Impact of Denoising on Deep-Learning-Based Automatic Segmentation Framework for Breast Cancer Radiotherapy Planning

Jung Ho Im 1,†, Ik Jae Lee 2,†, Yeonho Choi 3, Jiwon Sung 2, Jin Sook Ha 3 and Ho Lee 2,∗

1 CHA Bundang Medical Center, Department of Radiation Oncology, CHA University School of Medicine, Seongnam 13496, Korea; junghosta@naver.com
2 Department of Radiation Oncology, Yonsei University College of Medicine, Seoul 03722, Korea; ikjae412@yuhs.ac (I.J.L.); pokemon30@yuhs.ac (J.S.)
3 Department of Radiation Oncology, Gangnam Severance Hospital, Seoul 06273, Korea; yunhoc12@yuhs.ac (Y.C.); hjshjs83@yuhs.ac (J.S.H.)
∗ Correspondence: holee@yuhs.ac; Tel.: +82-2-2228-8109; Fax: +82-2-2227-7823
† These authors contributed equally to the work.

Simple Summary: We investigated the contouring data of organs at risk from 40 patients with breast cancer who underwent radiotherapy. The performance of denoising-based auto-segmentation was compared with manual segmentation and conventional deep-learning-based auto-segmentation without denoising. Denoising-based auto-segmentation achieved superior segmentation accuracy on the liver compared with AccuContour™-based auto-segmentation. This denoising-based auto-segmentation method could provide more precise contour delineation of the liver and reduce the clinical workload.

Abstract: Objective: This study aimed to investigate the segmentation accuracy of organs at risk (OARs) when denoised computed tomography (CT) images are used as input data for a deep-learning-based auto-segmentation framework. Methods: We used non-contrast enhanced planning CT scans from 40 patients with breast cancer. The heart, lungs, esophagus, spinal cord, and liver were manually delineated by two experienced radiation oncologists in a double-blind manner. The denoised CT images were used as input data for the AccuContour™ segmentation software to increase the signal difference between structures of interest and unwanted noise in non-contrast CT. The accuracy of the segmentation was assessed using the Dice similarity coefficient (DSC), and the results were compared with those of conventional deep-learning-based auto-segmentation without denoising. Results: The average DSC outcomes were higher than 0.80 for all OARs except for the esophagus. AccuContour™-based and denoising-based auto-segmentation demonstrated comparable performance for the lungs and spinal cord but showed limited performance for the esophagus. Denoising-based auto-segmentation for the liver was minimal but had statistically significantly better DSC than AccuContour™-based auto-segmentation (p < 0.05). Conclusions: Denoising-based auto-segmentation demonstrated satisfactory performance in automatic liver segmentation from non-contrast enhanced CT scans. Further external validation studies with larger cohorts are needed to verify the usefulness of denoising-based auto-segmentation.

Keywords: radiation therapy; contouring; organs at risk; deep-learning-based auto-segmentation; denoiser

1. Introduction

In radiotherapy planning, organs at risk (OARs) are manually delineated by physicians based on computed tomography (CT) scans. Accurate contouring of OARs is essential for precise radiotherapy. OARs are manually delineated and carefully reviewed by physicians.
However, it is a time-consuming process associated with an increased workload. The manual segmentation of OARs can take 1 h per patient due to the large number of axial slices. As such, atlas-based and deep-learning algorithms based on convolutional neural network auto-segmentation have been developed to alleviate the labor-intensive delineation of OARs [1–7]. Machine learning approaches, especially deep learning with multi-layered neural networks, have been actively applied to treatment planning in radiotherapy [8–19]. Many studies have investigated the deep-learning-based auto-segmentation of OARs for various disease sites [1,2,20–27]. The application of deep-learning algorithms based on convolutional neural networks has been proven effective and has shown high performance in delineating OARs [12–22]. Clinical practices use several commercially available deep-learning contouring products, including AccuContour™ (Manteia Medical Technologies Co. Ltd., Xiamen, China).

Contrast-enhanced planning CT is used for the delineation of target volumes and OARs, and intravenous CT contrast can enhance normal tissue visualization and delineation [28,29]. Due to safety concerns associated with CT contrast, contrast-enhanced planning CT cannot be performed for all patients [30]. Therefore, non-contrast CT images have been used, but the boundary between the OARs and neighboring structures may be indistinguishable due to the suboptimal image quality of non-contrast CT [28,29]. In real-world clinical practice, portions of automatically generated contours in non-contrast CT require manual corrections to make them clinically acceptable. The noise in non-contrast CT images disturbs the visualization of structures, which increases the uncertainty of image segmentation. Deep-learning-based auto-segmentation algorithms have to overcome these image-related problems.

An image processing technique with improved performance, such as a denoising algorithm that can remove noise while maintaining the edge structure, is needed to achieve high-contrast CT images [31–35]. In a previous study, we implemented an anisotropic total variation (ATV) denoiser to enhance the image quality of low-dose cone-beam CT [36]. In this study, we establish that the segmentation accuracy can be improved significantly when this denoising technique is applied to increase the signal difference between structures of interest and unwanted noise in non-contrast CT. Moreover, we investigate the accuracy of segmentation when denoised non-contrast CT images are used as input images for deep-learning-based auto-segmentation of OARs. We compare the performance of this approach with manual segmentation and conventional deep-learning-based auto-segmentation without denoising.

