IOP Publishing Phys. Med. Biol. 65 (2020) 065002 (9pp) https://doi.org/10.1088/1361-6560/ab6f99

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans

RECEIVED
7 May 2019

REVISED
2 January 2020

ACCEPTED FOR PUBLICATION
24 January 2020

PUBLISHED
11 March 2020

T G W Boers[1], Y Hu[2], E Gibson[2], D C Barratt[2], E Bonmati[2], J Krdzalic[3], F van der Heijden[1], J J Hermans[3] and H J Huisman[4]

[1] Faculty of Science and Technology, University of Twente, Enschede, The Netherlands
[2] Department of Medical Physics and Biomedical Engineering, University College London, London
[3] Department of Radiology and Nuclear Medicine, Radboud UMC, Nijmegen, The Netherlands
[4] Diagnostic Image Analysis Group, Radboud UMC, Nijmegen, The Netherlands

E-mail: tgw.boers@gmail.com

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

Abstract

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 computed tomography (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 network architecture. We hypothesize that a state-of-the-art U-net neural network architecture is better because it can produce a better initial segmentation and is likely to be extended to work in a similar interactive approach. 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 dice similarity coefficient analysis. Secondly, we assessed the performance gain in interactive use with three observers on segmentation quality and time. Average automated baseline performance was 78% (iUnet) versus 72% (FCN). Manual and semi-automatic segmentation performance was: 87% in 15 min. for manual, and 86% in 8 min. for iUNet. 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.

1. Introduction

The increasing incidence of pancreatic cancer (PC) will make it the second deadliest cancer in 2030 (Rahib *et al* 2014). This incidence has reached up to 2400 newly reported cases (Netherlands) in 2017 (integraal kankercentrum Nederland 2018). 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 (Åkerberg *et al* 2017). The difficulty is that the physical complaints of PC frequently appear in a late stage of the disease, turning the patient incurable (Ryan *et al* 2014).

Image-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 (Haenssle *et al* 2018), ophthalmology (Gulshan *et al* 2016, De Fauw *et al* 2018) and radiology (Ronneberger *et al* 2015, Rajpurkar *et al* 2017, Ma *et al* 2019). These innovative technologies should be adaptable for the automatic detection of PC in 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 datasets to train and develop AI, and for computer assisted interventional guidance. The quality and size of the training dataset are crucial for the performance of AI systems (Kavzoglu 2009, Gulshan *et al* 2016). Training data requires accurate outlines of organs and lesions of interest. Any ambiguities in the outline will affect performance in limited datasets. 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 (Roth *et al* 2015). Cutting edge technologies like (Wolz *et al* 2013) reached only 70% dice similarity coefficient(DSC) using multi atlas technology. Even recent state of the art deep learning techniques, like (Gibson *et al* 2018) are still limited to 78% DSC.

A potential solution seemed a recent interactive deep learning segmentation framework for brain CT, that helped to strongly improve initial automated segmentation with minimal user input. Wang *et al* (2018) 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 results using U-net have demonstrated state-of-the-art performance (Alom *et al* 2018). 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 (IRB) Committee approval from the Radboud UMC IRB2017-3976. 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.

2.2.1. D_1

The first dataset is used to train the neural network. This set is sourced from a public dataset (Gibson *et al* 2018), which contains 90 late venous phased abdominal CT images and a respective reference segmentation. These were drawn from two datasets: The Cancer Image Archive (TCIA) Pancreas-CT dataset and the Beyond the Cranial Vault (BTCV) Abdomen dataset. Both datasets are comprised of scans that contain non-pancreatic related pathologies.

2.2.2. D_2

The second dataset is used to validate our interactive U-net. Ten CT-scans were randomly selected from a dataset containing 1905 late venous phased abdominal CT scans, which were acquired in the year 2015 at the Radboud UMC. The image pixel spacing in the x and y-axis are 0.781 mm, and varies in the z-axis between 1 and 3 mm. This data is derived from patients who were treated in the oncology department at the time of scanning. The cohort consist of 941 males and 964 females, with a 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 $\{0, 1, 2,..., K\}$ with 0 being the background label and K denoting the number of labels included in the set. Here, in this work, we demonstrate results with $K = 1$ to distinguish background and pancreas classes. $\hat{Y}$ 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 (Marin *et al* 2010). 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 3D 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 $64 \times 64 \times 24$ voxels using trilinear interpolation. After preprocessing, the pixel spacing ranges between for the x-axis (1.7–2.7) mm, y-axis (0.79–1.8) mm and z-axis (3.34–4.7) mm.

