Liver Segmentation on a Variety of CT Images Based on Convolutional Neural Networks Combined with Connected Components

Hoang Hong Son[1], Pham Cam Phuong[2], Theo van Walsum[3], Luu Manh Ha[1,3,*]

1 AVITECH & FET, VNU University of Engineering and Technology, Hanoi, Vietnam

2 The Nuclear Medicine and Oncology center, Bach Mai hospital, Hanoi, Vietnam

3 BIGR, Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands

Abstract

Liver segmentation is relevant for several clinical applications. Automatic liver segmentation using convolutional neural networks (CNNs) has been recently investigated. In this paper, we propose a new approach of combining a largest connected component (LCC) algorithm, as a post-processing step, with CNN approaches to improve liver segmentation accuracy. Specifically, in this study, the algorithm is combined with three well-known CNNs for liver segmentation: FCN-CRF, DRIU and V-net. We perform the experiment on a variety of liver CT images, ranging from non-contrast enhanced CT images to low-dose contrast enhanced CT images. The methods are evaluated using Dice score, Haudorff distance, mean surface distance, and false positive rate between the liver segmentation and the ground truth. The quantitative results demonstrate that the LCC algorithm statistically significantly improves results of the liver segmentation on non-contrast enhanced and low-dose images for all three CNNs. The combination with V-net shows the best performance in Dice score (higher than 90%), while the DRIU network achieves the smallest computation time (2 to 6 seconds) for a single segmentation on average. The source code of this study is publicly available at *https://github.com/kennyha85/Liver-segmentation*.

Keywords: Liver segmentations, CNNs, Connected Components, Post processing.

1. Introduction

Liver cancer has one of the highest mortality rates for cancers worldwide [8], with a total of approximately 800,000 new cases annually. In general, the 5-year survival rate of liver cancer patient without treatment is less than 15% [13]. Liver cancer is more common in sub-Saharan Africa and Southeast Asia regions compared with Europe and United States. In some developing countries such as Vietnam, liver cancer is the most common type of cancer [12,20]. Liver radiofrequency ablation (RFA) has become a popular treatment for liver cancer due to its several advantages. This type of treatment is appropriate in the early stage or in cases of multiple tumors. RFA is a relatively low-risk minimally invasive procedure without producing toxic side-effects such as radioembolization and chemoembolization [30,31]. Furthermore, the liver of patients treated with RFA recovers in only a few days after receiving the intervention [32].

* Corresponding author. E-mail.: halm@vnu.edu.vn, luumanhha85@gmail.com

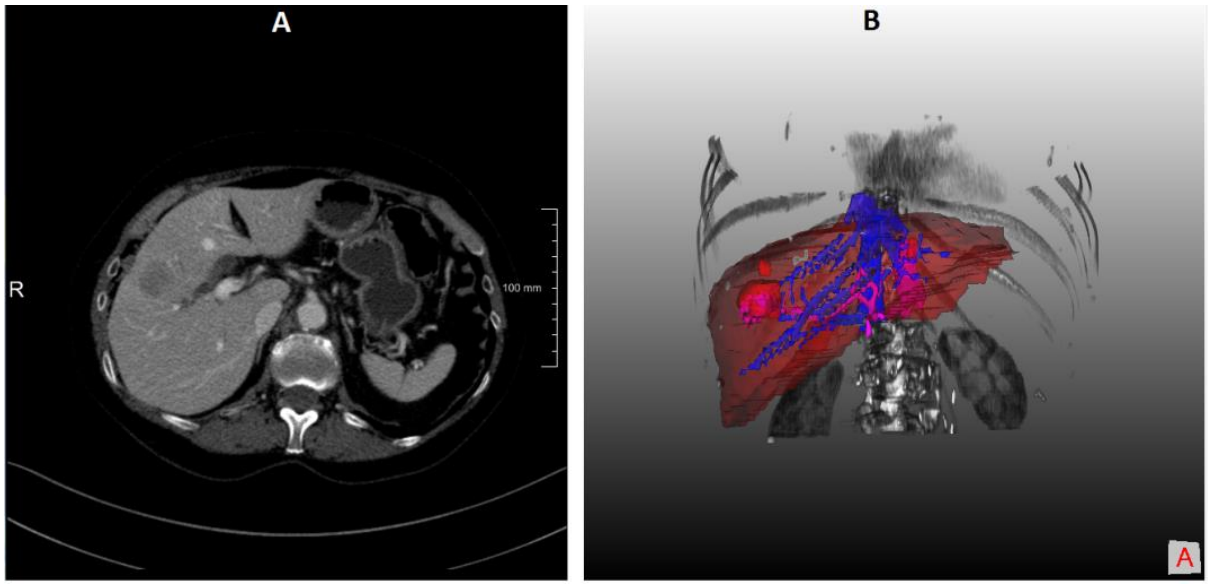

Figure 1. A typical contrast enhanced CT image of the liver (A) and the 3D segmentations of the liver, vessels and tumors (B). The volume rendering provides 3D visualization of the liver and the tumor in a RFA planning stage.

