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 investigated recently. In this paper we present a study on combining largest connected component (LCC) algorithm, as a post processing step, to liver segmentation using CNN approaches to improve the segmentation accuracy. In our work, we used a variety of liver CT images, ranging from non-contrast enhanced CT images to low-dose contrast enhanced CT images. We combined the algorithm with three well-known CNNs for liver segmentation: FCN-CRF, DRIU and V-net. The methods were evaluated using Dice score, Haudorff distance, mean surface distance, and false positive rate between the liver segmentation and the ground truth. The quantitative results showed that LCC algorithm statistically significantly improves the liver segmentation result of the liver segmentation on non-contrast enhanced and low-dose images for all three CNNs. The V-net showed the best performance in Dice score (higher than 90%) while the DRIU network achieved 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 is one of the highest mortality cancers worldwide [8], with 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, it is the most common cancer type [12,20]. Liver radiofrequency ablation (RFA) has become a regular treatment for liver cancer nowadays due to its several advantages. This type of treatment is suitable for the cases of multiple tumors or in the early stage. It is a relatively low-risk minimally invasive procedure without producing toxicity side-effects such as in radioembolization and chemoembolization [30,31]. In addition, the liver of patients treated with RFA recovers in just a few days after the intervention [32].

The CT imaging modality is often used for diagnostic of liver cancer and planning of the RFA treatment procedure for liver cancer. 3D liver segmentation on the CT images of the liver is relevant for RFA treatment of liver cancer. In

* Corresponding author. E-mail.: halm@vnu.edu.vn

the planning stage, the liver segmentation acts as a region of interest, which contains liver tumor and the liver vessels (see Figure 1). Firstly, visualization of the 3D liver segmentation provides information for radiologist to make decision of the ablator trajectory insertion. Secondly, 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, liver segmentation can be performed manually by a radiologist. This is generally a slice-by-slice approach, which requires a substantial amount of work, and is tedious and complicated, and because of the time required it does not match clinical workflow well. Therefore, liver segmentation using computer-based automatic and semiautomatic strategies has become a field of active research recently. However, the low contrast between the liver and nearby organs, liver movement due to breathing motion, differences in size, shape and voxel intensity inside the liver across different patients make the liver segmentation remains challenging tasks.

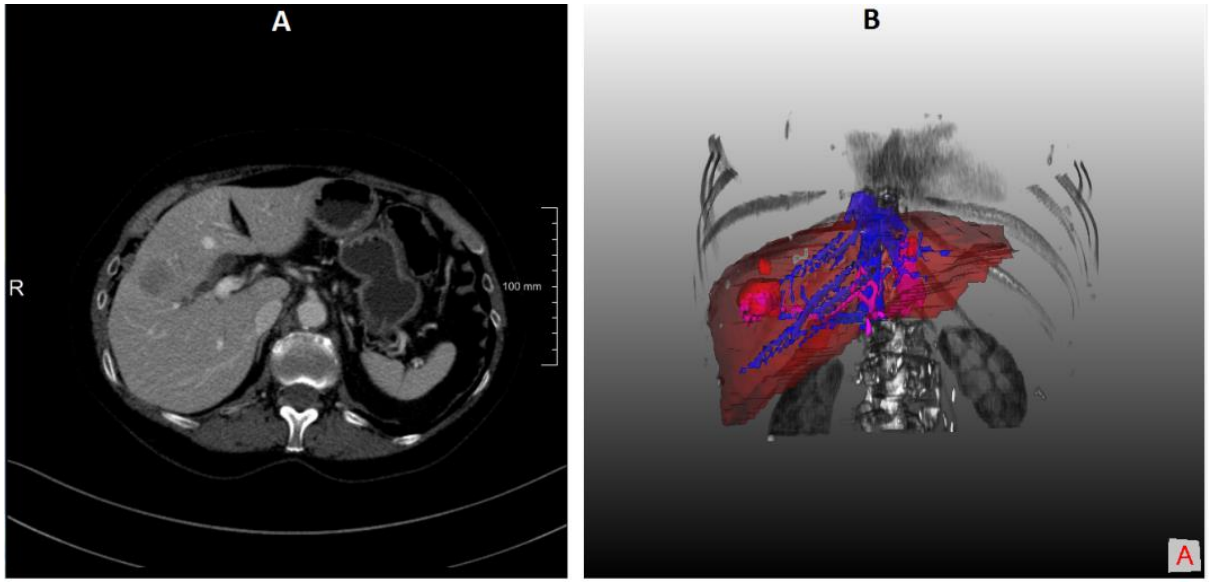

Figure 1. A typical contrast enhanced CT image of the liver (A) and the 3D segmentations of the liver, vessels and tumors (B). The volume rendering provides 3D visualization of the liver and the tumor in a RFA planning stage.

