

UNIVERSITY OF TWENTE

MASTER THESIS

Interactive Residual 3D U-net for the Segmentation of the Pancreas in Computed Tomography Scans

Author:

Tim G.W. BOERS

Supervisors:

Dr. F. VAN DER HEIJDEN

Dr. H.J. HUISMAN

Dr. J.J. HERMANS

F. VAN DEN NOORT MSc

R. HAARMAN MSc

*A thesis submitted in fulfillment of the requirements
for the degree of Master of Science*

in the

Field of Technical Medicine
University of Twente

April 4, 2019

UNIVERSITY OF TWENTE.

Declaration of Authorship

I, Tim G.W. BOERS, declare that this thesis titled, “Interactive Residual 3D U-net for the Segmentation of the Pancreas in Computed Tomography Scans” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Preface

The thesis is divided into two main parts. Part A will focus on the clinical background, theoretical background of deep learning and will also denote a prior experiment which did not end up in the paper. Part B contains the main product of this thesis and will be provided in paper format.

Contents

Declaration of Authorship	iii
Preface	v
A Thesis	1
1 Introduction	3
2 Basics of the pancreas	5
2.1 The Anatomy	5
2.2 Physiology of the Pancreas	6
2.3 Blood supply	6
3 Detection and Treatment of Pancreatic Cancer	7
3.1 Pancreatic cancer	7
3.2 Symptoms	7
3.3 Imaging of Pancreatic Cancer	9
3.4 Pitfalls of the Pancreas Imaging	9
3.5 Treatment of pancreatic carcinoma	10
4 Basics of Deep learning Methods	13
4.1 Introduction	13
4.2 Fully Convolutional networks	14
4.3 Encoder-Decoder Architecture: U-net	15
4.4 Transfer Learning	15
5 Experiment: Interactive Recurrent U-net	17
5.1 Introduction	17
5.2 Methods	17
5.3 Results	19
5.4 Discussion	19
6 Conclusion	21
B Paper	25
I Graphical User Interface	41

Part A

Thesis

Chapter 1

Introduction

Pancreatic cancer (PC) is starting to become a prominent problem among aging citizens in western civilizations, whilst innovations in imaging screening and intervention are currently still lacking for diagnosis and treatment purposes.

The median age at diagnosis of PC is 71 years, and is rarely diagnosed in persons younger than 40 years of age[1]. The incidence of all types of PC ranges from 1 to 10 cases per 100,000 people of which 85% are adenocarcinomas[1]. Approximately 60 to 70% of PCs are located in the head of the pancreas, and 20 to 25% are located in the body and tail of the pancreas.[1] In 2017 the incidence of PC has been increasing up to 2400 new cases[2] and with the current projections, PC is to become the second leading cause of cancer-related in the year 2030[3]. Efforts to find a cure remain unsuccessful, as the 5-year overall survival rate has remained stable at approximately 5% for the last 30 years[4].

The source of this problem is that the symptoms of PC frequently appear when the tumour has metastasized or is classified as locally advanced pancreatic cancer, turning the patient, in both cases, incurable.[1] This makes it incredibly hard to find an adequate therapy for PC, which is observable in the overall survival rate compared to other cancer types, see Figure 1.1. In order to increase the overall survival rate, interventions should be improved by visualization techniques allowing to discriminate between stages of the disease and effectively target it during treatment. Early detection and treatment of the disease is essential to improve the survival rate of these patients. However, large scale screening programmes for early detection are hard to implement due to a scarcity of medical experts. This upsurges the need for assistive tools that reduce the dependency on medical expert, which could render PC screening and intervention a viable option. In particular artificial intelligence (AI) shows great promise, to build semi-automatic segmentation tools.

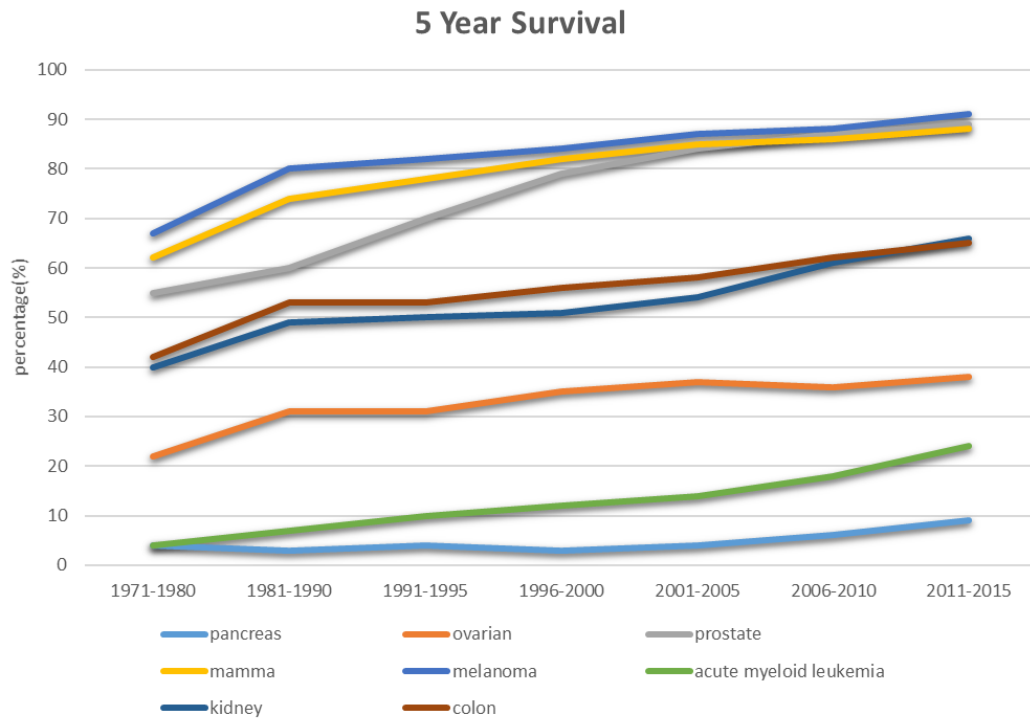


FIGURE 1.1: Trends in the 5 year overall survival in the Netherlands

Screening programmes can have a massive impact on the survival of the patient, and especially imaging based methods are suitable for use in a PC detection programme. Current figures show that the survival of patients undergoing palliative resection is about 12 to 18 months longer than that of patients with unresectable disease, which have a median survival of 4 to 6 months[5]. If resectable disease were curable, the survival of pancreatic cancer could even increase about fivefold [5]. Screening programmes can benefit the treatment outcome, but essential is that they are safe, inexpensive and highly accurate in the diagnoses of PC at a stage when it is not causing symptoms in the patient. Though the standard accepted modality, Computed Tomography (CT), for PC detection meets these requirement, its use is still limited. In practice this modality is bottle-necked by the scarcity of radiological experts. Thus screening programmes could benefit from the assistive tools that reduce the dependency on expert radiologists to screen for pancreatic cancer. AI techniques can be introduced to support the decision process. To assure these tools are working properly, thousands of reference standards are required for an extensive validation. The process of data collection can be accelerated by smart semi-automatic segmentation tools to elevate the dependency on the user.

In this thesis we set forth; the clinical background of pancreatic cancers, Artificial intelligence based techniques for the segmentation of the pancreas. We aim to utilize state of the art techniques to interactively segment the pancreas in late venous CT scans for diagnosis and intervention.

Chapter 2

Basics of the pancreas

2.1 The Anatomy

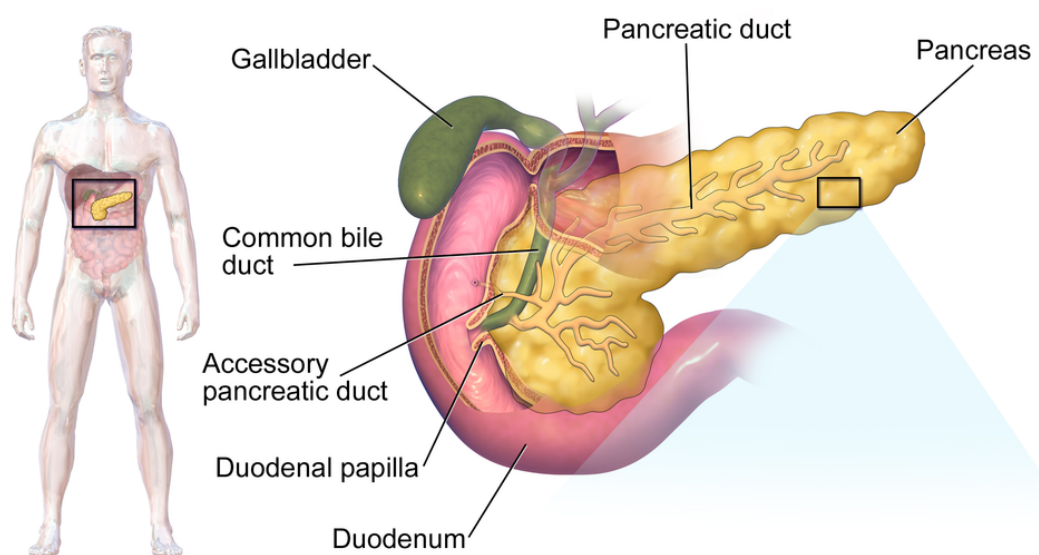


FIGURE 2.1: The anatomy of the pancreas[6]

The pancreas is a retroperitoneal organ of the digestive system that in humans lies in the upper left part of the abdomen, see Figure 2.1. The pancreas can be divided into four main parts: the head, neck, body and the tail of pancreas.

The head is the thickest part and is connected to segment D2 and D3 of the duodenum. The head is further extended by the uncinate process which lies to the right of the superior mesenteric vessels (SMV). The neck is the thinnest part of the pancreas and lies anterior to SMV. Its front upper surface supports the pylorus of the stomach. The body is the main part and lies to left of the SMV. The splenic vein lies in groove on posterior surface of body. The tail lies between layers of the splenorenal ligament in the splenic hilum. Branches of the splenic artery supply the body and tail via multiple branches including the dorsal pancreatic artery, greater pancreatic artery and transverse pancreatic artery.

The pancreas is made up of 2 types of glands, the exocrine and the endocrine glands. The exocrine gland secretes digestive enzymes. These enzymes are secreted into a network of ducts that join the main pancreatic duct, which runs the length of the pancreas. The endocrine gland consists of the islets of Langerhans and secretes hormones into the bloodstream.

2.2 Physiology of the Pancreas

The pancreas is a mixed gland, having both an endocrine and an exocrine function. As an exocrine gland it is involved in digestion by its secretion of pancreatic juice into a branching system of pancreatic ducts that extend throughout the gland. The main pancreatic duct empties in most of the individuals into the duodenum at the ampulla of Vater. In other individuals, the accessory pancreatic duct will effectuate this task. The juice contains bicarbonate, which neutralizes acid from the stomach by the secretion of bicarbonate entering the duodenum. Other digestive enzymes break down carbohydrates, proteins, and lipids in ingested food. The endocrine part is composed of hormonal tissue distributed along the pancreas in discrete clusters called the islets of Langerhans. These Islets are involved in the production of the hormones insulin and glucagon, which regulate the level of glucose in the blood, as well as somatostatin, and pancreatic polypeptide.