The CT imaging modality is often used for diagnosing liver cancer and planning the RFA treatment procedure for liver cancer. The 3D liver segmentation on the CT images of the liver is thus relevant for RFA treatment of liver cancer. In the planning stage, the liver segmentation acts as a region of interest, which contains the liver tumor and the liver vessels (see Figure 1). First, the visualization of the 3D liver segmentation provides adequate information to enable the radiologist to decide on the process of ablator insertion such that the trajectory of the insertion does not reach the critical parts such as bones, vessels and the kidneys. Second, the liver segmentation may also act as a mask region for liver registration using pre-operative, intra-operative and post-operative CT images of the RFA liver intervention [27,28]. Typically, the liver segmentation can be performed manually by a radiologist as a slice-by-slice approach. Because this manual approach requires tedious work and a substantial amount of time, it does not match the clinical workflow well. Therefore, liver segmentation using computer-based automatic and semiautomatic strategies has recently become an active research field. However, the noise due to lowering radiation dose, the low contrast between the liver and nearby organs, liver movement due to breathing motion, and the differences in size, shape and voxel intensity inside the liver across different patients present as current challenges to the implementation of 3D liver segmentation in the clinical setting. Several liver segmentation methods have been proposed in the literature and have high potential to be applied in clinical practice. In general, those methods can be classified into two main groups. The first group contains classical statistical and image-processing approaches such as region growing, active contour, deformable models, graph-cuts, statistical shape model [5,26]. These methods use hand-crafted features, and thus provide limited feature representation capability. The second group consists of Convolutional Neural Networks (CNNs), which have achieved remarkable success in many fields in the medical imaging domain such as object classification, object detection, and anatomical segmentation. Several CNN approaches have shown improved accuracy performance and are comparable to manual annotations by experts in oncology and radiology [1]. This success can be attributed to the ability of CNNs to learn a hierarchical representation of spatial information of CT images [7]. CNN approaches, how require large

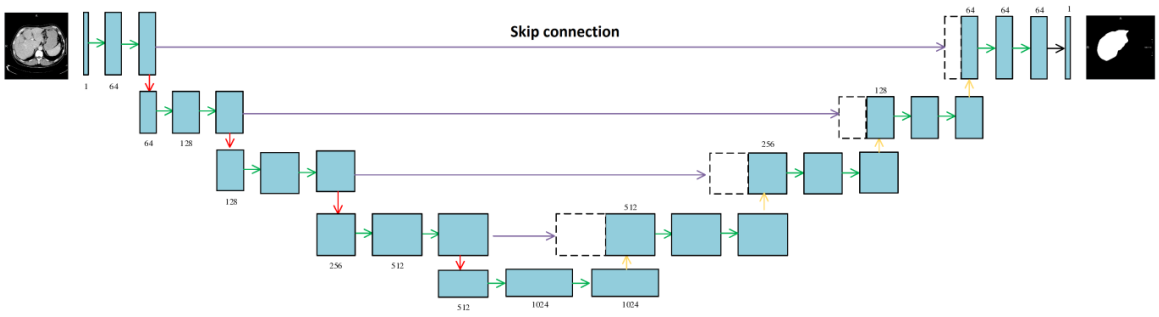

Figure 2. Illustration of 2D U-net architecture for liver segmentation using CT images with the inputs as a 2D image and the output as a predicted map of the liver. The networks contain four levels of the hierarchical representation. The skip connections provide linear combinations of the feature maps at the same level of up sampling and down sampling paths.

amount of data to train the models which is one of the main limitations in medical imaging research domain because medical image sharing is often limited due to privacy concerns.

In current liver segmentation, CNN-based segmentation algorithms have considerably outperformed the classical statistical/image-processing-based approaches [1,2,3,21]. U-net, one of the most well-known CNN architectures, introduced by Ronneberger et al. (2015), has received high rankings in several competitions in the field of medical image segmentation [1], and Christ et al. (2016) have successfully segmented the liver using a U-net architecture [3] (see Figure 2). Christ et al. (2017) further developed a fully convolutional neural network (FCN) based on the U-net architecture to segment the liver in both CT and MRI images, achieving a mean of Dice score of 94% with fewer than 100 training images [2]. Lu et al. (2015) have proposed a 3D CNN-GC method that combines a 3D fully convoluted neural network and graph cuts to achieve automatic liver segmentation in CT images with an accuracy of VOE of 9.4% on average [33]. Li et al. (2018) have also introduced the H-dense U-net for automatic liver segmentation, coupling intra-slice information using 2D dense U-net and inter-slice information using a 3D counterpart, and obtained the mean of DICE of 96.1% [4]. Bellver et al. (2017) have further improvised the original OVOS neural network, called DRIU, to segment the liver in CT images and achieved comparative results [6]. The number of publications relating to liver segmentation using a CNN has been increasing dramatically and most of them participate in the MICCAI grand challenge for liver segmentation (LiTS). Those CNNs, in general, can be classified into two categories: 2D Fully Convolutional Networks (2D FCNs) [2], [3], [6] and 3D Fully Convolutional Networks (3D FCNs) [4], [7], [18]. While 3D CNNs require greater computational complexity and consume more VRAM memory, the segmentation performance of 3D FCN versus 2D FCN still remains under debate [21].

As a machine learning classification family, CNNs perform convolutional filter image classification to segment the objects and as a result may contain several mis-classified voxels. Therefore, post-processing techniques may be applied to improve liver segmentation using CNNs. Conditional Random Forest (CRF) is a well-known method for post-processing of liver segmentation, but based on our previous study [29], CRF does not work well with CNN-based liver segmentation of low-dose/non-contrast CT images. Milletari et al. (2016) further states that "*post-processing approaches such as connected components analysis normally yield no improvement*" [7]. Considering the paucity of studies, it is necessary to elucidate how post-processing impacts the liver segmentation on CT images.

Given that the liver is the largest organ in the abdominal cavity, we hypothesize that the liver segmentation should be the largest connected component in the segmentations obtained from the CNNs. The main contribution of our study is that we propose a largest connected component

(LCC) algorithm to improve the liver segmentation in CT images using CNNs. To do this, we perform a full search for the largest connected component based on the connected component algorithm [22], and then we apply the algorithm on the liver segmentations generated by three well-known CNN architectures: U-net + CRF [2], DRIU [6] and V-net [7]. We evaluate the methods on three datasets: contrast enhanced CT images, low-dose contrast enhanced CT image, and low-dose, non-contrast enhanced CT image.

The next sections are organized as follows: the methods section briefly describes the three CNNs architectures and LCC method; next, the experiments section presents in detail the implementation of the CNNs architectures, the data used in the study, and the criteria to evaluate the performance of the proposed method. The results are illustrated in section 4, which is followed by a discussion of the results in section 5. The conclusion section summarizes the findings in this study.

