UNIVERSITY OF TWENTE.

Faculty of Science and Technology

Semantic segmentation of minimally invasive anti-reflux surgery video using U-NET Machine Learning

Julian R. Abbing, BSc

M.Sc. Thesis in Technical Medicine

University of Twente
Department of Surgery
Meander Medical Center
28th May 2020

Semantic segementation of minimally invasis anti-reflux surgery video using U-NET Machine Learning

J.R. Abbing, B.Sc.

This thesis is submitted to the University of Twente for the Master of Science degree in Technical Medicine

UNIVERSITY OF TWENTE.

meander MEDISCH CENTRUM

Graduation committee:
prof. dr. I.A.M.J.Broeders (Chairman & Medical supervisor)
dr. ir. F. van der Heijden (Technical supervisor)
drs. P.A.van Katwijk (Process supervisor)
dr. ir. J.F. Broenink (External Member)

University of Twente
Faculty of Science and Technology
Department of Surgery,
Meander Medical Center
Maatweg 3
3813 TZ Amersfoort

Preface

Dingen die men ziet, lijkt je eerder te begrijpen. Die gedachtegang lijkt in de maatschappij ook steeds gaande te zijn. Zo bestaat handleiding van Ikea voornamelijk uit afbeeldingen. Daarnaast kan een ziekenhuis niet zonder een radiologie afdeling. Tevens is er op wetenschappelijk gebied, zeker het kwantificeren van dat was je ziet steeds belangrijker. Deze gedachtegang is niet alleen zichtbaar in de maatschappij maar ook persoonlijk door te kiezen voor de Master track keuze (Medical Imaging and Interventions).

En de Master (track) bleek ook volledig in lijn met de eindopdracht van de middelbare school; *of men onder water met glazen lenzen beter zou kunnen zien, dan met een duikbril*. Eveneens een zicht/visie/beeldvormend en technisch onderwerp. Een onderwerp dat op al op technisch geneeskundige onderwerpen de mogelijk en onmogelijkheden is had belicht.

Een opleiding maar ook een studentenleven dat veel stageplekken en vele uitstapjes kende; Een Enschedese studententijd, de URaad, stages in Hardenberg, Zwolle, Utrecht, Leiden/Amsterdam, Straatsburg. Plus het wonen in deze verschillende steden. Niet te vergeten ook de studiereis met het backpacken door China.

Nu ruim 7 jaar verder heb ik me dan ook bezig kunnen houden met een AI binnen de chirurgie dat ongetwijfeld hetzelfde pad zal volgen als andere sectoren zoals de automotive of radiologie. Een werkgebied dat sterk moet vertrouwen op het zicht. Dit is daarnaast een onderzoeksgebied waar ikzelf me in de 6 jaar Technische Geneeskunde nog niet in had verdiept maar zeker van grote betekenis zal zijn in de medisch sector.

In het bijzonder wil ik daarvoor een aantal mensen bedanken. Deze stage, zou nooit mogelijk zijn geweest door de visie en gedrevenheid van prof. dr. I.A.M.J. Broeders en het Meander Medisch Centrum. Zonder de daarbij behorende Technische begeleiding en ideeën en blik n imaging onderwijs van dr. ir. Ferdi van der Heijden zou eveneens deze thesis niet compleet zijn geweest. De afgelopen drie jaar heeft Paul van Katwijk een belangrijke rol in professionele ontwikkeling gehad van vele TG'ers, waaronder ikzelf. We hebben altijd delen van elkaars ideeën over de opleiding, stage(s) maar ook privézaken toch kunnen ondersteunen.

Er zijn daarnaast een boel anderen (TG'ers, coassistenten, arts-assistenten, ect)

die er zeker aan hebben bijgedragen maar de twee mensen die ik absoluut niet mag overslaan zijn mijn ouders. Zij die alles van visie op studie/studenten zaken, verhuizen, halen/brengen naar het treinstation tot het lezen van de talloze concept verslagen altijd hebben ondersteund.

Ontzettend bedankt en veel plezier met het lezen van deze master thesis!

Julian

Summary

English

Introduction During anti-reflux surgery, there is a potential risk of (unintended) Nervus Vagus injury. Which estimated around 20%. A solution and our goal is to create an AI tool (Deep Learning) that can detect the Nervus Vagus and other anatomical structures in surgical videos. Addition of temporal features that might help in segmentation/detecting the actual nerve.
Method Five UNET algorithm structures are used as a basis for the training of 5 visible structures in the videoframes. These networks are trained on two datasets; a small clinical (105 frames, from 10 videos) for the actual goal and a larger automotive dataset (2121 frames, from 5 videos) for testing the functionality of the networks and testing the addition of dense optical flow (temporal features). The *Dense Optical Flow* is calculated and used as an input or extra input for the algorithms. The clinical dataset is, just like the automotive dataset, pixel-wise annotated for the target structures (liver, crus, Vagus Nerve, stomach/oesophagus, else). The target structures for the automotive dataset are; car, road, traffic signs, sidewalk, else. The five algorithms differ from each other on the input data; the red-green-blue (RGB) frame as only input, the *Dense Optical Flow* and a combinitation of both.
Results The UNETS are able to segment in all cases at least two structures in the automotive with an IoU 0.5 or higher. And even three structures in the RGB only algorithm. The RGB only algorithm is performing better in both datasets and has for each segmentable structure the highest IoU score compared to the other algorithms that also used dense optical flow. IoU scores for the clinical dataset are much lower but show a similar pattern. Only the ”else” structure reaches an IoU above 0.5 in the clinical dataset. The confusion matrix shows similar findings, and in none of the networks, the vagus nerve popped out (all algorithms score a 0.00 in the normalised confusion matrices for the Vagus Nerve). Visual inspection of the heatmap/probability maps of the Vagus Nerve of the RGB algorithm shows a relatively broad region where the vagus nerve indeed can be.
Conclusion Visualisation of semantic segmentation was possible on top of surgical

video frames. Semantic segmentation with UNETs trained on surgical images is possible. The addition of temporal features (Dense optical flow) of videos by combining the RGB data in the first layer of the UNET algorithm does not improve the semantic segmentation.

Contents

List of acronyms

AI	Artificial Intelligence
DL	Deep Learning
GERD	Gastroesophageal reflux disease
PPI's	proton-pump inhibitors
HHD	hiatal hernia diaphragmaticus
LES	lower oesophagal sphincter
DOP	Dense Optical Flow
VKITTI	Virtual KITTI
IOU	Intersection over Union
U-NET	U-(shaped)Network
RGB	RGB = Red, green and blue. A 3 channel input containing red, green and blue values. Those are commonly used for videorepresentation or pictures.
FCN	Fully Convolutional Network(s)
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
PNG	Portable Network Graphics, image data format

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Gastroesophageal reflux disease (GERD) is considered a benign condition of the stomach and oesophagus. [1] The primary medical treatment of GERD is the use of proton-pump inhibitors (PPI's), however, a 10 to 40% of the patients remain unresponsive [17]. Surgical treatment is a second treatment option. When PPI treatment does not show results in proven GERD, the recommended treatment is a fundoplication. Even if no HHD is present, but PPI treatment does not work, a fundoplication is recommended. [18] (Examples of a HHD are given in figure 1.1)

However, there is a potential risk of Vagus Nerve (Latin: *Nervus Vagus)* injury in fundoplications with HDD repair. [19] More important, Vagus Nerve injury has a significant negative effect on the reflux control postoperative and a significantly higher redo rate compared when there is no vagus injury post surgery [20]. Research of Van Rijn et al. (2016) ([20]) reported an incidence of 20% on unintended vagus injury. It should be mentioned that this long-term follow-up data of surgeries were collected between 1990 and 2000. Back then, the laparoscopic video systems were not as good as today. In this cohort of vagus injury (the study of Van Rijn et al. (2016)) over 50% had redo surgery and most of them because of recurrent reflux problems. Better knowledge per patient of the location during surgery or visualisation of this nerve might improve the outcomes.

Due to a dysfunctional closure of the lower oesophagal sphincter duodenal gastric material can enter the oesophagus and even higher anatomical structures. This reflux can cause apart from discomfort, damage and inflammation of those structures. Untreated, the inflammation and tissue changes can lead to aspiration, Barrett's oesophagus, stricture, esophagitis or an adenocarcinoma. A *higher* incidence in GERD is found in patients who have an HHD, obesity or delayed gastric emptying. [21]

Due to change of anatomy, an HHD reduces functionality of the lower oesopha-

gal sphincter (LES), which results in possible entering of stomach fluids into the oesophagus. An HHD is a protrusion of anatomical structures (other than the oesophagus) into the thoracic cavity due to a widened hiatus diaphragmaticus. [22] This causes the symptoms; pain, heartburn, bleeding, dysphagia, weight loss, vomiting and regurgitation. [23]

Those GERD-like symptoms are strongly related with the HHD but are not necessarily present with every HHD. With a hiatal hernia, the stomach can migrate partially or entirely to the thoracal cavity. An HHD has four different subtypes anatomically (see figure 1.1). The most common one is type 1 and does not imply a non-functional LES. Though non-functionality of the LES is also very size-dependent. A type 2 to 4 is likely to cause GERD symptoms. In type 2, the gastroesophageal junction is in the abdominal cavity although the gastric fundus slides into the hiatal hernia. In a type 3 HHD, the fundus of the stomach and the gastroesophageal junction are located in the thorax cavity instead of in the abdominal cavity. A type 4 (not visualised in figure 1.1) other anatomical structures migrate cranially to the hiatal hernia. [1]

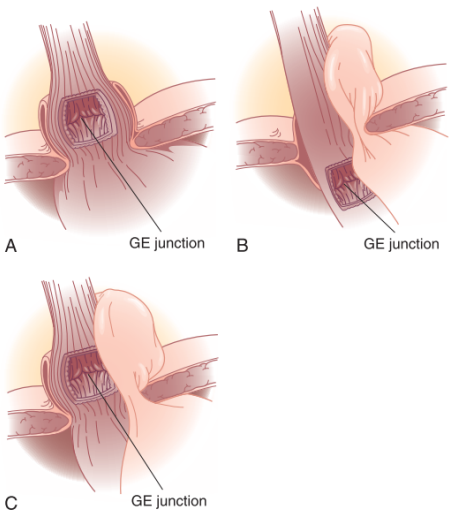


Figure 1.1: Type 1, 2 and 3 of hiatal herniations (HHD). **A** is a type 1 hernia (sliding hernia). **B** is a type 2 hernia (rolling hernia). **C** is a combination type of the type 1 and 2 (mixed hernia). [1]

The most common hiatal hernia is type 1; it covers 95% of all hiatal hernias. The other three subtypes together make the other 5%. The common symptoms of type 1 are the presence of GERD/reflux. [24] The other subtypes present themselves, on the other hand, more frequently with obstructive symptoms. [1] Also in type 1 hiatal hernia without reflux disease is also considered as no indication for surgery [22].

Artificial Intelligence for healthcare applications

In healthcare, the thrive to improve patient outcomes without raising the cost has always been the case. A new step in the digitisation in health care might support this need through big data and similar technologies like artificial intelligence (AI). AI tends to improve on diagnostics, patient therapy, prevention and support health care in making clinical decisions. A subtype of AI is machine learning. It can find correlations, associations, segmentation and generate new insights in vast amounts of data. AI is used in the automotive, finance and smart homes. In medicine, the first clinical setups show their great value; node detection in X-ray images, the prediction of outcomes in infectious diseases and ECG arrhythmia detection. Deep learning is also a type of AI and machine learning but relies on a small infrastructure which mimics the brain infrastructure. It is called deep because of stacked layers with multiple artificial 'neurons' that can be trained with existing data to make predictions or classifications. This is achieved by *learning* based on prelabelled data [25]

Due to the enormous variation between observed data (patients), the other regular learning methods are not sufficient anymore (i.e. selection on only colour differences). So the step from machine learning to deep learning is established. Deep learning can make automated predictions on large complex datasets. [26]

Also, in surgery, the first AI applications are built. An example is a previous work from M. Schuhmacher at the Meander Medical Center. He showed that surgical video could be used for object detection. He showed that with an object detection algorithm, trained on a clinical in hospital made dataset, is possible. The used data was colorectal surgery video frames. The accuracy was low and acceptable (43.7%), but improvement for clinical usage would be necessary as also stated in his thesis. He used a CNN (convolutional neural network, the YOLOv2 network) for autonomous structure recognition in the lower abdomen. Suggestions he made were; more training data, more complex network architectures like long short-term memory networks (LSTM). An LSTM is a recurrent neural network (RNN) which adds a specific 'memory' to the algorithm. In contrast, CNN does not have this 'memory'. [4] In this proposed study is to test new network structures on a clinical visualisation problem during anti-reflux surgery.

In 2010 by Zanjani et Al. described that the AI segementation algorithms used on videos still were based on the single frames and not on the video itself. Possible clues that could be hidden in previous frames was discarded. This is also the case in the YoloV2 algorithm that was used by M. Schuhmacher. That only used single frames. [27] This problem of not using the video but single frames, in combination with the recommendations of M. Schuhmacher, raises the question if a possible mechanical fingerprint or clue (in previous frames), can be used to improve the seg-

mentation of anatomical structures. In our case, the anatomical structures like a Vagus Nerve.

Although the first, AI algorithms are introduced in medicine, none of them are used in daily practise in interventional care like surgery. The Meander Medical Center tries to fill this gab. This is achieved by the research line AI and Surgery, of which the research of M. Schuhmacher as described was the first result. With that mindset, this research continues in the field of surgery and the use of AI. [4], [28]

1.2 Research aim and research hypothesis

The ultimate aim of this study is to develop a surgical tool that is able to assist during surgery in real-time recognition of the Vagus Nerve to help training residents and support new surgeons. In this research, multiple steps are considered to achieve a supportive algorithm for anatomical structure recognition. First, a laparoscopic fundoplication dataset has to be created that can be used for training and building these algorithms. Next, this data is used to create an algorithm using AI. Lastly, an application has to be developed, which visualizes the data real-time to the surgeon. To achieve these steps, the following research hypotheses and one sub question have to be answered:

- The use of Deep Learning network can be used for anatomical structure recognition/segmentation, such as the Vagus Nerve, on surgical video.
- The addition of movement (temporal information) of the surgical video does improve the segmentation of anatomical structures by a Deep Learning network.
- How can visual output from an algorithm be presented to the surgeon?

1.3 Outline

In the first chapter, the aim, clinical relevance and the link between the clinical problem and the technical approach are explained. In the second chapter, more background is given on the clinical aspects of the problem. A technical introduction with a basis of Artificial Intelligence (AI), Deep Learning (DL), and how movement properties can be acquired in video by DOP is explained in the third chapter. Also, the last research question about the representation of output representation to the surgeon is addressed in this chapter. The fourth chapter will combine the clinical background

and the technical background into a set of tests, how the output performance is measured and the way the outcomes are presented. The results are shown and explained in the fifth chapter and discussed in the following chapter. Finally, the conclusions are drawn. The last chapter also contains a set of specific recommendation for future work. In the appendices, the study protocol, metadata about the networks and raw output figures are included.

Chapter 2

Clinical background

In this chapter, the surgical aspects of the anti-reflux surgery the role and anatomy of the Vagus Nerve are discussed in-depth. Furthermore, the steps of the surgical procedure, including the aim of the surgical procedure, is made clear. At the end of this chapter, requirements are set for a possible algorithm to be clinically applicable for the surgeon.

2.1 Vagus Nerve

The name vagus in Latin means *"wandering"* and is a direct link to the behaviour of the path it follows through the human body. It has been studied since ancient times but at a stomach/gastric level. E. O. Schumov-Simanovskaja and the famous Nobel Laureate Pavlov (Ivan Petrovich Pavlov) studied gastric secretion by the vagus nerve in dogs. He was also the first to prove that the vagus nerve pas partially responsible for gastric acid secretion. The response was that the vagotomy (cutting off one ore more branches of the vagus) turned into a treatment for ulcer disease. [29] Around the hiatus diaphragmaticus (the opening in the diaphragm where the oesophagus passes through) as explained in chapter 1 are important structures such as the lungs, heart but also the vagus nerve trunks. The vagus nerve is a cerebral nerve (CN X, originates directly from the brain at both sides) and bundles into two nerve trunks. The anterior vagal trunk is derived in most cases from the left vagus nerve and enters the abdomen through the diaphragm on the anterior side of the oesophagus. The path of this anterior vagus nerve is via the lesser curvature of the stomach, and two branches leave to the hepatic-duodenal ligament (the hepatic and duodenal branch). The trunk continues via the lesser curvature to the anterior gastric branches and supplies the pyloric branch. On the posterior side of the oesophagus at the level of the diaphragm enters the posterior vagal trunk. This nerve bundle is mainly derived from the right vagus nerve (CN X). The posterior

vagus nerve also runs to the lesser curvature but has branches to the posterior and anterior surface of the stomach. Also, it continues the lesser curvature and gives off a celiac branch to the celiac plexus. From there, it continues to the lower abdominal organs. [30], [31]

The full path visualized in figure 2.1 and intraoperative human figures of the Vagus Nerve in figures 2.3 and 2.4.