2.3 Blood supply

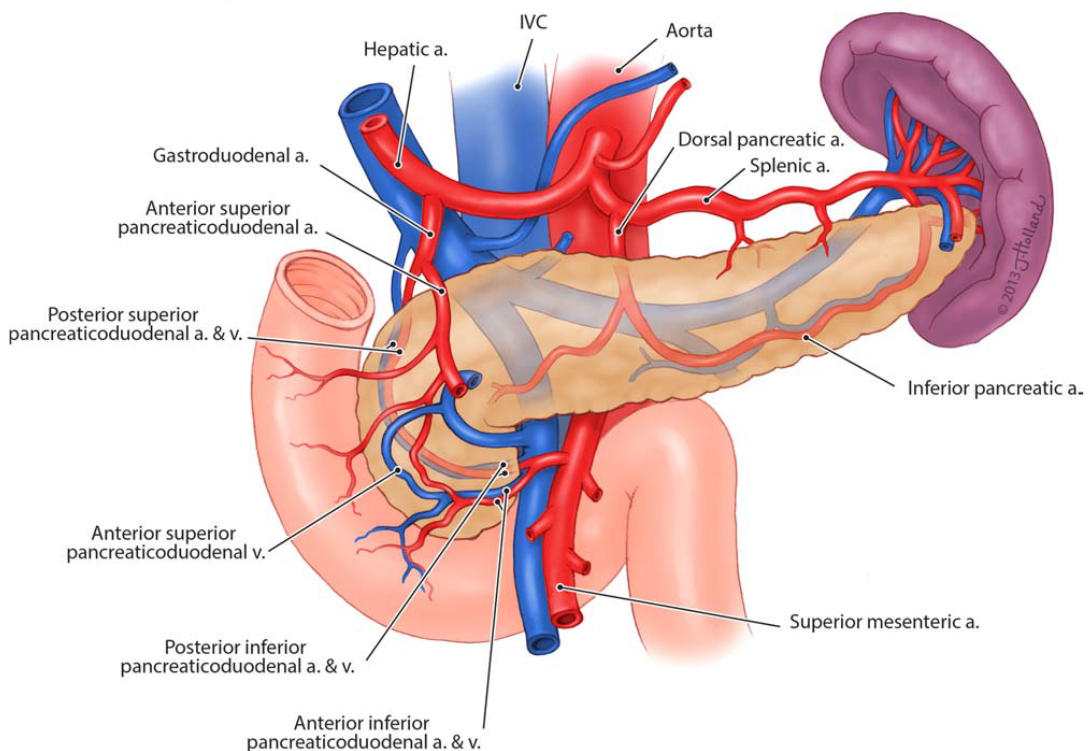


FIGURE 2.2: Blood supply of the pancreas[7]

The pancreas is primarily supplied of blood by two arteries; the pancreaticoduodenal and the splenic arteries, see Figure 2.2. The inferior and superior pancreaticoduodenal arteries supply to the head and multiple branches of the splenic artery supply the body and tail. Drainage of the blood will be flow from the pancreatic and pancreaticoduodenal veins into respectively the splenic, and the splenic and portal veins.

Chapter 3

Detection and Treatment of Pancreatic Cancer

3.1 Pancreatic cancer

Pancreatic cancer (PC) is a devastating and poorly understood cancer. PC can both involve the endocrine and exocrine tissue. The most common form of PC is a pancreatic adenocarcinoma, which affects the exocrine part of the pancreas. The disease occurs more often in the developed world, which had 68% of new cases in 2012. Pancreatic adenocarcinoma typically has poor outcomes, currently a median life expectancy is 6 – 10 and 3 – 6 months for patients presenting with locally advanced disease or metastatic disease, respectively [8]. The average percentage alive for at least one and five years after diagnosis being 25% and 5% respectively. In localized disease where the cancer is small (< 2 cm) the number alive at five years is approximately 20%. For those with neuroendocrine cancers the number alive after five years is much better at 65%, varying considerably with type.

Risk factors of PC include smoking, obesity, diabetes, and certain rare genetic conditions including multiple endocrine neoplasia type 1 and hereditary nonpolyposis colon cancer among others. About 25% of cases are attributable to tobacco smoking, while 5–10% of cases are linked to inherited genes.

3.2 Symptoms

The presenting signs and symptoms are related to the location within the pancreas.[1] Most commonly patients with PC initially present with abdominal pain, weight loss, asthenia and anorexia. Also, diabetes mellitus is common under PC patients, as it is present in at least 50% of patients with PC.[1] Courvoisier's sign is a common manifestation of tumours in the head of the pancreas, meaning painless jaundice presumably caused by obstruction of the pancreatic or biliary tract[9]

The secondary signs of PC are valuable for the image based diagnosis. These signs on medical imaging include pancreatic and common bile duct dilation and atrophy of the pancreas upstream to the tumour[10]. In Figures 3.1 - 3.3 three different stages of atrophy with ductal dilation are illustrated which are found in patients with PC.

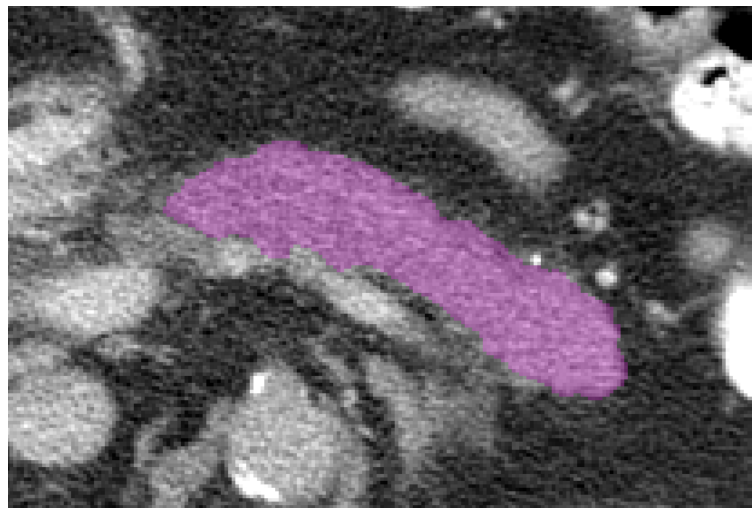


FIGURE 3.1: Segmented pancreas; normal parenchyma(purple), normal pancreatic duct(not visible)

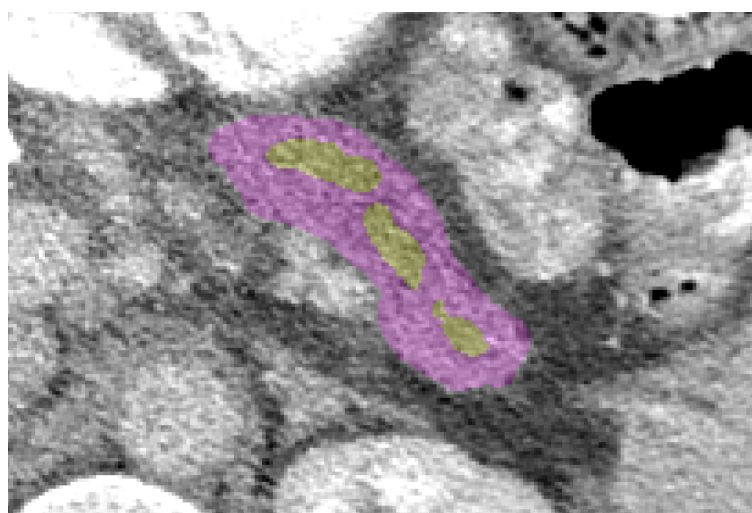


FIGURE 3.2: Segmented pancreas; normal parenchyma(purple), dilated pancreatic duct(yellow)

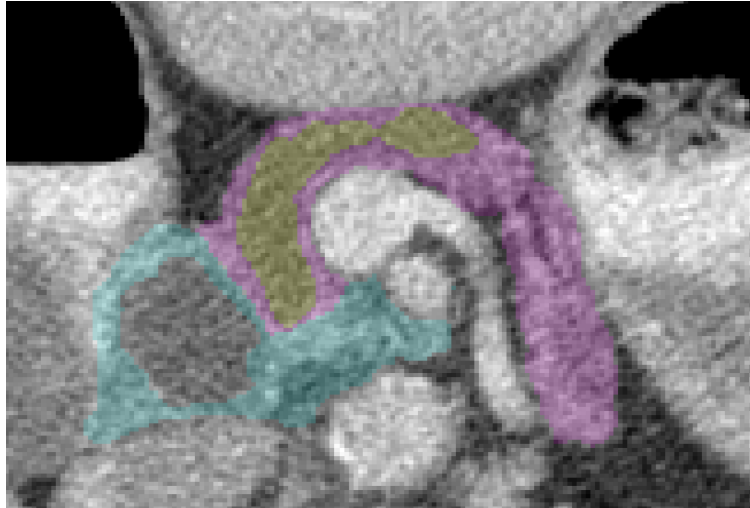


FIGURE 3.3: Segmented pancreas; atrophic parenchyma(purple), dilated pancreatic duct(yellow), tumour(blue) and parenchyma downstream from tumour(blue)

3.3 Imaging of Pancreatic Cancer

Computed Tomography(CT) is the work-horse of pancreatic imaging. Typically ductal adenocarcinomas appear as poorly defined masses with an extensive surrounding desmoplastic reaction. They enhance poorly compared to adjacent normal pancreatic tissue and thus appear hypodense on contrast enhanced arterial scans in 75-90% of the cases [11]. On delayed scans these tumours might turn isodense. The attenuation of the pancreatic tumour is correlated to its vascularity and changes as a function of time after intravenous contrast administration.[11] It is common practice to acquire multiple phases to get the best contrast between the tumours and the adjacent tissues.

CT plays an important role in determining a treatment plan for the patient. This modality correlates well with surgical findings in predicting the resectability with a positive predictive value of 89-100%[12]. The most important feature to assess locally is the relationship of the tumour to surrounding vessels i.e. Portomesenteric Veins, Superior Mesenteric Artery, Coeliac Axis and the Common Hepatic Artery. Local tumours are categorized as "resectable", "borderline resectable" and "irresectable" according to the involvement of these bloodvessels. If the tumor has no contact with its surrounding vessels it is deemed resectable. When the tumour is in contact with an artery, but is encased less than 90 degrees, the tumour is deemed borderline resectable. Lastly, if the encasement of the tumour around one of the arteries is more than 90 degrees and in contact with a blood-vessel, then the tumour is deemed irresectable. When the tumor is in contact with the Superior Mesenteric Vein or Portal Vein, the encasement can go up to 270 degrees before being deemed as irresectable.[13]

3.4 Pitfalls of the Pancreas Imaging

The pancreas is distinctively a hard organ to delineate on CT images, which makes it hard to assess the organ fully. The pancreas is surrounded by tissues with similar HU values, a clear border is often lacking between the organ and adjacent tissues.

