SurgAI3.8K: a labelled dataset of gynaecologic organs in laparoscopy, with application to automatic augmented reality surgical guidance

Running title: SurgAI3.8K laparoscopic image dataset

Title: SurgAI3.8K: a labelled dataset of gynaecologic organs in laparoscopy, with application to automatic augmented reality surgical guidance

Author names and affiliations: Sabrina Madad Zadeh M.D. (1,2), Tom François Ph.D. (2), Aurélie Comptour Ph.D. (3), Michel Canis M.D. Ph.D. (2,3), Nicolas Bourdel M.D. Ph.D. (2,3) and Adrien Bartoli Ph.D. (2,4)

1. Surgical Oncology Department, Centre Jean Perrin, 63011 Clermont-Ferrand, France
2. EnCoV, Institut Pascal, UMR 6602 CNRS/Université Clermont-Auvergne, Clermont-Ferrand, France
3. Department of Obstetrics and Gynecology, University Hospital Clermont-Ferrand, 63000 Clermont Ferrand, France
4. Department of Clinical Research and Innovation, University Hospital Clermont-Ferrand, 63000 Clermont Ferrand, France

Corresponding author: Nicolas Bourdel. nicolas.bourdel@gmail.com. Department of Obstetrics and Gynecology, University Hospital Clermont-Ferrand, 63000 Clermont Ferrand, France. +33676713113

IRB: IRB 2016-002773-35

Précis: We provide the surgical dataset SurgAI3.8K, train an Artificial Intelligence system to recognise gynaecologic organs and show its direct impact in an augmented reality surgical guidance software.

Abstract

Study Objective: We focus on explaining the concepts underlying Artificial Intelligence (AI), using Uteraug, a laparoscopic surgery guidance application based on Augmented Reality (AR), to provide concrete examples. AI can be used to automatically interpret the surgical images. We are specifically interested in the tasks of uterus segmentation and uterus contouring in laparoscopic images. A major difficulty with AI methods is their requirement for a massive amount of annotated data. We propose SurgAI3.8K, the first gynaecological dataset with annotated anatomy. We study the impact of AI on automating key steps of Uteraug.

Design: We constructed the SurgAI3.8K dataset with 3800 images extracted from 79 laparoscopy videos. We created the following annotations: the uterus segmentation, the uterus contours and the regions of the left and right fallopian tube junctions. We divided our dataset into a training and a test dataset. Our engineers trained a neural network from the training dataset. We then investigated the performance of the neural network compared to the experts on the test dataset. In particular, we established the relationship between the size of the training dataset and the performance, by creating size-performance graphs.

Setting: University

Patients: NA

Intervention: NA

Measurements and main results: The size-performance graphs show a performance plateau at 700 images for uterus segmentation and 2000 images for uterus contouring. The final segmentation scores on the training and test dataset were 94.6% and 84.9% (the higher, the better) and the final contour error were 19.5% and 47.3% (the lower, the better). These results allowed us to bootstrap Uteraug, achieving AR performance equivalent to its current manual setup.

Conclusion: We describe a concrete AI system in laparoscopic surgery with all steps from data collection, data annotation, neural network training, performance evaluation, to final application.

Keywords: *artificial intelligence, deep learning, laparoscopic surgery, gynaecological surgery, augmented reality, computer-assisted surgery*

1. Introduction

Computer-aided surgery systems require the computer to interpret surgical images automatically. In this respect, Artificial Intelligence (AI) has recently shown unprecedented performance in the technical literature, in particular via the deep learning approach (1). The key idea in deep learning is to train a neural network to replicate results created by experts. A neural network is an artificial object created in the computer's memory[1]. The concept of training can be understood as 'teaching', as it is also said that the neural network 'learns from data'. This indeed works by means of creating a dataset, which is an ensemble of images, where the expected results were manually annotated by experts. Our objective is to explain these concepts in detail using a concrete example of surgical application. Specifically, we show that a neural network can be used to automate Uteraug, a visual guidance software for gynecologic surgery developed by our team (2). This article results from the collaboration

[1] A neural network is an artificial object which solves a specific task. Alternatively, the expression neural networks may be found in the literature to encompass the set of methods related to deep learning.

between three expert surgeons (SMZ, MC and NB) and two scholars researching and engineering the techniques of AI and their application to surgery (TF and AB).