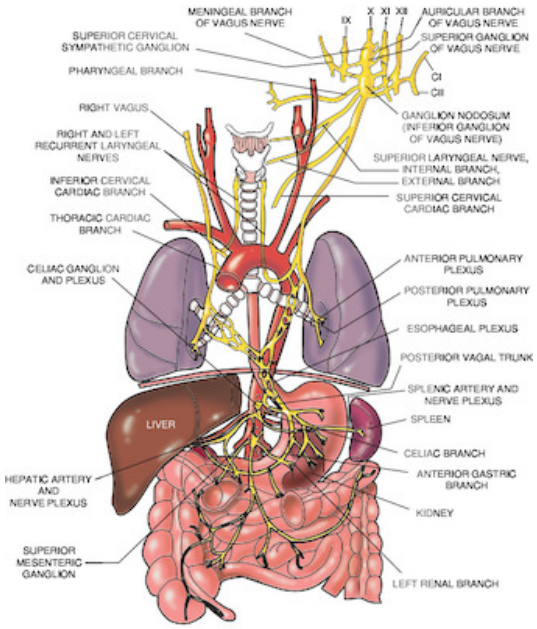


Figure 2.1: Vagus nerve (CN X): schema/pathway through the body [2]

Functionality of the nerve at the level of the stomach

The nerve has a sensory function for the gastrointestinal tract, but also the heart and the lower respiratory tract. Although the last ones are not specifically relevant in this study due to the level of possible injury. The sensory functionality starts from the anatomical track of the nerve and ends at the distal part of the colon [29], [31], [32]Also, it includes parasympathetic functionality of the respiratory tract (smooth muscles of the bronchi), the heart and the intestine/stomach. Note, 90% of the Vagus Nerve fibers of the bowel are afferent. The 10% left is efferent to give the parasympathetic signals to the abdominal viscera (liver kidneys, spleen, splenic flexure) as till the last third part of the colon. The rest of the colon receives parasympathetic signals from pelvic nerves. The vagus nerve also innervates the circular muscle at the end of the stomach (at the antrum, the pylorus) for gastric drainage [29] This means those structures are critical to for proper gastric motility, emptying and gastric secretion if the sensory function and parasympathetic function is disturbed. This clarifies the higher redo (reoperation) rate as stated in chapter 1 and also more evident if this is disturbed around the stomach. [20]

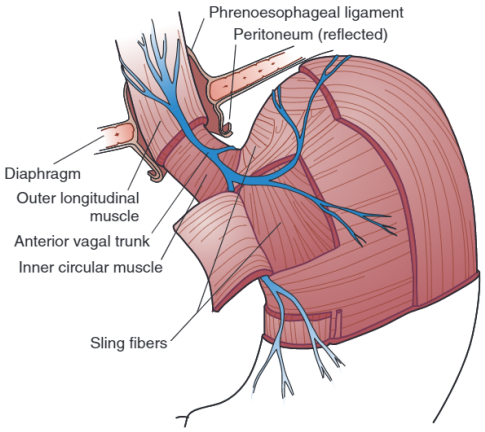


Figure 2.2: A drawing of the stomach with the anterior vagal trunk (*Nervus Vagus/Vagus Nerve*) . [1]

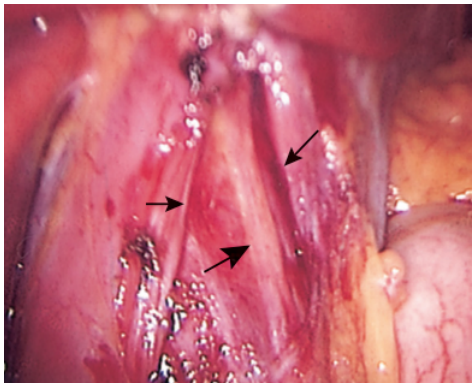


Figure 2.3: The big arrow points at the anterior Vagus Nerve surrounded by the crura left and right (smaller arrows). [3]

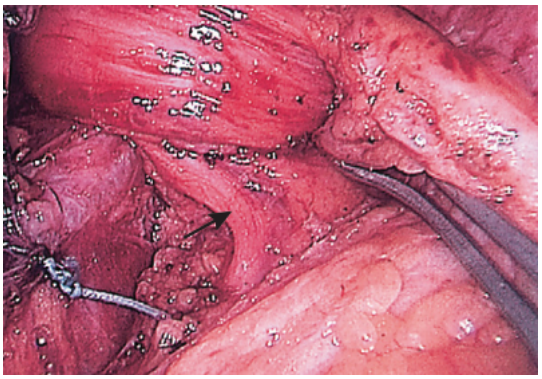


Figure 2.4: Here the arrow points at the posterior trunk of the Nervus Vagus/Vagus Nerve, with a clearly visible esophagus on top op it. [3]

2.2 Surgical procedure

In chapter 1 is written that a surgical procedure is only performed as a second treatment option. The anti-reflux surgery is also called a fundoplication due to the upper part of the stomach (fundus) that is used to compromise the functionality of the valve (LES, Lower oesophagal sphincter). This sphincter functionality is often too little, and otherwise, the GERD remains intact. The fundoplication in figure 2.5 is a visual representation of how the fundus (upper part of the stomach) is folded around the gastroesophageal junction and sutured to the surrounding structure. This creates the new supportive valve at the gastroesophageal junction. [1]

There are multiple slightly different anti-reflux surgeries available. But overall the same strategy is used: bring the stomach back in its original position beneath the diaphragm, make the hiatus (opening) in the diaphragm smaller and finally create a fundoplication to create a new supportive gastro-oesophagal valve to prevent gastric

fluids from flowing back upward into the oesophagus or even mouth. The last step is crucial to prevent this due to the widespread dysfunctional closure of this valve (the lower oesophagal sphincter at the gastroesophageal junction). [1], [21]

There are three types of fundoplications, visualized in figure 2.5: 180 degrees, 270 degrees and the 360 degrees fundoplications. [1] At the Meander Medical Centre (MMC) the number of performed full 360 degrees (Nissen) fundoplications is minimal due to more favourable outcomes of the partial fundoplications. [33]

The main surgical steps are:

1. Trocar placement and insufflation of the abdominal cavity [1], [34]
2. Left crural dissection after the pars flaccida is dissected. [21], [34]
3. Division of the short gastric vessels [1], [21], [34]
4. Right crural dissection [1], [34]
5. Mobilisation of esophagus [1], [34]
6. Approximation of the crura by a few stitches [1], [34]
7. (optional) Insertion of an orogastric tube with a (52 - 60) bougie. [1], [34] This specific step is not performed in the Meander Medical Centre and differs per surgeon. [21]
8. Actual folding and suturing the fundus of the stomach [34] At this point, the actual the biggest difference is made in the severity of the fundoplication as described above. [1], [21]
9. Desuflation of the abdominal cavity and closure of the entry ports. [3]

During the procedure from the second phase until the actual creation of the fundoplication, the vagus nerve might be visible and being injured. However, during these phases in surgery it is still hard to discover and see the actual nerve, which may support the unintended injury of the Vagus Nerve.

2.3 Clinical tool

The clinical tool should be applicable during surgery. Based on the difficulty of finding the actual nerve and the ”wandering” path of the nerve, the location on the screen is equally important as if it is on screen. As mentioned earlier, the study of M. Schuhmacher used an algorithm that visualised the resulted by so-called bounding boxes (see figure 2.6). Here the bounding box is visualized around the predicted

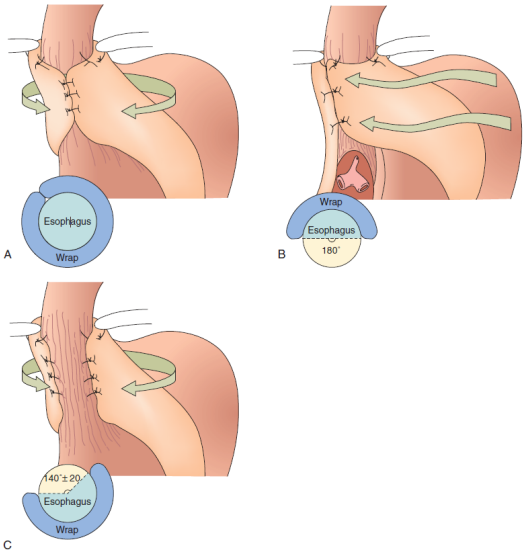


Figure 2.5: A visualisation of the different types of funduswraps that can be made. A represents a 360 degrees fundoplication (also called a Nissen fundoplication). B represents a partial anterior fundoplication (180 degrees, also called Thal or Dor). C is a partial posterior fundoplication ($\tilde{1}$40 degrees, also called Toupet). [1]

area. Optimal is this probably not for nerves, due to their difficult visible presentation. Because it is a small long structure. For instance, if a small tubular-shaped structure can be a large bounding box if the prediction is horizontal. Optimal would be a tool that can segment per pixel if it is a vagus nerve or not. In collaboration with residents and surgeons, it would also be beneficial if there is a region that can be shown where there is a high probability of finding the nerve. This because the nerve itself is not always visible.

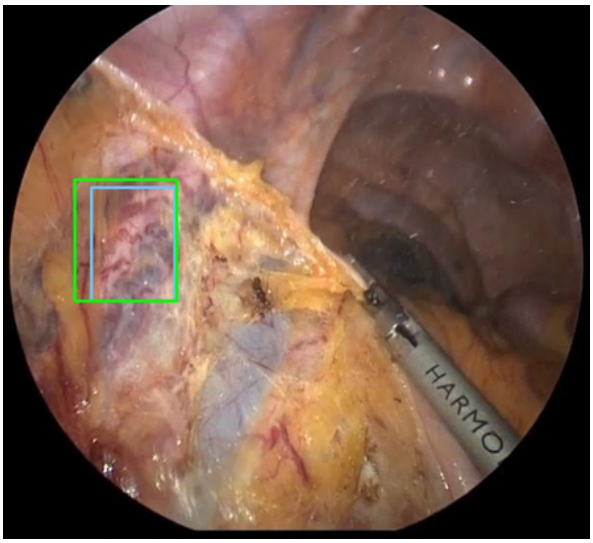


Figure 2.6: This is a visualisation example from a incomple prediction of a urether prediction algorithm made by M. Schuhmacher [4]

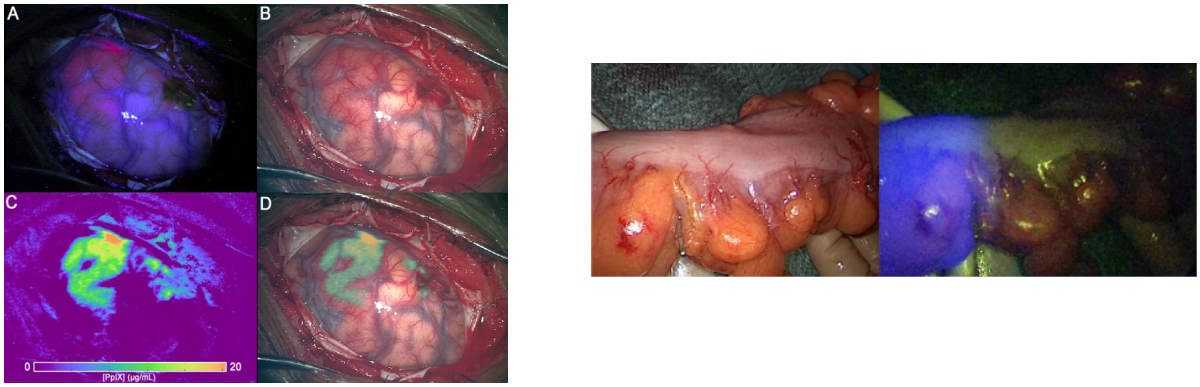


Figure 2.7: Two different types of surgery with multiple visualisations of the ICG fluorescence of the surgical area. This superimposing is only visible on a digital screen not with the naked eye. [5]

To give the information back with high probabilities of finding the Nerve back to the surgeon, there is the option to visualise the segmentation as a separate output on the screen. This would still be difficult matching the exact location of the nerve in the surgical field. If the information could be superimposed on the surgical video, both information can be used by the surgeon. If superimposing this information is applied, it uses the digital equivalent of the ICG visualisation. Switching between the ICG only and the normal Red-green-blue video as well superimposing the ICG as a heatmap on top of the red-green-blue video gives the best of both world to the surgeon. ICG (Indocyanine green) is a fluorescent that is used in different types of fluorescent guided surgery. [5] Superimposing could be a solution if hot spots where there is a high chance of hitting the vagus nerves. A so-called probability map or heatmap. This would lead to a certain probability map as an overlay during the surgical procedure.

To summarize how the clinical representation could be more beneficial than a bounding box are:

- A segmented nerve (a per-pixel knowledge)
- A probability map of the nerve (a per pixel predictive number that it a nerve)
- A superimpose of the output(the two points mentioned above) on top of the original surgical video

Summary of this chapter

In this chapter, the anatomical path and functionality of the nerve are explained. The surgical procedure is made clear and it is clarified which phases of the surgery

are related to the nerve and nerve injury. Finally, the optimal requirements of an algorithm output are set out to make requirements for our setup.

Chapter 3

Technical background

This chapter contains the technical background of AI and DL. More specific on how AI and DL can be used to do object detection and segmentation. Then we explain what temporal information is in the surgical videos, and how this is used as information for the DL algorithms.

3.1 Artificial Intelligence

The actual name *Artificial Intelligence* first used during a conference about ”the science and engineering of making intelligent machines by John McCarthy in 1956. In the subsequent decades’ research in this field became less until the nineties. Though exact definition differs from each other slightly, the general concept is AI a part of computer science which tries to make complex algorithms and machines that mimic characteristics of humans. This is not the *thinking and acting* of robots and algorithms as some people also describe as *Artificial Intelligence*. Different AI algorithms are used by many others during the day such as; text, speech analysis, facial recognition and even in cars. [35]

Though the term/definition of AI broad, the term *Machine Learning* is already more specific. Machine learning is an element of AI that specifically is build to find patterns in data. [9]

Machine learning roughly can be divided into two types; supervised learning and unsupervised learning. Unsupervised learning focuses on interpreting and grouping data based on the data only. This mainly results in the clustering of the input data. Supervised learning is used in creating predictive models with input and output data. The output can be a classification model or regression model. Supervised learning is chosen if the outcome should be a prediction or a classification. [36]Due to the task, that is aimed in this thesis, classification and prediction, we will from now on focus on supervised learning. Although there are also many options available within

the supervised learning such as support vector machines, decision trees, Bayes networks, Neural networks and Deep Learning. Two key terms are important to understand; Classes and features. A *Class* refer to the output that should be found. This can apply on full images but can also be something pixelwise (which is used for segmentation, a per-pixel prediction). A *feature* is an indicator within the data that can be used to separate the data. It is the task to find and calculate that feature to separate data in the classes that are asked. When these features are found, they can be used as an input for the earlier mentioned network (support vector machines, Bayes networks, decision trees, Neural networks and Deep Learning. Neural networks, as the name suggests, mimic the neural structure that can be found in brains, although highly simplified. A visualisation of an example of a simple neural network can be seen in figure 3.4. The basic functionality of this network is that every single input is for each feature that provided. Within al the circles the, called nodes, there is a *weight* that is multiplied with the feature. The output is given to the next nodes in the next layer (figure 3.4 the nodes in the hidden layer). In that next node, all inputs are added together and pushed through an *"activation function"* and also being multiplied by a weight before being passed through to the next layer. This interconnectivity and number of layers can be determined while building the network. The last layer, in figure 3.4 called output layer, the combination of outputs from the hidden layers is converted to the desired output format. This output is generated by an activation function. The type of output mainly determines the type of activation function in the last layer. The previously mentioned steps are when data flows through the network to generate output. However, the weights need to be defined in order to function properly and is achieved by *"learning"*. So data is necessary that contains a known input and output; this is called labelled data. For training the data will flow through the network and based on the error between the model output and the known output the *"weights"* are adjusted. The basic approach is with a big error a bigger adjustment should be made than with a small error. This can be applied many times until the algorithm reaches the satisfying outcome, does not learn any more, or is "overtraining/overfitting". The difficult thing is to adjust the right weight within the network. In the last decade, the size and complexity of those networks exploded. Due to the " depth" of the layers, this concept is called "Deep Learning". [9], [37] So what makes this specific type of networks possible? Well, tasks that are very difficult to accomplish with so-called "rule-based" programming or reasoning like structure recognition. [38]

Deep learning

The growth of deep learning was mainly because of the growth of computation power that decreased the learning time of the compiled/build networks. We will address two key concepts to understand the building of those Deep Learning networks; Layers

and architecture. A layer is already visualised in figure 3.4, all the input nodes form one layer, each other step of nodes within that image. Architecture is the macroscopic shape and type of nodes and layers that is used. Some basic deep learning layers that are broadly used are:

- **Fully connected layers:** This is more or less the network that is visualised in figure 3.4. This can be part of a bigger neural network with multiple other (different) layers. Often you can find this at the end of the complex network structure to combine multiple neurons/inputs to the desired output. [9]
- **Convolutional layers:** Convolutional layers are essential in pattern recognition tasks. For one but also multidimensional signals. For imaging-based solutions, it is quite common to place multiple convolutional layers within the network. The basic concept of a convolutional layer is that they can filter on a specific structure, and the filter itself is constructed during the learning process. So specific features that are visible. [9]

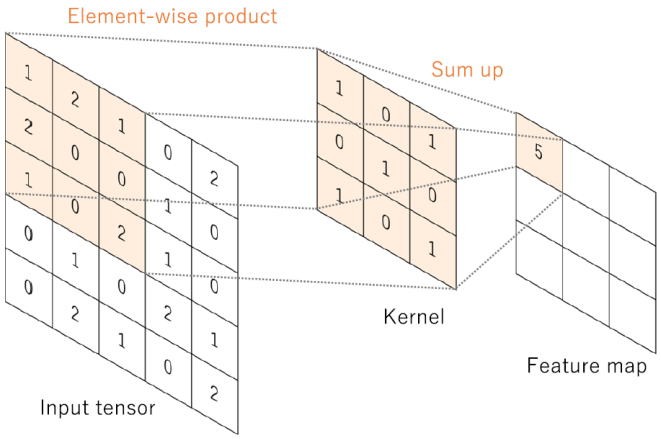


Figure 3.1: An input tensor (which can represent data from an other layer, or is the input image) that is used for a convolutional operation with a given filter kernel (size 3 x 3). The first output of the convolutional kernel is also given in the feature map. There is no padding and the stride is 1. [6]

- **Pooling layers:** pooling is a way of downscaling the output of the previous layer. Mostly in convolutional networks, multiple outputs of the convolutional layer are with pooling downscaled. Pooling groups a set of outcomes that are as wide as the dimensions of the ”window” and based on the type of pooling it passes only one outcome. The most commonly used is maxpool, here the outcome within the moving window will be passed through. [9]
- **Activation layers:** Deep Learning mainly is built on top of non-linearity of the system. This is mainly achieved by the activation layers. Sometimes the

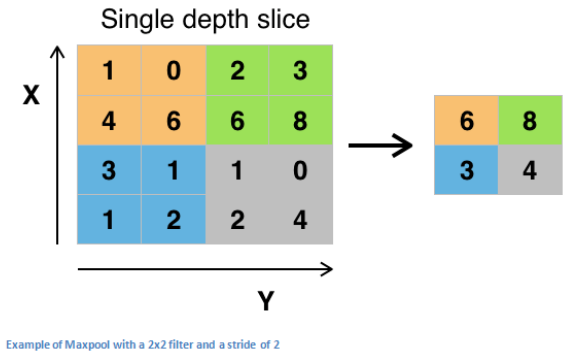


Figure 3.2: Max pooling example. An example of a large image on the left with pixel values that are maxpooled. So a small window (in this case 2x2) the max value is used to generate a smaller image on the right. [7]

activation layer itself is built within other layers. The functionality was copied from the biological behaviour of non-linear input-output of neurons. The closest Deep Learning activation layer that mimics the neuron activation was the hyperbolic tangent function, but nowadays, other better-performing functions are used to create this non-linearity. Also, specific activation functions can be selected for specific tasks and mainly determine the output of the neuron. [9]

- **Output layer:**The output layer is a very special node as written earlier. It can convert the different inputs to the desired output format that can be interpreted by us. For instance, some networks need a yes/no, left-right, cat/dog output. This is binary output although others need a specific number or even set of numbers. In figure 3.5 different output formats are presented. These output layers are based on special *"activation functions"*. [9]

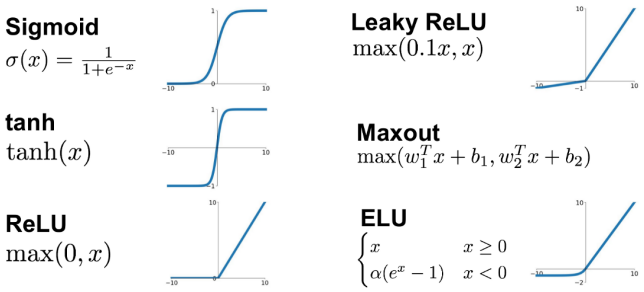


Figure 3.3: An set of activation functions for deep learning layers. [8]

- **Residual layer:** The best to describe these layers are additional layers. These layers are made to achieve better results for specific cases. The function of this layer is to bypass some input data to the next layer or even further in the network and compare the two paths (bypass vs no bypass). The non-bypass is forced to perform better than the no bypass path, which increases learning

in that specific part. This mainly allows to decrease the number of layers and so the number of trainable parameters and overfitting of the training data. [9]

The shown output types of Deep Learning in figure 3.5 also each need different annotations for training, validation and test data. For image classification are labels on image level needed, for object detection are object labelled (i.e. bounding boxes) needed, and for semantic segmentation or instance segmentation are on pixel-level labels needed. As a rule of thumb, each type is that is on a higher resolution level also requires more labelling time. [10]