Especially a clear border between the duodenum and the pancreas is absent, because they have similar densities and lack a sharp transition zone. Also commonly seen in patients is fat replacement of the pancreas, which is a benign process which can be secondary to diabetes, obesity and chronic pancreatitis.[14] Fat replacement is usually more severe in anterior aspect of head of pancreas. Due to fat replacement it can be hard to distinguish the lining of the pancreas with the abdominal fat.

Moreover pathologies associated with PC increases can mimic the phenotype of PC. Often seen is chronic pancreatitis, which is a known risk factor for PC. [15] The characteristics of chronic pancreatitis are very similar to a pancreatic tumour and is therefore easily mistaken for PC itself. In early findings chronic pancreatitis has a diagnostic appearance of decreased and delayed enhancement of administration of contrast, ductal dilatation with prominent side branches and ductal calcifications. Late findings might show either parenchymal atrophy or enlargement.

Additionally, pancreatic tumours can induce inflammation i.e. secondary pancreatitis. This commonly manifests itself as a swollen pancreas with poorly demonstrated borders. This could lead to impairing results on CT as degradation of the parenchyma tissue might not lead to visible atrophy, as the swelling could correct for the atrophic volume loss.[16]

3.5 Treatment of pancreatic carcinoma

Surgical resection is the only potentially curative treatment. Three main reasons for the poor prognosis of PC can be identified[5]. First, pancreatic cancer patients seldom exhibit disease-specific symptoms until late in the course of the illness, when the tumour has already metastasized or is classified as locally advanced pancreatic cancer, turning the patient, in both cases incurable. While the only hope of long-term survival in pancreatic cancer is a curative resection, by the time the diagnosis is made, only 15% to 20% of the patients, are eligible for surgery. Furthermore only up to 30% of patients undergoing resection will have positive resection margins, meaning that the tumor can not be resected. [1] Second, radical resection of PC is not curative. If resectable disease were curable, the survival of pancreatic cancer could increase about fivefold. Unfortunately, surgery is only palliative in the majority of patients undergoing resection[5]. The median survival of patients undergoing curative resection is only about 12 to 18 months longer than that of patients with unresectable disease, which is 4 to 6 months. Thus, the failure of radical resection of pancreatic cancer to effect a cure is another contributor to the poor survival in pancreatic cancer. Third, adjuvant therapy is only palliative. If adjuvant therapy could eradicate the residual cancer after curative resection, more patients undergoing pancreatic resection could hope for long-term survival. Such therapy would potentially also have an impact on locally unresectable disease. While some progress has been made in the chemotherapy of pancreatic cancer and newer biologic agents promise better results, overall adjuvant therapy only prolongs life and rarely cures pancreatic cancer. Single-agent gemcitabine is the standard first-line agent for the treatment of advanced PC.[17]. Currently, even the most effective chemotherapy regimens for metastatic or locally advanced disease are largely palliative and are capable of extending overall survival by only several months.[18] On top of that even in cases with localized disease, treatable with surgery followed by adjuvant chemotherapy, has a 5-year survival rate of only 24% [18].

Thus, it is clear that unless newer chemotherapeutic agents can downstage locally unresectable cancers to allow resection, the greatest impact on resectability and

survival would be through detection of early pancreatic cancer. Therefore, we believe that among the most compelling needs in pancreatic cancer research today are to devote more attention to screening programmes, give better consecutive therapy.

Chapter 4

Basics of Deep learning Methods

4.1 Introduction

Deep learning is a sub-part of machine learning, which is based on the learning of data representations. It utilizes Deep Neural Networks, which are connected systems that are able to learn to perform tasks by learning on examples without having a prior knowledge about the tasks. The term “deep” usually refers to the number of hidden layers in the neural network. Deep learning models are trained by using large sets of labeled data and neural network architectures that learn features directly from the data without the need for manual feature extraction.

4.1.1 Neuron model

Similar to our brains, the basic computational units of the mathematical model are the neurons. The neuron receives its input from the dendrites, the produced output is sent from their axons, see Figure 4.1.

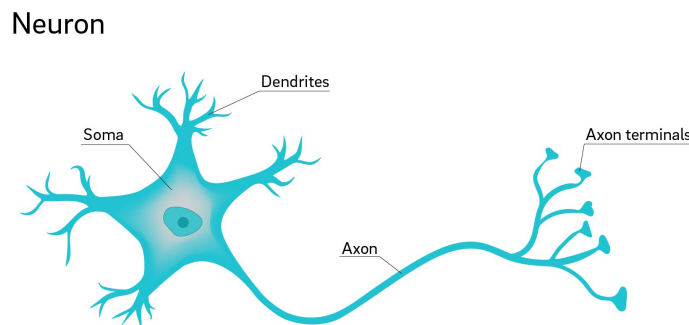


FIGURE 4.1: An example of a biological neuron

A synapse is the transition between the axon of one neuron and dendrite of another. In the mathematical model these synaptic strengths are learnable, modeled with weights w_i . Dendrites transport the input signals x_i to the cell body, here the inputs are summed. When the total sum reaches a certain threshold, the neuron is able to produce an output using its axons. The output production rate of a neuron is modeled by a non-linear activation function $f(x)$. The activation of a neuron is modeled as shown in equation 4.1. This equation models the decision-making, by

varying the weights and activation function affects the output.

$$a_i = f\left(\sum_i (w_i x_i) + b\right) \quad (4.1)$$

Here a_i represents activation of a neuron, $f(x)$ is the non-linear activation function, w_i is the synaptic strength weight, x_i is the neuron input and b is the bias of the neuron. As Rectified Linear Unit (ReLU) functions have proven to be computationally efficient, and are often used as non-linear activation function.

In a deep learning models, many activation functions, as in equation 4.1, are put together into one model, see Figure 4.2. The input of each node is represented by the arrows, and corresponds to x_i of our mathematical model. The product of each node is represented by the white circle, which is the output of the activation function.

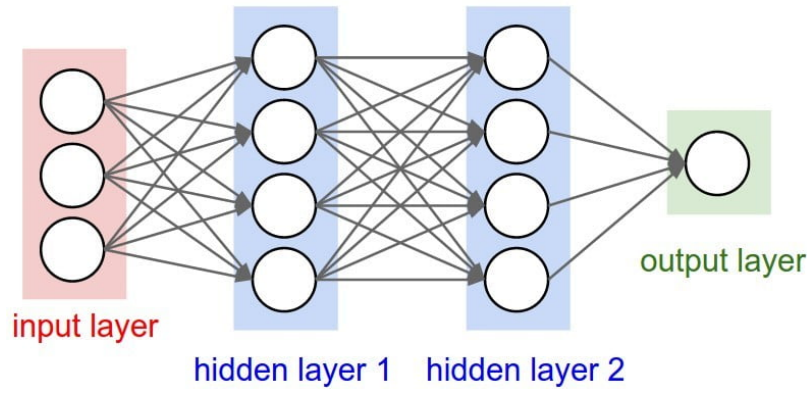


FIGURE 4.2: A fully connected neural network containing two hidden layers.

4.2 Fully Convolutional networks

Fully Convolutional Networks (FCNs) are built only from locally connected layers, such as convolution, pooling and upsampling. Fully connected layers are not used in this kind of architecture. This gives a reduction to the number of parameters and computation time. Moreover, the network can work regardless of the original image size, without requiring any fixed number of units at any stage, given that all connections are local.

Originally, FCNs started for the use on classification tasks, where the output to an image is a single class label. However, in many visual tasks, especially in biomedical image processing, the desired output should include localization, or semantic segmentation.

In the last two years, deep convolutional networks have outperformed the state of the art in many visual recognition tasks. While convolutional networks have already existed for a long time, their success was limited due to the size of the available training sets and the size of the considered networks. The breakthrough was due to a recent growth in computational power and the availability of data, enabling the supervised training of a large network.

These models must be trained on huge datasets, otherwise the training leads to overfitting of the neural networks. Overfitting means that the model tends to fit well to the training data, but is predictive on newly presented data. Various regularization techniques have been introduced to counter over-fitting, namely early stopping,

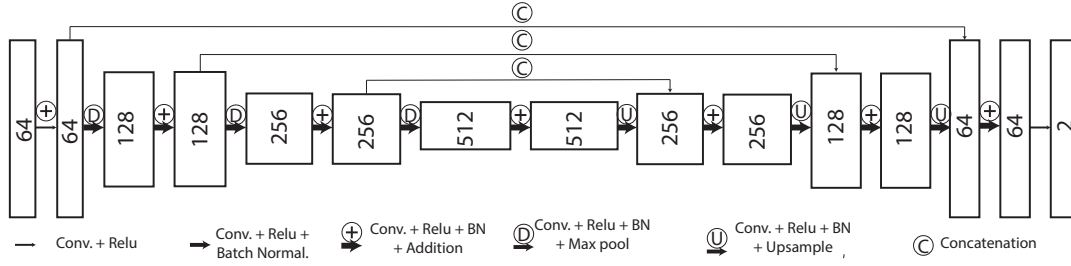


FIGURE 4.3: Residual 3D U-net architecture. Each box corresponds to a multi-channel feature map. The number of filters is denoted by the number in the box. The arrows denote the different operations.

weight decay, L1 and L2 regularizations, dropout and data augmentation. Data augmentation is essential to teach the network the desired invariance and robustness properties, when only few training samples are available.

In case of microscopical images we primarily need shift and rotation invariance as well as robustness to deformations and gray value variations. Especially random elastic deformations of the training samples is a key concept to train a segmentation network with very few annotated images.

4.3 Encoder-Decoder Architecture: U-net

For biomedical image segmentation purposes, Ronneberger et al[19] have designed a network called U-Net. The network does not include the usual stacking of layers, but a different approach is taken that consists of an encoder path followed by a symmetric decoder path for semantic segmentation. The name U-net refers to the shape in which the layers are sequenced. The theory behind this architecture is to combine lower and higher level feature maps through skip connections, that will improve the localization of high resolution features. The encoder path aims to encode the input image into feature representations at multiple different levels, into a low dimensional feature space.

The decoder path expands the feature dimensions to meet the same size with the corresponding concatenation blocks from the encoder input. Because of the up-sampled features the FCN is able to better localize and learn representations with following convolutions. This strategy allows the seamless semantic segmentation of arbitrarily large images by an overlap-tile strategy.

4.4 Transfer Learning

Transfer learning has emerged as the ability to apply knowledge and skills learned in prior tasks to be applied in different cases. Due to a scarcity of labeled data researchers have been looking for ways to decrease the dependence on huge data sets. Inspired by the fact that human beings can utilize previously-acquired knowledge to solve new but similar problems much more quickly and effectively. It is not only a way to ease to learning process of new models, but it also eases the need and effort to recollect the new training data.

As a FCN architecture could contain millions of parameters, directly learning so many parameters on a limited database is problematic. The key idea is that the

internal layers of the FCN can act as a generic extractor of mid-level image representation, which can be pre-trained on one dataset and then re-used on other target tasks. This is motivated by the observation that the earlier features of a FCNs contain more generic features, and later layers of the FCNs becomes progressively more specific to the details of the classes contained in the original dataset.

