HHS Public Access

Author manuscript

Med Phys. Author manuscript; available in PMC 2021 July 14.

Published in final edited form as:

Med Phys. 2020 September ; 47(9): 4316–4324. doi:10.1002/mp.14386.

CT-based multi-organ segmentation using a 3D self-attention U-net network for pancreatic radiotherapy

Yingzi Liu[*], **Yang Lei**[*], **Yabo Fu**, **Tonghe Wang**

Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

Xiangyang Tang

Department of Radiology and Imaging Sciences and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

Xiaojun Jiang, **Walter J. Curran**, **Tian Liu**, **Pretesh Patel**, **Xiaofeng Yang**[a)]

Department of Radiation Oncology and Winship Cancer Institute, Emory University, Atlanta, GA 30322, USA

Abstract

Purpose: Segmentation of organs-at-risk (OARs) is a weak link in radiotherapeutic treatment planning process because the manual contouring action is labor-intensive and time-consuming. This work aimed to develop a deep learning-based method for rapid and accurate pancreatic multi-organ segmentation that can expedite the treatment planning process.

Methods: We retrospectively investigated one hundred patients with computed tomography (CT) simulation scanned and contours delineated. Eight OARs including large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord and stomach were the target organs to be segmented. The proposed three-dimensional (3D) deep attention U-Net is featured with a deep attention strategy to effectively differentiate multiple organs. Performance of the proposed method was evaluated using six metrics, including Dice similarity coefficient (DSC), sensitivity, specificity, Hausdorff distance 95% (HD95), mean surface distance (MSD) and residual mean square distance (RMSD).

Results: The contours generated by the proposed method closely resemble the ground-truth manual contours, as evidenced by encouraging quantitative results in terms of DSC, sensitivity, specificity, HD95, MSD and RMSD. For DSC, mean values of 0.91 ± 0.03, 0.89 ± 0.06, 0.86 ± 0.06, 0.95 ± 0.02, 0.95 ± 0.02, 0.96 ± 0.01, 0.87 ± 0.05 and 0.93 ± 0.03 were achieved for large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord and stomach, respectively.

Conclusions: The proposed method could significantly expedite the treatment planning process by rapidly segmenting multiple OARs. The method could potentially be used in pancreatic adaptive radiotherapy to increase dose delivery accuracy and minimize gastrointestinal toxicity.

[a)]Author to whom correspondence should be addressed. xiaofeng.yang@emory.edu; Telephone: (404)-778-8622; Fax: (404)-778-4139.

[*]Co-first author.

Keywords

adaptive radiotherapy; multi-organ segmentation; pancreatic radiotherapy; treatment planning

1. INTRODUCTION

Pancreatic cancer is a highly aggressive malignancy with poor prognosis. In the US, while it only accounts for 3% of all cancer cases, pancreatic cancer is the fourth leading cause of cancer-related mortality. The current 5-yr overall survival rate of pancreatic cancer is only 9%.[1] Although surgery provides the best curative option, the majority of patients are unresectable at diagnosis because they present with either locally advanced pancreatic cancer (LAPC) or with distant metastasis.[2] Radiotherapy approaches, with or without chemotherapy, have been frequently used to treat the patients with LAPC. The challenge of controlling pancreatic cancer with radiation is the coexistence of tumor radio-resistance with the radio-sensitivity of the surrounding organs-at-risk (OARs) in the immediate neighborhood such as duodenum and stomach. Thus, high dose is needed but must be carefully controlled. Both standard fractionation in terms of simultaneous-integrated boost (SIB) and stereotactic body radiotherapy (SBRT) have been used for dose escalation. At the site of abdomen, there is considerable patient motion induced by respiration and organ peristalsis, with the complexity compounded by daily variability in stomach filling and bowel gas patterns. Due to the challenges induced by daily anatomical variation, pancreatic radiotherapy is often associated with severe gastrointestinal (GI) toxicities such as nausea, vomiting and diarrhea. Partly due to these GI toxicities, notable weight loss is common in patients who underwent pancreatic radiotherapy, which could in turn degrade the radiation dose delivery accuracy by overdosing the surrounding OARs within the radiation fields. Careful clinical practice of pancreatic radiotherapy treatment planning and delivery are thus crucial to reduce GI toxicities and achieve optimal treatment outcomes.

At the initial radiation treatment planning process, professional expertise is needed in organ contouring. But manual organ contouring is a labor-intensive and time-consuming process that places a heavy demand on the physician. Besides the initial treatment planning, online image guided adaptive pancreatic radiotherapy has been proposed to account for the inter-fractional anatomical variations,[3,4] such as MRI-guided[3,5] and computed tomography (CT)-on-rails-based[6,7] adaptive radiotherapy. For the latter approach, online CT-based multi-organ segmentation is required. In the off-line evaluation work carried out by Liu et al.[6] and Loi et al.,[7] CT-on-rails-based scans were acquired during the course of pancreatic SBRT[7] and standard-fractionated radiotherapy.[6] In their studies, a series of abdominal organs were delineated on the daily CT, either manually or using atlas-based auto-segmentation followed by manual modification. In both cases, their results demonstrated notable benefits of implementation online pancreatic adaptive radiotherapy in OAR dose sparing. However, slow manual/registration-based OAR contouring is impractical for online plan adaptation as it would decrease patient comfort, limit machine throughput and cost much medical resources. In addition, it has shown that human decision-making abilities could be degraded under time pressure,[8] possibly resulting in inaccurate organ delineation. It is thus of primary interest to develop a fully automated tool for simultaneous multi-organ segmentation to

overcome the temporal and logistical pressure. To summarize, either in the current general clinic practice of treatment planning or the future implementation of online adaptive radiotherapy, an automatic CT-based multi-organ segmentation tool that can rapidly and accurately delineate the upper abdominal organs is desirable to achieve optimal outcome of pancreatic radiotherapy.