2. Method

2.1 Convolution Neural network architectures

- *Fully Convolutional Network (FCN) combined with conditional random fields (CRF)*

The Fully Convolutional Network (FCN) combined with conditional random fields (CRF), proposed by Christ et al. (2017), contains two 2D U-net networks in a cascaded structure to sequentially segment both the liver and liver tumors [3]. U-net architecture is a well-known FCN that is able to learn a hierarchical representation of the image in the training stage [2]. In this study, we re-implement the first U-net network for the task of liver segmentation using CT images. The U-net architecture contains 19 layers in 4 levels and is divided into two parts: the encoder (also called "contracting path") and the decoder (also called "expanding path"). The encoder classifies the contextual information of all of the pixels in the input image via a process of hierarchical extractions, while the decoder provides the spatial information of the classified pixels to their corresponding location in the original image. Furthermore, the U-net skips several connections at different levels to provide information of the feature maps from the encoder section to the decoder section at the same levels. Embedding the skipped connections allows compensation of information about the objects that can be lost after each layer in the main path of U-net architecture.

The U-net input is 2D images and the output is a 2D probability map as the result of a soft prediction classifier for each pixel in the original images.

For the optimization process, weighted binary cross entropy *CE* is used as the objective loss function:

$$CE = -\frac{1}{N}\sum_i^N w_i t_i \log(s_i) , \quad (1)$$

where N is the number of pixels involved in the training stage; t_i is the ground truth value, which is either 0 or 1 when the pixel i is either background or foreground; S_i is the soft prediction score at the location pixel; i and w_i are the weights defining the degree of importance of the liver pixels. w_i is chosen as 1 over the foreground region size.

Subsequently, a 3D-dense conditional random field (CRF) is applied on the 2D probability maps, enabling the combination of both 3D spatial coherence and 2D appearance information from the slice-wise U-net segmentation [3].

- *V-Net: Fully CNNs for Volumetric Medical Image Segmentation*

While most CNNs utilize 2D convolution kernels to segment objects in 2D images, the V-net segments a 3D liver volume using 3D convolution kernels embedded in a fully convolutional neural network [4,7]. The V-net is more or less a 3D version of U-net and also contains two parts: the down-sampling path and the up-sampling path. The down-sampling path compresses the original 3D images into feature maps, while the up-sampling path extracts the feature maps until the final output reaches the original size of the input 3D image. Similar to U-net, the skipped connections from the encoding

to the decoding path at the same deep levels to provide spatial information of each layer and thus further improve the accuracy of the final segmentation prediction.

In this study, we utilize Dice loss as the objective function in the optimization process as suggested in the original work [7]:

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}, \quad (2)$$

where p_i and g_i are voxel values, either being 1 or 0, of the predicted liver segmentation and the ground truth, respectively, and *N* is the number of voxels of the two images in the same size.

- *DRIU: Deep retinal image understanding*

DRIU was introduced by Bellver et al. (2017) to segment the liver in abdominal contrast enhanced CT images [6]. The network architecture utilizes VGG-16 as the back-bone network, removing the last classification layers, i.e. the fully-connected layers, while maintaining other layers such as the fully convolutional layers, ReLU active function, and max-pooling layers. Similar to U-net, the DRIU architecture includes a contracting part and an expanding part containing several paired convolutional layers with the same size of feature map. The main difference from U-net is that the feature map at each level of the expanding part is achieved by up-sampling the feature map in the lower layer from the contracting part. In addition, in the expanding path, the output of DRIU is a combination of all feature maps at multiple scales by rescaling them to the original image size and then integrating them up into a single image. Thus, the segmentation contains information of the liver as a multiscale representation of the image. We also use weighted Binary Cross Entropy loss function for the optimization process.

2.2. Largest connected component (LCC)

In order to remove isolated regions of false segmentations of the liver generated by the CNNs, we propose to apply a connected component algorithm in the post-processing stage. We first apply a 3D connected component-labeling algorithm [22] and then perform a full searching for the largest connected component. Note that there should be a few connected components with the liver segmentation component as the largest one, given that the liver is the largest organ in the abdominal cavity. In the case that the largest component is not the liver, the neural network would not perform well and the segmentation should be treated as a failed case.

Table 1. The pseudocode of the largest connected component algorithm.

```
algorithm LCC(segmentation)

labels = list of connected component of segmentation
LCC_label = 0
Largest_CC_size = 0
for label in labels:
      if volume of label is larger than largest_CC_size
largest_CC_label = label
largest_CC_size = volume of label
Largest_LCC_segmentation = segmentation labeled by LCC_label
return Largest_LCC_segmentation
```

3. Data and Experiment setup

3.1. Clinical Data

In this study, we perform experiments using four datasets of CT images as in our previous study [29], which contains several variants of liver CT images: contrast enhanced, low-dose contrast enhanced, and low-dose non-contrast enhanced CT images. All of the confidential information in the datasets were anonymized by their own medical centers before taking part in this study. The parameters of the datasets are summarized in the Table 2.

The first dataset contains 115 contrast enhanced CT images from the Liver Tumour Segmentation (LiTS) challenge in the MICCAI grand challenge [34]. The images were acquired on a variety of CT scanners and protocols from multiple medical centers. We used LiTS dataset

Table 2. Parameters of the datasets in the study.

Dataset	Number of data	Resolution (mm)	Spacing (mm)	Number of slices	Voltage (kVP)
LiTS	115	0.55 - 1.0	0.45 - 6.0	74 - 986	-
Mayo	10	0.64 - 0.84	3.0	46 - 112	100
EMC_LD	15	0.56 - 0.89	2 - 5	27 -68	80 - 120
EMC_NC_LD	15	0.56 - 0.89	5	21 - 89	80 - 120

for training the three CNN models, like as previous done in Bellver et al. (2017) [6].