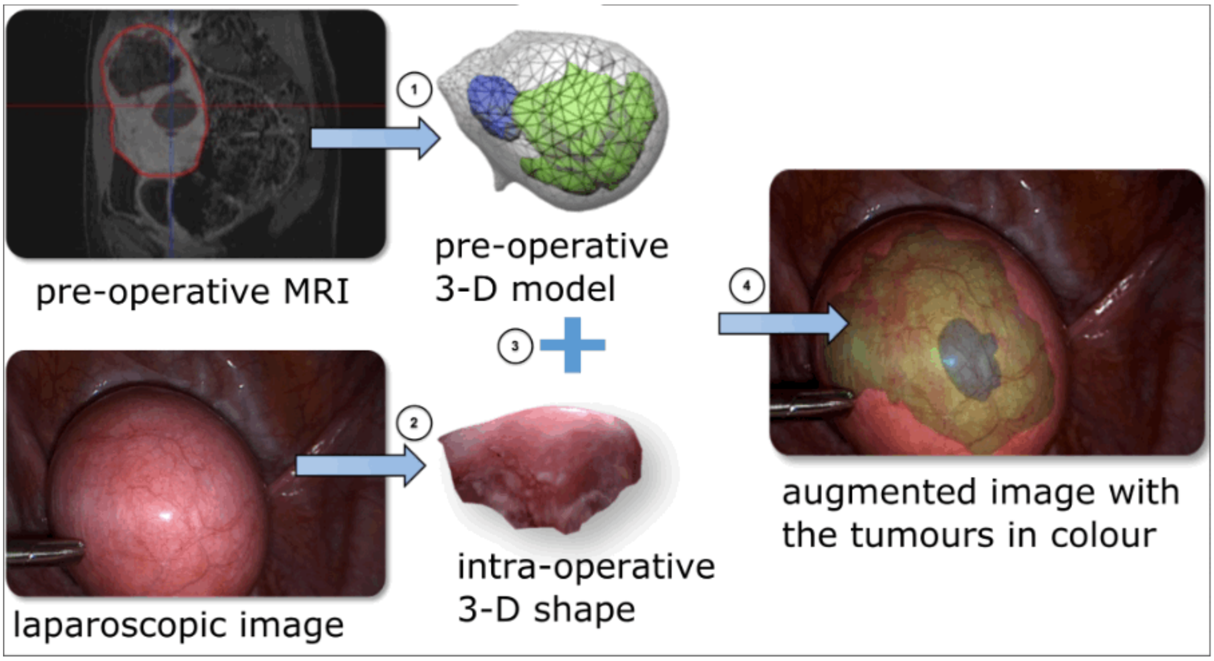

Figure 1: Sketch-up of Uteraug, the EnCoV team's AR system (4)**.** Example of a laparoscopic myomectomy assisted by AR. Step 1: the preoperative 3D model is reconstructed from pelvic MRI. Step 2: the intraoperative 3D model is reconstructed from a set of laparoscopy images. Step 3: the deformation between the pre- and intraoperative states of the uterus is computed. Step 4: the uterus is tracked automatically and augmented in real time with two myomas using custom colours, creating the effect of virtual transparency. Our proposed neural network has been the cornerstone to automate steps 2 and 4, dropping the need for surgeon attention to setup the system and dramatically increasing usability. In the longer run, our dataset and methods may be used to solve other problems in the implementation of computer-aided surgery support systems. With this in mind, we have included extra annotations in our dataset, namely the junctions between the uterus and fallopian tubes.

As we shall see, automating Uteraug requires us to solve two tasks. Uteraug implements a virtual transparency visualisation mode of the uterus by fusing preoperative 3D images with the laparoscopic images, as shown in figure 1. In its setup, Uteraug requires the surgeon to select the region occupied by the uterus and its contours in several laparoscopy images, which is a strong limitation in terms of clinical usability. In the context of AI, these two tasks are referred to as *uterus segmentation* and *uterus contouring*, which are illustrated in figures 2 and 3. Uterus segmentation consists in labelling each image pixel as being uterus or non-uterus. Uterus contouring consists in labelling each image pixel as being uterus contour or non-contour. Technically, a contour is a boundary between the image part containing the uterus and the rest of the image. These two tasks are extremely simple to solve for an expert in most cases. The human brain is indeed particularly well-equipped to recognise and delineate objects from images. However, in spite of its simplicity, labelling the extent of all pixels in an image is extremely time-consuming for an expert. A major advantage of a neural network is that, once properly trained, it can solve this type of task in a split second, typically processing several dozen images in a second, without any further expert supervision.

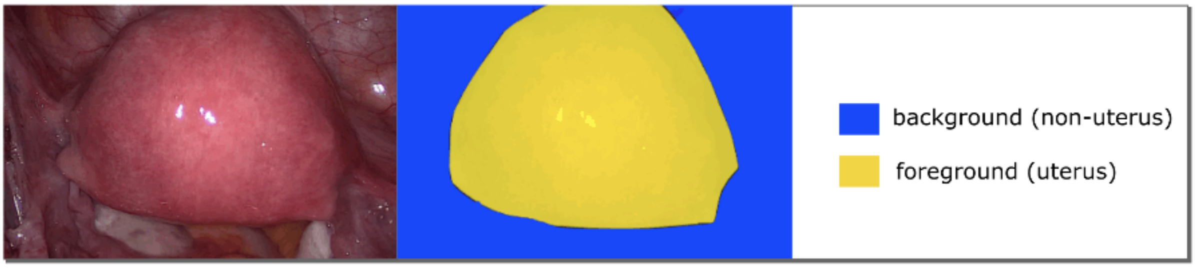

Figure 2: Uterus segmentation. Each image pixel receives one of two labels, namely uterus and non-uterus. The result is called a segmentation mask and is a binary image, which can be visualised with black and white or any other two colours.

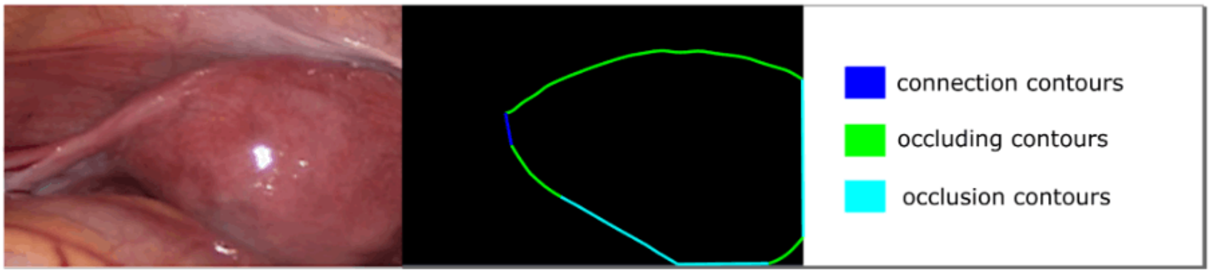

Figure 3: Uterus contouring. Each image pixel receives one of four labels, namely occluding contour (the visible boundaries of the uterus), occlusion contour (for instance, the boundary created by the sigmoid colon in front of the uterus), connection contour (the connection between occluding and occlusion contours) and non-contour. The result is called a contour mask and is a four-colour image, which can be visualised using any four colours. The uterus contouring task is to achieve the automatic detection of each type of contour.