Several liver segmentation methods have been proposed in the literature in the past, and these have a high potential to be applied in the clinical practice. In general, those methods can be classified in to 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 groups consist of Convolutional Neural Networks (CNNs) which recently have achieved remarkable success in many fields in medical imaging domain such as object classification, object detection and anatomical segmentation. Several CNN approaches have shown great performance with the accuracy is even comparable to manual annotations by experts in oncology and radiology [1]. The main reason for this success is that CNNs are able to learn a hierarchical representation of spatial information of images [7]. CNN approaches require large 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 fields 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 (CCN) 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 less than 100 training images [2]. Lu et al (2015) proposed a 3D CNN-GC method that combines a 3D fully convolutional neural network and graph cuts to achieve automatic liver segmentation in CT images with the accuracy of VOE of 9.4% on average [33]. Li et al. (2018) introduced the H-dense U-net for automatic liver segmentation, which combines 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) improvised the original OVOS neural network, called DRIU, to segment the liver in CT images and achieved comparative results [6].

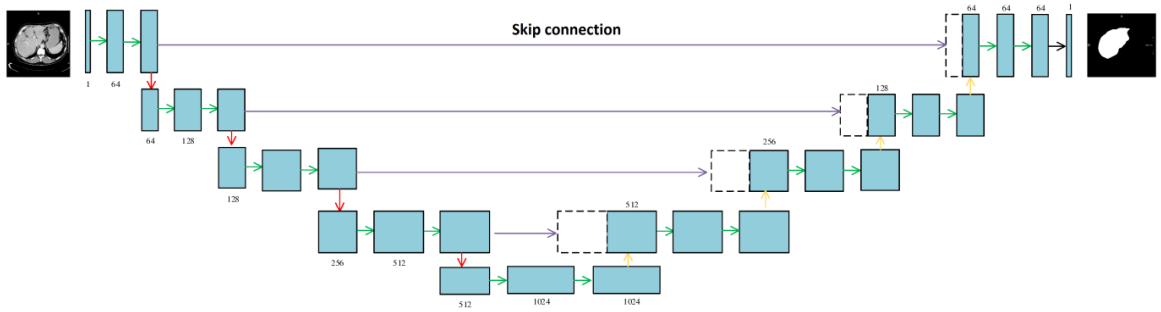

Figure 2: Illustration of 2D U-net architecture for liver segmentation using CT images with the inputs as a 2D image and the output as a predicted map of the liver. The networks contain four levels of the hierarchical representation. The skip connections provide linear combinations of the feature maps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

The number of publications relating to liver segmentation using a CNN has been increasing dramatically recently 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 neural networks (2D FCNs) [2], [3], [6] and 3D fully convolutional neural networks (3D FCNs) [4], [7], [18]. 3D CNNs require higher computational complexity and consume more VRAM memory, however, the segmentation performance of 3D FCN versus 2D FCN still remains under debate [21]. Also, as a machine learning classification family, CNNs perform convolutional filter image classification to segment the objects and as a consequence 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. However, based on our previous study [29], a CRF does not work well with CNN-based liver segmentation of low-dose/non-contrast CT images. In contrast, it was stated by Milletari et al. (2016) that "*post-processing approaches such as connected components analysis normally yield no improvement*" [7]. However, studies on how post-processing impacts on the liver segmentation on CT images are virtually missing.

Our hypothesis is that the liver is the largest organ in the abdominal, thus the liver segmentation should be the largest connected component in the segmentations obtain from the CNNs. Therefore, in this study, we investigate whereas the largest connected component method (LCC) can improve the liver

segmentation in CT images using CNNs. To do this, we apply LCC on the liver segmentation by three well known CNN architectures: U-net + CRF [2], DRIU [6] and V-net [7] and evaluate on three datasets: contrast enhanced CT images, low-dose contrast enhanced CT image and low-dose, non-contrast enhanced CT image to ensure solid conclusions.