Multi-organ segmentation has been the subject of extensive study.[9–19] A thorough review of CT-based multi-organ segmentation methods on different body sites can be found in the recent publication of Cerrolaza et al.'s review work.[20] In this paper, we reviewed some closely related work on multi-organ segmentation at the site of abdomen. There are some well-known challenges of CT-based abdominal multi-organ segmentation, including motion artifacts and low contrast organ boundaries such as at the pancreatic head and the 1st portion of duodenum. Additionally, the shapes, positions and appearances of the abdominal organs are subject to large variations due to respiratory motion, bowel movement and fillings. Previously, statistical-shape-models[12,13,17] and multi-atlas label fusion[14,15] were used for multi-organ segmentation, both relying on image registration to establish (atlas) anatomical correspondence. Statistical-shape-model-based methods are constructed from spatially aligned training atlas CT-contours pairs, producing statistical models which contain information about the expected shape and appearance of the organ of interest. The statistical-shape-models are then used to regularize the shapes of the target image segmentation results.[16] For multi-atlas label fusion methods, multiple atlases are registered to the target image, propagating the atlas contours to the target image using either averaging or statistical label fusion. Both methods are limited by the accuracies and performances of the underlying image registration algorithm. Moreover, inter-patient registration is much less accurate for abdominal imaging than other more stationary body sites such as brain, due to the highly variable inter-subject anatomies.

Recently, machine learning-based methods have also been proposed for the task of abdominal multi-organ segmentation. Models were trained to directly predict organ contours after learning the comprehensive relationship between the CT images and the corresponding contours. Compared to the atlas-based methods, machine learning-based methods would be more suitable for upper abdominal organ delineation because of their no need for inter-patient registration. For the abdominal multi-organ segmentation, machine learning-based methods can be mainly grouped into dictionary learning[21] and deep learning-based methods.[10,22,23] The main difference between dictionary learning- and deep learning-based methods is that the latter can learn useful features of the data by itself, eliminating the need for handcrafted features such as atlas selection in dictionary learning.[21] Deep learning-based methods are therefore registration-free, requiring neither explicit anatomical correspondences nor handcrafted image features. There are different categorization methods for deep-learning-based multi-organ segmentation depending on the properties such as network architecture, training process (supervised, semi-supervised, unsupervised or transfer learning), input size [patch-based, whole volume-based, two-dimensional (2D) or three-dimensional (3D)], etc. In terms of the architecture, it can be categorized as auto-encoder (AE), convolutional neural network (CNN) and fully convolutional network (FCN). More specifically, AE usually consists a single neural network encoder layer that transforms the input into a latent or compressed representation and a single neural network decoder layer

that restores the original input from the low-dimensional latent space. Some upgraded networks such as stacked AE (SAE)[24] and denoising AE (DAE)[25] were developed for better higher-level representation and to extract useful features. CNN derives its name from the type of hidden layers it consists of. The hidden layers of a CNN typically consist of convolutional layers, max pooling layers, batch normalization layers, dropout layers and fully connected layers, and normalization layers. The last layer of a CNN is typically a sigmoid or softmax layer for classification and tanh layer for regression. Convolution layers are the core of CNN and are used for feature extraction. More recently, 3D patch-based CNN has been proposed to reserve more spatial information.[26] Fully convolutional network is built based on the CNN network, whose last fully connected layer is replaced by a convolutional layer. Due to the major improvement of deconvolution kernels used to up-sample the feature map, FCN allows the model to have a dense voxel-wise prediction from the full size whole volume instead of a patch-wise classification as in traditional CNN. FCN is thus well-suited to the task of multi-organ segmentation. Although widely varied in the detailed network design, FCN-based methods have achieved state-of-the-art outcomes.[10,22,23] Ronneberger et al. developed U-Net based on FCN to extract contextual information from down-sampling path and more structural information obtained from multiple up-sampling layers.[27] To minimize the adverse effect of inter-patient variation in network training, Oktay et al. used attention gate to highlight salient features useful for the segmentation task.[28] In this work, we propose to use a 3D deep attention U-Net network with a deep supervised compound loss function to perform segmentation. There is no published study on abdomen multi-organ segmentation for pancreatic radiotherapy as far as the authors know. We retrospectively investigated one hundred patients with CT simulation scanned and contours delineated. Eight OARs including large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord, and stomach were target organs to be segmented. To overcome the low contrast between organs such as pancreas and duodenum, deep supervision was integrated into U-Net to force hidden layers learning deep features that can differentiate different organs uniformly. In addition, CT image quality may be affected by motion artifact and several convolution layers may also involve bias. In order to decrease these artifact, and further keep the U-Net focusing on differentiate multiple organs, attention gates were used to reduce the irrelevant features.

2. MATERIALS AND METHODS

2.A. Image acquisition and processing

The study cohort was composed of one hundred pancreatic cancer patients who underwent pancreatic radiotherapy at three clinical sites at our institution from the year 2015 to 2019. CT scans were acquired on four different CT simulators including SOMATOM Definition AS and SOMATOM Definition Edge (Siemens Healthcare, Erlangen, Germany), Brilliance Big Bore (Philips Healthcare, Best, The Netherlands) and LightSpeed RT (GE Medical Systems, Milwaukee, USA). Tube voltage of 100 kVp was used for eleven patients with intravenous contrast injection and 120 kVp were applied for the rest eighty-nine patients. The voxel sizes in left-right and anterior-posterior were around 1 mm (range from 0.90 to 1.56 mm) with 512 voxels in total, and the slice thicknesses in superior-inferior direction range from 1.5 to 3 mm with total scan lengths ranging from 216 to 712.5 mm.

Eight OARs including large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord and stomach were contoured on CT images by physicians in our department during the initial treatment planning process. In this study, two expert physicians went over all the contours and had an agreement with organ delineations for each patient. In other words, only one set of contours was delineated for each patient as a consensus achieved between two physicians. The OAR contours coupled with the corresponding CT images were used as the training dataset for our proposed deep learning-based method. Fivefold cross-validation was performed.