Depending on the size of the data-set it is possible to fine-tune all the layers of the FCNs, or to keep some of the earlier layers fixed and only fine-tune some higher-level portion of the network. The weights are fine-tuned of the pretrained network are fine-tuned by continuing the backpropagation.

Chapter 5

Experiment: Interactive Recurrent U-net

5.1 Introduction

Limitations of medical image segmentation using supervised learning includes data scarcity. Often in the thousands of large number of labels are required for training, but are most of the time not available. That might be due to a lack of medical experts, time or money. Due to these reasons, there is a significant demand for computer algorithms that can do segmentation quickly, accurately and preferably with as little human interaction as possible. FCNs have already proven that they are capable of automatically generating label maps, as well as the ability to learn dynamics within a system.

In this experiment we aim to jointly train an neural network that is able to learn to automatically segment an medical image, combined with the ability to take a prior generated segmentation and scribbles into account to refine the segmentation.

5.2 Methods

This experiment focuses on the joint training task network. This network, see Figure 5.1, is trained to generate a segmentation from a CT image, a segmentation, and a scribble map. In the first iteration, the segmentation and the scribbles will be matrices filled with zeros. After one iteration, the segmentation is returned back into the model to update to initial segmentation made by the Network. A network will then perform a following segmentation based on the CT image, the prior generated segmentation, and optionally scribbles in the scribble map.

5.2.1 Training protocol

Data handling is an important criterion for a robust FCN. Before the feature extraction we pre-processed the data with a few basic processing steps to reduce the input dimensionality. We also apply data augmentation to reduce overfitting of the model. The preprocessing step starts with by applying a gaussian filter with a sigma value of 0.75, to smooth the image. Then we rescale the image window from a range of -160 to 240 HU to a range of -1 and 1. Values below or above this range are clipped to -1 or 1 respectively. This window was chosen based on a basic soft tissue window[20]. Last, we crop the image based on a bounding box, which is automatically generated based on the reference standard. The bounding box is defined by the maximum and minimum index value corresponding to the segmentation in the X, Y and Z axis. Around this bounding box we expand a 5% margin with respect to the dimensions

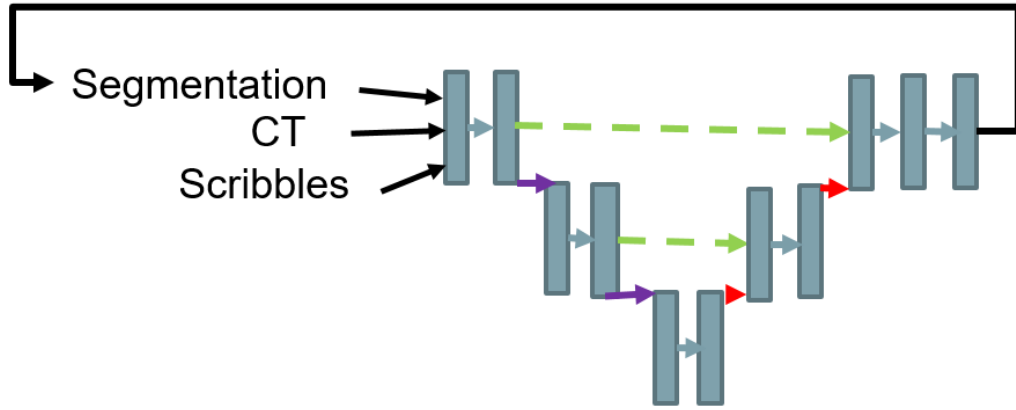


FIGURE 5.1: The Interactive Recurrent U-net. The arrows correspond to the flow of data. The blue arrows correspond to 3D convolutional layers, the purple arrows correspond to maxpool layers, the red arrows correspond to upscale layers, and the green arrows represent the skip-connections.

of the specific image. The resulting volume is ultimately cropped to $64 \times 64 \times 24$ volume.

During the training we generate smooth deformations using random displacement. We applied random rotations between -10° and $+10^\circ$, and translations of -3 to $+3$ voxels in each dimension at each iteration in order to generate plausible deformations during training. Per pixel displacements are computed using linear interpolation. On top of the random displacement we also add Gaussian noise with a sigma of 1%, based on the noise level found in acryl, which depicts similar HU values as soft human tissue.[21]

In order to optimize the amount of agreement between two binary regions in the training data, the Dice similarity coefficient (DSC) is used as loss function. We used a differentiable version, which has been proposed by Milletari et al.[22] for training our 3D U-Net model. We minimize the loss function for each class in K.

During the training, scribbles are automatically generated based on the difference between the predicted segmentation and the reference standard. From the difference the largest connected component was extracted, and processed with a 3D skeletanization operation, followed by a single 3d dilation operation. The structuring element used in the dilation was composed as a sphere with a radius of 1 pixel.

To ensure the model will not only converge on the refinement of small local errors, we introduced action replay in the training phase. Meaning that we reset the segmentation map and the scribble map, back to all zeros. This way the training puts more emphasis on the initial training task as well.

We will evaluate the performance of the model by performing a 5 fold cross validation. The performance will be measured in average Dice coefficient. To evaluate the beneficiary addition of the scribbles, we will compare the performance of the U-net when generation an initial segmentation, and the Dice coefficient score after 3 iterations.

5.3 Results

	Initial	Refined
Dice	78.4	78.8

TABLE 5.1: Average Dice scores of the 5 fold cross validation

5.4 Discussion

We have successfully trained a interactive Recurrent U-net, which is able to yield a DSC of 78.8%. The performance of refined segmentation is not significantly higher compared to the initial segmentation. It does not appear that the model is able to pick up on the scribbles and incorporate them into the models out. Because the model already had trouble picking up on the scribbles during the training stage, we did not feel the need to test our model on an independent data set. The motivation behind this is that the environment in the training stage is, and the validation stage should not differ a lot. If the neural network already has trouble picking up on the scribbles in this environment, testing on a separate data set is futile. Due to a time constraint we were not able to fully determine the cause of this issue, and further debug the code. We chose to further focus on another method explained in Part B of this thesis, as this method seemed more promising to us.

Chapter 6

Conclusion

In conclusion, the increasing incidence of pancreatic cancer will make it the second deadliest cancer in 2030. Due to the late presence of pancreas cancer symptoms many cases are no longer eligible for treatment. Imaging based early diagnosis and image guided treatment are emerging potential solutions to detect the disease. Artificial intelligence (AI) can help provide and improve widespread diagnostic expertise and accurate interventional image interpretation. However, due to the difficult anatomical shape and features, it is still hard to generate clinical applicable segmentations with the current AI technologies. A major limitation of the current AI technologies is that they require a lot of labeled data to generate proper segmentations, which is a lot of expert radiologist labour intensive, and therefore expensive. Therefore, we aim to develop a semi-automatic segmentation network based on deep learning that should allow us to quickly gather accurate labeled data. Early attempts that try to learn the dynamics of scribble based corrections of segmentations were not successful.

In part *B* of this thesis we will mainly focus on the interactively refinement of neural networks in order to optimize the network for case-specific segmentation. We will also briefly reiterate on the known difficulties of pancreas screening and automatic detection, which is further complemented with successful segmentation strategies for other organs, and which will become a basis for our tool.

Bibliography

- [1] David P. Ryan, Theodore S. Hong, and Nabeel Bardeesy. "Pancreatic Adenocarcinoma". In: *New England Journal of Medicine* 371.11 (2014), pp. 1039–1049. ISSN: 0028-4793. DOI: 10.1056/NEJMra1404198. URL: <http://www.nejm.org/doi/10.1056/NEJMra1404198>.
- [2] *Cijfers over kanker*. 2019. URL: https://www.cijfersoverkanker.nl/selecties/Dataset{_}1/img5c7d3987b1c03.
- [3] Lola Rahib, Benjamin D. Smith, Rhonda Aizenberg, et al. "Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states". In: *Cancer Research* 74.11 (2014), pp. 2913–2921. ISSN: 15387445. DOI: 10.1158/0008-5472.CAN-14-0155.
- [4] Daniel Åkerberg, Daniel Ansari, Roland Andersson, et al. "The Effects of Surgical Exploration on Survival of Unresectable Pancreatic Carcinoma : A Retrospective Case-Control Study". In: (2017), pp. 1–9. DOI: 10.4236/jbise.2017.101001.
- [5] Suresh T Chari. "Detecting Early Pancreatic Cancer : Problems and Prospects". In: (2007), pp. 284–294. DOI: 10.1053/j.seminoncol.2007.05.005.
- [6] H Gray, S Standring, H Ellis, et al. *Gray's anatomy : The Anatomical Basis of Clinical Practice*. Elsevier Health Sciences, 2005. ISBN: 0443071683.
- [7] A. Cesmebasi, J. Malefant, S. D. Patel, et al. "The surgical anatomy of the lymphatic system of the pancreas". In: *Clinical Anatomy* 28.4 (2015), pp. 527–537.
- [8] Megan B Wachsmann, Laurentiu M Pop, and Ellen S Vitetta. "Pancreatic ductal adenocarcinoma: a review of immunologic aspects." In: *Journal of investigative medicine : the official publication of the American Federation for Clinical Research* 60.4 (2012), pp. 643–63. ISSN: 1708-8267. DOI: 10.2310/JIM.0b013e31824a4d79. URL: <http://www.ncbi.nlm.nih.gov/pubmed/22406516{\%}5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3319488>.
- [9] Aysel Türkvatan, Aye Erden, Mehmet Akif Türkoğlu, et al. "Congenital variants and anomalies of the pancreas and pancreatic duct: Imaging by magnetic resonance cholangiopancreatography and multidetector computed tomography". In: *Korean Journal of Radiology* 14.6 (2013), pp. 905–913. ISSN: 12296929. DOI: 10.3348/kjr.2013.14.6.905.
- [10] D.A. Bluemke, J.L. Cameron, R.H. Hruban, et al. "Potentially resectable pancreatic adenocarcinoma: Spiral CT assessment with surgical and pathologic correlation". In: *Radiology* 197.2 (1995), pp. 381–385. ISSN: 00338419. DOI: 10.1148/radiology.197.2.7480681.
- [11] Gauri R. Varadhachary, Eric P. Tamm, James L. Abbruzzese, et al. "Borderline resectable pancreatic cancer: Definitions, management, and role of preoperative therapy". In: *Annals of Surgical Oncology* 13.8 (2006), pp. 1035–1046. ISSN: 10689265. DOI: 10.1245/ASO.2006.08.011.