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 details the implementation of the CNNs architectures, the data used in the study and the criteria for evaluating 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 summaries 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) is proposed by Christ et al. (2017), which contains two 2D U-net networks in a cascaded structure, to segment both the liver and liver tumors sequentially [3]. U-net architecture is well-known as a fully convolutional network (FCN) which is able to learn a hierarchical representation of the image in the training stage [2]. In this study, we reimplemented 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 divided into two parts: the encoder (so called "contracting path") and the decoder (so 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 several skip connections at different levels which provide information of the feature maps from the encoder section to the decoder section at the same levels The benefit of embedding the skips connection is that they compensate the part of information of the objects which can be loss after each layers 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:

$$CE = -\frac{1}{N}\sum_{i}^{N} w_i t_i \log(s_i) \qquad (1)$$

where N is the number of pixels involved in the training stage; t_i is ground truth value which is either 0 or 1 when the pixel i is either background or foreground; S_i is soft prediction score at the location pixel i; and w_i are the weights that define the degree of important 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 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 fully convolutional neural network [4,7]. The V-net is more or less a 3D version of U-net and it 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 skip connections from the encoding to the decoding path at the same deep levels provide spatial information of after 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]:

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

where p_i and g_i are voxel values, either being 1 or 0, of the predicted liver segmentation and the ground truth, correspondingly, 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, removes the last classification layers, i.e. the fully-connected layers, while it still maintains other layers such as 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 those contain several paired convolutional layers having the same size of feature map. The main difference from U-net is that the feature map at each level of expanding part is achieved by up-sampling the feature map in the lower layer from contracting part. In addition, in the expanding path, the output of DRIU is a combination all feature maps at multiple scales by rescaling them to the original image size and then adding them up into a single image. Thus, the segmentation contains information of the liver at multiscale representation of the image.

We also used weighted Binary Cross Entropy loss function for the optimization process as the following:

$$CE = -\frac{1}{N}\sum_i^N w_i t_i \log(s_i) \qquad (4)$$

2.2. Largest connected component (LCC)

In order to remove isolated regions of false segmentations of the liver, 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 we perform a full searching for the largest connected component. Note that the number of connected components should be not that many and the liver segmentation component should be the largest one because the liver is the largest organ in the abdominal. In addition, in case that the largest component is not the liver, the neural network did not perform well and the segmentation should be treated as failed case.