2.B. 3D deep attention U-Net network for pancreatic multi-organ segmentation

For a given pair of pancreatic CT image and its corresponding manual OAR contours, the manual contours were used as the learning targets of the CT image. The training data were 3D patches of CT-contour image pairs. The training patches covered 32 slices of the full 2D images along the superior-inferior direction (sliding window 512 × 512 × 32 voxels) with a step size of 16 slices between two neighboring patches. This study aimed to segment eight OARs, therefore the proposed method is nine-label-based (including background) segmentation model to simultaneously segment small bowel, large bowel, liver, stomach, spinal cord, left kidney, right kidney and duodenum. Figure 1 outlines the schematic workflow of our proposed deep attention U-Net method. The whole architecture includes a training stage and a segmentation stage. The training U-Net network consists of encoding and decoding paths. The encoding path is constructed by three down-sampling blocks, each block being composed of three convolution layers with the first layer followed by a max pooling layer. The decoding path is constructed by three up-sampling blocks, with each block being composed of a deconvolution layer to enhance the resolution followed by three convolutional layers. Long skip connection concatenates the feature maps from the encoding path and two former decoding paths. As is shown in the schematic flow-chart of attention gate demonstrated in Fig. 1(b), the feature maps extracted from the coarse scale are used in gating to disambiguate irrelevant and noisy responses in skip connections. U-Net with attention gate encourages each path to obtain both high-frequency information (such as textural information) and low-frequency information (such as structural information) to represent an image patch. At the end of the architecture, probability map of contours is generated by softmax operators. Binary masks are generated by using a threshold to the probability map, with each organ being assigned with a unique label.

To deal with limited training data, we employed a deep supervision strategy[29] with a compound loss function.[30] As shown in Fig. 1(a), deep supervision was used to supervise both the final prediction and each of the decoding stages. Feature maps extracted from the lower and middle stages pass through deconvolution, ReLU, and sigmoid functions to output images that have the same size as the input images. Compared with the attention U-Net proposed by Oktay et al.,[28] this deep supervision mechanism could make the residual information semantically meaningful for both the early and final stages in the architecture, which can reduce the convergence time and improve the segmentation performance. In the compound loss function, weighted cross-entropy and dice similarity coefficient (DSC) were combined. More specifically, cross-entropy is widely used in quantifying the dissimilarity between two probability distributions. Let $G = \{g_1, \ldots, g_N\}$ be the set of ground-truth labels

and $P = \{p_1, \ldots, p_N\}$ be the set of predicted labels, where N represents the number of voxels in a volume. Cross-entropy and its partial derivatives used within the gradient for the j-th voxel of the prediction can be directly defined as follows. C_E is short for cross-entropy.

$$C_E = -\sum_{i=1}^{N} g_i \log p_i - \sum_{i=1}^{N} (1 - g_i)\log(1 - p_i)$$

$$\frac{\partial C_E}{\partial p_j} = \frac{1 - g_j}{1 - p_j} - \frac{g_j}{p_j}$$

On the other hand, DSC is an overlap metric commonly used for assessing the quality of segmentation maps. DSC formulation and its partial derivatives providing the gradient for the j-th voxel of the prediction can be defined as follows, where D is short for DSC.

$$D = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N}\left(p_i^2 + g_i^2\right)}$$

$$\frac{\partial D}{\partial p_j} = \frac{2g_j\sum_{i=1}^{N}\left(p_i^2 + g_i^2\right) - 4p_j\sum_{i=1}^{N} p_i g_i}{\left(\sum_{i=1}^{N}\left(p_i^2 + g_i^2\right)\right)^2}$$

To enable the network to penalize dissimilarity (cross-entropy) and encourage similarity (DSC) between the prediction and training data simultaneously, cross-entropy and DSC were combined into the loss functions of all expanding stages, and the final loss function is defined as follows:

$$L_{\text{final}}(P, G) = \sum_{t=1}^{3} \gamma_t (L_{CE}(P, G) + \beta L_{DSC}(P, G))$$

L_{CE} and L_{DSC} represent cross-entropy and DSC loss function. Parameter β was used to balance the two losses. γ_t denotes the balancing weights of loss in each stage. γ_t should be provided according to different resolution of each stage of deep supervision. We used relatively low weight for the stage of low resolution and the highest weight for the final stage. When γ_t is set to be $\gamma_t = \{0.8^{t-1}\}$, $t \in \{1, 2, 3\}$, we can obtain the best performance in terms of DSC. The range of β was set to 0.5. In Lei et al.'s work on ultrasound-based prostate segmentation, the combination of logistic loss[27] and Dice loss[31] has been proved to be more accurate in prostate segmentation than using either loss function alone.

After training, the 3D patches of the testing CT images are fed into the final well-trained model to retrieve the predicted probability maps. These probability maps were then fused to the whole image. Averaging fusion method on overlapped regions were used for patch

fusion. The final segmentations were obtained by aggregating the multi-organ probability maps and a consolidation via averaging.

We used fivefold cross-validation for network training and testing. The 100 patients' datasets were divided into five groups, with 20 datasets in each group. One group was used as the testing dataset while the remaining 80 datasets were used as training sets. This process was repeated by five times to allow all 100 datasets to be tested once. Data augmentation such as flip, rotation, scaling was used to enlarge the variety of the training datasets. The augmentation includes rotation (0°, 30°, 60°, 90°), flipping (not flip, flip), scaling (1, 1.25, 1.5) and affine warp (original + 2 affine warp), resulting in 72 times more training datasets. The learning rate for Adam optimizer was set to 2e-4, and the model was trained and tested on a NVIDIA Tesla V100 GPU with 32 GB memory. A batch size of 20 was used. 10 GB CPU memory and 29 GB GPU memory was used for each batch optimization. The algorithm was implemented by python3.7 and Tensor-flow 1.14. The training was stopped after 200 epochs. After training was finished, the automatic generation of multi-organ contours can be done in <1 s.

2.C. Evaluation Matrices

The performance of our method was evaluated using six commonly used metrics, including DSC,[32] sensitivity,[32] specificity,[32] Hausdorff distance 95% (HD95),[33] mean surface distance (MSD),[33] and residual mean square distance (RMSD).[30] Dice similarity coefficient provides information of spatial overlap.

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

where X and Y are the manual and predicted masks, respectively.

Sensitivity/specificity measure the correct classification rate. They quantify the contour overlapping ratio inside the manual contour volume and the no-contour overlapping ratio outside the manual contour volume, respectively.

$$Sensitivity = \frac{|X \cap Y|}{|X|}$$

$$Specificity = \frac{|\bar{X} \cap \bar{Y}|}{|\bar{X}|}$$

HD95/MSD/RMSD measure the boundary similarity between the manual contour and the predicted contour. They quantify the maximum of 95th percentile, the average and the root of mean square deviation distance of all distances between points in X (the predicted contours) and the nearest point in Y (the manual contours), respectively.

$$HD95 = max_{k95\%}[d(X, Y), d(Y, X)].$$

$$MSD = \frac{1}{|X| + |Y|}\left(\sum_{x \in X} d(x, Y) + \sum_{y \in Y} d(y, X)\right).$$

$$RMSD = \sqrt{\frac{1}{|X| + |Y|}\left(\sum_{x \in X} d(x, Y)^2 + \sum_{y \in Y} d(y, X)^2\right)}$$

where $d(x, Y) = min_{y \in Y}\|x - y_2\|$. Doing this for all pixels in the surfaces gives the total surface distance between X and Y: (X, Y).