- [12] David S K Lu, Howard a Reber, Robert M Krasny, et al. "OBJECTIVE. major 1 1". In: *October June* (1997), pp. 1439–1443. ISSN: 0361-803X. DOI: 10.2214/ajr.168.6.9168704.
- [13] DPCG. *Dutch Pancreatic Cancer Group definitions for resectability of pancreatic adenocarcinoma*. 2012.
- [14] Namita S. Gandhi, Myra K. Feldman, Ott Le, et al. "Imaging mimics of pancreatic ductal adenocarcinoma". In: *Abdominal Radiology* (2017), pp. 1–12. ISSN: 23660058. DOI: 10.1007/s00261-017-1330-1.
- [15] Aram F Hezel, Alec C Kimmelman, Ben Z Stanger, et al. "Genetics and biology of pancreatic ductal adenocarcinoma Genetics and biology of pancreatic ductal adenocarcinoma". In: *Genes & Development* 1.30 (2016), pp. 355–385. ISSN: 08909369. DOI: 10.1101/gad.1415606. URL: <http://genesdev.cshlp.org/content/20/10/1218.full.pdf+html>.
- [16] Kiran K Busireddy, Mamdoh AlObaidy, Miguel Ramalho, et al. "Pancreatitis-imaging approach". In: *World Journal of Gastrointestinal Pathophysiology* 5.3 (2014), p. 252. ISSN: 2150-5330. DOI: 10.4291/wjgp.v5.i3.252. URL: <http://www.wjgnet.com/2150-5330/full/v5/i3/252.htm>.
- [17] Patrick C. Hermann, Stephan L. Huber, Tanja Herrler, et al. "Distinct Populations of Cancer Stem Cells Determine Tumor Growth and Metastatic Activity in Human Pancreatic Cancer". In: *Cell Stem Cell* 1.3 (2007), pp. 313–323. ISSN: 19345909. DOI: 10.1016/j.stem.2007.06.002.
- [18] J. J. Lee, R. M. Perera, H. Wang, et al. "Stromal response to Hedgehog signaling restrains pancreatic cancer progression". In: *Proceedings of the National Academy of Sciences* 111.30 (2014), E3091–E3100. ISSN: 0027-8424. DOI: 10.1073/pnas.1411679111. URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1411679111>.
- [19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (2015), pp. 1–8. ISSN: 16113349. DOI: 10.1007/978-3-319-24574-4_28. arXiv: 1505.04597. URL: <http://arxiv.org/abs/1505.04597>.
- [20] Daniele Marin, Rendon C Nelson, Lisa M Ho, et al. "Image Quality , and Radiation Dose during the Pancreatic Parenchymal Phase : Effect of a Purpose : Methods : Results :". In: 256.2 (2010).
- [21] Kristine Gulliksrud, Caroline Stokke, Anne Catrine, et al. "Physica Medica How to measure CT image quality : Variations in CT-numbers , uniformity and low contrast resolution for a CT quality assurance phantom". In: *Physica Medica* (2014), pp. 1–6. ISSN: 1120-1797. DOI: 10.1016/j.ejmp.2014.01.006. URL: <http://dx.doi.org/10.1016/j.ejmp.2014.01.006>.
- [22] Fausto Milletari, Nassir Navab, and Seyed Ahmad Ahmadi. "V-Net: Fully convolutional neural networks for volumetric medical image segmentation". In: *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016* (2016), pp. 565–571. DOI: 10.1109/3DV.2016.79. arXiv: 1606.04797.

Part B

Paper

Interactive Residual 3D U-net for the Segmentation of the Pancreas in Computed Tomography Scans

T.G.W. Boers¹, Y. Hu², E. Gibson², D.C. Barratt², E. Bonmati², J. Krdzalic³, F. van der Heijden¹, J.J. Hermans³, H.J. Huisman⁴

¹Faculty of Science and Technology, University of Twente, Netherlands

²Department of Medical Physics and Biomedical Engineering, University College London, London

³Department of Radiology and Nuclear Medicine, Radboud UMC, Netherlands

⁴Diagnostic Image Analysis Group, Radboud UMC, Netherlands

E-mail: tgw.boers@gmail.com

Abstract. *Introduction:* The increasing incidence of pancreatic cancer will make it the second deadliest cancer in 2030. Imaging based early diagnosis and image guided treatment are emerging potential solutions. Artificial intelligence (AI) can help provide and improve widespread diagnostic expertise and accurate interventional image interpretation. Accurate segmentation of the pancreas is essential to create annotated data sets to train AI, and for computer assisted interventional guidance. Automated deep learning segmentation performance in pancreas CT imaging is low due to poor grey value contrast and complex anatomy. A good solution seemed a recent interactive deep learning segmentation framework for brain CT that helped strongly improve initial automated segmentation with minimal user input. This method yielded no satisfactory results for pancreas CT, possibly due to a sub-optimal neural architecture. We hypothesize that a state-of-the-art U-net neural architecture is better because it can produce a better initial segmentation and is likely to be extended to work in a similar interactive approach. *Methods:* We implemented the existing interactive method, iFCN, and developed an interactive version of U-net method we call iUnet. The iUnet is fully trained to produce the best possible initial segmentation. In interactive mode it is additionally trained on a partial set of layers on user generated scribbles. We compare initial segmentation performance of iFCN and iUnet on a 100CT dataset using DSC analysis. Secondly, we assessed the performance gain in interactive use with three observers on segmentation quality and time. *Results:* Average automated baseline performance was 78% (iUnet) vs 72% (FCN). Manual and semi-automatic segmentation performance was: 87% in 15 min. for manual, and 86% in 8 min. for iUnet. *Discussion:* We conclude that iUnet provides a better baseline than iFCN and can reach expert manual performance significantly faster than manual segmentation in case of pancreas CT. Our novel iUnet architecture is modality and organ agnostic and can be a potential novel solution for semi-automatic medical imaging segmentation in general.

Keywords: Deep learning, pancreatic cancer, interactive segmentation, U-net

Submitted to: *Phys. Med. Biol.*

1. Introduction

The increasing incidence of pancreatic cancer (PC) will make it the second deadliest cancer in 2030[1]. This incidence has reached up to 2400 newly reported cases (Netherlands) in 2017[2]. Efforts to find a cure remain unsuccessful, as the 5-year overall survival rate continues to be stable at approximately 5% for the last 30 years[3]. The difficulty is that the physical complaints of PC frequently appear in a late stage of the disease, turning the patient incurable.[4]

Imaging based early diagnosis and image guided treatment are emerging potential solutions. Computed Tomography (CT) is routinely used for the diagnostic workup as well as followup in patients with PC. However, in up to 30%, the diagnosis of PC is delayed or a patient is wrongfully diagnosed with PC. Image guided treatment could provide precision targeting to enhance curative options.

Artificial intelligence (AI) can help provide and improve widespread diagnostic expertise and accurate interventional image interpretation. Recent advances have successfully been applied to imaging diagnostic tasks across dermatology[5], ophthalmology[6, 7] and radiology[8, 9]. These innovative technologies should be adaptable for the automatic detection of PC in Computed Tomography (CT) images. Potentially, AI could become a considerable aid in screening programs to detect the disease in an earlier stage, therefore increasing the effectiveness of therapy.

Accurate segmentation of the pancreas is essential to create annotated data sets to train and develop AI, and for computer assisted interventional guidance. The quality and size of the training data set are crucial for the performance of the AI system[10, 6]. Training data requires accurate outlines of organs and lesions of interest. Any ambiguities in the outline will affect performance in limited data sets. To really cover the wide range of pancreas shapes and surrounding tissue, several hundreds of CT images must be annotated which is labor intensive. Interventional image guidance requires accurate outlines of the pancreas and relevant anatomy.

Automated deep learning segmentation performance in pancreas CT imaging is low due to poor grey value contrast and complex anatomy. The difficulty arises due to a lack of contrast between pancreas parenchyma and bowel, especially with the duodenum. Moreover, large variations in size of the pancreas volume and large variation in peripancreatic fat tissue, on top of textural variations of the pancreas parenchyma, increase the difficulty as well[11]. Cutting edge technologies like Wolz et al [12] reached only 70% Dice Similarity Coefficient(DSC) using multi atlas technology. Even recent state of the art deep learning techniques, like Gibson et al[13] are still limited to 78% DSC.

A good solution seemed a recent interactive deep learning segmentation framework for brain CT that helped strongly improve initial automated segmentation with minimal user input. Wang et al.[14] proposed a semi-automated technique (iFCN), which utilizes Fully Convolutional Networks(FCN) that handles user interactions to interactively improve the initial segmentation.

The iFCN solution yielded no satisfactory results for pancreas CT, possibly due to a sub-optimal neural architecture. Wang utilized a simple dilated FCN, while the U-net architecture has been proven to be the state-of-the-art as of now[15]. Moreover, their framework still depends heavily on organ specific post-processing of the segmentation that took advantage of the sharp boundaries. However for the segmentation of the pancreas, this post processing step is inadequate as sharp distinguishable borders are not always present. Our experiments for automatic segmentation yielded a 68% DSC for the iFCN using this method.

We hypothesize that a state-of-the-art U-net neural architecture is better than iFCN because it can produce a better initial segmentation and is likely to be extended to work in a similar interactive approach.

2. Methods

2.1. Ethics and information governance

This work, and the local collection of data on implied consent, received national Research Ethics (REC) Committee approval from the Radboud UMC REC(2017-3976). De-identification was performed in line with the General Data Protection Regulation (EU) 2016/679.

2.2. Datasets and clinical taxonomy

The image data is derived from two independent datasets and will hence be distinguished independently.

D_1 : The first dataset is used to train the neural network. This set is sourced from a public dataset[13], which contains 90 late venous phased abdominal CT images and a respective reference segmentation. These were drawn from two data sets: The Cancer Image Archive (TCIA) Pancreas-CT data set and the Beyond the Cranial Vault (BTCV) Abdomen data set. Both datasets are comprised of scans that contain non-pancreatic related pathologies.

D_2 : The second dataset is used to validate our interactive U-net. This set consists of 10 cases. These cases were randomly selected from a data set containing 1905 late venous phased abdominal CT scans acquired in the year 2015 at the Radboud UMC. The patients included were all treated in the oncology department. The dataset consists of images from 941 males and 964 females. The mean age is 58.4 ± 13.3 years. Exclusion criteria are patients who were diagnosed with pancreas related pathologies.

During the training stage, the training set will be denoted as $T = \{X_i; Y_{ik}\}$, where X is the training image and Y is the reference label map, with i corresponding to a specific training case and k denotes the one-hot classification layer. The one-hot classification

label set k is $\{0, 1, 2, \dots, K\}$ with 0 being the background label and K denoting the number of labels included in the set. $\hat{Y}$ denotes the estimated label map produced by the trained FCN. $\hat{Y}'$ denotes the prediction with the scribbles incorporated.

2.3. Image Preprocessing

Data was preprocessed to fit the available computing facilities for the purpose of performing relevant experiments. Future algorithms should reduce the preprocessing requirement. Before the feature extraction we pre-processed the data with a few basic processing steps to reduce the input dimensionality. The preprocessing step starts with by applying a Gaussian filter with a sigma value of 0.75, to smooth the image for resampling. Then we rescale the image window from a range of -160 to 240 HU to a range of -1 and 1. Values below or above this range are clipped to -1 or 1 respectively. This window was chosen based on a basic soft tissue window[16]. Lastly, we crop the image based on a bounding box, which is automatically generated based on the reference standard segmentation. The bounding box is defined by the maximum and minimum index value corresponding to the segmentation in 3 dimensional axes. Around this bounding box we expand a 5% margin with respect to the dimensions of the specific image. The resulting volume is ultimately resampled to a volume of 64 x 64 x 24 voxels.