The second dataset consists of 10 CT images from the Mayo Clinic (Mayo), which were acquired by a Siemens CT scanner under a typical scanning protocol. The images are contrast enhanced portal-venous phase, and include several primary liver tumors. In order to reduce the redundant slices, the images were manually cropped in the *z* dimension such that the liver region is preserved.

The third and the fourth dataset are 15 contrast enhanced (EMC_LD) and 15 non-contrast enhanced CT images (EMC_NC_LD), respectively, which were randomly selected from Erasmus MC PACS in 2014 [27]. The images were acquired during radio frequency ablation intervention under low-dose protocol, resulting in noisy images due to the low radiation dose (see Figure 4).

The datasets from Erasmus MC and Mayo were manually annotated by two experts for ground truth, which is used in the evaluation section in this study, while the dataset from LiTS challenge already is publicly available with the liver segmentation ground truth segmented by several experts.

3.2. Implementation

We implement the algorithms in Python 3 using Tensorflow 1.18 and CUDA 9.1. The original source code for the FCN-CRF network, and the trained model from [2] are reused and modified to obtain a complete process of 3D liver segmentation. V-net and its trained model on the same LiTS dataset are re-implemented and based on the source code and introduction from Chen's website https://github.com/junqiangchen/LiTS—LiverTumor-Segmentation-Challenge. The DRIU network model is fine-tuned using the pre-trained model from Bellver et al [6]. The parameter settings are the same as suggested in the original work, including the batch size of 1; 15000 to 50000 iterations for a single channel; the initial learning rate of 10^{-8}; and SGD optimizer with momentum.

The LCC method is implemented in Python 3, using SITK library for connected components extraction. For further studies, the source code for the LCC method is publicly available at https://github.com/kennyha85/Liver-segmentation.

The study utilizes a Linux PC, Ubuntu 16.04, with Intel Core i9 9900K CPU, 8 cores, 3.6-5 GHz; NVIDIA Titan V GPU (11 GB RAM version), 64 GB DDR4, 2133 MHz Bus.

4. Evaluation and result

4.1. Evaluation metrics

In this study, we assess the performance of the combination of the CNNs with connected components using several criteria introduced in the MICCAI grand challenge. The algorithms yield binary liver segmentations, which are compared to the ground truth using Dice Score (***DSC***), Mean Surface Distance (***MSD***), Hausdoff Distance (***HD***), and False Positive Rate (***FPR***). We also evaluate the processing time of the methods. The evaluation metrics are described in more detail below.

4.1.1 Dice score (DSC)

Dice score is the overlap of the liver segmentation and the ground truth. Given a liver

Table 3. Median values of evaluation scores of LCC combined with the three CNN architectures. The numbers in brackets are quality of improvement compared to without using LCC. The last column are the minimum and maximum processing times. The bold number that they are the best scores.

Dataset	Methods	DSC (%)	HD (mm)	MSD (mm)	FPR (%)	Processing time (s)
Mayo	FCN+CRF+LCC	92.3 (2.1)	63.4 **(172)**	4.4 (0.6)	**3.1** (0.8)	7 - 8.2
	DRIU+LCC	92.6 (2.4)	34.6 (21)	2.2 **(2.3)**	8.1 (0.1)	**5.6 – 6.1**
	Vnet+LCC	**93.8 (3.4)**	**25.3** (91)	**1.6** (1.2)	6.7 **(3.9)**	6.6 - 9.8
EMC_LD	FCN+CRF+LCC	86.0 **(8.3)**	**35.1 (114)**	2.5 **(15.7)**	**13.5 (12)**	3.1 – 6.4
	DRIU+LCC	84.7 **(3.2)**	42.0 (106)	2.4 (12.2)	14.9 **(4.7)**	**2.6 – 5.3**
	Vnet+LCC	**90.4** (1.9)	38.2 (105)	**2.0** (8.2)	14.2 (3.2)	4.2 - 8.6
EMC_NC_LD	FCN+CRF+LCC	81.9 (2.4)	**51.5** (62)	3.6 **(12.1)**	23.3 (3.5)	3.6 – 7.7
	DRIU+LCC	87.2 (1.6)	66.1 **(66)**	4.9 (4.9)	8.8 (2.4)	**2.6 – 6.8**
	Vnet+LCC	**90.3 (4.1)**	51.7 (60)	**2.2** (1.9)	**7.8 (6.6)**	2.9 - 8.4

segmentation $\boldsymbol{X}$ and the ground truth $\boldsymbol{Y}$, ***DSC*** can be computed as:

$$\boldsymbol{DSC} = \frac{2|\boldsymbol{X} \cap \boldsymbol{Y}|}{|\boldsymbol{X} \cup \boldsymbol{Y}|}, \quad (3)$$

The maximum value of ***DSC*** is 1 when the segmentation $\boldsymbol{X}$ is perfectly matched the ground truth $\boldsymbol{Y}$. The ***DSC*** is 0 when $\boldsymbol{X}$ and $\boldsymbol{Y}$ do not have any voxel in common.

4.1.2 Mean Surface Distance (MSD)

Let ***S(X)*** denotes the set of surface voxels of the segmentation $\boldsymbol{X}$. The shortest distance of a voxel y to $\boldsymbol{S(X)}$ is defined as:

$$d\big(y, \boldsymbol{S}(\boldsymbol{X})\big) = min_{x \in S(X)} \|y - x\|, \quad (4)$$

where $\|.\|$ denotes the Euclidean distance. ***MSD*** is then computed by:

$$d_{\boldsymbol{MSD}}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{|S(X)|+|S(Y)|}\left(\sum_{x \in S(X)} d\big(x, \boldsymbol{S}(\boldsymbol{Y})\big) + \sum_{y \in S(Y)} d\big(y, \boldsymbol{S}(\boldsymbol{X})\big)\right), \quad (5)$$

4.1.3 Hausdorff Distance (HD)

Let ***S(X)*** and ***S(Y)*** be two boundaries of liver segmentation and ground truth, respectively. The Hausdorff distance $d_{HD}(\boldsymbol{S(X)},\boldsymbol{S(Y)})$ is the maximum distance between ***S(X)*** and ***S(Y)***, and is computed as follows:

$$d_{\boldsymbol{HD}}(\boldsymbol{S}(\boldsymbol{X}), \boldsymbol{S}(\boldsymbol{Y})) = \max\{\sup_{x \in S(X)} \inf_{y \in S(Y)} d(x, y), \sup_{y \in S(Y)} \inf_{x \in S(X)} d(x, y)\}, \quad (6)$$

where *sup* represents the supremum and *inf* denotes the infimum.