3. RESULTS

Figure 2 shows the 2D and 3D segmentation results from one representative patient using the proposed deep attention U-Net network, with side-by-side comparison to the corresponding CT and manual contours. Our method successfully delineated the eight OARs with good visual agreement to the manual delineation. Manual contours in the superior-inferior direction are often not smooth since manual contours were delineated in 2D in the axial plan. On the contrary, automatic segmentation results are smoother than manual contours since the contour was predicted in 3D.

The quantitative evaluation using DSC, sensitivity, specificity, HD95, MSD, and RMSD is summarized in Table I and Fig. 3. For DSC, mean values of 0.91 ± 0.03, 0.89 ± 0.06, 0.86 ± 0.06, 0.95 ± 0.02, 0.95 ± 0.02, 0.96 ± 0.01, 0.87 ± 0.05 and 0.93 ± 0.03 were achieved for large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord and stomach, respectively. In Fig. 3, some extreme outliers were plotted in separate figures showing with larger scales. Those outliers were generated by our method mainly due to: inter-observer variability of the training data, incomplete or over-prediction of organs in the most superior and inferior slices comparing to the manual contours, and error in prediction that can be caused by the algorithm itself and the different image acquiring protocols. The mean values for HD95 were 4.94 ± 2.57, 6.97 ± 5.98, 4.17 ± 1.74, 3.42 ± 2.37, 2.97 ± 2.18, 4.38 ± 2.32, 4.22 ± 6.82, 6.40 ± 6.99 mm for large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord and stomach, respectively. Some outliers of close to 40 mm were observed mainly due to the under- or over-prediction of the contours in the most superior and inferior slices. For some patients, a small extra volume of contour was generated by our algorithm that could significantly affect the HD95 value. Our MSD values were comparable to the state-of-the-art method.[10]

To demonstrate the effectiveness of each component that was incorporated in the U-Net-based network, an ablation study was performed. Specifically, the performances in terms of DSC and HD95 were compared among U-Net, U-Net with deep supervision (Unet-DS), Unet-DS with compound loss function (Unet-DS-CL), and Unet-DS-CL with attention gate (Unet-DS-CL-AG). Table II tabulates the results. As shown in the table, the adding features generally improve the performance of segmentation by showing higher DSC and smaller HD95. Table III lists *P*-values of the paired-two-tailed *t*-test between different variants of the

U-Net-based methods. As shown, significant differences ($P < 0.05$) were found between two variants of U-Net-based methods for most of the organs, which means the improvement is statistically significant.

4. DISCUSSION

This work automatically generated eight OARs including large bowel, small bowel, duodenum, left kidney, right kidney, liver, spinal cord, and stomach from CT images in one single forward network prediction. The overall DSC scores are encouraging and have proved that our method is suitable for future development in its implementation in clinic. Some outliers were present in our results. We checked those extreme cases slice-by-slice and summarized the reasons as follows: (a) The network failed to predict some organs in some slices. It can be caused by the variations in CT image qualities. The use of contrast during CT acquisition can improve the accuracy of both manual delineation and automatic segmentation. For example, most of the patients were injected with contrast in stomach. The network performed less accurately to generate the stomach contour for some patients without stomach contrast injection. (b) Due to image noise and artifacts, some spurious labeling was present in the segmentation results. (c) The CT image field ranges from 216 to 712.5 mm in superior-inferior direction. In the CT images with larger field, some of the organ (e.g. spinal cord) manual contours were missing at the most superior and inferior slices while they were automatically generated by our algorithm.

Despite some outliers, our multi-organ segmentation results were on average in good agreement with the manual contours. In this study, we have included eight OARs that are commonly used in pancreatic treatment planning in our clinic. Comparison between our method and the state-of-the-art methods demonstrated that our method achieved higher accuracy in bowel and duodenum segmentation and comparable accuracy in other organs. The segmentation accuracy of bowel and duodenum is important since they are the most concerned OARs in pancreatic radiotherapy. It is worth noting that image quality can affect both the manual delineation and the machine learning segmentation results. The CT images used by the other studies listed in Table II were acquired during the radiological process for disease diagnosis, which features finer voxel size and higher exposure dose than the CT scans used for the radiotherapy treatment planning purpose. This might partly explain why for some other organs such as liver and stomach, our results were not as good as the study done by Wang et al.[10] The other reasons include that different methods may excel in different organ segmentation tasks, the number of training data included, and the uniformity of contrast injection.

Automatic multi-organ segmentation can significantly shorten the process of treatment planning and plays an important role the development of pancreatic adaptive radiotherapy. For adaptive radiotherapy, besides CT-on-rails-based image guided system, there is potential use of CBCT-guided[34,35] adaptive imaging systems. In the CBCT-based scenario, the quality of CBCT images could be improved using correction methods to reduce image artifacts and restore clear organ boundaries, as well as to improve the HU accuracy that is comparable to the real CT. The corrected CBCT can thus be considered as the surrogate of real CT images.

For the future work, our deep attention U-Net method could be applied on the CBCT-based synthetic CT to segment OARs for adaptive radiotherapy.

The eight OARs included in this study were based on our institutional protocol for pancreatic radiotherapy. They were determined based on RTOG consensus panel 2012.[36] Some other organs such as spleen, pancreas and volumes of interest (VOIs) such as pancreatic lesion/tumor bed, portal vein (PV), superior mesenteric artery (SMA), and celiac artery (CA) were not included in this study. Future directions would include these organs and VOIs.

Since the application of machine learning in clinic has little tolerance for mistake, human inspection would still be required after the automatic generation of contours. But the total time required for inspection and modification is supposed to be much less than the time spent on manual contouring, providing that the model is with high accuracy and robustness.

