Image Segmentation on Laparoscopic Images using Deep Learning for Augmented Reality

Tom FRANCOIS - Supervisors: Adrien BARTOLI, Lilian CALVET and Damien SABOUL.

Université Clermont Auvergne, CHU Clermont-Ferrand, CNRS, SIGMA Clermont, Institut Pascal, F-63000 Clermont-Ferrand, France

Introduction

- Numerous diseases are nowadays treated using minimal invasive surgery to access the internal anatomy of the patient. Contrary to open surgery which requires a large incision of the abdominal wall to gain access to a lesion, laparoscopic surgery requires smaller incisions and is less traumatizing.

Figure 1: Laparoscopic surgery.

- One of the goals of medical Augmented Reality (AR) is to reveal the hidden anatomy, such as a tumour or major blood vessels within an organ. The EnCoV team has developped an AR system for gynecology to augment tumors inside the uterus.

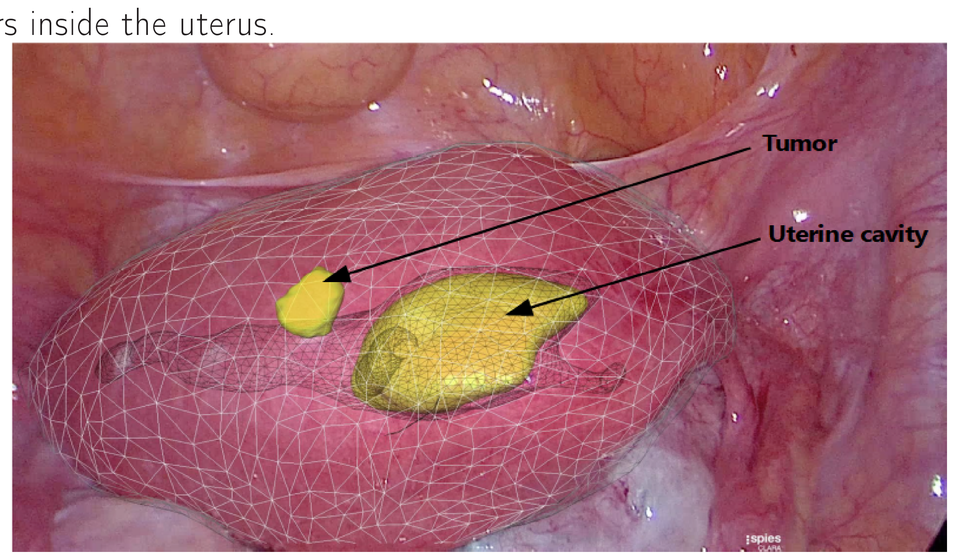

Figure 2: Augmentation of an uterus during a surgery.

- The current AR system requires manual inputs like annotating the live video to locate the organ. The goal on this project is to replace these manual steps using deep learning approaches. Convolutional Neural Networks (CNN) can now solve challenging computer vision issues.

Creation of a Custom Dataset

- No public dataset is available for uterus segmentation in laparoscopy.
- We are currently collecting data from surgeries performed by members of EnCoV. An annotated dataset containing laparoscopic images of gynecologic operations is being built based on these live videos. Still images are extracted and manually annotated. Compared to common object datasets (the COCO dataset contains more than 200K annotated images), ours is currently really small (350 annotated images).

First Results

- Transfer learning: Transfer learning is a way of reusing a model trained on a first dataset by finetuning on another specific dataset. In our case, we reused a model trained on the COCO Dataset [1] (persons, vehicles, animals, ...) and finetuned an instance segmentation system [2] with our custom dataset. These results of this approach can be seen in Figure 3. This approach is currently limited by the low quantity of annotated data available.

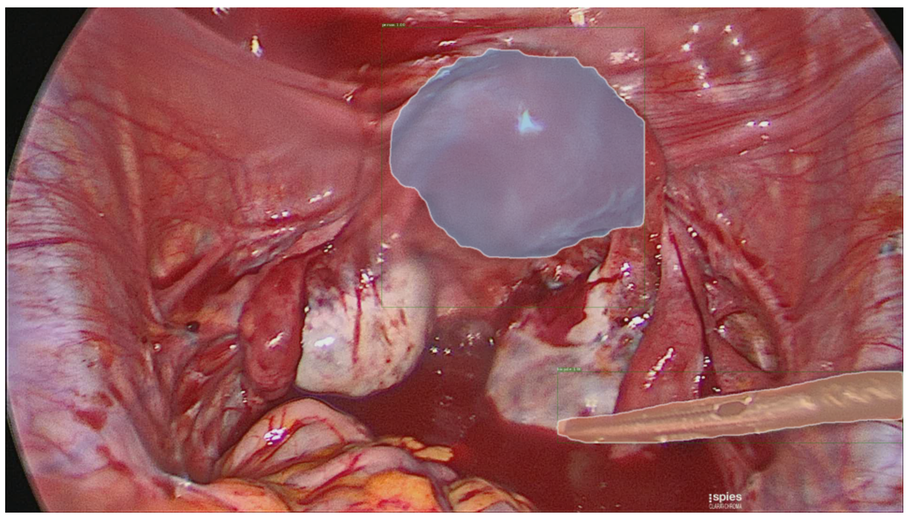

Figure 3: Automatic detection and segmentation of the uterus (blue) and surgical tool (yellow).