The **pseudocode** is as follows:

```
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 performed 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 parameter of the datasets are summarized in the Table 1.

The first dataset contains 115 contrast enhanced CT images from the Liver Tumour Segmentation (LiTS) challenge in MICCAI grand challenge [34]. The images were acquired on a variety of CT scanners and protocols from multiple medical centers. We used LiTS dataset for training the three CNN models as similar in Bellver et al. (2017) [6].

The second dataset is 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) correspondingly, 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 sections in this study, while the dataset from LiTS challenge already is publicly available with the liver segmentation ground truth segmented by several experts.

Table 1. Parameters of the datasets in the study.

Dataset	Number of data	Resolution (mm)	Spacing (mm)	Number of slices	Voltage (kVP)	CTDIvol (mGy)
LiTS	115	0.55 - 1.0	0.45 - 6.0	74 - 986	-	-
Mayo	10	0.64 - 0.84	3.0	46 - 112	100	18 - 21
EMC_LD	15	0.56 - 0.89	2 - 5	27 -68	80 - 120	4 - 12
EMC_NC_LD	15	0.56 - 0.89	5	21 - 89	80 - 120	4 - 9

3.2. Implementation

We implemented 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 were reimplemented and based on the source code and introduction from Chen's website https://github.com/junqiangchen/LiTS—LiverTumor-Segmentation-Challenge. The DRIU network model was fine-tuned trainning using the pretrained model from Bellver et al [6]. The parameter settings were as suggested in the original work, with 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 was 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 was carried on 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 challenges. The algorithms yield binary liver segmentations, which are compared to the ground truths using Dice Score

(***DSC***), Mean Surface Distance (***MSD***), Hausdoff Distance (***HD***), and False Positive Rate (***FPR***). In addition, 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 segmentation $\boldsymbol{X}$ and the ground truth $\boldsymbol{Y}$, ***DSC*** can be computed as:

$$\boldsymbol{DSC} = \frac{2|\boldsymbol{X} \cap \boldsymbol{Y}|}{|\boldsymbol{X} \cup \boldsymbol{Y}|} \quad (5)$$

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:

$$d\big(y, \boldsymbol{S}(\boldsymbol{X})\big) = min_{x \in S(X)} \|y - x\| \quad (6)$$

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

$$d_{\boldsymbol{MSD}}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{|S(X)|+|S(Y)|}\left(\sum_{x \in S(X)} d\big(x, \boldsymbol{S}(\boldsymbol{Y})\big) + \sum_{y \in S(Y)} d\big(y, \boldsymbol{S}(\boldsymbol{X})\big)\right) \quad (7)$$

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 $d_{HD}(\boldsymbol{S(X),S(Y)})$ is the maximum distance between ***S(X)*** and ***S(Y)***, and is computed as follows:

$$d_{\boldsymbol{HD}}(\boldsymbol{S}(\boldsymbol{X}), \boldsymbol{S}(\boldsymbol{Y})) = \max\{\sup_{x \in S(X)} \inf_{y \in S(Y)} d(x, y), \sup_{y \in S(Y)} \inf_{x \in S(X)} d(x, y)\} \quad (8)$$

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:

$$\boldsymbol{FPR}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{|\boldsymbol{X} \backslash \boldsymbol{Y}|}{|\boldsymbol{Y}|} \quad (9)$$