5. CONCLUSIONS

In this study, we have developed a deep attention U-Net method for multi-organ segmentation for pancreatic radiotherapy treatment planning. The contours generated by our method closely resemble the ground-truth manual contours, as evidenced by promising quantitative results in terms of DSC, sensitivity, specificity, HD95, MSD and RMSD. This method could significantly expedite the treatment planning process. It can also facilitate the development of pancreatic adaptive radiotherapy, potentially increase dose delivery accuracy, and minimize gastrointestinal toxicity.

ACKNOWLEDGMENTS

This research is supported in part by the National Cancer Institute of the National Institutes of Health under Award Number R01CA215718 and R01EB028324, and Emory Winship Cancer Institute Pilot grant.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA: Cancer J Clin. 2019;69:7–34. [PubMed: 30620402]
2. Brunner TB, Nestle U, Grosu A-L, Partridge M. SBRT in pancreatic cancer: what is the therapeutic window? Radiother Oncol. 2015;114:109–116. [PubMed: 25466369]
3. Boldrini L, Cusumano D, Cellini F, Azario L, Mattiucci GC, Valentini V. Online adaptive magnetic resonance guided radiotherapy for pancreatic cancer: state of the art, pearls and pitfalls. Radiat Oncol. 2019;14:71. [PubMed: 31036034]
4. Lamb J, Cao M, Kishan A, et al. Online adaptive radiation therapy: implementation of a new process of care. Cureus. 2017;9:e1618. [PubMed: 29104835]
5. Olberg S, Green O, Cai B, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy (MR-IGRT) of pancreatic cancer. Radiat Oncol. 2018;13:51. [PubMed: 29573744]
6. Liu F, Erickson B, Peng C, Li XA. Characterization and management of interfractional anatomic changes for pancreatic cancer radiotherapy. Int J Radiat Oncol Biol Phys. 2012;83:e423–e429. [PubMed: 22436785]
7. Loi M, Magallon-Baro A, Suker M, et al. Pancreatic cancer treated with SBRT: effect of anatomical interfraction variations on dose to organs at risk. Radiother Oncol. 2019;134:67–73. [PubMed: 31005226]

8. Tyran M, Jiang N, Cao M, et al. Retrospective evaluation of decision-making for pancreatic stereotactic MR-guided adaptive radiotherapy. Radiother Oncol. 2018;129:319–325. [PubMed: 30174107]

9. Dong X, Lei Y, Wang T, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys. 2019;46:2157–2168. [PubMed: 30810231]

10. Wang Y, Zhou Y, Shen W, Park S, Fishman EK, Yuille AL. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Med Image Anal. 2019;55:88–102. [PubMed: 31035060]

11. Li D, Zang P, Chai X, Cui Y, Li R, Xing L. Automatic multiorgan segmentation in CT images of the male pelvis using region-specific hierarchical appearance cluster models. Med Phys. 2016;43:5426–5436. [PubMed: 27782723]

12. Hyunjin P, Bland PH, Meyer CR. Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Trans Med Imaging. 2003;22:483–492. [PubMed: 12774894]

13. Linguraru MG, Pura JA, Pamulapati V, Summers RM. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. Med Image Anal. 2012;16:904–914. [PubMed: 22377657]

14. Wolz R, Chu C, Misawa K, Fujiwara M, Mori K, Rueckert D. Automated abdominal multi-organ segmentation with subject-specific atlas generation. IEEE Trans Med Imaging. 2013;32:1723–1730. [PubMed: 23744670]

15. Chu C, Oda M, Kitasaka T, et al. Multi-organ Segmentation Based on Spatially-Divided Probabilistic Atlas from 3D Abdominal CT Images. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013; 2013//; 2013; Berlin, Heidelberg

16. Heimann T, Meinzer H-P. Statistical shape models for 3D medical image segmentation: a review. Med Image Anal. 2009;13:543–563. [PubMed: 19525140]

17. Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y. Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. Med Image Anal. 2015;26:1–18. [PubMed: 26277022]

18. Lei Y, Liu Y, Dong X, et al. Automatic multi-organ segmentation in thorax CT images using U-Net-GAN. Vol 10950: SPIE; 2019.

19. Fu Y, Mazur TR, Wu X, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. Med Phys. 2018;45:5129–5137. [PubMed: 30269345]

20. Cerrolaza JJ, Picazo ML, Humbert L, et al. Computational anatomy for multi-organ analysis in medical imaging: a review. Med Image Anal. 2019;56:44–67. [PubMed: 31181343]

21. Tong T, Wolz R, Wang Z, et al. Discriminative dictionary learning for abdominal multi-organ segmentation. Med Image Anal. 2015;23:92–104. [PubMed: 25988490]

22. Gibson E, Giganti F, Hu Y, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans Med Imaging. 2018;37:1822–1834. [PubMed: 29994628]

23. Zhou X, Ito T, Takayama R, Wang S, Hara T, Fujita H. Three-Dimensional CT Image Segmentation by Combining 2D Fully Convolutional Network with 3D Majority Voting. Paper presented at: Deep Learning and Data Labeling for Medical Applications; 2016//, 2016; Cham.

24. Shin H, Orton MR, Collins DJ, Doran SJ, Leach MO. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell. 2013;35:1930–1943. [PubMed: 23787345]

25. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 2010;11:3371–3408.

26. Pezeshk A, Hamidian S, Petrick N, Sahiner B. 3-D convolutional neural networks for automatic detection of pulmonary nodules in chest CT. IEEE J Biomed Health Inform. 2019;23:2080–2090. [PubMed: 30418929]

27. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Paper presented at: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; 2015//, 2015; Cham.