4.1.4 False Positive Rate (FPR)

FPR is used to quantify the false positive segmentation i.e. the segmentation outside the ground truth. Given the segmentation $\boldsymbol{X}$ and the ground truth $\boldsymbol{Y}$, ***FPR*** of the segmentation can be computed as the following:

$$\boldsymbol{FPR}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{|\boldsymbol{X} \backslash \boldsymbol{Y}|}{|\boldsymbol{Y}|}, \quad (7)$$

where $|\boldsymbol{X}\backslash\boldsymbol{Y}|$ denotes number of voxels in $\boldsymbol{X}$ which do not overlap with $\boldsymbol{Y}$.

4.2. Quantitative Results

The median values of the evaluation scores of the liver segmentation predicted by using the three CNNs architecture combined with the LCC algorithm are summarized in the Table 3. All three of the CNNs successfully segment the liver in the Mayo and the EMC_LD dataset with Dice scores higher than 80% for every dataset. For the EMC_NC_LD dataset, each of the CNNs fails to segment one of the images, achieving Dice scores less than 50%. We use 50% to decide the threshold for failed cases. Based on Table 3, we

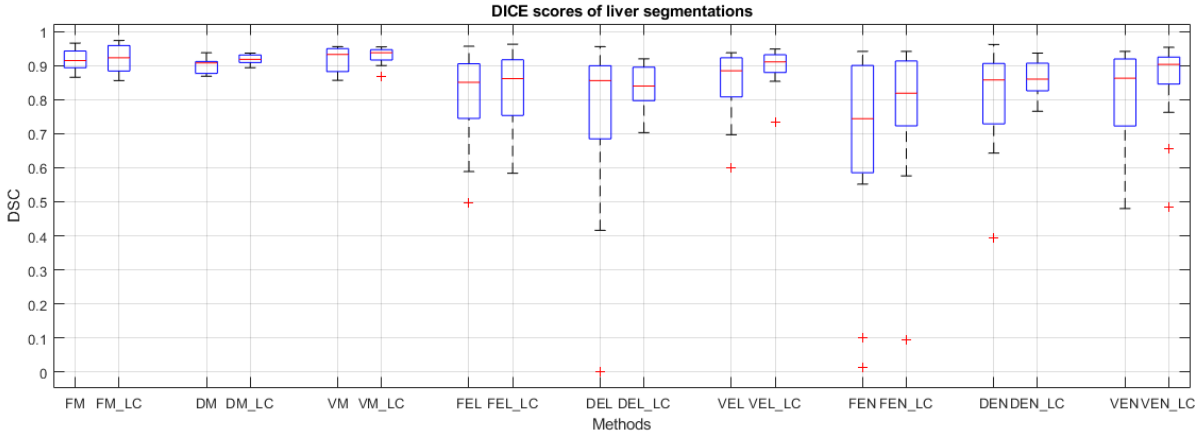

Figure 3. Scores of the three CNNs with and without LCC on the three datasets. The brief notations are described in the text.

can conclude that V-net + LCC perform the best with the medians of the Dice scores larger than 90%. Note that 90% Dice score is also the threshold for success used in other applications [14]. The minimum and maximum processing times, corresponding to the image size, are also reported in the last column of Table 3. Based on the statistics, we can conclude that the DRIU+LCC runs faster than V-net + LCC. Furthermore, the LCC takes less than a second for refining segmentations by the three CNNs on average. The maximum total processing time suggests the largest adding time that radiology technicians may have to take into account when they combine the methods to other processes. Note that the CT images are cropped to reduce the redundancy in a data preparation step (See section 3.1 Clincal Data).

Figure 3 is a box plot of the segmentation Dice scores of all of three CNNs on the three datasets with and without applying the LCC algorithm. The brief notations are descried as the following: FM (**F**CN+CRF on **M**ayo dataset), FM_LC (**F**CN+CRF with **L**CC on **M**ayo dataset), DM (**D**RIU on **M**ayo dataset), DM_LC (**D**RIU with **L**CC on **M**ayo dataset), VM (**V**net on **M**ayo dataset), VM_LC (**V**net with **L**CC on **M**ayo dataset), FEL (**F**CN+CRF on **E**MC **L**owdose dataset), FEL_LC (**F**CN+CRF with **L**CC on **E**MC **L**owdose dataset), DEL (DRIU on **E**MC **L**owdose dataset), DEL_LC (**D**RIU with **L**CC on **E**MC **L**owdose dataset), VEL (**V**net on **E**MC **L**owdose dataset), VEL_LC (**V**net with **L**CC on **E**MC **L**owdose dataset), FEN (**F**CN+CRF on **E**MC Lowdose **N**on-contrast enhanced dataset), FEN_LC (**F**CN+CRF with **L**CC on **E**MC Lowdose **N**on-contrast enhanced dataset), DEN (**D**RIU on **E**MC Lowdose **N**on-contrast enhanced dataset), DEN_LC (**D**RIU with **L**CC on **E**MC Lowdose **N**on-contrast enhanced dataset), VEN (**V**net on **E**MC Lowdose **N**on-contrast enhanced dataset), VEN_LC (**V**net with **L**CC on **E**MC Lowdose **N**on-contrast enhanced dataset). We also perform paired *T*-tests to assess the statistical significance of the difference between the results of the CNNs with and without the connected components method. The *p*-values of the t-tests for the evaluations scores of the pairs FM/FM_LC, DM/DM_LC, VM/VM_LV, FEL/FEL_LC, DEL/DEL_LC, VEL/VEL_LC, PEN/PEN_LC, DEN/DEN_LC and VEN/VEN_LC are summarized in Table 4. From Table 4, we can conclude that the LCC algorithm statistically significantly improves the segmentation results of all three CNNs in general.