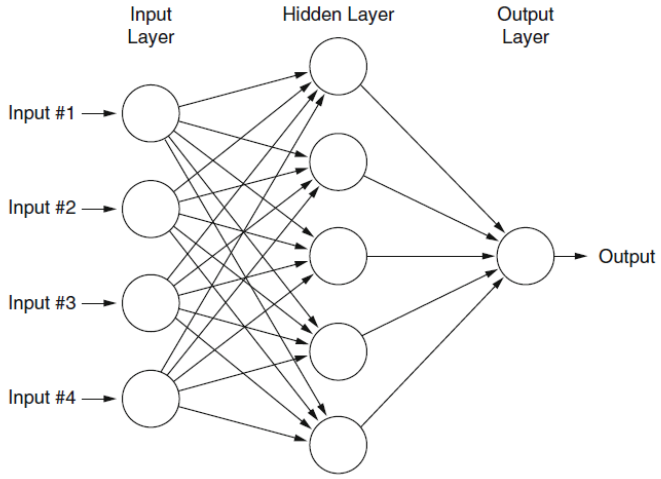
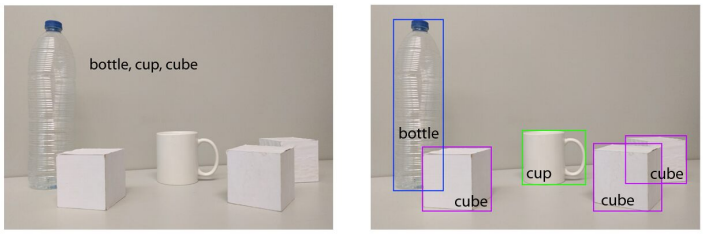
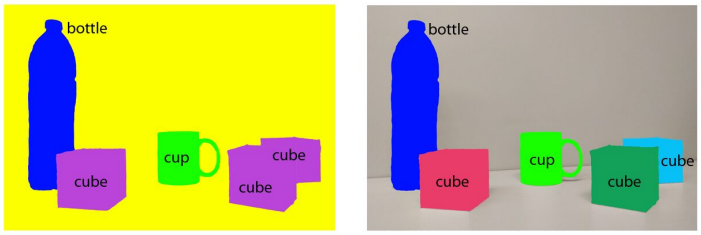


Figure 3.4: An scematic visualisation of an neural network. The network has multiple neuron (in deep learning called perceptrons). [9]



(a) Image classification (b) Object localization



(c) Semantic segmentation (d) Instance segmentation

Figure 3.5: Network outputs in Deep Learning networks can have different tasks. Every step needs different labelled data and other types of networks. Especially the activation function in the output layer is important. [10]

3.2 AI and healthcare

As shown in the research of Schuhmacher et al. (2018), not only the Vagus Nerve is also other structures that can be automatically detected of segmented can be useful. The opportunities are endless, even within surgery alone. Although Schuhmacher et al. (2018) showed an example of AI and interventional care, most AI tools in medicine are on the predictive tools for detection and diagnosis. Within surgery, AI might extent the eye of the surgeon but also objectify the performance of the surgeon or benchmark segments of the surgical intervention. So segmentation of other anatomical structures might be equally relevant. [4], [28]

Within radiology and pathology, AI is much further. From applications that diagnose TBC based on chest X-rays to mammography screening algorithms that is mentioned by the author as performing at the level of a radiologist. The impact on the physician in the field of image recognition and predictive analytics might be substantially based on some predictions that those algorithms perform more effective than humans. Although most physicians will not necessarily lose their job, due to some gradual change or specific areas that need traditional exams or they augment the use of these algorithms. [6], [35]

3.3 Deep Learning for structure recognition

Sufficient data is necessary for Deep learning due to the high number of parameters that need to be trained. Existing datasets that provide sufficient data with their ground truth can thereby be used to test functionality, test hypothesis and even compare results of networks that use the same datasets in for example contests (such as Kaggle.com).

When building datasets, there is always a risk of not providing/generating enough labelled data. Mainly because the labelling is quite labour intensive. In that case, the solution can also be the use of existing machine learning datasets to test the hypothesis and beside perform the process on the own made dataset to compare if results meet the expectations. [9], [39]

3.3.1 Network structure for semantic segementation

As described above the networks can be built like lego bricks with different layers, different inputs and different activation functions, all depending on what the network should do. AI gained momentum in using images due to a new approach that came in 1998. The trainable parameters came as convolutional filters. Those were introduced for the detection of handwritten characters. [40] Those convolutional layers

are also described above. These networks did the first step; classification of images not the segmentation of images. The transformation of fully connected layers with convolutional layers it is possible to perform an image classification (see example in figure 3.6). This is called a fully convolutional network. It is able to produce spatial heatmaps but no realistic semantic segmentation. The next step is to perform a reverse kind of operation, upsampling and deconvolutions to achieve the per pixel predictions (see figure 3.7. [10], [11]

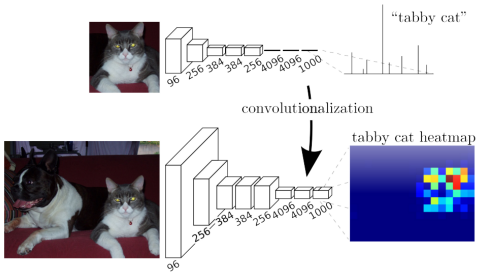


Figure 3.6: An image is feeded to a pre-trained FCN that is visualized by a set of white blocks that mimic the covolutional layers and their size. [11]

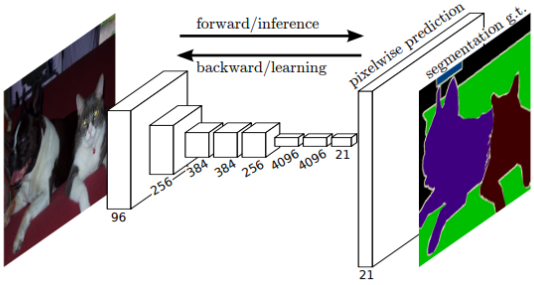


Figure 3.7: A set of deconvolutional layers and upsampling of the setup of figure 3.6 brings more semantic features back and gives a better per pixel prediction. [11]

This approach is also used by Ronneberger et al. (2015). They achieved a high-resolution output with limited training images. Their elegant network structure is U shaped and was highly symmetrical. It uses residual layers during upsampling from data directly from the left part of the U-NET to gain resolution back. This is done multiple times until the actual or desired resolution is achieved. The actual U-NET, which is used by Ronneberger et al. (2015) is visualized in figure 3.8. [9], [12]

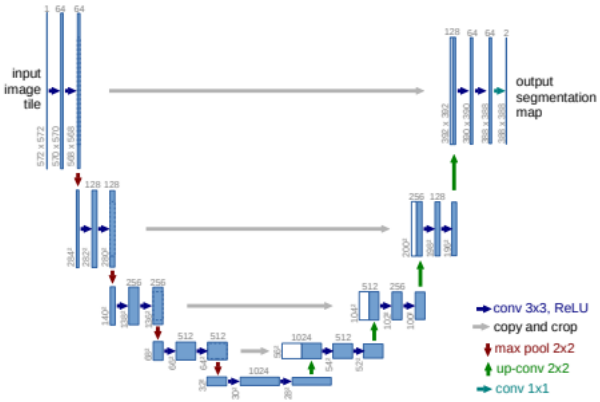


Figure 3.8: Basis U-NET that with an 1 channel 572x572 (pixel width height) that was used in the paper of Ronneberger 2015. [12]

Although there are many more network structures, like Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) they differ per needed applications in

this paper thesis we mainly describe to the used ones in imaging and the conducted research.

3.4 Dense optical flow

[12] Dense optical flow is not specifically part of DL or even AI. The concept of optical flow is the displacement of an object that is displayed. This optical flow is thereby a two-dimensional vector field that describes the displacement of the object between the images. If this is applied to each pixel in the image, it is called, dense optical flow. [41] Dense optical flow in video frames can be interpreted as multi-frame motion estimation. Zanjani et Al. (2010) already described that multi-label segmentation in videos in most cases is based on the individual frames while ignoring the dynamic information that could be stored in the video. In their approach to adding this extra temporal information was able to increase the segmentation performance. This motion clue was recently again confirmed in a similar, dense optical flow supported, approach by Rashed et Al. (2019). [27], [42], [43]

The used dense optical flow by Rashed et Al. (2019) was Farneback dense optical flow [43]. *Farneback dense optical flow* is able to make a motion estimate of each pixel of two consecutive image frames. The *Farneback dense optical flow* is performed in two steps; polynominal expansion and displacement estimation. The output will be a *displacement field*, a vector of each pixel that gives an estimate of the movement that the pixel made between two frames. [42] [44] The steps considered in Farneback dense optical flow algorithm to achieve a per-pixel displacement estimate (Dense Optical Flow):

1. Neighbourhoods of pixels are described with a polynomial (Polynomial expansion). A quadratic polynomial of a signal is given in equation 3.1.

$$f(\mathbf{x}) \sim \mathbf{x}^T \mathbf{A} x + \mathbf{b}^T \mathbf{x} + c \tag{3.1}$$

   [42] Where $\mathbf{A}$ is a symmetrical matrix, c a scalar and $\mathbf{b}$ a vector.

2. The two image frames differ from each other, as well the polynominals of the two frames. With the transform of the polynomials a displacement fields can be calculated. After some refinements the dense optical flow is given as a vector field. [42], [44]

The estimated flow given as a vector field of the two images (example given in figure 3.9. But can also be converted to a polar coordinates or the RGB color that is given in the figure. So this creates two options a RGB = Red, green and blue. A 3

channel input containing red, green and blue values. Those are commonly used for videorepresentation or pictures. (RGB) image represending the displacment (dense optical flow) or the vector field.

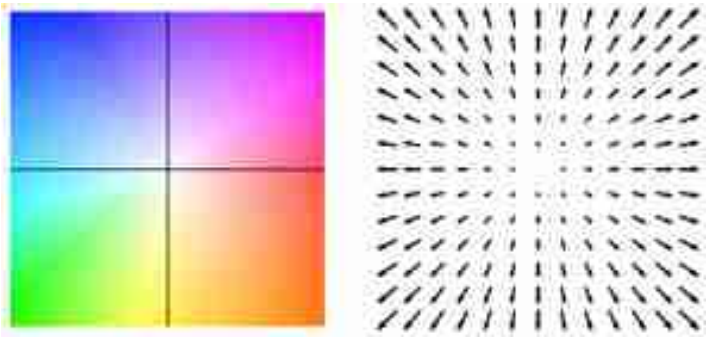


Figure 3.9: The flowfield, or vectorfield on the right and its RGB color that it represents in a polar coordinate system [13]

Summary of this chapter

In this chapter, the brought AI and Machine Learning are explained. Promising technologies in many industries, like automotive and healthcare. Machine learning can be applied to perform predictions or segmentation tasks. For those tasks is labelled data is preferred to train the algorithms. The training of this subdivision within machine learning is also called ”supervised learning”. The labelled data, input and output (i.e. data with the ground truth), is necessary. In the case of testing, building and training Deep Learning algorithms, sufficient data is necessary. Deep Learning networks can be built with different architectures, shapes and sizes. Also, multiple network structures are available for segmentation tasks. This mainly depends on the specific task and outcome that is asked or available data. Dense optical flow itself is not AI-related. It is a way of estimating the displacement of an object within two images on a per-pixel level. Previously conducted research suggests it might be a solution to improve the segmentation outcomes of Deep Learning networks.

Chapter 4

Method

In this chapter the actual setup to create a set of algorithms to test the idea's of segmentation of anatomical structures, use temporal information (movement information) and visualize them on a proper way to be clinically applicable for the surgeon or new surgeons/residents. In figure 4.1 the input, algorithms and output are summarised in an overview.

First, the datasets were used to be able to train AI networks. From the both dataset is also movement data calculated (*Dense Optical Flow*). Secondly, 5 different UNET structured network were adjusted to fit the provided datatypes of the datasets and trained with the provided datasets. After fine-tuning and retraining the networks are saved. In the end, the 5 trained are tested on a testdata set to provide performance data.

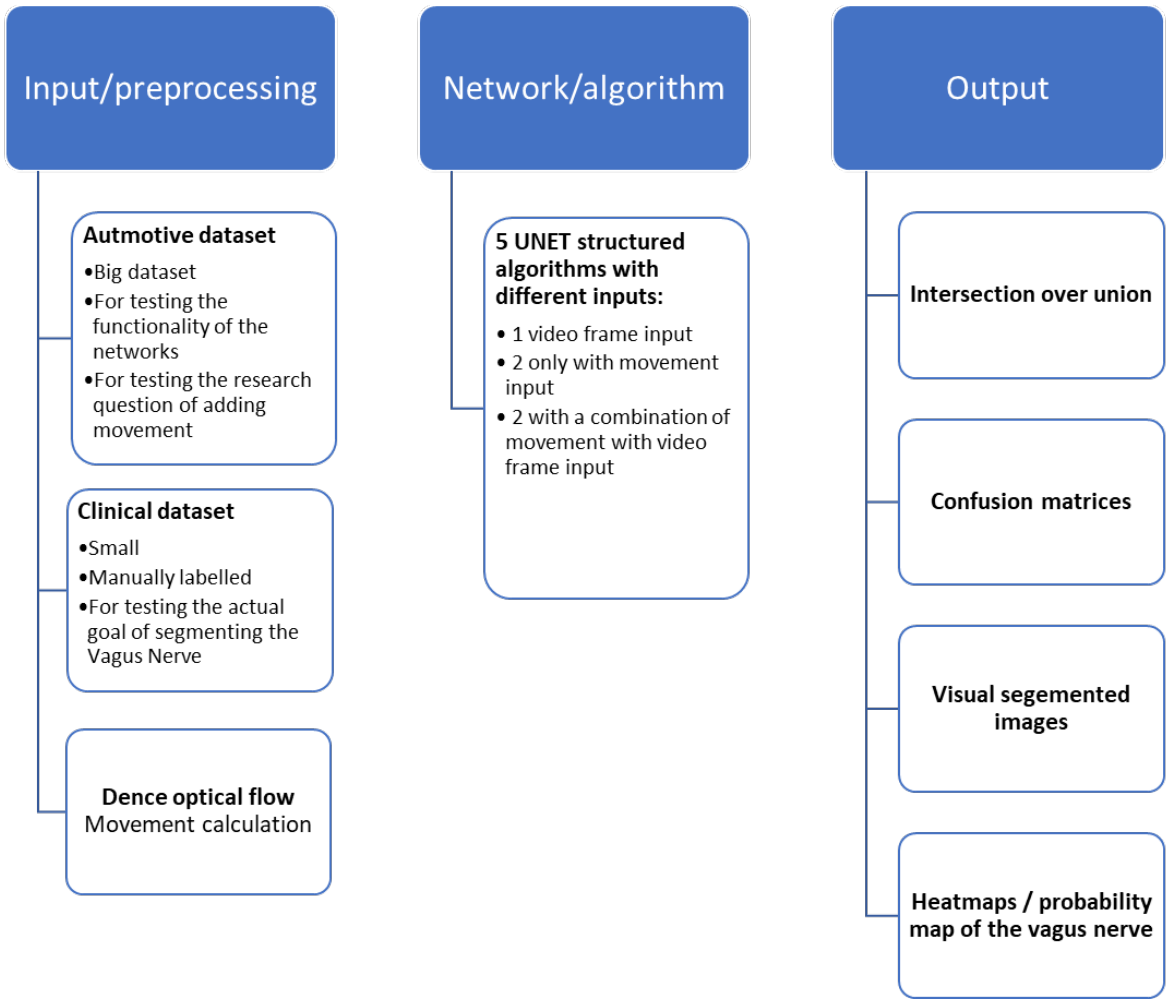


Figure 4.1: In this overview the input of the algorithms, the 5 algorithms and the outputs are schematically visualised.

4.1 Datasets and data retrieval

The size of an exiting, available datasets can easily be more substantial than what can be made available in a short time in a hospital. Also, no clinical dataset was available when we started building the networks. There is a potential risk that a clinical dataset that is too small to test the setup of adding temporal information to improve the semantic segementation outcome. To train and validate different UNets, two datasets were used.

First, an existing big database (VKITTI) was used to be able to train and test a self-developed algorithm. This is not a clinical dataset but an existing, non-medical automotive dataset. This was preferred because 1) no clinical dataset was available and 2) there is a potential risk that the manual created dataset is too small because it is a timeconsuming process.

Secondly, a clinical dataset was created to be able to train and validate the fine-tuned UNETs. In contrast to the VKITTI dataset, the clinically dataset was labelled manually. Mannually labbeling is timeconsuming and has the risk to be more inaccurate due to the fact that tissues can be hard to distinguish as well.

Different object structures were chosen in the two datasets, still similar structures were selected. The structures in the VKITTI which were most similar to the Vagus Nerve structure were selected: traffic signs and traffic lights. This was based on their object shape (thin) and availability of pixels of those classes. For the road structure a similar, almost always present structure is chosen in the clinical dataset; the liver. Those structures are probably easy to segment due to their availability in the datasets. The other structures that were chosen more on their overall availability and possible added value to segment for future surgical applications.

VKITTI Dataset

The VKITTI dataset is an automotive dataset. Which means it is made for the car industry to create and test algorithms like self-driving cars or safety systems. The virtual KITTI dataset is a synthetic dataset, which means it are no real pictures, but computer-based images in order to have a 100% knowledge of what each pixel is (i.e. car or road). The original dataset KITTI is an actually filmed dataset in a car mostly in Karlsruhe. With a real-to-virtual cloning method the actual images are converted to a virtual world with cars, trees, roadsigns on more or less the same place. The five environments, the five dash videos were mostly filmed in sunny condition, but with the virtual kitty approach, other weather conditions could be simulated. The virtual KITTI dataset also contains the five environments but also in 7 weather conditions. In our test setup, we only use the virtual KITTI data from sunny conditions. In figure 4.2 the real-to-virtual clone visualised. The advantage of using the virtual KITTI dataset over the original KITTI dataset is the available 100% true ground truth labels. [14]
In our setup we used:

- 5 different 'enviroments' all in sunny/normal weather conditions
- 1016 Trainings images (2 environments)
- 836 Validation images (2 environments)
- 269 Test images (1 environments)
- Used labels (multiple labels are converted to 5 labels: road, car, the "pedestrian" label are not pedestrians but the lampposts and traffic signs,

Figure 4.2: From top to bottom 5 frames one from each environment. Left are the kitti images and left are the virtual kitti images that are cloned from the real kitti images. [14]

- Class distribution of pixellabels in the dataset (road 24.4%, side 15.5%, ped. 1.6%, car 8.1%, else 50.3%)

Clinical laparoscopic dataset

The clinical dataset is made of selected video frames of the anti-reflux surgery. Due to the problem that the nerve is difficult to see, all surgical videos were reviewed and between the surgical phase of the left crural dissection and the actual folding and suturing of the fundus. Between these phases, there were possible events of a visible vagus nerve. This moment, ”event”, were cut out and used for labelling (gold standard). Ten seconds before this ”event” was saved as a video clip to generate the dense optical flow calculation and backup. All types of HHDs (ranging from type 1 to 4) were included robotic and conventional laparoscopic. Datasets for deep learning purposes need to be significant to be able to *'learn'* the network. Also, the dataset can be as diverse to learn from (anatomical) deviations.

- 10 surgical laparoscopic fundoplication surgeries wich made a total of 105 images.
- 73 trainings images (6 sugeries/patients)
- 26 validation images (3 surgeries/patients)
- 6 test images (1 surgery)

- labels Liver, Stomach, Esophagus, Crus, Nervus Vagus Anterior, Nervus Vagus posterior, the rest of the pixels is ”Else”. The labels Stomach, Esophagus are combined aswell for the two Nervus vagus labels are combined to the label *N.V.* (Nervus Vagus).