28. Oktay O, Schlemper J, Folgoc LL, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999; 2018.
29. Dou Q, Yu L, Chen H, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. Med Image Anal. 2017;41:40–54. [PubMed: 28526212]
30. Lei Y, Tian S, He X, et al. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Med Phys. 2019;46:3194–3206. [PubMed: 31074513]
31. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. Paper presented at: 2016 Fourth International Conference on 3D Vision (3DV); 2016.
32. Wang T, Lei Y, Tian S, et al. Learning-based automatic segmentation of arteriovenous malformations on contrast CT images in brain stereotactic radiosurgery. Med Phys. 2019;46:3133–3141. [PubMed: 31050804]
33. Wang B, Lei Y, Tian S, et al. Deeply supervised 3D fully convolutional networks with group dilated convolution for automatic MRI prostate segmentation. Med Phys. 2019;46:1707–1718. [PubMed: 30702759]
34. Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. Med Phys. 2019;46:3998–4009. [PubMed: 31206709]
35. Lei Y, Tang X, Higgins K, et al. Learning-based CBCT correction using alternating random forest based on auto-context model. Med Phys. 2019;46:601–618. [PubMed: 30471129]
36. Goodman KA, Regine WF, Dawson LA, et al. Radiation therapy oncology group consensus panel guidelines for the delineation of the clinical target volume in the postoperative treatment of pancreatic head cancer. Int J Radiat Oncol Biol Phys. 2012;83:901–908. [PubMed: 22483737]

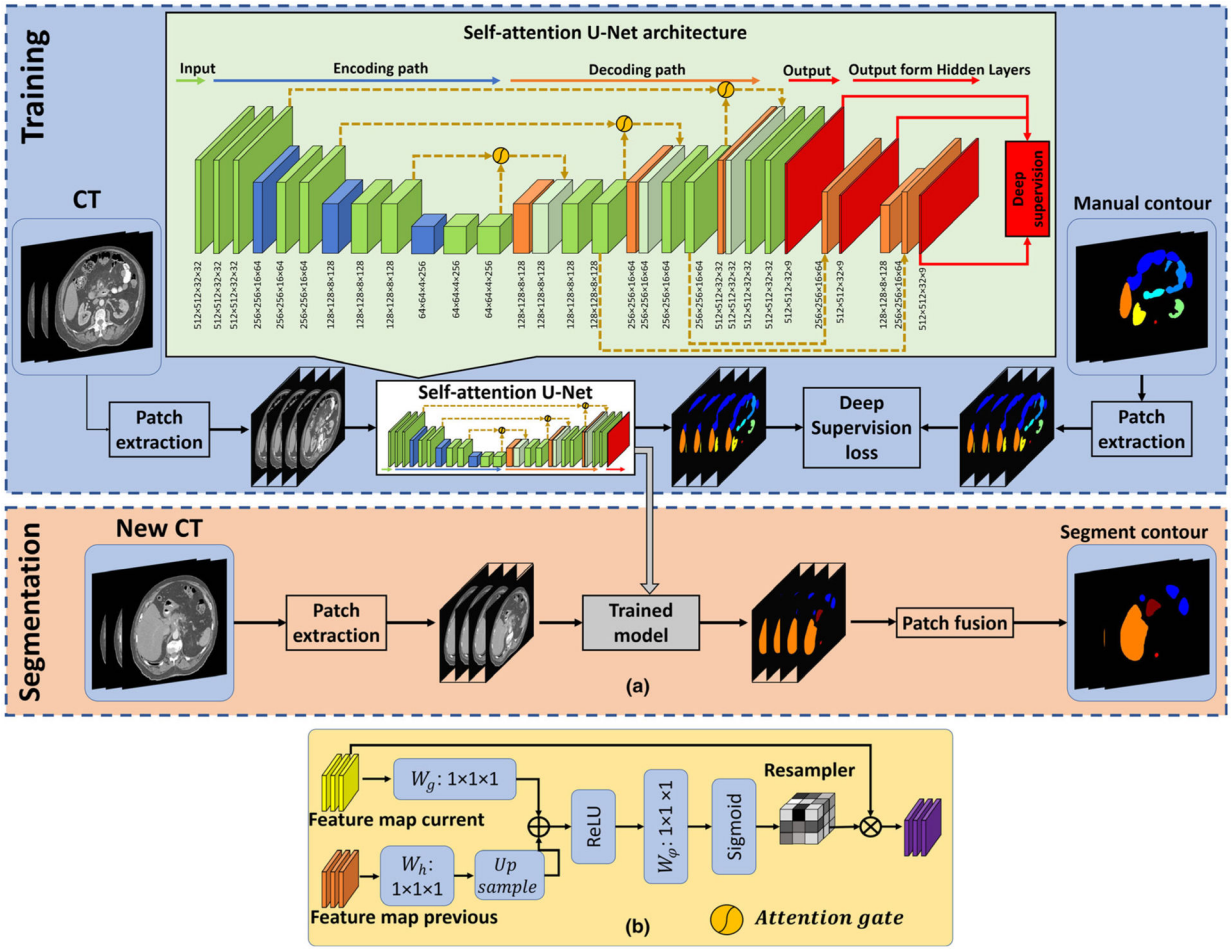

Fig. 1.

(a) Schematic flow-chart of the proposed algorithm for pancreatic multi-organ segmentation. The first row shows the training procedure of deep attention U-Net. The second row shows the segmentation of a new arrival patient. (b) Architecture of the attention gate.

