Reports of Practical Oncology and Radiotherapy 25 (2020) 981–986

Available online at www.sciencedirect.com

Reports of Practical Oncology and Radiotherapy

journal homepage: http://www.elsevier.com/locate/rpor

Original research article

Convolutional neural network-based automatic liver delineation on contrast-enhanced and non-contrast-enhanced CT images for radiotherapy planning

Naohiro Sakashita[a,*], Kiyonori Shirai[a], Yoshihiro Ueda[b], Ayuka Ono[a], Teruki Teshima[b]

[a] Department of Diagnostic and Interventional Radiology, Osaka International Cancer Institute, Osaka, Japan
[b] Department of Radiation Oncology, Osaka International Cancer Institute, Osaka, Japan

ARTICLE INFO

Article history:
Received 26 May 2020
Received in revised form 23 August 2020
Accepted 21 September 2020
Available online 2 October 2020

Keywords:
CT
Neural network
Radiation therapy planning

ABSTRACT

Aim: This study evaluated a convolutional neural network (CNN) for automatically delineating the liver on contrast-enhanced or non-contrast-enhanced CT, making comparisons with a commercial automated technique (MIM Maestro®).
Background: Intensity-modulated radiation therapy requires careful labor-intensive planning involving delineation of the target and organs on CT or MR images to ensure delivery of the effective dose to the target while avoiding organs at risk.
Materials and Methods: Contrast-enhanced planning CT images from 101 pancreatic cancer cases and accompanying mask images showing manually-delineated liver contours were used to train the CNN to segment the liver. The trained CNN then performed liver segmentation on a further 20 contrast-enhanced and 15 non-contrastenhanced CT image sets, producing three-dimensional mask images of the liver.
Results: For both contrast-enhanced and non-contrast-enhanced images, the mean Dice similarity coefficients between CNN segmentations and ground-truth manual segmentations were significantly higher than those between ground-truth and MIM Maestro software ($p < 0.001$). Although mean CT values of the liver were higher on contrast-enhanced than on non-contrast-enhanced CT, there were no significant differences in the Hausdorff distances of the CNN segmentations, indicating that the CNN could successfully segment the liver on both image types, despite being trained only on contrast-enhanced images.
Conclusions: Our results suggest that a CNN can perform highly accurate automated delineation of the liver on CT images, irrespective of whether the CT images are contrast-enhanced or not.

1. Background

The definition of target tumors and organs at risk (OARs) forms a very important part of the radiotherapy (RT) planning process. Target tumors and OARs are typically defined by manual delineation of their respective regions on computed tomography (CT) or magnetic resonance images (MRI) acquired for the RT simulation. Accurate delineation is essential for the planning of intensity modulated radiation therapy to ensure delivery of the effective dose to target tumors and avoid exposing OARs.[1–4]

Target tumors and OARs are generally manually delineated on each slice of the CT and/or MRI images, and the large number of these slices places a considerable work burden on RT planners. In particular, delineation of the liver region on abdominal images may require substantial time from the planner, because the liver is the biggest organ in the abdomen. Furthermore, it has been reported that delineation accuracy is influenced by interobserver variability.[5,6] For these reasons, software has been developed for the automatic delineation of tumors and organs.[7–9] However, the shapes of organs on images differ between patients, with the variations being caused by many factors including sex, height, weight, history of surgery, and respiration. Furthermore, the voxel values (e.g. the CT number in CT images) and contrast of organs depend on the scan parameters and use of contrast-enhancement agents.

Recently, machine learning using convolutional neural networks (CNNs) with deep learning architectures has attracted the interest of researchers working on automatic delineation for RT planning. CNNs can learn to automatically perform classification tasks on images by being given supervised input-output data pairs,

* Corresponding author at: Department of Diagnostic and Interventional Radiology, Osaka International Cancer Institute, 3-1-69, Otemae, Chuo-ku, Osaka-shi, Osaka, 541-8567, Japan.
E-mail address: naohiro.sakashita@oici.jp (N. Sakashita).

https://doi.org/10.1016/j.rpor.2020.09.005
1507-1367/

where the original image and desired output is provided, and then, once trained, they can produce the desired classification output for new unseen images. However, although some studies have reported using CNNs for organ delineation on medical images,[10–13] the criteria used were not uniform and the target organ datasets varied between studies, and this field, therefore, remains challenging.