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 translating, rotating and adding noise at each training epoch. The fractional translations ranged between −3 to +3 voxels. Image pixel translations are computed using trilinear interpolation. Random rotations ranged between −10° and +10°. Low level Gaussian noise was added with a uniform sigma range of 0–3 HU, based on the noise level found in acryl, which depicts similar HU values as soft human tissue (Gulliksrud *et al* 2014).

The segmentation performance is quantified by the DSC. We used a differentiable DSC version in the loss function, which has been proposed by Milletari *et al* (2016) 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:

$$\mathcal{L}_k = -\frac{2\sum_{i=1}^{N_v} \hat{Y}_{ik} Y_{ik}}{\sum_{i=1}^{N_v} \hat{Y}_i + \sum_{i=1}^{N_v} Y_{ik}}. \tag{1}$$

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

$$\mathcal{L}_{total} = \frac{1}{K}\sum_{k=0}^{K} \mathcal{L}_k. \tag{2}$$

2.5. Extending Unet for interactive training

In interactive mode a U-net framework can be made interactive by retraining a few iterations with user feedback by freezing all but a few specific layers. This is fast because it only requires a few iterations and one image to learn and it mitigates the risk of overfitting. This approach is a concept of transfer learning, where knowledge gained prior during training is updated with new data to find a more robust model (Weiss *et al* 2016). There are no general rules as to which layers to freeze. Generally only the last layer is retrained and this is what we propose as well. This is motivated by the observation that the first few layers are already trained to identify basic image level features, but the later layers become progressively more specific to actual segmentation. We also choose to add a layer in the deepest section for retraining in the (see figure 1), as these layers contain the largest receptive field. Esser *et al* (2018) demonstrated altering these layers (also referred to as latent space between the encoder and decoder of the U-net) they are able to control a large receptive field in the output.

2.6. Interactive training

During the interactive retraining of the network a previously described selection of iUnet layers are retrained (see figure 1) using scribbles. The flowchart for the interactive segmentation is illustrated in figure 3. 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 $\hat{Y}$, see figure 2. 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. This

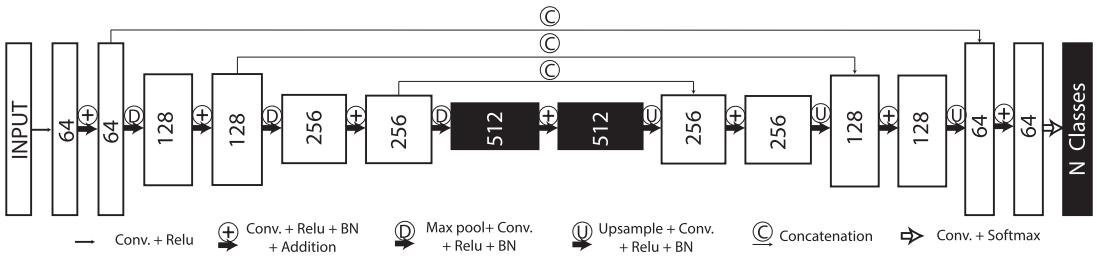
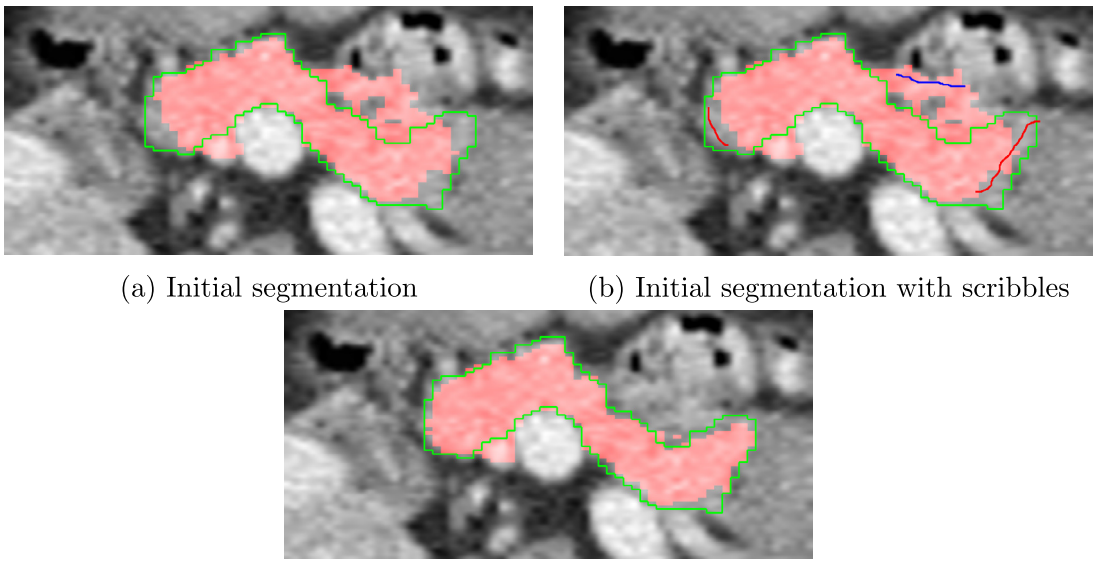