The Figure 4 is an example of 3D liver segmentations on a low-dose contrast enhanced CT image. In the second column, the liver segmentations by three CNNs include some false positive segmentations (in blue), which are eliminated by the LCC algorithm. Obviously, the difference in segmentation from three networks is not visible in the 2D view (right column). The 3D view in the first column visualizes the difference between the liver segmentations and the ground truth.

5. Discussion

Table 4. P-values of the T-tests for the proposed method with the corresponding original CNNs: The numbers are smaller than 0.05 indicating that the improvements are statistically significance.

Dataset	Methods	DSC	HD	MSD	FPR
Mayo	FM/FM_LC	0.021	0.019	0.002	0.001
	DM/DM_LC	0.002	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	VM/VM_LC	0.040	0.001	0.014	0.019
EMC_LD	FEL/FEL_LC	0.010	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	DEL/DEL_LC	0.016	$< 10^{-3}$	$< 10^{-3}$	0.118
	VEL/VEL_LC	0.027	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
EMC_NC_LD	FEN/FEN_LC	0.034	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	DEN/DEN_LC	0.055	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
	VEN/VEN_LC	0.019	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

In this study, we investigate the improvement in liver segmentation using CNNs approaches on CT images when they are combined with a connected component algorithm and the largest component in a post-processing step. We either re-implement or reuse the CNNs model trained with the LiTS dataset, testing them with other three datasets from two different medical centers with both standard and low dose protocols with and without contrast enhancement. Next, we apply the LCC algorithm on the liver segmentations by the CNNs approaches and quantitatively evaluate the results using well-known criteria for liver segmentation. Combination of the CNN approaches with the LCC algorithm statistically significantly improves the liver segmentation. The 3D visualization in the Figure 4 shows the improvements in a segmentation example. We also conclude that the FCN combined with conditional random forest method does not fully eliminate the isolated false positive segmentation. This can be explained by the fact that the CRF only examines inter-slice correlation of the segmentations, while the liver segmentation should be connected in 3D as one organ. From Figure 3, we can also conclude that the CNNs work better with the regular dose contrast enhanced CT images while most improvements by the LCC occur with the low-dose CT image. This may improve when more low dose images are included in the training stage. We refrained from adding more data in the training stage. In our opinion, while retraining CNNs network is a very "expensive" way of research, reusing the shared works and improving the result using "inexpensive" techniques is a reasonable approach to promote research results to practical application.

We also can see from Table 3 and Figure 3 that V-net combined with the LCC generally perform better than other methods. This confirms findings from Milletari et al. (2016) [7], which show that 3D segmentation approaches use inter-slice information and thus may improve segmentation accuracy. However, Table 3 also demonstrates that the 3D nature of the V-net leads to more computation time and requires more memory. These factors may limit its potential to be used in clinical practices that require very fast processing such as intra operation of liver RFA. Note that in our experiment, we already manually cropped the liver volume to avoid the redundancy while current CT scans in clinical practice may have hundreds of slices. A fast, automatic liver detection method may be beneficial for those cases to extract the region of interest while reducing the processing time. Although the LCC shows to be effective for liver segmentation, it still presents challenges. The LCC can only remove false positive segmentations, which are isolated from the main liver segmentation, and thus cannot get rid of false positive segmentations connected with the main part, or fill in missing parts. More advanced