A major difficulty with neural networks is their requirement for a massive amount of manually annotated data (3) to be trained for a specific task. Such data are gathered in a so-called dataset, which in practice requires surgeons to record surgeries and organise for experts to label the images. This requires attention and time, making existing datasets extremely valuable. We propose SurgAI3.8K, the first large gynaecologic dataset, comprising 3800 labelled images. A very important question, which is regularly asked when creating neural networks, regards the required quantity of data; otherwise said, the minimal size of the dataset required to achieve the desired task with the expected performance. It has been verified empirically that for most tasks, the larger the dataset, the better the performance. However, data is expensive, both in terms of collection and annotation. A sound way of determining when to stop data collection is to monitor the quantitative performance of the neural network as the dataset is being collected and annotated. Observing that the performance plateaus, and if the performance is sufficient for the target application, is generally a reasonable indication that data collection can be stopped. We show experimentally that our dataset

SurgAI3.8K is large enough to train a neural network with reliable performances for uterus segmentation and uterus contouring.

Finally, our neural network is shown to be a successful replacement of the surgeon for the manual tasks in Uteraug.

2. Methods

2.1. General Points

2.1.1 Deep Learning and Neural Networks

Deep learning is the scientific field which deals with large neural networks. A specificity of neural networks is that they do not require one to program the computer to explicitly perform a task. Rather, the neural networks are trained from a dataset containing information about the task, similarly to a human being taught to perform a task from examples. Indeed, the most common training paradigm is called *supervised training*, which requires the dataset to contain examples with their expected results. Training is generally a long process, requiring heavy computational power and the attention of expert engineers. However, it needs to be done only once, representing the first phase in the life cycle of a neural network. The second phase is called *prediction*. At this phase, the neural network is simply used to solve the target task for any new image used at input. The two essential phases to create and use a neural network are thus summarised as follow:

- Phase 1: training - the neural network is trained from many examples showing how the target task is solved by experts.
- Phase 2 : prediction - the neural network is used to predict the result, in other words to solve the task, for new cases without requiring the attention of experts.

The remainder of this paragraph formalises the concept of supervised training. It is slightly technical and may be skipped on a first reading. Formally, we denote an input as *X* -in the case at hand, X is a laparoscopy image- and the expected result as Y -in the case at hand, we choose Y to be the uterus segmentation for simplicity. The neural network is modelled by a mathematical function f, which has a fixed mathematical form representing the neural network designed by the engineer. Specifically, the neural network design specifies the number of artificial neurons being used and the way they are connected, similarly to biological neurons. Function f takes X as input and maps it to Y. A neural network with some prescribed architecture can be trained to solve many different tasks. This is because its behaviour is controlled by a set of parameters, contained in a variable p, which defines the firing rate of each of the artificial neurons the neural network is made of, again, similarly to biological neurons. Therefore, function f not only depends on X but also on p. The training process attempts to capture the relationship that exists between a laparoscopy image and a uterus segmentation by finding an optimal value for p. Technically, it finds the parameters in p which minimise the distance between Y, which is the expected uterus segmentation from the expert, and $f(X,p)$, which is the prediction of uterus segmentation made by the neural network for a given laparoscopy image X and parameters p. More specifically, the training process estimates p from the whole training dataset, generally containing many pairs (X,Y) of input laparoscopic image and expected uterus segmentation. Once p has been estimated from the training phase, the neural network is ready to be used on new data. This is the prediction phase, whereby the parameters p are frozen and the function $f(X,p)$ used to predict the uterus segmentation Y from a new, previously unseen, input laparoscopy image X.

The specificity of deep learning within the general world of machine learning and AI is related to the design of the neural networks it uses. These neural networks are based on artificial neurons, organised in layers connected to each other. The ‘deep’ qualifier comes from the large number of such layers, forming a so-called deep neural network. The neural network design defines its structure and its number of layers; it is also called the *neural network architecture* in the technical literature. The choice of the neural network architecture is critical to obtain reliable results. The architecture we chose for uterus segmentation and uterus contouring is discussed in section 2.2.

2.1.2 Annotation and Dataset Size

The annotations represent the expected results of the tasks that the neural network should learn, in other words, they are examples used to teach the neural network’s purpose. The annotation process is generally carried out manually. For common objects, the annotation can be done by anyone. In the medical field however, the required level of expertise reduces the number of reliable labellers. So, on the one hand, annotating data is time-consuming and on the other hand, the larger the training dataset, the better the final neural network performance. Determining the optimal size of the dataset is thus critical in practice, to best compromise feasibility and performance. We have proposed a methodology to address this problem based on creating a size-performance graph, described in section 2.3.1.