- Class distribution of pixellabels in the dataset (Liver 8.7%, Crus 5.0%, N.V. 0.15%, Stomach/esophagus 14.5%, Else 71.6%)

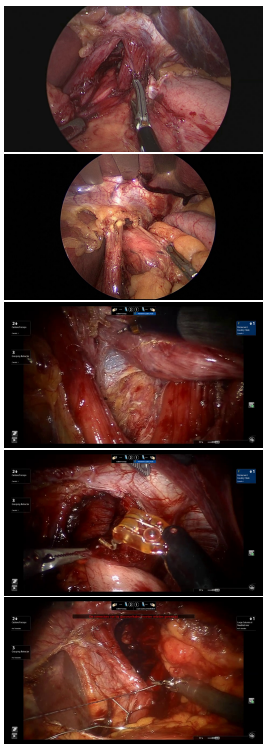


Figure 4.3: From top to bottom 5 random frames from the clinical dataset. The round field of view is caused by the video scope that is used in conventional laparoscopy. The square field of view is from the robot (the small blue squares is information of the robot that is projected on top of the surgicalvideo.

4.2 Labelling and preprocessing

For supervised learning, labelled data is required in order to train the networks for a specific task. Because the models are built for the data of the VKITTI dataset, and all the specific pixel values are used a similar labelled clinical dataset is preferred. The best solution to create a per-pixel-labelled-dataset. For the clinical dataset, this is conducted in two steps. Step one, annotation with a toolbox that allows easy and highspeed annotation. Step two is the output of that annotation tool and converting it to the same data format that is used in the VKITTI dataset. For the first step, the

image labelling toolbox of Matlab [15] is used. With a python script, the multidimensional arrays are converted to Portable Network Graphics, image data format (PNG) which is the same data format as the VKITTI data.

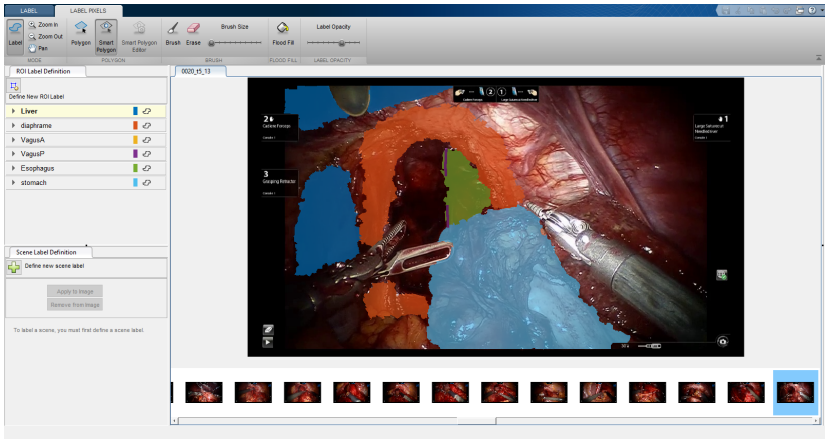


Figure 4.4: The first step: the image labeling of the clinical dataset with Matlab 2018a. [15]

DOP dataset

As written earlier, *Zanjani et Al.* (2010) described that multilabel segmentation in videos in most cases is based on the individual frames while ignoring the dynamic information that could be stored in the video. In their approach, they showed that within the dynamic behaviour segmentation is still possible. Although the VKITTI dataset also contains ground truth DOP data, this is not used for our setup. Because this is in an other format, also a 100% true, we chose to apply the same calculation of the movement that is applied on the clinical dataset. So differences would be minimal. In our setup, the DOP is calculated between two frames of which the last frame is also has ground truth labels. The difference between the first and second image was 0.1 for the VKITTI dataset and 0.4 seconds for clinical dataset. For the VKITTI data, this dataset is calculated between two frames. For the clinical dataset, this is calculated per sub video. Open CV is used to calculate the Farneback dense optical flow of those two frames. This data is stored in two ways; A vector array (two dimensional) and a polar (colour) version (a three-dimensional colour image). [41], [42]

4.3 Network architecture

Python 3.6 is used in combination with Keras to create, train, test and evaluate the networks [45]. The basis of the used network is a U-Net for the paper [12] In our setup, we choose a U-(shaped)Network (U-NET) for segmentation. *Ronneberger et al.* (2015) showed that a simple U shaped can perform with a rather small dataset to segment biomedical images (Microscopic cell structures). The main reason we chose this network is it can give a prediction per pixel. The output is either a binairy

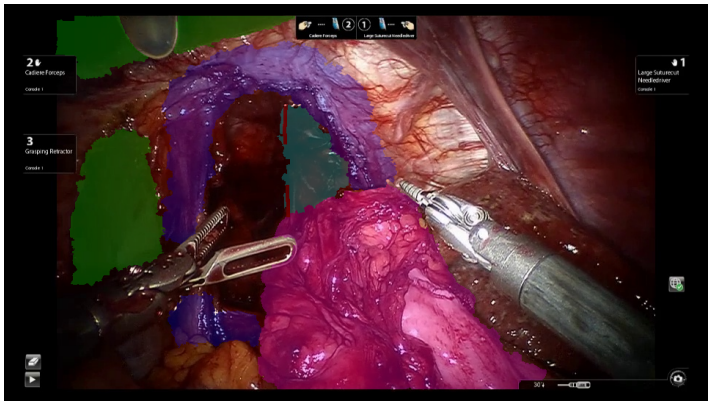


Figure 4.5: Image of the surgical procedure with with the labels visualised on top of the image as an semi-transparent overlay. The original image can be seen in figure 4.6 (Green for liver, blue for oesophagus, pink for stomach and purple for the Crus)

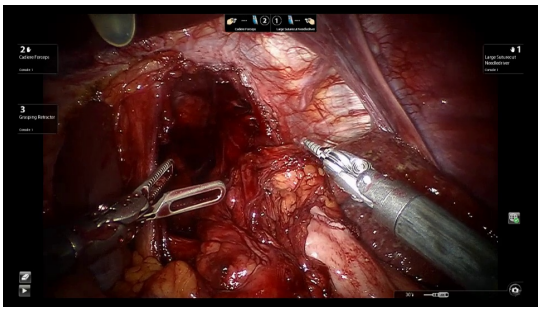


Figure 4.6: Image of the surgical procedure.

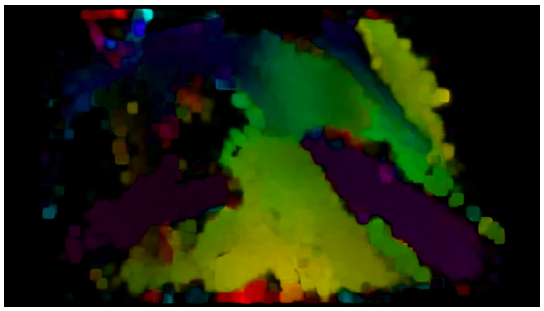


Figure 4.7: The corresponding DOP of figure 4.6. The colors represent movement by the colorscheme in figure 3.9

per pixel. [12] Next to that a U-NET and with modifications in the first and last layers they can be able to have a different number of input channels (for instance an RGB image is three channels R, G and B) and the output can be a different number of classes as output as well a variable output per pixel. Also, we know that a U-net can easily be trained on limited data U-NET.

4.3.1 U-NET

Based on the structure of *Ronneberger et al.* (2015), we modified the last layer that it can give five output values per pixel between 0 and 1. Each number will act as a prediction it belongs to one of those five classes. This is performed by the activation function softmax in the last layer. With this setup, it is possible to generate a heatmap per class but also post-process to generate the 5 class outcome.

The input has a number of channels that corresponds to the dimensions of the input data. An Red-Green-Blue (RGB) image has height x width x 3 (red, green, blue). If we add movement information either the vector array of the RGB (polar) data, we concatenate this to the RGB data. So an RGB image with a 2-dimensional

vector array will be an input array of height x width x 5 (3 for the RGB image and 2 for the vector array).

To summarize the basis modification on the basis U-NET.

- Output not binary (0,1) but between 0 and 1 and provide 5 output channels
- Input modifications for different shaped data inputs

U-NET input modifications. The name of the algorithm is based on the input:

1. ***DOP (RGB)*:** 3 input channels used RGB input
2. ***RGB video + DOP (vec)*:** 5 input channels used RGB org. + vector DOP
3. ***RGB video*:** 3 input channels used RGB original frames
4. ***RGB video + DOP (RGB)*:** 6 input channels used RGB original frames + RGB DOP
5. ***DOP (vec)*:** 2 input channels used vector DOP

In this thesis the names of the trained models trained have the names as described above.

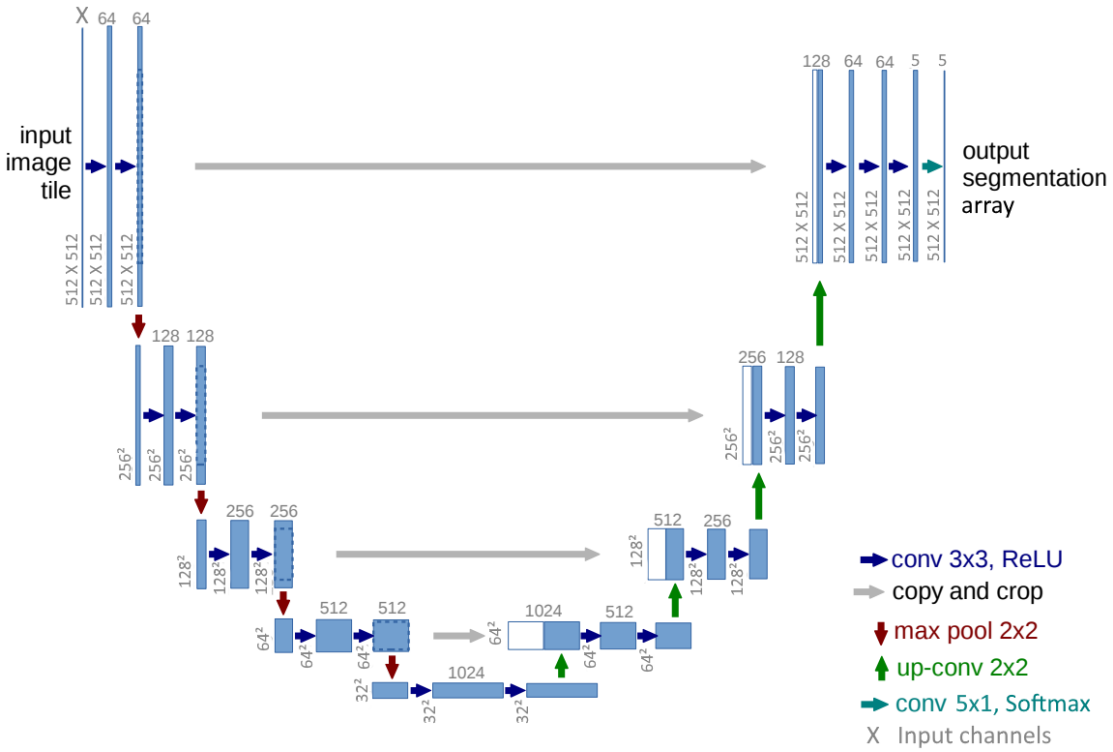


Figure 4.8: Overview of the UNET that is used. The X depends on the number of input channels given in the enumeration above. This network is a tailored and modified version of *Ronneberger et al.* (2015) [12]

4.3.2 Hyparameters and *Dense Optical Flow* parameters

To train the algorithms, different techniques were used to prepare/optimize the available datasets. Due to the different sizes of the dataset, we watched the clinical dataset manually. Also, the step sizes are smaller compared to the bigger VKITTI dataset. The training strategy with the kitti set is different,

Kitti training strategy

- batchsize: 5 stepsize: 50 maximum epochs: 20
- Early stopping on a plateau (patience 10 epochs)
- Automated reducing Learning rate (when monitored condition reaches a plateau with a patience of 4 epochs)
- Monitored paremeter for saving the network: lowest validation loss
- loss function: *kullback leibler divergence*

Clinical dataset training strategy

- batch size: 5 images, steps: 25 per epoch, epochs: 30
- No automated early stopping on plateau
- training manually watched; stopped if the loss reached a plateau
- Monitored parameter for saving the network: Accuracy, but manually watched
- loss function: *kullback Leibler divergence*

Training, validation and testing hardware

- Windows 10 pc with an NVidea GPU for accelation
- An anaconda virtual enviroment running Python 3.6 [46]
- A Keras with a Tensorflow backend [45]

4.4 Performance parameters/ evaluation metrics

The output of the algorithms or model is given in an accuracy and loss during training, but the actual performance may differ per class that is segmented. For instance, the algorithm may be better in segmenting a road instead of cars. All algorithms will be tested with a test dataset, and the output is used to calculate an Intersection over Union (Jaccard index) and a confusion matrix is made.

4.4.1 Jaccard index, Intersection over Union

An often used metric in medical imaging is the Jaccard index. A per pixel accuracy metric. The corresponding formula 4.1 shows the used definition in our setup. The idea is that the segmented area of the output of the algorithms overlaps with the segmented area of the ground truth, this intersection area is the intersection area (see figure 4.9. The intersection area is divided by the area that is formed by the ground truth with the algorithms output (union). In perfect condition the index would be a 1.0 but that is not realistic. A jaccard index of 0.5 and higher is considered as good. This is clearly visualised in 4.10 [16], [47]

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{4.1}$$

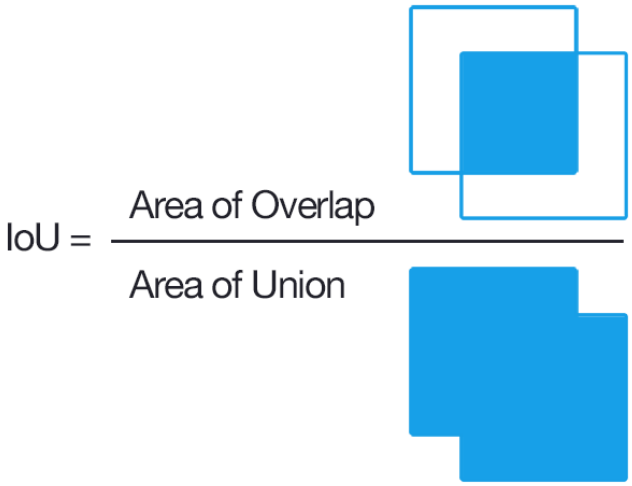


Figure 4.9: A visual representation of the mathematical representation of the intersection over union. [16]

Figure 4.10: A visual representation of the intersection over union. The green box represents the true area (ground truth) and the red box represents the predicted area. [16]

4.4.2 Confusion matrix

The IOU tells something about the performance of a specific class. A confusion matrix is able to show relations in multi-class classification outputs. It is not only able

to show if an output is assigned false, but also to which class it is assigned falsely. In order to compare the ground truth labels with the predictions by the network, confusion matrices are calculated based on a test dataset. This is performed to see if certain structures might be falsely assigned (too easy) to another class. An example of an confusion matrix is given in figure 4.11. [48]

(A)

Class	Y	N
Y	True positive (VP)	False negative (FN)
N	False positive (FP)	True negative (TN)

(B)

Class	1	2	3	4	Total
1	70	10	15	5	100
2	8	67	20	5	100
3	0	11	88	1	100
4	4	10	14	72	100

Figure 4.11: A small example of a 2 class and 4 class confusion matrix. In the four class confusion matrix the false output can be seen in other output classes. Vertical axis are the ground truths, horizontal are the predicted classes. [16]

4.4.3 Visual inspection

It is written above that the algorithm or model can give an output per class that represents a probability of a pixel belongs to that class. This can be visualized like a heatmap that is comparable with the ICG fluorescence of chapter 1 (figure 2.7).

This is also a subjective way for understanding the output and performance of the network. The outputs for a single class, the vagus nerve, will be presented for the different trained models. For the RGB model all classes will be shown as heatmaps.

4.5 Class weigthing

In our datasets, especially the clinical dataset, not all classes are equally available. This may lead to learning not equally each classes. So performance of one of the classes could be much higher than others. *Class weighting* weights a higher value for mistakes during training in less available classes. This forces the Deep Learning model to train more equal for all the classes although they are not equally present in the data. In Keras *class weighting*, to counter balance the class imbalance was not possible with our setup of the Networks. [45]

A possible option is to weight the output of the model after training. Due to a the lower availabible class, the output of the model is weighted heavier by a certain α. See the formula 4.2. Normally this is tackled during training by class weighting. In our dataset the Vagus Nerve is only 0.15% of the total pixels. This is only performed for the output of the Vagus Nerve and for only the basic RGB model. To understand

if performance will increase, the IOU and confusion matrices are plotted for different values of α

$$Weighted\ model\ output\ value_{Vagus\ Nerve} = \alpha * Model\ output\ value_{Vagus\ Nerve} \quad (4.2)$$

summary of this chapter

In this chapter, the setup of the experiment is explained; a set of deep learning networks are trained, tested and validated on two datasets with and without the movement information. The movement information is Dense Optical Flow. One of the two datasets, the VKITTI dataset, is already used in machine learning. The other dataset is made from specific episodes in the surgical intervention when the vagus nerve could be visible. The performance of the trained networks are tested and set out with a confusion matrix, a Jaccard index (intersection over union) per class and a heat map/probability map is visualized of a test image.

Chapter 5

Results

In this chapter, the output of the different networks are shown. The results of the RGB only model is also shown with the different heatmaps. The intersection over union (IOU) and the confusion matrices are presented for all proposed networks for both datasets as mentioned in chapter 4. Also, the per-pixel prediction of the Vagus Nerve is shown of all the proposed networks of a test image of the test dataset.

5.1 IOU

The IOU is the mean value for that class of the full test dataset. Every colour represents an model. The best scores are in both datasets the green (RGB only input). Combining video data with dense optical flow (Orange and Red bars) show less good results. Only using Dense Optical Flow as input allows still some segmentation but performs the worst. Visible are the differences in performance between the datasets. The smaller clinical dataset performs with a substantial smaller IoU. Both datasets show a high IoU in the ”else” label. The ratio of IoU between that ”else” label an the other labels is bigger in the small clinical dataset. An IoU above the 0.5 is mostly seen in the labels ”Else”, ”Road” and ”car” (for the RGB only model/algorithm). Exact values are available in the appendix A

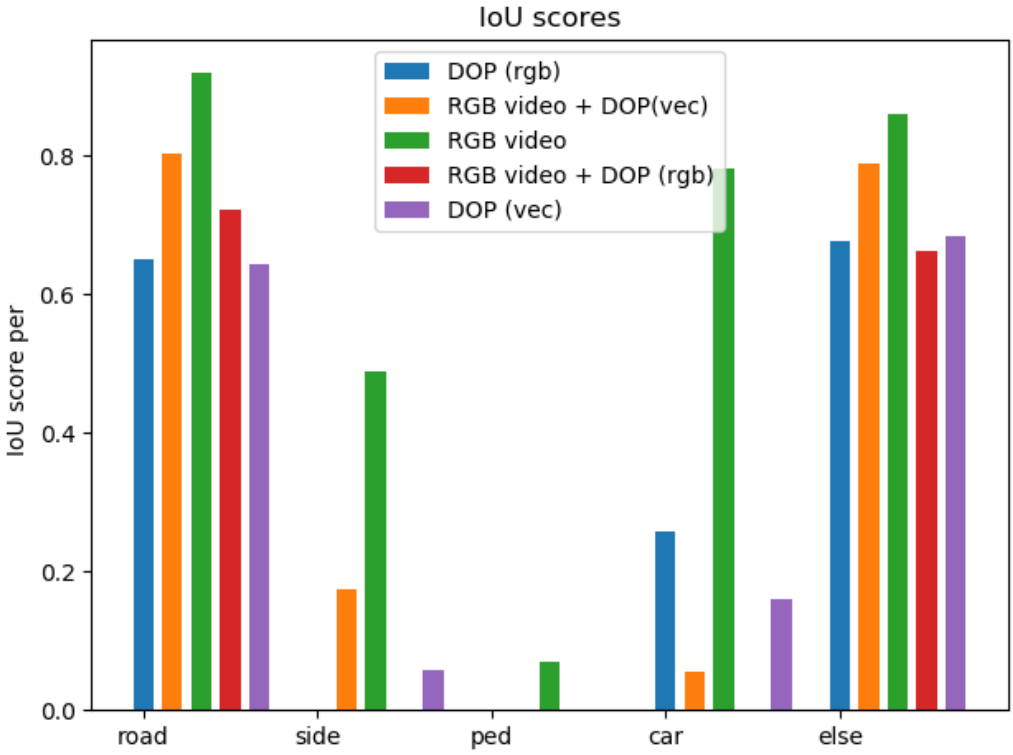


Figure 5.1: The IoU scores of a VKITTI test set with the different networks.