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 LCC algorithm are summarized in the Table 2. All of the three CNNs successfully segmented 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 CNNs had failed in segmenting one of the images, achieving Dice scores less than 50%. We use 50% to decide the threshold for failed cases. Based on Table 2, we can conclude that V-net + LCC performed the best with the medians of the Dice scores are larger than 90%. Note that 90% Dice score is 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 2. Based on the statistic, we can conclude that the DRIU+LCC runs literately faster than V-net + LCC. Furthermore, LCC took less than a second for refining segmentations by the three CNNs. 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 were cropped to reduce the redundancy in a data preparation step (See section 3.1 Clincal Data).

Table 2:

Median values of evaluation scores of LCC combined with the three CNN architectures. The last column are the minimum and maximum processing times. The bold number indicate that they are the best scores.

Dataset	Methods	DSC (%)	HD (mm)	MSD (mm)	FPR (%)	Processing time (s)
Mayo	FCN+CRF+LCC	92.3	63.4	4.4	**3.1**	7 - 8.2
	DRIU+LCC	92.6	34.6	2.2	8.1	**5.6 – 6.1**
	Vnet+LCC	**93.8**	**25.3**	**1.6**	6.7	6.6 - 9.8
EMC_LD	FCN+CRF+LCC	86.0	**35.1**	2.5	**13.5**	3.1 – 6.4
	DRIU+LCC	84.7	42.0	2.4	14.9	**2.6 – 5.3**
	Vnet+LCC	**90.4**	38.2	**2.0**	14.2	4.2 - 8.6
EMC_NC_LD	FCN+CRF+LCC	81.9	**51.5**	3.6	23.3	3.6 – 7.7
	DRIU+LCC	87.2	66.1	4.9	8.8	**2.6 – 6.8**
	Vnet+LCC	**90.3**	51.7	**2.2**	**7.8**	2.9 - 8.4

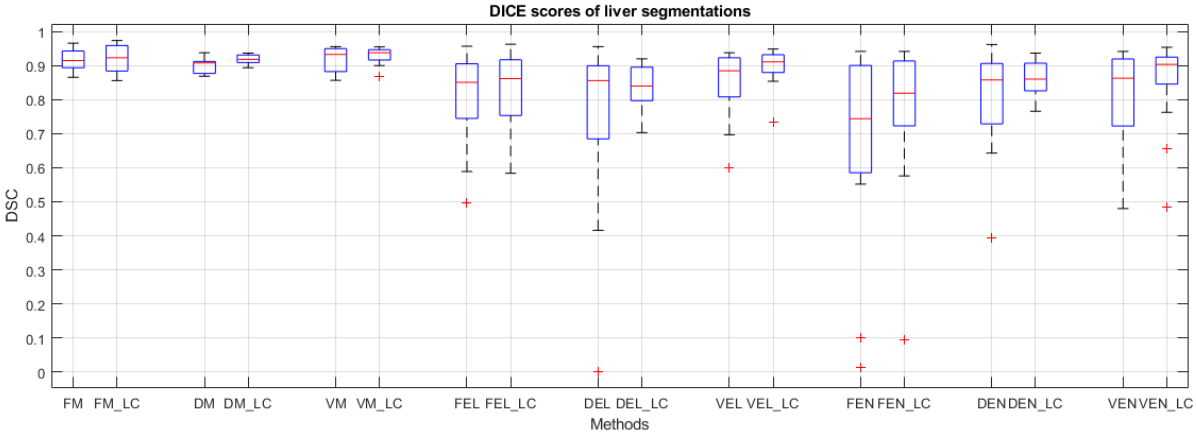

Figure 3: DICE scores of the three CNNs with and without LCC on the three datasets. The brief notations are descried as the followings: FM (**F**CN+CRF on **M**ayo dataset), FM_LC (**F**CN+CRF with **L**CC on **M**ayo dataset), DM (**D**RIU on **M**ayo dataset), DM_LC (**D**RIU with **L**CC on **M**ayo 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).

Figure 3 is a box plot of the segmentation Dice scores of all of three CNNs on the three datasets with and without applying LCC algorithm. Furthermore, we performed paired t-tests to assess the statistical significance of the difference between the results of the CNNs with and without using 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 3. From Table 3, we can conclude that 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. It can be seen in the second column that liver segmentations by three CNNs include

some false positive segmentations (in blue) which are eliminated by 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.

Table 3
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	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$

5. Discussion

In this study, we investigated the improvement in liver segmentation using CNNs approaches on CT images when they are combined with a connected component algorithm and selection of the largest component in a post-processing step. We either reimplemented or reused the CNNs model trained with the LiTS dataset, and tested them with other three datasets from two different medical centers with both standard and low dose protocols with and without contrast enhancement. Next, we applied LCC algorithm on the liver segmentations by the CNNs approaches and quantitatively evaluated the results using well-known criteria for liver segmentation evaluation.

Combination of the CNN approaches with LCC algorithm statistically significantly improved 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 looks at 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; most improvements occur with the low-dose CT image. This may improve when including more low dose images in the training. We refrained from doing so. 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 bring the research results to the practical application.