2.2. Architecture Design

The tasks at hand -uterus segmentation and uterus contouring- are strongly related. Specifically, knowing the segmentation is a strong cue to solve contouring, while knowing the contours should directly allow one to deduce the segmentation, as the inner part of the closed contours. Therefore, a natural question is whether we should strive to solve both tasks, or solve just contouring. Theoretically, this is a sound question, but in practice the contours are not guaranteed to be closed due to imperfect annotations and predictions, as seen for instance in the case of figure 3. Nevertheless, it remains true that both tasks are strongly related. This fact will be exploited by our technical solution, which uses a neural network architecture solving both tasks simultaneously. More specifically, our engineers chose an existing neural network architecture well-adapted to medical images called U-Net (4) and specialised it to the tasks at hand. In short, the proposed neural network has the following input and output specificities:

- Neural network inputs: the laparoscopy image.
- Neural network outputs: the segmentation and the contours.

The proposed methodology is applicable to any dataset containing contour annotations, whether it be a dataset of surgical images or other modalities such as radiological images.

2.3. Dataset Creation and Neural Network Training

2.3.1 Dataset Size

Finding the optimal dataset size is a challenging question because, as we have seen, it represents a trade-off between labelling effort and performance. The relationship to performance is easy to understand: an object has a visual appearance depending on several factors, including its position with respect to the camera and the background it lies on. The larger the number of examples which the neural network learns from, the better it will extrapolate to new data. However, beyond a certain quantity of examples, the addition of new examples will only lead to a marginal performance gain which is probably not worth the labelling effort. Hence, an optimal dataset size may be found as the best compromise between the labelling manpower availability and cost, and the incremental performance gain.

We propose to determine an optimal dataset size for uterus segmentation and uterus contouring by studying size-performance graphs. We measure performance using the so-called *test error*. The test error is an extremely simple, yet important notion. Once the neural network is trained, the test error is computed from the test dataset, containing data independent of the training dataset. In other words, the test error uses images which were not used for training, and for which the expected results are available, to compare the prediction of the neural network against the expert. It is thus customary for the engineers to split the dataset in two parts: the training dataset, which is typically about 80% of the dataset, which is used to train the neural network, and the test dataset, which is typically about 20% of the dataset, which is used to evaluate the performance independently.

Our methodology to construct the size-performance graphs is to train the neural network incrementally. We start with a small subset of the dataset of size 100 images, train the neural network and measure its performance. We then add a batch of 100 images to the training dataset and repeat the steps. We thus obtain the sought size-performance graph, into which we search for a performance plateau. The test dataset used to measure performance is fixed. It contains 581 images, representing approximately 15% of our 3800 images. The test and training datasets contain images from different procedures to prevent any patient overlap and to guarantee an unbiased performance evaluation.

2.3.2 Data Source, Extraction and Selection

We construct SurgAI3.8K, our proposed dataset, by extracting and labelling individual frames from 79 laparoscopy videos. The videos were recorded as part of a research protocol (IRB 2016-002773-35) (6,7). When creating a dataset, it is crucial to ensure data diversity. The dataset should be large, but it should also span the possible usage conditions. We have taken care of using videos capturing both intra-patient and inter-patient diversity. Intra-patient diversity is covered by including images with various uterus viewpoints, deformations and colour, as the latter evolves through the procedure. Inter-patient diversity is simply covered by including videos from different patients. In addition, we used videos from three types of procedures: hysterectomy, laparoscopic fertility exploration and endometriosis surgeries containing images of both normal and pathological cases. We used 79 videos from which we extracted our dataset of 3800 images. Technically, the videos were visualised with the multimedia player VLC, which allowed us to extract images at a regular time interval. The final images were then manually selected in order to fulfil the above diversity criteria. Manual selection by an expert is very important: it favours quality and diversity, whereas an automatic selection, for instance directly using the images extracted at a regular time interval, would focus on quantity only.

2.3.3 Annotation, Tools and Labellers

In our dataset, the annotations were designed to resolve uterus segmentation and uterus contouring. The uterus contours identify the relationships between the uterus and its neighbouring organs in terms of visual occlusions. Specifically, each of the pixels forming the uterus contours can be of one of three types: the occluding contour type, where the uterus ends by occluding another organ, the occlusion contour type, where the uterus is occluded by another organ or the image boundaries, and the connection contour type, where the uterus is connected to another organ. We have also labelled the junctions between the uterus and fallopian tubes specifically, to allow further usage. Overall, we thus specifically have the following annotations for each image of our dataset: the uterus segmentation, the uterus occluding contours, the uterus occlusion contours, the uterus connection contour, the region of the right fallopian tube junction and the region of the left fallopian tube junction. These annotations can be combined and arranged to create new labels. For instance, the fallopian tube junction region can be used together with the connection contour mask to create the uterus-fallopian tube junction mask. The selected images were transferred to the online annotation software Supervise.ly (8). They were then annotated by two expert gynaecological surgeons (SMZ and NB). Figure 4 shows an example of an annotated image under Supervise.ly.