2. Materials and Methods
2.1. Data and Delineation

Ethical approval for this study was obtained from the Institutional Review Board (IRB) of Yonsei University Health System, Gangnam Severance Hospital (Approval No.: 3-2021-0276). All methods were performed in accordance with the relevant guidelines and regulations. Due to the retrospective nature of this study, informed consent was waived by the IRB of Gangnam Severance Hospital. We used non-contrast planning CT scans of female patients with breast cancer who underwent modified radical mastectomy or breast-conserving surgery and received postoperative radiotherapy between 2019 and 2020 [37]. Forty patients were randomly chosen. The median age was 49 years old (range, 30–77 years), and the median body mass index was 22 kg/m$^2$ (range 17–32 kg/m$^2$). There were 22 patients with left breast cancer and 18 patients with right breast cancer. No patients had previously undergone surgical procedures for lung, heart, esophagus, spine, and upper abdominal organs at the time of conducting non-contrast planning CT. The CT images were acquired on a Siemens Sensation Open scanner (Siemens, Forchheim, Germany) using the following parameters: 120 kVp (scan voltage) and 3 mm slice thickness (layer thickness). Scans were conducted with tube current modulation, an adaptive method in which the current changes as the gantry is rotated. We obtained 81–123 slices per patient. All patients were scanned in the supine position with a customized arm support using a breast board.
In this study, the OARs included the heart, right and left lungs, esophagus, spinal cord, and liver. The contours were manually delineated by two experienced radiation oncologists. The radiation oncologist was blinded to the results of delineation of the OARs by other radiation oncologists.

2.2. Deep-Learning-Based Auto-Segmentation

Recently, various deep-learning-based auto-segmentation methods have been developed to assist with image segmentation tasks. Satisfactory organ segmentation results have been reported [1], and some commercial products have been implemented in clinics for CT-based automatic segmentation. In this study, a commercially available deep learning contouring software “AccuContour™” (Manteia Medical Technologies Co. Ltd., Xiamen, China) was used to generate the information required for treatment planning. It automatically segments the OARs, including the head-and-neck, thorax, abdomen, and pelvis for both males and females. AccuContour™ is based on the U-net model [38] pre-trained by the vendors. U-net is a fully convolutional network (FCN) based model with end-to-end scheme proposed for image segmentation [38–40]. In U-net, the network for obtaining the overall context information of the image and that for accurate localization are symmetrically configured. U-net is a model that applies up-sampling and skip architecture of concepts that are more extended than FCNs, resulting in the U-net’s structure demonstrating superior performance in several image segmentation problems by leveraging data augmentation with only a small amount of learning data.

The trained model data were collected from multiple centers, and data cleaning was performed. Initial contours generated by the deep-learning model were corrected by post-processing with graph-based models. The accuracy was further improved by combining local and global information from the image and initial segmentation results. With these procedures, the contouring workload was reduced from hours to less than a minute for each patient. This segmentation technique was applied to delineate the OARs.

2.3. Anisotropic Total Variation Denoiser-Based Auto-Segmentation

An ATV denoiser [36] was applied to the CT images to augment the intensity difference between the striking features and unwanted noise by combining the conduction coefficient used in the anisotropic diffusion filter [41]. The minimization of the ATV objective function implies that edges with high contrast relative to the surroundings are preserved, and noisy voxels with low contrast are smoothed [42].

The ATV objective function, \( R(V) \), can be expressed as follows:

\[
R(V) = \sum_j R(V_j) = \sum_j w_j D(V_j)
\]

where \( w_j \) is the anisotropic penalty with different weights for neighbors at the same distance and \( D(V_j) \) is the discrete gradient transform with backward difference in the \( j \)th indexed value of the CT images.

\[
D(V_j) = D(V_{(x,y)}) = \sqrt{(V_{(x,y)} - V_{(x-1,y)})^2 + (V_{(x,y)} - V_{(x,y-1)})^2}
\]

\[
w_j = \sum_{m \in N_j} \exp \left[ -\left( \frac{V_j - V_m}{\delta} \right)^2 \right]
\]

where index \( j \) identifies the index of voxel elements in the CT image, and \( V_{(x,y)} \) is the voxel element at the 2D position \((x, y)\). Equivalently, \( N_j \) represents the set of neighbors of the \( j \)th voxel element. We only considered four first-order neighbors in this study. Empirically, the most meaningful results were derived with the parameter \( \delta \) set to 80% of the cumulative distribution function histogram that accumulates the gradient at each voxel of a CT image.
The ATV objective function in Equation (1) was minimized using the steepest gradient descent method with an adaptive step size. It is expressed as follows.

\[ V_{t+1}^j = V_t^j - \lambda \nabla R(V_j) / |\nabla R(V)| \]  

(4)

\[ \lambda = \gamma \sqrt{\sum_j (V_t^j)^2} \]  

(5)

where \( \lambda \) is an adaptive parameter that reduces the smoothing degree as the iteration progresses [27,31]. The square root of all voxel elements updated in each step is used to change \( \lambda \) gradually to smaller values with an increase in the number of iterations. A scaling parameter \( \gamma \) was used to escape local minimization due to sudden changes. This value starts initially at 1.0 and decreases linearly by multiplication with a constant value (0.8) when \( R(V) \) in the current iteration step is greater than that in the preceding step. \( \nabla R(V) \) is the gradient of the objective function \( R(V) \) at the \( j \)th indexed pixel [32]. The root-sum-square of the gradient calculated at all the pixels, \( |\nabla R(V)| \), is required for the normalized gradient calculation [32]. The number of iterations is fine-tuned for the gradient descent optimizer. In this study, the optimal number of iterations was set to 20. The parameters used to optimize the denoising method were based on the manually adjusted analysis. The pseudo-code of the ATV denoiser is presented in Appendix A.

The proposed image processing pipeline includes three steps. The first step is denoising. The anisotropic total variation was used for each image set to smooth noisy pixels while preserving the intensity of the edges during denoising. For the second step, the denoised CT images were used as input data in the AccuContour\textsuperscript{TM} segmentation module based on the U-net model pre-trained by the vendors. The third step involves contour segmentation. Six auto-generated contour sets (heart, left lung, right lung, esophagus, spinal cord, and liver) for each CT image were generated using a deep-learning-based auto-segmentation framework (Figure 1).

2.4. Quantitative Analysis

The noise power spectrum (NPS) was calculated using open-source software (imQuest, Duke University, Durham, NC, USA) that uses the technology described in TG233 of the AAPM [43,44] to assess the image quality characteristics without and with ATV denoiser. The quantity of in-plane noise was evaluated using two-dimensional NPS. For CT images, NPS can be determined in structures with a homogenous area. In this study, the liver was selected for NPS calculations, and overall frequencies were compared using 1D profiles.