In RT planning, the use of contrast-enhancement agents is important for the delineation of organs. Contrast-enhanced CT (CECT) images have been reported to be very effective for tumor delineation,[14,15] but in some cases, contrast agents cannot be used.

The aim of this study was, therefore, to evaluate liver delineation on non-contrast enhanced CT (nCECT) images using CNNs trained on CECT image data, with comparisons also being made with conventional commercial delineation software.

2. Materials and methods

2.1. Image acquisition

This retrospective study was approved by our institution's ethical review board. CT images were collected to form the CECT and nCECT groups. 121 cases and fifteen cases that underwent pancreatic cancer RT planning were included in CECT and nCECT group, respectively.

Patients were instructed to fast for three hours before the CT scan, and to drink 3 mL of water-soluble gastrointestinal contrast agent diluted with 30 mL distilled water just before imaging, to make the lumen of the stomach easier to define. For the CECT cases, the patients were injected with 450 mgI/kg nonionic iodine contrast agent in the upper arm vein at a rate of 3.0 mL/s, and were then scanned after 45 s. Irrespective of whether contrast agent was used, the patients were instructed to exhale immediately prior to the CT scan and, then, not to breathe in during the CT scan. All scans were acquired on a multi detector CT scanner (GE LightSpeed, 16 slices, General Electric Co., Waukesha, WI) with 120 kVp tube voltage, 250–440 mA (13 noise index) current, 12×1.25 mm collimation, normalized pitch of 0.938, 0.5 s rotation time, 512×512 matrix, and 50-cm field of view. Transverse images were reconstructed at 2.5-mm slice thickness using a moderately soft reconstruction kernel (standard).

All image sets consisted of CT images and binary images showing the contours of the delineated liver. The livers were delineated by one of seven radiation oncologists with over five years of experience, with allocations to each oncologist made on a random basis. One hundred and one cases were randomly selected from the CECT group for use as training data, while the remaining 20 cases from the CECT group and all cases in the nCECT group were used as validation data.

2.2. Convolutional neural network training

An outline of the study's method is given in Fig. 1. First, to train the CNN efficiently, all CT images were compressed from 16 bit to 8 bit. All contour images were converted to mask images by filling the inside of the liver outline. Then, the CT images and their corresponding mask images were rotated to -30°, -15°, 15°, and 30° for data augmentation. These extra image sets were added to the original data to augment the training dataset. All image processing was performed using Image J (NIH, Bethesda, Maryland).

The CNN designed in our study has a structure similar to U-net,[16] and used the training dataset with mask images as supervised input data. The hyper parameters were determined using a grid search. Construction, training, and validation of the CNN were implemented using the Sony Neural Network Console (Sony Network Communication Inc, Tokyo, Japan).

2.3. Automatic segmentation

The trained CNN was used to perform automatic liver segmentation on the validation data consisting of 20 CECT and 15 nCECT acquisitions. The output images were not binary images, but 8-bit gray scale images, with the closer the intensity to 255, the higher the probability of the region being the liver. The output images were then binarized using a threshold of 128, with values above 128 being defined as the liver. The largest continuous region in three dimensions was then extracted to obtain the final automatic segmentation result.

For comparison purposes, automatic liver segmentation was also performed using commercial software (MIM Maestro®, MIM Software Inc, Cleveland, Ohio). In this software, users designate slice positions including the upper and lower end of the liver on the CT image, and the liver region is then automatically segmented by referring to the most similar example among several dozen liver segmentation results registered in advance. The user is able to select either contrast or non-contrast-enhanced data, and the software selects the appropriate method for the images to be analyzed.