2.4. Baseline Training

The baseline training is performed to find adequate network weights for the generation of an initial segmentation. This training involved realistic, 1000 fold augmentation of the data by randomly displacing, rotating and adding noise at each training epoch. The displacements ranged between -3 to $+3$ voxels. Image pixel displacements are computed using linear interpolation. Random rotations ranged between -10° and $+10^\circ$. Low level Gaussian noise was added with a uniform sigma range of 0-3HU, based on the noise level found in acryl, which depicts similar HU values as soft human tissue.[17]

The segmentation performance is quantified by the Dice similarity coefficient (DSC). We used a differentiable DSC version in the loss function, which has been proposed by Milletari et al.[18] for training the FCNs. We minimize the loss function for each of the K classes. The implementation of DSC in our loss function for class k is as follows:

$$\mathcal{L}_k = -\frac{2 \sum_{i=1}^{N_v} \hat{Y}_i Y_i}{\sum_{i=1}^{N_v} \hat{Y}_i + \sum_{i=1}^{N_v} Y_i} \quad (1)$$

The total loss is calculated as the mean over all classes:

$$\mathcal{L}_{total} = \frac{1}{K} \sum_{k=0}^K \mathcal{L}_k \quad (2)$$

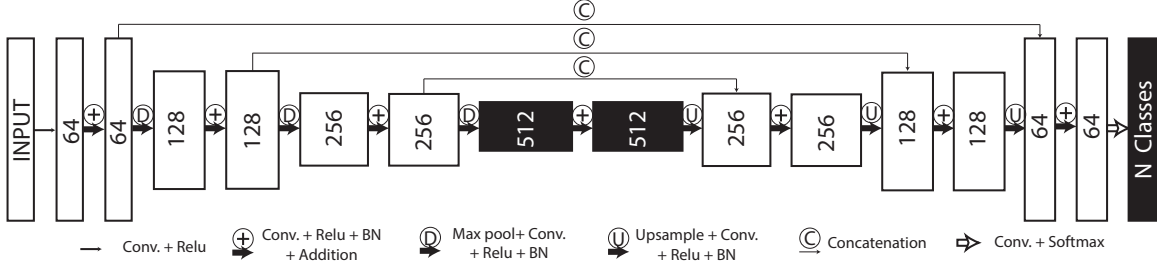


Figure 1: The iUnet architecture. The number in each box corresponds to the number of filter feature maps. The black boxes retrain during the interactive phase, the others remain fixed.

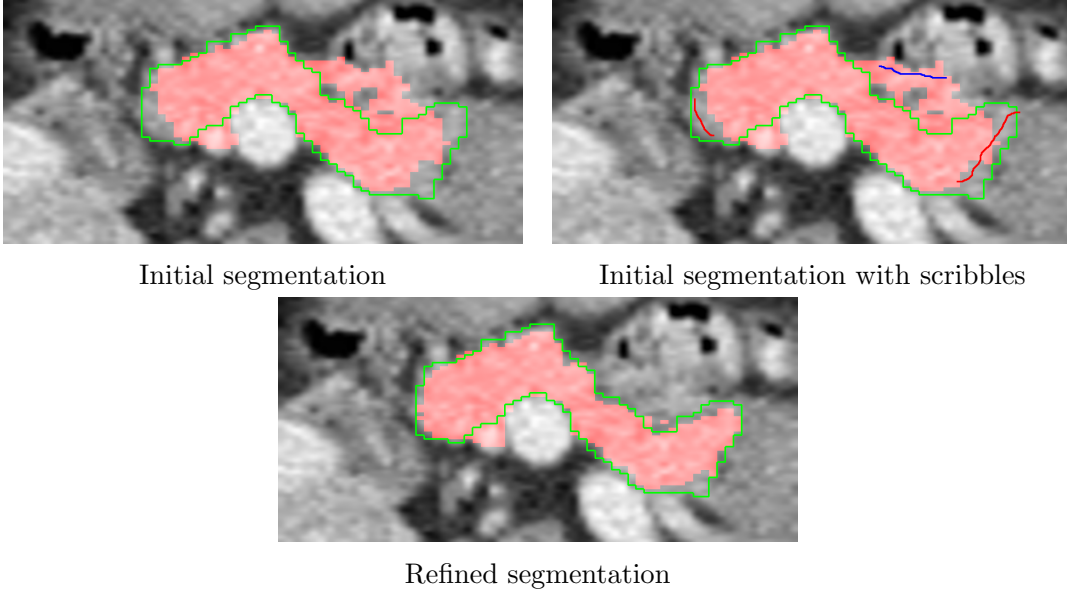


Figure 3: Example of the update process for the refinement of a segmentation. In all the images the red area depicts the segmentation create by our tool and the green delineation represents the ground truth. (a) displays the initial segmentation. (b) displays the initial segmentation including the scribbles. The red lines indicates the areas that need to be added to the segmentation. The blue line indicates the area that is falsely segmented and needs to be removed from the segmentation. (c) displays the result after refining the segmentation

2.5. iUnet layers

In the interactive U-net framework specific layers of the model are fine-tuned by minimizing an objective function to find a more appropriate conversion point for the unseen image. This fine-tuning is a concept of transfer learning, where knowledge gained prior during training is updated with new data to find a more robust model[19]. For transfer learning, it is challenging to determine which layers should be fine-tuned. Generally

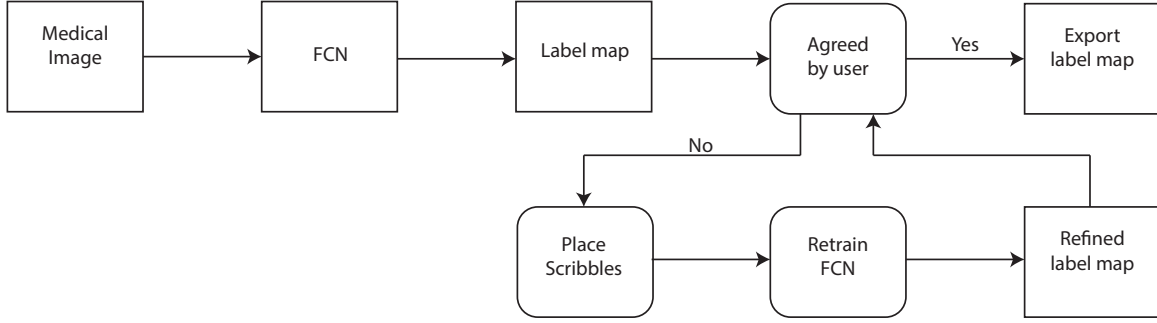


Figure 4: Framework for interactive segmentation

only the last layer is chosen for refinement. This is motivated by the observation that the first few layers are trained to identify more generic features of the task, but the later layers of the FCN become progressively more specific to the details of the classes contained in the original dataset. We however also choose to retrain the parameters within the deepest layers of the U-net, as these layers contain the largest receptive field. Esser et al. [20] demonstrated the dynamic performance of a variational U-net. By altering latent space between the encoder and decoder of the U-net they are able to control the output. Any additional information about the segmentation encoded in the latent space, which is not already contained in the prior, incurs a cost by providing new information on the likelihood $p(X|Y, \theta)$. With θ are the parameters of the FCN.

2.6. Interactive training

During the interactive retraining of the network a selection of iUnet layers are retrained (see Figure 1) using scribbles, to directly converge the network. The flowchart for the interactive segmentation is illustrated in Figure 4. The user generates an initial segmentation $\hat{Y}$ from the medical image X . With the initial segmentation obtained by the trained FCN, the user can provide a set of scribbles to provide new information to the iUnet to guide the update of $\hat{Y}$. The scribbles are denoted as S_k , with k denoting the corresponding label. In contrast to the standard training protocol that treats all pixels equally, now pixels are weighted based on a weight map. The initial weight map w starts with a volume that equals the size of the label map, Y , and a temporary value of 1 is assigned to all voxels. The user-provided scribbles are perceived as the true reference standard and should have a higher impact on the loss function, therefore receive a weight of 3. Lastly, the voxels surrounding the scribbles will be regarded as highly unreliable during the refinement, therefore receive a weighting of 0. This distance was determined by a threshold of the geodesic distance map, generated from the image, the scribbles and voxel indices. The function presented in Equation 3 is used to minimize the objective function. This loss function combines a DSC together with a voxel-specific weights and

a loss for volume difference.

$$\mathcal{L}_k = -\frac{2 \sum_{i=1}^{N_v} w_i \hat{Y}'_i \hat{Y}_i}{\sum_{i=1}^{N_v} \hat{Y}'_i + \sum_{i=1}^{N_v} \hat{Y}_i} + \lambda \left(\frac{\sum_{i=1}^{N_v} \hat{Y}_i - \hat{Y}'_i}{\sum_{i=1}^{N_v} \hat{Y}_i} \right)^2 \quad (3)$$

In order to predict multiple classes for segmentation, we calculate the total loss, which is depicted in formula 4.

$$\mathcal{L}_{total} = \frac{1}{K} \sum_{k=0}^K \mathcal{L}_k \quad (4)$$

2.7. Implementation

All models are implemented in Keras with the Tensorflow 1.12 backend. We use Adam optimization with an initial learning rate of 1e-4. We train the baseline iUNet for 2000 iterations, which takes about 12 hours, with a batch size of 8. The model is trained on a desktop running Windows 10 and leveraging a Nvidia RTX 2070 with CUDA 10.0 edition. The interactive training of the network was performed on a desktop with a Nvidia GTX 1080 running Ubuntu 16.04. A custom GUI, build in VTK and QT5, running via X-server were used to generate and optimize the segmentations.

2.8. Experiment 1: Baseline iFCN and iUNet comparison

The first experiment is set up to compare the automatic segmentation performance of iFCN and iUNet. The performance was quantified using DSC using a 5-fold cross validation on D_1 . This dataset is randomly divided into 5 equally sized folds. Four out of five folds of segmentations were used to train while the remaining fifth group was used as the development set. This strategy is repeated 5 times such that the DSC on the development set was computed. The accuracy reported in the paper is the average DSC obtained on the development set after each fold. To test the statical significance, we will also perform a Wilcoxon signed ranked test.

2.9. Experiment 2: iUnet validation

In the second experiment we compare manual expert segmentation to our iUnet segmentation method, based on segmentation quality and time to generate the segmentation.

A validation protocol is defined to compare the manual segmentation to the iUnet segmentation method. A team consisting of three radiological experts are appointed to perform the segmentations, and are referred to as readers. The ten cases in dataset D_2 are divided into two subsets containing 5 scans. To minimize the learning effect we will alternate the presented subsets and the segmentation method per reader. Dependant on the segmentation method, task specific constraints were set. The manual segmentation was performed using ITK-snap with full access to the features it provides. The time measurement of interactive time will start from the moment that the first annotation is

placed, and stopped after the segmentation is saved. There will be no time constraint to generate the segmentation. The interactive segmentation was performed in our custom made utility. Each reader will be given 2 test cases before the start of the validation to get familiar with the segmentation tool. At the start of the experiment, the reader will get the instruction to only focus on global features, as adjustments on voxel level would counteract the purpose of this tool. The recording of the time starts from the moment the reader generates the initial segmentation, and will stop when the segmentation is saved.