The manual contours drawn by two radiation oncologists were considered the ground truth in this study, against which the AccuContour\textsuperscript{TM}-based and denoising-based auto-segmentations were compared. To quantitatively evaluate the accuracy of AccuContour\textsuperscript{TM}-based and denoising-based auto-segmentations, the Dice similarity coefficient (DSC) was used to evaluate the performance of the proposed method. The DSC method calculates the overlapping results of two different volumes according to the following equation:

\[ \text{DSC} = \frac{2|A \cap B|}{|A| + |B|} \]  

(6)

where \( A \) is the manual segmentation volume, and \( B \) is the auto-segmentation volume (AccuContour\textsuperscript{TM} and denoising). DSC is a measure of overlap between two contours, from “0” to “1,” where “1” indicates a complete overlap. We considered a Dice score of 0.80 as an acceptable match [45].

Wilcoxon matched pairs signed-rank test was conducted, and statistical significance was defined as \( p < 0.05 \) for evaluating differences in the results of the DSC.
3. Results

Figure 2 shows the NPS curves without and with the ATV denoiser using planning CT images from 40 patients with breast cancer. Three square ROIs were placed at different positions in the liver area with uniform magnitude, as shown in Figure 2a. The ROIs were extended to five adjacent consecutive slices contained within the liver area. The average NPS peak frequency was obtained as 0.127 mm$^{-1}$ without denoiser and 0.035 mm$^{-1}$ for the ATV denoiser. The NPS peaks ranged from 209 to 957 HU$^2$ mm$^{-2}$ without the denoiser and 66 to 481 HU$^2$ mm$^{-2}$ for the ATV denoiser. As such, the NPS peak was on average lower with the ATV denoiser than without the denoiser. The peak spatial frequency values of NPS for ATV denoiser shifted to lower spatial frequencies in comparison to no denoiser. Numerically, the NPS average spatial frequencies were obtained as 0.142 mm$^{-1}$ for the ATV denoiser and 0.295 mm$^{-1}$ for no denoiser. Images with the ATV denoiser smoothed out with a lower noise amplitude, as indicated by the average frequencies of the NPS curves, resulting in a monotonous texture.
The results of DSC versus the manual contours from radiation oncologists 1 and 2 are shown in Tables 1 and 2, respectively. The average DSC outcomes were higher than 0.80 in all OARs, except for the esophagus. The AccuContour™-based and denoising-based auto-segmentations of the esophagus were below acceptable standards. In a comparison of AccuContour™-based and denoising-based auto-segmentation, the differences were not statistically significant for the lungs, esophagus, or spinal cord \((p > 0.05)\). The denoising-based auto-segmentations achieve superior segmentation accuracy on the liver and inferior segmentation accuracy on the heart compared with AccuContour™-based auto-segmentations \((p < 0.05)\).

![Image of liver area with ROIs](image)

**Figure 2.** Noise power spectrum (NPS) for evaluating the noise texture and magnitude. (a) Examples of ROIs selected for NPS calculation on each slice. NPS curves using 40 patient data: (b) 1 to 20 cases without denoiser, (c) 1 to 20 cases with ATV, (d) 21 to 40 cases without denoiser, and (e) 21 to 40 cases with ATV.
## Table 1. Comparison of the dice similarity coefficients (DSC) results generated from AccuContour™ and denoising-based auto-segmentations using the radiation oncologist 1’s manual contours as reference.

| Case   | Heart  | Right Lung | Left Lung | Esophagus | Spinal Cord | Liver  |
|--------|--------|------------|-----------|-----------|-------------|--------|
|        | Manteia | Denoiser   | Manteia   | Denoiser  | Manteia     | Denoiser|
| Case 1 | 0.964  | 0.960      | 0.982     | 0.983     | 0.981       | 0.981   |
| Case 2 | 0.951  | 0.947      | 0.980     | 0.980     | 0.979       | 0.979   |
| Case 3 | 0.958  | 0.952      | 0.982     | 0.982     | 0.979       | 0.979   |
| Case 4 | 0.931  | 0.928      | 0.971     | 0.970     | 0.975       | 0.975   |
| Case 5 | 0.908  | 0.894      | 0.971     | 0.971     | 0.975       | 0.975   |
| Case 6 | 0.930  | 0.929      | 0.982     | 0.982     | 0.973       | 0.974   |
| Case 7 | 0.978  | 0.977      | 0.951     | 0.951     | 0.954       | 0.955   |
| Case 8 | 0.929  | 0.926      | 0.971     | 0.972     | 0.970       | 0.970   |
| Case 9 | 0.938  | 0.936      | 0.978     | 0.978     | 0.972       | 0.972   |
| Case 10| 0.952  | 0.947      | 0.982     | 0.982     | 0.981       | 0.981   |
| Case 11| 0.945  | 0.944      | 0.979     | 0.979     | 0.977       | 0.978   |
| Case 12| 0.946  | 0.941      | 0.962     | 0.962     | 0.964       | 0.966   |
| Case 13| 0.952  | 0.950      | 0.984     | 0.984     | 0.981       | 0.982   |
| Case 14| 0.934  | 0.930      | 0.981     | 0.981     | 0.981       | 0.981   |
| Case 15| 0.930  | 0.926      | 0.975     | 0.976     | 0.973       | 0.973   |
| Case 16| 0.906  | 0.898      | 0.934     | 0.934     | 0.940       | 0.940   |
| Case 17| 0.939  | 0.932      | 0.977     | 0.977     | 0.972       | 0.972   |
| Case 18| 0.950  | 0.948      | 0.976     | 0.976     | 0.979       | 0.979   |
| Case 19| 0.910  | 0.887      | 0.981     | 0.981     | 0.981       | 0.981   |
| Case 20| 0.877  | 0.875      | 0.983     | 0.984     | 0.975       | 0.975   |
| Case 21| 0.919  | 0.921      | 0.981     | 0.981     | 0.978       | 0.978   |
| Case 22| 0.999  | 0.968      | 0.998     | 0.992     | 0.998       | 0.992   |
| Case 23| 0.956  | 0.954      | 0.977     | 0.977     | 0.980       | 0.979   |
| Case 24| 0.963  | 0.962      | 0.986     | 0.986     | 0.979       | 0.980   |
| Case 25| 0.919  | 0.920      | 0.987     | 0.987     | 0.979       | 0.978   |
| Case 26| 0.966  | 0.966      | 0.990     | 0.989     | 0.986       | 0.987   |
| Case 27| 0.963  | 0.961      | 0.992     | 0.992     | 0.989       | 0.989   |
| Case 28| 0.963  | 0.963      | 0.991     | 0.991     | 0.984       | 0.985   |
| Case 29| 0.967  | 0.967      | 0.985     | 0.985     | 0.976       | 0.978   |
| Case 30| 0.951  | 0.951      | 0.985     | 0.985     | 0.980       | 0.980   |
| Case 31| 0.950  | 0.950      | 0.986     | 0.986     | 0.984       | 0.984   |
| Case 32| 0.947  | 0.946      | 0.986     | 0.986     | 0.980       | 0.980   |
| Case 33| 0.956  | 0.956      | 0.962     | 0.962     | 0.975       | 0.975   |
### Table 1. Cont.