Figure 1. The iUnet architecture. The number in each box corresponds to the number of filter feature maps. The black boxes retrain during the interactive phase, the others remain fixed.



(a) Initial segmentation (b) Initial segmentation with scribbles

(c) Refined segmentation

Figure 2. Example of the update process for the refinement of a segmentation. In all the images the red area depicts the segmentation created 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 indicate 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. (a) Initial segmentation. (b) Initial segmentation with scribbles. (c) Refined segmentation.

weight map, w, equals the size of the label map, Y, and is initialized uniformly. The user-provided scribbles are considered improvement over the current segmentation and should have a higher impact on the loss function, therefore receive a weight of 3. Lastly, voxels at a small distance from the scribbles reflect a region of segmentation uncertainty, and therefore receive a weighting of 0. This distance was determined by a threshold (0.2) 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 is based on the DSC, weighted by a voxel-specific weight map w, and a loss for volume difference to optimize the objective function.

$$\mathcal{L}_k = -\frac{2\sum_{i=1}^{N_v} w_i \hat{Y}'_{ik} Y_{ik}}{\sum_{i=1}^{N_v} \hat{Y}'_{ik} + \sum_{i=1}^{N_v} Y_{ik}} + \lambda \left(\frac{\sum_{i=1}^{N_v} \hat{Y}_{ik} - \hat{Y}'_{ik}}{\sum_{i=1}^{N_v} \hat{Y}_{ik}} \right)^2. \tag{3}$$

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

$$\mathcal{L}_{total} = \frac{1}{K} \sum_{k=0}^{K} \mathcal{L}_k. \tag{4}$$

2.7. Implementation

All models are implemented in Keras with the Tensorflow 1.12 backend. We extended the U-net architecture to use 3D convolutional filters. Each layer uses a padding size of (1,1,1), in order to preserve the feature map size. We use Adam optimization with an initial learning rate of 1×10^{-4}. We train the baseline iUNet for 2000 iterations,

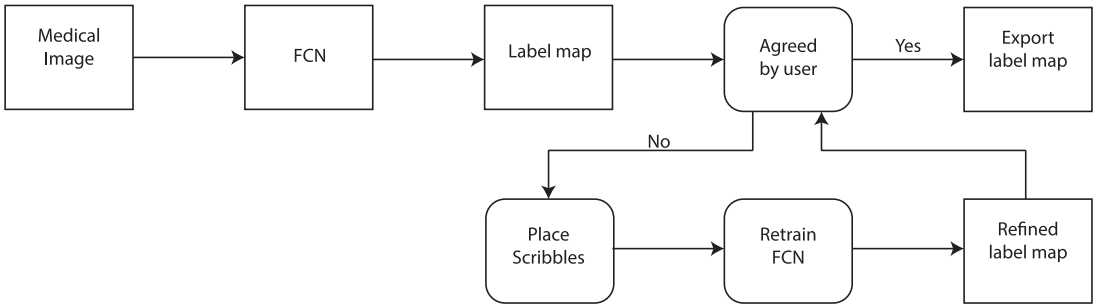


Figure 3. Framework for interactive segmentation.