2.4. Statistical analysis

Dice coefficients (DC) were calculated for the similarities between the ground-truth liver segmentations and the CNN and conventional software segmentation results. The DC is expressed by the following formula:

$$\mathrm{DC}(A, B) = \frac{2\left|A + B\right|}{\left|A\right| + \left|B\right|}$$

where A is the ground truth liver region (i.e., the original mask image) and B is the automatic segmentation region. DC is a number on a scale of 0–1. DC value of 0 means that the two structures do not share a volume or overlap, and the value of 1 means that the two structures have the same volume and overlap. The Dice coefficients DC_{CNN} and DC_{conv} were calculated for the results of the CNN and conventional methods, respectively, and were also calculated separately for the CECT and nCECT groups.

Next, the Hausdorff distances[17] between the ground truth liver contours and the contours of each automatic segmentation method were calculated separately for the CECT and nCECT groups. The Hausdorff distance was defined as the maximum value of the shortest distances from any point in set A to another set B, which in this study was the maximum distance from a point on the surface of one liver segmentation to the closest point on the surface of another segmentation. The Hausdorff Distance was used to evaluate the similarity of the contours in each slice, with d_{CNN} or d_{conv} being the distances between the ground-truth liver contours and the CNN or conventional software contours, respectively.

Wilcoxon's signed-rank test was used to compare the CNN and conventional methods in the CECT and nCECT groups. The Mann–Whitney U test was used to test the difference between the CECT and nCECT groups for the two methods.

A p value of less than 0.05 was considered to indicate a significant difference. Statistical analysis was performed using the R statistical package (version 3.5.1, R Foundation for Statistical Computing, Vienna, Austria.).

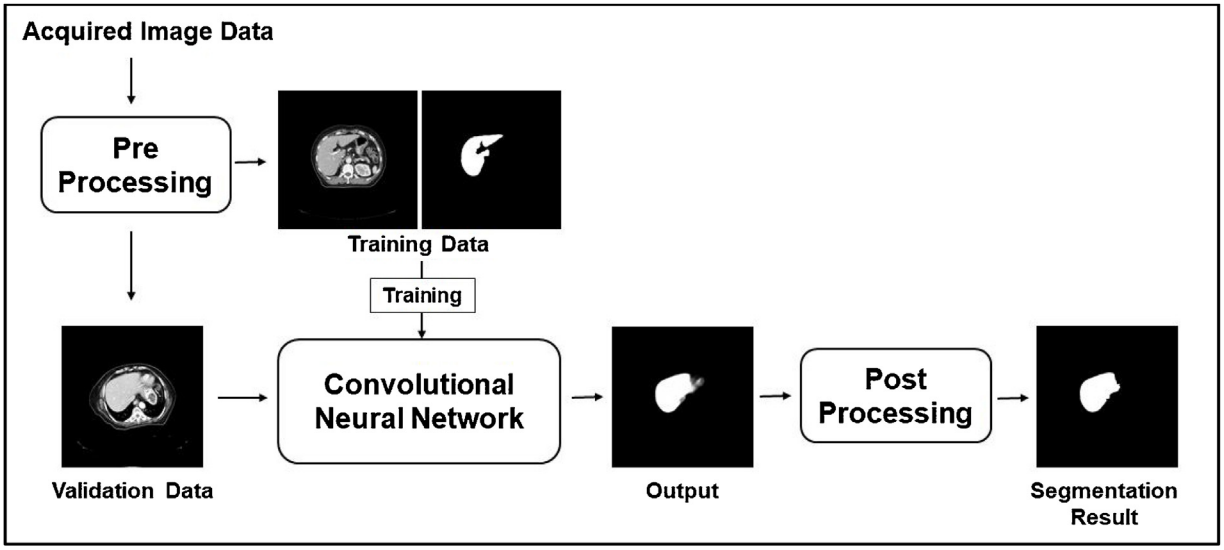

Fig. 1. Outline of the study methods.

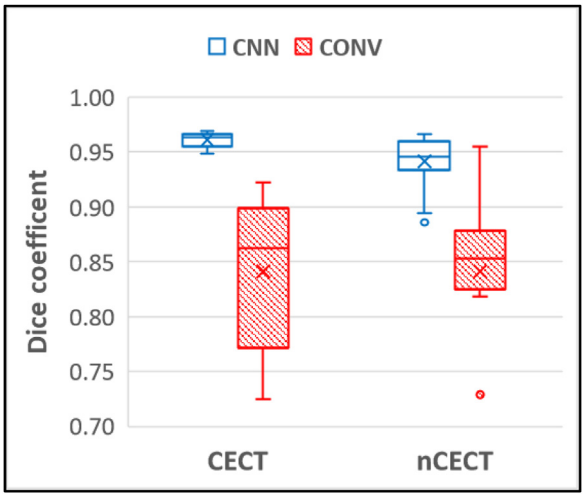

Fig. 2. Dice coefficient comparisons of the CNN and conventional segmentation methods.
The blue box plots show the results of the CNN method, and the red box plots show the results of the conventional method.