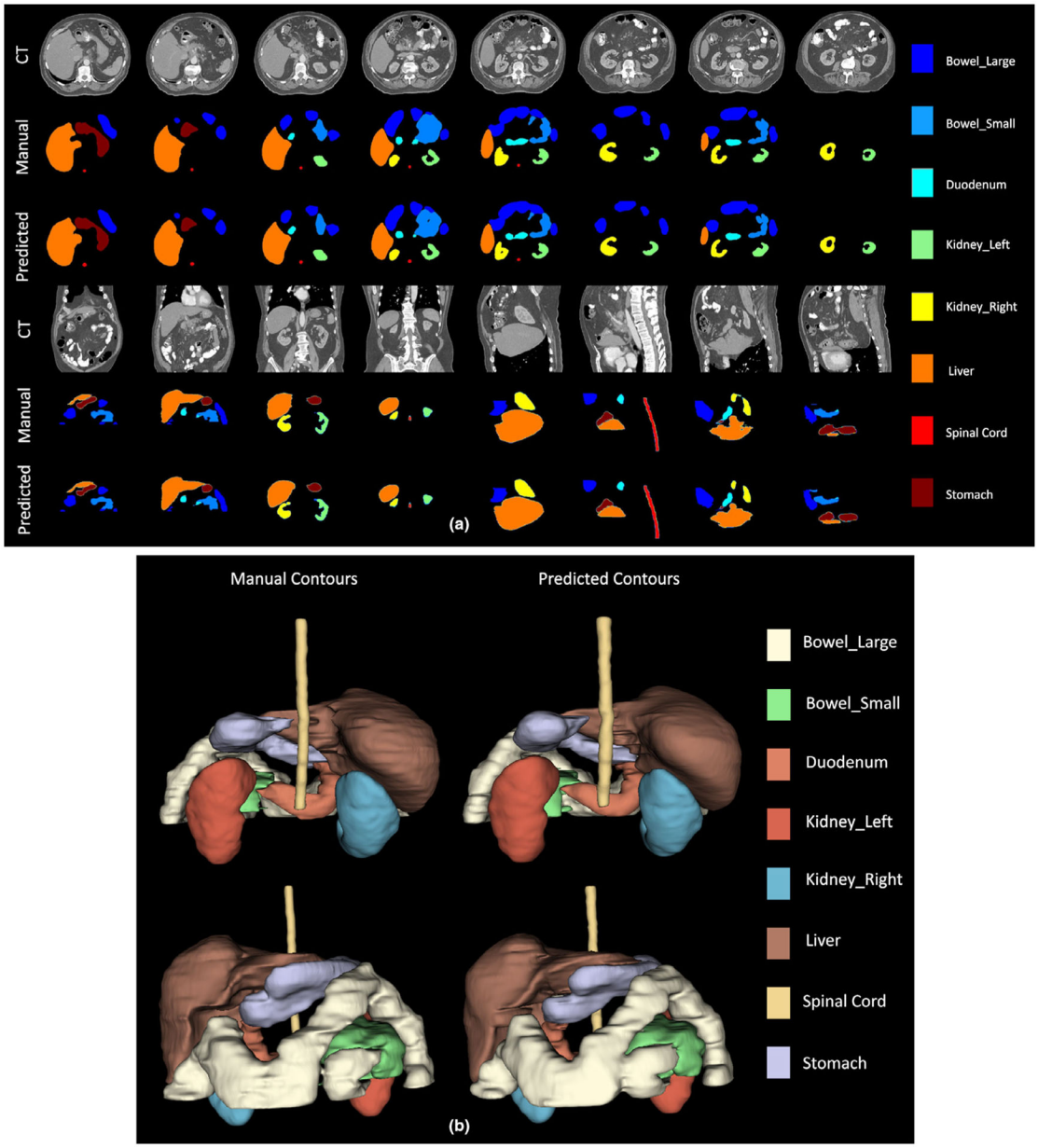

Fig. 2.

(a) Axial, coronal and sagittal two-dimensional visualization and (b) posterior and anterior three-dimensional visualization of the manual and predicted organs-at-risk contours of a representative patient.

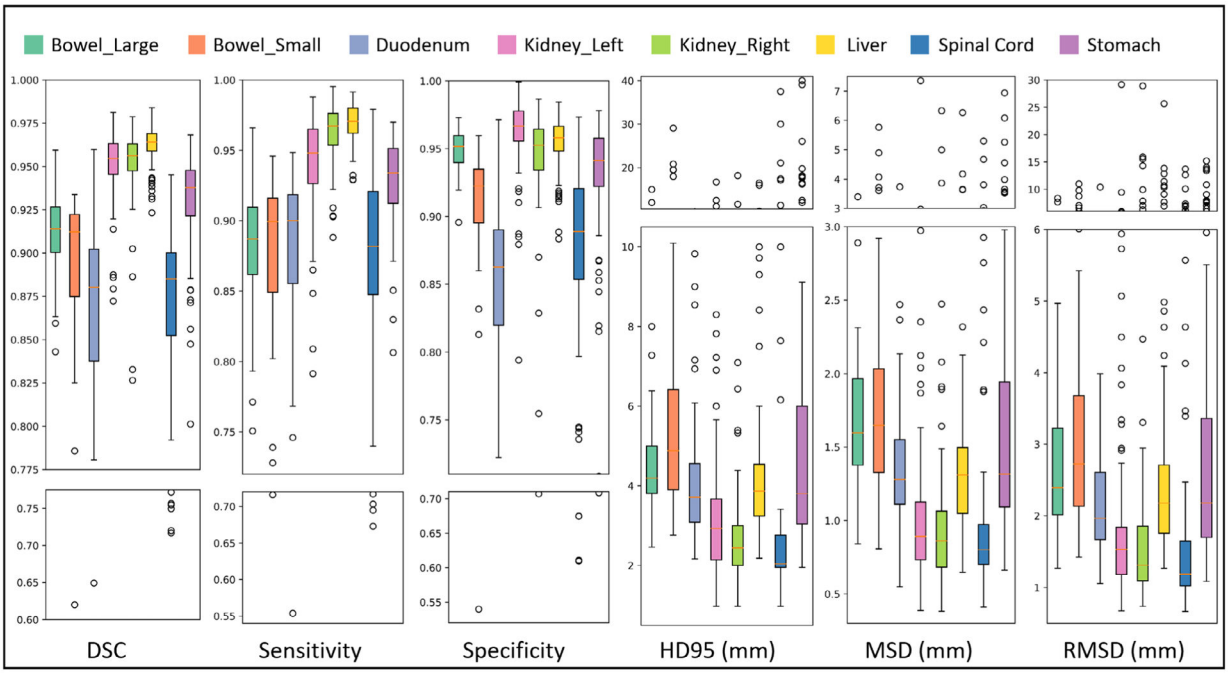

Fig. 3.

Box and whisker plots of the evaluation matrices including of dice similarity coefficient, sensitivity, specificity, Hausdorff distance 95% (HD95), mean surface distance (MSD), residual MSD (RMSD) between the manual contours and the contours predicted by our proposed method. The central orange line indicates the median value, and the borders of the box represent the 25th and 75th percentiles. The whiskers show inliers within 1.5 inter-quartile ranges. The outliers are plotted by the black “O” marker.

Table I.

Mean ± standard deviation, and median (minimum-maximum) value of dice similarity coefficient (DSC), sensitivity, specificity, Hausdorff distance (HD), mean surface distance (MSD), residual MSD (RMSD) between the manual contours and the contours predicted by our proposed method.