| Case | Heart | Right Lung | Left Lung | Esophagus | Spinal Cord | Liver |
|------|-------|------------|-----------|-----------|------------|-------|
|      | * Manteia | † Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser |
| Case 34 | 0.950 | 0.950 | 0.978 | 0.978 | 0.979 | 0.980 | 0.740 | 0.725 | 0.861 | 0.857 | 0.943 | 0.946 |
| Case 35 | 0.947 | 0.949 | 0.977 | 0.977 | 0.981 | 0.981 | 0.749 | 0.739 | 0.887 | 0.880 | 0.956 | 0.959 |
| Case 36 | 0.937 | 0.938 | 0.985 | 0.985 | 0.977 | 0.978 | 0.788 | 0.782 | 0.881 | 0.883 | 0.950 | 0.952 |
| Case 37 | 0.945 | 0.945 | 0.984 | 0.984 | 0.980 | 0.981 | 0.730 | 0.727 | 0.879 | 0.882 | 0.939 | 0.938 |
| Case 38 | 0.933 | 0.934 | 0.987 | 0.987 | 0.983 | 0.982 | 0.809 | 0.819 | 0.872 | 0.878 | 0.947 | 0.949 |
| Case 39 | 0.945 | 0.945 | 0.968 | 0.962 | 0.974 | 0.975 | 0.779 | 0.743 | 0.887 | 0.894 | 0.952 | 0.953 |
| Case 40 | 0.911 | 0.911 | 0.982 | 0.982 | 0.977 | 0.978 | 0.664 | 0.672 | 0.876 | 0.868 | 0.947 | 0.947 |
| Average | 0.943 | 0.940 | 0.979 | 0.978 | 0.977 | 0.977 | 0.719 | 0.717 | 0.858 | 0.868 | 0.938 | 0.943 |

*p*-value: 0.000 0.091 0.095 0.705 0.762 0.000

* AccuContour™ (Manteia Medical Technologies Co. Ltd., Xiamen, China)-based auto-segmentation, † Denoising-based auto-segmentation.

### Table 2.

Comparison of the dice similarity coefficients (DSC) results generated from AccuContour™ and denoising-based auto-segmentations using the radiation oncologist 2’s manual contours as reference.