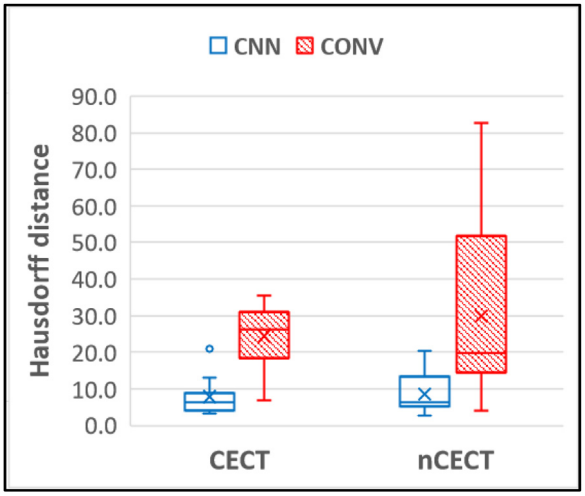

Fig. 3. Hausdorff distance comparisons of the CNN and conventional methods.
The blue box plots show the results of the CNN method, and the red box plots show the results of the conventional method.

3. Results

The mean DC_{CNN} values (and standard deviation) of the CECT and nCECT groups were $0.961 \pm 0.01$and 0.949 ± 0.02, respectively (Fig. 2), while the corresponding DC_{conv} values were 0.844 ± 0.06 and 0.841 ± 0.08. The DC_{CNN} values were significantly higher than the DC_{conv} values in both the CECT and nCECT groups ($p < 0.001$). The DC_{CNN} values of the CECT images were significantly higher than those of the nCECT images ($p = 0.003$), although the DC_{conv} values showed no significant difference between the CECT and nCECT image sets ($p = 0.95$).

The mean d_{CNN} values (and standard deviation) of the CECT and nCECT groups were 7.87 ± 4.28 and 8.59 ± 4.97, respectively (Fig. 3), while the corresponding d_{conv} values were 24.5 ± 7.9 and 29.9 ± 21.98. Statistical analysis of the Hausdorff distance showed that the similarity of the liver contours was significantly higher with the CNN than with the conventional software, in both the CECT and nCECT groups ($p < 0.001$). However, there were no significant differences between the CECT and nCECT Hausdorff distances for the CNN segmentation ($p = 0.82$) or conventional software method ($p = 0.85$).

Examples of the CECT and nCECT results are shown in Figs. 4 and 5, respectively. In both figures, the CNN segmentation results of the upper, middle, and lower parts of the liver are compared with the segmentation results of the ground-truth and conventional method.

4. Discussion

In this study, we show that automatic liver segmentation of abdominal CECT images can be successfully performed by a CNN trained on CECT images. The DC between the ground-truth results and those of the CNN was 0.96, indicating relatively high accuracy, and these results were higher than those of the conventional software. The Hausdorff distance evaluation also showed the same trend as the DC results. In Figs. 4 and 5, compared to the conventional method, the CNN method accurately separated and delineated the boundary between the rib and the liver. Even when comparing the time spent on output, mean delineation time of the liver on CT images by the radiotherapists was about 10 min,