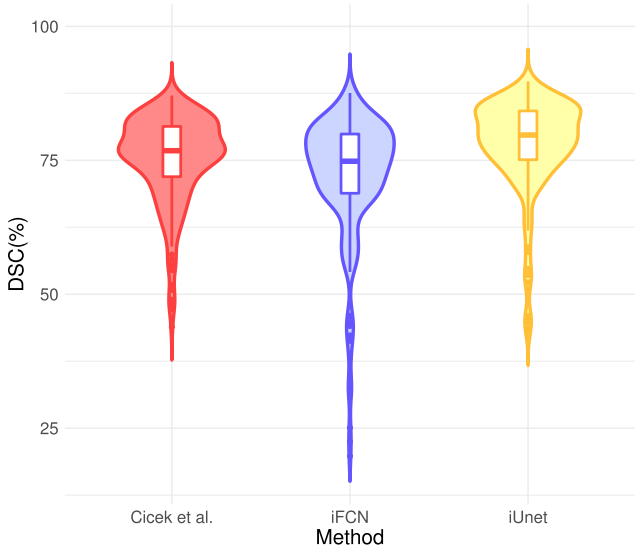


Figure 4. Performance of the Cicek *et al* U-net (Çiçek *et al* 2016), Dilated FCN network architecture and our 3D U-net in a five-fold cross validation using dataset D1.

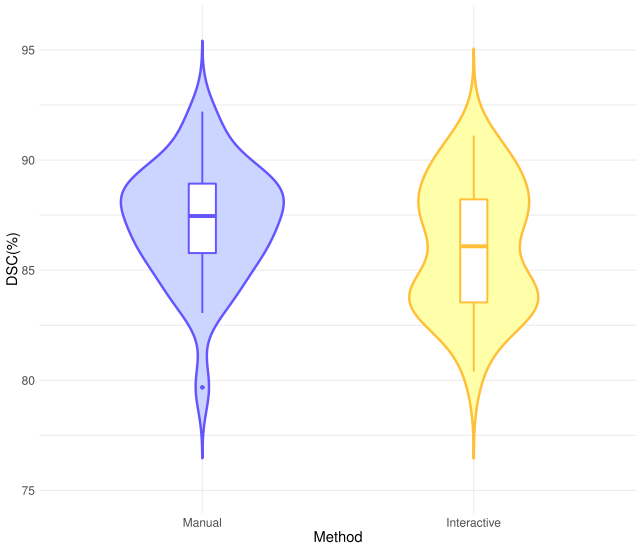


Figure 5. Boxplot of the DSC of the manual segmentation versus the iteractive segmentation on dataset D_2.

which takes about 12 h, 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.

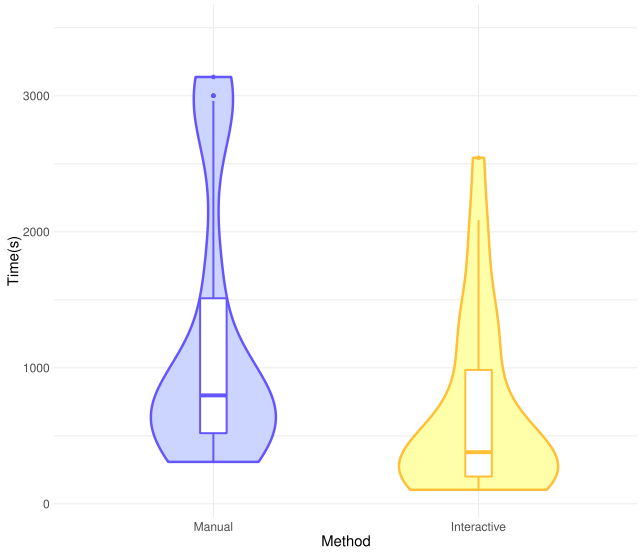


Figure 6. Boxplot of the time required to produce a segmentation for dataset D_2.

