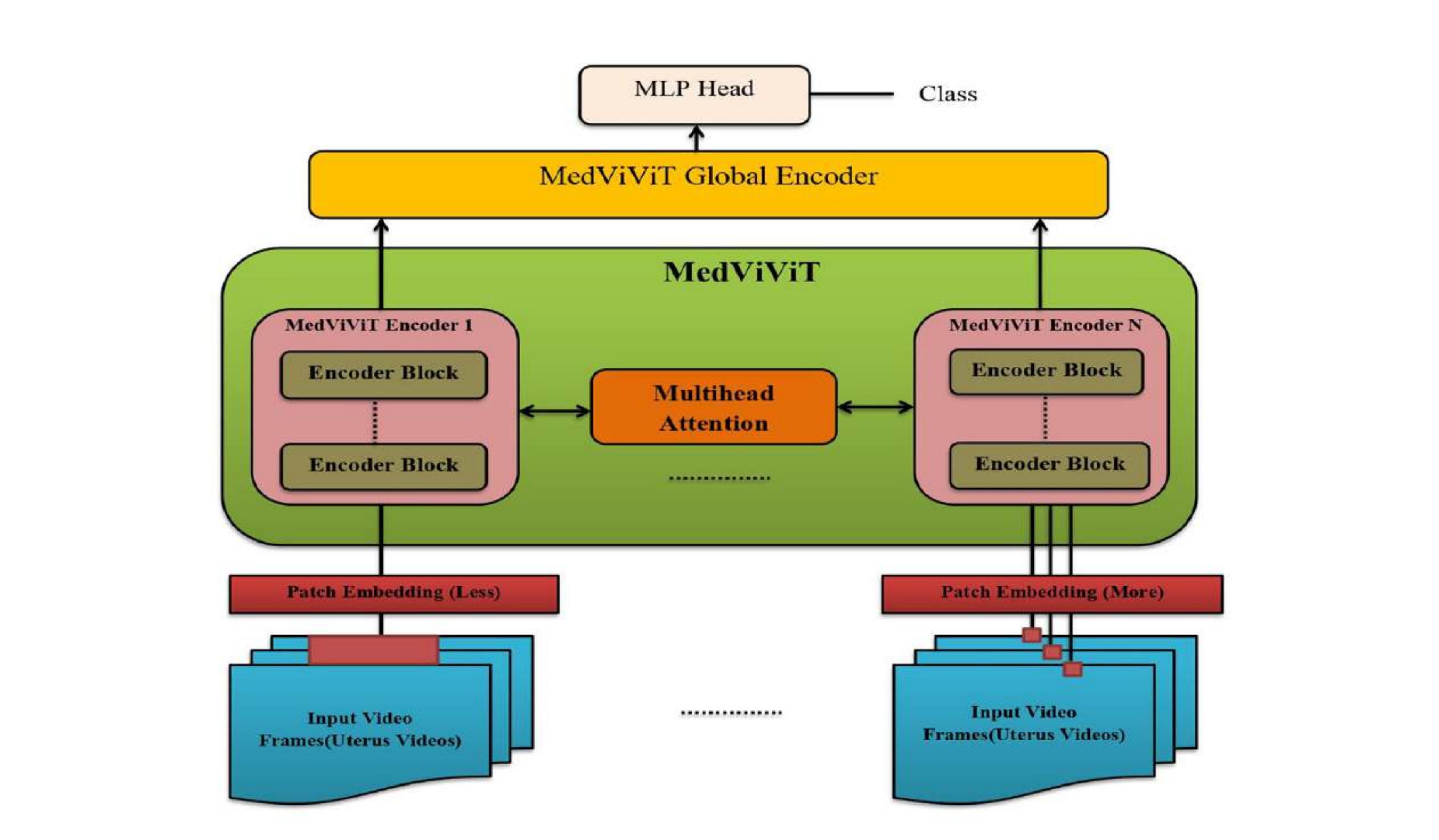


Figure : Architecture diagram of MedViViT Model

Task 1: Proposed MedViViT Model Design

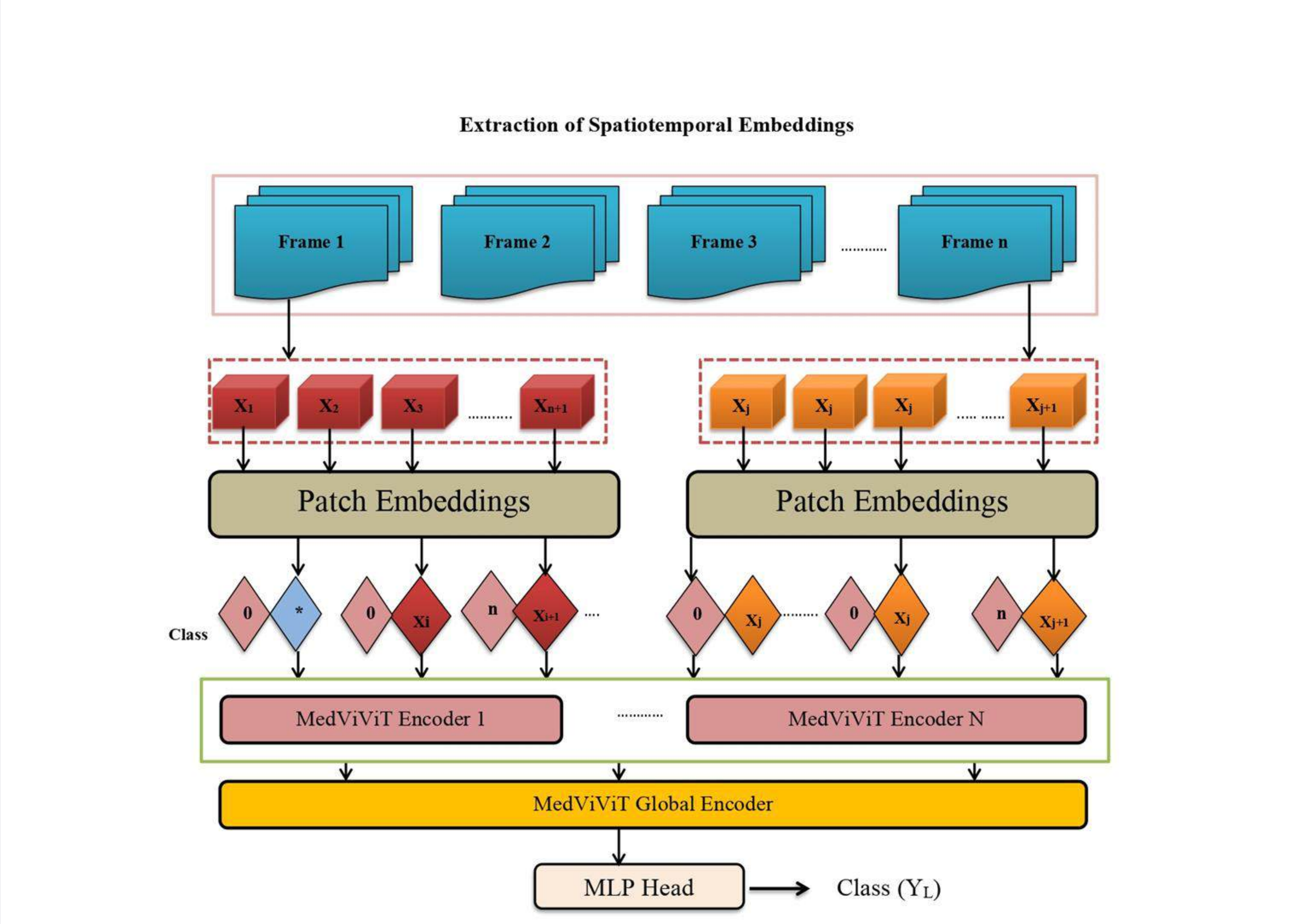


Figure : Feature Extraction using MedViViT Model

Task 1: Proposed MedViViT Model Design

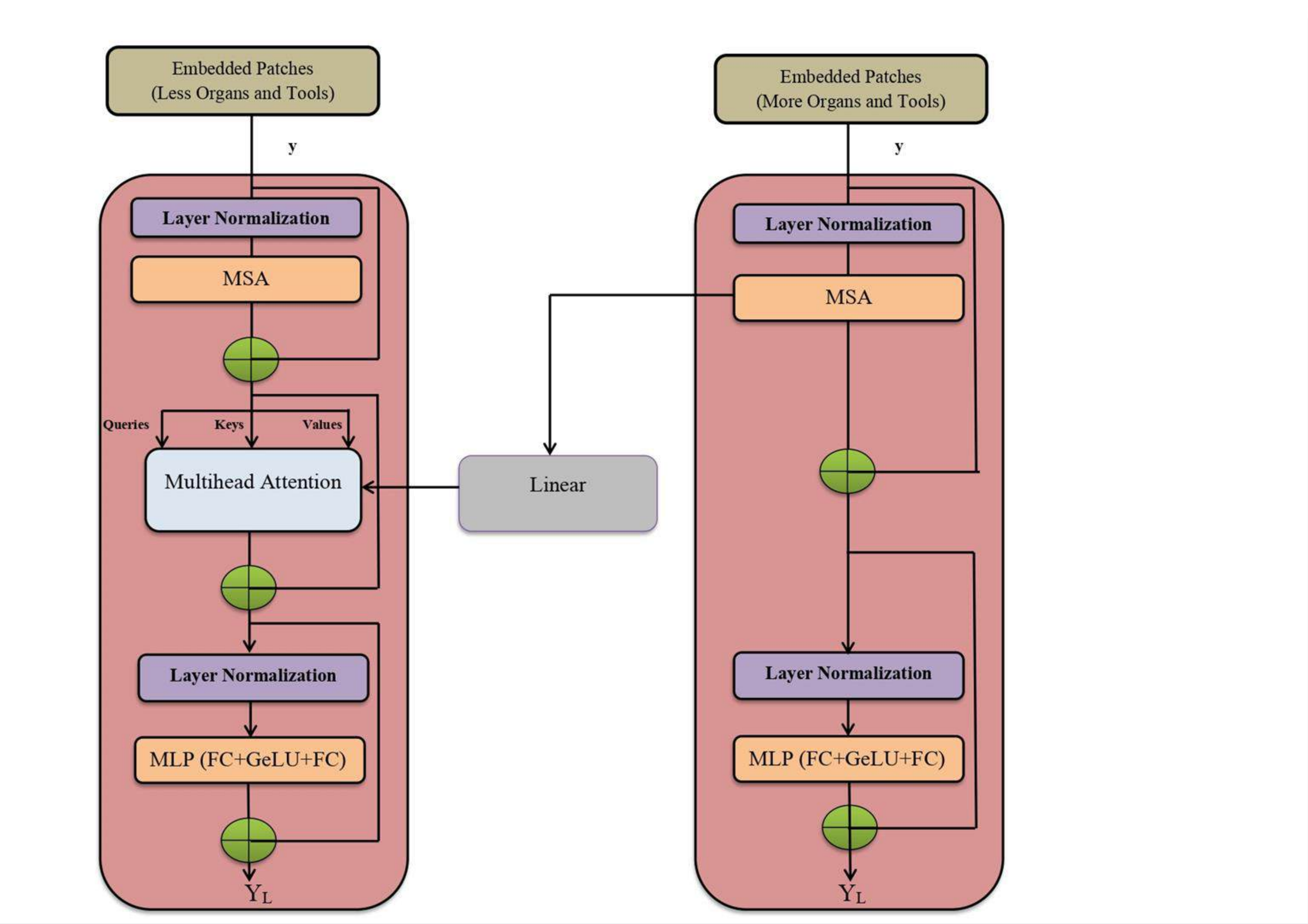


Figure : Attention Mechanism of MedViViT Model

Proposed MedViViT Model Design

- MedViViT Hyper Parameter Tuning

	Selection Criteria	**Choice Made**
Optimization Parameters		
Batch Size	64	64
Epochs	50-200	100
L2 Penalty	L2 {0.1,0.01,0.001,0.0001}	0.1
Dropout	0.0 - 1.0	0.1
Momentum	0.0 - 1.0	0.9
Data augmentation		
Random crop probability	1.0 - 1.5	1.0
Random flip probability	0.1 - 0.5	0.5
Maximum Scale	1.0 - 1.5	1.33
Minimum Scale	0.1 - 1.0	0.99

Task 2: Customized YOLOv8 for Tool Segmentation

- In YOLOv8, the segmentation heads consist of convolutional layers followed by upsampling layers (such as transposed convolution or nearest-neighbor upsampling). The output of each segmentation head is a set of feature maps with the same spatial resolution as the input image, where each pixel represents the probability or confidence of belonging to a particular class or object category.