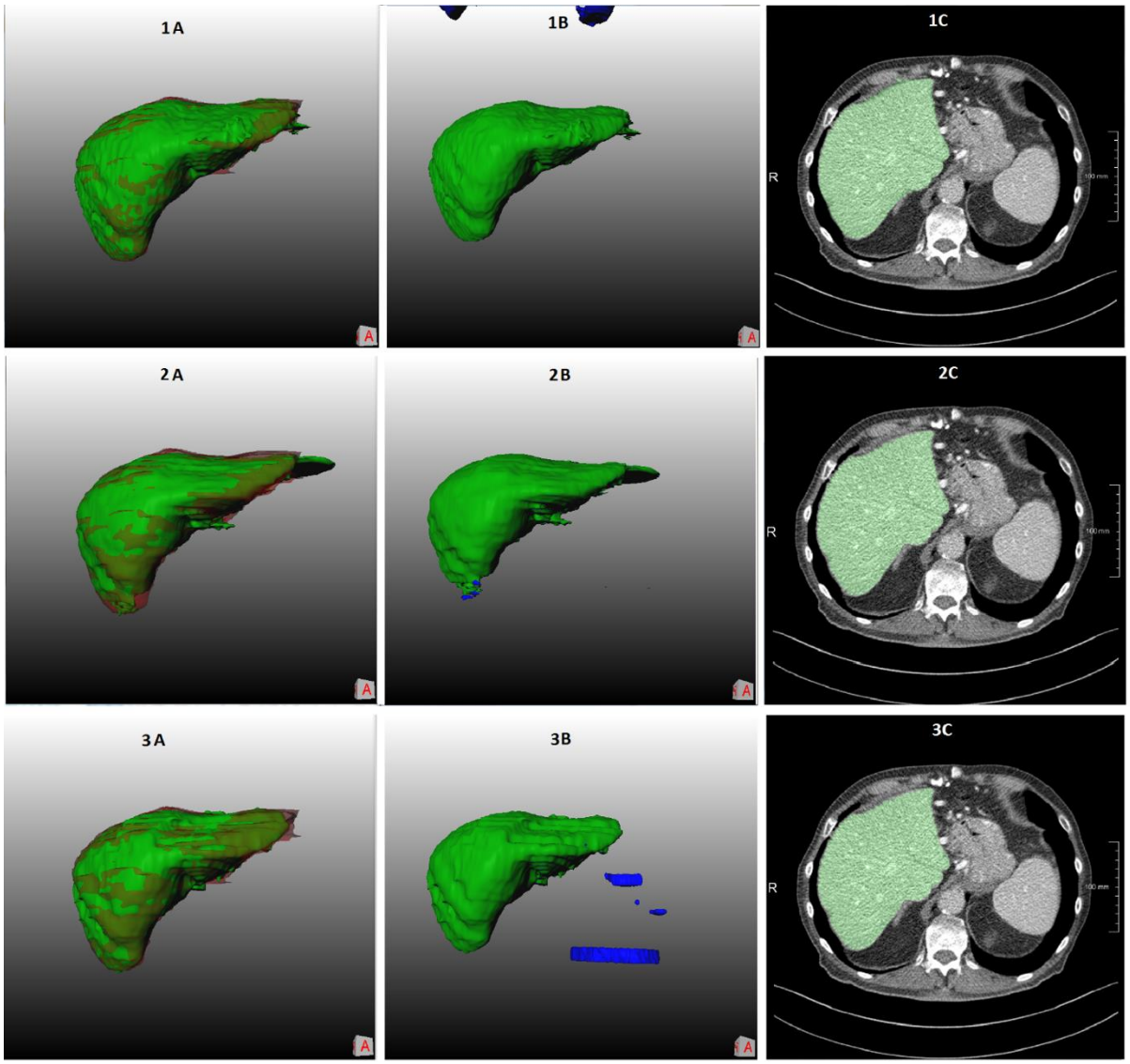

Figure 4. Example of 3D liver segmentations by the three CNNs on a low-dose contrast enhanced CT image. The first raw is segmentations by FCN, the second one is by DRIU and the last one is by V-net. The first column contains the liver segmentations using with LCC (green) and the ground truth (red), the second column illustrates the raw liver segmentation from the CNNs (blue) overlapped by the segmentation after post processing, and the last column is the final 2D liver segmentations on 2D CT slice of the liver.

segmentation methods, such as level set and graph-cuts, may further improve the smoothing on the surface of the liver, since they can embed and model liver shape and curvature information. Thus, the precise liver surface segmentation needs to be further investigated. Perhaps, subsequent studies may use data sharing to utilize more data in the training stage. While data sharing is currently challenging due to administrative procedures and privacy concerns, data-augmentation research directions could help enrich the training data pools.

There are some limitations in our study. First, we only use 10 contrast enhanced CT, 15 low-dose contrast enhanced CT, and 15 low-dose non-contrast enhanced CT from two medical centers for evaluating the methods. Nevertheless, we assume that the images from other medical centers will yield similar results as those in this study. Second, the training dataset for the CNNs does not include low-dose CT

images, resulting in poor performance with the EMC dataset. However, while investigating to improve the CNNs with more dataset in the training stage is not the main purpose of our research, we believe that adding low-dose CT images may improve the segmentation results. The improvement may be limited due to effects of the low-dose noise on the image quality. A noise removal CNN network combined with the current CNNs may be a more effective approach to improve the liver segmentation. Third, there have been several other variants of CNNs for liver segmentation that have achieved adequate results [4,18,19,23,24,25]. However, as pixel classification based methods, these CNNs may contain mis-classification parts and may likely benefit as well from post-processing methods such as the LCC.

6. Conclusion

In this paper, we present our work on improving liver segmentation for CNN based approaches using LCC algorithm. Experiments are performed with three well-known CNN architectures and with retrained or reused trained models. We evaluate three datasets from two different medical centers with regular contrast enhanced CT image and both contrast and non-contrast enhancement of low-dose image. The quantitative evaluation results show that LCC statistically significantly improves the liver segmentation accuracy of the CNNs, while maintaining the processing time of less than 10 seconds in total for all of the networks, including the LCC processing time of less than a second. In our study, we find that V-net combined with the LCC achieves a Dice score of approximately 94%, which is comparable to other state of the art methods. We believe that with the current development of CNN-based approach research, the liver segmentation using CNNs has a high potential to be applied in the clinical practice soon.

Acknowledgments

This work has been supported by VNU University of Engineering and Technology under project number CN 18.03. We would like to thank Mayo Clinical for supporting us their data. We also would like to thank NVIDIA for their aid of a graphics hardware unit.

Reference

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
2. Christ, P. F., Ettlinger, F., Grün, F., Elshaera, M. E. A., Lipkova, J., Schlecht, S, & Rempfler, M. (2017). Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks. arXiv preprint arXiv:1702.05970.
3. Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., & Sommer, W. H. (2016). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 415-423). Springer, Cham.
4. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. W., & Heng, P. A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE transactions on medical imaging, 37(12), 2663-2674.
5. Gotra, A., Sivakumaran, L., Chartrand, G., Vu, K. N., Vandenbroucke-Menu, F., Kauffmann, C., & Tang, A. (2017). Liver segmentation: indications, techniques and future directions. Insights into imaging, 8(4), 377-392.
6. Bellver, M., Maninis, K. K., Pont-Tuset, J., Giró-i-Nieto, X., Torres, J., & Van Gool, L. (2017). Detection-aided liver lesion segmentation using deep learning. arXiv preprint arXiv:1711.11069.
7. Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 Fourth International Conference on 3D Vision (3DV) (pp. 565-571). IEEE.
8. McGlynn, K. A., Petrick, J. L., & London, W. T. (2015). Global epidemiology of hepatocellular carcinoma: an emphasis on demographic and