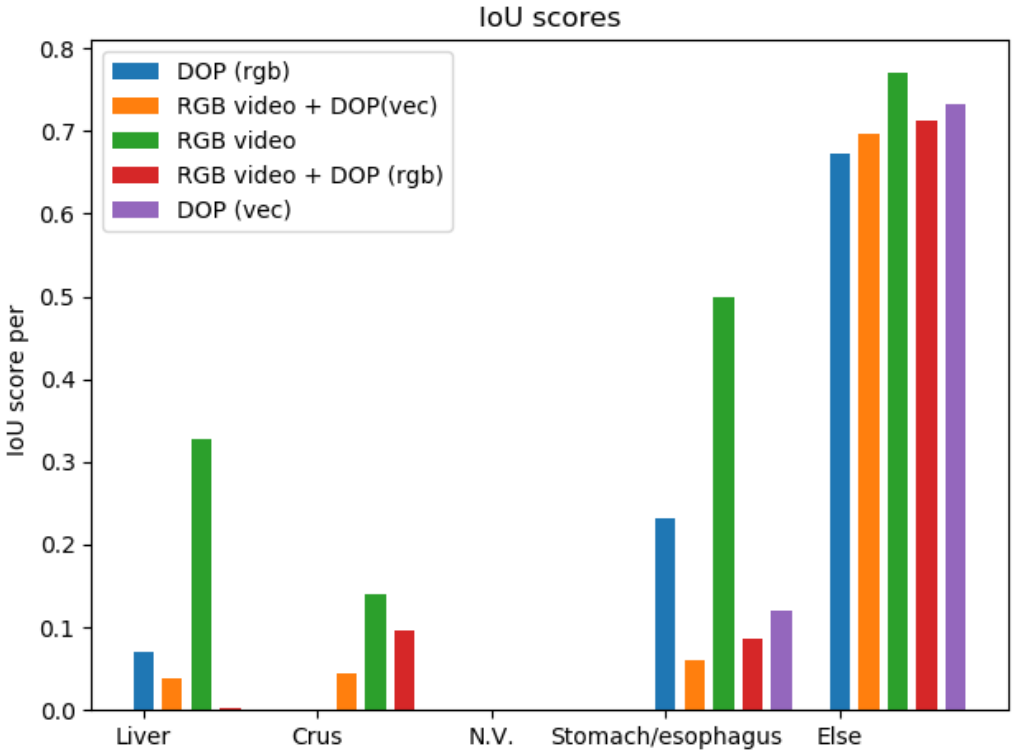


Figure 5.2: The IoU scores of a clinical test set with the different networks.

5.2 Visual inspection of outputs and output of the basic segmentation network

In figures 5.3- 5.6 an output of the model input of the basic RGB video input of the model, with all five label outputs visualized as a heatmap. Plus, the combined output. The highest output channel per pixel is the assigned class. Note in figure 5.4 the liver is not visualised by a transparent label. The original image that is given to the RGB only model.

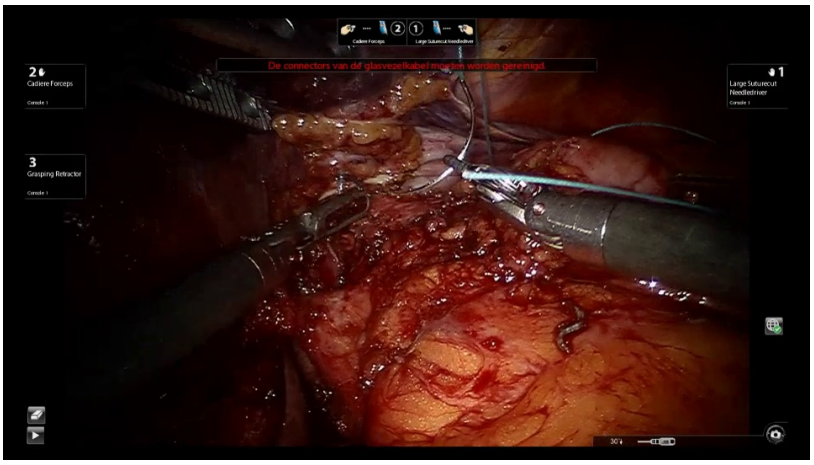


Figure 5.3: A raw RGB input frame of the clinical test dataset. On this image are two surgical tools visible (on the left and the right side).

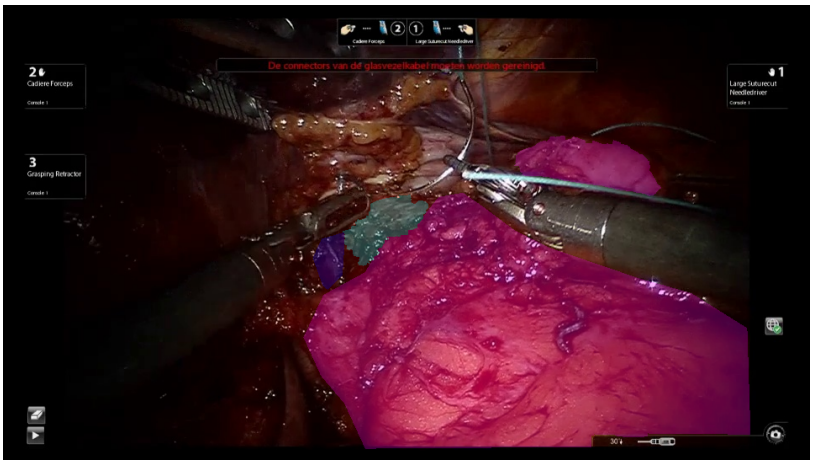


Figure 5.4: Input image of the test set with the ground truth labels as an overlay. Only the liver on the left was not labelled manually, unfortunately. (stomach is pink, the esophagus is blue, and the crus is purple-blue.

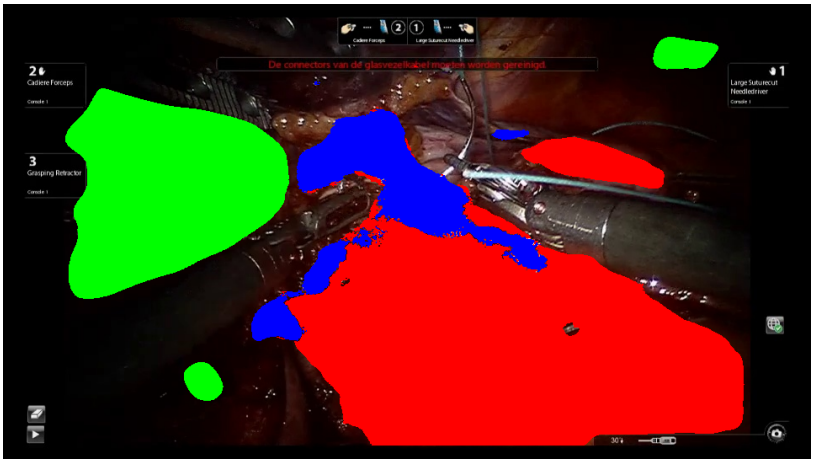


Figure 5.5: The output/prediction visualised on by hard colours on top of the input figure 5.7 by the RGB based model. The green colour represents the predicted liver. The Vagus Nerve is not predicted by the model, and the red colour represents the predicted stomach and oesophagus, the blue colour represents the prediction of the crus, the label else is transparent

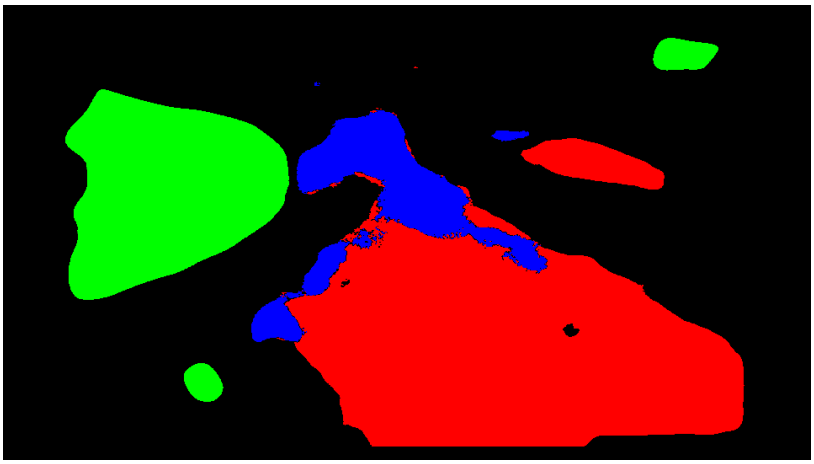
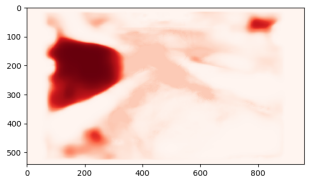
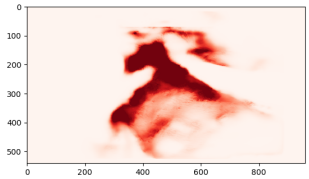


Figure 5.6: The output/prediction visualised by hard colours only of the of figure 5.7 by the RGB based model. The green colour represents the predicted liver. The Vagus Nerve is not predicted by the model, and the red colour represents the predicted stomach and oesophagus, the blue colour represents the prediction of the crus, the label else is black

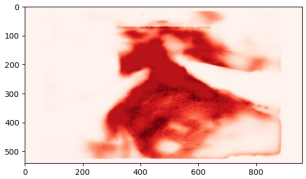
In figure 5.7 5 output channels are visualised as the probability maps/heatmaps. This visualisation is from the same image as in figure 5.3. Some structures are easier to distinguish for the human eye, such as the liver in the left sub image. Also, the region of the vagus nerve is broad.



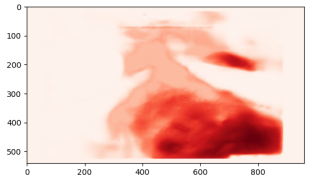
Liver
output



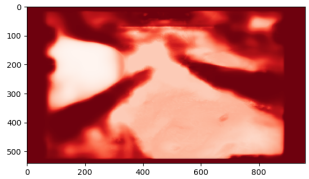
Crus
output



Nervus Vagus (Vagus Nerve)
output



stomach/esophagus
output



Else
output

Figure 5.7: The separate outputs of figure 5.4 by the model visualized as heatmap. White represents a low value, so a low prediction

5.3 Confusion Matrices

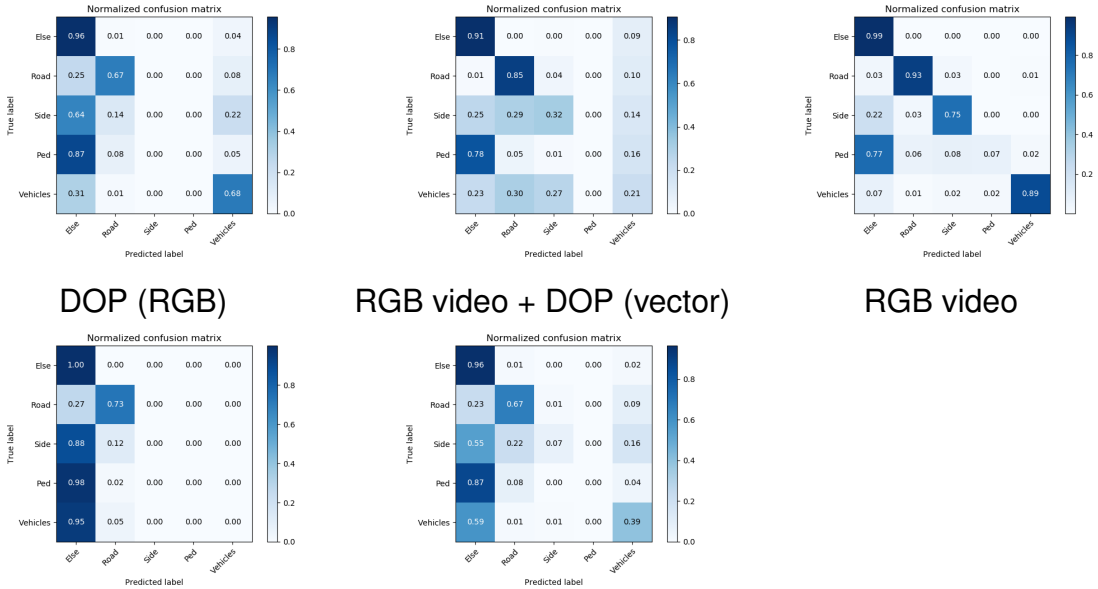
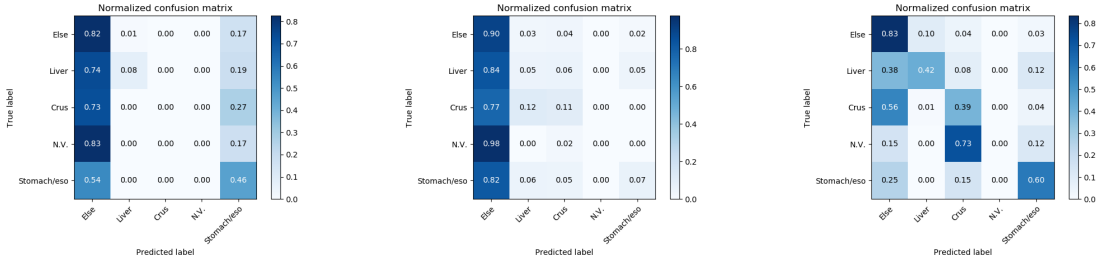


Figure 5.8: Confusion matrices of the different moddels trained on the VKITTI dataset. Those results are based on the test set.

The normalisation values are performed per horizontal row in the matrix. The blue colours are linked to the blue values of the normalisation. So interpreted horizontally how the true label is distributed over the predicted labels. In figure 5.8 the confusion matrices for each model trained on the VKITTI data are shown. The data used for generating the results are from the test data. The confusion matrices show similar outcomes as in the IOU. The smaller objects, like lampposts and traffic signs (Ped.) is mostly predicted in the else group. Overall the sidewalk (side) and lampposts and traffic signs (Ped.) are predicted very poorly. The RGB video model is doing the best overall when looking at the different classes because the highest values are scored to the correct class (except Ped.).

In figure 5.9, the confusion matrices for each model trained on the clinical data are shown. Compared to the other confusion matrices in of the VKITTI dataset, these results are inferior. Only the RGB video shows on the diagonal corresponding truelabel - predicted label outputs. For all the models, no Vagus Nerves were detected. Most of the pixels that were true label Vagus Nerve were predicted as ”Else” of as ”Crus”.



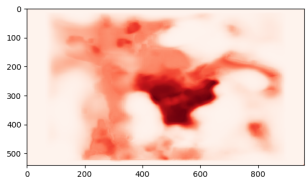
DOP (RGB) RGB video + DOP (vector) RGB video

RGB video + DOP (RGB) DOP (vector)

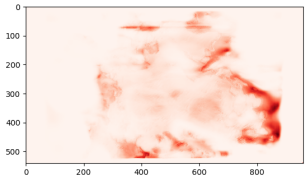
Figure 5.9: Confusion matrices of the different models trained on the clinical dataset. Those results are based on the test set.

5.4 Vagus nerve heatmap as an output of the models

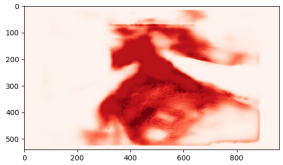
Visualisation for nerve detection



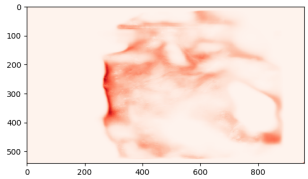
DOP (RGB)



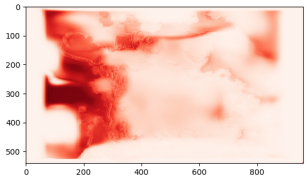
RGB video + DOP (vector)



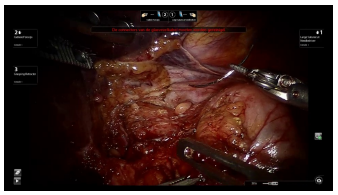
RGB video



RGB video + DOP (RGB)



DOP (vector)



Original image

Figure 5.10: Heatmaps of the vagus nerve which is given by the different networks/models on the same sample/test image. The input image with labels (including the Vagus Nerve) is given in figure 5.11

In the 6 subfigures of figure 5.10 are all the different outputs for the Vagus Nerve prediction visualized as heatmaps. A high probability is a more red color per pixel. There is no similarity visible between the images predicting the same structure and of the same image. The expected region is around the yellow line of the labelled ground truth visible in 5.11. Visual inspection shows the only image with a similar outcome is the RGB video model.

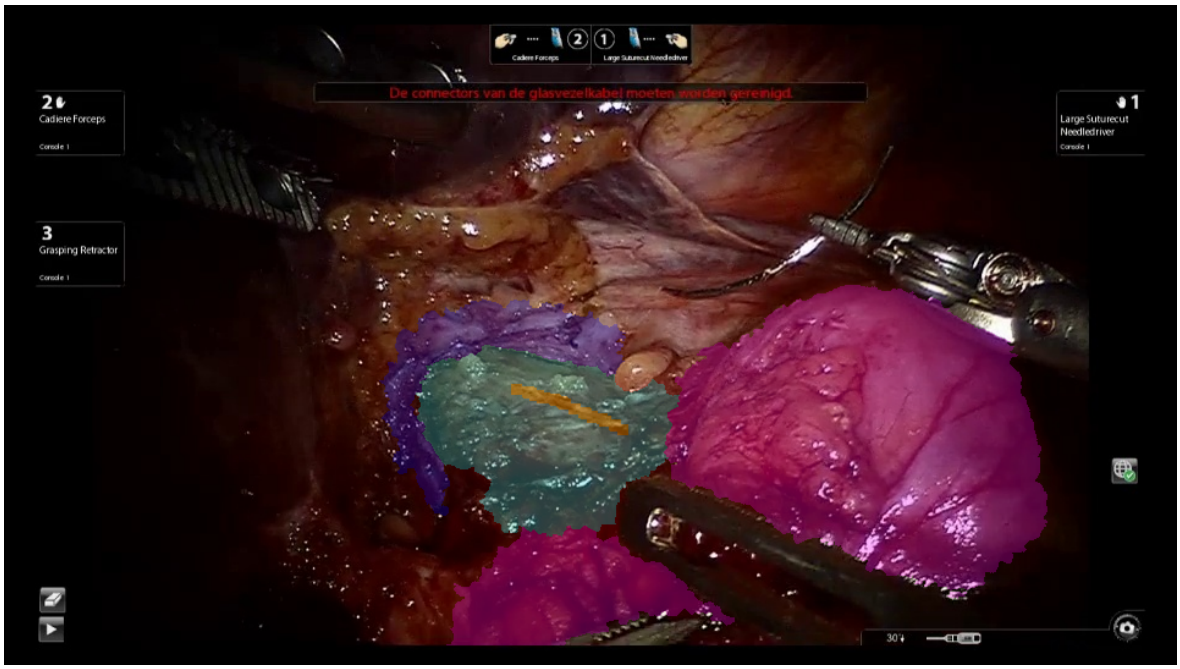


Figure 5.11: One of the images from the test dataset. The anatomical structures are visualised on top with transparant colors; yellow for vagus nerve, blue for crus, purple for stomach and light blue for esophagus. This is one of the images in the test dataset and used to visualize the outputs given in figure 5.10

5.5 Weighting of the model output value of the Vagus Nerve

Weighting the output value for the Vagus Nerve of the RGB model. This is performed with the formula 4.2 with different α (ranging from 1.0 till 1.5).