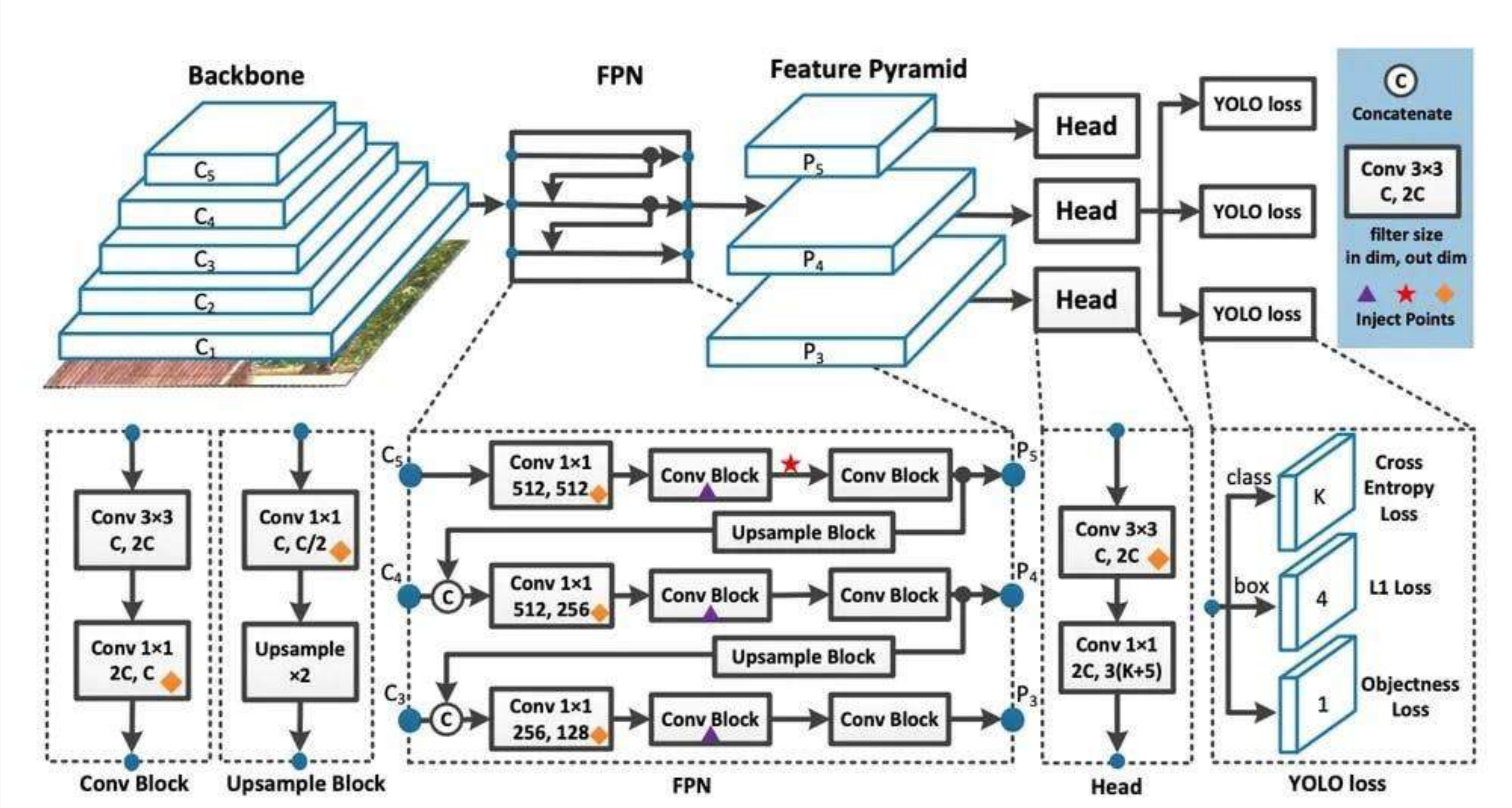


Figure : Architecture diagram of YOLOv8 Model

PHASE 01

Preparation

PHASE 02

Dividing Ligament and Peritoneum

PHASE 03

Dividing Uterine Vessels and Ligament

PHASE 05

Specimen Removal

PHASE 06

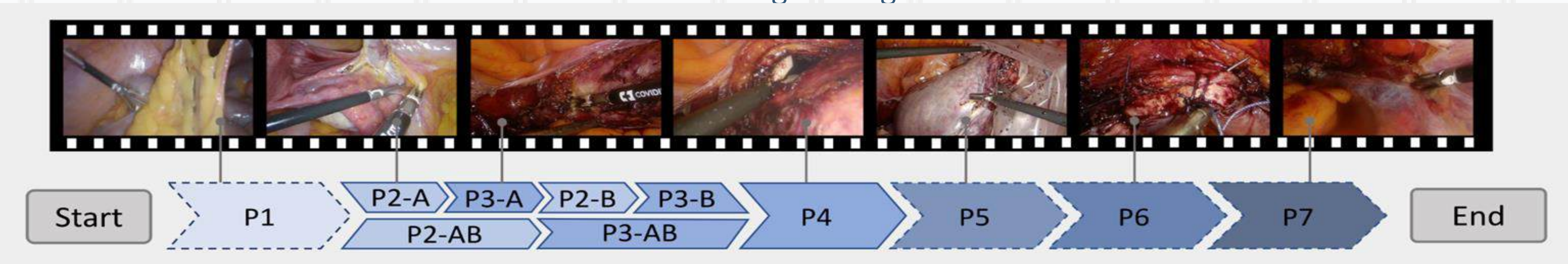
Suturing

PHASE 07

Washing

PHASE 04

Transecting the Vagina



UTERUS SURGERY WORKFLOW RECOGNITION

UTERUS SURGERY TOOLS(4 tools)

Grasping forceps, LigaSure, Dissecting and grasping forceps, Electric hook

Visual Q&A

Question 01

What is the phase at 25th second?

Answer 01

At the 25th second of the surgical video, the phase is still in the Preparation Phase, which lasts from 0 to 91 seconds. During this phase, parts such as the uterus and ovaries are detected using a laparoscope as a tool. .