| Case | Heart | Right Lung | Left Lung | Esophagus | Spinal Cord | Liver |
|------|-------|------------|-----------|-----------|------------|-------|
|      | * Manteia | † Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser |
| Case 1 | 0.964 | 0.960 | 0.981 | 0.981 | 0.980 | 0.980 | 0.685 | 0.742 | 0.749 | 0.785 | 0.953 | 0.955 |
| Case 2 | 0.937 | 0.934 | 0.985 | 0.985 | 0.982 | 0.982 | 0.696 | 0.730 | 0.872 | 0.844 | 0.932 | 0.934 |
| Case 3 | 0.964 | 0.962 | 0.984 | 0.984 | 0.980 | 0.981 | 0.638 | 0.597 | 0.865 | 0.867 | 0.949 | 0.956 |
| Case 4 | 0.912 | 0.909 | 0.973 | 0.972 | 0.979 | 0.979 | 0.741 | 0.697 | 0.857 | 0.864 | 0.927 | 0.929 |
| Case 5 | 0.907 | 0.895 | 0.977 | 0.977 | 0.977 | 0.976 | 0.721 | 0.625 | 0.848 | 0.848 | 0.941 | 0.947 |
| Case 6 | 0.927 | 0.927 | 0.987 | 0.987 | 0.980 | 0.980 | 0.577 | 0.693 | 0.887 | 0.892 | 0.884 | 0.920 |
| Case 7 | 0.986 | 0.989 | 0.996 | 0.997 | 0.995 | 0.996 | 0.808 | 0.839 | 0.870 | 0.873 | 0.849 | 0.859 |
| Case 8 | 0.925 | 0.922 | 0.964 | 0.963 | 0.973 | 0.973 | 0.658 | 0.689 | 0.872 | 0.851 | 0.954 | 0.953 |
| Case 9 | 0.934 | 0.933 | 0.979 | 0.978 | 0.974 | 0.974 | 0.702 | 0.747 | 0.851 | 0.852 | 0.952 | 0.952 |
| Case 10 | 0.916 | 0.909 | 0.984 | 0.984 | 0.984 | 0.983 | 0.732 | 0.756 | 0.869 | 0.852 | 0.957 | 0.959 |
| Case 11 | 0.938 | 0.939 | 0.986 | 0.986 | 0.974 | 0.974 | 0.683 | 0.683 | 0.786 | 0.749 | 0.912 | 0.915 |
| Case 12 | 0.949 | 0.943 | 0.977 | 0.977 | 0.967 | 0.969 | 0.707 | 0.655 | 0.778 | 0.747 | 0.925 | 0.925 |
| Case 13 | 0.935 | 0.935 | 0.985 | 0.985 | 0.982 | 0.982 | 0.665 | 0.692 | 0.844 | 0.823 | 0.921 | 0.925 |
| Case 14 | 0.935 | 0.933 | 0.984 | 0.984 | 0.979 | 0.979 | 0.633 | 0.695 | 0.823 | 0.790 | 0.902 | 0.927 |
| Case 15 | 0.907 | 0.909 | 0.983 | 0.983 | 0.976 | 0.976 | 0.660 | 0.648 | 0.835 | 0.841 | 0.921 | 0.880 |
| Case 16 | 0.926 | 0.936 | 0.985 | 0.985 | 0.979 | 0.979 | 0.681 | 0.637 | 0.781 | 0.787 | 0.953 | 0.956 |
| Case 17 | 0.939 | 0.932 | 0.977 | 0.977 | 0.972 | 0.972 | 0.697 | 0.697 | 0.804 | 0.822 | 0.949 | 0.949 |
| Case 18 | 0.950 | 0.951 | 0.982 | 0.982 | 0.982 | 0.982 | 0.734 | 0.735 | 0.870 | 0.880 | 0.932 | 0.936 |
| Case  | Heart | Right Lung | Left Lung | Esophagus | Spinal Cord | Liver |
|-------|-------|------------|-----------|-----------|-------------|-------|
|       | * Manteia | † Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser | Manteia | Denoiser |
| Case 19 | 0.890 | 0.867 | 0.985 | 0.985 | 0.983 | 0.981 | 0.630 | 0.604 | 0.860 | 0.865 | 0.857 | 0.869 |
| Case 20 | 0.850 | 0.848 | 0.986 | 0.986 | 0.979 | 0.978 | 0.738 | 0.743 | 0.885 | 0.860 | 0.949 | 0.950 |
| Case 21 | 0.954 | 0.955 | 0.991 | 0.991 | 0.986 | 0.986 | 0.730 | 0.731 | 0.853 | 0.835 | 0.946 | 0.947 |
| Case 22 | 0.999 | 0.968 | 0.995 | 0.990 | 0.995 | 0.989 | 0.902 | 0.752 | 0.790 | 0.827 | 0.941 | 0.952 |
| Case 23 | 0.956 | 0.954 | 0.975 | 0.975 | 0.978 | 0.978 | 0.622 | 0.632 | 0.838 | 0.816 | 0.953 | 0.955 |
| Case 24 | 0.963 | 0.962 | 0.984 | 0.984 | 0.977 | 0.978 | 0.726 | 0.724 | 0.838 | 0.820 | 0.956 | 0.956 |
| Case 25 | 0.924 | 0.925 | 0.986 | 0.985 | 0.978 | 0.980 | 0.767 | 0.763 | 0.860 | 0.864 | 0.962 | 0.965 |
| Case 26 | 0.966 | 0.966 | 0.986 | 0.986 | 0.983 | 0.984 | 0.813 | 0.821 | 0.868 | 0.861 | 0.960 | 0.961 |
| Case 27 | 0.963 | 0.961 | 0.990 | 0.990 | 0.987 | 0.988 | 0.636 | 0.627 | 0.840 | 0.837 | 0.959 | 0.959 |
| Case 28 | 0.963 | 0.963 | 0.991 | 0.991 | 0.984 | 0.985 | 0.785 | 0.790 | 0.853 | 0.862 | 0.951 | 0.951 |
| Case 29 | 0.967 | 0.967 | 0.984 | 0.984 | 0.974 | 0.976 | 0.765 | 0.768 | 0.847 | 0.827 | 0.934 | 0.936 |
| Case 30 | 0.951 | 0.951 | 0.984 | 0.984 | 0.979 | 0.980 | 0.661 | 0.671 | 0.857 | 0.840 | 0.938 | 0.939 |
| Case 31 | 0.932 | 0.932 | 0.986 | 0.986 | 0.983 | 0.983 | 0.668 | 0.650 | 0.852 | 0.843 | 0.895 | 0.897 |
| Case 32 | 0.939 | 0.938 | 0.985 | 0.985 | 0.978 | 0.978 | 0.791 | 0.778 | 0.857 | 0.855 | 0.912 | 0.911 |
| Case 33 | 0.956 | 0.956 | 0.960 | 0.960 | 0.975 | 0.975 | 0.710 | 0.709 | 0.857 | 0.839 | 0.948 | 0.950 |
| Case 34 | 0.949 | 0.948 | 0.976 | 0.976 | 0.976 | 0.977 | 0.739 | 0.723 | 0.876 | 0.861 | 0.947 | 0.949 |
| Case 35 | 0.935 | 0.934 | 0.982 | 0.982 | 0.980 | 0.980 | 0.749 | 0.739 | 0.880 | 0.874 | 0.953 | 0.955 |
| Case 36 | 0.936 | 0.936 | 0.984 | 0.984 | 0.976 | 0.976 | 0.793 | 0.786 | 0.860 | 0.840 | 0.951 | 0.952 |
| Case 37 | 0.945 | 0.945 | 0.982 | 0.982 | 0.978 | 0.979 | 0.669 | 0.662 | 0.812 | 0.821 | 0.938 | 0.936 |
| Case 38 | 0.944 | 0.945 | 0.986 | 0.986 | 0.981 | 0.981 | 0.809 | 0.819 | 0.870 | 0.874 | 0.950 | 0.951 |
| Case 39 | 0.955 | 0.955 | 0.965 | 0.960 | 0.974 | 0.975 | 0.682 | 0.663 | 0.884 | 0.892 | 0.953 | 0.953 |
| Case 40 | 0.911 | 0.911 | 0.980 | 0.980 | 0.977 | 0.977 | 0.624 | 0.630 | 0.874 | 0.875 | 0.948 | 0.948 |
| Average | 0.940 | 0.938 | 0.982 | 0.982 | 0.979 | 0.980 | 0.711 | 0.709 | 0.847 | 0.841 | 0.935 | 0.938 |