We also can see from Table 2 and Figure 3 that V-net combined with LCC generally performed better than the other methods. This result confirms results of Milletari et al. (2016) [7], showing that 3D segmentation approaches use inter-slice information and thus may improve the segmentation accuracy. However, Table 2 also illustrates that the 3D nature of the V-net leads to more computation time. In addition, there is more memory required when using this CNN. 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 manual 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 the liver while reducing the processing time.

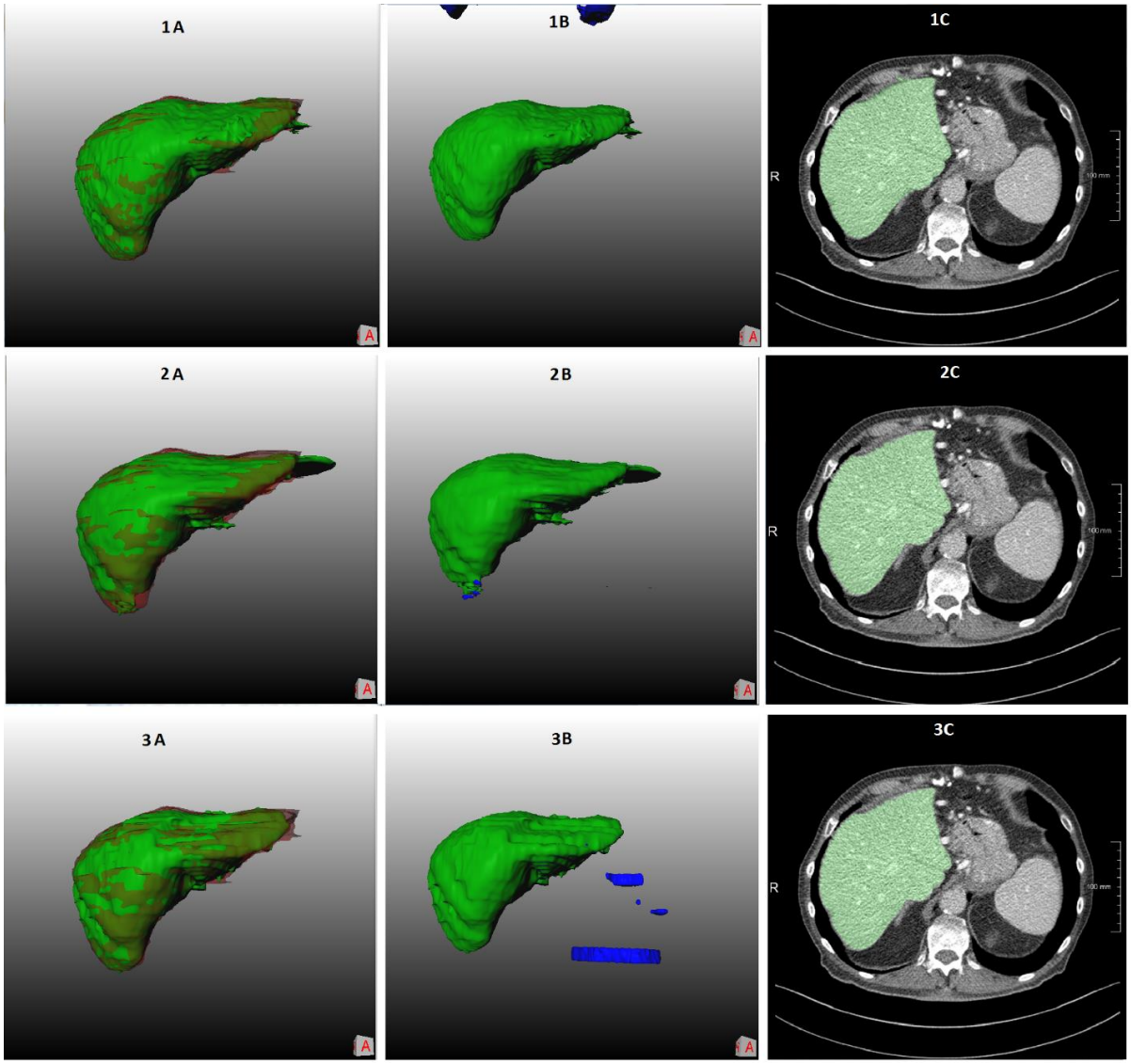

Figure 4: Example of 3D liver segmentations by the three CNNs on a low-dose contrast enhanced CT image. The first raw is segmentations by FCN, the second one is by DRIU and the last one is by V-net. The first column contains the liver segmentations using with LCC (green) and the ground truth (red), the 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

Although LCC showed to be effective for liver segmentation, it does not solve all problems. LCC can only remove false positive segmentations, which are isolated to the main liver segmentation, and thus it cannot get rid of false positive segmentations connected with the main part, or fill in missing parts. Some more advanced segmentation methods, such as level set and graph-cuts, may further improve the smoothing on the surface of the liver, as these methods are able to embed and model liver shape and curvature information. The precise liver

surface segmentation thus remains a topic for further studies. Subsequent studies may, e.g. utilized more data in the training stage. We strongly believe that data sharing is a relevant for this. Currently, while data sharing still is difficult because of 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 used 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 the similar results in this study. Second, the training dataset for the CNNs does not include low-dose CT images, which led to the poor performance with the EMC dataset. However, while investigating to improve the CNNs with more dataset in the training stage is not our main purpose of this research, we believe that adding low-dose CT image may improve the segmentation results but not that much because the low-dose noise absolutely affects the image quality. A noise removal CNN network combining with the current CNNs may be a good approach to improve the liver segmentation. Third, there have been several other variants of CNNs for liver segmentation already published and achieving good results [4,18,19,23,24,25]. Yet, these CNN approaches are pixel classification based methods and thus they may contain mis-classification parts and may likely benefit as well from post processing methods such as LCC.

6. Conclusion

In this paper, we presented a work on improving liver segmentation for CNN based approaches using LCC algorithm. Experiments were performed with three well-known CNN architectures and with retrained or reused trained models. We perform the evaluation on three datasets from two different medical centers with regular contrast enhanced CT image and both contrast and non-contrast enchantment 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 less than 10 seconds in total for all of the networks with the LCC processing time of less than a second. In our study, we found that V-net combined with LCC achieved a Dice score of approximately 94%, which is comparable to other state of the art methods. We believe with the current development of CNN-based approach researches, 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, October). 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, October). 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, June). 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, October). 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.