Question 02

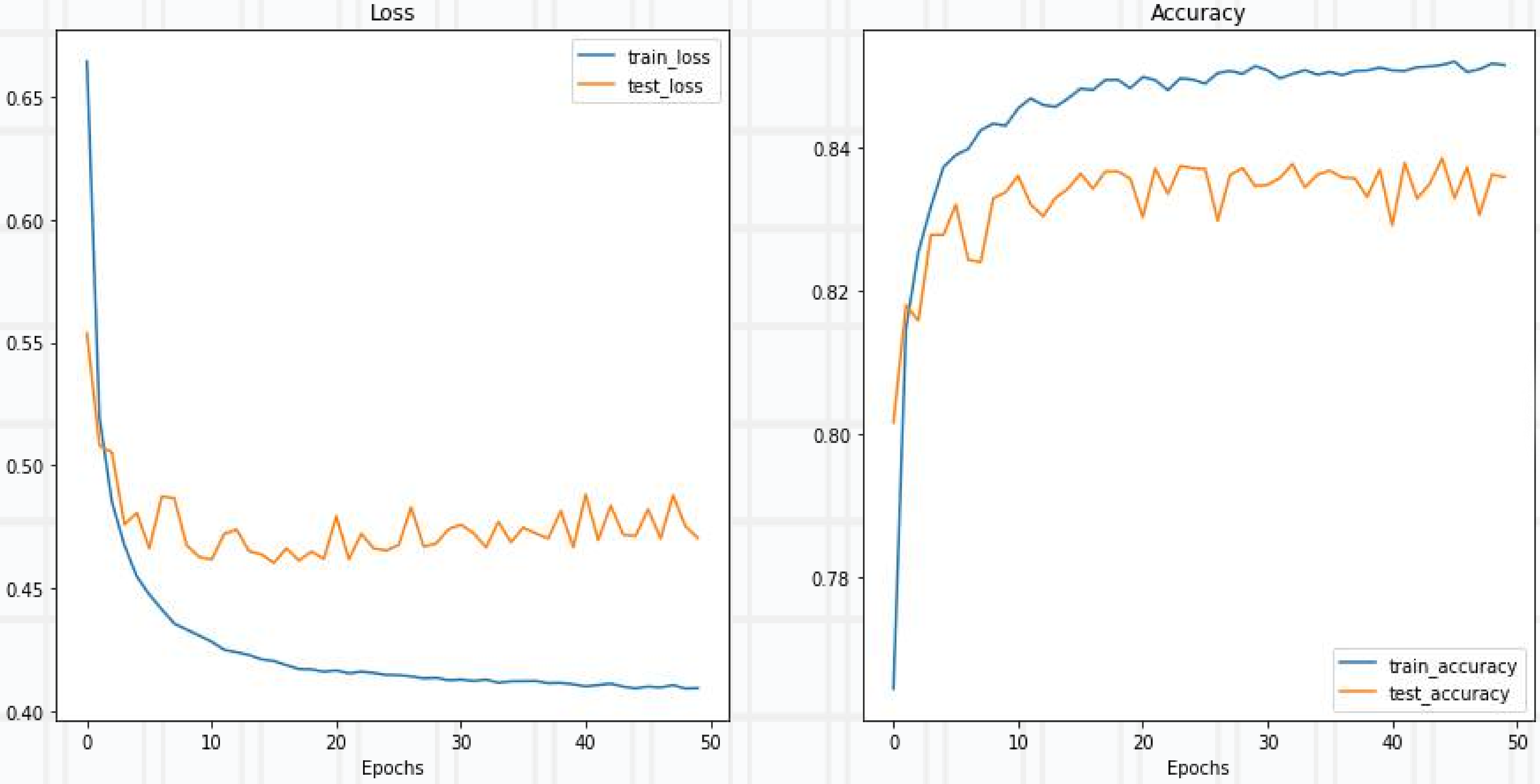
In which phases is the Scalpel detected?

Answer 02

The scalpel is detected in two phases of the surgical video. Dividing ligament and peritoneum Phase (92-102 seconds). The scalpel is used to divide the ligament and peritoneum, moving from the bottom right to the center of the image.

Sample questions

Performance of MedViViT



Accuracy : 83.77%

Loss : 0.40%

Figure : Accuracy and Loss Graph for MedViViT Model

Inference

```python
from going_modular.going_modular.predictions import pred_and_plot_image


custom_image_path = "test/382.jpg"
class_names = ['phase_1', 'phase_2', 'phase_3', 'phase_4', 'phase_5', 'phase_6', 'phase_7']

pred_and_plot_image(model=MedViVit,
                    image_path=custom_image_path,
                    class_names=class_names)
```