*p-value* 0.008 0.091 0.097 0.984 0.082 0.000

* AccuContour™ (Manteia Medical Technologies Co. Ltd., Xiamen, China)-based auto-segmentation, † Denoising-based auto-segmentation.
4. Discussion

In this study, we compared the auto-contouring results in five organ structures using the commercial deep-learning contouring program AccuContour™ with those obtained from an anisotropic total variation denoiser. Both the AccuContour™-based and denoising-based auto-segmentation were considered to yield an acceptable accuracy for generating contours of the heart, lungs, spinal cord, and liver. However, these techniques yielded limited performance for the esophagus.

Deep-learning algorithms based on convolutional neural networks and AccuContour™ have yielded satisfactory performance outcomes for the automatic segmentation of OARs. However, some parts of automatic segmentation of the liver in non-contrast CT required manual corrections to make them clinically acceptable (Figure 3). In non-contrast CT images, it might be difficult to delineate the fuzzy boundaries between the liver and adjacent organs, owing to low soft tissue contrast between the liver and its surrounding organs. In this study, the auto-segmentation results showed a significant improvement in the DSC when using denoising-based auto-segmentations of the liver, compared to using AccuContour™-based auto-segmentation. An ATV denoiser could enhance the image quality of CT by removing noisy areas, and this may lead to improved segmentation boundaries. Figure 3 shows that some parts surrounding the gall bladder, pancreas, duodenum, large vessel, or kidney are included in AccuContour™-based auto-segmentation of the liver. However, denoising-based auto-segmentations could delineate the liver accurately by distinguishing the surrounding organs of CT images. These results indicate that the performance of denoising-based auto-segmentation is superior to that of AccuContour™-based auto-segmentation. All deep-learning-based auto-segmentations should be carefully reviewed and approved by the radiation oncologists before use for a treatment plan. In some CT slices, major or minor errors of deep-learning-based auto-contour are present, and correction is required. The denoising-based auto-segmentation might convert “large or minor errors” of conventional deep-learning-based auto-segmentation to “minor errors” (a small amount of editing needed) or “no correction”. The denoising-based auto-segmentation could be a practical tool for reducing the clinical workload of radiotherapy planning.

The esophagus is one of the most challenging OARs in thoracic organ auto-segmentation. In this study, the performance of AccuContour™-based and denoising-based auto-segmentation was below a satisfactory level for the esophagus. Previous studies have reported that the DSCs of deep-learning-based auto-segmentation do not exceed 0.8 for the esophagus [40,46–49]. Due to the absence of a consistent intensity contrast between the esophagus and neighboring tissues in non-contrast CT images, the boundaries between the esophagus and surrounding soft tissues are not well-defined. Figure 4 shows that some parts of the surrounding pulmonary vessel were included in AccuContour™-based and denoising-based auto-segmentation of the esophagus. In addition, the appearance of the esophagus varies depending on whether it is filled with air or not. Figure 5 shows that the air-filled regions of the esophagus were not included in AccuContour™-based and denoising-based auto-segmentations. The segmentation results for the esophagus obtained from the denoising-based auto-segmentation may still be inaccurate and unsatisfactory.
AccuContour™-based auto-segmentation over-contoured (A) gallbladder, (B) pancreas and portal vein, (C) pancreas and duodenum, (D) duodenum, (E) pancreas, duodenum, and kidney, and (F) gallbladder.

It has been demonstrated that auto-segmentations for the heart and lungs yield higher DSCs, with an average of over 0.9 [45, 48–51]. This study also showed that the average DSC outcomes of the heart and lungs are higher than 0.9. High-contrast edges and distinct structural boundaries of the heart and lungs were detected easily in both the AccuContour™-based and denoising-based auto-segmentation. Therefore, AccuContour™-based or denoising-based auto-segmentation for the heart and lungs can be used without major adjustments. The denoising-based auto-segmentations achieve inferior segmentation accuracy on the heart compared with AccuContour™-based auto-segmentations. The denoising-based auto-segmentation volumes are slightly larger than manual contours (Figure 6A,D) or smaller than manual contours (Figure 6B,C). Denoising-based auto-segmentation did not improve the accuracy of auto-segmentation of the heart.
This study had several limitations. Although 40 patients were randomly selected, selection bias in terms of the CT samples may be present. Since the results of this study were generated by only one proprietary software and CT scan, selection bias may have impacted our results.

Further, the contouring bias of the physicians may impact our results. Therefore, further external validation studies involving multiple experts, hospitals, and a larger sample size are needed to overcome these limitations. Moreover, denoising-based auto-segmentation should be compared using software other than AccuContour\textsuperscript{TM}.

However, in several deep-learning-based automatic segmentation studies, a single experienced radiation oncologist delineated the organ at risk or the clinical target volume [10,11,14,50]. In this study, two radiation oncologists delineated the organ at risk because it was thought that inter-observer variability may exist. The results of the Dice similarity coefficient of the five organs at risk from radiation oncologist 1 and 2 were consistent, so the results of this study are expected to be relatively reliable.
It has been demonstrated that auto-segmentations for the heart and lungs yield higher DSCs, with an average of over 0.9 [45,48–51]. This study also showed that the average DSC outcomes of the heart and lungs are higher than 0.9. High-contrast edges and distinct structural boundaries of the heart and lungs were detected easily in both the AccuContour™-based and denoising-based auto-segmentation. Therefore, AccuContour™-based or denoising-based auto-segmentation for the heart and lungs can be used without major adjustments. The denoising-based auto-segmentations achieve inferior segmentation accuracy on the heart compared with AccuContour™-based auto-segmentations. The denoising-based auto-segmentation volumes are slightly larger than manual contours (Figure 6A,D) or smaller than manual contours (Figure 6B,C). Denoising-based auto-segmentation did not improve the accuracy of auto-segmentation of the heart.