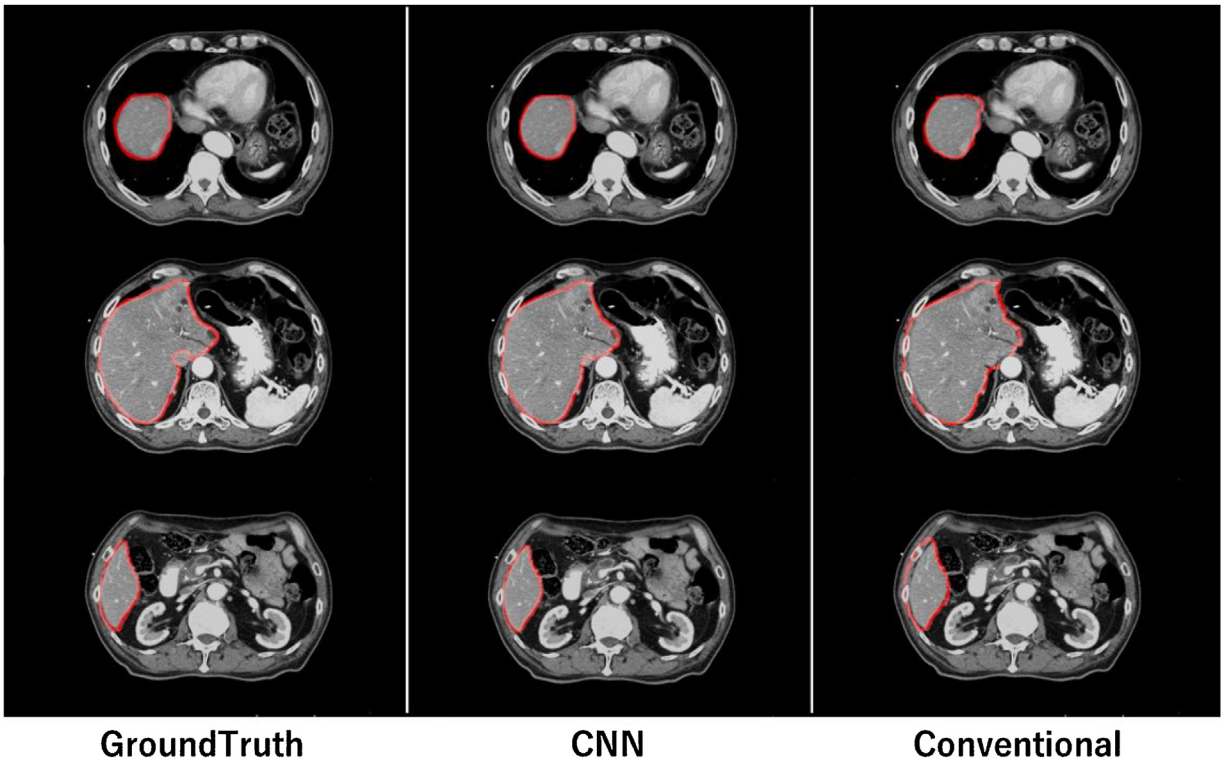

Fig. 4. Comparison of the CNN and conventional method on contrast-enhanced CT.
All images are composites of the original CECT and segmentation result images. The red lines represent the contour of the liver delineated by the radiation oncologists (ground truth) and automatically segmented by the CNN or conventional method. From top to bottom, examples of upper, middle, and lower parts of the liver on contrast-enhanced CT.