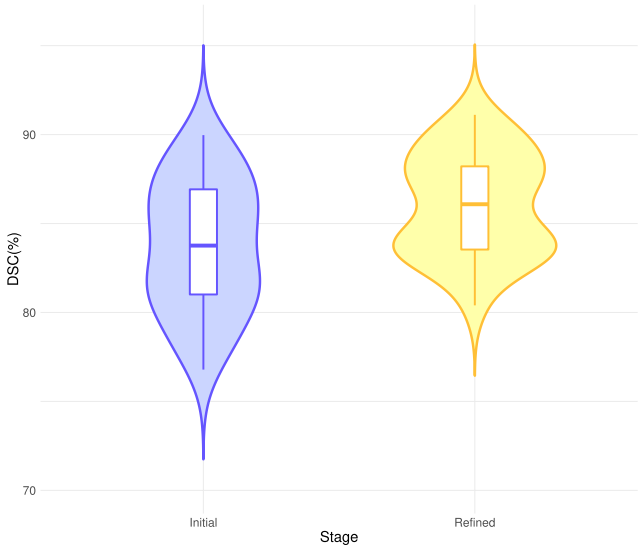


Figure 7. Performance of the fully automatic segmentation versus the interactive segmentation on dataset D_2.

Table 1. DSC values among three 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)			

2.8. Experiment 1: baseline 3D U-net, iFCN and iUNet comparison

In the first experiment we compare the automatic segmentation performance of the (Çiçek *et al* 2016) *3D U-net*, iFCN and our proposed iUNet. The performance is quantified in DSC using a five-fold cross validation on D_1. The dataset is randomly divided into 5 equally sized folds. Four out of five folds of segmentations are used to train while the remaining fifth group is used as the development set. This strategy is repeated 5 times such that the DSC on the development set is computed. The accuracy reported in the paper is the average DSC obtained on the development set after each fold. To substantiate the statistical significance, we also perform a Wilcoxon signed-rank test.

2.9. Experiment 2: iUnet validation

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

A validation protocol is defined to compare the manual segmentation to the iUnet segmentation method. A team consisting of three radiological experts were appointed to perform the segmentations, and are referred to as readers. The ten cases in dataset D_2 are divided into two subsets containing five scans. To minimize the learning effect we alternate the subsets and the segmentation method per reader. Dependant on the segmentation method, specific constraints are defined. The manual segmentation was performed using ITK-snap with full access to the segmentation features. The time measurement of interactive time starts from the moment that the first annotation is placed, and stopped after the segmentation is saved. There is no time constraint to create the segmentation. The interactive segmentation was performed using our custom made framework. Each reader is given two test cases before the start of the validation to get familiar with this tool. At the start of the experiment, the reader gets the instruction to only focus on global features, as adjustments on specific-voxel level would counteract the purpose of the tool. The recording of the time starts from the moment the reader generates the initial segmentation, and is stopped after the segmentation is saved.

A consensus standard reference is produced in order to 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, the performance is determined 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 the satisfactory score of the observer.

3. Results

Experiment 1 resulted in a DSC of 75.3% $\pm$ 8.1% for the Cicek *et al* 3D U-net, a DSC of 72.3% $\pm$ 11.4% for dilated FCN and a DSC of 78.1% $\pm$ 8.7% for the iUnet.

The Wilcoxon signed-rank test found that $p < 0.001$ in the comparison of the manual segmentation versus the interactive U-net, figure 4.

The Wilcoxon signed-rank test found that $p = 0.017$ in the comparison of the manual segmentation versus the interactive U-net, figure 5.

The Wilcoxon signed-rank test found that $p < 0.001$ in the comparison of the manual segmentation time versus the interactive U-net segmentation time, figure 6.

The Wilcoxon signed-rank test found that $p < 0.001$ in the comparsison of the manual initial segmentation versus the interactively refined segmentation, figure 7.

4. Discussion

The results demonstrate that iUnet achieves expert performance in nearly twice the speed of expert radiologists. Observers with iUnet assistance reached 86.0 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 U-Net that showed a significantly higher segmentation performance than iFCN.

In comparison with the most common automatic segmentation method we observe the following. As expected, the standard Cicek UNet (DSC 75.3% $\pm$ 8.1%) showed improved performance over baseline iFCN (72.3% $\pm$ 11.4%). The slightly worse performance than our baseline iUnet (78.1% $\pm$ 8.7%) can be attributed to fine tuning to the problem at hand. Thereby we need to note that we similarly optimized iFCN. Furthermore, both U-net architectures are more robust as demonstrated by the smaller performance deviations and the violin distribution plot showing less outliers. These observations confirm our choice for the development of an interactive version of UNet.

The iUnet reaches a higher DSC for the automatic segmentation on a 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* (2015) 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. Size difference in the data sets plays a minor role in range with the statistical test range.