A consensus standard reference is produced in order measure the segmentation performance. This consensus reference standard is produced by two readers based on the full resolution images. These segmentations are cropped to the same dimensions as the scans in the validation set and are used to measure the DSC relative to a produced manual and interactive segmentations.

The data is analyzed in three methods. First, we will find the baseline for the manual segmentation by the comparison of the observer-to-observer differences to determine the agreement between the readers. A pairwise comparison approach between each label is performed and reported separately in DSC.

Second we compare the consensus reference standard to the manual, automatic and iUnet segmentation method, recorded in DSC. The DSCs are depicted in a violin plot to give a full picture on the variance within the data.

Finally, we will perform a Wilcoxon signed rank test to determine the significance of the improvement of the segmentation quality over the invested time and satisfactory score of the observer.

3. Results

The baseline experiment resulted in a DSC of $78.1\% \pm 8.7\%$ for the u-net and a dice score of $72.3\% \pm 11.4\%$ for dilated FCN.

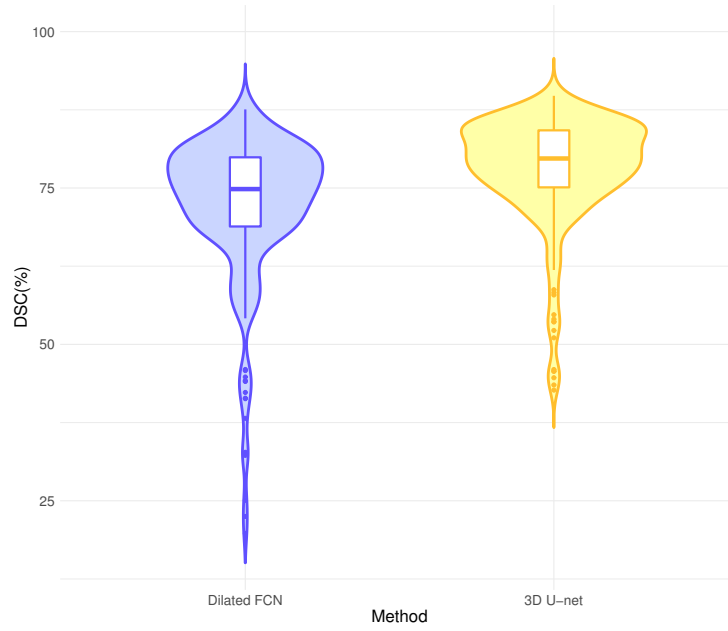


Figure 5: Performance of Dilated network architecture compared to the residual 3D U-net

The Wilcoxon signed ranked test found that $p < 0.001$ in the comparison of the manual segmentation vs the interactive U-net

Table 1: DSC values among 3 observers for manual pancreas segmentation

Dice coefficient	Kappa value		
	observer		
	1	2	3
observer 1	1	0.87	0.85
observer 2		1	0.86
observer 3			1
Average intra observer DSC is 0.86 (95% CI; 0.85-0.87)			

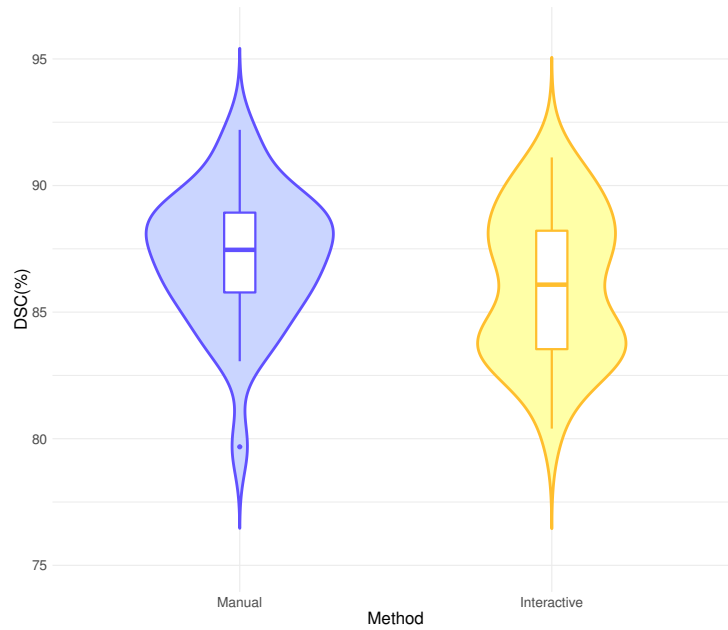


Figure 6: Boxplot of the DSC of the manual segmentation vs the interactive segmentation

The Wilcoxon signed ranked test found that $p = 0.017$ in the comparison of the manual segmentation vs the interactive U-net.

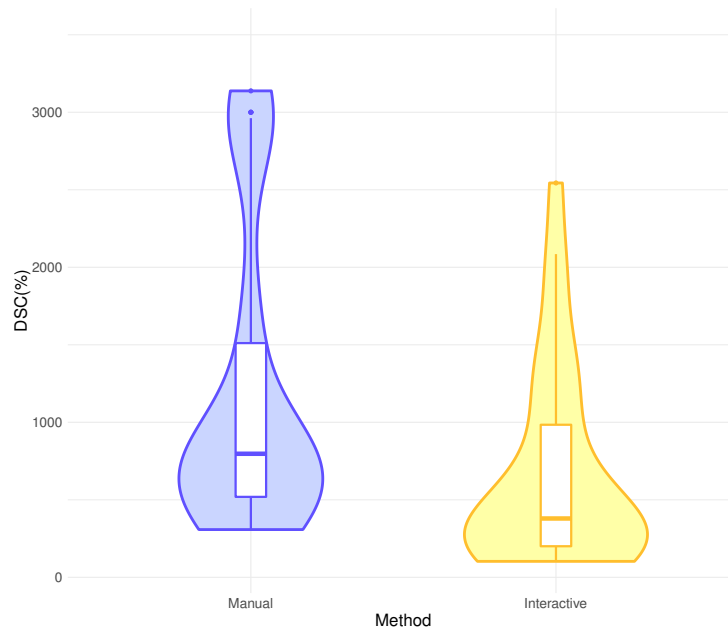


Figure 7: Boxplot of the time required to produce a segmentation

The wilcoxon signed ranked test found that $p < 0.001$ in the comparison of the manual segmentation time vs the interactive U-net segmentation time.

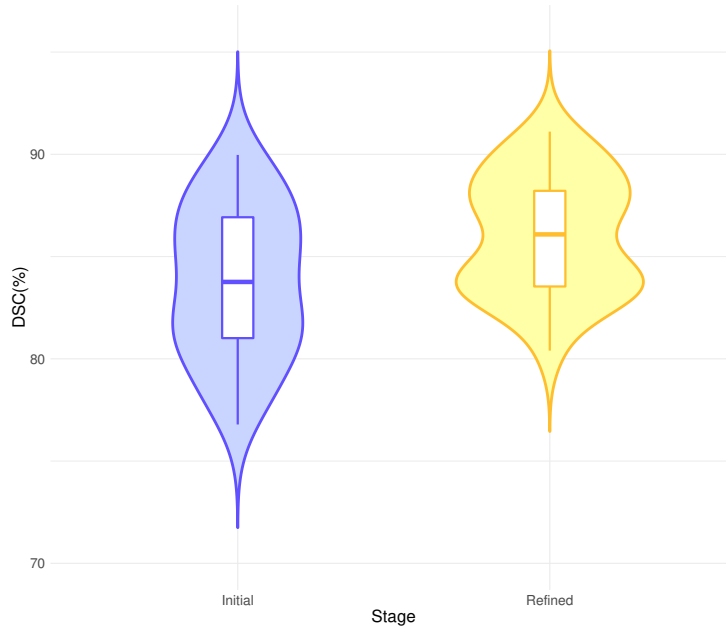


Figure 8: Performance of the fully automatic segmentation vs the interactive segmentation

The Wilcoxon signed ranked test found that $p < 0.001$ in the comparison of the manual initial segmentation vs the interactively refined segmentation.

4. Discussion

The results demonstrate that iUnet achieves expert performance in nearly twice the speed of expert radiologists. Observers with iUnet assistance reached 86 DSC versus 87.5 DSC in manual mode, but in a median 48.4% time reduction. This was achieved by developing an interactive version of UNet that showed a significantly higher baseline performance than iFCN.

The average automated baseline performance was $78.1\% \pm 8.7\%$ (iUnet) vs $72.3\% \pm 11.4\%$ (iFCN). We found that in all the cases the U-net performs better on the initial segmentation, and are also more reproducible judging by the smaller deviation. In the comparison, the iUnet architecture seems more stable, as less outliers are found for the resulting segmentation.

The iUnet reaches a higher DSC for the automatic segmentation on validation set compared to the training set. After comparison of the images, we see that the pancreas in D_2 are mostly all strongly delineated with adipose tissue. The poorer performance on our training set is likely associated with the lesser amount of visceral fat present in this patient, causing the boundaries between pancreas and surrounding tissues to be less well defined. Roth et al.[21] found that the body-mass-index has a strong influence on the difficulty of the segmentation task, and we speculate that this is the main reason

for the higher DSC on the validation set.

The initial iUnet has trouble picking up on the contours of the pancreas in certain anatomical regions, especially at the uncinate process. Visually, this area looks similar to the adjacent tissue. Also, this area contains a lot of anatomical variation in between subjects, which makes it hard for the network to learn a general feature.

Overall the performance of the interactive segmentation reaches expert level. We found an inter-observer variation, which on average has a 86% agreement. So reasonably, this is the maximum average DSC we expect to find for the interactive segmentation as well.

We were able to define two major contributors to the interobserver variability. The first cause is that due to the low resolution images, it is visually hard to delineate similar tissues, as crucial land marks are lacking. Secondly, the border of the pancreas contains a lot of partial volume artifacts, leading to disagreements whether the partial volume artifacts were parts of the pancreas. Due to the low resolution images this is a substantial amount. The outer-edge contains on average 25.1% of the volume of the segmentation.

The interactive experiments identified that the adaptability of the iUnet strongly depends on the training protocol, i.e the introduction of strong data augmentation in the training gives the U-net more flexibility to pick up on specific features. To further increase iUnet’s ability to generalize, it is of interest to use a larger training set with more patients, since a large training set with a wide variety helps to learn common features among different subjects. We speculate that we could improve the performance gain even more, if we train the decoder of the network on a huge similar dataset separately prior to the training of the decoder with the dataset for our specific use case.

Caution needs to be taken with the amount of layers that are opened. The algorithm disregards the adjacent area due to a lack of constraints. The algorithm has enough freedom to learn the scribbles by itself, causing the algorithm to diverge. After the first iteration we see an initial drop of the DSC.

After the first iteration the algorithm tends to under-segment the pancreas. This is however easily corrected by drawing additional scribbles so the iUnet converges to yield a more appropriate segmentation. To counteract this phenomenon we hypothesize that it will be beneficiary to add in a probabilistic representation in a latent space variable, to encourage a disentangled distribution over the generative factors $q(z|x)$ to be closer to an isotropic Gaussian $N(0, I)$.