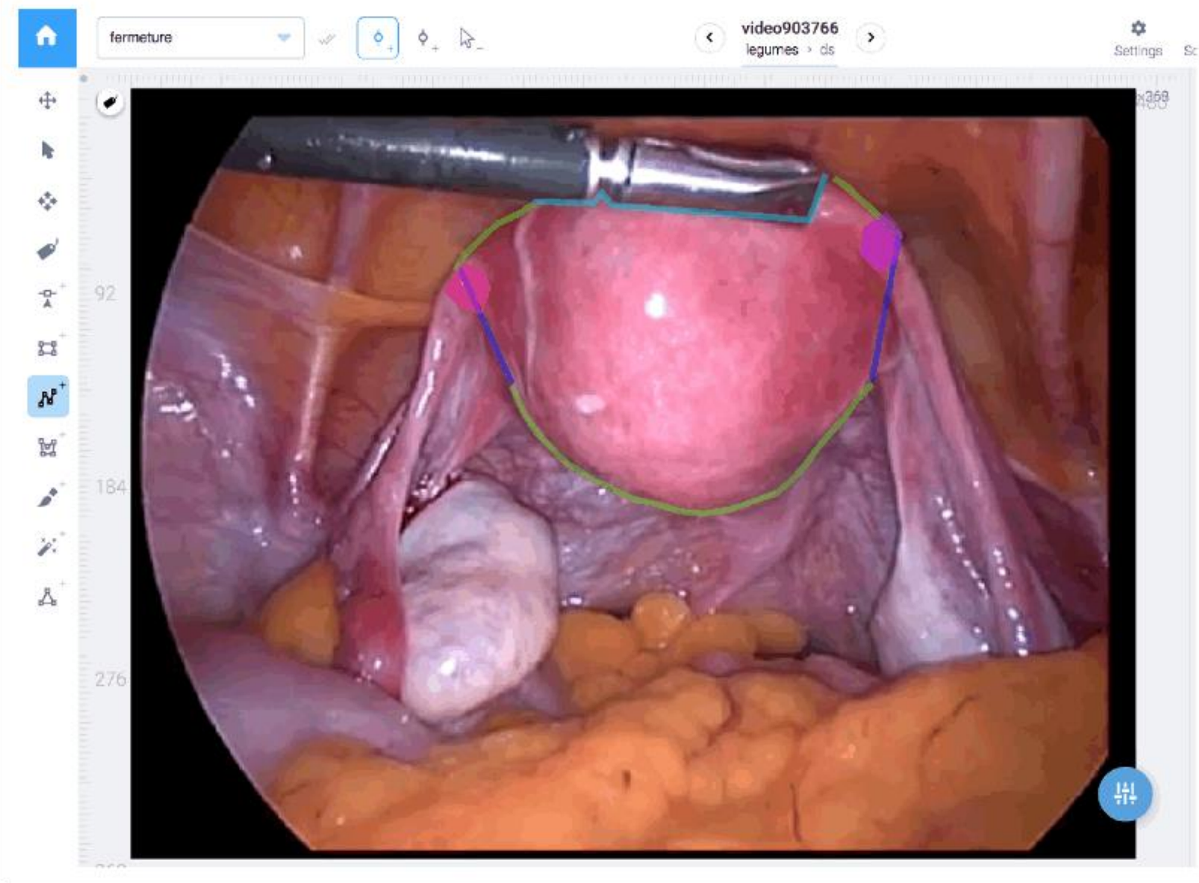

Figure 4: Example of a laparoscopic image annotated with the online annotation software, Supervise.ly. In green, the occluding contours, in light blue the occlusion contours, in dark blue the connection contours and in purple and pink the right and left uterus-fallopian tube junctions.

3. Results

<u>3.1 Neural Network Implementation and Evaluation</u>

Our engineers (TF and AB) implemented the neural network using the programming language Python with *Facebook's PyTorch software toolbox* running on a standard desktop PC computer. They evaluated the neural network with so-called evaluation metrics, quantifying the discrepancy between the predicted and the expert annotations, for the images from the test dataset. Recall that the test dataset contains patient data which were not used to train the neural network. It thus allows one to perform an independent evaluation. We now describe the evaluation metrics.

Uterus segmentation. Segmentation is a well-studied task for which there exist simple and commonly accepted evaluation metrics. Specifically, we use the Intersection over Union (IoU), which vastly dominates the evaluation of segmentation in the literature. As illustrated in figure 5, the IoU represents the percentage of overlap between the expert segmentation and the neural network predicted segmentation. We use the average of the IoU over all the images from the test dataset. The IoU is a measure of agreement between the experts and the neural network, thus the higher, the better. The IoU ranges from 100%, which is a perfect result, to 0%, which is a very bad result.

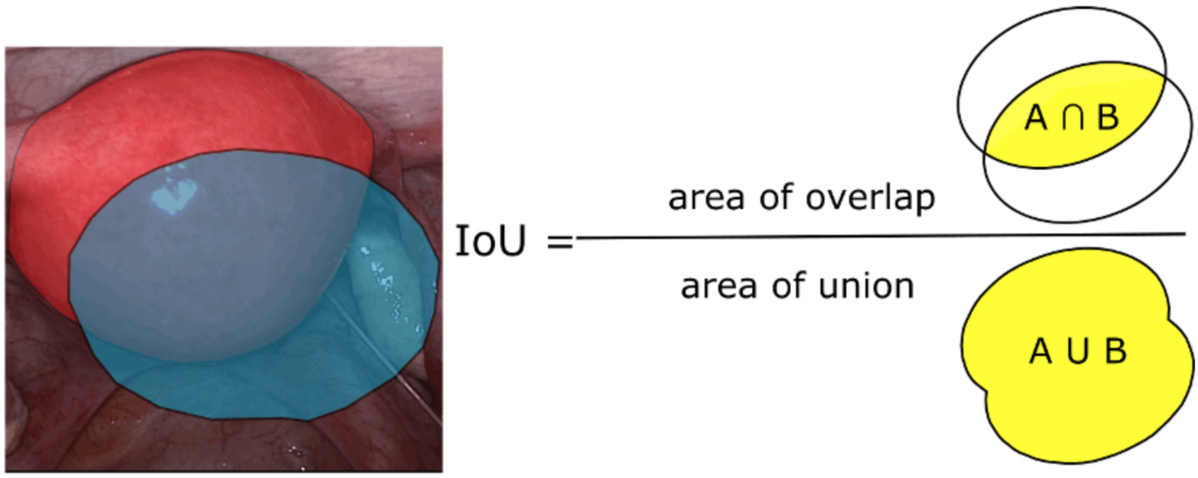

Figure 5: Intersection over Union explanatory diagram. The IoU between segmentations A and B is defined as the ratio between the area of the intersection and the area of the union of A and B. In the example image, A in red is the expert segmentation and B in blue is the neural network predicted segmentation, leading to an IoU of 39%. The IoU is usually expressed as a percentage and varies between 100% for a perfect segmentation and 0% for an extremely poor segmentation.