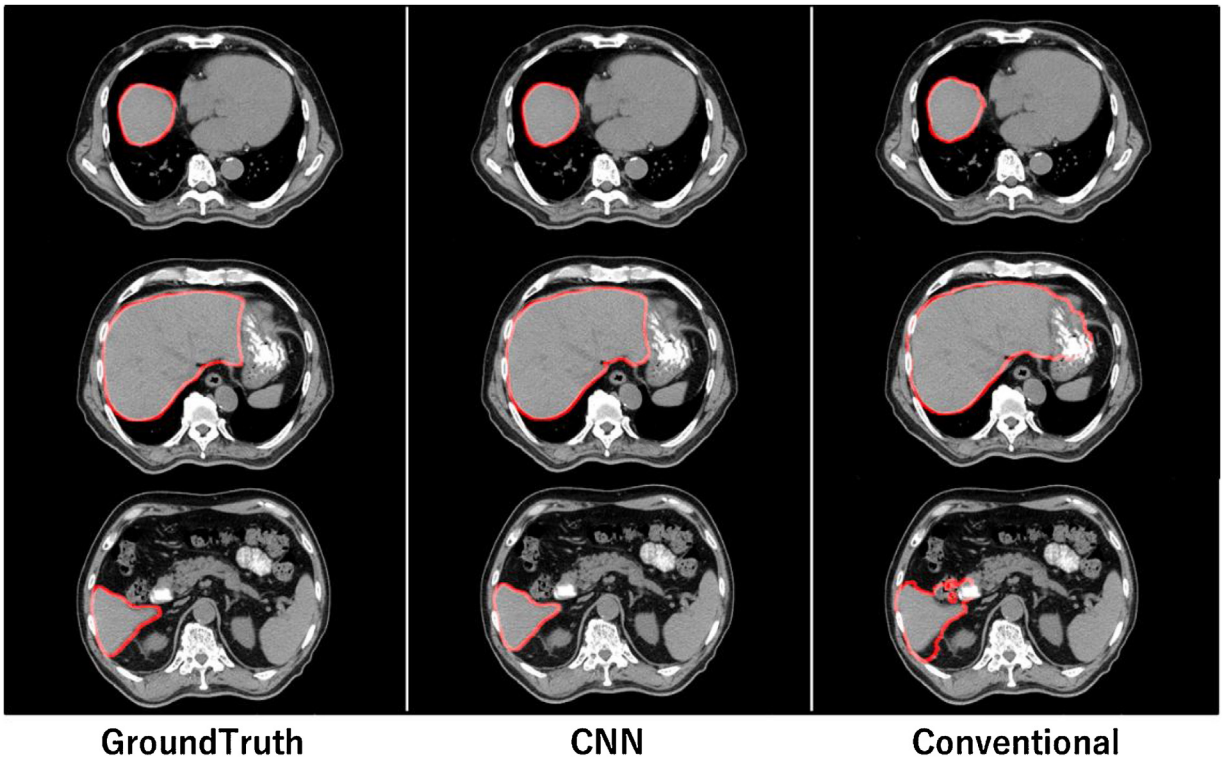

Fig. 5. Comparison of the CNN and conventional method on non-contrast-enhanced CT (nCECT).
All images are composites of the original nCECT and segmentation result images. The red lines represent the contours of the liver delineated by radiation oncologists (ground truth) and automatically segmented by CNN or the conventional method. From top to bottom, examples of upper, middle, and lower parts of the liver on nCECT.

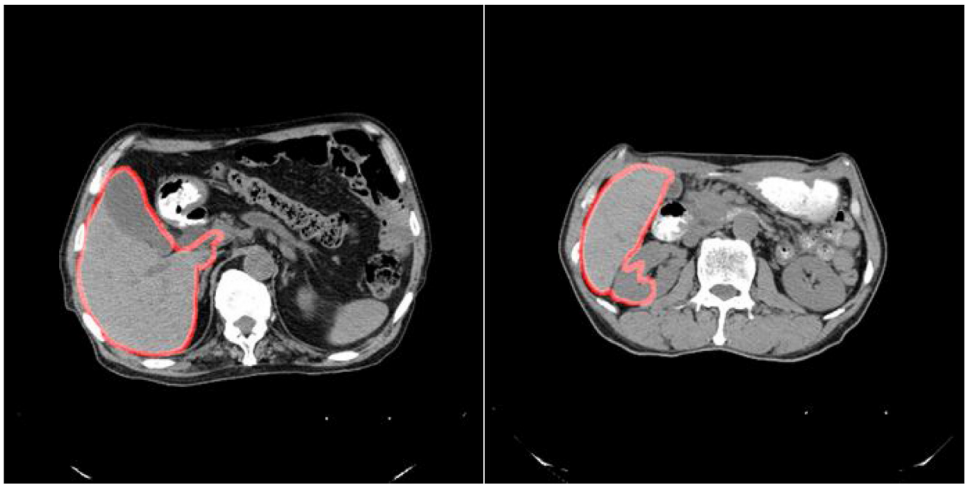

Fig. 6. Examples that were not successfully segmented by our CNN method.
The red lines represent the contours of the liver automatically segmented by the CNN method. Parts of a gallbladder and kidney were accidentally included in the CNN liver segmentations.

the segmentation time of the conventional method was about several minutes, while the CNN was able to segment in a few seconds per case. These results suggest that CNN-based automatic liver segmentation has the potential to replace the role of conventional segmentation. The output data of our CNN model were little affected by measurement bias, because the CNN was trained using ground-truth liver segmentations performed by seven radiotherapists who were randomly allocated different image sets.