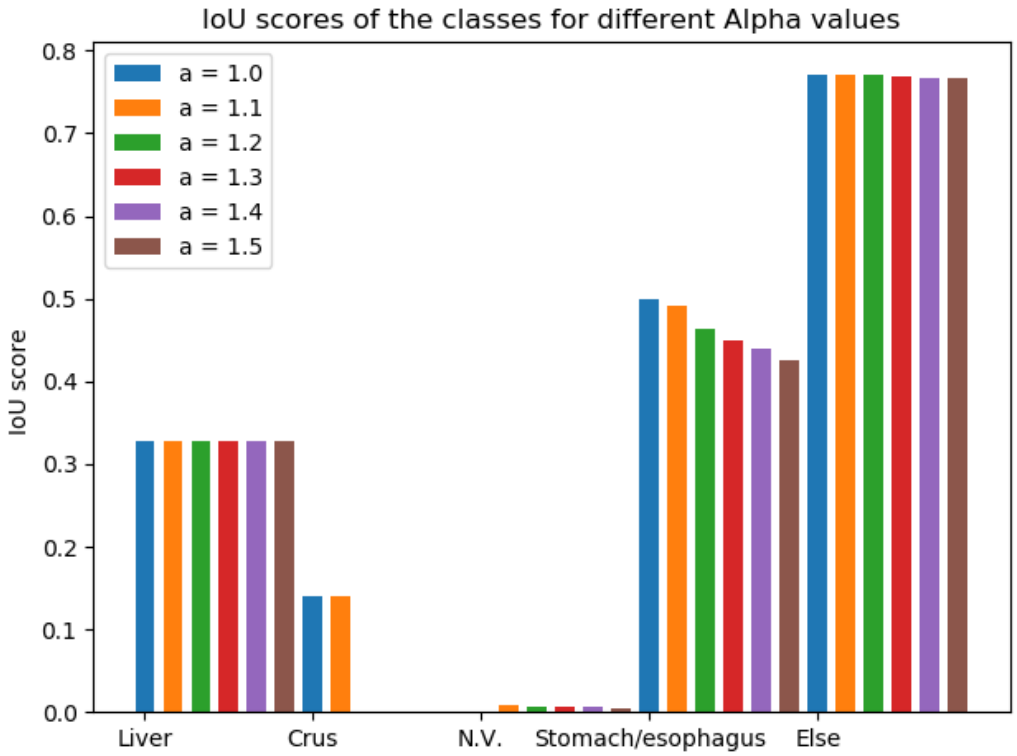
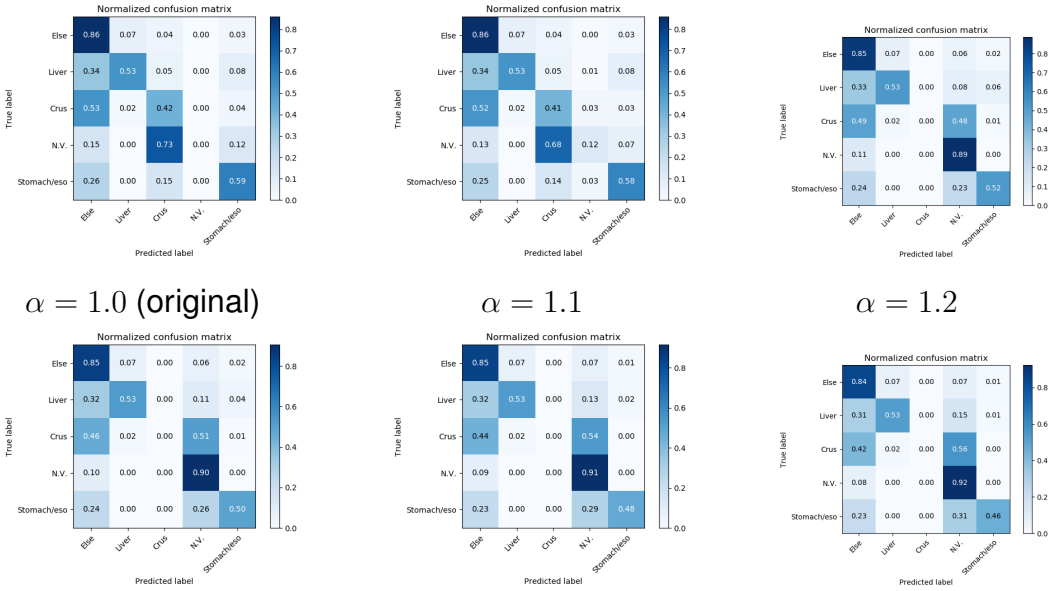


Figure 5.12: The IOU of the test data with different weigthing (α) of the Vagus Nerve output. Applied only on the RGB model.

The IOU in figure 5.12 shows no output for the Vagus Nerve with a normal model output ($\alpha = 1$). The Nerve becomes present when enhanced by 10% ($\alpha = 1.1$). Afther raising further the outcomes for *Stomach/Esophagus* further decline and the *Crus/Diaphragm* drops to zero, although the *Vagus Nerve* does not rise much. This is also visible in the confusion matrices in figure‘5.13. The values for Vagus Nerve make the biggest increase $\alpha = 1.2$, but the *Crus/diaphragm* vanishes. A further increase drops the label *Esosphagus/Stomach* and shows that the pixels probably are now labelled as Vagus Nerve.



$\alpha = 1.0$ (original) $\alpha = 1.1$ $\alpha = 1.2$

$\alpha = 1.3$ $\alpha = 1.4$ $\alpha = 1.5$

Figure 5.13: Confusion matrices of the test data with different weigthing (α) of the Vagus Nerve output. Applied only on the RGB model.

5.6 Summary of this chapter

In this chapter the results of the different trained networks are shown. For the IOU labels Else scored in both trained models high. The VKITTI trained model also showed useful outcomes for the label road. The smaller clinical dataset overall scored much lower. Numbers in IOU decreased also when the dense optical flow as added. Although the Dense Optical flow input still showed some performance in the IOU. For the clinical trained models, labels scored lower IOU's and the Vagus Nerve in the clinical dataset showed even zero as output in the IOU. Visual inspection of the RGB model on the clinical data suggested a a working model for some of the labels, but no visible Vagus Nerve in the outputs. This was also visible in the confusion matrices. The heatmap/probability map of the Vagus Nerve for the different models showed in visual inspection also only for the RGB model a plausible output. Even after heavier weighing the output of the Vagus Nerve output, to counter balance the class imbalance, the Vagus Nerve prediction in the IOU and confusion matrices rose slightly. Weigthing more the 20% did vanish the Crus label and higher weighing mostly lowered the outcomes for the label stomach/esophagus and showed no big increase in the Vagus label.

Chapter 6

Discussion and Conclusions

6.1 Discussion

The first hypothesis if Deep Learning algorithms can be used for anatomical structure recognition/segmentation, such as the Vagus Nerve on surgical video is yes. In our setup it was not possible to segment the Vagus Nerve. But, although the dataset is small, the stomach/oesophagus was possible to segmented by the simplest U-Net.

The network is able to create a probability/heat map per class-based of structures, but the clinical dataset contains too little data to give a reliable outcome of the nerve location. Though the VKITTI dataset trained network suggests that this might be possible with more data, at least for other structures such as the stomach/oesophagus.

The second hypothesis if the addition of movement (temporal information of the surgical video improves the segmentation of anatomical structures by a Deep Learning network is no. There is information that potentially can be used for better segmentation. Based on the results that if the only dense optical flow is used still ok segmentation results can be acquired. This suggests explicitly that there is spatial information available in the DOP images. The combination of frames (RGB) and the temporal information with a U-Net is not successful and lowers the outcome results. The visualisation as a heatmap for the probability that the Vagus Nerve is in the surgical field is still debatable. The outcome for this structure is little, and no clear conclusion can be made if this is clinically the right way to give this type of information back to the surgeon.

The visual output of the models can be presented to a surgeon as a heatmap, which answers the last research question partially. However, the usability as well as the performance or accuracy of the heatmap is highly questionable and should be further investigated.

Dataset and data

The first aim was to build a model that was able to segment images with a probability map per class as output, this in combination with the possibility to generate movement information. The second aim was to use this knowledge in building a clinical dataset. Although the datasets are in some ways comparable, they are also very different. The labels of the VKITTI dataset are available with a 100% ground truth because the full dataset is generated by a Game Engine. The labels of the clinical dataset are made by humans. This has many consequences. First vague labels such as the crus and Nervus Vagus/Vagus Nerve, can easily be false. So training becomes more difficult and if these ”errors” are present in the validation or test set the model is wrong. Also, the different tissues are difficult to distinguish from each other, which makes the dataset even more difficult to create but also difficult to train on. The easiest structure for labelling was the liver, so labelling is probably performed the best of all clinical labels. Also, training is performed easier. Also, this is one of the classes that are available. The problem of labelling errors/uncertainty will always exist in human-made datasets, but a good solution is to perform labelling by multiple people and do an average of all the people (possibly also with the exclusion of outliers). This makes a dataset more valuable.

Next difference is the size of the two datasets. Deep Learning relies on sufficient data to train (learn). The VKITTI dataset was 20 times larger than the clinical dataset. The biggest drawback is that labelling consumes a lot of time, when creating an new dataset. It is suggested that there is too little data in the clinical dataset to be able to train a network sufficient enough, at least the labels that are less frequent in the images. Another drawback that is the issue that both datasets have an imbalance of labels. For instance, the number of pixels with label Pedestrian (in fact the lampposts and road signs) compared to the label road. So the training of the label road is possibly faster due to the availability of the label. This can be solved by class weights. So a mistake in a less frequent available label can be weighted heavier weighted that a frequent label such as road. This is performed within the loss function. So the total loss of the model. Unfortunately, the package that is used for training, Keras, was not able to perform class weights to the loss function with the output dimensions that were used in this thesis.

So the label/class imbalance is probably a good explanation of high IOU for the better training of the class road and very low outputs for the other classes. The overall low outputs (in the IOU and the confusion matrix) could be a combination of class imbalance and the limited size of the dataset.

A good point for creating this dataset is diversity. The clinical data contains multiple types of fundoplication surgery. This makes training in a small dataset difficult, but if the dataset is large enough, the trained model will be more accurate in all types

of fundoplication surgeries. So increasing the dataset is highly recommended.

Training

The first tests with the training of the small clinical dataset on the algorithms overfitting were reached easily. The steps and epochs were reduced to the limited number that is mentioned in the method. As mentioned above the dataset was too small, although class prediction for the RGB (video) only was good on all the classes (except the Vagus Nerve).

The overtraining was also reached on longer training of the VKITTI dataset. The results when approaching overtraining were that a higher accuracy and low loss did not show a IOU that was above 0.10 for all the predicted labels other than one (other than else). This indicates the overtraining of one class, possibly due to class imbalance. Unfortunately, classes could not be weighted during training, so the chosen solution was shorter training to prevent this over training event. This happened much faster in the VKITTI data than the limited clinical dataset.

The clinical dataset was so small that training was watched manually, also to prevent overtraining. A critical note for the validation set is that three patients also for validation is minimal. A good solution is to increase the potential of this small dataset to use cross-validation. Unfortunately, this was not possible with the data generator that was designed by ourselves. This design problem also arose with the use of class weighting.

As written above, no classes weighting was implemented due to no support of Keras for these multiple output variables. However, other improvements should be considered to train these models better next time; implementation of cross-validation and image optimisation. Due to manual building, the data generators no implementation of rotation, sheer and cropping was applied.

Also, these implementation issues created another potential problem, the a priori chance that a particular pixel has a specific class. So the road is always beneath, in the middle of the image, so the model can be falsely trained that there is always road there, this can be prevented by a, shifting, rotating and cropping the image, when preparing the data for deep learning.

The most significant recommendation that follows from this design is that, although the outcome presentation is favourable, the use of the Deep Learning package in python does not have the right tools for this type type of labelling. Next time, the dataset should be labelled differently with also different type of outcomes to be able to use the full potential of the possibilities within Keras package. Alternatively, the data generator, loss function/class weighting should be built. It should also be considered to use a labelling standard for surgical video to be able to use the full

potential from different Deep Learning networks in the future and usability for other research subjects.

Network and Dense Optical Flow

It can easily be spotted that the addition of DOP lowers the results of the Deep Learning networks. It can be a possibility that the addition of two different types to a convolutional network eighter one of the two types of data is noise for the model. This is visible that the addition of DOP RGB and DOP vector to the RGB frames the IoU is lower for all classes. This in contrast that the DOP vector and DOP RGB in the KITTI dataset were still ok in the segmentation for the label road. The answer might be in the way the first step of the network uses and combines the data.

Good to mention is that it is highly feasible that segmentable information is available within the DOP information when looking at the outcomes of the VKITTI trained models. Because segmentation was still possible even with IOU outcomes higher than 0.5.

Other network structures that still combine the dense optical flow with images should be further investigated. Detach the RGB information from the motion seems necessary in the first part of the network. This was also visible in the two independent encoding paths (one for RGB and one for the Flow data) and one decoding path to come till one semantic segmentation outcome in the paper of Rashed et al. (2019). [43] Because this approach seems to work other suggestions with multi encoders arise and should be further investigated.

There is also a difference in dense optical flow data when looking at the type of movement that is present in the two datasets, although no differences in outcome based on the different types is visible. The VKITTI dataset is a dashcam simulated video (the camera is moving), so the most significant movement is always around the edges. The clinical dataset is a helicopter view video, so roughly only the anatomical structures move apart from some little movement of the camera itself. It did not became clear in this research what the effect of the different types of movement was on the semantic segmentation.

Vagus detection

Nerve prediction is shown in the results as an low outcomes in the confusion matrices and IOU metrics. Though looking at the visualisation in the heatmap suggests the region of the Nerve in the right area. So the nerve prediction is not wrong, but the other classes score a higher prediction value. In the IOU and confusion matrix this is not seen. Most of the pixels that were true label Vagus Nerve, were predicted

as ”Else” of as ”Crus”. But not the surrounding oesophagus or stomach. Which you could expect when it is common surrounding tissue! So it can distinguish slightly the difference.

After weighting the output heavier of the Vagus Nerve prediction only a slight increase in performance is seen. But this increase is little and does not increase substantial after weighting the output more than 20%. Suggesting there is too little Vagus Nerve data to learn properly. Concluding it is a difficult/impossible class to learn based on this small data is not possible.

In our view no performance parameter is needed for the heatmap of the vagus region. The IOU and confusion matrices represent well the performance of the model. Although showing the output in a heatmap suggests a lot of the performance of the model, but interpretation is still subjective. Superimposing and heatmaps, are a way of visualisation that might has an added value for the clinic of those type of modeloutput. But models need to be far more precise to rely on during surgery.

Potential future of AI and Surgery

The future of AI and surgery probably will grow. Although, the near future might not be in the direct clinical decision support like the models that are shown in this thesis. More likely are AI models that track, time and benchmark surgical performance. Measuring parameters such as the duration of steps during procedures might give basic insight in the learning curve of residents or surgeons among each other. Currently this is something that is only performed by a human with a stopwatch in research setting. With the help of AI this could be something monitored always and for everybody. Timing duration is only one parameter, but the options are endless like; tracking movement of tools on screen, track the onset of bleeding. The actual surgery, performed by a robot without a human interaction is a step further. Much much further. This will not only need the detection performed by AI but also the decision making and the performing the needed action. It is probably not going to happen, although the other options of AI within surgery still remain. In conclusion the possibilities are endless. No matter what, this will change the way how we educate surgery and how we perform surgery.

6.2 Conclusions and recommendations

Conclusions

A first step is made in the ultimate goal of assisting or help training residents or surgeons. The way the results can be displayed, data is labelled, and output is generated can be further developed with this labelling and training approach. Also, this research showed that a pixel-wise segmentation of anatomical surgical video is possible with a U-NET.

These results proof and show further possibilities of the use of semantic segmentation networks, although other U-Net setups might be a better solution in combining the two different data types. Data types like *Dense Optical Flow* and spational data.

The use of how to combine the video frame data with the DOP within a Deep Learning network should be further investigated, but show a promising future in Deep Learning and the medical field.

Recommendations

In order to perform more accurate result, the clinical dataset should be increased to similar amounts of images as the VKITTIdataset. The class imbalance within the data is not solvable, but the weight of an error that is made in a less available label could be accounted much heavier than a highly available label error. This is called class weighting and might improve learning outcomes of labels such as the Vagus Nerve. This is not possible with Keras at the time of writing with this output. A (tailored) loss function that allows class weighting is highly recommended in this type of semantic segmentation.

Both datasets itself could also be further optimised by a more complex data generator. Options that enlarge the limited data such as tilting, mirroring and cropping could further improve the trainingdata.

Create or use a labelling standard for semantic segmentation/machine learning. In that case, the usage of other models of clinical data would be more translational and probably also easier to apply in the hospital.

The last recommendation is more (complex) networks testing, like the two encoding arms networks (described by Rashed et al. (2019) [43]). But also other complex two arm/multi-channel input structures.

Bibliography

[1] R. B. Yates, B. K. Oelschlager, and C. A. Pellegrini, *Ch 42 Gastroesophageal Reflux Disease and Hiatal Hernia*, 2016. [Online]. Available: http://dx.doi.org/10.1016/B978-0-323-29987-9.00042-4

[2] "nerve — Taber's Medical Dictionary." [Online]. Available: https://www.tabers.com/tabersonline/view/Tabers-Dictionary/730433/all/nerve

[3] S. R. Evans and E. A. David, "Laparoscopic Nissen Fundoplication," in *Surgical Pitfalls*. Elsevier Inc., 1 2009, pp. 175–185.

[4] M. Schuhmacher, "Autonomous anatomical structure recognition using machine learning," Tech. Rep., 2018.

[5] J. T. Elliott, A. V. Dsouza, S. C. Davis, J. D. Olson, K. D. Paulsen, D. W. Roberts, and B. W. Pogue, "Review of fluorescence guided surgery visualization and overlay techniques," *Biomedical Optics Express*, vol. 6, no. 10, p. 3765, 10 2015.

[6] R. Yamashita, M. Nishio, R. Kinh, G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology." [Online]. Available: https://doi.org/10.1007/s13244-018-0639-9

[7] "A Beginner's Guide To Understanding Convolutional Neural Networks Part 2 Adit Deshpande Engineering at Forward — UCLA CS '19." [Online]. Available: https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/

[8] "Complete Guide of Activation Functions - Towards Data Science." [Online]. Available: https://towardsdatascience.com/complete-guide-of-activation-functions-34076e95d044

[9] B. J. Erickson, "Deep learning and machine learning in imaging: Basic principles," in *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*. Springer International Publishing, 1 2019, pp. 39–46.

[10] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation," 4 2017. [Online]. Available: http://arxiv.org/abs/1704.06857

[11] E. Shelhamer, J. Long, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Tech. Rep.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Springer Verlag, 2015, pp. 234–241.