Uterus contouring. In contrast to segmentation, the evaluation metric for contour detection is challenging to design and has not been standardised yet. This is because a contour is a precisely localised thin image part. A prediction is thus rarely perfect, in the sense that it never perfectly reproduces the annotation. Consequently, even if a prediction lies close to the annotation and is thus acceptable, it will in most cases have a very low IoU. Our engineers proposed the *contour error*, which addresses the problem using a tolerance distance between the contour points, as explained in our previous paper (5). The final contour error is a measure of discrepancy between the experts and the neural network, thus the lower, the better. The contour error ranges from 0%, which is a perfect result, to 100%, which is a very bad result.

3.2 Training Results and Dataset Size

The dataset consists of 3800 annotated images, whose characteristics are given in table 1. The dataset was split into a training set of 3234 images and a test set of 581 images. Figure 6 shows the uterus segmentation performance. We observed a steep improvement from 100 to 700 training images and a much slower improvement beyond. The final IoU on the training and test datasets were 94.6% and 84.9% (the higher, the better). Similarly, for uterus contouring, we observed a steep improvement from 100 to about 2000 training images and a much slower improvement beyond. The final contour error on the training and test datasets were 19.5% and 47.3% (the lower, the better).

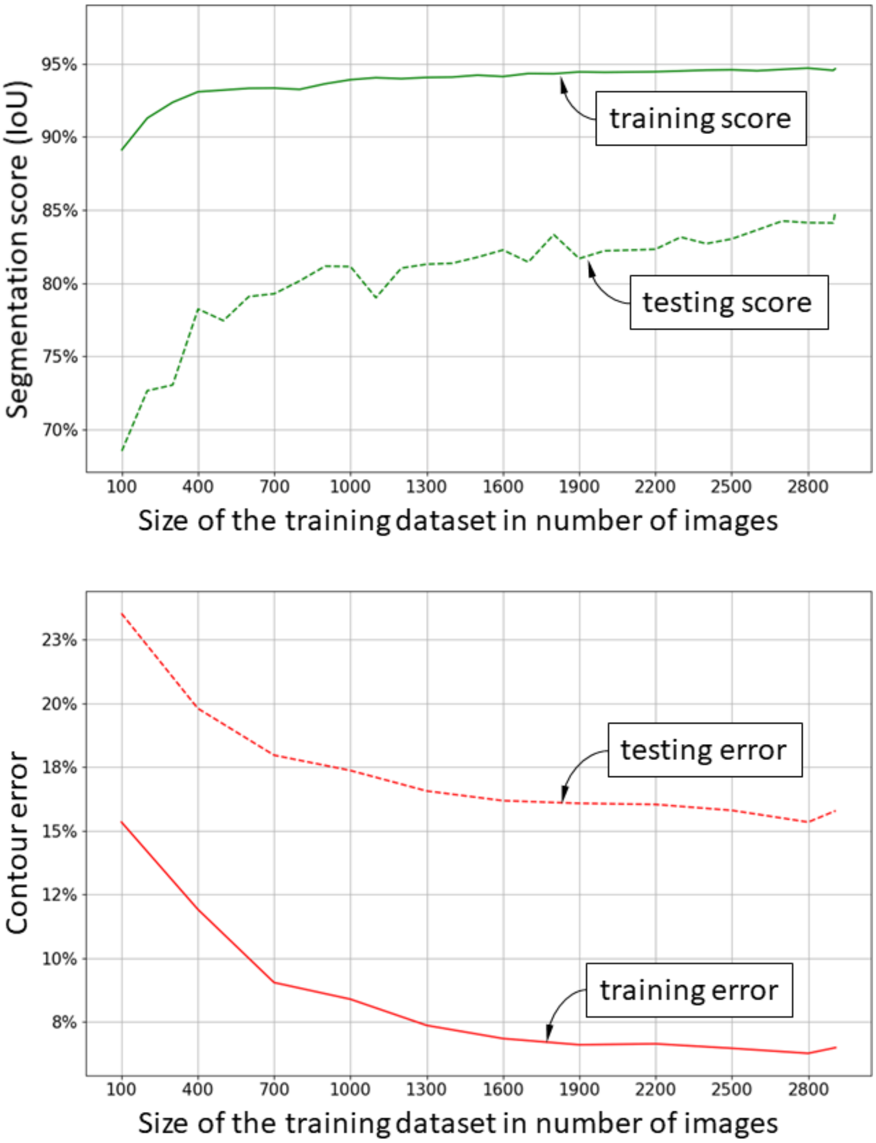

Figure 6: Size-performance graphs for uterus segmentation and uterus contouring. The curves show the performance as the training set is increased by adding batches of 100 images. Recall the higher, the better for the segmentation score (the IoU) and the lower, the better for the contour error.