Although CECT images were used as training data for the CNN, the trained CNN also successfully delineated the liver on nCECT images. The mean CT values of the liver on CECT and nCECT were 182 HU and 160 HU, respectively. The DC values of the nCECT were lower than those of the CECT images. When focusing only on the liver contour, the CNN Hausdorff distances of the CECT and nCECT groups were not significantly different. One of the reasons for this is that the automatic segmentation by the CNN does not only depend on the CT values of the liver, but is also affected by the shape and positional relationship of the liver in the body.

If the CNN was trained with nCECT images, the accuracy of the automatic segmentation would probably be even higher, although the use of nCECT data for training is subject to some problems. One of these problems is the difficulty in collecting sufficient nCECT data. In our institution, a CECT examination is the first choice for pancreatic cancer RT planning. Although some other institutions use non-contrast CT for RT calculations and contrast CT for RT contouring, in the dose calculation algorithm named anisotropic analytical algorithm (AAA; Varian Medical Systems, Palo Alto, CA, USA), the difference between the calculation results of non-contrast and contrast CT is very small,[18] and we rather consider the effect of organ shift due to the difference in breath holding between them. A CECT examination is necessary to accurately evaluate the size of the tumor just before performing the RT planning, because it is possible that the examination date of the original diagnostic imaging and the RT starting date may be far apart. An nCECT examination is generally only selected when the patient is at risk of adverse effects from iodine-based contrast agents, and as patients at risk of such adverse effects are rare in Japan,[19] it was not possible to obtain enough cases for the CNN training within the time period surveyed. We considered the possibility of using diagnostic CT images as a case of the nCECT group. However, these images were excluded in this study because of the influence of geometrical differences between the CT images for RT simulation and diagnostics (CT scanner cradle shape, breath-holding method, image field of view, slice thickness, etc.). On nCECT, differences in the CT values of each organ in the abdomen are smaller than on CECT, and the accuracy of automatic contouring tends generally to decrease in comparison with CECT. However, we showed that high accuracy segmentation can be achieved on nCECT images using the CNN.

There have been several previous reports on automatic segmentation of the liver using CNNs, with studies performed under various conditions. Landman et al. described a CNN that learned abdominal MRI images of 36 cases, and then automatically extracted 9 cases, with a DC of 0.913 for segmentation of the liver.[12] A study by Hu et al. reported a DC of 0.971 for liver segmentation on 42 abdominal CT images using deep 3D CNN.[13] High accuracy was found for the CNN-based automatic segmentation of the liver in all previous relevant studies and, in this respect, our study found similar DC values. However, it should be noted that the results cannot be directly compared, because the conditions for selecting cases and acquiring the images used as training data differ between studies. Furthermore, there have been no reports describing the automatic segmentation of the liver on nCECT images using a CNN trained on CECT images.

A possible limitation of our study is that there were no cases where the liver regions contained tissues differing from normal liver tissue, such as liver cysts and cancer. If cases with these diseases were included in the validation data, the accuracy of the automatic segmentation might have been reduced. In addition, while the upper and middle part of the liver could be segmented accurately, in the lower part, the gallbladder, kidney, and intestine were occasionally wrongly included in the liver region (Fig. 6). This was partly caused by the greater variations in the other adjacent organs in the lower part of the liver in comparison with the upper and middle sections. To solve this problem, it will be necessary to further increase the training data.

This study suggests that highly accurate delineation of the liver for radiation treatment planning can be obtained automatically in a few seconds using the CNN, irrespective of whether the CT images are CECT or nCECT images; this allows the burden of manual delineation on the treatment planner to be reduced. In the future, development of a CNN model for the automatic segmentation of not only the liver, but also other abdominal organs and lymph nodes, will be required.

Financial disclosure statement

There is no conflict of interest and a statement indicating any source of funding or financial interest for all authors.