[13] "What is Optical Flow and why does it matter in deep learning mc.ai." [Online]. Available: https://mc.ai/what-is-optical-flow-and-why-does-it-matter-in-deep-learning/

[14] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "VirtualWorlds as Proxy for Multi-object Tracking Analysis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December. IEEE Computer Society, 12 2016, pp. 4340–4349.

[15] MATLAB, *version 9.4.0.813654 (R2018a)*. Natick, Massachusetts: The MathWorks Inc., 2018.

[16] "Intersection over Union (IoU) for object detection - PyImageSearch." [Online]. Available: https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/

[17] F. Baldi, "PPI-Refractory GERD: an Intriguing, Probably Overestimated, Phenomenon," *Current Gastroenterology Reports*, 2015.

[18] R. Yadlapati, M. F. Vaezi, M. F. Vela, S. J. Spechler, N. J. Shaheen, J. Richter, B. E. Lacy, D. Katzka, P. O. Katz, P. J. Kahrilas, C. P. Gyawali, L. Gerson, R. Fass, D. O. Castell, J. Craft, L. Hillman, and J. E. Pandolfino, "Management options for patients with GERD and persistent symptoms on proton pump inhibitors: recommendations from an expert panel," *American Journal of Gastroenterology*, vol. 113, no. 7, pp. 980–986, 2018.

[19] K. Yolsuriyanwong, E. Marcotte, M. Venu, and B. Chand, "Impact of vagus nerve integrity testing on surgical management in patients with previous operations with potential risk of vagal injury," 2018.

[20] S. van Rijn, N. F. Rinsma, M. Y. A. van Herwaarden-Lindeboom, J. Ringers, H. G. Gooszen, P. J. J. van Rijn, R. A. Veenendaal, J. M. Conchillo, N. D. Bouvy, and A. A. M. Masclee, "Effect of Vagus Nerve Integrity on Short and Long-Term Efficacy of Antireflux Surgery." *The American journal of gastroenterology*, vol. 111, no. 4, pp. 508–15, 4 2016. [Online]. Available: http://insights.ovid.com/crossref?an=00000434-201604000-00021http://www.ncbi.nlm.nih.gov/pubmed/26977759

[21] S. C. Liao, H. Z. Yeh, C. W. Ko, H. C. Lien, and C. S. Chang, "The management of gastroesophageal reflux disease: An update," pp. 381–390, 2010.

[22] G. P. Kohn, R. R. Price, S. R. DeMeester, J. Zehetner, O. J. Muensterer, Z. Awad, S. K. Mittal, W. S. Richardson, D. Stefanidis, and R. D. Fanelli, "Guidelines for the management of hiatal hernia," *Surgical Endoscopy*, vol. 27, no. 12, pp. 4409–4428, 12 2013. [Online]. Available: http://link.springer.com/10.1007/s00464-013-3173-3

[23] N. K. Altorki, D. Yankelevitz, and D. B. Skinner, "Massive hiatal hernias: The anatomic basis of repair," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 115, no. 4, pp. 828–835, 4 1998. [Online]. Available: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L28196593%0Ahttp://dx.doi.org/10.1016/S0022-5223(98)70363-0%0Ahttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=00225223&id=doi:10.1016%2FS0022-5223%2898%2970363-0&atitle=Massive+hiat

[24] P. J. Kahrilas and I. Hirano, "Diseases of the Esophagus," in *Harrison's Principles of Internal Medicine*, 19th ed. McGraw-Hill Global Education, 2016, ch. 347, pp. 1900 – 1911.

[25] N. Noorbakhsh-Sabet, R. Zand, Y. Zhang, and V. Abedi, "Artificial Intelligence Transforms the Future of Healthcare," *The American Journal of Medicine*, 2019.

[26] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," in *Lecture Notes in Computational Vision and Biomechanics*, 2018.

[27] F. G. Zanjani, "Improving Semantic Video Segmentation by Dynamic Scene Integration," 2010.

[28] "Collaborative autonomy in the Operating Room: Verb Surgical and Democratized Surgery Technology and Operations Management." [Online]. Available: https://digital.hbs.edu/platform-rctom/submission/collaborative-autonomy-in-the-operating-room-verb-surgical-and-democratized-surgery/

[29] R. Câmara and C. J. Griessenauer, “Anatomy of the Vagus Nerve,” in *Nerves and Nerve Injuries*, 2015.

[30] K. L. Moore, D. F. Arthur, and A. M. R. Agur, “Stomach,” in *Clincally orriented anatomy*, 6th ed. Philadelphia: Lippincott Williams and Wilkins, 2010, ch. Chapter 2, pp. 230 – 238.

[31] ——, “Vagus Nerve (CN X),” in *Clincally orriented anatomy*, 6th ed. Philadelphia: Lippincott Williams and Wilkins, 2010, ch. 9 Summary, pp. 1073 – 1078.

[32] H. J. Binder, “Organisation of the gastrointestinal system,” in *Medical Physiology, 2e Updated Edition: with STUDENT CONSULT Online Access*, second edi ed., W. F. Boron and E. L. Boulpaep, Eds. Elsevier Health Sciences, 2012, ch. 41, p. 1350.

[33] A. C. Mertens, R. C. Tolboom, H. Zavrtanik, W. A. Draaisma, and I. A. M. J. Broeders, “Morbidity and mortality in complex robot-assisted hiatal hernia surgery: 7-year experience in a high-volume center.” *Surgical endoscopy*, 10 2018. [Online]. Available: http://link.springer.com/10.1007/s00464-018-6494-4http://www.ncbi.nlm.nih.gov/pubmed/30350095

[34] K. Seeras and M. A. Siccardi, *Nissen Fundoplication (Anti-reflux Procedure)*. StatPearls Publishing, 2 2019. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/30137806

[35] A. S. Ahuja, “The impact of artificial intelligence in medicine on the future role of the physician,” *PeerJ*, vol. 2019, no. 10, 2019.

[36] “What Is Machine Learning? — How It Works, Techniques & Applications - MATLAB & Simulink.” [Online]. Available: https://nl.mathworks.com/discovery/machine-learning.html

[37] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” pp. 436–444, 5 2015.

[38] B. M. ter Haar Romeny, “A deeper understanding of deep learning,” in *Artificial Intelligence in Medical Imaging: Opportunities, Applications and Risks*. Springer International Publishing, 1 2019, pp. 25–38.

[39] “Kaggle: Your Machine Learning and Data Science Community.” [Online]. Available: https://www.kaggle.com/

[40] Y. Lecun, E. Bottou, Y. Bengio, and P. Haffner, “Gradient-Based Learning Applied to Document Recognition,” Tech. Rep., 1998.

[41] “OpenCV: Optical Flow.” [Online]. Available: https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html

[42] G. Farnebäck, “Two-Frame Motion Estimation Based on Polynomial Expansion,” Tech. Rep. [Online]. Available: http://www.isy.liu.se/cvl/

[43] H. Rashed, S. Yogamani, A. El-Sallab, P. Krizek, and M. El-Helw, “Optical Flow augmented Semantic Segmentation networks for Automated Driving,” 1 2019. [Online]. Available: http://arxiv.org/abs/1901.07355

[44] “Introduction to Motion Estimation with Optical Flow.” [Online]. Available: https://nanonets.com/blog/optical-flow/

[45] F. Chollet and and others, “Keras,” 2015. [Online]. Available: https://keras.io

[46] Computer software., “Anaconda Software Distribution.” 2016. [Online]. Available: https://anaconda.com

[47] D. Vázquez, J. Bernal, . F. Javier Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, . M. Drozdzal, and A. Courville, “A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images,” Tech. Rep.

[48] P. Diez, “Introduction,” in *Smart Wheelchairs and Brain-computer Interfaces: Mobile Assistive Technologies*. Elsevier, 1 2018, pp. 1–21.

Appendix A

Overview of IoU output metrics of the different models

Table A.1: The IOU results of the different models trained, validated and tested with the virtual KITTI dataset. The IOU results are created with the test data.

Model name\Class	Road	Side	Ped (road signs)	Car	Else
DOP (rgb)	$6.493*10^{-1}$	$4.620*10^{-4}$	0.0	$2.567*10^{-1}$	$6.776*10^{-1}$
RGB + DOP (vector)	0.8030	0.1751	0.0000	0.0564	0.7878
RGB	0.9200	0.4878	0.0706	0.7822	0.8586
RGB + DOP (rgb)	$7.214*10^{-1}$	$2.417*10^{-7}$	0.0000	0.0000	$6.624*10^{-1}$
DOP (vector)	0.6421	0.0573	0.0000	0.1605	0.6841

Table A.2: The IOU results of the different models trained, validated and tested with clinical dataset. The IOU results are created with the test data.

Model name\Class	Liver	Crus	Vagus Nerve(s)	Stomach/ Esophagus	Else
DOP (rgb)	0.0707	0.0000	0.0000	0.2313	0.6720
RGB + DOP (vector)	0.0379	0.0437	0.0000	0.0596	0.6969
RGB	0.3285	0.1405	0.0000	0.4984	0.7710
RGB + DOP (rgb)	0.0025	0.0967	0.0000	0.0855	0.7135
DOP (vector)	0.0000	0.0000	0.0000	0.1207	0.7316

Appendix B

Overview of the training of the models

Onderschriften corresponderen niet met daadwerkelijke resultaten. Afbeeldingen zijn wel correct.

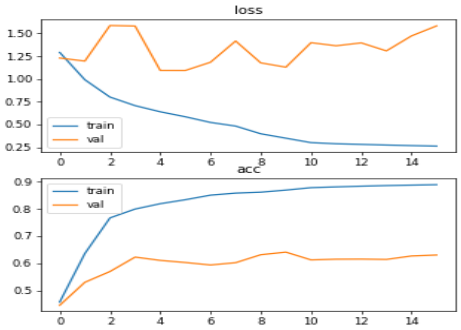


Figure B.1: Training input: DOP (RGB) on the VKITTI dataset. Saved Epoch: 5 based on lowest validation loss, accuracy 0.833, loss 0.584, validation accuracy 0.604, validation loss 1.0929

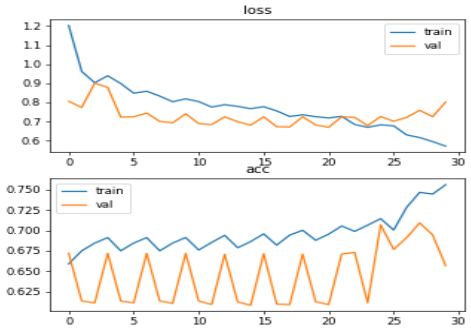


Figure B.2: Training input:DOP (RGB) on the clinical dataset. Saved Epoch: 29 based on highest acc, accuracy 0.756, loss 0.572, validation Accuracy 0.657, validation loss 0.802

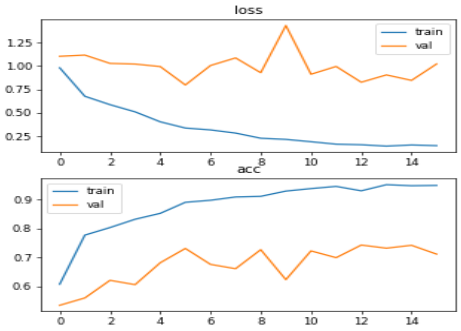


Figure B.3: Training input: DOP (RGB) + RGB (vector) on the VKITTI dataset. Saved Epoch: 5 based on lowest validation loss, accuracy 0.891, loss 0.337, validation accuracy 0.731, validation loss 0.797

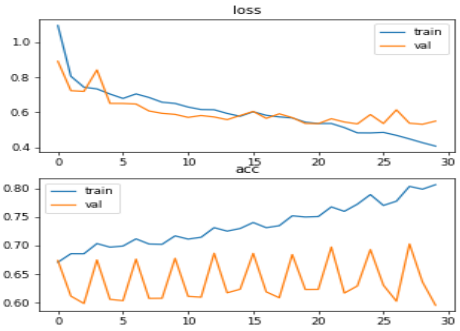


Figure B.4: Training input: DOP (RGB) + RGB (vector) on the clinical dataset. Saved Epoch: 29 based on highest accuracy, accuracy 0.806, loss 0.408, validation Accuracy 0.596, validation loss 0.550

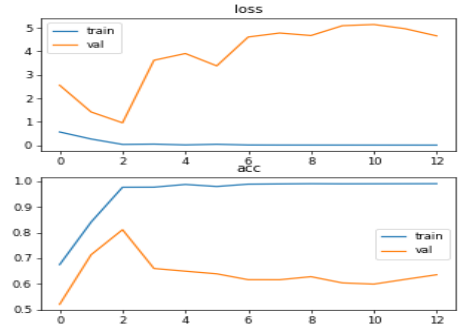


Figure B.5: Training input: RGB on the VKITTI dataset. Saved Epoch: 2 based on lowest validation loss, accuracy 0.976, loss 0.043, validation accuracy 0.811, validation loss 0.961

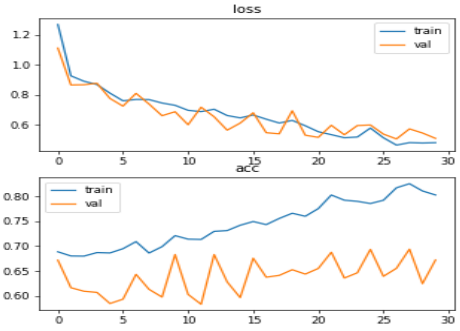


Figure B.6: Training input: RGB on the clinical dataset. Saved Epoch: 27 based on highest accuracy, accuracy 0.826, loss 0.483, validation Accuracy 0.694, validation loss 0.575

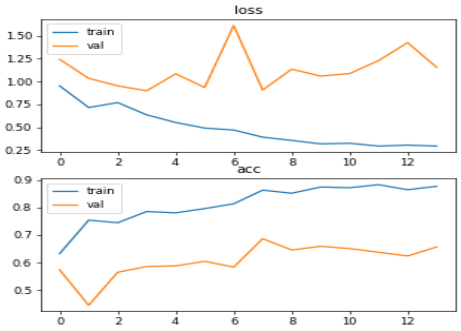


Figure B.7: Training input: RGB frames + DOP (RGB) on the VKITTI dataset. Saved Epoch: 3 based on lowest validation loss, accuracy 0.786, loss 0.636, validation accuracy 0.585, validation loss 0.899

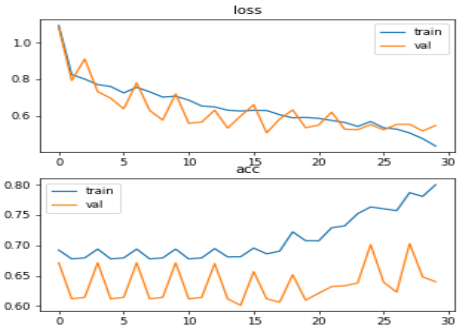


Figure B.8: Training input: RGB frame + DOP (RGB) on the clinical dataset. Saved Epoch: 29 based on highest accuracy, accuracy 0.800, loss 0.434, validation Accuracy 0.640, validation loss 0.546

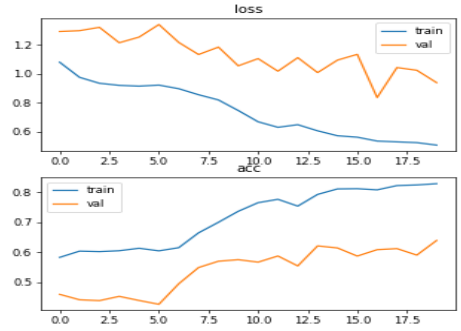


Figure B.9: Training input: DOP (vector) on the VKITTI dataset. Saved Epoch: 16 based on lowest validation loss, accuracy 0.807, loss 0.534, validation accuracy 0.608, validation loss 0.836

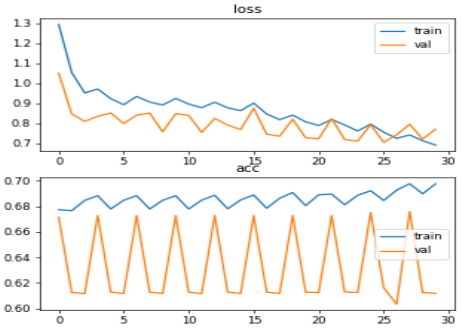


Figure B.10: Training input: DOP (vector) on the clinical dataset. Saved Epoch: 27 based on highest accuracy, accuracy 0.698, loss 0.742, validation Accuracy 0.676, validation loss 0.796

Appendix C

Study Protocol

Study protocol: Machine Learning visualisation algorithms on anti-reflux surgery video

Prof. Dr. I.A.M.J. BROEDERS *Surgeon/Professor of robotics and minimally invasive surgery*
J.R. ABBING BSc. *Investigator/Student*

I. SUMMARY

Nederlands:
In de gezondheidszorg is altijd een drijfveer de zorg te innoveren voor betere uitkomsten zonder de kosten te doen stijgen. Een nieuwe stap van de digitalisatie in de gezondheidszorg zou de vraag naar technologieën als artificial intelligence (AI, kunstmatige intelligentie) en big data waarschijnlijk doen stijgen. Binnen de chirurgie zijn dan ook al de eerste toepassingen gemaakt en getest. Deze technologie bevat ook mogelijkheden voor ons probleem. Bij anti-reflux chirurgie is er een risico tot het doornemen of beschadigen van de Nervus Vagus. Dit letsel bij dit type operaties is wordt dan ook geschat op 20%. Een oplossing en tevens ons doel is het creëren van een AI toepassing (Deep learning) die de zenuw én andere anatomische structuren kan detecteren op basis van alleen de laparoscopische videobeelden. Het verwachtte resultaat is een toepassing die de zenuw kan detecteren met een hoge mate van nauwkeurigheid. De te gebruiken laparoscopische videobeelden worden retrospectief gebruikt en alleen indien deze volledig geanonimiseerd zijn; daarnaast worden ook geen andere parameters/informatie uit het elektronisch patienten dossier verkregen.

English:
In healthcare, the thrive to improve patient outcomes without raising the cost has always been the case. A new step in the digitisation in health care might support this need through big data and similar technologies like artificial intelligence (AI). Also, in surgery, the first AI applications are build and tested. This might hold a solution for our problem. During anti-reflux surgery, there is a potential risk of Nervus Vagus injury. Here the rate of unintended Nervus Vagus injury is estimated around 20%. A solution and our goal is to create an AI tool (deep learning) that can detect the Nervus Vagus and other anatomical structures in those surgical videos. The expected result is an AI tool that can detect with a certain accuracy the Nervus Vagus, but this also depends on the size of the dataset. The video data is used retrospectively and only used if it is fully anonymised; also no other information is obtained from the patient record systems.

II. INTRODUCTION

Artificial Intelligence for healthcare applications: In healthcare, the thrive to improve patient outcomes without raising the cost has always been the case. A new step in the digitisation in health care might support this need through big data and similar technologies like artificial intelligence (AI). AI tends to improve on diagnostics, patient therapy, prevention and support health care in making clinical decisions. A subtype of AI is machine learning. It can find correlations, associations, segmentation and generate new insights in very large amounts of data. AI is used in the automotive, finance and smart homes. In medicine, the first clinical setups show their great value; node detection in X-ray images, the prediction of outcomes in infectious diseases and ECG arrhythmia detection. Deep learning is also a type of AI and machine learning but relies on a small infrastructure which mimics the brain infrastructure. It is called deep because of stacked layers with multiple artificial 'neurons' that can be trained with existing data to make predictions or classifications. This is achieved by *learning* based on prelabelled data [1]

Due to the enormous variation between observed data (patients), the other regular learning methods are not sufficient anymore (i.e. selection on only colour differences). So the step from machine learning to deep learning is established. Deep learning can make automated predictions on very large complex datasets. [2]

Also, in surgery, the first AI applications are built. An example is the previous work from M. Schuhmacher at the Meander Medical Center. He showed that surgical video could be used for object detection. The accuracy was acceptable (43.7%), but improvement for clinical usage would be necessary as also stated in his thesis. He used a CNN (convolutional neural network, the YOLOv2 network) for autonomous structure recognition in the lower abdomen. Suggestions he made were; more training data, more complex network architectures like long short-term memory networks (LSTM). An LSTM is a recurrent neural network (RNN) which adds a specific 'memory' to the algorithm. In contrast, CNN does not have this 'memory'.[3] In this proposed study is to test new network structures on a clinical visualisation problem during anti-reflux surgery.