	DSC	Sensitivity	Specificity	HD95 (mm)	MSD (mm)	RMSD (mm)
Bowel Large	0.91 ± 0.03	0.88 ± 0.05	0.95 ± 0.02	4.94 ± 2.57	1.67 ± 0.55	2.90 ± 1.57
	0.91 (0.84, 0.96)	0.89 (0.75, 0.97)	0.95 (0.90, 0.97)	4.19 (2.46, 15.00)	1.60 (0.84, 3.40)	2.39 (1.27, 8.37)
Bowel Small	0.89 ± 0.06	0.88 ± 0.06	0.90 ± 0.07	6.97 ± 5.98	1.99 ± 1.08	3.55 ± 2.29
	0.91 (0.62, 0.93)	0.90 (0.72, 0.95)	0.92 (0.54, 0.96)	4.88 (2.76, 29.09)	1.65 (0.81, 5.76)	2.73 (1.43, 11.00)
Duodenum	0.86 ± 0.06	0.88 ± 0.07	0.85 ± 0.06	4.17 ± 1.74	1.39 ± 0.54	2.31 ± 1.41
	0.88 (0.65, 0.96)	0.90 (0.55, 0.95)	0.86 (0.72, 0.97)	3.71 (2.17, 9.83)	1.28 (0.55, 3.74)	1.97 (1.06, 10.42)
Kidney Left	0.95 ± 0.02	0.94 ± 0.04	0.96 ± 0.03	3.42 ± 2.37	1.06 ± 0.79	2.18 ± 3.15
	0.95 (0.87, 0.98)	0.95 (0.79, 0.99)	0.97 (0.79, 1.00)	2.93 (0.98, 16.73)	0.89 (0.39, 7.35)	1.53 (0.67, 29.16)
Kidney Right	0.95 ± 0.02	0.96 ± 0.02	0.94 ± 0.04	2.97 ± 2.18	1.05 ± 0.86	2.52 ± 4.02
	0.96 (0.83, 0.98)	0.97 (0.89, 1.00)	0.95 (0.71, 0.99)	2.44 (0.98, 18.20)	0.86 (0.38, 6.33)	1.31 (0.74, 28.92)
Liver	0.96 ± 0.01	0.97 ± 0.01	0.95 ± 0.02	4.38 ± 2.32	1.45 ± 0.80	3.26 ± 3.48
	0.96 (0.92, 0.98)	0.97 (0.93, 0.99)	0.96 (0.88, 0.98)	3.87 (2.18, 16.48)	1.31 (0.65, 6.26)	2.18 (1.27, 25.65)
Spinal Cord	0.87 ± 0.05	0.87 ± 0.07	0.88 ± 0.07	4.22 ± 6.82	1.07 ± 0.84	2.05 ± 2.40
	0.89 (0.72, 0.95)	0.88 (0.67, 0.98)	0.89 (0.61, 0.97)	2.04 (0.98, 37.52)	0.80 (0.41, 5.29)	1.18 (0.67, 13.74)
Stomach	0.93 ± 0.03	0.93 ± 0.03	0.93 ± 0.04	6.40 ± 6.99	1.77 ± 1.19	3.46 ± 3.13
	0.94 (0.80, 0.97)	0.93 (0.81, 0.97)	0.94 (0.71, 0.98)	3.81 (1.95, 40.02)	1.32 (0.66, 6.94)	2.18 (1.08, 15.21)

Table II.

Statistical comparison in terms of dice similarity coefficient (DSC) and Hausdorff distance 95% (HD95) among methods with and without deep supervision (DS), compound loss (DL) and attention gate (AG) incorporating to the U-Net network.

	U-Net		Unet-DS		Unet-DS-CL		Unet-DS-CL-AG	
	DSC	HD95 (mm)	DSC	HD95 (mm)	DSC	HD95 (mm)	DSC	HD95 (mm)
Bowel large	0.88 ± 0.10	6.47 ± 6.47	0.90 ± 0.04	5.17 ± 5.17	0.89 ± 0.03	5.95 ± 5.95	0.91 ± 0.03	4.94 ± 2.57
Bowel small	0.85 ± 0.15	8.19 ± 8.19	0.86 ± 0.08	8.45 ± 8.45	0.88 ± 0.06	9.64 ± 9.64	0.89 ± 0.06	6.97 ± 5.98
Duodenum	0.82 ± 0.07	6.61 ± 6.61	0.83 ± 0.09	7.14 ± 7.14	0.83 ± 0.09	6.52 ± 6.52	0.86 ± 0.06	4.17 ± 1.74
Kidney left	0.94 ± 0.03	6.45 ± 6.45	0.93 ± 0.04	8.63 ± 8.63	0.94 ± 0.02	7.18 ± 7.18	0.95 ± 0.02	3.42 ± 2.37
Kidney right	0.94 ± 0.03	3.70 ± 3.70	0.94 ± 0.05	3.08 ± 3.08	0.95 ± 0.02	2.89 ± 2.89	0.95 ± 0.02	2.97 ± 2.18
Liver	0.95 ± 0.02	9.44 ± 9.44	0.96 ± 0.02	7.43 ± 7.43	0.96 ± 0.02	4.70 ± 4.70	0.96 ± 0.01	4.38 ± 2.32
Spinal cord	0.84 ± 0.06	6.52 ± 6.52	0.86 ± 0.05	5.63 ± 5.63	0.86 ± 0.05	4.53 ± 4.53	0.87 ± 0.05	4.22 ± 6.82
Stomach	0.90 ± 0.11	10.98 ± 11.04	0.91 ± 0.11	10.08 ± 10.15	0.91 ± 0.11	8.23 ± 8.28	0.93 ± 0.03	6.40 ± 6.99

Table III.

P-value comparison in terms of dice similarity coefficient (DSC) and Hausdorff distance 95% (HD95) between proposed two variants of U-Net-based methods.

	U-Net vs Unet-DS		Unet-DS vs Unet-DS-CL		Unet-DS-CL vs Unet-DS-CL-AG	
	DSC	**HD95**	**DSC**	**HD95**	**DSC**	**HD95**
Bowel large	<0.001	<0.001	0.894	0.004	<0.001	<0.001
Bowel small	0.052	0.018	<0.001	<0.001	0.029	0.018
Duodenum	0.067	<0.001	0.258	<0.001	<0.001	<0.001
Kidney left	0.027	<0.001	0.046	<0.001	0.027	<0.001
Kidney right	0.294	0.002	0.137	0.018	0.530	0.127
Liver	0.008	<0.001	0.850	<0.001	0.284	0.035
Spinal cord	<0.001	<0.001	0.332	<0.001	0.039	0.041
Stomach	0.004	<0.001	0.425	<0.001	<0.001	<0.001