Current Approach

- The AR system need to register the preoperative data, that contain the 3D model of the uterus and the tumors (b), with the intraoperative video stream of the surgery (c) (see Figure 4). One of the cues we are using to perform this is the occluding contours of the uterus. We want to develop a deep learning algorithm to extract these contours from frames extracted of the video stream.

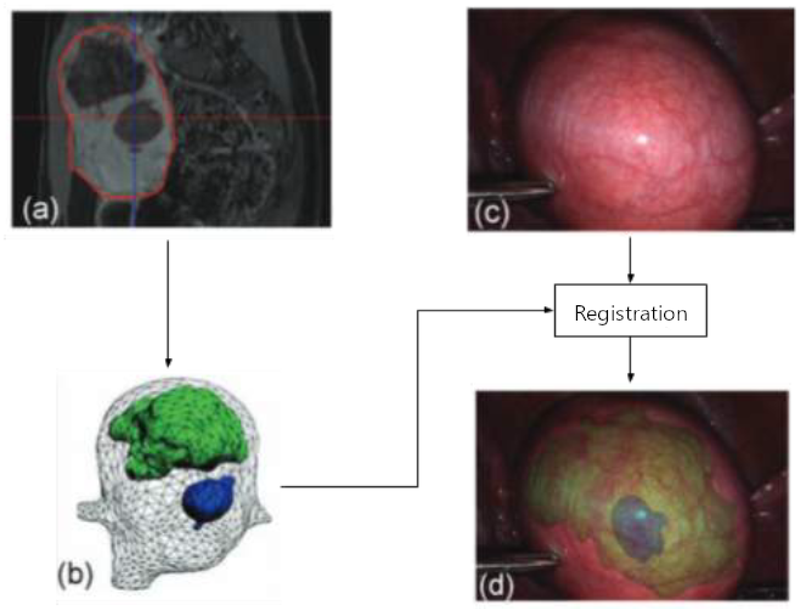

Figure 4: Diagram of the process of our AR system

- In order to pretrain our CNN, we use an external contours extractor to create the annotated data. Once our model is trained, we finetune this to detect different type of contours : occluding contours of the uterus, surgical tool contours, etc. (see Figure 5).

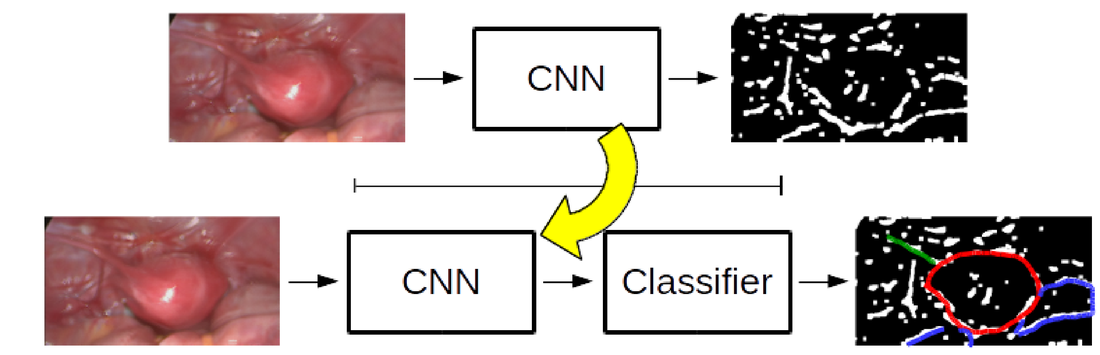

Figure 5: The top row illustrate a first learning phase to train a CNN to extract contours from laparoscopic images. The bottom row presents a second learning phase using the previous CNN models to classify contours (uterus occluding contours, surrounding organ contours, etc).

Conclusion

Using deep learning methods on the augmented reality system could help clinicians perform their surgery more precisely and eventually faster.

Next steps

- Increase the size of our laparoscopic dataset.
- Complete and improve our solution to automatically classify and extract contours of the organ to perform the 3D registration between preoperative and intraoperative data.

Acknowledgments

- This research is conducted within a CIFRE convention between Almerys (Be-Studys) and EnCoV (Institut Pascal).

References

[1] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick.
Microsoft coco: Common objects in context.

[2] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick.
Mask r-cnn.

[3] Carl Doersch, Abhinav Gupta, and Alexei A. Efros.
Unsupervised visual representation learning by context prediction.