Conflict of interest

None.

References

1. Jouglar E, et al. Can we spare the pancreas and other abdominal organs at risk? A comparison of conformal radiotherapy, helical tomotherapy and proton beam therapy in pediatric irradiation. *PLoS One*. 2016;11(10):e0164643.
2. Chen AM, Chin R, Beron P, Yoshizaki T, Mikaeilian AG, Cao M. Inadequate target volume delineation and local–Regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer: Local–Regional recurrence after IMRT for HPV-positive oropharynx cancer. *Radiother Oncol*. 2017;123(3):412–418. Jun.
3. Baldini EH, et al. Retroperitoneal sarcoma target volume and organ at risk contour delineation agreement among NRG sarcoma radiation oncologists. *Int J Radiat Oncol Biol Phys*. 2015;92(5):1053–1059.
4. Segedin B, Petric P. Uncertainties in target volume delineation in radiotherapy - are they relevant and what can we do about them? *Radiol Oncol*. 2016;50(3):254–262. Sep.
5. Loo SW, Martin WMC, Smith P, Cherian S, Roques TW. Interobserver variation in parotid gland delineation: A study of its impact on intensity-modulated radiotherapy solutions with a systematic review of the literature. *Br J Radiol*. 2012;85(1016):1070–1077. Aug.
6. Caldwell CB, et al. Observer variation in contouring gross tumor volume in patients with poorly defined non-small-cell lung tumors on CT: The impact of18FDG-hybrid PET fusion. *Int J Radiat Oncol Biol Phys*. 2001;51(4):923–931. Nov.
7. Yuan Y, Chao M, Sheu RD, Rosenzweig K, Lo YC. Tracking fuzzy borders using geodesic curves with application to liver segmentation on planning CT. *Med Phys*. 2015;42(7):4015–4026. Jun.
8. Park H, Bland PH, Meyer CR. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Trans Med Imaging*. 2003;22(4):483–492. Apri.
9. Li D, et al. Augmenting atlas-based liver segmentation for radiotherapy treatment planning by incorporating image features proximal to the atlas contours. *Phys Med Biol*. 2017;62(1):272–288. Jan.
10. Lu F, Wu F, Hu P, Peng Z, Kong D. Automatic 3D liver location and segmentation via convolutional neural network and graph cut. *Int J Comput Assist Radiol Surg*. 2017;12(2):171–182. Feb.
11. Yasaka K, Akai H, Abe O, Kiryu S. Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: A preliminary study. *Radiology*. 2017;286(3):170706. Mar.
12. Landman BA, et al. Fully convolutional neural networks improve abdominal organ segmentation. *Med Imaging 2018: Image Process*. 2018;10574:100.
13. Hu P, Wu F, Peng J, Liang P, Kong D. Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. *Phys Med Biol*. 2016;61(24):8676–8698. Dec.
14. Godfrey DJ, Patel BN, Adamson JD, Subashi E, Salama JK, Palta M. Triphasic contrast enhanced CT simulation with bolus tracking for pancreas SBRT target delineation. *Pract Radiat Oncol*. 2017;7(6):e489–e497. Nov.
15. Jensen NKG, et al. Dynamic contrast enhanced CT aiding gross tumor volume delineation of liver tumors: An interobserver variability study. *Radiother Oncol*. 2014;111(1):153–157. Apr.
16. Su H, Xing F, Kong X, Xie Y, Zhang S, Yang L. U-net:robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. In: *Advances in computer vision and pattern recognition.*; 2017:257–278, no. 9783319429984.
17. Bridson MR, Haefliger A. *Metric spaces of non-positive curvature*. Springer; 1999.
18. Ohira S, et al. Treatment planning based on water density image generated using dual-energy computed tomography for pancreatic cancer with contrast-enhancing agent: Phantom and clinical study. *Med Phys*. 2018;45(11):5208–5217. Nov.
19. Katayama H, Yamaguchi K, Kozuka T, Takashima T, Seez P, Matsuura K. Adverse reactions to ionic and nonionic contrast media. A report from the Japanese Committee on the Safety of Contrast Media. *Radiology*. 1990;175(3):621–628. Jun.