In the future, we need to focus on the definition of clinically acceptable segmentation accuracy, which have yet to be defined depending on the use-case, e.g. guiding abdominal interventions. We do expect that the segmentation time will improve as a result of improvements in GPU technology. The availability of larger amounts of GPU memory will allow the processing of whole CT volumes at higher resolution. Future work can potentially augment the tissue-segmentation map with multiple labels per pixel to encode local tissue features, or with additional channels that encode continuous features. Lastly, bounding boxes in validation are provided by the user, but they could potentially

be obtained by automatic detection to further increase efficiency.

5. Conclusion

We conclude that iUnet provides a better baseline than iFCN and can reach expert manual performance significantly faster than manual segmentation in case of pancreas CT. Our novel iUNet architecture is modality and organ agnostic and can be potential novel solution for semi-automatic medical imaging segmentation in general.

References

- [1] L. Rahib, B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, and L. M. Matrisian. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the united states. *Cancer Research*, 74(11):2913–2921, 2014.
- [2] <https://www.cijfersoverkanker.nl/>, accessed 2019-02-02.
- [3] D. Åkerberg, D. Ansari, R. Andersson, and B. Tingstedt. The Effects of Surgical Exploration on Survival of Unresectable Pancreatic Carcinoma : A Retrospective Case-Control Study. pages 1–9, 2017.
- [4] D. P. Ryan, T. S. Hong, and N. Bardeesy. Pancreatic Adenocarcinoma. *New England Journal of Medicine*, 371(11):1039–1049, 2014.
- [5] H. A. Haenssle, C. Fink, R. a Schneiderbauer, and I. Zalaudek. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, (May):1836–1842, 2018.
- [6] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, and D. R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. 94043, 2016.
- [7] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, , and O. Ronneberger. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9):1342–1350, 2018.
- [8] J. Rajpurkar, P. and Irvin, K. Zhu, and A. Y. Ng. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. pages 3–9, 2017.
- [9] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 1–8, 2015.
- [10] T. Kavzoglu. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling and Software*, 24(7):850–858, 2009.
- [11] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers. Deep convolutional networks for pancreas segmentation in CT imaging. 2015.
- [12] R. Wolz, C. Chu, K. Misawa, K. Fujiwara, M. and Mori, and D. Rueckert. Automated Abdominal Multi-Organ Segmentation With Subject-Specific Atlas Generation. 32(9):1723–1730, 2013.
- [13] E. Gibson, F. Giganti, Y. Hu, and D. C. Barratt. Multi-organ Abdominal CT Reference Standard Segmentations. feb 2018.
- [14] G. Wang, W. Li, M. A. Zuluaga, and T. Vercauteren. Interactive Medical Image Segmentation using Deep Learning with Image-specific Fine-tuning. pages 1–12, 2018.
- [15] Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *Chaos, Solitons and Fractals*, 38(5):1411–1422, 2018.
- [16] D. Marin, R. C. Nelson, L. M. Ho, and R. Youngblood. Image Quality , and Radiation Dose during the Pancreatic Parenchymal Phase : Effect of a Purpose : Methods : Results :. 256(2), 2010.
- [17] K. Gulliksrud, C. Stokke, A. Catrine, and T. Martinsen. *Physica Medica* How to measure CT

- image quality : Variations in CT-numbers , uniformity and low contrast resolution for a CT quality assurance phantom. *Physica Medica*, pages 1–6, 2014.
- [18] F. Milletari, N. Navab, and S. A. Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, pages 565–571, 2016.
- [19] Karl Weiss, Taghi M. Khoshgoftaar, and Ding Ding Wang. *A survey of transfer learning*, volume 3. Springer International Publishing, 2016.
- [20] P. Esser, E. Sutter, and B. Ommer. A Variational U-Net for Conditional Appearance and Shape Generation. 2018.
- [21] H. R. Roth, A. Farag, L. Lu, E. B. Turkbey, and R. M. Summers. Deep convolutional networks for pancreas segmentation in CT imaging. (May), 2015.

Data availability

The clinical data used for the validation was collected at Radboud UMC. Data were used with both local and national permissions. They are not publicly available, as restrictions apply to their use.

Appendix I

Graphical User Interface

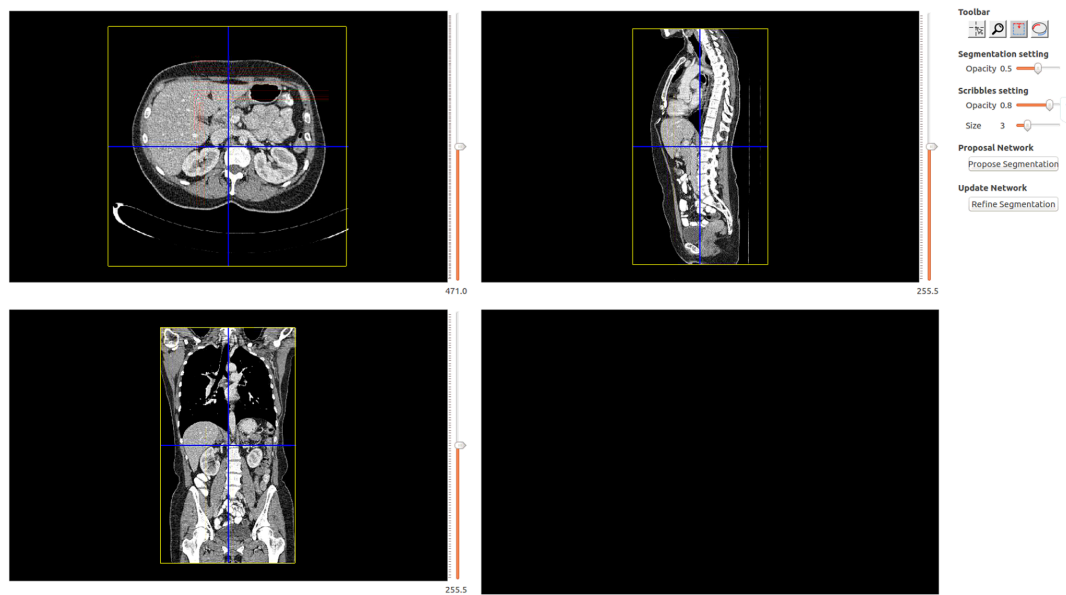


FIGURE I.1: A display of the graphical user interface used during the testing of our framework. The GUI is able to load a full resolution CT scan, and display them in three different anatomical axes. The GUI includes scrolling, zooming and translation features. The blue lines, correspond to the crosshair, which pinpoint a location in all anatomical axes. The yellow block is the bounding box. Everything contained within this bounding box will be downscaled to the resolution of the neural network, when generating an initial segmentation.

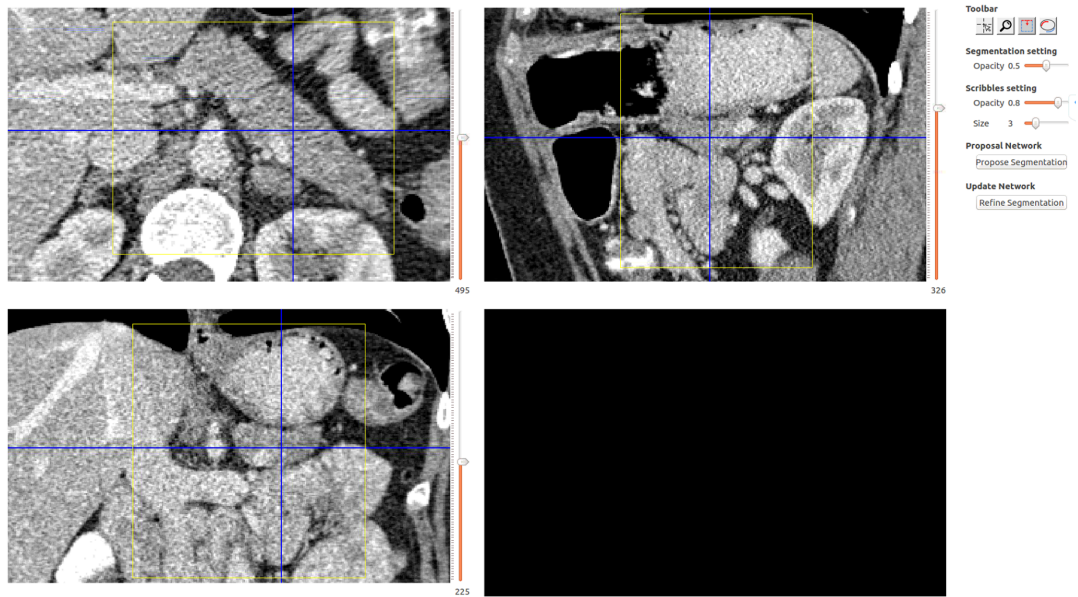


FIGURE I.2: This image displays the GUI, with a zoomed in projection of the pancreas. A bounding box is set to closely fit the pancreas.

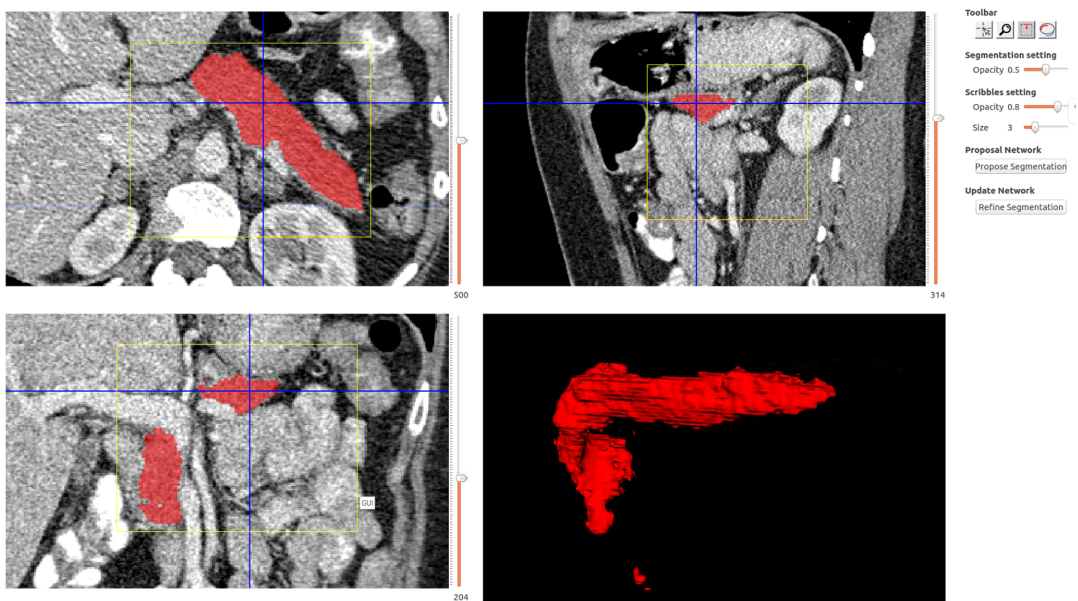


FIGURE I.3: A projection of the pancreas, with the initial label map. On the right-bottom screen, a 3d renderization of the initial segmentation is displayed.