regional variability. Clinics in liver disease, 19(2), 223-238.
9. Foltz, G. (2014). Image-guided percutaneous ablation of hepatic malignancies. In Seminars in interventional radiology(Vol. 31, No. 02, pp. 180-186). Thieme Medical Publishers.
10. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. IEEE transactions on medical imaging, 35(5), 1299-1312.
11. Gotra, A., Sivakumaran, L., Chartrand, G., Vu, K. N., Vandenbroucke-Menu, F., Kauffmann, C.,Tang, A. (2017). Liver segmentation: indications, techniques and future directions. Insights into imaging, 8(4), 377–392. doi:10.1007/s13244-017-0558-1
12. Hong, T. T., Phuong Hoa, N., Walker, S. M., Hill, P. S., & Rao, C. (2018). Completeness and reliability of mortality data in Viet Nam: Implications for the national routine health management information system. PloS one, 13(1), e0190755. doi:10.1371/journal.pone.0190755
13. Mohammadian, M., Mahdavifar, N., Mohammadian-Hafshejani, A., & Salehiniya, H. (2018). Liver cancer in the world: epidemiology, incidence, mortality and risk factors. World Cancer Res J, 5(2), e1082.
14. Luu, H. M., Moelker, A., Klein, S., Niessen, W., & van Walsum, T. (2018). Quantification of nonrigid liver deformation in radiofrequency ablation interventions using image registration. Physics in Medicine & Biology, 63(17), 175005.
15. Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdzal, M., Turcotte, S., Pal, C. J., & Tang, A. (2017). Deep learning: a primer for radiologists. Radiographics, 37(7), 2113-2131.
16. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
17. Meine, H., Chlebus, G., Ghafoorian, M., Endo, I., & Schenk, A. (2018). Comparison of U-net-based Convolutional Neural Networks for Liver Segmentation in CT. arXiv preprint arXiv:1810.04017.
18. Novikov, A. A., Major, D., Wimmer, M., Lenis, D., & Bühler, K. (2018). Deep Sequential Segmentation of Organs in Volumetric Medical Scans. IEEE transactions on medical imaging.
19. Huo, Y., Terry, J. G., Wang, J., Nair, S., Lasko, T. A., Freedman, B. I.,& Landman, B. A. (2019). Fully Automatic Liver Attenuation Estimation combing CNN Segmentation and Morphological Operations. Medical physics.
20. Pham, T., Bui, L., Kim, G., Hoang, D., Tran, T., & Hoang, M. (2019). Cancers in Vietnam—Burden and Control Efforts: A Narrative Scoping Review. Cancer Control, 26(1), 1073274819863802..
21. Meine, H., Chlebus, G., Ghafoorian, M., Endo, I., & Schenk, A. (2018). Comparison of U-net-based Convolutional Neural Networks for Liver Segmentation in CT. arXiv preprint arXiv:1810.04017..
22. Samet, H., & Tamminen, M. (1988). Efficient component labeling of images of arbitrary dimension represented by linear bintrees. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(4), 579-586.
23. Gruber, N., Antholzer, S., Jaschke, W., Kremser, C., & Haltmeier, M. (2019). A Joint Deep Learning Approach for Automated Liver and Tumor Segmentation. arXiv preprint arXiv:1902.07971.
24. Chen, S., Ma, K., & Zheng, Y. (2019). Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv preprint arXiv:1904.00625.
25. Tang, W., Zou, D., Yang, S., & Shi, J. (2018, October). DSL: Automatic Liver Segmentation with Faster R-CNN and DeepLab. In International Conference on Artificial Neural Networks (pp. 137-147). Springer, Cham.
26. Heimann, T., Van Ginneken, B., Styner, M. A., Arzhaeva, Y., Aurich, V., Bauer, C.,. & Bello, F. (2009). Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE transactions on medical imaging, 28(8), 1251-1265.
27. Luu, H. M., Klink, C., Niessen, W., Moelker, A., & van Walsum, T. (2016). Non-rigid registration of liver CT images for CT-guided ablation of liver tumors. PloS one, 11(9), e0161600.
28. Gunay, G., Luu, M. H., Moelker, A., van Walsum, T., & Klein, S. (2017). Semiautomated registration of pre-and intraoperative CT for image-guided percutaneous liver tumor ablation interventions. Medical physics, 44(7), 3718-3725.
29. Hoang, H. S., Pham, C. P., Franklin, D., van Walsum, T., & Luu, M. H. (2019, September). An Evaluation of CNN-based Liver Segmentation Methods using Multi-types of CT Abdominal Images from Multiple Medical Centers. In 2019 19th International Symposium on Communications and Information Technologies (ISCIT) (pp. 20-25). IEEE
30. Borner, M., Castiglione, M., Triller, J., Baer, H. U., Soucek, M., Blumgart, L., & Brunner, K. (1992). Arena: Considerable side effects of chemoembolization for colorectal carcinoma

metastatic to the liver. Annals of oncology, 3(2), 113-115.

31. Memon, K., Lewandowski, R. J., Kulik, L., Riaz, A., Mulcahy, M. F., & Salem, R. (2011). Radioembolization for primary and metastatic liver cancer. In Seminars in radiation oncology (Vol. 21, No. 4, pp. 294-302). WB Saunders.
32. Gory, I., Fink, M., Bell, S., Gow, P., Nicoll, A., Knight, V.,. & Kemp, W. (2015). Radiofrequency ablation versus resection for the treatment of early stage hepatocellular carcinoma: a multicenter Australian study. Scandinavian journal of gastroenterology, 50(5), 567-576.
33. Lu, F., Wu, F., Hu, P., Peng, Z., & Kong, D. (2017). Automatic 3D liver location and segmentation via convolutional neural network and graph cut. International journal of computer assisted radiology and surgery, 12(2), 171-182.
34. Bilic, P., Christ, P. F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q.,. & Kadoury, S. (2019). The liver tumor segmentation benchmark (lits). arXiv preprint arXiv:1901.04056.