Pred: phase_2 | Prob: 0.639

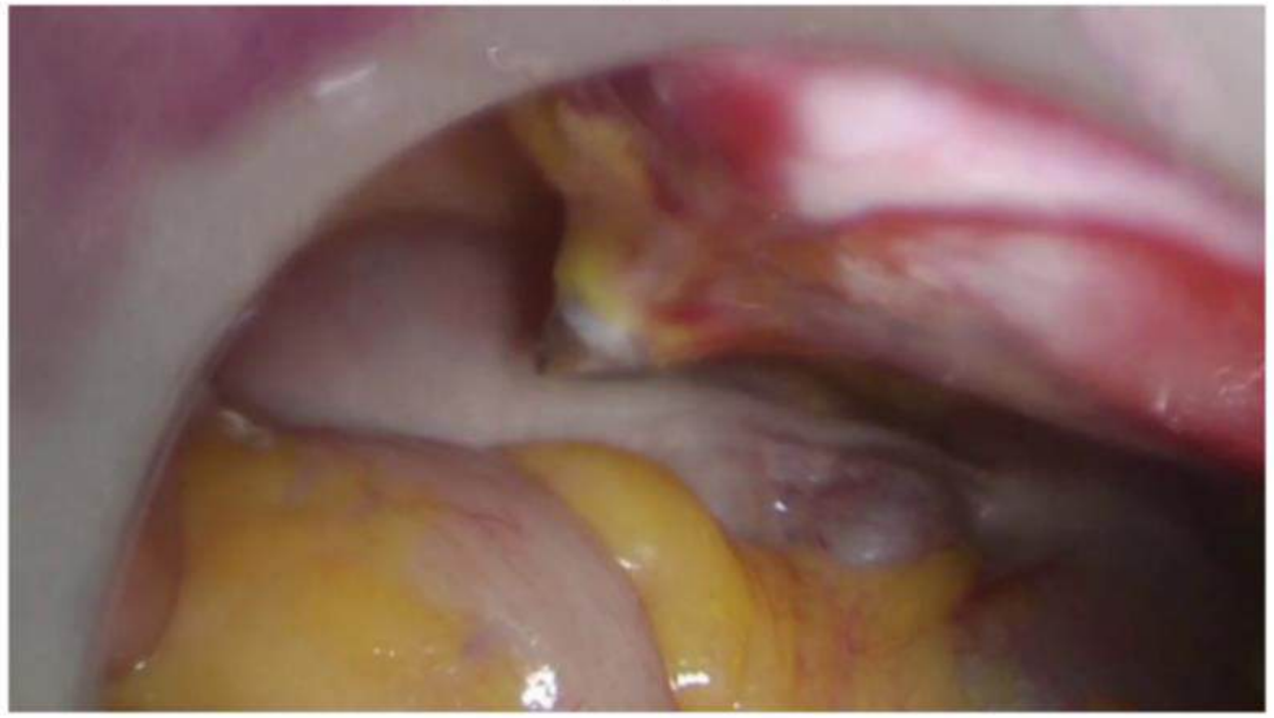


Figure : Uterus Phase / Workflow Recognition

Yolov8 Performance Analysis

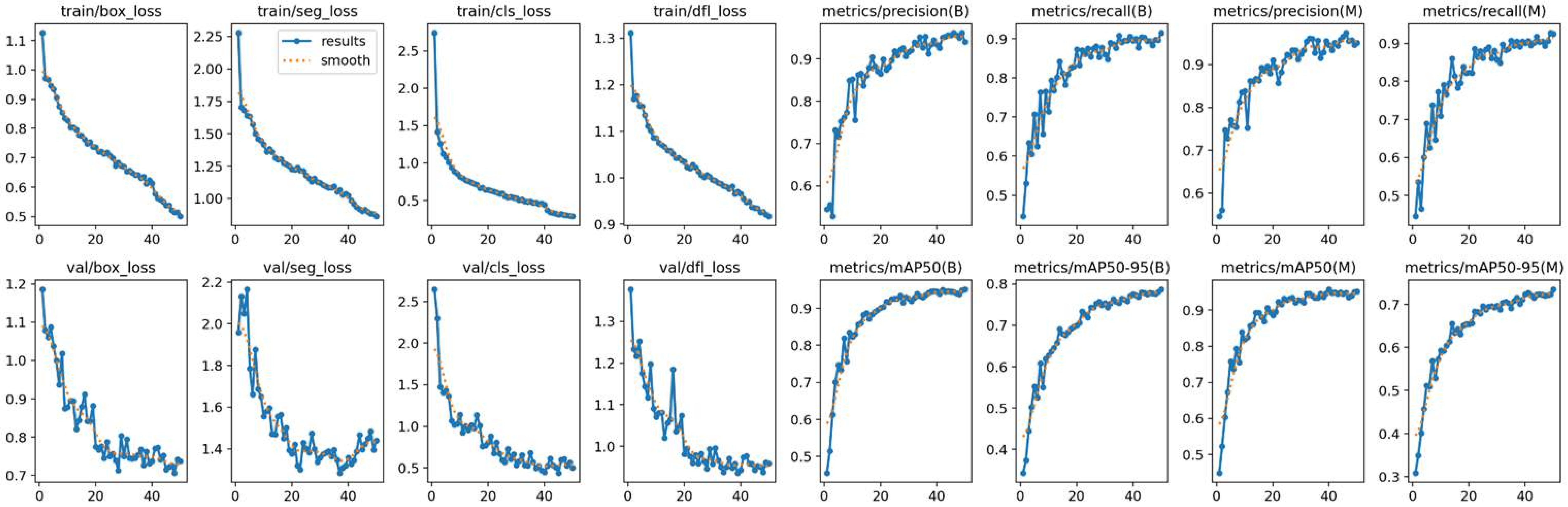


Figure : Precision, Recall, Mean Average Precision Metrics for Performance Analysis