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 the adjacent tissue. Also, this area contains a lot of anatomical variation 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 an 86.0% agreement (table 1). 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 landmarks 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 (Ronneberger *et al* 2015). 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 for retraining. The algorithm might disregard the adjacent area due to a lack of constraints. Therefore 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 has 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 a 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 should be provided by the user, but they could potentially be obtained by automatic detection to increase efficiency further.

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 can potentially be a novel solution for semi-automatic medical imaging segmentation in general.

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.

References

Åkerberg D, Ansari D, Andersson R and Tingstedt B 2017 The effects of surgical exploration on survival of unresectable pancreatic carcinoma: a retrospective case-control study *J. Biomed. Sci. Eng.* **10** 1–9

Alom Z, Hasan M, Yakopcic C, Taha T M and Asari V K 2018 Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation *Chaos Solitons Fractals* **38** 1411–22

Çiçek O, Abdulkadir A, Lienkamp S S, Brox T and Ronneberger O 2016 3d U-Net: learning dense volumetric segmentation from sparse annotation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Berlin: Springer) pp 424–32

De Fauw J, Ledsam J R, Romera-Paredes B and Ronneberger O 2018 Clinically applicable deep learning for diagnosis and referral in retinal disease *Nat. Med.* **24** 1342–50

Esser P, Sutter E and Ommer B 2018 A variational U-Net for conditional appearance and shape generation (arXiv:1804.04694)

Gibson E, Giganti F, Hu Y and Barratt D C 2018 Multi-organ abdominal CT reference standard segmentations *Zenodo* (https://doi.org/10.1109/TMI.2018.2806309)

Gulliksrud K, Stokke C, Catrine A and Martinsen T 2014 How to measure CT image quality: variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom *Phys. Medica* **30** 521–6
Gulshan V, Peng L, Coram M, Stumpe M C, Wu D and Webster D R 2016 Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs *JAMA* **316** 2402–10
Haenssle H A, Fink C, Schneiderbauer R A and Zalaudek I 2018 Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists *Ann. Oncol.* **29** 1836–42
Integraal kankercentrum Nederland 2018 www.cijfersoverkanker.nl/ (Accessed: 2 February 2019)
Kavzoglu T 2009 Increasing the accuracy of neural network classification using refined training data *Environ. Modelling Softw.* **24** 850–8
Ma X, Hadjiiski L M, Wei J, Chan H P, Cha K H, Cohan R H, Caoili E M, Samala R, Zhou C and Lu Y 2019 U-net based deep learning bladder segmentation in CT urography *Med. Phys.* **46** 1752–65
Marin D, Nelson R C, Ho L M and Youngblood R 2010 Detection of pancreatic tumors, image quality, and radiation dose during the pancreatic parenchymal phase: effect of a low-tube-voltage, high-tube-current CT technique–preliminary results *Radiology* **256** 450–9
Milletari F, Navab N and Ahmadi S A 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation *Proc. 2016 4th Int. Conf. on 3D Vision, 3DV 2016* (Piscataway, NJ: IEEE) pp 565–71
Rahib L, Smith B D, Aizenberg R, Rosenzweig A B, Fleshman J M and Matrisian L M 2014 Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States *Cancer Res.* **74** 2913–21
Rajpurkar P, Irvin J, Zhu K and Ng A Y 2017 CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning (arXiv:1711.05225)
Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation (arXiv:1505.04597)
Roth H R, Farag A, Lu L, Turkbey E B and Summers R M 2015 Deep convolutional networks for pancreas segmentation in CT imaging *Proc. SPIE* **9413** 94131G
Ryan D P, Hong T S and Bardeesy N 2014 Pancreatic adenocarcinoma *New Engl. J. Med.* **371** 1039–49
Wang G *et al* 2018 Interactive medical image segmentation using deep learning with image-specific fine tuning *IEEE Trans. Med. Imaging* **37** 1562–73
Weiss K, Khoshgoftaar T M and Wang D D 2016 *A Survey of Transfer Learning* vol 3 (New York: Springer)
Wolz R, Chu C, Misawa K, Fujiwara M, Mori K and Rueckert D 2013 Automated abdominal multi-organ segmentation with subject-specific atlas generation *IEEE Trans. Med. Imaging* **32** 1723–30