This study had several limitations. Although 40 patients were randomly selected, selection bias in terms of the CT samples may be present. Since the results of this study were generated by only one proprietary software and CT scan, selection bias may have impacted our results. Further, the contouring bias of the physicians may impact our results. Therefore, further external validation studies involving multiple experts, hospitals, and a larger sample size are needed to overcome these limitations. Moreover, denoising-based auto-segmentation should be compared using software other than AccuContour™. However, in several deep-learning-based automatic segmentation studies, a single experienced radiation oncologist delineated the organ at risk or the clinical target volume [10,11,14,50]. In this study, two radiation oncologists delineated the organ at risk because it was thought that inter-observer variability may exist. The results of the Dice similarity coefficient of the five organs at risk from radiation oncologist 1 and 2 were consistent, so the results of this study are expected to be relatively reliable.

Figure 5. Example cases showing (A) radiation oncologist 1’s manual contour (red), (B) radiation oncologist 2’s manual contour (blue), (C) denoising-based auto-segmentation (yellow), and (D) AccuContour™-based auto-segmentation (green) for the esophagus.

Figure 6. Example cases (A–D) showing radiation oncologist 1’s manual contour (red), radiation oncologist 2’s manual contour (blue), AccuContour™-based auto-segmentation (green), and denoising-based auto-segmentation (yellow) for the heart.
5. Conclusions

CT images are subject to noise that can affect the boundary between the adjacent organs, providing potentially limited contrast. Reducing the noise level in CT images provides the best visualization of structures, which increases the accuracy of image segmentation.

By combining the denoising algorithm in the deep-learning based auto-segmentation, the denoising-based auto-segmentation results of the liver from non-contrast CT scans were slightly superior to those of commercial conventional deep-learning-based auto-segmentation and had greater similarities with the ground truth. This denoising-based auto-segmentation could provide a more precise contour delineation of the liver, thus reducing the clinical workload. The results of this study require validation through further studies using a higher sample size, which will compare denoising-based auto-segmentation using software other than AccuContour™.

Author Contributions: Conceptualization, H.L. and I.J.L.; methodology, J.H.I. and H.L.; software, H.L.; validation, I.J.L. and H.L.; formal analysis, J.H.I.; investigation, H.L. and J.H.I.; data acquisition, Y.C., J.S. and J.S.H.; data curation, Y.C.; writing-original draft preparation, J.H.I.; writing-review and editing, H.L.; supervision, I.J.L. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1A2C2011556) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (HI19C1330).

Institutional Review Board Statement: Ethical approval for this study was obtained from the Institutional Review Board of Yonsei University Health System, Gangnam Severance Hospital.

Informed Consent Statement: The need for patient consent was waived due to the retrospective nature of the study design.

Data Availability Statement: All data generated or analyzed during this study are included in the article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

For n ← 1 to all images do
For j ← 1 to all voxels do
    Calculate $D(V_j)$ using Equation (2)
    Create gradient CDF histogram using $D(V_j)$
End For
δ ← Value at 80% of gradient CDF
End For
For n ← 1 to all images do
R(V) ← 0, r ← 1, $r_{red} ← 0.8$
For j ← 1 to all voxels do
    $w_j ← 0$
    For m ← 1 to $N_j$ do
        $w_j ← w_j + \exp\left[\frac{\left(V_j - V_m\right)^2}{2}\right]\]
    End For
    Calculate $D(V_j)$ using Equation (2)
    $R(V_j) ← w_j D(V_j)$
    $R(V) ← R(V) + R(V_j)$
End For
For t ← 1 to 20 do
    $|\nabla R(V)| ← 0$
For \( j \leftarrow 1 \) to all voxels do
\[
\lambda \leftarrow \frac{\sum V_j}{V}
\]
\[
\lambda \leftarrow \lambda \times r
\]
Calculate \( \nabla R(V_j) \)
\[
|\nabla R(V_j)| \leftarrow |\nabla R(V)| + (\nabla R(V_j))^2
\]
End For
\[
|\nabla R(V)| \leftarrow \sqrt{|\nabla R(V)|^2}
\]
For \( j \leftarrow 1 \) to all voxels do
\[
V'_j \leftarrow V_j + \lambda \nabla R(V_j) / |\nabla R(V)|
\]
Calculate \( D(V'_j) \) using Equation (2)
End For
Calculate \( R(V') \) using \( w_r \) and \( D(V'_j) \)
While \( R(V'_j) > R(V) \) do
\[
r \leftarrow r \times r_{red}
\]
\[
\lambda \leftarrow \lambda \times r
\]
For \( j \leftarrow 1 \) to all voxels do
\[
V'_j \leftarrow V_j + \lambda \nabla R(V_j) / |\nabla R(V)|
\]
Calculate \( D(V'_j) \) using Equation (2)
End For
Calculate \( R(V') \)
End While
Update \( V'_j \) to \( V_j \)
End For