Anti-Reflux Surgery; GERD and Hiatal Hernia diaphragmaticus

Gastroesophageal reflux disease (GERD) is considered a benign condition of the stomach and oesophagus.[4] The primary medical treatment of GERD is the use of proton-pump inhibitors (PPI's), though a 10 to 40% of the patients remains unresponsive [5]. Surgical treatment is a second treatment option.
When PPI treatment does not show results in proven GERD, the recommended treatment is a fundoplication. Even if no hiatal hernia diaphragmaticus (HHD) is present, but PPI treatment does not work a fundoplication is recommended. [6]However, there is a potential risk of Nervus Vagus injury in fundoplications with HDD repair. [7] More important, Nervus Vagus (Nervus Vagus) injury has a significant negative effect on the reflux control postoperative and a significantly higher redo rate compared when there is no vagus injury post surgery[8]. Research of Van Rijn et al. (2016) ([8]) reported an incidence of 20% on unintended vagus injury. It should be mentioned that this long-term follow-up data of surgeries were collected between 1990 and 2000. Back then, the laparoscopic video systems were not as good as today. In this cohort of vagus injury (the study of Van Rijn et al. (2016)) over 50% had redo surgery and most of them because of recurrent reflux problems. Better knowledge per patient of the location

during surgery or visualisation of this nerve might improve the outcomes.

Due to a dysfunctional closure of the lower oesophagal sphincter duodenal gastric material can enter the oesophagus and even higher anatomical structures. This reflux can cause apart from discomfort, damage and inflammation of those structures. Untreated, the inflammation and tissue changes can lead to aspiration, Barrett's oesophagus, stricture, esophagitis or an adenocarcinoma. A *higher* incidence in GERD is found in patients who have a hiatal hernia (HHD, hiatal hernia diaphramatica), obesity or delayed gastric emptying.[9]

Due to change of anatomy, a HHD reduces the functionality of the lower oesophagal sphincter (LES) which results in possible entering of stomach fluids into the oesophagus. An HHD is a protrusion of anatomical structures (other than the oesophagus) into the thoracic cavity due to a widened hiatus diaphragmaticus.[10] This causes the symptoms; pain, heartburn, bleeding, dysphagia, weight loss, vomiting and regurgitation.[11]

Those GERD-like symptoms are strongly related with the HHD but are not necessarily present with every HHD. With a hiatal hernia, the stomach can migrate partially or entirely to the thoracal cavity. An HHD has four different subtypes anatomically (see figure 1). The most common one is type 1 and does not imply a non-functional LES. Though non-functionality is also very size-dependent. A type 2 to 4 is likely to cause GERD symptoms. In type 2, the gastroesophageal junction is in the abdominal cavity although the gastric fundus slides into the hiatal hernia. In a type 3 HHD, the fundus of the stomach and the gastroesophageal junction are located in the thorax cavity instead of in the abdominal cavity. A type 4 (not visualised in figure 1) other anatomical structures migrate cranially to the hiatal hernia.[4]

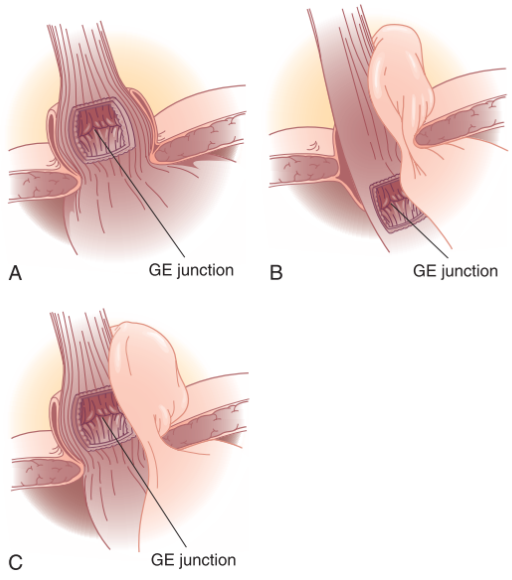


Figure 1: Type 1, 2 and 3 of hiatal herniations. **A** is a type 1 hernia (sliding hernia). **B** is a type 2 hernia (rolling hernia). **C** is a combination type of the type 1 and 2 (mixed hernia).[4]

The most common hiatal hernia is type 1; it covers 95% of all hiatal hernias. The other three subtypes together make the other 5%. The common symptoms of type 1 are the presence of GERD/reflux. [12] The other subtypes present themselves, on the other hand, more frequently with obstructive symptoms. [4] Also in type 1 hiatal hernia without reflux disease is also considered as no indication for surgery [10].

A. *Visualisation with Artificial Intelligence*

A logic step for improvement of this intervention is a visualisation of the (path of the) Nervus Vagus perioperative to decrease the Nervus Vagus injury and improve outcome. See figure 2 , 3 and figure 4 for the Nervus Vagus anterior and the Nervus Vagus posterior during surgery. [13] Complications in reoperations are frequent, although the risk of complication is lower in expertise centres. [14, 15] Hashimi et al. (2015) state that complications occur twice as much in redo surgery. [14] Because the visualisation of the nerve might be a step for improvement of the surgical outcome a high-end visualisation algorithm is suggested. Previous work on using Machine Learning on anatomical images (conducted in the Meander Medical Center) was done by Michiel Schuhmacher. His work showed that the concept of object detection worked on surgical videos. [3] This Machine learning approach might hold a solution in the Nervus Vagus visualisation problem.

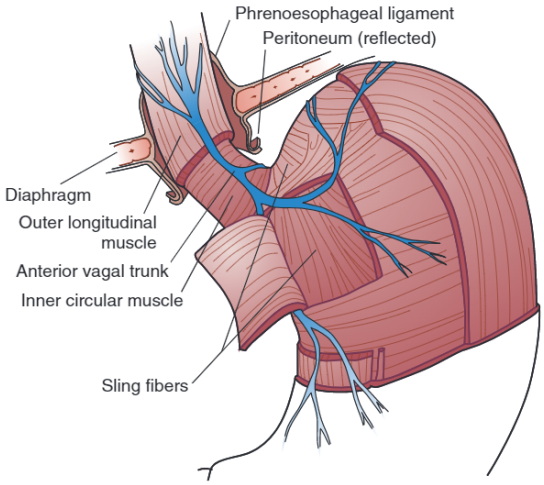


Figure 2: A drawing of the stomach with the anterior vagal trunk (*Nervus Vagus*) .[4]

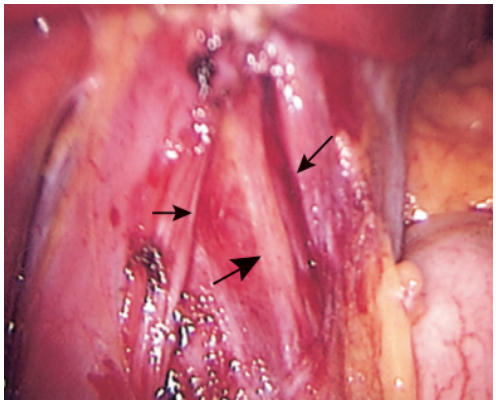


Figure 3: The big arrow points at the anterior Nervus Vagus surrounded by the crura left and right (smaller arrows). [13]

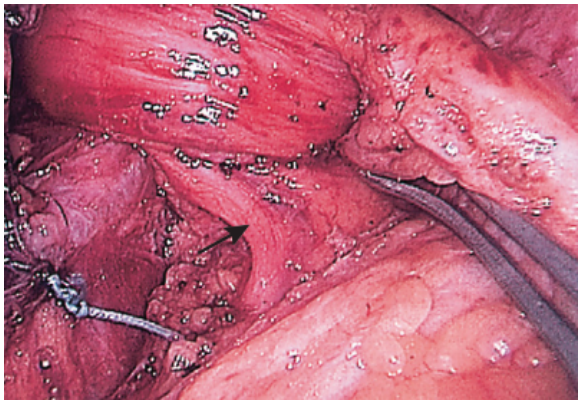


Figure 4: Here the arrow points at the posterior trunk of the Nervus Vagus, with a clearly visible esophagus on top op it. [13]

B. Objectives

Primary objective: To create a tool, based on deep learning, that can visualise the different anatomical structures such as the oesophagus, Nervus Vagus, diaphragm and liver.

Secondary objective: To test different artificial deep learning network structures that can visualise different classes (i.e. visualise only the Nervus Vaguss). Different inputs other than video frames (stills) but also the movement of pixels using as input in order to probably increase the accuracy.

III. Study population

A. study population

population base One subject group of laparoscopic fundoplication surgery between 01-01-2018 and 31-12-2019. No information other than the anonymised video is obtained from the EPD. The fundoplication surgery data are from type 1 to 4 The distinction between robotic surgery and conventional laparoscopic surgery can be made based on the video data itself. Datasets for deeplearning puroses need to be significant to be able to *'learn'* the network. Also, the dataset can be as diverse to learn from (anatomical) deviations. The estimated number of patients needed is 80 fudoplication video's, based on the fact that the surgical videos have multiple segments where the nerve will be visible. Those separate events can all be used as 'learn data' and will keep the number of needed surgical video's low.

Inclusion criterea are

- fundoplication surgery; robotic and conventional laparoscopy
- Had surgery between 01-01-2018 and 31-12-2019

Exclusion criterea

- fundoplication video data that cannot be anonymized (on-screen text, patient data or date)

B. Collecting the video data

1) Patient had surgery
2) A (legally) access to the patient file (i.e. the surgeon or assistant collects the video of the procedure and analyses if the video needs to be excluded).
3) The included video is uploaded to a disk drive, which is secured. (Recommened encryption is used op the IA department, if they do not have a recommended encryption software Veracrypt or Bitlocker is used).
4) The data/disk drive is handed over to the research group.

IV. Methods

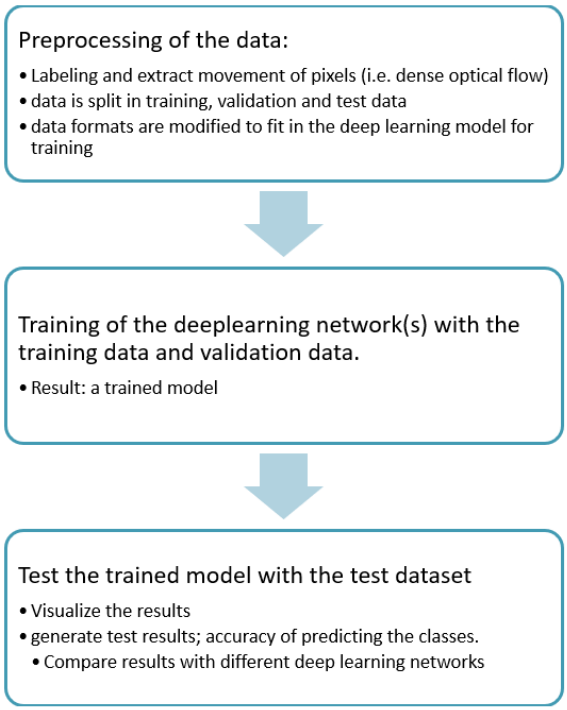


Figure 5: Study procedure

A. Main study parameters/end points

Main study parameters are the accuracy parameters of predicting classes in the images. Classes like the stomach, oesophagus, Nervus Vagus, diaphragm and liver. Secondary parameters will be the difference in the accuracy of these structures between the different Deep learning networks.

B. study procedures

There is no interference or change to interventions. Data is collected retrospective. After data is received by the research group, the dataset is modified for deep learning. This is achieved by selecting video frames and label these images manually. This labelling is drawing regions on the video frames which belong to a certain anatomical structure. These labels are the ground truth of what a certain pixel is. For example, a pixel, or set of pixels, belongs to the stomach. An example of labelling and the output from a trained network is given in figure **??** Also, extra information from frame-to-frame is generated. So-called dense optical flow, the estimated movement of a pixel colour to another region in the next frame.[16] The workflow is visualised in figure 5. An example of a labelled image of a dataset (Virtual Kitti dataset[17]) and the predicted output generated by an modified UNET[18] can be seen in figure 6.

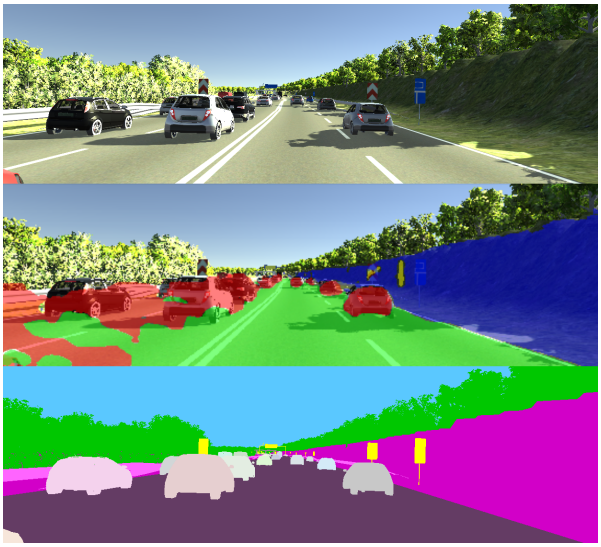


Figure 6: The input, output and ground truth of data from a modified U-NET[18] on the Virtual KITTI dataset[17]. The image below is a labelled image of the image on top.

V. Privacy and WMO

For the usage of data that is stored in the hospital patient record systems apply some strict regulations and laws. Two of them consider the use of medical data and privacy; *Wet medisch-wetenschappelijk onderzoek met mensen (WMO)* and the *Algemene verordening gegevensbescherming (AVG or the English version: GDPR)*

The WMO states if ethical approval is needed by the METC (medical ethical review committee). This is the case if both rules apply to the study:[19]

1) Er sprake is van medisch-wetenschappelijk onderzoek. English: It concerns medical, scientific research
2) Personen worden onderworpen aan handelingen of aan hen wordt een bepaalde gedragswijze opgelegd. English: The patients/participants are subject to procedures or are required to follow rules of behaviour.

Only the first rule applies to our study, based on that our study is not *WMO plichtig*.

1) Informed concent/Privacy: By dutch law (AVG) a *persoonsgegevens*/personal data is: All personal data from an identified or identifiable natural person. It is considered as information directly about someone or can be traced back to someone. [20] The data is fully anonymised (not pseude anonymised) and thereby it is not a *persoonsgegeven*. Because no persoonsgegevens/personal data is used, the data can/is legally allowed to be collected without informed consent of the patients.

In our conclusion for this study, no approval is needed from the METC, and anonymised video data can be used if this protocol is followed and no informed consent is needed.

REFERENCES

[1] Nariman Noorbakhsh-Sabet, Ramin Zand, Yanfei Zhang, and Vida Abedi. Artificial Intelligence Transforms the Future of Healthcare. *The American Journal of Medicine*, 2019.

[2] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing: Overview, challenges and the future. In *Lecture Notes in Computational Vision and Biomechanics*. 2018.

[3] Michiel Schuhmacher. Autonomous anatomical structure recognition using machine learning. Technical report, 2018.

[4] Robert B Yates, Brant K Oelschlager, and Carlos A Pellegrini. *Ch 42 Gastroesophageal Reflux Disease and Hiatal Hernia*. 2016.

[5] Fabio Baldi. PPI-Refractory GERD: an Intriguing, Probably Overestimated, Phenomenon. *Current Gastroenterology Reports*, 2015.

[6] Rena Yadlapati, Michael F. Vaezi, Marcelo F. Vela, Stuart J. Spechler, Nicholas J. Shaheen, Joel Richter, Brian E. Lacy, David Katzka, Philip O. Katz, Peter J. Kahrilas, C. Prakash Gyawali, Lauren Gerson, Ronnie Fass, Donald O. Castell, Jenna Craft, Luke Hillman, and John E. Pandolfino. Management options for patients with GERD and persistent symptoms on proton pump inhibitors: recommendations from an expert panel. *American Journal of Gastroenterology*, 113(7):980–986, 2018.

[7] Kamthorn Yolsuriyanwong, Eric Marcotte, Mukund Venu, and Bipan Chand. Impact of vagus nerve integrity testing on surgical management in patients with previous operations with potential risk of vagal injury, 2018.

[8] S van Rijn, N. F. Rinsma, M Y A van Herwaarden-Lindeboom, J. Ringers, H. G. Gooszen, P J J van Rijn, R. A. Veenendaal, J. M. Conchillo, N. D. Bouvy, and Adrian A M Masclee. Effect of Vagus Nerve Integrity on Short and Long-Term Efficacy of Antireflux Surgery. *The American journal of gastroenterology*, 111(4):508–15, 4 2016.

[9] Szu Chia Liao, Hong Zen Yeh, Chung Wang Ko, Han Chung Lien, and Chi Sen Chang. The management of gastroesophageal reflux disease: An update, 2010.

[10] Geoffrey Paul Kohn, Raymond Richard Price, Steven R. DeMeester, Jörg Zehetner, Oliver J Muensterer, Ziad Awad, Sumeet K Mittal, William S Richardson, Dimitrios Stefanidis, and Robert D Fanelli. Guidelines for the management of hiatal hernia. *Surgical Endoscopy*, 27(12):4409–4428, 12 2013.

[11] Nasser K Altorki, David Yankelevitz, and David B Skinner. Massive hiatal hernias: The anatomic basis of repair. *The Journal of Thoracic and Cardiovascular Surgery*, 115(4):828–835, 4 1998.

[12] Peter J. Kahrilas and Ikuo Hirano. Diseases of the Esophagus. In *Harrison's Principles of Internal Medicine*, chapter 347, pages 1900 – 1911. McGraw-Hill Global Education, 19 edition, 2016.

[13] Stephen R. T. Evans and Elizabeth A. David. Laparoscopic Nissen Fundoplication. In *Surgical Pitfalls*, pages 175–185. Elsevier, 2009.

[14] Samad Hashimi and Ross M. Bremner. Complications Following Surgery for Gastroesophageal Reflux Disease and Achalasia, 2015.

[15] Peter Funch-Jensen, Anette Bendixen, Maria Gerding Iversen, and Henrik Kehlet. Complications and frequency of redo antireflux surgery in Denmark: A nationwide study, 1997-2005. *Surgical Endoscopy and Other Interventional Techniques*, 2008.

[16] Alexander Mordvintsev and K Abid. Optical Flow — OpenCV-Python Tutorials 1 documentation.

[17] Adrien Gaidon, Qiao Wang, Yohann Cabon, and Eleonora Vig. VirtualWorlds as Proxy for Multi-object Tracking Analysis. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2016-December, pages 4340–4349. IEEE Computer Society, 12 2016.

[18] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351, pages 234–241. Springer Verlag, 2015.

[19] Uw onderzoek: WMO-plichtig of niet? — Onderzoekers — Centrale Commissie Mensgebonden Onderzoek — https://www.ccmo.nl/onderzoekers/wet-en-regelgeving-voor-medisch-wetenschappelijk-onderzoek/uw-onderzoek-wmo-plichtig-of-niet.

[20] Wat zijn persoonsgegevens? — Autoriteit Persoonsgegevens — https://autoriteitpersoonsgegevens.nl/nl/over-privacy/persoonsgegevens/wat-zijn-persoonsgegevens.