Inference

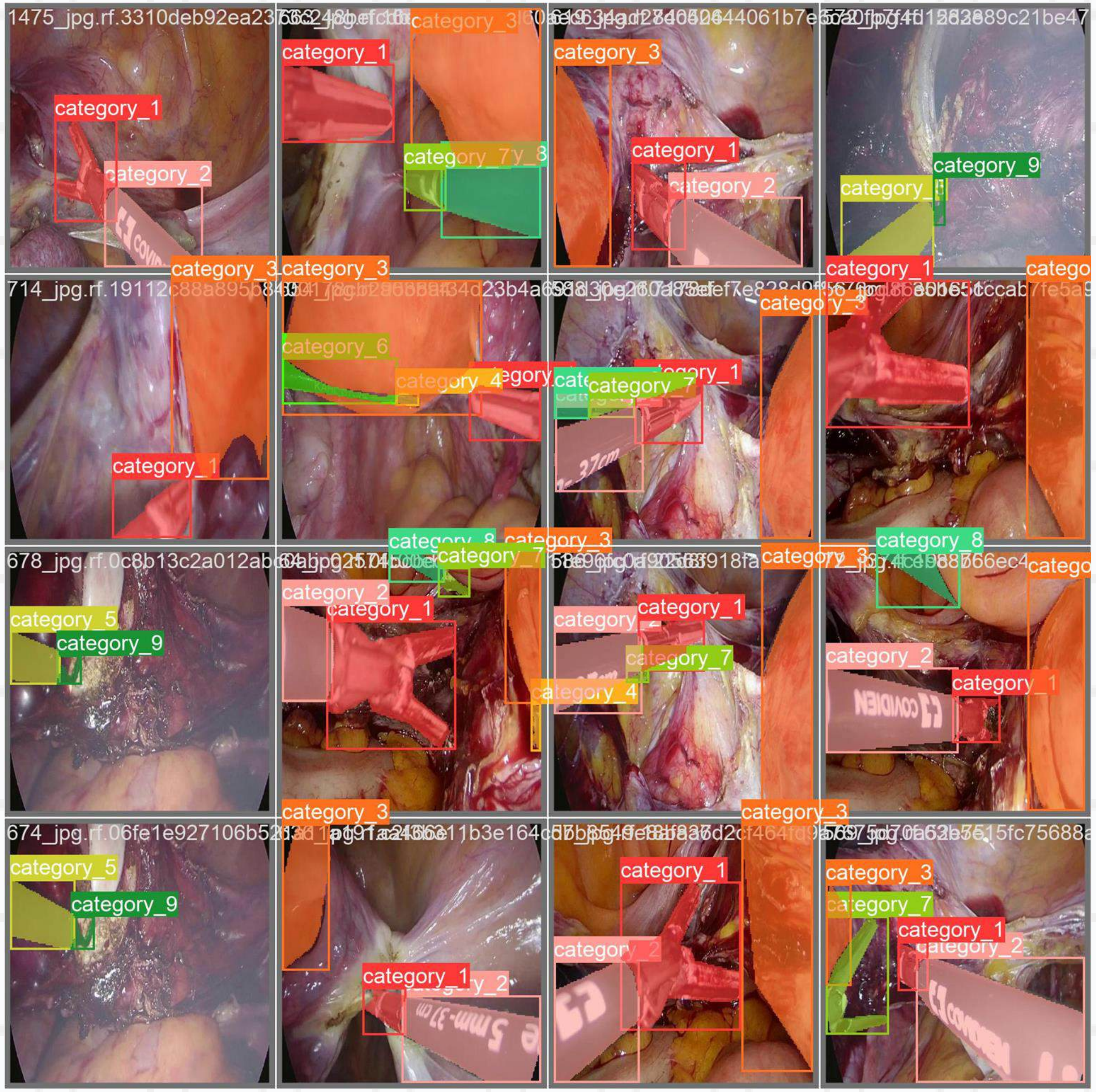


Figure : Instrument or Key tool Segmentation

STAKEHOLDERS

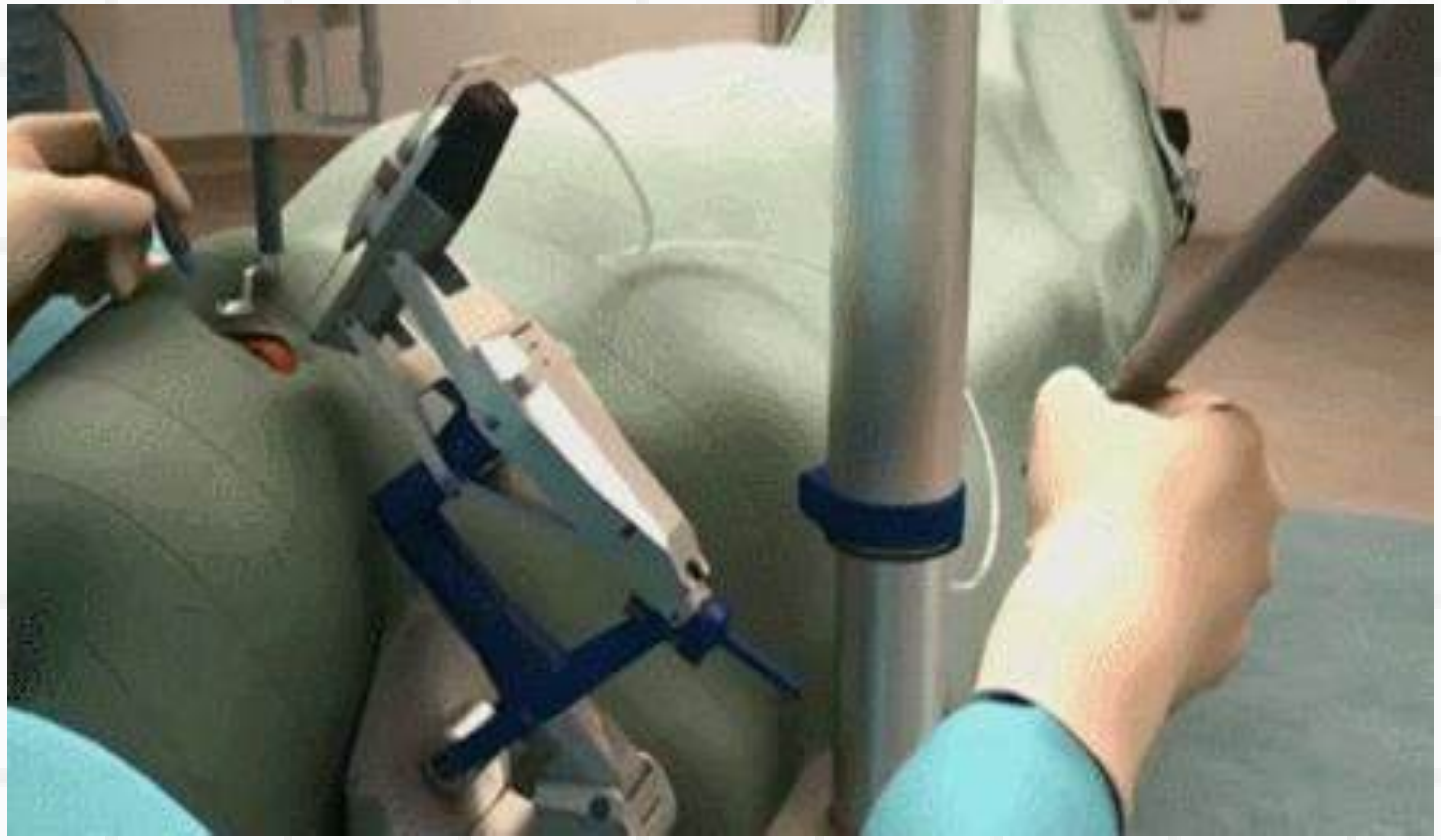
STAKEHOLDER 01

Medical Institutions and Hospitals that perform Robotic Surgery

STAKEHOLDER 02

Junior Doctors who are Studying Surgery

Stakeholder 1: Applications



- **Minimally Invasive Surgery:** Robotics and CAI are extensively used in minimally invasive surgical procedures, such as robotic-assisted laparoscopic surgery and robot-guided endoscopic procedures. These technologies enable surgeons to perform complex surgeries with greater precision, dexterity, and control, leading to reduced trauma, faster recovery times, and improved patient outcomes.

- **Telemedicine and Remote Surgery:** Robotics and CAI enable telemedicine platforms and remote surgery systems, allowing healthcare providers to deliver medical care and perform surgical procedures from a distance. This is particularly beneficial in remote or underserved areas where access to specialized healthcare services may be limited.

Stakeholder 2: Applications

- **Training and Education:** Junior doctors often require extensive training and education to develop their surgical skills and knowledge. A Medical ChatGPT system could serve as a valuable educational tool, providing real-time guidance, explanations, and feedback on surgical procedures, techniques, and decision-making.

- **Decision Support:** During surgical procedures, junior doctors may encounter challenging situations or unfamiliar cases where they require immediate assistance or advice. ChatGPT could act as a virtual mentor, offering guidance, suggestions, and evidence-based recommendations to support junior doctors in making informed decisions and navigating complex surgical scenarios.

PRODUCT MARKETING

STRATEGY 01

We can organize live demonstrations, workshops, and training sessions to showcase the capabilities of Surgeons's Eye at medical institutions that perform surgery

STRATEGY 02