The dataset	
Data	images
Number of patients	79
Number of videos	79
Laparoscopic gynaecological procedures	Hysterectomies (48) Rectovaginal endometriosis nodule (21) Laparoscopic fertility exploration (10)
Classes	Uterus segmentations (3800) Occluding contours (3809) Occlusion contours (2629) Closure contours (3711) Left uterus-fallopian tube junction (1364) Right uterus-fallopian tube junction (1368)

Table 1. Characteristics of the proposed SurgAI3.8K dataset.

Figure 7 illustrates the uterus segmentation results for the cases of 100, 1300 and 2900 training images. Visual inspection confirms that, as suggested by the IoU values, while a strong agreement holds between the second and third cases, a substantial difference can be seen between them and the first case.

Figure 7: Uterus segmentation results. Training was performed with (a) 100, (b) 1300 and (c) 2900 images. True Positives are in green, i.e. pixels labelled as ‘uterus’ by the expert and predicted as ‘uterus’ by the neural network, False Positives are in red, i.e. pixels not labelled as ‘uterus’ by the expert and predicted as ‘uterus’ by the neural network and False Negatives are in blue, i.e. pixels labelled as ‘uterus’ by the expert and not predicted as ‘uterus’ by the neural network. The IoU is given in the bottom right corner.

Figure 8 illustrates the occluding contours of the uterus for the cases of 100, 1300 and 2900 training images. Visual inspection confirms that, as suggested by the contour error, while a strong agreement holds between the second and third cases, a substantial difference can be seen between them and the first case.

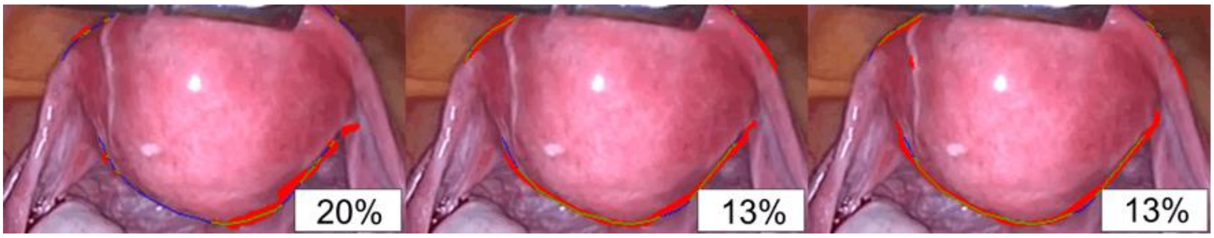

Figure 8: Uterus contouring results. The images specifically show the occluding contours of the uterus. Training was performed with (a) 100, (b) 1300 and (c) 2900 images. True Positives are in green, i.e. pixels labelled as 'occluding contour' by the expert and predicted as 'occluding contour' by the neural network, False Positives are in red, i.e. pixels not labelled as 'occluding contour' by the expert and predicted as 'occluding contour' by the neural network and False Negatives are in blue, i.e. pixels labelled as 'occluding contour' by the expert and not predicted as 'occluding contour' by the neural network. The contour error is given in the bottom right corner.

4. Discussion

4.1 Summary of Contributions

Introducing the method of AI. This article contributed a pedagogical introduction of the fundamental notions of AI, through the use of a concrete application of image processing towards automating a surgery guidance application, as described below. It introduced the notions of neural networks, their design, their training from the training dataset and their performance evaluation from the test dataset. It then shows that the key factor is the availability of a dataset with expert annotations. This introduction is intended to give surgeons an understanding of how AI works and the ability to use it wisely. For instance, it is clear that if a neural network was not trained on a sufficiently large dataset, its performance may be poor in some cases. Requesting the amount of cases a neural network was trained from and what the test error was may thus become a natural question for the surgeons to ask before adopting an AI-based technology.

The dataset. This work contributed to the construction of a dataset of annotated laparoscopic images for the uterus segmentation, uterus contouring and fallopian tube junctions detection tasks. Compared to the existing datasets for the automatic recognition of common, non-medical objects, it is modest in size. Nonetheless, it has variability, obtained by including multiple patients, different times of surgery, different view angles, and has quality manual annotations. Several datasets for the automatic detection of surgical tools exist in the literature, but very few works deal with the automatic detection of anatomical structures. Prior to this work, we carried out a feasibility study on a reduced dataset and a simple task of detecting pelvic organs (uterus, ovary, surgical tools) (9). This prior work showed that AI techniques could feasibly solve this type of task. Leibetseder et al. have carried out several studies in laparoscopic surgery, but their work focuses on the classification of surgical images (10), merely indicating the presence or absence of an organ in an image. Recently, they have published work aiming at endometriosis detection in laparoscopic surgery images (11). They have published a dataset of 25K images with half of the images with endometriosis lesions and the other half without. Only 300 images are manually annotated with specific endometriosis lesion contours, which are known to require a high degree of expertise in laparoscopic anatomy. More recently, the same team has published endometriosis lesion detection using Mask-RCNN (12), which was a significant step forward. Concretely, the results are bounding boxes containing the lesions. In contrast, our results provide the detailed contour of the anatomical structures, representing a much richer piece of information than a bounding box. In addition, several studies have been carried out on the automatic detection of surgical tools (13–15). Surgical tools differ significantly from the abdominal anatomy and do not necessarily require expert annotation. There is no large-scale dataset for the automatic detection of gynaecological organs in laparoscopic surgery to date. Nevertheless, the proposed dataset could be extended by labelling the other anatomical structures visible in the image, which may potentially improve the performance of AI. We leave this extension for future work.