References

1. Cardenas, C.E.; Yang, J.; Anderson, B.M.; Court, L.E.; Brock, K.B. Advances in auto-segmentation. In *Seminars in Radiation Oncology*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 185–197.
2. Kosmin, M.; Ledsam, J.; Romera-Paredes, B.; Mendes, R.; Moinuddin, S.; de Souza, D.; Gunn, L.; Kelly, C.; Hughes, C.; Karthikesalingam, A. Rapid advances in auto-segmentation of organs at risk and target volumes for personalized adaptive radiation therapy in head and neck area. *Front. Oncol.* 2021, 13, 702. [CrossRef] [PubMed]
3. Ayyalusamy, A.; Vellaiyan, S.; Subramanian, S.; Ilamurugu, A.; Satpathy, S.; Nauman, M.; Katta, G.; Madineni, A. Auto-segmentation of head and neck organs at risk and nodal target volumes: A systematic review. *Diagnostics* 2020, 10, 959. [CrossRef]
4. Daisne, J.-F.; Blumhofer, A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: A clinical validation. *Radiat. Oncol.* 2013, 8, 1–11. [CrossRef] [PubMed]
5. Tong, N.; Gou, S.; Yang, S.; Ruan, D.; Sheng, K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med. Phys.* 2018, 45, 4558–4567. [CrossRef]
6. Kiljunen, T.; Akram, S.; Niemelä, J.; Löytyniemi, E.; Seppälä, J.; Heikkilä, J.; Vuolukka, K.; Kääriäinen, O.-S.; Heikkilä, V.-P.; Lehtío, K. A deep learning-based automated ct segmentation of prostate cancer anatomy for radiation therapy planning-a retrospective multicenter study. *Diagnostics* 2020, 10, 959. [CrossRef]
7. Windisch, P.; Koechli, C.; Rogers, S.; Schröder, C.; Förster, R.; Zwahlen, D.R.; Bodis, S. Machine learning for the detection and segmentation of benign tumors of the central nervous system: A systematic review. *Cancers* 2022, 14, 2676. [CrossRef]
8. Kim, N.; Chun, J.; Chang, J.S.; Lee, C.G.; Keum, K.C.; Kim, J.S. Feasibility of continual deep learning-based segmentation for personalized adaptive radiation therapy in head and neck area. *Cancers* 2021, 13, 702. [CrossRef]
9. Yoo, S.K.; Kim, T.H.; Chun, J.; Choi, B.S.; Kim, H.; Yang, S.; Yoon, H.I.; Kim, J.S. Deep-learning-based automatic detection and segmentation of brain metastases with small volume for stereotactic ablative radiotherapy. *Cancers* 2022, 14, 2555. [CrossRef]
10. Comelli, A.; Dahiyi, N.; Stefano, A.; Vernuccio, F.; Fortogheese, M.; Cutaia, G.; Bruno, A.; Salvaggio, G.; Yezzi, A. Deep learning-based methods for prostate segmentation in magnetic resonance imaging. *Appl. Sci.* 2021, 11, 782. [CrossRef] [PubMed]
11. Liu, X.; Li, K.-W.; Yang, R.; Geng, L.-S. Review of deep learning based automatic segmentation for lung cancer radiotherapy. *Front. Oncol.* 2021, 11, 2599. [CrossRef] [PubMed]
40. Dong, X.; Lei, Y.; Wang, T.; Thomas, M.; Tang, L.; Curran, W.J.; Liu, T.; Yang, X. Automatic multiorgan segmentation in thorax ct images using u-net-gan. Med. Phys. **2019**, *46*, 2157–2168. [CrossRef]

41. Perona, P.; Malik, J. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **1990**, *12*, 629–639. [CrossRef]

42. Wang, J.; Li, T.; Xing, L. Iterative image reconstruction for cbct using edge-preserving prior. *Med. Phys.* **2009**, *36*, 252–260. [CrossRef]

43. Greffier, J.; Hamard, A.; Pereira, F.; Barrau, C.; Pasquier, H.; Beregi, J.P.; Frandon, J. Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for ct: A phantom study. *Eur. Radiol.* **2020**, *30*, 3951–3959. [CrossRef] [PubMed]

44. Samei, E.; Bakalyar, D.; Boedeker, K.L.; Brady, S.; Fan, J.; Leng, S.; Myers, K.J.; Popescu, L.M.; Ramirez Giraldo, J.C.; Ranallo, F. Performance evaluation of computed tomography systems: Summary of aapm task group 233. *Med. Phys.* **2019**, *46*, e735–e756. [CrossRef] [PubMed]

45. Chung, S.Y.; Chang, J.S.; Choi, M.S.; Chang, Y.; Choi, B.S.; Chun, J.; Keum, K.C.; Kim, J.S.; Kim, Y.B. Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiat. Oncol.* **2021**, *16*, 1–10. [CrossRef] [PubMed]

46. Gibson, E.; Giganti, F.; Hu, Y.; Bonmati, E.; Bandula, S.; Gurusamy, K.; Davidson, B.; Pereira, S.P.; Clarkson, M.J.; Barratt, D.C. Automatic multi-organ segmentation on abdominal ct with dense v-networks. *IEEE Trans. Med. Imaging* **2018**, *37*, 1822–1834. [CrossRef]

47. Vu, C.C.; Siddiqui, Z.A.; Zamdborg, L.; Thompson, A.B.; Quinn, T.J.; Castillo, E.; Guerrero, T.M. Deep convolutional neural networks for automatic segmentation of thoracic organs-at-risk in radiation oncology–use of non-domain transfer learning. *J. Appl. Clin. Med. Phys.* **2020**, *21*, 108–113. [CrossRef]

48. Zhu, J.; Zhang, J.; Qiu, B.; Liu, Y.; Liu, X.; Chen, L. Comparison of the automatic segmentation of multiple organs at risk in ct images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncol.* **2019**, *58*, 257–264. [CrossRef] [PubMed]

49. Men, K.; Geng, H.; Biswas, T.; Liao, Z.; Xiao, Y. Automated quality assurance of oar contouring for lung cancer based on segmentation with deep active learning. *Front. Oncol.* **2020**, *10*, 986. [CrossRef]

50. Choi, M.S.; Choi, B.S.; Chung, S.Y.; Kim, N.; Chun, J.; Kim, Y.B.; Chang, J.S.; Kim, J.S. Clinical evaluation of atlas-and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiother. Oncol.* **2020**, *153*, 139–145. [CrossRef]

51. Feng, X.; Qing, K.; Tustison, N.J.; Meyer, C.H.; Chen, Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3d images. *Med. Phys.* **2019**, *46*, 2169–2180. [CrossRef]