We can leverage social media platforms to engage with healthcare professionals, share relevant content, and participate in conversations related to proposed Surgeons's Eye medical innovation.

STRATEGY 03

We can use targeted digital advertising campaigns on platforms such as Google Ads and LinkedIn to reach healthcare professionals and decision-makers in medical institutions for marketing proposed Surgeons's Eye product.

Conclusion

- In conclusion, the innovation of Surgeon's Eye represents a significant advancement in the field of medical surgery, facilitated by the development of a cutting-edge Chat system integrated with the proposed MedViViT model and YOLO model. This innovative solution addresses critical challenges faced by medical professionals by providing real-time assistance in surgical workflow recognition and tool recognition through natural language interaction.

- The deployment of Surgeon's Eye in operating theatres can streamline surgical workflows, improve surgical outcomes, and enhance patient safety. Medical professionals can benefit from real-time guidance and assistance during procedures, reducing the likelihood of errors and complications.

Novelty or Innovation

1. Intelligent Video Scene Understanding by applying Instance Segmentation and Classification for surgical vidoes for surgical phase, gesture, tool and anomaly detection using Customized yoloV8 MODEL AND proposed MedViViT transformer.
2. Expansion of Dataset to create huge Repository with three surgery videos on Ophthalmology, Laparoscopic, Cholecystectomy, Laparoscopic Hysterectomy for researchers.
3. Intelligent ChatBot for Visual Question Answering integrated with Llamma 3 called surgeons eye.

Future Work

- In future work, we envision expanding the Surgeons' Eye dataset to include additional surgical procedures and tasks, further enhancing its utility as a resource for advancing learning-based automation in CAI.
- Moreover, efforts to collaborate with medical institutions and professionals can facilitate the collection of more diverse and comprehensive data, ultimately driving innovation and improvements in computer-assisted minimally invasive surgery.

DEMO

Q&A