Dataset size. An important contribution of this work is to analyse the size of the laparoscopic image dataset required for the automatic segmentation and contouring of the uterus in a laparoscopic image. It contributes to answering an essential question regarding the use of AI techniques, which regards the number of training images necessary and sufficient to achieve the desired performance. This question has not been resolved to date in the literature. We bring an answer regarding uterus segmentation and contouring by means of observing the size-performance graphs. We may hypothesise that for other organs presenting the same type of inter-patient variability on a laparoscopic image, it may be possible to consider the creation of a dataset with the same order of magnitude in size to obtain similar results.

4.2 Contribution of the Neural Network to Augmented Reality

The objective of solving uterus segmentation and uterus contouring was to replace the manual annotation required at the initialisation phase of Uteraug, the AR based laparoscopic surgery assistance software developed by our team in prior work. Steps 2 and 4 in the initialisation phase include manual interactions: the selection of the region formed by the image of the uterus and the selection of the different contours of the uterus. Thanks to our dataset and trained neural network, we have automatised these manual tasks. We have integrated our neural network to Uteraug and evaluated the quality of the manual and automatic solutions by comparing the surgeons' annotations with the neural network. We have also measured the time saved by the automatic annotation during the initialisation phase. These results have been published by our team in (6). The automatic annotation achieves almost identical AR results to manual annotation in terms of quality. The time is reduced by 3 minutes and 56 seconds compared to manual annotation, on average, which represents a 97.4% reduction, increasing the software usability and presumably its acceptability. The manual interactions required at the initialisation phase of Uteraug would harm its clinical use in the operating room without trained staff. By automating this phase, AI offers the possibility to extend its use to a larger number of surgeons.

5. Conclusion

We have described a concrete example of how AI can be used in surgery, which conveys the basic concepts of AI in a pedagogical way, with illustrations given on the concrete case of augmented reality in laparoscopy.

Disclosures: Drs Sabrina Madad Zadeh, Tom Francois, Aurélie Comptour, Michel Canis, Nicolas Bourdel and Adrien Bartoli report no conflict of interest.

Source of funding: no funding

References:

1. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. ArXiv13126229 Cs [Internet]. 23 févr 2014 [cité 20 sept 2020]; Disponible sur: http://arxiv.org/abs/1312.6229

2. Collins T, Pizarro D, Gasparini S, Bourdel N, Chauvet P, Canis M, et al. Augmented Reality Guided Laparoscopic Surgery of the Uterus. IEEE Trans Med Imaging. janv 2021;40(1):371-80.

3. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA, USA: MIT Press; 2016. 800 p. (Bach F. Adaptive Computation and Machine Learning series).

4. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, éditeurs. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer International Publishing; 2015. p. 234-41. (Lecture Notes in Computer Science).

5. François T, Calvet L, Madad Zadeh S, Saboul D, Gasparini S, Samarakoon P, et al. Detecting the occluding contours of the uterus to automatise augmented laparoscopy: score, loss, dataset, evaluation and user study. Int J Comput Assist Radiol Surg. juill 2020;15(7):1177-86.

6. Bar-Shavit Y, Jaillet L, Chauvet P, Canis M, Bourdel N. Use of indocyanine green in endometriosis surgery. Fertil Steril. 1 juin 2018;109(6):1136-7.

7. Bourdel N, Jaillet L, Bar-Shavit Y, Comptour A, Pereira B, Canis M, et al. Indocyanine green in deep infiltrating endometriosis: a preliminary feasibility study to examine vascularization after rectal shaving. Fertil Steril. août 2020;114(2):367-73.

8. Supervisely - Web platform for computer vision. Annotation, training and deploy [Internet]. 2020 [cité 20 sept 2020]. Disponible sur: https://supervise.ly/

9. Madad Zadeh S, Francois T, Calvet L, Chauvet P, Canis M, Bartoli A, et al. SurgAI: deep learning for computerized laparoscopic image understanding in gynaecology. Surg Endosc. 29 janv 2020;

10. Lapgyn4 | Proceedings of the 9th ACM Multimedia Systems Conference [Internet]. 2020 [cité 20 sept 2020]. Disponible sur: https://dl.acm.org/doi/abs/10.1145/3204949.3208127

11. Leibetseder A, Kletz S, Schoeffmann K, Keckstein S, Keckstein J. GLENDA: Gynecologic Laparoscopy Endometriosis Dataset. In: Ro YM, Cheng WH, Kim J, Chu WT, Cui P, Choi JW, et al., éditeurs. MultiMedia Modeling. Cham: Springer International Publishing; 2020. p. 439-50. (Lecture Notes in Computer Science)

12. He K, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In 2017 [cité 8 mars 2022]. p. 2961-9. Disponible sur: https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html

13. Zhang B, Wang S, Dong L, Chen P. Surgical Tools Detection Based on Modulated Anchoring Network in Laparoscopic Videos. IEEE Access. 2020;8:23748-58.

14. EndoVisSub-Instrument - Grand Challenge [Internet]. grand-challenge.org. 2020 [cité 20 sept 2020]. Disponible sur: https://endovissub-instrument.grand-challenge.org/

15. García-Peraza-Herrera LC, Li W, Fidon L, Gruijthuijsen C, Devreker A, Attilakos G, et al. ToolNet: Holistically-nested real-time segmentation of robotic surgical tools. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2017. p. 5717-22.